

**Univerzita Palackého v Olomouci**  
**Lékařská fakulta**  
Ústav klinické a molekulární patologie  
Laboratoř molekulární patologie



**Vybrané epigenetické mechanismy zahrnuté do nádorové transformace karcinomů hlavy a krku a karcinomů plic**

Disertační práce

Mgr. Veronika Žižková

Studijní program: Lékařská biologie

Forma studia: Prezenční

Olomouc 2021

Školitel: doc. MUDr. *et* MVDr. Jozef Škarda, Ph.D. *et* Ph.D.

**Prohlášení:**

Prohlašuji, že jsem tuto disertační práci vypracovala samostatně pod vedením doc. MUDr. *et* MVDr. Jozefa Škardy, Ph.D. *et* Ph.D. a uvedla jsem všechny použité prameny a literaturu. Vznik této práce byl podpořen grantem IGA MZ ČR NT/13569-4 a Nadačním fondem UP.

Ve Šternberku, dne 7. 2. 2021

.....

## Souhrn

Nádorová onemocnění představují se svojí zvyšující se incidencí v současné době celosvětový problém. Kancerogeneze je vysoce komplexní proces, při kterém dochází k maligní transformaci buněk na základě vnitřních a vnějších faktorů. Tyto faktory se mohou vzájemně ovlivňovat a kombinovat. V průběhu kancerogeneze dochází také k řadě genetických i epigenetických změn vyskytujících se ve všech kategoriích nádorových onemocnění – sporadické, familiární i hereditární. Nádory hlavy a krku, stejně jako karcinomy plic, jsou epiteloví maligní nádorová onemocnění, která jsou ve všech průmyslově rozvinutých zemích řazena k nejzávažnějším zdravotním problémům.

Úvodní část disertační práce se zabývá popisem incidence nádorů hlavy a krku a karcinomů plic, jejich diagnostikou a také faktory, které vedou k jejich vzniku. Další kapitoly se zabývají obecnými postupy a možnostmi léčby jak karcinomu hlavy a krku, tak i karcinomů plic.

Cílem mé disertační práce bylo studium vybraných epigenetických mechanismů zahrnutých do nádorové transformace dlaždicového karcinomu hlavy a krku (HNSCC) a karcinomů plic. U HNSCC jsem se zaměřila na porovnání epigenetického profilu nádorové tkáně a shodných vzorků uvulopalatofaryngoplastické tkáně a na vytipování transkripčních faktorů vhodných k determinaci HPV<sup>+</sup> a HPV<sup>-</sup> HNSCC. U nemalobuněčných karcinomů plic jsem detekovala expresi vybraných miRNA (miR-21, miR-23b, miR-126, miR-205 a RNU6B) a následně korelovala získané výsledky s aktivitou proteinů ovlivňující mnohočetnou lékovou rezistenci – P-gp, MRP a LRP/MVP.

## **Summary**

Cancer diseases with their increasing incidence are currently a global problem. Carcinogenesis is a highly complex process in which malignant transformation of cells occurs based on internal and external factors. These factors can interact and be combined. During carcinogenesis, there are also a number of genetic and epigenetic changes occurring in all categories of cancer - sporadic, familial and hereditary. Head and neck cancers, as well as lung cancers, are epithelial malignancies that are among the most serious health problems in all industrialized countries.

The introductory part of the dissertation deals with the description of the incidence of head and neck tumors and lung cancers, their diagnosis and also the factors that lead to their occurrence. The next chapters deal with general procedures and treatment options for both head and neck and lung cancers.

The aim of my dissertation was to study selected epigenetic mechanisms involved in the tumor transformation of squamous cell carcinoma of the head and neck (HNSCC) and lung carcinomas. In HNSCC, I focused on the comparison of the epigenetic profile of tumor tissue and identical samples of uvulopalatopharyngoplastic tissue and on the identification of transcription factors suitable for the determination of HPV<sup>+</sup> and HPV<sup>-</sup> HNSCC. In non-small cell lung cancers, I detected the expression of selected miRNAs (miR-21, miR-23b, miR-126, miR-205 and RNU6B) and subsequently correlated the obtained results with the activity of proteins influencing multiple drug resistance - P-gp, MRP and LRP/MVP.

## **Poděkování**

Ráda bych poděkovala svému vedoucímu doc. MUDr. *et* MVDr. Jozefu Škardovi, Ph. D. *et* Ph. D. za čas a cenné rady, které mi poskytl během tvorby této disertační práce. Dále děkuji ostatním kolegům z Ústavu klinické a molekulární patologie, LF UP a rovněž také kolegům z americké Univerzity Johns Hopkins v Baltimore, USA. Mé další díky patří všem, kteří mi v průběhu mnohaletého studia pomáhali a podporovali mě jak v pracovním, tak i osobním životě.



# OSNOVA

1. Úvod .....	8
2. Karcinom hlavy a krku .....	11
2.1 Léčba nádorů hlavy a krku .....	14
2.1.1 Chirurgická léčba .....	14
2.1.2 Radioterapie .....	14
2.1.3 Chemoterapie .....	14
2.1.4 Biologická léčba .....	15
3. Nádorové onemocnění plic .....	16
3.1 Současná léčba nemalobuněčného karcinomu plic .....	20
3.1.1 Chemoterapie NSCLC .....	22
3.1.2 Radioterapie NSCLC .....	22
3.1.3 Cílená biologická léčba NSCLC .....	22
3.1.4 Imunoterapie NSCLC .....	23
3.1.5 Současná léčba malobuněčného karcinomu plic .....	24
4. Epigenetika .....	25
4.1 Methylace DNA .....	25
4.2 Demethylace DNA .....	29
4.3 Modifikace histonů .....	29
4.4 Acethylace histonů .....	30
4.5 RNA interference .....	31
4.5.1 Historie a charakteristika mikroRNA .....	31
4.5.2 Biogeneze mikroRNA .....	32
4.5.3 Úloha miRNA v nádorech .....	38
5. Cíle práce .....	40
6. Materiál a metodika .....	41
7. Výsledky .....	54
8. Diskuze .....	61
9. Závěr .....	64
10. Literatura .....	66
11. Seznam použitých zkratk .....	85

12.	Přehled publikací autora .....	87
13.	Publikovaná abstrakta .....	90
14.	Seznam přednášek/posterů přednesených uchazečem na veřejných odborných fórech .....	93

# 1. Úvod

Nádorová onemocnění představují se svojí zvyšující se incidencí v současné době celosvětový problém. Kancerogeneze (karcinogeneze, tumorigeneze či ontogeneze) je vysoce komplexní proces, při kterém dochází na základě vnitřních (spontánní mutace, genomová nestabilita) a vnějších faktorů (fyzikální, chemické a biologické) k maligní transformaci buněk a následně ke vzniku nádoru. Tyto faktory se mohou navzájem ovlivňovat a kombinovat. Existuje řada genetických a epigenetických změn, ke kterým může docházet v průběhu kancerogeneze a vyskytují se ve všech překrývajících se kategoriích nádorových onemocnění: sporadické (cca 70 % případů), familiární (15-25 %) a hereditární (5-10 %).

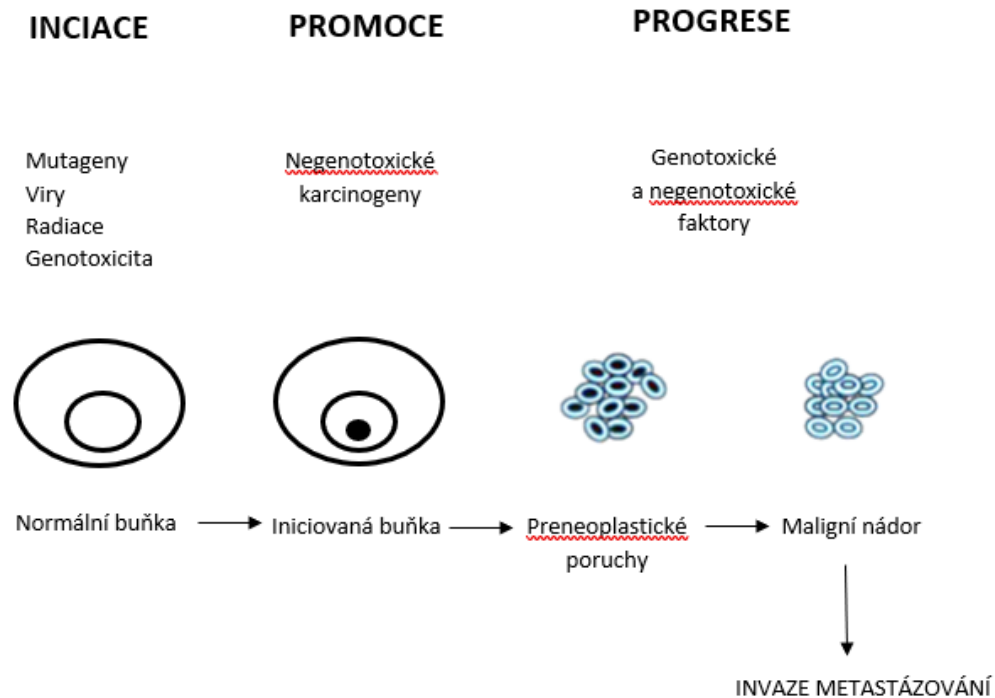
Kancerogeneze je mnohostupňový proces (obr. 1), který prochází několika stadiemi. Prvním z nich je stadium iniciační, kdy dochází k prvotním změnám v buňce způsobené mutací určitého kritického genu. Toto stadium je časově krátké a nevratné.

Další fází je fáze promoční (rozvojová) - může trvat několik let, desetiletí, dokonce i celý život pacienta. Je známo, že promoční faktory nemohou vznik nádorového onemocnění vyvolat, jen se podílejí na stimulaci latentního iniciovaného buněčného klonu k ještě intenzivnějšímu buněčnému dělení. Jako promoční faktory jsou považovány například nemutagenní látky v cigaretovém kouři, které se podílí na stimulaci iniciovaného klonu bronchiálního/plicního karcinomu.

Třetím stadiem kancerogeneze je progrese, během které je charakteristické nahromadění mutací dalších genů, změny počtu chromosomů a jejich přestavby, což je spojováno se zvyšující rychlostí proliferace, invazivitou a vznikem metastáz. V této fázi kancerogeneze se buňky nevratně mění z preneoplastické fáze v neoplastickou a benigní stadia se mění v maligní. Kancerogenezi je možno charakterizovat jako multifaktoriální a mnohostupňový proces, odpovídající morfologické, histopatologické a klinické progresi onemocnění.



## FÁZE KANCEROGENEZE

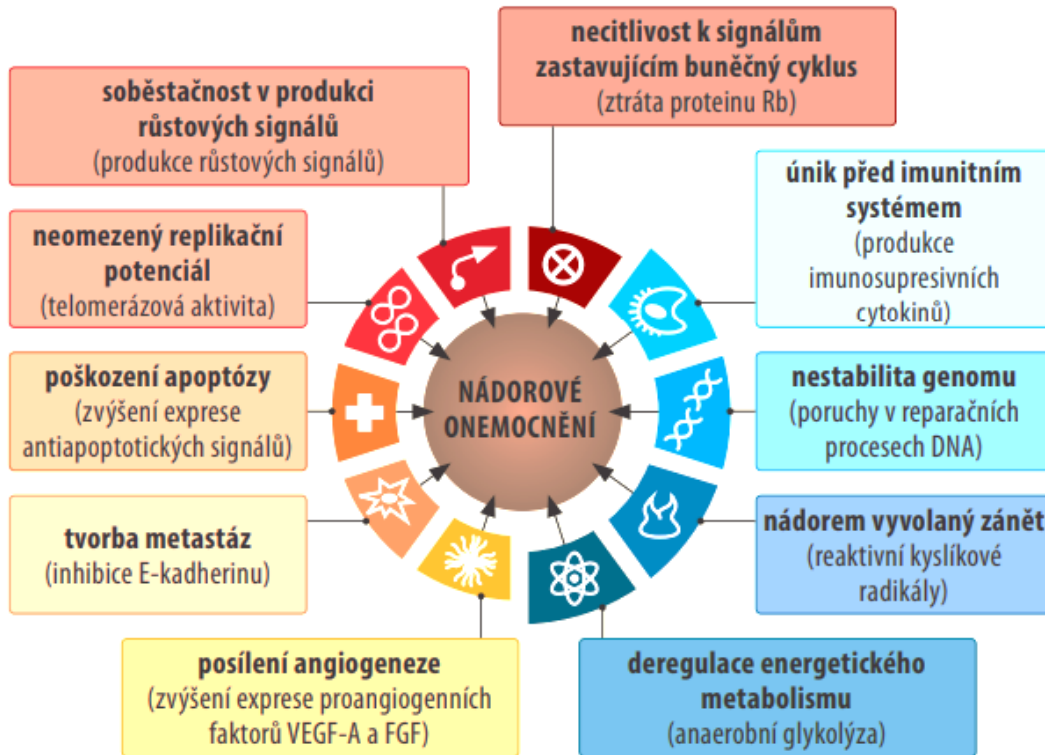


Obr. 1: Schema mnohostupňového procesu kancerogeneze

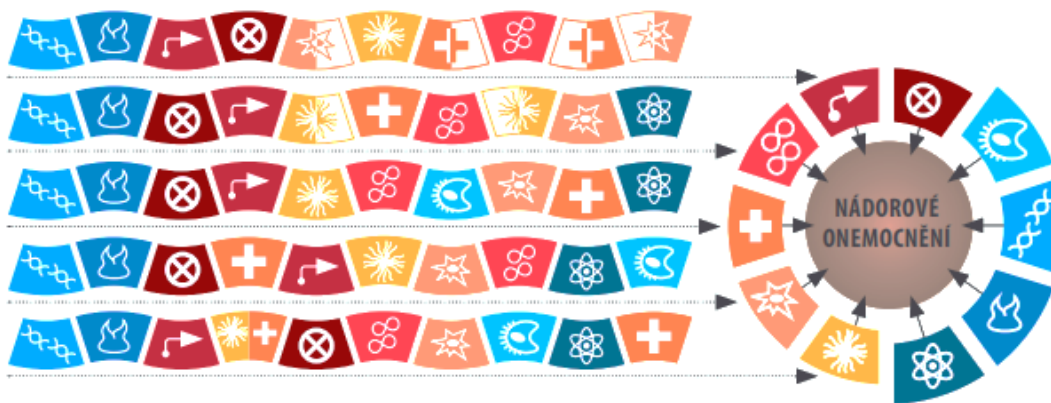
Nádorové onemocnění vzniká v zásadě poruchou homeostasy na úrovni buněk a buněčných populací. Nádorová populace je charakterizována následujícími znaky – ztrátou kontaktní inhibice a kontroly růstu, ztrátou schopnosti terminální diferenciace, snížením nebo úplnou ztrátou mezibuněčné komunikace a ztrátou schopnosti apoptosy.

V roce 2000 bylo popsáno 6 základních fyziologických změn v buňkách maligních nádorů: soběstačnost v produkci růstových faktorů, neomezený replikační potenciál, necitlivost k signálům zastavující buněčný cyklus, poškození apoptosy, indukce angiogeneze a tvorba metastáz (Hanahan *et* Weinberg, 2000). Takto definované charakteristiky byly revidovány v roce 2011 a doplněny o další čtyři znaky – schopnost deregulace metabolismu nádorových buněk, únik před imunitním systémem, genomová nestabilita a mutace, nádorový zánět

(obr. 2). Počet zásahů, pořadí a konkrétní postižené geny jsou pro každý nádor individuální (obr. 3, Hanahan *et* Weinberg, 2000).



Obr. 2: Hlavní znaky maligního nádoru (upraveno dle Hanahan *et* Weinberg, 2011)



Obr. 3.: Individuální průběh kancerogeneze (upraveno dle Hanahan *et* Weinberg, 2000)

## 2. Karcinom hlavy a krku

Mezi nádory hlavy a krku jsou řazeny nádory polykacího a dýchacího ústrojí, postihující rty, dutinu ústní, nos, vedlejší nosní dutiny, hrtan, hltan, slinné žlázy a místní lymfatickou tkáň, dále se v této oblasti setkáváme s kožními nádory, nádory v oblasti ucha a očníce a s nádory štítné žlázy (Adam *et al.*, 2004).

Výskyt nádorového onemocnění hlavy a krku je spojován s expozicí určitých rizikových faktorů. Nejčastěji se uvádí souvislost se závislostí na alkoholu a nikotinu – zvláště pak jejich kombinace, nedostatečnou ústní hygienou a špatnými dietními návyky. Tyto rizikové faktory jsou spojovány s ¾ nemocných. Jako další rizikový faktor se ukazuje vliv virového onemocnění – tři dekády zpět se onemocnění karcinomu hlavy a krku často spojovalo s Epstein-Baarové virem (EBV), avšak během posledních pár let je se vznikem karcinomů hlavy a krku spojován human papillomavirus (HPV) (Rothenberg *et Ellisen*, 2012; Vorlíček *et al.*, 2006). Jsou známy studie, které prokázaly, že 20-25 % dlaždicových karcinomů hlavy a krku je kauzálně spojeno s chronickou infekcí vysoce rizikovým kmenem HPV. Nejvyšší prevalence je tradičně detekována u tonsilárních karcinomů (50-60 %) (Muñoz *et al.*, 2006; Parkin *et Bray*, 2006; Zheng *et Baker*, 2006).

V zemích Evropské unie představují karcinomy hlavy a krku čtvrté nejčastější nádorové onemocnění u mužů. Obecně zde byl pozorován vzrůstající trend incidence od severských zemí směrem ke Středozemnímu moři. U žen je výskyt karcinomů hlavy a krku výrazně nižší než u mužů, avšak poslední dobou se počet pacientek s tímto onemocněním zvyšuje, kvůli narůstajícímu procentu kuřaček v populaci. Celosvětově se pak udává incidence 550 000 až 600 000 nových případů za rok, přičemž nejpočetnější skupinou jsou novotvary orofaryngu, hrtanu a dutiny ústní. Zhruba 300 000 pacientů tomuto onemocnění každoročně podlehnou (Ferlay *et al.*, 2010; Young *et al.*, 2015).

Lidský papilomavirus je DNA virus, který je zodpovědný za onkogenní transformaci hostitelské buňky. Genom toho viru kóduje kromě genů nezbytných k přežití také mimo jiné i proteiny E6 a E7, které souvisí s přeměnou normální buňky v buňku nádorovou. Proteiny E6 inaktivují nádorový supresor *p53*, který aktivuje telomerasu, inaktivuje *p16<sup>INK4A</sup>*, což je inhibitor cyklin-dependentních kinas a integruje s dalšími proteiny v hostitelské buňce. Proteiny rodiny E7 inhibují nádorový supresor *Rb*, inaktivuje *p21<sup>WAF1</sup>* a *p27<sup>KIP1</sup>*, které

v buňce vystupují jako inhibitory komplexů cyklin/cyklin-dependentních kinas a inhibuje signální dráhu TGF- $\beta$ . Oba proteiny se tak společně podílejí na zvýšení replikační kapacity, posílení buněčné proliferace, na inhibici apoptosy, inhibici terminální diferenciaci, změně morfologie a tím související zvýšení genetické nestability (Tan *et al.*, 1995; Weinberger *et al.*, 2006).

Během posledních 20 let došlo k velkému věkovému posunu, co se charakteristiky pacienta s orofaryngeálním karcinomem týká. Dříve toto onemocnění bylo indikováno převážně u starších osob, manuálně pracujících, řadících se do skupiny alkoholiků a kuřáků s řadou přidružených chorob. V dnešní době však je takovýto pacient mladšího věku, vzdělaný, dobrého tělesného stavu, nekuřák, abstinents a ve velké většině se jedná o HPV pozitivního jedince (Ryerson *et al.*, 2008; Reidy *et al.*, 2011).

V několika publikacích však byla potvrzena lepší prognosa pro HPV pozitivní pacienty, jelikož HPV pozitivní nádory lépe odpovídají na chemoterapii i chemoradioterapii a obecně HPV pozitivita je spojována s lepší prognosou ve srovnání s pacienty stejného stadia orofaryngeálního karcinomu HPV negativní (Weinberger *et al.*, 2009; Mellin *et al.*, 2002; Licitra *et al.*, 2006; Paz *et al.*, 1997; Rischin *et al.*, 2009; Lindel *et al.*, 2001; Li *et al.*, 2003). Klinický obraz u nádorů hlavy a krku je ve většině případů charakterizován postupnou, nenápadnou symptomatologií. Nejčastějším rysem u karcinomu laryngu je chrapot, který přetrvává několik měsíců, u karcinomu nasofaryngu si pacienti stěžují na pocit ucpaného nosu či zaléhání ucha, u karcinomu dutiny ústní se projevuje dysfagií a odynofagií. Velmi častým symptomem je nebolestivá rezistence na krku, kterou si pacient nahmatá sám a při postupném růstu a zvětšování ho dovede k lékaři. Až 2/3 nádorů hlavy a krku jsou diagnostikovány až ve stadiu pokročilého onemocnění, což významně ovlivňuje léčebné a prognostické výsledky (Smilek *et al.*, 2015).

V rámci diagnostiky nádorů hlavy a krku je základním krokem podrobná a důsledná anamnéza sledující celkovou symptomatologii v dlouhém časovém období včetně endoskopie s odběrem vzorku. Endoskopická technika umožňuje prohlédnout oblasti, které jsou během běžného vyšetření nedostupné a tím i lépe zmapovat rozsah onemocnění (Šlampa *et al.*, 2016). K tomu, aby se zjistila míra prorůstání nádorů do okolních tkání bývají využívány zobrazovací metody, jako například ultrasonografie, výpočetní tomografie, nukleární magnetická rezonance nebo pozitronová emisní tomografie. Principem

ultrasonografie je zpracování průchodu a odrazu ultrazvukového signálu od měkkých tkání. Výhodou pro pacienta je neinvazivnost a snadná dostupnost (Čelakovský *et al.*, 2012).

Výpočetní tomografie (CT, computed tomography) je rychlé a jednoduché vyšetření, při kterém však bývá použita intravenózní jodová kontrastní látka, díky které je určeno prvotní morfologické zobrazení nádoru. Vyšetření má však právě kvůli kontrastní látce nezanedbatelnou radiační zátěž, avšak CT metoda dokáže posoudit také rozsah a prorůstání nádorů do okolních tkání.

Nukleární magnetická rezonance (NMR, nuclear magnetic resonance) je nejpřesnější metoda využívána k zobrazení měkkých tkání a jejich patologie. NMR má pro diagnostiku karcinomu hlavy a krku zásadní význam, jelikož lze díky ní odhalit míru prorůstání nádoru intrakraniálně pod spodinu lebeční, posouzení infiltrace mozkových obalů, případně i samotné mozkové tkáně (Šlampa *et al.*, 2016).

Pozitronová emisní tomografie (PET, positron emission tomography) umožňuje zobrazit nádorové tkáně na základě její vyšší metabolické aktivity oproti jiným tkáním. U PET se využívá detekce metabolicky zpracované glukosy, která je označena radionuklidem (18F-FDG, 18-fluorodeoxyglukosa). Limitní rozlišovací schopnost pro PET je 5-7 mm. Velmi často se využívá kombinace zobrazovacích metod PET a CT například u pacientů, kde jsou známy metastáze, avšak je potřeba detekovat primární nádor (Čelakovský *et al.*, 2012).

## **2.1 Léčba nádorů hlavy a krku**

Existuje několik možností, jak se dá karcinom hlavy a krku léčit, avšak o způsobu léčby rozhodují jisté faktory – pokročilost nádoru (hodnoceno dle TNM klasifikace), histologický typ a lokalizace nádoru, věk nemocného a jeho celkový zdravotní stav (Adam, 2004).

### **2.1.1 Chirurgická léčba**

Chirurgická léčba je základní modalitou u pacientů se solidním nádorem hlavy a krku. Limituje ji však například vysoký věk pacienta nebo umístění nádoru v místě, které nelze operovat. Chirurgický zákrok podstupují až 2/3 pacientů s nádorem hlavy a krku. V praxi se lékaři snaží o co nejradikálnější odstranění a dosažení maximálního léčebného účinku za minimálního funkčního poškození (Vorlíček, 2006).

### **2.1.2 Radioterapie**

Léčba ionizujícím zářením je účinnou v léčbě onemocnění s místním šířením. U pacientů s karcinomem hlavy a krku s časným stadiem onemocnění jsou výsledky léčby srovnatelné s výsledky dosaženými chirurgickou resekcí. Radioterapie je preferována před chirurgií například u nádorové lokalizace v hrtanu kvůli možnosti zachování plné funkčnosti orgánu. U středních velikostí nádorů je radioterapie používána jako adjuvantní, tedy po chirurgickém zákroku s cílem, co největší místní kontroly (Spurný et al., 2002).

### **2.1.3 Chemoterapie**

Tento typ léčby je u pacientů s karcinomem hlavy a krku využíván okrajově, omezen na paliativní podání u pacientů s opakujícím se onemocněním nebo se vzdálenými metastázemi. Možný přínos chemoterapie je využíván ve zmírnění příznaků, způsobených tlakem rostoucího nádoru, proto je vždy na zvážení lékaře podání v souvislosti s toxicitou léčiva. V indikaci metastatického a recidivujícího onemocnění bylo vyzkoušeno celá řada cytostatik – metotrexát, 5-fluorouracil, bleomycin, mitomycin C, cisplatina, karboplatina

a v posledních letech také taxany, gemcitabin, hydroxyurea, ifosfamid nebo vinorelbin. Kombinace jednotlivých léčiv se užívá u mladších pacientů s karcinomem hlavy a krku s dobrým zdravotním stavem (Lo Nigro *et al.*, 2017)

#### **2.1.4 Biologická léčba**

Nejčastější léčbou využívanou u pacientů s karcinomem hlavy a krku je tedy chirurgický zákrok, radioterapie, případně chemoterapie. Biologická léčba se v těchto případech začala využívat až v posledních letech, a to u recidivujících a metastatických nádorů, především jako úleva pacienta před neselektivitou a toxicitou léčiv. Významným cílem protinádorové léčby je, vzhledem ke svému vlivu na buněčné dělení, diferenciaci a metastazování, receptor pro epidermální růstový faktor (EGFR). K léčbě pokročilých nádorů hlavy a krku jsou aktuálně podávány biologické látky – cetuximab, gefitinib, erlotinib a lepatinib. Nejzajímavější výsledky byly dosaženy u pacientů s karcinomem hlavy a krku, jež dostávali cetuximab, který zvyšuje protinádorový účinek současně podávaných cytostatik a záření. Ve studii dle Bonnera se u neléčených pacientů s nádory orofaryngu, hypofaryngu a laryngu v pokročilém stadiu bez možnosti chirurgického zákroku, kteří vykazovali dobrý zdravotní stav, nasadila léčba radioterapií anebo radioterapií v kombinaci s cetuximabem. Výsledky ukázaly významné zlepšení při použití právě kombinovaného postupu (Bonner *et al.*, 2006).

### 3. Nádorové onemocnění plic

Karcinom plic neboli zhoubný nádor plic, je epitelové maligní nádorové onemocnění, které postihuje tkáň plic i průdušek a je řazen ve všech průmyslově rozvinutých zemích k nejzávažnějším zdravotním problémům. V České republice je karcinom plic nejčastěji se vyskytující zhoubný nádor u mužů a jde o druhý až třetí nejčastěji se vyskytující zhoubný nádor u žen. I když se mortalita mužů od 90. let minulého století mírně snižuje a za posledních 25 let klesla asi o 15 %, u žen se naopak rok od roku zvyšuje. Karcinom plic patří do skupiny zhoubných nádorů, které vznikají nekontrolovatelným bujením a šířením nádorových buněk. Již v časném stadiu vývoje mají nádorové buňky schopnost cestovat krevním řečištěm ke vzdáleným orgánům a vytvářet tak dceřiná ložiska, tzv. metastáze. Imunitní systém pacienta zpravidla není schopen některé nádorové buňky včas rozpoznat a vytvořit si proti nim účinnou obranu. (Le Chevalier, 2011).

V současnosti platná WHO klasifikace rozeznává několik typů epitelových maligních nádorů plic (Tab. I). Všechny tyto karcinomy zahrnují jeden nebo více subtypů. Z hlediska biologických vlastností se však v praxi používá zjednodušené rozdělení bronchogenních karcinomů na dva typy – malobuněčný (small cell lung cancer – SCLC) a nemalobuněčný (non-small cell lung cancer – NSCLC) (Kolek, 2010).

Malobuněčný epitelový karcinom představuje okolo 15-20 % z celkového počtu histologických typů plicní rakoviny. Je charakterizován rychlým růstem s výrazným sklonem k častému metastázování do kostí, jater, centrálního nervového systému a nadledvin. Z počátku jsou SCLC velmi sensitivní na radioterapii a chemoterapii. Jejich chemosensitivita a radiosenzitivita však po čase přechází v radiorezistenci a chemorezistenci. Tento typ karcinomu obvykle není vhodný k chirurgickému odstranění (Skříčková *et al.*, 2008a).

Nemalobuněčný karcinom se vyskytuje mnohem častěji než malobuněčný typ (Tab II). Jeho míra výskytu je odhadována na 80-85 % případů (Skříčková *et al.*, 2008b). NSCLC obvykle vykazuje pomalejší růst a metastázuje v pozdějších obdobích. Možnosti chirurgické resekce jsou v tomto případě obecně větší než u malobuněčného karcinomu, jelikož ložiska metastáz vznikají později. V pokročilejších stádiích se k léčbě využívá chemoterapie a radioterapie. Citlivost nádoru na chemoterapii a radioterapii je většinou nižší než u malobuněčného karcinomu (Klein, 2006).



**Tab I: WHO klasifikace karcinomů plic**

<b><u>Skvamocelulární karcinom</u></b>	<b><u>Velkobuněčný karcinom</u></b>
Papilární	Velkobuněčný neuroendokrinní karcinom
Světlobuněčný	Bazaloidní karcinom
Malobuněčný	Lymfoepiteliomu podobný karcinom
Bazaloidní	Světlobuněčný karcinom
<b><u>Adenokarcinom</u></b>	Velkobuněčný karcinom s rhabdoidním fenotypem
Adenokarcinom, smíšený subtyp	<b><u>Malobuněčný karcinom</u></b>
Acinární adenokarcinom	Kombinovaný malobuněčný karcinom
Papilární adenokarcinom	<b><u>Adenoskvamózní karcinom</u></b>
Adenokarcinom s lepidickým typem růstu	Sarkomatoidní karcinom
<ul style="list-style-type: none"> <li>● mucinózní</li> </ul>	Pleomorfní karcinom
<ul style="list-style-type: none"> <li>● nemucinózní</li> </ul>	Vřetenobuněčný karcinom
<ul style="list-style-type: none"> <li>● smíšený mucinózní/nemucinózní</li> </ul>	Obrovskobuněčný karcinom
Solidní adenokarcinom s hlenotvorbou	Karcinosarkom
<ul style="list-style-type: none"> <li>● fetální adenokarcinom</li> </ul>	Pulmonární blastom
<ul style="list-style-type: none"> <li>● mucinózní adenokarcinom</li> </ul>	<b><u>Karcinom typu slinných žláz</u></b>
<ul style="list-style-type: none"> <li>● mucinózní cystadenokarcinom</li> </ul>	Mukoepidermoidní karcinom
<ul style="list-style-type: none"> <li>● adenokarcinom z prstenčitých buněk</li> </ul>	Adenoidně cystický karcinom
<ul style="list-style-type: none"> <li>● světlobuněčný adenokarcinom</li> </ul>	Epitelově-myoepitelový karcinom

Plicní nádor se může skládat z komponent SCLC a zároveň i jiného histologického typu. Takový nádor nese označení kombinovaný karcinom (Skřičková *et al.*, 2018).

V posledních letech se k léčbě rakoviny využívá také tzv. biologická (cílená) léčba. U pacientů s NSCLC jsou využívány monoklonální protilátky nebo nízkomolekulární tyrosin-kinasové inhibitory, avšak tato léčba je finančně náročná a účinná jen u vybrané skupiny pacientů (Goldstraw *et al.*, 2011).

Dle WHO se ke správné typizaci nádorů využívá “histopatologický grading” a “staging” (Dvořák *et al.*, 2004).

Histopatologický grading určuje stupeň diferenciacie nádoru a má prognostický význam:

GX	stupeň diferenciacie nelze hodnotit
G1	dobře diferencovaný
G2	středně diferencovaný
G3	nízce diferencovaný
G4	nediferencovaný

Staging definuje stadium onemocnění na základě tzv. TNM klasifikace, kde “T” označuje velikost nádoru (tumor), “N” regionální mízní uzliny (nodes) a “M” přítomnost/nepřítomnost vzdálených metastas (metastasis) (Planchard *et al.*, 2018). TNM klasifikace karcinomů plic se využívá pro malobuněčné i nemalobuněčné karcinomy a pro bronchopulmonární karcinoidy. Není používána pro sarkomy a ostatní vzácné nádory.

Tab II: Klinická stadia NSCLC (TNM klasifikace z roku 2009)

		<b>N0</b>	<b>N1</b>	<b>N2</b>	<b>N3</b>	<b>M1</b>
<b>T1</b>	a	IA	IIA	IIIA	IIIB	IV
	b					
<b>T2</b>	a	IB / IIA	IIA / IIB	IIIA	IIIB	IV
	b					
<b>T3</b>		IIB	IIIA	IIIA	IIIB	IV
<b>T4</b>		IIIA	IIIA	IIIB	IIIB	IV
<b>M1</b>	a	IV	IV	IV	IV	IV
	b					

Tab III: TNM klasifikace brochogenního karcinomu z roku 2009

<b>T – primární tumor</b>	
<b>TX</b>	Primární tumor nelze hodnotit, ale byla prokázána přítomnost nádorových buněk ve sputu nebo bronchiálním výplachu, ačkoliv nádor nebyl prokázán zobrazovacími vyšetřeními ani bronchoskopicky
<b>T0</b>	Bez příznaků přítomnosti primárního tumoru
<b>Tis</b>	Tumor <i>in situ</i>
<b>T1</b>	Tumor dosahující maximální velikosti 3 cm nebo méně v největším rozměru a je obklopen plicní tkání nebo viscerální pleurou. Bronchoskopicky nepřesahuje lobární bronchus, tzn. nádor se nešíří do hlavního bronchu. T1a $\leq 2$ cm v největším rozměru; T1b $> 2$ cm a zároveň $\leq 3$ cm v největším rozměru
<b>T2</b>	Tumor, který má jednu z následujících charakteristik, co se týká velikosti či rozsahu: <ul style="list-style-type: none"> <li>• tumor je <math>&gt; 3</math> cm a zároveň <math>\leq 7</math> cm</li> <li>• postihuje hlavní bronchus do vzdálenosti <math>\geq 2</math> cm distálně od kariny</li> <li>• invaze viscerální pleury</li> <li>• podmiňuje atelektázu nebo obstrukční pneumonii v rozsahu menším, než je celá plice</li> </ul> T2a $> 3$ cm a zároveň $\leq 5$ cm v největším rozměru; T2b $> 5$ cm a zároveň $\leq 7$ cm v největším rozměru
<b>T3</b>	Tumor $> 7$ cm nebo tumor jakékoli velikosti se šířením do hrudní stěny, bránice, frenického nervu, mediastinální pleury, parietálního perikardia hlavního bronchu ve vzdálenosti $< 2$ cm od kariny, kterou však nepostihuje. Označení také pro tumor podmiňující atelektázu nebo obstrukční bronchopneumonii kompletně celé plice či tumor, který vytváří nejméně jeden oddělený nádorový uzel ve stejném laloku
<b>T4</b>	Tumor jakékoli velikosti se šířením do mediastina, srdce, velkých cév, průdušnice, nervus laryngeus recurrens, jícnu, obratlových těl, bifurkace kariny. Tumor vytváří nejméně jeden oddělený nádorový uzel v jiném ipsilaterálním laloku
<b>N – regionální lymfatické uzliny</b>	
<b>NX</b>	Regionální lymfatické uzliny není možné hodnotit
<b>N0</b>	Bez přítomnosti metastáz v regionálních lymfatických uzlinách
<b>N1</b>	Metastázy v ipsilaterálních peribronchiálních a/nebo ipsilaterálních hilových uzlinách a intrapulmonálních uzlinách včetně postižení přímého prorůstání primárního tumoru
<b>N2</b>	Metastázy v ipsilaterálních mediastinálních a/nebo subkarinních lymfatických uzlinách
<b>N3</b>	Metastázy jsou přítomny v kontralaterálních mediastinálních uzlinách, v kontralaterálních hilových uzlinách, v ipsilaterálních nebo kontralaterálních sklaenových nebo supraklavikulárních lymfatických uzlinách
<b>M – vzdálené metastázy</b>	
<b>MX</b>	Vzdálené metastázy nelze hodnotit
<b>M0</b>	Vzdálené metastázy nejsou přítomné
<b>M1</b>	M1a: separátní nádorový nodul(y) v kontralaterálním laloku nebo maligní pleurální (nebo perikardiální) výpotek; M1b: vzdálené metastázy

### 3.1 Současná léčba nemalobuněčného karcinomu plic

Léčba plicního karcinomu závisí na histologickém typu a na jeho rozsahu. Důležitým faktorem je také fakt, zda je nádorem postižena pouze plicní tkáň, či došlo k rozšíření nádoru i do mízních uzlin nebo ke vzniku metastáz ve vzdálených orgánech. Při plánování léčby se také zohledňuje celkový stav pacienta včetně případných přidružených onemocnění. U malobuněčného plicního karcinomu je hlavní léčebnou modalitou chemoterapie – je to dáno vysokou chemosenzitivitou nádoru. U nemalobuněčného karcinomu plic je chemosenzitivita nižší a až do konce 80. let minulého století byla zpochybňována její účinnost. Právě v 80. letech bylo v několika studiích a analýzách prokázáno, že u pacientů dochází k prodloužení přežívání vlivem účinné chemoterapie ve srovnání s léčbou podpůrnou. V tomto období se začala chemoterapie využívat v širším měřítku, a to především u pacientů, kteří spadají do třetího a čtvrtého klinického stadia (Perry, 2008). Pacienty s nemalobuněčným karcinomem plic je možno podle závažnosti onemocnění rozdělit do několika typů stadií – od stadia IA až po stadium IV.

#### Klinické stadium IA, IB

V tomto stadiu je chirurgický resekcí zákrok považován za dostatečně radikální, a není zde indikována žádná další léčba. Využívá se lobektomie, což je odstranění laloku některého z orgánů, bilobektomie – odnětí dvou plicních laloků, či pneumonektomie – chirurgické odnětí celé plíce. V případech, kdy není možno provést chirurgické řešení, je indikována radioterapie, především stereotaktická. U pacientů se stadiem IIB a nádorem větším než 4 cm, by měla být zvažována adjuvantní léčba (Skříčková *et al.*, 2016; Skříčková *et al.*, 2017).

#### Klinická stadia IIA, IIB, IIIA

V případě klinického stadia IIA je indikována chirurgická léčba s následnou adjuvantní chemoterapií. Pokud nemocný není schopen chirurgického zákroku, bývá využívána k léčbě

souběžná nebo sekvenční chemoterapie. Při kontraindikaci chemoterapie se indikuje radioterapie.

U pacientů se stadiem IIIA je posuzována operabilita tumoru multidisciplinárním týmem a následně po chirurgické léčbě je indikována adjuvantní chemoterapie (Skřičková *et al.*, 2016; Skřičková *et al.*, 2017).

#### Klinické stadium IIIB

Standardní léčbou lokálně pokročilých stádií IIIB byla po mnoho let samostatná radioterapie. V posledním desetiletí řada studií ale potvrdila přínos kombinované léčby. Chirurgická léčba se využívá minimálně, a to jen v těch případech, kdy pomocí indukční chemoterapie nebo chemoradioterapie došlo k významnému zmenšení nádoru (Klein, 2006; Stolz *et al.*, 2010; Becker *et al.*, 2005).

#### Klinické stadium IV

Léčebný záměr u stadia IV je individuální a vždy zohledňuje aktuální zdravotní stav pacienta. U nemocných s dobrým celkovým stavem je vhodné podat kombinovanou chemoterapii. Radioterapie je využívána ke zmírnění symptomů způsobených samotným nádorem, ale i metastázemi (Skřičková *et al.*, 2008a). Chirurgická léčba se ve stadiu IV neprovádí. Výjimkou jsou však ti pacienti, kteří mají izolovanou metastázi pouze v jednom orgánu (v mozku, nadledvině nebo v plíci) a primární nádor i metastázi je možno kompletně chirurgicky odstranit. Nachází-li se metastáze v jiném orgánu, chirurgická léčba není vhodná (Zatloukal, 2005).

Velmi často dochází ke kombinaci různých léčebných metod. V léčbě nádorových onemocnění často nastává situace, že je možné použít buď různé léčebné metody nebo různé kombinace cytostatik, přičemž však nelze předem stanovit, který zvolený postup bude pro pacienta účinnější (Terry, 2007).

### 3.1.1 Chemoterapie NSCLC

Chemoterapie znamená podávání léků, které vznikly produkcí chemické syntézy. V onkologické léčbě se pod tímto pojmem rozumí podávání léků s cytotoxickým účinkem, ať už jsou původu syntetického nebo se jedná o deriváty látek získaných z rostlin či plísní (Sculier *et Moro-Sibilot*, 2009). V několika klinických studiích bylo prokázáno, že podávání chemoterapeutik vede ke zlepšení kvality života a prodloužení přežití o několik měsíců. Často se využívá v chemoterapeutickém režimu dvojkombinace platinového derivátu (cisplatiny nebo karboplatiny) a jednoho z následujících cytostatik: docetaxel, gemcitabin, paklitaxel, vinorelbin. Monoterapie bývá využívána u pacientů s kontraindikací k podání cisplatiny či karboplatiny a také u starších nemocných (70-75 let věku) (Ginsberg, 2002; Pallis *et al.*, 2014).

### 3.1.2 Radioterapie NSCLC

Radioterapie se využívá jako kurativní nebo paliativní metoda u všech pacientů s NSCLC. Samotná radioterapie se indikuje u pacientů ve stadiu I a II, u kterých není možná chirurgická resekce nádoru nebo těch, kteří tento způsob léčby odmítli (Skříčková *et al.*, 2017).

### 3.1.3 Cílená biologická léčba NSCLC

Biologické preparáty fungují u nádorových buněk jiným mechanismem než standardní chemoterapie – selektivně totiž zasahují do jejich intracelulárních pochodů. V cílené biologické léčbě se využívají nízkomolekulární látky, které po navázání do receptorových míst nádorových buněk změny jejich schopnosti jako je inhibice apoptosy, schopnost tvorby cév a vlastního zásobování nádoru živinami nebo schopnost nádorové buňky metastázovat. Biologická cílená léčba NSCLC je velkým trendem a stále se hledají nové nadějně preparáty s co největší účinností.

K léčbě pokročilého stadia NSCLC se u nás běžně využívají dva preparáty – erlotinib a gefitinib, které slouží jako inhibitory tyrosinkinasy (TKI). Druhou generací léčiv TKI je afatinib, který dle studií významně prodlužuje dobu do progresu v porovnání s nejlepší

standardní chemoterapií. Na rozdíl od erlotinibu a gefitinibu však afatinib působí jiným způsobem. Jeho mechanismus je založen na ireverzibilní blokádě rodiny receptorů ERBB, které jsou součástí signální dráhy umožňující nádorovým buňkám růst, metabolizovat a migrovat (Rossel *et al.*, 2012; Yang *et al.*, 2012).

Mezi další léčebné preparáty spojené s cílenou biologickou léčbou u pacientů s NSCLC patří bevacizumab – protilátka blokující receptor vaskulárního endoteliálního růstového faktoru (VEGFR), přičemž právě zvýšená exprese VEGFR je negativním prognostickým faktorem vedoucí ke zhoršení přežívání pacientů s NSCLC (Reck *et al.*, 2009).

Mezi preparáty, které ovlivňují angiogenezi patří nintedanib a ramucirumab. Nintedanib blokuje tři receptory pro růstový faktor: receptory pro vaskulární endotelový faktor (VEGFR1-3), receptory pro růstový faktor odvozený od trombocytů (PDGFR- $\alpha$  a PDGFR- $\beta$ ) a receptory pro fibroblastový růstový faktor (FGFR1-3). Ramucirumab je plně humánní monoklonální protilátka s vysokou vazebnou afinitou k extracelulární vazebné doméně receptoru typu 2 pro VEGFR (VEGFR2), čímž znemožňuje receptorovou vazbu a aktivaci signální dráhy všemi aktivujícími ligandy (Reck, 2013; Hanna, 2013; Perol *et al.*, 2014).

U cca 5 % pacientů s NSCLC je přítomen specifický fúzní onkogen EML4-ALK a proto je podáván těmto pacientům lék krizotinib, což je selektivní inhibitor ALK (anaplastické lymfomové kinas) (Shaw *et Salomon*, 2011).

### **3.1.4 Imunoterapie NSCLC**

Nádory mají různé mechanismy, které jim umožňují uniknout kontrole imunitního systému – například redukcí antigenů a snížením exprese molekul třídy I MHC a kostimulačních molekul, což vede k poruše rozpoznávání a aktivaci T lymfocytů a k navození imunotolerance. Také sekrece různých cytokinů ovlivňuje dozrávání dendritických buněk a vytváří tak imunosupresivní prostředí. Nádorové buňky se také mohou stát rezistentními vůči účinkům cytotoxických T lymfocytů tím, že aktivují apoptosu (Kelly *et al.*, 2010).

Imunoterapie je založená na využití imunitního systému ke kontrole a následné eliminaci nádorového onemocnění. První imunoterapeutické léky, např. interleukin 2, interferon a vakcíny první generace, nebyly v praxi zcela úspěšné. Posun nastal až s poznatky, že léčiva založená na monoklonálních protilátkách umožňují cíleně ovlivňovat kontrolní body

(checkpointy) imunitní reakce (Fong *et al.*, 2008). Velkým posunem byl objev léčiva s protilátkou anti CTLA-4 (cytotoxický T lymfocytový antigen 4) ipilimumab, u kterého aktuálně probíhají klinická hodnocení fáze I-III ve všech stadiích NSCLC (Al-Shibli *et al.*, 2008). Z dalších léčiv v rámci imunoterapie u pacientů s NSCLC je využíván nivolumab a pembrolizumab jakožto protilátky proti receptoru programované buněčné smrti a atezolizumab a durvalumab jako protilátky proti ligandu receptoru programované buněčné smrti (Spigel *et al.*, 2015; Reck *et al.*, 2016; Garassino *et al.*, 2016).

### **3.1.5 Současná léčba malobuněčného karcinomu plic**

Malobuněčný karcinom plic je charakteristický rychlým a agresivním růstem, časným metastazováním a je více citlivý k chemoterapii a radioterapii. Pro malobuněčný karcinom se stejně jako u NSCLC používá TNM klasifikace a pacienti s malobuněčným karcinomem plic se dají rozdělit do dvou stadií – limitované stadium a extenzivní stadium.

Limitované stadium (Limited disease, LD) je definováno jako onemocnění s lokalizací pouze v jednom plicním křídle s/bez postižení ipsilaterálních nebo kontralaterálních mediastinálních nebo supraklavikulárních uzlin a s/bez ipsilaterálního výpotku, které může být ozářeno v rámci jednoho ozařovacího pole. Extenzivní onemocnění (Extensive disease, ED) zahrnuje všechny ostatní formy malobuněčného karcinomu plic. (Češka *et al.*, 2010).

Pro pacienty v LD stadiu je jako léčba využívána kombinace chemoterapie a radioterapie s tím, že radioterapie by měla být nasazena co nejdříve po zahájení chemoterapie, ideálně souběžně s prvním či druhým cyklem chemoterapie. Operativní zákrok se využívá u LD pacientů pouze u stadia I-IIA. U ED stadia je primární volbou systémová chemoterapie, indikace ozáření je vyhrazeno pro symptomatickou progresi v oblasti primárního nádoru či metastáz (Almquist *et al.*, 2016; Corso *et al.*, 2015; Niho *et al.*, 2008).



## 4. Epigenetika

Epigenetické modifikace DNA umožňují regulovat genovou expresi bez nutnosti změny genetického kódu v dané buňce. Epigenetické změny jsou spontánní, reverzibilní a dědičné z jedné generace na další. Ovlivňují buněčnou diferenciaci, morfogenezi, proměnlivost i adaptabilitu organismu. Mezi epigenetické mechanismy jsou řazeny methylace DNA; posttranslační kovalentní modifikace histonů zahrnující acetylaci, methylaci, fosforylaci, ubiquitinylation a sumoylaci nebo postranskripční regulace aktivity genů pomocí miRNA.

Genetický materiál je v eukaryotních organismech lokalizován v jádře buněk a má podobu chromatinu, který umožňuje díky strukturálním a dynamickým změnám ovlivňovat různé procesy buněčného jádra, jako je aktivace, či inhibice genomové exprese. Tento komplex je označován jako epigenom, který se svojí strukturou liší v jednotlivých buňkách téhož organismu, na rozdíl od genomu, který je identický v každé buňce daného jedince.

### 4.1 Methylace DNA

Jedná se o starobylou epigenetickou modifikaci, která byla popsána u archebakterií, u bakterií i eukaryot. U některých organismů však byla schopnost methylace DNA v průběhu evoluce druhotně ztracena, jako například u *Saccharomyces cerevisiae*, či *Caenorhabditis elegans* (Buck-Koehntop *et* Defosse, 2013). Methylace DNA u bakterií je využívána jako ochranný proces jejich vlastní DNA před rozpoznáním restrikčními enzymy, které mají za úkol štěpit cizorodou nemethylovanou bakteriofágovou DNA (Albu *et al.*, 2012).

Methylace DNA je první a nejlépe prostudovaný epigenetický mechanismus. První modifikovaná nukleotidová báze byla objevena v roce 1948 pomocí papírové chromatografie v DNA z telecího brzlíku a jednalo se o 5-methylcytosin (5mC) (Hotchkiss, 1948). Methylace DNA popisuje modifikaci nukleových bází cytosinu a adeninu kovalentní adicí methylového zbytku. U adeninu dochází k připojení methylové skupiny na šestý uhlík (N6-methyl adenin, 6mA), u cytosinu buď na čtvrtý (N4-methyl cytosin, 4mC) nebo na pátý uhlík, kdy vzniká 5mC (obr. 4) (Kim *et al.*, 2009; Albu *et al.*, 2012). Zatímco varianty 6mA a 4mC se vyskytují pouze u prokaryot a minority eukaryot, varianta 5mC je jedna z nejčastějších epigenetických modifikací u eukaryot vůbec (Fu *et al.*, 2015; Zhang *et al.*, 2015;

Greer *et al.*, 2015; Zemach *et al.*, 2010). Výskyt 5mC v genomu různých druhů organismů je variabilní.

U bezobratlých je přítomnost 5mC často minimální, u obratlovců se udává, že je methylováno okolo 8 procent cytosinových bází a u rostlin zastoupení 5mC činí až 30 % (Adams *et Burdon*, 1985). U savců jsou 5mC lokalizovány v oblastech nazývaných CpG-ostrovy, což jsou sekvence DNA, kde je cytosin (C) propojen s guanosem (G) pomocí fosfátové vazby (p). Distribuce CpG-ostrovů v lidském genomu není rovnoměrná. Udává se, že v genomu savců se vyskytuje asi 35 tisíc takovýchto CpG ostrovů, které jsou dle výskytu rozděleny na startovní CpG oblasti, které se vyskytují převážně v transkripčních oblastech a úzce tak ovlivňují regulaci exprese přítomných genů nebo na nestartovní CpG oblasti, které se nacházejí mimo transkripční místa. Startovní CpG oblasti mají v sekvenci větší zastoupení cytosinu i guaninu a bývají také výrazně delší než nestartovní CpG oblasti (Ponger *et al.*, 2001). Přibližně 50-70 % methylovaných dinukleotidů CpG se u člověka nachází v oblastech heterochromatinu. Naproti tomu euchromatinové oblasti obsahující CpG zůstávají v genomu nemethylované a jejich výskyt je lokalizován převážně v promotorech genů. Výjimku tvoří geny zahrnuté v inaktivaci chromozomu X, genovém imprintingu nebo tkáňově specifické diferenciaci, kde jsou euchromatinové oblasti methylovány (Nishimura *et Paszkowski*, 2007).

K tomu, aby mohlo dojít k přidání methylové skupiny z S-adenosyl-L-metioninu k řetězci DNA, jsou potřeba specifické enzymy – DNA methyltransferasy. Jedná se o velice konzervovanou skupinu enzymů, která se podílí na přepisování epigenomu. U savců se vyskytují 3 hlavní typy methyltransferas. První je DNA methyltransferasa 1 (DNMT1), která má vysokou afinitu k hemimethylované DNA, během replikace udržuje replikační vzor v nově syntetizovaném vlákne a je zodpovědná za epigenetickou dědičnost (Li *et al.*, 1992). Druhou skupinu tvoří DNMT3a a DNMT3b, které mají v buňce funkci *de novo* methylace v raném vývoji savců a jsou nezbytné pro vytvoření DNA methylačních vzorů (Okano, 1998; Okano *et al.*, 1999).

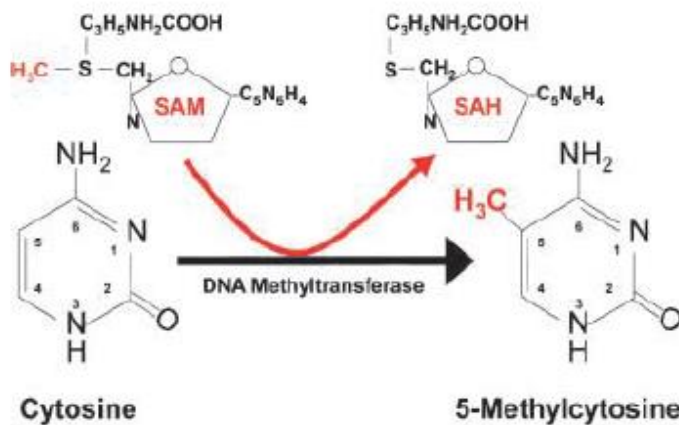
Strukturně jsou si DNMT1, DNMT3a a DNMT3b velmi podobné. Společným znakem je přítomnost 10 konzervovaných aminokyselin na C-konci, na kterém se nachází doména s katalytickou funkcí. N-konec těchto polypeptidů slouží jako regulační doména.

I přes veškerou podobnost jsou však tyto 3 enzymy exprimovány odlišně, a i jejich funkce v buňce jsou různé (Yen *et al.*, 1992; Xie *et al.*, 1999).

Velikostně je největší methyltransferasa DNMT1 s molekulovou hmotností 184 kDa, která se v adultních somatických buňkách vyskytuje v mnohem vyšších koncentracích než výše zmiňované zbylé DNMTs (Bestor, 2000). Oproti tomu DNMT3a a DNMT3b s možností *de novo* methylace se vyskytují převážně v kmenových buňkách, které dávají vzniknout buňkám s odlišným vzorem genetické informace (Okano *et al.*, 1999). Oba enzymy jsou si do jisté míry podobné a mohou se nahrazovat ve funkci, ale jejich načasování exprese a tkáňová specifita je různá. DNMT3A je exprimován ve všech buňkách, oproti tomu exprese DNMT3b byla detekována kromě buněk kostní dřeně a štítné žlázy výjimečně (Aoki *et al.*, 2001; Suetake *et al.*, 2003; Xie *et al.*, 1999). Velikost těchto dvou enzymů se pohybuje od 100 – 130 kDa. (Lee *et al.*, 2009; Peng *et al.*, 2011).

V lidském genomu se vyskytuje také DNMT3L (DNA methyltransferasa 3-like), která vykazuje strukturální podobnost na C- i N- konci s DNMT3a i DNMT3b, u níž enzymatická aktivita nebyla prokázána. Interaguje však s DNMT3a a DNMT3b a v zárodečných liniích se tak podílí na určení methylačních vzorů. Exprese DNMT3L byla prokázána majoritně v embryonálních buňkách, přítomnost v adultních somatických buňkách se potvrdila pouze v buňkách brzlíku (Suetake *et al.*, 2004; Aapola, 2000).

Dále se také v literatuře vyskytuje zmínka o DNMT2, která je homologní ke všem u obratlovců vyskytujícím se DNMTs, avšak stejně jako DNMT3L nevykazuje žádnou methylační aktivitu a její funkce proto zůstává neodhalena (Jeltsch *et al.*, 2006).



Obr. 4: Konverze cytosinu na 5-methylcytosin pomocí DNA methyltransferasy, převzato dle Attwood *et al.*, 2002

Methylace DNA je epigenetický mechanismus, který ovlivňuje mnoho zásadních buněčných procesů. Příkladem může být lyonizace jednoho z chromosomů X v samičích buňkách savců v časném stadiu embryonálního vývoje. Díky methylaci DNA je inaktivní chromosom v heterochromatinovém, tzn. kondenzovaném a neaktivním stavu a v buňce tak dochází ke správné kompenzaci dávky genů lokalizovaných na chromosomu X (Vyskot, 1999). Methylace DNA ovlivňuje také genomický imprinting, což je genetický jev, který představuje rozpor s Mendelovskými pravidly dědičnosti, který tvrdí, že funkce genu nezávisí na tom, zda byl zděděn od matky nebo otce. Manifestace genu z pohledu genomického imprintingu je totiž výsledkem různého stupně methylace cytosinu maternálního nebo paternálního chromosomu, což souvisí s kondenzací chromatinu a následně s aktivitou, případně inaktivitou konkrétního genu (Eden *et Cedar*, 1995). Mimoto se methylace DNA uplatňuje také při replikaci, rekombinaci a opravě DNA a supresi transponovatelných DNA elementů (von Känel *et Huber*, 2013).

## 4.2 Demethylace DNA

Demethylace DNA je proces, při kterém dochází k odstranění methylové skupiny z 5-methylcytosinu a může probíhat pasivně, kdy nově vznikající řetězec DNA po replikaci není methylován, nebo aktivně na replikaci nezávislým způsobem. Aktivní demethylace byla zaznamenána v zygotách, pluripotentních zárodečných buňkách, neuronech či T lymfocytech (Chen *et Riggs*, 2009). Pasivní demethylace DNA bývá využívána například v léčbě myelodysplastického syndromu s využitím 5-azacytidinu, což je chemický analog cytidinu a zapříčiňuje hypomethylace a demethylace DNA inhibicí DNA methyltransferas (Borodovsky *et al.*, 2013).

## 4.3 Modifikace histonů

Histony jsou vysoce konzervované bazické, kladně nabitě nukleoproteiny, které společně s DNA tvoří nukleosom, základní stavební jednotku chromatinu. Histony se dělí do pěti skupin – H1, H2a, H2b, H3 a H4. Histony H2A, H2B a H1 mají vysoké množství lysinových zbytků, kdežto histony H3 a H4 jsou bohaté na arginin. Oktamery histonů H2a, H2b, H3 a H4 se podílejí na formování nukleosomu a histon H1 pomáhá kondenzovat nukleosomy do chromatinových vláken (Moss *et al.*, 1977; Snustad *et Simmons*, 2016). C-konce histonů směřují do jádra nukleosomu, naopak N-konce z nukleosomu ven, kdy právě v N-terminálních oblastech se majoritně vyskytují kladně nabitě aminokyselinové zbytky argininu a lysinu, které jsou cílem postranslačních modifikací (Verbsky *et Richards*, 2001). Mezi postranslační modifikace histonů se řadí acethylace, methylace, fosforylace, ubikvitinylace anebo sumoylace, které mění strukturu chromatinu a společně s methylací DNA se podílejí na regulaci genové exprese (Kouzarides, 2007; Snustad *et Simmons*, 2016). Struktura chromatinu může být dvojitá – euchromatinová a heterochromatinová. Euchromatin je méně kondenzovaný, méně barvitelný, replikuje se dříve než heterochromatin a jedná se o transkripčně aktivní oblast genomu obsahující hyperacetylované histony. V interfázi je většinou despiralizován. Heterochromatin v eukaryotických genomech obsahuje hypoacetylované histony a je geneticky inaktivní. Heterochromatin se v buňkách může vyskytovat v podobě fakultativní, což souvisí s epigenetickou represí některých

chromosomálních oblastí nebo konstitutivní, který se vyskytuje v oblastech centromer a subtelomerových oblastech (Ansari *et al.*, 2016; Shukla *et al.*, 2013)

#### 4.4 Acethylace histonů

Acethylace histonů je posttranslační proces, při kterém dochází k přenosu acetyllové skupiny z acetylkoenzymu A na  $\text{NH}_3^+$  skupinu lysinu na jejich N-terminálním konci za využití enzymů rodiny histon acetyltransferas (HATs). Acethylace histonů je reversibilní a ke zpětné reakci je nutná přítomnost histondeacethylas (HDAC) (Kuo *et al.*, 1998). Obecně je tento epigenetický jev spojován se snížením pozitivního náboje histonového aminoterminálního konce, což způsobuje sníženou interakci mezi histonem a DNA a vede ke zvýšené transkripční aktivitě. DNA má v tomto případě strukturu euchromatinu a je obnažena pro transkripční faktory (Hong *et al.*, 1993; Brownell *et al.*, 1996). Naopak snížená acethylace histonů souvisí se sníženou transkripční aktivitou až s úplným potlačením DNA (Braunstein *et al.*, 1993).

HATs se dělí na dvě základní skupiny, HAT typu A a HAT typu B. Enzymy HAT typu A jsou lokalizovány v jádře, regulují genovou expresi pomocí acethylace nukleosomálních histonů pomocí přítomné bromodomény. Skupina HAT typu B se vyskytuje v cytoplazmě a je vysoce afinitní k volným, nově vzniklým histonům před začleněním do chromatinu (Brownell *et al.*, 1996). HATs se mohou dělit dle své struktury a toho, jaké histony acethylují na další tři skupiny acetyltransferas – GNAT, MYST a p300/CBP. Tato diversifikace zahrnuje HAT typu A i HAT typu B (Sharma *et al.*, 2016). GNAT je rodina enzymů, která acethyluje histony, nehistonové proteiny, transkripční faktory a protein p53 (Vetting *et al.*, 2005). Jako první enzym z této skupiny byl nalezen GCN5, který acethyluje lysinové zbytky 14 u histonu H3 a lysinové zbytky 12 a 16 u histonu H4 (Kuo *et al.*, 1996). K MYST acetyltransferasám bývá řazena Sas2, Sas3 nebo Esa1, kdy pro Esa1 byly objeveny dva substráty a to propionyl-CoA a mnohem častěji využívaný acetyl-CoA (Berndsen *et al.*, 2007). Substrátem pro skupinu p300/CBP jsou čtyři typy histonů (H2A, H2B, H3, H4), ale největší aktivitu projevuje u histonů H3 a H4.

## 4.5 RNA interference

RNA interference představuje vysoce konzervovaný molekulární mechanismus, který je využíván u zvířat, rostlin, hub i virů k posttranskripčnímu, sekvenačně specifickému umlčování genové exprese. V jednotlivých organismech bývá využívána při řízení diferenciaci buněk, morfologickém vývoji nebo ovlivňování stavu chromatinu v buňce, který reguluje aktivitu genů. Předpokládá se, že RNAi vznikla jako obrana eukaryotických buněk proti transposonům a virům. Funkčními jednotkami RNAi jsou jednořetězcové molekuly RNA o délce okolo 20 páru bazí, které jsou schopny se párovat na základě komplementarity svých bazí a vytvářet dvouřetězcové molekuly. Tato komplementarita zaručuje, že k efektu RNAi nebude docházet náhodně, ale bude namířen konkrétně k určitým cílům (Appasani *et al.*, 2005). Existuje několik typů molekul RNA, jež se v buňkách podílí na RNAi. Mezi ně například mikroRNA (miRNA), malé interferující RNA (small interfering RNA, siRNA), malé vlásenkové RNA (small hairpin RNA, shRNA), bifunkční sh-RNA (bifunction shRNA, bi-shRNA) či piRNA (PIWI-interacting RNA), které vznikají v jádře po transkripci příslušných genů.

### 4.5.1 Historie a charakteristika mikroRNA

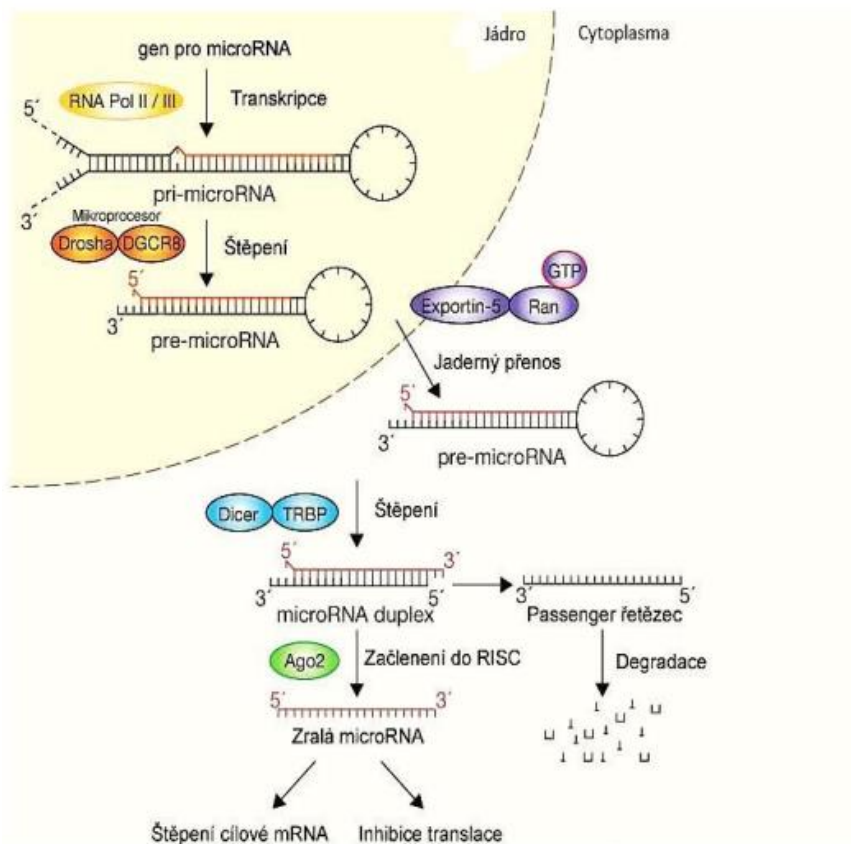
MikroRNA (miRNA) jsou nejvíce prozkoumanou skupinou malých RNA zodpovědných za RNAi. Do dnešní doby bylo prokázáno, že se podílí na regulaci exprese více než 40 % savčích genů kódujících proteiny (Lewis *et al.*, 2003). Jako první miRNA byla objevena v roce 1993 miRNA u hlístice *Caenorhabditis elegans* (*C. elegans*) a byla pojmenována jako lin-4 (lineage-4). Bylo zjištěno, že LIN-4 protein se podílí na represi LIN-14 proteinu, který zodpovídá při diferenciaci za přechod z prvního larválního stadia do stadia následujícího (Lee *et al.*, 1993; Wightman *et al.*, 1993; Feinbaum *et al.*, 1999; Olsen *et al.*, 1999). O několik let později byla u *C. elegans* objevena další miRNA, let-7 (lethal-7). Tato miRNA je exprimována v buněčném cyklu později než lin-4 a je zodpovědná při diferenciaci za přechod jedince ze čtvrtého larválního stadia do stadia imaga (Reinhart *et al.*, 2000; Slack *et al.*, 2000). Při downregulaci této miRNA se buňky nediferencují správně, neustále se dělí a získávají tak vlastnosti podobné nádorovým buňkám (Kent *et al.*, 2006). Zajímavostí je,

že *let-7* gen stejně jako *let-7* RNA byly objeveny ve stejném roce i u člověka, *Drosophily melanogaster* (*D. melanogaster*) a dalších bilaterálních zvířat (Pasquinelli *et al.*, 2000). Do dnešní doby bylo popsáno 1982 miRNA u člověka, 274 miRNA u *D. melanogaster*, 330 miRNA u *C. elegans*, 398 miRNA u *Danio rerio*, 907 miRNA u *Gallus gallus* a 664 miRNA u *Arabidopsis thaliana* (*A.thaliana*) (www.miRBase.org). Na základě rozmístění v genomu je možno geny pro miRNA rozdělit do několika skupin. U rostlin jsou geny pro miRNA lokalizovány převážně v mimogenových oblastech, kdežto u živočichů je lze najít v genových oblastech, často v intronech kódujících genů (Rodriguez *et al.*, 2004). Geny pro miRNA jsou u člověka rozmístěny náhodně po všech chromosomech kromě chromosomu Y. Je také zajímavé, že polovina genů pro miRNA u člověka byla nalezena v blízkosti dalšího genu pro miRNA (Lagos-Quintana *et al.*, 2001; Lau *et al.*, 2001). Takovéto klastrové geny bývají transkribovány současně za vzniku polycistronního primárního transkriptu (Lee *et al.*, 2002). Obvykle se v jednom klastru vyskytují dva až tři geny, zatímco na chromosomu 13 se vyskytuje takovýchto genů v klastru celkem sedm (Calin *et al.*, 2004; He *et al.*, 2005).

#### **4.5.2 Biogeneze mikroRNA**

Molekuly miRNA vznikají v jádře po transkripci genu pro miRNA na primární transkript (pri-miRNA), který je následně našťipán pomocí komplexu Drosha na prekurzorovou miRNA (pre-miRNA). Pre-miRNA je transportována z jádra do cytoplasmy pomocí exportinu-5 a v cytoplasmě je podrobena enzymatickým úpravám. Vznikne tak jedno vláknový řetězec RNA, který se pak na základě komplementarity naváže na cílové molekuly mRNA. Podle míry komplementarity může být mRNA buď kompletně degradována, či je pouze zamezeno translaci, aniž by však došlo k jejímu našťipení (Wahid *et al.*, 2010) (Obr. 5).





Obr. 5: Schema biogeneze miRNA. V jádře je gen pro miRNA přepsán na pri-miRNA, který je pomocí mikropresorového komplexu zpracován na pre-miRNA. Pre-miRNA je transportována z jádra do cytoplasmy pomocí exportinu-5 a zde je Dicerem rozštěpena na krátké dsRNA. „Passanger“ řetězec je oddělen a následně degradován, vedoucí řetězec je začleněn do RISC a společně rozštěpí cílovou mRNA, či způsobí inhibici její translace (upraveno dle Winter *et al.*, 2009).

## RNA polymerasa

RNA polymerasa II nebo III je využívána v jádře ke vzniku primárního transkriptu miRNA, který se následně skládá do vlásenky. Pro pri-miRNA vzniklé přepisem RNA polymerasy II je typická přítomnost 7-methylguanosinové čepičky na 5' konci a polyadenylové sekvence na 3' konci. Avšak jiná charakteristika platí pro miRNA vzniklé transkripcí RNA

polymerasou III. V tomto případě je na jedné straně transkriptu charakteristická přítomnost repetitivní sekvence Alu (Borchert *et al.*, 2006).

### **Mikroprocesorový komplex**

Z několika set nukleotidů dlouhé pri-miRNA se pomocí několika enzymů zvaných Mikroprocesorový komplex tvoří v jádře prekursorová miRNA o délce 60-70 nukleotidů. Mikroprocesorový komplex se skládá z několika důležitých komponent. Prvním z nich je enzym Drosha, který má vlastnosti RNasy III a je schopný rozlišit pri-miRNA od ostatních RNA v jádře, které mají shodnou vlásenkovou strukturu (Han *et al.*, 2004). Další složkou tohoto komplexu u člověka je jako kofaktor sloužící doména proteinu DGCR8 (DiGeorge syndrom critical region in gene 8). Ta se specificky váže na cílovou dvou řetězcovou molekulu pri-RNA a určuje místo, kde dojde k jejímu rozštěpení prostřednictvím enzymu Drosha. U *D. melanogaster* či *C. elegans* najdeme místo DGCR8 kofaktor Pasha. Dicer má však ale stejnou enzymovou specifitu (Lee *et al.*, 2003; Han *et al.*, 2004; Denli *et al.*, 2004; Gregory *et al.*, 2004, Landthaler *et al.*, 2004; Wang *et al.*, 2007).

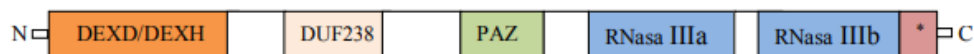
### **Exportin-5**

Transport pre-miRNA z jádra do cytoplasmy je zprostředkován přes jaderné póry pomocí proteinu exportin-5 (EXP5). Součástí EXP5 je GTP-vázající kofaktor Ran. Pokud je ke kofaktoru Ran navázán GTP, je inaktivní. K aktivaci je nezbytná hydrolyza Ran-GTP pomocí cytoplasmatického enzymu Ran-GTPasa. Díky tomu dojde k rozštěpení makroergní vazby, odštěpení fosfátu z Ran-GTP za vzniku Ran-GDP. V tento okamžik je EXP5 aktivován a může dojít k transportu pre-miRNA z jádra do cytoplasmy (Gwizdek *et al.*, 2003).

### **Dicer**

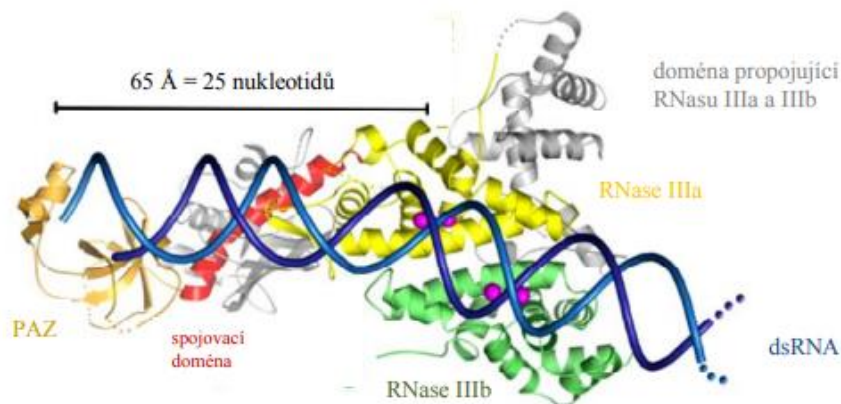
V cytoplasmě je pre-miRNA naštěpena komplexem Dicer na dvou řetězcové úseky miRNA o délce cca 22 nukleotidů. Je odštěpena nepárující se smyčka vlásenky, ale je zanechán

dvounukleotidový přesah na 3' konci (Lee *et al.*, 2003). Dicer je ATP-dependentní multienzymatický komplex složený ze čtyř odlišných domén (Obr. 6). Na N-konci se nachází RNAsová helikasová doména DEXD/DEXH a doména neznámé funkce DUF238, následovaná PAZ (PIWI-argonaute-zwille) doménou, dvěma RNasami III a C-konec je tvořen vazebnou doménou pro dsRNA (Carthew *et al.*, 2009). Hlavní funkci v komplexu Dicer zastává doména endoribonukleasa III, díky níž dochází k naštěpení fosfodiesterové vazby uvnitř řetězců dsRNA a vznikají tak molekuly miRNA s fosfátovou skupinou na 5' konci, hydroxylovou skupinou na 3' konci a s dvounukleotidovým přesahem na 3' koncích (Elbashir *et al.*, 2001). V lidských buňkách je Dicer spojován se dvěma typy proteinů obsahující dsRNA vazebné domény. Prvním z nich je TRBP (trans-activation response RNA-binding protein) a dalším je PRKRA (protein kinase, interferon-inducible double stranded RNA-dependent activator), který je znám také pod zkratkou PACT. Funkce těchto proteinů ještě nebyla zcela prokázána, ale předpokládá se, že mají určitou souvislost s tvorbou komplexu DICER a nepřímo se tak podílí na RNAi (Chendrimada *et al.*, 2005; Haase *et al.*, 2005, Lee *et al.*, 2006).



Obr. 6: Schematické znázornění primární struktury lidského multienzymatického komplexu Dicer. Dicer je tvořen RNasovou helikasovou doménou, doménou DUF238 s neznámou funkcí, PAZ doménou, dvěma RNasami III a hvězdičkou je označena vazebná doména pro dsRNA (Carthew *et al.*, 2009).

U většiny organismů lze najít pouze jeden typ Diceru, který zodpovídá za naštípání veškerých forem RNA spojovaných s RNAi (Obr. 7). Avšak u *D. melanogaster* najdeme tyto komplexy dva a u *A. thaliana* dokonce čtyři (Blevins *et al.*, 2006). V takovýchto případech má každý z typů Diceru svoji specifickou funkci. Například u *D. melanogaster*, kde byly pojmenovány jako Dicer-1 a Dicer-2, Dicer-1 katalyzuje štěpení pouze prekurzorů miRNA, za vzniku miRNA, zatímco Dicer-2 štěpí všechny ostatní dlouhé dvouřetězcové molekuly RNA (Tomari *et al.*, 2005; Lee *et al.*, 2004).



Obr. 7: Model krystalové struktury komplexu Dicer (upraveno dle MacRea *et al.*, 2006)

### RISC loading complex

Po naštěpení Dicerem se využívá další komplex, RISC (RNA-induced silencing complex). Nejprve je však našťípaná, dvouřetězová RNA integrována do RISC loading komplexu (RLC). V tento okamžik je nutné, aby došlo k rozpletení řetězce, protože RLC akceptuje pouze jedno vlákno. Každý dvouřetězec molekul RNAi se skládá z vedoucího řetězce a z tzv. „passanger“ řetězce. Tento „passanger“ řetězec je odvinut helikasou v přítomnosti ATP a následně degradován. „Passanger“ řetězec bývá v literatuře označován hvězdičkou (např. hsa-miR-21\*) a obecně se za něj považuje ten řetězec, který má termodynamicky stabilnější 5' konec. Jednořetězec, jenž je termodynamicky labilnější, je začleněn do komplexu RISC a vzniká tzv. miRNP (microRNA ribonucleoprotein complex). Tento řetězec se následně zformuje do podoby zralé miRNA (Schwarz *et al.*, 2003; Mourelatos *et al.*, 2002).

### RISC

RISC je multienzymatický komplex, jehož hlavní funkcí je rozštěpení nebo translační inhibice cílové mRNA, kterou si vyhledal na základě komplementarity s asociovanou jedno vláknovou miRNA. Základními podjednotkami komplexu RISC jsou Argonaute proteiny

(Ago) a vazebné domény PAZ a PIWI, kterými jsou jednotlivé Ago proteiny propojeny. Ago proteiny byly poprvé rozpoznány a popsány u *A. thaliana*, která měla defektní gen pro protein Argonaute1 a vykazovala specifickou poruchu v architektuře rostliny (Bohmert *et al.*, 1998). Do dnešní doby byly proteiny rodiny Ago nalezeny i u některých prokaryot, avšak primárně se vyskytují u všech eukaryot, avšak u některých můžou i sekundárně chybět, jako například u *S. cerevisiae* (Drinnenberg *et al.*, 2009). Z tohoto poznatku je možno usuzovat, že se jedná o velice konzervované a pro organismy důležité proteiny. Proteiny Ago obsahují dvě RNA vazebná místa, z nichž první z nich je PAZ doména a druhá PIWI doména, dále pak N a Mid doména (Obr. 8). PAZ doména slouží k vazbě dvou nukleotidových přesahů na 3' konci miRNA a PIWI je zodpovědná za navázání 5' konce miRNA. PIWI doména má dle krystalografické studie podobné vlastnosti jako RNasa H (Cerutti *et al.*, 2000; Song *et al.*, 2004).



Obr. 8: Schematické znázornění struktury proteinu Ago. Proteiny Argonaute jsou tvořeny N doménou, doménou PAZ, Mid doménou a doménou PIWI (Carthew *et al.*, 2009)

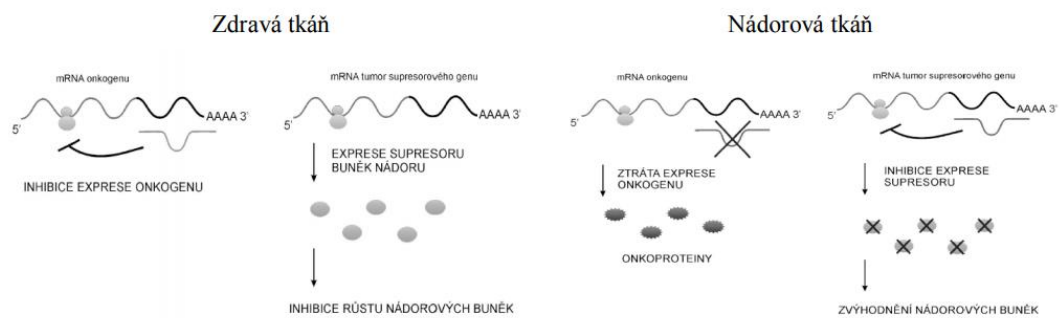
Samotné štěpení cílové mRNA při dosažení absolutní komplementarity s miRNA zajišťuje u živočichů protein Argonaute2 (Ago2), který byl objeven při purifikaci RISC komplexu u *D. melanogaster*. V lidských buňkách byly rozpoznány celkem čtyři typy proteinu Ago, ale pouze u Ago2 byla prokázána katalytická aktivita a schopnost štěpit fosfodiesterové vazby v cílové mRNA (Liu *et al.*, 2004). Pokud však komplementarita mezi oběma řetězci RNA, miRNA a mRNA, není absolutní, dochází k translační inhibici mRNA. V tomto případě se RISC s asociovanou miRNA naváže na cílovou sekvenci a znemožní její translaci na ribosomech. Ke spuštění translační inhibice postačuje hybridizace mezi druhou až sedmou nebo devátou bází miRNA s cílovou mRNA (Lewis *et al.*, 2003). Doposud však mechanismus translační inhibice komplexem RISC nebyl u savců ještě zcela prozkoumán.

### 4.5.3 Úloha miRNA v nádorech

Existuje stále více a více důkazů o tom, že miRNA mají možnost na post-transkripční úrovni kontrolovat a současně ovlivňovat proliferaci, diferenciaci i apoptosu a účastnit se tak procesů kancerogeneze, chemorezistence, či radiorezistence (Ranade *et al.*, 2010; Cho, 2011).

miRNA je možno rozdělit z hlediska jejich cílových kódujících mRNA na miRNA s onkogenním charakterem nebo miRNA s charakterem nádorového supresoru (Obr. 9). miRNA s charakterem onkogenu umlčují translaci nádorových supresorů. Hladiny exprese takovýchto miRNA jsou v nádorových buňkách oproti normálním buňkám zvýšeny, což má za důsledek snížení hladin exprese nádorových supresorů v nádorových buňkách (Kumar *et al.*, 2007). Naopak hladiny tumor-supresorových miRNA umlčující translaci onkogenů bývají v nádorových buňkách několikanásobně sníženy (Johnson *et al.*, 2005). Pro nádorové buňky je rovněž charakteristická celkově nižší hladina exprese miRNA ve srovnání s buňkami zdravé tkáně (Lu *et al.*, 2005).

Klasickým příkladem onkogenní miRNA je miR-155, jejíž až stonásobně zvýšené hladiny doprovází velkou skupinu lymfoidních malignit (Burkittova, Hodgkinova i difúzního velkobuněčného B-lymfocytárního lymfomu) (Kluiver *et al.*, 2005; Kluiver *et al.*, 2006; Eis *et al.*, 2005). Příkladem miRNA s funkcí nádorového supresoru je rodina miRNA označována jako let-7, která negativně reguluje onkogen RAS. Do regulace dalšího významného onkogenu MYC je zapojen například klastr miR-17-92, miR-142 nebo miR-155 (Chen, 2005; Esquela-Kerscher *et Slack*, 2006).



Obr. 9: Obecné schéma kontroly nádorového bujení pomocí molekul miRNA (převzato od Lochmanové a Bartoše, 2008)

## 5. Cíle práce

1. Shrnout aktuální poznatky týkající se charakteristiky karcinomů hlavy a krku a karcinomů plic a vypracovat rešerši.
2. Stanovit expresi miRNA (miR-21, miR-23b, miR-126, miR-205 a RNU 6B) ve vzorcích pacientů s nemalobuněčným karcinomem plic a korelovat získané výsledky s aktivitou proteinů ovlivňující mnohočetnou lékovou rezistenci – P-gp, MRP a LRP/MVP.
3. Porovnat epigenetické profily dlaždicových karcinomů hlavy a krku se shodnými vzorky uvulopalatofaryngoplastické tkáně a zjistit aktivitu/inaktivitu tumor supresorových genů v souvislosti s methylací a/nebo somatickou mutací.
4. Vytipovat transkripční faktory u dlaždicových karcinomů hlavy a krku související s determinací HPV<sup>+</sup> a HPV<sup>-</sup> nádorů.



## **6. Materiál a metodika**

**USA:**

### **Kultivace buněk**

V rámci experimentální části byly použity buněčné linie nádorů hlavy a krku – Ho1N1, HSC2, SKN3, UM-SCC-047 a UM-SCC-22B. Ve vodní lázni bylo předehřáto vhodné kultivační médium ke kultivaci příslušné buněčné linie na 37 °C. Následně byla vytažena z tekutého dusíku kryozkumavka obsahující vybranou buněčnou linii a vložena na 1-2 minuty do vodní lázně. V laminárním boxu bylo napipetováno 10 ml předehřátého kultivačního média do 15 ml falkonky. Pomalu, kapku po kapce, se přidal 1 ml obsahující buněčnou linii z kryozkumavky do falkonky a ta byla centrifugována při 800 rpm 3 min. V laminárním boxu byl opatrně odsát supernatant a pelet potřepáním promíchán. Do falkonky bylo připipetováno 5 ml nahřátého média, celý objem byl přenesen do kultivační lahve (T-75) a bylo připipetováno dalších 7 ml nahřátého média. Takto připravená buněčná linie byla zkontrolována pod mikroskopem a přenesena do inkubátoru (37 °C, 5 % CO<sub>2</sub>), přičemž médium v kultivační lahvi bylo třeba vyměnit každý následující den nebo každý druhý den dle toho, jak moc rychle buněčné linie rostly.

### **Dělení buněk**

Při zaplnění dna kultivační lahve buňkami příslušné buněčné linie, muselo dojít k dělení buněk do více kultivačních lahví. V laminárním boxu bylo aspirováno médium z kultivační lahve, buňky byly opláchnuty 10 ml PBS (pH 7,2-7,4) (pro T-75) a PBS byl hned z kultivační lahve odpipetován. K buňkám byly přidány 2 ml trypsin-EDTA (0,05 % trypsin) a kultivační láhev se vrátila na 5 min do inkubátoru. Po uplynutí doby inkubace byly buňky pozorovány v mikroskopu – byly kulaté, nepřichycené ke dnu kultivační lahve a plavaly volně v médiu. Následně bylo přidáno do kultivační lahve 6 ml předehřátého kultivačního média obohaceným o sérum (např. 10 % FBS v DMEM) a buňky byly rozděleny dle své schopnosti růstu do příslušného počtu nových kultivačních lahví, kam se následně připipetovalo 12-15

ml příslušného kultivačního média vhodného k růstu konkrétní buněčné linie. Kultivační lahve byly následně inkubovány v inkubátoru (37 °C, 5 % CO<sub>2</sub>). Po jednodenní kultivaci bylo vyměněno kultivační médium.

### **siRNA transfekční protokol**

Po nárůstu buněk v kultivačních lahvích se přechází k dalšímu kroku experimentu – k transfekčnímu protokolu za využití siRNA. V laminárním boxu bylo aspirováno médium z kultivační lahve, buňky byly opláchnuty 10 ml PBS (pH 7,2-7,4) (pro T-75) a PBS byl hned z kultivační lahve odpipetován. K buňkám byly přidány 2 ml trypsin-EDTA (0,05 % trypsin) a kultivační láhev se vrátila na 5 min do inkubátoru. Po uplynutí doby inkubace byly buňky pozorovány v mikroskopu – byly kulaté, nepřichycené ke dnu kultivační lahve a plavaly volně v médiu. Následně bylo přidáno do kultivační lahve 6 ml předeřhřátého kultivačního média obohaceným o sérum (např. 10 % FBS v DMEM) a buňky byly rozděleny dle své schopnosti růstu do příslušného počtu nových kultivačních lahví, kam se následně připipetovalo 12-15 ml příslušného kultivačního média vhodného k růstu konkrétní buněčné linie. Kultivační lahve byly následně inkubovány v inkubátoru (37 °C, 5 % CO<sub>2</sub>).

Zbytek suspenze, která nebyla využita pro další dělení buněk, bylo přepipetováno do 15 ml falconky a zcentrifugováno při 1000 rpm po dobu 5 minut. Následně bylo odpipetován supernatant a pipetou byl rozbit pelet. Bylo přidáno 10 ml příslušného média vhodného k právě používané buněčné kultuře a objem byl důkladně propipetován. Z této suspenze bylo odebráno 20 µl na sklíčko k měření na cellometru za účelem zjištění přesné koncentrace buněk a falconka se zbylou suspenzí byla odložena do ledu. Dle vzorce:

objem, který potřebuju x počet buněk, který chci mít v jedné jamce x objem, který mám  
k dispozici

---

koncentrace buněk, kterou mi změřil cellometr

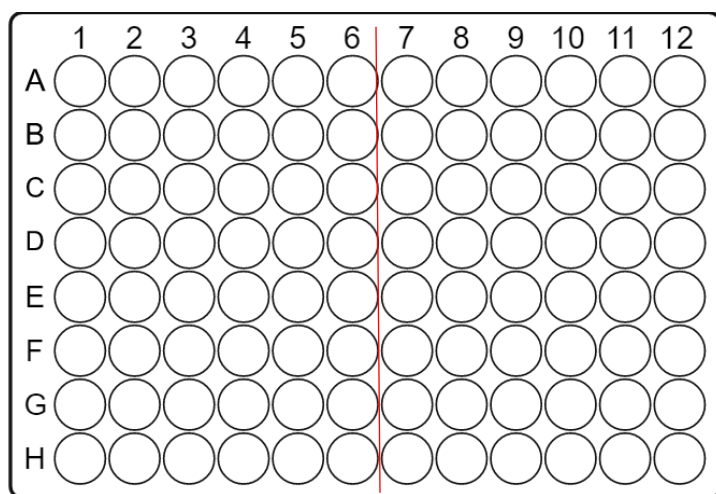
bylo smíchán příslušný objem suspenze obsahující buněčnou kulturu a čerstvé příslušné médium vhodné k růstu konkrétní buněčné linie. Pro 96 jamkovou kultivační destičku se

následně využilo 100 µl takto naředěné suspenze na jamku. Takto připravené čtyři 96 jamkové destičky byly umístěny přes noc do inkubátoru (37 °C, 5 % CO<sub>2</sub>).

Po 20hodinové inkubaci bylo odsáto z jamek kultivační médium a připipetováno 100 µl předeřátého kultivačního média Opti-MEM™ (Opti-MEM™ s redukováným sérem, ThermoFisher), kterým byly buňky promyty. Toto médium bylo ihned odpipetováno a opět připipetováno 100 µl k buňkám do jamek. Takto připravené 96 jamkové destičky byly umístěny do inkubátoru (37 °C, 5 % CO<sub>2</sub>).

Po další 4hodinové inkubaci byla měřena viabilita buněk v jednotlivých jamkách v jedné z 96 jamkových destiček. Bylo k tomu využito médium DMEM (DMEM – Dulbecco's Modified Eagle Medium, ThermoFisher) a barva alamarBlue™ (alamarBlue™ Cell Viability Reagent, ThermoFisher) v poměru 1:10. Nejprve byl z jamek odsát Opti-MEM™, následně bylo k buňkám připipetováno 100 µl naředěné směsi DMEM a alamarBlue™ a 5 jamek se využilo jako kontrolních. Za hodinu od tohoto kroku byla změřena viabilita buněk na cellometru.

Bylo připraveny 2 sady po 5 ks 1,5 ml zkumavek, přičemž do prvního setu se napipetovalo 175 µl Opti-MEM™ a 5,25 µl RNAiMAX. Do druhého setu se napipetovalo 175 µl Opti-MEM™ a 5,25 µl připravených siRNA k transfekci. Inkubace probíhala 5 min při RT. Následně byl přepipetován objem druhého setu k příslušným zkumavkám z prvního setu. Do každé jamky bylo připipetováno 10 µl transfekčního objemu dle následujícího schématu:



1. Koncentrace buněk      2. Koncentrace buněk

- Non-target siRNA – řádek B
- 1. typ siRNA A – řádek C
- 1. typ siRNA B – řádek D
- 2. typ siRNA A – řádek E
- 2. typ siRNA B – řádek F
- Kontrolní jamky k měření viability - řádek G

96 jamkové destičky byly inkubovány při 37 °C, 5 % CO<sub>2</sub> po dobu 24-48 hod pro mRNA analýzu nebo 48-96 hod pro analýzu proteinů.

Následující 3 dny se ve stejný čas přidalo do jedné z 96 jamkových destiček 100 µl mixu DMEM + barva alamarBlue™ v poměru 1:10 a opět byla detekována viabilita buněk v jednotlivých jamkách po transfekci příslušnou siRNA.

Po měření viability buněk 48 hod po transfekci se z jamek odpipetoval mix DMEM + barva alamarBlue™, připipetovalo se 100 µl 1x PBS k promytí buněk a ihned se objem zase odpipetoval. Toto promytí buněk 1x PBS se ještě jedenkrát zopakovalo a po odpipetování 1x PBS z jamek, bylo přidáno 100 µl vychlazeného QIAZolu (QIAzol Lysis Reagent, Qiagen) a buňky se nechaly zlyzovat po dobu 5 min při RT. Po uplynutí doby lýze byl sesbíráán objem z jamek se stejnou aplikací transfekčních siRNA do 1,5 ml zkumavky a doplněn do 1000 µl QIAZolem. Takto připravený objem byl následně využit z izolaci RNA.

### **Izolace RNA z buněk transfekovaných siRNA**

K 1000 µl zlyzovaných buněk v QIAZolu bylo přidáno v digestoři 200 µl chloroformu a zkumavka byla vortexována po dobu 60 s. Celý objem byl přepipetován do “phase-lock-gel” zkumavky a zvortexován po dobu 20 s. Inkubace probíhala 5 min při RT. Zkumavka byla následně zcentrifugována při 11 000 rpm, 4 °C. Vrchní fáze byla přepipetována do gDNA eliminující zkumavky s filtrem a zcentrifugována 2 min při 13 000 rpm. K získanému objemu bylo připipetováno stejné množství 70% ethanolu, mix byl aplikován do RNeasy mini kolonky (QIAGEN) a zcentrifugován při 10 000 rpm po dobu 15 s. Filtrovací kolonka byla přemístěna do nové 1,5 ml zkumavky a byla promyta 700 µl pufru RW1 při centrifugaci 10 000 rpm po dobu 15 s. Filtrovací kolonka byla přemístěna do nové 1,5 ml zkumavky a byla dvakrát po sobě promyta 500 µl pufru RPE při centrifugaci 10 000 rpm po dobu 15 s. Filtrovací kolonka byla přemístěna do nové 1,5 ml zkumavky byla zcentrifugována při 14 000 rpm po dobu 2 min. Filtrovací kolonka byla přemístěna do nové 1,5 ml zkumavky a byla eluována RNA s využitím 40 µl nuclease-free vody (inkubace 2 min, pak centrifugace při 10 000 rpm po dobu 1 min). Eluát obsahoval vyizolovanou RNA.

## Měření koncentrace RNA

Koncentrace RNA byla měřena spektrofotometricky na Nanodropu ND-1000 (NanoDrop Technologies, Wilmington, Delaware, USA). Pro měření byl použit 1  $\mu$ l eluátu s vyizolovanou RNA. Jako reference byl využit 1  $\mu$ l nuclease-free vody. Kvalita RNA byla posuzována podle hodnot poměrů absorbance při vybraných vlnových délkách 260/280 a 260/230, přitom poměr 260/280 by se měl optimálně pohybovat v rozmezí 1,8-2,0.

## Syntéza komplementární DNA (cDNA)

Komplementární cDNA byla získána v jednokrokové reverzní transkripci (RT) s použitím kitu High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems, Foster City, CA, USA). Nejprve bylo potřeba si připravit master mix, a to smícháním 10x Reverse Transcription Buffer, 100 mM dNTPs Mix, 10x RT Random Primers a MultiScribe™ Reverse Transcriptase (Tab. 4). Příslušný objem všech složek pro odpovídající počet vzorků byl napipetován do zkumavky, krátce zvortexován a rozpipetován po 5,8  $\mu$ l do PCR mikrozkuvek. Do každé zkumavky bylo přidáno 14,2  $\mu$ l naředěné RNA na množství 150 ng. Poté byla mikrozkuvka krátce zvortexována, zcentrifugována a vložena do termocykléru, kde proběhla reverzní transkripce (Tab. 5).

**Tab. 4: Složení mastermixu pro syntézu cDNA**

Chemikálie	Množství pro 1 vzorek [ $\mu$ l]
10x Reverse Transcription Buffer	2,0
100 mM dNTPs	0,8
10X RT Random Primers	2,0
MultiScribe™ Reverse Transcriptase	1,0
Celkem	5,8

**Tab. 5: Podmínky reverzní transkripce**

Reverzní transkripce	Teplota [°C]	Čas [min]
1. krok	25	10
2. krok	37	120
3. krok	85	5
Chlazení	4	∞

**Real-time PCR**

Jednotlivé komponenty nutné pro provedení real-time PCR byly 2x TaqMan® Universal PCR Master Mix (Applied Biosystems), nuclease-free voda, příslušné TaqMan® MicroRNA Assays (Applied Biosystems) a cDNA produkt, který byl naředěn ze získané koncentrace 7,5 ng/μl na 5 ng/μl (Tab. 6). Připravený PCR mix v mikrozkušavkách byl krátce zvortexován, zcentrifugován a rozpipetován do tří jamek 96 jamkové mikrotitrační destičky po 7 μl. Poté byla 96 jamková mikrotitrační destička utěsněna fólií, zcentrifugována při 1 000 rpm, 2 min při RT a vložena do ABI 7000 Real Time PCR (Applied Biosystems), kde proběhla real-time PCR (Tab. 7).

**Tab. 6: Složení PCR mixu pro real-time PCR**

Chemikálie	Množství pro 1 vzorek [μl]
TaqMan® Universal PCR Master Mix (2x)	3,5
Nuclease-free voda	0,35
TaqMan® MicroRNA Assay	1,15
cDNA produkt (naředěn na 5 ng/μl)	2
Celkem	7

**Tab. 7: Podmínky real-time PCR**

<b>Real-time PCR</b>	<b>Teplota [°C]</b>	<b>Čas</b>	<b>Počet cyklů</b>
Počáteční denaturace	95	10 min	
Denaturace	95	15 s	40
Annealing + elongace	60	60 s	
Chlazení	4	∞	

## ČR

### MATERIÁL

Jako materiál pro experimentální část dizertační práce byly použity ve formalínu fixované a do parafínu zalité vzorky nemalobuněčných karcinomů plic, které byly získány od pacientů operovaných v letech 1996 – 2002 na 1. Chirurgické klinice Lékařské fakulty Univerzity Palackého a Fakultní nemocnice Olomouc (Tab. 8). Vzorek každého případu byl hodnocen dvěma patologi. Celkem bylo použito 62 vzorků od pacientů s karcinomem plic, které pocházely od 55 mužů a 7 žen. Věkové rozpětí pacientů bylo 33 – 78 let a v této kohortě bylo 30 tumorů klasifikováno jako adenokarcinom (ADC), 26 jako skvamózní karcinom (SCC) a 6 jako velkobuněčný karcinom (LCC). Z toho 18 pacientů bylo v klinickém stádiu I, či II a 33 pacientů v klinickém stádiu III, či IV. Pro zbylých 11 pacientů nebylo klinické stadium zřejmé.

**Tab. 8: Charakteristika pacientů**

Počet všech pacientů	Pohlaví		Věk		Klinické stadium		Typ nádoru		
	Ženy	Muži	<60	≥60	I+II	III+IV	ADC	SCC	LCC
62	55	7	31	31	18	33	30	26	6

### METODY

#### Sestavení tkáňových mikroerejí

Tkáňové mikroereje byly získány z 62 vzorků primárních plicních tumorů. Správná selekce reprezentativních okrsků tkáně byla provedena na základě vizuálního srovnání bloku se standardním hematoxilínovým preparátem. Pomocí speciální punkční jehly o průměru 2 mm byly získány dva až čtyři tkáňové válečky z donorového bloku a následně byly přeneseny do recipientního bloku pomocí manuálního přístroje Galileo TMA CK3500 (BioRep, Milan, Italy). Zároveň byly do bloku také umístěny kontrolní válečky normální plicní tkáně.



