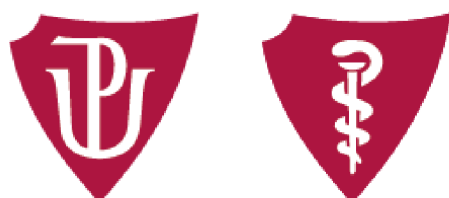


**Univerzita Palackého v Olomouci**

**Lékařská fakulta**



**Dizertační práce**

**Variabilita imunitních genů u pacientů  
se sarkoidózou**

**Mgr. Kateřina Sikorová**

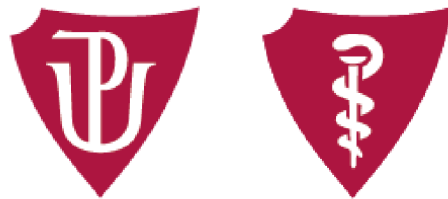
**Školitel: Prof. MUDr. Martin Petřek, CSc.**

**Ústav patologické fyziologie**

**Obor: Fyziologie a patologická fyziologie**

**Olomouc, 2022**

**Palacký University Olomouc**  
**Faculty of Medicine and Dentistry**



**Dissertation thesis**

**Immune gene variability in Sarcoidosis  
patients**

**Mgr. Kateřina Sikorová**

**Supervisor: Prof. MUDr. Martin Petřek, CSc.**

**Department of Pathological Physiology**

**Field of study: Physiology and Pathological Physiology**

**Olomouc, 2022**

„Prohlašuji, že jsem předloženou dizertační práci vypracovala samostatně pod vedením prof. MUDr. Martina Petřka, CS.c. a za použití citované literatury.“

V Olomouci .....

.....

Mgr. Kateřina Sikorová

## **PODĚKOVÁNÍ**

Chtěla bych poděkovat prof. MUDr. Martinu Petřkovi, CSc. za příležitost studovat fyziologii a patologickou fyziologii na Lékařské fakultě. Děkuji za vedení studií, poskytnutý biologický materiál a laboratorní zázemí. Děkuji za ochotu, trpělivost a za odborné konzultace včetně spolupráce při publikování.

Děkuji Mgr. Lence Kocourkové a Mgr. Veronice Žižkové Ph.D., které se se mnou podílely na přípravě biologického materiálu a vlastním měření. Děkuji za ochotu a trpělivost při vysvětlování, za pomoc při optimalizaci metod a za vedení při analýzách. Ing. Rastislavu Slavkovskému Ph.D. děkuji za pomoc při získávání dat z přístroje MiSeq Illumina.

## **GRANTOVÁ PODPORA**

Studium jednonukleotidových polymorfismů a HLA variant u pacientů se sarkoidózou bylo podpořeno grantovým projektem LO1304, IGA UP: LF 2016\_009, IGA UP: LF2017\_014, IGA UP: LF2018\_015, IGA UP: LF2019\_009, IGA UP: LF2020\_004, IGA UP: LF2021\_014, IGA UP: LF2022\_005 a CZ.02.1.01/0.0/0.0/16\_019/0000868 (ENOCH). Výzkum na české populaci byl podpořen grantem ministerstva zdravotnictví NV18-05-00134 a sběr vzorků v korejské populaci byl podpořen grantem NRF-2019R1A2C2008541.

## Souhrn

Sarkoidóza je autoimunitní onemocnění postihující primárně plíce, ale i široké spektrum dalších orgánů. U větší části pacientů je přítomna její mírnější forma, může však dojít i k progresi onemocnění a vzniku plicní fibrózy. Včasné odhalení této skupiny pacientů, či naopak identifikace pacientů, u kterých nasazení léčby není potřeba, by mohlo vést k lepšímu managementu onemocnění. Etiologie sarkoidózy stále není zcela objasněná. Velmi podstatnou úlohu v ní ovšem mají především geny imunitního systému. Tato práce byla zaměřena na typizaci HLA variant a jednonukleotidových polymorfismů v genech spojených především s imunitním systémem. Sarkoidóza, jak z předešlých studií vyplývá, je velmi heterogenní onemocnění a jeho výskyt, ale i projevy závisí na etnickém původu pacientů. Byly analyzovány dvě středoevropské populace – česká a polská, dále pak řecká populace a korejská populace. Variabilita HLA lokusů byla studována i ve zdravé chorvatské populaci.

Analýza HLA lokusů byla provedena pomocí NGS typizace s vysokým rozlišením na 4 pole. Jednonukleotidové polymorfismy byly identifikovány pomocí metody MassArray, která je založená na prodloužení o jednu bázi a následnou analýzu MALDI-TOF.

V této práci byly replikovány výsledky předchozích studií, ale i objeveny nové polymorfismy a HLA varianty asociované se vznikem, či průběhem sarkoidózy. Byla prokázána také spojitost etnického původu s výskytem HLA variant a polymorfismů v daných populacích. Ve středoevropských populacích byla významná spojitost polymorfismů spjatých s MHC oblastí s příznivější prognózou sarkoidózy. U řecké a české populace byla potvrzena spojitost polymorfismu v genu *ANXA11* jako protektivního pro vznik sarkoidózy. V řecké populaci, a především pak v populaci korejské, byl naopak znatelný silný vliv polymorfismů v genu *TGF-β* na vznik a průběh sarkoidózy.

Byla popsána diverzita HLA variant u východních Chorvatů, která poukazuje na charakteristické detaily alel a haplotypů, které odpovídají složité populační historii studované skupiny Chorvatů a této geografické oblasti. Analýzou HLA lokusů byly v české populaci identifikovány jako rizikové pro vznik sarkoidózy varianty HLA-DQB1\*06:02, HLA-DQB1\*06:04 a DRB1\*11:01. S příznivou prognózou byly spojené varianty, které jsou součástí 8.1 ancestrálního haplotypu. Se závažnějším průběhem byly asociovány varianty HLA-DQB1\*05:03, DRB1\*14:54 a HLA-DQA1\*01:04.

V korejské populaci byly s rizikem sarkoidózy asociované varianty HLA-C\*03:04, HLA-DRB1\*12:01 a HLA-DRB1\*14:54. Naopak s příznivější prognózou byly u korejské populace spojeny varianty HLA-DRB1\*12:01 či HLA-DQA1\*01:02.

## Summary

Sarcoidosis is an autoimmune disease primarily affecting the lungs but also a wide range of other organs. A milder form of sarcoidosis is present in most patients, but the progression of the disease and pulmonary fibrosis can also occur. Early detection of this group of patients or, conversely, identification of patients in whom treatment is not needed could lead to better disease management. The etiology of sarcoidosis is still not fully understood, but immune system genes play a crucial role. In this study, we focused on HLA variant typing and single nucleotide polymorphism typing in genes mainly associated with the immune system. As previous studies have shown, sarcoidosis is heterogeneous; its incidence and manifestations depend on ethnicity. We analyzed two Central European populations - Czech and Polish, and the Greek and Korean populations. We also studied the variation of HLA loci in the healthy Croatian population.

HLA locus typing was performed using high-resolution NGS typing on 4 fields. Single nucleotide polymorphisms were identified using the *MassArray* method based on single base extension followed by MALDI-TOF analysis.

In this work, we replicated the results of previous studies and discovered new polymorphisms and HLA variants associated with the onset or course of sarcoidosis. We also demonstrated the importance of ethnic influences on the prevalence of HLA variants and polymorphisms in the populations. In Central European populations, polymorphisms associated with the MHC region were significantly associated with a more favorable prognosis of sarcoidosis. In the Greek and Czech populations, we confirmed the protective association of polymorphisms in the ANXA11. In contrast, in the Greek population and the Korean population, polymorphisms in the TGF- $\beta$  gene strongly influenced the development and progression of sarcoidosis.

HLA diversity in eastern Croats was described, which shows characteristic details of alleles and haplotypes that correspond to the complex population history of the Croatian group studied and this geographical region. By analysis of HLA loci, we identified HLA-DQB1\*06:02, HLA-DQB1\*06:04, and DRB1\*11:01 as risk loci for sarcoidosis in the Czech population. Variants that are part of the 8.1 ancestral haplotype were associated with a favorable prognosis. A more severe course was associated with HLA-DQB1\*05:03, DRB1\*14:54, and HLA-DQA1\*01:04 variants. In the Korean population, variants HLA-C\*03:04, HLA-DRB1\*12:01, HLA-DRB1\*14:54, and HLA-DQA1\*05:08, for example, were associated with sarcoidosis risk. HLA-DRB1\*12:01 or HLA-DQA1\*01:02 variants have been associated with a better prognosis in the Korean population.