## **Izolace totální RNA z parafinových bločků**

Z parafinových bločků byly pomocí skalpelu vyřezány jednotlivé vzorky plicní tkáně a byla z nich izolována totální RNA pomocí RecoverAll™ Total Nucleic Acid Isolation Kitu (Applied Biosystems, Foster City). Vzorky byly podrobeny deparafinizaci, štěpení proteasou, izolaci nukleových kyselin, štěpení nukleasou a finální purifikaci. Plicní tkáň o hmotnosti minimálně 35 mg byla vložena do 1,5 ml zkumavky a byl k ní připipetován 1 ml 100% xylenu. Zkumavka byla vložena do termobloku na 3 min při 50 °C, aby došlo k rozpuštění parafínu. Následovala centrifugace po dobu 2 min, 20 000 rpm při RT. Poté byl xylén ze zkumavky odstraněn a byl připipetován 1 ml 100% ethanolu. Zkumavka byla krátce zvortexována a zcentrifugována po dobu 2 min, 20 000 rpm při RT. Ethanol byl odpipetován, byl znovu připipetován 1 ml 100% ethanolu a následovala centrifugace za stejných podmínek, 20 000 rpm, 2 min při RT. Ethanol byl opět odstraněn. Po vyschnutí obsahu zkumavky (15-45 min) bylo k peletu přidáno 200 µl štěpícího pufru, 4 µl proteasy. Vzorek byl inkubován 15 min při 50 °C a dalších 15 min při 80 °C v termobloku. K naštěpenému vzorku tkáně bylo přidáno 240 µl izolačního aditiva, 550 µl 100% ethanolu a byl pipetováním promíchán. Filtrovací kolonka byla vložena do sběrné zkumavky a na ni bylo napipetováno 700 µl vzorku. Poté proběhla centrifugace při 10 000 rpm, RT po 30 s a objem, který prošel skrz filtrovací kolonku byl odstraněn. Zbytek vzorku byl přenesen na filtrovací kolonku, opět následovala centrifugace při 10 000 rpm, RT po 30 s a objem, který zůstal ve sběrné zkumavce byl odstraněn. Na filtrovací kolonku bylo napipetováno 700 µl promývacího pufru Wash 1 a zkumavka byla zcentrifugována při 10 000 rpm, RT po 30 s. Objem, který prošel přes filtrovací kolonku byl odpipetován. Poté bylo napipetováno 500 µl promývacího pufru Wash 2/3, centrifugace za stejných podmínek a následovalo opětovné odstranění prošlé tekutiny. Filtrovací kolonka spolu se sběrnou zkumavkou byly ještě jednou zcentrifugovány při 10 000 rpm, RT po dobu 30 s, aby byla z filtru odstraněna veškerá tekutina. Doprostřed filtrovací kolonky bylo připipetováno 6 µl 10x DNase pufru, 4 µl DNasy, 50 µl nuclease-free water a následovala inkubace 30 min při RT. Po inkubaci bylo na filtrovací kolonku napipetováno 700 µl promývacího pufru Wash 1, následovala inkubace po dobu 30-60 s při RT a zkumavka byla zcentrifugována při 10 000 rpm, RT po 30 s. Objem, který prošel přes filtrovací kolonku byl odpipetován. Poté bylo 30 napipetováno dvakrát po sobě 500 µl

promývacího pufru Wash 2/3, inkubace 30-60 s při RT a centrifugace za stejných podmínek. Následovalo odstranění prošlé tekutiny. Nakonec byla zcentrifugována sběrná zkumavka spolu s filtrovací kolonkou po dobu 1 min při 10 000 rpm, RT. Filtrovací kolonka byla přemístěna do nové sběrné zkumavky a do středu filtru bylo připipetováno 60  $\mu$ l elučního roztoku. Vzorek se inkuboval 1 min při RT a poté následovala centrifugace po dobu 1 min, 20 000 rpm při RT. Eluát obsahoval vyizolovanou RNA.

### **Stanovení koncentrace RNA**

Koncentrace RNA byla měřena spektrofotometricky na Nanodropu ND-1000 (NanoDrop Technologies, Wilmington, Delaware, USA). Pro měření byl použit 1  $\mu$ l eluátu s vyizolovanou RNA. Jako reference byl využit 1  $\mu$ l elučního roztoku. Kvalita RNA byla posuzována podle hodnot poměrů absorbance při vybraných vlnových délkách 260/280 a 260/230, přitom poměr 260/280 by se měl optimálně pohybovat v rozmezí 1,8-2,0.

### **Syntéza komplementární DNA (cDNA)**

Komplementární cDNA byla získána v jednokrokové reverzní transkripci (RT) s využitím napoolovaných RT primerů získané smícháním stejných objemů primerů (miR-21, miR-23, miR-126, miR-205 a RNU6B) (Applied Biosystems) a TaqMan® MicroRNA Reverse Transcription Kitu (Applied Biosystems). Nejprve bylo potřeba připravit master mix, a to smícháním komponent – Nuclease-free vody, 10x Reverse Transcription Buffer, 100 mM dNTPs, MultiScribe™ Reverse Transcriptase a RNase Inhibitor (Tab. 9). Příslušný objem všech složek pro odpovídající počet vzorků byl napipetován do zkumavky, krátce zvortexován a rozpipetován po 7  $\mu$ l do PCR mikrozkuavek. Do každé mikrozkuavky bylo přidáno 5  $\mu$ l vzorku RNA o koncentraci 10 ng/ $\mu$ l a 3  $\mu$ l napoolovaných RT primerů. Poté byla mikrozkuavka krátce zvortexována, zcentrifugována a vložena do termocykléru, kde proběhla reverzní transkripce (Tab. 10).

**Tab. 9: Složení mastermixu pro syntézu cDNA**

Chemikálie	Množství pro 1 vzorek [μl]
Nuclease-free voda	4,16
10x Reverse Transcription Buffer	1,50
100 mM dNTPs	0,15
Multiscribe™ Reverse Transcriptase	1,00
RNAse Inhibitor, 20U/μl	0,19
Celkem	7

**Tab. 10: Podmínky reverzní transkripce**

Reverzní transkripce	Teplota [°C]	Čas [min]
1. krok	16	30
2. krok	42	30
3. krok	85	5
Chlazení	4	∞

### **Preamplifikace s využitím napoolovaných miRNA assays**

Na přípravu preamplifikace byl potřeba TaqMan® PreAmp Master Mix, jednotlivé vzorky cDNA a napoolované TaqMan® MicroRNA Assays (Applied Biosystems) (Tab. 11). Ty byly vytvořeny smícháním stejného objemu všech TaqMan® MicroRNA Assays (miR-21, miR-23, miR-126, miR-205 a RNU6B) (Applied Biosystems) dohromady do jedné zkumavky a doplněním objemu na 500 μl TE pufrem. Připravené mikrozkumavky byly krátce zvortexovány, zcentrifugovány vloženy do termocykléru, kde proběhla preamplifikace (Tab. 12).

**Tab. 11: Složení PCR mixu pro preamplifikaci**

Chemikálie	Množství pro 1 vzorek [μl]
TaqMan® PreAmp Master Mix (2x)	12,50
cDNA	6,25
Napoolované TaqMan® MicroRNA Assays	6,25
Celkem	25,00

**Tab. 12: Podmínky preamplifikace**

Preamplifikace	Teplota [°C]	Čas	Počet cyklů
Počáteční denaturace	95	10 min	
Denaturace	95	15 s	14
Annealing + elongace	60	4 min	
Chlazení	4	∞	

**Real-time PCR s preamplifikovanými vzorky**

Jednotlivé komponenty nutné pro provedení real-time PCR byly TaqMan® Universal PCR Master Mix (Applied Biosystems), nuclease-free voda, příslušné TaqMan® MicroRNA Assays (Applied Biosystems) a PCR produkt získaný při preamplifikaci (Tab. 13). Ten bylo potřeba nejprve naředit 1:20 pomocí TE pufru. Připravený PCR mix v mikrozkušavkách byl krátce zvortexován, zcentrifugován a rozpipetován do tří jamek 96 jamkové mikrotitrační destičky po 10 μl. Poté byla 96 jamková mikrotitrační destička utěsněna fólií, zcentrifugována při 1 000 rpm, 2 min při RT a vložena do LightCycleru® 480 Real-Time PCR Systém (Roche, Branford, CT, USA), kde proběhla real-time PCR (Tab. 14).

**Tab. 13: Složení PCR mixu pro real-time PCR**

<b>Chemikálie</b>	<b>Množství pro 1 vzorek - triplikát [μl]</b>
TaqMan® Universal PCR Master Mix (2x)	15
Nuclease-free voda	6
TaqMan® MicroRNA Assay	1,5
PCR produkt z preamplifikace (zředěno 1:20)	7,5
Celkem	30

**Tab. 14: Podmínky real-time PCR**

<b>Real-time PCR</b>	<b>Teplota [°C]</b>	<b>Čas</b>	<b>Počet cyklů</b>
Počáteční denaturace	95	10 min	
Denaturace	95	15 s	40
Annealing + elongace	60	60 s	
Chlazení	4	∞	

## 7. Výsledky

Epigenetics 9:7, 1031–1046; July 2014; © 2014 Landes Bioscience

RESEARCH PAPER

# Key tumor suppressor genes inactivated by “greater promoter” methylation and somatic mutations in head and neck cancer

Rafael Guerrero-Preston<sup>1,2,†,\*</sup>, Christina Michailidi<sup>1,†</sup>, Luigi Marchionni<sup>3</sup>, Curtis R Pickering<sup>4</sup>, Mitchell J Frederick<sup>4</sup>, Jeffrey N Myers<sup>4</sup>, Srinivasan Yegnasubramanian<sup>5</sup>, Tal Hadar<sup>1</sup>, Maartje G Noordhuis<sup>1,6</sup>, Veronika Zizkova<sup>1,7</sup>, Elana Fertig<sup>8</sup>, Nishant Agrawal<sup>1,5</sup>, William Westra<sup>1</sup>, Wayne Koch<sup>1</sup>, Joseph Califano<sup>1,9</sup>, Victor E Velculescu<sup>5</sup>, and David Sidransky<sup>1,\*</sup>

<sup>1</sup>Department of Otolaryngology and Head and Neck Surgery; Johns Hopkins University; School of Medicine; Baltimore, MD USA; <sup>2</sup>Department of Obstetrics and Gynecology; University of Puerto Rico School of Medicine; Rio Piedras, Puerto Rico; <sup>3</sup>Department of Oncology Biostatistics; Johns Hopkins University; School of Medicine; Baltimore, MD USA; <sup>4</sup>Department of Head and Neck Surgery; University of Texas M.D. Anderson Cancer Center; Houston, TX USA; <sup>5</sup>Sidney Kimmel Comprehensive Cancer Center; Johns Hopkins University; Baltimore, MD USA; <sup>6</sup>Department of Otorhinolaryngology-Head and Neck Surgery; University of Groningen; University Medical Center; Groningen, The Netherlands; <sup>7</sup>Laboratory of Molecular Pathology and Institute of Molecular and Translational Medicine; Faculty of Medicine and Dentistry; Palacky University; Olomouc, Czech Republic; <sup>8</sup>Division of Biostatistics and Bioinformatics; Department of Oncology; Sidney Kimmel Comprehensive Cancer Center; Baltimore, MD USA; <sup>9</sup>Milton J. Dance Head and Neck Center; Greater Baltimore Medical Center; Baltimore, MD USA

<sup>†</sup>These authors contributed equally to this work.

Keywords: Head and Neck Squamous Cell Carcinoma, Tumor Suppressor Genes, DNA methylation, somatic mutations, integration analysis

Tumor supresorové geny (TSGs) jsou obvykle inaktivovány somatickými mutacemi a/nebo methyloací promotorů. Doposud však ani „high-throughput“ genomické studie nezjistily příčinu inaktivace TSGs oběma mechanismy. V této práci byla provedena integrovaná molekulární analýza založená na sekvenování methylační vazebné domény (MBD-seq), 450K methylační arraye, sekvenování celého exomu a celogenomová genová expresi v primárním karcinomu dlaždicových buněk hlavy a krku (HNSCC) a shodných vzorcích uvulopalatofaryngoplastické tkáně (UPPP). Bylo odhaleno 186 downregulovaných genů nesoucích metylaci specifického promotoru pro karcinom včetně *PAX1* a *PAX5* a dále bylo identifikováno 10 klíčových tumor supresorových genů (*GABRB3*, *HOXC12*, *PARP15*, *SLCO4C1*, *CDKN2A*, *PAX1*, *PIK3AP1*, *HOXC6*, *PLCB1* a *ZIC4*) inaktivovaných buď methyloací promotoru a/nebo somatickou mutací. Mezi novými objevenými geny s duálním mechanismem inaktivace jsme zjistili vysokou frekvenci genomových a epigenomických změn v transkripčních faktorech genové rodiny PAX, které selektivně ovlivňují kanonické dráhy NOTCH a TP53 související s buněčným osudem, přežíváním buněk a udržování genomu. Naše výsledky zdůrazňují důležitost hodnocení TSGs na genomické

i epigenomické úrovni k odhalení klíčové role u HNCSS deregulované simultánní metylací promotoru a somatickými mutacemi.

## NF- $\kappa$ B and stat3 transcription factor signatures differentiate HPV-positive and HPV-negative head and neck squamous cell carcinoma

Daria A. Gaykalova<sup>1</sup>, Judith B. Manola<sup>2</sup>, Hiroyuki Ozawa<sup>3</sup>, Veronika Zizkova<sup>1,4</sup>, Kathryn Morton<sup>1</sup>, Justin A. Bishop<sup>1,5</sup>, Rajni Sharma<sup>5</sup>, Chi Zhang<sup>1,6</sup>, Christina Michailidi<sup>1</sup>, Michael Considine<sup>7,8</sup>, Marietta Tan<sup>1</sup>, Elana J. Fertig<sup>7,8</sup>, Patrick T. Hennessey<sup>1,11</sup>, Julie Ahn<sup>1</sup>, Wayne M. Koch<sup>1</sup>, William H. Westra<sup>1,5</sup>, Zubair Khan<sup>1</sup>, Christine H. Chung<sup>1,3</sup>, Michael F. Ochs<sup>7,8,9</sup> and Joseph A. Califano<sup>1,10</sup>

<sup>1</sup> Department of Otolaryngology—Head and Neck Surgery, Johns Hopkins Medical Institutions, Baltimore, MD

<sup>2</sup> Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, MA

<sup>3</sup> Department of Oncology, Johns Hopkins Medical Institutions, Baltimore, MD

<sup>4</sup> Laboratory of Molecular Pathology, Institute of Molecular and Translational Medicine, Palacky University, Olomouc, Czech Republic

<sup>5</sup> Department of Pathology, Johns Hopkins Medical Institutions, Baltimore, MD

<sup>6</sup> University of Virginia, Charlottesville, VA

<sup>7</sup> Division of Oncology Biostatistics, Johns Hopkins Medical Institutions, Baltimore, MD

<sup>8</sup> Department of Oncology, Johns Hopkins Medical Institutions, Baltimore, MD

<sup>9</sup> Department of Mathematics and Statistics, The College of New Jersey, Ewing, NJ

<sup>10</sup> Milton J. Dance Head and Neck Center, Greater Baltimore Medical Center, Baltimore, MD

<sup>11</sup> Mid-Michigan Ear, Nose, and Throat, East Lansing, MI

S využitím „high-throughput“ analýz a databáze TRANSFAC<sup>®</sup> byly charakterizovány transkripční faktory (TF) u podskupin nádorů hlavy a krku pomocí interferenčních analýz exprese cílových genů, určující efekt DNA methylace. Tímto způsobem byly odlišeny HPV<sup>+</sup> a HPV<sup>-</sup> dlaždicové karcinomy hlavy a krku (HNSCC) v závislosti na aktivitě vybraných transkripčních faktorů – AP1, STAT, NF- $\kappa$ B a p53. Imunohistochemické barvení následně potvrdilo HPV<sup>-</sup> charakteristiku HNSCC a funkční studie ukazují, že pro tento fenotyp je možno efektivně využívat anti-NF- $\kappa$ B a anti-STAT terapie. Zjištěné informace korelují s již dříve objeveným spojením STAT, NF- $\kappa$ B a AP1 u HNSCC. Identifikovali jsme pět nejlépe hodnocených párových biomarkerů ze STAT, NF- $\kappa$ B a AP1 drah, které umožňují odlišit HPV<sup>+</sup> od HPV<sup>-</sup> HNSCC na základě aktivity TF a zároveň také validovali tyto markery v TCGA (The Cancer Genome Atlas) a na nezávislých validačních kohortách. Na základě získaných výsledků a znalostí se předpokládá, že nový přístup k analýze drah TF by mohl vnést lepší vhled do terapeutického cílení v rámci jednotlivých podskupin pacientů s heterogenním onemocněním, jako je HNSCC.



## The Relationship of Mir-21, Mir-126 and Mir-205 to P-Glycoprotein, Mrp1 and Lrp/Mvp in Non-Small Cell Lung Cancer

Zizkova V<sup>1†</sup>, Skarda J<sup>1,2†</sup>, Janikova M<sup>1,2\*</sup>, Luzna P<sup>3</sup>, Radova L<sup>2</sup>, Kurfurstova D<sup>1</sup> and Kolar Z<sup>1,2</sup>

<sup>1</sup>Department of Clinical and Molecular Pathology and Laboratory of Molecular Pathology, Palacky University Olomouc and University Hospital Olomouc, Czech Republic

<sup>2</sup>Institute of Molecular and Translational Medicine, Palacky University Olomouc and University Hospital Olomouc, Czech Republic

<sup>3</sup>Department of Histology and Embryology, Palacky University Olomouc and University Hospital Olomouc, Czech Republic †Authors' contributions: Veronika Zizkova and Jozef Skarda contributed equally to this work

\*Corresponding author: Janikova M, Department of Clinical and Molecular Pathology and Laboratory of Molecular Pathology, Institute of Molecular and Translational Medicine, Palacky University Olomouc and University Hospital Olomouc, Hnevotinska 3, 775 15 Olomouc, Czech Republic.

Received: May 20, 2015; Accepted: June 15, 2015; Published: July 29, 2015

Proteinové transportéry P-gp, MRP1 a LRP/MVP se podílejí na vzniku multidrogové rezistence (MDR) u nemalobuněčného karcinomu plic (NSCLC). Jejich exprese je posttranskripčně regulována mikroRNA (miRNA). Dysregulace miR-21, miR-126 a miR-205 se často vyskytuje v NSCLC. Cílem této studie bylo zjistit, zda je úroveň miRNA spojena s expresí výše uvedených proteinů zapojených do MDR a zda mohou být použity jako prognostické a diagnostické markery. Analyzovali jsme miR-21, miR-126 a miR-205 v různých histologických podtypech NSCLC. Jejich exprese poté korelovala s klinicko-patologickými charakteristikami, jako je přežití bez progresu (PFS), celkové přežití (OS) a různé histologické podtypy NSCLC a s expresí P-gp, MRP1 a LRP/MVP. Nejistili jsme žádný významný vztah mezi expresí miR-21 a miR-126 a klinicko-patologickými parametry. Hladiny miR-205 však byly významně zvýšeny u spinocelulárních karcinomů ( $p < 10^{-6}$ ) ve srovnání s jinými histologickými podtypy NSCLC. Hladina miR-205 navíc nepřímě korelovala s expresí P-gp u pacientů s NSCLC ( $p = 0,03$ ). Výsledky této studie naznačují, že miR-205 může být použit jako diagnostický marker a jeho downregulace může indikovat vznik rezistence na léky zprostředkované P-gp u pacientů s NSCLC.

## Prognostic significance of miR-23b in combination with P-gp, MRP and LRP/MVP expression in non-small cell lung cancer

M. JANIKOVA<sup>1,2,4</sup>, V. ZIZKOVA<sup>1,4</sup>, J. SKARDA<sup>1,2,\*</sup>, G. KHARAISHVILI<sup>1,2</sup>, L. RADOVA<sup>3</sup>, Z. KOLAR<sup>1,2</sup>

<sup>1</sup>Department of Clinical and Molecular Pathology and Laboratory of Molecular Pathology; <sup>2</sup>Institute of Molecular and Translational Medicine, Faculty of Medicine and Dentistry, Palacky University Olomouc and University Hospital Olomouc; <sup>3</sup>CEITEC MU, Brno, Czech Republic

\*Correspondence: [jojos@email.cz](mailto:jojos@email.cz)

<sup>4</sup>Contributed equally to this work.

Nedávno byla miR-23b označena jako slibným nádorovým biomarkerem, avšak jeho funkce v karcinomu plic nebyla ještě zcela stanovena. Pacienti stále nereagují dobře na dostupnou léčbu, pravděpodobně kvůli expresi proteinů multirezistence (MDR), jako jsou P-gp, MRP a LRP/MVP. Cílem této studie bylo zjistit roli miR-23b u nemalobuněčného karcinomu plic (NSCLC) a jeho vztah k výsledku pacienta v souvislosti s transportními proteiny MDR. Imunohistochemicky jsme vyhodnotili expresi P-gp, MRP a LRP/MVP a kvantifikovali relativní hladiny miR-23b v 62 vzorcích pacientů s NSCLC. Prognostický význam proteinů miR-23b a MDR byl testován pomocí Kaplan-Meierovy a Coxovy regresní analýzy. Naše výsledky ukázaly, že miR-23b je většinou downregulován ve vzorcích NSCLC (57/62) a že jeho upregulace v nádorech souvisí s delším přežitím bez progresu (PFS;  $P=0,065$ ) a celkovým přežitím (OS;  $P=0,048$ ). Coxův model proporcionálního rizika odhalil, že riziko úmrtí nebo relapsu u pacientů s NSCLC s downregulací miR-23b se zvyšuje spolu s expresí LRP/MVP a obě rizika se snižují s upregulací miR-23b (HRPFS=4,342; PPF=0,022; HROS=4,408; POS=0,015). Naše nálezy naznačují, že miR-23b, zejména v kombinaci s expresí LRP/MVP, může sloužit jako vhodný prognostický biomarker pro pacienty s NSCLC.

## Integrative computational analysis of transcriptional and epigenetic alterations implicates *DTX1* as a putative tumor suppressor gene in HNSCC

Daria A. Gaykalova<sup>1</sup>, Veronika Zizkova<sup>1,2</sup>, Theresa Guo<sup>1</sup>, Ilse Tiscareno<sup>1</sup>, Yingying Wei<sup>3,10</sup>, Rajita Vatapalli<sup>1,4</sup>, Patrick T. Hennessey<sup>1,5</sup>, Julie Ahn<sup>1</sup>, Ludmila Danilova<sup>3,11</sup>, Zubair Khan<sup>1</sup>, Justin A. Bishop<sup>1,6</sup>, J. Silvio Gutkind<sup>7</sup>, Wayne M. Koch<sup>1</sup>, William H. Westra<sup>1,6</sup>, Elana J. Fertig<sup>3</sup>, Michael F. Ochs<sup>3,8</sup>, Joseph A. Califano<sup>1,9</sup>

<sup>1</sup>Department of Otolaryngology—Head and Neck Surgery, Johns Hopkins Medical Institutions, Baltimore, Maryland, USA

<sup>2</sup>Institute of Molecular and Translational Medicine, Faculty of Medicine and Dentistry, Palacky University, Olomouc, Czech Republic

<sup>3</sup>Division of Oncology Biostatistics, Department of Oncology, Johns Hopkins Medical Institutions, Baltimore, Maryland, USA

<sup>4</sup>Department of Urology, Northwestern University, Chicago, Illinois, USA

<sup>5</sup>Department of Otolaryngology, Mid-Michigan Ear Nose and Throat, East Lansing, Michigan, USA

<sup>6</sup>Department of Pathology, Johns Hopkins Medical Institutions, Baltimore, Maryland, USA

<sup>7</sup>Department of Pharmacology, UC San Diego Moores Cancer Center, La Jolla, California, USA

<sup>8</sup>Department of Mathematics and Statistics, The College of New Jersey, Ewing, New Jersey, USA

<sup>9</sup>Department of Surgery, UC San Diego, Moores Cancer Center, La Jolla, California, USA

<sup>10</sup>Department of Statistics, The Chinese University of Hong Kong, NT, Shatin, Hong Kong

<sup>11</sup>Laboratory of Systems Biology and Computational Genetics, Vavilov Institute of General Genetics, Russian Academy of Sciences, Moscow, Russia

**Correspondence to:** Joseph A. Califano, email: jcalifano@ucsd.edu  
Daria A. Gaykalova, email: dgaykal1@jhmi.edu

**Keywords:** HNSCC, *DTX1*, expression, methylation, integration

**Received:** September 07, 2016

**Accepted:** January 16, 2017

**Published:** January 27, 2017

Každoročně je celosvětově diagnostikováno více než půl milionu nových případů dlaždicových karcinomů hlavy a krku (HNSCC), avšak celkové přežití pacientů po 5 letech je pouze 50 %. V poslední době se zlepšily a zrychlily „high-throughput“ technologie využívané k charakterizaci HNSCC. Využili jsme statistiku „high-throughput“ dat exprese genů a identifikovali tak 76 nejlépe hodnocených kandidátů se signifikantním rozdílem exprese v nádorech ve srovnání s normálními tkáněmi. Dále jsme také získali 15 epigeneticky regulovaných kandidátů díky zaměření se na podmnožinu genů s negativní korelací mezi genovou expresí a metylací promotoru. Různá exprese a methylace vybraných tří kandidátů (*BANK1*, *BIN2* a *DTX1*) byly potvrzeny v nezávislé kohortě HNSCC získané z nemocnice Johns Hopkins a TCGA (The Cancer Genome Atlas). Dále byla provedeno také hodnocení NOTCH regulátoru *DTX1*, který byl downregulován hypermethylnací promotoru

v nádorech a ukazuje to na skutečnost, že snížená exprese *DTX1* v HNSCC může být asociována s aktivací signální dráhy NOTCH a zvyšovat migrační potenciál.

## 8. Diskuze

Nádory hlavy a krku, stejně jako karcinomy plic, jsou epitelová maligní nádorová onemocnění, která jsou ve všech průmyslově rozvinutých zemích řazeny k nejzávažnějším zdravotním problémům. K léčbě se běžně využívá chirurgická resekce, chemoterapie, či radioterapie, avšak v posledních letech se začaly využívat také mechanismy k přímé regulaci genové exprese.

Právě identifikace nových biomarkerů u pacientů s nemalobuněčným karcinomem plic (NSCLC) je v poslední době prioritou. Dle dostupných statistik se cca 11 % pacientů s NSCLC nedožívá déle než 5 let od stanovení diagnózy (D'Addario *et al.*, 2010). Důvodem může být skutečnost, že většina pacientů bývá diagnostikována až v pokročilém stadiu onemocnění NSCLC (hlavně III a IV), kdy už nádor získává svoji heterogenitu a dochází k rozvoji mnohočetné lékové rezistence (MDR). Do dnešní doby nebyly identifikovány žádné biomarkery, které by jednoznačně určovaly nástup aktivace MDR mechanismů a cíleně tak optimalizovaly léčbu pro každého pacienta.

Jedním z cílů mé disertační práce bylo zjistit expresi vybraných miRNA (miR-21, miR-23b, miR-126 a miR-205) u jednotlivých vzorků pacientů s NSCLC a tuto expresi dále korelovat s klinicko-patologickými charakteristikami, jako je přežití bez progresu (PFS), celkové přežití (OS) v různých histologických podtypech NSCLC a s expresí P-gp, MRP1 a LRP/MVP. Nebyl zjištěn žádný významný vztah mezi expresí miR-21 a miR-126 a klinicko-patologickými parametry. Tato skutečnost rozporuje publikacím, ve kterých exprese miR-21 a miR-126 potvrdila možnost prognostického významu v souvislosti s klinicko-patologickými charakteristikami u pacientů s NSCLC (Trussardi-Regnier *et al.*, 2007; Oue *et al.*, 2009; Markou *et al.*, 2008; Gao *et al.*, 2011). Důvodem může být málopočetný soubor pacientů použitý v rámci našeho experimentu.

Existují různé práce, které se zabývaly souvislostí exprese miR-21 a MDR v buněčných liniích, ale pouze pár, které studují vliv exprese miR-21 přímo u pacientů s NSCLC. Ve většině prací vykazuje miR-21 pro pacienty s NSCLC charakteristiku negativního prognostického markeru. (Markou *et al.*, 2008; Gao *et al.*, 2011; Gao *et al.*, 2012; Wang *et al.*, 2012).

Expresí miR-23b byla ve většině námi použitých vzorků NSCLC downregulována (57/62), stejně tak downregulace miR-23b byla potvrzena v karcinomech prostaty, karcinomu močového měchýře a karcinomu štítné žlázy (Majid *et al.*, 2012; Majid *et al.*, 2013; Dettmer *et al.*, 2014). Zvýšená hladina exprese miR-23b v karcinomech je v některých publikacích brána jako pozitivní faktor pro přežití, v jiných publikacích je naopak spojována s horším přežitím pacienta (Jin *et al.*, 2013; Chen *et al.*, 2012; Zaman *et al.*, 2012). Předpokládá se, že miR-23b může mít onkogenní charakter, anebo tumor supresorový charakter v závislosti na typu rakovinného onemocnění. Coxův model proporcionálního rizika v našem souboru pacientů s NSCLC odhalil, že riziko úmrtí nebo relapsu u pacientů s NSCLC s downregulací miR-23b se zvyšuje spolu s expresí LRP/MVP a obě rizika se snižují s upregulací miR-23b (HRPFS = 4,342, PPF5 = 0,022; HROS = 4,408, POS = 0,015). Naše nálezy naznačují, že miR-23b, zejména v kombinaci s expresí LRP/MVP, může sloužit jako vhodný prognostický biomarker pro pacienty s NSCLC.

V rámci mé disertační práce bylo zjištěno, že miR-126 byla downregulována u pacientů s NSCLC, což je v souladu s jinými články zabývající se touto tematikou (Crowford *et al.*, 2008; Sun *et al.*, 2010).

Hladiny miR-205 byly významně zvýšeny u spinocelulárních karcinomů ( $p < 10^{-6}$ ) ve srovnání s jinými histologickými podtypy NSCLC. Hladina miR-205 navíc nepřímo korelovala s expresí P-gp u pacientů s NSCLC ( $p=0,03$ ). Ze získaných výsledků se dá uvažovat, že miR-205 může být u pacientů s NSCLC použita jako diagnostický marker a její downregulace může indikovat vznik rezistence na léky zprostředkované P-gp stejně jako bylo potvrzeno v jiných odborných pracích (Lebanony *et al.*, 2009; Bishop *et al.*, 2010; Zhang *et al.*, 2012; Miranda *et al.*, 2006).

V rámci experimentální části mé disertační práce jsem se zabývala také studiem epigenetických mechanismů zapojených do nádorové transformace dlaždicového karcinomu hlavy a krku (HNSCC). Byla provedena rozsáhlá genomická a epigenomická analýza, která byla zaměřena u pacientů s HNSCC na identifikaci tumor supresorových genů inaktivovaných methylocí promotorů a/nebo somatickými mutacemi a porovnání s tumor supresorovými geny s mutací u HNSCC. Během této analýzy bylo objeveno 10 downregulovaných genů jak somatickou mutací, tak methylocí protomoru. Právě tyto geny by mohly být klíčovými hráči v problematice onkogeneze. Ze získaných výsledků je zřejmé,

že *PAX1*, *ZIC4*, *PLCB1* a *PAX5* dosahovaly největších rozdílů na úrovni methylace DNA u HNSCC, ve srovnání s kontrolními vzorky. *PAX1* a *PAX5* byly vybrány jako klíčové geny, patřící do stejné rodiny transkripčních faktorů, které byly nejvíce methylovány a downregulovány v námi vybrané kohortě vzorků HNSCC.

HPV<sup>+</sup> status představuje po několik dekád příhodnější prognózu pro pacienty s HNSCC, oproti HPV<sup>-</sup> (Weinberger *et al.*, 2006). Jednou z priorit právě u pacientů s HNSCC je objevení nových biomarkerů, které by sloužily k brzké detekci HPV<sup>-</sup> pacientů. Hlavními cíli detekce biomarkerů u HNSCC jsou transkripční faktory, které jsou přímo zapojené do epigenetických změn a mechanismů vedoucí k nádorové transformaci. V rámci mé disertační práce došlo s využitím „high-throughput“ analýz a databáze TRANSFAC<sup>®</sup> k objevení skutečnosti, že u HPV<sup>-</sup> pacientů HNSCC dochází ke koordinované dysregulaci několika zásadních signálních drah jako je NF- $\kappa$ B a STAT. Naše výsledky tak souhlasí s výsledky v jiných publikacích při použití cDNA microarray analýz HNSCC buněčných linií (Yan *et al.*, 2007; Yan *et al.*, 2008). Na základě získaných výsledků a znalostí se předpokládá, že nový přístup k analýze drah TF by mohl vnést lepší vhled do terapeutického cílení v rámci jednotlivých podskupin pacientů s heterogenním onemocněním, jako je HNSCC.

## 9. Závěr

Nádory hlavy a krku, stejně jako karcinomy plic, jsou epitelová maligní nádorová onemocnění, která jsou ve všech průmyslově rozvinutých zemích řazeny k nejzávažnějším zdravotním problémům. K léčbě se běžně využívá chirurgická resekce, chemoterapie, či radioterapie, avšak v posledních letech se začaly využívat také mechanismy k přímé regulaci genové exprese.

V experimentální části mé disertační práce jsem se zaměřila na studium vybraných epigenetických mechanismů zahrnutých do nádorové transformace dlaždicového karcinomu hlavy a krku (HNSCC) a nemalobuněčného karcinomu plic (NSCLC).

U pacientů s NSCLC jsem se zaměřila na stanovení exprese vybraných miRNA (miR-21, miR-23b, miR-126, miR-205) a korelaci získaných výsledky s klinicko-patologickými charakteristikami, jako je přežití bez progresu (PFS), celkové přežití (OS) v různých histologických podtypech NSCLC a s expresí P-gp, MRP1 a LRP/MVP. Nebyl zjištěn žádný významný vztah mezi expresí miR-21 a miR-126 a klinicko-patologickými parametry. Hladiny miR-205 však byly významně zvýšeny u spinocelulárních karcinomů ( $p < 10^{-6}$ ) ve srovnání s jinými histologickými podtypy NSCLC. Hladina miR-205 navíc nepřímo korelovala s expresí P-gp u pacientů s NSCLC ( $p = 0,03$ ). Výsledky této studie naznačují, že miR-205 může být použit jako diagnostický marker a jeho downregulace může indikovat vznik rezistence na léky zprostředkované P-gp u pacientů s NSCLC. Také bylo prokázáno, že exprese miR-23b je ve většině námi použitých vzorků NSCLC downregulována (57/62) a up-regulace miR-23b v nádorech souvisí s delším přežitím bez progresu (PFS;  $P = 0,065$ ) a celkovým přežitím (OS;  $P = 0,048$ ). Coxův model proporcionálního rizika odhalil, že riziko úmrtí nebo relapsu u pacientů s NSCLC s downregulací miR-23b se zvyšuje spolu s expresí LRP/MVP a obě rizika se snižují s upregulací miR-23b (HRPFS=4,342; PPF=0,022; HROS=4,408; POS =0,015). Naše nálezy naznačují, že miR-23b, zejména v kombinaci s expresí LRP/MVP, může sloužit jako vhodný prognostický biomarker pro pacienty s NSCLC.

U HNSCC jsem porovnávala epigenetické profily nádorové tkáně a shodných vzorků uvulopalatofaryngoplastické tkáně a snažila se vytipovat transkripční faktory (TF) vhodné k determinaci HPV<sup>+</sup> a HPV<sup>-</sup> HNSCC. V rámci mé disertační práce bylo odhaleno 186



downregulovaných genů nesoucích methylaci pro karcinom specifického promotoru včetně *PAX1* a *PAX5* a dále bylo identifikováno 10 klíčových tumor supresorových genů (*GABRB3*, *HOXC12*, *PARP15*, *SLCO4C1*, *CDKN2A*, *PAX1*, *PIK3AP1*, *HOXC6*, *PLCB1* a *ZIC4*) inaktivovaných buď methylací promotoru a/nebo somatickou mutací. S využitím „high-throughput“ analýz a databáze TRANSFAC® byly charakterizovány TF (AP1, STAT, NF- $\kappa$ B a p53) vhodné k determinaci HPV<sup>+</sup> a HPV<sup>-</sup> HNSCC. Na základě získaných výsledků a znalostí se předpokládá, že nový přístup k analýze drah TF by mohl vnést lepší vhled do terapeutického cílení v rámci jednotlivých podskupin pacientů s heterogenním onemocněním, jako je HNSCC.

## 10. Literatura

- **Aapola U., Shibuya K., Scott H.S., Ollila J., Vihinen M., Heino M., Shintani A., Kawasaki K., Minoshima S., Krohn K., Antonarakis S.E., Shimizu N., Kudoh J., Peterson P.** (2000): Isolation and initial characterization of a novel Zinc finger gene, DNMT3L, on 21q22.3, Related to the Cytosine-5- Methyltransferase 3 Gene Family. *Genomics*, 65, 293-298
- **Adam Z., Vorlíček J., Vaníček J. a kol.** (2004) Diagnostické a léčebné postupy u maligních chorob. 2., aktualiz. a dopl. vyd. Praha: *Grada Publishing*, 684 s. ISBN 80-247-0896-5
- **Adams R.L.P. et Burdon R.H.** (1985): *Molecular Biology of DNA Methylation*. Springer-Verlag, New York
- **Albu R.F., Jurkowski T.P., Jeltsch A.** (2012): The *Caulobacter crescentus* DNA-(adenine-N6)-methyltransferase CcrM methylates DNA in a distributive manner. *Nucleic Acids Res* 40, 1708-1716
- **Almquist D., Mosalpuria K., Ganti A.K.** (2016): Multimodality therapy for limited-stage small-cell lung cancer. *Journal of Clinical Oncology* 12: 111–117
- **Al-Shibli K.I., Donnem T., Al-Saad S., Persson M., Bremnes R.M., Busund L.T.** (2008): Prognostic effect of epithelial and stromal lymphocyte infiltration in non-small cell lung cancer. *Clinical Cancer Research* 14: 5220–5227
- **Ansari J., Shackelford R.E., El-Osta H.** (2016): Epigenetics in non-small cell lung cancer: from basics to therapeutics. *Translational Lung Cancer Research* 5: 155-171
- **Aoki A., Suatake I., Miyagawa J., Fujio T., Sasaki H., Tajima S.** (2001): Enzymatic properties of de novo-type mouse DNA (cytosine-5) methyltransferases. *Nucleic Acids Research*, 29, 3506-3512
- **Appasani K., Fire A., Nirenberg M.** (2005): *RNA Interference Technology from basic science to drug development*. Cambridge University Press, New York, 561 s
- **Attwood J.T., Yung R.L., Richardson B.C.** (2002): DNA methylation and the regulation of gene transcription. *Cellular and Molecular Life Science* 59, 241-257
- **Becker H.D., Hohenberger W., Junginger T., Schlag P.M.** (2005): *Chirurgická onkologie*. Grada publishing, Praha, 880 s

- **Berndsen C.E., Albaugh B.N., Tan S., Denu J.M.** (2007): Catalytic mechanism of a MYST family histone acetyltransferase. *Biochemistry* 46, 623-629
- **Bestor T.H.** (2000): The DNA methyltransferases of mammals. *Human Molecular Genetics* 9, 2395-2402
- **Bishop J.A., Benjamin H., Cholakh H., Chajut A., Clark D.P., Westra W.H.** (2010): Accurate classification of non-small cell lung carcinoma using a novel microRNA-based approach. *Clinical Cancer Research* 16: 610-619
- **Blevins T., Rajeswaran R., Shivaprasad P.V., Beknazariants D., Si-Ammour A., Park H.S., Vazquez F., Robertson D., Meins Jr F., Hohn T., Pooggin M.M.** (2006): Four plant Dicers mediate viral small RNA biogenesis and DNA virus induces silencing. *Nucleic Acids Research* 21: 6233-6246
- **Blower P.E., Chung J.H., Verducci J.S., Lin S., Park J.K., Dai Z., Liu C.G., Schmittgen T.D., Reinhold W.C., Croce C.M., Weinstein J.N., Sadee W.** (2008): MicroRNAs modulate the chemosensitivity of tumor cells. *Molecular Cancer Therapeutics* 7: 1-9
- **Bonner J.A., Harari P.M., Giralt J., Azarnia N., Shin D.M., Cohen R.B., Jones Ch.U., Sur R., Raben D., Jassem J., Ove R., Kies M.S., Baselga J., Youssoufian H., Amellal N., Rowinsky E.K., Ang K.K.** (2006): Radiotherapy plus cetuximab for squamous-cell carcinoma of the head and neck. *The New England Journal of Medicine* 354: 567-578
- **Bohmert K., Camus I., Bellini C., Bouchej D., Caboche M., Benning C.** (1998): AGO1 defines a novel locus of Arabidopsis controlling leaf development. *EMBO Journal* 17: 170-180
- **Borchert G.M., Langer W., Davidson B.L.** (2006): RNA polymerase III transcribes human microRNAs. *Nature Structural & Molecular Biology* 13: 1097-1101
- **Borodovsky A., Salmasi V., Turcan S., Fabius A.W., Baia G.S., Eberhart C.G., Weingart J.D., Gallia G.L., Baylin S.B., Chan T.A., Riggins G.J.** (2013): 5-azacytidine reduces methylation, promotes differentiation and induces tumor regression in a patient-derived IDH1 mutant glioma xenograft. *Oncotarget* 4, 1737-1747

- **Braunstein M., Rose A.B., Holmes S.G., Allis C.D. Broach J.R.** (1993): Transcriptional silencing in yeast is associated with reduced nucleosome acetylation. *Genes & Development* 7, 592-604
- **Brownell J.E. et Allis C.D.** (1996): Special HATs for special occasions: linking histone acetylation to chromatin assembly and gene activation. *Current Opinion in Genetics & Development* 6, 176–184
- **Buck-Koehntop B.A. et Defossez P.A.** (2013): On how mammalian transcription factors recognize methylated DNA. *Epigenetics* 8, 131-137
- **Calin G.A., Sevignani C., Dumitru C.D., Hyslop T., Noch E., Yendamuri S., Shimizu M., Rattan S., Bullrich F., Negrini M., Croce C.M.** (2004): Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. *Proceedings of the National Academy of Sciences of the United States of America* 101: 2999-3004
- **Carthew R.W. and Sontheimer E.J.** (2009): Origins and mechanisms of miRNAs and siRNAs. *Cell* 136: 642-655
- **Cerutti L., Mian N., Bateman, A.** (2000): Domains in gene silencing and cell differentiation proteins: the novel PAZ domain and redefinition of the PIWI domain. *Trends in Biochemical Science* 25: 481-482
- **Corso Ch.D., Rutter Ch.E., Park H.S., Lester-Coll, N.H., Kim A.W., Wilson L.D., Husain, Z.A., Lilenbaum R.C., Yu, J.B., Decker, R.H.** (2015) Role of chemoradiotherapy in elderly patients with limited-stage small-cell lung cancer. *Journal of Clinical Oncology* 33: 4240–4246
- **Crawford M., Brawner E., Batte K., Yu L., Hunter M.G., Otterson G.A., Nuovo G., Marsh C.B., nana-Sinkam S.P.** (2008): MicroRNA-126 inhibits invasion in non-small cell lung carcinoma cell lines. *Biochemical and Biophysical Research Communications* 373: 607-612
- **Čelakovský P., Betka J., Plzák J., Chrobok V. a kol.** (2012): *Krční metastázy*. Havlíčkův Brod: *Tobiáš*, 317 s. Medicína hlavy a krku. ISBN 978-80-7311-131-1
- **Češka, R., et al.** (2010): *Interna: Cytologie a obecná histologie*. 1. vydání. Praha: *Triton*, ISBN 978-80-7387-423-0

- **D'Addario G., Fruh M., Reck M., Baumann P., Klepetko W., Felip E. (2010):** ESMO Guidelines Working Group. Metastatic non-small-cell lung cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Annals of Oncology* 21:116-119.
- **Denli A.M., Tops B.B.J., Plasterk R.H.A., Ketting R.F., Hannon G.J. (2004):** Processing of primary microRNAs by the Microprocessor complex. *Nature* 432: 231-235
- **Dettmer M.S., Perren A., Moch H., Komminoth P., Nikiforov Y.E., Nikiforova M.N. (2014):** MicroRNA profile of poorly differentiated thyroid carcinomas: new diagnostic and prognostic insights. *Journal of Molecular Endocrinology* 52: 181–189
- **Drinnenberg I.A., Weinberg D.E., Xie K.T., Mower J.P., Wolfe K.H., Fink G.R., Bartel D.P. (2009):** RNAi in budding yeast. *Science* 326: 544-550
- **Dvořák K., Lukáš Z., Fabián P., Doubek M., Dvořáková D., Adam Z. (2004):** Patomorfologie maligních chorob. In: Adam Z., Vorlíček J, editors. *Obecná onkologie*. Brno: *Masarykova Univerzita*, 2004; 50-71
- **Eden, S. et Cedar, H. (1995):** Action at a distance. *Nature* 375, 16-17
- **Eis, P. S., Tam, W., Sun, L., Chadburn, A., Li, Z., Gomez, M.F., Lund, E., Dahlberg, J.E. (2005):** Accumulation of miR-155 and BIC RNA in human B cell lymphomas. *Proceedings of the National Academy of Sciences of United States of America* 102: 3627-3632
- **Elbashir S.M., Martinez J., Patkaniowska A., Lendeckel W., Tuschl T. (2001):** Functional anatomy of siRNAs for mediating efficient RNAi in *Drosophila melanogaster* embryo lysate. *EMBO Journal* 20: 6877–6888
- **Esquela-Kerscher, A. et Slack, F.J. (2006):** Oncomirs – microRNAs with a role in cancer. *Nature Reviews Cancer* 6: 259-269
- **Feinbaum R. et Ambros V. (1999):** The timing of lin-4 RNA accumulation controls the timing of postembryonic developmental events in *Caenorhabditis elegans*. *Developmental Biology* 210: 87-95
- **Ferlay J., Shin H.R., Bray F., Forman D., Mathers C., Parkin D.M. (2010)** Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *International Journal of Cancer* 127: 2893–2917

- **Fong L. et Small E.J.** (2008): Anti-cytotoxic T-lymphocyte antigen-4 antibody: the first in an emerging class of immunomodulatory antibodies for cancer treatment. *Journal of Clinical Oncology* 26: 5275–5283
- **Fu Y., Luo G.Z., Chen K., Deng X., Yu M., Han D., Hao Z., Liu J., Lu X., Dore L.C., Weng X., Ji Q., Mets L., He C.** (2015): N6-Methyldeoxyadenosine Marks Active Transcription Start Sites in *Chlamydomonas*. *Cell* 16, 879–892
- **Gao W., Shen H., Liu L., Xu J., Xu J., Shu Y.** (2011): MiR-21 overexpression in human primary squamous cell lung carcinoma is associated with poor patient prognosis. *Journal of Cancer Research and Clinical Oncology* 137: 557-566
- **Gao W., Lu X., Liu L., Xu J., Feng D., Shu Y.** (2012): MiRNA-21: a biomarker predictive for platinum-based adjuvant chemotherapy response in patients with non-small cell lung cancer. *Cancer Biology Therapy* 13: 330-340
- **Garassino M.C., Rizvi N., Besse B., Jänne, P., Christoph D., Peters, S., Toh Ch.K., Kurata, T., Costa E.C., Koczywas M., Felip E., Chaft J., Qiu J., Kowanetz M., Coleman, S., Mocchi, S., Sandler, A., Gettinger, S., Johnson M., Iess, S.** (2016): Atezolizumab as 1L therapy for advanced NSCLC in PD-L1-selected patients: updated ORR, PFS and OS data from the BIRCH study. *Journal of Thoracic Oncology* 12: 1 Supplement
- **Ginsberg R.J.** (2002): Lung Cancer. *BC Decker Hamilton*, London, 175 s
- **Greer E.L., Blanco M.A., Gu L., Sendinc E., Liu J., Aristizábal-Corrales D., Hsu C.H., Aravind L., He C., Shi Y.** (2015): DNA Methylation on N6-Adenine in *C. elegans*. *Cell* 161, 868–878
- **Gregory R.I., Yan K.P., Amuthan G., Chendrimada T., Doratotaj B., Cooch N., Shiekhattar R.** (2004): The Microprocessor complex mediates the genesis of microRNAs. *Nature* 432: 235-240
- **Goldstraw P., Ball D., Jett J.R., Le Chevalier T., Lim E., Nicholson A.G., Shepherd F.A.** (2011) Non-small-cell lung cancer. *Lancet* 378: 1727-1740
- **Gwizdek C., Ossareh-Nazari B., Brownawell A.M., Doglio A., Bertrand E., Macara I.G., Dargemont C.** (2003): Exportin-5 mediates nuclear export of minihelix-containing RNAs. *The Journal of Biological Chemistry* 278: 5505-5508

- **Haase A.D., Jaskiewitz I., Zhang H., Laine S., Sack R., Gatignol A., Filipowicz W.** (2005): TRBP, a regulator of cellular PKR and HIV-1 virus expression, interacts with Dicer and functions in RNA silencing. *EMBO Reports* 6: 961-967
- **Han J., Lee Y., Yeom K.H., Kim Y.K., Jin H., Kim V.N.** (2004): The Drosha-DGCR8 complex in primary microRNA processing. *Genes & Development* 18: 3016-3027
- **Hanahan D. et Weinberg R. A.** (2000): The hallmarks of cancer. *Cell* 100: 57-70
- **Hanahan D. et Weinberg R. A.** (2011): Hallmarks of cancer: the next generation. *Cell* 144:646-674
- **Hanna N.H.** (2013): LUME-Lung 2: a multicentre, randomized, double-blind, phase III study of nintedanib plus pemetrexed vs. placebo plus pemetrexed in patients with advanced non-squamous non-small cell lung cancer (NSCLC) after failure of first-line chemotherapy. *Journal of Clinical Oncology* 31: 18 Supplement
- **He L., Thomson J.M., Hemann M.T., Hernando-Monge E., Mu D., Goodson S., Powers S., Cordon-Cardo C, Lowe S.W., Hannon G.J., Hammond S.M.** (2005): A microRNA polycistron as a potential human oncogene. *Nature* 435(7043): 828-833
- **Hotchkiss R.D.** (1948): The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography. *Biological Chemistry* 168: 315-332
- **Hong L., Schroth G.P., Matthews H.R., Yau P., Bradbury E.M.** (1993): Studies of the DNA binding properties of histone H4 amino terminus. Thermal denaturation studies reveal that acetylation markedly reduces the binding constant of the H4 “tail” to DNA. *The Journal of Biological Chemistry* 268, 305-314
- **Chen, C.Z.** (2005): MicroRNAs as oncogenes and tumor suppressors. *The New England Journal of Medicine* 353: 1768-1771
- **Chen L., Han L., Zhang K., Shi Z., Zhang J., Zhang A., Wang Y., Song Y., Li Y., Jiang T., Pu P., Jiang Ch. Kang Ch.** (2012): VHL regulates the effects of miR-23b on glioma survival and invasion via suppression of HIF-1 $\alpha$ /VEGF and  $\beta$ -catenin/Tcf-4 signaling. *Neuro Oncology* 14: 1026-1036
- **Chen Z.X. et Riggs A.D.** (2009): DNA methylation and demethylation in mammals. *The Journal of Biological Chemistry* 286: 18347-18353

- **Chendrimada T.P., Gregory R.I., Kumaraswamy E., Norman J., Cooch N., Nishikura K., Shiekhattar R.** (2005): TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. *Nature* 436: 740-744
- **Cho, W.C.S.** (2011): MicroRNAs in cancer translation research. *Springer*, New York, 571
- **Jeltsch A., Nellen W., Lyko F.** (2006): Two substrates are better than one: dual specificities for Dnmt2 methyltransferases. *Trends in Biochemical Sciences* 31, 306-308
- **Jin L., Wessely O., Marccuson E.G., Ivan C., Calin G.A., Alahari S.K.** (2013): Prooncogenic factors miR-23b and miR-27b are regulated by Her2/Neu, EGF, and TNF- $\alpha$  in breast cancer. *Cancer Research* 73: 2884–2896
- **Johnson, S.M., Grosshans, H., Shingara, J., Byrom, M., Jarvis, R., Cheng, A., Labourier, E., Reinert, K.L., Brown, D., Slack, F.J.** (2005): RAS is regulated by the let-7 microRNA family. *Cell* 120: 635-47
- **Kelly R.J., Gulley J.L., Giaccone G.** (2010): Targeting the immune system in non-small cell lung cancer: bridging the gap between promising concept and therapeutic reality. *Clinical Lung Cancer* 11: 228–237
- **Kent O.A. et Mendell J.T.** (2006): A small piece in the cancer puzzle: microRNA as tumor suppressors and oncogenes. *Oncogene* 25: 6188-6196
- **Kim J.K., Samaranayake M., Pradhan S.** (2009): Epigenetic Mechanisms in Mammals. *Cellular and Molecular Life Science* 66, 596-612
- **Klein J.** (2006): Chirurgie karcinomu plic. *Grada publishing*, Praha, 220 s
- **Kluiver, J., Poppema, S., de Jong, D., Blokzijl, T., Harms, G., Jacobs, S., Kroesen, B.J., van den Berg, A.** (2005): BIC and miR-155 are highly expressed in Hodgkin, primary mediastinal and diffuse large B cell lymphomas. *The Journal of Pathology* 207: 243-249
- **Kluiver, J., Haralambieva, E., de Jong, D., Blokzijl, T., Jacobs, S., Kroesen, B.J., Poppema, S., van den Berg, A.** (2006): Lack of BIC and microRNA miR-155 expression in primary cases of Burkitt lymphoma. *Genes, Chromosomes and Cancer* 45: 147-153



- **Kumar, M.S., Lu, J., Mercer, K.L., Golub, T.R., Jacks, T.** (2007): Impaired microRNA processing enhances cellular transformation and tumorigenesis. *Nature Genetics* 39: 673-677
- **Kuo M.H. et Allis C.D.** (1998): Roles of histone acetyltransferases and deacetylases in gene regulation. *Bioessays* 20, 615-626
- **Kuo M.H., Brownell J.E., Sobel R.E., Ranalli T.A., Cook R.G., Edmondson D.G., Roth S.Y., Allis C.D.** (1996): Transcription-linked acetylation by Gcn5p of histones H3 and H4 at specific lysines. *Nature* 383, 269-272
- **Kolek, V.** (2010): Karcinom plic - současná léčebná strategie z pohledu pneumoonkologa. *Postgraduální medicína* 3:338-352
- **Kouzarides T.** (2007): SnapShot: Histone-modifying enzymes. *Cell* 131, 822
- **Lau N.C., Lim L.P., Weinstein E.G., Bartel D.P.** (2001): An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science* 294: 858-862
- **Landthaler M., Yalcin A., Tuschl T.** (2004): The human DiGeorge syndrome critical region gene 8 and its *D. melanogaster* homolog are required for miRNA biogenesis. *Current Biology* 14: 2162-2167
- **Lagos-Quintana M., Rauhut R., Lendeckel W., Tuschl T.** (2001): Identification of novel genes coding for small expressed RNAs. *Science* 294: 853-858
- **Le Chevalier T.** (2011): Non-small cell lung cancer: the challenges of the next decade. *Frontiers in Oncology* 1: 1-4
- **Lebanony D., Benjamin H., Gilad S., Ezagouri M., Dov A., Ashkenazi K., Gefen N., Izraeli S., Rechavi G., Pass H., Nonaka D., Li J., Spector Y., Rosenfeld N., Chajut A., Cohen D., Aharonov R., Mansukhani M.** (2009): Diagnostic assay based on hsa-miR-205 expression distinguishes squamous from nonsquamous non-small-cell lung carcinoma. *Journal of Clinical Oncology* 27: 2030-2037
- **Lee B. et Muller M.T.** (2009): SUMOylation enhances DNA methyltransferase 1 activity. *Biochemical Journal* 421, 449-461
- **Lee R.C., Feinbaum R.L., Ambros V.** (1993): The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* 75: 843-854

- **Lee Y., Ahn C., Han J., Choi H., Kim J., Yim J., Lee J., Provost P., Radmark O., Kim S., Kim V.N.** (2003): The nuclear Rnase III Drosha initiates microRNA processing. *Nature* 425: 415-419
- **Lee Y., Hur I., Park S.Y., Kim Y.K., Suh M.R., Kim V.N.** (2006): The role of PACT in the RNA silencing pathway. *EMBO Journal* 25: 522-532
- **Lee Y., Jeon K., Lee J.T., Kim S., Kim V.N.** (2002): MicroRNA maturation: stepwise processing and subcellular localization. *EMBO Journal* 21: 4663–4670
- **Lee Y.S., Nakahara K., Pham J.W., Kim K., He Z., Sontheimer E.J., Carthew R.W.** (2004): Distinct roles for Drosophila Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. *Cell* 117: 69-81
- **Lewis B.P., Shih I.H., Jones-Rhoades M.W., Bartel D.P., Burge C.B.** (2003): Prediction of mammalian microRNA targets. *Cell* 115: 787-798
- **Li E., Bestor T.H., Jaenisch, R.** (1992): Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* 69, 915–926
- **Li W., Thompson C.H., O'Brien C.J., McNeil E.B., Scolyer R.A., Cossart Y.E., Veness M.J., Murray Walker D., Morgan G.J., Rose B.R.** (2003): Human papillomavirus positivity predicts favourable outcome for squamous carcinoma of the tonsil. *International Journal of Cancer* 106: 553–558
- **Licitra L., Perrone F.D., Bossi P., Suardi S., Mariani L., Artusi R., Oggionni M., Rossini Ch., Cantù G., Squadrelli M., Quattrone P., Locati L.D., Bergamini C., Olmi P., Pirotti M.A., Pilotti S.** (2006): High-risk human papillomavirus affects prognosis in patients with surgically treated oropharyngeal squamous cell carcinoma. *Journal of Clinical Oncology* 24: 5630–5636
- **Lindel K., Beer K.T., Laissue J., Greiner R.H., Aebersold D.M.** (2001): Human papillomavirus positive squamous cell carcinoma of the oropharynx: a radiosensitive subgroup of head and neck carcinoma. *Cancer* 921: 805–813
- **Liu J., Carmell M.A., Rivas F.V., Mardse C.G., Thomson J.M., Song J.J., Hammond S.M., Joshua-Tor L., Hannon G.J.** (2004): Argonaute2 is the catalytic engine of mammalian RNAi. *Science* 305: 1437-1441

- **Lo Nigro, Denaro N., Merlotti A., Merlano M. (2017):** Head and neck cancer: improving outcomes with a multidisciplinary approach. *Cancer Management and Research* 9: 363-371
- **Lochmanová, J. a Bartoš, M. (2008):** Interferující ribonukleové kyseliny a molekulární patofyziologie vybraných onemocnění. *Časopis lékařů českých* 147: 607-615
- **Lu, J., Getz, G., Miska, E.A., Alvarez-Saavedra, E., Lamb, J., Peck, D., Sweet-Cordero, A., Ebert, B.L., Mak, R.H., Ferrando, A.A., Downing, J.R., Jacks, T., Horvitz, H.R., Golub, T.R. (2005):** MicroRNA expression profiles classify human cancers. *Nature* 435: 834-838
- **MacRea I.J., Zhou K., Li F., Repic A., Brooks A.N., Cande W.Z., Adams P.D., Doudna J.A. (2006):** Structural basis for double-stranded RNA processing by Dicer. *Science* 311: 195-198
- **Majid S., Dar A.A., Saini S., Arora S., Shahryari V., Zaman M.S., Chang I., Yamamura Y., Tanaka Y., Deng G., Dahiya R. (2012):** miR-23b represses proto-oncogene Src kinase and functions as methylation-silenced tumor suppressor with diagnostic and prognostic significance in prostate cancer. *Cancer Research* 72: 6435-6446
- **Majid S., Dar A.A., Saini S., Deng G., Chang I., Greene K., Tanaka Y., Dahiya R., Yamamura S. (2013):** MicroRNA-23b functions as a tumor suppressor by regulating Zeb1 in bladder cancer. *PLoS One* 8: e67686
- **Markou A., Tsaroucha E.G., Kaklamanis L., Fotinou M., Georgoulas V., Lianidou E.S. (2008):** Prognostic value of mature microRNA-21 and microRNA-205 overexpression in non-small cell lung cancer by quantitative real-time RT-PCR. *Clinical Chemistry* 54: 1696-1704
- **Mellin H., Dahlgren L., Munck-Wikland E., Lindholm J., Rabbani H., Kalantari M., Dalianis T. (2002):** Human papillomavirus type 16 is episomal and high viral load may be correlated to better prognosis in tonsillar cancer. *International Journal of Cancer* 102: 152–158

- **Miranda K.C., Huynh T., Tay Y., Ang Y.S., Tam W.L., Thomson M., Lim B., Rigoutsos I.** (2006): A pattern-based method for the identification of MicroRNA binding sites and their corresponding heteroduplexes. *Cell* 126: 1203-1217
- **Moss T., Stephens R.M., Crane-Robinson C., Bradbury E.M.** (1977): A nucleosome-like structure containing DNA and the arginine-rich histones H3 and H4. *Nucleic Acids Research* 4, 2477-2485
- **Mourelatos Z., Dostie J., Paushkin S., Sharma A., Charroux B., Abel L., Rappsilber J., Mann M., Dreyfuss G.** (2002): MiRNPs: a novel class of ribonucleoproteins containing numerous microRNAs. *Genes & Development* 16: 720-728
- **Muñoz N., Castellsagué X., Berringtonde González A., Gissmann L.** (2006): Chapter 1: HPV in the etiology of human cancer. *Vaccine* 24: S1-S10
- **Niho S., Kubota K., Yoh K., Goto K., Ohmatsu H., Nikei K., Saijo, N., Nishiwaki, Y.** (2008): Clinical outcome of chemoradiation therapy in patients with limited-disease small cell lung cancer with ipsilateral pleural effusion. *Journal of Thoracic Oncology* 3: 723–727
- **Nishimura T. et Paszkowski J.** (2007): Epigenetic transitions in plants not associated with changes in DNA or histone modification. *Biochimica et Biophysica Acta* 1769, 393-398
- **Okano M.** (1998): Dnmt2 is not required for de novo and maintenance methylation of viral DNA in embryonic stem cells. *Nucleic Acids Research*, 26, 2536-2540
- **Okano M., Bell D.W., Haber D.A., Li E.** (1999): DNA methyltransferases Dnmt3a and Dnmt3b Are essential for de novo methylation and mammalian development. *Cell*, 99, 247-257
- **Olsen P.H. et Ambros V.** (1999): The lin-4 regulatory RNA controls developmental timing in *Caenorhabditis elegans* by blocking LIN-14 protein synthesis after the initiation of translation. *Developmental Biology* 216: 671-680
- **Oue T., Yoneda A., Uehara S., Yamanaka H., Fukuzawa M.** (2009): Increased expression of multidrug resistance-associated genes after chemotherapy in pediatric solid malignancies. *Journal of Pediatric Surgery* 44: 377-380

- **Pallis A.G., Gridelli C., Wedding U., Faivre-Finn C., Veronesi G., Jaklitsch M., Luciani A., O'Brien M.** (2014): Management of elderly patients with NSCLC; updated expert's opinion paper: EORTC Elderly Task Force, Lung Cancer Group and International Society for Geriatric Oncology. *Annals of Oncology* 25: 1270–1283
- **Parkin D.M. et Bray F.** (2006): Chapter 2: The burden of HPV-related cancers. *Vaccine* 24: S11-S25
- **Pasquinelli A.E., Reinhart B.J., Slack F., Martindale M.Q., Kuroda M.I., Maller B., Hayward B.M., Ball E.E., Degnan B., Müller P., Spring J., Srinivasan A., Fishman M., Finnerty J., Corbo J., Levine M., Leahy P., Davidson E., Ruvkun G.** (2000): Conservation across animal phylogeny of the sequence and temporal regulation of the 21 nucleotide let-7 heterochronic regulatory RNA. *Nature* 408: 86-89
- **Paz I.B., Cook N., Odom-Maryon T., Xie Y., Wilczynski S.P.** (1997): Human papillomavirus (HPV) in head and neck cancer. An association of HPV 16 with squamous cell carcinoma of Waldayer's tonsillar ring. *Cancer* 79: 595–604.
- **Peng L., Yuan Z., Ling H., Fukasawa K., Robertson K., Olashaw N., Koomen J., Chen J., Lane W.S., Seto, E.** (2011): SIRT1 deacetylates the DNA methyltransferase 1 (DNMT1) protein and alters its activities. *Molecular and Cellular Biology* 31, 4720–4734
- **Perol M., Ciuleanu T.E., Arrieta O., Prabhaskar K., Syrigos, K.N., Park T.G., Kowalyszyn R.D., Pikiel J., Czyzewicz G., Orlov S., Lewanski C.R., Alexandris E., Zimmerman A., Chouaki N., John W.J., Yurasov S., Garon E.B.** (2014): REVEL: a randomized, double-blind, phase III study of docetaxel (DOC) and ramucirumab (RAM; IMC-1121B) versus DOC and placebo (PL) in the second-line treatment of stage IV non-small cell lung cancer (NSCLC) following disease progression after one prior platinum-based therapy. *Journal of Clinical Oncology* 32: 5 Supplement
- **Perry M.C.** (2008): The chemotherapy source book. *Lippincott Williams & Wilkins*, New York, 779 s
- **Planchard D., Popat S., Kerr K., Novello S., Smit E.F., FAIVRE-Finn C., Mok T.S., Reck M., Van Schil P.E., Hellmann M.D., Peters S., ESMO Guidelines**

**Working Group** (2018): Metastatic non-small-cell lung cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Annals of Oncology* 29: 4 Supplement

- **Ponger L., Duret L., Mouchiroud D.** (2001): Determinants of CpG islands: expression in early embryo and isochore structure. *Genome Research*, 11:1854-1860
- **Ranade, A.R., Cherba, D., Sridhar, S., Richardson, P., Webb, C., Paripati, A., Bowles, B., Weiss, G.J.** (2010): MicroRNA 92a-2\*: a biomarker predictive for chemoresistance and prognostic for survival in patients with small cell lung cancer. *Journal of Thoracic Oncology* 5: 1273-1278
- **Reck M., von Pawel J., Zatloukal P., Ramlau R., Gorbounova V., Hirsh V., Leigh N., Mezger J., Archer V., Moore N., Manegold Ch.** (2009): Phase III trial of cisplatin plus gemcitabine with either placebo or bevacizumab as first-line therapy for nonsquamous non-small-cell lung cancer. *Journal of Clinical Oncology* 27: 1227–1234
- **Reck M.** (2013): Nintedanib (BIBF 1120) plus docetaxel in NSCLC patients progressing after first-line chemotherapy: LUME Lung 1, a randomized, double-blind phase III trial. *Journal of Clinical Oncology* 31: 18 Supplement
- **Reck M., Rodríguez-Abreu D., Robinson A.G., Hui R., Csósz T., Fülöp, A., Gottfried M., Peled, N., Tafreshi A., Cuffe S., O'Brien M., Rao, S., Hotta K., Leiby M.A., Lubiniecki G., Shentu Y., Rangwala R., Brahmer J.R.** (2016): Pembrolizumab versus chemotherapy for PD-L1-positive non-small-cell lung cancer. *The New England Journal of Medicine* 375: 1823–1833
- **Reidy J., McHugh E., Stassen L.F.A** (2011): A review of the relationship between alcohol and oral cancer. *The Surgeon* 9: 278-283
- **Reinhart B.J., Slack F.J., Basson M., Bettinger J.C., Pasquinelli A.E., Rougvie A.E., Horvitz H.R., Ruvkun G.** (2000): The 21 nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature* 403: 901-906
- **Rischin D., Young R., Fischer R., Fox S., Le Q., Peters L., Choi J., O'Sullivan, Giralt J., McArthur G.** (2009): Prognostic significance of HPV and p16 status in patients with oropharyngeal cancer treated on a large international phase III trial. *Journal of Clinical Oncology* 27: 15 Supplement

- **Rodriguez A., Griffiths-Jones S., Ashurst J.L., Bradley A.** (2004): Identification of mammalian microRNA host genes and transcription units. *Genome Research* 14: 1902-1910
- **Rosell R., Carcereny E., Gervais R., Vergnenegre A., Massuti B., Felip E., Palmero R., Garcia-Gomez P., Pallares C., Sanchez J.M., Porta R., Cobo M., Garrido P., Longo F., Moran T., Insa A., De Marinis F., Corre R., Bover I., Illiano A., Dansin E., et al.** (2012): Erlotinib versus standard chemotherapy as first-line treatment for European patients with advanced EGFR mutation-positive non-small-cell lung cancer (EORTAC): a multicentre, open-label, randomised phase 3 trial. *The Lancet Oncology* 13: 239–246
- **Rothemberg M. et Ellisen L.W.** (2012): The molecular pathogenesis of head and neck squamous cell carcinoma. *Journal of Clinical Investigation* 6: 1951-1957
- **Ryerson A.B., Peters E.S., Coughlin S.S., Chen V.W., Gillison M.L., Reichman M.E., Wu X., Chaturvedi A.K., Kawaoka K.** (2008): Burden of potentially human papillomavirus-associated cancers of the oropharynx and oral cavity in the USA 1998–2003. *Cancer* 113, 10 Supplement, 2901–2909
- **Sculier J.P. et Moro-Sibilot D.** (2009): First- and second-line therapy for advanced nonsmall cell lung cancer. *European Respiratory Journal* 33: 915-930
- **Sharma S., Poetz F., Bruer M., Ly-Hartig T.B., Schott J., Séraphin B., Stoecklin G.** (2016): Acetylation-Dependent Control of Global Poly(A) RNA Degradation by CBP/p300 and HDAC1/2. *Molecular Cell* 63, 927-938
- **Shaw A.T et Solomon B.** (2011): Targeting anaplastic lymphoma kinase in lung cancer. *Clinical Cancer Research* 17: 2081–2086
- **Shukla S., Khan S., Tollefsbol T.O., Meeran S.M.** (2013): Genetics and Epigenetics of Lung Cancer: Mechanisms and Future Perspectives. *Current Cancer Therapy Reviews* 9: 97-110
- **Schwarz D.S., Hutvagner G., Du T., Xu Z.S., Aronin N., Zamore P.D.** (2003): Asymmetry in the assembly of the RNAi enzyme complex. *Cell* 115: 199-208
- **Skřičková J., Čoupek P., Babičková L., Tomišková M., Kaplanová J., Princ D., Kadlec B., Špelda S.** (2008a): Léčebné postupy u nemalobuněčného karcinomu plic. *Klinická Onkologie* 21: 317-329

- **Skříčková J., Tomášková M., Babičková L., Kaplanová J.** (2008b): Diagnostika a léčba karcinomu plic ve stáří. *Česká Geriatrická Revue* 6(I): 20-30
- **Skříčková J., Venclíček O., Kadlec B., Tomášková M., Jakubíková L., Špeldová J. Dušek. L., Mužík J.** (2016): Nenalobuněčný karcinom plic. *Postgraduální medicína* 18: 21–27
- **Skříčková J., Kolek V. a kol.** (2017): Základy moderní pneumoonkologie (2. vyd.). *Maxdorf*, Praha
- **Skříčková J., Kadlec B., Venclíček O., Merta Z.** (2018): Karcinom plic. *Časopis českých lékařů* 157: 226-236
- **Slack F.J., Basson M., Liu Z., Ambros V., Horvitz H.R., Ruvkun G.** (2000): The lin-41 RBCC gene acts in the Caenorhabditis elegans heterochronic pathway between the let-7 regulatory RNA and the LIN-29 transcription factor. *Molecular Cell* 5: 659-669
- **Smilek P., Plzák J., Klozar J. a kol.** (2015): Karcinomy dutiny ústní a hltanu. Havlíčkův Brod: *Tobiáš*, 377 s. ISBN 978-80-7311-153-3
- **Snustad D.P. et Simmons M.J.** (2016): Principles of Genetics, 7th Edition. *Wiley*, ISBN 978-11-191-4236-2
- **Spigel D.R., Reckamp K.L., Rizvi N.A., Poddubskaya E., West H.J., Ernst W., Eberhardt E., Baas, P., Antonia S.J., Pluzanski A., Vokes E.E., Holgado E., Waterhouse D.M., Ready N., Gainor J.F., Aren O.R., Horn, L., Paz-Ares, L., Baudelet Ch., Lestini B.J., Brahmer J.R.** (2015): A phase III study (CheckMate 017) of nivolumab (NIVO; anti-programmed death-1 [PD-1]) vs docetaxel (DOC) in previously treated advanced or metastatic squamous (SQ) cell non-small cell lung cancer (NSCLC). *Journal of Clinical Oncology* 33: 15 Supplement
- **Spurný V., Mechl Z., Červená R.** (2002): Současné možnosti chemoterapie a radioterapie nádorů oblasti hlavy a krku. *Praktický lékař* 82: 541-545
- **Song J.J., Smith S.K., Hannon G.J., Joshua-Tor L.** (2004): Crystal structure of Argonaute and its implications for RISC slicer activity. *Science* 305: 1434–1437
- **Stolz A., Pafko P. a kolektiv** (2010): Komplikace v plicní chirurgii. *Grada publishing*, Praha, 240 s



- **Suetake I., Miyazaki J., Murakami C., Takeshima H., Tajima S.** (2003): Distinct enzymatic properties of recombinant mouse DNA methyltransferases Dnmt3a and Dnmt3b. *Journal of Biochemistry* 133, 737-744
- **Suetake I., Shinozaki F., Miyagawa J., Takeshima H., Tajima, S.** (2004): DNMT3L stimulates the DNA methylation activity of Dnmt3a and Dnmt3b through a direct interaction. *Journal of Biological Chemistry* 279, 27816-27823
- **Sun Y., Bai Y., Zhang F., Wang Y., Guo Y., Guo L.** (2010): miR-126 inhibits non-small cell lung cancer cells proliferation by targeting EGFL7. *Biochemical and Biophysical Research Community* 391: 1483-1489
- **Šlampa P., Smilek P. a kol.** (2016): Nádory hlavy a krku: přehled diagnostiky a léčby maligních nádorů horních dýchacích a polykacích cest, hrtanu, slinných žláz a kůže. Praha: *Mladá fronta*, 377 s. ISBN 978-80-204-3743-3
- **Tan T.M. et Ting R.C.** (1995): *In vitro* and *in vivo* inhibition of human papillomavirus type 16 E6 and E7 genes. *Cancer Research* 55: 4599–4605
- **Terry P.B.** (2007): Lung disorders. *Johns Hopkins Health*, Maryland, 76 s
- **Tomari Y. et Zamore P.D.** (2005): Perspective: machines for RNAi. *Genes & Development* 19: 517-529
- **Trussardi-Regnier A., Millot J.M., Gorisse M.C., Delvincourt C., Prevost A.** (2007): Detection of drug-resistance genes using single bronchoscopy biopsy specimens. *Oncology Reports* 18: 703-708
- **Verbsky M.L. et Richards E.J.** (2001): Chromatin remodeling in plants. *Current Opinion in Plant Biology* 4, 494-500
- **Vetting M.W., de Carvalho L.P.S., Yu M., Hegde S.S., Magnet S., Roderick S.L., Blanchard, J.S.** (2005): Structure and functions of the GNAT superfamily of acetyltransferases. *Archives of Biochemistry and Biophysics* 433, 212-226
- **von Känel T. et Huber A.R.** (2013): DNA methylation analysis. *Swiss Med. Wkly* 143, w13799
- **Vorlíček, J., Abrahámová J., Vorlíčková, H. a kol.** (2006) Klinická onkologie pro sestry. Praha: *Grada*, ISBN 80-247-1716-6
- **Vyskot B.** (1999): Přehled vývojové biologie a genetiky. Vydal Ústav molekulární genetiky AV ČR, Praha.

- **Wahid F., Shehzad A., Khan T., Kim Y.Y.** (2010): MicroRNA: Synthesis, mechanism, function, and recent clinical trials. *Biochimica et Biophysica Acta* 1803: 1231-1243
- **Wang Y., Medvid R., Melton C., Jaenisch R., Bielloch R.** (2007): DGCR8 is essential for microRNA biogenesis and silencing of embryonic stem cell self-renewal. *Nature Genetics* 39: 380-385
- **Wang Q., Wang S., Wang H., Li P., Ma Z.** (2012): MicroRNAs: novel biomarkers for lung cancer diagnosis, prediction and treatment. *Experimental Biology and Medicine (SAGE Journals)* 237: 227-235
- **Weinberger P.M., Yu Z., Haffty B.G., Kowalski D., Harigopal M., Brandsma J., Sasaki C., Joe J., Camp R.L., Rimm D.L., Psyrri A.** (2006): Molecular classification identifies a subset of human papillomavirus-associated oropharyngeal cancers with favourable prognosis. *Journal of Clinical Oncology* 121: 2465–2472
- **Weinberger P.M., Yu Z., Kountourakis P., Sasaki C., Haffty B.G., Kowalski D., Merkley M.A., Rimm D.L., Camp R.L., Psyrri A.** (2009): Defining molecular phenotypes of human papillomavirus-associated oropharyngeal squamous cell carcinoma: validation of three-class hypothesis. *Otolaryngology-Head and Neck Surgery* 141: 382–389
- **Weinberger P.M., Yu Z., Haffty B.G., Kowalski D., Harigopal M., Brandsma J., Sasaki C., Joe J., Camp R.L., Rimm D.L., Psyrri A.** (2006): Molecular classification identifies a subset of human papillomavirus--associated oropharyngeal cancers with favorable prognosis. *Journal of Clinical Oncology* 24:736-747
- **Wightman B., Ha I., Ruvkun G.** (1993): Posttranscriptional regulation of the heterochronic gene *lin-14* by *lin-4* mediates temporal pattern-formation in *C. elegans*. *Cell* 75: 855-86
- **Winter J., Jung S., Keller S., Gregory R.I., Diederichs S.** (2009): Many roads to maturity: microRNA biogenesis, pathways and their regulation. *Nature Cell Biology* 11: 228-234
- **Xie S., Wang Z., Okano M., Nogami M., Li Y., He W., Li E.** (1999): Cloning, expression and chromosome locations of the human DNMT3 gene family. *Gene* 236, 87-95

- **Yan B., Yang X., Lee T.L., Friedman J., Tang J., van Waes C., Chen Z.** (2007): Genome-wide identification of novel expression signatures reveal distinct patterns and prevalence of binding motifs for p53, nuclear factor-kappaB and other signal transcription factors in head and neck squamous cell carcinoma. *Genome biology* 8: R78
- **Yan B., Chen G., Saigal K., Yang X., Jensen S.T., Van Waes C., Stoeckert C.J., Chen Z.** (2008): Systems biology-defined NF-kappaB regulons, interacting signal pathways and networks are implicated in the malignant phenotype of head and neck cancer cell lines differing in p53 status. *Genome biology* 9: R53.
- **Yang J.Ch., Schuler M.H., Yamamoto N., O'Byrne K.J., Hirsh V., MokSaravut T., Geater L., Orlov S.V., Tsai Ch.M., Boyer M.J., Su W.Ch., Bennouna J., Kato T., Gorbunova V., Lee K.H., Shah R.N.H., Massey D., Lorence R.M., Shahidi M., Segui L.V.** (2012): LUX-Lung 3: a randomized, open-label, phase III study of afatinib vs cisplatin/pemetrexed as 1st-line treatment for patients with advanced adenocarcinoma of the lung harboring EGFR-activating mutations. *Journal of Clinical Oncology* 30: 18 Supplement
- **Yen R.C., Vertino P.M., Nelkin B.D., Yu J.J., El-Deiry W., Cumaraswamy A., Baylin S.B.** (1992): Isolation and characterization of the cDNA encoding human DNA methyltransferase. *Nucleic Acids Research* 20, 2287-2291
- **Young D., Xiao C.C., Murphy B., Moore M., Fakhry C., Day T.A.** (2015): Increase in head and neck cancer in younger patients due to human papillomavirus (HPV). *Oral Oncology* 51: 727-730
- **Zaman M.S., Thamminana S., Shahryari V., Chiyomaru T., Deng G., Saini S., Majid S., Fukuhara S., Chang I., Arora S., Hirata H., Ueno K., Singh K., Tanaka Y., Dahiya R.** (2012): Inhibition of PTEN gene expression by oncogenic miR-23b-3p in renal cancer. *PLoS One* 7: e50203
- **Zatloukal P.** (2005): Chemoterapie karcinomu plic. *Farmakoterapie* 1: 76-80
- **Zemach A., McDaniel I.E., Silva P., Zilberman D.** (2010): Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science* 328, 916–919
- **Zheng Z. et Baker C.C.** (2006): Papillomavirus genome structure, expression, and post-transcriptional regulation. *Frontiers Bioscience* 11: 2286-2302

- **Zhang G., Huang H., Liu D., Cheng Y., Liu X., Zhang W., Yin R., Zhang D., Zhang P., Liu J., Li C., Liu B., Luo Y., Zhu Y., Zhang N., He S., He C., Wang H., Chen D.** (2015): N6-Methyladenine DNA Modification in *Drosophila*. *Cell* 161, 893-906
- **Zhang Y.K., Zhu W.Y., He J.Y., Chen D.D., Huang Y.Y., Le H.B., Liu X.G.** (2012): miRNAs expression profiling to distinguish lung squamous-cell carcinoma from adenocarcinoma subtypes. *Journal of Cancer Research of Clinical Oncology* 138: 1641-1650

**Internetové zdroje:**

[www.genecards.org](http://www.genecards.org)

[www.miRBase.org](http://www.miRBase.org)

**Fiat FIAT**

**12345**

## 11. Seznam použitých zkratek

<b>AGO</b>	Argonaute protein
<b>ALK</b>	anaplastická lymfomová kinasa
<b>ATP</b>	adenosintrifosfát
<b>bi-shRNA</b>	bifunkční shRNA
<b>cDNA</b>	komplementární DNA
<b>CT</b>	výpočetní tomografie
<b>CTLA-4</b>	cytotoxický T lymfocytový antigen 4
<b>DGCR8</b>	DiGeorge syndrom critical region in gene 8
<b>DNMT1</b>	DNA methyltransferasa 1
<b>DNMT3L</b>	DNA methyltransferasa 3-like
<b>dsRNA</b>	dvouvláknová RNA
<b>EBV</b>	Epstein-Baarové virus
<b>ED</b>	extenzivní onemocnění
<b>EGFR</b>	epidermální růstový faktor
<b>EXP5</b>	exportin-5
<b>GDP</b>	guanosin-difosfát
<b>GTP</b>	guanosin-trifosfát
<b>HAT</b>	histoacetyltransferasa
<b>HDAC</b>	histondeacetylasa
<b>HPV</b>	lidský papilomavirus
<b>LD</b>	limitované stadium
<b>NMR</b>	nukleární magnetická rezonance
<b>NSCLC</b>	nemalobuněčný karcinom plic
<b>MAP</b>	mitogenem aktivovaná proteinkinasa
<b>mRNA</b>	mediátorová RNA
<b>miRNA</b>	mikroRNA
<b>miRNP</b>	microRNA ribonucleoprotein
<b>PAZ</b>	PIWI-argonaute-zwille doména

<b>PCR</b>	polymerasová řetězová reakce
<b>PET</b>	pozitronová emisní tomografie
<b>piRNA</b>	PIWI-interagující miRNA
<b>pre-miRNA</b>	prekurzorová miRNA
<b>PRKRA</b>	protein kinase, interferon-inducible double stranded RNA dependent aktivátor
<b>RISC</b>	RNA-induced silencing complex
<b>RLC</b>	RISC loading komplex
<b>RNA</b>	ribonukleová kyselina
<b>RNAi</b>	RNA interference
<b>RT</b>	pokožková teplota / reverzní transkripce
<b>SCLC</b>	malobuněčný karcinom plic
<b>shRNA</b>	malá vlásenková RNA
<b>siRNA</b>	malá interferující RNA
<b>TKI</b>	inhibitor tyrosinkinasy
<b>TRBP</b>	trans-activation response RNA-binding protein
<b>VEGFR</b>	vaskulární endoteliální růstový faktor

## 12. Přehled publikací autora

### Práce související s disertační prací

#### **Původní vědecké publikace *in extenso* v daném oboru uveřejněné v časopisech s IF**

Guerrero-Preston R., Michailidi C., Marchionni L., Pickering C.R., Frederick M.J., Myers J.N., Yegnasubramanian S., Hadar T., Noordhuis M.G., **Zizkova V.**, Fertig E., Agrawal N., Westra W., Koch W., Califano J., Velculescu V.E., Sidransky D. (2014): Key tumor suppressor genes inactivated by "greater promoter" methylation and somatic mutations in head and neck cancer. *Epigenetics* 9(7):1031-46. IF 2.888

Gaykalova D.A., Manola J.B., Ozawa H., **Zizkova V.**, Morton K., Bishop J.A., Sharma R., Zhang C., Michailidi C., Considine M., Tan M., Fertig E.J., Hennessey P.T., Ahn J., Koch W.M., Westra W.H., Khan Z., Chung C.H., Ochs M.F., Califano J.A. (2015): NF- $\kappa$ B and STAT3 transcription factor signatures differentiate HPV-positive and HPV-negative Head and neck squamous cell carcinoma. *International Journal of Cancer* 137(8):1879-89. IF 5.949

**Zizkova V.**, Skarda J., Janikova M., Luzna P., Radova L., Kurfurstova D., Kolar Z. (2015): The relationship of miR-21, miR-126 and miR-205 to P-glycoprotein, MRP1 and LRP/MVP in non-small cell lung cancer. *Austin Journal of Cancer and Clinical Research* 2(5):1042. IF 1.80

Janikova M., Zizkova V., Skarda J., Kharraishvili G., Radova L., Kolar Z. (2016): Prognostic significance of miR-23b in combination with P-gp, MRP and LRP/MVP expression in non-small cell lung cancer. *Neoplasma* 63(4):576-87. IF 2.10

Gaykalova D.A., **Zizkova V.**, Guo T., Tiscareno I., Wei Y. Vatapalli R., Hennessey P.T., Ahn J., Danilova L., Khan Z., Bishop J.A., Gutkind J.S., Koch W.M., Westra W.H., Fertig E.J., Ochs M.F., Califano J.A. (2017): Integrative computational analysis of transcriptional

and epigenetic alterations implicates DTX1 as a putative tumor suppressor gene in HNSCC. *Oncotarget*; 8(9): 15349-15363. IF 4.79

## **Ostatní publikace**

### **Původní vědecké publikace *in extenso* v daném oboru uveřejněné v časopisech s IF**

Kishore A., **Zizkova V.**, Kocourkova L., Petrek M. (2015): A Dataset of 26 Candidate Gene and Pro-Inflammatory Cytokine Variants for Association Studies in Idiopathic Pulmonary Fibrosis: Frequency Distribution in Normal Czech Population. *Frontiers in Immunology* 6:476. IF 2.827

Kishore A., **Zizkova V.**, Kocourková L., Petrková J., Bouros E., Nunes H., Lošťáková V., Müller-Quernheim J., Zissel G., Kolek V., Bouros D., Valeyre D., Petrek M. (2016): Association Study for 26 Candidate Loci in Idiopathic Pulmonary Fibrosis Patients from Four European Populations. *Frontiers in Immunology* 7:274. IF 3.034

Petřek M., Kocourková L., **Žižková V.**, Nosek Z., Táborský M., Petrková J. (2016): Characterization of three CYP2C19 gene variants by Mass Array and Point of care techniques - experience from a Czech centre. *Archivum Immunologiae et Therapiae Experimentalis* 64, 99-107. IF 2.242

Sikorova K., Kishore A., Rapti A., Adam K., Kocourkova L., **Zizkova V.**, Charikiopoulou M., Kalianos A., Bouros E., Bouros D., Petrek M. (2020): Association of *TGF-β3* and *ANXA11* with pulmonary sarcoidosis in Greek population. *Expert Review of Respiratory Medicine* 14(10): 1065-1069. IF 2.432

Tokić S., **Žižkova V.**, Štefanić M., Glavaš-Obrovac L., Marczy S., Samardžija M., Sikorova K., Petrek M. (2020): HLA-A, -B, -C, -DRB1, -DQA1, and -DQB1 allele and haplotype frequencies defined by next generation sequencing in a population of East Croatia blood donors. *Scientific Reports - Nature* 10(1): 5513. IF 3.998



Zakharyan R., Ghazaryan H., Kocourkova L., Chavushyan A., Mkrtychyan A., **Zizkova V.**, Arakelyan A., Petrek M. (2020): Association of Genetic Variants of Dopamine and Serotonin In Schizophrenia. *Archives of Medical Research* 51(1): 13-20. IF 2.093

### 13. Publikovaná abstrakta

Janíková M., Škarda J., **Žižková V.**, Lužná P., Radová L. (2011): Význam miR-21, miR-126 a miR-205 u pacientů s nemalobuněčným karcinomem plic. Edukační sborník XXXV. Brněnské onkologické dny a XXV. Konference pro nelékařské zdravotnické pracovníky, Brno, ISBN 978-80-86793-17-7

Janíková M., Škarda J., **Žižková V.**, Lužná P., Radová L. (2011): MiR-21, miR-126 and miR-205 expression in NSCLC patients. Sborník abstrakt. Konference: Nová léčiva závažných lidských onemocnění. Kouty nad Desnou

Janíková M., Škarda J., **Žižková V.**, Lužná P., Radová L. (2011): Expres miR-21, miR-126 a miR-205 u pacientů s NSCLC. Sborník abstrakt. Konference vědeckých prací studentů DSP. Olomouc. ISBN 978-80-244-2847-5

Janíková M., Škarda J., **Žižková V.**, Lužná P., Radová L. (2011): Importance of miR-21, miR-126 and miR-205 detection in archived FFPE samples in NSCLC patients. Biopreserv & Biobanking; 9:299, Abstract No. HR-07. ESBB's Inaugural Conference, Marseille, France. IF 1,294

Janíková M., Škarda J., **Žižková V.**, Lužná P., Radová L. (2011): The Importance of miR-21, miR-126, and miR-205 in NSCLC Patients. Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub.; 155: S2. Konference projektu 303/09/H048, Drnovice. ISSN 1213-8118. IF 0,702

Janikova M., **Zizkova V.**, Skarda J., Luzna P., Radova L. (2012): Correlation of Specific MiRNAs Expression With Survival and Drug Resistance Related Protein Expression in Non-small Cell Lung Cancer. *Eur J Cancer*. 2012; 48: S5: S200-S201, Abstract No. 836. 22<sup>nd</sup> Biennial Congress of the European-Association-for-Cancer-Research, Barcelona, Spain. IF 5,061

Skarda J., Janikova M., **Zizkova V.** (2012): Regulation of multidrug resistance in NSCLC by miRNAs. *Virchows Arch.* 2012; 461: S1:S160-S161. 24<sup>th</sup> European Congress of Pathology, Prague, Czech Republic. IF 2,676

Skarda J., Janikova M., **Zizkova V.** (2012): The role of miRNA expressions in multidrug resistance in NSCLC. *Lung Cancer.* 2012; 77: S1:S14-S15. 13<sup>th</sup> Central European Lung Cancer Conference, Prague, Czech Republic. IF 3,392

Janikova M., **Zizkova V.**, Luzna P., Radova L., Skarda J. (2012): The role of miR-21, miR-126 and miR-205 in P-glycoprotein, MRP1 and LRP/MVP mediated drug resistance in non-small cell lung cancer. *Sborník abstrakt. 3<sup>rd</sup> ICCTI Workshop – MikroRNA v onkologii, Brno.*

Janikova M., **Zizkova V.**, Skarda J., Radova L. (2012): The role of selected miRNA in MDR caused by P-gp, MRP1 and LRP/MVP in NSCLC patients. *Biomedical Papers of Medical Faculty University Palacky Olomouc Czech Republic*; 156: S3:121. Konference projektu 303/09/H048, Drnovice. ISSN 1213-8118. IF 0,990

Škarda J., Janíková M., **Žižková V.**, Kolek V., Kolář Z. (2013): Korelace exprese miRNA s expresí proteinu odpovědných za reparaci DNA u nemalobunečných karcinomů plic. *Sborník abstrakt, XXXVII. Brněnské onkologické dny a XXXVII. Konference pro nelékařské zdravotnické pracovníky, Brno.* ISBN 978-80-904596-9-4

**Žižková V.**, Janikova M., Luzna P., Skarda J., Radova L., Kolek V., Kolar Z. (2013): The relationship of selected miRNAs to P-glycoprotein, MRP1 and LRP/MVP mediated drug resistance in non-small cell lung cancer. *Sbornik abstrakt. The 9<sup>th</sup> Symposium & Workshop on Molecular Pathology and Histo(cyto)chemistry. The 99<sup>th</sup> Seminar of the Czech Division of the International Academy of Pathology. The 5<sup>th</sup> Olomouc Days of Histology Laboratory Technicians, Olomouc.* ISBN 978-80-7471-022-3

Janíková M., **Žižková V.**, Radová L., Škarda J., Kolář Z. (2013): Identifikácia nových biomarkerov MDR u pacientov s NSCLC pomocou prístupu *in silico*. Sborník abstrakt. XVII. celostátní konference DNA diagnostiky, Dolní Morava.

Janíková M., Žižková V., Radová L., Škarda J., Kolář Z. (2014): Identification of new MDR biomarkers in NSCLC patients using *in silico* approach. Sborník abstrakt: Konference Chemické biologie a genetiky. Kouty nad Desnou.

Gaykalova D.A., **Zizkova V.**, Tiscareno I., Wei Y. Vatapalli R., Hennessey P.T., Ahn J., Danilova L., Khan Z., Bishop J.A., Koch W.M., Westra W.H., Ochs M.F., Califano J.A. (2015): DTX1 is an epigenetically regulated tumor suppressor gene discovered by integrative analysis of epigenetic and transcriptional alterations in HNSCC. 106<sup>th</sup> Annual Meeting of the American Association for Cancer Research; Apr 18-22; Philadelphia, PA, USA: AACR; *Cancer Research* 75 (15 Suppl): Abstract nr 815. doi:10.1158/1538-7445.AM2015-815. IF 9.88

Gaykalova D.A., Kelley D., Guo T., Bohrson C., Tiscareno I., **Zizkova V.**, Considine M., Danilova L., Flam, E., Bishop J., Ahn J., Merritt S., Goldsmith M., Zhang C., Koch W., Westra W., Khan Z., Ochs M., Wheelan S., Fertig E., Califano J. (2016): The discovery of novel GSN alternative splicing in head and neck squamous cell carcinoma. 107<sup>th</sup> Annual Meeting of the American Association for Cancer Research; Apr 16-20; New Orleans, LA, USA: AACR; *Cancer Research* 76 (14 Suppl): Abstract nr 2880. doi: 10.1158/1538-7445.AM2016-2880. IF 9.88

## **14. Seznam přednášek/posterů přednesených uchazečem na veřejných odborných fórech**

**Zizkova V.** (2018): Farmakogenetika v praxi. The 14<sup>th</sup> Symposium of Molecular Pathology and Histo(cyto)chemistry. The 104<sup>th</sup> Olomouc Diagnostic Seminar of The Czech Division of the International Academy of Pathology. The 8<sup>th</sup> Olomouc Days of Histologic Technicians, Olomouc. ISBN 978-80-7471-234-0

# Key tumor suppressor genes inactivated by “greater promoter” methylation and somatic mutations in head and neck cancer

Rafael Guerrero-Preston<sup>1,2,†,\*</sup>, Christina Michailidi<sup>1,†</sup>, Luigi Marchionni<sup>3</sup>, Curtis R Pickering<sup>4</sup>, Mitchell J Frederick<sup>4</sup>, Jeffrey N Myers<sup>4</sup>, Srinivasan Yegnasubramanian<sup>5</sup>, Tal Hadar<sup>1</sup>, Maartje G Noordhuis<sup>1,6</sup>, Veronika Zizkova<sup>1,7</sup>, Elana Fertig<sup>8</sup>, Nishant Agrawal<sup>1,5</sup>, William Westra<sup>1</sup>, Wayne Koch<sup>1</sup>, Joseph Califano<sup>1,9</sup>, Victor E Velculescu<sup>5</sup>, and David Sidransky<sup>1,\*</sup>

<sup>1</sup>Department of Otolaryngology and Head and Neck Surgery; Johns Hopkins University; School of Medicine; Baltimore, MD USA; <sup>2</sup>Department of Obstetrics and Gynecology; University of Puerto Rico School of Medicine; Río Piedras, Puerto Rico; <sup>3</sup>Department of Oncology Biostatistics; Johns Hopkins University; School of Medicine; Baltimore, MD USA; <sup>4</sup>Department of Head and Neck Surgery; University of Texas M.D. Anderson Cancer Center; Houston, TX USA; <sup>5</sup>Sidney Kimmel Comprehensive Cancer Center; Johns Hopkins University; Baltimore, MD USA; <sup>6</sup>Department of Otorhinolaryngology-Head and Neck Surgery; University of Groningen; University Medical Center; Groningen, The Netherlands; <sup>7</sup>Laboratory of Molecular Pathology and Institute of Molecular and Translational Medicine; Faculty of Medicine and Dentistry; Palacky University; Olomouc, Czech Republic; <sup>8</sup>Division of Biostatistics and Bioinformatics; Department of Oncology; Sidney Kimmel Comprehensive Cancer Center; Baltimore, MD USA; <sup>9</sup>Milton J. Dance Head and Neck Center; Greater Baltimore Medical Center; Baltimore, MD USA

<sup>†</sup>These authors contributed equally to this work.

**Keywords:** Head and Neck Squamous Cell Carcinoma, Tumor Suppressor Genes, DNA methylation, somatic mutations, integration analysis

Tumor suppressor genes (TSGs) are commonly inactivated by somatic mutation and/or promoter methylation; yet, recent high-throughput genomic studies have not identified key TSGs inactivated by both mechanisms. We pursued an integrated molecular analysis based on methylation binding domain sequencing (MBD-seq), 450K Methylation arrays, whole exome sequencing, and whole genome gene expression arrays in primary head and neck squamous cell carcinoma (HNSCC) tumors and matched uvulopalatopharyngoplasty tissue samples (UPPPs). We uncovered 186 down-regulated genes harboring cancer specific promoter methylation including *PAX1* and *PAX5* and we identified 10 key tumor suppressor genes (*GABRB3*, *HOXC12*, *PARP15*, *SLCO4C1*, *CDKN2A*, *PAX1*, *PIK3AP1*, *HOXC6*, *PLCB1*, and *ZIC4*) inactivated by both promoter methylation and/or somatic mutation. Among the novel tumor suppressor genes discovered with dual mechanisms of inactivation, we found a high frequency of genomic and epigenomic alterations in the *PAX* gene family of transcription factors, which selectively impact canonical *NOTCH* and *TP53* pathways to determine cell fate, cell survival, and genome maintenance. Our results highlight the importance of assessing TSGs at the genomic and epigenomic level to identify key pathways in HNSCC, deregulated by simultaneous promoter methylation and somatic mutations.

## Introduction

Head and neck squamous cell carcinoma (HNSCC) is the sixth most common cancer worldwide with an approximate 50% five-year survival rate.<sup>1</sup> Two groups that independently studied the genetic origins of HNSCC reported inactivating mutations in *NOTCH1*.<sup>1,2</sup> This was the first strong evidence of *NOTCH1* mutations in solid tumors; analysis of the mutations suggested that *NOTCH1* might act as a tumor suppressor gene (TSG) in HNSCC.<sup>1,2</sup> Notwithstanding this important finding, and contrary to original expectations, these and other detailed analyses of HNSCC did not uncover a great number of recurrent somatic mutations in novel genes.<sup>3,4</sup>

The number of known mutations and specific mutational hotspots in HNSCC tumors only partially explains their biological complexity and limits the development of novel diagnostic markers and therapeutic agents. *TP53* was again identified as the most commonly mutated gene in HNSCC and, while mutant *TP53* has been associated with poor survival,<sup>5</sup> the most important biologic consequences of this alteration have been elusive. Moreover, it was also known that overall and disease-specific survival is higher in patients with HPV-associated HNSCC tumors,<sup>6</sup> and that this distinct molecular and pathologic subtype displays an average of 4 somatic mutations per tumor, while HPV-negative HNSCC tumors harbor 20.<sup>1</sup> HPV-associated HNSCCs have a distinctly different molecular landscape from HPV-negative HNSCCs. *TP53* is rarely identified as mutated in HPV-positive HNSCC patients because the HPV E6 oncoprotein silences

\*Correspondence to: David Sidransky; Email: dsidrans@jhmi.edu; Rafael Guerrero-Preston; Email: rguerre3@jhmi.edu  
Submitted: 04/21/2014; Accepted: 04/25/2014; Published Online: 05/01/2014  
<http://dx.doi.org/10.4161/epi.29025>

*TP53*, protecting the cells from apoptosis and senescence, while the HPV E7 oncoprotein deregulates the cell cycle.<sup>7,8</sup> *CDKN2A*, a principal cyclin-dependent kinase inhibitor that decelerates the cell cycle, is lost in HPV-negative HNSCC<sup>9</sup> and amplified in HPV-positive HNSCC.<sup>10</sup> HNSCCs also exhibit many chromosomal abnormalities, including amplifications of the 11q13 region containing the *cyclin D1* gene and the 7p11 region encoding *EGFR*, which lead to proto-oncogene activation.<sup>1</sup>

Aberrant DNA methylation of CpGs in the proximity of predicted transcription start sites (TSS) often leads to alterations in gene function and pathway deregulation in human cancer. Epigenetic events linked to TSG inactivation through promoter methylation, are more frequent events than somatic mutations in cancer, and may be driving tumorigenic initiation and progression. Promoter methylation of *CDK2NA*, *HOXA9*, *NID2*, *EDNRB*, *KIF1A*, and *DCC* have previously been identified and characterized in HNSCC.<sup>11-13</sup>

To date, recent high-throughput methylation studies<sup>14,15</sup> have not focused on the relative contribution of TSGs inactivation by DNA methylation and somatic mutations in oncogenesis and may have severely underestimated the true frequency of inactivation of key TSGs and signaling pathways. We tested the hypothesis that TSG promoter methylation predominantly occurs in genes or pathways with well-known somatic mutations and/or deletions in most HNSCC tumors, including *TP53*, *CDK2NA*,<sup>16</sup> and, more recently, *NOTCH1*<sup>1,2</sup> and *FAT1*,<sup>17</sup> as well as in recently described genes with low frequency mutations.<sup>18</sup>

## Results

### Patient selection

One hundred and eight (108) head and neck squamous cell carcinoma (HNSCC) and 35 uvulopalatopharyngealplasty (UPPP) patients were consented for this study at Johns Hopkins Medical Institutions hospitals and MD Anderson Cancer Center. The study was approved by the Ethics Committee of each participating hospital, as well as by the Johns Hopkins Institutional Review Board. The 143 samples were divided into Discovery and Prevalence cohorts. The Discovery cohort consisted of 32 HNSCC and 16 UPPP samples. The Prevalence cohort consisted of 76 HNSCC and 19 UPPP samples. The 32 tumor samples selected for the Discovery cohort were accrued in Hopkins (n=17) and in MD Anderson (n=15). The 35 UPPP samples were accrued at Hopkins.

### Johns Hopkins component

Fresh-frozen surgically resected HNSCC (n = 93) and UPPP (n = 35) tissue samples were obtained from patients at Johns Hopkins Medical Institutions in Baltimore. HNSCC tissue was analyzed by frozen section histology to estimate neoplastic cellularity. In order to enrich the samples for neoplastic cells, normal tissue was removed from the samples using macro-dissection based on the frozen section histology. HPV tumor status was determined for oropharyngeal tumors per standard clinical care using in situ hybridization. Hybridization was performed using the HPV III Family16 probe set that captures HPV genotypes

16, 18, 33, 35, 45, 51, 52, 56, and 66. HPV16-positive controls included an HPV16-positive oropharyngeal cancer and the SiHa and CaSki cell lines. HPV tumor status was also determined by E6/E7 PCR primer amplification.

### MD Anderson component

Fresh-frozen surgically resected tumor samples (n = 15) were obtained from consented patients treated for HNSCC at the University of Texas MD Anderson Cancer Center, under an Institutional Review Board approved protocol. Frozen tissue was embedded in optimal cutting temperature compound and cryosections from the top and middle of specimens were stained with hematoxylin and eosin prior to being evaluated by a pathologist for the presence of >60% tumor nuclei content or absence of tumor (i.e., normal). Samples that passed this criterion were sectioned all the way through and washed once in PBS prior to isolating genomic DNA using an ArchivePure DNA purification kit.

**Table 1** describes the clinical attributes of the combined 143 patient samples from the Discovery and Validation cohorts, as well as of the 289 samples from The Cancer Genome Atlas that were used to confirm the *TP53*mut/*PAX5* met and the *NOTCH1*mut/*PAX1*met associations. The 32 HNSCC patient samples selected for Discovery in this epigenetic study were the same used by Agrawal et al. for the genetic study of HNSCC<sup>1</sup>. This study design allowed us to analyze mutation, methylation, and expression data from the same HNSCC patients. We queried the methylome of these 32 HNSCC patients and frequency matched 16 UPPP normal controls using the Human Methylation 450K Beadchip (Illumina). However, the HNSCC transcriptome was queried with different platforms. In Hopkins the GeneST1.0 arrays (Affymetrix) were used to query the HNSCC and UPPP transcriptome and in MD Anderson they queried the HNSCC transcriptome with the Human Exon 1.0ST arrays (Affymetrix). The integrated methylation/mRNA expression analysis we are reporting was performed using transcriptome data from the 32 samples (16 HNSCC and 16 UPPP) accrued at Hopkins because we did not find a satisfactory method of combining GeneST1.0 and Human Exon 1.0ST array data. Nor did we have transcriptome and methylome data from normal patient samples collected at MD Anderson to compare with the results obtained with patients accrued at Hopkins.

### Characterization of the HNSCC methylome using MBD-seq

In the first part of our study, we used a methylated DNA binding domain based sequencing (MBD-seq) approach similar, in principle, to what has been previously described,<sup>19</sup> but with significant modifications (Fig. 1). This analysis was performed on a subset of tumors from the same Discovery Cohort in which Agrawal et al. discovered and mapped mutations in HNSCC,<sup>1</sup> comprising ten patients and ten frequency matched normal controls (uvulopalatopharyngoplasty tissue samples [UPPP]). The ten tumor samples were obtained before chemotherapy or radiation treatment, ensuring that the changes we identified are truly reflective of tumor biology, and were micro-dissected to achieve a neoplastic cellularity of greater than 60%. Following MBD-seq, an average of 48.9 million 50 bp reads was obtained for each

**Table 1.** Summary of clinical attributes of samples in Discovery, Prevalence, and TCGA cohorts

	Discovery			Prevalence			TCGA	
	Normal (n = 16)	HPV+ (n = 3)	HPV- (n = 29)	Normal (n = 19)	HPV+ (n = 4)	HPV- (n = 72)	HPV+ (n = 35)	HPV- (n = 244)
<b>Sex</b>								
Male	9(56%)	3(100%)	18(62%)	14(74%)	3(75%)	52(72%)	31(89%)	172(70%)
<b>Race</b>								
Non-Latino White	9(56%)	3(100%)	22(76%)	11(58%)	3(75%)	62(86%)	33(94%)	209(86%)
African American/ Black	7(44%)	0	4(14%)	7(37%)	1(25%)	7(10%)	2(6%)	24(10%)
Other	0	0	3(10%)	1(5%)	0	3(4%)	0	11(5%)
<b>Smoking</b>								
Yes	0	1(33%)	23(79%)	0	2(50%)	26(38%)	25(75%)	195(80%)
No	16(100%)	2(67%)	6(21%)	19(100%)	2(50%)	12(18%)	10(29%)	41(17%)
Unknown	0	0	0		0	34(47%)	0	8(3%)
<b>Tumor Site</b>								
Oral Cavity		0	25(86%)		0	35(49%)	12(34%)	160(66%)
Oropharynx		3(100%)	2(7%)		3(75%)	21(29%)	21(60%)	12(5%)
Larynx		0	2(7%)		0	10(14%)	1(3%)	71(29%)
Hypopharynx		0	0		0	4(5%)	1(3%)	1(0.4%)
Unknown		0	0		1(25%)	2(3%)	0	0
<b>T stage</b>								
I		0	1(4%)		0	0	3(9%)	17(7%)
II		3(100%)	6(21%)		0	1(1%)	10(29%)	63(26%)
III		0	12(41%)		0	2(3%)	2(6%)	54(22%)
IV		0	9(31%)		4(100%)	63(88%)	9(26%)	87(36%)
Unknown		0	1(4%)		0	6(8%)	11(31%)	23(9%)

sample, with an average of 68% of these reads aligning to the hg19 genome, 67% of which were aligning uniquely (Data set 1).

Our purpose was to study alterations in the promoter region of TSGs and for this study we focused on the “greater promoter,” a region that encompasses the well-studied proximal promoter that harbors CpG Islands (1500 bases up and downstream from the TSS) and the distal promoter (6000 bp upstream of the TSS), which includes recently identified CpG Island Shores<sup>20</sup> and Shelves (Fig. 2A).<sup>21</sup> (From this point on, wherever the term promoter is used it refers to the greater promoter region as defined here).

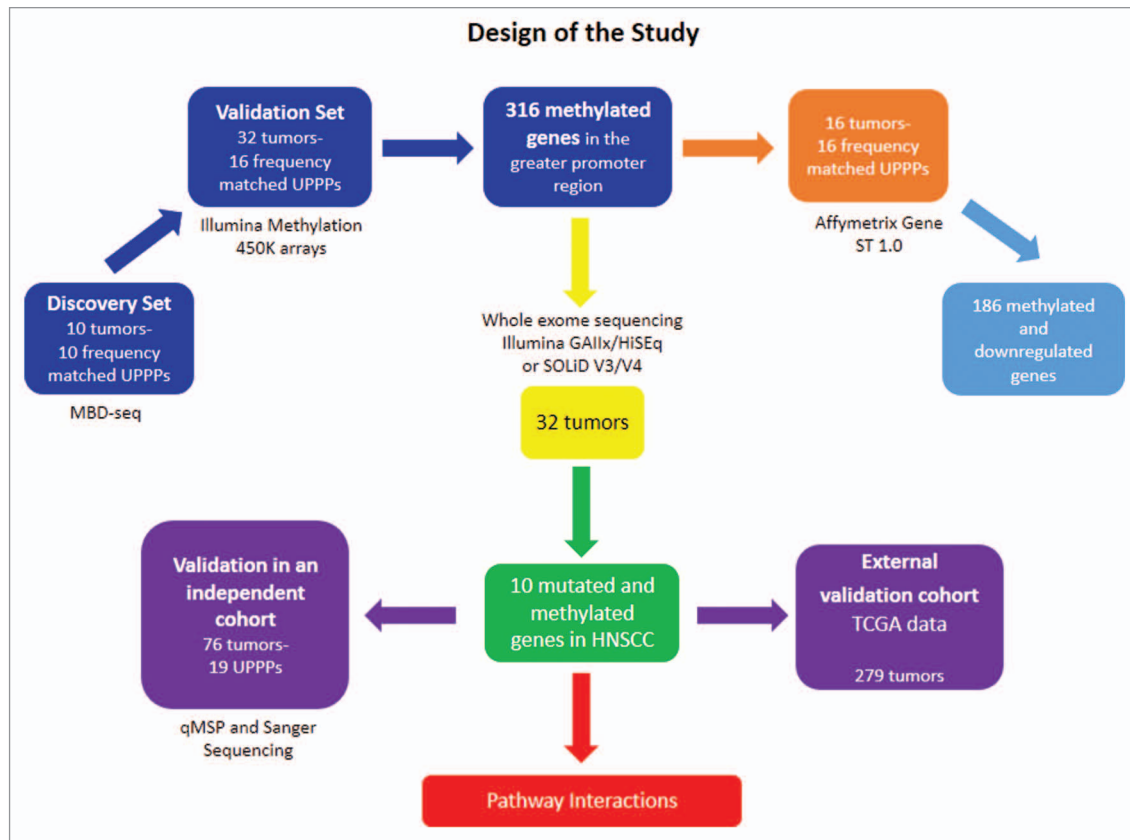
We used two independent and highly validated analytical approaches, model-based analysis of ChIP-seq (MACS)<sup>22</sup> and Bump hunting,<sup>23</sup> to identify methylation changes across the HNSCC genome. MACS identified 9648 alterations, 60% of which were gains in methylation events (Fig. 2B). The vast majority of the methylome changes identified by MACS (73%) were observed outside of the promoter region. Within the promoter region, we observed mostly (89%) gain of methylation events. The majority of the methylation loss identified by MACS (93%) occurred outside the promoter region. These genome-wide methylation motifs were integrated with the differentially methylated

regions (DMRs) identified by Bump hunting to obtain the first detailed next-gen analysis of the HNSCC promoter methylome (Data set 2; Tables S1 and S2).

#### Confirmation of methylated sites using 450K arrays and integration with expression arrays

To validate the MBD-seq results, we evaluated genome-wide differential methylation with the Infinium HumanMethylation450K Beadchip (450K array) in 41 cancer (including the 10 samples sequenced with MBD-seq) and 16 UPPP samples. The intersection of genome-wide methylation sequencing and methylation array screens uncovered 316 genes, which harbor promoter methylation in HNSCC (Table S3). To determine the extent of correlation between differential methylation and mRNA expression patterns, we performed mRNA expression microarray analysis (Affymetrix Gene ST 2.0) using 16 tumor and 16 UPPP samples from the samples included in the cohort analyzed with the 450K array platform. We found close to 60% concordance between concurrent greater promoter methylation and gene downregulation. 186 methylated genes were found to harbor methylation and downregulation; *PAX1* and *PAX5* exhibited the greatest expression loss (Fig. 3A; Table S4A). Table S4B shows the relationship between methylation frequency





**Figure 1.** Representation of the workflow of the study. The figure ascribes all the platforms and techniques used in the discovery and the two independent validation sets. The number of samples recruited its time are also depicted in brackets.

of the significantly downregulated candidate genes with  $FC < -2$  and the expression levels of these candidate genes for the 16 tumor samples queried with mRNA arrays in the Discovery cohort. We have also included **Figure S1** to show the expression level box plots for these candidate genes in the 16 HNSCC and 16 UPPP samples queried in the Discovery cohort.

#### Integration of promoter methylation with somatic mutation profiles

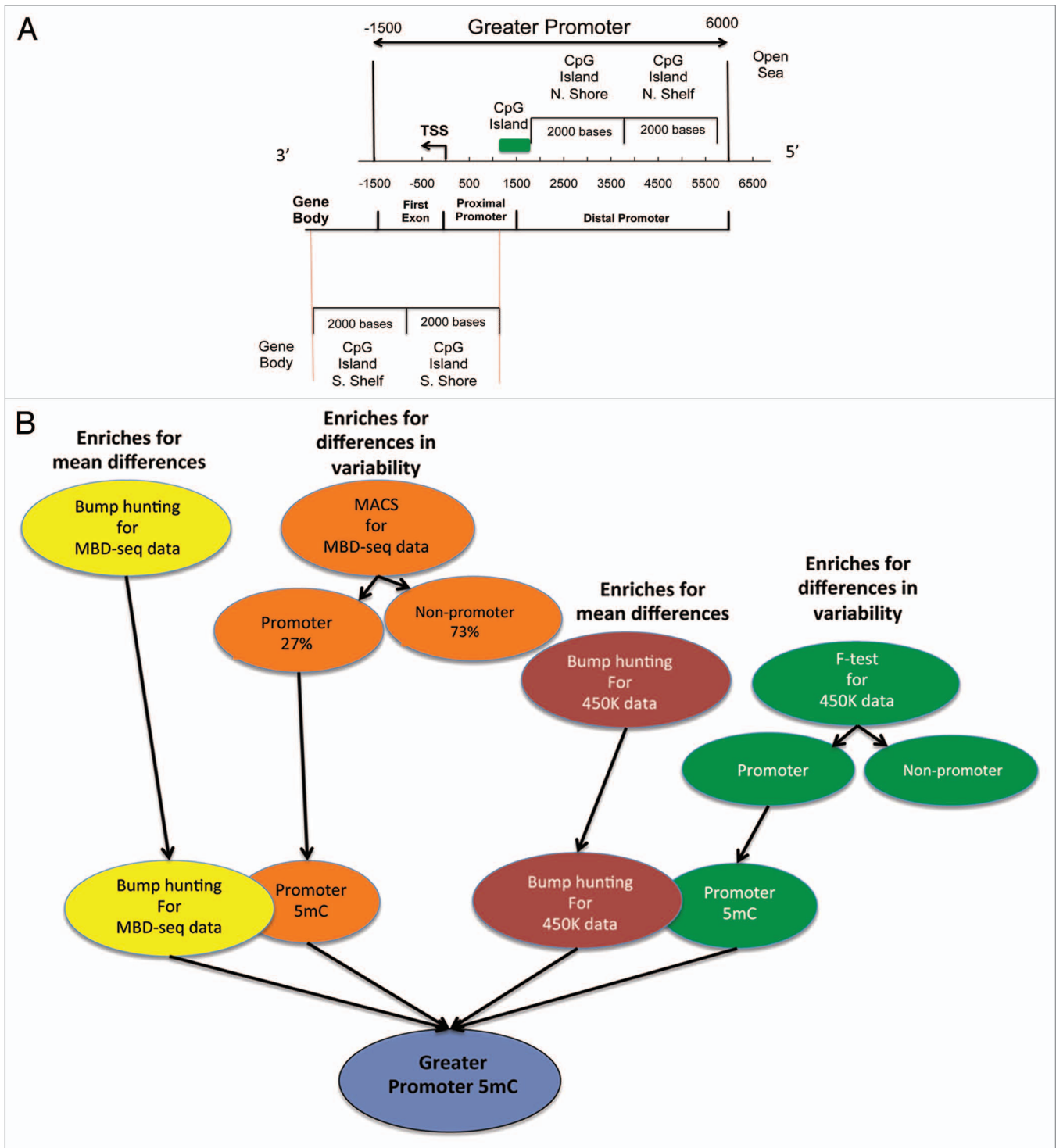
We subsequently intersected the promoter methylome with the mutational landscape of HNSCC<sup>1</sup> and identified concurrent promoter methylation and somatic mutations in ten tumor suppressor genes: *GABRB3*, *HOXC12*, *PARP15*, *SLCO4C1*, *CDKN2A*, *PAX1*, *PIK3API*, *HOXC6*, *PLCB1*, and *ZIC4* (**Fig. 3B**). *CDKN2A* (p16), one of the most frequently altered tumor suppressor genes in human cancer by mutation, methylation, and/or deletion was confirmed on this list. The rest of the genes displayed a very low mutation frequency in HNSCC, together with frequent inactivation by promoter methylation.

#### Unbiased genome-wide and gene set enrichment analyses

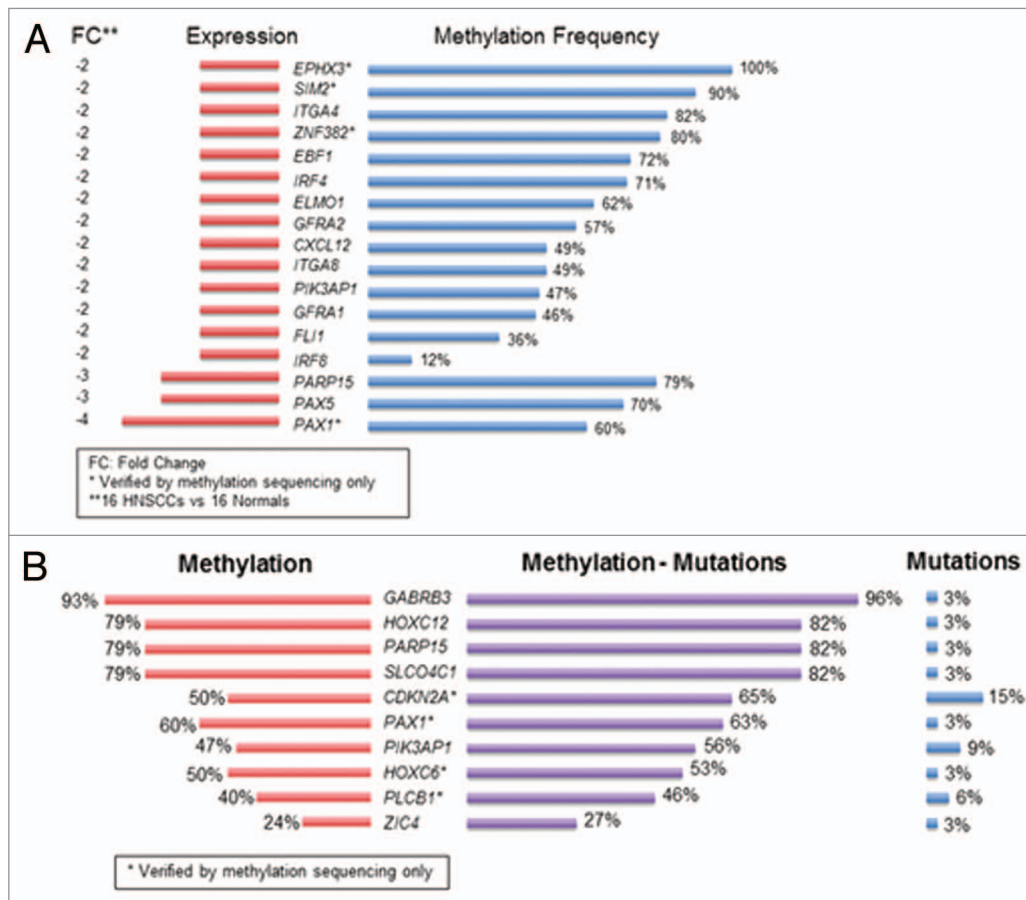
Unbiased genome-wide analyses were performed to visualize and interpret the large amount of data produced by the sequencing and microarray experiments. Unsupervised hierarchical clustering of the differential methylation events in HNSCC revealed a genome-wide loss of methylation. More than half of the genes in cancer have lost methylation when compared with normal samples at the genome-wide level. This massive loss of

methylation often suggests the acquisition of plasticity and de-differentiation associated with stem cells<sup>24</sup> (**Fig. S2A**), together with specific gains in promoter methylation (**Fig. S2B**). To better understand the relationship between genome-wide DNA methylation loss and copy number alterations we examined the correlation between genome-wide DNA methylation frequency and copy number alterations. The Spearman correlation coefficient between methylation frequency and copy number loss was 0.02. The Spearman correlation coefficient between methylation frequency and copy number gain was  $-0.004$ . The lack of correlation between DNA methylation and genomic aberrations in genes for which both methylome and copy number alterations are available in the Discovery cohort can be seen in **Table S5**. The genome-wide DNA methylation and genomic aberrations in tumor samples for which both methylome and copy number alterations are available in the Discovery cohort can be seen in **Figure S3**.

Analysis of Functional Annotation (AFA)<sup>25</sup> was then used to integrate the HNSCC methylation, mutation, and expression landscapes and detect alterations in cellular signaling pathways, protein-protein interaction networks, and gene ontology (GO) in HNSCC. AFA revealed that pathways involved in development, differentiation, adhesion, proliferation, and biological/cellular/transcriptional regulation are impacted by concurrent promoter methylation, mutations, and differential gene expression in HNSCC (**Figs. S4 and S5**; **Table S6**). AFA also showed



**Figure 2. (A)** Illustration defining the Greater Promoter region. Using a functional genomic distribution viewpoint we define five CpG genomic locations in relation to their distance to the Transcription Start Site: Proximal promoter, distal promoter, first exon, gene body and intergenic locations. From a CpG content and neighborhood context viewpoint we define four CpG genomic locations in relation to their distance to the nearest CpG Island: CpG Island, CpG Island Shore, CpG Island Shelf, Open Sea and Gene Body. The Greater promoter window is fixed in relation to the TSS. Therefore, the location of CpG Islands will influence the number of significant sequencing reads and 450K probes per gene that are included in our analysis; **(B)** Workflow for identification of differential methylation in the greater promoter of HNSCC, using next-generation MBD-sequencing and 450K methylation platforms. Schematic description of the analytic pipeline developed to unveil the HNSCC methylome. This pipeline enriches for mean genome-wide differences in CpG methylation as also for genome-wide differences in CpG methylation variability at each chromosomal location, for both, methylation sequencing and methylation 450K array data.



**Figure 3.** Integrative analysis of co-localized promoter methylation and somatic mutations with concurrent expression changes. (A) Methylated genes with fold change differences in expression greater than 2; (B) Genes with co-localized promoter methylation and somatic mutations in HNSCC. Methylation frequency is represented by the red color. Mutation frequency is represented by the blue color. The purple color represents the combined frequency of methylation and mutation events in the Discovery cohort.

that pathways involved in immune system development, and cell differentiation, proliferation, growth and renewal are impacted by concurrent epigenetic and genetic alterations in HNSCC (Fig. S6).

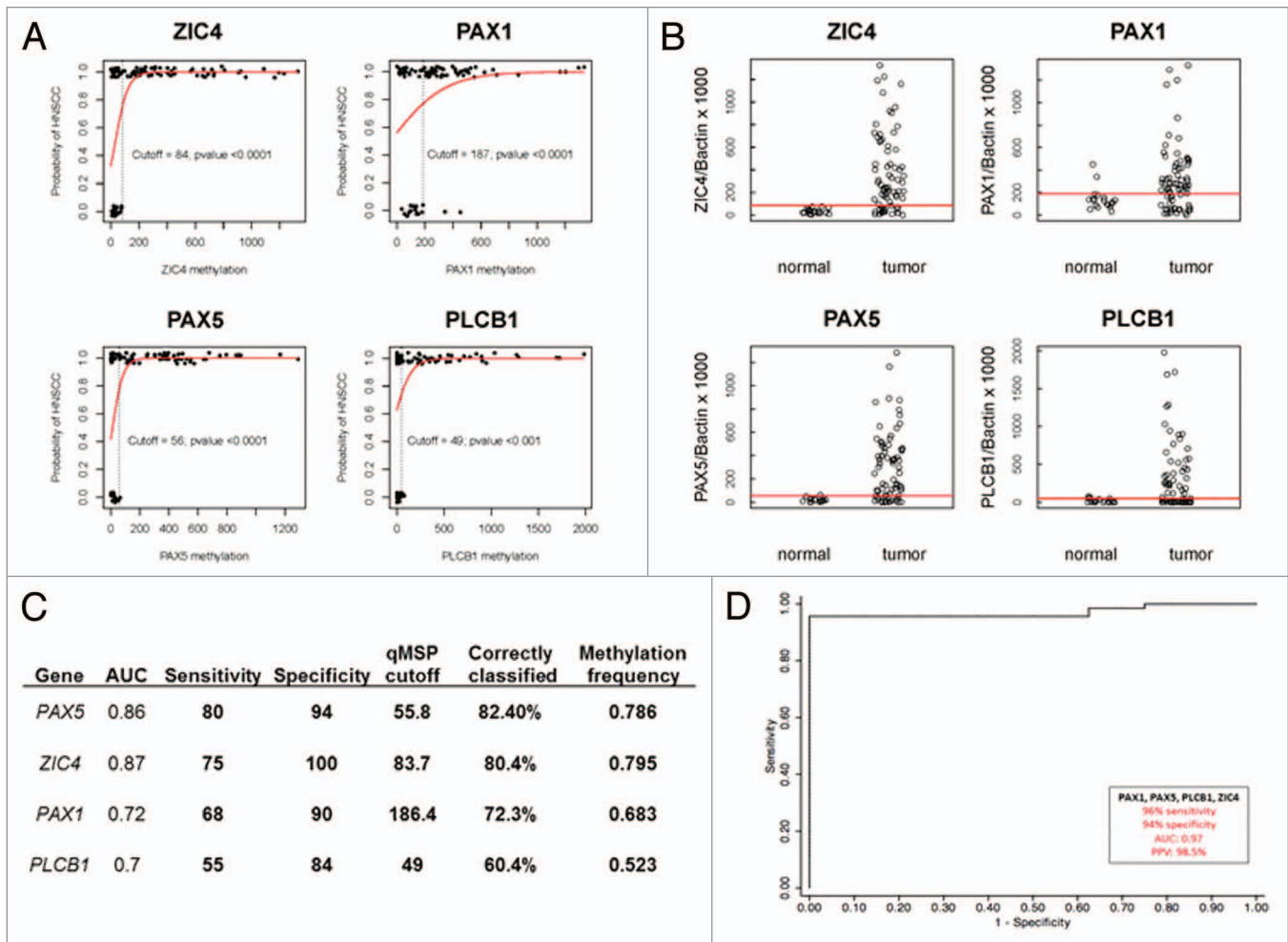
#### Methylation-Mutations Interactions with Risk Factors in the Discovery Cohort

When attempting to search for any associations between the mutated and/or methylated genes and HNSCC risk factors (TP53 mutations, HPV status, and smoking history) in the Discovery Cohort, we observed interesting patterns for the *PAX1* and *PAX5* genes. *PAX1* was methylated in all HPV negative tumors whereas zero methylation events of this gene were observed in HPV positive tumors. *PAX1* was also methylated in most patients with a history of tobacco exposure (71%), while only 33% of patients without tobacco exposure history exhibited *PAX1* methylation. Most HPV negative tumors (83%) showed *PAX5* methylation compared with 25% of HPV positive tumors. On the contrary, tumors from patients with a history of tobacco exposure (57%) had similar frequency of *PAX5* methylation to patients with no smoking history (67%). We also observed concurrent genomic and epigenomic associations with viral and tobacco exposures. Patients with *TP53* mutations also had *PAX1*

promoter methylation, history of tobacco exposure, and were HPV negative. Most (83%) of the patients with *TP53* mutations had evidence of *PAX5* methylation (Fig. S7).

#### Validation of *PAX1*, *ZIC4*, *PLCB1*, and *PAX5* promoter methylation with quantitative methylation specific PCR and TCGA data

We performed qMSP for 3 genes, *PAX1*, *ZIC4*, and *PLCB1*, in 76 tumor samples previously used by Agrawal in a HNSCC deep sequencing study<sup>1</sup> and 19 UPPP samples (Table S7 provides qMSP primers and probes information). All 3 genes appeared on our top ten list of TSG inactivated by both mechanisms and we confirmed a high frequency of tumor specific methylation in the Validation cohort. *PAX1* was methylated in 68% of the cancer cases, *ZIC4* in 80% and *PLCB1* in 52%. We then tested *PAX5* because it is in the *PAX* gene family and appeared on the list of the most frequently inactivated genes by promoter methylation and downregulation of expression. *PAX5* was methylated in 77% of the HNSCC cases in the validation cohort, a number very similar to the 70% methylation frequency identified in the discovery set. We found that *PAX1* ( $P < 0.0001$ ), *ZIC4* ( $P < 0.0001$ ), *PLCB1* ( $P < 0.001$ ), and *PAX5* ( $P < 0.0001$ ) methylation distinguished tumor from UPPP samples (Figs. 4A and B).



**Figure 4.** qMSP results for *PAX1*, *PAX5*, *ZIC4* and *PLCB1*. (A) Graphical expression of the logistic regression,  $\Pr(HNSCC = 1) = \text{logit}^{-1}(b_0 + b_1 \times \text{methylation})$  in tissue from 76 participants with data overlay. The predictor *methylation* is the qMSP value for each case (1) and each control (0). Cutoff methylation values for *PAX1*, *PAX5*, *ZIC4* and *PLCB1* are shown by the vertical dotted line. Probability of HNSCC is shown in red; (B) Scatterplots of quantitative MSP analysis of candidate genes promoters in the Validation screen cohort, which consisted of 76 HNSCC tumor tissue samples and 19 normal tissue samples obtained from uvulopharyngopalatoplasty (UPPP) procedures performed in non-cancer patients. The relative level of methylated DNA for each gene in each sample was determined as a ratio of MSP for the amplified gene to ACTB and then multiplied by 1000 [(average value of duplicates of gene of interest / average value of duplicates of ACTB) x 1000] for *PAX1*, *PAX5*, *ZIC4*, and *PLCB1*. Red line denotes cutoff value; (C) Sensitivity, Specificity and AUC results for qMSP analysis; (D) Receiver Operator Characteristics (ROC) curve for promoter methylation of *PAX1*, *PAX5*, *ZIC4* and *PLCB1* genes in the validation cohort. The figure shows that for this four gene panel the qMSP results have 96% sensitivity, 94% specificity, a 0.97 AUC and a Positive Predictive Value of 98.5%.

Receiver Operator Characteristic (ROC) curve analysis revealed that *PAX1* had 68% sensitivity, 90% specificity and a 0.72 AUC; *ZIC4* had 73% sensitivity, 100% specificity, and a 0.87 AUC; *PLCB1* had 55% sensitivity, 84% specificity, and a 0.70 AUC; and *PAX5* had 80% sensitivity, 94% specificity, and a 0.86 AUC (Fig. 4C). A gene panel combining promoter methylation results for these four genes had 96% sensitivity, 94% specificity, a 0.97 AUC, and a Positive Predictive Value of 98.5% (Fig. 4D). A chi-square test of independence revealed an association between methylation in *PLCB1* and tumor site,  $P < 0.01$ . Tumors of the oral cavity and oropharynx were the most frequently methylated.

All samples harboring *CDKN2A* mutations had *PAX1* methylation ( $P < 0.0001$ ) as did most of *TP53* mutated samples ( $P < 0.01$ ). More than half of *NOTCH1* (61.5%,  $P < 0.0001$ ) mutated

samples also exhibited *PAX1* promoter methylation. All the samples with mutations in *FBXW7* ( $P < 0.0001$ ), and most of the samples with mutations in *TP53* (79%,  $P < 0.0001$ ), and *NOTCH1* (92%,  $P < 0.0001$ ) were methylated in the *PAX5* promoter even after controlling for HPV-status and history of tobacco use (Table S8A).

We further corroborated our qMSP and somatic mutation results by analyzing The Cancer Genome Atlas (TCGA) publicly available data from 279 HNSCC patients (<https://tcga-data.nci.nih.gov>). *PAX5* promoter methylation was associated with *TP53* mutations ( $P = 0.02$ ), while *PAX1* promoter methylation was associated with *NOTCH1* mutations ( $P < 0.0001$ ), even after controlling for HPV-status and tobacco use. This evidence suggests a frequent occurrence of previously unreported interactions

between *PAX1* and *PAX5* promoter methylation and exonic mutations in *NOTCH1* and *TP53* in HNSCC, respectively (Table S8B).

To provide additional evidence of expression downregulation of *PAX1* and *PAX5* in HNSCC we compared mRNA expression in HNSCC and UPPP samples. Differential transcript levels for *PAX1* and *PAX5* were confirmed by quantitative RT-PCR in some of the RNA samples used for microarray analysis and RNA samples from an independent set (Fig. S8A and B). Expression levels were studied in 13 HNSCC and 17 UPPP samples that were readily available. The relative expression levels showed consistency with the results obtained for *PAX1* and *PAX5* in the Discovery cohort with genome-wide methylation and mRNA expression platforms.

We also performed *PAX5* knock-in and knock out studies in *p53* wild type (*p53wt*) and *p53* mutated (*p53mut*) HNSCC cell lines to assess the role of *PAX5* as a tumor suppressor gene connected to the *p53* pathway in HNSCC. We studied the functional consequences of *PAX5* induction in 022(*p53wt*) and 22A(*p53mut*) HNSCC cell lines. Following transfection with (Myc-DDK-tagged)-Human paired box 5, 022 and 22A cells exhibited a dramatic decrease in cell proliferation and *PAX5* expression levels were significantly increased, when measured 48h after transient transfection (Fig. S9).

Conversely, 22B(*p53mut*) HNSCC cells showed modest increase in cell proliferation compared with the control, when *PAX5* was inhibited by siRNA following transfection with *PAX5* siRNAs. Expression levels were significantly decreased, 48h after transient transfection (Fig. S10).

#### Proposed pathway interplay

After gene set enrichment analysis revealed that concurrent promoter methylation, mutations, and differential gene expression impacted cell differentiation, proliferation, growth and renewal pathways we performed a review study of the most important, potentially impacted cancer pathways in HNSCC. We found evidence of a strong interplay between somatic mutations in *p53* and *NOTCH1* and gene downregulation associated to *PAX1* and *PAX5* promoter methylation in HNSCC.

#### *p53-PAX5*

Our literature search revealed that *p53-PAX5* interactions are implicated in apoptotic and/or proliferating signals. *p53* is a downstream target for *PAX* proteins.<sup>26,27</sup> The human *p53* gene harbors a *PAX* binding site within its un-translated first exon that is conserved throughout evolution, which suggests the importance of this interaction. Frequent promoter methylation of *PAX5* has been reported in ductal carcinoma in situ, invasive breast cancer, and neuroendocrine carcinomas.<sup>28,29</sup> Furthermore, *PAX5* has been reported to function as a tumor suppressor gene in hepatocellular carcinoma<sup>30</sup> and gastric cancer,<sup>31</sup> directly binding to the *p53* promoter (Fig. S11). *PAX5* also plays an important role in the commitment of lymphoid progenitors to the B lymphocyte lineage.<sup>32</sup> The mechanism through which *PAX5* acts in B-cell differentiation is well established.<sup>33-35</sup> Recent studies identified some of these interactions also in solid tumors, but little has been shown so far (Fig. 5A).<sup>36-38</sup> In our Discovery cohort we identified high frequency of *PAX5* promoter methylation in HNSCC,

coinciding with low expression levels, a finding that supports a TSG function.

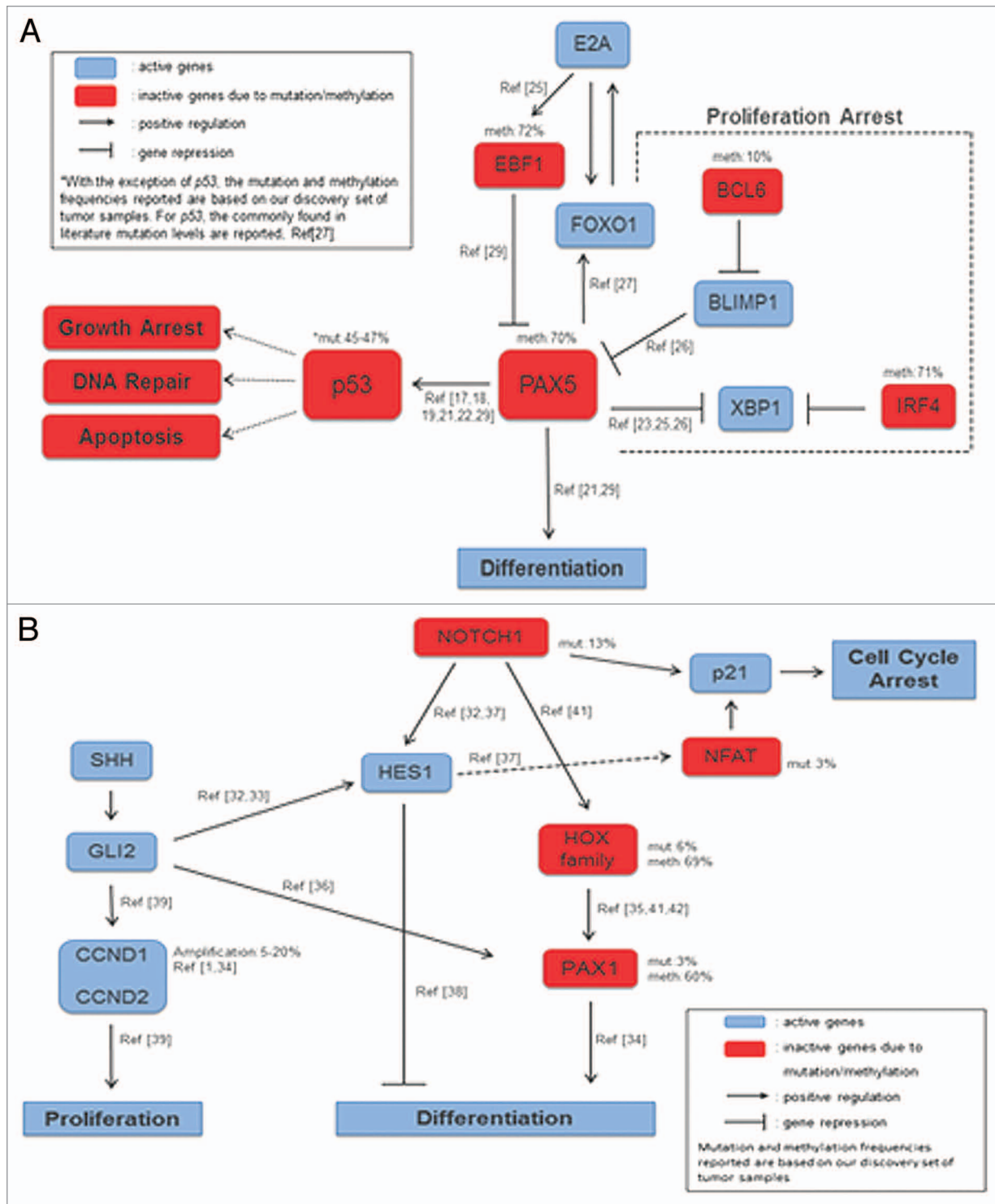
#### *NOTCH1-PAX1*

Agrawal et al. revealed inactivating mutations in *NOTCH1* gene, depicting its importance in HNSCC, proposing also a tumor suppressor function in this particular type of tumors.<sup>1</sup> When *NOTCH1* acts as a TSG it inhibits proliferation and promotes entry to differentiation.<sup>39</sup> Targeting the Notch proliferation pathway is really difficult. Preliminary data in cell lines showed that downregulation of *NOTCH1* in *NOTCH1* mutant cell lines leads to a modest effect of cell cycle acceleration and anti-*NOTCH1* agents are only effective in *NOTCH1* wild type cell lines.<sup>40</sup> In our study, *PAX1* was found to be the most frequently methylated and downregulated gene. In addition to that, several *HOX* family genes, some of which are known to interact with *NOTCH* signaling, were prominent in our list of methylated genes in HNSCC (Fig. S12). Our evidence suggests an interaction between *NOTCH1* and *PAX1* through the *HOX* family of transcription factors<sup>39,41-43</sup> and also with the Hedgehog pathway through *Hes1*, in which the well-established *CCND1* amplification plays an important role.<sup>44-48</sup> *PAX1* plays a role in sclerotome differentiation and has been shown to interact with homeobox genes which play a prominent role in normal development and the control of cell proliferation.<sup>49</sup> Retinoid acid (RA) signaling acts via Hox gene pathways,<sup>50,51</sup> some of which are able to regulate *PAX1* through canonical *NOTCH1* expression. These interactions are described in Figure 5B.

## Discussion

We have conducted the first comprehensive integrated genomic and epigenomic analysis in HNSCC, focusing on identifying TSG genes that demonstrate concurrent promoter methylation with downregulation of expression and somatic mutations. Recent studies in HNSCC, published as we were finalizing this manuscript, focused on therapeutic pathways affected by somatic mutations and copy number alterations,<sup>52-54</sup> but only described the clustering effects of DNA methylation at a global genome wide level.<sup>52</sup> We performed the first detailed genome wide analysis of the HNSCC methylome that studied expression alterations associated with differential methylation patterns in the greater promoter region, focusing on key TSGs that are also inactivated by somatic mutations. The main focus of this paper was to identify the number of tumor suppressor genes differentially methylated in the greater promoter region and mutated in HNSCC, and examine their combined impact in expression downregulation. We controlled for chromosomal deletions using CNV data previously generated in the Discovery cohort, to distinguish if expression alterations were related to methylation events or to deletions in specific regions of the chromosome.

Many genes with dense promoter methylation and downregulation of expression are candidate TSGs. However, genes inactivated by both somatic mutations and promoter methylation are likely to be key drivers of oncogenesis. Our analysis identified 10 downregulated genes inactivated by somatic mutation and



**Figure 5.** Proposed genomic and epigenomic interactions in HNSCC: **(A)** Proposed partial pathway interplay of *p53* and *PAX5* in HNSCC. Downregulation of *PAX5* leads to differentiation. When methylated, *PAX5* an upstream target of *p53*, fails to activate the later which is also silenced due to mutations, and thus DNA repair, Apoptosis, and Growth Arrest pathways are inactive; **(B)** *PAX1*-*NOTCH1* interplay through crosstalk of Hedgehog and Notch pathways in cell differentiation and proliferation signals. *Notch1* induces *p21* expression, either directly through the canonical pathway or indirectly through *Hes1* and *NFAT* activation, leading in both cases to cell cycle arrest. Active *Notch1* targets either the *Hox* family or *Hes1*. *Hes1* is active and will block differentiation. The *HOX* family of transcription factors, downstream targets of Notch signaling, is frequently silenced, thus blocking the activation of *PAX1* which is also downregulated in HNSCC and will not promote differentiation. *PAX1* expression can also be induced by *Shh* through *Gli2*, which is active. Finally, proliferation is promoted through *Gli2* interaction with *CCND1* and *CCND2*.

promoter methylation. We selected *PAX1*, *ZIC4*, *PLCB1*, and *PAX5* for further validation and we were indeed able to detect stark differences in DNA methylation levels between cancer and UPPP samples. *ZIC4* and *PLCB1* were the genes with the lowest methylation frequency in HNSCC, exhibiting also low mutation frequency, and yet, we confirmed *ZIC4* and *PLCB1* promoter methylation differences between normal and cancer samples in the Validation cohort.

*PAX1* and *PAX5* were highlighted as key genes in this study as they belong to the same family of transcription factors and were both found to be methylated and downregulated in HNSCC. *PAX1* and *PAX5* are genes involved in differentiation/proliferation signals and the gene set enrichment analysis performed clearly depicted these signaling pathways as deregulated in HNSCC in our discovery set of tumors.

We observed interesting relationships among the most commonly mutated and methylated genes in HNSCC: 61.5% of the *NOTCH1* mutated samples also exhibited *PAX1* methylation and 79% of the samples carrying *TP53* mutations were also methylated in the *PAX5* gene promoter. External validation in 279 primary HNSCC samples from the TCGA project verified our initial findings. Combined, this evidence suggests the frequent occurrence of previously unreported interactions between *PAX1* and *PAX5* promoter methylation and exonic mutations in *NOTCH1* and *TP53* in HNSCC.

The greater promoter *PAX1* and *PAX5* methylation levels reported in this manuscript were obtained with three different platforms that use diverse technology, chemistry, sample preparation and data analysis pipelines to obtain methylation values. We have combined these three platforms into a very robust methylation detection pipeline.

*PAX1* and *PAX5* methylation levels listed in the Discovery cohort are both, from MBD-seq and the 450K BeadChip assay. The 450K results were used to confirm the MBD-seq results in the subset of ten Discovery samples that were sequenced and, by proxy, validate the methylation results of the other 22 Discovery cohort samples that were only queried with the 450K arrays. The levels of *PAX1* and *PAX5* methylation levels listed for the Prevalence cohort were obtained with quantitative methylation specific PCR (qMSP). The levels of *PAX1* and *PAX5* listed for TCGA were obtained with the 450K BeadChip assay. The difference in *PAX1* methylation between HPV positive and HPV negative tumors was significant in the Discovery cohort. The difference between the two in the Prevalence cohort was not significant ( $P = 0.225$ ), perhaps due to the small sample size of HPV positive patients.

*PAX* genes, a family of nine transcription factors which act as cell lineage specific regulators of the tissues where they are normally expressed, are now also recognized as important factors in cancer progression. *PAX* genes, similarly to the *NOTCH* gene family, may play previously unrecognized fundamental roles in balancing proliferation and differentiation signals, two conceptually opposite cellular processes in canonical cancer research. The *PAX* family of genes may ultimately, following Waddington's epigenetic landscape metaphor, be part of an epigenomic mediated switch between cancer initiation and cancer maintenance

pathways, which stochastically drive cancer progression, immune system avoidance, acquisition of tumor resistance, and establishment of metastatic disease. Loss of *NOTCH1* function due to mutation, or mutation/methylation-dependent silencing of downstream genes, such as *PAX1* or the *HOX* family genes, is likely to abrogate normal cell differentiation.

Clinically, *TP53* mutation has been shown time and again to be among the worst molecular alterations in patients with HNSCC.<sup>36,55</sup> Patients that harbor p53 mutant tumors are more likely to relapse after complete resection and radiation therapy.<sup>5</sup> We confirmed a high frequency of *PAX* methylated tumors in 279 HNSCC tumor samples from the TCGA cohort and found that tumors which already harbor a p53 mutation, also harbor *PAX5* promoter methylation.

Together, our results support the notion that differential promoter methylation and somatic mutations are the main cause of gene inactivation and pathway deregulation in HNSCC. We have unveiled hitherto unknown interactions between mutated and methylated genes that are associated with gene expression alterations in HNSCC. Characterization of the complete HNSCC methylome has contributed insights into the clustering of specific genetic and epigenetic events in the greater promoter region and highlights the importance of understanding the relative contribution of each to the overall frequency of TSG inactivation. We plan future in vitro studies focused on the functional consequences of the inactivation of these high frequency genes and will further explore the above proposed gene interactions. Understanding the complete contribution of genomic and epigenomic alterations to specific genes and pathways in cancer will reveal novel high frequency specific markers for better risk assessment (as we observed in the TCGA cohort above) and will highlight the true frequency of therapeutic pathways to better target the disease at the molecular level.

## Materials and Methods

### Participants

#### *Patient selection*

Head and Neck Squamous Cell Carcinoma ( $n = 91$ ) and uvulopalatopharyngealplasty (UPPP) patients ( $n = 35$ ) were consented for this study at the Johns Hopkins Medical Institutions hospitals and MD Anderson Cancer Center. The study was approved by the Ethics Committee of each participating hospital, as well as by the Johns Hopkins Institutional Review Board.

#### Johns Hopkins component

Fresh-frozen surgically resected tissue and matched blood were obtained from patients at Johns Hopkins Medical Institutions in Baltimore. Tissue was analyzed by frozen section histology to estimate neoplastic cellularity. In order to enrich the samples for neoplastic cells, normal tissue was removed from the samples using macro-dissection based on the frozen section histology. HPV tumor status was determined for oropharyngeal tumors per standard clinical care using in situ hybridization. Hybridization was performed using the HPV III Family16 probe set that captures HPV genotypes 16, 18, 33, 35, 45, 51, 52, 56, and 66.

HPV16-positive controls included an HPV16-positive oropharyngeal cancer and the SiHa and CaSki cell lines. HPV tumor status was also determined by E6/E7 PCR primer amplification.

#### MD Anderson Component

Fresh-frozen surgically resected tumor and matched non-malignant adjacent tissue were obtained from consented patients treated for HNSCC at the University of Texas MD Anderson Cancer Center, under an Institutional Review Board approved protocol. Frozen tissue was embedded in optimal cutting temperature compound and cryosections from the top and middle of specimens were stained with hematoxylin and eosin prior to being evaluated by a pathologist for the presence of >60% tumor nuclei content or absence of tumor (i.e., normal). Samples that passed this criterion were sectioned all the way through and washed once in PBS prior to isolating genomic DNA using an ArchivePure DNA purification kit.

#### **Methylated binding domain sequencing (MBD-seq)**

##### *Preparation of libraries*

Tissue samples were digested with 1% SDS and 50  $\mu\text{g}/\text{mL}$  proteinase K (Boehringer Mannheim) at 48 °C overnight, followed by phenol/chloroform extraction and ethanol precipitation of DNA. Two micrograms of DNA were sonicated to a modal size of ~150–250 bp, and end-repaired using the NEBNext SOLiD DNA library preparation kit end-repair module following the manufacturer's protocol (New England Biolabs). After column-purification (using the Qiagen PCR purification kit), SOLiD P1 and P2 adapters lacking 5' phosphate groups (Life Technologies) were ligated using the NEBNext adaptor ligation module and column-purified, and subjected to nick-translation by treating with Platinum Taq polymerase to remove the nick.

##### *Affinity enrichment and capture of methylated DNA fragments*

The resulting library was divided into two fractions, a total input fraction, and an enriched methylated fraction. The enriched methylated fraction was then subjected to affinity enrichment of methylated DNA fragments by using 6xHis-MBD2-MBD polypeptides immobilized on magnetic beads as described previously.<sup>56,57</sup> The resulting enriched methylated fraction and the total input fraction were then subjected to library amplification using the NEBNext amplification module according to the manufacturer's protocols, using 4–6 cycles for the total input, and 10–12 cycles for the enriched methylated fractions. Library fragments that were between 200–300 bp were size selected after agarose gel electrophoresis.

##### *Massively parallel sequencing of MBD-seq libraries*

The libraries were then subjected to emulsion PCR and bead enrichment following the SOLiD emulsion PCR protocol (Life Technologies). The resulting beads were then deposited on the SOLiD flow cell and subjected to massively parallel 50 bp single-read sequencing on a SOLiD v4.0 sequencer octet segment, with one octet segment for the total input and another one for the enriched methylated fraction. The details of the sequencing output for each sample from the sequencing run (number of tags, coverage, etc.) are provided in Table S9. Reads were aligned to hg19 using default settings in bioscope v1.3, with the exception of the bam output method, which was changed to alignment score.

#### **Bioinformatics analysis of MBD-seq data**

##### *MACS analysis and identification of differential methylation*

For the purposes of the analysis we divided the genome into two broad regions: the greater promoter, was defined as the region encompassing 6000 bases upstream and 1500 bases downstream from the transcription start site (TSS). From the functional genome distribution standpoint the greater promoter region, includes CpG sites in the proximal promoters, 1500 bases upstream from the described TSS, and 1500 bases downstream from the TSS, in the 5' untranslated region and exon1. From the CpG content and neighborhood context the differentially methylated CpGs in HNSCC were located in CpG islands, CpG shores (regions 2000 bp upstream and downstream of but not inside CpG islands), CpG shelves (regions 2000 bp upstream and downstream of but not inside the shores), or as isolated CpGs in the area of the genome now defined as Open Sea. Methylated regions were identified as peaks of aligned sequencing tags in the enriched compared with total input fraction using MACS v1.4 software,<sup>22,58,59</sup> which allows identification of peaks after accounting for both global and local biases using the total input fraction. Peaks of methylation were identified for each sample separately.

To identify differentially methylated regions we first used stringent parameters to define presence and absence of methylation in EACH sample as follows: we used a low MACS  $P$  value cut-off ( $P < 10^{-6}$ ) to identify regions that are methylated, and another cut-off (MACS  $P > 10^{-2}$ ) to identify those regions that have very little evidence for methylation.

Next, for any given comparison of group A vs. group B (e.g., Group A = all tumors; Group B = all normals), we identified all regions that showed absence of methylation in all samples of Group A and presence of methylation in at least one sample from group B. All such overlapping regions across samples with peak calls in Group B were then merged, and the number of such samples and the lowest p-value of the peaks for these samples were recorded as the aggregate differentially methylated region. This analysis therefore yields regions in which all samples in Group A have absence of methylation peaks, and at least one sample in Group B has a methylation peak. The converse comparisons (i.e., absence of methylation in Group B, with at least one sample having a methylation peak in Group A) were also performed to obtain regions showing both gain of- and loss of-methylation.

##### *Methylation bump hunting for identification of differentially methylated regions*

Differential methylation was also identified using an independent approach called bump hunting that has been previously used to identify differential peaks in methylation data. Methylation bump hunting is a data analysis pipeline that effectively models measurement error, removes batch effects, detects regions of interest and attaches statistical uncertainty to regions identified as differentially methylated.<sup>23</sup> These methods are implemented in the *bumphunter* Bioconductor package and described in more detail in the section "Bioinformatics for 450K data." Reported functionally relevant findings have been generally associated with genomic regions rather than single CpGs, either CpG islands,<sup>60</sup> CpG island shores,<sup>20</sup> genomic blocks,<sup>61</sup> or generic 2-kb regions.<sup>62</sup>



Epigenomic bumps may have greater variability in size and shape than MBD-seq peaks.

#### *Integration of MACS and bumphunting results*

To identify the promoter regions that are differentially methylated in HNSCC when compared with normal oral mucosa, we intersected the list of methylated probes that discriminated between tumor and normal tissue identified with MACS with the list of methylated regions that discriminated between tumor and normal tissue identified with bumphunting. We used R (v3.00) to analyze the correlation of methylated promoter regions and HNSCC etiological factors.

#### **Verification of MBD-seq results with HumanMethylation450K DNA BeadChip assay**

##### *450K array description and sample preparation*

The 450K is a two-color array that detects cytosine methylation at 485,512 methylation loci, mostly at CpGs, but also at a small number of cytosine residues outside of the CpG context, using bisulfite converted DNA. For each individual CpG two different signals are detected. One signal measures the amount of methylated DNA (Meth) and the other one measures the amount of unmethylated DNA (Unmeth). The Meth and Unmeth signals are measured with two different assays called “Type I” design or a “Type II” design. A  $\beta$  ( $\beta$ ) value is generated by both Type I and Type II design probes to denote the methylation level of the CpG loci using the ratio of intensities between Meth and Unmeth ( $\beta$  value = methylation intensity / methylation + unmethylated intensity of the given CpG locus). Each methylation locus is interrogated by one of these designs. For a type I locus the Meth and Unmeth signals are measured by two paired probes, with a given locus using either the red or green signal from these probes. Type II loci are assayed using a single probe, with Meth and Unmeth signals derived from the green and red channels respectively. In addition to the methylation loci, the array contains a small number of control probes and 65 probes measuring common SNPs, intended for sample tracking. Type II probes use only one probe per methylation locus and hence allows more loci on the array, at a fixed array size. However, due to the chemistry used by the type II probe design, type II probes can only tolerate up to three CpGs within the 50bp probe. The type I design tolerates more CpGs within the 50bp probe, but assumes that all methylation loci in the probed sequence are in the same state.<sup>63</sup>

Bisulfite modification of genomic DNA (2  $\mu$ g) was performed with EpiTect Bisulfite Kit (QIAGEN) according to the manufacturer’s protocol. We hybridized bisulfite converted DNA from normal (UPPP) tissue ( $n = 16$ ) and Head and Neck Squamous Cell Carcinoma (HNSCC) tissue ( $n = 31$ ) samples to the 450K array.

##### *Bioinformatics for 450K data*

Bioinformatics strategies were used for background correction, normalization, and data analysis of differentially methylated genomic regions between tumor, and normal tissue. As with the analysis of methylation sequencing data we used two different analytic pipelines: a pipeline designed to capture the variability in methylated signals across the arrays using an F-test<sup>11</sup> and the bumphunting pipeline.<sup>23</sup> We used the *minfi* and *bumphunter* packages found in Bioconductor to perform background

correction, normalization, and data analysis of differentially methylated genomic regions between tumor and normal tissue. The *minfi* package provides tools for analyzing Illumina’s Methylation arrays, with a special focus on the new 450k array for humans, and includes methods for preprocessing, quality assessment, and detection of differentially methylated regions from the kilobase to the megabase scale.<sup>64</sup> The *bumphunter* package is meant to work on data with several biological replicates, similar to the *lmFit* function in *limma*. While *bumphunter* is written using genomic data as an illustrative example, most of it is generalizable to other data types (with some one-dimensional location information).

##### F-test analytic pipeline

The selection of significantly methylated CpGs in the Illumina 450K Infinium assay data was performed in a stepwise manner. The 450.idat files were preprocessed and background corrected with the *preprocessIllumina* function in *minfi* to obtain  $\beta$  values. An F-test was performed across all 47 samples to identify CpGs with a significant difference in  $\beta$  values between normal and malignant tissue. Since the empirical  $P$  values were calculated genome-wide, adjustment for multiple testing was performed. Q-values were computed from the empirical  $P$  values using the Benjamin and Hochberg correction. Probes with q-values less than 0.05 were deemed statistically significant and were included in the CpG list. We then selected only those CpGs that showed a methylation difference of at least 0.25 between cancer and normal tissues and a  $\beta$  value of at least 0.3 in cancer, as previously described.<sup>11</sup> All bioinformatics analyses were performed using R version 3.0.

##### 450K-bumphunting analytic pipeline

We performed pre-processing with the *minfi* package, which applies a version of subset quantile normalization to the Meth and Unmeth intensities separately. The distribution of type I and type II probes is forced to be the same by first quantile normalizing the type II probes across samples and then interpolating a reference distribution to which the type I probes are normalized. For the probes on the X and Y chromosomes the males and females are normalized separately. Sex is determined by the *getSex* function using copy number information. The stratified quantile normalization method is implemented by the *preprocessQuantile* function (the function does no background correction and removes zeros using the *fix2MethOutlier* function). This algorithm relies on the assumptions necessary for quantile normalization and involves both within- and between- sample normalization.<sup>64</sup>

##### *Integration of F-test and 450K-bumphunting results*

To identify the promoter regions that are differentially methylated in HNSCC when compared with normal oral mucosa, we intersected the list of methylated probes that discriminated between tumor and normal tissue identified with the F-test with the list of methylated regions that discriminated between tumor and normal tissue identified with 450K-bumphunting.

##### *E. mRNA expression arrays*

Total RNA was isolated from normal (UPPP) tissue ( $n = 16$ ) and Head and Neck Squamous Cell Carcinoma (HNSCC) tissue ( $n = 16$ ) samples by using Tri-reagent. cDNA was made, and

hybridized to Affymetrix GeneST1.0 Arrays (Affymetrix) according to manufacturer's instructions. Six of these samples were also sequenced with MDB-seq. The data obtained from CEL files was background corrected with RMA, quantile normalized before an ANOVA was used to determine the Fold Change difference in log-transformed intensities between Tumor and Normal samples.

#### Analysis of functional annotation

Enrichment analysis of functional themes (Analysis of Functional Annotation, AFA) was performed to capture biological processes over-represented in the various conditions under investigation. This unbiased computational approach, conceptually similar to Gene Set Enrichment Analysis (GSEA),<sup>65</sup> enables the interpretation of genome-wide data through the identification and visualization of information encompassing distinct biological concepts, and was previously used successfully to integrate and interpret both differential gene expression and methylation data.<sup>25,66</sup> A chi-square test of independence was applied to test whether each Functional Gene Set (FGS) was over-represented in any of the gene list associated with any of the investigated contrasts/conditions (e.g., gene associated with methylated promoters in HNSCC). In the present study, individual, non-redundant genes, as annotated in the NCBI Entrez gene database (R/Bioconductor package *org.Hs.eg.db* version 2.4.6) were used as the total gene space, and contingency tables were used to identify gene sets over-represented in the investigated conditions. Correction for multiple hypothesis testing was obtained separately for each FGS collection, by applying the Benjamini and Hochberg method<sup>67</sup> as implemented in the mult-test R/Bioconductor package. Overall, this approach is analogous to Gene Set Enrichment Analysis (GSEA),<sup>65,68,69</sup> and has already been successfully applied in other studies.<sup>66,70</sup> The heat-maps' color bar represents the negative log<sub>10</sub> False Discovery Rate (FDR). For each gene set collection the sets for which at least one condition showed FDR < 0.01 were reported. The top 150 conditions were reported when too many gene sets were retrieved.

AFA was used to compare biological themes enriched in the following gene lists: (1) mutated genes; (2) genes with hyper-methylated promoters; (3) genes showing hyper-methylation outside the greater promoter region; (4) genes upregulated in cancer when compared with normal; and (5) genes downregulated in cancer when compared with normal. Gene set enrichment was assessed by testing for gene set over-representation with a chi-square test, because each gene list was obtained from diverse analyses using different methods. In the AFA analysis each gene was counted only once, while using the totality of the genes annotated to the NCBI Entrez gene database as the background gene space. For this reason there was no confounding with the number of probes or the length of the genes (i.e., each gene was counted only once irrespective to the number of probes or its length). The GO categories reported in Table S6 encompass different biological processes and contain a fairly large number of distinct genes. Such genes show variable length, from short to long transcripts, and all possible and disparate genomic arrangements, with genes located in both "gene-rich" and "gene-poor" genomic regions.

#### Validation of genomic and epigenomic alterations

##### *Quantitative methylation specific PCR (qMSP) in the prevalence cohort*

qMSP was used for validation in a Prevalence cohort of 76 tumors in which we had previously identified somatic mutations in *TP53*, *NOTCH1*, *CDKN2A*, *PIK3CA*, *FBXW7*, and *HRAS* and in 19 UPPP normal control tissue samples. We examined promoter methylation in three of the genes that were included in our final list of mutated and methylated genes and that were methylated in at least 40% of the Discovery Cohort samples: *PAX1*, *PAX5*, *PLCB1*, and *ZIC4*. Bisulfite-modified DNA was used as a template for fluorescence-based real-time PCR, as previously described.<sup>71</sup>

##### *Contingency tables of mutational and methylation events in TCGA data set*

Publicly available HNSCC Illumina 450K methylation and exome sequencing data was downloaded from the TCGA website (<http://cancergenome.nih.gov>) and the cBioPortal for Cancer Genomics ([www.cbioportal.org/public-portal/](http://www.cbioportal.org/public-portal/)) using R (v3.0.0). Publicly available exome mutation data for *TP53* and *NOTCH1* and  $\beta$  values for all *PAX1* and *PAX5* 450K array probes were extracted for all HNSCC samples analyzed by the TCGA project that had paired methylation and mutation data for the genes of interest ( $n = 279$ ). Only the 450K probes located in TSS1500, TSS200, and 1st exon, as per the manufacturer's annotation, were used to create the contingency tables for methylation and mutation analyses. Contingency tables were used to examine the association between exonic mutations of *TP53*, *CDKN2A*, *HRAS*, *FBXW7*, and *NOTCH1* and promoter methylation of *ZIC4*, *PAX1*, *PAX5*, and *PLCB1*. The MacNemar test for paired data was implemented in R (v3.0.0) to evaluate the association between mutations and promoter methylation.

##### *Quantitative real-time reverse transcription PCR*

HNSCC RNA samples from the Discovery cohort and from an independent cohort were assessed for *PAX5*, *PAX1*, and *GAPDH* expression levels using quantitative real-time reverse transcription (RT)-PCR (TaqMan). Reverse transcription was performed with random hexamer primers and Superscript II Reverse Transcriptase (Invitrogen Corp.) according to manufacturer's instructions. Quantitative RT-PCR was then performed on the Applied Biosystems 7900 Sequence Detection Instrument (Applied Biosystems) using TaqMan expression assays (Life Technologies).

##### Functional studies in cell lines

Human HNSCC cell lines with known *p53* status were cultured to determine *PAX5* methylation and expression levels. Human HNSCC cell lines 022 (*p53* wt), 22A (*p53* mut), and 22B (*p53* mut) were selected for functional studies based on their expression and methylation results (Fig. S13). Cell growth conditions were maintained at 37 °C in an atmosphere of 5% CO<sub>2</sub>.

##### Transient transfection, PAX5 inhibition or overexpression, and cell proliferation assay

We knocked down *PAX5* in the 22B cell line using the ON-TARGETplus Pool of siRNAs against *PAX5* and a non-targeting Pool of siRNAs (Thermo Scientific) as control. Cells were seeded in 96-well plates and allowed to grow until approximately

70% confluence. Cells were transfected with siRNA using Lipofectamine RNAiMAX Reagent (Invitrogen) and cell metabolic activity was determined every 24 h using the CCK-8 colorimetric assay (Dojindo). Values are mean  $\pm$  SEM for pentaplicates of cultured cells. The transfection efficiency was confirmed by quantitative real-time RT-PCR as described above and normalized to *GAPDH* at 48-h time point. Pax5 knockdown in 22B cells showed a modest increase in cell proliferation compared with the control.

Forced expression of *PAX5* in 022 and 22A cell lines was performed. The pCMV-Entry-EV (control) and the (Myc-DDK-tagged)-Human paired box 5 (*PAX5*) from ORIGENE were transfected into cells (FUGENE HD Promega) which were then seeded in 96-well plates and allowed to grow until approximately 70% confluence. Cell metabolic activity was determined every 24 h using the CCK-8 colorimetric assay (Dojindo). Values are mean  $\pm$  SEM for pentaplicates of cultured cells. The transfection efficiency was confirmed by quantitative real-time RT-PCR as described above and normalized to *GAPDH* at the 48-h time point. 022 and 22A cells exhibited a dramatic decrease in cell proliferation after forced expression of *PAX5* 48h after transfection compared to controls.

#### Disclosure of Potential Conflicts of Interest

J.A.C. is the Director of Research of the Milton J. Dance Head and Neck Endowment. V.E.V. is a co-founder of Inostics

and Personal Genome Diagnostics and is a member of their Scientific Advisory Boards. V.E.V. owns Inostics and Personal Genome Diagnostics stock, which is subject to certain restrictions under University policy. The Johns Hopkins University in accordance with its conflict of interest policies is managing the terms of this arrangement.

#### Acknowledgments

The authors wish to thank Rafael Irizarry for his supervision in our use and interpretation of *minfi* and *bumphunter* analytic platforms. We also want to thank A. Jaffe, M. Haffner, T. Mosbrugger, S. Wheelan, L. Danilova, C. Talbot and A. Jedlicka for their expert assistance.

#### Financial Support

National Cancer Institute grants U01CA84986 and K01CA164092 and CA121113; National Institute of Dental and Craniofacial Research grants P50DE019032 Head and Neck Cancer SPORE, and RC2 DE20957, and the Commonwealth Fund supported this research. The funders have no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

#### Supplemental Materials

Supplemental materials may be found here: [www.landesbioscience.com/journals/epigenetics/article/29025](http://www.landesbioscience.com/journals/epigenetics/article/29025)

#### References

1. Agrawal N, Frederick MJ, Pickering CR, Bettegowda C, Chang K, Li RJ, Fakhry C, Xie TX, Zhang J, Wang J, et al. Exome sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in NOTCH1. *Science* 2011; 333:1154-7; PMID:21798897; <http://dx.doi.org/10.1126/science.1206923>
2. Stransky N, Eglöf AM, Tward AD, Kostic AD, Cibulskis K, Sivachenko A, Kryukov GV, Lawrence MS, Sougnez C, McKenna A, et al. The mutational landscape of head and neck squamous cell carcinoma. *Science* 2011; 333:1157-60; PMID:21798893; <http://dx.doi.org/10.1126/science.1208130>
3. Westra WH. The changing face of head and neck cancer in the 21st century: the impact of HPV on the epidemiology and pathology of oral cancer. *Head Neck Pathol* 2009; 3:78-81; PMID:20596995; <http://dx.doi.org/10.1007/s12105-009-0100-y>
4. Kim GB, Wang Z, Liu YY, Stavrou S, Mathias A, Goodwin KJ, Thomas JM, Neville DM. A fold-back single-chain diabody format enhances the bioactivity of an anti-monkey CD3 recombinant diphtheria toxin-based immunotoxin. *Protein Eng Des Sel* 2007; 20:425-32; PMID:17693455; <http://dx.doi.org/10.1093/protein/gzm040>
5. Poeta ML, Manola J, Goldwasser MA, Forastiere A, Benoit N, Califano JA, Ridge JA, Goodwin J, Kenady D, Saunders J, et al. TP53 mutations and survival in squamous-cell carcinoma of the head and neck. *N Engl J Med* 2007; 357:2552-61; PMID:18094376; <http://dx.doi.org/10.1056/NEJMoa073770>
6. Ang KK, Harris J, Wheeler R, Weber R, Rosenthal DI, Nguyen-Tân PF, Westra WH, Chung CH, Jordan RC, Lu C, et al. Human papillomavirus and survival of patients with oropharyngeal cancer. *N Engl J Med* 2010; 363:24-35; PMID:20530316; <http://dx.doi.org/10.1056/NEJMoa0912217>
7. Kelloff GJ, Lippman SM, Dannenberg AJ, Sigman CC, Pearce HL, Reid BJ, Szabo E, Jordan VC, Spitz MR, Mills GB, et al.; AACR Task Force on Cancer Prevention. Progress in chemoprevention drug development: the promise of molecular biomarkers for prevention of intraepithelial neoplasia and cancer--a plan to move forward. *Clin Cancer Res* 2006; 12:3661-97; PMID:16778094; <http://dx.doi.org/10.1158/1078-0432.CCR-06-1104>
8. Myers MF, Chang MH, Jorgensen C, Whitworth W, Kassim S, Litch JA, Armstrong L, Bernhardt B, Faucett WA, Irwin D, et al. Genetic testing for susceptibility to breast and ovarian cancer: evaluating the impact of a direct-to-consumer marketing campaign on physicians' knowledge and practices. *Genet Med* 2006; 8:361-70; PMID:16778598; <http://dx.doi.org/10.1097/01.gim.0000223544.68475.6c>
9. Whitworth A. New research suggests access, genetic differences play role in high minority cancer death rate. *J Natl Cancer Inst* 2006; 98:669; PMID:16705120; <http://dx.doi.org/10.1093/jnci/djj223>
10. Svatek RS, Lee JJ, Roehrborn CG, Lippman SM, Lotan Y. The cost of prostate cancer chemoprevention: a decision analysis model. *Cancer Epidemiol Biomarkers Prev* 2006; 15:1485-9; PMID:16896037; <http://dx.doi.org/10.1158/1055-9965.EPI-06-0221>
11. Guerrero-Preston R, Soudry E, Acero J, Orera M, Moreno-López L, Macía-Colón G, Jaffe A, Berdasco M, Ili-Gangas C, Brebi-Mieville P, et al. NID2 and HOKA9 promoter hypermethylation as biomarkers for prevention and early detection in oral cavity squamous cell carcinoma tissues and saliva. *Cancer Prev Res (Phila)* 2011; 4:1061-72; PMID:21558411; <http://dx.doi.org/10.1158/1940-6207.CAPR-11-0006>
12. Carvalho AL, Chuang A, Jiang WW, Lee J, Begum S, Poeta L, Zhao M, Jerónimo C, Henrique R, Nayak CS, et al. Deleted in colorectal cancer is a putative conditional tumor-suppressor gene inactivated by promoter hypermethylation in head and neck squamous cell carcinoma. *Cancer Res* 2006; 66:9401-7; PMID:17018594; <http://dx.doi.org/10.1158/0008-5472.CAN-06-1073>
13. Demokan S, Chang X, Chuang A, Mydlarz WK, Kaur J, Huang P, Khan Z, Khan T, Ostrow KL, Brait M, et al. KIF1A and EDNRB are differentially methylated in primary HNSCC and salivary rinses. *Int J Cancer* 2010; 127:2351-9; PMID:20162572; <http://dx.doi.org/10.1002/ijc.25248>
14. Andrades P, Asiedu C, Ray P, Rodriguez C, Goodwin J, McCarn J, Thomas JM. Islet yield after different methods of pancreatic Liberase delivery. *Transplant Proc* 2007; 39:183-4; PMID:17275501; <http://dx.doi.org/10.1016/j.transproceed.2006.10.016>
15. Settle K, Posner MR, Schumaker LM, Tan M, Suntharalingam M, Goloubeva O, Strome SE, Haddad RI, Patel SS, Cambell EV 3<sup>rd</sup>, et al. Racial survival disparity in head and neck cancer results from low prevalence of human papillomavirus infection in black oropharyngeal cancer patients. *Cancer Prev Res (Phila)* 2009; 2:776-81; PMID:19641042; <http://dx.doi.org/10.1158/1940-6207.CAPR-09-0149>
16. Leemans CR, Braakhuis BJ, Brakenhoff RH. The molecular biology of head and neck cancer. *Nat Rev Cancer* 2011; 11:9-22; PMID:21160525; <http://dx.doi.org/10.1038/nrc2982>
17. Morris LG, Kaufman AM, Gong Y, Ramaswami D, Walsh LA, Turcan Ş, Eng S, Kannan K, Zou Y, Peng L, et al. Recurrent somatic mutation of FAT1 in multiple human cancers leads to aberrant Wnt activation. *Nat Genet* 2013; 45:253-61; PMID:23354438; <http://dx.doi.org/10.1038/ng.2538>
18. Scholtens D, Vidal M, Gentleman R. Local modeling of global interactome networks. *Bioinformatics* 2005; 21:3548-57; PMID:15998662; <http://dx.doi.org/10.1093/bioinformatics/bti567>

19. Serre D, Lee BH, Ting AH. MBD-isolated Genome Sequencing provides a high-throughput and comprehensive survey of DNA methylation in the human genome. *Nucleic Acids Res* 2010; 38:391-9; PMID:19906696; <http://dx.doi.org/10.1093/nar/gkp992>
20. Irizarry RA, Ladd-Acosta C, Wen B, Wu Z, Montano C, Onyango P, Cui H, Gabo K, Rongione M, Webster M, et al. The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat Genet* 2009; 41:178-86; PMID:19151715; <http://dx.doi.org/10.1038/ng.298>
21. Mena E, Turkbey B, Mani H, Adler S, Valera VA, Bernardo M, Shah V, Pohida T, McKinney Y, Kwarteng G, et al. 11C-Acetate PET/CT in localized prostate cancer: a study with MRI and histopathologic correlation. *J Nucl Med* 2012; 53:538-45; PMID:22343504; <http://dx.doi.org/10.2967/jnumed.111.096032>
22. Feng J, Liu T, Zhang Y. Using MACS to identify peaks from ChIP-Seq data. *Current protocols in bioinformatics / editorial board, Andreas D Baxevanis [et al. 2011; Chapter 2:Unit 2 14.*
23. Jaffe AE, Murakami P, Lee H, Leek JT, Fallin MD, Feinberg AP, Irizarry RA. Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies. *Int J Epidemiol* 2012; 41:200-9; PMID:22422453; <http://dx.doi.org/10.1093/ije/dyr238>
24. Zhang X, Cruz FD, Terry M, Remotti F, Matushansky I. Terminal differentiation and loss of tumorigenicity of human cancers via pluripotency-based reprogramming. *Oncogene* 2013; 32:2249-60, e1-21; PMID:22777357; <http://dx.doi.org/10.1038/ncr.2012.237>
25. Kortenhorst MS1, Wissing MD, Rodríguez R, Kachhap SK, Jans JJ, Van der Groep P, Verheul HM, Gupta A, Aiyetan PO, van der Wall E, et al. Analysis of the genomic response of human prostate cancer cells to histone deacetylase inhibitors. *Epigenetics*. 2013; 8:907-20; PMID:23880963; <http://dx.doi.org/10.4161/epi.25574>
26. Stuart ET, Haffner R, Oren M, Gruss P. Loss of p53 function through PAX-mediated transcriptional repression. *EMBO J* 1995; 14:5638-45; PMID:8521821
27. O'Brien P, Morin P Jr., Ouellette RJ, Robichaud GA. The Pax-5 gene: a pluripotent regulator of B-cell differentiation and cancer disease. *Cancer Res* 2011; 71:7345-50; PMID:22127921; <http://dx.doi.org/10.1158/0008-5472.CAN-11-1874>
28. Moelans CB, Verschuur-Maes AH, van Diest PJ. Frequent promoter hypermethylation of BRCA2, CDH13, MSH6, PAX5, PAX6 and WT1 in ductal carcinoma in situ and invasive breast cancer. *J Pathol* 2011; 225:222-31; PMID:21710692; <http://dx.doi.org/10.1002/path.2930>
29. Torlakovic E, Slipicevic A, Robinson C, DeCoteau JF, Alfsen GC, Vyberg M, Chibbar R, Flores VA. Pax-5 expression in nonhematopoietic tissues. *Am J Clin Pathol* 2006; 126:798-804; PMID:17050077; <http://dx.doi.org/10.1309/JEC7JMW9YRM74RNO>
30. Liu W, Li X, Chu ES, Go MY, Xu L, Zhao G, Li L, Dai N, Si J, Tao Q, et al. Paired box gene 5 is a novel tumor suppressor in hepatocellular carcinoma through interaction with p53 signaling pathway. *Hepatology* 2011; 53:843-53; PMID:21319196; <http://dx.doi.org/10.1002/hep.24124>
31. Li X, Cheung KF, Ma X, Tian L, Zhao J, Go MY, Shen B, Cheng AS, Ying J, Tao Q, et al. Epigenetic inactivation of paired box gene 5, a novel tumor suppressor gene, through direct upregulation of p53 is associated with prognosis in gastric cancer patients. *Oncogene* 2012; 31:3419-30; PMID:22105368; <http://dx.doi.org/10.1038/ncr.2011.511>
32. Cobaleda C, Schebesta A, Delogu A, Busslinger M. Pax5: the guardian of B cell identity and function. *Nat Immunol* 2007; 8:463-70; PMID:17440452; <http://dx.doi.org/10.1038/ni1454>
33. Mullighan CG, Goorha S, Radtke I, Miller CB, Coustan-Smith E, Dalton JD, Girtman K, Mathew S, Ma J, Pounds SB, et al. Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. *Nature* 2007; 446:758-64; PMID:17344859; <http://dx.doi.org/10.1038/nature05690>
34. Mandel EM, Grosschedl R. Transcription control of early B cell differentiation. *Curr Opin Immunol* 2010; 22:161-7; PMID:20144854; <http://dx.doi.org/10.1016/j.coi.2010.01.010>
35. Todd DJ, Lee AH, Glimcher LH. The endoplasmic reticulum stress response in immunity and autoimmunity. *Nat Rev Immunol* 2008; 8:663-74; PMID:18670423; <http://dx.doi.org/10.1038/nri2359>
36. Chen Z, Xiao Y, Zhang J, Li J, Liu Y, Zhao Y, Ma C, Luo J, Qiu Y, Huang G, et al. Transcription factors E2A, FOXO1 and FOXP1 regulate recombination activating gene expression in cancer cells. *PLoS One* 2011; 6:e20475; PMID:21655267; <http://dx.doi.org/10.1371/journal.pone.0020475>
37. Palmisano WA, Crume KP, Grimes MJ, Winters SA, Toyota M, Esteller M, Joste N, Baylin SB, Belinsky SA. Aberrant promoter methylation of the transcription factor genes PAX5 alpha and beta in human cancers. *Cancer Res* 2003; 63:4620-5; PMID:12907641
38. Lagergren A, Manetopoulos C, Axelson H, Sigvardsson M. Neuroblastoma and pre-B lymphoma cells share expression of key transcription factors but display tissue restricted target gene expression. *BMC Cancer* 2004; 4:80; PMID:15544702; <http://dx.doi.org/10.1186/1471-2407-4-80>
39. Bolós V, Grego-Bessa J, de la Pompa JL. Notch signaling in development and cancer. *Endocr Rev* 2007; 28:339-63; PMID:17409286; <http://dx.doi.org/10.1210/er.2006-0046>
40. Mani S, Szymańska K, Cuenin C, Zaridze D, Balassiano K, Lima SC, Matos E, Daudt A, Koifman S, Filho VW, et al. DNA methylation changes associated with risk factors in tumors of the upper aerodigestive tract. *Epigenetics* 2012; 7:270-7; PMID:22430803; <http://dx.doi.org/10.4161/epi.7.3.19306>
41. Wall DS, Mears AJ, McNeill B, Mazerolle C, Thurig S, Wang Y, Kageyama R, Wallace VA. Progenitor cell proliferation in the retina is dependent on Notch-independent Sonic hedgehog/Hes1 activity. *J Cell Biol* 2009; 184:101-12; PMID:19124651; <http://dx.doi.org/10.1083/jcb.200805155>
42. Landsman L, Parent A, Hebrok M. Elevated Hedgehog/Gli signaling causes beta-cell dedifferentiation in mice. *Proc Natl Acad Sci U S A* 2011; 108:17010-5; PMID:21969560; <http://dx.doi.org/10.1073/pnas.1105404108>
43. Forastiere A, Koch W, Trotti A, Sidransky D. Head and neck cancer. *N Engl J Med* 2001; 345:1890-900; PMID:11756581; <http://dx.doi.org/10.1056/NEJMra001375>
44. Manley NR, Capecchi MR. The role of Hoxa-3 in mouse thymus and thyroid development. *Development* 1995; 121:1989-2003; PMID:7635047
45. Rodrigo I, Hill RE, Balling R, Münsterberg A, Imai K. Pax1 and Pax9 activate Bapx1 to induce chondrogenic differentiation in the sclerotome. *Development* 2003; 130:473-82; PMID:12490554; <http://dx.doi.org/10.1242/dev.00240>
46. Mammucari C, Tommasi di Vignano A, Sharov AA, Neilson J, Havrdra MC, Roop DN, Botchkarev VA, Crabtree GR, Dotto GP. Integration of Notch 1 and calcineurin/NFAT signaling pathways in keratinocyte growth and differentiation control. *Dev Cell* 2005; 8:665-76; PMID:15866158; <http://dx.doi.org/10.1016/j.devcel.2005.02.016>
47. Sang L, Roberts JM, Coller HA. Hijacking HES1: how tumors co-opt the anti-differentiation strategies of quiescent cells. *Trends Mol Med* 2010; 16:17-26; PMID:20022559; <http://dx.doi.org/10.1016/j.molmed.2009.11.001>
48. Mill P, Mo R, Fu H, Grachtchouk M, Kim PC, Dlugosz AA, Hui CC. Sonic hedgehog-dependent activation of Gli2 is essential for embryonic hair follicle development. *Genes Dev* 2003; 17:282-94; PMID:12533516; <http://dx.doi.org/10.1101/gad.1038103>
49. Cillo C, Cantile M, Faiella A, Boncinelli E. Homeobox genes in normal and malignant cells. *J Cell Physiol* 2001; 188:161-9; PMID:11424082; <http://dx.doi.org/10.1002/jcp.1115>
50. Schubert M, Yu JK, Holland ND, Escriva H, Laudet V, Holland LZ. Retinoic acid signaling acts via Hox1 to establish the posterior limit of the pharynx in the chordate amphioxus. *Development* 2005; 132:61-73; PMID:15576409; <http://dx.doi.org/10.1242/dev.01554>
51. Koop D, Holland ND, Sémon M, Alvarez S, de Lera AR, Laudet V, Holland LZ, Schubert M. Retinoic acid signaling targets Hox genes during the amphioxus gastrula stage: insights into early anterior-posterior patterning of the chordate body plan. *Dev Biol* 2010; 338:98-106; PMID:19914237; <http://dx.doi.org/10.1016/j.ydbio.2009.11.016>
52. Turkbey B, Mani H, Aras O, Ho J, Hoang A, Rastinehad AR, Agarwal H, Shah V, Bernardo M, Pang Y, et al. Prostate cancer: can multiparametric MR imaging help identify patients who are candidates for active surveillance? *Radiology* 2013; 268:144-52; PMID:23468576; <http://dx.doi.org/10.1148/radiol.13121325>
53. Turkbey B, Mani H, Aras O, Rastinehad AR, Shah V, Bernardo M, Pohida T, Daar D, Benjamin C, McKinney YL, et al. Correlation of magnetic resonance imaging tumor volume with histopathology. *J Urol* 2012; 188:1157-63; PMID:22901591; <http://dx.doi.org/10.1016/j.juro.2012.06.011>
54. Shah V, Turkbey B, Mani H, Pang Y, Pohida T, Merino MJ, Pinto PA, Choyke PL, Bernardo M. Decision support system for localizing prostate cancer based on multiparametric magnetic resonance imaging. *Med Phys* 2012; 39:4093-103; PMID:22830742; <http://dx.doi.org/10.1118/1.4722753>
55. Dal Molin M, Matthaei H, Wu J, Blackford A, Debeljak M, Rezaee N, Wolfgang CL, Butturini G, Sallia R, Bassi C, et al. Clinicopathological correlates of activating GNAS mutations in intraductal papillary mucinous neoplasm (IPMN) of the pancreas. *Ann Surg Oncol* 2013; 20:3802-8; PMID:23846778; <http://dx.doi.org/10.1245/s10434-013-3096-1>
56. Bock C, Tomazou EM, Brinkman AB, Müller F, Simmer F, Gu H, Jäger N, Gnirke A, Stunnenberg HG, Meissner A. Quantitative comparison of genome-wide DNA methylation mapping technologies. *Nat Biotechnol* 2010; 28:1106-14; PMID:20852634; <http://dx.doi.org/10.1038/nbt.1681>
57. Yegnasubramanian S, Wu Z, Haffner MC, Esopi D, Aryee MJ, Badrinath R, He TL, Morgan JD, Carvalho B, Zheng Q, et al. Chromosome-wide mapping of DNA methylation patterns in normal and malignant prostate cells reveals pervasive methylation of gene-associated and conserved intergenic sequences. *BMC Genomics* 2011; 12:313; PMID:21669002; <http://dx.doi.org/10.1186/1471-2164-12-313>
58. Zhang Y, Liu T, Meyer CA, Eeckhoutte J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown N, Li W, et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 2008; 9:R137; PMID:18798982; <http://dx.doi.org/10.1186/gb-2008-9-9-r137>
59. Feng J, Liu T, Qin B, Zhang Y, Liu XS. Identifying ChIP-seq enrichment using MACS. *Nat Protoc* 2012; 7:1728-40; PMID:22936215; <http://dx.doi.org/10.1038/nprot.2012.101>

60. Jaenisch R, Bird A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet* 2003; 33(Suppl):245-54; PMID:12610534; <http://dx.doi.org/10.1038/ng1089>
61. Hansen KD, Timp W, Bravo HC, Sabuncian S, Langmead B, McDonald OG, Wen B, Wu H, Liu Y, Diep D, et al. Increased methylation variation in epigenetic domains across cancer types. *Nat Genet* 2011; 43:768-75; PMID:21706001; <http://dx.doi.org/10.1038/ng.865>
62. Lister R, Pelizzola M, Downen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo QM, et al. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* 2009; 462:315-22; PMID:19829295; <http://dx.doi.org/10.1038/nature08514>
63. Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, Le JM, Delano D, Zhang L, Schroth GP, Gunderson KL, et al. High density DNA methylation array with single CpG site resolution. *Genomics* 2011; 98:288-95; PMID:21839163; <http://dx.doi.org/10.1016/j.ygeno.2011.07.007>
64. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, Irizarry RA. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* 2014; PMID:24478339; <http://dx.doi.org/10.1093/bioinformatics/btu049>
65. Kim SY, Volsky DJ. PAGE: parametric analysis of gene set enrichment. *BMC Bioinformatics* 2005; 6:144; PMID:15941488; <http://dx.doi.org/10.1186/1471-2105-6-144>
66. Tyekucheva S, Marchionni L, Karchin R, Parmigiani G. Integrating diverse genomic data using gene sets. *Genome Biol* 2011; 12:R105; PMID:22018358; <http://dx.doi.org/10.1186/gb-2011-12-10-r105>
67. Benjamini Y, Drai D, Elmer G, Kafkafi N, Golani I. Controlling the false discovery rate in behavior genetics research. *Behav Brain Res* 2001; 125:279-84; PMID:11682119; [http://dx.doi.org/10.1016/S0166-4328\(01\)00297-2](http://dx.doi.org/10.1016/S0166-4328(01)00297-2)
68. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005; 102:15545-50; PMID:16199517; <http://dx.doi.org/10.1073/pnas.0506580102>
69. Daniel VC, Marchionni L, Hierman JS, Rhodes JT, Devereux WL, Rudin CM, Yung R, Parmigiani G, Dorsch M, Peacock CD, et al. A primary xenograft model of small-cell lung cancer reveals irreversible changes in gene expression imposed by culture in vitro. *Cancer Res* 2009; 69:3364-73; PMID:19351829; <http://dx.doi.org/10.1158/0008-5472.CAN-08-4210>
70. Ross AE, Marchionni L, Vuica-Ross M, Cheadle C, Fan J, Berman DM, Schaeffer EM. Gene expression pathways of high grade localized prostate cancer. *Prostate* 2011; PMID:21360566; <http://dx.doi.org/10.1002/pros.21373>
71. Hoque MO, Begum S, Topaloglu O, Chatterjee A, Rosenbaum E, Van Criekinge W, Westra WH, Schoenberg M, Zahurak M, Goodman SN, et al. Quantitation of promoter methylation of multiple genes in urine DNA and bladder cancer detection. *J Natl Cancer Inst* 2006; 98:996-1004; PMID:16849682; <http://dx.doi.org/10.1093/jnci/djj265>



Published in final edited form as:

*Int J Cancer*. 2015 October 15; 137(8): 1879–1889. doi:10.1002/ijc.29558.

## NF- $\kappa$ B and STAT3 Transcription Factor Signatures Differentiate HPV-positive and HPV-negative Head and Neck Squamous Cell Carcinoma

Daria A. Gaykalova<sup>1</sup>, Judith B. Manola<sup>2</sup>, Hiroyuki Ozawa<sup>3</sup>, Veronika Zizkova<sup>1,4</sup>, Kathryn Morton<sup>1</sup>, Justin A. Bishop<sup>1,5</sup>, Rajni Sharma<sup>5</sup>, Chi Zhang<sup>1,6</sup>, Christina Michailidi<sup>1</sup>, Michael Considine<sup>7</sup>, Marietta Tan<sup>1</sup>, Elana J. Fertig<sup>7</sup>, Patrick T. Hennessey<sup>1</sup>, Julie Ahn<sup>1</sup>, Wayne M. Koch<sup>1</sup>, William H. Westra<sup>1,5</sup>, Zubair Khan<sup>1</sup>, Christine H. Chung<sup>1,3</sup>, Michael F. Ochs<sup>7,8</sup>, and Joseph A. Califano<sup>1,9</sup>

<sup>1</sup>Department of Otolaryngology—Head and Neck Surgery, Johns Hopkins Medical Institutions, Baltimore, Maryland, USA

<sup>2</sup>Department of Biostatistics & Computational Biology, Dana-Farber Cancer Institute, Boston, Massachusetts, USA

<sup>3</sup>Department of Oncology, Johns Hopkins Medical Institutions, Baltimore, Maryland, USA

<sup>4</sup>Laboratory of Molecular Pathology, Institute of Molecular and Translational Medicine, Palacky University, Olomouc, Czech Republic

<sup>5</sup>Department of Pathology, Johns Hopkins Medical Institutions, Baltimore, Maryland, USA

<sup>6</sup>University of Virginia, Charlottesville, Virginia, USA

<sup>7</sup>Division of Oncology Biostatistics, Department of Oncology, Johns Hopkins Medical Institutions, Baltimore, Maryland, USA

<sup>8</sup>Department of Mathematics and Statistics, The College of New Jersey, Ewing, New Jersey, USA

<sup>9</sup>Milton J. Dance Head and Neck Center, Greater Baltimore Medical Center, Baltimore, Maryland, USA

### Abstract

Using high-throughput analyses and the TRANSFAC database, we characterized TF signatures of head and neck squamous cell carcinoma (HNSCC) subgroups by inferential analysis of target gene expression, correcting for the effects of DNA methylation and copy number. Using this discovery pipeline, we determined that human papillomavirus-related (HPV+) and HPV– HNSCC differed significantly based on the activity levels of key TFs including AP1, STATs, NF- $\kappa$ B, and p53. Immunohistochemical analysis confirmed that HPV– HNSCC is characterized by co-activated STAT3 and NF- $\kappa$ B pathways, and functional studies demonstrate that this phenotype can be

**CORRESPONDING AUTHOR:** Joseph A. Califano, M.D., Department of Otolaryngology-Head and Neck Surgery, Johns Hopkins Medical Institutions, 1550 Orleans Street, Room 5N.04, Baltimore, Maryland 21231, jcalifa@jhmi.edu, Phone: 410-502-2692, Fax: 410-614-1411.

### CONFLICT OF INTEREST:

The authors declare no conflict of interest

effectively targeted with combined anti-NF- $\kappa$ B and anti-STAT therapies. These discoveries correlate strongly with previous findings connecting STATs, NF- $\kappa$ B, and AP1 in HNSCC. We identified 5 top-scoring pair biomarkers from STATs, NF- $\kappa$ B and AP1 pathways that distinguish HPV+ from HPV- HNSCC based on TF activity, and validated these biomarkers on TCGA and on independent validation cohorts. We conclude that a novel approach to TF pathway analysis can provide insight into therapeutic targeting of patient subgroup for heterogeneous disease such as HNSCC.

### Keywords

HNSCC; HPV; STAT3; NF- $\kappa$ B; Transcription factor

---

## INTRODUCTION

Head and neck squamous cell carcinoma (HNSCC) is the fifth most common cancer worldwide.<sup>1</sup> HNSCC has been traditionally associated with tobacco and alcohol exposure. A subset of high-risk human papilloma virus (HPV)-related HNSCC often found in tobacco non-exposed individuals has been described as a distinct clinicopathologic entity.<sup>2</sup> HPV- HNSCC is typically associated with poorer outcomes, while HPV+ HNSCC has a more favorable prognosis.<sup>3</sup> For patients with locally advanced HNSCC, multi-modality treatments including surgery, radiotherapy and/or cytotoxic chemotherapy have significantly improved, but five-year overall HNSCC survival remains poor.<sup>3</sup> Both genetic and epigenetic aberrations have been shown to play a role in HNSCC development, treatment response and survival.<sup>4</sup> However, the only targeted agent approved by the FDA for the treatment of HNSCC is cetuximab, a monoclonal antibody directed against epidermal growth factor receptor (EGFR).<sup>5</sup>

The development of targeted therapy is challenged by diverse genetic and epigenetic alterations heterogeneously distributed in HNSCC. The data suggest that genetic abnormalities affect a limited number of mitogenic pathways, such as PI3K and NOTCH.<sup>6-8</sup> However, genome-wide expression array data demonstrate that HNSCC is characterized by expression dysregulation for a majority of genes.<sup>9, 10</sup> Transcription factors are a primary determinant in the regulation of gene expression. Unfortunately, direct comprehensive analysis of the dysregulation of key TFs is challenging with existing screening techniques for several reasons: it is difficult to analyze changes in TFs expression due to their transient and low level expression; activation of most TFs requires post-translational modifications, protein cleavage, formation of active transcriptional complex or protein translocation from cytoplasm to nucleus.<sup>11</sup> Therapeutic targeting of TF is complicated due to the low level and cycle-dependent expression of TF, diversity of posttranscriptional modifications and redundancy of the downstream gene regulation. Despite that, many laboratories have demonstrated efficiency of several anti-TF therapies, including agents such as bortezomib [targeting NF- $\kappa$ B through proteasome inhibition],<sup>12</sup> BEZ-235 [a dual PI3K/mTOR inhibitor],<sup>6</sup> and decoy oligonucleotides [STAT3 and NF- $\kappa$ B inhibitors].<sup>13, 14</sup>

Using high throughput array datasets and the TRANSFAC database, we have developed an inferential method describing TF activity by the gene expression of their targets.<sup>15</sup> Using this method we demonstrate that HPV+ and HPV- HNSCC are differentiated by the activity of key TFs including STATs and NF- $\kappa$ B. Approximately 25% of primary HNSCC, almost exclusively HPV-, have these TFs and their targets coordinately upregulated. After validation of this TF alteration pattern in separate clinical cohorts, we further confirmed that HNSCC with key TF pathway alteration are sensitive to therapy directed at these key TF pathways, providing evidence that TF pathway analysis can provide insight into therapeutic targeting.

## MATERIALS AND METHODS

The detailed methods can be found in “Supporting Information”

### Tissue samples

We used three independent cohorts of HNSCC and normal specimens, overall: 195 HNSCC and 63 non-cancer affected patients (Tables S1, S2, and S3). Every participant signed a written informed consent before participating in this study. This study was approved by Johns Hopkins Medicine Internal Review Board (JHM IRB), and performed under approved research protocol NA\_00036235. TMA tissues were collected under ECOG/RTOG-approved protocols. We have also used publicly available data (<http://cancergenome.nih.gov/cancersselected/headandneck>) for the TCGA-HNSCC cohort that includes 279 HNSCC and 50 control tissues.

### DNA and RNA preparation from microdissected tissues

DNA was isolated by phenol-chloroform extraction after proteinase K digestion<sup>16</sup> and RNA was isolated with the mirVana miRNA Isolation Kit (Ambion) per manufacturer’s recommendations.

### Arrays

Two  $\mu$ g of RNA or DNA were run on Affymetrix HuEx1.0 GeneChips (RNA), Illumina Infinium HumanMethylation27 BeadChips (bisulfite converted DNA), and Affymetrix Genome-wide SNP 6.0 Array (DNA). All high throughput data sets are available in GEO superSeries GSE33232. The data can be freely downloaded from <http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE33232>.

### HPV analyses

HPV status of oropharyngeal SCC tumors was tested by *in situ* hybridization (ISH) for high-risk HPV,<sup>17</sup> by quantitative PCR (qPCR).<sup>16</sup> P16 expression was checked by immunohistochemical staining.<sup>17</sup>

### Reverse Transcription (RT) and quantitative Real Time PCR (qRT-PCR)

One  $\mu$ g of RNA from the validation cohort was reverse transcribed using the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems). QRT-PCR was performed on



Taqman 7900HT (Applied Biosystems, Foster City, CA) with gene-specific expression assays (Applied Biosystems) using standard qRT-PCR conditions.

### Immunohistochemistry

Immunostaining of 5  $\mu$ m cuts was carried out on a Bond-Leica autostaining system (Leica Microsystems, Buffalo Grove, IL) using a standard immunohistochemistry (IHC) protocol with 15 min incubation against total NF- $\kappa$ B p65 - RELA (#8242, dilution 1:400) or against total STAT3 (#4904, 1:200 dilution) antibodies (Cell Signaling). Staining was quantified by Aperio software.

### Cell Culture

**Cell Lines and Cell Culture Conditions**—Human HNSCC cell lines Ho1N1, HSC2, and SKN3 were purchased from the Japanese Collection of Research Bioresource. Each cell line was authenticated using a short tandem repeat analysis kit, GenePrint 10 (Promega), as directed at the Johns Hopkins University Core Facility. Cell grew at 37°C in 5% CO<sub>2</sub>.

**Transient Transfection and Cell Proliferation Assay**—The expression of STAT1, STAT3 and RELA genes were knocked down by ON-TARGETplus siRNA SMARTpool (Thermo Scientific) using RNAiMAX Transfection reagent (Life Technologies). Cell proliferation was measure by CCK-8 kit (Dojindo).

**Chemical Treatment**—Ho1N1, HSC2 and SKN3 cell lines were treated by either 0.1% Dimethyl sulfoxide (DMSO) alone or by 0.1% DMSO with different concentrations of Bortezomib (#681238/8, Kindly provided by NIH, Bethesda, MD), Bay 11-7085 (#B5681, Sigma-Aldrich, St. Louis, MO), Cucurbitacin I (#1571, Tocris, Bristol, UK), or and SH-4-54 (#S7337, Selleckchem, Houston, TX). All cells grew in high glucose DMEM medium with 10% FBS and 1% Penicillin-Streptavidin. Cell proliferation was measure by the CCK-8 kit as described above.

### Statistical analyses

**Preparation of TF target gene sets**—We applied robust multi-array average (RMA) analysis to an expression array dataset for the discovery cohort. We annotated each TF with a list of its experimentally validated targets described in the Transcription Factor Database [TRANSFAC Professional].<sup>18</sup> TFs with less than 5 targets were removed. From each TF gene set, we removed genes that were expected to have significantly reduced expression either due to increased methylation ( $\beta > 0.15$ ) or DNA copy loss (CNV < 1.2) on a per tumor sample basis, creating tumor-specific TF gene sets.<sup>15</sup> Methylation arrays were preprocessed using R scripts that provided the estimated b value of percentage methylation. SNP-chip arrays were preprocessed using corrected robust linear models with maximum likelihood-based distances (CRLMM) to provide copy number estimates.<sup>19</sup> Genes were removed from the TF gene set on a patient-specific basis using the resulting methylation and copy number estimates.