## Obsah

1	Úvod.....	4
1.1	Sarkoidóza .....	4
1.2	Epidemiologie.....	4
1.3	Etiologie sarkoidózy .....	6
1.4	Léčba .....	8
1.5	Patofyziologie sarkoidózy .....	9
1.6	Průběh nemoci .....	12
1.7	Mimoplicní postižení .....	13
1.8	Diagnostika.....	14
1.9	Biomarkery .....	15
1.9.1	Celogenomové asociační studie .....	17
1.10	HLA typizace.....	20
1.11	Vyšetření polymorfismů .....	23
2	Cíle práce .....	27
3	Materiál a metody .....	28
3.1	Biologický materiál .....	28
3.1.1	Česká populace.....	28
3.1.2	Polská populace.....	29
3.1.3	Řecká populace .....	30
3.1.4	Korejská populace .....	31
3.1.5	Chorvatská populace .....	32
3.2	Izolace DNA .....	32
3.2.1	Přístrojová izolace DNA .....	32
3.2.2	Izolace DNA vysolovací metodou .....	32
3.2.3	Měření koncentrace DNA .....	32
3.3	Analýza technologií MassArray® .....	33
3.3.1	Návrh primerů .....	33
3.3.2	Příprava produktů pro genotypizaci .....	35
3.4	End point genotypizace .....	37
3.5	NGS .....	38
3.6	Statistické zpracování dat .....	40
4	Výsledky .....	42



4.1	HLA analýza v české populaci .....	42
4.1.1	Asociace HLA variant s klinickými fenotypy sarkoidózy .....	43
4.1.1.1	Varianty HLA u pacientů s Löfgrenovým syndromem (LS).....	43
4.1.1.2	Srovnání distribuce variant HLA v kontextu různých stadií na základě rtg. nálezu.....	43
4.1.1.3	Varianty HLA v kontextu klinického průběhu onemocnění (2 roky sledování).....	43
4.1.1.4	Varianty HLA v kontextu léčby kortikosteroidy .....	44
4.1.1.5	Varianty HLA u pacientů s mimoplicní manifestací sarkoidózy .....	44
4.1.1.6	Analýza vazebné nerovnováhy .....	44
4.2	SNP analýza v české populaci .....	45
4.2.1	Asociace SNPs s klinickými fenotypy sarkoidózy.....	47
4.2.1.1	Srovnání na základě rtg stadií.....	47
4.2.1.2	Srovnání pacientů s LS .....	48
4.2.1.3	Srovnání na základě průběhu onemocnění .....	49
4.2.1.4	Srovnání na základě potřeby léčby .....	50
4.2.1.5	Srovnání na základě přítomností mimoplicních lokalizací.....	51
4.2.1.6	Korelace výskytu polymorfismů s výskytem daných HLA variant .....	52
4.3	SNP analýza v polské populaci .....	52
4.3.1	Asociace SNP s klinickými fenotypy sarkoidózy .....	54
4.3.1.1	Srovnání na základě rtg. stadií.....	54
4.3.1.2	Srovnání na základě průběhu onemocnění .....	54
4.3.1.3	Srovnání na základě přítomnosti LS.....	55
4.4	HLA analýza v korejské populaci .....	56
4.4.1	Asociace HLA variant s klinickými fenotypy sarkoidózy .....	58
4.4.1.1	Srovnání distribuce variant HLA v kontextu různých stadií na základě rtg. nálezu.....	58
4.4.1.2	Srovnání HLA variant u pacientů s mimoplicní manifestací sarkoidózy .....	59
4.4.1.3	Varianty HLA v kontextu klinického průběhu onemocnění (2 roky sledování).....	59
4.5	SNP analýza v korejské populaci .....	60
4.5.1	Asociace SNP s klinickými fenotypy sarkoidózy .....	61
4.6	SNP analýza v řecké populaci .....	63
4.6.1	Asociace SNP s klinickými fenotypy sarkoidózy .....	64

4.6.2	Analýza HLA u chorvatských zdravých jedinců.....	67
5	Diskuze .....	69
5.1	HLA varianty asociované se sarkoidózou v české populaci.....	69
5.2	SNP asociované se sarkoidózou v české populaci.....	72
5.3	SNP asociované se sarkoidózou v polské populaci .....	74
5.4	HLA varianty asociované se sarkoidózou v korejské populaci .....	75
5.5	SNP asociované se sarkoidózou v korejské populaci .....	77
5.6	SNP asociované se sarkoidózou v řecké populaci .....	78
5.7	Analýza zdravé chorvatské populace .....	80
5.8	Spojitosť výskytu polymorfismů s geografickou polohou .....	81
6	Závěr .....	83
7	Seznam publikací a konferenčních příspěvků.....	85
7.1	Seznam publikací spojených s dizertační prací .....	85
7.2	Seznam ostatních publikací .....	85
7.3	Kapitola v monografii.....	85
7.4	Konferenční příspěvky .....	85
8	Seznam zkratk .....	87
9	Seznam literatury .....	91
10	Internetové odkazy k statistickým nástrojům a databázím.....	101
11	Přílohy .....	

# 1 Úvod

Intersticiální onemocnění jsou skupinou onemocnění charakteristickou zánětlivými a fibrotickými změnami intersticia nebo interalveolárních sept plic, které mají za následek nepoddajnost plic. Do této skupiny onemocnění řadíme i sarkoidózu.

## 1.1 Sarkoidóza

Sarkoidóza je známá již přes sto let. První případ sarkoidózy byl zaznamenán v roce 1877 v Londýně (Jain *et al.*, 2020).

Dle mezinárodní klasifikace nemocí řadíme sarkoidózu mezi poruchy mechanismu imunity. Rozděluje ji dále dle místa postižení na sarkoidózu plic, mízních uzlin, kůže a dalších, případně kombinovaných lokalizací.

Jedná se o onemocnění velmi symptomaticky rozmanité, a proto jsou jeho příčiny i etiologie stále předmětem výzkumu (Valeyre *et al.*, 2014). Onemocnění postihuje především plíce. Téměř polovina pacientů nemá žádné příznaky, u některých ovšem může onemocnění vyústit až v selhávání orgánů způsobené fibrózou a pulmonální hypertenzí (Starshinova *et al.*, 2020).

Stěžejním prvkem v diagnostice sarkoidózy je přítomnost nekaseifikujícího granulomu. K jeho tvorbě dochází především v plicích a lymfatických uzlinách, ale také na kůži a v játrech. K tvorbě granulomů dochází agregací hypertrofovaných makrofágů (Valeyre *et al.*, 2014). Neznámý antigen je zpracován aktivovanými makrofágy a podněcuje imunitní odpověď regulovanou T-lymfocyty a makrofágy. Toto zapříčiňuje uvolnění cytokinů, chemokinů a dalších mediátorů, které se dále podílejí na progresi sarkoidózy.

## 1.2 Epidemiologie

Sarkoidóza je rozšířená celosvětově, ale četnost jejího výskytu je odlišná mezi jednotlivými skupinami v závislosti na rase, regionu, ale i věku a pohlaví (Jeon *et al.*, 2020). Nejvyšší incidenci tohoto onemocnění, což je počet nových případů onemocnění za rok, pozorujeme u Afroameričanů (17.8 případů na 100 000 obyvatel) a v severní Evropě (Švédsko 11.5 případů na 100 000 obyvatel) (Arkema *et Cozier*, 2020; Yoon *et al.*, 2018). Tato práce byla zaměřena na populaci Česka, Polska, Řecka a Jižní Koreje.

V české populaci je incidence sarkoidózy 4.4 případů na 100 000 obyvatel (Arkema *et Cozier*, 2020) a prevalence sarkoidózy je 63.1 případů na 100 000 obyvatel (Kolek *et al.*, 1994). Nejvíce zasaženy jsou u českých pacientů, mimo plic, periferní lymfatické uzliny, vyskytuje se hyperkalciurie, zasažení kůže, jater, kostí, sleziny a očí (Žurková *et al.*, 2014).

Podle studií provedených na polské populaci se incidence sarkoidózy pohybuje mezi 4-7.5 případů na 100 000 obyvatel. Sarkoidóza je zde více rozšířená u mužů, které v průměru postihuje v nižším věku než ženy. Věk jedinců hospitalizovaných se sarkoidózou v posledních letech stoupá. Podíl žen mezi pacienty se sarkoidózou naopak klesá. Mimo plíce (32.1 % případů) je v této populaci časté především postižení lymfatických uzlin (18 % případů), kombinace postižení plic se zasažením lymfatických uzlin (25.8 % případů), případně zasažení kůže (1.4 % případů) (Bogdanet *et al.*, 2019; Kowalska *et al.*, 2014).

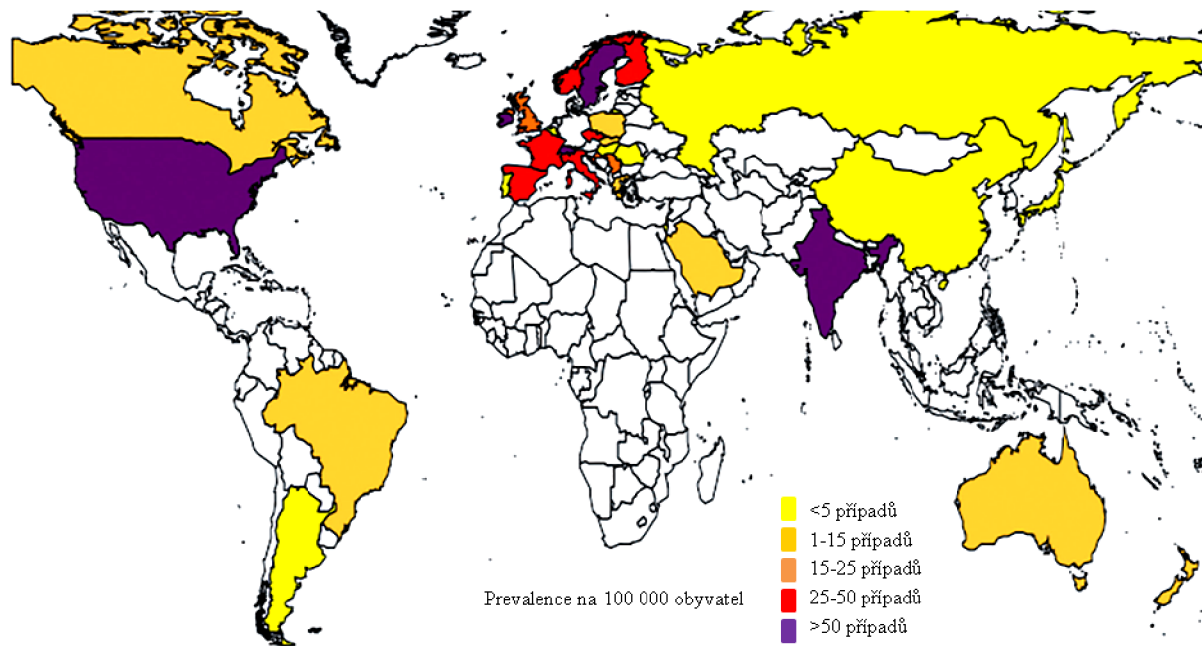
V Řecku patří sarkoidóza mezi nejrozšířenější onemocnění ze skupiny intersticiálních plicních onemocnění. Její incidence je zde 1.2 případů na 100 000 obyvatel, prevalence 5.7 případů na 100 000 obyvatel (Obr. 1) (Karakatsani *et al.*, 2009; Hilberg *et al.*, 2022).

V Asii je obecně pozorována nižší incidence sarkoidózy. Například v Japonsku je to 1 případ na 100 000 obyvatel (prevalence 3.7 případů na 100 000 obyvatel) (Arkema *et Cozier*, 2020).

V Jižní Koreji, incidence sarkoidózy v posledních letech vzrůstá. V roce 2009 byla incidence 0.85 případů na 100 000 obyvatel a v roce 2015 vzrostla incidence na hodnotu 0.97 případů na 100 000 obyvatel. Částečně je výše uvedené způsobeno zlepšením diagnostických postupů (Kim *et al.*, 2021; Yoon *et al.*, 2018). Mezi diagnostikovanými lehce převládají ženy. Ty ovšem mají o něco vyšší míru přežití než muži. Nejvíce zasaženým orgánem jsou i v této populaci plíce, následované postižením kůže, očí a mimoplicních lymfatických uzlin. Zasaženo bývá i srdce a nervový systém (Kim *et al.*, 2021; Jeon *et al.*, 2020). Zhruba jedna třetina pacientů trpí hyperkalcemií nebo hyperkalciurií (Kim *et al.*, 2021).

Vysoká mortalita byla u postižených sarkoidózou zaznamenána především u mladistvých do 19 let věku. U mladých dospělých je pak mortalita poměrně nízká a stoupá přímo úměrně s věkem. Nejrizikovějším faktorem pro mortalitu u tohoto onemocnění je zasažení srdce (Jeon *et al.*, 2020).

**Obrázek 1: Prevalence sarkoidózy v jednotlivých zemích podle hlavních epidemiologických studií z let 1958- 2017 (převzato z Brito- Zerón *et al.*, (2019))**



### 1.3 Etiologie sarkoidózy

Příčina vzniku sarkoidózy není stále zcela známá. Existuje několik příčin, které při koexistenci s rizikovými faktory mohou vést ke vzniku tohoto onemocnění (Seve *et al.*, 2021).

Neexistuje důkaz, že by sarkoidóza byla infekčním onemocněním. Jedná se spíše o nepřiměřenou reakci imunitního systému při rozpoznání molekulárních struktur typických pro povrch patogenních buněk (PAMP – *pathogen-associated molecular patterns*), mrtvých a částečně degradovaných mykobakterií a propionibakterií. Sarkoidózu mohou spouštět i další organické a anorganické látky (Valeyre *et al.*, 2014).

V souvislosti se vznikem sarkoidózy se uvažuje také o infekčních agens, jako jsou mykobakterie, protože vznik granulomů je hlavní imunitní odpovědí na tyto agens. Hlavním kandidátem na původce infekcí-indukované sarkoidózy je *Mycobacterium tuberculosis*. V souvislosti se sarkoidózou byly identifikovány další infekční agens, fungi nebo také borellie či *Propionibacterium acnes* (Jain *et al.*, 2020, Judson, 2020).

Doposud nebyl identifikován žádný autoantigen pro sarkoidózu, přestože v jejím průběhu dochází k prezentaci autoantigenů molekulami HLA II. třídy receptorům T-lymfocytů. Mezi kandidáty na tuto pozici patří vimentin. Jedná se o peptid přítomný v pojivové tkáni, který se účastní mezibuněčných interakcí a ovlivňuje funkci imunitního

systemu. Vimentin má vliv například na vznik revmatoidní artritidy a mohl by se podílet i na vzniku sarkoidózy (Starshinova *et al.*, 2020; Jain *et al.*, 2020).

Mnoho studií uvádí, že genetické faktory by mohly mít zásadní vliv na vznik a progresi sarkoidózy. Ve spojitosti se sarkoidózou bylo identifikováno několik lokusů genů například *BTNL2*, *ANXA11*, *IL23R*, *IL12B*, *TAP2*, *TNF $\alpha$*  nebo *RAB23* či lokusy pro HLA (Fischer *et al.*, 2015). Touto problematikou se zabývaly i celogenomové studie, které potvrdily spojitost *HLA* i non-*HLA* genů (geny zapojené v apoptóze, regulaci imunitní odpovědi a aktivaci T-lymfocytů) se vznikem sarkoidózy (Schurmann *et al.*, 2001; Judson, 2020).

Výskyt sarkoidózy v rodinné anamnéze zvyšuje pětikrát více riziko vzniku mezi sourozenci nemocného. U dvojčat je toto číslo ještě vyšší (Rybicki *et al.*, 2001; Jain *et al.*, 2020).

V neposlední řadě mají velký vliv na vznik a rozvoj sarkoidózy i environmentální faktory. Riziková je inhalace bioaerosolů (např. borovicový pyl), které jsou nejaktivnější především v jarním období. U anorganických látek je riziková expozice oxidu křemičitému a jiným prachům z kovů. Mezi rizikové faktory patří některá povolání, kdy dochází k expozici specifickým látkám. Zařazujeme mezi ně zemědělské profese (vystavení se pylům, prachu či pesticidům) nebo dělnické profese v továrnách (vystavení se kovovým prachům, minerálním vláknům) (Jain *et al.*, 2020, Judson, 2020; Arkema *et Cozier*, 2020).

Mezi další rizikové faktory, především pro závažnější průběh sarkoidózy, řadíme etnický původ.

U Afroameričanů je vysoká incidence a prevalence sarkoidózy, což poukazuje na vrozenou genetickou predispozici (Hena, 2020). Úmrtnost pacientů se sarkoidózou je v afroamerické populaci velmi vysoká a ve srovnání s bělošskými pacienty umírají v nižším věku (Mirsaeidi *et al.*, 2015). U afroamerické populace je také vyšší náchylnost k mimoplicnímu postižení. Sarkoidóza v této populaci nejčastěji postihuje kůži (s výjimkou výskytu nodózního erytému), kostní dřeň, játra, extrathorakální lymfatické uzliny a oči (Hena, 2020). Genetické predispozice se v afroamerické populaci projevují především v oblasti kodující *HLA* na chromozomu 6. Nejsilnější asociace u *HLA-DRB1* jsou populačně specifické. U Afroameričanů se projevuje především varianta *DRB1\*03:02* a *DQB1\*06:02* spojená jak s vyšším rizikem vzniku, tak progresí sarkoidózy (Hena, 2020; van Moorsel *et al.*, 2022). S predispozicí k sarkoidóze u afroamerické populace je také spojená oblast 5q11.2 (Hena, 2020).

Naopak u bělošské populace se častěji vyskytuje porucha metabolismu vápníku. Konkrétně v evropské populaci je také vyšší výskyt nodózního erytému (Hena, 2020;

Morimoto *et al.*, 2008). Alela HLA-DRB1\*03:01 se častěji vyskytuje v populacích evropského původu, naopak v populacích afrického a asijského původu se většinou nevyskytuje (van Moorsel *et al.*, 2022).

U studií s thajskými pacienty se sarkoidózou byla velká převaha postižených žen. Projevovalo se to i vysokým výskytem uveitidy v této populaci (téměř 40 %), přičemž mužské pohlaví je významným ochranným faktorem při vzniku uveitidy (Tripipitsiriwat *et al.*, 2020). Dvě nedávné studie z Číny zjistily prevalenci uveitidy pouze necelých 6 %. Vysoký podíl očního postižení u pacientů se sarkoidózou byl také v Japonské populaci (30.9 %) (Morimoto *et al.*, 2008).

U japonských pacientů se také s vysokou četností objevovaly srdeční sarkoidózní granulomy. Průběh srdeční sarkoidózy je variabilní a pohybuje se od benigních arytmií nebo srdeční blokády vysokého stupně až po náhlou smrt. Úmrtnost související se srdeční sarkoidózou byla zjištěna u 46.9 % (Morimoto *et al.*, 2008).

Mezi rizikové faktory patří i věk nad 40 let a koexistence dalšího onemocnění například *lupus perio* nebo chronická uveitida. Rizikovými faktory jsou i kouření či obezita (Arkema *et Cozier*, 2018).

## 1.4 Léčba

Lék na sarkoidózu jako takovou, který by toto onemocnění zcela vyléčil, neexistuje. V současné době používané léky ovlivňují granulomatózní proces a jeho klinické projevy. Převážná část pacientů se sarkoidózou nepotřebuje léčbu a je jen dlouhodobě sledována pro prevenci případného zhoršení stavu. Sarkoidóza často míří ke spontánní regresi bez toho, aby zapříčinila jakékoliv trvalé poškození zasažených orgánů (Jain *et al.*, 2020; Ungprasert *et al.*, 2019).

K nasazení terapie vedou tři hlavní důvody, a to riziko závažné dysfunkce nebo kritického orgánového postižení, riziko úmrtí a přítomnost symptomů závažně zhoršujících kvalitu života (Valeyre *et al.*, 2014).

Léky první volby při léčbě sarkoidózy jsou kortikosteroidy, především Prednisolon. Kortikosteroidy ovšem mají velké množství nežádoucích účinků. U většiny pacientů je navíc nutnost dlouhodobého užívání. Po jejich vysazení či snížení dávky dojde až u 30 % pacientů k relapsu onemocnění. Kortikosteroidy v inhalační formě se podávají především u symptomů podobným astmatu a na zmírnění kašle (Ungprasert *et al.*, 2019; Rahaghi *et al.*, 2020). Role kortikosteroidů v léčbě chronické plicní sarkoidózy zůstává stále kontroverzní.

Dalším používaným léčivem jsou cytostatika a imunomodulátory. Tyto se nasazují při špatné odpovědi na léčbu kortikosteroidy nebo při nutnosti dlouhodobého užívání kortikosteroidů, kdy dochází k riziku steroidní toxicity. Nejčastěji jsou v této skupině užívané hydrochloroquine, azathioprine, methotrexat či mycophenolate mofetil (Rahaghi *et al.*, 2020; Jain *et al.*, 2020; Ungprasert *et al.*, 2019).