**TF activity ranking**—The resulting 1,325 tumor-specific TF target gene sets, independently corrected in the discovery and TCGA cohort for DNA-methylation and CNV-

dependent expression loss, were used to compare samples from HPV+ and HPV– groups to the pooled values of non-cancer samples for each individual target gene by pseudo-t test, where each individual tumor sample was given a score based on the probability that it came from a normal distribution with mean and standard deviation equal to the expression of that gene in the normal samples. Wilcoxon rank-sum tests were applied to these pseudo-t statistics to compare the expression of targets of each TF relative to the expression of targets not regulated by the TF in each sample relative to normal. Wilcoxon p-values of all TF gene sets were used to rank TFs.

**Top Scoring Pair (TSP)**—TSP was applied to the combined list of 72 target genes of STAT1, STAT3, NF- $\kappa$ B and AP1 pathways (Table S4). TSP aimed to discover five independent pairs of genes, whose changes in relative level best separate HPV+ and HPV– samples from the discovery cohort. These relative changes were used to score patients in TCGA and a novel validation cohort, where measurements were made by sequencing and RT-PCR methods.

**Correlation of IHC staining with clinical data**—Protein staining scored by Aperio software was correlated with clinical data of ECOG-TMA cohort. Wilcoxon rank sum tests were used to compare two groups and Kruskal-Wallis tests were used to compare more than 2 groups. Markers' co-staining was assessed by linear regression.

## RESULTS

### Cohort assembly and array analysis

To determine gene expression, methylation and copy number in primary HNSCC, we employed Affymetrix HuEx1.0 GeneChips Array, Illumina Infinium HumanMethylation27 BeadChips, and Affymetrix Genome-wide SNP 6.0 Array for 44 HNSCC (multiple primary sites, broad stage distribution) and 25 normal samples from a discovery cohort described earlier [see Table S1 and <sup>8,9</sup>].

### Array data annotation and preparation of TF target gene sets

The information regarding the target genes for each TF was obtained from the TRANSFAC 2010.4 Professional database.<sup>18</sup> We retained TFs with a minimum of five experimentally validated targets, creating sets of target genes for 1,325 human TFs of the total 2,600 human TF described in TRANSFAC. The decreased expression signals from methylation and copy number-lost target genes would lower the calculated activity of TF, without necessarily lowering the net activity of the TF in a particular tumor sample. To correct for estimated changes in TF activity due to the effects of DNA methylation and copy number changes on target gene expression, we removed genes from each individual sample analysis that had a possibility of reduced expression regardless of TF activity by potential promoter methylation-driven silencing ( $\beta > 0.15$ ) or by genetic loss of homo- or heterozygosity [Copy Number  $< 1.2$  in Copy Number Variation (CNV) estimation]. These genes were identified by Methylation and SNP arrays used for the same discovery cohort of samples. Methylation and CNV-corrected 1,325 TF target gene sets for each sample were used for further analysis (Figure 1).

## HPV positive and HPV negative patients are different by their TF signature

Methylation and CNV-corrected TF target gene sets were analyzed, where the expression of each target gene was compared for each individual samples to the pool of normal samples. In order to elucidate TF driven differences between HPV+ and HPV- samples, we compared gene expression differences between HPV- and HPV+ samples for the corrected list of gene targets of each TF. We used a Wilcoxon rank sum gene set test to rank the TFs based on the expression changes in their targets. While a single target gene could be either upregulated or downregulated in the particular sample or patient group, the determination of TF activity significance is agnostic to direction of dysregulation in TF gene targets, assuming the presence of both upregulated and downregulated targets in its list. We isolated the fifty TFs with the lowest p-values comparing expression of genes in HPV- and HPV+ groups for further analysis. Of note, due to some overlap in nomenclature of specific TFs annotated in TRANSFAC, some TFs belong to the same TF pathway. Thus, the majority of the top fifty ranked TFs could be assigned to a limited number of key TF pathways, including NF- $\kappa$ B [(RelA-p65)2, NF-kappaB1, NF-kappaB1-p50:RelA-p65], AP-1 [c-Fos:c-Jun, Fra-1, ATF, ATF-1, NF-AT1], retinoic acid signaling [PPARalpha:RXR-alpha, RXR-alpha:PPARalpha, RXR-alpha:PPARgamma, LXR-alpha:RXR-alpha, RORalpha1, RXR-alpha], STAT [STAT1, STAT1alpha, STAT5, STAT6, STAT1:STAT1 and STAT3:STAT3], NOTCH [E12, HES-1], p53 [p53], and RB [E2F-4] pathways (Table S5). Please note, that while we detected canonical NF- $\kappa$ B pathway members, we did not detect any members of NF- $\kappa$ B alternative pathway (NFKB2 or RELB).<sup>20</sup> Our data suggest that HPV+ and HPV- HNSCC demonstrate gene dysregulation related specifically to distinct alterations in TF activity (Table S5). These data were recapitulated using TCGA-HNSCC cohort (Table S6). These results support prior reports of overall dysregulation of several common TF pathways in head and neck cancer.<sup>21</sup>

In order to determine robustness to different CNV and methylation thresholds, we repeated this analysis while varying threshold values, such as  $\beta > 0.25$ ,  $\beta > 0.35$  and  $\beta > 0.45$  for DNA methylation, and  $\text{CNV} < 1$ ,  $\text{CNV} < 0.8$ , and  $\text{CNV} = 0$  for DNA copy number. The majority of the top TFs with diverse thresholds were found in the top 50 TF list with originally applied biologically relevant  $\beta > 0.15$  and  $\text{CNV} < 1.2$  conditions, with most TFs involved in STATs, NF- $\kappa$ B, AP1 and retinoic acid pathways (data not shown). However, we noted that p53 was lost during threshold variations. The p53 pathway is known to be affected in HNSCC via two mechanisms, one – via direct HPV E6 oncoprotein-induced degradation of p53 protein, second – via multiple *TP53* mutations.<sup>22, 23</sup> Point mutations of *TP53* genes as well as chromosomal loss<sup>23</sup> and lack of adjustment for CNV correction on p53 itself may have altered the significance of p53 TF pathway dysregulation in this analysis.

## Co-activation of NF- $\kappa$ B and STAT3 in HPV- HNSCC

To investigate the coordinated dysregulation and the direction of this dysregulation (upregulation or downregulation) for key TFs, we performed immunohistochemical (IHC) staining experiments on a primary HNSCC tissue microarray (TMA) cohort including 100 HNSCC (consisting of 90 HPV- and 10 HPV+) from a prior clinical trial (ECOG E4393/RTOG 9614) and 13 control samples (Table S2).<sup>23</sup> We performed an analysis of STAT3 and NF- $\kappa$ B protein expression for these samples (Figure 2). The protein staining was scored

using Aperio software only in the tumor cells in each sample, where staining was independently quantified for the whole cell (reflecting overall protein expression) or for the nuclei (reflecting protein activation and translocation to nuclei) (Table S7, Figure S1). We have evaluated the correlation of protein staining with clinical data, such as age, gender, race, cancer site, as well as with disease, smoking, and HPV statuses. HPV status has the strongest correlation with protein staining. In particular HPV– tumors have increased protein expression and significant protein activation as defined by greater than average cellular and nuclear staining of both STAT3 and NF- $\kappa$ B, as compared to HPV+ (Figure 3), supporting the conclusion that these two groups are differentiated by the activity of TFs such as NF- $\kappa$ B and STATs (Tables S5 and S6). Correlation of nuclear protein staining with HPV was supported by univariate and multivariate analyses (Table S8). The co-staining of NF- $\kappa$ B and STAT3 was evaluated and showed significant co-activation of NF- $\kappa$ B and STAT3 in primary HNSCC (Table S9, Spearman correlation coefficient 0.30,  $p=0.003$ ). We found significant correlation of STAT3 and NF- $\kappa$ B activation (co-staining in nuclei) (Table S10). We have calculated the median of nuclear and cellular staining of STAT3 and NF- $\kappa$ B for the entire TMA cohort of 100 HNSCC samples (Table S10). We have noticed that 50 of 90 HPV– patients (55.6%) had NF- $\kappa$ B nuclear staining above the median staining intensity (12.85) while none of HPV+ samples had NF- $\kappa$ B staining above the median (Table S10). Similarly, 49 HPV– patients (54.4%) had STAT3 nuclear staining above the median (23.17) while only one HPV+ sample had STAT3 staining above the median ( $p=0.0012$  and  $0.0157$  for NF- $\kappa$ B and STAT3, Table S10). 30 of 90 HPV– patients (33.3%) had both NF- $\kappa$ B and STAT3 nuclear staining above the median, while no HPV+ patient was found to have this signature ( $p=0.0302$ ). See table S10 for details.

### **Target genes of transcription factors STATs, NF- $\kappa$ B and AP1 distinguish HPV+ and HPV– HNSCC**

To define expression biomarkers related to TF activity, we defined the TF target genes that reflect the dysregulation of these pathways and examined the ability of these genes to distinguish HPV+ and HPV– HNSCC. We combined the lists of all targets for NF- $\kappa$ B, STAT1, STAT3 and AP1 factors (Table S4) and applied top scoring pair [TSP]<sup>24, 25</sup> analysis to this combined list. TSP was applied to the expression array data of 72 combined target genes from the discovery cohort. TSP allowed the discovery of 5 gene pairs among 72 targets that discriminate HPV+ and HPV– (Figure S1), namely: CCND1, CEBPD, ICAM1, IFG1R, IL6ST, IRF1, JAG1, JAK3, NOS3, and SOCS3. Of note, 5 out of 10 of these paired genes belong to more than one of the above dysregulated pathways (Table S4). The five TSPs distinguished HPV+ and HPV– samples from the discovery cohort with a  $p$ -value of  $9.5 \times 10^{-6}$  and an Odds Ratio (OR) of 44 [5.0, 2185, 95% CI] (Table 1).

To validate our discovery of TF differential dysregulation, we adopted RNA-Seq data available for the TCGA HNSCC (The Cancer Genome Atlas) cohort. This cohort consists of 279 tumor samples, including 35 HPV+ and 244 HPV– samples. Validation of these genes as discriminators between HPV+ and HPV– within the TCGA cohort also demonstrated significant separation with a  $p$ -value of  $6.7 \times 10^{-8}$  OR = 8.9 [3.6 26, 95% CI] (Table 1).

### Validation of transcription factor target genes by quantitative real time PCR (qRT-PCR)

In order to validate our results and evaluate the discriminative ability of the top five gene pairs from the TSP analysis, we assembled an independent validation cohort of 61 HNSCC (including 43 HPV– and 18 HPV+ samples) and 28 control uvulopalatopharyngoplasty (UPPP) samples (Table S3). The TSPs based on expression of the top five gene pairs were analyzed by qRT-PCR, which was able to separate HPV+ and HPV– HNSCC from this validation cohort with a p-value of 0.0006, OR = 9.6 [2.2 60] (Table 1). Using qRT-PCR data for 5 top scoring gene pairs from TSP, this validation cohort (Table S3) was separated into three groups by unsupervised hierarchical clustering (Figure 4 heat-map was built for the purpose of visualization). We found that 25 out of 61 samples were clustered into a separate group that was enriched by HPV+ patients. That group contains 16 out of 18 HPV+ patients and is distinct from two other primarily HPV– groups by downregulation of CCND1 and partial downregulation of IRF1, ICAM1, IGF1R and CCND1 ( $p < 3.5 \times 10^{-5}$ ). These data suggest that STATs, NF- $\kappa$ B and AP1 pathways are upregulated in HPV– HNSCC patients as compared to HPV+ patients, in concordance with our immunohistochemical data. Two groups of primarily HPV– patients can be distinguished by relative expression of TSP genes, with one group showing higher expression of CCND1, CEBPD, ICAM1, IRF1, IL6ST, JAG1, JAK3, NOS3, and SOCS3. These data suggest that dysregulation of STATs, AP1 and NF- $\kappa$ B pathways is coordinated and that these pathways are co-activated for a subset of HNSCC patients (25%, 15 out of total 61 samples), predominantly HPV– patients ( $p < 0.05$  by Fisher's exact test applied to hierarchically clustered groups, Figure 4). As HPV– tumors have been associated with poor clinical outcome, the detected gene signatures can potentially aid clinical assessment of HNSCC prognostic subgroups.

### NF- $\kappa$ B and STAT co-activated HPV– HNSCC cells are growth-inhibited by combined targeted therapy

We chose Ho1N1 and HSC2 from total 34 cell lines described in <sup>26</sup> after expression analysis of key TFs and their significant targets (detected via TSP analysis); these cell lines were selected as representative of STAT3–NF- $\kappa$ B over-expressing or base-line expressing tumors based on expression pattern. To evaluate the clinical relevance and potential therapeutic susceptibility of NF- $\kappa$ B and STATs co-activation in HPV– HNSCC, we knocked down the expression of these TFs *in vitro* by RNAi techniques. Ho1N1 cells are characterized by over-expression of *STAT1*, *STAT3*, *RELA* and *NFKB1* and their signature targets as defined by our TSP analysis (*IRF1*, *CEBPD*, *CCND1*, *ICAM1*, *JAG1*, *JAK3*, and *NOS3*). Since our data demonstrate that expression of signature target genes discovered as five TSPs reliably describes the activity of key TF pathways, we conclude that Ho1N1 has hyper-activated STAT3 and NF- $\kappa$ B pathways. Conversely HSC2 showed low baseline expression of STATs, NF- $\kappa$ B genes and their significant targets. Of note, *JAK3* expression was not detected in HSC2 and the expression of *JAG1* and *NOS3* was not significantly different in these two cell lines (Figure S2).

The siRNA SMARTpools for *STAT1*, *STAT3* and *RELA* allowed significant inhibition of target gene expression in each cell line (Figure S3). Growth inhibition was associated with induction of the cell death by downregulation of NF- $\kappa$ B, shown for wt *TP53* head and neck

cell lines (HPV-).<sup>27</sup> Interestingly, significant growth inhibition was noticed only in the STAT and NF- $\kappa$ B activated Ho1N1 cell line, but not in the HSC2 cell line (Figure 5). Combined downregulation of two genes had additive effect on cell growth, suggesting that these pathways were downregulated independently. The combined inhibition of STAT1 and STAT3 expression had the strongest effect on Ho1N1 cell line proliferation (62% cell growth inhibition, as compared to control scrambled siRNA pool); and second strongest growth inhibition was archived with double STAT3-RELA inhibition (56% cell growth inhibition). The Ho1N1 cell line was the most sensitive to chemical NF- $\kappa$ B and STAT3 inhibitors (Bortezomib, Bay 11-7085, Cucurbitacin and SH-4-54; Figures S4 and S5, and References<sup>12, 28-30</sup>), supporting our siRNA experiments. NF- $\kappa$ B and STAT3 inhibitors were not used in combination due to the high cellular toxicity of the individual agents. No effect of siRNA inhibition was noted in the low STAT-NF- $\kappa$ B expressing HSC2 cell line (Figure 5b). HSC2 was the most resistant to chemical NF- $\kappa$ B and STAT3 inhibitors (Figures S4 and S5), supporting our siRNA experiments. SKN3 cells, characterized by intermediate expression of STATs, NF- $\kappa$ B and their targets (Figure S2), were used as a validating control. This cell line had intermediate cell growth inhibitory effect of STAT and NF- $\kappa$ B downregulation (Figure S6), with maximal cell growth inhibition with use of combined STAT3 and RELA siRNA (42% cell growth inhibition, Figure S6a). SKN3 has an intermediate response to both NF- $\kappa$ B inhibitors (Figure S4). As for STAT3 inhibitors, SKN3 was sensitive to Cucurbitacin and resistant to SH-4-54 (Figure S5), suggesting an overall intermediate response to STAT3 inhibitors. The diverse response of SKN3 to STAT3 inhibitors may be explained by high toxicity of Cucurbitacin and low specificity of SH-4-54.

## DISCUSSION

Over the past decade, there has been recognition that HNSCC includes favorable prognosis HPV+ and poor prognosis HPV- patients.<sup>31</sup> The discovery of novel therapeutic agents for poor prognosis HPV- HNSCC has been challenging, despite use of high throughput sequencing efforts to define novel therapeutic targets. TF are significant drivers of gene expression variation, and characterization of TF alterations facilitates insight into HNSCC biology; however, reliable changes in TF activity are hard to directly detect due to the complexity of TF signaling. Using three high throughput platforms and an innovative integrative strategy that compensates for genetic and epigenetic changes in TF targets, we were able to annotate highly specific target genes to each known TF and infer whole-genome TF activity by the expression of those target genes in each particular tumor sample.

We have found that HPV- patients have strong coordinated dysregulation of several pathways, including coordinated activation of NF- $\kappa$ B and STATs pathways. Our results are supported by similar results shown by other investigators using a cDNA microarray analysis of 10 HPV- HNSCC cell lines.<sup>32, 33</sup> We confirmed our results on multiple validation datasets, including an ECOG TMA cohort, a TCGA-HNSCC cohort and independent HNSCC validation cohort, by protein and expression assays, as well as by confirmation of expression alteration for direct targets of specific TFs. The discovered co-activation of NF- $\kappa$ B and STAT3 in primary HNSCC tissue is also in agreement with prior published IHC data.<sup>34</sup>

We have also shown that co-activation of NF- $\kappa$ B and STAT3 provide an avenue for selective therapeutic targeting in cell line models. We have demonstrated that these alterations occur in a substantial subset of HNSCCs. For example, 15 out of total 61 HNSCC patients (25%) or 14 out of total 43 HPV– HNSCC patients (33%) have higher expression of direct targets of STATs, NF- $\kappa$ B and AP1 TFs, such as CCND1, CEBPD, ICAM1, IRF1, JAG1, JAK3 and NOS3 (Figure 4). The direct dependence of these targets on TFs activity strongly correlates with the previous data showing NF- $\kappa$ B target activation in HNSCC.<sup>27</sup>

Prior reports have described changes in the activity of the several TFs, like STAT1, STAT3, NF- $\kappa$ B and AP1, in HPV– HNSCC, as well as in HPV mixed HNSCC population and HPV + cervix patients.<sup>35-39</sup> For example, HPV16 E7 was shown to inhibit NF- $\kappa$ B activity.<sup>40</sup> Additional TF pathway activity alterations in HNSCC have been shown for p53,<sup>22</sup> RB1,<sup>23</sup> and retinoic X receptor<sup>41</sup> cascades. There are many possibilities for these pathways to interact in HNSCC. Many of these TFs have common upstream regulators: including dependence of NF- $\kappa$ B, STATs and AP1 activation on IKK $\alpha$ , IKK $\beta$ , and on cytokine signals.<sup>20, 42</sup> They also share common downstream targets: for example, STAT1 and NF- $\kappa$ B regulate CD40 expression<sup>43, 44</sup>; AP1 and NF- $\kappa$ B regulate IL6, IL8 and VEGF expression.<sup>34</sup> Finally, these TFs can also affect their own expression and expression of each other: NF- $\kappa$ B plays a role in paracrine-dependent activation of STAT3 pathway.<sup>34, 42</sup> More shared targets can be found in Table S4 and in Reference<sup>32</sup>. Prior publications also suggest that co-activation of NF- $\kappa$ B and STAT3, especially in the context of dysfunctional p53 will enhance BCL-XL expression and lead to HNSCC cell survival.<sup>45</sup> While direct ChIP-Seq experiments are not feasible to perform on the primary HNSCC samples due to the limited tumor availability, we expect overlap in TF binding to their target genes, inconsistent with the literature cited above.

The overlapping gene targets of p53, p63 and p73 TFs and their ability to distinguish HPV+ and HPV– patients is summarized in Table S11. Notably, the mutation status of TP53 gene has a mixed downstream effect on the activity of its target genes (Table S12). The TP53 mutation status was available for 37 out of 44 HNSCC samples in the discovery cohort from previous deep-sequencing study<sup>46</sup>, where 25 samples were found to have wild-type TP53 and 7 samples were found to have disruptive TP53 mutations (2 samples with nonsense mutations and 5 samples with missense hot spot mutations: R175H, Y220C, R248Q, R273C, and R282W). The diverse effect of disruptive TP53 mutations can be explained by downregulating effect of nonsense mutations and activating effect of gain-of-function hot spot mutations. Such results correlate with previously published data, summarized in Reference<sup>47</sup>.

In general, TFs were considered “undruggable” for many years, but with improved understanding of transcription factor biology and new insight into pathway analysis, as well as novel technologies in the development of targeting agents and agent delivery, there is a broad venue for treatment based on TFs and their pathways.<sup>48, 49</sup> We have demonstrated that therapeutic agents directed against STATs and NF- $\kappa$ B, such as Cucurbitacin<sup>29</sup>, Bortezomib<sup>12</sup>, Bay 11-7085<sup>28</sup> and SH-4-54<sup>30</sup> have high toxicity and limited specificity (See Figures S4 and S5 for details). However, small peptides and peptidomimetics have shown lower cellular toxicity and better specificity to decrease transcription activity of STAT3, NF-

$\kappa$ B, MYC and others.<sup>50</sup> Oligonucleotide-based therapy, including RNA interference agents and single-stranded decoys have demonstrated specific downregulation of targeted pathways, and are actively used in preclinical and early clinical studies to facilitate AP1, RB1, ERs, NF- $\kappa$ B, and STATs downregulation.<sup>13, 14, 51</sup>

The correlation of our findings with previous published data suggests that the inferential technique we have developed to define TF activity is robust and a useful tool for discovery of genome-wide changes in TF activity. While we evaluated only the most prominent TF pathways, such as STATs, NF- $\kappa$ B and AP1, many more TFs were found to be significantly and differentially activated in two HNSCC groups. The use of an inferential strategy to deduce activity of specific TFs also has the potential for application in therapeutic targeting, as we were able to use a list of ten target genes that helped us to validate TF signatures experimentally. This strategy could potentially be adapted for clinical practice to identify patients with dysregulated (upregulated or downregulated) pathways. Characterization of a limited list of TF target genes allowed us to accurately characterize TF signatures experimentally, and could potentially be applied to define patients with tumors that are susceptible to specific TF pathway targeting agents as we have done in cell line systems, albeit with potential limitations in translating cell line responses to primary tumor responses to therapy. In addition, we used a correction strategy to infer TF signatures despite heterogeneity of genetic and epigenetic alterations that could potentially mask TF pathway alteration. This strategy has potential to clarify pathway alterations in specific individual tumors within tumor types that exhibit significant heterogeneity in terms of genetic and epigenetic alterations.

In summary, the use of high throughput integrated expression, methylation, and genome copy number analysis to infer TF activity in HNSCC provides insight into biologic variability of behavior and treatment response for HPV+ and HPV- HNSCC patients, and may potentially be used to direct targeted therapy based on TF pathway analysis.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## ACKNOWLEDGEMENT

This manuscript is based on a web database application provided by Research Information Technology Systems (RITS)—<https://www.rits.onc.jhmi.edu/>.

### FINANCIAL SUPPORT:

The work is supported by American Head and Neck Society (AHNS) 306481 (DAG), NIDCR/NIH Challenge Grant RC1DE020324 (JAC), NIDCR/NCI P50DE019032 Head and Neck Cancer SPORE (JAC), NIH/NLM R01 LM011000 (MFO), and NIDCR/NIH R01 DE013152 (WMK).

## REFERENCES

1. Siegel R, Ma J, Zou Z, Jemal A. Cancer statistics, 2014. *CA: a cancer journal for clinicians*. 2014; 64:9–29. [PubMed: 24399786]



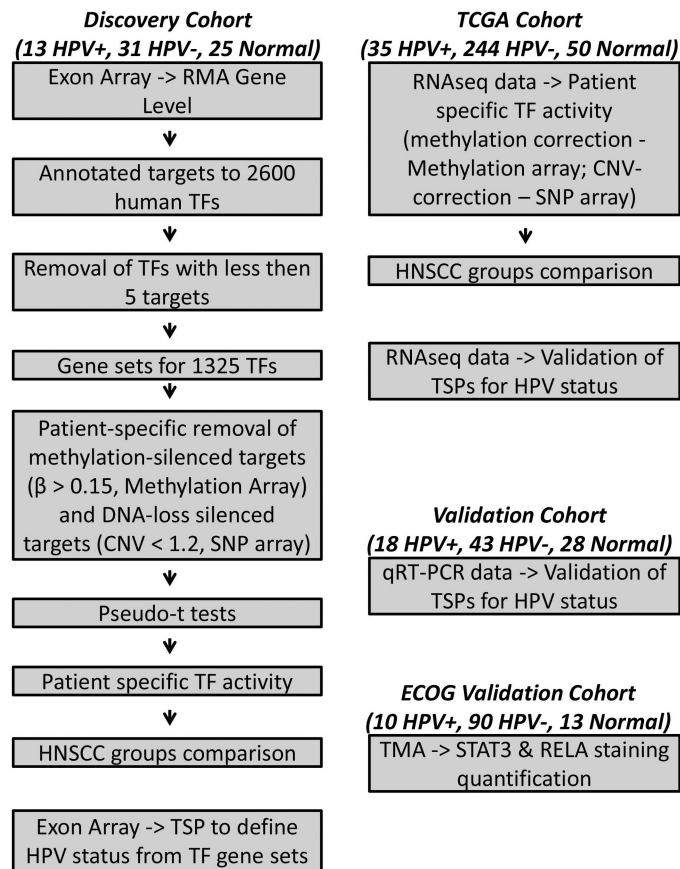
2. Psyrri A, Boutati E, Karageorgopoulou S. Human papillomavirus in head and neck cancers: biology, prognosis, hope of treatment, and vaccines. *Anticancer Drugs*. 2011; 22:586–90. [PubMed: 21403517]
3. Chaturvedi AK, Engels EA, Pfeiffer RM, Hernandez BY, Xiao W, Kim E, Jiang B, Goodman MT, Sibug-Saber M, Cozen W, Liu L, Lynch CF, et al. Human papillomavirus and rising oropharyngeal cancer incidence in the United States. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2011; 29:4294–301. [PubMed: 21969503]
4. Smith IM, Mydlarz WK, Mithani SK, Califano JA. DNA global hypomethylation in squamous cell head and neck cancer associated with smoking, alcohol consumption and stage. *International journal of cancer Journal international du cancer*. 2007; 121:1724–8. [PubMed: 17582607]
5. Bonner JA, Harari PM, Giralt J, Cohen RB, Jones CU, Sur RK, Raben D, Baselga J, Spencer SA, Zhu J, Youssoufian H, Rowinsky EK, et al. Radiotherapy plus cetuximab for locoregionally advanced head and neck cancer: 5-year survival data from a phase 3 randomised trial, and relation between cetuximab-induced rash and survival. *The lancet oncology*. 2010; 11:21–8. [PubMed: 19897418]
6. Lui VW, Hedberg ML, Li H, Vangara BS, Pendleton K, Zeng Y, Lu Y, Zhang Q, Du Y, Gilbert BR, Freilino M, Sauerwein S, et al. Frequent mutation of the PI3K pathway in head and neck cancer defines predictive biomarkers. *Cancer discovery*. 2013; 3:761–9. [PubMed: 23619167]
7. Agrawal N, Frederick MJ, Pickering CR, Bettegowda C, Chang K, Li RJ, Fakhry C, Xie TX, Zhang J, Wang J, Zhang N, El-Naggar AK, et al. Exome sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in NOTCH1. *Science*. 2011; 333:1154–7. [PubMed: 21798897]
8. Sun W, Gaykalova DA, Ochs MF, Mambo E, Arnaoutakis D, Liu Y, Loyo M, Agrawal N, Howard J, Li R, Ahn S, Fertig E, et al. Activation of the NOTCH pathway in head and neck cancer. *Cancer research*. 2014; 74:1091–104. [PubMed: 24351288]
9. Fertig EJ, Markovic A, Danilova LV, Gaykalova DA, Cope L, Chung CH, Ochs MF, Califano JA. Preferential activation of the hedgehog pathway by epigenetic modulations in HPV negative HNSCC identified with meta-pathway analysis. *PloS one*. 2013; 8:e78127. [PubMed: 24223768]
10. Chung CH, Parker JS, Karaca G, Wu J, Funkhouser WK, Moore D, Butterfoss D, Xiang D, Zanation A, Yin X, Shockley WW, Weissler MC, et al. Molecular classification of head and neck squamous cell carcinomas using patterns of gene expression. *Cancer cell*. 2004; 5:489–500. [PubMed: 15144956]
11. Tootle TL, Rebay I. Post-translational modifications influence transcription factor activity: a view from the ETS superfamily. *Bioessays*. 2005; 27:285–98. [PubMed: 15714552]
12. Allen C, Saigal K, Nottingham L, Arun P, Chen Z, Van Waes C. Bortezomib-induced apoptosis with limited clinical response is accompanied by inhibition of canonical but not alternative nuclear factor- $\kappa$ B subunits in head and neck cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research*. 2008; 14:4175–85. [PubMed: 18593997]
13. Gill JS, Zhu X, Moore MJ, Lu L, Yaszemski MJ, Windebank AJ. Effects of NF $\kappa$ B decoy oligonucleotides released from biodegradable polymer microparticles on a glioblastoma cell line. *Biomaterials*. 2002; 23:2773–81. [PubMed: 12059028]
14. Sen M, Paul K, Freilino ML, Li H, Li C, Johnson DE, Wang L, Eiseman J, Grandis JR. Systemic Administration of a Cyclic Signal Transducer and Activator of Transcription 3 (STAT3) Decoy Oligonucleotide Inhibits Tumor Growth without Inducing Toxicological Effects. *Molecular medicine*. 2014; 20:46–56. [PubMed: 24395569]
15. Rathi KS, Gaykalova DA, Hennessey P, Califano JA, Ochs MF. Correcting Transcription Factor Gene Sets for Copy Number and Promoter Methylation Variations. *Drug Dev Res*. 2014
16. Carvalho AL, Henrique R, Jeronimo C, Nayak CS, Reddy AN, Hoque MO, Chang S, Brait M, Jiang WW, Kim MM, Claybourne Q, Goldenberg D, et al. Detection of promoter hypermethylation in salivary rinses as a biomarker for head and neck squamous cell carcinoma surveillance. *Clinical cancer research : an official journal of the American Association for Cancer Research*. 2011; 17:4782–9. [PubMed: 21628494]
17. Singhi AD, Califano J, Westra WH. High-risk human papillomavirus in nasopharyngeal carcinoma. *Head Neck*. 2012; 34:213–8. [PubMed: 21484924]

18. Matys V, Fricke E, Geffers R, Gossling E, Haubrock M, Hehl R, Hornischer K, Karas D, Kel AE, Kel-Margoulis OV, Kloos DU, Land S, et al. TRANSFAC: transcriptional regulation, from patterns to profiles. *Nucleic Acids Res.* 2003; 31:374–8. [PubMed: 12520026]
19. Scharpf RB, Irizarry RA, Ritchie ME, Carvalho B, Ruczinski I. Using the R Package crlmm for Genotyping and Copy Number Estimation. *J Stat Softw.* 2011; 40:1–32. [PubMed: 22523482]
20. Nottingham LK, Yan CH, Yang X, Si H, Coupar J, Bian Y, Cheng TF, Allen C, Arun P, Gius D, Dang L, Van Waes C, et al. Aberrant IKKalpha and IKKbeta cooperatively activate NF-kappaB and induce EGFR/AP1 signaling to promote survival and migration of head and neck cancer. *Oncogene.* 2014; 33:1135–47. [PubMed: 23455325]
21. Rampias T, Sasaki C, Weinberger P, Psyri A. E6 and E7 gene silencing and transformed phenotype of human papillomavirus 16-positive oropharyngeal cancer cells. *J Natl Cancer Inst.* 2009; 101:412–23. [PubMed: 19276448]
22. Haraf DJ, Nodzenski E, Brachman D, Mick R, Montag A, Graves D, Vokes EE, Weichselbaum RR. Human papilloma virus and p53 in head and neck cancer: clinical correlates and survival. *Clinical cancer research : an official journal of the American Association for Cancer Research.* 1996; 2:755–62. [PubMed: 9816227]
23. Poeta ML, Manola J, Goldwasser MA, Forastiere A, Benoit N, Califano JA, Ridge JA, Goodwin J, Kenady D, Saunders J, Westra W, Sidransky D, et al. TP53 mutations and survival in squamous-cell carcinoma of the head and neck. *N Engl J Med.* 2007; 357:2552–61. [PubMed: 18094376]
24. Geman D, d'Avignon C, Naiman DQ, Winslow RL. Classifying gene expression profiles from pairwise mRNA comparisons. *Stat Appl Genet Mol Biol.* 2004; 3
25. Leek JT. The tspair package for finding top scoring pair classifiers in R. *Bioinformatics.* 2009; 25:1203–4. [PubMed: 19276151]
26. Hatakeyama H, Cheng H, Wirth P, Counsell A, Marcrom SR, Wood CB, Pohlmann PR, Gilbert J, Murphy B, Yarbrough WG, Wheeler DL, Harari PM, et al. Regulation of heparin-binding EGF-like growth factor by miR-212 and acquired cetuximab-resistance in head and neck squamous cell carcinoma. *PloS one.* 2010; 5:e12702. [PubMed: 20856931]
27. Lee TL, Yang XP, Yan B, Friedman J, Duggal P, Bagain L, Dong G, Yeh NT, Wang J, Zhou J, Elkahlon A, Van Waes C, et al. A novel nuclear factor-kappaB gene signature is differentially expressed in head and neck squamous cell carcinomas in association with TP53 status. *Clinical cancer research : an official journal of the American Association for Cancer Research.* 2007; 13:5680–91. [PubMed: 17908957]
28. Wang Q, Huber N, Noel G, Haar L, Shan Y, Pritts TA, Ogle CK. NF-kappaBeta inhibition is ineffective in blocking cytokine-induced IL-8 production but P38 and STAT1 inhibitors are effective. *Inflamm Res.* 2012; 61:977–85. [PubMed: 22618201]
29. Sun J, Blaskovich MA, Jove R, Livingston SK, Coppola D, Sebti SM. Cucurbitacin Q: a selective STAT3 activation inhibitor with potent antitumor activity. *Oncogene.* 2005; 24:3236–45. [PubMed: 15735720]
30. Haftchenary S, Luchman HA, Jouk AO, Veloso AJ, Page BD, Cheng XR, Dawson SS, Grinshtein N, Shahani VM, Kerman K, Kaplan DR, Griffin C, et al. Potent Targeting of the STAT3 Protein in Brain Cancer Stem Cells: A Promising Route for Treating Glioblastoma. *ACS medicinal chemistry letters.* 2013; 4:1102–7. [PubMed: 24900612]
31. Weinberger PM, Yu Z, Haffty BG, Kowalski D, Harigopal M, Brandsma J, Sasaki C, Joe J, Camp RL, Rimm DL, Psyri A. Molecular classification identifies a subset of human papillomavirus--associated oropharyngeal cancers with favorable prognosis. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology.* 2006; 24:736–47. [PubMed: 16401683]
32. Yan B, Chen G, Saigal K, Yang X, Jensen ST, Van Waes C, Stoeckert CJ, Chen Z. Systems biology-defined NF-kappaB regulons, interacting signal pathways and networks are implicated in the malignant phenotype of head and neck cancer cell lines differing in p53 status. *Genome biology.* 2008; 9:R53. [PubMed: 18334025]
33. Yan B, Yang X, Lee TL, Friedman J, Tang J, Van Waes C, Chen Z. Genome-wide identification of novel expression signatures reveal distinct patterns and prevalence of binding motifs for p53, nuclear factor-kappaB and other signal transcription factors in head and neck squamous cell carcinoma. *Genome biology.* 2007; 8:R78. [PubMed: 17498291]

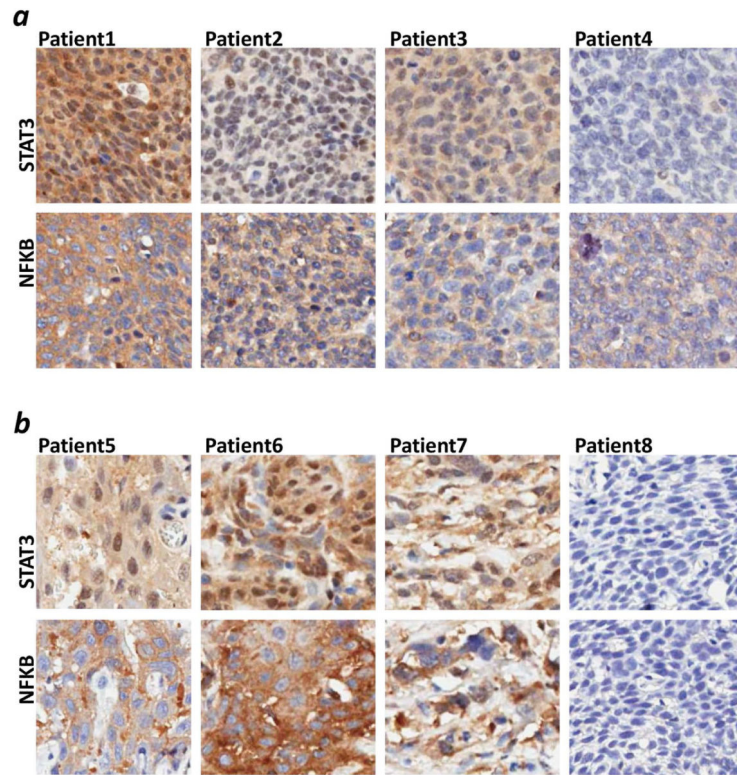
34. Squarize CH, Castilho RM, Sriuranpong V, Pinto DS Jr, Gutkind JS. Molecular cross-talk between the NF-kappaB and STAT3 signaling pathways in head and neck squamous cell carcinoma. *Neoplasia*. 2006; 8:733–46. [PubMed: 16984731]
35. Li Q, Verma IM. NF-kappaB regulation in the immune system. *Nat Rev Immunol*. 2002; 2:725–34. [PubMed: 12360211]
36. Dauer DJ, Ferraro B, Song L, Yu B, Mora L, Buettner R, Enkemann S, Jove R, Haura EB. Stat3 regulates genes common to both wound healing and cancer. *Oncogene*. 2005; 24:3397–408. [PubMed: 15735721]
37. Cho HJ, Kang JH, Kwak JY, Lee TS, Lee IS, Park NG, Nakajima H, Magae J, Chang YC. Ascofuranone suppresses PMA-mediated matrix metalloproteinase-9 gene activation through the Ras/Raf/MEK/ERK- and Ap1-dependent mechanisms. *Carcinogenesis*. 2007; 28:1104–10. [PubMed: 17114644]
38. Ondrey FG, Dong G, Sunwoo J, Chen Z, Wolf JS, Crowl-Bancroft CV, Mukaida N, Van Waes C. Constitutive activation of transcription factors NF-(kappa)B, AP-1, and NF-IL6 in human head and neck squamous cell carcinoma cell lines that express pro-inflammatory and pro-angiogenic cytokines. *Molecular carcinogenesis*. 1999; 26:119–29. [PubMed: 10506755]
39. Song JI, Grandis JR. STAT signaling in head and neck cancer. *Oncogene*. 2000; 19:2489–95. [PubMed: 10851047]
40. Vandermark ER, Deluca KA, Gardner CR, Marker DF, Schreiner CN, Strickland DA, Wilton KM, Mondal S, Woodworth CD. Human papillomavirus type 16 E6 and E 7 proteins alter NF-kB in cultured cervical epithelial cells and inhibition of NF-kB promotes cell growth and immortalization. *Virology*. 2012; 425:53–60. [PubMed: 22284893]
41. Sato O, Kuriki C, Fukui Y, Motojima K. Dual promoter structure of mouse and human fatty acid translocase/CD36 genes and unique transcriptional activation by peroxisome proliferator-activated receptor alpha and gamma ligands. *J Biol Chem*. 2002; 277:15703–11. [PubMed: 11867619]
42. Bancroft CC, Chen Z, Yeh J, Sunwoo JB, Yeh NT, Jackson S, Jackson C, Van Waes C. Effects of pharmacologic antagonists of epidermal growth factor receptor, PI3K and MEK signal kinases on NF-kappaB and AP-1 activation and IL-8 and VEGF expression in human head and neck squamous cell carcinoma lines. *International journal of cancer Journal international du cancer*. 2002; 99:538–48. [PubMed: 11992543]
43. Lee SJ, Qin H, Benveniste EN. Simvastatin inhibits IFN-gamma-induced CD40 gene expression by suppressing STAT-1alpha. *J Leukoc Biol*. 2007; 82:436–47. [PubMed: 17507688]
44. McArdle MA, Finucane OM, Connaughton RM, McMorrow AM, Roche HM. Mechanisms of obesity-induced inflammation and insulin resistance: insights into the emerging role of nutritional strategies. *Frontiers in endocrinology (Lausanne)*. 2013; 4:52.
45. Lee TL, Yeh J, Friedman J, Yan B, Yang X, Yeh NT, Van Waes C, Chen Z. A signal network involving coactivated NF-kappaB and STAT3 and altered p53 modulates BAX/BCL-XL expression and promotes cell survival of head and neck squamous cell carcinomas. *International journal of cancer Journal international du cancer*. 2008; 122:1987–98. [PubMed: 18172861]
46. Gaykalova DA, Mambo E, Choudhary A, Houghton J, Buddavarapu K, Sanford T, Darden W, Adai A, Hadd A, Latham G, Danilova LV, Bishop J, et al. Novel insight into mutational landscape of head and neck squamous cell carcinoma. *PloS one*. 2014; 9:e93102. [PubMed: 24667986]
47. Freed-Pastor WA, Prives C. Mutant p53: one name, many proteins. *Genes & development*. 2012; 26:1268–86. [PubMed: 22713868]
48. Redmond AM, Carroll JS. Defining and targeting transcription factors in cancer. *Genome biology*. 2009; 10:311. [PubMed: 19664186]
49. Yeh JE, Toniolo PA, Frank DA. Targeting transcription factors: promising new strategies for cancer therapy. *Curr Opin Oncol*. 2013; 25:652–8. [PubMed: 24048019]
50. Turkson J, Kim JS, Zhang S, Yuan J, Huang M, Glenn M, Haura E, Sebt S, Hamilton AD, Jove R. Novel peptidomimetic inhibitors of signal transducer and activator of transcription 3 dimerization and biological activity. *Mol Cancer Ther*. 2004; 3:261–9. [PubMed: 15026546]
51. Sen M, Thomas SM, Kim S, Yeh JI, Ferris RL, Johnson JT, Duvvuri U, Lee J, Sahu N, Joyce S, Freilino ML, Shi H, et al. First-in-human trial of a STAT3 decoy oligonucleotide in head and neck tumors: implications for cancer therapy. *Cancer discovery*. 2012; 2:694–705. [PubMed: 22719020]

**WHAT'S NEW**

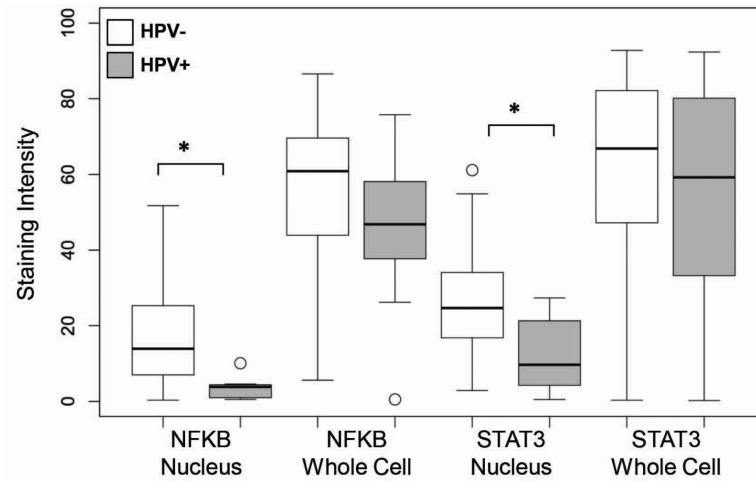
This is the first comprehensive genome wide analysis of transcription factors (TFs) in primary head and neck squamous cell carcinoma (HNSCC). TF activity was estimated globally using a novel inferential approach that accounts for gene silencing and loss of hetero- and homo-zygosity. This approach also generated biomarkers of TF activity linked to HPV status, which was validated in two additional cohorts using different molecular measurement technologies. Using the employed inferential approach we developed a top-scoring-pair biomarkers based panel. Using this panel we defined therapeutic response in cell line models. These data indicate that TF based analyses of HNSCC have potential for defining therapeutic responsiveness to TF based targeted therapies.

**Figure 1.**

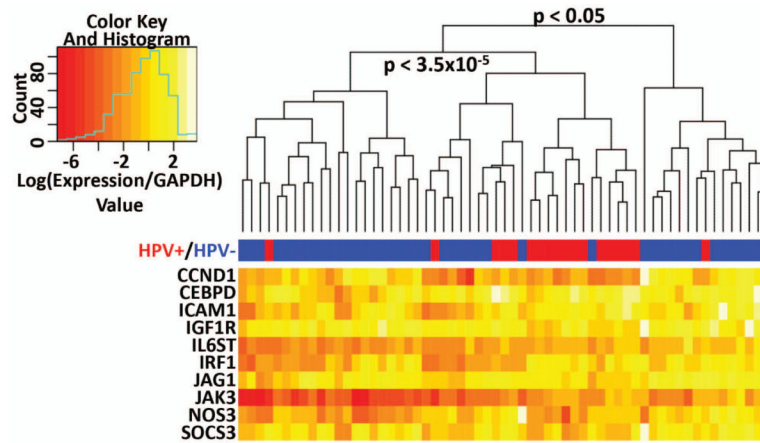
Experimental set up. *Discovery cohort*, 44 HNSCC (13 HPV+ and 31 HPV-) and 25 UPPP. Exon Array data for discovery cohort samples was normalized by RMA. 2600 human TF were depicted from TRANSFAC database and were annotated by highly relevant experimentally validated target genes. The list of TF was reduced by removal of TF with less than 5 targets. For the remaining 1325 TFs, the target genes from the TF gene sets were removed if their expression was silenced by hypermethylation or DNA copy loss (using methylation array and SNP assay data) on individual sample basis. Resulting TF target gene sets were then used to compare individual samples to pool of normal by pseudo-t test to evaluate patient specific TF activity. Control-normalized TF gene sets were compared in HPV+ vs HPV- groups. RMA-normalized exon array data alone was used in TSP to define HPV status from TF gene sets for API1, STAT1, STAT3 and NF-κB pathways. *TCGA cohort*, 279 HNSCC (35 HPV+ and 244 HPV-) and 50 matched normal. RNA-Seq data was used to evaluate individual gene expression. Target genes from the 1325 TF gene sets were used and were corrected for genes silenced by hypermethylation or DNA copy loss (using methylation array and SNP assay data) and compared for HPV+ and HPV- patients on an individual sample basis similar to the discovery cohort. RNA-Seq was used to validate TSP analysis. *Validation cohort*, 61 HNSCC (18 HPV+ and 43 HPV-) and 28 normal UPPP, was used for validation of TSP by qRT-PCR. *ECOG Validation Cohort*, 100 HNSCC (10 HPV+, 90 HPV-) and 13 normal controls, was used in IHC experiments.



**Figure 2.** IHC staining of NF- $\kappa$ B and STAT3. TMA containing 100 HNSCC and 13 control tissues was stained against total NF- $\kappa$ B (RELA) and STAT3 antibodies. Representative HNSCC samples of NF- $\kappa$ B and STAT3 co-staining are shown. (a) HPV+ samples. (b) HPV- samples. Samples are arranged from stronger staining and better co-staining on the left to no staining on the right. Both markers co-stain in cellular nuclei of cancer tissues (Spearman correlation 0.30,  $p=0.003$ ).



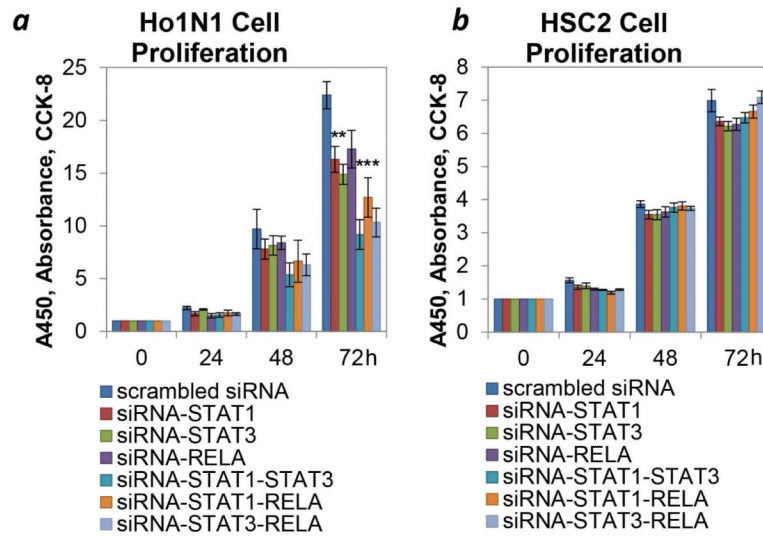
**Figure 3.** Association of protein level with HPV infection. The level of antibody staining was quantified by the Aperio in the whole cell (total protein expression) or in the nuclei (protein activation) for NF- $\kappa$ B and STAT3. The staining intensity was averaged for three tissue replicates. (\*  $p < 0.05$ ).



**Figure 4.**

Separation of HNSCC by the expression of the five top-scoring gene pairs. The expression of the top five gene pairs discovered by TSP was analyzed by qRT-PCR on 61 HNSCC and 28 UPPP samples from the validation cohort. Logarithm-converted values of GAPDH-relative expression for these genes were used for heatmap preparation in R script. Tumor samples were separated on three groups by unsupervised hierarchical clustering. The middle group of 25 HNSCC contains 16 out of total 18 HPV+ patients. P-values were calculated by Fisher exact test.





**Figure 5.** Cooperative downregulation of STATs and NF- $\kappa$ B pathways in HNSCC cell lines. Ho1N1 (a, over-expresser) and HSC2 (b, baseline) cell lines were used in experiments with single and combined downregulation of STAT1, STAT3 and NF- $\kappa$ B subunit – RELA. Cell proliferation of Ho1N1 and HSC2 with or without of gene specific knockdown of STAT1, STAT3 and RELA was measured by CCK8. Values are mean  $\pm$  SEM, the experiments were performed in pentaplicates. (\* indicates a significant difference ( $p < 0.05$ ) between the control (scrambled) and targeted RNAi experiments).

**Table 1**

Top 5 pairs of target genes discriminate HPV– and HPV+ patients from different cohorts

	<b>Discovery Cohort</b>	<b>TCGA</b>	<b>Validation Cohort</b>
p value	$9.5 \times 10^{-6}$	$6.7 \times 10^{-8}$	0.0006
odds ratio [95% CI]	44 [5.0 - 2185]	8.9 [3.6 26]	9.6 [2.2 60]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

## Research Article

# The Relationship of MiR-21, MiR-126 and MiR-205 to P-Glycoprotein, MRP1 and LRP/MVP in Non-Small Cell Lung Cancer

Zizkova V<sup>1†</sup>, Skarda J<sup>1,2†</sup>, Janikova M<sup>1,2\*</sup>, Luzna P<sup>3</sup>, Radova L<sup>2</sup>, Kurfurstova D<sup>1</sup> and Kolar Z<sup>1,2</sup>

<sup>1</sup>Department of Clinical and Molecular Pathology and Laboratory of Molecular Pathology, Palacky University Olomouc and University Hospital Olomouc, Czech Republic

<sup>2</sup>Institute of Molecular and Translational Medicine, Palacky University Olomouc and University Hospital Olomouc, Czech Republic

<sup>3</sup>Department of Histology and Embryology, Palacky University Olomouc and University Hospital Olomouc, Czech Republic

†Authors' contributions: Veronika Zizkova and Jozef Skarda contributed equally to this work

\*Corresponding author: Janikova M, Department of Clinical and Molecular Pathology and Laboratory of Molecular Pathology, Institute of Molecular and Translational Medicine, Palacky University Olomouc and University Hospital Olomouc, Hnevotinska 3, 775 15 Olomouc, Czech Republic

Received: May 20, 2015; Accepted: June 15, 2015;

Published: July 29, 2015

## Abstract

Protein transporters P-gp, MRP1 and LRP/MVP participate in the emergence of multidrug resistance (MDR) in non-small cell lung cancer (NSCLC). Their expression is post-transcriptionally regulated by microRNAs (miRNAs). Dysregulation of miR-21, miR-126 and miR-205 is often found in NSCLC. The aim of this study was to determine whether the level of miRNAs is associated with expression of the above mentioned proteins involved in MDR and whether they can be used as prognostic and diagnostic markers. We analysed miR-21, miR-126 and miR-205 in various histological subtypes of NSCLC. Their expression was then correlated with clinico-pathological characteristics, such as progression free survival (PFS), overall survival (OS) and different histological subtypes of NSCLC and, with expression of P-gp, MRP1 and LRP/MVP. We found no significant relationship between miR-21 and miR-126 expression and clinico-pathological parameters. However, miR-205 levels were significantly increased in squamous cell carcinomas ( $p < 10^{-6}$ ) compared with other histological subtypes of NSCLC. Additionally, the level of miR-205 inversely correlated with P-gp expression in NSCLC patients ( $p = 0.03$ ). Results of this study suggest that miR-205 could be used as a diagnostic marker and its downregulation may indicate the emergence of P-gp mediated drug resistance in NSCLC patients.

**Keywords:** microRNA; NSCLC; P-gp; MRP1; LRP/MVP

## Abbreviations

ADC: Adenocarcinoma; EGFR: Epidermal Growth Factor Receptor; FFPE: Formalin-fixed and Paraffin-embedded; HER3: v-erb-b2 Erythroblastic Leukemia Viral Oncogene Homolog 3; LCC: Large Cell Carcinoma; LRP/MVP: Lung Resistance-related Protein/major Vault Protein; MDR: Multidrug Resistance; miRNA: microRNAs; MRP1: Multidrug Resistance-associated Protein 1; NSCLC: Non-small Cell Lung Cancer; OS: Overall Survival; PFS: Progression Free Survival; P-gp: P-glycoprotein; PTEN: Phosphatase and Tensin Homologue; RNAi: RNA Interference; RT-PCR: Reverse Transcription Polymerase Chain Reaction; SCC: Squamous Cell Carcinoma; SCLC: Small Cell Lung Cancer; TMA: Tissue Microarrays; VEGF: Vascular Endothelial Growth Factor; ZEB: Zinc Finger E-box Binding Homeobox

## Introduction

A major cause of cancer mortality worldwide is lung cancer and non-small cell lung cancer (NSCLC) represents approximately 80-85% of all cases. Many NSCLC patients do not respond to therapy due to the emergence of multidrug resistance (MDR). For treatment of advanced forms of NSCLC is frequently treated using radiotherapy, chemotherapy or biological therapies. In recent years, RNA interference (RNAi) has also been introduced [1-3]. RNAi is a molecular mechanism of gene silencing which inhibits gene expression at post-transcriptional level. In human it is govern by small (~22 nt)

non-coding, endogenous, single-stranded RNAs, called microRNAs (miRNAs). Based on complementarity, these molecules bind to the target mRNA, which is either completely degraded or prevented from translation, without any split [3]. Currently, according to miRBase 21, 2588 mature human miRNAs are known to regulate many protein-coding genes [4-7]. MiRNAs play an important role in many biological processes, such as proliferation, apoptosis, development, differentiation, DNA damage response and other processes. They can also affect carcinogenesis, chemoresistance and radioresistance and they are involved in diverse regulatory pathways [8-12]. For these reasons, miRNAs often dysregulated in tumours could be classified as a class of oncogenes or tumour suppressor genes and, they might be used as markers for monitoring carcinogenesis [8, 13].

The human microRNA-21 gene (*hsa-miR-21*), located on chromosome 17q23.2, has been shown to be dysregulated in e.g. breast, colon, pancreatic, stomach, prostate, ovarian and lung cancer [14-16]. Overexpression of miR-21 in lung cancer in never-smokers is probably connected with activated epidermal growth factor receptor (EGFR) signalling [17]. miR-21 supports cell proliferation because it inhibits the tumour suppressor gene Phosphatase and tensin homologue (*PTEN*). This leads to constitutively active signalling through the PI3K/Akt pathway followed by K-Ras signalling which supports survival and proliferation of tumour cells [18-20]. miR-21 also participates in other processes, such as differentiation, cell cycle progression, apoptosis, tumour invasion and DNA-damage repair

processes [21-23]. There is also evidence that the Akt2-dependent pathway activated by hypoxia can support tumour resistance via miR-21 induction. Chemoresistance accompanied by miR-21 upregulation has been found in breast and ovarian carcinomas, pancreatic cancer, prostate cancer and glioblastomas [24-28].

The expression of human microRNA-126 gene (*hsa-miR-126*), located on chromosome 9q34.3, has been shown to be dysregulated in hepatocarcinomas, breast, colorectal, cervical and lung cancer [29-31,14,15]. MiR-126 is the regulator of Vascular endothelial growth factor A (*VEGF-A*) and therefore, it has an important role in angiogenesis [29,32,33]. It is also involved in proliferation, differentiation and metastasis [31,34,35]. On the other hand, downregulation of miR-126 decreases the cytotoxic effect of gefitinib in adenocarcinomas cell lines. This can lead to the emergence of resistance to gefitinib [36].

The human microRNA-205 gene (*hsa-miR-205*), located on chromosome 1q32.2, has been shown to be upregulated in head and neck cancer, bladder cancer and in squamous cell carcinomas [37,38]. Likewise miR-21, miR-205 inhibits tumour suppressor gene *PTEN* and also regulates Zinc finger E-box binding Homeobox 1 and 2 (*ZEB1* and *ZEB2*) regulating tumour invasion [18,39]. miR-205 regulation of *v-erb-b2* erythroblastic leukemia viral oncogene homolog 3 (*HER3*) receptor activating Akt can sensitize breast cancer to treatment using the tyrosine-kinase inhibitors gefitinib and lapatinib [40]. The pro-apoptotic effect of chemotherapeutic agents was also observed in prostate cancer cells with induced expression of miR-205 [41].

The following study on miR-21, miR-126 and miR-205, whose genes are located in regions frequently amplified (*hsa-miR-21* and *hsa-miR-205*) or deleted (*hsa-miR-126*) in lung cancer was based on the work of Yanaihara et al. [14]. We correlated levels of these miRNAs with clinico-pathological status of NSCLC patients and with expression of known transporter proteins involved in MDR, such as P-glycoprotein (P-gp), Multidrug resistance-associated protein 1 (MRP1) and Lung resistance-related protein/Major vault protein (LRP/MVP). These proteins are able to efflux anti-cancer drugs from the cells which is one of the main mechanisms of MDR [2].

## Materials and Methods

### Clinical assessment and patients

Formalin-fixed and paraffin-embedded (FFPE) surgical tissue samples of lung cancer from years 1996-2000 were obtained from the archives of the Department of Clinical and Molecular Pathology, Faculty of Medicine and Dentistry, Palacky University and University Hospital, Olomouc. The group of patients comprised of 65 patients with lung cancer after approval by the ethics committee of University Hospital and Faculty of Medicine and Dentistry, Palacky University Olomouc. The cohort consists of 65 patients: 56 male and 9 female and age range 33 to 78 years. 31 tumours were classified as adenocarcinomas (ADCs), 26 as squamous cell carcinomas (SCCs), 6 as large cell carcinomas (LCCs) and 2 as small cell lung cancer (SCLC) patients. 19 patients were in stage I/II and 34 patients in stage III/IV. For the rest of the patients, the stage was unknown. The characteristics of patients are shown in Table 1. 21 patients had received chemotherapy. 18 of them underwent platinum based chemotherapy regime and the rest of patients were treated with

different chemotherapeutics, such as fluorouracil, doxorubicin or taxanes. The clinico-pathological parameters progression free survival (PFS) and overall survival (OS) were monitored over than 15 years.

### Tissue microarray construction

Tumour tissue microarrays (TMA) were constructed using 65 formalin-fixed and paraffin-embedded primary lung cancer specimens. The tissue areas for sampling were selected based on visual alignment with the corresponding H&E stained section. Two to four tissue cores were taken from a tumour block and were replaced into a recipient paraffin block with a tissue microarrayer Galileo TMA CK3500 (BioRep, Milan, Italy).

### Immunohistochemical staining of P-gp, MRP1 and LRP/MVP

Indirect immunohistochemistry was used. The TMA sections were deparaffinized. Then LRP/MVP antigen was unmasked in citrate buffer (pH 6) and MRP1 in Target Retrieval Solution, High pH (10x) (Dako, Denmark). The sections for detection of P-gp were not pre-treated because the monoclonal antibody UIC2 (1864, Immunotech, Marseille, France), used at a dilution of 1:50, targets the extracellular epitope of P-gp protein [42]. The Monoclonal Antibody to MRP1 (human) (MRPm5) (801-012-C250, Alexis, Lausen, Switzerland) at a dilution of 1:25 and Monoclonal Antibody to LRP [MVP] (human) (LMR5) (801-014-C025, Alexis, Lausen, Switzerland) at a dilution 1:20 were used. Visualisation was made by EnVisionTM+ Dual Link (Dako, Denmark). Nuclei were counterstained with haematoxylin. The preparations were observed under an optical microscope and images were captured with a DP71 camera (Olympus, Tokyo, Japan). The expression of P-gp, MRP1 and LRP/MVP were semi quantitatively assessed in stained tissue sections by estimation of the percentage of positive cells as very low ( $\leq 10\%$ ), low ( $\leq 30\%$ ), moderate ( $\leq 60\%$ ) or high ( $> 60\%$ ).

### Isolation of total RNA

Total RNA isolation from corresponding tissue cores obtained by tissue microarrayer Galileo TMA CK3500 [43] was performed using the RecoverAll™ Total Nucleic Acid Isolation Kit for FFPE (Applied Biosystems, Foster City, CA, USA) according to the manufacturer's instructions. All preparation and handling steps of RNA were performed under RNase-free conditions. The concentration of total RNA was measured using NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, Delaware, USA) and then RNA was stored at  $-80^{\circ}\text{C}$  until use.

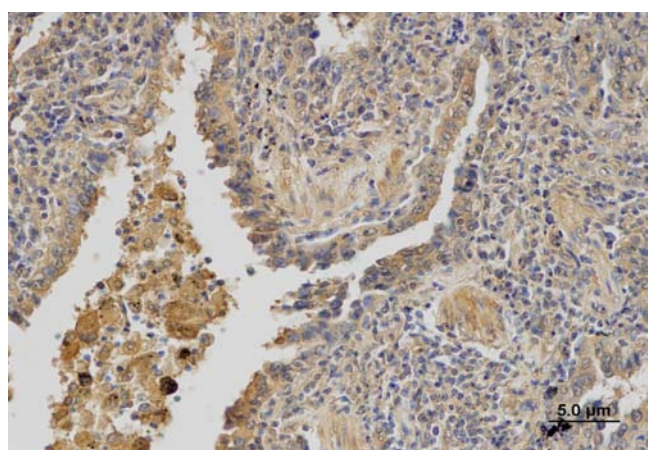
### Reverse transcription

TaqMan® MicroRNA Reverse Transcription Kit (Applied Biosystems, Foster City, CA, USA) was used according to the manufacturer protocol. 10 ng of total RNA was used for reverse transcription in total volume of reaction 15  $\mu\text{l}$ . Pooled gene-specific primers of miR-21, miR-126, miR-205 and RNU6B (RNA, U6 small nuclear 2) (Applied Biosystems, Foster City, CA, USA) were added to the reverse transcription polymerase chain reaction (RT-PCR). RT-PCR product was then preamplified according to protocol using TaqMan® PreAmp Master Mix (2x) (Applied Biosystems, Foster City, CA, USA) with pooled TaqMan® MicroRNA Assays (Applied Biosystems, Foster City, CA, USA) of miR-21, miR-126, miR-205 and RNU6B.

**Table 1:** Characteristics of patients.

All patients	Sex		Age		Grade		Histological subtype of NSCLC			SCLC
	female	male	<60	≥60	I/II	III/IV	ADC	SCC	LCC	
65	9	56	28	37	19	34	31	26	6	2

NSCLC – non-small cell lung cancer, SCLC – small cell lung cancer, ADC – adenocarcinoma, SCC – squamous cell carcinoma, LCC – large cell carcinoma

**Figure 1a:** Immunohistochemical detection of P-gp.

### Real-time PCR for miRNAs quantification

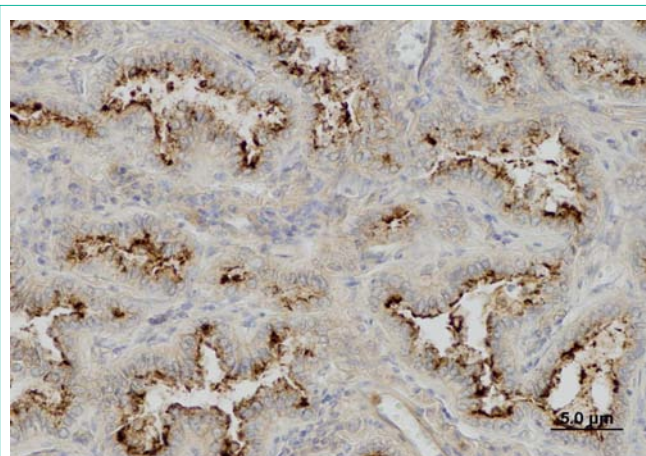
The pre-PCR product was diluted in TE buffer (1:20). The reaction consisted of TaqMan® Universal PCR Master Mix (2x) (Applied Biosystems, Foster City, CA, USA), nuclease-free water and corresponding TaqMan® MicroRNA Assay and diluted pre-PCR product. Each sample was analysed in triplicate and volume per each reaction was 10 μl. Real-time PCR was performed using LightCycler® 480 (Roche, Branford, CT, USA) according to protocol. As an endogenous control, RNU6B was used. The control sample was prepared by mixing the same amounts of commercially available Human Lung Total RNAs (AM7968, Lot. 0904002 and 1203010, Applied Biosystems, Foster City, CA, USA) and Total RNAs, Lung, Human (540019, Lot. 0006051745 and 0006079356, Agilent Technologies, Santa Clara, CA, USA).

### Statistical analysis

The relative level of expression of assessed miRNAs was calculated using the following equation: relative gene expression =  $2^{-(\Delta C_t \text{ sample} - \Delta C_t \text{ control})}$ . The difference among controls was determined using ANOVA (analysis of variance) test. For the correlation of expression levels of tested miRNAs and proteins with histological subtypes a Kruskal-Wallis test was used. PFS and OS were determined using Kaplan-Meier analysis. The significance of P-gp, MRP1 and LRP/MVP levels in relation to the higher or lower miRNAs expression in NSCLC were determined by the Mann-Whitney U test.  $p \leq 0.05$  was considered statistically significant.

## Results

TMA blocks from 65 patients were immunohistochemically analysed for P-gp, MRP1 and LRP/MVP expression. The tissues from TMAs were also used for the detection of expression levels of miR-21, miR-126 and miR-205.

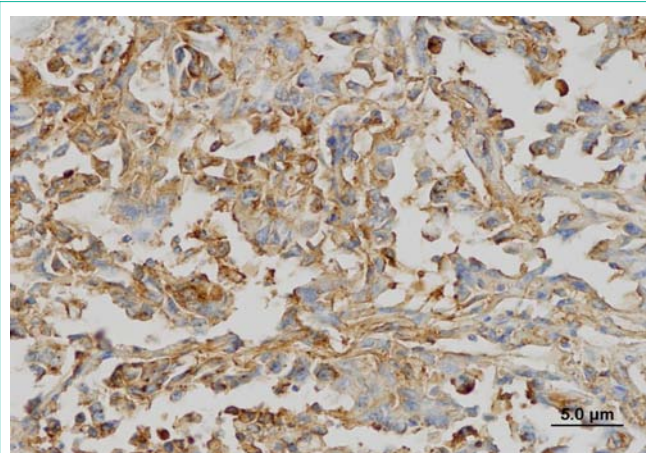
**Figure 1b:** Immunohistochemical detection of MRP1.

### Expression of P-gp, MRP1 and LRP/MVP in NSCLC

The expression of P-gp, MRP1 and LRP/MVP was assessed in 62 patients. Three patients were excluded from the analysis; two of them for the diagnosis of small-cell lung carcinoma and one for sampling from the metastasis of ADC into adrenal gland. The P-gp was detected in 39 (62.9 %) patients, in 18 (29 %) its expression was high (Figure 1a). Expression of MRP1 was found in 48 (77.4 %) patients. It was high in 26 (41.9 %) patients (Figure 1b). The LRP/MVP was detected in 37 (59.7 %) patients; high expression was in 20 (32.3 %) patients (Figure 1c). Their expression in different histological subtypes is summarised in Table 2. There was no significant difference in P-gp, MRP1 and LRP/MVP expression between ADC, SCC and LCC specimens. For the next analysis, very low expression of proteins was considered as a negative expression while higher expression levels (low, moderate and high) were considered as a positive expression.

### Expression of miR-21, miR-126 and miR-205 in NSCLC

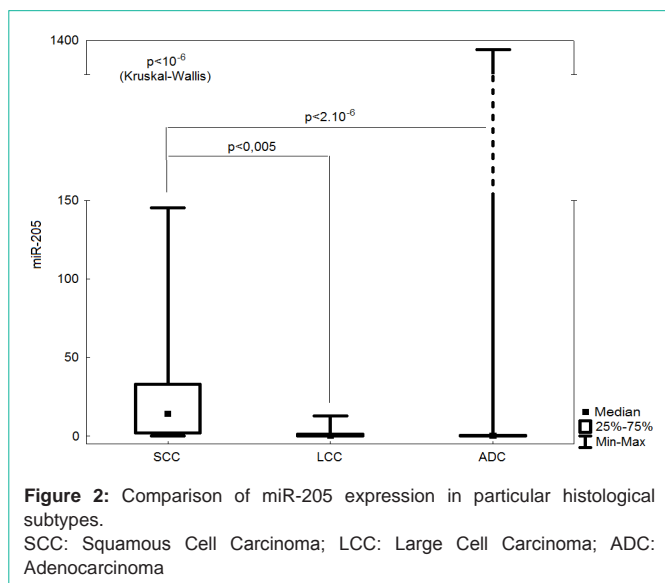
The relative quantification of mature miR-21, miR-126 and miR-205 in 62 lung cancer samples was performed. Overexpression of miR-21 was found in 7 (11.3 %) samples, miR-126 was upregulated in 3 (4.8 %) samples and miR-205 was overexpressed in 30 (48.4 %) samples. There was no significant difference in miR-21 and miR-126 expression between ADC, SCC and LCC specimens. The expression

**Figure 1c:** Immunohistochemical detection of LRP/MVP.

**Table 2:** Expression of P-gp, MRP1 and LRP/MVP in different histological subtypes of NSCLC.

Protein	Intensity	ADC [No. (%)]	SCC [No. (%)]	LCC [No. (%)]	Total [No. (%)]
P-gp	very low	9 (30.0)	12 (46.2)	2 (33.3)	23 (37.1)
	low	7 (23.3)	6 (23.1)	3 (50.0)	16 (25.8)
	moderate	3 (10.0)	2 (7.7)	0 (0)	5 (8.1)
	high	11 (36.7)	6 (23.1)	1 (16.7)	18 (29.0)
	<b>Total</b>	<b>30</b>	<b>26</b>	<b>6</b>	<b>62</b>
MRP1	very low	6 (20.0)	7 (26.9)	1 (16.7)	14 (22.6)
	low	9 (30.0)	7 (26.9)	1 (16.7)	17 (27.4)
	moderate	4 (13.3)	1 (3.8)	0 (0)	5 (8.1)
	high	11 (36.7)	11 (42.3)	4 (66.7)	26 (41.9)
	<b>Total</b>	<b>30</b>	<b>26</b>	<b>6</b>	<b>62</b>
LRP/MVP	very low	11 (36.7)	13 (50.0)	1 (16.7)	25 (40.3)
	low	4 (13.3)	2 (7.7)	3 (50.0)	9 (14.5)
	moderate	3 (10.0)	5 (19.2)	0 (0)	8 (12.9)
	high	12 (40.0)	6 (23.1)	2 (33.3)	20 (32.3)
	<b>Total</b>	<b>30</b>	<b>26</b>	<b>6</b>	<b>62</b>

P-gp: P-glycoprotein; MRP1: Multidrug Resistance-associated Protein 1, LRP/MVP: Lung Resistance-related Protein/major Vault Protein; ADC: Adenocarcinoma; SCC: Squamous Cell Carcinoma; LCC: Large Cell Carcinoma



of miR-205 was significantly higher in SCCs than in other histological subtypes ( $p < 10_{-6}$ ) (Figure 2).

**Analysis of relationship between P-gp, MRP1, LRP/MVP and miRNAs status and, PFS and OS**

PFS and OS were used as the clinico-pathological parameters for evaluating the response of patients to the therapy and their survival status. The surveillance analysis did not show the significant differences between the expression of P-gp, MRP1, LRP/MVP and tested miRNAs and, PFS and OS. Expression of these proteins was subsequently correlated with the expression levels of miR-21, mir-126 and miR-205.

**Correlation of miRNAs status with P-gp, MRP1 and LRP/MVP expression**

Neither miR-21 nor miR-126 levels correlated with P-gp, MRP1 and LRP/MVP expression. Also the level of miR-205 was not linked with MRP or LRP/MVP expression. The only significant negative correlation was found between miR-205 and P-gp expression ( $p = 0.03$ ) (Figure 3).

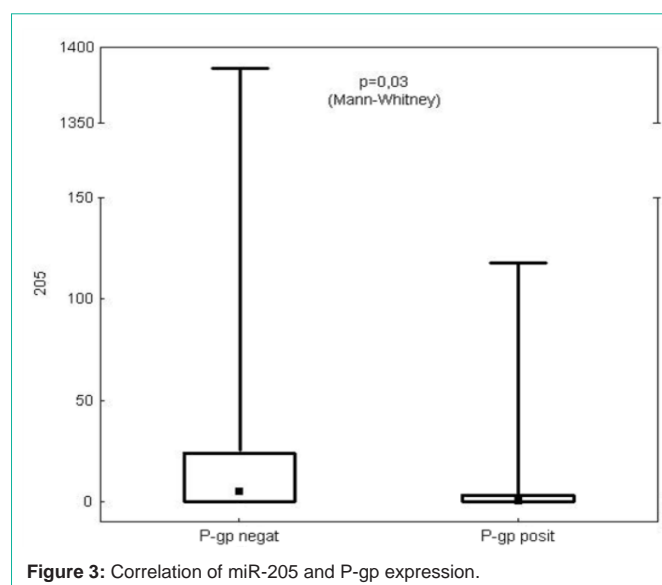
**Discussion**

In the present pilot study, data from 62 formalin-fixed and paraffin-embedded primary lung cancer samples were analysed. They were assessed against commercially available control lung RNAs obtained from non-lung cancer patients because the presence of other factors influencing miRNAs expression (e.g. inflammation) in adjacent non-tumour tissues in lung cancer patients.

In this study, no correlation of P-gp, MRP1 and LRP/MVP expression and, miR-21 and miR-126 levels with PFS and OS was found. This is in contrast to other studies in which expression of these proteins and miRNAs appeared to have prognostic value [44-47,13].

The failure of the therapy may be responsible for the shorter PFS and OS of cancer patients probably due to the emergence of MDR. P-gp, MRP1 and LRP/MVP are considered as prognostic markers with regard to the emergence of MDR associated with their expression during chemotherapy [44,45]. Results of this pilot study were not consistent with previous studies. It is probably caused by little cohort of analysed patients because our previous unpublished results obtained from larger group of patients (more than 200) were in agreement with this claim.

In many studies using cell lines, involvement of miR-21 in MDR has been described [24,48-50]. A few studies only have been carried out on NSCLC patient samples. In the majority of publications, miR-21 overexpression is a suggested negative prognostic marker for NSCLC [46,47,13] but in some, this significance was not validated. The downregulation of miR-21 has also been observed and Wang et al. concluded that this may be associated with sensitivity of NSCLC patients to radiotherapy. In this study, the levels of



miR-21 in radiosensitive patient's samples were assessed against radioresistant patient's samples [51]. There were no available data on the radioresistance / radiosensitivity status of our cohort of patients. However, due to sample collection mainly before therapy, patient radiosensitivity might have been a reason for miR-21 downregulation as downregulation of miR-21 was detected in 55 (88.7 %) samples.

Recently, different expression level of miR-21 was observed in distinct tumour areas of colorectal cancer patients. Nielsen et al. found 6-fold higher level of miR-21 in stromal cells compared to that in cancer cells [52]. For the purpose of this study, TMA cores for RNA isolation and quantification were taken from the tumour areas containing cancer cells, not stromal cells. This might explain the finding of lower levels of miR-21 mostly detected in our patients.

Downregulation of miR-126 is commonly observed in NSCLC [34,53] and our results are in accordance with this. In addition, it has been proposed that miR-126 might have a tumour-suppressor effect during tumorigenesis [31].

Our results suggest miR-205 could be used as a diagnostic marker for distinguishing SCCs from ADCs and other histological subtypes ( $p < 10_{-6}$ ). This is consistent with other studies showing the role of miR-205 in NSCLC diagnosis [54-56]. In addition, the relationship of miR-205 and P-gp expression has been observed in this study ( $p = 0.03$ ). So far, there is no verified explanation for this phenomenon but RNA22-HSA algorithm predicted P-gp as one of the targets of miR-205 ([http://cm.jefferson.edu/rna22v1.0-homo\\_sapiens](http://cm.jefferson.edu/rna22v1.0-homo_sapiens)) [57].

P-gp is the main executor of so-called typical MDR. It is responsible for the efflux of some anti-cancer drugs from cells which are one of the major mechanisms of drug resistance. The emergence of the resistance can cause relapse and therefore, its earlier detection might prevent MDR to fully develop. In some cases, typical MDR can be overcome using compounds, such as Cyclosporine A or Tamoxifen [58,59,2]. We present initial results that indicate the possible role of miR-205 in typical MDR caused by P-gp. Validation of these results in a larger cohort of cases is however required.

## Conclusion

The recognition of MDR in NSCLC patients is important for decision to change the ineffective chemotherapy to different effective treatment. Detection of molecules regulating expression of proteins involved in MDR may help to reveal these processes earlier. MiR-205 seems to regulate P-gp, which is one of the executors of typical MDR. Therefore, downregulation of miR-205 might be a sign of higher P-gp expression and thus, activation of MDR mechanisms.

## Acknowledgement

We would like to thank Jiri Klein and Vitezslav Kolek for kind help with collection and characterisation of patients. We would like to thank also Eva Sedlakova for performance of immunohistochemical staining. This work was supported by grants: IGA MZCR NT13569, LF\_2014\_003 and NPU I LO1304.

## References

- D'Addario G, Früh M, Reck M, Baumann P, Klepetko W, Felip E; ESMO Guidelines Working Group. Metastatic non-small-cell lung cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol.* 2010; 21 Suppl 5: v116-119.
- Noskova V, Hajdich M, Mihal V, Cwierzka K. Mechanizmy mnohocetné lékové rezistence a jejich význam pro klinickou praxi, I. typická MDR. *Klin Onkol.* 2010; 13: 4-9.
- Paranjape T, Choi J, Weidhaas JB. MicroRNA as Potential Diagnostics and Therapeutics. In: *MicroRNAs in Development and Cancer*. Slack FJ, editor. Imperial College Press. 2011; 213-236.
- Griffiths-Jones S. The microRNA Registry. *Nucleic Acids Res.* 2004; 32 (Database issue): D109-111.
- Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res.* 2006; 34: D140-144.
- Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ. miRBase: tools for microRNA genomics. *Nucleic Acids Res.* 2008; 36 (Database issue): D154-158.
- Kozomara A, Griffiths-Jones S. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res.* 2011; 39: D152-157.
- Ventura A, Kumar MS, Jacks T. Roles of microRNAs in cancer and development. In: *MicroRNAs: From Basic Science to Disease Biology*. Appasani K, editor. Cambridge University Press. 2008; 322-337.
- O'Connell RM, Baltimore D. MicroRNAs and hematopoietic cell development. *Curr Top Dev Biol.* 2012; 99: 145-174.
- Mansfield JH, McGlenn E. Evolution, expression, and developmental function of Hox-embedded miRNAs. *Curr Top Dev Biol.* 2012; 99: 31-57.
- Wan G, Mathur R, Hu X, Zhang X, Lu X. miRNA response to DNA damage. *Trends Biochem Sci.* 2011; 36: 478-484.
- Ueno K, Hirata H, Hinoda Y, Dahiya R. Frizzled homolog proteins, microRNAs and Wnt signaling in cancer. *Int J Cancer.* 2013; 132: 1731-1740.
- Wang Q, Wang S, Wang H, Li P, Ma Z. MicroRNAs: novel biomarkers for lung cancer diagnosis, prediction and treatment. *Exp Biol Med (Maywood).* 2012; 237: 227-235.
- Yanaiharu N, Caplen N, Bowman E, Seike M, Kumamoto K, Yi M, et al. Unique microRNA molecular profiles in lung cancer diagnosis and prognosis. *Cancer Cell.* 2006; 9: 189-198.
- Koturbash I, Zemp FJ, Pogribny I, Kovalchuk O. Small molecules with big effects: the role of the microRNAome in cancer and carcinogenesis. *Mutat Res.* 2011; 722: 94-105.
- Sarkar FH, Li Y, Wang Z, Kong D, Ali S. Implication of microRNAs in drug resistance for designing novel cancer therapy. *Drug Resist Updat.* 2010; 13: 57-66.
- Seike M, Goto A, Okano T, Bowman ED, Schetter AJ, Horikawa I, et al. miR-21 is an EGFR-regulated anti-apoptotic factor in lung cancer in never-smokers. *Proc Natl Acad Sci U S A.* 2009; 106: 12085-12090.
- Zhang JG, Wang JJ, Zhao F, Liu Q, Jiang K, Yang GH, et al. MicroRNA-21 (miR-21) represses tumor suppressor PTEN and promotes growth and invasion in non-small cell lung cancer (NSCLC). *Clin Chim Acta.* 2010; 411: 846-852.
- Hatley ME, Patrick DM, Garcia MR, Richardson JA, Bassel-Duby R, van Rooij E, et al. Modulation of K-Ras-dependent lung tumorigenesis by MicroRNA-21. *Cancer Cell.* 2010; 18: 282-293.
- Frezzetti D, De Menna M, Zoppoli P, Guerra C, Ferraro A, Bello AM, et al. Upregulation of miR-21 by Ras in vivo and its role in tumor growth. *Oncogene.* 2011; 30: 275-286.
- Gao W, Xu J, Liu L, Shen H, Zeng H, Shu Y. A systematic-analysis of predicted miR-21 targets identifies a signature for lung cancer. *Biomed Pharmacother.* 2012; 66: 21-28.
- Gabriely G, Wurdinger T, Kesari S, Esau CC, Burchard J, Linsley PS, et al. MicroRNA 21 promotes glioma invasion by targeting matrix metalloproteinase regulators. *Mol Cell Biol.* 2008; 28: 5369-5380.
- Valeri N, Gasparini P, Braconi C, Paone A, Lovat F, Fabbri M, et al.

- MicroRNA-21 induces resistance to 5-fluorouracil by down-regulating human DNA MutS homolog 2 (hMSH2). *Proc Natl Acad Sci U S A*. 2010; 107: 21098-21103.
24. Polytarchou C, Iliopoulos D, Hatzia Apostolou M, Kottakis F, Maroulakou I, Struhl K, et al. Akt2 regulates all Akt isoforms and promotes resistance to hypoxia through induction of miR-21 upon oxygen deprivation. *Cancer Res*. 2011; 71: 4720-4731.
  25. Mei M, Ren Y, Zhou X, Yuan XB, Han L, Wang GX, et al. Downregulation of miR-21 enhances chemotherapeutic effect of taxol in breast carcinoma cells. *Technol Cancer Res Treat*. 2010; 9: 77-86.
  26. Hwang JH, Voortman J, Giovannetti E, Steinberg SM, Leon LG, Kim YT, et al. Identification of microRNA-21 as a biomarker for chemoresistance and clinical outcome following adjuvant therapy in resectable pancreatic cancer. *PLoS One*. 2010; 5: e10630.
  27. Shi GH, Ye DW, Yao XD, Zhang SL, Dai B, Zhang HL, et al. Involvement of microRNA-21 in mediating chemo-resistance to docetaxel in androgen-independent prostate cancer PC3 cells. *Acta Pharmacol Sin*. 2010; 31: 867-873.
  28. Shi L, Chen J, Yang J, Pan T, Zhang S, Wang Z. MiR-21 protected human glioblastoma U87MG cells from chemotherapeutic drug temozolomide induced apoptosis by decreasing Bax/Bcl-2 ratio and caspase-3 activity. *Brain Res*. 2010; 1352: 255-264.
  29. Wang S, Aurora AB, Johnson BA, Qi X, McAnally J, Hill JA, et al. The endothelial-specific microRNA miR-126 governs vascular integrity and angiogenesis. *Dev Cell*. 2008; 15: 261-271.
  30. Negrini M, Calin GA. Breast cancer metastasis: a microRNA story. *Breast Cancer Res*. 2008; 10: 203.
  31. Feng R, Chen X, Yu Y, Su L, Yu B, Li J, et al. miR-126 functions as a tumour suppressor in human gastric cancer. *Cancer Lett*. 2010; 298: 50-63.
  32. Heusschen R, van Gink M, Griffioen AW, Thijssen VL. MicroRNAs in the tumor endothelium: novel controls on the angioregulatory switchboard. *Biochim Biophys Acta*. 2010; 1805: 87-96.
  33. Sandler A, Gray R, Perry MC, Brahmer J, Schiller JH, Dowlati A, et al. Paclitaxel-carboplatin alone or with bevacizumab for non-small-cell lung cancer. *N Engl J Med*. 2006; 355: 2542-2550.
  34. Crawford M, Brawner E, Batte K, Yu L, Hunter MG, Otterson GA, et al. MicroRNA-126 inhibits invasion in non-small cell lung carcinoma cell lines. *Biochem Biophys Res Commun*. 2008; 373: 607-612.
  35. Huang F, Fang ZF, Hu XQ, Tang L, Zhou SH, Huang JP, et al. Overexpression of miR-126 promotes the differentiation of mesenchymal stem cells toward endothelial cells via activation of PI3K/Akt and MAPK/ERK pathways and release of paracrine factors. *Biol Chem*. 2013; 394: 1223-1233.
  36. Zhong M, Ma X, Sun C, Chen L. MicroRNAs reduce tumor growth and contribute to enhance cytotoxicity induced by gefitinib in non-small cell lung cancer. *Chem Biol Interact*. 2010; 184: 431-438.
  37. Jiang J, Lee EJ, Gusev Y, Schmittgen TD. Real-time expression profiling of microRNA precursors in human cancer cell lines. *Nucleic Acids Res*. 2005; 33: 5394-5403.
  38. Gottardo F, Liu CG, Ferracin M, Calin GA, Fassan M, Bassi P, et al. MicroRNA profiling in kidney and bladder cancers. *Urol Oncol*. 2007; 25: 387-392.
  39. Greene SB, Gunaratne PH, Hammond SM, Rosen JM. A putative role for microRNA-205 in mammary epithelial cell progenitors. *J Cell Sci*. 2010; 123: 606-618.
  40. Iorio MV, Casalini P, Piovon C, Di Leva G, Merlo A, Triulzi T, et al. microRNA-205 regulates HER3 in human breast cancer. *Cancer Res*. 2009; 69: 2195-2200.
  41. Bhatnagar N, Li X, Padi SK, Zhang Q, Tang MS, Guo B, et al. Downregulation of miR-205 and miR-31 confers resistance to chemotherapy-induced apoptosis in prostate cancer cells. *Cell Death Dis*. 2010; 1: e105.
  42. Kennedy BG, Mangini NJ. P-glycoprotein expression in human retinal pigment epithelium. *Mol Vis*. 2002; 8: 422-430.
  43. Cardano M, Diaferia GR, Falavigna M, Spinelli CC, Sessa F, DeBlasio P, et al. Cell and tissue microarray technologies for protein and nucleic acid expression profiling. *J Histochem Cytochem*. 2013; 61: 116-124.
  44. Trussardi-Regnier A, Millot JM, Gorisse MC, Delvincourt C, Prevost A. Detection of drug-resistance genes using single bronchoscopy biopsy specimens. *Oncol Rep*. 2007; 18: 703-708.
  45. Oue T, Yoneda A, Uehara S, Yamanaka H, Fukuzawa M. Increased expression of multidrug resistance-associated genes after chemotherapy in pediatric solid malignancies. *J Pediatr Surg*. 2009; 44: 377-380.
  46. Markou A, Tsaroucha EG, Kaklamanis L, Fotinou M, Georgoulis V, Lianidou ES. Prognostic value of mature microRNA-21 and microRNA-205 overexpression in non-small cell lung cancer by quantitative real-time RT-PCR. *Clin Chem*. 2008; 54: 1696-1704.
  47. Gao W, Shen H, Liu L, Xu J, Xu J, Shu Y, et al. MiR-21 overexpression in human primary squamous cell lung carcinoma is associated with poor patient prognosis. *J Cancer Res Clin Oncol*. 2011; 137: 557-566.
  48. Gao W, Lu X, Liu L, Xu J, Feng D, Shu Y, et al. MiRNA-21: a biomarker predictive for platinum-based adjuvant chemotherapy response in patients with non-small cell lung cancer. *Cancer Biol Ther*. 2012; 13: 330-340.
  49. Blower PE, Chung JH, Verducci JS, Lin S, Park JK, Dai Z, et al. MicroRNAs modulate the chemosensitivity of tumor cells. *Mol Cancer Ther*. 2008; 7: 1-9.
  50. Rui W, Bing F, Hai-Zhu S, Wei D, Long-Bang C. Identification of microRNA profiles in docetaxel-resistant human non-small cell lung carcinoma cells (SPC-A1). *J Cell Mol Med*. 2010; 14: 206-214.
  51. Wang XC, Du LQ, Tian LL, Wu HL, Jiang XY, Zhang H, et al. Expression and function of miRNA in postoperative radiotherapy sensitive and resistant patients of non-small cell lung cancer. *Lung Cancer*. 2011; 72: 92-99.
  52. Nielsen BS, Jørgensen S, Fog JU, Søkilde R, Christensen IJ, Hansen U, et al. High levels of microRNA-21 in the stroma of colorectal cancers predict short disease-free survival in stage II colon cancer patients. *Clin Exp Metastasis*. 2011; 28: 27-38.
  53. Sun Y, Bai Y, Zhang F, Wang Y, Guo Y, Guo L, et al. miR-126 inhibits non-small cell lung cancer cells proliferation by targeting EGFL7. *Biochem Biophys Res Commun*. 2010; 391: 1483-1489.
  54. Lebanony D, Benjamin H, Gilad S, Ezagouri M, Dov A, Ashkenazi K, et al. Diagnostic assay based on hsa-miR-205 expression distinguishes squamous from nonsquamous non-small-cell lung carcinoma. *J Clin Oncol*. 2009; 27: 2030-2037.
  55. Bishop JA, Benjamin H, Cholakh H, Chajut A, Clark DP, Westra WH. Accurate classification of non-small cell lung carcinoma using a novel microRNA-based approach. *Clin Cancer Res*. 2010; 16: 610-619.
  56. Zhang YK, Zhu WY, He JY, Chen DD, Huang YY, Le HB, et al. miRNAs expression profiling to distinguish lung squamous-cell carcinoma from adenocarcinoma subtypes. *J Cancer Res Clin Oncol*. 2012; 138: 1641-1650.
  57. Miranda KC, Huynh T, Tay Y, Ang YS, Tam WL, Thomson AM, et al. A pattern-based method for the identification of MicroRNA binding sites and their corresponding heteroduplexes. *Cell*. 2006; 126: 1203-1217.
  58. Claudio JA, Emerman JT. The effects of cyclosporin A, tamoxifen, and medroxyprogesterone acetate on the enhancement of adriamycin cytotoxicity in primary cultures of human breast epithelial cells. *Breast Cancer Res Treat*. 1996; 41: 111-122.
  59. Urasaki Y, Ueda T, Nakamura T. Circumvention of daunorubicin resistance by a new tamoxifen derivative, toremifene, in multidrug-resistant cell line. *Jpn J Cancer Res*. 1994; 85: 659-664.



## Prognostic significance of miR-23b in combination with P-gp, MRP and LRP/MVP expression in non-small cell lung cancer

M. JANIKOVA<sup>1,2,‡</sup>, V. ZIZKOVA<sup>1,‡</sup>, J. SKARDA<sup>1,2,\*</sup>, G. KHARAISHVILI<sup>1,2</sup>, L. RADOVA<sup>3</sup>, Z. KOLAR<sup>1,2</sup>

<sup>1</sup>Department of Clinical and Molecular Pathology and Laboratory of Molecular Pathology; <sup>2</sup>Institute of Molecular and Translational Medicine, Faculty of Medicine and Dentistry, Palacky University Olomouc and University Hospital Olomouc; <sup>3</sup>CEITEC MU, Brno, Czech Republic

\*Correspondence: [jojso@email.cz](mailto:jojso@email.cz)

‡Contributed equally to this work.

Received May 14, 2015 / Accepted November 13, 2015

Recently, miR-23b has emerged as a promising new cancer biomarker but its role in lung cancer has not been established yet. Patients still do not respond well to available treatments, probably due to expression of multidrug resistance (MDR) proteins, such as P-gp, MRP and LRP/MVP. The aim of this study was to determine the role of miR-23b in non-small cell lung cancer (NSCLC) and its relationship to the patient outcome together with MDR transporter proteins. We immunohistochemically evaluated expression of P-gp, MRP and LRP/MVP and quantified the relative levels of miR-23b in 62 NSCLC patients' samples. The prognostic significance of miR-23b and MDR proteins was tested by Kaplan-Meier and Cox-regression analysis. Our results showed that miR-23b is mostly downregulated in NSCLC samples (57/62) and that its upregulation in tumors is connected with longer progression-free survival (PFS;  $P = 0.065$ ) and overall survival (OS;  $P = 0.048$ ). The Cox proportional hazard model revealed that the risk of death or relapse in NSCLC patients with miR-23b downregulation increases together with LRP/MVP expression and both risks decrease with miR-23b upregulation ( $HR_{PFS} = 4.342$ ,  $P_{PFS} = 0.022$ ;  $HR_{OS} = 4.408$ ,  $P_{OS} = 0.015$ ). Our findings indicate that miR-23b, especially in combination with LRP/MVP expression, might serve as a suitable prognostic biomarker for NSCLC patients.

*Key words: non-small cell lung cancer (NSCLC), MiR-23b, P-gp, MRP, LRP/MVP, prognosis*

Lung cancer is a major cause of cancer mortality worldwide and non-small cell lung cancer (NSCLC) represents approximately 80-85% of all cases [1]. The problem is that many NSCLC patients do not respond to the therapy due to the emergence of multidrug resistance (MDR). The most studied proteins underlying this phenomenon are P-glycoprotein (P-gp/ABCB1/MDR1), Multidrug resistance-related protein 1 (MRP/ABCC1/MRP1) and Lung-resistance related protein/Major vault protein (LRP/MVP).

P-gp and MRP belong to the family of ATP binding cassette (ABC) transporters. P-gp actively transports hydrophobic, positively charged and neutral molecules, while MRP pumps negatively charged molecules (mainly glutathione conjugates) from the cell to the extracellular space [2-5]. They are localized in the cytoplasmic membrane and they are normally expressed in organs, such as lung, intestine or the hemato-encephalic barrier which are in contact with environment and xenobiotics (toxins and drugs). The physiological function of P-gp and

MRP is to export toxic metabolites and xenobiotics entering the cells from the environment and to protect the cells from damage [6-10]. These ABC-transporters are aberrantly expressed in tumor cells. Activation of their expression in cancer cells is responsible for the export of diverse chemotherapeutics from the cells [5, 7, 8, 11, 12].

LRP/MVP is a ribonucleoprotein particle able to transport drugs such as platinum derivatives from the nucleus to the cytoplasm. After that, these compounds can be transported outside from the cells by ABC-transporters [13-15].

Expression of P-gp, MRP and LRP/MVP is often observed in NSCLC samples and it is believed that these molecules play a major role in the emergence of multidrug resistance [5, 7, 8, 16].

Recent studies have shown that miRNAs play an important role in development, differentiation, DNA damage response, proliferation, apoptosis and other processes. They are also involved in diverse regulatory pathways and they can affect

carcinogenesis, chemoresistance and radioresistance [17-21]. For these reasons, miRNAs whose expression may be increased or decreased in tumors could be called a class of oncogenes or tumor suppressors and they might be used as biomarkers for monitoring carcinogenesis [17, 22]. Radiotherapy, chemotherapy and biological therapy are frequently used for treating advanced forms of NSCLC but in recent years, RNA interference (RNAi) has also been introduced [1, 23]. RNAi is a highly conserved molecular mechanism of post-transcriptional gene silencing. Functional units of RNAi in humans are mainly small (~22 nucleotides) non-coding, endogenous, single-stranded RNAs, called microRNAs (miRNAs). Based on complementarity, miRNAs bind to the target mRNA which is either completely degraded or prevented from translation [23]. Currently, according to miRBase 21, 2588 mature human miRNAs are known to regulate many protein-coding genes [24].

Recently, miR-23b has emerged as a promising molecule that may be involved in tumor development and metastasis. It behaves as a tumor-suppressor or as an oncogenic miRNA in time- and tissue-dependent manner [25]. Owing to its dual role, miR-23b in tumors targets different molecules and regulates oncogenes supporting proliferation and metastasis formation, such as Src kinase, AKT, SNAIL and ZEB1, or tumor-suppressors having an apoptotic and anti-metastatic effect (for example Nischarin, VHL and PTEN) [26-31]. Downregulation of miR-23b was observed in prostate, bladder and thyroid cancer. Patients with higher levels of miR-23b had better prognosis than those with lower levels [26, 28, 32]. Upregulation of miR-23b was detected in renal, breast cancer and gliomas [29-31] and its upregulation was connected with poor patient's outcome.

Prognostic significance of miR-23b has not been studied in NSCLC yet. For this reason, we quantified level of miR-23b in NSCLC patient samples. Identification of its significance could be crucial for its usage as a biomarker. We compared its expression with the expression of known transporter proteins P-gp, MRP and LRP/MVP involved in the emergence of MDR. Finally, we created a regression model for assessing patient risk of death or relapse.

## Patients and methods

**Patients and their clinical data.** Formalin-fixed and paraffin-embedded (FFPE) tissue samples of lung cancer from 1996-2000 were obtained from the archives of the Department of Clinical and Molecular Pathology, Faculty of Medicine and Dentistry, Palacky University Olomouc and University Hospital Olomouc. The sample comprised 62 patients with lung cancer, 55 males and 7 females (age range 33 to 78 years), after approval by the ethics committee of University Hospital and Faculty of Medicine and Dentistry, Palacky University Olomouc. In this cohort, 30 tumors were classified as adenocarcinomas (ADCs), 26 as squamous cell carcinomas (SCCs) and 6 as large cell carcinomas (LCCs). 18 patients were in stage

I/II and 33 patients in stage III/IV. For the rest of the patients, the stage was unknown. 22 patients received chemotherapy. 17 of them underwent platinum based chemotherapy regime and the rest were treated with different chemotherapeutics, such as fluorouracil, doxorubicin or taxanes. The characteristics of patients are summarised in Table 1. The clinicopathological parameters, progression free survival (PFS) and overall survival (OS), were monitored over 175 months.

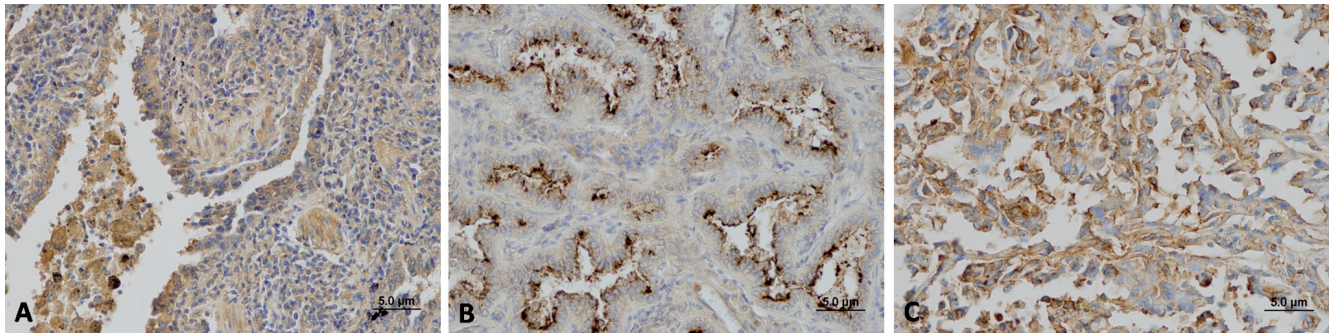
**Tissue microarray construction.** Tumor tissue microarrays (TMAs) were constructed using 62 formalin-fixed and paraffin-embedded primary lung cancer specimens. The tissue areas for sampling were selected by pathologists on the basis of visual alignment with corresponding H&E stained section and were taken from two to four tissue cores from a tumor block and were replaced into recipient paraffin block using a tissue microarrayer Galileo TMA CK3500 (BioRep, Milan, Italy).