Třetí možností pro ty, kdo nereagují nebo z nějakého důvodu nesnesou glukokortikoidy a cytostatika, jsou monoklonální protilátky proti cytokinům. Nejčastěji používané jsou inhibitory TNF- $\alpha$ , například Infliximab či Thalidomid (Ungprasert *et al.*, 2019; Jain *et al.*, 2020).

Při léčbě sarkoidózy se mohou používat i antimalarika či antimikrobiální léky zvláště u postižení kůže nebo gastrointestinálního traktu. Při vyčerpání všech léčebných možností přichází v úvahu i transplantace (Valeyre *et al.*, 2014).

## 1.5 Patofyziologie sarkoidózy

Charakteristickým znakem sarkoidózy je vývoj a akumulace granulomů, které jsou umístěny podél lymfatických cest (Iannuzzi *et al.*, 2007). Granulom se skládá z jádra tvořeného aktivovanými a vysoce diferencovanými mononukleárními fagocyty (epiteloidními buňkami a Langerhasovými buňkami). Charakteristická je i přítomnost inkluzí, například Schaumanovy inkluze (Drent *et al.*, 2021). V centrální části granulomu se nacházejí CD4<sup>+</sup> T-lymfocyty. V periferní části je granulom obklopen CD8<sup>+</sup> T-lymfocyty, regulačními T-lymfocyty, B-lymfocyty a fibroblasty (Polverino *et al.*, 2020; Drent *et al.*, 2021).

První fáze imunitní odpovědi na přítomnost neznámého antigenu je výsledkem působení genetických, epigenetických a environmentálních faktorů ve spolupráci s vlivy vnějšího prostředí. Jedinec se specifickým genetickým základem má tendence k dysfunkční imunitní odpovědi a ke tvorbě výše zmíněných granulomů (Drent *et al.*, 2021).

Neznámý antigen také aktivuje dendritické buňky a alveolární makrofágy (Polverino *et al.*, 2020). Při kontaktu s buňkou má neznámý antigen schopnost interagovat s buněčnými transmembránovými *Toll-like* receptory (TLR2, TLR4) a *NOD-like* receptory (NLR), což vede k aktivaci transkripčního faktoru *Nuclear Factor-Kappa B* (NF $\kappa$ B). Tyto transmembránové receptory mají klíčovou roli v nespecifické imunitní odpovědi. Rozpoznávají konzervované struktury molekul pocházejících z mikrobů a poškozených buněk (Tarasidis *et Arce*, 2020).

Při expozici neznámému antigenu dochází ke stimulaci transkripčních faktorů a následné produkci cytokinů, expresi chemokinových receptorů a chemokinů (Obr. 2).



Imunitní odpověď poté buď eliminuje spouštěcí antigen a dojde k ukončení imunitní odpovědi, nebo antigen přetrvává a následuje formování granulomu (Drent *et al.*, 2021).

Tento proces také přiláká buňky specifické imunitní odpovědi  $CD4^+$  a  $CD8^+$  T-lymfocyty do místa zasažení (Drent *et al.*, 2021). Nejtypičtějším imunologickým nálezem v tekutině z bronchoalveolární laváže (BAL) pacientů se sarkoidózou je převaha  $CD4^+$  T-lymfocytů, proto je u nich zvýšený poměr  $CD4^+/CD8^+$  T-lymfocytů (Shen *et al.*, 2016).

$CD4^+$  T-lymfocyty se účastní tvorby granulomů, naboru monocytů/granulocytů a komunikace s B-lymfocyty, což vede k tvorbě protilátek.

Hlavní rysem akutního onemocnění je zánětlivý proces řízený Th1/Th17/regulačními T-lymfocyty (Tregs), na němž se podílejí makrofágy jako antigen prezentující buňky i jako klíčové efektory (Polverino *et al.*, 2020).

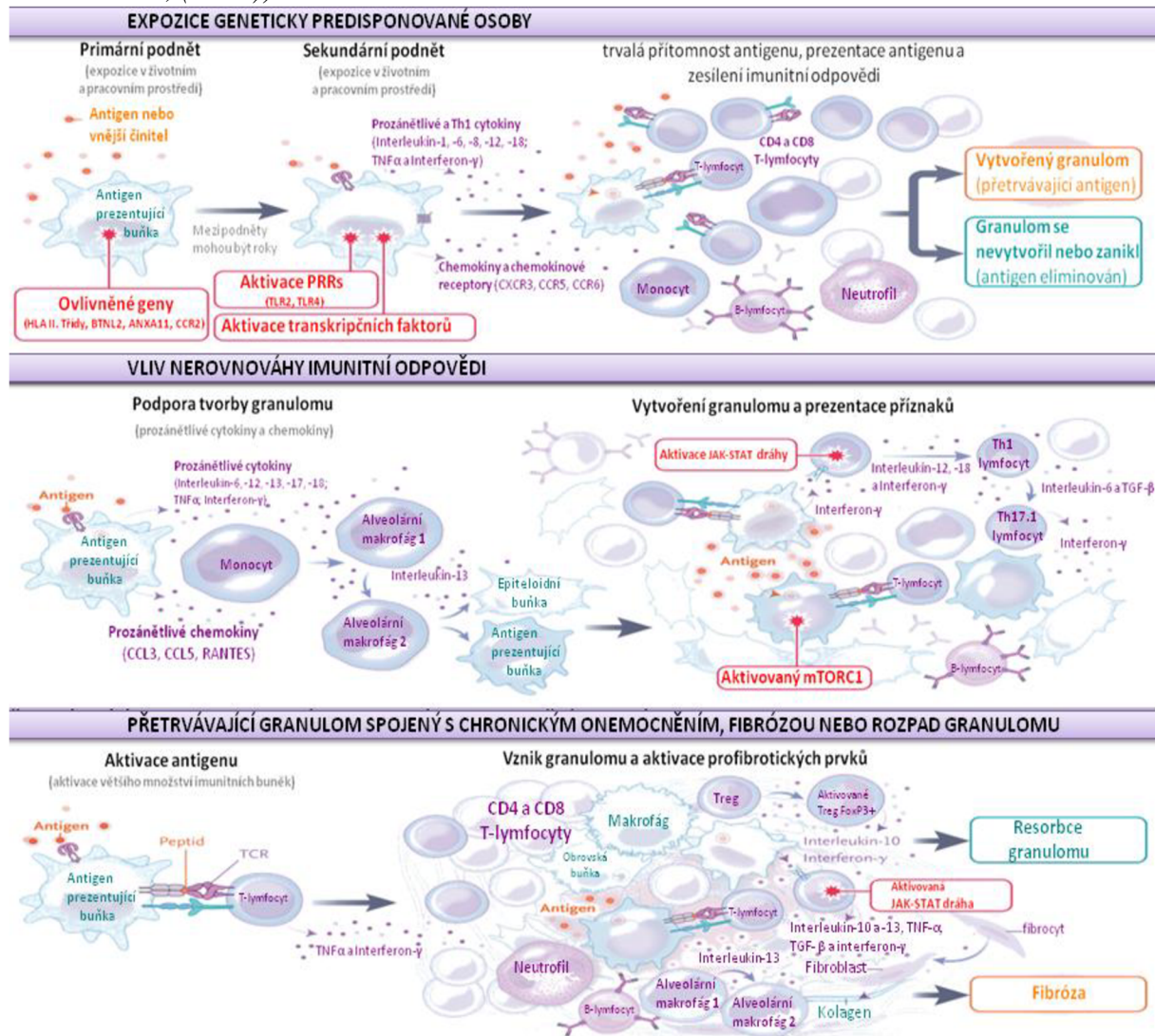
Ukázalo se, že hladina  $IFN-\gamma$  u sarkoidózy je značně zvýšená, což ukazuje na primární Th1 odpověď (Katchar *et al.*, 2003). Infikované makrofágy interagují s T-lymfocyty pomocí MHC II a zároveň i produkují IL-12. Obě tyto skutečnosti stimulují indukci diferenciaci  $CD4^+$  T-lymfocytů směrem k Th1 lymfocytům. Th1 lymfocyty amplifikují lokální imunitní odpověď (Iannuzzi *et al.*, 2007). Th1 lymfocyty stimulují přeměnu makrofágů na aktivované makrofágy a produkují  $TNF-\alpha$ , IL-2 a  $IFN-\gamma$ , které tvoří obranyschopnost proti mikroorganismům (Tarasidis *et Arce*, 2020). Předpokládá se, že zmíněné cytokinové prostředí stimuluje aktivaci klasických prozánětlivých makrofágů (M1) (Wojtan *et al.*, 2016).

Alveolární makrofágy jsou spolu s plicními T-lymfocyty hlavním zdrojem interferonu gama ( $IFN-\gamma$ ). Předpokládá se, že role  $IFN-\gamma$  spočívá v inhibici apoptózy makrofágů prostřednictvím exprese inhibitoru p21/Waf1 cyklin-dependentní kinázy, jejíž zvýšená aktivita byla pozorována u sarkoidních granulomů (Zhou *et Arce*, 2020).  $IFN-\gamma$  také aktivuje Janus kinázu a tím JAK-STAT dráhu (Calender *et al.*, 2020; Drent *et al.*, 2021). Diferenciace na Th17 lymfocyty je indukována  $TGF-\beta$  a IL-6, který je produkován především mononukleárními fagocyty. Buňky Th17 exprimují prozánětlivé IL-17, IL-21, IL-22, které mají klíčovou roli v plasticitě tvorby a udržování granulomu (Drent *et al.*, 2021). Předpokládá se, že k přetrvávajícím granulomatózním reakcím u sarkoidózy přispívá neschopnost imunitních regulačních mechanismů omezit trvání zánětu Th1-Th17.

U chronického onemocnění vlivem IL-12 a IL-18 (produkovaného mononukleárními fagocyty a indukujícího produkci  $IFN-\gamma$ ) dochází k nižší aktivaci alveolárních makrofágů. Aktivace Th1 je neúplná kvůli nedostatečným hladinám IL-12 a IL-18 (Drent *et al.*, 2021).

Během formování granulomu může vlivem zvýšené produkce makrofágových fibrogenních cytokinů (např. TGF- $\beta$ ) docházet k posunu diferenciaci k Th2 lymfocytům, které produkují protizánětlivé cytokiny IL-4, -10, -13. Toto má klíčovou úlohu při vzniku fibrózy. IL-13 podporuje vznik makrofágů (M2). Ty diferencují v epiteloidní buňky a antigen prezentující buňky (Drent *et al.*, 2021). Je také aktivovaná dráha mTORC1 (*mammalian target of rapamycin complex 1*) v makrofázích, podporující nadměrnou tvorbu granulomů (Obr. 2) (Calender *et al.*, 2020; Polverino *et al.*, 2020). Alveolární makrofágy aktivované v souvislosti s cytokiny Th2 produkují vysoké hladiny fibronektinu a chemokinu CCL18. Ten zvyšuje produkci kolagenu plicními fibroblasty, což následně zvyšuje uvolňování CCL18 makrofágy. Dochází ke vzniku pozitivní zpětnovazebné smyčky vedoucí k plicní fibróze (Iannuzzi *et al.*, 2007). Přetrvávání antigenu nebo jeho nedegradovatelných zbytků udržuje zánět s Th2 a M2 buňkami, což vede k fibrotické remodelaci dolních cest dýchacích nebo jiné další zapojené tkáni (Obr. 2)(Müller-Quernheim *et al.*, 2012).

**Obrázek 2: Imunologická kaskáda způsobující tvorbu sarkoidního granulomu (převzato z Drent *et al.*, (2021))**



## 1.6 Průběh nemoci

Mezi obecné příznaky sarkoidózy řadíme horečku, vyčerpání, únavu, ztrátu hmotnosti a noční pocení. Sarkoidóza postihuje především plíce, které jsou zasaženy u více jak 90 % pacientů (Jain *et al.*, 2020). Mezi nejčastější projevy sarkoidózy patří suchý kašel, dušnost a bolest na hrudi (Starshinova *et al.*, 2020). Plicní sarkoidózu můžeme rozdělit do 5 stadií posuzovaných na základě rentgenu (rtg) hrudníku (Seve *et al.*, 2021). Stadium bez jakéhokoliv viditelného plicního postižení se označuje jako stadium 0, toto stadium se vyskytuje zhruba u 5-15 % pacientů. U sarkoidózy ve stadiu 1 je většinou přítomná symetrická hilová lymfadenopatie a v tomto stadiu je zhruba 45-65 % pacientů se sarkoidózou. Stadium 2 zahrnuje postižení uzlin, ale i plicního parenchymu a toto stadium

je evidováno u přibližně 30-40 % pacientů. Ve stadiu 3 je postižen pouze plicní parenchym bez známek adenopatie či fibrózy, toto je evidováno u 10-15 % pacientů. Ve stadiu 4 dochází k fibrotickým změnám, které jsou přítomny u zhruba 5 % pacientů se sarkoidózou (Silva *et al.*, 2015).

Častým příznakem jsou změny plicních funkcí jako snížení hodnot statických plicních objemů, například snížení FVC (usilovná vitální kapacita) a také snížení poměru usilovně vydechnutého objemu za jednu sekundu (FEV1) / FVC vyskytující se především z důvodu přítomnosti granulomu či fibrotických změn v peribronchiu. Pro posouzení závažnosti fibrotických změn plic byl vytvořen kompozitní fyziologický index, který tyto parametry využívá.

U sarkoidózy byly popsány dva syndromy charakteristické akutním průběhem. Jedná se o Heerfordtův syndrom (charakteristický uveitidou, parotitidou, horečkou a paralýzou obličeje) a Löfgrenův syndrom (LS).

Löfgrenův syndrom byl poprvé popsán v roce 1952. Ve většině případů se prezentuje horečkou, oboustrannou hilovou lymfadenopatií, nodózním erytémem, oboustrannou artritidou kotníků a periartikulárním zánětem. Bolestivost kloubů může být tak značná, že ovlivňuje i chůzi (Rubio-Rivas *et al.*, 2020; Kolek *et al.*, 2017). LS se nejčastěji vyskytuje v evropské populaci, především ve Švédsku. Tato forma sarkoidózy je většinou spojena s dobrou prognózou onemocnění (Seve *et al.*, 2021).

## **1.7 Mimoplicní postižení**

Mimoplicní postižení se objevuje až ve čtvrtině případů (Starshinova *et al.*, 2020). U 20 % pacientů se vyskytuje většinou nebolestivé postižení periferních lymfatických uzlin, projevující se jejich otokem (Jain *et al.*, 2020).

U sarkoidózy dochází často k postižení kůže a to ve 20-30 %. Nejčastějším projevem je nodózní erytém. Časté je také silné pocení, malouzlová nebo velkouzlová kožní sarkoidóza, subkutanní sarkoidóza či sarkoid v jizvě či v tetováních projevující se zčervenáním a zduřením starých jizev. Charakteristická je také přítomnost *Lupus pernio* (Ungrasert *et al.*, 2019; Kolek *et al.*, 2017).

U zhruba u 30 % pacientů sarkoidóza postihuje oči. Zasažení očí je častější u Afroameričanů a japonské populace. Nejčastější formou postižení očí je uveitida (Seve *et al.*, 2021; Morimoto *et al.*, 2008). Může se projevovat jako konjunktivitida, katarakta, glaukoatrofie optického nervu nebo může docházet k adhezím mezi duhovkou a čočku (Kolek *et al.*, 2017).

V jedné čtvrtině případů se objevuje postižení srdce (Jain *et al.*, 2020). Hlavním patologickým procesem je granulomatózní zánět myokardu, který vede k arytmií a kardiopatii (Ungprasert *et al.*, 2019). Arytmie mohou být zachyceny 24hodinovým EKG. Granulomatózní infiltraci srdce lze potvrdit pozitronovou emisní scintigrafií. Při podezření na zasažení myokardu se provádí magnetická rezonance. Potvrzení diagnózy probíhá endomyokardiální biopsií s průkazem granulomů (Kolek *et al.*, 2017).

Neurologické postižení nebo neurosarkoidóza se vyskytují u méně než 10 % pacientů. Postiženy jsou hlavně hlavové nervy. Vyskytují se i periferní neuropatie. Může docházet i k aseptické meningitidě nebo encefalopatii (Kolek *et al.*, 2017; Jain *et al.*, 2020). Postižení mozku a míchy je většinou vážné a prognosticky nejisté.

Sarkoidóza může napadat i kosti, což se projevuje především artritidou, artralgií, osteoporózou, osteopenií a uzlovitými a cystickými lézemi.

Dále sarkoidóza postihuje játra, slinivku, ledviny nebo endokrinní žlázy (Jain *et al.*, 2020).

## 1.8 Diagnostika

Diagnostika sarkoidózy se odvíjí od stěžejních bodů, kterými jsou odpovídající klinická prezentace, přítomnost nenekrotizujícího granulomatozního zánětu a vyloučení jiné alternativní příčiny onemocnění. Některé klinické prezentace jsou pro sarkoidózu vysoce specifické. Jedná se o Löfgrenův syndrom, *lupus pernio* a Heerfordtův syndrom. Silnou asociaci se sarkoidózou vykazuje i oboustranná hilová adenopatie (Crouser *et al.*, 2020).

V minulosti bylo nejvíce případů, a to i bezpříznakových, odhaleno díky rentgenu plic, který byl u pacienta indikován z jiných příčin (Seve *et al.*, 2021). Na základě těchto zjištění dělíme sarkoidózu do pěti výše zmíněných stadií.

Charakteristickými znaky výpočetní tomografie (CT) u pacientů se sarkoidózou jsou klinická symetrická hilová lymfadenopatie, kalcifikace lymfatických uzlin, konsolidace a opacity mléčného skla, noduly (mikroskopické, makroskopické), voštinovitá přestavba plic, *air-trapping* a fibrotické změny (Silva *et al.*, 2015).

Další podstatnou složkou diagnostiky je vyšetření bioptických či BAL vzorků pro identifikaci přítomnosti granulomatozního zánětu (Drent *et al.*, 2021). Pro sarkoidózu je zde charakteristická přítomnost granulomu se soustředně uspořádanými imunitními buňkami. Sarkoidní granulom je většinou nenekrotizující (Crouser *et al.*, 2020). Diagnózu sarkoidózy podporuje i zvýšený poměr CD4<sup>+</sup>/CD8<sup>+</sup> T-lymfocytů v BAL, který může být užitečný i pro následnou diferenciální diagnostiku. Dalším navazujícím krokem je totiž

vyloučení jiných možných diagnóz, zejména těch zapříčiňujících granulomatózní zánět (Drent *et al.*, 2021; Crouser *et al.*, 2020).

V diagnostice se využívají i radionuklidové metody, například Ga-scintigrafie, či pozitronová emisní tomografie (PET). PET napomáhá identifikaci míst vhodných pro biopsii, rozpoznání srdeční sarkoidózy či vyšetření aktivity u závažných stavů sarkoidózy, kdy není přítomen jiný biologický důkaz aktivity (Jeny *et al.*, 2020).

Magnetická rezonance hrudníku se ukázala užitečná především při hodnocení lymfadenopatie (Chung *et al.*, 2014).

V souvislosti se sarkoidózou bylo navrženo mnoho biomarkerů. Jen malá část z nich je ovšem klinicky spolehlivá a rutinně se využívá. Přehled hlavních biomarkerů jak využívaných, tak těch, které jsou ve fázi výzkumu, je uveden v samostatné kapitole níže.