**Immunohistochemical staining of P-gp, MRP and LRP/MVP.** The TMA sections were deparaffinized and an indirect immunohistochemistry technique was used. Then LRP/MVP antigen was unmasked in citrate buffer (pH 6) and MRP in Target Retrieval Solution, High pH (10x) (Dako, Glostrup, Denmark). The sections for detection of P-gp were not pre-treated because the ABCB1 / P Glycoprotein Mouse anti-Human Monoclonal (UIC2) Antibody (LS-C58240-100, LifeSpan Biosciences, Inc., Seattle, WA, USA) targets the extracellular epitope of P-gp protein [33]. This antibody was diluted in a ratio 1:200. Both Monoclonal Antibody to MRP1 (human) (MRPm5) (ALX-801-012-C125, Enzo Life Sciences, Inc., Farmingdale, NY, USA) and Monoclonal Antibody to MVP/LRP (human), mAb (MVP-37) (LMR5) (ALX-801-026-C125, Enzo Life Sciences, Inc., Farmingdale, NY, USA)

**Table 1. Characteristics of patients.**

Characteristics	Subgroups	Number	%
Gender	Male	55	88.7
	Female	7	11.3
Age in years	< 60	31	50.0
	≥ 60	31	50.0
Histological subtype	ADC	30	48.4
	SCC	26	41.9
	LCC	6	9.7
Stage at diagnosis	I	13	21.0
	II	5	8.1
	IIIA	20	32.3
	IIIB	5	8.1
	IV	8	12.9
	Not specified	11	17.7
Chemotherapy	Yes	22	35.5
	No	31	50.0
	Not specified	9	14.5
Total		62	100.0

ADC – adenocarcinoma, SCC – squamous cell carcinoma, LCC – large cell carcinoma.



**Figure 1.** Immunohistochemical expression of P-gp, MRP and LRP/MVP. Representative images of immunohistochemical staining of P-gp (A), MRP (B) and LRP/MVP (C) in NSCLC samples. Magnification 200x.

were used at a dilution 1:200. Visualisation was made by EnVision™+ Dual Link (Dako, Glostrup, Denmark) and nuclei were counterstained with haematoxylin. Stained sections were observed under an optical microscope and images were captured with a DP71 camera (Olympus, Tokyo, Japan). The expression of P-gp, MRP and LRP/MVP were semi quantitatively assessed by two pathologists by estimation of the percentage of positive cells as very low ( $\leq 0.1$ ), low ( $\leq 0.3$ ), moderate ( $\leq 0.6$ ) or high ( $> 0.6$ ).

**Isolation of total RNA.** Total RNA isolation from corresponding tissue cores obtained by tissue microarrayer Galileo TMA CK3500 [34] was performed using the RecoverAll™ Total Nucleic Acid Isolation Kit for FFPE (Applied Biosystems, Foster City, CA, USA) according to the manufacturer's instructions. The concentration of total RNA was measured using NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, Delaware, USA) and then RNA was stored at  $-80^{\circ}\text{C}$  until use.

**Reverse transcription.** TaqMan® MicroRNA Reverse Transcription Kit (Applied Biosystems, Foster City, CA, USA) was used according to manufacturer's protocol. For reverse transcription, 10 ng of total RNA was used and total volume of reaction was 15  $\mu\text{l}$ . Pooled gene-specific primers for miR-23b and RNA, U6 small nuclear 2 (RNU6B) (Applied Biosystems, Foster City, CA, USA) were added to the reverse transcription polymerase chain reaction (RT-PCR). RT-PCR product was then pre-amplified according to protocol using TaqMan® PreAmp Master Mix (2x) (Applied Biosystems, Foster City, CA, USA) with pooled TaqMan® MicroRNA Assays (Applied Biosystems, Foster City, CA, USA) for miR-23b and RNU6B.

**Real-time PCR for miR-23b quantification.** The pre-amplified PCR product was diluted in DEPC water (1:20). Each sample was analysed in triplicate and the volume per each reaction was 10  $\mu\text{l}$ . The reaction consisted of TaqMan® Universal PCR Master Mix (2x) (Applied Biosystems, Foster City, CA, USA), DEPC water, corresponding TaqMan® MicroRNA Assay and diluted pre-amplified PCR product. Real-time PCR was performed using LightCycler® 480 (Roche, Branford,

CT, USA) with appropriate software according to protocol and RNU6B was used as an endogenous control. The non-tumor control sample was prepared by pooling commercially available Human Lung Total RNAs (AM7968, Lot. 0904002 and 1203010, Applied Biosystems, Foster City, CA, USA) and Total RNAs, Lung, Human (540019, Lot. 0006051745 and 0006079356, Agilent Technologies, Santa Clara, CA, USA). The relative expression level of miR-23b was calculated using the following equation: relative gene expression =  $2^{-(\Delta\text{Ct sample} - \Delta\text{Ct control})}$ .

**Statistical analysis.** Statistical analysis was performed using software SPSS, version 20.0. Pearson's Chi-square was used to assess the relationships between categorical variables. Mann-Whitney test was used to evaluate difference between groups. Survival analysis was done by Kaplan-Meier method with the log-rank test applied for comparison of curves. Overall survival (OS) was defined as time from the surgery until patient's death, progression free survival (PFS) as time from surgery until disease relapse or patient's death. Cox proportional hazard model was created for assessing the risk of patient's death and/or relapse of disease.  $P < 0.05$  was considered as statistically significant.

## Results

TMA blocks from 62 patients were immunohistochemically analysed for P-gp, MRP and LRP/MVP expression. The tissues from TMAs were also used for the quantification of miR-23b expression.

**Expression of P-gp, MRP and LRP/MVP in NSCLC.** The expression of P-gp, MRP and LRP/MVP was assessed in 62 patients (Figures 1, 2). The P-gp was detected in 40 (64.5 %) patients, MRP in 50 (80.6 %) patients and expression of LRP/MVP was found in 35 (56.4 %) patients. Expression of these proteins in different histological subtypes is summarised in Table 2. There was no significant difference in P-gp, MRP or LRP/MVP expression between ADC, SCC and LCC specimens ( $P_{\text{P-gp}} = 0.313$ ,  $P_{\text{MRP}} = 0.432$ ,  $P_{\text{LRP/MVP}} = 0.108$ ) (Figure 3). For the next analysis, very low expression of proteins was considered

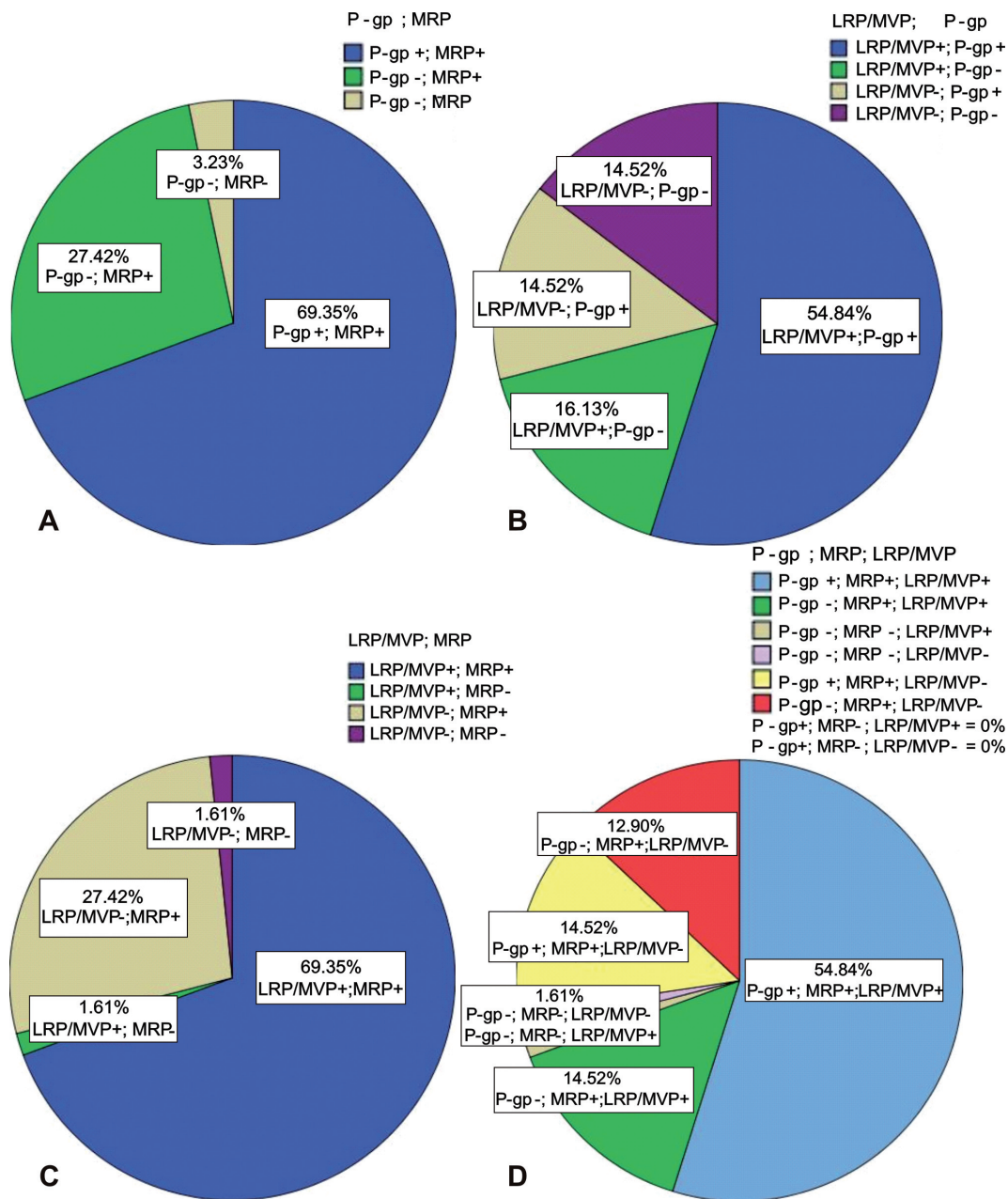


Figure 2. Combined expression of P-gp, MRP and LRP/MVP. Pie charts show percentage of combined expression of P-gp and MRP (A), LRP/MVP and P-gp (B), LRP/MVP and MRP (C) and all three analysed protein markers P-gp, MRP and LRP/MVP (D).

as a negative expression while low, moderate and high were considered as a positive expression.

**Expression of miR-23b in NSCLC.** The relative quantification of mature miR-23b in 62 lung cancer samples was performed. Higher expression of miR-23b compared to non-tumor control RNA was detected in 5 (8.1%) samples and lower expression in 57 (91.9 %) samples. There was no significant difference in miR-23b expression between specimens from different histological subtypes (ADC, SCC, LCC;

$P = 0.711$ ; Table 3). For the next analysis, higher expression of miR-23b ( $2^{-\Delta\Delta Ct} > 1$ ) was marked as “up” (means upregulation) and lower expression ( $2^{-\Delta\Delta Ct} < 1$ ) was marked as “down” (means downregulation).

**Relationship of miR-23b, P-gp, MRP and LRP/MVP expression with clinicopathological parameters.** Progression free survival and overall survival were used as the clinicopathological parameters for evaluating the prognostic role of these variables in NSCLC patients. Patient outcome was monitored

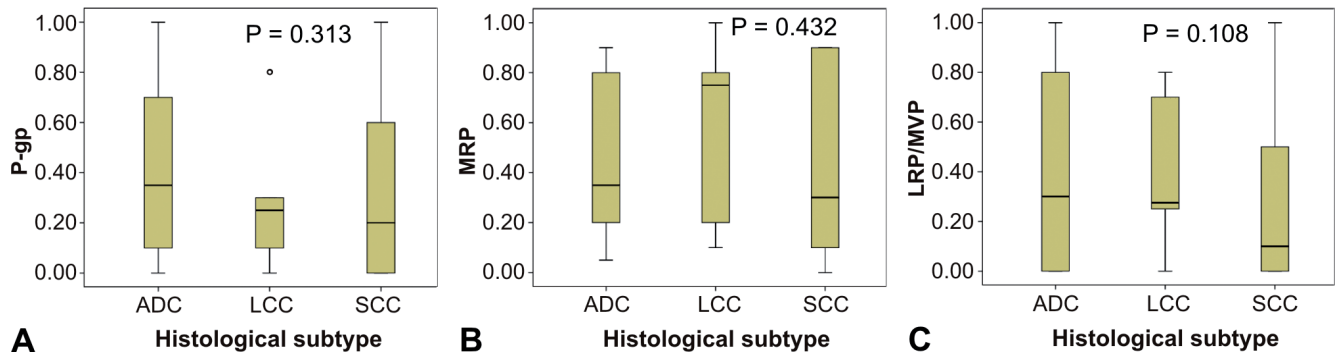


Figure 3. Analysis of P-gp, MRP and LRP/MVP expression in different histological subtypes. Box plots show the proportion of P-gp (A), MRP (B) and LRP/MVP (C) expression in different histological subtypes of NSCLC.

ADC – adenocarcinoma, SCC – squamous cell carcinoma, LCC – large cell carcinoma.

Table 2. Summary of expression of P-gp, MRP and LRP/MVP in different histological subtypes.

Protein	Histological subtype	Intensity				Statistics	
		Very low ( $\leq 0.1$ )	Low ( $\leq 0.3$ )	Moderate ( $\leq 0.6$ )	High ( $> 0.6$ )	Pearson $\chi^2$	P
P-gp	ADC	8	7	3	12	2.324	0.313
	SCC	12	6	2	6		
	LCC	2	3	0	1		
	Total [No. (%)]	22 (35.5)	16 (25.8)	5 (8.1)	19 (30.6)		
MRP	ADC	4	11	4	11	1.679	0.432
	SCC	7	7	1	11		
	LCC	1	1	0	4		
	Total [No. (%)]	12 (19.4)	19 (30.6)	5 (8.1)	26 (41.9)		
LRP/MVP	ADC	11	4	3	12	4.457	0.108
	SCC	15	2	5	4		
	LCC	1	3	0	2		
	Total [No. (%)]	27 (43.5)	9 (14.5)	8 (12.9)	18 (29.0)		

Comparison of negative ( $\leq 0.1$ ) and positive ( $\geq 0.1$ ) expression of MDR proteins in different histological subtypes. MDR – multidrug resistance, P-gp – P-glycoprotein, MRP – Multidrug resistance-related protein 1, LRP/MVP – Lung-resistance related protein/Major vault protein, ADC – adenocarcinoma, SCC – squamous cell carcinoma, LCC – large cell carcinoma.

over 175 months. Individual data are showed in Table 4. Expression of P-gp, MRP and LRP/MVP in patients with regard to miR-23b expression levels is summarised in Table 5. The Kaplan-Meier analysis showed a trend to association between PFS ( $P = 0.065$ ) and statistically significant relationship be-

Table 3. Expression of miR-23b in different histological subtypes.

Histological subtype	miR-23b		Total
	down	up	
ADC	27	3	30
SCC	24	2	26
LCC	6	0	6
Total	57	5	62

Summary of miR-23b relative expression in different histological subtypes based on  $2^{-\Delta\Delta Ct}$  values.

Down – downregulation ( $2^{-\Delta\Delta Ct} < 1$ ), up – upregulation ( $2^{-\Delta\Delta Ct} > 1$ ).

tween OS ( $P = 0.048$ ) and miR-23b expression (Figure 4A, B). Survival analysis showed no significant association between expression of P-gp, MRP and LRP/MVP and patient's outcome (P-gp:  $P_{PFS} = 0.993$ ,  $P_{OS} = 0.982$ ; MRP:  $P_{PFS} = 0.911$ ,  $P_{OS} = 0.945$ ; LRP/MVP:  $P_{PFS} = 0.908$ ,  $P_{OS} = 0.966$ ; Figure 4C-H). Based on these results, overexpression of miR-23b seems to be good prognostic factor in the case of NSCLC patients.

**Cox regression hazard model.** The relationship between miR-23b and P-gp, MRP and LRP/MVP was first analysed by Chi-Square and Mann-Whitney tests. Neither of them showed any significant association between the variables (Table 6).

The Cox proportional hazard model was created to assess the relationship between the risk of death or relapse and variables of interest such as miR-23b and LRP/MVP. The association between the risk and the above variables were statistically significant (PFS:  $P = 0.022$ , HR = 4.342, CI (95%) = 1.231-15.318, Wald = 5.211; OS:  $P = 0.015$ , HR = 4.408, CI (95%) = 1.356-16.350, Wald = 5.949) and model fit parameters

**Table 4. Individual data on tumor classification and expression levels of analysed markers.**

Patient No.	Gender	Histological subtype	Classification			P-gp	MRP	LRP/MVP	miR-23b
			T	N	M				
1	M	ADC				0.0	0.9	0.0	0.00
3	M	SCC	2	2	0	0.0	0.0	0.0	0.00
4	M	LCC	2	0	0	0.1	0.8	0.1	0.17
5	M	ADC	1	0	0	0.3	0.9	0.2	0.18
6	M	ADC	2	0	0	0.0	0.9	0.7	0.09
7	M	ADC	1	0	0	0.3	0.8	0.1	0.09
8	F	SCC	3	2	0	0.6	0.3	0.9	0.56
9	M	SCC	3	1	0	0.0	0.1	0.5	1.73
10	M	SCC				0.3	0.9	0.0	0.00
11	M	SCC	2	0	0	0.0	0.9	1.0	0.00
12	M	ADC	2	0	0	0.0	0.6	0.0	0.04
13	M	SCC	2	2	0	0.6	0.1	0.0	0.07
14	M	SCC	1	0	0	0.0	0.9	0.0	0.01
15	M	ADC	4	2	0	0.3	0.2	0.7	0.15
16	M	ADC	2	2	0	1.0	0.3	1.0	0.06
17	M	ADC	3	2	0	0.0	0.2	1.0	0.36
19	M	SCC	3	1	0	0.0	0.3	0.0	0.15
20	M	SCC				0.0	0.8	0.0	0.05
21	M	SCC	2	0	0	0.0	0.8	0.4	0.04
22	M	SCC	2	2	0	0.0	0.0	0.1	0.03
23	M	SCC				0.0	0.9	0.0	0.11
24	M	ADC	1	0	0	0.0	0.8	0.0	3.38
25	F	ADC	4	2	1	0.8	0.1	0.2	0.09
26	M	ADC	2	0	1	0.0	0.3	1.0	0.23
27	M	SCC	4	0	0	0.0	0.1	0.5	0.14
28	M	ADC	3	2	1	0.8	0.2	0.5	0.13
29	M	ADC				0.2	0.8	0.0	0.15
30	F	ADC	2	2	0	0.7	0.2	0.0	0.15
31	M	ADC				1.0	0.8	0.9	0.17
32	M	ADC	1	0	0	0.1	0.8	0.2	0.04
33	M	LCC	2	1	0	0.2	1.0	0.3	0.10
34	M	SCC	2	2	0	0.0	0.1	0.0	0.18
35	M	SCC	2	1	0	0.3	0.9	0.0	0.26
36	M	SCC	3	1	0	1.0	0.2	0.8	0.24
37	M	ADC	2	2	0	0.3	0.2	0.9	0.03
38	M	SCC	2	2	0	0.7	0.3	0.0	0.15
39	M	LCC	1	2	0	0.3	0.1	0.8	0.49
40	M	LCC	1	1	0	0.0	0.7	0.7	0.25
41	M	SCC	2	2	0	0.0	0.2	0.1	0.05
42	F	ADC	1	0	0	0.4	0.9	0.1	0.27
43	M	ADC	4	1	0	0.9	0.4	0.8	21.95
44	M	ADC	2	2	0	0.9	0.1	0.0	0.97
45	F	ADC				0.1	0.9	0.4	0.11
46	F	SCC	3	1	0	0.8	0.3	1.0	0.50
47	M	ADC	2	1	0	0.2	0.8	0.1	0.23
48	M	SCC	1	0	0	0.2	0.9	0.5	0.31
49	F	ADC	4	2	0	0.7	0.2	0.0	0.06
50	M	ADC	1	0	1	0.7	0.1	0.2	0.39
51	M	SCC	1	2	0	0.8	0.3	0.6	0.28
52	M	SCC				0.2	0.8	0.1	0.20
53	M	ADC	1	0	0	0.5	0.6	0.4	0.22
54	M	ADC	3	1	0	0.7	0.3	0.8	1.26
55	M	SCC	2	0	1	0.8	0.1	1.0	5.15
56	M	ADC	x	3	1	0.6	0.1	0.7	0.18
57	M	ADC	1	0	1	0.7	0.2	0.0	0.57
59	M	ADC	3	0	0	0.3	0.6	0.8	0.25
60	M	LCC	3	3	0	0.8	0.2	0.3	0.16
61	M	ADC	3	2	0	0.8	0.2	0.7	0.68
62	M	LCC				0.3	0.8	0.3	0.16
63	M	SCC	3	2	2	0.9	0.4	1.0	0.55
65	M	SCC				0.3	0.8	0.3	0.27
66	M	SCC				0.2	0.9	0.2	0.24

Individual data of 62 patients on gender, histological subtype, TNM classification and expression levels of analysed markers. Expression level of protein markers P-gp, MRP and LRP/MVP is showed as percentage of positive cells (0.0 min – 1.0 max). Relative expression of miR-23b is showed as  $2^{-\Delta\Delta Ct}$  value. Value  $2^{-\Delta\Delta Ct} > 1$  means upregulation and  $2^{-\Delta\Delta Ct} < 1$  means downregulation of miR-23b. M – man, F – female.

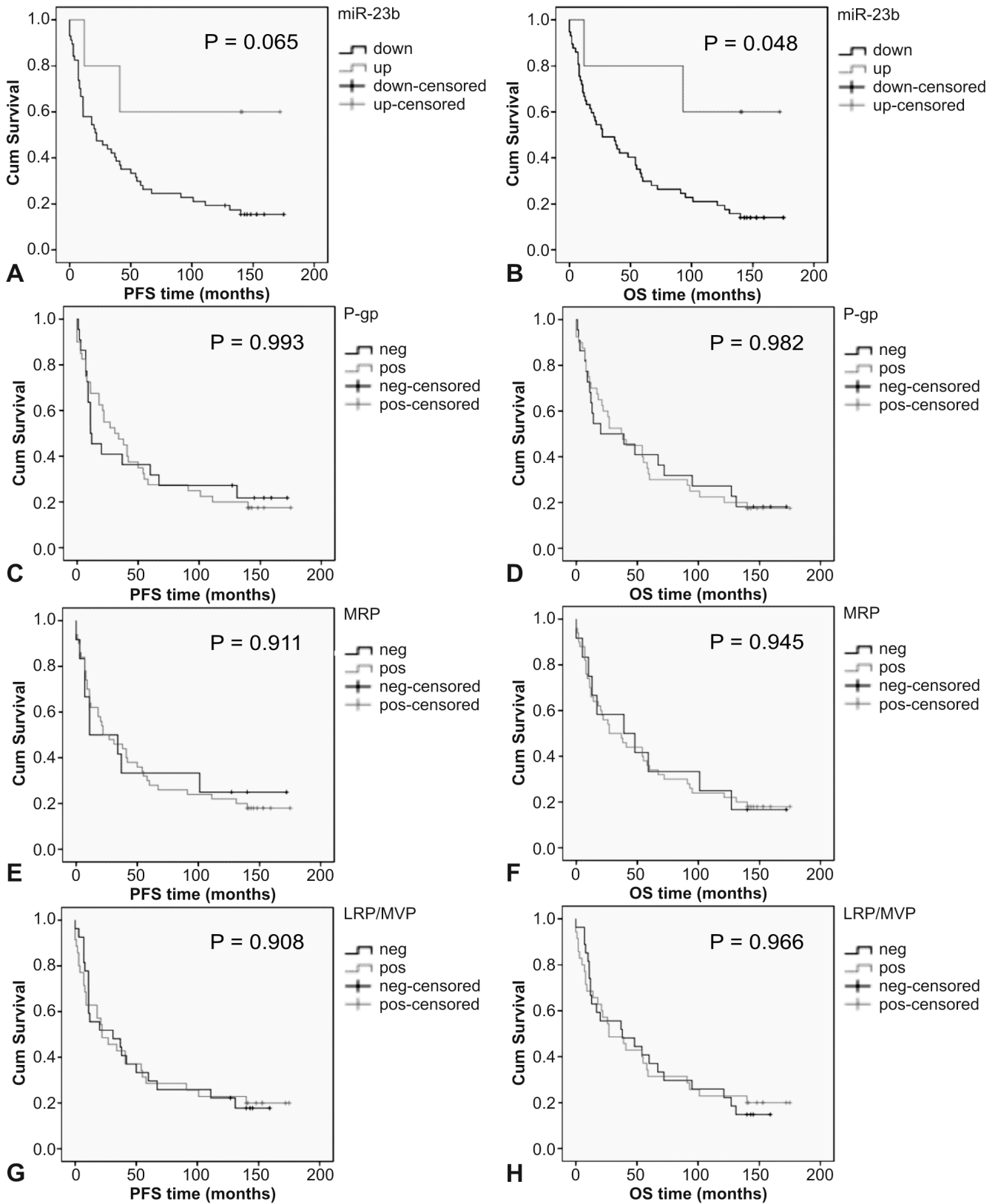


Figure 4. Survival analysis of miR-23b, P-gp, MRP and LRP/MVP in NSCLC patients. Kaplan-Meier curves represent PFS and OS of NSCLC patients and depend on the expression of miR-23b (A, B), P-gp (C, D), MRP (E, F) and LRP/MVP (G, H). PFS – progression-free survival, OS – overall survival.

meet the required criteria (PFS:  $P = 0.020$ ,  $\chi^2 = 5.400$ ; OS:  $P = 0.013$ ,  $\chi^2 = 6.185$ ). As a result, there was increased risk of death or relapse in patients with miR-23b downregulation and LRP/MVP expression from 0.1 to 1; and there was decreased risk of death or relapse in patients who had upregulated miR-23b (Figure 5). Association between risks and miR-23b with P-gp and MRP expression was not significant for modelling. Importantly, univariate Cox regression analysis of miR-23b indicated that there was a trend of association between miR-23b and risk of death or relapse in NSCLC patients (PFS:  $P = 0.087$ , HR = 3.451, CI (95%) = 0.836-14.246, Wald = 2.932; OS:  $P = 0.067$ , HR = 3.765, CI (95%) = 0.912-15.552, Wald = 3.357). Therefore, the combined model of miR-23b and LRP/MVP defines well the risk of death or relapse in NSCLC patients.

We also found that patients expressing P-gp from 0.1 to 1 with miR-23b downregulation had higher risk of cancer-related death compared to patients with miR-23b upregulation (RR = 2.595, CI = 0.521-12.920,  $\chi^2 = 5.430$ ,  $P = 0.02$ ; Figure 6). However, as mentioned above, combined expression of miR-23b with P-gp did not meet the criteria for modelling and no Cox regression model for these variables was prepared.

Our results showed that miR-23b, especially in combination with LRP/MVP, may serve as a prognostic biomarker for NSCLC patients. A larger cohort of patient samples is needed to be analyzed to confirm the results.

**Discussion**

The initial goal of this preliminary study was to determine the significance of miR-23b in NSCLC. Previous studies describe its dual role as an oncogenic or tumor-suppressor miRNA depending on cancer type [25]. The relative quantification of miR-23b in this study showed that miR-23b is mostly downregulated in NSCLC patient samples (57 out of 62 samples). Its downregulation has also been described in prostate, bladder and thyroid cancer. Higher levels of miR-23b are a good prognostic factor according to some authors [26, 28, 32]. The survival analysis data are in agreement with these studies showing that upregulation of miR-23b in tumors with its predominant downregulation is connected with better patient outcome ( $P_{PFS} = 0.065$ ;  $P_{OS} = 0.048$ ). There are also studies that present miR-23b as an oncogenic miRNA, where its upregulation in different tumors is associated with worse patient outcome [29-31]. The split role of miR-23b in different cancer types can be caused by targeting different molecules.

In tumors with miR-23b predominant downregulation it is known as a tumor-suppressor and targets mainly proteins responsible for proliferation and metastasis formation, such as AKT, Src kinase, ZEB1 or SNAIL [26-28]. In tumors with miR-23b upregulation, it functions as an oncogene and targets mainly proteins involved in apoptosis and metastasis suppression, such as Nischarin, VHL or PTEN [29-31].

Park *et al.* identified participation of miR-23b in chemoresistance [35]. Therefore, we decided to detect immunohistochemical expression of well-known proteins associated with MDR, namely P-gp, MRP and LRP/MVP. Expression of P-gp was detected in 40 (64.5 %), MRP in 50 (80.6 %) and LRP/MVP in 35 (56.4 %) patients. Our survival analysis showed no relationship between expression of these proteins and PFS and OS (P-gp:  $P_{PFS} = 0.993$ ,  $P_{OS} = 0.982$ ; MRP:  $P_{PFS} = 0.911$ ,  $P_{OS} = 0.945$ ; LRP/MVP:  $P_{PFS} = 0.908$ ,  $P_{OS} = 0.966$ ). Similar observations were also published by other authors [36-39]. On the other hand, there are also studies supporting the negative prognostic role of these MDR proteins in cancer patients [40-44]. Thus, we decided to establish Cox proportional hazard model for testing the possible prognostic significance of combined expression of these proteins and miR-23b and for assessing risk of death and relapse in our patients.

First, we found that patients expressing P-gp with miR-23b downregulation had higher risk of cancer-related death compared to patients with miR-23b upregulation (RR = 2.595,  $P = 0.02$ ). More importantly, our Cox proportional hazard model showed that the risk of death or relapse in patients with miR-23b downregulation increases together with expression of LRP/MVP and decreases in patients with miR-23b upregulation (PFS:  $P = 0.022$ , HR = 4.342; OS:  $P = 0.015$ , HR = 4.408). On the other hand, univariate analysis of miR-23b alone did not show any statistically significant association with patient's risk (PFS:  $P = 0.087$ , HR = 3.451; OS:  $P = 0.067$ , HR = 3.765). We are probably the first group interested in the prognostic role of miR-23b in combination with other known biomarkers in NSCLC. But available reports on the possible role of miR-23b in different cancer types support our observations. These authors found that miR-23b upregulation is associated with better patient's prognosis in tumors where it occurs as a tumor-suppressor [26, 28]. These preliminary results need to be validated in a larger cohort of cases.

**Table 5. Summary of expression of P-gp, MRP and LRP/MVP in patients with regard to miR-23b up- and downregulation.**

		P-gp		MRP		LRP/MVP	
		negat	posit	negat	posit	negat	posit
miR-23b	down	20	37	10	47	26	31
	up	2	3	2	3	1	4
<b>Total</b>		22	40	12	50	27	35

**Table 6. Correlation analysis of expression of miR-23b and P-gp, MRP and LRP/MVP.**

Protein	Chi-Square test		Mann-Whitney test		
	$\chi^2$ value	P	P	Z	
miR-23b	P-gp	0.048	0.826	0.599	-0.526
	MRP	1.485	0.223	0.320	0.995
	LRP/MVP	1.227	0.268	0.115	-1.576

The relationship between miR-23b and P-gp, MRP and LRP/MVP analysed by Chi-Square and Mann-Whitney tests.



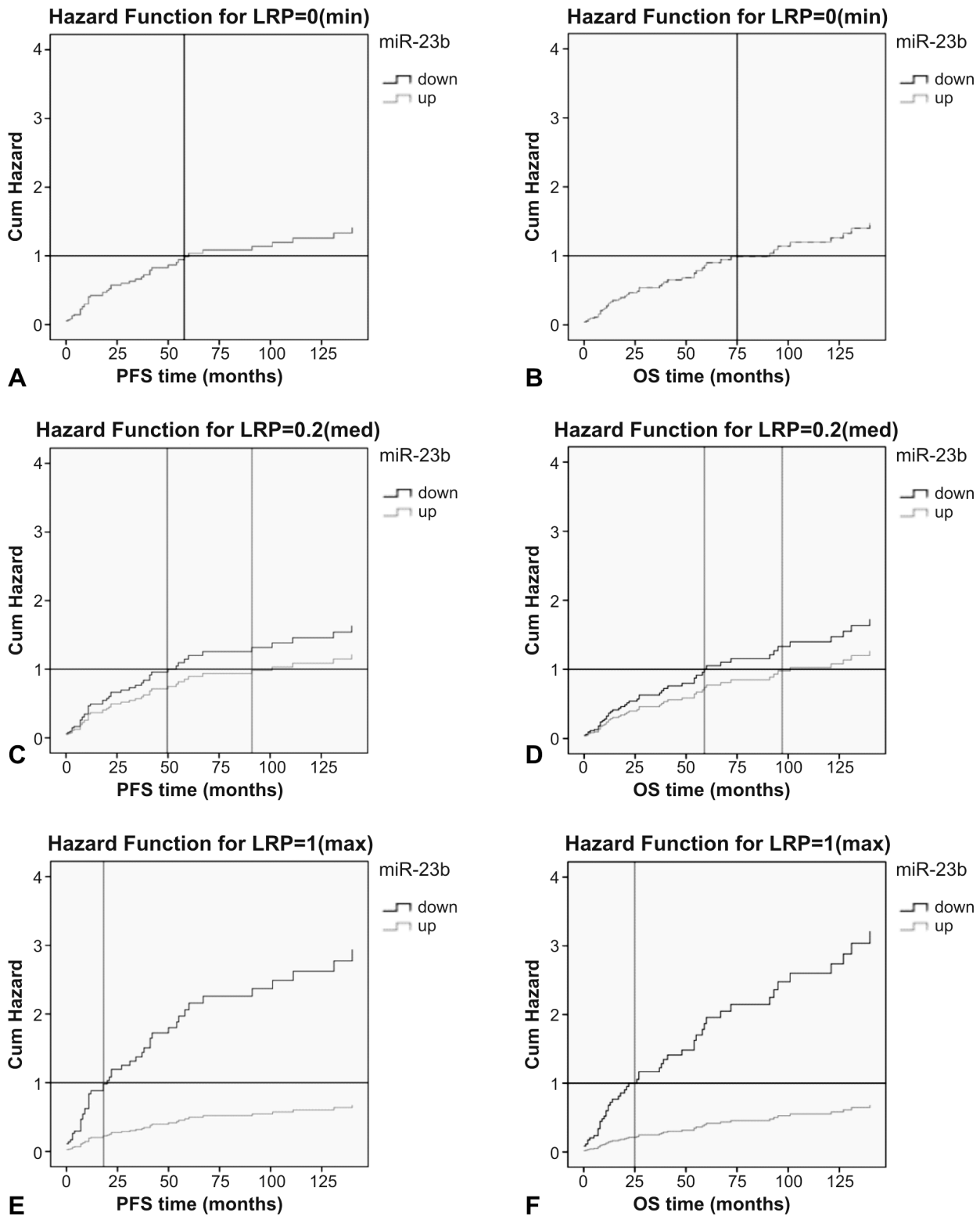


Figure 5. Cox regression hazard model of miR-23b and LRP/MVP expression in NSCLC patients. Hazards of death and relapse between miR-23b down- and upregulation groups are the same when LRP/MVP expression is 0 (A, B). Hazard is higher in the group with miR-23b downregulation in comparison to the group with miR-23b upregulation when LRP/MVP expression is 0.2 (median) (C, D). The difference between the hazards of miR-23b down- and upregulation groups is the highest when LRP/MVP expression is 1 (max) (E, F).

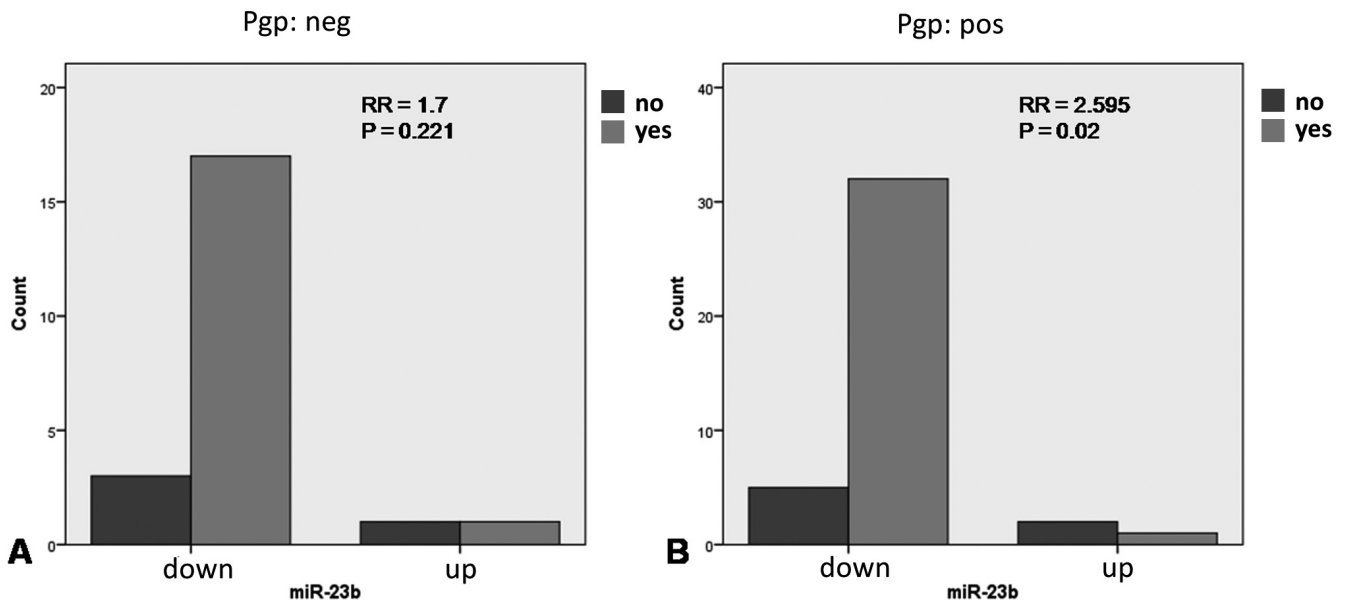


Figure 6. Odds ratio of P-gp and miR-23b expression in NSCLC patients. Patients with miR-23b downregulation had higher risk of cancer-related death compared to patients with miR-23b upregulation in the case of P-gp expression from 0.1 to 1.

**Conclusion**

In this study, we determined the possible role of miR-23b which is very often deregulated in different cancers and its role in NSCLC has not been fully established yet. The results suggest that miR-23b may function as a tumor-suppressor in NSCLC as it was often downregulated in advanced tumors and patients had better outcome when it was upregulated. Our results also showed that detection of miR-23b levels together with P-gp and LRP/MVP expression may be suitable for assessing the prognosis and risk in NSCLC patients. According to the regression model, the risk of death or relapse in NSCLC patients with downregulated miR-23b increases together with LRP/MVP expression and decreases in patients with upregulated miR-23b.

**Acknowledgements:** We would like to thank Jiri Klein and Vitezslav Kolek for kind help with collection and characterization of patients. We would like to thank also Eva Sedlakova for the performance of immunohistochemical staining. This work was supported by grants: IGA MZCR NT13569, LF\_2014\_003 and NPU I LO1304.

**References**

[1] D’ADDARIO G, FRUH M, RECK M, BAUMANN P, KLEPETKO W et al. ESMO Guidelines Working Group: Metastatic non-small-cell lung cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol* 2010; 21(Suppl 5): v116–v119. <http://dx.doi.org/10.1093/annonc/mdq189>

[2] KARTNER N, RIORDAN JR, LING V Cell surface P-glycoprotein associated with multidrug resistance in mammalian cell

lines. *Science* 1983; 221: 1285–1288. <http://dx.doi.org/10.1126/science.6137059>

[3] COLE SP, BHARDWAJ G, GERLACH JH, MACKIE JE, GRANT CE et al. Overexpression of a transporter gene in a multidrug-resistant human lung cancer cell line. *Science* 1992; 258: 1650–1654. <http://dx.doi.org/10.1126/science.1360704>

[4] COLE SP, DEELEY RG Transport of glutathione and glutathione conjugates by MRP1. *Trends Pharmacol Sci* 2006; 27: 438–446. <http://dx.doi.org/10.1016/j.tips.2006.06.008>

[5] NOSKOVA V, HAJDUCH M, MIHAL V, CWIERTKA K [Mechanisms of multidrug resistance and their clinical implications, I. typical MDR]. *Klin Onkol* 2000; 13: 4–9. [Article in Czech]

[6] DEAN M, RZHETSKY A, ALLIKMETS R The human ATP-binding cassette (ABC) transporter superfamily. *Genome Res* 2001; 11: 1156–1166. <http://dx.doi.org/10.1101/gr.GR-1649R>

[7] FOJO AT, UEDA K, SLAMON DJ, POPLACK DG, GOTTESMAN MM et al. Expression of a multidrug-resistance gene in human tumors and tissues. *Proc Natl Acad Sci U S A* 1987; 84: 265–269. <http://dx.doi.org/10.1073/pnas.84.1.265>

[8] FLENS MJ, ZAMAN GJ, VAN DER VALK P, IZQUIERDO MA, SCHROEIJERS AB et al. Tissue distribution of the multidrug resistance protein. *Am J Pathol* 1996; 148: 1237–1247.

[9] DEMEULE M, REGINA A, JODOIN J, LAPLANTE A, DAGENAIS C et al. Drug transport to the brain: key roles for the efflux pump P-glycoprotein in the blood-brain barrier. *Vascul Pharmacol* 2002; 38: 339–348. [http://dx.doi.org/10.1016/S1537-1891\(02\)00201-X](http://dx.doi.org/10.1016/S1537-1891(02)00201-X)

[10] NIES AT, JEDLITSCHKY G, KONIG J, HEROLD-MENDE C, STEINER HH et al. Expression and immunolocalization of the multidrug resistance proteins, MRP1-MRP6 (ABCC1-

- ABCC6), in human brain. *Neuroscience* 2004; 129: 349–360. <http://dx.doi.org/10.1016/j.neuroscience.2004.07.051>
- [11] NOOTER K, BOSMAN FT, BURGER H, VAN WINGERDEN KE, FLENS MJ et al. Expression of the multidrug resistance-associated protein (MRP) gene in primary non-small-cell lung cancer. *Ann Oncol* 1996; 7: 75–81. <http://dx.doi.org/10.1093/oxfordjournals.annonc.a010484>
- [12] Burger H, Nooter K, ZAMAN GJ, SONNEVELD P, VAN WINGERDEN KE et al. Expression of the multidrug resistance-associated protein (MRP) in acute and chronic leukemias. *Leukemia* 1994; 8: 990–997.
- [13] SCHEFFER GL, WIJNGAARD PL, FLENS MJ, IZQUIERDO MA, SLOVAK ML et al. The drug resistance-related protein LRP is the human major vault protein. *Nat Med* 1995; 1: 578–582. <http://dx.doi.org/10.1038/nm0695-578>
- [14] KICKHOEFER VA, RAJAVEL KS, SCHEFFER GL, DALTON WS, SCHEPER RJ et al. Vaults are up-regulated in multidrug-resistant cancer cell lines. *J Biol Chem* 1998; 273: 8971–8974. <http://dx.doi.org/10.1074/jbc.273.15.8971>
- [15] KEDERSHA NL, ROME LH Isolation and characterization of a novel ribonucleoprotein particle: large structures contain a single species of small RNA. *J Cell Biol* 1986; 103: 699–709. <http://dx.doi.org/10.1083/jcb.103.3.699>
- [16] IZQUIERDO MA, SCHEFFER GL, FLENS MJ, GIACCONE G, BROXTERMAN HJ et al. Broad distribution of the multidrug resistance-related vault lung resistance protein in normal human tissues and tumors. *Am J Pathol* 1996; 148: 877–887.
- [17] VENTURA A, KUMAR MS, JACKS T Roles of microRNAs in cancer and development. In: Appasani K, editor. *MicroRNAs: From Basic Science to Disease Biology*. New York: Cambridge University Press, 2008: 322–337.
- [18] O'CONNELL RM, BALTIMORE D MicroRNAs and hematopoietic cell development. *Curr Top Dev Biol* 2012; 99: 145–174. <http://dx.doi.org/10.1016/B978-0-12-387038-4.00006-9>
- [19] MANSFIELD JH, MCGLINN E Evolution, expression, and developmental function of Hox-embedded miRNAs. *Curr Top Dev Biol* 2012; 99: 31–57. <http://dx.doi.org/10.1016/B978-0-12-387038-4.00002-1>
- [20] WAN G, MATHUR R, HU X, ZHANG X, LU X miRNA response to DNA damage. *Trends Biochem Sci* 2011; 36: 478–484. <http://dx.doi.org/10.1016/j.tibs.2011.06.002>
- [21] UENO K, HIRATA H, HINODA Y, DAHIYA R Frizzled homolog proteins, microRNAs and Wnt signaling in cancer. *Int J Cancer* 2013; 132: 1731–1740. <http://dx.doi.org/10.1002/ijc.27746>
- [22] WANG Q, WANG S, WANG H, LI P, MA Z MicroRNAs: novel biomarkers for lung cancer diagnosis, prediction and treatment. *Exp Biol Med (Maywood)* 2012; 237: 227–235. <http://dx.doi.org/10.1258/ebm.2011.011192>
- [23] PARANJAPE T, CHOI TJ, WEIDHAAS JB MicroRNA as Potential Diagnostics and Therapeutics. In: Slack FJ, editor. *MicroRNAs in Development and Cancer*. London: Imperial College Press, 2011: 213–236.
- [24] GRIFFITHS-JONES S The microRNA Registry. *Nucleic Acids Res* 2004; 32(Database issue): D109–111.
- [25] DONADELLI M, DANDO I, FIORINI C, PALMIERI M Regulation of miR-23b expression and its dual role on ROS production and tumor development. *Cancer Lett* 2014; 349: 107–113. <http://dx.doi.org/10.1016/j.canlet.2014.04.012>
- [26] MAJID S, DAR AA, SAINI S, ARORA S, SHAHRYARI V et al. miR-23b represses proto-oncogene Src kinase and functions as methylation-silenced tumor suppressor with diagnostic and prognostic significance in prostate cancer. *Cancer Res* 2012; 72: 6435–6446. <http://dx.doi.org/10.1158/0008-5472.CAN-12-2181>
- [27] CASTILLA MÁ, MORENO-BUENO G, ROMERO-PEREZ L, VAN DE VIJVER K, BISCUOLA M et al. Micro-RNA signature of the epithelial-mesenchymal transition in endometrial carcinosarcoma. *J Pathol* 2011; 223: 72–80. <http://dx.doi.org/10.1002/path.2802>
- [28] MAJID S, DAR AA, SAINI S, DENG G, CHANG I et al. MicroRNA-23b functions as a tumor suppressor by regulating Zeb1 in bladder cancer. *PLoS One* 2013; 8: e67686. <http://dx.doi.org/10.1371/journal.pone.0067686>
- [29] JIN L, WESSELY O, MARCUSSEON EG, IVAN C, CALIN GA et al. Prooncogenic factors miR-23b and miR-27b are regulated by Her2/Neu, EGF, and TNF- $\alpha$  in breast cancer. *Cancer Res* 2013; 73: 2884–2896. <http://dx.doi.org/10.1158/0008-5472.CAN-12-2162>
- [30] CHEN L, HAN L, ZHANG K, SHI Z, ZHANG J et al. VHL regulates the effects of miR-23b on glioma survival and invasion via suppression of HIF-1 $\alpha$ /VEGF and  $\beta$ -catenin/Tcf-4 signaling. *Neuro Oncol* 2012; 14: 1026–1036. <http://dx.doi.org/10.1093/neuonc/nos122>
- [31] ZAMAN MS, THAMMINANA S, SHAHRYARI V, CHIYOMARU T, DENG G et al. Inhibition of PTEN gene expression by oncogenic miR-23b-3p in renal cancer. *PLoS One* 2012; 7: e50203. <http://dx.doi.org/10.1371/journal.pone.0050203>
- [32] DETTMER MS, PERREN A, MOCH H, KOMMINOTH P, NIKIFOROV YE et al. MicroRNA profile of poorly differentiated thyroid carcinomas: new diagnostic and prognostic insights. *J Mol Endocrinol* 2014; 52: 181–189. <http://dx.doi.org/10.1530/JME-13-0266>
- [33] KENNEDY BG, MANGINI NJ P-glycoprotein expression in human retinal pigment epithelium. *Mol Vis* 2002; 8: 422–430.
- [34] CARDANO M, DIAFERIA GR, FALAVIGNA M, SPINELLI CC, SESSA F et al. Cell and tissue microarray technologies for protein and nucleic acid expression profiling. *J Histochem Cytochem* 2013; 61: 116–124. <http://dx.doi.org/10.1369/0022155412470455>
- [35] PARK YT, JEONG JY, LEE MJ, KIM KI, KIM TH et al. MicroRNAs overexpressed in ovarian ALDH1-positive cells are associated with chemoresistance. *J Ovarian Res* 2013; 6: 18. <http://dx.doi.org/10.1186/1757-2215-6-18>
- [36] DINGEMANS AM, VAN ARK-OTTE J, VAN DER VALK P, APOLINARIO RM, SCHEPER RJ et al. Expression of the human major vault protein LRP in human lung cancer samples and normal lung tissues. *Ann Oncol* 1996; 7: 625–630. <http://dx.doi.org/10.1093/oxfordjournals.annonc.a010681>
- [37] ARTS HJ, KATSAROS D, DE VRIES EG, MASSOBRIO M, GENTA F et al. Drug resistance-associated markers P-glycoprotein, multidrug resistance-associated protein 1, multidrug

- resistance-associated protein 2, and lung resistance protein as prognostic factors in ovarian carcinoma. *Clin Cancer Res* 1999; 5: 2798–2805.
- [38] POHL G, FILIPITS M, SUCHOMEL RW, STRANZL T, DE-PISCH D et al. Expression of the lung resistance protein (LRP) in primary breast cancer. *Anticancer Res* 1999; 19: 5051–5055.
- [39] MIYATAKE K, GEMBA K, UEOKA H, NISHII K, KIURA K et al. Prognostic significance of mutant p53 protein, P-glycoprotein and glutathione S-transferase-pi in patients with unresectable non-small cell lung cancer. *Anticancer Res* 2003; 23: 2829–2836.
- [40] CHEN ZJ, LE HB, ZHANG YK, QIAN LY, SEKHAR KR et al. Lung resistance protein and multidrug resistance protein in non-small cell lung cancer and their clinical significance. *J Int Med Res* 2011; 39: 1693–1700. <http://dx.doi.org/10.1177/147323001103900511>
- [41] BERGER W, SETINEK U, HOLLAUS P, ZIDEK T, STEINER E et al. Multidrug resistance markers P-glycoprotein, multidrug resistance protein 1, and lung resistance protein in non-small cell lung cancer: prognostic implications. *J Cancer Res Clin Oncol* 2005; 131: 355–363. <http://dx.doi.org/10.1007/s00432-004-0653-9>
- [42] IZQUIERDO MA, VAN DER ZEE AG, VERMORKEN JB, VAN DER VALK P, BELIËN JA et al. Drug resistance-associated marker Lrp for prediction of response to chemotherapy and prognoses in advanced ovarian carcinoma. *J Natl Cancer Inst* 1995; 87: 1230–1237. <http://dx.doi.org/10.1093/jnci/87.16.1230>
- [43] FILIPITS M, DRACH J, POHL G, SCHUSTER J, STRANZL T et al. Expression of the lung resistance protein predicts poor outcome in patients with multiple myeloma. *Clin Cancer Res* 1999; 5: 2426–2430.
- [44] FILIPITS M, STRANZL T, POHL G, HEINZL H, JAGER U et al. Drug resistance factors in acute myeloid leukemia: a comparative analysis. *Leukemia* 2000; 14: 68–76. <http://dx.doi.org/10.1038/sj.leu.2401634>

## Integrative computational analysis of transcriptional and epigenetic alterations implicates *DTX1* as a putative tumor suppressor gene in HNSCC

Daria A. Gaykalova<sup>1</sup>, Veronika Zizkova<sup>1,2</sup>, Theresa Guo<sup>1</sup>, Ilse Tiscareno<sup>1</sup>, Yingying Wei<sup>3,10</sup>, Rajita Vatapalli<sup>1,4</sup>, Patrick T. Hennessey<sup>1,5</sup>, Julie Ahn<sup>1</sup>, Ludmila Danilova<sup>3,11</sup>, Zubair Khan<sup>1</sup>, Justin A. Bishop<sup>1,6</sup>, J. Silvio Gutkind<sup>7</sup>, Wayne M. Koch<sup>1</sup>, William H. Westra<sup>1,6</sup>, Elana J. Fertig<sup>3</sup>, Michael F. Ochs<sup>3,8</sup>, Joseph A. Califano<sup>1,9</sup>

<sup>1</sup>Department of Otolaryngology—Head and Neck Surgery, Johns Hopkins Medical Institutions, Baltimore, Maryland, USA

<sup>2</sup>Institute of Molecular and Translational Medicine, Faculty of Medicine and Dentistry, Palacky University, Olomouc, Czech Republic

<sup>3</sup>Division of Oncology Biostatistics, Department of Oncology, Johns Hopkins Medical Institutions, Baltimore, Maryland, USA

<sup>4</sup>Department of Urology, Northwestern University, Chicago, Illinois, USA

<sup>5</sup>Department of Otolaryngology, Mid-Michigan Ear Nose and Throat, East Lansing, Michigan, USA

<sup>6</sup>Department of Pathology, Johns Hopkins Medical Institutions, Baltimore, Maryland, USA

<sup>7</sup>Department of Pharmacology, UC San Diego Moores Cancer Center, La Jolla, California, USA

<sup>8</sup>Department of Mathematics and Statistics, The College of New Jersey, Ewing, New Jersey, USA

<sup>9</sup>Department of Surgery, UC San Diego, Moores Cancer Center, La Jolla, California, USA

<sup>10</sup>Department of Statistics, The Chinese University of Hong Kong, NT, Shatin, Hong Kong

<sup>11</sup>Laboratory of Systems Biology and Computational Genetics, Vavilov Institute of General Genetics, Russian Academy of Sciences, Moscow, Russia

**Correspondence to:** Joseph A. Califano, **email:** jcalifano@ucsd.edu  
Daria A. Gaykalova, **email:** dgaykal1@jhmi.edu

**Keywords:** HNSCC, *DTX1*, expression, methylation, integration

**Received:** September 07, 2016

**Accepted:** January 16, 2017

**Published:** January 27, 2017

### ABSTRACT

Over a half million new cases of Head and Neck Squamous Cell Carcinoma (HNSCC) are diagnosed annually worldwide, however, 5 year overall survival is only 50% for HNSCC patients. Recently, high throughput technologies have accelerated the genome-wide characterization of HNSCC. However, comprehensive pipelines with statistical algorithms that account for HNSCC biology and perform independent confirmatory and functional validation of candidates are needed to identify the most biologically relevant genes. We applied outlier statistics to high throughput gene expression data, and identified 76 top-scoring candidates with significant differential expression in tumors compared to normal tissues. We identified 15 epigenetically regulated candidates by focusing on a subset of the genes with a negative correlation between gene expression and promoter methylation. Differential expression and methylation of 3 selected candidates (*BANK1*, *BIN2*, and *DTX1*) were confirmed in an independent HNSCC cohorts from Johns Hopkins and TCGA (The Cancer Genome Atlas). We further performed functional evaluation of NOTCH regulator, *DTX1*, which was downregulated by promoter hypermethylation in tumors, and demonstrated that decreased expression of *DTX1* in HNSCC tumors maybe associated with NOTCH pathway activation and increased migration potential.

## INTRODUCTION

As the fifth most common cancer, Head and Neck Squamous Cell Carcinoma (HNSCC) is responsible for 600,000 new cases and over 300,000 deaths per year worldwide [1, 2]. Nonetheless, the majority of HNSCC patients are diagnosed at an advanced stage due to the asymptomatic course of early stage disease and the absence of the routine screening techniques [3–6]. Development of low toxicity targeted therapeutics and biomarkers for early detection could improve survival rate and quality of life for HNSCC patients.

Traditionally, research groups have focused on the roles of individual genes in HNSCC to develop candidate biomarkers for diagnosis and treatment selection [7]. The process of single-gene investigation is time and labor intensive, while high-throughput profiling enables more rapid discovery of targetable disease-specific modifications. Thus, recent high throughput profiling of HNSCC identified number of targetable disease-specific modifications, such as genetic mutations and differentially expressed genes in HNSCC primary tissues or cell lines compared to normal samples [8–16]. In addition, epigenetic based expression alterations are noted to drive key biologic processes in HNSCC [17, 18]. Integration of high throughput data from expression and methylation platforms may enhance accurate discovery of cancer-driving genes so they can be used as disease biomarkers or therapy targets [19].

Therefore, a multi-platform high throughput analyses of gene expression and DNA methylation in primary HNSCC and normal samples and outlier statistics [20] were utilized to rank candidate genes and prioritize genes with the most prominent abnormalities in tumor samples that were absent in normal samples. Next, candidate genes were validated in separate clinical cohorts. Finally, the functional role of a lead candidate, *DTX1*, in HNSCC cell migration was demonstrated. *DTX1* expression was found to be increased in samples with decreased NOTCH pathway activity, suggesting that *DTX1* can serve as a biomarker of NOTCH pathway inhibition. The promoter DNA was exclusively methylated in tumors, suggesting that it can also serve as a HNSCC biomarker. This discovery was made possible due to employment of well-considered statistical approaches, cross cohort validation, and complementary detection tools.

## RESULTS

### Candidate genes with differential expression in HNSCC detected by outlier statistics

In order to identify relevant gene candidates in HNSCC, we used gene expression array data from a discovery cohort of 44 HNSCC primary tumors and 25 non-cancer normal tissue samples described in our

previous publications (Supplementary Table 1 and [21–23]). Notably, the clinical differences between tumor and control population of the discovery cohort were identified and discussed earlier [21–24].

The novelty of the current study was that we applied outlier analysis adopted from Ochs et al. [20] to rank and prioritize cancer-related alterations in HNSCC samples relative to normal controls (Figure 1). Outlier analysis is adapted for the study of heterogeneous samples, such as HNSCC primary tissues, because it is sensitive to alterations that maybe present in only a subset of samples. Given the high sensitivity of whole-genome gene expression analysis, thousands of differentially expressed genes can be detected while comparing tumor and normal samples. The heterogeneity of genetic and epigenetic alterations in solid tumors has presented challenges in using conventional statistical approaches, such as *t*-tests or signal-to-noise tests. There are several most-commonly accepted methods employed in cancer research for analysis of high-throughput data of heterogeneous cancers, including Cancer Outlier Profile Analysis (COPA)-based methodology, which compares outliers to an empirical null [25–28]. Outlier-based analysis has provided a mechanism to define significant, but diverse, alterations in cancers [20]. To eliminate low-signal outliers, this work implemented COPA-based statistics with a rank sum outlier approach as well as set a minimum level for the calling of an outlier. Such outlier analysis was recently successfully implemented and validated in a wet-lab setting for the discovery of tumor-specific signatures from DNA methylation array data for HNSCC [29].

Based on the number of outlier samples and the relative signal intensity, 76 of the top ranking candidates were chosen for further analysis (Supplementary Table 2). Overall, 50 candidate genes demonstrated increased gene expression and 26 candidate genes demonstrated decreased gene expression in tumor samples. (Supplementary Table 2). Notably, the standard and more stringent *t*-test demonstrated that 70 out of 76 genes (92%) had statistically significant difference in gene expression between normal and tumor samples.

### Negative correlation between expression and methylation identifies candidate genes that are epigenetically silenced in HNSCC

High gene promoter methylation if often associates with decreased gene expression and can result in epigenetic silencing [17, 18]. Therefore, DNA methylation array data was integrated with gene expression analysis (Figure 1) to identify gene expression changes potentially driven by methylation, as well as to eliminate biases from individual high-throughput platform [23]. The Illumina 27 DNA methylation array was utilized, which contained 27,000 probes covering approximately 15,000 genes, including gene promoter probes for 19 of the

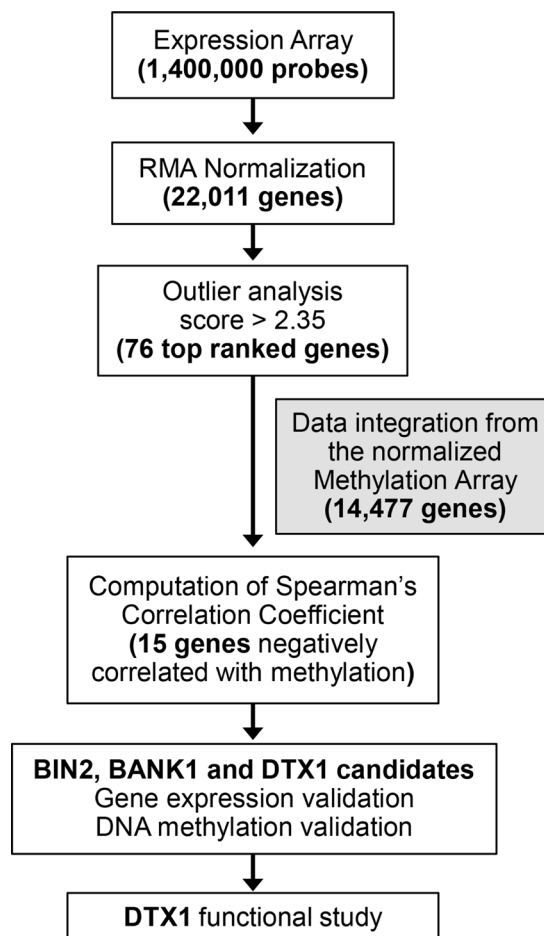
76 candidates. Spearman correlation coefficients were calculated (Supplementary Table 2), and 15 out of 19 genes had negative correlation between expression and methylation. The description of these 15 candidates can be found in Table 1.

Of the list of 15 candidates, *BANK1* (a scaffold protein) and *BIN2* (a bridging integrator protein) had the highest combined outlier score and negative Spearman coefficient. *DTX1* was chosen for its relatively high outlier score and its regulatory role in the NOTCH pathway, which is commonly dysregulated in HNSCC [12–14, 22]. All 3 genes have been implicated as potential cancer drivers in other non-head and neck solid tumors [30–33]. Notably, all 3 genes were downregulated and hypermethylated in tumor samples compared to normal controls (Figure 2). Therefore these 3 genes were selected for further validation. Other candidates, such as *CD79B*, *MAP4K1*, *GRAP*, *TNFRSF13C*, and *INA*, had comparable scores and are candidates for further study, as they have

also been implicated in carcinogenesis of other cancer types [34–38].

### Independent validation of differential expression and methylation of *BANK1*, *BIN2* and *DTX1*

To confirm the differential expression and methylation of *BANK1*, *BIN2* and *DTX1*, an independent validation cohort was assembled of 61 HNSCC primary tumors and 28 UPPP samples, with similar clinical characteristics as the discovery cohort (Supplementary Table 1). *BANK1*, *BIN2* and *DTX1* gene expression was evaluated by qRT-PCR, and DNA methylation was evaluated by bisulfite sequencing (Figure 3, Supplementary Table 3). Gene expression was significantly decreased in tumor tissues for all 3 genes (*t*-test *p*-values:  $9.4 \times 10^{-6}$ ,  $7.1 \times 10^{-6}$ , and 0.0013 for *DTX1*, *BANK1*, and *BIN2*, respectively), and DNA methylation was present in significantly more tumor samples (Fisher exact test



**Figure 1: Experimental flow.** Expression array probes, 1.4M total, were normalized using RMA package. Gene level estimates were produced by choosing the highest mean expression levels among all probes linked to the same gene for expression, yielding 22,011 genes. We applied outlier analysis [20] to the gene expression data set, containing 22,011 genes. The outlier score cut-off for expression data was set at 2.3, resulting in prioritizing 76 top scoring expression candidates. Spearman gene expression-methylation for these 76 candidates was calculated via integration of normalized methylation array data available for the samples. Fifteen out of 76 candidates were found to have a negative Spearman coefficient. Differential expression and methylation of *BIN2*, *BANK1* and *DTX1* were validated in the validation and TCGA-HNSCC cohorts. Functional study was performed for *DTX1*.

**Table 1: Fifteen candidate genes with negative expression-methylation correlation**

#	Gene	Description	Expression-Methylation Correlation, Spearman coefficient	Outlier score	Alteration in different tumor types	Reference
1	<i>ATP2A3</i>	ATPase	-0.120	2.45	HNSCC, lung, colon, cancers of central nervous system	[50]
2	<i>ATP8A1</i>	ATPase	-0.195	2.60	Lung cancer	[73]
3	<i>BANK1</i>	Scaffold protein	-0.420	4.97	Lymphoma, colorectal cancer	[30, 32]
4	<i>BIN2</i>	Bridging integrator	-0.693	2.94	myeloproliferative neoplasm	[31]
5	<i>CD79B</i>	immunoglobulin-beta protein	-0.556	3.16	Myeloma, CLL	[74, 75]
6	<i>CYP11B1</i>	Cytochrome	-0.183	2.50	Smoking related cancers, ovarian cancer	[51, 52]
7	<i>DTXI</i>	Notch-pathway regulator	-0.274	3.40	thymic tumor, glioblastoma, osteoblastoma	[46, 47]
8	<i>FZD3</i>	Frizzled receptor	-0.161	3.02	Colorectal, non-melanoma skin cancer, CLL	[53-55]
9	<i>GRAP</i>	cytoplasmic signaling protein	-0.444	2.93	medullary thyroid carcinoma	[34]
10	<i>INA</i>	Neurofilament	-0.333	2.69	colorectal cancer, adenomas	[35]
11	<i>MAP4K1</i>	MAP kinase	-0.590	2.58	Bladder, colorectal cancer	[36, 38]
12	<i>ORAOV1</i>	Oral cavity oncogene	-0.064	4.00	oral SCC	[56]
13	<i>PDE5A</i>	phosphodiesterase	-0.10	3.51	melanoma	[76]
14	<i>TNFRSF13C</i>	TNF receptor	-0.295	3.12	non-Hodgkin lymphoma	[37]
15	<i>VAV1</i>	proto-oncogene, a member of guanine nucleotide exchange factors	-0.208	2.64	neuroblastoma, lung, pancreatic cancer	[77, 78]

Genes are in alphabetic order. Spearman coefficient and Outlier score are provided.

*p*-values: 0.105, < 0.0001, and 0.0006 for *DTXI*, *BANK1*, and *BIN2*, respectively). Utilization of the independent cohort of HNSCC samples analyzed by complementary methodology (Figure 3) enhanced the rigor of the data through technical validation and eliminated potential sample biases. Notably, for *DTXI* differential methylation did not reach statistical significance (*p* = 0.105). However, *DTXI* showed almost no methylation in normal tissues, resulting in high tumor specificity. Therefore, *DTXI* was maintained as a gene candidate for further study.

High throughput gene expression and DNA methylation analysis was also recently performed by TCGA (The Cancer Genome Atlas), including 222 matched HNSCC tumors and 50 normal samples (Supplementary Table 1 and [39]). TCGA used RNA-Seq for gene expression evaluation and Illumina Infinium

HumanMethylation450 BeadChip platform for DNA methylation analysis. We used the TCGA dataset for *BANK1*, *BIN2* and *DTXI* validation (Figure 4). TCGA was not used for initial discovery because use of adjacent normal tissue from cancer patients was a concern. In the HNSCC population, high rate of tobacco and alcohol exposure lead to field of cancerization effect, as well as genetic and epigenetic changes can be seen in tumor adjacent apparently normal tissue [40, 41].

Nonetheless, data from TCGA was able to provide additional independent validation. Within TCGA, *DTXI* was found to have significantly decreased expression (*p*-value =  $8 \times 10^{-6}$ ) and promoter hypermethylation (*p*-value =  $3.11 \times 10^{-20}$ ), validating prior results. However *BANK1* shared promoter hypermethylation (*p*-value =  $6.36 \times 10^{-19}$ ) without significant changes in



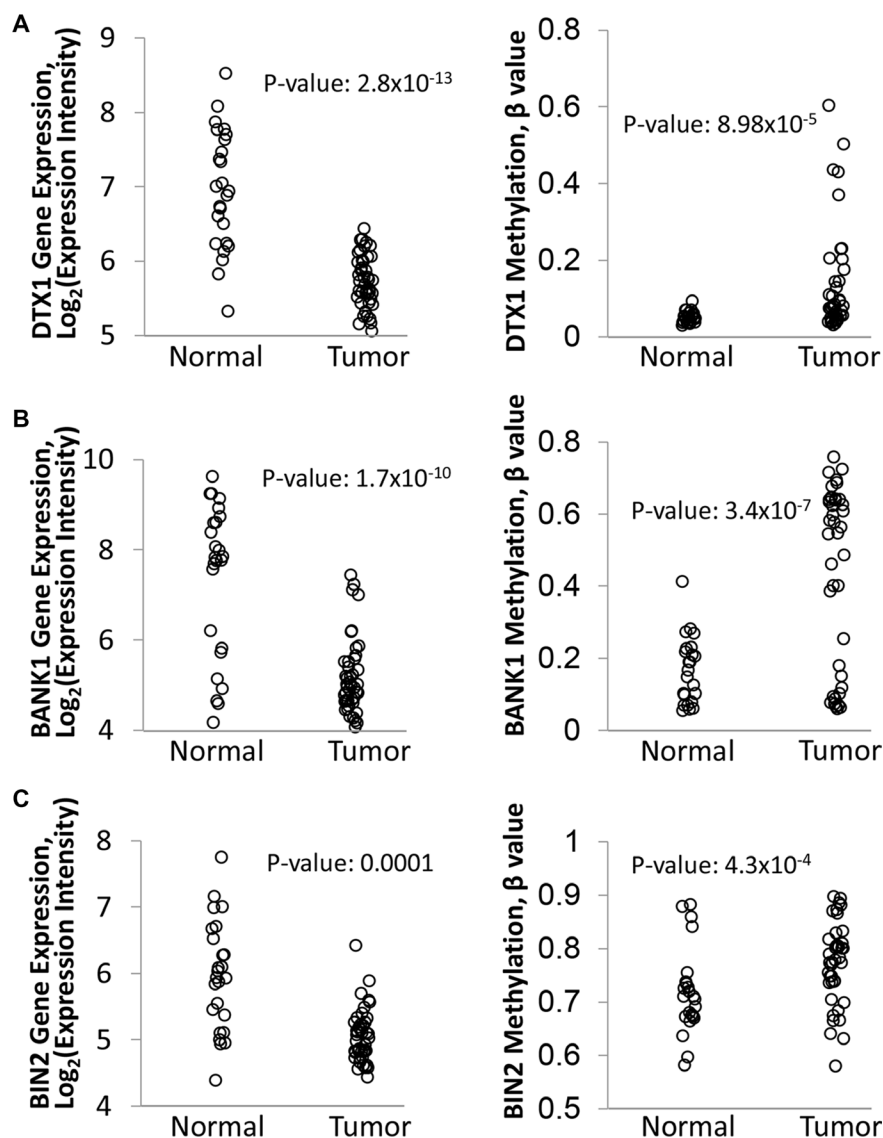
gene expression ( $p$ -value = 0.38). For *BIN2* neither gene expression downregulation nor hypermethylation changes were validated in TCGA. Since *DTX1* showed the best performance in TCGA cross validation, we focused our further functional validation only on the *DTX1* gene.

### The role of *DTX1* in the NOTCH pathway for HNSCC

*DTX1* is a regulator of the NOTCH pathway [42, 43]. Since the NOTCH pathway is dysregulated in HNSCC [12, 13, 22], we evaluated the role of *DTX1* relative to other NOTCH pathway genes (KEGG database and [22]). Overall, we analyzed gene expression for 44 NOTCH related genes for 44 HNSCC and 25 normal controls from the discovery cohort using Affymetrix Exon

Array data (Supplementary Figure 1). Genes were sorted by unsupervised hierarchical clustering. *DTX1* was noted to cluster together with *DTX3* (NOTCH regulator), *MFNG* (NOTCH modifier) and *DLL3* (NOTCH ligand).

Interestingly, separation of HNSCC samples by *DTX1* expression separated the tumor samples into two subsets of samples, with different expression of NOTCH pathway genes in each group (Supplementary Figure 1). To confirm this observation, we separated 44 tumors into two equal groups by *DTX1* expression: lower *DTX1* expression ( $n = 22$ ) and higher *DTX1* expression ( $n = 22$ ) and compared the expression of NOTCH genes in each group (Supplementary Table 4). Indeed, 34 out of 43 NOTCH pathway genes excluding *DTX1* (79%) were significantly differentially expressed between *DTX1* low and high expressed tumors. According to Supplementary



**Figure 2:** Differential expression and methylation of *DTX1* (A), *BANK1* (B) and *BIN2* (C) in the original discovery cohort. Gene expression (left) was evaluated by Affymetrix HuEx1.0 GeneChip. DNA methylation (right) was evaluated by Illumina Infinium HumanMethylation27 BeadChip platform, and the data was normalized and processed as described in methods.  $P$ -value were calculated by  $t$ -test.

Figure 1 and Supplementary Table 4, the majority of NOTCH pathway genes were significantly upregulated in the group of samples with lower *DTX1* expression ( $n = 28$  genes, including *NOTCH1-3*, *HES1*, *HEY1*, and *JAG1*) relative to *DTX1* high expression samples. Fifteen genes had lower expression in the *DTX1* lower-expression group relative to *DTX1* high expression samples, including *DLLs*, *DTXs* and *NOTCH4* genes. Gene set analysis confirmed that the set of NOTCH pathway genes were significantly overexpressed in *DTX1* low samples relative to *DTX1* high ( $p$ -value of 0.0012, Supplementary Figure 2) and relative to Normal samples ( $p$ -value of 0.0024, Supplementary Figure 2). These findings confirm that downregulation of *DTX1* results in a strong difference in activation of the NOTCH pathway in HNSCC samples.

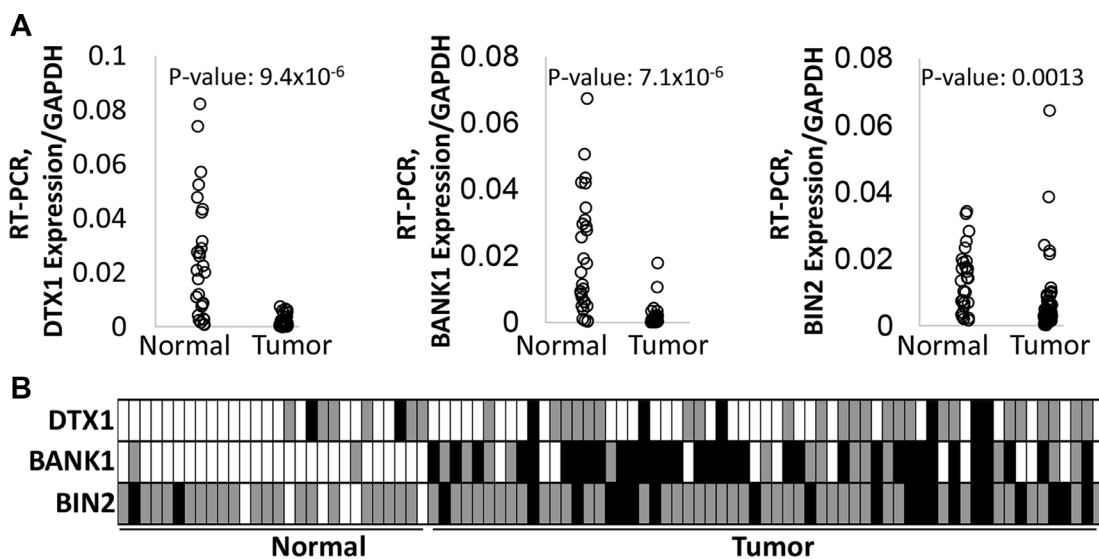
We further evaluated relative expression of *DTX1* with NOTCH downstream targets (Supplementary Figure 3). *DTX1* expression co-clustered with the expression of *GATA4* (regulated by *HEY1*) [44] and *NEUROG3* (negatively regulated by *HES1*) [45] by unsupervised hierarchical clustering. The set of NOTCH downstream targets were not significantly differentially expressed between *DTX1* sample groups or relative to normal samples with gene set analysis. Nonetheless, *GATA4* and *NEUROG3* were all downregulated in samples with low *DTX1* expression relative to high *DTX1* expression ( $p$ -values of  $7 \times 10^{-7}$  and  $6 \times 10^{-10}$ , respectively; Supplementary Table 5). In addition, *DTX1* low samples had significantly higher expression of *HES1* relative to *DTX1* high samples ( $p$ -value of 0.01; Supplementary Table 5), consistent with *NEUROG3* downregulation [45], but significant changes were not observed from gene set statistics (data not shown).

## Functional role of *DTX1* dysregulation in HNSCC

Since *DTX1* was downregulated in HNSCC tumor samples, (Figures 2–4) it is expected to have tumor-suppressor properties. Overexpression and silencing of *DTX1* expression *in vitro* did not affect cell proliferation (3 immortalized normal keratinocyte and 6 HNSCC cell lines were tested, data not shown). On the other hand, recent data suggest that *DTX1* may play a role in inhibition of invasion in osteosarcoma [33]. In order to evaluate if *DTX1* could modify the invasiveness of HNSCC cells as well, we performed matrigel cell migration assays. UM-SCC-047 and UM-SCC-22B were selected due to their increased mobility relative to other HNSCC cell lines necessary for invasion assay. Base-line *DTX1* expression analysis determined that UM-SCC-047 had relatively lower gene expression and, therefore, was used for ectopic *DTX1* expression (Supplementary Figure 4A). On the other hand, UM-SCC-22B had a higher *DTX1* expression rate and, therefore, was used for knock-down RNAi experiments. Upregulation of *DTX1* via ectopic expression led to significant decrease of UM-SCC-047 cell invasion ( $p = 0.014$ , Figure 5A and Supplementary Figure 5A), while downregulation of *DTX1* by RNAi led to strong enhancement of cell invasiveness in UM-SCC-22B cell line ( $p = 0.004$ , Figure 5B and Supplementary Figure 5B).

## DISCUSSION

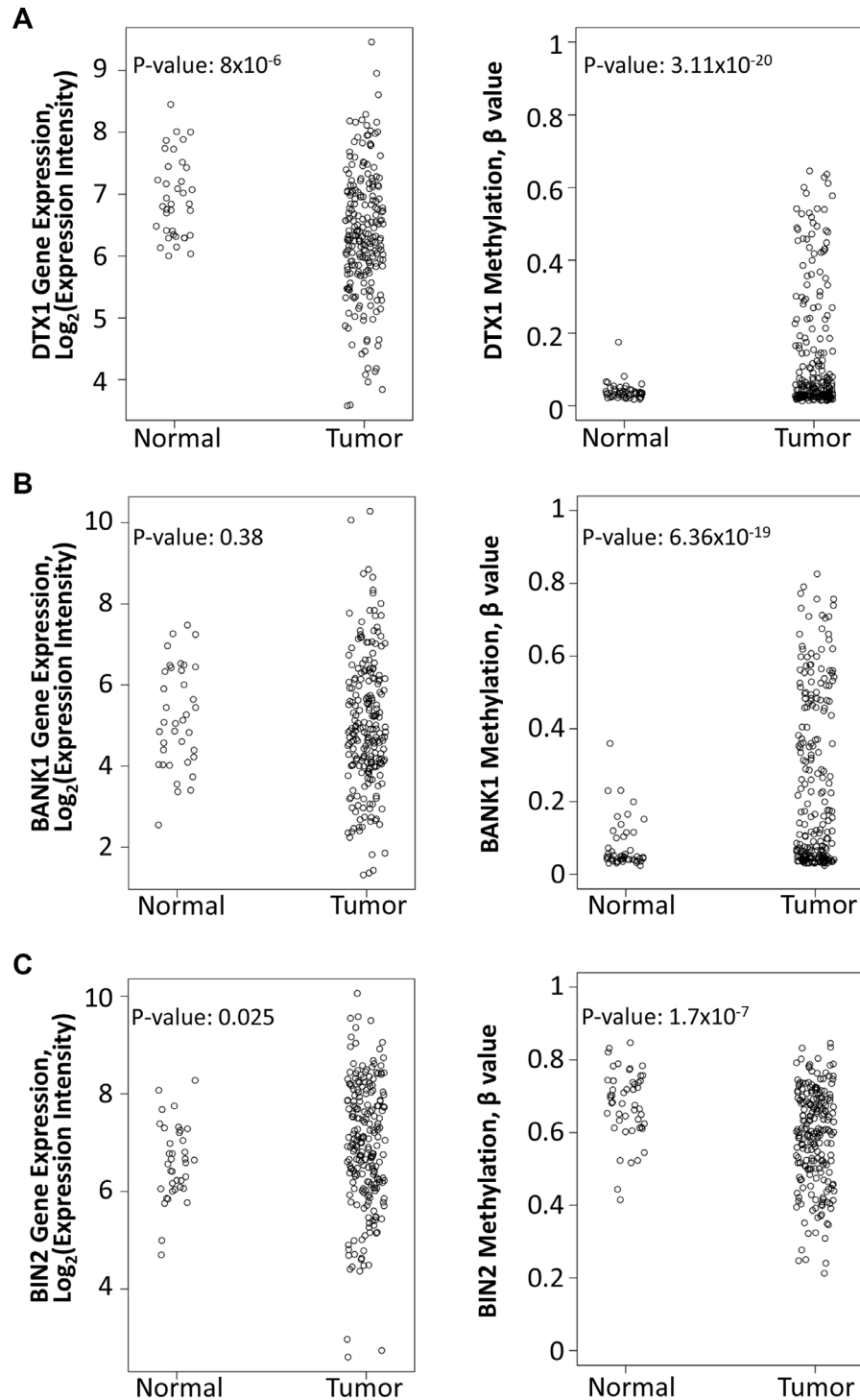
Five year survival for HNSCC is only 50%, and there is a clear need for identification of novel cancer-



**Figure 3: Differential expression and methylation of *DTX1*, *BANK1* and *BIN2* in the validation cohort.** Gene expression (A) was evaluated by quantitative RT-PCR.  $P$ -values were calculated by  $t$ -test. DNA methylation (B) was evaluated by bisulfite sequencing. Box color-code: white—unmethylated (hypomethylated); grey—hemimethylated, black—hypermethylated.  $P$ -values were calculated by Fisher exact test, as unmethylated signal vs methylated signal (hemi- or hypermethylated) in two groups.  $P$ -values for *DTX1* = 0.105, for *BANK1* < 0.0001, for *BIN1* = 0.0006.

specific therapeutic targets. This study integrated DNA methylation and gene expression to define cancer-related alterations. We performed genomic analysis with thorough confirmatory and functional validation, which identified *DTX1* as a potential regulator of migration in HNSCC.

The integrated analysis and outlier statistics allowed us to correlate gene expression with promoter methylation and discriminate candidates that were biologically relevant. Indeed, all 15 high-value candidates identified through these methods (Table 1) have been described as cancer drivers in other cancer types. Thus, *DTX1*



**Figure 4:** Differential expression and methylation of *DTX1* (A), *BANK1* (B) and *BIN2* (C) in the TCGA-HNSCC cohort. Gene expression (left) was evaluated by RNA-Seq. DNA methylation (right) was evaluated by Illumina Infinium HumanMethylation450 BeadChip platform, and the data was normalized and processed as described in methods. *P*-values were calculated by *t*-test.

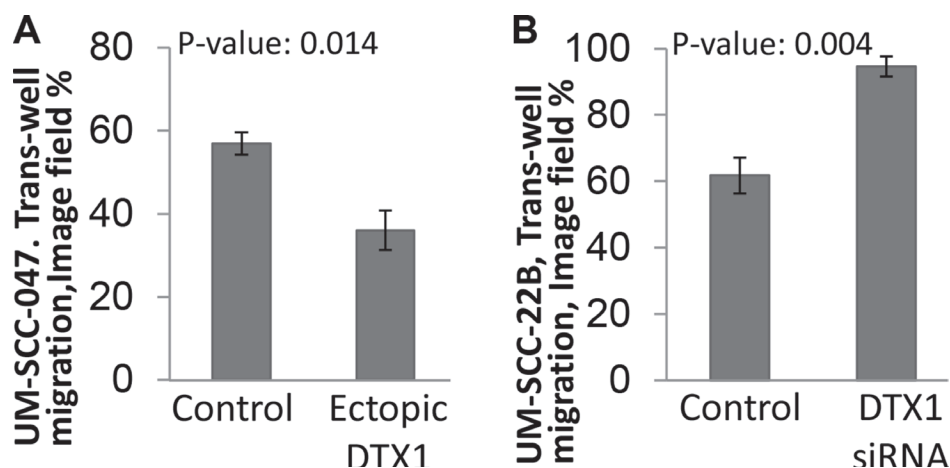
is a regulator of the NOTCH pathway; this pathway was recently found both downregulated in HNSCC via *NOTCH1* mutations and upregulated in HNSCC via amplifications of NOTCH's ligands and receptors [12–14, 22]. *DTX1* is upregulated in thymic tumors and in glioblastoma [46, 47], and it inhibits osteoblastoma cell invasion [33]. Moreover, multiple *DTX1* polymorphisms were found in non-small cell lung and B-cell precursor acute lymphoblastic leukemia patients [48, 49], which most likely associate with inactivation of *DTX1* and further NOTCH activation in these cancers. *BANK1* is downregulated in lymphoma and colorectal cancers [30, 32]. Overexpressed *BIN2* fusions were detected in myeloproliferative neoplasms [31]. Out of other candidates, *ATP2A3* has been shown to be mutated and downregulated in HNSCC, lung, colon and central nervous system cancers [50]. Several polymorphism of *CYP1B1* are found in many cancer types, including tobacco-related cancers; *CYP1B1* is also overexpressed in ovarian cancer [51, 52]. Wnt-pathway receptor *FZD3* is strongly expressed in colorectal and non-melanoma skin cancer and during chronic lymphocytic leukemia, CLL [53–55]. *ORAOV1* is an overexpressed marker of oral SCC [56]. Information about other genes can be found in Table 1.

In order to confirm the role of identified candidate genes, the 3 candidates were independently validated in multiple cohorts. As acknowledged by Mirghani et al [57], high throughput data is often poorly validated, with biases inherent to a single institution cohort or single methodology. To improve the candidate discovery pipeline, we employed both single-institution and multi-institutional HNSCC cohorts, and utilized diverse detection platforms: Illumina exon array, RNA-Seq and qRT-PCR for gene expression; and Methylation arrays

27 and 450, as well as bisulfite sequencing for DNA methylation. Moreover the 3 genes were confirmed to be hypomethylated and upregulated in healthy tissues of non-cancer patients and in non-cancer tumor-adjacent tissues of HNSCC patients with minor exclusions. While only 3 exemplary candidate genes out of the total 15 candidates were evaluated, the others may be expected to have strong differential expression and methylation in HNSCC regardless of sample cohort and detection tool.

Analysis of gene expression of NOTCH pathway members suggested that *DTX1* downregulation in HNSCC is correlated with downregulation of some NOTCH pathway genes including *DTX3*, *DLL3*, and *MFNG*, as well as genes downstream of NOTCH (*NEUROG3* and *GATA4*). Furthermore, *DTX1* expression was inversely correlated with *HES1* expression. These results confirm prior data that *HES1* is a negative regulator of *NEUROG3* and *DTXs* expression [33, 45]. *DTX1* carries a putative SH3-binding domain and binds to the intracellular domain of NOTCH (ICN) [42]. One mechanism by which *DTX1* may negatively regulate NOTCH is through ubiquitination; *DTX1* may be an E3 ubiquitin ligase, as it contains a RING finger and two WWE domains [58]. Zhang and colleagues [33] proposed that *DTX1* may bind and ubiquitinate NOTCH's ICN, leading to negative regulation of NOTCH pathway, which was consistent with our findings (Supplementary Figure S6). *HEY1* also negatively regulates *GATA4* [44], which was downregulated in parallel with *DTX1*.

Overall, *DTX1* was downregulated in the entire HNSCC discovery cohort, together with 20 other NOTCH pathway genes including *DLLs*, *NOTCH1*, *NOTCH2* and *NOTCH4* (Supplementary Table 4). *JAG1*, *JAG2*, *HEY1* and *HES1* had relatively higher expression in



**Figure 5: *DTX1* blocks HNSCC invasiveness.** Migration assay was performed using UM-SCC-047 (A) or UM-SCC-22B (B) cells using transient transfection. The image of cells that had invaded through matrigel (Supplementary Figure 5) was processed and quantified in Photoshop. Both UM-SCC-047 and UM-SCC-22B cells had similar 60% invasion when treated with control constructs (empty vector for ectopic expression or non-targeting siRNA pool for RNAi). The migration of each cell was dysregulated significantly by ectopic *DTX1* overexpression (a, UM-SCC-047 cells) or by transient *DTX1* downregulation (b, UM-SCC-22B cells). *P*-value were calculated by *t*-test for experiments performed in triplicate. Transfection efficiency for each experiment was confirmed by qRT-PCR (Supplementary Figure 4).

tumor samples compared to the pool of normal controls (Supplementary Table 4). This data suggested that *DTX1* expression can serve as biomarker of NOTCH pathway inhibition. While the NOTCH pathway is activated in many cancer types, *DTX1* is seen to be downregulated in osteoblastoma and HNSCC, while it is upregulated in glioblastoma [22, 33, 46, 59]. In addition, the NOTCH pathway was also found to be downregulated in thyroid cancer and in subgroup of HNSCC patients [22, 60]. Notably, genetic alterations of *NOTCH1* were recently found to dysregulate the NOTCH pathway in HNSCC [12, 13]. However, we did not find any correlation of *DTX1* expression with presence of *NOTCH1* mutations [14, 22, 61, 62]. Interestingly, 1/3 of NOTCH pathway genes are co-downregulated with *DTX1* (such as *DLLs*, *DTXs*, and *NOTCH4*), while the rest of the NOTCH pathway genes were upregulated in the low *DTX1* expression subgroup (genes including *NOTCH1-3*, *HES1*, *HEY1*, and *JAG1*, Supplementary Tables 4–5). These results correlate with recent discovery that NOTCH pathway has complex gene interactions and dual function, where it is activated in some tumors while inactivated in others [22, 61]. Notably, there were more HPV-related (HPV+) patients in the lower *DTX1* expression group ( $n = 9$  HPV+ patients), compared to the higher *DTX1* expression group ( $n = 4$  HPV+ patients, Fisher's exact test  $p$ -value = 0.185). There were 4 times more oral cavity patients in the higher *DTX1* expression group ( $n = 8$ , 0.069). Unfortunately, those or other correlations with clinical characteristics did not reach statistical significance. This is the first report that has demonstrated that *DTX1* blocks HNSCC migration, which is in agreement with recently published data suggesting that *DTX1* blocks osteosarcoma invasiveness [33]. Results of current analysis suggests that in HNSCC, downregulation of *DTX1* by DNA methylation leads to more aggressive behavior of HNSCC cells.

Since the expression of *NEUROG3* and *GATA4* (downstream of *HES1* and *HEY1* respectively) was downregulated in parallel with *DTX1*, we speculate, that *NEUROG3* and *GATA4* have a negative effect on cell migration (Supplementary Figure 6). The molecular mechanism by which *DTX1* blocks HNSCC cell migration needs to be further evaluated. Additional analysis of role of *DTX1* on cell proliferation did not show any significant changes of cell growth depending of *DTX1* expression (data not shown).

We have to acknowledge several limitations of our study: 1) clinical characteristics between tumor and non-tumor groups do not match in patients in both discovery and validation cohorts (Supplementary Table 1), due to peculiarity of UPPP and HNSCC populations [21–24]. Nonetheless the employed UPPP population helped revealing strong cancer-specific signatures of HNSCC in previous studies [21–24]. Moreover, employment

of TCGA's control population with matched clinical characteristics confirmed our original discovery of hypermethylation and downregulation of candidate genes, especially the leading candidate, *DTX1*, in tumor samples. 2) Utilization of older generation DNA methylation array data (Illumina Infinium HumanMethylation27 BeadChip) narrowed down the list of candidates by lack of available promoter-methylation data for several genes. Nonetheless this is one of the largest HNSCC cohorts, after the employed TCGA, with publicly available matched gene expression and promoter methylation data [21–24]. We see this study as confirmation of our pipeline for discovery of biologically-relevant candidates and smaller number of candidate genes helped us focus only on the limited number of genes within a limited time frame. 3) Not all candidates were functionally evaluated within given time frame, but will be used for further independent analyses. The complete list of high priority candidates discovered during this project will become available for the research community for their prospective studies.

Thousands of alterations can be detected by different independent high throughput platforms, given their high sensitivity. Integration of gene expression and DNA methylation high throughput data focused study to a limited list of relevant genes with potential roles in HNSCC carcinogenesis. Employment of well-considered statistical approaches, cross cohort validation, and complementary detection tools allowed us to discover an epigenetically regulated tumor suppressor gene, *DTX1*, which controls HNSCC cell migration.

## MATERIALS AND METHODS

### Specimen cohort assembly

We used two independent cohorts of specimens, each composed of primary head and neck squamous cell carcinoma (HNSCC) tissue specimens and control specimens comprising normal mucosal samples from uvulopalatopharyngoplasty (UPPP) surgeries of non-cancer affected patients. The discovery cohort comprised of 44 HNSCC and 25 normal UPPP samples, as described in previous publications [22, 23]. The validation cohort comprised of 61 HNSCC and 28 normal UPPP samples, is reported for the first time. The demographic and clinicopathological characteristics of patients from the discovery and validation cohorts are listed in Supplementary Table 1. The clinical differences between tumor and control populations in the discovery cohort were previously identified and acknowledged [21–24].

All tissue samples were obtained from the Johns Hopkins Tissue Core, as a part of the Head and Neck Cancer Specialized Program of Research Excellence (HNC-SPORE). These samples were acquired under Internal Review Board-approved research protocol

#NA\_00036235. Informed consent was obtained from all patients recruited under this protocol prior to participation in the study.

### Tissue processing

Two Johns Hopkins Hospital Pathologists (WHW and JAB) independently confirmed that all primary tumor samples were consistent with HNSCC. After this, all tumor tissues were microdissected to yield at least 80% tumor purity. All tissue specimens were stored at  $-140^{\circ}\text{C}$  until a cut and extraction were performed. For each extraction, a 0.35 mm thick cut of tissue was used.

### RNA preparation

RNA was isolated from 0.35 mm thick frozen tissue cuts with the mirVana miRNA Isolation Kit (Ambion, Forster City, CA) at room temperature as per manufacturer's recommendations. The concentration of the isolated RNA was quantified using the NanoDrop spectrophotometer (Thermo Fisher Scientific, Waltham, MA).

### DNA preparation

Similarly, 0.35 mm thick frozen tissue cuts were digested in 1% SDS (Sigma-Aldrich, St. Louis, MO) and 50  $\mu\text{g}/\text{ml}$  proteinase K (Invitrogen, Carlsbad, CA) solution at  $48^{\circ}\text{C}$  for 48 hours. The DNA was purified by phenol-chloroform extraction and ethanol precipitation as previously described [63]. DNA was resuspended in LoTE buffer, and the DNA concentration was quantified using the NanoDrop spectrophotometer.

### High throughput transcriptional and methylation profiling data

High throughput data of gene expression and DNA methylation was obtained from the discovery cohort which was previously published using the methods described previously [22, 23]. The gene expression data for the discovery cohort was obtained from Affymetrix HuEx1.0 GeneChips (containing 1.4 million probes) and is available on the NCBI Gene Expression Omnibus (GEO) public repository (GEO33205). DNA methylation data for the discovery cohort was obtained using Illumina Infinium HumanMethylation27 BeadChip (probing 27,578 CpG dinucleotides) is also publically available (GEO33202). The data from both platforms can be downloaded from the combined superSeries GSE33232.

### Reverse transcription and quantitative real time PCR

Validation of gene expression was performed in the independent validation cohort using reverse transcription

and quantitative real-time PCR. One microgram of RNA from each sample in the validation cohort was reverse transcribed using the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems, Forster City, CA). Quantitative real-time PCR (qRT-PCR) was performed using gene-specific expression assays (Supplementary Table 3) and Universal PCR Master Mix on a 7900HT real-time PCR machine (Applied Biosystems) as per manufacturer's recommendations. Expression of the gene of interest was quantified in triplicate relative to *GAPDH* expression using the  $2^{-\Delta\Delta\text{CT}}$  method [64]. We also confirmed that *GAPDH* expression was not significantly different in normal and HNSCC samples.

### Bisulfite treatment and bisulfite genomic sequencing

Validation of DNA methylation was performed in the validation cohort through bisulfite genomic sequencing. The EpiTect Bisulfite Kit (Qiagen, Valencia, CA) was used to convert unmethylated cytosines to uracil in genomic DNA. Touch-down PCR was performed on bisulfite-converted DNA with primers designed to span areas of CpG islands for each gene promoter (Supplementary Table 3). [65]. The PCR products were purified using the QIAquick 96 PCR Purification Kit (Qiagen). The purified PCR product from bisulfite-converted DNA for each sample and gene was sequenced (Genewiz, South Plainfield, NJ). Relative heights of C and T peaks measured on sequencing were then used to assign the DNA methylation status as unmethylated, methylated or hemimethylated.

### TCGA sample selection

The TCGA project for HNSCC was recently completed and published [39]. Overall, the analysis included  $n = 279$  HNSCC samples. Such a cohort was different from the employed Johns Hopkins cohorts by increased number of oral cavity samples, samples with TNM stage II–III, and decreased HPV-positive samples. In order to “match” clinical characteristics of the analyzed TCGA cohort, we removed oral cavity samples with TNM stage II and III. We did not do any additional manipulation of the TCGA cohort to keep the “matching” procedure simple and randomized, as well as to avoid any potential biases due to non-random sample selection. Since oral cavity samples are predominately HPV-negative, this helped to increase relative percentage of HPV-positive HNSCC samples within the new “matched” TCGA population ( $n = 222$  total). Notably, the new “matched” TCGA cohort did not significantly change other clinical characteristics that were relevant to the Johns Hopkins cohort. This “matched” TCGA cohort (Supplementary Table 1) was used for validation purposes (Figure 4).

## Cell culture

### Cell lines and cell culture conditions

Human HNSCC cell lines UM-SCC-047 and UM-SCC-22B were provided by Dr. Thomas Carey (University of Michigan) for the functional experiments. Each cell line was authenticated using a Short Tandem Repeat (STR) Identifier kit (Applied Biosystems). Cells were grown on high-glucose DMEM media (Clontech, Mountain View, CA), supplemented by 10% fetal bovine serum (FBS) and 1% Penicillin-Streptomycin at 37°C in 5% CO<sub>2</sub>.

### Transient transfection

For knockdown assays, the expression of *DTX1* gene was downregulated by ON-TARGETplus siRNA SMARTpool RNA (L-006525-00-0005, Thermo Scientific, Waltham, MA) using RNAiMAX transfection reagent (Life Technologies, Carlsbad, CA). Non-targeting SMARTpool RNA (D-001810-10-05, Life Technologies) was used as a control. The transfection efficiency was confirmed by qRT-PCR.

The ectopic overexpression of *DTX1* was achieved with pCMV6-Entry-DTX1 plasmid (RC208338, Origene, Rockville, MD) using FuGENE Extreme 9 transfection reagent (Roche, Nutley, NJ). Empty pCMV6-Entry (PS100001, Origene) was used as a control. Transfection efficiency was confirmed by *DTX1*-specific qRT-PCR using the same qRT-PCR technique as described earlier for tissue RNA analysis.

### Matrigel invasion assay

We performed the Matrigel invasion assay to assess the migration and invasion ability of transfected cells with over expression and under expression of *DTX1* using techniques described in previous publications [66]. In short, 8- $\mu$ m pore filter inserts in 24-well plates (Sigma-Aldrich) coated with Matrigel (BD Biosciences, San Jose, CA) was used. Cells were transfected for 24 hours and then were trypsinized, washed three times with serum-free DMEM media and resuspended in serum-free DMEM to obtain the concentration of 10<sup>6</sup> cells/ml. An aliquot of 100  $\mu$ l of cells were plated onto each insert. Chemo-attractant media with 10% FBS (600  $\mu$ l) was added to the bottom of a 24-well plate. Each insert, with cell suspensions, was placed into the individual well with chemo-attractant media. After 24 hours of incubation at 37°C in 5% CO<sub>2</sub>, the inserts were removed from the media. Cells on the upper surface of the insert that did not invade through the membrane were removed with a cotton swab. The cells that had migrated to the lower surface of the membrane, facing the chemo-attractant media, were fixed by 10% formaldehyde and stained by 1% crystal violet. The membranes with fixed and stained cells were removed, mounted onto slides and photographed by microscopy at 4 $\times$  magnification. Each experiment was performed in triplicate.

## Matrigel migration quantification

The 4 $\times$  magnified images of the insert membrane were analyzed using Adobe Photoshop SC6 (Adobe Systems, McLean, VA). Stained cell-occupied image area (purple) was selected by the “Color Range” tool with 70% fuzziness. The number of pixels in the entire image and the number of pixels within areas occupied by cells were calculated by the “Histogram” tool of Photoshop. The percentage of image field occupied by cells was calculated as total number of pixels occupied by cells relative to the total number of pixels. Triplicate images were analyzed for each experiment and the mean of percentage of the cell-occupied image field was calculated.

## Statistical analysis

### Expression array normalization

The gene expression data from GEO33205 used Robust Multiarray Average (RMA) implemented in the Bioconductor oligo package [67, 68] for normalization, as previously described [21, 22]. The gene level expression estimates were calculated as the mean expression levels among all core probes linked to the same gene, yielding 22,011 genes.