Po určení diagnózy mohou následovat další vyšetření pro zjištění dalších zasažených orgánů a nespecifických symptomů nemoci, které ovlivňují kvalitu života pacienta (Drent *et al.*, 2021).

## 1.9 Biomarkery

Laboratorní vyšetření u sarkoidózy jsou užitečná pro stanovení diagnózy a charakterizaci závažnosti onemocnění. Základní biochemické hodnoty a krevní obraz vykazují jen nespecifické změny. Časté jsou u sarkoidózy hypergamaglobulinemie a zvýšení imunoglobulinů IgG, IgA a IgM. Zhruba u poloviny nemocných bývají v akutní fázi onemocnění zvýšeny cirkulující imunokomplexy.

Před zavedením fibroptické bronchoskopie a BAL se jako diagnostický test sarkoidózy používal Kveim-Siltzbachův test, který má dnes již jen historický význam (Ahmadzai *et al.*, 2014).

Mezi dnes používané biomarkery sarkoidózy patří Angiotensin-konvertující enzym (ACE), Neopterin, Lysozym a některé další níže uvedené.

ACE, který přeměňuje angiotenzin I na angiotenzin II, je zvýšený v séru a v BAL u pacientů se sarkoidózou. Jeho koncentrace korelují s radiologickými stadii (Jain *et al.*, 2020). ACE může být zvýšen u jiných granulomatózních poruch (Bergantini *et al.*, 2019).

Neopterin jako biomarker aktivity imunitního systému a oxidačního stresu, je uvolňován v odpovědi na IFN- $\gamma$ . Je ukazatelem aktivity Th1 lymfocytů. Jeho vyšší hodnoty byly detekovány v séru a moči u pacientů se sarkoidózou (Ahmadzai *et al.*, 2014).

Lysozym je produkován systémem monocytů, makrofágů a epiteloidních buněk zapojených do tvorby granulomu. Pacienti se sarkoidózou vykazují zvýšený výskyt koncentrace lysozymu

v aktivní fázi. Lysozym je považován spíše za prognostický indikátor než jako diagnostický nástroj (Bergantini *et al.*, 2019).

Nově byl jako marker navržen sérový amyloid A (SAA). Jeho úloha je ve tvorbě granulomu a prezentaci antigenu. Bylo zjištěno, že distribuce SAA v sarkoidních granulomech se liší od jiných granulomatózních onemocnění (Tarasidis *et Arce*, 2020).

U více než 40 % pacientů se sarkoidózou dojde k hyperkalciurii, a hyperkalcémie se vyskytuje u 5–10 % pacientů. Hyperkalcémie u sarkoidózy je způsobena zvýšením sérového 1.25-dihydroxy vitamínu D3 (kalcitriol), který zvyšuje koncentrace vápníku v séru. Monitorování koncentrace vápníku v séru a 24hodinové měření vápníku v moči jsou důležité u všech pacientů se sarkoidózou (Gwadera *et al.*, 2019).

Nové studie se zaměřují na biomarkery spojené s B-lymfocyty, jako jsou např. *B-Cell Activating Factor* (BAFF) a CD19. Vysoké hladiny BAFF byly nalezeny v séru pacientů se sarkoidózou a korelovaly se zvýšenými hladinami ACE, lysozymu a IFN- $\gamma$  v séru těchto pacientů (Tarasidis *et Arce* 2020).

U sarkoidózy byla hodnocena i široká škála potenciálních biomarkerů v kondenzátu vydechaného vzduchu. Z kondenzátu vydechaného vzduchu byla určována hladina zánětlivých cytosinů jako TNF- $\alpha$  či TGF- $\beta$ , jejichž hladina byla u pacientů se sarkoidózou zvýšená. Dřívější výzkumy týkající se těkavých organických látek ukázaly, že mají slibný potenciál a mohou poskytnout jedinečný dechový otisk. Ten by mohl nejen pomoci při diagnostice, ale také k identifikaci osob ohrožených chronickým onemocněním nebo recidivou a také při monitorování odpovědi na léčbu (Terrington *et al.*, 2019).

Dalším významným biomarkerem sarkoidózy je poměr CD4<sup>+</sup>/CD8<sup>+</sup> lymfocytů v bronchalveolární laváži. Analýza BAL tekutiny je užitečným vyšetřením pro diagnózu plicní sarkoidózy, protože při absenci jiných příčin odhalí lymfocytózu se zvýšeným poměrem CD4<sup>+</sup>/CD8<sup>+</sup> lymfocytů, obvykle > 3.5 (Jain *et al.*, 2020; Ahmadzai *et al.*, 2014).

Dalšími potenciálními biomarkery jsou chemokiny. Zvýšená produkce CCL18 vede k plicní fibróze. U pacientů se sarkoidózou byla také pozorována vysoká hladina CXCL9 a CXCL10 v séru. Svůj vliv na patogenezi sarkoidózy by mohla mít i variabilita počtů kopií segmentů DNA (CNV *copy number variation*). U pacientů se sarkoidózou byl detekován výskyt těchto změn v genu *BNTL2* a komponentě komplementu – C4 (Kishore *et Petrek*, 2013).

Genetické predispozice k sarkoidóze jsou spojené s geny MHC komplexu. Bylo identifikováno mnoho variant, ať už protektivních (HLA-DPB1\*04:02, HLA-DQB1\*02:01, HLA-DQB1\*05:01) pro sarkoidózu, či rizikových pro progresi onemocnění

(HLA-DQB1\*06:02, HLA-DRB1\*11:01, HLA-DRB1\*15:01), případně spojených s určitým klinickým fenotypem, například spojitost HLA-DRB1\*03:01 s výskytem LS (Kishore *et Petrek*, 2013).

Genetické studie prokázaly, že jedinci s určitými alelami HLA mají vyšší riziko specifických autoimunitních onemocnění. Zejména je tato asociace častá u evropské populace. Výrazná asociace se zde našla s geny *HLA-B8*, *DR3*, které jsou součástí ancestrálního haplotypu (AH) 8.1 HLA-A1, Cw7, B8, TNFAB\*a2b3, TNFN\*S, C2\*C, Bf\*s, C4A\*Q0, C4B\*1, DRB1\*0301, DRB3\*0101, DQA1\*0501, DQB1\*0201. Tento haplotyp, který je častější v severní Evropě, je také spojován u zdravých osob s řadou dysfunkcí imunitního systému. Bylo navrženo, že 8.1 AH modifikuje imunitní reaktivitu, a tím ovlivňuje více imunopatologických onemocnění. 8.1 AH ovlivňuje humorální a buněčnou imunitní odpověď. U jeho nositelů je převaha cytokinů typu II. Alely DR a DQ zapojené do 8.1 AH zřejmě podmiňují specifitu produkovaných autoprotilátek. Některé alely, které jsou součástí tohoto haplotypu, byly nalezeny asociovány se sarkoidózou (Price *et al.*, 1999; Candore *et al.*, 2002; Grunewald, 2010).

### 1.9.1 Celogenomové asociační studie

Lepším pochopením etiologie komplexních onemocnění, jako je právě sarkoidóza, a nalezení nových potenciálních biomarkerů se zabývá genetická epidemiologie. Tato studuje genetické faktory při určování zdraví a nemoci v populacích a jejich interakci s faktory prostředí. Zkoumá, jak kombinace těchto faktorů určuje vznik a vývoj nemoci. Přítomnost konkrétního genotypu a vnější prostředí jsou rizikovým faktorem významně zvyšujícím pravděpodobnost vzniku onemocnění u osob, na něž působí. S rozvojem pokročilých molekulárních technik jsou prováděny celogenomové asociační studie.

Celogenomové asociační studie (*genome-wide association study* – GWAS) zkoumají celý genom. Testují tisíce genetických variant a hledají ty, které jsou spojené s konkrétním znakem či nemocí. GWAS zkoumají i CNV nebo sekvenční varianty, ale nejčastěji studovanou variantou v rámci těchto studií jsou jednonukleotidové polymorfismy.

Výsledky těchto studií mohou být použity pro předpoklad rizika vzniku určitých onemocnění. Skóre polygenních rizik (*Polygenic risk score* PRS) vypovídá o náchylnosti jedince k určitému znaku nebo onemocnění. Získává se srovnáním genomu jedince s daty získanými z celogenomových asociačních studií o velikosti účinku této kombinace.



V budoucnu by toto mohlo sloužit jako biomarker v klinické praxi. Nejvíce se na něm v současnosti pracuje v souvislosti s rakovinou prsu, diabetem druhého typu či ischemickou chorobou srdeční (Petřková *et al.*, 2018; Uffelmann *et al.*, 2021).

Díky GWAS byly například odhaleny jednonukleotidové polymorfismy v několika genech spojených s výskytem sarkoidózy, konkrétně *ANXA11*, *BTNL2*, *NOTCH4*, *NOD2*, *TNF $\alpha$*  (Hofmann *et al.*, 2011).

Tyto studie pomáhají také porozumění genetické struktuře znaků. Zahrnují odhad počtu kauzálních variant odpovídajících velikosti jejich účinků a jejich četnosti. Umožňují odhadnout heritabilitu, neboli podíl variability, která se projevuje v průběhu vývoje znaku. To lze vysvětlit genetickou variabilitou daného znaku v populaci.

GWAS studie by mohly napomoci redukovat čas a náklady potřebné na vývoj léků, také časně odhalit jakou reakci bude mít daný jedinec na určitý druh terapie (Uffelmann *et al.*, 2021).

Katalog GWAS byl založen NHGRI (*National Human Genome Research Institute*) v roce 2008 v reakci na rychlý nárůst počtu publikovaných asociačních studií. Tyto studie poskytují bezprecedentní příležitost zkoumat vliv běžných variant na komplexní onemocnění (<https://www.ebi.ac.uk/gwas/>).