### DNA methylation array normalization

For promoter methylation data available at GEO33202, beta values (percent methylation) were estimated from unmethylated (U) and methylated (M) measurements on a probe level basis:  $\beta = M/(M+U)$  [23, 24]. The gene level estimates were produced by choosing the highest methylation levels among all probes linked to the same gene yielding, 14,477 individual genes.

### Outlier analysis

HNSCC is a heterogeneous disease with cancer-related changes detected in only a small portion of the samples [12–14, 39]. Unfortunately, such changes are poorly detected by conventional statistical approaches such as the *t*-test. A standard method employed in cancer research for outlier analysis is Cancer Outlier Profile Analysis (COPA) and its derivatives [28, 69], which generate statistics by comparing the outlier distributions to an empirical null generated by permutation of class labels. However, these methods have limitations when counting outliers since the distribution of medians and median absolute deviations permits outliers to be called in cases where the deviations are biologically insignificant. Therefore, we recently implemented a rank sum outlier approach, modified from Ghosh [70], where a minimum change levels was set for the calling of an outlier [20]. Such methods allowed us to eliminate many outliers where change is not biologically meaningful (e.g., a gene expression change of less than 10%, or 2.35 log fold change, between any two samples). The outlier statistics was used exclusively for discovery purposes.

To discover the genes with changes in expression, we applied the outlier statistics described above in reference [20] to the array gene expression data set, containing the 22,011 genes for each of 44 HNSCC tissue samples from the discovery cohort. The signals from 25 normal samples from the same cohort were used to establish the empirical null level for each gene. We calculated outlier score for both left-tail (10th percentile) and right-tail (90th percentile) cases, which allowed us to define outliers that were downregulated and upregulated, respectively [20, 70]. The outlier statistics yielded an outlier score, which quantified the number of tumors with gene expression values that were outliers from the distribution defined by normal as defined in previously published work [20]. Each of the 22,011 genes was assigned its outlier score and ranked from the largest to the smallest. Outlier analysis does not have a cut-off value of significance, such as 0.05 for *p*-value; therefore for the convenience we manually selected the genes with the top 76 outlier scores as candidates (manual outlier score cut-off set = 2.3, Supplementary Table 2). All outlier analyses were performed with custom scripts in R adapted from [20].

#### The pathway enrichment analysis

The pathway enrichment analysis utilized gene expression data was used for 43 NOTCH pathway genes (KEGG database and [22] without *DTX1*) with available Affymetrix Exon Array data. Differential expression analysis of NOTCH pathway and downstream genes was performed with empirical Bayes moderated *t*-statistics using limma R package 2.12.0 [71]. Contrasts were formulated to define the difference between different gene expression in groups of patient samples: tumor samples with higher *DTX1* expression, tumor samples with lower *DTX1* expression, and non-cancer samples. *P*-values for differential expression statistics were reported after FDR adjustment with Benjamini-Hochberg correction [72] among 43 NOTCH pathway and 9 downstream genes, including *DTX1* in both sets. Genes with FDR adjusted *p*-values below 0.05 were called significantly differentially expressed. Pathway-level statistics were computed by applying the limma function *geneSetTest* to the empirical Bayes moderated *t*-statistics for each contrast with the alternative hypothesis of “either” specifying that genes in the set are up or down regulated as a group.

#### Expression-methylation correlation

We utilized a correlation analysis to associate changes in gene expression with epigenetic regulation. Specifically, we computed Spearman correlation coefficients between gene-level estimates of DNA methylation and expression for each candidate gene inferred from outlier statistics. Candidate genes with negative correlations between expression and methylation were selected. All correlation analyses were performed using limma R package 2.12.0 [71].

#### *P*-value calculation

Log transform of gene expression values from array, normalized qRT-PCR gene expression values, DNA methylation  $\beta$ -values, and image field percentage values were compared for HNSCC and control samples using the Student *t*-test. We used *t*-test for all our validation step as it is more stringent than outlier test, used exclusively for discovery purposes. Bisulfite sequencing results were compared for HNSCC and control samples using the two-tailed Fisher exact test.

#### CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

#### GRANT SUPPORT

The work is supported by P30CA006973 (Cancer Center Core), NIDCR/NCI P50CA019032 Head and Neck Cancer SPORE (JAC and DAG), NIDCR/NIH Challenge Grant RC1DE020324 (JAC), NIDCR/NIH R21DE025398 (DAG), NIDCR/NIH R01DE023347 (JAC), K25CA141053 and R01CA177669 (EJF), NIH/NLM R01LM011000 (MFO), and NIDCR/NIH R01DE013152 (WMK).

#### REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. *CA Cancer J Clin.* 2016; 66:7–30.
2. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. *CA Cancer J Clin.* 2015; 65:87–108.
3. Argiris A, Karamouzis MV, Raben D, Ferris RL. Head and neck cancer. *Lancet.* 2008; 371:1695–709.
4. Mignogna MD, Fedele S, Lo Russo L. The World Cancer Report and the burden of oral cancer. *European journal of cancer prevention.* 2004; 13:139–42.
5. Howlader N, Ries LA, Mariotto AB, Reichman ME, Ruhl J, Cronin KA. Improved estimates of cancer-specific survival rates from population-based data. *J Natl Cancer Inst.* 2010; 102:1584–98.
6. Tan M, Myers JN, Agrawal N. Oral cavity and oropharyngeal squamous cell carcinoma genomics. *Otolaryngologic clinics of North America.* 2013; 46:545–66.
7. Rocco JW, Leong CO, Kuperwasser N, DeYoung MP, Ellisen LW. p63 mediates survival in squamous cell carcinoma by suppression of p73-dependent apoptosis. *Cancer Cell.* 2006; 9:45–56.
8. Han J, Kioi M, Chu WS, Kasperbauer JL, Strome SE, Puri RK. Identification of potential therapeutic targets in human head & neck squamous cell carcinoma. *Head & neck oncology.* 2009; 1:27.



9. Chung CH, Parker JS, Karaca G, Wu J, Funkhouser WK, Moore D, Butterfoss D, Xiang D, Zanation A, Yin X, Shockley WW, Weissler MC, Dressler LG, et al. Molecular classification of head and neck squamous cell carcinomas using patterns of gene expression. *Cancer Cell*. 2004; 5:489–500.
10. Carles A, Millon R, Cromer A, Ganguli G, Lemaire F, Young J, Wasyluk C, Muller D, Schultz I, Rabouel Y, Dembele D, Zhao C, Marchal P, et al. Head and neck squamous cell carcinoma transcriptome analysis by comprehensive validated differential display. *Oncogene*. 2006; 25:1821–31.
11. Leethanakul C, Knezevic V, Patel V, Amornphimoltham P, Gillespie J, Shillitoe EJ, Emko P, Park MH, Emmert-Buck MR, Strausberg RL, Krizman DB, Gutkind JS, Head, et al. Gene discovery in oral squamous cell carcinoma through the Head and Neck Cancer Genome Anatomy Project: confirmation by microarray analysis. *Oral Oncol*. 2003; 39:248–58.
12. Agrawal N, Frederick MJ, Pickering CR, Bettgowda C, Chang K, Li RJ, Fakhry C, Xie TX, Zhang J, Wang J, Zhang N, El-Naggar AK, Jasser SA, et al. Exome sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in NOTCH1. *Science*. 2011; 333:1154–7.
13. Stransky N, Egloff AM, Tward AD, Kostic AD, Cibulskis K, Sivachenko A, Kryukov GV, Lawrence MS, Sougnez C, McKenna A, Shefler E, Ramos AH, Stojanov P, et al. The mutational landscape of head and neck squamous cell carcinoma. *Science*. 2011; 333:1157–60.
14. Gaykalova DA, Mambo E, Choudhary A, Houghton J, Buddavarapu K, Sanford T, Darden W, Adai A, Hadd A, Latham G, Danilova LV, Bishop J, Li RJ, et al. Novel insight into mutational landscape of head and neck squamous cell carcinoma. *PLoS One*. 2014; 9:e93102.
15. Lui VW, Hedberg ML, Li H, Vangara BS, Pendleton K, Zeng Y, Lu Y, Zhang Q, Du Y, Gilbert BR, Freilino M, Sauerwein S, Peyser ND, et al. Frequent mutation of the PI3K pathway in head and neck cancer defines predictive biomarkers. *Cancer discovery*. 2013; 3:761–9.
16. Lui VW, Peyser ND, Ng PK, Hritz J, Zeng Y, Lu Y, Li H, Wang L, Gilbert BR, General IJ, Bahar I, Ju Z, Wang Z, et al. Frequent mutation of receptor protein tyrosine phosphatases provides a mechanism for STAT3 hyperactivation in head and neck cancer. *Proc Natl Acad Sci USA*. 2014; 111:1114–9.
17. Vanderkraats ND, Hiken JF, Decker KF, Edwards JR. Discovering high-resolution patterns of differential DNA methylation that correlate with gene expression changes. *Nucleic Acids Res*. 2013; 41:6816–27.
18. Irizarry RA, Ladd-Acosta C, Wen B, Wu Z, Montano C, Onyango P, Cui H, Gabo K, Rongione M, Webster M, Ji H, Potash JB, Sabunciyan S, et al. The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nature genetics*. 2009; 41:178–86.
19. Artemov A, Aliper A, Korzinkin M, Lezhnina K, Jellen L, Zhukov N, Roumiantsev S, Gaifullin N, Zhavoronkov A, Borisov N, Buzdin A. A method for predicting target drug efficiency in cancer based on the analysis of signaling pathway activation. *Oncotarget*. 2015; 6:29347–56. doi: 10.18632/oncotarget.5119.
20. Ochs MF, Farrar JE, Considine M, Wei Y, Meshinchi S, Arceci RJ. Outlier Analysis and Top Scoring Pair for Integrated Data Analysis and Biomarker Discovery. *IEEE/ACM transactions on computational biology and bioinformatics*. 2013.
21. Li R, Ochs MF, Ahn SM, Hennessey P, Tan M, Soudry E, Gaykalova DA, Uemura M, Brait M, Shao C, Westra W, Bishop J, Fertig EJ, et al. Expression microarray analysis reveals alternative splicing of LAMA3 and DST genes in head and neck squamous cell carcinoma. *PLoS One*. 2014; 9:e91263.
22. Sun W, Gaykalova DA, Ochs MF, Mambo E, Arnaoutakis D, Liu Y, Loyo M, Agrawal N, Howard J, Li R, Ahn S, Fertig E, Sidransky D, et al. Activation of the NOTCH pathway in head and neck cancer. *Cancer Res*. 2014; 74:1091–104.
23. Fertig EJ, Markovic A, Danilova LV, Gaykalova DA, Cope L, Chung CH, Ochs MF, Califano JA. Preferential activation of the hedgehog pathway by epigenetic modulations in HPV negative HNSCC identified with meta-pathway analysis. *PLoS One*. 2013; 8:e78127.
24. Rathi KS, Gaykalova DA, Hennessey P, Califano JA, Ochs MF. Correcting transcription factor gene sets for copy number and promoter methylation variations. *Drug development research*. 2014; 75:343–7.
25. Afsari B, Geman D, Fertig EJ. Learning dysregulated pathways in cancers from differential variability analysis. *Cancer Inform*. 2014; 13:61–7.
26. van Wieringen WN, van de Wiel MA, van der Vaart AW. A Test for Partial Differential Expression. *J Am Stat Assoc*. 2008; 103:1039–1049.
27. Ghosh D. Genomic outlier detection in high-throughput data analysis. *Methods Mol Biol*. 2013; 972:141–53.
28. MacDonald JW, Ghosh D. COPA—cancer outlier profile analysis. *Bioinformatics*. 2006; 22:2950–1.
29. Gaykalova DA, Vatapalli R, Wei Y, Tsai HL, Wang H, Zhang C, Hennessey PT, Guo T, Tan M, Li R, Ahn J, Khan Z, Westra WH, et al. Outlier Analysis Defines Zinc Finger Gene Family DNA Methylation in Tumors and Saliva of Head and Neck Cancer Patients. *PLoS One*. 2015; 10:e0142148.
30. Han M, Liew CT, Zhang HW, Chao S, Zheng R, Yip KT, Song ZY, Li HM, Geng XP, Zhu LX, Lin JJ, Marshall KW, Liew CC. Novel blood-based, five-gene biomarker set for the detection of colorectal cancer. *Clin Cancer Res*. 2008; 14:455–60.
31. Hidalgo-Curtis C, Apperley JF, Stark A, Jeng M, Gotlib J, Chase A, Cross NC, Grand FH. Fusion of PDGFRB to

- two distinct loci at 3p21 and a third at 12q13 in imatinib-responsive myeloproliferative neoplasms. *Br J Haematol*. 2010; 148:268–73.
32. Yan J, Nie K, Mathew S, Tam Y, Cheng S, Knowles DM, Orazi A, Tam W. Inactivation of BANK1 in a novel IGH-associated translocation t(4;14)(q24;q32) suggests a tumor suppressor role in B-cell lymphoma. *Blood cancer journal*, 2014. 4:e215.
  33. Zhang P, Yang Y, Nolo R, Zweidler-McKay PA, Hughes DP. Regulation of NOTCH signaling by reciprocal inhibition of HES1 and Deltex 1 and its role in osteosarcoma invasiveness. *Oncogene*. 2010; 29:2916–26.
  34. Ludwig L, Oswald F, Hoang-Vu C, Dralle H, Hildt E, Schmid RM, Karges W. Expression of the Grb2-related RET adapter protein Grap-2 in human medullary thyroid carcinoma. *Cancer Lett*. 2009; 275:194–7.
  35. Lind GE, Danielsen SA, Ahlquist T, Merok MA, Andresen K, Skotheim RI, Hektoen M, Rognum TO, Meling GI, Hoff G, Bretthauer M, Thiis-Evensen E, Nesbakken A, et al. Identification of an epigenetic biomarker panel with high sensitivity and specificity for colorectal cancer and adenomas. *Molecular cancer*, 2011. 10:85.
  36. Wang Y, Luo H, Li Y, Chen T, Wu S, Yang L. hsa-miR-96 up-regulates MAP4K1 and IRS1 and may function as a promising diagnostic marker in human bladder urothelial carcinomas. *Molecular medicine reports*. 2012; 5:260–5.
  37. Wang SS, Purdue MP, Cerhan JR, Zheng T, Menashe I, Armstrong BK, Lan Q, Hartge P, Krickler A, Zhang Y, Morton LM, Vajdic CM, Holford TR, et al. Common gene variants in the tumor necrosis factor (TNF) and TNF receptor superfamilies and NF- $\kappa$ B transcription factors and non-Hodgkin lymphoma risk. *PLoS One*. 2009; 4:e5360.
  38. Kaiser S, Park YK, Franklin JL, Halberg RB, Yu M, Jessen WJ, Freudenberg J, Chen X, Haigis K, Jegga AG, Kong S, Sakthivel B, Xu H, et al. Transcriptional recapitulation and subversion of embryonic colon development by mouse colon tumor models and human colon cancer. *Genome Biol*. 2007; 8:R131.
  39. Cancer Genome Atlas, N. Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature*. 2015; 517:576–82.
  40. Homann N, Nees M, Conradt C, Dietz A, Weidauer H, Maier H, Bosch FX. Overexpression of p53 in tumor-distant epithelia of head and neck cancer patients is associated with an increased incidence of second primary carcinoma. *Clin Cancer Res*. 2001; 7:290–6.
  41. Ha PK, Califano JA. The molecular biology of mucosal field cancerization of the head and neck. *Critical reviews in oral biology and medicine*. 2003; 14:363–9.
  42. Matsuno K, Eastman D, Mitsiades T, Quinn AM, Carcanciu ML, Ordentlich P, Kadesch T, Artavanis-Tsakonas S. Human deltex is a conserved regulator of Notch signalling. *Nature genetics*. 1998; 19:74–8.
  43. Yamamoto N, Yamamoto S, Inagaki F, Kawaichi M, Fukamizu A, Kishi N, Matsuno K, Nakamura K, Weinmaster G, Okano H, Nakafuku M. Role of Deltex-1 as a transcriptional regulator downstream of the Notch receptor. *J Biol Chem*. 2001; 276:45031–40.
  44. Fischer A, Klattig J, Kneitz B, Diez H, Maier M, Holtmann B, Englert C, Gessler M. Hey basic helix-loop-helix transcription factors are repressors of GATA4 and GATA6 and restrict expression of the GATA target gene ANF in fetal hearts. *Molecular and cellular biology*. 2005; 25:8960–70.
  45. Lee JC, Smith SB, Watada H, Lin J, Scheel D, Wang J, Mirmira RG, German MS. Regulation of the pancreatic pro-endocrine gene neurogenin3. *Diabetes*. 2001; 50:928–36.
  46. Huber RM, Rajski M, Sivasankaran B, Moncayo G, Hemmings BA, Merlo A. Deltex-1 activates mitotic signaling and proliferation and increases the clonogenic and invasive potential of U373 and LN18 glioblastoma cells and correlates with patient survival. *PLoS One*. 2013; 8:e57793.
  47. Lin YW, Deveney R, Barbara M, Iscove NN, Nimer SD, Slape C, Aplan PD. OLIG2 (BHLHB1), a BHLH transcription factor, contributes to leukemogenesis in concert with LMO1. *Cancer Res*. 2005; 65:7151–8.
  48. Lindqvist CM, Lundmark A, Nordlund J, Freyhult E, Ekman D, Carlsson Almlof J, Raine A, Overnas E, Abrahamsson J, Frost BM, Grander D, Heyman M, Palle J, et al. Deep targeted sequencing in pediatric acute lymphoblastic leukemia unveils distinct mutational patterns between genetic subtypes and novel relapse-associated genes. *Oncotarget*. 2016; 7:64071–64088. doi: 10.18632/oncotarget.11773.
  49. Xu Y, Wang Y, Liu H, Kang X, Li W, Wei Q. Genetic variants of genes in the Notch signaling pathway predict overall survival of non-small cell lung cancer patients in the PLCO study. *Oncotarget*. 2016; 7:61716–61727. doi: 10.18632/oncotarget.11436.
  50. Korosec B, Glavac D, Volavsek M, Ravnik-Glavac M. ATP2A3 gene is involved in cancer susceptibility. *Cancer Genet Cytogenet*. 2009; 188:88–94.
  51. Roos PH, Bolt HM. Cytochrome P450 interactions in human cancers: new aspects considering CYP1B1. *Expert opinion on drug metabolism & toxicology*. 2005; 1:187–202.
  52. McFadyen MC, Cruickshank ME, Miller ID, McLeod HL, Melvin WT, Haites NE, Parkin D, Murray GI. Cytochrome P450 CYP1B1 over-expression in primary and metastatic ovarian cancer. *Br J Cancer*. 2001; 85:242–6.
  53. Pourreyaon C, Reilly L, Proby C, Panteleyev A, Fleming C, McLean K, South AP, Foerster J. Wnt5a is strongly expressed at the leading edge in non-melanoma skin cancer, forming active gradients, while canonical Wnt signalling is repressed. *PLoS One*. 2012; 7:e31827.
  54. Kaucka M, Plevova K, Pavlova S, Janovska P, Mishra A, Verner J, Prochazkova J, Krejci P, Kotaskova J, Ovesna P, Tichy B, Brychtova Y, Doubek M, et al. The planar cell polarity pathway drives pathogenesis of chronic lymphocytic leukemia by the regulation of B-lymphocyte migration. *Cancer Res*. 2013; 73:1491–501.

55. Caldwell GM, Jones CE, Soon Y, Warrack R, Morton DG, Matthews GM. Reorganisation of Wnt-response pathways in colorectal tumorigenesis. *Br J Cancer*. 2008; 98:1437–42.
56. Jiang L, Zeng X, Yang H, Wang Z, Shen J, Bai J, Zhang Y, Gao F, Zhou M, Chen Q. Oral cancer overexpressed 1 (ORAOV1): a regulator for the cell growth and tumor angiogenesis in oral squamous cell carcinoma. *Int J Cancer*. 2008; 123:1779–86.
57. Mirghani H, Ugolin N, Ory C, Lefevre M, Baulande S, Hofman P, St Guily JL, Chevillard S, Lacave R. A predictive transcriptomic signature of oropharyngeal cancer according to HPV16 status exclusively. *Oral Oncol*. 2014.
58. Jackson PK, Eldridge AG, Freed E, Furstenthal L, Hsu JY, Kaiser BK, Reimann JD. The lore of the RINGs: substrate recognition and catalysis by ubiquitin ligases. *Trends in cell biology*. 2000; 10:429–39.
59. Yao K, Rizzo P, Rajan P, Albain K, Rychlik K, Shah S, Miele L. Notch-1 and notch-4 receptors as prognostic markers in breast cancer. *International journal of surgical pathology*. 2011; 19:607–13.
60. Ferretti E, Tosi E, Po A, Scipioni A, Morisi R, Espinola MS, Russo D, Durante C, Schlumberger M, Screpanti I, Filetti S, Gulino A. Notch signaling is involved in expression of thyrocyte differentiation markers and is down-regulated in thyroid tumors. *J Clin Endocrinol Metab*. 2008; 93:4080–7.
61. Izumchenko E, Sun K, Jones S, Brait M, Agrawal N, Koch W, McCord CL, Riley DR, Angiuoli SV, Velculescu VE, Jiang WW, Sidransky D. Notch1 mutations are drivers of oral tumorigenesis. *Cancer Prev Res (Phila)*. 2015. 8:277-86.
62. Pickering CR, Zhou JH, Lee JJ, Drummond JA, Peng SA, Saade RE, Tsai KY, Curry JL, Tetzlaff MT, Lai SY, Yu J, Muzny DM, Doddapaneni H, et al. Mutational landscape of aggressive cutaneous squamous cell carcinoma. *Clin Cancer Res*. 2014; 20:6582–92.
63. Shao C, Bai W, Junn JC, Uemura M, Hennessey PT, Zaboli D, Sidransky D, Califano JA, Ha PK. Evaluation of MYB promoter methylation in salivary adenoid cystic carcinoma. *Oral Oncol*. 2011; 47:251–5.
64. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods*. 2001; 25:402–8.
65. Li LC, Dahiya R. MethPrimer: designing primers for methylation PCRs. *Bioinformatics*. 2002; 18:1427–31.
66. Shao C, Tan M, Bishop JA, Liu J, Bai W, Gaykalova DA, Ogawa T, Vikani AR, Agrawal Y, Li RJ, Kim MS, Westra WH, Sidransky D, et al. Suprabasin is hypomethylated and associated with metastasis in salivary adenoid cystic carcinoma. *PLoS One*. 2012; 7:e48582.
67. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol*. 2004; 5:R80.
68. Carvalho BS, Irizarry RA. A framework for oligonucleotide microarray preprocessing. *Bioinformatics*. 2010; 26:2363–7.
69. Tibshirani, R. Hastie T. Outlier sums for differential gene expression analysis. *Biostatistics*. 2007; 8:2–8.
70. Ghosh, D. Discrete nonparametric algorithms for outlier detection with genomic data. *J Biopharm Stat*. 2010; 20:193–208.
71. Smyth GK. Limma: linear models for microarray data, in *Bioinformatics and Computational Biology Solutions using R and BioConductor*, R. Gentleman, R.V. Carey, S. Dudoit, R. Irizarry, and W. Huber, Editors. 2005, Springer: New York. 397–420.
72. Benjamini, Y. Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J R Statist Soc B*. 1995; 57:289-300.
73. Cancer Genome Atlas Research, N. Comprehensive genomic characterization of squamous cell lung cancers. *Nature*. 2012; 489:519–25.
74. Huang X, Takata K, Sato Y, Tanaka T, Ichimura K, Tamura M, Oka T, Yoshino T. Downregulation of the B-cell receptor signaling component CD79b in plasma cell myeloma: a possible post transcriptional regulation. *Pathology international*. 2011; 61:122–9.
75. Dornan D, Bennett F, Chen Y, Dennis M, Eaton D, Elkins K, French D, Go MA, Jack A, Junutula JR, Koeppen H, Lau J, McBride J, et al. Therapeutic potential of an anti-CD79b antibody-drug conjugate, anti-CD79b-vc-MMAE, for the treatment of non-Hodgkin lymphoma. *Blood*. 2009; 114:2721–9.
76. Arozarena I, Sanchez-Laorden B, Packer L, Hidalgo-Carcedo C, Hayward R, Viros A, Sahai E, Marais R. Oncogenic BRAF induces melanoma cell invasion by downregulating the cGMP-specific phosphodiesterase PDE5A. *Cancer Cell*. 2011; 19:45–57.
77. Fernandez-Zapico ME, Gonzalez-Paz NC, Weiss E, Savoy DN, Molina JR, Fonseca R, Smyrk TC, Chari ST, Urrutia R, Billadeau DD. Ectopic expression of VAV1 reveals an unexpected role in pancreatic cancer tumorigenesis. *Cancer Cell*. 2005; 7:39–49.
78. Hornstein I, Pikarsky E, Groysman M, Amir G, Peylan-Ramu N, Katzav S. The haematopoietic specific signal transducer Vav1 is expressed in a subset of human neuroblastomas. *J Pathol*. 2003; 199:526–33.



# A dataset of 26 candidate gene and pro-inflammatory cytokine variants for association studies in idiopathic pulmonary fibrosis: frequency distribution in normal Czech population

Amit Kishore<sup>1</sup>, Veronika Žižková<sup>1</sup>, Lenka Kocourková<sup>1</sup> and Martin Petřek<sup>1,2\*</sup>

<sup>1</sup> Department of Pathological Physiology, Laboratory of Immunogenomics, Faculty of Medicine and Dentistry, Palacký University, Olomouc, Czech Republic, <sup>2</sup> Institute of Molecular and Translational Medicine, Faculty of Medicine and Dentistry, Palacký University, Olomouc, Czech Republic

## OPEN ACCESS

### Edited by:

Masaaki Murakami,  
Hokkaido University, Japan

### Reviewed by:

Takayuki Yoshimoto,  
Tokyo Medical University, Japan  
Daisuke Kamimura,  
Hokkaido University, Japan

### \*Correspondence:

Martin Petřek,  
Department of Pathological  
Physiology, Faculty of Medicine and  
Dentistry, Palacký University,  
Hněvotínská Street 3, Olomouc  
77515, Czech Republic  
martin.petrek@fnol.cz

### Specialty section:

This article was submitted to  
Inflammation, a section of the journal  
Frontiers in Immunology

**Received:** 14 August 2015

**Accepted:** 02 September 2015

**Published:** 22 September 2015

### Citation:

Kishore A, Žižková V, Kocourková L  
and Petřek M (2015) A dataset of 26  
candidate gene and pro-inflammatory  
cytokine variants for association  
studies in idiopathic pulmonary  
fibrosis: frequency distribution in  
normal Czech population.  
*Front. Immunol.* 6:476.  
doi: 10.3389/fimmu.2015.00476

**Keywords:** MUC5B, cytokines, single nucleotide polymorphism, susceptibility, sequenom MassARRAY, idiopathic pulmonary fibrosis, Czech normal population, association study

## Introduction

Idiopathic pulmonary fibrosis (IPF) is a specific form of chronic, progressive fibrosing interstitial pneumonia with poor diagnosis and a median survival of 2–3 years from initial diagnosis (1, 2). The cellular inflammation drives the fibrotic response in lung and plays a major role in IPF pathogenesis (3). Inflammatory cells (majorly, type 2 alveolar epithelial cells) release TGF- $\beta$ , the key mediator of pulmonary fibrosis, that regulates several profibrotic cytokines/chemokines, their receptors, receptor subunits, and growth factors inducing process of epithelial–mesenchymal transition (EMT) (3, 4). Among the pro-inflammatory and profibrotic cytokines involved in IPF pathogenesis, interleukin (IL)-1 (4), IL-1 $\beta$  (5), IL-4, IL-5 (6), IL-6R $\alpha$  (4), angiogenic IL-8/CXCL-8 (7), IL-13, its receptor IL-13 R $\alpha$ 2 (8), and IL-33 (9) have been implicated in accelerated inflammation and irreversible damage to lung architecture with loss of alveolar-capillary barrier basal membrane leading to persistent fibrosis. Genes encoding these factors exhibit nucleotide variation that could affect the severity of immune/inflammatory reactions and extent of any subsequent dysregulated fibroproliferative activity in disease development. Furthermore, variants in mucin-encoding genes (10–13) and in genes for pathogen-associated molecular patterns (PAMPs) receptors of innate immunity known as toll-like receptors (TLRs) (14, 15) have been also implicated in IPF immunopathogenesis and related to rapid progression of the disease. Investigations of these “candidate” gene variant(s) e.g., in case-control association studies may, therefore, provide novel insight into underlying mechanism of IPF susceptibility/disease outcome and, further, may aid to develop novel diagnostic approaches and eventually therapeutic interventions based on genetic information (16).

A candidate gene study typically involves genotyping 5–50 single nucleotide polymorphisms (SNPs) within gene(s) for its coding and non-coding/regulatory regions (17). Irrespective of the number of tested gene variants; for a standard conductance, data collection, and transparent reporting of a genetic association study, the recommendations of STrengthening the REporting of Genetic Association studies (STREGA) and STrengthening the Reporting of OBServational studies in Epidemiology (STROBE) should be considered (16). In case-control studies, knowledge of frequency distribution of candidate gene loci/variants among normal (healthy control) population(s)

is necessary and could be useful also for genetically related population(s) to determine the gene variants associated with disease and/or its clinical course.

The role of inflammatory and profibrotic mechanisms involving gene variation has been investigated in IPF and spectrum of susceptible polymorphic gene variants, including those in genes of immune reactions and signaling processes, have been recently reported from both genome-wide association studies (GWAS) and population-based case-control studies (14, 18, 19) performed mostly in US Caucasians, and also in some other ethnicities. The nominated gene variants, summarized in **Table 1**, are of different functions and implicate some yet-unanticipated pathways in IPF pathogenesis, including endoplasmic reticulum stress and unfolded protein response, cellular senescence, DNA-damage response, and already known Wnt- $\beta$ -catenin signaling (20). The distribution of SNPs may greatly differ with populations (ethnicities), for example, a high frequency of *MUC5B* rs35705950\*T allele in IPF cases is observed among European-Americans (14–34%) (21, 22), while its low frequency is characteristic for Asians, such as Chinese (3.3%) (23), Japanese (3.4%) (24), and Korean (1.0%) cohorts (11). Similarly, *MUC2* rs7934606\*A allele exhibit frequency of 41% in Europeans and 1% in Asians (1000 Genomes Project Phase 3 allele frequencies). Furthermore, in context of

participation of more than one gene in IPF pathogenesis, it will be important to analyze multiple susceptible gene variants. The approach of analyzing common and rare genetic factors in IPF susceptibility may provide novel insights into IPF and it could also be helpful in identifying population-specific rare variants, predominant panel of candidate gene variants for IPF risk and in understanding the basis of variable disease severity or progression among different populations.

No complex data have been yet reported on IPF-related variants in Slavonic populations, including Czechs. Starting our investigations of plausible multiple IPF susceptibility polymorphisms primarily in Czech and also related populations, we adopted allele-specific MALDI-TOF mass spectrometry-based SNPs genotyping assay for determination of gene variation in the relevant targets. Several IPF susceptible SNPs in genes of various functional categories were multiplexed, and in the first phase genotyped in probands from normal (healthy) Czech population using Sequenom MassARRAY platform. In the current dataset manuscript, we, besides genotyping methodology, report the genotype, allele, and phenotype (carriage rate) frequencies for plausible IPF susceptibility variants among normal population of Czech Republic of Western Slavonic (Caucasian) ancestry.

**TABLE 1 | List of candidate SNPs investigated in the study.**

S.No.	Location	Gene	SNP ID	Position	Region	Functional category	Reference
1	2q14.1	<i>IL-1 <math>\alpha</math></i>	rs1800587	112,785,383	5'-flanking region	Pro-inflammatory cytokine	(25)
2	2q14.1	<i>IL-1 <math>\beta</math></i>	rs16944	112,837,290	Promoter	Pro-inflammatory cytokine	(25)
3	2q14.1	<i>IL-1 <math>\beta</math></i>	rs1143634	112,832,813	Exon (ds)	Pro-inflammatory cytokine	(25)
4	2p21	<i>PRKCE</i>	rs628877	45,698,441	Intron	Cellular signaling pathways	–
5	3q26	<i>LRRRC34</i>	rs6793295	169,800,667	Exon (dn)	Protein-protein interaction	(26)
6	3q22.1	<i>TF</i>	rs1799899	133,756,968	Exon (dn)	Iron ion transport	–
7	4q13.3	<i>IL-8</i>	rs4073	73,740,307	Promoter	Pro-inflammatory cytokine	(18)
8	4q22.1	<i>FAM13A</i>	rs2609255	88,890,044	Intron	GTPase activator activity	(26)
9	4q35	<i>TLR3</i>	rs3775291	186,082,920	Exon (dn)	Innate immunity	(14)
10	5p15	<i>TERT</i>	rs2736100	1,286,401	Intron 2	Telomerase maintenance	(26)
11	5q31.1	<i>IL-13</i>	rs1800925	132,657,117	Promoter	Pro-inflammatory cytokine	(27)
12	5q31.1	<i>IL-4</i>	rs2243248	132,672,952	Promoter	Pro-inflammatory cytokine	(19)
13	5q31.1	<i>IL-4</i>	rs2243250	132,673,462	Promoter	Pro-inflammatory cytokine	(19)
14	5q31.1	<i>IL-4</i>	rs2070874	132,674,018	Promoter	Pro-inflammatory cytokine	(19)
15	6p21.2	<i>CDKN1A</i>	rs2395655	36,677,919	Promoter	Cell cycle regulation	(19)
16	6p21.2	<i>CDKN1A</i>	rs733590	36,677,426	Promoter	Cell cycle regulation	(19)
17	6p24	<i>DSP</i>	rs2076295	7,562,999	Intron	Binds intermediate filaments	(26)
18	10q24	<i>OBFC1</i>	rs11191865	103,913,084	Intron	Telomerase maintenance	(26)
19	11p15.5	<i>MUC2</i>	rs7934606	1,100,037	Intron	Mucus production in lungs	(26)
20	11p15.5	<i>MUC5B</i>	rs35705950	1,219,991	Promoter	Mucus production in lungs	(12, 21)
21	11p15.5	<i>TOLLIP</i>	rs5743890	1,304,599	Intron	Innate immunity	(21)
22	13q34	<i>ATP11A</i>	rs1278769	112,882,313	3' UTR	Phospholipid translocation	(26)
23	16p12.1	<i>IL-4R <math>\alpha</math></i>	rs1801275	27,363,079	Exon (dn)	Pro-inflammatory cytokines	(19)
24	17p13.1	<i>TP53</i>	rs12951053	7,674,089	Intron	Cell cycle regulation	(28)
25	17p13.1	<i>TP53</i>	rs12602273	7,679,695	Intron	Cell cycle regulation	(28)
26	17q21	<i>MAPT</i>	rs1981997	45,979,401	Intron	Microtubule-associated tau	(26)
27	17q23.3	<i>ACE</i>	rs4277405	63,471,557	Promoter	Fibrotic pathway	(29)
28	17q23.3	<i>ACE</i>	rs4459609	63,471,587	Promoter	Fibrotic pathway	(29)
29	19p13	<i>DPP9</i>	rs12610495	4,717,660	Intron	Serine protease encoding	(26)

*ACE*, angiotensin-converting enzyme; *ATP11A*, ATPase, class VI, type 11A; *AZGP1*, alpha-2-glycoprotein 1, zinc-binding; *CDKN1A*, cyclin-dependent kinase inhibitor 1A, the gene encoding p21; *COMP*, cartilage oligomeric matrix protein; *DPP9*, dipeptidyl-peptidase 9; *DSP*, desmoplakin; *FAM13A*, family with sequence similarity 13, member A; *IL-13*, interleukin-13; *IL-8*, interleukin-18; *LRRRC34*, leucine rich repeat containing 34; *MAPT*, microtubule-associated protein tau; *MUC2*, mucin 2; *MUC5B*, mucin 5B; *OBFC1*, oligonucleotide/oligosaccharide-binding fold containing 1; *PKCE*, protein kinase C, epsilon); *SFTP*, surfactant protein; *SPPL2C*, signal peptide peptidase like 2C; *TERT*, telomerase reverse transcriptase; *TF*, transferrin; *TGF- $\beta$* , transforming growth factor- $\beta$ ); *TNF- $\alpha$* , tumor necrosis factor- $\alpha$ ); *TLR3*, Toll-like receptor; *TOLLIP*, toll interacting protein; *TP53*, tumor protein 53; *dn*, synonymous mutation; *dn*, non-synonymous mutation; *UTR*, untranslated region.

## Materials and Methods

### Characteristics of the Study Group

Ninety-six unrelated healthy subjects (45 males, 51 females), free of any disease as assessed by physician's enquiry about their personal and family history, were enrolled. The mean age  $\pm$  standard deviation was  $34.5 \pm 8.9$  years and ranged from 18–57 years. All were Caucasians, and as assessed by surname and tracking personal history of Czech (Western Slavonic) ancestry living in Moravian region of the Czech Republic. All probands were informed about the purpose of the study and provided informed consent; the study was realized with the approval of the institutional ethics committee. Genomic DNA was isolated from peripheral blood leukocytes by standard salting out method (30).

A list of reference single nucleotide polymorphism database ID (rsSNP) was prepared for IPF-associated gene variants identified from available literature (GWAS and case-control studies) (Table 1). SNP genotyping using Sequenom iPLEX and MALDI-TOF-based MassARRAY platform allows analysis of up to 40 SNPs in a single well reaction. Primers were designed using Assay Design Suite v2.0 (ADS v 2.0) available under online tools of Agena Bioscience (<https://www.mysequenom.com/Tools>).

### Assay Design, PCR Amplification, and Genotyping

For online genotyping assay design of multiplexed SNPs using ADS v2.0, an input list of SNP IDs (rsSNP of each target) was provided and following steps were followed: (1) the sequence for each rsSNP was retrieved from database and formatted accordingly, (2) the proximal SNPs for each rsSNP were identified from database, and (3) optimal primer areas were identified that result in a unique amplicons containing a target for the extension primer. To avoid extension primer rejection due to insufficient known bases, the proximal base was replaced with inosine. (4) PCR and extension primers were designed and checked for false priming, hairpin/dimer formation. The primer multiplexes with mass separation of analytes (alleles) were created. The PCR primer plex was prepared for PCR amplification, and extension primer mix was prepared for single base extension (SBE)-based iPLEX reaction. The SBE pool plex consists of multiplexed primers that anneal adjacent to polymorphic site for each reaction present together in the multiplexed assay pool. Thus, several individual DNA polymorphisms with their corresponding SNP sites could be analyzed in a single reaction. Due to the inverse relationship between peak intensity and extend primer mass, the extension primers in iPLEX assays were adjusted for concentration to ensure the possible equal intensity of extension primers.

For multiplexed PCR amplification and genotyping assay, protocol described in literature was followed (31). In brief, multiplexed PCR amplification was performed using 10 ng DNA template. A cleanup reaction for amplicons plex was performed with shrimp alkaline phosphatase (SAP) mix. These SAP cleaned amplicons plex were further subjected to SBE reaction with iPLEX mix containing extend primer plex. The iPLEX reaction products were treated with a cationic resin (SpectroCLEAN, Sequenom) to remove salts, such as  $\text{Na}^+$ ,  $\text{K}^+$ , and  $\text{Mg}^{2+}$  ions. The desalted iPLEX extend amplicons were dispensed on SpectroCHIPs using

MassARRAY® Nanodispenser RS 1000 station and analyzed on matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS)-based MassARRAY analyzer for target allele(s). As assay control of SNP genotyping, a duplicate, positive, and negative control samples were included in each assay plate. In each assay, the specific ddNTP incorporated at target site could be identified with peak representing increase in mass of extend primer. The assay peak spectrum and call cluster plot resulting from MALDI-TOF MS analysis were analyzed with MassARRAY Typer 4.0.20 software that traces primer masses to the assayed alleles.

### Statistical Analysis

Each SNP was tested for Hardy–Weinberg Equilibrium (HWE) by Pearson's goodness-of-fit, Chi-square ( $\chi^2$ ) test. SNPs within HWE ( $p > 0.05$ ) and sufficiently common (Minor Allele Frequency, MAF  $> 5\%$ ) in the general population were included. Phenotype frequency (carriage rate) was calculated as proportion of individuals carrying one or two copies of a particular allele on one or both (maternal or paternal) chromosomes.

## Results

Using online Assay Design Suite v2.0 for primer designing, two plexes were generated. Plexes I and II consisted of 22 and 7 SNPs, respectively. For a successful assay, inosine was used instead of proximal SNP in the extend primer sequence for rs2736100, rs2243250, and rs4277405 of plex I.

In assay control of multiplexed SNP genotyping, SNP rs2395655 in *CDKN1A* showed low assay-success rate ( $< 95\%$ ) and two SNPs *DSP* rs2076295 and *TOLLIP* rs5743890 were found as positive in no template control. These SNPs failed the quality control assessments and were removed from further analysis (Table 2). The genotyping assays success rates for all other analyzed SNPs were 98–100%. In our Czech healthy control population, all analyzed SNPs were in HWE, except for *IL-4R $\alpha$*  rs1801275 that exhibited minor deviation ( $p = 0.04$ ) reflecting a small anomaly, so the locus was not excluded from analysis (Table 2).

Among analyzed SNPs in cytokines, their receptor and subunits, *IL-4* rs2243248 exhibited highest genotype (TT = 0.85), allele frequency (T = 0.93) and carriage rate (T = 1.00). Besides cytokines, we also report allelic frequency of rs3775291 in *TLR3*, an innate immune system receptor. In present report, four out of 26 SNPs, viz., *TP53* rs12951053, *TP53* rs12602273, *TF* rs1799899, and *IL-4* rs2243248 showed complete absence of their respective homozygous genotype CC, GG, AA, and GG, and exhibited high phenotype frequency (1.00) for allele A, C, G, and T, respectively (Table 2). For *MUC5B* rs35705950\*T risk-allele, allelic and phenotype frequencies were found as 9% and 17%, respectively.

The genotype frequency and allele frequency for the 26 analyzed SNPs are available online at ALlele FREquency Database with Sample UID: SA004336Q (<http://alfred.med.yale.edu/alfred/pophetgraph.asp?sampleuid=SA004336Q&cutoff=0.25>) and will be publicly available at dbSNP database with the release of dbSNP Build (B144) ([http://www.ncbi.nlm.nih.gov/SNP/snp\\_viewTable.cgi?handle=LIGP](http://www.ncbi.nlm.nih.gov/SNP/snp_viewTable.cgi?handle=LIGP)).

**TABLE 2 | Genotype distribution and phenotype frequency of candidate SNPs in normal Czech population.**

S.No.	Genetic variant	Assay rate (%)	Genotype	n = 96	HWE	Allele	Phenotype frequency
1	<i>IL-1 α</i> rs1800587	100	CC	39	0.86	C	0.88
			CT	45		T	0.59
			TT	12			
2	<i>IL-1 β</i> rs16944	100	GG	46	0.35	G	0.88
			GA	38		A	0.52
			AA	12			
3	<i>IL-1 β</i> rs1143634	100	CC	56	0.49	C	0.93
			CT	33		T	0.42
			TT	7			
4	<i>PRKCE</i> rs628877	99 <sup>a</sup>	GG	61	0.11	G	0.93
			GT	27		T	0.36
			TT	7			
5	<i>LRRC34</i> rs6793295	100	TT	45	0.23	T	0.94
			TC	45		C	0.53
			CC	6			
6	<i>TF</i> rs1799899	100	GG	81	0.41	G	1.00
			GA	15		A	0.16
			AA	0			
7	<i>IL-8</i> rs4073	100	AA	19	0.50	A	0.73
			AT	51		T	0.80
			TT	26			
8	<i>FAM13A</i> rs2609255	100	TT	53	0.89	T	0.94
			GT	37		G	0.45
			GG	6			
9	<i>TLR3</i> rs3775291	100	GG	41	0.88	G	0.89
			GA	44		A	0.57
			AA	11			
10	<i>TERT</i> rs2736100	100	TT	29	0.91	T	0.80
			GT	48		G	0.70
			GG	19			
11	<i>IL-13</i> rs1800925	100	CC	49	0.68	C	0.91
			CT	38		T	0.49
			TT	9			
12	<i>IL-4</i> rs2243248	98 <sup>b</sup>	TT	80	0.44	T	1.00
			GT	14		G	0.15
			GG	0			
13	<i>IL-4</i> rs2243250	98 <sup>b</sup>	CC	64	0.24	C	0.95
			CT	25		T	0.32
			TT	5			
14	<i>IL-4</i> rs2070874	100	CC	66	0.91	C	0.97
			CT	27		T	0.31
			TT	3			
15	<i>CDKN1A</i> rs733590	100	TT	36	0.71	T	0.86
			CT	47		C	0.63
			CC	13			
16	<i>OBFC1</i> rs11191865	100	GG	33	0.70	G	0.81
			AG	45		A	0.66
			AA	18			
17	<i>MUC2</i> rs7934606	100	GG	41	0.16	G	0.93
			AG	48		A	0.57
			AA	7			
18	<i>MUC5B</i> rs35705950	100	GG	80	0.16	G	0.98
			GT	14		T	0.17
			TT	2			

(Continued)

TABLE 2 | Continued

S.No.	Genetic variant	Assay rate (%)	Genotype	n = 96	HWE	Allele	Phenotype frequency
19	<i>ATP11A</i> rs1278769	100	GG	48	0.79	G	0.91
			GA	39		A	0.50
			AA	9			
20	<i>IL-4R<math>\alpha</math></i> rs1801275	100	AA	68	0.04	A	0.94
			AG	22		G	0.29
			GG	6			
21	<i>TP53</i> rs12951053	100	AA	81	0.41	A	1.00
			CA	15		C	0.16
			CC	0			
22	<i>TP53</i> rs12602273	100	CC	81	0.41	C	1.00
			GC	15		G	0.16
			GG	0			
23	<i>MAPT</i> rs1981997	100	GG	67	0.81	G	0.97
			GA	26		A	0.30
			AA	3			
24	<i>ACE</i> rs4277405	100	TT	36	0.51	T	0.88
			CT	48		C	0.63
			CC	12			
25	<i>ACE</i> rs4459609	100	AA	37	0.83	A	0.86
			CA	46		C	0.61
			CC	13			
26	<i>DPP9</i> rs12610495	100	AA	54	0.69	A	0.93
			GA	35		G	0.44
			GG	7			

<sup>a</sup>One sample for *PRKCE* rs628877.

<sup>b</sup>Two samples each for *IL-4* rs2243250 and *IL-4* rs2243248 were not genotyped.

HWE, Hardy-Weinberg equilibrium.

## Discussion

The present dataset reports the genotype distribution, genotype, allele, and phenotype frequency of 26 gene variants involved in immune-related pathomechanisms of IPF in normal Czech population using Sequenom MassARRAY based genotyping platform. Besides the relevance to the delineation of immunogenetic component of IPF, the knowledge of frequency distribution of gene variants in normal populations is of considerable importance for their evaluation as genetic markers in susceptibility, manifestation, prognosis, and potentially treatment of diseases in different populations (32).

A SNP rs35705950 in the putative promoter of *MUC5B* has been shown to exhibit strong association with both familial interstitial pneumonia and IPF (33). The observed rs35705950\*T risk-allele frequency of 9% in normal Czech population was in concordance with other reports in normal Caucasians of European-American descents, as 9–11% in American (33), 10% in UK Caucasians (34), 11% in French (22), and 4.3% in Germans (24) populations among Europeans. Interestingly, the *MUC5B* promoter polymorphism is observed less frequently in normal Asian populations, such as 0.8% in Japanese (24), 0.7% in Chinese (23), and <1% in Koreans (11). Overall, mucin glycoprotein encoding *MUC5B* has role in normal lung function by regulating immune function, microbial population, airway infection, and mucociliary clearance in lungs (35, 36).

Among analyzed cytokines, *IL-4* has significant role in IPF pathogenesis by regulating fibroblast functions, such as chemotaxis, proliferation, collagen synthesis, myofibroblast differentiation, and  $T_h1/T_h2$  equilibrium (19). The angiogenic *IL-8* was shown as predictive for early stage of IPF (37) and as poor IPF survival (38). Additionally, *IL-13* and *IL-13* pathway markers (39) and the innate immune signaling receptor *TLR3* have been suggested as potential markers of rapidly progressive form of IPF. Several recent studies have suggested that defective TLRs are linked to dysregulated fibrogenesis and have key role in myofibroblast activation, increased profibrotic cytokines, collagen deposition, fibrosis, and tissue destruction and, thus, promoting the progression of disease during the later phase of IPF (14, 15, 40, 41).

Of the four variants that exhibited absence of homozygous genotypes in this data report: (1) the frequency of *TP53* rs12951053 CC genotype has been reported as 6% in Caucasian HC (28), 1.2% in European and Africans and relatively higher in Asian (11.9% in Han Chinese and 11.6% in Japanese) populations ([http://snp-nexus.org/temp/snpnexus\\_10220/results.html](http://snp-nexus.org/temp/snpnexus_10220/results.html)); (2) For *TP53* rs12602273, CC genotype frequency has been reported as 3% in Caucasian healthy controls (28); (3) For *TF* rs1799899, AA genotype frequency has been reported as 0.6% in European, and 0.0% among African, Han Chinese, and Japanese populations ([http://snp-nexus.org/temp/snpnexus\\_10168/results.html](http://snp-nexus.org/temp/snpnexus_10168/results.html)); and (4) For *IL-4* rs2243248, low GG genotype frequency (*IL-4*, -1098 G/T) has been reported in another independent study for



characterization of *IL-4* gene polymorphism in a relatively small cohort of IPF patients of same ethnicity (19).

The present findings are widely applicable in IPF genetics research in other related populations as well. In a current research initiatives in immunogenetics by HLA-NET network, a working group for population definitions and sampling strategies in population genetics analyses strongly recommend the usage of geographical and/or cultural criteria (with anthropological considerations) to describe human populations instead of *a priori* misclassifications of racial and ethnic groups (42). In this context, Central Europe populations have been demonstrated as similar and genetically homogeneous (32, 43, 44). Therefore, the present findings are relevant for IPF gene case-control studies not only in Czech but also in neighboring populations, namely Slovak and Polish, and also in Germans and Austrians, as we could recently exemplify in preliminary investigations of immune-related IPF susceptible variants in Czech and German population cohorts (10, 13).

## References

- Raghu G, Collard HR, Egan JJ, Martinez FJ, Behr J, Brown KK, et al. An official ATS/ERS/JRS/ALAT statement: idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management. *Am J Respir Crit Care Med* (2011) **183**:788–824. doi:10.1164/rccm.2009-040GL
- Spagnolo P, Sverzellati N, Rossi G, Cavazza A, Tzouvelekas A, Crestani B, et al. Idiopathic pulmonary fibrosis: an update. *Ann Med* (2015) **47**:15–27. doi:10.3109/07853890.2014.982165
- Agostini C, Gurrieri C. Chemokine/cytokine cocktail in idiopathic pulmonary fibrosis. *Proc Am Thorac Soc* (2006) **3**:357–63. doi:10.1513/pats.200601-010TK
- Zhang J, Tian XJ, Zhang H, Teng Y, Li R, Bai F, et al. TGF-beta-induced epithelial-to-mesenchymal transition proceeds through stepwise activation of multiple feedback loops. *Sci Signal* (2014) **7**:ra91. doi:10.1126/scisignal.2005304
- Strieter RM, Mehrad B. New mechanisms of pulmonary fibrosis. *Chest* (2009) **136**:1364–70. doi:10.1378/chest.09-0510
- Furuie H, Yamasaki H, Suga M, Ando M. Altered accessory cell function of alveolar macrophages: a possible mechanism for induction of Th2 secretory profile in idiopathic pulmonary fibrosis. *Eur Respir J* (1997) **10**:787–94.
- Antoniou KM, Tzouvelekas A, Alexandrakis MG, Sfridakis K, Tsiligianni I, Rachiotis G, et al. Different angiogenic activity in pulmonary sarcoidosis and idiopathic pulmonary fibrosis. *Chest* (2006) **130**:982–8. doi:10.1378/chest.130.4.982
- Barnes JC, Lumsden RV, Worrell J, Counihan IP, O'Beirne SL, Belperio JA, et al. CXCR3 is required for the IL-13 mediated upregulation of IL-13Ralpha2 in pulmonary fibroblasts. *Am J Respir Cell Mol Biol* (2014) **53**:217–25. doi:10.1165/rcmb.2013-0433OC
- Luzina IG, Pickering EM, Kopach P, Kang PH, Lockatell V, Todd NW, et al. Full-length IL-33 promotes inflammation but not Th2 response in vivo in an ST2-independent fashion. *J Immunol* (2012) **189**:403–10. doi:10.4049/jimmunol.1200259
- Kishore A, Zissel G, Zizkova V, Mueller-Quernheim J, Petrek M. Association of mucin (MUC2, MUC5B) gene variants with idiopathic pulmonary fibrosis (IPF) in a German population: a pilot study using MassARRAY technology. *Tissue Antigens* (2015) **85**:340. doi:10.1111/tan.12557
- Peljto AL, Selman M, Kim DS, Murphy E, Tucker L, Pardo A, et al. The MUC5B promoter polymorphism is associated with IPF in a Mexican cohort but is rare among Asian ancestries. *Chest* (2015) **147**:460–4. doi:10.1378/chest.14-0867
- Peljto AL, Zhang Y, Fingerlin TE, Ma SF, Garcia JG, Richards TJ, et al. Association between the MUC5B promoter polymorphism and survival in patients with idiopathic pulmonary fibrosis. *JAMA* (2013) **309**:2232–9. doi:10.1164/ajrccm-conference.2015.191.1\_MeetingAbstracts.A4382
- Petrek M, Kishore A, Zizkova V. Genetic association study for idiopathic pulmonary fibrosis in a Czech population: results from a pilot study using

## Conclusion

The present data on a spectrum of 26 gene variants including 10 variants of immune and inflammatory response (cytokines/chemokines and TLR) and their frequency distribution in normal Czech (Western Slavonic, Caucasian) population has wider application as standard control along with cases in association studies for IPF. It is also relevant in other fibrotic lung diseases among Czech and genetically related/neighboring population(s) and in the wider context for further delineation of the role of immune and inflammatory reactions in this debilitating disease.

## Acknowledgments

Grant support: CZ.1.07/2.3.00/30.0041, LO1304, and IGA PU LF\_2015\_020.

- MassARRAY technology. *Am J Respir Crit Care Med* (2015) **191**:A4382. doi:10.1164/ajrccm-conference.2015.191.1\_MeetingAbstracts.A4382
- O'Dwyer DN, Armstrong ME, Trujillo G, Cooke G, Keane MP, Fallon PG, et al. The toll-like receptor 3 L412F polymorphism and disease progression in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* (2013) **188**:1442–50. doi:10.1164/rccm.201304-0760OC
  - O'Dwyer DN, Armstrong ME, Kooball M, Donnelly SC. Targeting defective toll-like receptor-3 function and idiopathic pulmonary fibrosis. *Expert Opin Ther Targets* (2015) **19**:507–14. doi:10.1517/14728222.2014.988706
  - Little J, Higgins JP, Ioannidis JP, Moher D, Gagnon F, von Elm E, et al. Strengthening the reporting of genetic association studies (STREGA): an extension of the STROBE statement. *Eur J Epidemiol* (2009) **24**:37–55. doi:10.1007/s10654-008-9302-y
  - Balding DJ. A tutorial on statistical methods for population association studies. *Nat Rev Genet* (2006) **7**:781–91. doi:10.1038/nrg1916
  - Ahn MH, Park BL, Lee SH, Park SW, Park JS, Kim DJ, et al. A promoter SNP rs4073T>A in the common allele of the interleukin 8 gene is associated with the development of idiopathic pulmonary fibrosis via the IL-8 protein enhancing mode. *Respir Res* (2011) **12**:73. doi:10.1186/1465-9921-12-73
  - Vasakova M, Sterclova M, Matej R, Olejar T, Kolesar L, Skibova J, et al. IL-4 polymorphisms, HRCT score and lung tissue markers in idiopathic pulmonary fibrosis. *Hum Immunol* (2013) **74**:1346–51. doi:10.1016/j.humimm.2013.07.011
  - Kropski JA, Lawson WE, Young LR, Blackwell TS. Genetic studies provide clues on the pathogenesis of idiopathic pulmonary fibrosis. *Dis Model Mech* (2013) **6**:9–17. doi:10.1242/dmm.010736
  - Noth I, Zhang Y, Ma SF, Flores C, Barber M, Huang Y, et al. Genetic variants associated with idiopathic pulmonary fibrosis susceptibility and mortality: a genome-wide association study. *Lancet Respir Med* (2013) **1**:309–17. doi:10.1016/S2213-2600(13)70045-6
  - Borie R, Crestani B, Dieude P, Nunes H, Allanore Y, Kannengiesser C, et al. The MUC5B variant is associated with idiopathic pulmonary fibrosis but not with systemic sclerosis interstitial lung disease in the European Caucasian population. *PLoS One* (2013) **8**:e70621. doi:10.1371/journal.pone.0070621
  - Wang C, Zhuang Y, Guo W, Cao L, Zhang H, Xu L, et al. Mucin 5B promoter polymorphism is associated with susceptibility to interstitial lung diseases in Chinese males. *PLoS One* (2014) **9**:e104919. doi:10.1371/journal.pone.0104919
  - Horimasu Y, Ohshimo S, Bonella F, Tanaka S, Ishikawa N, Hattori N, et al. MUC5B promoter polymorphism in Japanese patients with idiopathic pulmonary fibrosis. *Respirology* (2015) **20**:439–44. doi:10.1111/resp.12466
  - Hutyrova B, Pantelidis P, Drabek J, Zurkova M, Kolek V, Lenhart K, et al. Interleukin-1 gene cluster polymorphisms in sarcoidosis and idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* (2002) **165**:148–51. doi:10.1164/ajrccm.165.2.2106004

26. Fingerlin TE, Murphy E, Zhang W, Peljto AL, Brown KK, Steele MP, et al. Genome-wide association study identifies multiple susceptibility loci for pulmonary fibrosis. *Nat Genet* (2013) **45**:613–20. doi:10.1038/ng.2609
27. Ding M, Sheng H, Shen W, Zhen J, Xi L, Zeng J, et al. Association of the SNP rs1800925(C/T) in the interleukin-13 gene promoter with pulmonary function in Chinese Han patients with idiopathic pulmonary fibrosis. *Cell Biochem Biophys* (2013) **67**:905–9. doi:10.1007/s12013-013-9580-1
28. Korthagen NM, van Moorsel CH, Barlo NP, Kazemier KM, Ruven HJ, Grutters JC. Association between variations in cell cycle genes and idiopathic pulmonary fibrosis. *PLoS One* (2012) **7**:e30442. doi:10.1371/journal.pone.0030442
29. Uh ST, Kim TH, Shim EY, Jang AS, Park SW, Park JS, et al. Angiotensin-converting enzyme (ACE) gene polymorphisms are associated with idiopathic pulmonary fibrosis. *Lung* (2013) **191**:345–51. doi:10.1007/s00408-013-9469-1
30. Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* (1988) **16**:1215. doi:10.1093/nar/16.3.1215
31. Bradic M, Costa J, Chelo IM. Genotyping with Sequenom. *Methods Mol Biol* (2011) **772**:193–210. doi:10.1007/978-1-61779-228-1\_11
32. Kubistova Z, Mrazek F, Tudos Z, Kriegova E, Ambruzova Z, Mytilineos J, et al. Distribution of 22 cytokine gene polymorphisms in the healthy Czech population. *Int J Immunogenet* (2006) **33**:261–7. doi:10.1111/j.1744-313X.2006.00609.x
33. Seibold MA, Wise AL, Speer MC, Steele MP, Brown KK, Loyd JE, et al. A common MUC5B promoter polymorphism and pulmonary fibrosis. *N Engl J Med* (2011) **364**:1503–12. doi:10.1056/NEJMoa1013660
34. Stock CJ, Sato H, Fonseca C, Banya WA, Molyneaux PL, Adamali H, et al. Mucin 5B promoter polymorphism is associated with idiopathic pulmonary fibrosis but not with development of lung fibrosis in systemic sclerosis or sarcoidosis. *Thorax* (2013) **68**:436–41. doi:10.1136/thoraxjnl-2012-201786
35. Roy MG, Livraghi-Butrico A, Fletcher AA, McElwee MM, Evans SE, Boerner RM, et al. Muc5b is required for airway defence. *Nature* (2014) **505**:412–6. doi:10.1038/nature12807
36. Fletcher AA, Evans CM. Regulation of microbial populations and immune functions by Muc5b. *Ann Am Thorac Soc* (2014) **11**:S80–S. doi:10.1513/AnnalsATS.201308-263MG
37. Stercova M, Vasakova M, Metlicka M, Pavlicek J, Kolesar L, Striz I. Angiostatic versus angiogenic chemokines in IPF and EAA. *Respir Med* (2009) **103**:1651–6. doi:10.1016/j.rmed.2009.05.012
38. Richards TJ, Kaminski N, Baribaud F, Flavin S, Brodmerkel C, Horowitz D, et al. Peripheral blood proteins predict mortality in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* (2012) **185**:67–76. doi:10.1164/rccm.2011101-0058OC
39. Murray LA, Zhang H, Oak SR, Coelho AL, Herath A, Flaherty KR, et al. Targeting interleukin-13 with tralokinumab attenuates lung fibrosis and epithelial damage in a humanized SCID idiopathic pulmonary fibrosis model. *Am J Respir Cell Mol Biol* (2014) **50**:985–94. doi:10.1165/rcmb.2013-0342OC
40. Hogaboam CM, Trujillo G, Martinez FJ. Aberrant innate immune sensing leads to the rapid progression of idiopathic pulmonary fibrosis. *Fibrogenesis Tissue Repair* (2012) **5**:S3. doi:10.1186/1755-1536-5-S1-S3
41. Trujillo G, Meneghin A, Flaherty KR, Sholl LM, Myers JL, Kazerooni EA, et al. TLR9 differentiates rapidly from slowly progressing forms of idiopathic pulmonary fibrosis. *Sci Transl Med* (2010) **2**:57ra82. doi:10.1126/scitranslmed.3001510
42. Sanchez-Mazas A, Vidan-Jeras B, Nunes JM, Fischer G, Little AM, Bekmane U, et al. Strategies to work with HLA data in human populations for histocompatibility, clinical transplantation, epidemiology and population genetics: HLA-NET methodological recommendations. *Int J Immunogenet* (2012) **39**:459–72. doi:10.1111/j.1744-313X.2012.01113.x
43. Lundmark PE, Liljedahl U, Boomsma DI, Mannila H, Martin NG, Palotie A, et al. Evaluation of HapMap data in six populations of European descent. *Eur J Hum Genet* (2008) **16**:1142–50. doi:10.1038/ejhg.2008.77
44. Heath SC, Gut IG, Brennan P, McKay JD, Bencko V, Fabianova E, et al. Investigation of the fine structure of European populations with applications to disease association studies. *Eur J Hum Genet* (2008) **16**:1413–29. doi:10.1038/ejhg.2008.210

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Kishore, Žížková, Kocourková and Petřek. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Association Study for 26 Candidate Loci in Idiopathic Pulmonary Fibrosis Patients from Four European Populations

Amit Kishore<sup>1</sup>, Veronika Žižková<sup>1</sup>, Lenka Kocourková<sup>1</sup>, Jana Petrková<sup>1</sup>, Evangelos Bouros<sup>2</sup>, Hilario Nunes<sup>3</sup>, Vladimíra Lošťáková<sup>4</sup>, Joachim Müller-Quernheim<sup>5</sup>, Gernot Zissel<sup>5</sup>, Vitezslav Kolek<sup>4</sup>, Demosthenes Bouros<sup>6</sup>, Dominique Valeyre<sup>3</sup> and Martin Petrek<sup>1,7\*</sup>

<sup>1</sup>Laboratory of Immunogenomics, Department of Pathological Physiology, Faculty of Medicine and Dentistry, Palacký University, Olomouc, Czech Republic, <sup>2</sup>Laboratory of Pharmacology, University Hospital Alexandroupolis, Democritus University of Thrace, Athens, Greece, <sup>3</sup>Université Paris 13, COMUE Sorbonne Paris Cité, Bobigny, Paris, France, <sup>4</sup>Department of Respiratory Medicine, Faculty of Medicine and Dentistry, Palacký University, Olomouc, Czech Republic, <sup>5</sup>Department of Pneumology, Center for Medicine, Medical Center, University of Freiburg, Freiburg, Germany, <sup>6</sup>Academic Department of Pneumology, Hospital for Diseases of the Chest 'Sotiria', Medical School, University of Athens, Athens, Greece, <sup>7</sup>Faculty of Medicine and Dentistry, Institute of Molecular and Translational Medicine, Palacký University and Faculty Hospital, Olomouc, Czech Republic

## OPEN ACCESS

### Edited by:

Pietro Ghezzi,  
Brighton and Sussex Medical School,  
UK

### Reviewed by:

Hideki Ogura,  
Yale University, USA  
John W. Holloway,  
University of Southampton, UK

### \*Correspondence:

Martin Petrek  
martin.petrek@fnol.cz

### Specialty section:

This article was submitted  
to Inflammation,  
a section of the journal  
Frontiers in Immunology

**Received:** 29 January 2016

**Accepted:** 29 June 2016

**Published:** 11 July 2016

### Citation:

Kishore A, Žižková V, Kocourková L,  
Petrková J, Bouros E, Nunes H,  
Lošťáková V, Müller-Quernheim J,  
Zissel G, Kolek V, Bouros D,  
Valeyre D and Petrek M (2016)  
Association Study for 26 Candidate  
Loci in Idiopathic Pulmonary Fibrosis  
Patients from Four European  
Populations.  
Front. Immunol. 7:274.  
doi: 10.3389/fimmu.2016.00274

Idiopathic pulmonary fibrosis (IPF) affects lung parenchyma with progressing fibrosis. In this study, we aimed to replicate *MUC5B* rs35705950 variants and determine new plausible candidate variants for IPF among four different European populations. We genotyped 26 IPF candidate loci in 165 IPF patients from four European countries, such as Czech Republic ( $n = 41$ ), Germany ( $n = 33$ ), Greece ( $n = 40$ ), France ( $n = 51$ ), and performed association study comparing observed variant distribution with that obtained in a genetically similar Czech healthy control population ( $n = 96$ ) described in our earlier data report. A highly significant association for a promoter variant (rs35705950) of mucin encoding *MUC5B* gene was observed in all IPF populations, individually and combined [odds ratio (95% confidence interval);  $p$ -value as 5.23 (8.94–3.06);  $1.80 \times 10^{-11}$ ]. Another non-coding variant, rs7934606 in *MUC2* was significant among German patients [2.85 (5.05–1.60);  $4.03 \times 10^{-4}$ ] and combined European IPF cases [2.18 (3.16–1.50);  $3.73 \times 10^{-5}$ ]. The network analysis for these variants indicated gene–gene and gene–phenotype interactions in IPF and lung biology. With replication of *MUC5B* rs35705950 previously reported in U.S. populations of European descent and indicating other plausible polymorphic variants relevant for IPF, we provide additional reference information for future extended functional and population studies aimed, ideally with inclusion of clinical parameters, at identification of IPF genetic markers.

**Keywords:** *MUC5B*, *MUC2*, cytokines, idiopathic pulmonary fibrosis, sequenom MassARRAY, single nucleotide polymorphism, association study, network analysis

## INTRODUCTION

Idiopathic pulmonary fibrosis (IPF) is a chronic, progressive form of fibrosing interstitial pneumonia of unknown cause that predominantly affects lung parenchyma, leading to progressive worsening of dyspnea and lung function (1). In pathobiological mechanisms of IPF, role for gene variation has been implicated and spectrum of susceptible/protective polymorphic gene variants, including those

in loci governing immune and inflammatory reactions and signaling processes, has been recently reported from genome-wide association studies (GWAS) or population-based case-control investigations (2–13); notably distribution of nominated gene variants varied among populations of different ancestry (Table S1 in Supplementary Material). Further, recent bioinformatics approaches yielded a genomic model that accurately predicted high- and low-risk IPF patients using a list of 118 IPF prognostic predictor genes, many of those with immune-, also T-cell-related functions (14). In aggregate, these reports implicated involvement of multiple genetic factors in IPF development and emphasized the need for their evaluation in different populations to decipher the plausible pathobiological mechanism of IPF.

In context of the above efforts, we have recently identified and reported 26 IPF-associated candidate loci (15). Besides characterizing their major functions, e.g., in regulating production of mucins (*MUC5B* and *MUC2*) or of pro-inflammatory cytokines (IL-1, IL-8) and also in cell signaling and innate immunity processes (*TLR3* and *TOLLIP*) involved in inflammatory and profibrotic pathways (<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4585032/table/T1/>), we have described the approach for their simultaneous investigation using a novel mass spectrometry based matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) multiplexed genotyping assay and reported their gene frequencies in healthy Czech (European) population (15). There, we also suggested the wider application of the data from this report for association studies among genetically homogenous populations following the recommendation of STrengthening the REporting of Genetic Association studies (STREGA) (16). Suggestion to include this particular control population in the present IPF association study in different European populations was based on reports of genetic similarity among Europeans, for example from results of genotyping 6000 individuals as control samples for >300,000 single nucleotide polymorphisms (SNPs) in a GWAS (17) and findings of a HapMap study (18). Further, in our decision we also reflected recommendations of HLA-NET network group for usage of geographical and/or cultural criteria to describe human populations (19) and also our own observations of substantial degree of homogeneity in distribution of immune-related gene variants within European populations, including Greeks (20).

Hereby, we report the results from a multicenter association gene study in which we determined the status (genotype distribution, genotype, and allele frequencies with carriage rate) of 26 IPF candidate loci. Here, we have performed a comparative study for association of these IPF candidate genetic variants among

four different European (Czech, German, Greek, and French) populations and have amended it by network prediction for gene-gene/gene-phenotype interactions in IPF and lung biology. We suggest that future extended and replicative studies following hereby described approach could enable better understanding of IPF pathogenesis, and if further supported by patient laboratory and clinical data, it could help to nominate novel disease markers.

## MATERIALS AND METHODS

### Characteristics of IPF Cases

In this study, 165 IPF patients from four European populations comprising 41 Czechs (Centre: University Hospital, Olomouc), 33 Germans (University Medical Center, Freiburg), 40 Greeks (Medical School University, Athens), and 51 French (University Hospital, Paris), representing Central, Southern, and Western Europe, were enrolled (Table 1).

All subjects were unrelated, white, and of European origin living in specified countries of Europe and speaking their respective national languages. The IPF cases were diagnosed as per ATS/ERS/JRS/ALAT guidelines (1, 21, 22) with typical clinical features and abnormalities on chest high-resolution computed tomography (HRCT) scans, abnormal lung function tests with reduced diffusing capacity of the lung for CO (DL<sub>CO</sub>), and/or restrictive pulmonary deficit, exclusion of other known causes of interstitial lung disease (ILD). For comparisons of genotype, allele frequency, and carriage rate (phenotype frequency) of analyzed genomic variants and case-control association study, we have utilized the data on distribution of these variants in 96 Czech healthy controls (15). Genomic DNA was isolated from peripheral blood leukocytes by standard salting out method (23). Informed consent was obtained from all study participants. The study was performed with approval of institutional ethical committees at respective centers (Ethics Committee of University Hospital and Medical Faculty of Palacky University, Olomouc, Czech Republic; Ethics Committee of the University Hospital Freiburg, Germany; Ethics Committee Hospital for Diseases of the Chest, Athens, Greece; and Comité Consultatif de protection des Personnes dans la recherche biomédicale-hôpital Robert Ballanger, France).

### Assay Design, PCR Amplification and Genotyping

The details of the panel comprising 26 IPF candidate loci and genotyping procedure in IPF cases have been described

**TABLE 1 | Subjects characteristics under study comprising IPF cases from four different populations.**

	Control subjects <sup>a</sup>	European IPF cases				Total IPF
		Czech	German	Greek	French	
N	96	41	33	40	51	165
Age, mean ± SD	34.45 ± 8.94	59.78 ± 10.12	66.25 ± 12.28	72.15 ± 7.71	72.19 ± 11.97	67.97 ± 11.60
Age, range	18–57	42–81	36–85	51–88	36–92	36–92
Males:females	45:51	23:18	26:7	32:8	44:7	125:40

<sup>a</sup>Healthy control subjects from our previous data report (15).

previously (15). In brief, a total of 26 SNPs reported as associated with IPF in literature were selected. These SNPs were located within the genes of different functional categories (mucus production, pro-inflammatory cytokines, chemokines, innate immune response, telomerase maintenance, cell surface remodeling, GTPase activator activity, cell-cycle regulators, phospholipid translocators, desmoplakin production, etc.). For PCR amplification and single base extension (SBE) reaction, the primer pairs along with extension primers were designed using Assay design suite v2.0. These primers were multiplexed and genotyped using Sequenom MassARRAY platform integrating iPLEX® SBE reaction and MassARRAY® technology (Agena Bioscience, San Diego, CA, USA) based MALDI-TOF MS assay. The assay consists of an initial locus-specific PCR amplification followed by SBE using mass-modified dideoxynucleotide terminators of an oligonucleotide primer that anneals immediately upstream of the target polymorphic site. The distinct mass of extended primer traces the alternative alleles using MassARRAY Typer 4.0.20. For quality control (QC) step, we determined data missing rate per individuals and missing rate per SNP. Also, for QC of SNP genotyping, positive and negative template control samples were included in each assay plate. Any assay found as positive in negative template control were removed from the study.

## Statistical Analysis

Each SNP was tested for Hardy–Weinberg equilibrium (HWE) by Pearson's Chi-square ( $\chi^2$ ) test or Fisher exact test, as applicable. SNPs within HWE ( $p > 0.05$ ) and sufficiently common [minor allele frequency (MAF)  $> 5\%$ ] in studied population were included. Carriage rate (phenotype frequency) was calculated as number of individuals carrying one (or two) copies of a particular allele on one or both (maternal and paternal) chromosomes. Association of SNPs minor alleles with IPF susceptibility were evaluated by Fisher's exact test providing odds ratio (OR), 95% confidence interval (CI), and level of significance ( $p$ ). For Bonferroni correction of multiple comparison (number of test = 100; 20 SNPs for four individual and the combined populations), a stringent approach with  $p$ -value  $< 0.05/100$  ( $5 \times 10^{-4}$ ) was considered as significant.

## Network Analysis

Prediction of gene–gene network for plausible candidate variants ( $p < 0.05$ ) and their interaction with IPF and other phenotypes, such as lung disease, lung injury, and lung function was performed using Phenolyzer, a tool for phenotype-based prioritization of candidate genes in human diseases (24). The candidate genes and their relationship with IPF and related phenotypes were investigated in several databases to determine and score relevant seed genes. The seed genes are then expanded to include related genes, on the basis of several types of gene–gene relationship components, such as exhibiting a protein–protein interaction, sharing a gene family or biological pathway, or transcriptionally regulating or being regulated by another gene. Finally, these different types of scores from seed gene ranking and gene–gene relationships are integrated to generate a ranked candidate gene list, together with gene–gene and gene–phenotypes interactions used to normalize the scores in range 0–1 (24).

## RESULTS

The characteristics of IPF patients and healthy control subjects included in this study are presented in **Table 1**. The proportions of IPF male cases were higher than of female cases. Following QC steps for missing rates, the genotyping data from all individual passed the QC with (i) missing frequency per individual 0.038 ( $N_{\text{MISS}} = 1$ ) to 0.077 ( $N_{\text{MISS}} = 2$ ) (a single IPF case from Greek population that failed the assay was not included in this study), and (ii) genotype missing frequency ( $F_{\text{MISS}}$ /assay error rate; **Table 2**) 0.004 to 0.015. Five SNPs showed departure from HWE (Table S2 in Supplementary Material), namely, *PRKCE* rs628877 ( $p = 0.02$  in combined IPF cases with  $F_{\text{MISS}} = 0.008$ ), *IL-4* rs2243250 ( $p = 0.01$ ,  $F_{\text{MISS}} = 0.0146$  in Czech IPF and  $p = 1.2 \times 10^{-4}$ ,  $F_{\text{MISS}} = 0.011$  in combined IPF cases), *IL-4* rs2070874 ( $p = 0.02$ ,  $F_{\text{MISS}} = 0$  in Czech IPF), *IL-4R $\alpha$*  rs1801275 ( $p = 0.04$ ,  $F_{\text{MISS}} = 0$  in Czech healthy controls;  $p = 0.03$ ,  $F_{\text{MISS}} = 0$  in Greek IPF and  $p = 0.01$ ,  $F_{\text{MISS}} = 0$  in combined IPF cases), and *MAPT* rs1981997 ( $p = 0.03$ ,  $F_{\text{MISS}} = 0.007$  in French IPF cases). Further, with MAF threshold check, *TP53* rs12951053 was found with MAF = 0.04,  $F_{\text{MISS}} = 0$  in Greek IPF cases. Thus, keeping the QC stringent conditions, these six SNPs were removed from further analysis, for which 20 variants remained (**Table 2**).

The primary analysis ( $p < 0.05$ ) using allelic (multiplicative) genetic model revealed a total of nine SNPs for IPF susceptibility. Among these, three SNPs were shared among different IPF populations: first, rs35705950\*T within promoter region of *Mucin5B* (*MUC5B*) was highly significant among all the IPF populations – Czech [OR (95% CI);  $p$ : 3.77 (7.47–1.9);  $1.62 \times 10^{-4}$ ]; German [4.83 (9.79–2.39);  $1.55 \times 10^{-5}$ ]; Greek [5.46 (10.82–2.76);  $1.13 \times 10^{-6}$ ]; French [6.77 (12.65–3.62);  $5.28 \times 10^{-10}$ ]; and combined IPF cases [5.23 (8.94–3.06);  $1.80 \times 10^{-11}$ ] (**Table 2**). Second, rs7934606\*A within intron region of *Mucin2* (*MUC2*) was significant for German [2.85 (5.05–1.60);  $4.03 \times 10^{-4}$ ], Greek [2.45 (4.19–1.43);  $1.43 \times 10^{-3}$ ], French [2.36 (3.86–1.44);  $7.03 \times 10^{-4}$ ], and combined IPF [2.18 (3.16–1.50);  $3.73 \times 10^{-5}$ ] cases. The third variant, rs1799899\*A located in exon region of *Transferrin* (*TF*) gene was significant for Germans [2.89 (6.47–1.30);  $1.12 \times 10^{-2}$ ], French [2.20 (4.65–1.04);  $4.58 \times 10^{-2}$ ], and in combined IPF [2.06 (3.78–1.12);  $1.87 \times 10^{-2}$ ] cases. Prediction of loss-/gain-of-function for functional variant *TF* rs1799899 suggested it as a probably damaging mutation (Polyphen score: 0.869 and SIFT score: 0.07).

The other six significant SNPs were featured in individual populations of which three were associated with IPF in Czech: (i) rs12602273\*G in *Tumor protein 53* (*TP53*) [2.43 (5.3–1.11);  $3.10 \times 10^{-2}$ ]; (ii) rs4277405\*C in *Angiotensin converting enzyme* (*ACE*) [0.54 (0.96–0.30);  $3.71 \times 10^{-2}$ ]; and (iii) rs4459609\*C in *ACE* [0.54 (0.96–0.30);  $3.71 \times 10^{-2}$ ] and one each in Greek: *Telomerase reverse transcriptase* (*TERT*) rs2736100\*G [0.49 (0.83–0.29);  $1.12 \times 10^{-2}$ ]; French: *ATPase, type 11A* (*ATP11A*) rs1278769\*A [0.51 (0.92–0.28);  $2.51 \times 10^{-2}$ ]; and in total IPF cases: *Interleukin-1  $\alpha$*  (*IL-1 $\alpha$* ) rs1800587\*T [0.68 (0.99–0.46);  $4.94 \times 10^{-2}$ ] (**Table 2**).

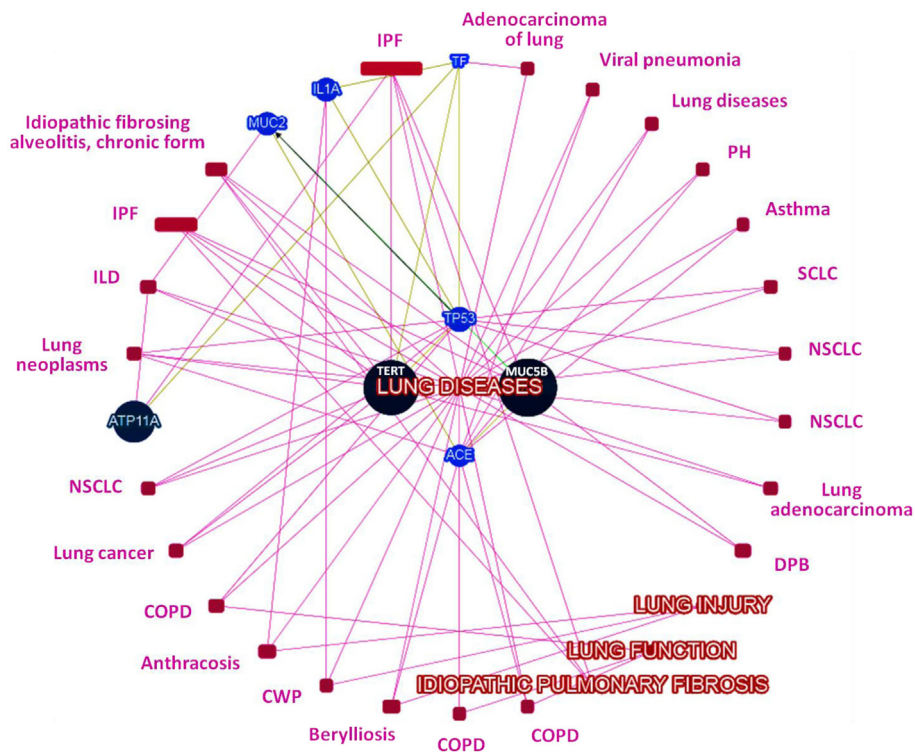
These findings of allelic model were in concordance with Pearson's  $\chi^2$  test and additive test of logistic regression analysis

TABLE 2 | Allelic model of association for IPF risk among four European populations.

S. No.	Gene	SNP	Minor allele	Czechs		Germans		Greeks		French		Total		Assay error rate
				OR (95% CI)	p	OR (95% CI)	p	OR (95% CI)	p	OR (95% CI)	p	OR (95% CI)	p	
1	<i>IL-1α</i>	rs1800587	T	0.65 (1.16–0.37)	0.16	0.78 (1.42–0.42)	0.45	0.56 (1.01–0.31)	0.06	0.74 (1.25–0.44)	0.30	0.68 (0.99–0.46)	$4.94 \times 10^{-2}$	0
2	<i>L-1β</i>	rs16944	A	1.15 (1.98–0.67)	0.67	0.79 (1.46–0.42)	0.54	1.01 (1.76–0.58)	1.00	1.14 (1.9–0.69)	0.61	1.03 (1.51–0.71)	0.92	0
3	<i>IL-1β</i>	rs1143634	T	1.06 (1.93–0.59)	0.88	1.34 (2.49–0.72)	0.42	0.71 (1.37–0.37)	0.34	1.06 (1.84–0.61)	0.89	1.02 (1.54–0.68)	1.00	0
4	<i>LRRC34</i>	rs6793295	C	0.87 (1.55–0.49)	0.66	0.96 (1.77–0.52)	1.00	0.79 (1.43–0.44)	0.46	1.35 (2.24–0.81)	0.29	1.00 (1.48–0.68)	1.00	0
5	<i>TF</i>	rs1799899	A	1.46 (3.47–0.61)	0.48	2.89 (6.47–1.3)	$1.12 \times 10^{-2}$	1.88 (4.3–0.82)	0.17	2.20 (4.65–1.04)	$4.58 \times 10^{-2}$	2.06 (3.78–1.12)	$1.87 \times 10^{-2}$	0
6	<i>IL-8</i>	rs4073	A	1.28 (2.14–0.76)	0.36	1.16 (2.03–0.66)	0.67	1.16 (1.95–0.69)	0.60	1.11 (1.8–0.69)	0.71	1.17 (1.67–0.82)	0.41	0
7	<i>FAM13A</i>	rs2609255	G	1.01 (1.82–0.56)	1.00	0.79 (1.54–0.4)	0.51	1.04 (1.88–0.57)	0.88	1.46 (2.47–0.86)	0.17	1.09 (1.64–0.73)	0.68	0
8	<i>TLR3</i>	rs3775291	A	0.94 (1.62–0.54)	0.89	0.61 (1.16–0.32)	0.17	0.72 (1.29–0.41)	0.32	0.72 (1.22–0.43)	0.24	0.75 (1.1–0.51)	0.14	0
9	<i>TERT</i>	rs2736100	G	1.01 (1.70–0.60)	1.00	0.85 (1.51–0.48)	0.67	0.49 (0.83–0.29)	$1.12 \times 10^{-2}$	1.23 (1.99–0.76)	0.46	1.23 (1.76–0.86)	0.28	0
10	<i>IL-13</i>	rs1800925	T	0.66 (1.22–0.35)	0.23	0.65 (1.27–0.34)	0.26	1.08 (1.91–0.61)	0.88	0.83 (1.43–0.48)	0.59	0.80 (1.2–0.54)	0.30	0.008
11	<i>IL-4</i>	rs2243248	G	0.47 (1.69–0.13)	0.29	0.59 (2.13–0.16)	0.57	1.78 (4.19–0.75)	0.24	1.20 (2.88–0.5)	0.66	1.02 (2.01–0.52)	1.00	0.008
12	<i>CDNK1A</i>	rs733590	C	1.21 (2.05–0.72)	0.50	0.76 (1.38–0.42)	0.46	1.15 (1.95–0.67)	0.68	1.01 (1.65–0.62)	1.00	1.03 (1.49–0.72)	0.93	0
13	<i>OBFC1</i>	rs11191865	A	1.67 (2.81–0.99)	0.06	1.75 (3.07–0.99)	0.06	1.07 (1.8–0.63)	0.89	1.43 (2.31–0.88)	0.18	1.44 (2.06–1.01)	0.06	0
14	<i>MUC2</i>	rs7934606	A	1.41 (2.41–0.83)	0.21	<b>2.85 (5.05–1.60)</b>	$4.03 \times 10^{-4}$	2.45 (4.19–1.43)	$1.43 \times 10^{-3}$	2.36 (3.86–1.44)	$7.03 \times 10^{-4}$	<b>2.18 (3.16–1.50)</b>	$3.73 \times 10^{-5}$	0.004
15	<i>MUC5B</i>	rs35705950	T	<b>3.77 (7.47–1.9)</b>	$1.62 \times 10^{-4}$	<b>4.83 (9.79–2.39)</b>	$1.55 \times 10^{-5}$	<b>5.46 (10.82–2.76)</b>	$1.13 \times 10^{-6}$	<b>6.77 (12.65–3.62)</b>	$5.28 \times 10^{-10}$	<b>5.23 (8.94–3.06)</b>	$1.80 \times 10^{-11}$	0.015
16	<i>ATP11A</i>	rs1278769	A	1.10 (1.92–0.63)	0.77	0.53 (1.06–0.26)	0.08	1.02 (1.79–0.57)	1.00	0.51 (0.92–0.28)	$2.51 \times 10^{-2}$	0.76 (1.13–0.51)	0.18	0
17	<i>TP53</i>	rs12602273	G	2.43 (5.3–1.11)	$3.10 \times 10^{-2}$	1.40 (3.6–0.54)	0.45	1.13 (2.89–0.44)	0.81	1.57 (3.5–0.71)	0.29	1.63 (3.03–0.87)	0.14	0
18	<i>ACE</i>	rs4277405	C	0.54 (0.96–0.30)	$3.71 \times 10^{-2}$	1.08 (1.92–0.61)	0.88	1.11 (1.9–0.65)	0.78	1.12 (1.83–0.68)	0.71	0.94 (1.36–0.65)	0.78	0
19	<i>ACE</i>	rs4459609	C	0.54 (0.96–0.30)	$3.71 \times 10^{-2}$	1.08 (1.92–0.61)	0.88	1.11 (1.9–0.65)	0.78	1.08 (1.76–0.66)	0.80	0.93 (1.34–0.64)	0.71	0
20	<i>DPP9</i>	rs12610495	G	1.01 (1.82–0.56)	1.00	1.56 (2.85–0.86)	0.15	1.01 (1.84–0.55)	1.00	1.14 (1.96–0.66)	0.68	1.15 (1.72–0.77)	0.54	0.008

Bold values are significant with Bonferroni corrected  $p < 5 \times 10^{-4}$ .

OR: Odds ratio; CI: Confidence interval.



**FIGURE 1 |** Network of genes with candidate loci and their interaction with IPF, lung injury, lung function, and lung disease phenotypes (ILD, interstitial lung disease; CWP, coal worker's pneumoconiosis; COPD, chronic obstructive pulmonary disease; DPB, diffuse panbronchiolitis; NSCLC, non-small cell lung cancer; SCLC, small cell lung cancer; PH, pulmonary hypertension).

with similar panel and population distribution/stratification of significant risk variants, and an exceptional addition of rs11191865\*A ( $p = 0.046$ ) in *Oligonucleotide/oligosaccharide-binding fold containing 1 (OBFC1)* exclusively in total IPF set under  $\chi^2$  test.

Network analysis among genes with significant variants showed *MUC5B* and *MUC2* in same gene family (Figure S1 in Supplementary Material), transcriptional interaction among *MUC2* and *TP53* (Figure S2 in Supplementary Material), and all identified candidate genes belonged to the same biosystem (Figure S3 in Supplementary Material). Further prediction based on their records in several databases (CLINVAR, ORPHANET, OMIM, DISGENET, GWAS, HTRI, GENE\_FAMILY, and BIOSYSTEM), gene-gene interaction and interaction with phenotypes (IPF, lung disease, lung injury, and lung function) has (Figure 1) prioritized *MUC5B* as highest-ranked genes and indicated *MUC5B*, *TERT*, and *ATP11a* as seed genes, while *TP53*, *IL1A*, *MUC2*, *ACE*, and *TF* as predictive genes based on their normalized scores (Figure 2).

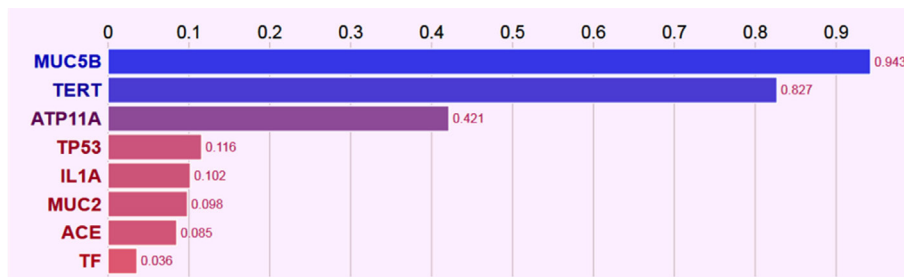
Applying rather conservative Bonferroni correction for multiple comparisons (100 tests, see Materials and Methods), *MUC5B* rs35705950\*T remained as highly significant in all IPF populations, Czech ( $p = 1.62 \times 10^{-4}$ ), German ( $1.55 \times 10^{-5}$ ), Greek ( $1.13 \times 10^{-6}$ ), French ( $5.28 \times 10^{-10}$ ), and in combined IPF population ( $1.80 \times 10^{-11}$ ); whereas *MUC2* rs7934606\*A was significant in German ( $4.03 \times 10^{-4}$ ) and in combined IPF ( $3.73 \times 10^{-5}$ ) (Table 2). The inheritance hypothesis for Bonferroni

significant variants, *MUC5B* rs35705950 and *MUC2* rs7934606, under allelic model for IPF association were also tested for other models and the variant rs35705950 was found in concordance to the dominant model indicating allele T for increased risk (Table 3).

For the highly significant variants, we also report the allele frequency of rs35705950\*T risk allele to range from 0.09 to 0.41 and rs7934606\*A from 0.32 to 0.58 among the analyzed four European populations (Table S2 in Supplementary Material). Among IPF cases, the allele frequency and carriage rate for rs35705950\*T were lowest in Czech (0.28 and 0.49, respectively) and highest in French (0.41 and 0.65, respectively); while rs7934606\*A has lowest frequencies in Czech (0.40 and 0.63, respectively) and highest in Germans (0.58 and 0.91, respectively).

## DISCUSSION

This study is a first report of comparative distribution of genotypes and alleles of 26 candidate gene variants implicated in mucin production, cell-cycle regulation, pro-inflammatory, and profibrotic signaling pathways pertinent to IPF pathobiology among patients from four populations across Europe (Czech, German, Greek, and French). Employing stringent statistical approach ( $p < 5 \times 10^{-4}$ ), we report a high association between *MUC5B* rs35705950\*T and IPF susceptibility in all the four analyzed populations among Czechs: [OR (95% CI)  $p$ ] 3.77 (7.47–1.9)



**FIGURE 2 |** Bar plot of seed genes (*MUC5B*, *TERT*, and *ATP11a*) and predicted genes (*TP53*, *IL1A*, *MUC2*, *ACE*, and *TF*) normalized scores based on the gene-gene and gene-phenotype interactions.

**TABLE 3 |** Dominant model of inheritance for highly significant IPF susceptible risk variants with Bonferroni correction among the four European populations.

Genetic variants	Czech	German	Greek	French	Total IPF
	OR (95% CI); <i>p</i>	OR (95% CI); <i>p</i>	OR (95% CI); <i>p</i>	OR (95% CI); <i>p</i>	OR (95% CI); <i>p</i>
<i>MUC5B</i> rs35705950*T (GT + TT vs. GG)	<b>4.76 (2.11–10.75);</b> <b>2.21 × 10<sup>-4</sup></b>	<b>8.75 (3.59–21.30);</b> <b>9.04 × 10<sup>-7</sup></b>	<b>6.25 (2.67–14.60);</b> <b>1.97 × 10<sup>-5</sup></b>	<b>9.167 (4.18–20.12);</b> <b>9.6 × 10<sup>-9</sup></b>	<b>7.01 (3.77–13.06);</b> <b>6.21 × 10<sup>-11</sup></b>
<i>MUC2</i> rs7934606*A (GA + AA vs. GG)	1.29 (0.61–2.74); NS	<b>7.45 (2.13–26.12);</b> <b>2.73 × 10<sup>-4</sup></b>	2.48 (1.06–5.80); 4.83 × 10 <sup>-2</sup>	2.71 (1.24–5.92); 1.16 × 10 <sup>-2</sup>	2.47 (1.43–4.26); 1.27 × 10 <sup>-3</sup>

Bold values are significant with Bonferroni corrected  $p < 5 \times 10^{-4}$ .

OR, Odds ratio; CI, Confidence interval; NS, Non-significant.

$1.62 \times 10^{-4}$ , Germans:  $4.83 (9.79-2.39) 1.55 \times 10^{-5}$ , Greeks:  $5.46 (10.82-2.76) 1.13 \times 10^{-6}$ , and French:  $6.77 (12.65-3.62) 5.28 \times 10^{-10}$  along with all patients from this study as a whole:  $5.23 (8.94-3.06) 1.80 \times 10^{-11}$ . Thus, we replicate previous findings and provide an insight for predominant association of gel-forming mucin-encoding gene variants with IPF in Europeans and populations of European descent. The role of mucin variants was also supported with high significance of *MUC2* rs7934606 among Germans [2.85 (5.05–1.60);  $4.03 \times 10^{-4}$ ] and combined European IPF cases [2.18 (3.16–1.50);  $3.73 \times 10^{-5}$ ]. The identified nine variants among *MUC5B*, *MUC2*, *TF*, *TP53*, *ACE*, *TERT*, *ATP11A*, and *IL-1 $\alpha$*  (significant at  $p < 0.05$ ) were interactive at gene-gene and gene-phenotype level in network analysis. The present study for inter-population comparison is imperative with respect of comparing the distributions of plausible IPF-associated gene variants and, thus, may provide starting point(s) for further investigations of biological (functional implication of nominated loci) and translational (relationship with laboratory/clinical parameters) aspects.

Among mucin gene cluster (*MUC6*, *MUC2*, *MUC5AC*, and *MUC5B*) on chromosome 11q15.5, *MUC5B* is the most predominant in the normal distal airway epithelium and is widely associated with sporadic IPF and familial interstitial pneumonia (9, 25–27). The *MUC5B* promoter-variant rs35705950 SNP predicted change in transcription factors binding sites (disruption of E2F and creation of HOX9 and PAX2) and the risk T-allele has been strongly associated with increased *MUC5B* expression in the lung tissue. The relative high frequency of rs35705950\*T risk allele (IPF cases: 0.28–0.41 and healthy controls: 0.09) observed

in this study are in concordance with several reports, including populations of European descent (Table S1 in Supplementary Material). To explain the mucin-expressing structures in IPF pathogenesis, Seibold et al. analyzed ciliated, basal, and alveolar type II cells in lung tissue and reported predominant expression of *MUC5B* in pseudostratified mucociliary epithelium comprised of basal epithelial cells and mucus cells in distal airway. Further, it is suggested that mucociliary dysfunction in the distal airway causing honeycomb cyst may play a role in the development of progressive fibroproliferative lung disease (25, 27). Although rs35705950 is reported as highly associated with IPF in Europeans and in populations of European descent, interestingly, it is weakly associated in East-Asians, such as Chinese, Japanese, and Koreans (Table S1 in Supplementary Material). Another mutation in the same gene family (Figure S1 in Supplementary Material) identified *MUC2* rs79834606 as significantly associated in IPF ( $p = 3.8 \times 10^{-6}$ ) with MAF (A-allele) 0.54 in IPF cases and 0.41 in controls (9), which is in accordance to our present findings (IPF: 0.40–0.58 and controls: 0.32) (Table S2 in Supplementary Material). Similarly, MAF of 0.52 for rs79834606\*A was reported among IPF cases in a GWAS (3).

Here, we newly identified rs1799899 located in the *TF* gene as IPF risk variant in Germans, French, and in combined European IPF cases, significant at primary analysis ( $0.05 > p > 5 \times 10^{-4}$ ). The SNP rs1799899 marks a Gly/Ser change located in *TF* gene encoding a glycoprotein involved in iron ion transport and removal of certain organic matter and allergens from serum. As this is the first implication of *TF* gene in context of IPF, in parallel to the studies of its possible functional role, this association requires



replication (16). The other six IPF risk variants identified in this study were featured among individual populations (Table 2). Among Czech IPF cases, (i) SNP rs12602273 located in intron of *TP53* that regulates cell-cycle arrest; while, (ii) rs4277405 and (iii) rs4459609 located in promoter region of *ACE-II* suggested with critical profibrotic role in IPF (28, 29) were observed. The MAF of rs12602273\*G reported as 0.07 in IPF case and 0.08 in healthy population of the Netherlands (11) was comparable to our current findings (IPF: 0.09–0.12 and controls: 0.08). For Greek IPF cases, (iv) a common variant rs2736100 within intron of *TERT* and (v) rs1278769 in 3'-UTR of *ATP11A* were significant. Several studies have reported the telomerase gene mutation causing short telomerase as risk factor and poor survival in IPF (2, 30, 31). These variants, rs2736100\*G and rs1278769\*A, were initially reported in a GWAS study for IPF cases with MAF of 0.43 and 0.20, respectively (MAF = 0.41–0.63 and 0.18–0.30, respectively in this study) (3). Besides, (vi) rs1800587 in 5'-flanking region of pro-inflammatory cytokine *IL-1 $\alpha$*  was lesser significant among the combined IPF cases. Earlier, a study in Czech population has reported rs1800587\*T MAF of 0.30 in IPF cases and 0.32 in healthy controls (4), similar to our current finding (IPF: 0.27–0.29 and controls: 0.36).

While *MUC5B* rs35705950 is highly replicable, the other probable candidate variants (including new report of *TF* rs1799899 in IPF) reported in this study must be replicated in accordance to the guidelines for conductance of genetic association studies (16), by other independent studies. However, already at this stage, our analysis of the gene–gene and gene–phenotype interaction networks suggests that these candidate variant genes are pertinent to IPF and lung function biology (Figure 1) and provides support for biological plausibility of observed variants.

Apart from using Czech population control data for comparisons within the European context, which was noted and reasoned for in the Section “Introduction,” another limitation of this study is that a single national center was included for each of four European populations, where a relatively small number of IPF cases were genotyped, which reduces the power of our present findings. However, with a conservative incidence range of 3–9 cases per 100,000 per year for Europe (32), the enrolled sample size (165 IPF cases) in the present study considerably represents the disease among Europeans; moreover, it is comparable with sizes of other reported studies (Table S1 in Supplementary Material). Off note, our primary aim was to report our findings of the candidate IPF genetic variants so that these could be replicated in other centers (16) and investigated further. In this regard, investigations of relationships with clinical parameters such as lung functions will follow.

Our findings provide evidence that gene variants involved in mucin production (*MUC5B* and *MUC2*) do increase IPF risk among the four European populations, two of which (Czech and Greek) have not been studied before in this regard at all. Additionally, the nominated variants in *TF* and other variants of *TP53*, *ACE*, *TERT*, *ATP11A*, and *IL-1 $\alpha$*  may also contribute to IPF susceptibility. Despite our panel of 26 gene variants was designed across pertinent pathobiological pathways, it did not include others from the wide range of plausible IPF-associated

SNPs regulating immune and fibrotic functions, such as master regulator TGF- $\beta$ , TNF- $\alpha$ , full spectrum of TLRs, MHC (HLA) variants, and also SNPs in regulatory microRNAs (miRNAs) (4, 8, 33, 34). Therefore, these variants should be prioritized in the future studies aimed at extending a profile general and population-specific IPF gene biomarkers, including exploration of their functions, so that our view on the role of gene variation in origin of IPF and its further development more closely approaches the reality.