Mezi první celogenomové studie zabývající se i sarkoidózou patří Franke *et al.*, (2008), který zkoumal sarkoidózu i Crohnovu nemoc. Obojí jsou chronická zánětlivá onemocnění neznámého původu, sdílející mnoho společných znaků jako jsou tvorba granulomů, aktivace T-lymfocytů a produkce prozánětlivých cytokinů. Tato studie odhalila polymorfismus v lokusu *C10orf67* asociovaný s predispozicí k oběma onemocněním. Z téhož roku je i celogenomová studie Hofmann *et al.*, (2008), která studovala přes 300 000 jednonukleotidových polymorfismů a jejich spojitost se sarkoidózou. Nejsilnější asociace se sarkoidózou byla na základě jeho výsledků objevená v oblasti genu *ANXA11* pro anexin A11, který má komplexní a nezbytnou funkci v několika biologických drahách včetně proliferace a apoptózy.

Další celogenomové asociační studie u sarkoidózy se zabývaly mapováním určitých úseků chromozomů a hledání asociovaných variant v těchto úsecích.

Mezi tyto práce patří Hofmann *et al.*, (2011), který pomocí přesného mapování pokryl oblast chromozomu 6p12.1, kde byly nalezeny kandidátní polymorfismy především v genu *RAB23*.

Fischer *et al.*, (2012) mapoval oblast chromozomu 11q13.1 a Hofmann *et al.*, (2013), oblast chromozomu 12q13.3-q14.1. Tento region byl již dříve zmiňovaný jako asociovaný

s revmatoidní artritidou (Barton *et al.*, 2008). Studie tohoto regionu u sarkoidózy odhalila polymorfismy v genu *OS9*, a možnou úlohu tohoto genu, jako rizikového faktoru pro sarkoidózu, jelikož by mohl přispívat k autoimunitní reakci a mít tak vliv na imunopatogenezi sarkoidózy (Hofmann *et al.*, 2013).

Obecně se sarkoidózou zabýval ještě Fischer *et al.*, (2015). Specifikoval genetické varianty potenciálně funkčně relevantní pro patogenezi sarkoidózy. Varianta rs3184504 na chromozomu 12q24.12 vyvolává změnu aminokyseliny (Arg262Trp) v proteinu SH2B3. Tento protein se podílí na proliferaci B-lymfocytů. SNP rs3184504\*A je spojen se silnější aktivací rozpoznávání *NOD2* v odpovědi na lipopolysacharid a muramyldipeptid, což naznačuje, že SH2B3 může hrát roli v ochraně před bakteriálními infekcemi. Další SNP rs12069782 v regionu 1p31.3 se nachází v promotorové oblasti genu *IL23R* a mohl by ovlivnit jeho expresi. Byly nalezeny také SNP na chromozomu 2q33.2, které by mohly ovlivňovat expresi mnoha genů, mezi nimi *BMP2*, který se podílí na regulaci zánětlivých procesů v plicích.

Alelická variabilita SNP rs34536443 (*TYK2* Pro1104Ala) je známá tím, že ovlivňuje polarizaci Th2 lymfocytů. Tento posun je také pozorován u těžké sarkoidózy s progredující plicní fibrózou.

Často je zmiňována spojitost variability v oblasti HLA se sarkoidózou. Fischer *et al.*, (2015) popsal asociaci v genu *BTNL2*. Nejsilněji asociované varianty se nacházejí v promotorové oblasti genu *BTNL2*. Mohou tedy ovlivňovat expresi *BTNL2* tím, že mění vazby transkripčních faktorů. Druhá asociace sarkoidózy s HLA oblastí byla identifikována v blízkosti *HLA-B* – *HLA-B\*08:01*, přičemž hlavní SNP rs4143332 byl významně spojen s akutní sarkoidózou a Löfgrenovým syndromem.

Nejen objevováním, ale i případnou funkcí polymorfismů se zabývá také recentní studie Meguro *et al.*, (2020). Provedli první celogenomovou asociační studii u pacientů se sarkoidózou ve východoasijské – japonské populaci. Identifikovali tři lokusy, které vykazovaly signifikantní asociaci se sarkoidózou. Výskyt polymorfismu v genu *CCL24* může mít za následek přiklonění se k imunitní odpovědi Th1 typu. Dalším lokusem byl *STYXLI-SRRM3*. Polymorfismus v tomto genu ovlivňuje endogenní hladinu steroidů a může být jak rizikovým faktorem pro vznik onemocnění, tak může napomáhat k spontánní remisi onemocnění. Dalším asociovaným byl polymorfismus v lokusu *C1orf141-IL23R*, který byl identifikován i u dalších komplexních onemocnění jako je Crohnova choroba či ulcerózní kolitida.

Další studie se zabývala jednotlivými fenotypy u sarkoidózy, především LS (Rivera *et al.*, 2016) či sarkoidózou očí u Afroameričanů (Garman *et al.*, 2021), kde byla zjištěna asociace především v HLA lokusech.

Další studie se zabývá vlivem kouření na sarkoidózu. U pacientů s určitou genetickou variantou může kouření modulovat riziko pro vznik sarkoidózy, a to jak pro pacienty s LS, tak bez LS (Rivera *et al.*, 2019).

*Liu et al.*, (2013) studoval souvislost s primární sklerozující cholangitidou i s dalšími imunitními onemocněními včetně sarkoidózy.

Recentnější GWAS studie zmiňující sarkoidózu jsou pak zaměřené na sekvenace velkých souborů mnoha jedinců s různými komplexními onemocněními, mezi které se řadí i sarkoidóza. Práce se ale dále nezabývají možnou funkcí či vlivem těchto variant na onemocnění (Sakaue *et al.*, 2021; Backman *et al.*, 2021; Jiang *et al.*, 2021).

## 1.10 HLA typizace

Geny pro HLA jsou důležitou komponentou imunitního systému. Nacházejí se v nejvíce variabilní oblasti lidského genomu, MHC, lokalizovaném na krátkém raménku chromozomu 6 (Bravo-Egana *et al.*, 2021). HLA molekuly interagují s T-lymfocyty v thymu a sekundárních lymfatických orgánech a regulují imunitní odpověď.

HLA typizace je využívána zejména při transplantaci orgánů, kdy je přesná typizace dárce a příjemce nezbytná pro určení stupně neshody HLA mezi dárce a příjemcem. Další využití, nejzásadnější medicínské využití HLA typizace, je při transplantaci hematopoetických kmenových buněk, ale také k evaluaci při transfúzi krevních destiček. Využívá se i ve farmakogenomice, pro zjištění asociace s genetickými onemocněními, především těmi autoimunitními či nádorovými (Madden *et al.*, Chabot-Richards, 2019).

Metody HLA typizace byly původně založeny na detekci exprimovaných HLA molekul na povrchu separovaných T-lymfocytů (HLA I. třídy) a B-lymfocytů (HLA II. třídy). Používal se panel sér se známými anti-HLA protilátkami. Data získaná touto metodou měla nízké rozlišení a u některých pacientů bylo obtížné získat dostatečný počet lymfocytů (Howell *et al.*, 2010).

S nástupem PCR (*polymerase chain reaction*), se začaly pro typizaci využívat metody založené právě na PCR (Obr. 3). Jedná se především o SSP (analýza se sekvenčně specifickými primery). Využívá primerů, které se specificky vážou na amplifikovanou komplementární cílovou sekvenci. Pomocí těchto primerů je ovšem možné amplifikovat jen omezené množství alel. Tato metoda byla poprvé použita pro amplifikaci genů pro HLA-DR.

Často se používá pro vyřešení nejednoznačných výsledků získaných jinými metodami. Poskytuje výsledky s vysokou přesností (Bravo-Egana *et al.*, 2021).

Další PCR metodou je Real time PCR, kdy se pro typizaci používá TaqMan assay nebo analýza křivky tání.

Metoda SSOP (sekvenčně specifické oligonuklotidové sondy) je založena na hybridizaci (Obr. 3). Tato metoda využívá oligonukleotidové sondy, které hybridizují s komplementární cílovou genomickou sekvencí. Sondy jsou komplementární k limitovanému množství známých sekvencí. Sonda hybridizovaná s produktem je detekovaná pomocí chemiluminiscenčních či kolorimetrických metod a takto je určen typ HLA. Nejčastější adaptace této metody je v současné době využívána v průtokové cytometrii či například v technologii Luminex (Madden *et Chabot-Richards*, 2019; Bravo-Egana *et al.*, 2021).

Zlatým standardem pro typizaci HLA je v současné době sekvenování. Jeho začátek je u Sangerova sekvenování, kdy byly amplifikovány specifické lokusy genů pro HLA, obvykle v exonu 2, 3 a 4 pro HLA I. třídy a exonu 2 pro HLA II. třídy. Tato metoda poskytuje vysoké rozlišení na dvě pole, je ovšem obvykle prováděna po jednom vzorku (Madden *et Chabot-Richards*, 2019).

Metody založené na Sangerově sekvenování jsou hlavní technikou, která se používá k typizaci HLA u pacientů při testování histokompatibility. V zásadě všechny typizace založené na sekvenování využívají primer po amplifikaci DNA a fluorescenčně značené dideoxynukleotidy, které fungují jako molekuly ukončující řetězec a tím ukončují PCR amplifikaci během sekvenování. Různé sekvenovací technologie pak mohou být použity k identifikaci nukleotidu v místě ukončení PCR. Tato metoda typizace poskytuje výhodu vyššího rozlišení výsledků než při SSP nebo SSOP, protože sekvenování dokáže určit každý pár bází v HLA genu, který je typizován. Bohužel ve srovnání s metodou založenou na PCR je náročnější na práci i čas (Cornaby *et al.*, 2021).

Sekvenování nové generace (NGS- *Next generation sequencing*) je metoda umožňující paralelní sekvenování. Geny pro HLA jsou amplifikovány a fragmentovány na segmenty, které jsou následně sekvenovány. Z každé alely je generováno mnoho sekvencí, umožňujících kompletní charakteristiku obou alel a přesnou typizaci až na čtyři pole (Madden *et Chabot-Richards*, 2019).