## CONCLUSION

The present study confirms and further extends strong association of *MUC5B* promoter region variant (rs35705950) with IPF disease among Europeans. In addition, it suggests further IPF-associated polymorphisms: *MUC2* (rs79834606) and *TF* (rs1799899) variants in general, and the other six (*TP53* rs12602273, *ACE-II* rs4277405, *ACE-II* rs4459609, *TERT* rs2736100, *ATP11A* rs1278769, and *IL-1 $\alpha$*  rs1800587) in individual European populations. Their further investigation for disease association among extended patient cohorts is, therefore, warranted. At the same time, findings of our present study represent reference information to be utilized for future extended functional and population studies as well as for translational research of the nominated variants aiming at characterization of biomarkers and/or novel therapeutic targets.

## ETHICS STATEMENT

The study was approved by the ethical committee of the University Hospital, Olomouc; University Medical Center, Freiburg; Medical School University, Athens; and University Hospital, Paris. After approval of the study by ethical committee, the consent of human participants visiting the centres was taken in written for usage of biological sample for research purpose in future.

## AUTHOR CONTRIBUTIONS

MP, GZ, JM-Q, JP, VK, DB, and DV conceived and designed the work; AK, VZ, LK, JP, EB, HN, VL, GZ, VK, and MP contributed to data acquisition; AK performed data analysis and interpretation; AK and MP drafted the manuscript after its revision for important intellectual context by all authors; MP and AK finalized the article. All authors have read and approved the final manuscript and agreed to be accountable for all aspects of the work.

## FUNDING

Grant support: CZ.1.07/2.3.00/30.0041, LO1304, IGA\_PU\_LF\_2015\_020, 2016\_009, National Strategic Reference Framework (Greece) 09SYN-12-680, Hellenic scientific society for rare diseases and orphan drugs.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at <http://journal.frontiersin.org/article/10.3389/fimmu.2016.00274>


## REFERENCES

- Raghu G, Collard HR, Egan JJ, Martinez FJ, Behr J, Brown KK, et al. An official ATS/ERS/JRS/ALAT statement: idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management. *Am J Respir Crit Care Med* (2011) 183:788–824. doi:10.1164/rccm.2009-040GL
- Mushiroda T, Wattanapokayakit S, Takahashi A, Nukiwa T, Kudoh S, Ogura T, et al. A genome-wide association study identifies an association of a common variant in TERT with susceptibility to idiopathic pulmonary fibrosis. *J Med Genet* (2008) 45:654–6. doi:10.1136/jmg.2008.057356
- Fingerlin TE, Murphy E, Zhang W, Peljto AL, Brown KK, Steele MP, et al. Genome-wide association study identifies multiple susceptibility loci for pulmonary fibrosis. *Nat Genet* (2013) 45:613–20. doi:10.1038/ng.2609
- Hutyrova B, Pantelidis P, Drabek J, Zurkova M, Kolek V, Lenhart K, et al. Interleukin-1 gene cluster polymorphisms in sarcoidosis and idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* (2002) 165:148–51. doi:10.1164/ajrccm.165.2.2106004
- Ahn MH, Park BL, Lee SH, Park SW, Park JS, Kim DJ, et al. A promoter SNP rs4073T>A in the common allele of the interleukin 8 gene is associated with the development of idiopathic pulmonary fibrosis via the IL-8 protein enhancing mode. *Respir Res* (2011) 12:73. doi:10.1186/1465-9921-12-73
- O'Dwyer DN, Armstrong ME, Trujillo G, Cooke G, Keane MP, Fallon PG, et al. The toll-like receptor 3 L412F polymorphism and disease progression in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* (2013) 188:1442–50. doi:10.1164/rccm.201304-0760OC
- Vasakova M, Sterclova M, Matej R, Olejar T, Kolesar L, Skibova J, et al. IL-4 polymorphisms, HRCT score and lung tissue markers in idiopathic pulmonary fibrosis. *Hum Immunol* (2013) 74:1346–51. doi:10.1016/j.humimm.2013.07.011
- Zhang HP, Zou J, Xie P, Gao F, Mu HJ. Association of HLA and cytokine gene polymorphisms with idiopathic pulmonary fibrosis. *Kaohsiung J Med Sci* (2015) 31:613–20. doi:10.1016/j.kjms.2015.10.007
- Seibold MA, Wise AL, Speer MC, Steele MP, Brown KK, Loyd JE, et al. A common MUC5B promoter polymorphism and pulmonary fibrosis. *N Engl J Med* (2011) 364:1503–12. doi:10.1056/NEJMoa1013660
- Kishore A, Zissel G, Zizkova V, Mueller-Quernheim J, Petrek M. Association of mucin (MUC2, MUC5B) gene variants with idiopathic pulmonary fibrosis (IPF) in a German population: a pilot study using MassARRAY technology. *Tissue Antigens* (2015) 85:340. doi:10.1111/tan.12557
- Korthagen NM, van Moorsel CH, Barlo NP, Kazemier KM, Ruven HJ, Grutters JC. Association between variations in cell cycle genes and idiopathic pulmonary fibrosis. *PLoS One* (2012) 7:e30442. doi:10.1371/journal.pone.0030442
- Petrek M, Kishore A, Zizkova V. Genetic association study for idiopathic pulmonary fibrosis in a Czech population: results from a pilot study using massarray technology. *Am J Respir Crit Care Med* (2015) 191:A4382–A. doi:10.1164/ajrccm-conference.2015.191.1\_MeetingAbstracts.A4382
- Uh ST, Kim TH, Shim EY, Jang AS, Park SW, Park JS, et al. Angiotensin-converting enzyme (ACE) gene polymorphisms are associated with idiopathic pulmonary fibrosis. *Lung* (2013) 191:345–51. doi:10.1007/s00408-013-9469-1
- Huang Y, Ma SF, Vij R, Oldham JM, Herazo-Maya J, Broderick SM, et al. A functional genomic model for predicting prognosis in idiopathic pulmonary fibrosis. *BMC Pulm Med* (2015) 15:147. doi:10.1186/s12890-015-0142-8
- Kishore A, Zizkova V, Kocourkova L, Petrek M. A dataset of 26 candidate gene and pro-inflammatory cytokine variants for association studies in idiopathic pulmonary fibrosis: frequency distribution in normal Czech population. *Front Immunol* (2015) 6:476. doi:10.3389/fimmu.2015.00476
- Little J, Higgins JP, Ioannidis JP, Moher D, Gagnon F, von Elm E, et al. Strengthening the reporting of genetic association studies (STREGA): an extension of the STROBE statement. *Eur J Epidemiol* (2009) 24:37–55. doi:10.1007/s10654-008-9302-y
- Heath SC, Gut IG, Brennan P, McKay JD, Bencko V, Fabianova E, et al. Investigation of the fine structure of European populations with applications to disease association studies. *Eur J Hum Genet* (2008) 16:1413–29. doi:10.1038/ejhg.2008.210
- Lundmark PE, Liljedahl U, Boomsma DI, Mannila H, Martin NG, Palotie A, et al. Evaluation of HapMap data in six populations of European descent. *Eur J Hum Genet* (2008) 16:1142–50. doi:10.1038/ejhg.2008.77
- Sanchez-Mazas A, Vidan-Jeras B, Nunes JM, Fischer G, Little AM, Bekmane U, et al. Strategies to work with HLA data in human populations for histocompatibility, clinical transplantation, epidemiology and population genetics: HLA-NET methodological recommendations. *Int J Immunogenet* (2012) 39:459–72. doi:10.1111/j.1744-313X.2012.01113.x
- Kubistova Z, Mrazek F, Tudos Z, Kriegova E, Ambruzova Z, Mytilineos J, et al. Distribution of 22 cytokine gene polymorphisms in the healthy Czech population. *Int J Immunogenet* (2006) 33:261–7. doi:10.1111/j.1744-313X.2006.00609.x
- American Thoracic Society. Idiopathic pulmonary fibrosis: diagnosis and treatment. International consensus statement. American Thoracic Society (ATS), and the European Respiratory Society (ERS). *Am J Respir Crit Care Med* (2000) 161:646–64. doi:10.1164/ajrccm.161.2.ats3-00
- American Thoracic Society, European Respiratory Society. American Thoracic Society/European Respiratory Society International Multidisciplinary Consensus Classification of the Idiopathic Interstitial Pneumonias. This joint statement of the American Thoracic Society (ATS), and the European Respiratory Society (ERS) was adopted by the ATS board of directors, June 2001 and by the ERS Executive Committee, June 2001. *Am J Respir Crit Care Med* (2002) 165:277–304. doi:10.1164/ajrccm.165.2.ats01
- Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* (1988) 16:1215. doi:10.1093/nar/16.3.1215
- Yang H, Robinson PN, Wang K. Phenolyzer: phenotype-based prioritization of candidate genes for human diseases. *Nat Methods* (2015) 12:841–3. doi:10.1038/nmeth.3484
- Seibold MA, Smith RW, Urbanek C, Groshong SD, Cosgrove GP, Brown KK, et al. The idiopathic pulmonary fibrosis honeycomb cyst contains a mucociliary pseudostratified epithelium. *PLoS One* (2013) 8:e58658. doi:10.1371/journal.pone.0058658
- van der Vis JJ, Snetelaar R, Kazemier KM, Ten Klooster L, Grutters JC, van Moorsel CH. Effect of Muc5b promoter polymorphism on disease predisposition and survival in idiopathic interstitial pneumonias. *Respirology* (2016) 21(4):712–7. doi:10.1111/resp.12728
- Yang IV, Fingerlin TE, Evans CM, Schwarz MI, Schwartz DA. MUC5B and idiopathic pulmonary fibrosis. *Ann Am Thorac Soc* (2015) 12(Suppl 2):S193–9. doi:10.1513/AnnalsATS.201503-110AW
- Uhal BD, Kim JK, Li X, Molina-Molina M. Angiotensin-TGF-beta 1 crosstalk in human idiopathic pulmonary fibrosis: autocrine mechanisms in myofibroblasts and macrophages. *Curr Pharm Des* (2007) 13:1247–56. doi:10.2174/138161207780618885
- Li X, Molina-Molina M, Abdul-Hafez A, Uhal V, Xaubet A, Uhal BD. Angiotensin converting enzyme-2 is protective but downregulated in human and experimental lung fibrosis. *Am J Physiol Lung Cell Mol Physiol* (2008) 295:L178–85. doi:10.1152/ajplung.00009.2008
- Alder JK, Chen JJ, Lancaster L, Danoff S, Su SC, Cogan JD, et al. Short telomeres are a risk factor for idiopathic pulmonary fibrosis. *Proc Natl Acad Sci U S A* (2008) 105:13051–6. doi:10.1073/pnas.0804280105
- Wei R, Li C, Zhang M, Jones-Hall YL, Myers JL, Noth I, et al. Association between MUC5B and TERT polymorphisms and different interstitial lung disease phenotypes. *Transl Res* (2014) 163:494–502. doi:10.1016/j.trsl.2013.12.006
- Hutchinson J, Fogarty A, Hubbard R, McKeever T. Global incidence and mortality of idiopathic pulmonary fibrosis: a systematic review. *Eur Respir J* (2015) 46:795–806. doi:10.1183/09031936.00185114
- Kishore A, Borucka J, Petrakova J, Petrek M. Novel insights into miRNA in lung and heart inflammatory diseases. *Mediators Inflamm* (2014) 2014:259131. doi:10.1155/2014/259131
- Rajasekaran S, Rajaguru P, Sudhakar Gandhi PS. MicroRNAs as potential targets for progressive pulmonary fibrosis. *Front Pharmacol* (2015) 6:254. doi:10.3389/fphar.2015.00254

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Kishore, Žižková, Kocourková, Petrakova, Bouros, Nunes, Lošťáková, Müller-Quernheim, Zissel, Kolek, Bouros, Valeyre and Petrek. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Characterization of Three *CYP2C19* Gene Variants by MassARRAY and Point of Care Techniques: Experience from a Czech Centre

Martin Petrek<sup>1,2,3</sup>  · Lenka Kocourkova<sup>1,3</sup> · Veronika Zizkova<sup>1,3</sup> · Zdenek Nosek<sup>3</sup> · Milos Taborsky<sup>4</sup> · Jana Petrakova<sup>3,4</sup>

Received: 29 July 2016 / Accepted: 20 October 2016 / Published online: 12 January 2017  
© L. Hirszfeld Institute of Immunology and Experimental Therapy, Wrocław, Poland 2016

**Abstract** Distribution of cytochrome P450 2C19 enzyme gene (*CYP2C19*) variants affecting metabolism of clopidogrel was determined in 526 Czech patients after percutaneous coronary intervention using MassARRAY genotyping and compared to distribution in other populations of European descent. Fifty-three (10%) patients underwent parallel determination of *CYP2C19* genotypes from buccal swabs by a point of care technique with 100% concordance to the main genotyping platform. Observed *CYP2C19* genotypes were related to clopidogrel metabolism phenotypes and discussed in population context. Hereby, presented methodologies provide accurate *CYP2C19* genotyping results in a relatively short time of one up to 12 h and may, therefore, find the relevant place in the field of genotype-guided antiplatelet therapy.

**Keywords** *CYP2C19* · Pharmacogenetics · Clopidogrel · Genotyping · Point of care

## Introduction

Pharmacogenetics (PGx)-study of variations in DNA sequences affecting drug response and/or toxicity (Yip et al. 2015)-has been gradually gaining its place in personalised patient care in various fields of medicine (Jani et al. 2015; Ortega et al. 2015; Petrek 2015; Tangamornsuksan et al. 2015). Clopidogrel, as the drug most commonly prescribed together with aspirin after percutaneous coronary intervention (PCI) to prevent atherothrombotic events in patients with acute coronary syndrome, has an important place among the drugs in which PGx tests proved to have adequate clinical validity. Available data, including those from meta-analyses, indicate that patients on clopidogrel carrying particular variants of the cytochrome P450 2C19 enzyme gene (*CYP2C19*) may be at an increased risk of adverse cardiovascular events compared with noncarriers (Hulot et al. 2010; Mega et al. 2010a, b; Sofi et al. 2011; Zabalza et al. 2012). Indeed, FDA black box warning suggests consideration of alternative therapy in patients identified as *CYP2C19* poor metabolizers and notes that these patients could be identified by genotyping (US Food and Drug Administration 2010). Recommendations regarding *CYP2C19* genotyping in context of clopidogrel therapy have also been made by Clinical Pharmacogenetics Implementation Consortium (Scott et al. 2013). There have been a number of academic institutions which have implemented hospital-wide reporting of *CYP2C19* genotypes in their electronic medical records, also to facilitate health care provider decision on treatment by alternative platelet inhibitors, such as prasugrel and ticagrelor (Knauer et al. 2015). Recently, there also have been reports on cost-effectiveness of genotype-guided antiplatelet therapy (Jiang and You 2015; Kazi et al. 2014).

✉ Martin Petrek  
martin.petek@fnol.cz

✉ Jana Petrakova  
jana.petrakova@fnol.cz

<sup>1</sup> Laboratory of Cardiogenomics-LEM, University Hospital Olomouc, Olomouc, Czech Republic

<sup>2</sup> Faculty of Medicine and Dentistry, Institute of Molecular and Translational Medicine, Palacký University Olomouc, Olomouc, Czech Republic

<sup>3</sup> Department of Pathological Physiology, Faculty of Medicine and Dentistry, Palacký University Olomouc, Olomouc, Czech Republic

<sup>4</sup> Department of Internal Medicine-Cardiology, Faculty of Medicine and Dentistry, Palacký University and University Hospital Olomouc, Olomouc, Czech Republic

Wider application of PGx testing for the *CYP2C19* variants underlying altered responsiveness to clopidogrel has been hindered by time-lag between specimen sampling and results' reporting which, in usual routine, involves 3 or more days. However, in patients endangered by complications, the genotype knowledge is desirable at the time of prescription and any therapeutic adjustment should be decided in a very limited time frame (Knauer et al. 2015). In this context, we aimed at facilitating our genotyping by MassARRAY technology, by which turn-around-time of 12 h could be achieved, and additionally by utilisation of a commercial point of care (POC) genotyping technique producing result within 1 h. Our findings, including population data on *CYP2C19* variants relevant for clopidogrel metabolism in Czechs and other Europeans, are summarised in the following report.

## Materials and Methods

### Characteristics of the Study Group

Five hundred and twenty-six patients (138 females, 388 males) after PCI from University Hospital Olomouc underwent routine examination of *CYP2C19* in the period from March 2013 to February 2016. Briefly, a sample of peripheral blood was obtained for genotyping by MassARRAY<sup>®</sup> technology (AgenaBio, San Diego, USA). The mean patient age was 66 years and ranged 24–93 years. Randomly selected patient subset ( $n = 53$ , 14 females, 39 males) was tested in parallel using a POC technique-Spartan RX System (Spartan Bioscience Inc., Ottawa, Ontario, Canada) utilising a buccal swab specimen. The mean age in this patient subset was 57 years and it ranged 31–77 years. All subjects were unrelated subjects living in the Czech Republic, i.e., of Central European descent, speaking Czech language. Signed informed consent explaining the nature of the examination and containing approval with usage of the routine test results for scientific purpose was obtained before specimen collection from each subject.

### *CYP2C19* Genotyping by MassARRAY

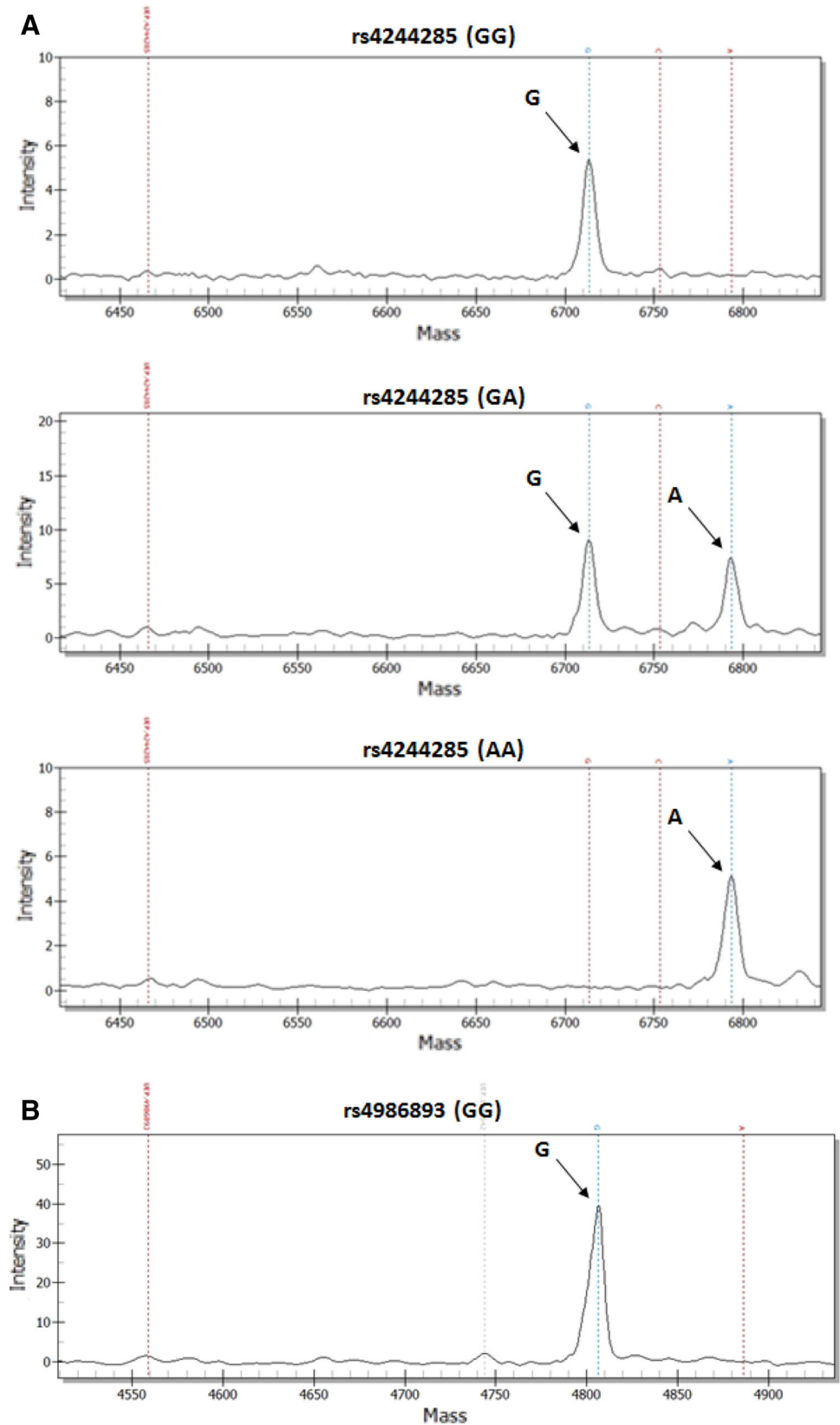
Genomic DNA was isolated from 0.5 ml of peripheral blood drawn into K<sub>3</sub>EDTA-coated tubes using Arrow automated extraction system (Isogen Life Science, De Meern, The Netherlands). Methodical details on MassARRAY procedure have previously been described, in general (Stacey et al. 2009), and also regarding our laboratory procedure (Kishore et al. 2015). In brief, the primer pairs (Assay design suite v2.0; <https://agenacx.com/Tools>) were multiplexed. Genotyping was performed using

Sequenom MassARRAY platform integrating iPLEX<sup>®</sup> SBE (single base extension reaction) and MassARRAY<sup>®</sup> technology (Agena Bioscience, San Diego, CA, USA)-based MALDI-TOF MS assay. The assay consists of an initial locus-specific PCR amplification followed by SBE using mass-modified dideoxynucleotide terminators of an oligonucleotide primer that anneals immediately upstream of the target polymorphic site. The distinct mass of extended primer traces the alternative alleles using MassARRAY Typer 4.0.20. For quality control (QC) step, we determined data missing rate per individuals and missing rate per single-nucleotide polymorphism (SNP). In addition, for QC of SNP genotyping, positive and negative template control samples were included in each assay plate. Any assay found as positive in negative template control was removed from the study. The correctness of our *CYP2C19* genotyping has been repeatedly verified in three annual intervals (2013–2015) by successful participation in the external proficiency testing scheme “Molecular Genetics set 06” (775) provided by INSTAND, Düsseldorf, Germany.

### *CYP2C19* Genotyping Using POC Technique

In a subset of our patients ( $n = 53$ ), *CYP2C19*\*2,\*3 and \*17 variants were determined by Spartan RX System (Spartan Bioscience Inc., Ottawa, Ontario, Canada)-a POC system based on processing of buccal swabs in a closed, semi-automated amplification system. According to the manufacturer, Spartan RX Analyser employs optical detection and subsequent automated analysis for evaluation of fluorescent signals from oligonucleotide probes binding to PCR amplicons and thus determinates the genotype; it consists of Spartan RX *CYP2C19* Assay and Spartan RX Platform. Three sample collection kits, each for one of the three *CYP2C19* tested variants, were used to collect buccal swab samples. The collection kit, provided by the manufacturer, consists of a pouch containing a buccal swab and a reagent tube. The reagent tube contains chemicals for DNA extraction, PCR amplification, and fluorescent detection of the specific *CYP2C19* allele. Collection procedure and subsequent sample processing were performed according to Spartan RX System Operator's Manual (Document Number 01001920\_1.5, 1–37). Briefly, before sample collection, patient rinsed the mouth by a small amount of water, then a buccal swab from the inside of patient's cheek was collected and samples were immediately processed in a nearby laboratory according to the manufacturer's instructions. Briefly, the reagents and buccal samples inside reagent tubes were mixed by taping on their bottom. Next, tubes were inserted into Spartan RX Analyser and in less than an hour, the test completed with automatic print-out of the results.

**Fig. 1** Mass spectra for genotypes of *CYP2C19* variants rs4244285 ( $G > A$ ), *CYP2C19\*2* (a), rs4986893 ( $G > A$ ), *CYP2C19\*3* (b), and rs12248560 ( $C > T$ ), *CYP2C19\*17* (c); representative examples. For genotype designations, refer to top of the panels; peaks denote particular alleles, single peaks represent homozygous, and double peaks represent heterozygous combinations



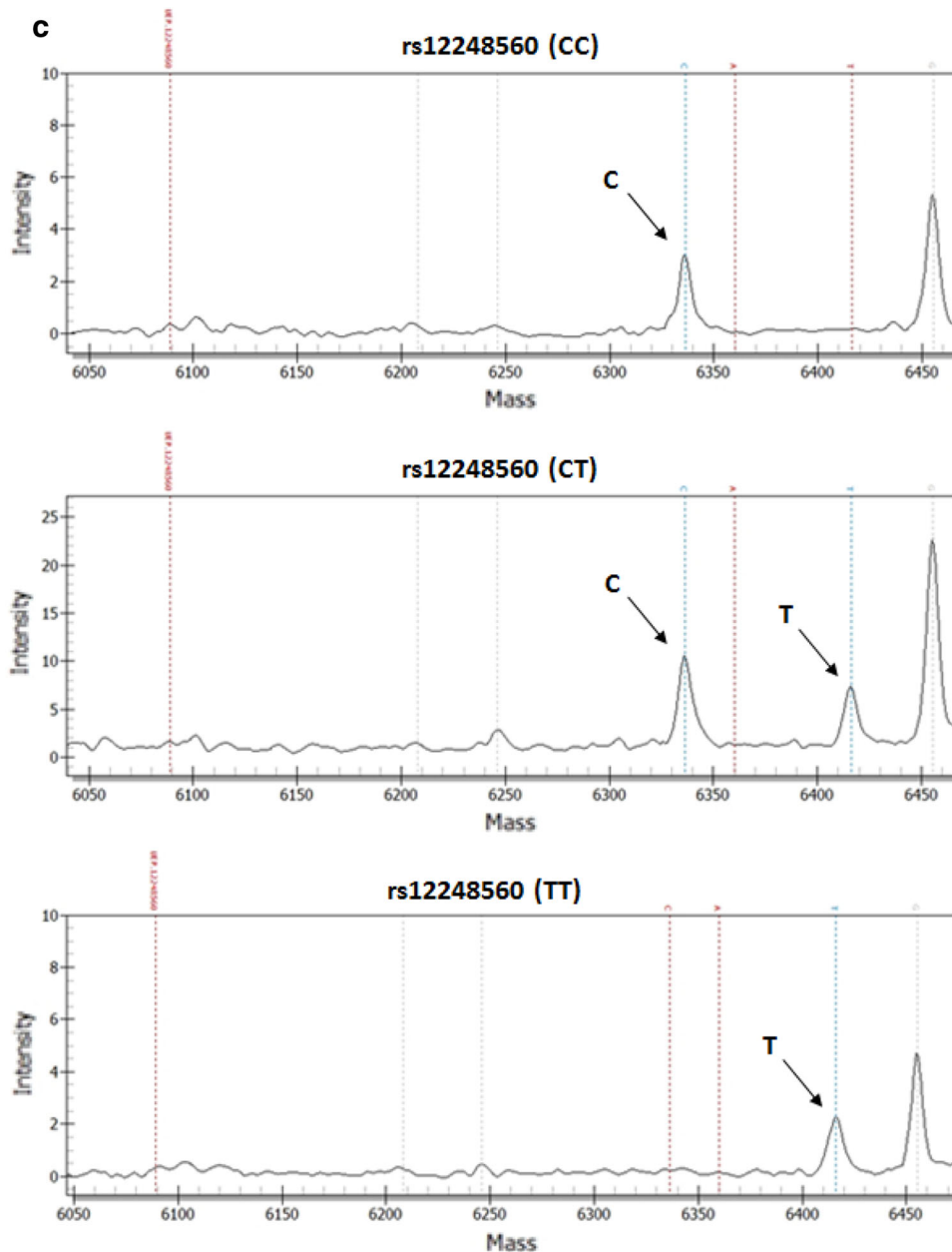


Fig. 1 continued

### Statistical Analysis

Each of the three SNPs was tested for Hardy–Weinberg equilibrium (HWE) by Pearson's Chi-square ( $\chi^2$ ) test ( $p > 0.05$ ). Genotype and allele frequencies were determined by direct counting. Carriage rate (phenotype frequency) was calculated as number of individuals carrying one (or two) copies of a particular allele on one or both (maternal and paternal) chromosomes.

### Results

#### *CYP2C19* Variant Determination by MassARRAY

All tested *CYP2C19* SNPs could be determined with following success rates: *CYP2C19*\*2, 99.8%; \*3, 99.8%; \*17, 99.2%; and, overall, 99.6%. Representative examples of the spectra from MassARRAY are shown in Fig. 1a–c; the distribution of respective genotypes for each of the SNPs is

illustrated by call cluster plots from MassARRAY (Fig. 2a–c). The turn-around-time, from blood sampling to the result print-out, was 12 h in urgent (*statim*) regimen, otherwise from 24 to 36 h.

The distribution of genotypes and alleles of the *CYP2C19* variants in our cohort is summarised in Table 1 (left columns); carriage rates for variant alleles were 0.26 for *CYP2C19*\*2 A and 0.44 for T allele of *CYP2C19*\*17. For *CYP2C19*\*2 and \*17 variants, the distributions of genotypes were within HWE ( $p > 0.05$ ); regarding *CYP2C19*\*3, due to the absence of the variant A allele,  $\chi^2$  value could not be defined.

The observed *CYP2C19* variant frequencies in Czechs were fully comparable with those reported for other populations of European descent (Table 1, middle and right columns); the reference population data were obtained from The 1000 Genomes Project (The 1000 Genomes Project Consortium 2015).

### ***CYP2C19* Variant Determination by POC Technique**

The success rate for determination of *CYP2C19* variants was 84% (for specific SNPs: *CYP2C19*\*2: 87%, \*3: 83%, \*17: 81%). In 10 out of 53 tested samples, the results could not be obtained: there was four times failure during the amplification procedure; three times the report classified results as inconclusive; and in three cases, only two of the three variants could be determined.

Comparison of the genotypes provided by POC with those yielded from parallel examination by MassARRAY showed absolute (100%) concordance. The turn-around-time (from buccal swab sampling to the result print-out) was 60 min.

### **Relevance for Clopidogrel Metabolism**

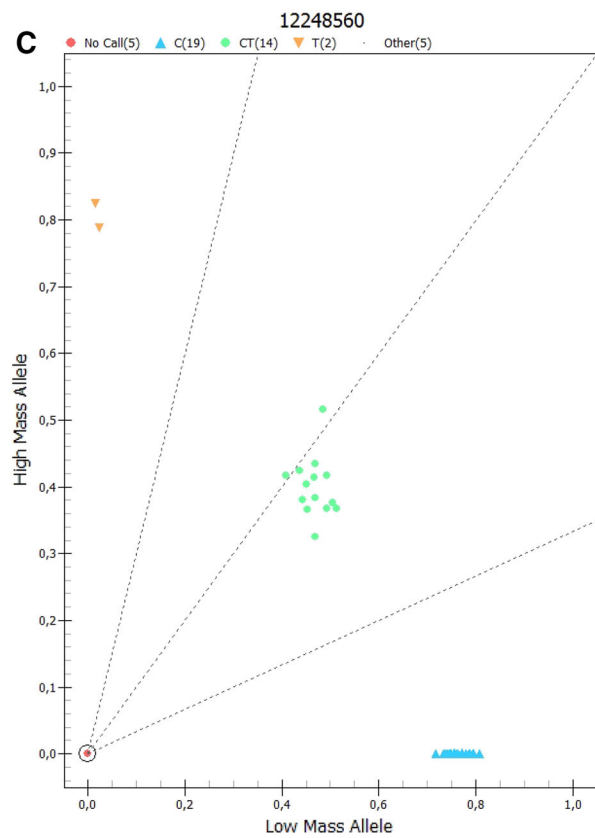
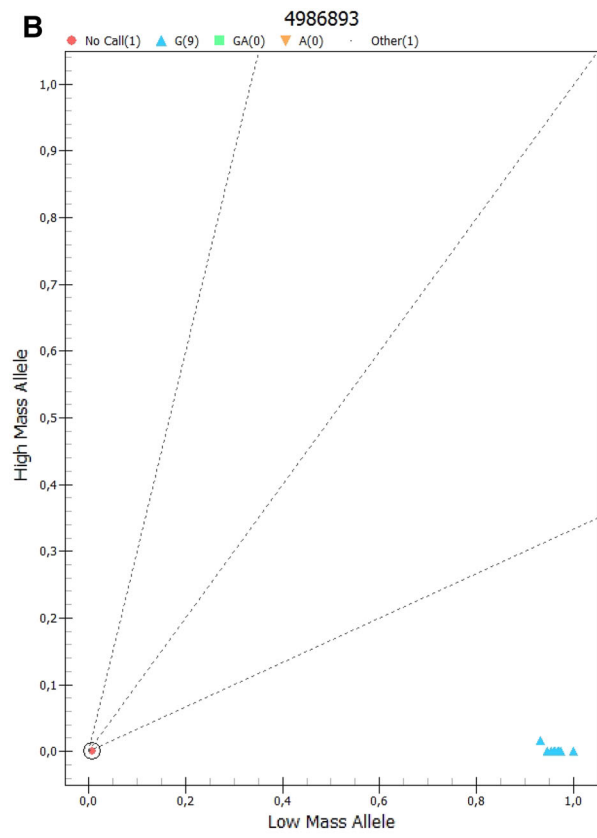
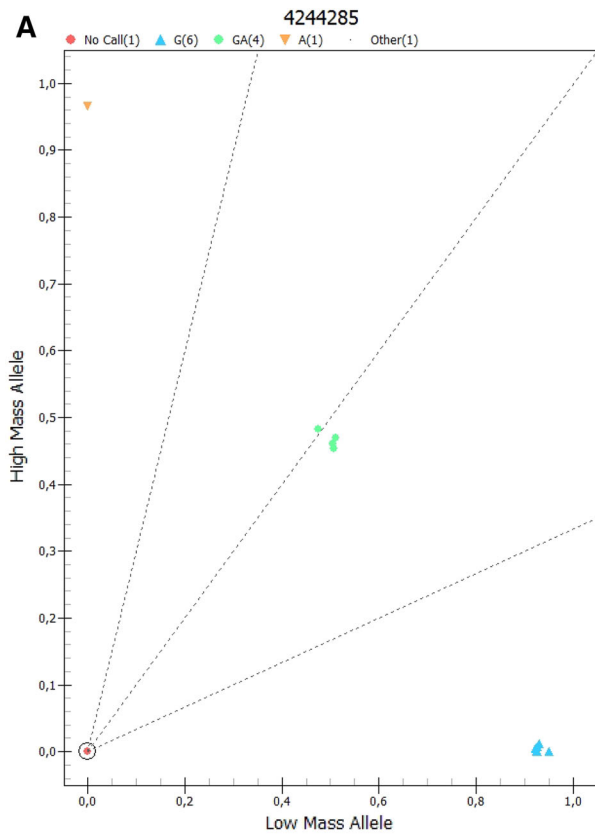
To move from the laboratory genotyping to applied area, the obtained genotypes were divided according to their relevance for clopidogrel metabolism and thus predicted risk of cardiovascular complications (Table 2, left part); the division was based on the classification by Desai et al. (2013). Again, apart from reporting results regarding our study group, we also provide comparisons of the distribution of the particular genotypes in Czechs to other cohorts of distinct ethnicity, more specifically to those of Northern/Southern Europeans and U.S. of European descent (Table 2, middle and right parts). From these comparisons, no substantial differences emerged.

## **Discussion**

Hereby, we present genotyping data on *CYP2C19*\*2, \*3, and \*17 variants relevant for clopidogrel metabolism, which were obtained by MassARRAY methodology and in part also by a POC technique in Czech, i.e., Central European cohort counting more than 500 subjects. Both utilised tests proved to be able to determine specific genotypes in a short-time frame (1–12 h) and may, therefore, represent the way towards progress in clinical applications of *CYP2C19* genotyping for management of patients requiring treatment by clopidogrel or alternative platelet inhibitors.

When we compared our Czech data with those reported for other populations of European descent, there were no major differences in the frequencies of individual variants or in the distribution of genotypes relevant for clopidogrel metabolism. In this regard, approximately, a quarter of our patients carried the loss-of-function allele *CYP2C19*\*2 A, i.e., were poor clopidogrel metabolizers. The A allele is associated with reduced conversion of clopidogrel to active metabolite and these patients are in increased risk of cardiovascular complication including stent thrombosis (Mega et al. 2010b). Another *CYP2C19* loss-of-function allele (\*3 A) was not observed in our cohort which was in line with its extremely rare occurrence in Europeans; this allele is present only in Asians with frequency from 1 to 6% (The 1000 Genomes Project Consortium 2015). *CYP2C19*\*17 TT homozygote individuals were presented in 5%; these “ultra-rapid metabolizers” may have increased risk of bleeding events (Li et al. 2012). Similar to other studies (e.g., Desai et al. 2013), approximately one-third of our patients possessed *CYP2C19* variants phenotypically expressed by alteration of clopidogrel metabolism. Hereby, reported availability of reasonably short-time identification of these variants would, therefore, allow appropriate patient monitoring including consideration of treatment by alternative drugs.

Regarding current initiatives on relocating genotyping from the laboratory to the bedside (Beitelshees 2012; Beitelshees et al. 2015; Knauer et al. 2015), our data may add a small piece into the emerging mosaic. So far, the solid data in area of POC clopidogrel genotyping have been reported by Roberts et al. (2012) whose report suggested that POC testing after PCI can be done effectively at the bedside and be beneficial for treatment of identified *CYP2C19*\*2 carriers with prasugrel; So et al. (2016) recently added report of a POC extended to contain *ABCB1* testing. The presence of Spartan RX and that of another commercial POC test on the market (Nanosphere Verigen System, see <http://www.nanosphere.us/products/cyp2c19-test>) also provided impetus for development of further





**Fig. 2** Call cluster plots showing distribution of *CYP2C19* genotypes; representative examples. **a** rs4244285 ( $G > A$ ), *CYP2C19*\*2; blue symbols represent genotype GG; green symbols GA; orange symbols AA. **b** rs4986893 ( $G > A$ ), *CYP2C19*\*3; blue symbols represent genotype GG, other genotypes not present. **c** rs12248560 ( $C > T$ ), *CYP2C19*\*17; blue symbols represent genotype CC; green symbols CT; orange symbols TT

alternative POC techniques, such as a microarray-based POC *CYP2C19* genotyping test (Chae et al. 2013) or a recently described chip technique (Marziliano et al. 2015). Before these techniques are readily applied to clinics, there is urgent need for further improvement and standardization including development of relevant external quality assessment schemes as in other fields where genotyping has already been routinely applied to patient care (e.g., Haselmann et al. 2016). In this context, there has been report about specific *CYP2C19* QC outside Europe (Lin et al. 2015).

The strength of our report lies not only within the great number of genotyped subjects, but also regards our main genotyping methodology, which has been controlled by our continuous participation in the external proficiency exercise and the robustness of which we have also documented in other SNP genotyping applications (Kishore et al. 2015, 2016). Regarding limitations of our study, it should be noted that we have involved patients for determination of genotype distribution; however, also our own data confirm that the distribution in patients reflects that in normal reference population (The 1000 Genomes Project Consortium 2015). Otherwise, we did not include other variants known to affect clopidogrel metabolism, such as *ABCB1* (Mega et al. 2010a); however, *ABCB1* gene was not covered by the utilised POC system at the time of our analyses and thus will not enable comparisons.

Importantly, only general knowledge of the Spartan RX POC test methodology is another limiting factor. While with our MassARRAY protocol, we had under

**Table 1** Genotype and allele frequencies of *CYP2C19* variants in Czechs in comparison to other European populations

Allele	SNP	Genotype frequencies	Genotype frequencies (EUR)	Genotype frequencies (CEU)	Allele frequencies	Allele frequencies (EUR)	Allele frequencies (CEU)		
<i>CYP2C19</i> *2	19154 $G > A$ rs4244285	GG	0.745	0.722	0.758	G	0.864	0.855	0.869
		GA	0.238	0.266	0.222	A	0.136	0.145	0.131
		AA	0.017	0.012	0.020				
<i>CYP2C19</i> *17	806 $C > T$ rs12248560	CC	0.559	0.596	0.586	C	0.753	0.776	0.778
		CT	0.388	0.360	0.384	T	0.247	0.224	0.222
		TT	0.053	0.044	0.030				
<i>CYP2C19</i> *3	636 $G > A$ rs4986893	GG	1.000	1.000	1.000	G	1.000	1.000	1.000
		GA	0.000	0.000	0.000	A	0.000	0.000	0.000
		AA	0.000	0.000	0.000				

EUR European, CEU Utah residence with Northern and Western European Ancestry

For the reference population data, see The 1000 Genomes Project Consortium (2015)

**Table 2** Distribution of *CYP2C19* genotypes according to their relevance for clopidogrel metabolism phenotypes in Czech and other European cohorts

Clopidogrel metabolism phenotype	<i>CYP2C19</i> genotype	Population (frequency %)					
		Czech ( $n = 526$ )	Greek ( $n = 283$ ) (Ragia et al. 2009)	Danish ( $n = 276$ ) (Pedersen et al. 2010)	Faroese ( $n = 311$ ) (Pedersen et al. 2010)	Norwegian ( $n = 328$ ) (Pedersen et al. 2010)	American-USA ( $n = 499$ ) (Desai et al. 2013)
Poor metabolizer	*2/*2	1.7	2.1	2.2	3.2	1.3	3.0
Intermediate metabolizer	*1/*2	23.8	17.7	18.5	23.5	20.1	26.3
Extensive metabolizer	*1/*1, *1/*17	69.2	72.8	67.0	62.4	66.0	65.7
Ultra metabolizer	*17/*17	5.3	3.2	5.1	3.5	4.9	3.2

Classification of metabolism phenotypes according to Desai et al. (2013). For numbers of subjects, see “ $n$ ” next to population designations

control all methodical steps, including reagents' and primer-plexes preparation and handling, with the "closed" nature of the POC system, these and other technical details were unknown, and genotyping was beyond our control from the time-point of inserting the swab into the analyser. Similar situation occurred also in the interpretation phase: while the POC analyser just prints the test outcome, with the MassARRAY, scientists can see the range of parameters including peaks corresponding to amplicons and if relevant they can perform necessary validations. Thus, any explanation of the observed lower success rate of POC genotyping would not be based on solid grounds and will be purely speculative; the same refers to the reasoning for particular statements of the POC system, such as "failure during the amplification procedure" or "inconclusive result" as the QC step of the POC data is not known to us. What we, however, can recommend based on our experience is to carefully respect all steps of the preanalytical phase, especially handling of the swab-stick, performing swab procedure itself, and keeping temperatures and times as described in the SPARTAN RX protocol.

In conclusion, our aggregate data on distribution of main pharmacogenetically relevant *CYP2C19* variants for clopidogrel metabolism obtained by two hereby utilised specific methodologies suggest that there are opportunities for providing genotyping results in a reasonable time frame that can facilitate their application in clinical patient management.

**Acknowledgements** Part of this work was performed when Laboratory of Cardiogenomics was affiliated to the Department of Clinical and Molecular Pathology, University Hospital Olomouc. This work was partially supported by Grants: IGA PU: LF\_2013\_009, 2014\_012, 2015\_020, and LO1304.

## References

- Beitelshees AL (2012) Personalised antiplatelet treatment: a RAPIDly moving target. *Lancet* 379:1680–1682
- Beitelshees AL, Voora D, Lewis JP (2015) Personalized antiplatelet and anticoagulation therapy: applications and significance of pharmacogenomics. *Pharmacogenom Pers Med* 8:43–61
- Chae H, Kim M, Koh YS et al (2013) Feasibility of a microarray-based point-of-care *CYP2C19* genotyping test for predicting clopidogrel on-treatment platelet reactivity. *Biomed Res Int* 2013:154073
- Desai NR, Canestaro WJ, Kyrychenko P et al (2013) Impact of *CYP2C19* genetic testing on provider prescribing patterns for antiplatelet therapy after acute coronary syndromes and percutaneous coronary intervention. *Circ Cardiovasc Qual Outcomes* 6:694–699
- Haselmann V, Geilenkeuser WJ, Helfert S (2016) Thirteen years of an international external quality assessment scheme for genotyping: results and recommendations. *Clin Chem* 62:1084–1095
- Hulot JS, Collet JP, Silvain J et al (2010) Cardiovascular risk in clopidogrel-treated patients according to cytochrome P450 2C19\*2 loss-of-function allele or proton pump inhibitor coadministration: a systematic meta-analysis. *J Am Coll Cardiol* 56:134–143
- Jani M, Barton A, Ho P (2015) Pharmacogenetics of treatment response in psoriatic arthritis. *Curr Rheumatol Rep* 17(7):44. doi:10.1007/s11926-015-0518-z
- Jiang M, You JH (2015) Review of pharmacoeconomic evaluation of genotype-guided antiplatelet therapy. *Expert Opin Pharmacother* 16:771–779
- Kazi DS, Garber AM, Shah RU et al (2014) Cost-effectiveness of genotype-guided and dual antiplatelet therapies in acute coronary syndrome. *Ann Intern Med* 160:221–232
- Kishore A, Žižková V, Kocourková L et al (2015) A dataset of 26 candidate gene and pro-inflammatory cytokine variants for association studies in idiopathic pulmonary fibrosis: frequency distribution in normal Czech population. *Front Immunol* 22:476
- Kishore A, Žižková V, Kocourková L et al (2016) Association study for 26 candidate loci in idiopathic pulmonary fibrosis patients from four European populations. *Front Immunol* 7:274
- Knauer MJ, Diamandis EP, Hulot JS et al (2015) Clopidogrel and *CYP2C19*: pharmacogenetic testing ready for clinical prime time? *Clin Chem* 61:1235–1240
- Li Y, Tang HL, Hu YF et al (2012) The gain-of-function variant allele *CYP2C19\*17*: a double-edged sword between thrombosis and bleeding in clopidogrel-treated patients. *J Thromb Haemost* 10:199–206
- Lin G, Yi L, Zhang K et al (2015) Implementation of cell samples as controls in national proficiency testing for clopidogrel therapy-related *CYP2C19* genotyping in China: a novel approach. *PLoS One* 10:e0134174
- Marziliano N, Notarangelo MF, Cereda M et al (2015) Rapid and portable, lab-on-chip, point-of-care genotyping for evaluating clopidogrel metabolism. *Clin Chim Acta* 451:240–246
- Mega JL, Close SL, Wiviott SD et al (2010a) Genetic variants in *ABCB1* and *CYP2C19* and cardiovascular outcomes after treatment with clopidogrel and prasugrel in the TRITON-TIMI 38 trial: a pharmacogenetic analysis. *Lancet* 376:1312–1319
- Mega JL, Simon T, Collet JP et al (2010b) Reduced-function *CYP2C19* genotype and risk of adverse clinical outcomes among patients treated with clopidogrel predominantly for PCI: a meta-analysis. *JAMA* 304:1821–1830
- Ortega VE, Meyers DA, Bleeker ER (2015) Asthma pharmacogenetics and the development of genetic profiles for personalized medicine. *Pharmacogenom Pers Med* 8:9–22
- Pedersen RS, Brasch-Andersen C, Sim SC et al (2010) Linkage disequilibrium between the *CYP2C19\*17* allele and wildtype *CYP2C8* and *CYP2C9* alleles: identification of *CYP2C* haplotypes in healthy Nordic populations. *Eur J Clin Pharmacol* 66:1199–1205
- Petrek M (2015) Personalized medicine in sarcoidosis: predict responders and nonresponders. *Curr Opin Pulm Med* 21:532–537
- Ragia G, Arvanitidis KI, Tavridou A et al (2009) Need for reassessment of reported *CYP2C19* allele frequencies in various populations in view of *CYP2C19\*17* discovery: the case of Greece. *Pharmacogenomics* 10:43–49
- Roberts JD, Wells GA, LeMay MR et al (2012) Point-of-care genetic testing for personalization of antiplatelet treatment (rapid gene): a prospective, randomised, proof of concept trials. *Lancet* 379:1705–1711
- Scott SA, Sangkuhl K, Stein CM et al (2013) Clinical Pharmacogenetics Implementation Consortium guidelines for *CYP2C19* genotype and clopidogrel therapy: 2013 update. *Clin Pharmacol Ther* 94:317–323
- So DY, Wells GA, McPherson R et al (2016) A prospective randomized evaluation of a pharmacogenomic approach to antiplatelet therapy among patients with ST-elevation

- myocardial infarction: the RAPID STEMI study. *Pharmacogenom J* 16:71–78
- Sofi F, Giusti B, Marcucci R et al (2011) Cytochrome P450 2C19\*2 polymorphism and cardiovascular recurrences in patients taking clopidogrel: a meta-analysis. *Pharmacogenomics J* 11:199–206
- Stacey G, Ziaugra L, Tabbaa D (2009) SNP genotyping using the sequenom MassARRAY iPLEX platform. *Curr Protoc Hum Genet* 60:2.12.1–2.12.18
- Tangamornsuksan W, Lohitnavy O, Kongkaew C et al (2015) Association of HLA-B\*5701 genotypes and abacavir-induced hypersensitivity reaction: a systematic review and meta-analysis. *J Pharm Pharm Sci* 18:68–76
- The 1000 Genomes Project Consortium (2015) A global reference for human genetic variation. *Nature* 526:68–74
- US Food and Drug Administration (2010) FDA drug safety communication: reduced effectiveness of Plavix (clopidogrel) in patients who are poor metabolizers of the drug (online). <http://www.fda.gov/Drugs/DrugSafety/PostmarketDrugSafetyInformationforPatientsandProviders/ucm203888.htm>. Accessed 25 July 2016
- Yip V, Hawcutt DB, Pirmohamed M (2015) Pharmacogenetic markers of drug efficacy and toxicity. *Clin Pharmacol Ther* 98:61–70
- Zabalza M, Subirana I, Sala J et al (2012) Meta-analyses of the association between cytochrome CYP2C19 loss- and gain-of-function polymorphisms and cardiovascular outcomes in patients with coronary artery disease treated with clopidogrel. *Heart* 98:100–108

## Association of *TGF-β3* and *ANXA11* with pulmonary sarcoidosis in Greek population

Katerina Sikorova , Amit Kishore , Angeliki Rapti , Kalliopi Adam , Lenka Kocourkova , Veronika Zizkova , Maria Charikiopoulou , Anastasios Kalianos , Evangelos Bouros , Demosthenes Bouros & Martin Petrek

To cite this article: Katerina Sikorova , Amit Kishore , Angeliki Rapti , Kalliopi Adam , Lenka Kocourkova , Veronika Zizkova , Maria Charikiopoulou , Anastasios Kalianos , Evangelos Bouros , Demosthenes Bouros & Martin Petrek (2020) Association of *TGF-β3* and *ANXA11* with pulmonary sarcoidosis in Greek population, Expert Review of Respiratory Medicine, 14:10, 1065-1069, DOI: [10.1080/17476348.2020.1784729](https://doi.org/10.1080/17476348.2020.1784729)

To link to this article: <https://doi.org/10.1080/17476348.2020.1784729>



Published online: 07 Jul 2020.



Submit your article to this journal [↗](#)



Article views: 20



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 1 View citing articles [↗](#)

ORIGINAL RESEARCH



## Association of *TGF-β3* and *ANXA11* with pulmonary sarcoidosis in Greek population

Katerina Sikorova<sup>a</sup>, Amit Kishore<sup>a\*</sup>, Angeliki Rapti<sup>b</sup>, Kalliopi Adam<sup>c</sup>, Lenka Kocourkova<sup>d</sup>, Veronika Zizkova<sup>a</sup>, Maria Charikiopoulou<sup>b</sup>, Anastasios Kalianos<sup>b</sup>, Evangelos Bouros<sup>e</sup>, Demosthenes Bouros<sup>e</sup> and Martin Petrek<sup>a</sup>

<sup>a</sup>Department of Pathological Physiology and Institute of Molecular and Translational Medicine, Faculty of Medicine and Dentistry, Palacky University Olomouc, Olomouc, Czech Republic; <sup>b</sup>Sarcoidosis Center, General Hospital of Chest Diseases of Athens “Sotiria”, Athens, Greece; <sup>c</sup>Department of Immunology and Histocompatibility, Laiko General Hospital, Athens, Greece; <sup>d</sup>Laboratory of Cardiogenomics, University Hospital Olomouc, Olomouc, Czech Republic; <sup>e</sup>Medical School, National and Kapodistrian University of Athens First Academic Department of Pneumology, Interstitial Lung Diseases Unit Hospital for Diseases of the Chest “Sotiria”, Athens, Greece

### ABSTRACT

**Background:** In sarcoidosis, the direction and intensity of immunological reactions involved in disease pathophysiology is affected by variation in the genes coding for effector and regulatory molecules with immune functions. This study, therefore, investigates polymorphic variants in genes involved in inflammation, immune reactions, and granuloma formation in context of their plausible association with sarcoidosis, with specific focus on Greek population.

**Methods:** A total of 18 single-nucleotide polymorphisms (SNPs) were genotyped in Greek patients with pulmonary sarcoidosis ( $n = 103$ ) and in healthy Greek control subjects ( $n = 100$ ) using multiplexed MassARRAY (MassARRAY<sup>®</sup>) iPLEX assay based on MALDI-TOF mass spectrometry.

**Results:** *TGF-β3* rs3917200\*G variant was associated with sarcoidosis (OR: 3.04 [95% CI: 1.98–4.69],  $p = 2.76 \times 10^{-7}$ ). Further, *ANXA11* rs1049550\*A variant was associated with sarcoidosis (OR: 0.59 [0.39–0.89],  $p = 0.01$ ).

**Conclusions:** This first study of genetic variation of immune-related genes in Greek patients with sarcoidosis brings to attention a novel disease ‘susceptibility’ factor: *TGF-β3* rs3917200\*G allele. It also confirms previously reported ‘protective’ association between sarcoidosis and functional variant *ANXA11* rs1049550\*A. Further work is required to validate these findings and to expand investigation of their plausible relationship with clinical course of the disease.

### ARTICLE HISTORY

Received 26 May 2020  
Accepted 16 June 2020

### KEYWORDS

*ANXA11*; genetic factors; Greek; polymorphisms; sarcoidosis; *TGF-β3*

## 1. Introduction

Sarcoidosis is a disease with features of altered immune response, often with multisystemic involvement. Presentation of this disease is heterogeneous; the mostly presented phenotype is pulmonary manifestation, which is widespread mostly in populations of European descent. Course of this disease is also very variable. In some patients, there is often speedy resolution but in approximately 15%–20% of cases, there is progress to one of the most dangerous courses, irreversible pulmonary fibrosis, where the risk of deterioration, even death is high [1–3].

Etiology of sarcoidosis is still unknown. The genetic factors, altered immunoregulation, and inflammation leading to granuloma formation have been suggested to have a crucial role in sarcoidosis pathogenesis [4]. Further, patients with sarcoidosis often present with hypercalcemia [1].

In this context, this study aimed to determine the implication of gene variants coding for immune-related molecules such as cytokines, proteins, and receptors involved in antigen presentation and immunoregulation processes, and also in calcium signalization, to characterize their possible roles as genetic factors sarcoidosis, including those contributing to disease course.

According to study in Greek centers enrolling 60% of the adult population [5], sarcoidosis is the leader between lung diseases (ILD) in Greece: one of five ILD cases presented with sarcoidosis [5]. As there has been rather limited information on immunogenetic background of Greek patients with sarcoidosis, we focused our study on this particular population.

## 2. Patients and methods

### 2.1. Patients and control subjects

This case-control study comprised 103 sarcoidosis cases (mean age  $\pm$  standard deviation [SD],  $56.6 \pm 11.6$  years; age range, 28–85 years; male [M]: female [F] ratio, 30:73) and 100 control subjects ( $43.5 \pm 11.9$  years 21–72 years, M:F, 30:70); all subjects were unrelated and of Greek origin. The patients were diagnosed according to the ATS, ERS, and WASOG statement on sarcoidosis [6] at the General Hospital for Chest Diseases, Athens; the healthy control subjects were enrolled at the Laiko General Hospital, Athens (Department of Immunology and Histocompatibility) as healthy blood donors. Both patients and control subjects provided informed consent with participation in the study, which

**Article highlights**

- We present the first detailed study of genetic variation of immune-related genes in Greek patients with sarcoidosis.
- We confirm previously reported 'protective' association between functional variant *ANXA11* rs1049550\*A (Arg230Cys).
- We report a novel, plausible 'susceptibility' factor - *TGF-β3* rs3917200\*G.
- Further, polymorphism rs1800629 in *TNFA* gene is associated with the more advanced CXR stages in our cohort of Greek sarcoidosis patients.
- Further work is required to validate these findings and to investigate in detail their relationship with clinical course of the disease.

was approved by institutional review boards. The study was carried out from 11 October 2017 to 13 November 2019.

Sarcoidosis cases were followed for a minimum of 2 years after their diagnosis to observe the development of the disease and response to the treatment. Patients were treated by corticosteroids ( $n = 57$ ); in some cases in combination with methotrexate ( $n = 5$ ) this therapy was used in all cases combined with pulmonary hypertension and also in some cases with extrapulmonary cardiac involvement or in inhalation form in combination with long-acting  $\beta_2$ -agonist ( $n = 2$ ); in five cases the treatment was stopped. The remaining 46 patients with sarcoidosis were without any treatment. After 2 years from the diagnosis, the patients were evaluated, i.e. for the disease status as follows: (a) stabilized patients ( $n = 87$ ), (b) remission ( $n = 13$ ), and (c) progression ( $n = 3$ ), for patients with deterioration of the condition. Two of three progressing patients had extrapulmonary cardiac involvement.

Patients were also evaluated based on the chest X-ray (CXR) stages (thoracic involvement) using Scadding's criteria: stage 1 that involved bilateral hilar lymphadenopathy (BHL) only ( $n = 29$ ), stage 2 for BHL with pulmonary infiltrates ( $n = 59$ ), stage 3 with parenchymal infiltrates only ( $n = 4$ ), and stage 4 with pulmonary fibrosis ( $n = 5$ ); in six patients the CXR stage was not determined.

Some patients have also extrathoracic manifestation such as cardiopathy based on magnetic resonance imaging findings ( $n = 12$ ), cutaneous sarcoidosis ( $n = 13$ ), uveitis ( $n = 8$ ), peripheral lymphadenopathy ( $n = 4$ ), arthropathy ( $n = 3$ ), hypercalcemia ( $n = 3$ ), erythema nodosum ( $n = 2$ ), pulmonary hypertension ( $n = 2$ ), and neurosarcoidosis ( $n = 1$ ).

## 2.2. Genotyping

DNA was isolated from peripheral blood using arrow automated extraction system (Isogen Life Science, PW De Meern, Utrecht, the Netherlands). For polymorphism analysis, a total of 18 plausible gene variants (Table 1) were genotyped using MassARRAY® method. These 18 gene variants were selected based on the previous reports for associated with sarcoidosis [7–14]. These polymorphisms were located in genes involved in antigen presentation, immunoregulation, calcium signalization, cytokines, or genes otherwise connected to immune functions.

Genotyping was performed using multiplexed MassARRAY iPLEX assay followed by MALDI-TOF mass spectrometry (Agena Bioscience, CA, USA). The steps for assay design using Assay Design Suite v2.0, polymerase chain reaction

**Table 1.** List of candidate single-nucleotide polymorphisms (SNPs) investigated in the study.

Gene	SNP ID	Minor allele	Chr.	Function
<i>ANXA11</i>	rs1049550	A	10	Ca <sup>2+</sup> dependent phospholipid-binding
<i>TNF-α</i>	rs1800629	A	6	Cytokine (inflammation)
<i>TGF-β2</i>	rs1891467	G	1	Cytokine (cell signaling and SMAD activation)
<i>GREM1</i>	rs1919364	G	15	Cytokine, regulating tissue differentiation
<i>HLA-DRA, HLA-DRB5</i>	rs1964995	G	6	Immunity and antigen presentation
<i>NOD2</i>	rs2066844	T	16	Stimulates an immune reaction
<i>BTNL2</i>	rs2076530	G	6	Regulates activated T cells
<i>HLA-DQA1</i>	rs2187668	T	6	Immunity and antigen presentation
<i>CACFD1</i>	rs3124765	T	9	Calcium channel activity
<i>C6orf10</i>	rs3129927	C	6	Immune diseases risk
<i>ATF6B</i>	rs3130288	A	6	ER stress and signal transduction
<i>HLA-DRA</i>	rs3135394	G	6	Immunity and antigen presentation
<i>TAP2</i>	rs3819717	A	6	Membrane transport and antigen presentation
<i>HLA-DRB1</i>	rs3830135	A	6	Immunity and antigen presentation
<i>TGF-β3</i>	rs3917165	T	14	Cytokine and lung development
	rs3917200	G	14	Cytokine and lung development
<i>HLA-DPB1</i>	rs9277357	G	6	Immunity and antigen presentation
<i>LRRC16A</i>	rs9295661	C	6	Megakaryocyte and platelet production

SNP ID, designation of the particular SNP within the reference sequence (rs) nomenclature; Chr., chromosomal localization; ER, endoplasmic reticulum.

(PCR) amplification of multiplexed iPLEX assay and MALDI-TOF MS-based genotyping have been described in our previous study [15]. In brief, 10 ng of template DNA was combined with the designed primers (Assay design suite v2.0; <https://www.mysequenom.com/Tools>) and master mix. The amplified targeted PCR products were cleaned by SAP enzymatic digestion, followed by a single base extension (SBE) reaction. The SBE product was applied on Spectro-chip by nanodispenser and genotyped by MALDI-TOF MS-based MassARRAY platform (Agena Bioscience, San Diego, CA, USA). In the assay, positive and negative controls were also included. Results were obtained by MassARRAY Typer 4.0.20 and manually validated using call cluster plot.

The single-nucleotide polymorphism (SNP) rs3917200 in the *TGF-β3* gene was genotyped by TaqMan assay using LightCycler 480 System (Roche, Branford, CT, USA). For this reaction, 10 ng of template DNA, LightCycler 480 Probes Master (Roche) and designed TaqMan SNP genotyping assay (ThermoFisher Scientific, USA) were used. The cycling profile was as follows: 5 min at 95°C; 45 cycles of 10 s at 95°C, 45 s at 60°C, 1 s at 72°C (with acquisition in each cycle), and then 30 s at 40°C. The data were obtained by endpoint analysis.

## 2.3. Statistical analysis

The association analysis was performed using allelic model by two-tailed Fisher exact probability test to estimate  $p$ -value,

odds ratio (OR) and 95% confidence interval (CI). Each SNP was tested for Hardy–Weinberg equilibrium (HWE) ( $> 0.01$ ) using Pearson's chi-square ( $\chi^2$ ) test or Fisher's exact test. The statistical significance was considered as  $p < 0.05$ . Bonferroni correction of multiple comparisons was performed as  $p_{\text{corr}} = p \times 18$ .

The statistical analysis was performed using <http://shesis-plus.bio-x.cn/SHEsis.html> and <https://www.snpstats.net/start.htm> in combination with <http://vassarstats.net/odds2x2.html>.

The calculation of the statistical power of the study was performed using <http://osse.bii.a-star.edu.sg/calculation2.php>.

### 3. Results

We analyzed 18 SNPs in 103 Greek sarcoidosis cases and in 100 healthy control subjects. One polymorphism, namely rs3819717 in *TAP2* gene, was not found in HWE in the group of control subjects and it was, therefore, excluded from further analysis.

The polymorphism rs3917200\*G in the gene for cytokine *TGF- $\beta$ 3* was associated with sarcoidosis on primary level (OR: 3.04 [95% CI: 1.98–4.69],  $p = 2.76 \times 10^{-7}$ ) and also after correction for multiple comparison ( $p_{\text{corr}} = 4.97 \times 10^{-6}$ ) (see Table 2). The statistical power of the study sample size for the rs3917200\*G was 96.2%.

The polymorphism rs1049550\*A located in the gene *ANXA11* was found associated with sarcoidosis (OR: 0.59 [95% CI: 0.39–0.89],  $p = 0.01$ ). However, this association did not attain significance after correction for multiple comparison ( $p_{\text{corr}} > 0.05$ ) (see Table 2).

Other 16 investigated polymorphisms were not associated with sarcoidosis compared to healthy controls.

For subanalysis based on CXR stages, the patients were subdivided into two groups: the first one comprising patients with CXR-1, the second one consisting of patients with CXR-2 and higher. Next, these groups were compared with healthy control group. An association on primary level as well as after Bonferroni correction was observed between the polymorphism rs3917200\*G in *TGF- $\beta$ 3* and patients with CXR-1 (OR: 3.26 [95%

**Table 3.** Allelic model of association for sarcoidosis subgroup analysis, based on the chest X-ray (CXR) stage in Greek patients with sarcoidosis.

Gene	SNP	CXR Stg. 1 (n = 29) vs. controls (n = 100)		CXR Stg. 2–4 (n = 68) vs. controls (n = 100)	
		p-value	OR	p-value	OR
<i>ANXA11</i>	rs1049550	0.29	0.69	<b>0.01</b>	0.53
<i>TNFA</i>	rs1800629	0.57	1.34	<b>0.02</b>	2.50
<i>TGF-<math>\beta</math>2</i>	rs1891467	0.63	1.18	0.71	0.89
<i>GREM1</i>	rs1919364	0.30	1.37	0.65	0.89
<i>HLA-DRA, HLA-DRB5</i>	rs1964995	0.10	0.34	0.61	0.79
<i>NOD2</i>	rs2066844	0.24	2.37	0.76	1.24
<i>BTNL2</i>	rs2076530	0.64	1.18	0.47	0.84
<i>HLA-DQA1</i>	rs2187668	1.00	0.84	0.84	1.12
<i>CACFD1</i>	rs3124765	0.43	1.38	0.19	1.48
<i>C6orf10</i>	rs3129927	0.32	2.32	0.10	3.58
<i>ATF6B</i>	rs3130288	0.32	2.32	0.16	3.10
<i>HLA-DRA</i>	rs3135394	0.32	2.32	0.10	3.58
<i>HLA-DRB1</i>	rs3830135	0.53	0.54	0.63	0.72
<i>TGF-<math>\beta</math>3</i>	rs3917165	<b><math>5.00 \times 10^{-3}</math></b>	4.96	0.56	1.48
<i>TGF-<math>\beta</math>3</i>	rs3917200	<b><math>1.74 \times 10^{-4}</math></b> §	3.26	<b><math>9.52 \times 10^{-6}</math></b> †	2.95
<i>HLA-DPB1</i>	rs9277357	0.60	0.79	0.53	1.20
<i>LRRC16A</i>	rs9295661	1.00	NA	0.40	2.22

§SNP rs3917200 CXR Stg. 1 vs. controls  $p_{\text{corr}} = 3.13 \times 10^{-3}$ ; †SNP rs3917200 CXR Stg. 2–4 vs. controls  $p_{\text{corr}} = 1.71 \times 10^{-4}$ ; NA: not analyzed (in one of the analyzed groups there were present just homozygotes for one allele, not for the other allele and no heterozygotes); significant values are in bold.

CI: 1.77–6.02],  $p_{\text{corr}} = 3.13 \times 10^{-3}$ ) and also with CXR-2 and higher (OR: 2.95 [95% CI: 1.84–4.75],  $p_{\text{corr}} = 1.71 \times 10^{-4}$ ) (see Table 3).

Further an association on primary level was observed between SNP rs3917165\*T in *TGF- $\beta$ 3* gene (OR: 4.96 [95% CI: 1.65–14.95],  $p = 5.00 \times 10^{-3}$ ) and patients with CXR-1, compared to healthy control group (Table 3).

In the case of polymorphism rs1049550 (*ANXA11*), protective association was found with CXR-2 and higher (OR: 0.53 [95% CI: 0.33–0.84],  $p = 0.01$ ); no association of this polymorphism was observed in the group of patients with CXR-1 (Table 3).

Finally, an association was observed between SNP rs1800629\*A in *TNFA* gene (OR: 2.50 [95% CI: 1.20–5.21],  $p = 0.02$ ) and patients with CXR-2 and higher, compared to healthy control group (Table 3).

Other subgroups based on distinct clinical phenotypes, presence of extrathoracic involvement could not be analyzed due to small numbers of cases.

**Table 2.** Allelic model of association for sarcoidosis risk in Greek population.

Sarcoidosis (n = 103) vs. controls, (n = 100)			
Gene	SNP	p	OR [95% CI]
<b><i>ANXA11</i></b>	rs1049550	<b>0.01</b>	<b>0.59 [0.39–0.89]</b>
<i>TNFA</i>	rs1800629	0.06	1.99 [0.99–4.01]
<i>TGF-<math>\beta</math>2</i>	rs1891467	1.00	1.00 [0.66–1.53]
<i>GREM1</i>	rs1919364	0.84	1.04 [0.70–1.54]
<i>HLA-DRA, HLA-DRB5</i>	rs1964995	0.21	0.64 [0.34–1.20]
<i>NOD2</i>	rs2066844	0.60	1.48 [0.52–4.23]
<i>BTNL2</i>	rs2076530	0.92	0.98 [0.65–1.48]
<i>HLA-DQA1</i>	rs2187668	1.00	0.97 [0.47–1.99]
<i>CACFD1</i>	rs3124765	0.16	1.47 [0.88–2.43]
<i>C6orf10</i>	rs3129927	0.14	3.00 [0.80–11.25]
<i>ATF6B</i>	rs3130288	0.22	2.68 [0.70–10.25]
<i>HLA-DRA</i>	rs3135394	0.14	3.00 [0.80–11.25]
<i>HLA-DRB1</i>	rs3830135	0.51	0.71 [0.29–1.72]
<i>TGF-<math>\beta</math>3</i>	rs3917165	0.11	2.31 [0.87–6.14]
<b><i>TGF-<math>\beta</math>3</i></b>	rs3917200	<b><math>2.76 \times 10^{-7}</math></b> §	<b>3.04 [1.98–4.69]</b>
<i>HLA-DPB1</i>	rs9277357	0.91	1.04 [0.66–1.63]
<i>LRRC16A</i>	rs9295661	1.00	1.45 [0.24–8.76]

§SNP rs3917200  $p_{\text{corr}} = 4.97 \times 10^{-6}$ ; OR: odds ratio; CI: confidence interval; Bold values are significant.

### 4. Discussion

We analyzed a panel of 18 SNPs previously reported to impact events having the main role in sarcoidosis pathogenesis such as immune and inflammatory processes, fibrotization and calcium signalization. In this context, this is the first study of genetic variation of functional immune-related gene variants performed in Greek patients with sarcoidosis. The main observation was the strong association between sarcoidosis and the rs3917200\*G variant in *TGF- $\beta$ 3* gene. Further, this study confirmed that *ANXA11* rs1049550\*A variant, previously reported in other population, was associated with sarcoidosis as a whole also in our Greek patients.

*TGF- $\beta$ 3* gene has been suggested to have antifibrotic effect in the wound healing [16,17], the polymorphisms in this gene can

disrupt this mechanism. In this context, polymorphism rs3917200\*G in the gene *TGF-β3* was previously described in association with fibrotic processes [11]. In this study from German population, a trend was observed connecting minor allele in this SNP as a factor contributing to formation of lung fibrosis in sarcoidosis patients [11]. Similar observations regarding this *TGF-β3* SNP were reported in a Dutch sarcoidosis population [13]. In our study, we found SNP rs3917200 in *TGF-β3* gene strongly associated with sarcoidosis among Greek sarcoidosis cases, with more pronounced occurrence of G allele in the patients. We may, therefore, speculate that the SNP rs3917200 in *TGF-β3* might be a possible marker for progressing disease in sarcoidosis, characterized by involvement of fibrotization processes [3]. Extended study with increased proportion of the most advanced cases is however required to confirm this speculation.

*ANXA11* gene is involved in calcium signalization and apoptosis. It effects granuloma formation and also the maintenance of granulomatous inflammation [2]: disruption of this mechanism could counter-effect sarcoidosis granuloma development and thus may possibly contribute to prevention of sarcoidosis onset. The non-synonymous mutation rs1049550 located in exon 6 of *ANXA11* results in amino acid change from arginine to cytosine localized on evolutionary conserved position 230 of the first annexin domain [18], with subsequent *ANXA11* protein structure disruption. The protective (i.e. OR < 1) association of this polymorphism with sarcoidosis was first reported within a genome-wide association study performed in German population [18], and was subsequently replicated in Czech [19], later also Portuguese [20] and African and European Americans [21] population, in the latter an association for the disease persistence was observed.

We also observed an association of the polymorphism rs1800629 in *TNFA* gene with more advanced CXR stages (CXR 2–4), as compared to healthy controls. This association between this particular polymorphism and sarcoidosis in general has been previously reported in other, Czech population [14] and also in a meta-analysis [22]. However, these studies did not associate this SNP with the advanced CXR stages of sarcoidosis. The TNF-α as product of *TNFA* gene has an important role in the pathogenesis of sarcoidosis in granuloma formation and polymorphism rs1800629 could potentially influence the transcription/production of this gene and thus contribute to disease development [23,24].

One may argue that the limitation of the study is a low number of patients investigated. However, this is relative due to a lower prevalence of this disease in the population of Greece. The sarcoidosis prevalence in Greece is 5.89/100 000 [5] compared to the global sarcoidosis prevalence which ranges between 4.7 and 64/100 000 [3]. Further work, also in other populations, namely of same or similar ethnic origins is, therefore, required to validate these findings – this could apply for example to Greek Cypriots or the Greeks from the main diasporas around the world [25]. Greater groups will also enable to investigate a possible relationship of genetic markers with the clinical course of the disease. For future investigations it will be also interesting to explore if any of identified gene variants

correlate with the biomarkers that already have been practical use such as ACE, IL-2 r or chitotriosidase [26,27].

## 5. Conclusion

This case-control study in Greek population confirmed the protective association of functional variant *ANXA11* rs1049550\*A with sarcoidosis previously reported for Germans and Czechs. Polymorphism rs1800629 in *TNFA* gene was also associated with more advanced CXR stages in our cohort of Greek sarcoidosis patients. Most importantly, we identified a novel plausible disease risk factor, *TGF-β3* rs3917200\*G allele, which represents a target for replication studies in Greek and/or other populations.

## Acknowledgments

The authors thank Prof. Dr. Joannis Mytilineos (Ulm, Germany) for his kind advice regarding Greek healthy control population.

## Author contributions

MP, AR, KA, MCh, AK, EB, and DB contributed to conception and design, and provision of study materials or patients.. KS, AK, DB, LK, and VZ contributed to collection, assembly of data, data analysis, and interpretation. MP, KS, AR, KA, MCh, AK, EB, DB, AK, LK, and VZ contributed to manuscript writing and final approval of the manuscript.

## Declaration of interest

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

## Reviewer disclosures

Peer reviewers on this manuscript have no relevant financial or other relationships to disclose.

## Funding

This work was supported by Palacky University [grant number IGA PU LF 2020\_004]; Czech Ministry of Health [grant number NV18-05-00134]; and Czech government [grant number CZ.02.1.01/0.0/0.0/16\_019/0000868, ENOCH].

## Data availability statement

The data that support the findings of this study are available from the corresponding author [martin.petrek2@fnol.cz](mailto:martin.petrek2@fnol.cz), on reasonable request.

## Previous presentations

Part of this work has been presented as an abstract at the International Conference on Sarcoidosis and Interstitial Lung Diseases – WASOG 2018 (7–9 June), Heraklion, Greece.



## References

Papers of special note have been highlighted as either of interest (\*) or of considerable interest (\*\*\*) to reader

- Dubrey S, Shah S, Hardman T, et al. Sarcoidosis: the links between epidemiology and aetiology. *Postgrad Med J*. 2014;90(1068):582–589.
- Mirsaeidi M, Gidfar S, Vu A, et al. Annexins family: insights into their functions and potential role in pathogenesis of sarcoidosis. *J Transl Med*. 2016;14:89.
- Valeyre D, Prasse A, Nunes H, et al. Sarcoidosis. *Lancet*. 2014;383(9923):1155–1167.
- A paper from The Lancet seminars series reviewing sarcoidosis epidemiology, pathophysiology, diagnosis and treatment.**
- Mortaz E, Masjedi MR, Tabarsi P, et al. Immunopathology of sarcoidosis. *Iran J Allergy Asthma Immunol*. 2014;13(5):300–306.
- Karakatsani A, Papakosta D, Rapti A, et al. Epidemiology of interstitial lung diseases in Greece. *Respir Med*. 2009;103(8):1122–1129.
- Hunninghake GW, Costabel U, Ando M, et al. Statement on sarcoidosis. *Am J Respir Crit Care Med*. 1999;160(2):736–755.
- Morais A, Lima B, Alves H, et al. Associations between sarcoidosis clinical course and ANXA11 rs1049550 C/T, BTNL2 rs2076530 G/A, and HLA class I and II alleles. *Clin Respir J*. 2018;12(2):532–537.
- Rivera NV, Ronninger M, Shchetynsky K, et al. High-density genetic mapping identifies new susceptibility variants in sarcoidosis phenotypes and shows genomic-driven phenotypic differences. *Am J Respir Crit Care Med*. 2016;193(9):1008–1022.
- An extensive study of susceptibility genes in sarcoidosis showing that clinical phenotype may in part reflect the genotype; as exemplified by proposal of distinct molecular mechanisms for Löfgren syndrome and non-Löfgren sarcoidosis.**
- Fischer A, Ellinghaus D, Nutsua M, et al. Identification of immune-relevant factors conferring sarcoidosis genetic risk. *Am J Respir Crit Care Med*. 2015;192(6):727–736.
- A comprehensive study of immune-related genes as susceptibility factors for sarcoidosis which besides HLA-B, HLA-DPB1, BTNL2, IL23R loci confirmed ANXA11 polymorphism as a sarcoidosis risk factor, the patients were enrolled from several ethnically different populations.**
- Heron M, Van Moorsel CHM, Grutters JC, et al. Genetic variation in GREM1 is a risk factor for fibrosis in pulmonary sarcoidosis. *Tissue Antigens*. 2011;77(2):112–117.
- Pabst S, Fränken T, Schönau J, et al. Transforming growth factor- $\beta$  gene polymorphisms in different phenotypes of sarcoidosis. *Eur Respir J*. 2011;38(1):169–175.
- Sato H, Williams HRT, Spagnolo P, et al. CARD15/NOD2 polymorphisms are associated with severe pulmonary sarcoidosis. *Eur Respir J*. 2010;35(2):324–330.
- Kruit A, Grutters JC, Ruven HJT, et al. Transforming growth factor- $\beta$  gene polymorphisms in sarcoidosis patients with and without fibrosis. *Chest*. 2006;129(6):1584–1591.
- Mrázek F, Holla LI, HutYROVA B, et al. Association of tumour necrosis factor- $\alpha$ , lymphotoxin- $\alpha$  and HLA-DRB1 gene polymorphisms with Löfgren's syndrome in Czech patients with sarcoidosis. *Tissue Antigens*. 2005;65(2):163–171.
- Kishore A, Žižková V, Kocourková L, et al. A dataset of 26 candidate gene and pro-inflammatory cytokine variants for association studies in idiopathic pulmonary fibrosis: frequency distribution in normal Czech population. *Front Immunol*. 2015;6:476.
- The methodology for single nucleotide polymorphism genotyping by MassArray technology, which was used also in the present study, is described here in detail.**
- Gilbert RWD, Vickaryous MK, Vilorio-Petit AM. Signalling by transforming growth factor beta isoforms in wound healing and tissue regeneration. *J Dev Biol*. 2016;4(2):21.
- Shah M, Foreman DM, Ferguson MW. Neutralisation of TGF- $\beta$ 1 and TGF- $\beta$ 2 or exogenous addition of TGF- $\beta$ 3 to cutaneous rat wounds reduces scarring. *J Cell Sci*. 1995;108(Pt 3):985–1002.
- Hofmann S, Franke A, Fischer A, et al. Genome-wide association study identifies ANXA11 as a new susceptibility locus for sarcoidosis. *Nat Genet*. 2008;40(9):1103–1106.
- Mrázek F, Stahelova A, Kriegova E, et al. Functional variant ANXA11 R230C: true marker of protection and candidate disease modifier in sarcoidosis. *Genes Immun*. 2011;12(6):490–494.
- Morais A, Lima B, Peixoto M, et al. Annexin A11 gene polymorphism (R230C variant) and sarcoidosis in a Portuguese population. *Tissue Antigens*. 2013;82(3):186–191.
- Levin AM, Iannuzzi MC, Montgomery CG, et al. Association of ANXA11 genetic variation with sarcoidosis in African Americans and European Americans. *Genes Immun*. 2013;14(1):13–18.
- Medica I, Kastrin A, Maver A, et al. Role of genetic polymorphisms in ACE and TNF- $\alpha$  gene in sarcoidosis: a meta-analysis. *J Hum Genet*. 2007;52(10):836–847.
- Ziegenhagen MW, Benner UK, Zissel G, et al. Sarcoidosis: TNF- $\alpha$  release from alveolar macrophages and serum level of sIL-2R are prognostic markers. *Am J Respir Crit Care Med*. 1997;156(5):1586–1592.
- Ziegenhagen MW, Müller-Quernheim J. The cytokine network in sarcoidosis and its clinical relevance. *J Intern Med*. 2003;253(1):18–30.
- Wikipedia [Internet]. Greek diaspora; 2020 [last updated 18 May 2020]. cited 2020 May 19]. Available from: [https://en.wikipedia.org/wiki/Greek\\_diaspora](https://en.wikipedia.org/wiki/Greek_diaspora)
- Culver DA, Judson MA. New advances in the management of pulmonary sarcoidosis. *BMJ*. 2019;367:l5553.
- An extensively referenced review of advances in sarcoidosis diagnosis, prognosis and treatment which also discusses role of serum and other emerging biomarkers including genetics. Research questions pertinent for future research on sarcoidosis pathomechanisms and management are posed.**
- Cameli P, Gonnelli S, Bargagli E, et al. The role of urinary calcium and chitotriosidase in a cohort of chronic sarcoidosis patients. *Respiration*. 2020;99(3):207–212.

OPEN

# HLA-A, -B, -C, -DRB1, -DQA1, and -DQB1 allele and haplotype frequencies defined by next generation sequencing in a population of East Croatia blood donors

Stana Tokić<sup>1\*</sup>, Veronika Žižkova<sup>4</sup>, Mario Štefanić<sup>2\*</sup>, Ljubica Glavaš-Obrovac<sup>1</sup>, Saška Marczi<sup>3</sup>, Marina Samardžija<sup>3</sup>, Katerina Sikorova<sup>4</sup> & Martin Petrek<sup>4\*</sup>

Next-generation sequencing (NGS) is increasingly used in transplantation settings, but also as a method of choice for in-depth analysis of population-specific HLA genetic architecture and its linkage to various diseases. With respect to complex ethnic admixture characteristic for East Croatian population, we aimed to investigate class-I (HLA-A, -B, -C) and class-II (HLA-DRB1, -DQA1, -DQB1) HLA diversity at the highest, 4-field resolution level in 120 healthy, unrelated, blood donor volunteers. Genomic DNA was extracted and HLA genotypes of class I and DQA1 genes were defined in full-length, -DQB1 from intron 1 to 3' UTR, and -DRB1 from intron 1 to intron 4 (Illumina MiSeq platform, Omixon Twin algorithms, IMGT/HLA release 3.30.0\_5). Linkage disequilibrium statistics, Hardy-Weinberg departures, and haplotype frequencies were inferred by exact tests and iterative Expectation-Maximization algorithm using PyPop 0.7.0 and Arlequin v3.5.2.2 software. Our data provide first description of 4-field allele and haplotype frequencies in Croatian population, revealing 192 class-I and class-II alleles and extended haplotypic combinations not apparent from the existing 2-field HLA reports from Croatia. This established reference database complements current knowledge of HLA diversity and should prove useful in future population studies, transplantation settings, and disease-associated HLA screening.

Croatia is a Mediterranean, crescent-shaped south European country bordering Slovenia in the northwest, Hungary in the northeast, Serbia in the east, Bosnia and Herzegovina and Montenegro in the southeast, and Italy along the maritime border. Croatia consists of three major geomorphologic areas, which can be further broken down into five traditional districts based on history, topography, and economy; Istria and Dalmatia in the northern and southern Croatian littoral, Gorski Kotar in country's mountainous area, central continental Croatia, and Slavonia in the Pannonian basin in the east (Fig. 1). Slavonia territory was originally populated by the southern branch of the Indo-European Slavic populations in the 7<sup>th</sup> century<sup>1</sup>, and has been a witness of significant population admixture ever since, including the Hungarian migration to Slavonia in 10<sup>th</sup> century, and the influx of Islamic and Orthodox Balkan and Asian populations during the Ottoman conquest in 16<sup>th</sup> century, causing at the same time, the continuous shift of Catholics from Bosnia to Slavonia during several centuries<sup>2</sup>. Under the auspices of Habsburg monarchy, the settlement of Germans and Austrians in Slavonian urban areas peaks between 18<sup>th</sup> and 19<sup>th</sup> century, while Orthodox Vlachs from Bosnia, immigrating Czechs, Slovaks, Ukrainians, Italians, and

<sup>1</sup>Department of Medical Chemistry, Biochemistry and Clinical Chemistry, Faculty of Medicine, University of Osijek, J. Huttlera 4, HR-31000, Osijek, Croatia. <sup>2</sup>Department of Nuclear Medicine and Oncology, Faculty of Medicine, University of Osijek, J. Huttlera 4, HR-31000, Osijek, Croatia. <sup>3</sup>Department of Laboratory Diagnostics and Clinical Transfusion Medicine, Clinical Institute of Transfusion Medicine, Osijek University Hospital, J. Huttlera 4, HR-31000, Osijek, Croatia. <sup>4</sup>Department of Pathological Physiology, Faculty of Medicine and Dentistry, Palacký University, Hnevotinska 3, 775 15, Olomouc, Czech Republic. \*email: [stokic@mefos.hr](mailto:stokic@mefos.hr); [mstefanic@mefos.hr](mailto:mstefanic@mefos.hr); [Martin.Petrek@fnol.cz](mailto:Martin.Petrek@fnol.cz)



**Figure 1.** Map of geographical location of Croatia with representation of Dalmatia, Istria, Gorski Kotar, central and eastern Croatian regions, five of which (colored in light grey) participated in sample collection. In detail: Osijek-Baranja county (n = 80), Vukovar-Syrmia county (n = 22), Brod-Posavina county (n = 9), Požega-Slavonia county (n = 4) and Virovitica-Podravina county (n = 5)<sup>56</sup>.

Croatians from Gorski Kotar populate rural settlements<sup>3</sup>. These historic and more recent 20<sup>th</sup> century migration events, encompassing emigration of Germans and Austrians from Slavonia, and the settlement of Balkan War veterans from Serbia, Croatian immigrants from Dalmatia, Herzegovina, and most recently also from north Bosnia, have shaped the genetic diversity of East Croatian population.

Population migration and admixture have an important role in the evolution and diversifying selection of hypervariable human leukocyte antigen (HLA) molecules involved in innate and adaptive immune responses<sup>4</sup>. Two types of HLA molecules, class I and class II, are codominantly expressed on the surface of nucleated and antigen presenting cells, and facilitate peptide antigen presentation, self-tolerance and immune surveillance through activation of CD8 and CD4 T lymphocytes, respectively. The biological function of HLA molecules is correlated with high variability of human major histocompatibility (MHC) proteins, which are encoded by the highly polymorphic gene complex located in the short arm of chromosome 6<sup>5</sup>. Up to date, 18,691 class I, 7,065 class II and 202 non-HLA allelic variants (IPD-IMGT/HLA database, release 3.38., <https://www.ebi.ac.uk/ipd/imgt/hla/stats.html>)<sup>6,7</sup> have been identified in various human populations and the highest degree of variability was noticed within exon regions encoding phenotype and peptide-binding preferences of class I (HLA-A, -B, -C) and class II (HLA-DR, -DQ) HLA molecular isotypes.

Considering important role of HLA molecules in immune response, the HLA gene complex has been extensively studied in the context of allogeneic transplantation, inflammatory, infectious and autoimmune disease associations<sup>8</sup>. The worldwide application of molecular, PCR-based techniques in the clinical settings (sequence-specific primer (SSP), sequence-specific oligonucleotide (SSO), Sanger sequencing-based typing (SBT) and most recently, next-generation sequencing (NGS)), also enabled development of the population specific HLA typing data repositories (The Allele Frequency Net Database)<sup>9</sup>, and enhanced assessments of populations migration, diversity and regional HLA specificity. Significant differences in frequency of common and well documented alleles recently described within European sub-regions, support the role of geographical dispersion

in development of region specific HLA heritage<sup>10</sup>. The inter- and intra-population HLA comparisons made today are however, mostly constrained to a partial, exon description of HLA genetic variability (1<sup>st</sup> and 2<sup>nd</sup> fields of HLA nomenclature), whereas synonymous (3<sup>rd</sup> field) and non-coding (4<sup>th</sup> field) nucleotide variations remain largely unexplored. In such settings, it remains challenging to accurately build contiguous long haplotypes of uniformly high resolution even for the largest sample cohorts. The implementation of NGS technologies in HLA research and routine clinical work, enables however, elucidation of full-length HLA gene sequences, permitting an in-depth characterisation of population HLA diversity. In this context NGS can overcome limitations of traditional typing techniques, thereby sustaining optimal HLA matching of donor-recipient pairs for organ and particularly, haematopoietic stem cell transplantation (HSCT)<sup>11</sup>, improved estimates of population structure and HLA associated disease risk<sup>12–14</sup>, and better understanding of demographic history and geographic origin of a given population<sup>15</sup>.

Up to date, HLA allelic and haplotype diversity of a general Croatian population, irrespective of specific geographical preferences, was estimated in several large cohorts originating from Croatia<sup>16–19</sup> and emigrant population in Germany<sup>20</sup>, as well as few, more isolated populations from particular geographical locations such as island Krk<sup>21</sup>, island Hvar<sup>22</sup>, Istrian city of Rijeka<sup>23</sup> and Gorski Kotar<sup>24</sup>. Moreover, several non-frequent, rare and very rare HLA-A, -B and -DRB1 alleles and haplotypes have been characterized among the unrelated volunteer donors from the Croatian Bone Marrow Donor Registry (CBMDR)<sup>25</sup>. In addition, DPB1 allelic diversity was recently evaluated in 82 Croatian patients who underwent HSCT<sup>26</sup>. These previous studies were however, based on lower resolution (1<sup>st</sup> and 2<sup>nd</sup> field) HLA typing of selected HLA loci in individuals originating from various Croatian regions, providing partial insight into the HLA diversity of a general, but not East Croatia population.

The aim of the present study was thus to investigate and describe extended allelic and haplotype diversity of HLA-A, -B, -C, -DQA1, -DQB1 and DRB1 loci at high-resolution, 4<sup>th</sup> field level, using high-throughput NGS technique for HLA typing of 120 healthy, unrelated blood donor volunteers from east Croatia.

## Results

**NGS sequencing results.** We evaluated 120 donors (120 donors × 6 loci × 2 alleles = 1440 alleles), 9 of whom were excluded from further analysis due to low-performing samples on quality control check [low read count (≤2500 bp for class I and ≤5000 bp for class II genes) and/or low key exon coverage depth (≤30)]. The coverage of each locus in the remaining samples (111 donors, 1332 alleles) was calculated by Twin as the percentage of gene regions covered by reads compared to the whole allele sequence (coverage %). The Omixon Holotype primer positioning allowed only partial amplification of 3'UTRs and 5'UTRs, and covered exons 2–6 of DQB1 and exons 2–4 of DRB1 loci (Fig. 2).

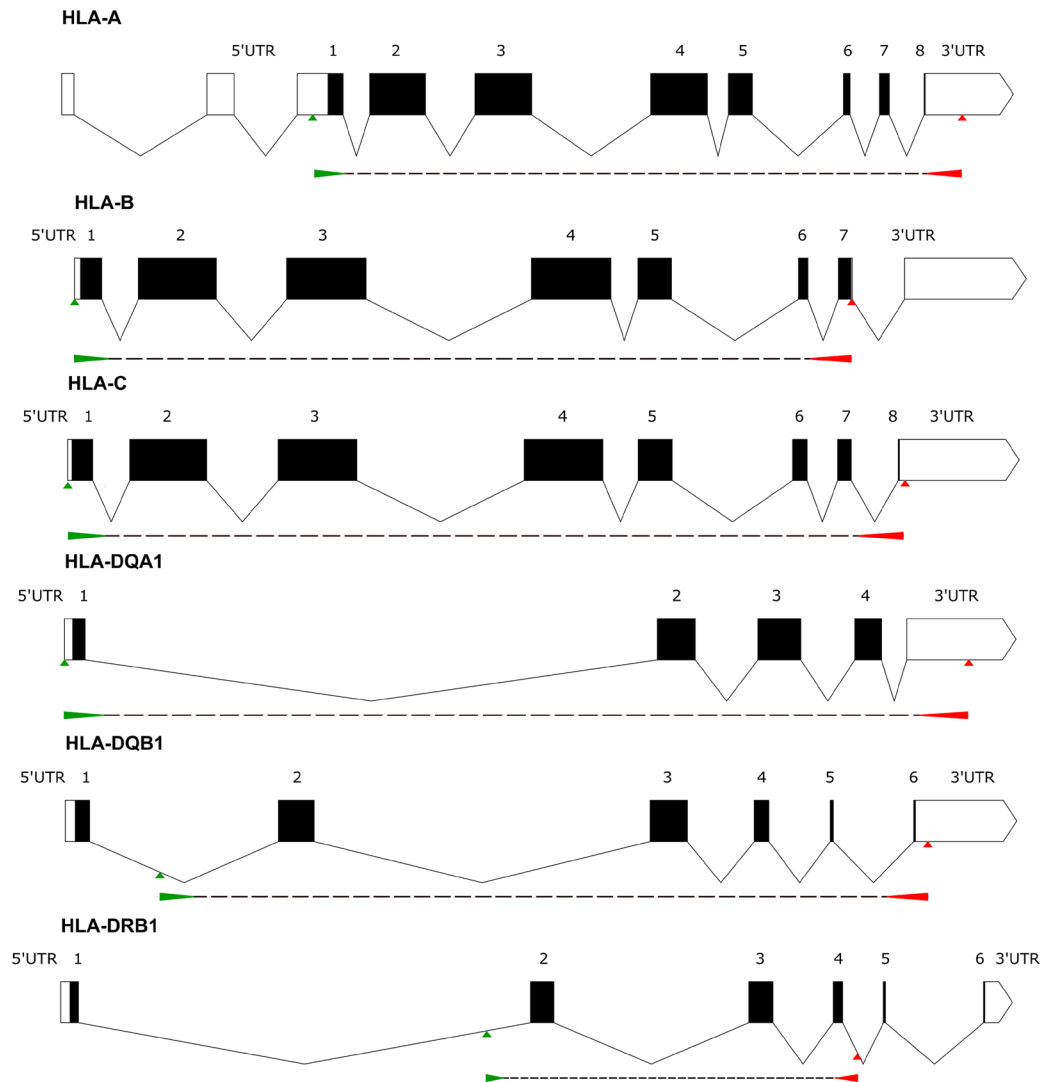
Ambiguous allele calls, which remained unresolvable due to inherent product limitations of the Omixon Holotype Kit (missing data on SNPs or INDEL variations within the unsequenced 3' UTR, 5' UTR and intron 1/exon 1 regions), were reported as ambiguous (i.e. DQB1\*06:01:01/15), or up to the third field level only (i.e. DQB1\*05:03:01), and are enlisted in Supplementary Table 1. Nevertheless, all amplified exon/intron regions of each allele in the remaining samples were fully covered (detection %), with an average coverage depth of >140 reads per nucleotide position (Supplementary Table 2). On average, 54359 (Supplementary Table 3) high-quality reads were produced per sample, of which 91% were subsequently used for final consensus generation after removing noise reads and PCR crossover artefacts. The average fragment size was 259 bp, and the average read length 208 bp.

**Linkage disequilibrium and HWE estimates.** Genotype frequencies of HLA-A, -B, -C, -DRB1, -DQA1 and DQB1 loci did not deviate from Hardy-Weinberg expectations (Supplementary Table 4). Strong linkage was confirmed between class I, HLA-A, -B, and -C, as well as class II, HLA-DQA1, -DQB1 and DRB1 loci (Supplementary Table 5 and Fig. 3).

**Class I allelic frequencies.** The observed frequencies of class I HLA alleles in eastern Croatian population are reported in Table 1. HLA-A genotyping analysis uncovered 25 different alleles in our sample population, five of which with frequency ≥10%. The most commonly observed *HLA-A\*02:01:01* group alleles (32.43%) were followed by *HLA-A\*01:01:01:01* (12.16%), *HLA-A\*03:01:01:01* (11.71%), *HLA-A\*24:02:01:01* (11.71%) and *HLA-A\*11:01:01:01* (9.91%). Together, these allelic variants comprised 77.47% of HLA-A allelic diversity in our sample population. Nearly half of the tested individuals (n = 59) were positive for *HLA-A\*02:01:01* allelic variants, 13 of which were homozygous carriers.

As expected, the highest polymorphism was observed within the HLA-B region, encompassing 39 different allelic specificities. The *HLA-B\*51:01:01* (8.56%), *HLA-B\*35:01:01* (8.11%) and *HLA-B\*08:01:01* (8.11%) were the most frequent alleles, and together with the *HLA-B\*07:02:01* (6.76%), *HLA-B\*35:03:01* (6.76%) and *HLA-B\*18:01:01* (5.41%) accounted for majority (43.7%) of HLA-B allelic diversity in our population. Notably, *HLA-B\*15* (gene frequency, 6.75%), *HLA-B\*35* (17.57%) and *HLA-B\*44* (9.91%) allelic families exhibited the highest variability, each encompassing at least 4 allelic members with different peptide binding motifs. Of interest, two rare alleles (<http://www.allelefrequencies.net>, Rare allele detector, score < 4) of undefined C/WD status were also uncovered, namely *HLA-B\*15:01:06* (0.45%) and *HLA-B\*44:02:01:03* (0.45%).

The HLA-C genotyping uncovered 26 different allelic specificities within 13 allelic families. The most polymorphic was *HLA-C\*07* group with 6 alleles, covering 27.48% of HLA-C allelic variability. Moreover, 50 subjects, were positive for either *HLA-C\*07:01:01* G (14.41%) or *HLA-C\*07:02:01* G (8.11%) allelic variants<sup>6,7</sup>. Nonetheless, the most common individual HLA-C representative was the *HLA-C\*04:01:01* (13.96%) allele, followed by *HLA-C\*02:02:02* (9.46%), *HLA-C\*12:03:01* (7.21%) and *HLA-C\*03:03:01* (5.41%). Rare, but well-documented *HLA-C\*07:01:09* allele of undefined C/WD status was also found.

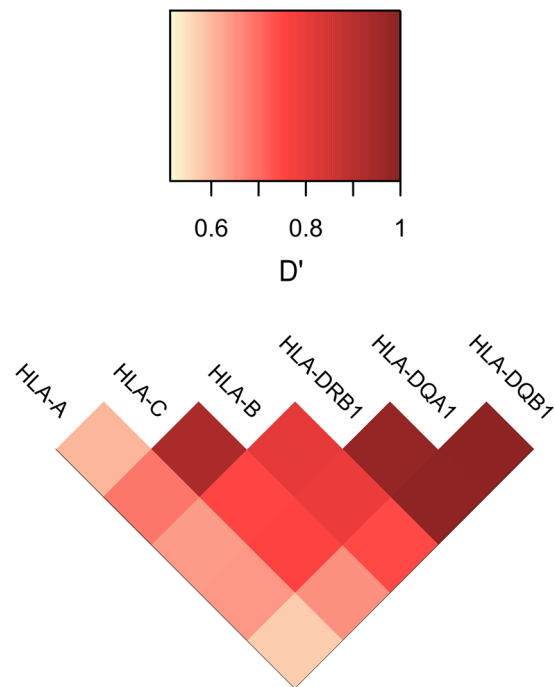


**Figure 2.** Schematic representation of HLA-A, -B, -C, -DQA1, -DQB1 and -DRB1 gene regions covered by Omixon Hologroup PCR primers are depicted using Exon-Intron Graphic Maker v.4. (<http://wormweb.org/exonintron>)<sup>57</sup>. Start positions of forward and reverse primers are marked by green and red arrows, respectively.

**Class II allelic frequencies.** Detailed analysis of class II HLA alleles is presented in Table 2. Among class II genes, DRB1 exhibited greatest allelic diversity through 40 alleles, of which *HLA-DRB1\*16:01:01* (13.51%) and *HLA-DRB1\*01:01:01* (10.81%) were most commonly present (>10%). Of interest, *HLA-DRB1\*16:01:01* allele was twice as frequent as the *HLA-DRB1\*15:01:01* (6.76%), or *HLA-DRB1\*03:01:01:01* (5.41%). Together with *HLA-DRB1\*07:01:01* (9.91%) and *HLA-DRB1\*11:01:01:01* (4.5%), these 6 DRB1 alleles were responsible for 50.9% of DRB1 allelic variability. Notably, *HLA-DRB1\*16* (13.96%) and *HLA-DRB1\*11* (13.5%) were the most common allelic families in our donor population. One of the DRB1\*11 family members was a rare DRB1\*11:01:08 allele, spotted only once. The highest individual variability, however, was detected within the DRB1\*04 allelic group, which was comprised out of 9 different allelic members, together covering 8.11% of DRB1 allelic variability in our population.

The DQA1 analysis uncovered 37 individual alleles, most common being *HLA-DQA1\*01:02:02* (15.77%), *HLA-DQA1\*03:01:01* (7.21%) and *HLA-DQA1\*02:01:01:01* (6.76%). The highest allelic diversity was observed within DQA1\*01 and DQA1\*05 allelic families, together accounting for 77.47% of DQA1 allelic variability in our sample population. The most common DQA1\*05 representatives were DQA1\*05:01:01:02 (5.41%) and DQA1\*05:05:01:09 (4.5%) alleles. Two rare alleles of undefined C/WD status were also uncovered, namely, *HLA-DQA1\*01:10* and *HLA-DQA1\*04:01:02:01*.

Among 25 different HLA-DQB1 alleles observed, the most frequent alleles accounting for 40.99% of DQB1 allelic variability were *DQB1\*05:02:01:01* (15.77%), *DQB1\*03:01:01:03* (14.86%) and *DQB1\*05:01:01:03* (10.36%). These were followed by *HLA-DQB1\*02:01:01* (7.66%), *HLA-DQB1\*03:02:01:01* (7.21%) and *HLA-DQB1\*02:02:01:01* (6.76%) variants. Two most polymorphic allelic families, with at least 5 different allelic specificities in each, were DQB1\*05 and DQB1\*06 allelic groups. Of interest, *HLA-DQB1\*03:01:01:19* rare allele of undefined C/WD status was spotted in one donor.



**Figure 3.** Heat map of HLA-A~B~C~DRB1~DQA1~DQB1 linkage disequilibrium (LD) expressed in terms of the  $D'$  measure. The upper colour-key represents the range of  $D'$ . Bright yellow colours represent lower  $D'$  values, while dark red colours demonstrate high haplotype LD. R (v3.6.0), packages: gplots, RColorBrewer (<http://www.R-project.org>)<sup>58</sup>.

**Haplotype frequencies.** A complete list of predicted three and six locus haplotypes is given in Supplementary Tables 6–8. In total, 126 HLA-A~C~B haplotypes were observed in our sample population (Supplementary Table 6), and the list of 20 most commonly linked allelic combinations (>1%) is available in Table 3. Only one HLA-A~C~B haplotype, the ancestral European combination A\*01:01:01:01~C\*07:01:01~B\*08:01:01, was present in more than 5% of tested individuals. The second, third and fourth most frequent class I haplotypes were A\*11:01:01:01~C\*04:01:01:01~B\*35:01:01 (3.6%) A\*03:01:01:01~C\*07:02:01~B\*07:02:01 (3.11%) and A\*02:01:01~C\*06:02:01:01~B\*57:01:01:01 (2.7%), respectively. The remaining three-loci class I haplotypes were observed  $\leq 5$  times in our data set.

The HLA-DRB1~DQA1~DQB1 haplotype diversity is presented in Supplementary Table 7 and those observed  $\geq 3$  times are enlisted in Table 4. Among 75 different haplotypes, DRB1\*16:01:01~DQA1\*01:02:02~DQB1\*05:02:01:01 (13.51%) was the most frequent in our population. The next high-ranking haplotypes, were DRB1\*07:01:01~DQA1\*02:01:01:01~DQB1\*02:02:01:01 (5.32%), DRB1\*03:01:01:01~DQA1\*05:01:01:02~DQB1\*02:01:01 (4.96%), DRB1\*15:01:01~DQA1\*01:02:01:01~DQB1\*06:02:01:01 (4.50%) and DRB1\*01:01:01~DQA1\*01:01:01:01~DQB1\*05:01:01:03 (4.05%).

Of interest, rare DRB1\*01:01:01~DQA1\*01:02:01:03~DQB1\*05:04, and two tentative DRB1\*13:01:01:02~DQA1\*01:10~DQB1\*06:03:01:01 and DRB1\*04:05:01:03~DQA1\*05:05:01:06~DQB1\*03:01:01:03 haplotypes (Supplementary Table 8) with no previous entries within the Allele Frequency database so far, were spotted once.

Among 181 six-loci haplotypes (Supplementary Table 8), 10 appeared in three or more copies (Table 5), and the most frequent was the ancestral haplotype A\*01:01:01:01~C\*07:01:01~B\*08:01:01~DRB1\*03:01:01:01~DQA1\*05:01:01:02~DQB1\*02:01:01 (3.6%). Three six-loci combinations (A\*02:01:01~C\*07:01:01~B\*18:01:01~DRB1\*11:04:01~DQA1\*05:05:01:09~DQB1\*03:01:01:03, A\*03:01:01:01~C\*07:02:01~B\*07:02:01~DRB1\*16:01:01~DQA1\*01:02:02~DQB1\*05:02:01:01, and A\*11:01:01:01~C\*04:01:01:01~B\*35:01:01~DRB1\*16:01:01~DQA1\*01:02:02~DQB1\*05:02:01:01), which have already been reported in Italian population<sup>27</sup>, Turkish minority in Germany<sup>20</sup>, and Croatian families<sup>18</sup>, were also commonly (1.8%) observed. Haplotypes characteristic for South European populations of Slavic background or admixture were also found<sup>17</sup>, namely A\*02:01:01~C\*02:02:02~B\*27:02:01:01~DRB1\*16:01:02~DQA1\*01:02:02~DQB1\*05:02:01:01 (1.8%), and even more unique A\*02:01:01~C\*02:02:02~B\*27:05:02~DRB1\*01:01:01~DQA1\*01:01:01:03~DQB1\*05:02:01:01 (0.45%), A\*02:01:01~C\*07:02:01~B\*27:05:02~DRB1\*01:01:01~DQA1\*01:01:01:01~DQB1\*05:01:01:03 (0.45%) and A\*02:01:01~C\*07:02:01~B\*27:05:02~DRB1\*01:01:01~DQA1\*01:01:01:03~DQB1\*03:01:01:03 (0.45%). Nevertheless, a limited power in identifying low-frequency haplotypes should be noticed in our sample, where majority of haplotypic combinations were observed only once, and most frequent variants cover only 17.57% of our population haplotypic diversity.

	Allele HLA-A	n	AF	cF	Allele HLA-B	n	AF	cF	Allele HLA-C	n	AF	cF
1	A*02:01:01	72	32.43%	32.43%	B*51:01:01	19	8.56%	8.56%	C*04:01:01:01	31	13.96%	13.96%
2	A*01:01:01:01	27	12.16%	44.59%	B*08:01:01	18	8.11%	16.67%	C*07:01:01	25	11.26%	25.23%
3	A*03:01:01:01	26	11.71%	56.31%	B*35:01:01	18	8.11%	24.77%	C*02:02:02	21	9.46%	34.68%
4	A*24:02:01:01	26	11.71%	68.02%	B*07:02:01	15	6.76%	31.53%	C*07:02:01	18	8.11%	42.79%
5	A*11:01:01:01	22	9.91%	77.93%	B*35:03:01	15	6.76%	38.29%	C*12:03:01	16	7.21%	50.00%
6	A*32:01:01:01	8	3.60%	81.53%	B*18:01:01	12	5.41%	43.69%	C*06:02:01:01	15	6.76%	56.76%
7	A*23:01:01:01	5	2.25%	83.78%	B*15:01:01:01	9	4.05%	47.75%	C*03:03:01	12	5.41%	62.16%
8	A*25:01:01:01	5	2.25%	86.04%	B*27:05:02	9	4.05%	51.80%	C*01:02:01	11	4.96%	67.12%
9	A*31:01:02:01	5	2.25%	88.29%	B*44:02:01:01	9	4.05%	55.86%	C*05:01:01	11	4.96%	72.07%
10	A*26:01:01:01	4	1.80%	90.09%	B*27:02:01:01	8	3.60%	59.46%	C*07:04:01	11	4.96%	77.03%
11	A*30:01:01	3	1.35%	91.44%	B*52:01:01	8	3.60%	63.06%	C*04:01:01:06	9	4.05%	81.08%
12	A*68:01:02:02	3	1.35%	92.79%	B*55:01:01	8	3.60%	66.67%	C*12:02:02	8	3.60%	84.69%
13	A*24:02:01:05	2	0.90%	93.69%	B*57:01:01:01	7	3.15%	69.82%	C*15:02:01:01	7	3.15%	87.84%
14	A*29:02:01:01	2	0.90%	94.59%	B*13:02:01:01	6	2.70%	72.52%	C*03:04:01	5	2.25%	90.09%
15	A*66:01:01:01	2	0.90%	95.50%	B*44:27:01	6	2.70%	75.23%	C*07:01:02	4	1.80%	91.89%
16	A*03:01:01:03	1	0.45%	95.95%	B*44:03:01	5	2.25%	77.48%	C*08:02:01:01	3	1.35%	93.24%
17	A*03:01:01:05	1	0.45%	96.40%	B*15:17:01:01	4	1.80%	79.28%	C*14:02:01	3	1.35%	94.60%
18	A*26:01:01:06	1	0.45%	96.85%	B*35:02:01	4	1.80%	81.08%	C*07:18	2	0.90%	95.50%
19	A*26:08	1	0.45%	97.30%	B*38:01:01	4	1.80%	82.88%	C*15:05:02	2	0.90%	96.40%
20	A*29:01:01:01	1	0.45%	97.75%	B*39:01:01	4	1.80%	84.69%	C*17:03:01	2	0.90%	97.30%
21	A*31:01:02:04	1	0.45%	98.20%	B*14:02:01:01	3	1.35%	86.04%	C*03:02:02:01	1	0.45%	97.75%
22	A*33:01:01:01	1	0.45%	98.65%	B*40:01:02	3	1.35%	87.39%	C*03:04:02	1	0.45%	98.20%
23	A*33:03:01:02	1	0.45%	99.10%	B*40:02:01	3	1.35%	88.74%	C*06:02:01:02	1	0.45%	98.65%
24	A*68:01:01:02	1	0.45%	99.55%	B*41:02:01	3	1.35%	90.09%	C*07:01:09	1	0.45%	99.10%
25	A*68:02:01:01	1	0.45%	100.00%	B*58:01:01	3	1.35%	91.44%	C*16:01:01	1	0.45%	99.55%
26					B*18:03	2	0.90%	92.34%	C*16:02:01	1	0.45%	100.00%
27					B*35:08:01:01	2	0.90%	93.24%				
28					B*37:01:01:01	2	0.90%	94.14%				
29					B*49:01:01	2	0.90%	95.05%				
30					B*56:01:01	2	0.90%	95.95%				
31					B*07:05:01	1	0.45%	96.40%				
32					B*07:06:01	1	0.45%	96.85%				
33					B*15:01:06	1	0.45%	97.30%				
34					B*15:10:01	1	0.45%	97.75%				
35					B*27:02:01:04	1	0.45%	98.20%				
36					B*44:02:01:03	1	0.45%	98.65%				
37					B*44:05:01	1	0.45%	99.10%				
38					B*50:01:01:01	1	0.45%	99.55%				
39					B*54:01:01	1	0.45%	100.00%				

**Table 1.** HLA-A, HLA-B and HLA-C allele (AF) and cumulative frequencies (cF) in the sample of 111 unrelated blood donor volunteers from East Croatia.

## Discussion

This study represents the first report of HLA diversity in an east Croatian population of healthy blood donors, as studied by the next generation sequencing of 6 HLA genes, providing extensive exon/intron coverage with minimum ambiguity. For the first time, the allele frequencies and the extended six gene haplotypes of Croats were examined at the 4-field resolution level and compared to the largest repository of HLA class I and class II data in Croatia, the Croatian bone marrow donor registry (CBMDR).

The comparison of HLA-A allele frequencies between our and CBMDR inventory did not reveal significant differences, and the ranking hierarchy of the most common A\*02:01:01, A\*01:01:01:01, A\*03:01:01:01, and A\*24:01:01:01 alleles was also the same. Greater differences in frequency rate between general and east Croatian population were, however, noticed among HLA-B allelic variants. The HLA-B\*51:01 allelic group was the most frequent in both general and east Croatian cohort, but the frequency rank of the remaining HLA-B allelic variants was different, which was particularly evident for our 6<sup>th</sup> ranking B\*18:01 allele group, previously reported as the 2<sup>nd</sup> most frequent allelic variant in the CBMDR (8.16%)<sup>17</sup> and one Croatian family study (8.27%)<sup>18</sup>. Among five different B\*18 alleles detected in the Croats so far<sup>28</sup>, we observed only two, the B\*18:01:01 (5.14%) and the B\*18:03 (0.90%), resulting in a B\*18:01 distribution closer to those reported in Armenians<sup>29</sup>, Germans<sup>30</sup>, Austrian and the Turkish minority in Germany<sup>20</sup>, Bulgarian Roma individuals<sup>31</sup>, and Iranian Kurds<sup>32</sup>. Moreover, the frequency of the 2<sup>nd</sup> most common HLA-B allelic group in our cohort, the HLA-B\*35:01 (8.11%, ranked

	Allele HLA-DRB1	n	AF	cF	Allele HLA-DQA1	n	AF	cF	Allele HLA-DQB1	n	AF	cF
1	DRB1*16:01:01	30	13.51%	13.51%	DQA1*01:02:02	35	15.77%	15.77%	DQB1*05:02:01:01	35	15.77%	15.77%
2	DRB1*01:01:01	24	10.81%	24.32%	DQA1*03:01:01	16	7.21%	22.97%	DQB1*03:01:01:03	33	14.86%	30.63%
3	DRB1*07:01:01	22	9.91%	34.23%	DQA1*02:01:01:01	15	6.76%	29.73%	DQB1*05:01:01:03	23	10.36%	40.99%
4	DRB1*15:01:01	15	6.76%	40.99%	DQA1*01:03:01:02	13	5.86%	35.59%	DQB1*02:01:01	17	7.66%	48.65%
5	DRB1*03:01:01:01	12	5.41%	46.40%	DQA1*05:01:01:02	12	5.41%	40.99%	DQB1*03:02:01:01	16	7.21%	55.86%
6	DRB1*11:01:01:01	10	4.50%	50.90%	DQA1*01:02:01:01	10	4.50%	45.50%	DQB1*02:02:01:01	15	6.76%	62.61%
7	DRB1*11:04:01	10	4.50%	55.41%	DQA1*05:05:01:09	10	4.50%	50.00%	DQB1*05:03:01	13	5.86%	68.47%
8	DRB1*14:54:01:02	9	4.05%	59.46%	DQA1*01:01:01:01	9	4.05%	54.05%	DQB1*06:02:01:01	13	5.86%	74.32%
9	DRB1*08:01:01	8	3.60%	63.06%	DQA1*05:05:01:04	9	4.05%	58.11%	DQB1*06:03:01:01	13	5.86%	80.18%
10	DRB1*13:01:01:02	8	3.60%	66.67%	DQA1*01:01:01:03	7	3.15%	61.26%	DQB1*03:03:02:01	7	3.15%	83.33%
11	DRB1*13:01:01:01	7	3.15%	69.82%	DQA1*01:03:01:01	7	3.15%	64.41%	DQB1*04:02:01:01	7	3.15%	86.49%
12	DRB1*15:02:01:01	7	3.15%	72.97%	DQA1*02:01:01:02	7	3.15%	67.57%	DQB1*06:01:01/15	7	3.15%	89.64%
13	DRB1*11:01:01:03	6	2.70%	75.68%	DQA1*01:04:01:01	6	2.70%	70.27%	DQB1*06:04:01	4	1.80%	91.44%
14	DRB1*13:02:01:02	5	2.25%	77.93%	DQA1*05:05:01:01	6	2.70%	72.97%	DQB1*03:01:01:05	3	1.35%	92.79%
15	DRB1*03:01:01:03	4	1.80%	79.73%	DQA1*01:02:01:04	5	2.25%	75.23%	DQB1*05:01:01:05	3	1.35%	94.14%
16	DRB1*04:01:01:03	3	1.35%	81.08%	DQA1*01:04:01:02	5	2.25%	77.48%	DQB1*06:03:01:02	3	1.35%	95.50%
17	DRB1*04:02:01	3	1.35%	82.43%	DQA1*05:01:01:03	5	2.25%	79.73%	DQB1*03:03:02:02	2	0.90%	96.40%
18	DRB1*04:03:01:01	3	1.35%	83.78%	DQA1*01:01:01:02	4	1.80%	81.53%	DQB1*03:01:01:01	1	0.45%	96.85%
19	DRB1*10:01:01:03	3	1.35%	85.14%	DQA1*01:05:01	4	1.80%	83.33%	DQB1*03:01:01:02	1	0.45%	97.30%
20	DRB1*11:03:01	3	1.35%	86.49%	DQA1*04:01:01:03	4	1.80%	85.14%	DQB1*03:01:01:10	1	0.45%	97.75%
21	DRB1*12:01:01/12:10	3	1.35%	87.84%	DQA1*05:05:01:02	4	1.80%	86.94%	DQB1*03:01:01:19	1	0.45%	98.20%
22	DRB1*13:03:01	3	1.35%	89.19%	DQA1*05:05:01:05	4	1.80%	88.74%	DQB1*03:02:01:02	1	0.45%	98.65%
23	DRB1*04:01:01:01	2	0.90%	90.09%	DQA1*01:01:01:05	3	1.35%	90.09%	DQB1*05:01:01:01	1	0.45%	99.10%
24	DRB1*04:05:01:03	2	0.90%	90.99%	DQA1*01:02:01:05	3	1.35%	91.44%	DQB1*05:01:01:02	1	0.45%	99.55%
25	DRB1*04:08:01	2	0.90%	91.89%	DQA1*03:03:01:01	3	1.35%	92.79%	DQB1*05:04	1	0.45%	100.00%
26	DRB1*09:01:02	2	0.90%	92.79%	DQA1*03:02:01:01	2	0.90%	93.69%				
27	DRB1*14:01:01	2	0.90%	93.69%	DQA1*04:02	2	0.90%	94.60%				
28	DRB1*16:02:01:02	2	0.90%	94.60%	DQA1*05:05:01:06	2	0.90%	95.50%				
29	DRB1*01:02:01	1	0.45%	95.05%	DQA1*05:05:01:08	2	0.90%	96.40%				
30	DRB1*03:01:01:02	1	0.45%	95.50%	DQA1*01:01:02	1	0.45%	96.85%				
31	DRB1*04:01:01:02	1	0.45%	95.95%	DQA1*01:02:01:03	1	0.45%	97.30%				
32	DRB1*04:04:01	1	0.45%	96.40%	DQA1*01:03:01:06	1	0.45%	97.75%				
33	DRB1*04:15	1	0.45%	96.85%	DQA1*01:04:01:03	1	0.45%	98.20%				
34	DRB1*08:04:01	1	0.45%	97.30%	DQA1*01:04:01:04	1	0.45%	98.65%				
35	DRB1*10:01:01:01	1	0.45%	97.75%	DQA1*01:10	1	0.45%	99.10%				
36	DRB1*11:01:08	1	0.45%	98.20%	DQA1*04:01:02:01	1	0.45%	99.55%				
37	DRB1*13:05:01	1	0.45%	98.65%	DQA1*05:05:01:03	1	0.45%	100.00%				
38	DRB1*14:05:01:02	1	0.45%	99.10%								
39	DRB1*14:54:01:01	1	0.45%	99.55%								
40	DRB1*16:01:02	1	0.45%	100.00%								

**Table 2.** HLA-DRB1, HLA-DQA1 and HLA-DQB1 allele (AF) and cumulative frequencies (cF) in the sample of 111 unrelated blood donor volunteers from East Croatia.

5<sup>th</sup> in the CBMDR), was more similar to those observed in Turks<sup>20</sup>, Serbians<sup>9</sup>, and Italians<sup>27</sup>. A strong influence of south-eastern European populations on the HLA makeup of eastern Croats was further supported by the high prevalence of the HLA-B\*27:02:01:01 variant, which fits well into the B\*27:02 frequency gradient diminishing from the Middle East towards the Central and West European countries. In support, the observed B\*27:02:01:01 frequency (3.6%) seems to be in close agreement with the B\*27:02 cline extending across the south-eastern Tunisian (5.8%)<sup>33</sup>, Bulgarian (4.6%)<sup>34</sup>, CBMDR (2.14%)<sup>17</sup>, Czech (1.9%)<sup>9</sup>, and Polish (1.5%)<sup>35</sup> populations. The B\*44:27:01 allele, an east European marker considered a rare variant according to the “Rare Alleles Detector” tool<sup>9</sup>, was also noticed in our cohort at a relatively high frequency (2.7%), contrasting observations from Croatian HSCT patients (1.18%)<sup>19</sup>, Czech National Marrow Donors Registry (0.69%)<sup>9</sup>, as well as Polish (0.8%)<sup>35</sup>, English (0.19%)<sup>36</sup>, and Argentinian blood donors (0.07%)<sup>37</sup>. As minor allele within the functionally identical B\*44:02 G group, the B\*44:27 variant has been reported at a relative ratio frequency of >5% among the B\*44:02:01G-positive Bulgarian (36.82%), Hungarian (9.4%), Slovenian (25.60%) and Portuguese (6.17%) individuals<sup>38</sup>. In our sample, the B\*44:27:01 relative ratio frequency within the B\*44:02:01G-positive individuals (37.5%) sets the local estimate at the upper boundary.

Several ranking differences were detected between the general Croatian and our population at the HLA-C loci as well. The 1<sup>st</sup> (C\*07:01, AF = 21.77%), and the 2<sup>nd</sup> (C\*04:01, AF = 15.59%) highest ranking alleles from



	HLA-A~B~C			Observed (n)	HF	cF
1.	A*01:01:01:01	B*08:01:01	C*07:01:01	15.00	6.76%	6.76%
2.	A*11:01:01:01	B*35:01:01	C*04:01:01:01	8.00	3.60%	10.36%
3.	A*03:01:01:01	B*07:02:01	C*07:02:01	6.89	3.11%	13.47%
4.	A*02:01:01	B*57:01:01:01	C*06:02:01:01	6.00	2.70%	16.17%
5.	A*02:01:01	B*27:02:01:01	C*02:02:02	5.00	2.25%	18.42%
6.	A*02:01:01	B*35:01:01	C*04:01:01:06	5.00	2.25%	20.67%
7.	A*02:01:01	B*35:03:01	C*04:01:01:01	5.00	2.25%	22.93%
8.	A*02:01:01	B*44:02:01:01	C*05:01:01	5.00	2.25%	25.18%
9.	A*03:01:01:01	B*35:03:01	C*04:01:01:01	5.00	2.25%	27.43%
10.	A*02:01:01	B*44:27:01	C*07:04:01	4.00	1.80%	29.23%
11.	A*11:01:01:01	B*52:01:01	C*12:02:02	4.00	1.80%	31.03%
12.	A*02:01:01	B*27:05:02	C*02:02:02	3.89	1.75%	32.79%
13.	A*02:01:01	B*18:01:01	C*07:01:01	3.00	1.35%	34.14%
14.	A*02:01:01	B*52:01:01	C*12:02:02	3.00	1.35%	35.49%
15.	A*03:01:01:01	B*15:01:01:01	C*07:04:01	3.00	1.35%	36.84%
16.	A*23:01:01:01	B*44:03:01	C*04:01:01:01	3.00	1.35%	38.19%
17.	A*24:02:01:01	B*13:02:01:01	C*06:02:01:01	3.00	1.35%	39.54%
18.	A*24:02:01:01	B*35:02:01	C*04:01:01:06	3.00	1.35%	40.90%
19.	A*24:02:01:01	B*51:01:01	C*15:02:01:01	3.00	1.35%	42.25%
20.	A*02:01:01	B*07:02:01	C*07:02:01	2.88	1.30%	43.54%

**Table 3.** The most frequent (>1%) HLA~A~B~C haplotypes in East Croatia blood donor volunteers (n = 111).

No.	HLA-DQA1~DQB1~DRB1			Observed (n)	HF	cF
1.	DQA1*01:02:02	DQB1*05:02:01:01	DRB1*16:01:01	30.00	13.51%	13.51%
2.	DQA1*02:01:01:01	DQB1*02:02:01:01	DRB1*07:01:01	11.82	5.32%	18.84%
3.	DQA1*05:01:01:02	DQB1*02:01:01	DRB1*03:01:01:01	11.00	4.96%	23.79%
4.	DQA1*01:02:01:01	DQB1*06:02:01:01	DRB1*15:01:01	10.00	4.50%	28.30%
5.	DQA1*01:01:01:01	DQB1*05:01:01:03	DRB1*01:01:01	9.00	4.05%	32.35%
6.	DQA1*01:01:01:03	DQB1*05:01:01:03	DRB1*01:01:01	7.00	3.15%	35.50%
7.	DQA1*01:03:01:01	DQB1*06:01:01/15	DRB1*15:02:01:01	7.00	3.15%	38.66%
8.	DQA1*01:03:01:02	DQB1*06:03:01:01	DRB1*13:01:01:01	7.00	3.15%	41.81%
9.	DQA1*01:04:01:01	DQB1*05:03:01	DRB1*14:54:01:02	6.00	2.70%	44.51%
10.	DQA1*05:05:01:09	DQB1*03:01:01:03	DRB1*11:04:01	5.00	2.25%	46.76%
11.	DQA1*01:01:01:02	DQB1*05:01:01:03	DRB1*01:01:01	4.00	1.80%	48.57%
12.	DQA1*01:02:01:04	DQB1*06:04:01	DRB1*13:02:01:02	4.00	1.80%	50.37%
13.	DQA1*01:03:01:02	DQB1*06:03:01:01	DRB1*13:01:01:02	4.00	1.80%	52.17%
14.	DQA1*04:01:01:03	DQB1*04:02:01:01	DRB1*08:01:01	4.00	1.80%	53.97%
15.	DQA1*05:01:01:03	DQB1*02:01:01	DRB1*03:01:01:03	4.00	1.80%	55.77%
16.	DQA1*02:01:01:02	DQB1*03:03:02:01	DRB1*07:01:01	3.82	1.72%	57.49%
17.	DQA1*02:01:01:01	DQB1*03:03:02:01	DRB1*07:01:01	3.18	1.43%	58.93%
18.	DQA1*02:01:01:02	DQB1*02:02:01:01	DRB1*07:01:01	3.18	1.43%	60.36%
19.	DQA1*01:01:01:05	DQB1*05:01:01:03	DRB1*01:01:01	3.00	1.35%	61.71%
20.	DQA1*01:02:01:05	DQB1*06:02:01:01	DRB1*15:01:01	3.00	1.35%	63.06%
21.	DQA1*03:01:01	DQB1*03:02:01:01	DRB1*04:01:01:03	3.00	1.35%	64.41%
22.	DQA1*03:01:01	DQB1*03:02:01:01	DRB1*04:02:01	3.00	1.35%	65.77%
23.	DQA1*03:01:01	DQB1*03:02:01:01	DRB1*04:03:01:01	3.00	1.35%	67.12%
24.	DQA1*05:05:01:02	DQB1*03:01:01:03	DRB1*11:01:01:01	3.00	1.35%	68.47%
25.	DQA1*05:05:01:04	DQB1*03:01:01:03	DRB1*11:01:01:01	3.00	1.35%	69.82%
26.	DQA1*05:05:01:05	DQB1*03:01:01:03	DRB1*13:03:01	3.00	1.35%	71.17%
27.	DQA1*05:05:01:09	DQB1*03:01:01:03	DRB1*11:01:01:03	3.00	1.35%	72.52%

**Table 4.** The most frequent (>1%) HLA-DQA1~DQB1~DRB1 haplotypes in East Croatia blood donor volunteers (n = 111).

No.	HLA-A~B~C~DQA1~DQB1~DRB1						n	HF	cF
1	A*01:01:01:01	B*08:01:01	C*07:01:01	DQA1*05:01:01:02	DQB1*02:01:01	DRB1*03:01:01:01	8.00	3.60%	3.60%
2	A*02:01:01	B*18:01:01	C*07:01:01	DQA1*05:05:01:09	DQB1*03:01:01:03	DRB1*11:04:01	4.00	1.80%	5.41%
3	A*02:01:01	B*27:02:01:01	C*02:02:02	DQA1*01:02:02	DQB1*05:02:01:01	DRB1*16:01:01	4.00	1.80%	7.21%
4	A*03:01:01:01	B*07:02:01	C*07:02:01	DQA1*01:02:02	DQB1*05:02:01:01	DRB1*16:01:01	4.00	1.80%	9.01%
5	A*11:01:01:01	B*35:01:01	C*04:01:01:01	DQA1*01:02:02	DQB1*05:02:01:01	DRB1*16:01:01	4.00	1.80%	10.81%
6	A*02:01:01	B*13:02:01:01	C*06:02:01:01	DQA1*02:01:01:01	DQB1*02:02:01:01	DRB1*07:01:01	3.00	1.35%	12.16%
7	A*02:01:01	B*44:27:01	C*07:04:01	DQA1*01:02:02	DQB1*05:02:01:01	DRB1*16:01:01	3.00	1.35%	13.51%
8	A*02:01:01	B*52:01:01	C*12:02:02	DQA1*01:03:01:01	DQB1*06:01:01/15	DRB1*15:02:01:01	3.00	1.35%	14.87%
9	A*11:01:01:01	B*35:01:01	C*04:01:01:01	DQA1*01:01:01:01	DQB1*05:01:01:03	DRB1*01:01:01	3.00	1.35%	16.22%
10	A*24:02:01:01	B*13:02:01:01	C*06:02:01:01	DQA1*02:01:01:01	DQB1*02:02:01:01	DRB1*07:01:01	3.00	1.35%	17.57%

**Table 5.** HLA-A~B~C~DRB1~DQA1~DQB1 extended haplotypes with estimated haplotype (HF) and cumulative frequency (cF) of 1% or more in East Croatian blood donor volunteers (n = 111).

the CBMDR inventory were ranked 2<sup>nd</sup> (14.41%) and 1<sup>st</sup> (18.01%) in east Croatian blood donors, respectively. The ranking sequence in our cohort continued with C\*02:02:02 (9.46%), and C\*07:02:01 (8.11%) allele group, which have been deemed 4<sup>th</sup> and 5<sup>th</sup> highest ranking alleles in the CBMRD. Overall, the ranking hierarchy of HLA-C allele groups in our sample was more similar to the one reported for Greece, and the Turkish minority in Germany<sup>20</sup>.

Compared to class I, class II loci exhibited a higher level of heterozygosity at the 4-field level (molecular variation in introns and UTR regions), allowing us to see three or more different allelic variants within the particular 3-field G allelic group. For instance, the highest allelic variability was found within the DQA1\*05:05:01 G (8 different alleles), DQA1\*01:01:01 G (5), DQA1\*01:02:01 G (5), DQA1\*01:04:01 G (5), DQB1\*03:01:01 G (6) and DQB1\*05:01:01 G (4), whereas up to 3 allelic variants were revealed within the DRB1\*03:01:01 G, DRB1\*04:01:01 G, DRB1\*11:01:01 G and DQA1\*03:01:01 G group<sup>6,7</sup>.

Compared to Croatian general population and neighbouring nations, the frequency and order of DRB1\*16:01 (13.96%) and DRB1\*03:01 (7.66%), the 1<sup>st</sup> and 5<sup>th</sup> most common DRB1 allele groups in our cohort, were opposite to CBMDR allelic hierarchy (9.41% (4<sup>th</sup>) and 10.01% (1<sup>st</sup>), respectively), but similar to East European (Bulgaria 15.5%, 8.18%, respectively)<sup>34</sup>, and Mediterranean populations (Macedonia 14.9%, 6.7% and Greece 11.5%, 6.5%, respectively)<sup>39–41</sup>. At the same time, the frequency ratio of DRB1\*11:01:01:01 and DRB1\*11:04:01 alleles in our sample was in concordance with the CBMDR data, confirming equal frequency rate of these two alleles across Croatia, in contrast to higher DRB1\*11:01 prevalence in the north, and DRB1\*11:04 in the south of Europe<sup>17</sup>. In line with the high prevalence and LD patterns of DRB1\*16:01:01, DQA1\*01:02:02 (15.77%) and DQB1\*05:02:01:01 (15.77%) alleles were the most commonly observed DQA1 and DQB1 variants at the 4-field level among east Croatian blood donors. However, looking at the 2-field level, DQA1\*01:02 (24.32%), DQA1\*05:05 (17.1%) and DQB1\*03:01 (18.01%) allelic families were found at a higher frequency, emphasizing significant allelic heterogeneity within the particular allelic group. The most striking observation was a high frequency of the DQA1\*05:05 allele (17.1%), a common South-European allelic variant (EFI CWD catalogue v.1.0.) that has been reported at much lower frequency only in the autochthonous Croatian population from the Gorski Kotar (2.4%)<sup>24</sup>, but not in the CBMDR database. Thus far, the DQA1\*05:05 allele has been reported for a handful of European populations<sup>9</sup>, and the frequency observed in our cohort is currently surpassed only by the prevalence observed among Czechs (20.6%)<sup>42</sup>, north Italians (30.5%)<sup>9</sup> and Greeks (32.5%)<sup>43</sup>. Additional differences at the DQB1 locus were noticed after comparing DQB1\*05:02 and DQB1\*02:01 frequencies, with DQB1\*05:02 being more prevalent in eastern (east vs. general CRO; 15.77% vs. 8.53%) and DQB1\*02:01 in general Croatian population (east vs. general CRO; 7.66 vs. 12.94%), partly reflecting a positive, North-West to South-East gradient of DQB1\*05:02 frequencies. In contrast, an inverse distribution was observed for DQB1\*02:01 across Europe, with the highest DQB1\*05:02 and DQB1\*02:01 prevalence observed in Greeks (19.8%)<sup>44</sup> and Englishmen (33%)<sup>45</sup>, respectively.

The haplotype distribution in our cohort is in close agreement with previously reported class I and class II associations in CBMDR and Croatian families, with few interesting differences found in our population of eastern Croats. Next to the pan-European (A\*01:01:01:01~C\*07:01:01~B\*08:01:01~DRB1\*03:01:01:01~DQA1\*05:01:01:02~DQB1\*02:01:01) and Mediterranean (A\*02:01:01~C\*07:01:01~B\*18:01:01~DRB1\*11:04:01~DQA1\*05:05:01:09~DQB1\*03:01:01:03) haplotypes, two notable exceptions were extended A\*03:01~C\*07:02~B\*07:02~DRB1\*16:01 and A\*11:01~B\*35:01~C\*04:01~DRB1\*16:01 variants, both commonly observed in Macedonian families<sup>46</sup>, as well as Austrian, Croatian, Bosnian-Herzegovinian, Italian, Romanian, Greek and Turkish minority in Germany<sup>20</sup>, but not in the CBMDR<sup>17</sup> or Croatian family study<sup>18</sup> (Table 5), where A\*03:01~B\*07:02 was more frequently found with DRB1\*15:01, and A\*11:01~B\*35:01 with DRB1\*01:01. Moreover, DRB1\*11:01, -\*11:04, -\*12:01 and -\*13:03 alleles were more frequently associated with DQA1\*05:01~DQB1\*03:01 haplotype in general, but not east Croatian population, where above-mentioned DRB1 alleles were in strong linkage disequilibrium with DQA1\*05:05~DQB1\*03:01 combination (Supplementary Table 8). Similar linkage patterns were, however, noticed by comparing extended A\*02:01~B\*27:02~C\*02:02~DRB1\*16:01~DQA1\*01:02~DQB1\*05:02 (1.8%) and A\*02:01~B\*27:05~C\*02:02~DRB1\*01:01~DQA1\*01:01~DQB1\*05:02 (0.45%) haplotypes with the CBMDR and Southeast European haplotype inventory available from the German Bone Marrow Donor Registry (DKMS)<sup>20</sup>. Compared to our database,

the A\*02:01~B\*27:02~C\*02:02~DRB1\*16:01 haplotype was more frequent only in Bulgarians (2.73%)<sup>34</sup>, but more similar to CBMDR (0.74%)<sup>17</sup>, Polish (1%), Bosnian and Herzegovinian (0.9%), and Croatian (0.834%) minority in Germany<sup>20</sup>.

The A\*02:01~B\*27:05~C\*02:02~DRB1\*01:01, was observed at lower frequency in Polish (0.27%), Austrian (0.12%), Bosnian and Herzegovinian (0.065%), Greek (0.026%) and Romanian (0.067%) minority in DKMS<sup>20</sup>, compared to CBMDR (0.85%) and our cohort (0.45%), possibly supporting local origin of this haplotype. Moreover, it was interesting to see a relatively high prevalence of the extended A\*02:01~B\*44:27~C\*07:04~DQA1\*01:02~DQB1\*05:02~DRB1\*16:01 (1.35%) haplotype, for whom there is no population data in Allele Frequency Net database. Recent study performed in B\*44-positive Croatian subjects reported a very strong and almost exclusive linkage disequilibrium between B\*44:27 and C\*07:04 alleles, whereas B\*44:02, -\*44:03 and -\*44:05 were more commonly seen in association with C\*05:01, -\*04:01 and -\*02:02, respectively<sup>47</sup>. This is in complete agreement with the B\*44 extended haplotypes found in our cohort [A\*02:01:01~B\*44:02:01:01~C\*05:01:01~DRB1\*13:01:01:01 (0.9%), A\*23:01:01:01~B\*44:03:01~C\*04:01:01:01~DRB1\*07:01:01 (1.35%), and A\*02:01:01~B\*44:05:01~C\*02:02:02~DRB1\*01:01:01 (0.45%)]. Our data further coincided with the DKMS reports on the A\*24:02~C\*15:02~B\*51:01~DRB1\*16:01~DQB1\*05:02 and A\*24:02~C\*06:02~B\*13:02~DRB1\*07:01~DQB1\*02:02 haplotype frequency in Croatian, Greece, Bosnian and Herzegovinian and Polish minority in Germany<sup>20</sup>. In addition, several extended variants of common Mediterranean and Southeastern European haplotypes were also observed in our population; but at a lower frequency; namely the A\*02:01:01~C\*12:03:01~B\*51:01:01~DRB1\*11:01:01:01~DQB1\*03:01:01:03 (0.45%) and A\*24:02:01:01~C\*04:01:01:06~B\*35:02:01~DRB1\*11:04:01~DQB1\*03:01:01:02 (0.45%), which were most frequent in Greece<sup>41</sup>, Albania<sup>48</sup>, Italian and Turkish minority in DKMS<sup>20</sup> (reduced haplotype variant); and the A\*02:01:01~C\*02:02:01~B\*51:01:01~DRB1\*13:01:01:01~DQB1\*03:01:01:19 (0.45%) variant, most frequently found in Bulgarians (reduced haplotype variant)<sup>34</sup>. Significant influence of Central and Western European countries on the east Croatian HLA profile is nonetheless also evident through higher prevalence of two extended, common European haplotypes, the A\*11:01:01~C\*04:01:01:01~B\*35:01:01~DRB1\*01:01:01~DQA1\*01:01:01~DQB1\*05:01:01:03 (1.35%) and A\*02:01:01~C\*06:02:01:01~B\*13:02:01:01~DRB1\*07:01:01~DQA1\*02:01:01:01~DQB1\*02:02:01:01 (1.35%), which occur at similar frequency in Swedish<sup>9</sup>, Polish<sup>35</sup>, CBMDR<sup>17</sup>, Austrian, Italian and the Portuguese minority population from the DKMS inventory<sup>20</sup>.

In conclusion, the present study provides an in-depth characterisation of HLA diversity in eastern Croats, revealing distinctive allele and haplotype detail consistent with the complex population history of the studied geographic region. The data complement and refine the existing estimates of HLA diversity in the Croatian population, increase population and geographic coverage by NGS data, and add granularity to clinically and genetically relevant HLA data. The study represents a useful reference for population and HLA-disease association studies; however, larger sample size and sequence coverage, particularly for the DQB1 and DRB1 genes, remain a prerequisite for the future studies.

## Materials and Methods

**Subjects.** The study collection consisted of 120 healthy, unrelated, blood donor volunteers (34 female, 86 male, 20–61 years of age, median age 36 years) originating from five eastern Croatia counties; Osijek-Baranja county (n = 80), Vukovar-Srijem county (n = 22), Brod-Posavina county (n = 9) Požega-Slavonia county (n = 4) and Virovitica-Podravina county (n = 5). All participants were recruited during voluntary blood donations in county Red Cross branches or at the Clinical Institute of Transfusion Medicine, Osijek University Hospital. Prior to the blood sampling, completed health questionnaire forms were collected from all donors, to select individuals with no personal or family history of various autoimmune and cardiovascular diseases, stroke or carcinoma. Informed consent in written form was collected from all subjects. All investigations were conducted in accordance with the 1964 Declaration of Helsinki and subsequent legal instruments. Ethical approval was provided by the University Hospital Osijek Ethics Committee (No. 25–1:831–3/2015).

**DNA extraction and quantification.** Genomic DNA was extracted from 200 µl peripheral blood samples mixed with EDTA, using High Pure PCR Template Preparation Kit (Roche Diagnostics, Mannheim, Germany) according to the instructions in the manufacturer leaflet. Quantity and quality of isolated genetic material was verified by OD<sub>260</sub>/OD<sub>280</sub> > 1.8 and OD<sub>260</sub>/OD<sub>230</sub> > 1.6 measurements performed on IMPLEN NanoPhotometer P-Class P-330 (IMPLEN GmbH, Munich, Germany).

**Long-range PCR amplification, pooling and clean-up of PCR products.** HLA genotypes for HLA-A, -B, -C, -DRB1, -DQA1, and -DQB1 loci were determined using high-resolution Omixon Holotype HLA 96/7 and 24/7 (Omixon Biocomputing Ltd, Budapest, Hungary) configuration kits on Illumina MiSeq next-generation sequencing platform. The Omixon Holotype PCR primers allowed amplification of the entire HLA-A (5'UTR nucleotide position: -78; 3'UTR nucleotide position: 3081), -B (5'UTR: -35; 3'UTR: 2680), -C (5'UTR: -122; 3'UTR: 2915), and -DQA1 (5'UTR: -281; 3'UTR: 5750) loci. Class II sequence analysis evaluated nucleotides from intron 1 to 3' UTR of HLA-DQB1 (intron 1: 645; 3'UTR: 6469), whereas -DRB1 locus was sequenced from intron 1 (nucleotide position 4753) to intron 4 (nucleotide position 9135) (Fig. 2). The HLA genotyping workflow was initiated by long-range PCR amplification of class I and class II HLA loci in a separate, sample and locus specific 25 µl reactions, comprising 2.5 µl of PCR buffer, 1.25 µl of dNTP mix, 2 µl of locus specific primers, 0.4 µl of LR PCR enzyme, and 5 µl of genomic DNA (~30 ng/µl). Combined DQB1 enhancer (5.6 µl/sample) was added to the DQB1 master mix only. The conditions for class I gene amplification on Mastercycler nexus thermal cycler (Eppendorf, Hamburg, Germany) were set as follows: 95 °C for 3 min, followed by 35 cycles of 95 °C for 15 s, 65 °C for 30 s and 68 °C for 5 min, and a final incubation at 68 °C for 10 min. For class II genes, the conditions were: 95 °C for 3 min, 35 cycles of 93 °C for 15 s, 60 °C for 30 s and 68 °C for 9 min, followed by final extension at

68 °C for 10 min. Amplicon size was validated by 2% agarose gel electrophoresis and DNA quantitated on EnSpire Multimode plate reader (PerkinElmer, Waltham, MA, USA) using QuantiFluor fluorescent dsDNA staining system (Promega, Madison, Wisconsin, USA). All six amplicons from one individual sample were pooled into a final 35 µl volume on a fresh 96-well PCR plate, and purified from residual primers and unincorporated nucleotides with the use of ExoSAP-IT enzyme mix (Affymetrix Inc., Santa Clara, CA, USA).

**Library construction, normalisation and sequencing on MiSeq.** Library preparation, in the next few steps, included fragmentation of each six-locus amplicon pool, fragment end repair and ligation with sample-specific indexed adaptors. Equal aliquots of indexed sample-specific libraries were subsequently combined into a 900 µl pooled library volume and mixed with 900 µl of the AMPure XP beads (Beckman Coulter, Beverly, Massachusetts, USA) to carry out magnetic bead-based library cleanup. Pooled library fragments ranging between 650 and 1300 bps in size were subsequently selected on Pippin Prep instrument (Sage Science, Beverly, Massachusetts, USA). The concentration of the size selected library was determined on LightCycler 480 II (Roche Diagnostics, Mannheim, Germany) real-time PCR instrument using KAPA Sybr Fast qPCR Master Mix (KAPA Biosystems, Boston, Massachusetts, USA) and DNA standards ranging from 0.02 pM to 20 pM concentrations. Prior to sequencing, library was diluted to a 2 nM concentration, loaded on a MiSeq flow cell (Illumina, San Diego, CA, USA) and sequenced in a single 500 cycle (V2) paired-end sequencing run. Collected reads were exported in fastq format and analysed with the Omixon Twin software v3.0.0. and the IPD-IMGT/HLA database Release 3.30.0\_5 (November 2017).

**Data analysis.** The best matching alleles were selected according to the alignment statistics (described in section 4.6), and homology to alleles available in the IMGT/HLA 3.30.0\_5 database<sup>6,7</sup>. If more than one allele call was available for a specific locus, the ambiguity was resolved by re-analysis of increased number of reads processed from the input files. The remaining ambiguous allele calls (presented in Supplementary Table 1) were referenced against the “Omixon Holotype HLA and Omixon HLA Twin known product limitations” (missing data on SNPs or INDEL variations within the unsequenced 3' UTR, 5' UTR and intron 1/exon 1 regions), and were hence reported as ambiguous (i.e. DQB1\*06:01:01/15) or up to the third field level only (i.e. DQB1\*05:03:01). The Common and Well-Documented (CWD) allele catalogue (version 2.0.), and “Rare Allele Detector” tool ([www.allelefreqencies.net](http://www.allelefreqencies.net)), were used for the identification of rare HLA alleles. Nine out of 120 samples were excluded from this study due to Omixon Twin quality control failure. HLA-A, -B, -C, DRB1, -DQA1, -and -DQB1 loci were successfully sequenced in all remaining samples (n = 111).

**Quality control (QC) metrics.** The Omixon Twin software combines statistical alignment and *de novo* assembly algorithms for robust allele calling. The default minimum number of reads required for reliable locus mapping was set at  $\geq 2500$  for class I, and  $\geq 5000$  for class II loci. A read length of 200 bp or greater was a prerequisite for passing QC criteria, and together with additional quality metrics (read quality, noise ratio, consensus phasing, allele imbalance, crossmapping reads, mismatch count) assured the accuracy and confidence of allele assignments. The minimum exon/intron coverage threshold supporting the consensus sequence at the weakest position was set at  $\geq 30$  reads.

**Statistics.** Allelic frequencies were determined by direct counting. Arlequin version 3.5.2.2<sup>49</sup> was used to calculate expected and observed heterozygosity, exact deviations from Hardy-Weinberg equilibrium (a modified version of the Markov-chain random walk algorithm described by Guo and Thomson,  $10^6$  steps in Markov chain,  $10^5$  dememorization steps)<sup>50</sup>, and maximum-likelihood haplotype frequencies (an iterative Expectation-Maximization algorithm, convergence criterion  $\epsilon = 10^{-7}$ , maximum number of iterations = 1000, 50 random initial conditions)<sup>51</sup>. A series of linkage disequilibrium (LD) measures ( $D^{52,53}$ ,  $Wn^{54}$ ) was provided for each pair of loci by using the Pypop 0.7.0 software<sup>55</sup>. The empirical P-values were obtained by permutation testing (1000 randomizations).

Received: 20 October 2019; Accepted: 3 March 2020;

Published online: 26 March 2020

## References

- Rendečić Miočević, D. *Iliri i Antički Svijet. (The Illyrians and the Ancient World., ed. Mimica, I) Književni krug: Split, Croatia, 1–920 (1989).*
- Mujadžević, D. The other Ottoman serhat in Europe: Ottoman territorial expansion in Bosnia and Croatia in 16th century. *Ankara Univ. J. Center South East Eur. Studies.* **1**, 99–111 (2012).
- Pavičić, S. *Slavonija u svojem naselnom razvitku od trinaestog stoljeća do danas. (Slavonia In Its Settlement Development From The Thirteenth Century To The Present., ed. V. Radauš) Zbornik radova I. znanstvenog sabora Slavonije i Baranje, Jugoslavenska akademija znanosti i umjetnosti, 191–236 (1970).*
- Sanchez-Mazas, A., Buhler, S. & Nunes, J. M. A new HLA map of Europe: Regional genetic variation and its implication for peopling history, disease-association studies and tissue transplantation. *Hum. Hered.* **76**, 162–177 (2013).
- Choo, S. Y. The HLA system: genetics, immunology, clinical testing, and clinical implications. *Yonsei. Med. J.* **48**, 11–23 (2007).
- Robinson, J. *et al.* The IPD and IPD-IMGT/HLA Database: allele variant databases. *Nucleic Acids Res.* **43**, D423–431 (2015).
- Robinson, J., Malik, A., Parham, P., Bodmer, J. G. & Marsh, S. G. E. IMGT/HLA - a sequence database for the human major histocompatibility complex. *Tissue Antigens* **55**, 280–28 (2000).
- Trowsdale, J. & Knight, J. C. Major Histocompatibility Complex Genomics and Human Disease. *Annu Rev. Genomics Hum. Genet.* **14**, 301–323 (2013).
- Gonzalez-Galarza, F. F. *et al.* Allele frequency net 2015 update: new features for HLA epitopes, KIR and disease and HLA adverse drug reaction associations. *Nucleic Acid Res.* **28**, D784–D788 (2015).
- Sanchez-Mazas, A. *et al.* Common and well-documented HLA alleles over all of Europe and within European sub-regions: A catalogue from the European Federation for Immunogenetics. *HLA* **89**, 104–113 (2017).

11. Pröll, J. *et al.* Sequence capture and next generation resequencing of the MHC region highlights potential transplantation determinants in HLA identical haematopoietic stem cell transplantation. *DNA Res.* **18**, 201–210 (2011).
12. Zhao, L. P. *et al.* Next-generation sequencing reveals that HLA-DRB3, -DRB4, and -DRB5 may be associated with islet autoantibodies and risk for childhood type 1 diabetes. *Diabetes.* **65**, 710–718 (2016).
13. Carapito, R., Radosavljevic, M. & Bahram, S. Next-generation sequencing of the HLA locus: methods and impacts on HLA typing, population genetics and disease association studies. *Hum. Immunol.* **77**, 1016–1023 (2016).
14. Kishore, A. & Petrek, M. Next-generation sequencing based HLA typing: deciphering immunogenetic aspects of sarcoidosis. *Front Genet.* **9**, 503 (2018).
15. Sanchez-Mazas, A. & Meyer, D. The relevance of HLA sequencing in population genetics studies. *J. Immunol. Res.* **2014**, 971818 (2014).
16. Grubic, Z., Zunec, R., Cecuk-Jelicic, E., Kerhin-Brkljacic & Kastelan, A. Polymorphism of HLA-A, -B, -DRB1, -DQA1, and -DQB1 haplotypes in a Croatian population. *Eur. J. Immunogenet.* **27**, 47–51 (2000).
17. Grubic, Z. *et al.* HLA-A, HLA-B and HLA-DRB1 allele and haplotype diversity among volunteer bone marrow donors from Croatia. *Int. J. Immunogenet.* **41**, 211–221 (2014).
18. Grubic, Z., Maskalan, M., Svilicic, D., Stingl Jankovic, K. & Zunec, R. Determination of HLA-A, -B, and -DRB1 allele and haplotype frequencies in the croatian population based on a family study. *Arch. Immunol. Ther. Exp. (Warsz).* **64**(Suppl 1), 83–88 (2016).
19. Grubic, Z. *et al.* The effect of HLA allele and haplotype polymorphisms on donor matching in hematopoietic stem cell transplantation – Croatian experience. *Hum. Immunol.* **77**, 1120–1127 (2016).
20. Pingel, J. *et al.* High-resolution HLA haplotype frequencies of stem cell donors in Germany with foreign parentage: how can they be used to improve unrelated donor searches? *Hum. Immunol.* **74**, 330–340 (2013).
21. Martinović, I. *et al.* Application of HLA class II polymorphism analysis to the study of the population structure of the island Krk, Croatia. *Hum Biol.* **69**, 819–829 (1997).
22. Grubić, Z. *et al.* HLA class II gene and haplotype diversity in the population of the island of Hvar, Croatia. *Coll. Antropol.* **22**, 157–168 (1998).
23. Crnić-Martinović, M. *et al.* HLA class I and class II polymorphism in the population of Rijeka, Croatia. *Coll. Antropol.* **26**, 69–75 (2002).
24. Crnić-Martinović, M. *et al.* HLA class II polymorphism in autochthonous population of Gorski kotar, Croatia. *Coll. Antropol.* **31**, 853–858 (2007).
25. Grubic, Z., Burek Kamenaric, M., Maskalan, M., Stingl Jankovic, K. & Zunec, R. Nonfrequent but well-documented, rare and very rare HLA alleles observed in the Croatian population. *Tissue Antigens* **84**, 560–564 (2014).
26. Burek Kamenaric, M. *et al.* HLA-DPB1 matching in unrelated hematopoietic stem cell transplantation program contributes to a higher incidence of disease relapse. *Hum. Immunol.* **78**, 665–671 (2017).
27. Rendine, S. *et al.* Estimation of human leukocyte antigen class I and class II high-resolution allele and haplotype frequencies in the Italian population and comparison with other European populations. *Hum. Immunol.* **73**, 399–404 (2012).
28. Burek Kamenaric, M., Maskalan, M., Drabbels, J., Golubic Cepulic, B. & Grubic, Z. Identification of the novel HLA-B\*18:37:02 allele in a Croatian individual. *HLA.* **91**, 299–300 (2018).
29. Matevosyan, L. *et al.* HLA-A, HLA-B, and HLA-DRB1 allele distribution in a large Armenian population sample. *Tissue Antigens* **78**, 21–30 (2011).
30. Schmidt, A. H. *et al.* Estimation of high-resolution HLA-A, -B, -C, -DRB1 allele and haplotype frequencies based on 8862 German stem cell donors and implications for strategic donor registry planning. *Hum. Immunol.* **70**, 895–902 (2009).
31. Meyer, D. *et al.* 13th IHWS anthropology/human genetic diversity joint report. Chapter 4. Single locus polymorphism of classical HLA genes. (ed. Hansen J. A.) Immunobiology of the human MHC, International Histocompatibility Working Group Press, 1–52 (2006).
32. Amirzargar, A. *et al.* Kurds HLA Genes: Its Implications in Transplantation and Pharmacogenomics. *Open Med. J.* **2**, 43–47 (2015).
33. Hajje, A. *et al.* HLA class I and class II polymorphism in a population from south-eastern Tunisia (Gabes Area). *Int. J. Immunogenet.* **38**, 191–199 (2011).
34. Ivanova, M. *et al.* HLA polymorphism in Bulgarians defined by high-resolution typing methods in comparison with other populations. *Tissue Antigens.* **60**, 496–504 (2002).
35. Nowak, J. *et al.* Allele and extended haplotype polymorphism of HLA-A, -C, -B, -DRB1 and -DQB1 loci in Polish population and genetic affinities to other populations. *Tissue Antigens.* **71**, 193–205 (2008).
36. Davey, S., Ord, J., Navarrete, C. & Brown, C. HLA-A, -B and -C allele and haplotype frequencies defined by next generation sequencing in a population of 519 English blood donors. *Hum. Immunol.* **78**, 397–398 (2017).
37. Montero-Martin, G. *et al.* High-resolution characterization of allelic and haplotypic HLA frequency distribution in a Spanish population using high-throughput next-generation sequencing. *Hum. Immunol.* **80**, 429–436 (2019).
38. Vidan-Jeras, B. *et al.* Resolution of HLA-B\*44:02:01G, -DRB1\*14:01:01G and -DQB1\*03:01:01G reveals a high allelic variability among 12 European populations. *Tissue Antigens* **84**, 459–464 (2014).
39. Arnaiz-Villena, A. *et al.* HLA genes in Macedonians and the sub-Saharan origin of the Greeks. *Tissue Antigens.* **57**, 118–127 (2001).
40. Petlichkovski, A. *et al.* High-resolution typing of HLA-DRB1 locus in the Macedonian population. *Tissue Antigens.* **64**, 486–491 (2004).
41. Papassavas, E. C. *et al.* MHC class I and class II phenotype, gene, and haplotype frequencies in Greeks using molecular typing data. *Hum Immunol.* **61**, 615–623 (2000).
42. Zajacova, M., Kotrbova-Kozak, A. & Cerna, M. HLA-DRB1, -DQA1 and -DQB1 genotyping of 180 Czech individuals from the Czech Republik pop 3. *Hum Immunol.* **77**, 365–366 (2016).
43. Kokaraki, G. *et al.* Major histocompatibility complex class II (DRB1\*, DQA1\*, and DQB1\*) and DRB1\*04 subtypes' associations of Hashimoto's thyroiditis in a Greek population. *Tissue Antigens.* **73**, 199–205 (2009).
44. Reveille, J. D. *et al.* HLA-class II alleles and C4 null genes in Greeks with systemic lupus erythematosus. *Tissue Antigens* **46**, 417–421 (1995).
45. Wu, Z. *et al.* Molecular analysis of HLA-DQ and -DP genes in caucasoid patients with Hashimoto's thyroiditis. *Tissue Antigens.* **43**, 116–119 (1994).
46. Kirijas, M., Genadijeva Stavrik, S., Senev, A., Efinška Mladenovska, O. & Petlichkovski, A. HLA-A, -B, -C and -DRB1 allele and haplotype frequencies in the Macedonian population based on a family study. *Hum. Immunol.* **79**, 145–153 (2018).
47. Stingl Jankovic, K. *et al.* The study of HLA-B\*44-C haplotype polymorphism in the Croatian population. *Mol. Exp. Biol. Med.* **1**, 42–46 (2017).
48. Sulcebe, G. *et al.* HLA allele and haplotype frequencies in the Albanian population and their relationship with the other European populations. *Int. J. Immunogenet.* **36**, 337–343 (2009).
49. Excoffier, L., Laval, G. & Schneider, S. Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol. Bioinform. Online.* **23**, 47–50 (2007).
50. Guo, S. W. & Thompson, E. A. Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics.* **48**, 361–372 (1992).
51. Excoffier, L. & Slatkin, M. Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. *Mol. Biol. Evol.* **12**, 921–927 (1995).

52. Lewontin, R. C. The interaction of selection and linkage. I. General considerations; heterotic models. *Genetics*. **49**, 49–67 (1964).
53. Hedrick, P. W. Gametic disequilibrium measures: proceed with caution. *Genetics*. **117**, 331–341 (1987).
54. Cramer, H. *Mathematical Models Of Statistics*. Princeton University Press, Princeton, NJ, p. 575 (1946).
55. Lancaster, A. K., Single, R. M., Solberg, O. D., Nelson, M. P. & Thomson, G. PyPop update—a software pipeline for large-scale multilocus population genomics. *Tissue Antigens*. **69**, 192–197 (2007).
56. Adapted from Regije Hrvatske, Wikipedia. Retrieved December 12, 2019, from [https://bs.wikipedia.org/wiki/Regije\\_Hrvatske#/media/Datoteka:Croatia\\_Macroregions.svg](https://bs.wikipedia.org/wiki/Regije_Hrvatske#/media/Datoteka:Croatia_Macroregions.svg). Available for reuse under the Creative Commons Attribution-Share Alike 3.0 (CC BY-SA 3.0) Unported license (<https://creativecommons.org/licenses/by-sa/3.0/>), work by Tomobe03 using file: Croatia location map.svg by NordNordWest, CC BY-SA 3.0, <https://commons.wikimedia.org/w/index.php?curid=19320911> (2012).
57. Bhatla, N. Exon-Intron Graphic Maker. Available at, <http://wormweb.org/exonintron> (2012).
58. R Core Team R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, <http://www.R-project.org> (2013).

## Acknowledgements

This research was funded by the Intramural Research Programme of the Josip Juraj Strossmayer University (grant number VIF2017-MEFOS-10 and VIF2018-MEFOS-8) and the Palacky University Olomouc (grant number RVO 61989592, IGA PU: LF\_2019\_009/2020\_004). We thank Kristina Krizmanić, Tea Stipanović and Anđelka Bugarin for excellent technical support.

## Author contributions

S.T. and M.P. developed the research concept; S.T. and V.Ž. performed the experiments; S.T., V.Ž., K.S. and M.Š. validated results; S.T. and M.Š. performed formal analysis of NGS data; S.T. and S.M. collected samples; M.S. recruited patients; S.T. and M.P. wrote the main manuscript; M.Š. produced figures; M.P. and M.Š. reviewed and edited the main manuscript; L.J.G.O. and M.P. provided financial support.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-020-62175-9>.

**Correspondence** and requests for materials should be addressed to S.T., M.Š. or M.P.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

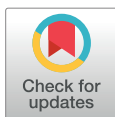


**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020



ELSEVIER



## ORIGINAL ARTICLE

## Association of Genetic Variants of Dopamine and Serotonin Receptors with Schizophrenia

Roksana Zakharyan,<sup>a,b</sup> Hovsep Ghazaryan,<sup>c</sup> Lenka Kocourkova,<sup>d</sup> Andranik Chavushyan,<sup>c</sup>  
Artur Mkrtchyan,<sup>e</sup> Veronika Zizkova,<sup>d</sup> Arsen Arakelyan,<sup>a,b</sup> and Martin Petrek<sup>d</sup><sup>a</sup>*Institute of Molecular Biology NAS RA, Yerevan, Armenia*<sup>b</sup>*Russian-Armenian, University, Yerevan, Armenia*<sup>c</sup>*Andranik Chavushyan, Institute of Molecular Biology NAS RA, Yerevan, Armenia*<sup>d</sup>*Department of Pathological Physiology, Faculty of Medicine and Dentistry, Palacky University Olomouc, Olomouc, Czech Republic*<sup>e</sup>*Department of Psychiatry, National Institute of Health, MH RA, Yerevan, Armenia*

Received for publication May 3, 2019; accepted December 16, 2019 (ARCMED\_2019\_393).

**Background.** Several studies indicated that antipsychotic treatment response and side effect manifestation can be different due to inter-individual variability in genetic variations.

**Aim of the study.** Here we perform a case-control study to explore a potential association between schizophrenia and variants within the antipsychotic drug molecular targets (DRD1, DRD2, DRD3, HTR2A, HTR6) and metabolizing enzymes (CYP2D6, COMT) genes in Armenian population including also analysis of their possible relationship with disease clinical symptoms.

**Methods.** A total of 18 SNPs was studied in patients with schizophrenia ( $n = 78$ ) and healthy control subjects ( $n = 77$ ) using MassARRAY genotyping. Results: We found that two studied genetic variants, namely DRD2 rs4436578\*C and HTR2A rs6314\*A are underrepresented in the group of patients compared to healthy subjects. After the correction for multiple testing, the rs4436578\*C variant remained significant while the rs6314\*A reported borderline significance. No significant differences in minor allele frequencies for other studied variants were identified. Also, a relationship between the genotypes and age of onset as well as disease duration has been detected.

**Conclusions.** The DRD2 rs4436578\*C genetic variant might have protective role against schizophrenia, at least in Armenians. © 2019 IMSS. Published by Elsevier Inc.

**Key Words:** Association, Schizophrenia, Genetic variant, Antipsychotic treatment, Dopamine, Serotonin, Metabolizing enzyme.

## Introduction

Several studies and clinical practice indicated that treatment response to antipsychotic drugs and the occurrence of side effects even in case of identical treatment can be different (1). This is mainly due to large inter-individual variability in genetic variations playing a substantial role among the factors that influence treatment efficacy (disease progression and outcome) (2). Studies of siblings and twins showed that variations in genes coding for molecular

targets and metabolizing enzymes of antipsychotic medications can significantly influence clinical outcome and side effects manifestation (2–8). In particular, one of drug metabolizing enzymes, catechol-O-methyltransferase (COMT), was considered as a key regulator of prefrontal cortex dopamine availability and also explored in relation to cognitive and executive functioning altered in schizophrenia (9–12). Moreover, it was shown that dopamine (D) receptor polymorphisms can contribute to positive symptom reduction, while negative symptom improvement was influenced by polymorphisms of genes involved in 5-HT neurotransmission (13). In addition, dopamine D1, D3, D4, and D5 receptor polymorphisms were associated with treatment response and increased risk for treatment tolerance (5). Based on these and other findings several

Address reprint requests to: Roksana Zakharyan, Senior Researcher Laboratory of Human Genomics and Immunomics Institute of Molecular Biology NAS RA, Russia; Phone: +37495542066; FAX: +37410282061; E-mail: r\_zakharyan@mb.sci.am

antipsychotic treatment predictive tests such as AmpliChip® CYP450 Test, DMET™ Plus Panel, the Genecept™ Assay, the Genomas HILomet PhyzioType™ System, and the GeneSight® Test based on the genes involved in drug metabolism such as cytochrome P450 family enzymes, were validated and introduced (14). These tests are mainly designed based on the knowledge about polymorphisms and their rates in populations present in the 1000 Genomes and HapMap projects, and thus, the accuracy of these genetic tests in other, specific populations is unknown. Importantly, up to date there is no published data about the prevalence of those mutations in Armenians.

Here we perform a case-control study to explore a potential association of genetic variants within the antipsychotic drug molecular targets (*DRD1*, *DRD2*, *DRD3*, *HTR2A*, *HTR6*) and metabolizing enzymes (*CYP2D6*, *COMT*) with schizophrenia in Armenian population and their possible relationship with disease clinical symptoms.

## Materials and Methods

### Study Population

In this study 78 patients with schizophrenia and 76 healthy control subjects were enrolled. Clinical and demographic data of the subjects enrolled in shown in Table 1. All patients were unrelated individuals of Armenian nationality recruited from the National Center of Mental Health Care of the Ministry of Health of the Republic of Armenia (MH RA) and Department of Psychiatry of the Yerevan State Medical University. The diagnosis of paranoid schizophrenia was made by two independent experienced psychiatrists, according to the presence of the relevant symptoms (ICD-10 code: F20.0) (15). All patients were treated with typical and/or atypical antipsychotics. Healthy control subjects were recruited among the blood donors of the Hematology Center after prof. R. Yolyan MH RA. Exclusion criteria for all study subjects included any serious neurological, endocrine, oncological, inflammatory, autoimmune, cerebrovascular, cardiovascular, metabolic or other disorder. All subjects gave their informed consents to participate

**Table 1.** Clinical and demographic data of subjects involved in the current study

Parameter	SCZ	Controls
Number of subjects involved	78	76
Age (M ± σ), years	48.9 ± 11.7	38.1 ± 10.8
Minimal age	20	23
Maximal age	70	65
Male/female	60/18	33/44
Age of disease first manifestation (M ± σ), years	29.0 ± 9.8	-
Duration of disease (M ± σ), years	19.9 ± 12.0	-
Family history of schizophrenia	36/42	-

in the study, which was approved by the Ethical Committee of the Institute of Molecular Biology of the National Academy of Sciences RA (IRB #00004079).

### Collection of Blood Samples and Genomic DNA Extraction

A total of 5 mL of the EDTA anticoagulated venous blood was collected from each patient and healthy control subject. Genomic DNA was isolated from peripheral blood leukocytes according to the modification of salting out method with a simple introduction of chloroform step and stored at -30°C until further use at the Laboratory of Human Genomics and Immunomics, Institute of Molecular Biology of the National Academy of Sciences RA. For all extracted DNA samples measurements of concentrations and purity have been performed. Only highly purified samples with A260/A280 ratio within the 1.8–2.0 range and concentration more or equal to 10 ng/ul were enrolled in this study. The working concentration was adjusted if applicable to 10 ng/ul using dilution of initial samples.

### SNP Selection in Candidate Genes

Single nucleotide polymorphisms (SNPs) within the genes coding for dopamine (*DRD1*, *DRD2*, *DRD3*) and serotonin receptors (*HTR2A*, *HTR6*) as well as antipsychotic drug metabolizing enzymes *CYP2D6* and *COMT* have been selected based on either SNP-drug response association reported in the Clinical Pharmacogenetics Implementation Consortium (PharmKB, <http://www.pharmgkb.org/page/cpic>) database or based upon previously determined associations with treatment response in candidate-gene and/or genome wide association studies (1–3,5–8,16,17). Initial selection resulted in 42 SNPs (Supplementary Table 1).

We performed initial evaluation of minor allele frequency (MAF) for these 42 polymorphisms based on the GWAS dataset on 99 available Armenian subjects (18). A total of 21 SNPs were not annotated in microarray chip used in the initial publication, so we obtained MAF data for them from 1000 Genomes European population using LDlink application for exploring population-specific variant and haplotype structure (19). Next, we selected SNPs with MAF greater or equal 0.1. Finally, using SNPclip module of the LDlink software we have pruned the list of variants to retain the most representative ones ( $R^2$  threshold for LD pruning = 0.5) resulting in 12 SNPs. To this list, 6 SNPs were further added based on functional effect demonstration using various functional assays (20). Finally, an assay control panel containing 4 SNPs of the *CYP2D6* gene, namely, rs28371725 (with common allele *CYP2D6\*41*), rs16947 (common allele *CYP2D6\*17*), rs3892097 (common allele *CYP2D6\*4*), and rs1065852 (common allele *CYP2D6\*10*) was also used. The resulting SNP panel and their minor allele frequencies in 1000



Genomes or healthy Armenians are presented in the [Supplementary Table 2](#).

#### Assay Design, PCR Amplification and Genotyping

The panel construction and genotyping were performed according to the protocol described elsewhere (21). Briefly, an input list of 18 selected SNP IDs (rsSNP of each target) was provided and the following steps were performed: (a) the sequence for each rsSNP was retrieved from database and formatted accordingly, (b) the proximal SNPs for each rsSNP were identified from database, (c) optimal primer areas were identified that result in a unique amplicons containing a target for the extension primer. To avoid extension primer rejection due to insufficient known bases, the proximal base was replaced with inosine. (d) Finally, PCR and extension primers were designed and checked for false priming, hairpin/dimer formation. The primer multiplexes with mass separation of analytes (alleles) were created. For PCR amplification and single base extension (SBE) reaction, the primer pairs along with the extension primers were designed using Assay design suite v2.0. All primers used in this study are presented in [Supplementary Table 3](#).

These primers were multiplexed and genotyped using Sequenom MassARRAY platform integrating iPLEX® SBE reaction and MassARRAY® technology (Agena Bioscience, San Diego, CA, USA). The assay consists of an initial locus-specific PCR amplification followed by SBE using mass-modified dideoxynucleotide terminators of an oligonucleotide primer that anneals immediately upstream of the target polymorphic site. The distinct mass of extended primer traces the alternative alleles using MassARRAY Typer 4.0.20. For quality control (QC) step, we determined data missing rate per individuals and missing rate per SNP. Also, for QC of SNP genotyping, positive and negative template control samples were included in each assay plate. As a positive control a commercially available DNA was used, while purified water serves as a negative control.

#### Statistical Analysis

The distribution of genotypes in cases and controls for each SNP was tested for correspondence ( $p > 0.05$ ) to Hardy-Weinberg equilibrium (HWE) by Pearson's  $\chi^2$  test. Disease trait SNP association was tested using logistic regression model with age and gender included as covariates. Analysis assuming additive, dominant and recessive models was performed. For multiple testing adjustment Bonferroni, and Benjamini & Hochberg metrics were used. All calculations were performed using PLINK whole genome association analysis toolset version 1.09 (22). Logistic regression analysis accounting for additive, dominant and recessive models was calculated with following options:

Assuming Additive Model:

```
/plink --noweb --file data/data --no-parents --no-sex
--allow-no-sex --adjust --logistic --covar data/data.cov
```

Assuming dominant model:

```
/plink --noweb --file data/data --no-parents --no-sex
--allow-no-sex --adjust --logistic --dominant --covar
data/data.cov
```

Assuming recessive model:

```
/plink --noweb --file data/data --no-parents --no-sex
--allow-no-sex --adjust --logistic --recessive --covar
data/data.cov
```

Where data.ped, data.map and data.fam contains genotypes and case-control status information, and data.cov contained gender and age information as covariates. Adjusted  $p$  \*/-values  $< 0.05$  were considered significant. The power of association under different models was calculated using “genpwr” R package (23).

In order to evaluate the association between studied SNPs and treatment prescription, drug dosages were converted into haloperidol and risperidone equivalents using a standardized method for comparing exposure to various antipsychotic medications of first- and second-generation (24). Quantitative trait association with drugs as well as age of onset and disease duration was performed using PLINK quantitative trait linear association (22). Linkage disequilibrium (LD) between investigated loci were calculated using SNP analyzer 2.0.

## Results

### Case-control Association Study

Here we studied antipsychotic drug response-associated 22 SNPs in the drug molecular targets and metabolizing enzymes genes in patients with schizophrenia and healthy subjects.

Genotyping was not successful and low signal was obtained for four SNPs (*COMT* rs165599, *COMT* rs933271, *DRD2* rs1076560, and *HTR2A* rs6311) out of all studied. The distribution of genotypes of 17 SNPs complied with HWE both in the patients with and controls, while *CYP2D6* rs1135840 was not ([Supplementary Table 4](#)). Genotyping of control panel was successful in all samples.

A logistic regression analysis was performed using three different models (additive, recessive, dominant). Two out of 17 studied SNPs were associated with schizophrenia ([Table 2](#)). Further information for 15 remaining genetic variants is shown in [Supplementary Tables 5–7](#).

After multiple corrections only *DRD2* gene rs4436578\*C variant remained significant while the rs6314\*A showed borderline significance. We found that this variant is associated with schizophrenia in Armenians and might have a protective effect against the disease.

LD values for all studied SNPs are presented in the [Supplementary Table 8](#). According to the data obtained,

**Table 2.** Association table for the *DRD2* gene rs4436578 and *HTR2A* genetic variants using additive, dominant and recessive models adjusted for age and gender

Chr.N	SNP	Location	Minor	Model	NMISS	OR	SE	95% CI	Stat	<i>p</i>	BONF	FDR_BH
11	rs4436578	113436043	C	additive	154	0.23	0.45	0.10–0.56	–3.27	0.0001	0.02	0.02
11	rs4436578	113436043	C	dominant	154	0.20	0.49	0.08–0.51	–3.33	0.0009	0.02	0.02
11	rs4436578	113436043	C	recessive	154	0.30	1.80	0.01–10.36	–0.66	0.51	1.00	1.00
13	rs6314	46834899	A	additive	154	0.28	0.49	0.11–0.73	–2.59	0.01	0.17	0.09
13	rs6314	46834899	A	dominant	154	0.23	0.55	0.08–0.67	–2.68	0.01	0.13	0.07
13	rs6314	46834899	A	recessive	154	0.27	1.60	0.01–6.29	–0.81	0.42	1.00	1.00

Odds ratios (OR), standard error (SE), 95 CI%, Bonferroni and FDR-corrected *p*-values are presented.

strong linkage between the rs6280 and rs167771 (chr.3), rs1079597 and rs1079598 (chr.11), rs2020917 and rs5993883 (chr.22). The strongest association for the rs1065852 and rs3892097 (chr.22) was detected.

#### Association of Prescribed Treatment with the Genotypes Obtained

The patients enrolled were treated with a variety of drugs including quietapine, triptazin, haloperidol, clozapine, fluphenazine, levomepromazine combined with anxiolytics. To make abovementioned correlations between standard homozygotes and minor allele carriers (if applicable) for studied SNPs, all dosages prescribed by psychiatrist(s) were converted to haloperidol equivalents using a standardized method for comparing exposure to various antipsychotic medications of first- and second-generation (24). Our analysis showed that only when comparing the mean treatment dosage of haloperidol for rs6280 CC homozygotes vs. T allele carriers, a significant difference was obtained ( $2.49 \pm 0.64$  mg/d vs.  $1.04 \pm 0.36$  mg/d,  $p = 0.039$ ). Further, the mean dosages prescribed were converted to risperidone equivalents and potential associations with genotypes (standard homozygotes vs. minor allele carriers and standard allele carriers vs. minor homozygotes) using the same method was explored. None of the genotypes obtained was associated with the converted risperidone dosages ( $p > 0.05$ ). According to the literature data available, there is no study on the relationship between the rs6280 genotypes and treatment response to haloperidol in schizophrenia patients. However, it was shown that schizophrenia patients with CC genotype had poorer response to clozapine compared with TT genotype (25). Also, this genetic variant was associated with the *DRD1* rs686 polymorphism in the same study (25). In patients with autism, Correia CT, et al. for the first time reported that the rs6280 C allele had greater symptom improvement and increased response to risperidone (26) whereas another study indicated decreased intolerability of methylphenidate (27). Also, predictors of clinical improvement with risperidone therapy among the *HTR2A* c.–1438G >A, *DRD3* Ser9Gly, *HTR2C* c.995G >A, and *ABCBI* 1236C >T polymorphisms were found (26). In case of treatment with olanzapine, CC was previously associated with greater

positive symptom improvement and positive symptom remission in schizophrenia (28). The C allele carriers had decreased risk of gastrointestinal toxicity in patients with Parkinson's disease treated with levodopa (29) while in Chinese patients this genotype was associated with decreased response to pramipexole (30). In Slovenian patients the *DRD2* gene rs1799732 (studied here) and *DRD3* gene rs6280 variants weren't associated neither with clinical symptoms nor the occurrence of treatment-resistant schizophrenia during antipsychotic treatment (31). These differences might appear due to the fact that relatively low dosages of the prescribed antipsychotics, converted to risperidone or variability between populations.

#### Association with Clinical Data

Relationship between the clinical-demographic data of patients and studied parameters has been performed. Correlation analysis carried out has identified that the genotypes are linked to the disease onset and duration. In particular, we found that the rs6314\*A minor allele and rs5030655\*del are positively correlated with disease duration ( $p = 0.03$  and  $0.11$ , respectively). Positive correlations between the age of schizophrenia onset and lorazepam prescribed ( $p = 0.033$ ), family history of schizophrenia ( $p = 0.045$ ) and the *5HTR6* rs1805054\*T allele ( $p = 0.009$ ) were detected.

#### Discussion and Conclusions

Pharmacogenetic approaches are widely used in the developed countries to achieve the most optimal treatment response reducing disease symptoms and minimizing manifestation of side effects (1). Such genetic factors as variants for drug metabolizing enzymes and molecular targets, play an important role in individual response and treatment efficacy (2). Therefore, studies focused on the identification of pharmacogenetic markers that contribute and predict antipsychotic treatment response might assist the development of genetically-driven approaches for moving to a personalized medicine particularly in schizophrenia research.

In this study we assessed the potential association of schizophrenia with 18 SNPs within the antipsychotic drug

molecular targets and enzymes and found that the *DRD2* gene rs4436578\*C allele might have a protective effect against this disease and the *HTR2A* gene rs6314\*A variant might have borderline significance.

Growing evidence suggest that molecular mechanisms responsible for dopamine transmission are altered in a number of psychiatric disorders including schizophrenia (32). Dopamine dysregulation together with glutamate, GABA, acetylcholine, and serotonin alterations are involved in schizophrenia during different stages of the disease (32). According to the dopamine theory of schizophrenia, hyperactivity of DRD2 receptor neurotransmission in subcortical and limbic brain regions is responsible for positive symptoms of this disease, while negative and cognitive symptoms are related to the hypoactivity of DRD1 receptor neurotransmission in the prefrontal cortex (33). Besides, DRD2 antagonism was considered as an essential pharmacological target for antipsychotic drugs, therefore, affecting its therapeutic efficacy depending on their inhibitory potency (34). The pioneer study by Carlsson and Lindqvist propose dopamine receptor blockade by chlorpromazine and haloperidol (35). The increased subcortical release of dopamine leads to the manifestation of disease positive symptoms such as hallucinations and delusions, most probably, through a disturbed cortical pathway (36). A recently performed gene-gene interaction analysis has identified a significant four-way interaction of SNPs included *DRD2* gene rs4436578 with verbal fluency in healthy subjects (37).

It has been reported that among others the *DRD2* gene rs4436578 SNP has an effect on antipsychotic drug-induced weight gain in long-term treated schizophrenia patients and that the rs4436578 CC together with C allele containing haplotype have been considered as a risk factor for weight gain (38,39).

So far, Kaur G, et al. (2017) has studied the effect of the *DRD2*, *5HT2A*, and *CYP2D6* gene polymorphisms on response to risperidone in Indian patients with schizophrenia (40). It has been shown that the presence of certain genotypes of the *DRD2* and *5HT2A* in patients are more likely associated with non-response to risperidone (40). Also, it has been shown that the *DRD2* Taq1A (rs1800497 C > T) genotype has no effect on the number of cigarettes smoked per day by patients with schizophrenia but female smokers with GG genotype smoked more compared to A allele carriers (41). Another SNP, C957T (rs6277), within the same gene has been linked to the executive functioning and early life stress. In particular, it has been detected that rs6277 CC homozygotes have significantly reduced performance in spatial working memory, planning, and those with early life stressful experience exhibit impaired performance in the processing test (42). Alladi CG, et al. (2017) has studied schizophrenia patients treated with risperidone during 4–8 weeks and has found that the *DRD2* gene –141C (rs1799732) Del carriers are significantly associated with increased prolactin levels in

response to treatment (43). Later with bigger sample size and additional polymorphisms enrolled they showed no association of Taq1A and *HTR2A* –1438 G/A, 102T/C and *HTR2C* –759 C/T with antipsychotic response to risperidone (44). For the same polymorphism Takase M, et al. (2017) has recently found a relationship between Del carriers and high-dose pharmacotherapy in schizophrenia (45).

Previously, a rs1076560 regulatory polymorphism of the *DRD2* gene has been identified as a genetic risk factor for schizophrenia and has been involved in schizophrenia intermediate phenotypes in the meta-analysis (46). Concerning the second positive finding of this study, a meta-analysis has indicated the rs6311 (–1438A/G) polymorphism of the *HTR2A* gene as a risk factor for schizophrenia but not major depressive or bipolar disorders (47). Importantly, the dependence of association from the ethnic origin of the study population has been detected (47). No association of another SNP (rs6313, T102C) of the same gene has been found in Chinese large samples with schizophrenia, bipolar disorder, and major depressive disorder (48). While no allele or genotype association has been detected, the combination of the *HTR2A* gene rs6311/rs6313 polymorphisms shows to predict negative symptoms performance in case of aripiprazole treatment in schizophrenia patients (49). Recently, rs6313\*C allele carriers were considered as having higher risk of schizophrenia development in Greek patients with schizophrenia (50). Nisenbaum LK, et al. (2016) has found that the *HTR2A* gene rs7330461 TT genotype is consistently associated with an increased treatment response to pomaglutamethad methionil compared to the AA genotype in Caucasian patients with schizophrenia (51). So far, rs1124491 and rs9567733 SNPs of the *DRD2* and *HTR2A* genes, respectively, have been linked to the risk of antipsychotic treatment-induced extrapyramidal syndrome (52). Pelka-Wysiecka J, et al. (2013) has shown that more (31 vs. 17%) Val/Val homozygous of the *COMT* gene rs4680 SNP were found in the non-deficit schizophrenia (NDS) compared to the deficit schizophrenia subgroup suggesting that these genetic factors must be taken into account when analyzing the reasons for the differentiation of schizophrenia subtypes (53). Study of *COMT-DRD4* interaction indicated that schizophrenia patients carrying Met allele and one or two *DRD4* 120 bp alleles had better response to clozapine (54). So far, an association of the *CYP2D6*\*4 1846G > A genetic variant and *CYP2D6* AA genotype with tardive dyskinesia in chronic schizophrenia patients has been detected (55).

To date, a number of studies were focused on the assessment of disease onset as a potential predictor of schizophrenia on clinical outcome. However, contradictory data were obtained, suggesting both positive and negative impact (56–58). Menezes NM, et al. later concluded that age of onset have no effect on clinical outcome (59). However, in original long-term studies, earlier age at onset has been associated with lower probability of symptomatic remission (57) and having more hospital admissions after

over 10 years since onset (60). Even with these inconsistent findings, it is not uncommon that in clinical textbooks onset early in life is regarded as a predictor for a poorer outcome.

Concerning the pharmacogenetic results, we have found that in case of *DRD3* rs6280, *DRD1* rs265981, and *DRD2* rs1079598 genotypes the prescribed treatment was in concordance with the recommended ones according to the clinical database. The application of pharmacogenetic approach widely used in the developed countries nowadays in mostly under development in low and middle income countries such as Armenia. Here, psychiatrists usually use trial-error approach in order to find the most appropriate treatment maximally suppressing disease symptoms and minimizing the manifestation of side effects. The comparison of the prescribed and recommended treatments described above suggested that, in fact, psychiatrists recommend the most optimal treatment but using a time-consuming approach instead of genetically-driven one.

Impairments in the dopamine system result from dopamine dysfunctions in the substantia nigra, ventral tegmental region, striatum, prefrontal cortex, and hippocampus. The “original dopamine hypothesis” states that hyperactive dopamine transmission results in schizophrenic symptoms. Blockage of dopamine receptors by chlorpromazine and haloperidol is the basic concept in psychiatry, and it has been shown that the positive symptoms of schizophrenia (hallucinations, delusions) are related to increased subcortical release of dopamine (61,62). The *HTR2A* gene rs6314\*A (Tyr) allele has been previously showed lowered antipsychotic binding affinity and decreased drug potency. Further, it has been detected that Tyr allele was significantly associated with poor response to clozapine treatment compared to the His allele (63–65). So far, Terzic T, et al. (2016) have evaluated the potential relationship between the rs6280 SNP and treatment prescribed to patients with psychopathological problems. Genotypes CT+TT are not associated with response to antipsychotics in people with schizophrenia as compared to genotype CC (31).

Further studies in larger sample size are required to identify pharmacogenetic markers of schizophrenia that might contribute to the development and implementation of genetically-driven treatment approaches in psychiatry in Armenia in the future.

In our study, we found underrepresentation of the *DRD2* rs4436578\*C and *HTR2A* rs6314\*A variants in patients with schizophrenia, however, after multiple corrections only rs4436578\*C remains significant and might have a protective role against the disease. These results contribute to current puzzle of schizophrenia causal variants. A relationship between the genotypes and age of onset as well as disease duration were detected.

It should be noted that this study has some limitations. Namely, relatively small size of control and diseased subjects was enrolled. Second, the limited number of SNPs was selected based on either SNP-drug response association

reported in PharmKB database or based upon previously determined associations with treatment response in candidate-gene and/or genome wide association studies mentioned in the manuscript. Finally, the important role of other genetic variants without the same genes not studied here must be additionally tested to extend the existing knowledge on contributory genetic factors of schizophrenia.

## Declaration of Absence of Conflicts of Interest

Authors declare no conflict of interests.

## Acknowledgments

This work was supported in part by RVO: 61989592 and projects reg. no. IGA PU: LF\_2017\_014/2019\_009, targeted research project “Genetic mapping of Armenian pathogenome (I-18/TN)”, State Committee of Science, Ministry of Education and Science, Republic of Armenia. This work was partly made possible by a research grant from the Armenian National Science and Education Fund (PI: Andranik Chavushyan, ANSEF # molbio-5342) based in New York, USA.

## Supplementary Data

Supplementary data related to this article can be found at <https://doi.org/10.1016/j.arcmed.2019.12.011>.

## References

1. Pouget JG, Shams TA, Tiwari AK, et al. Pharmacogenetics and outcome with antipsychotic drugs. *Dialogues Clin Neurosci* 2014; 16:555–566.
2. Ravyn D, Ravyn V, Lowney R, et al. CYP450 pharmacogenetic treatment strategies for antipsychotics: a review of the evidence. *Schizophr Res* 2013;149:1–14.
3. Bertolino A, Caforio G, Blasi G, et al. COMT Val158Met polymorphism predicts negative symptoms response to treatment with olanzapine in schizophrenia. *Schizophr Res* 2007;95:253–255.
4. Dudova I, Hrdlicka M. Successful use of olanzapine in adolescent monozygotic twins with catatonic schizophrenia resistant to electroconvulsive therapy: case report. *Neuro Endocrinol Lett* 2008;29: 47–50.
5. Lieberman JA, Stroup TS, McEvoy JP, et al. Effectiveness of antipsychotic drugs in patients with chronic schizophrenia. *N Engl J Med* 2005;353:1209–1223.
6. Meltzer HY, Massey BW. The role of serotonin receptors in the action of atypical antipsychotic drugs. *Curr Opin Pharmacol* 2011;11: 59–67.
7. Rebollo-Mesa I, Picchioni M, Shaikh M, et al. COMT (Val(158/108) Met) genotype moderates the impact of antipsychotic medication on verbal IQ in twins with schizophrenia. *Psychiatr Genet* 2011;21: 98–105.
8. Reynolds GP. The impact of pharmacogenetics on the development and use of antipsychotic drugs. *Drug Discov Today* 2007;12: 953–959.
9. Bosia M, Bechi M, Marino E, et al. Influence of catechol-O-methyltransferase Val158Met polymorphism on neuropsychological

- and functional outcomes of classical rehabilitation and cognitive remediation in schizophrenia. *Neurosci Lett* 2007;417:271–274.
10. Bosia M, Zanoletti A, Spangaro M, et al. Factors affecting cognitive remediation response in schizophrenia: the role of COMT gene and antipsychotic treatment. *Psychiatry Res* 2014;217:9–14.
  11. Ira E, Zannoni M, Ruggeri M, et al. COMT, neuropsychological function and brain structure in schizophrenia: a systematic review and neurobiological interpretation. *J Psychiatry Neurosci* 2013;38:366–380.
  12. Szoke A, Schurhoff F, Meary A, et al. Lack of influence of COMT and NET genes variants on executive functions in schizophrenic and bipolar patients, their first-degree relatives and controls. *Am J Med Genet B Neuropsychiatr Genet* 2006;141B:504–512.
  13. Reynolds GP, Yao Z, Zhang X, et al. Pharmacogenetics of treatment in first-episode schizophrenia: D3 and 5-HT2C receptor polymorphisms separately associate with positive and negative symptom response. *Eur Neuropsychopharmacol* 2005;15:143–151.
  14. Changasi AH, Shams TA, Pouget JG, et al. Genetics of antipsychotic drug outcome and implications for the clinician: into the limelight. *Transl Dev Psychiatry* 2014;2:24663.
  15. Organization WH. The international statistical classification of diseases and related health problems, 1992;. Geneva: World Health Organization.
  16. Buijnzeel D, Suryadevara U, Tandon R. Antipsychotic treatment of schizophrenia: an update. *Asian J Psychiatr* 2014;11:3–7.
  17. Zhang JP, Malhotra AK. Pharmacogenetics and antipsychotics: therapeutic efficacy and side effects prediction. *Expert Opin Drug Metab Toxicol* 2011;7:9–37.
  18. Haber M, Mezzavilla M, Xue Y, et al. Genetic evidence for an origin of the Armenians from Bronze Age mixing of multiple populations. *Eur J Hum Genet* 2016;24:931–936.
  19. Lobbstaal J, Leurgans M, Arntz A. Inter-rater reliability of the Structured Clinical Interview for DSM-IV Axis I Disorders (SCID I) and Axis II Disorders (SCID II). *Clin Psychol Psychother* 2011;18:75–79.
  20. Kang SG, Na KS, Lee HJ, et al. DRD2 genotypic and haplotype variation is associated with improvements in negative symptoms after 6 weeks' amisulpride treatment. *J Clin Psychopharmacol* 2015;35:158–162.
  21. Kishore A, Zizkova V, Kocourkova L, et al. A dataset of 26 candidate gene and pro-inflammatory cytokine variants for association studies in idiopathic pulmonary fibrosis: frequency distribution in normal Czech population. *Front Immunol* 2015;6:476.
  22. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a toolset for whole-genome association and population-based linkage analysis. *Am J Hum Genet* 2007;81:559–575.
  23. Moore C, Jacobson S. *genpwr: Power calculations under genetic model misspecification. R package version 1.0.1, 2019;. <https://CRAN.R-project.org/package=genpwr>. Accessed November 28, 2019.*
  24. Andreasen NC, Pressler M, Nopoulos P, et al. Antipsychotic dose equivalents and dose-years: a standardized method for comparing exposure to different drugs. *Biol Psychiatry* 2010;67:255–262.
  25. Hwang R, Souza RP, Tiwari AK, et al. Gene-gene interaction analyses between NMDA receptor subunit and dopamine receptor gene variants and clozapine response. *Pharmacogenomics* 2011;12:277–291.
  26. Correia CT, Almeida JP, Santos PE, et al. Pharmacogenetics of risperidone therapy in autism: association analysis of eight candidate genes with drug efficacy and adverse drug reactions. *Pharmacogenomics J* 2010;10:418–430.
  27. McCracken JT, Badashova KK, Posey DJ, et al. Positive effects of methylphenidate on hyperactivity are moderated by monoaminergic gene variants in children with autism spectrum disorders. *Pharmacogenomics J* 2014;14:295–302.
  28. Adams DH, Close S, Farnen M, et al. Dopamine receptor D3 genotype association with greater acute positive symptom remission with olanzapine therapy in predominately caucasian patients with chronic schizophrenia or schizoaffective disorder. *Hum Psychopharmacol* 2008;23:267–274.
  29. Rieck M, Schumacher-Schuh AF, Altmann V, et al. Association between DRD2 and DRD3 gene polymorphisms and gastrointestinal symptoms induced by levodopa therapy in Parkinson's disease. *Pharmacogenomics J* 2018;18:196–200.
  30. Liu YZ, Tang BS, Yan XX, et al. Association of the DRD2 and DRD3 polymorphisms with response to pramipexole in Parkinson's disease patients. *Eur J Clin Pharmacol* 2009;65:679–683.
  31. Terzic T, Kastelic M, Dolzan V, et al. Genetic polymorphisms in dopaminergic system and treatment-resistant schizophrenia. *Psychiatr Danub* 2016;28:127–131.
  32. Brisch R, Saniotis A, Wolf R, et al. The role of dopamine in schizophrenia from a neurobiological and evolutionary perspective: old fashioned, but still in vogue. *Front Psychiatry* 2014;5:47.
  33. Toda M, Abi-Dargham A. Dopamine hypothesis of schizophrenia: making sense of it all. *Curr Psychiatry Rep* 2007;9:329–336.
  34. Seeman P, Lee T. Antipsychotic drugs: direct correlation between clinical potency and presynaptic action on dopamine neurons. *Science* 1975;188:1217–1219.
  35. Baumeister AA. The chlorpromazine enigma. *J Hist Neurosci* 2013;22:14–29.
  36. O'Donnell P, Grace AA. Dysfunctions in multiple interrelated systems as the neurobiological bases of schizophrenic symptom clusters. *Schizophr Bull* 1998;24:267–283.
  37. Zhang S, Zhang M, Zhang J. Association of COMT and COMT-DRD2 interaction with creative potential. *Front Hum Neurosci* 2014;8:216.
  38. Hong CJ, Liou YJ, Bai YM, et al. Dopamine receptor D2 gene is associated with weight gain in schizophrenic patients under long-term atypical antipsychotic treatment. *Pharmacogenet Genomics* 2010;20:359–366.
  39. Muller DJ, Zai CC, Sicard M, et al. Systematic analysis of dopamine receptor genes (DRD1-DRD5) in antipsychotic-induced weight gain. *Pharmacogenomics J* 2012;12:156–164.
  40. Kaur G, Gupta D, Chavan BS, et al. Identification of genetic correlates of response to risperidone: findings of a multicentric schizophrenia study from India. *Asian J Psychiatr* 2017;29:174–182.
  41. Hirasawa-Fujita M, Bly MJ, Ellingrod VL, et al. Genetic variation of the Mu opioid receptor (OPRM1) and dopamine D2 receptor (DRD2) is related to smoking differences in patients with schizophrenia but not bipolar disorder. *Clin Schizophr Relat Psychoses* 2017;11:39–48.
  42. Naito H, Takeda T, Hanai J. Failure to detect dopaminergic neuroblastoma in a mass-screening system. *J Pediatr Surg* 1995;30:1317–1318.
  43. Alladi CG, Mohan A, Shewade DG, et al. Risperidone-induced adverse drug reactions and role of DRD2 (-141 C Ins/Del) and 5HTR2C (-759 C>T) genetic polymorphisms in patients with schizophrenia. *J Pharmacol Pharmacother* 2017;8:28–32.
  44. Alladi CG, RajKumar RP, Adithan S, et al. Dopamine (DRD2) and serotonin (HTR2A, 2C) receptor gene polymorphisms do not influence early response to risperidone in South Indian patients with schizophrenia. *Fundam Clin Pharmacol* 2019;33:355–364.
  45. Takase M, Kanahara N, Oda Y, et al. The impacts of dopamine D2 receptor polymorphism and antipsychotic dosage on dopamine supersensitivity psychosis in schizophrenia. *Schizophr Res* 2017;190:182–183.
  46. Luykx JJ, Broersen JL, de Leeuw M. The DRD2 rs1076560 polymorphism and schizophrenia-related intermediate phenotypes: A systematic review and meta-analysis. *Neurosci Biobehav Rev* 2017;74:214–224.

47. Gu L, Long J, Yan Y, et al. HTR2A-1438A/G polymorphism influences the risk of schizophrenia but not bipolar disorder or major depressive disorder: a meta-analysis. *J Neurosci Res* 2013;91:623–633.
48. Tan J, Chen S, Su L, et al. Association of the T102C polymorphism in the HTR2A gene with major depressive disorder, bipolar disorder, and schizophrenia. *Am J Med Genet B Neuropsychiatr Genet* 2014; 165B:438–455.
49. Yildiz SH, Akilli A, Bagcioglu E, et al. Association of schizophrenia with T102C (rs6313) and 1438 A/G (rs6311) polymorphisms of HTR2A gene. *Acta Neuropsychiatr* 2013;25:342–348.
50. Baritaki S, Rizos E, Zafiroopoulos A, et al. Association between schizophrenia and DRD3 or HTR2 receptor gene variants. *Eur J Hum Genet* 2004;12:535–541.
51. Nisenbaum LK, Downing AM, Zhao F, et al. Serotonin 2A receptor SNP rs7330461 association with treatment response to pomalglumetad methionil in patients with schizophrenia. *J Pers Med* 2016;6. <https://doi.org/10.3390/jpm6010009>.
52. Mas S, Gasso P, Lafuente A, et al. Pharmacogenetic study of antipsychotic induced acute extrapyramidal symptoms in a first episode psychosis cohort: role of dopamine, serotonin and glutamate candidate genes. *Pharmacogenomics J* 2016;16:439–445.
53. Pelka-Wysiecka J, Wronski M, Jasiewicz A, et al. BDNF rs 6265 polymorphism and COMT rs 4680 polymorphism in deficit schizophrenia in Polish sample. *Pharmacol Rep* 2013;65:1185–1193.
54. Rajagopal VM, Rajkumar AP, Jacob KS, et al. Gene-gene interaction between DRD4 and COMT modulates clinical response to clozapine in treatment-resistant schizophrenia. *Pharmacogenet Genomics* 2018; 28:31–35.
55. Ivanova SA, Filipenko ML, Vyalova NM, et al. CYP1A2 and CYP2D6 gene polymorphisms in schizophrenic patients with neuroleptic drug-induced side effects. *Bull Exp Biol Med* 2016;160:687–690.
56. Altamura AC, Bassetti R, Sassella F, et al. Duration of untreated psychosis as a predictor of outcome in first-episode schizophrenia: a retrospective study. *Schizophr Res* 2001;52:29–36.
57. Juola P, Miettunen J, Veijola J, et al. Predictors of short- and long-term clinical outcome in schizophrenic psychosis: the Northern Finland 1966 Birth Cohort study. *Eur Psychiatry* 2013;28: 263–268.
58. Stefanopoulou E, Lafuente AR, Fonseca AS, et al. Global assessment of psychosocial functioning and predictors of outcome in schizophrenia. *Int J Psychiatry Clin Pract* 2011;15:62–68.
59. Menezes NM, Arenovich T, Zipursky RB. A systematic review of longitudinal outcome studies of first-episode psychosis. *Psychol Med* 2006;36:1349–1362.
60. Rabinowitz J, Levine SZ, Hafner H. A population based elaboration of the role of age of onset on the course of schizophrenia. *Schizophr Res* 2006;88:96–101.
61. Iversen SD, Iversen LL. Dopamine: 50 years in perspective. *Trends Neurosci* 2007;30:188–193.
62. Madras BK. History of the discovery of the antipsychotic dopamine D2 receptor: a basis for the dopamine hypothesis of schizophrenia. *J Hist Neurosci* 2013;22:62–78.
63. Arranz MJ, Collier DA, Munro J, et al. Analysis of a structural polymorphism in the 5-HT2A receptor and clinical response to clozapine. *Neurosci Lett* 1996;217:177–178.
64. Arranz MJ, Munro J, Owen MJ, et al. Evidence for association between polymorphisms in the promoter and coding regions of the 5-HT2A receptor gene and response to clozapine. *Mol Psychiatry* 1998;3:61–66.
65. Masellis M, Basile V, Meltzer HY, et al. Serotonin subtype 2 receptor genes and clinical response to clozapine in schizophrenia patients. *Neuropsychopharmacology* 1998;19:123–132.