Současné přístupy NGS k typizaci HLA obvykle využívají obohacení izolované DNA pacienta o HLA lokusy, aby bylo zajištěno vysoké pokrytí. Bylo prokázáno, že typizace pomocí NGS je nezbytná pro rozlišení HLA na dvě pole, je zásadní pro identifikaci protilátek specifických pro dárce, a to jak před transplantací, tak i po ní (Cornaby *et al.*, 2021).

První dostupnou NGS platformou byla Roche 454, která využívá emulzní PCR a pyrosekvenování, kdy dochází k detekci pyrofosfátu uvolněného při inkorporaci nukleotidů. Její limitací je právě pyrosekvenační technologie, která problematicky určuje možné inserce a delece.

Další NGS platformou je IonTorrent systém využívající emulzní PCR. Po inkorporaci báze dojde k uvolnění protonu a je zaznamenán signál odpovídající počtu inkorporovaných bází (Bravo-Egana *et al.*, 2021).

Také Illumina, využívající můstkovou PCR, se řadí mezi NGS platformy. Při použití této metody jsou oligonucleotidy fixovány na povrch flow cell a vytvořené klastry jsou sekvenovány, při začleňování fluorescenčně značených nukleotidů. Platformy Illumina se využívají pro analýzu DNA, RNA a proteinů jak v oblasti výzkumu nemocí, vývoji léků a molekulárních testů, tak mají rovněž i klinické použití. Jedna z jejich platform je MiSeq. Ten má například v porovnání s IonTorrent vyšší celkový výstup bází. Každá báze má vyšší skóre kvality i vyšší přesnost při sekvenování přes problematické homopolymerní oblasti. Díky možnosti sekvenovat napříč introny i exony a řešit nejednoznačnosti je vhodný pro vysoce polymorfní geny jako HLA (Profaiser *et al.*, 2016).

V současné době existuje i takzvané sekvenování třetí generace, které je specifické produkovaním *long reads*. Jde o čtení dlouhých fragmentů DNA, nikoliv o rozbití DNA na malé segmenty. K těmto technologiím patří Nanopore od Oxford Nanopore Technologies a *Single molecule real time (SMRT) sequencing* od Pacific Bioscience (Bravo-Egana *et al.*, 2021).

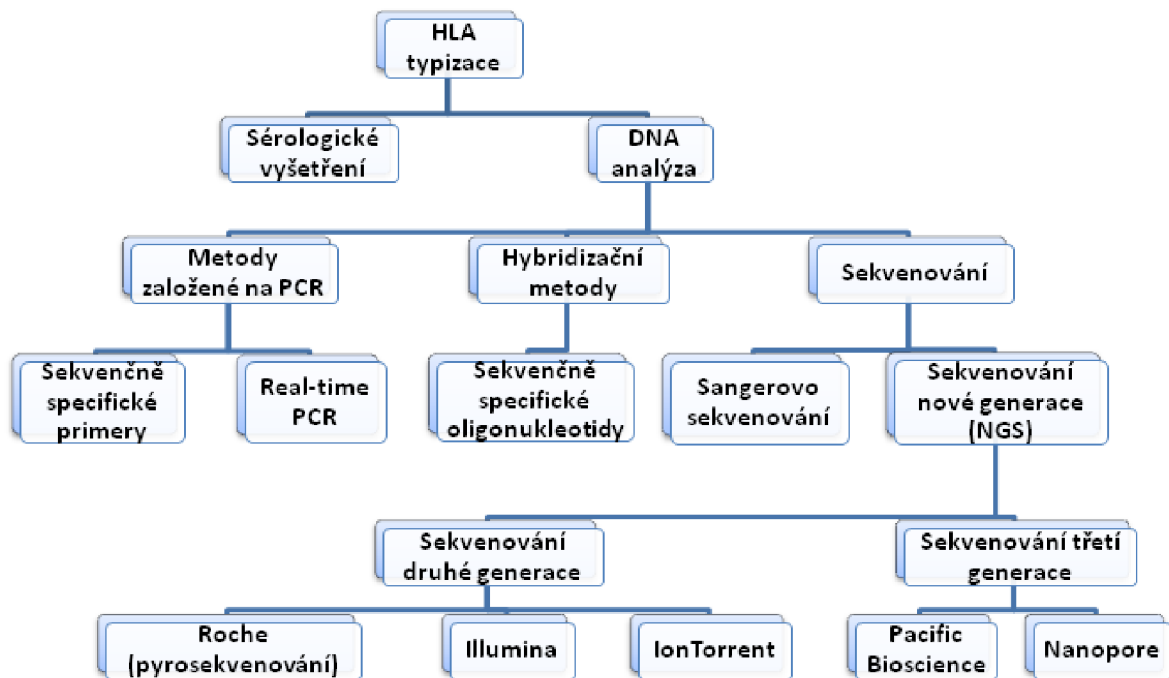
Zdokonalené metody NGS jsou dalším přínosem pro klinickou využitelnost v typizaci HLA. Patří mezi ně metody sekvenování třetí generace, jako je Oxford Nanopore Technologies a SMRT (Cornaby *et al.*, 2021). Další metodou testovanou pro identifikaci HLA alel u dárců je například long-read RNA sequencing, které poskytuje přesné výsledky v relativně krátkém čase (Cornaby *et al.*, 2022).

Typizace pomocí NGS je snadno přijímána transplantací komunitou, jak pro transplantace solidních orgánů, tak pro transplantace krvetvorných kmenových buněk. I když jedním z jeho současných omezení je časová náročnost. Několik připravovaných technologií a postupů nabízí potenciální řešení, jak tuto nevýhodu zmírnit.

Typizace HLA je mimo toto v současné době stále častěji využívána v asociačních studiích k identifikaci genetických variant spojených se zvýšeným rizikem nebo odolností vůči autoimunitním a infekčním onemocněním, a ve farmakogenomice. Řada HLA alel

je spojena se zajímavou škálou nežádoucích účinků, jako je například riziko zvýšené toxicity léků u pacientů s určitými alelami (Cornaby *et al.*, 2021).

**Obrázek 3** Přehled metod HLA typizace (převzato z Bravo-Egana *et al.*, (2021))



### 1.11 Vyšetření polymorfismů

SNP jsou jedním z nejvýznamnějších genetických polymorfismů a jsou využívány jako markery při studiu genetických aspektů onemocnění i ve farmakogenomice.

Jednonukleotidové polymorfismy jsou používány jako molekulární markery v mnoha studiích. Především u komplexních onemocnění, kde onemocnění způsobuje kombinace více rizikových variant. Mezi tyto onemocnění patří například diabetes, autoimunitní onemocnění, rakovina či Alzheimerova choroba.

SNP jsou významnými markery i ve farmakogenomice. Zásadní vliv mají varianty v genech kódující enzymy podílející se na metabolismu léčiv či varianty u genů, na které cílí účinek těchto léčiv, například u kardiovaskulárních onemocnění využívaný *SLCO1B1* (Kim *et Misra*, 2007; Petřková *et al.*, 2018).

Ideální metoda pro detekci SNP by neměla být finančně náročná na vývoj a optimalizaci, reakce by měla být dostatečně robustní pro diagnostiku i ne úplně kvalitních vzorků DNA. Analýza dat musí být nenáročná, ale přesná a reakční formát by měl být flexibilní (Kwok, 2001).

Genotypizační strategii můžeme rozdělit do dvou základních kroků – rozlišení alel a detekce alel.

Rozlišení alel můžeme rozdělit na extenzi pomocí specifických primerů, hybridizaci, ligaci a enzymatické štěpení.

U extenze pomocí specifických primerů můžeme použít sekvenační přístup. Určujeme sekvenci amplifikované DNA, nebo alelicky specifickou PCR, kdy DNA polymeráza amplifikuje cílovou DNA, pouze pokud je primer dokonale komplementární k sekvenci, na kterou nasedá (Kwok, 2001).

Hybridizační metoda využívá různé teplotní stability DNA k rozlišení perfektně sedícího páru templátová DNA – sonda (Kim *et* Misra, 2007).

Ligační přístup používá specifické ligační enzymy k dosažení rozlišení alel. DNA ligáza je vysoce specifická při opravování jednovláknových poškození molekul DNA. Dva sousední oligonukleotidy připojené k templátové DNA jsou ligovány pouze, když jsou perfektně shodné s templátem. Přestože je ligace nejvíce specifická a nejsnadněji optimalizovatelná ze všech mechanismů rozlišení alel, je to nejpomalejší reakce a často vyžaduje největší množství modifikovaných sond (Kwok, 2001).

Rozlišení alel pomocí enzymatického štěpení je založené na schopnosti určité třídy enzymů štěpit DNA v místě specifické sekvence či struktury. Enzymy mohou být použity pro rozlišení alel, pokud se v místě rozpoznávaném enzymem vyskytuje daný SNP a změna toto rozpoznávané místo ovlivňuje (Kim *et* Misra, 2007).

Detekce rozlišených alel může probíhat detekcí světla emitovaného produktem reakce, měřením hmotnosti produktu anebo měřením změny elektrických vlastností v době tvorby produktu.

Detekce luminiscence je využívána například u pyrosekvenování. Luminiscence je emitována při ATP dependentní luciferázové reakci. ATP je produkováno při extenční reakci, kdy k luminiscenci dochází po přidání báze (Kim *et* Misra, 2007). Detekci fluorescence můžeme rozdělit do několika dalších sekcí.

Přímá detekce fluorescence se používá pro zachycení fluorescenčních značících molekul na pevném nosiči a monitorování tvorby dsDNA s DNA interkalačním barvivem, které fluoreskuje jenom v přítomnosti dsDNA. Přímá detekce fluorescence je velmi všestranná a může být do určité míry provedena v multiplexu (Kwok, 2001).

Metody FRET (*fluorescence resonance energy transfer*) se velmi často využívá k detekci v homogenních genotypizačních testech. K přenosu fluorescenční rezonanční

energie (FRET) dochází mezi fluorescenčními barvivými, když emisní spektrum jednoho z barviv (donoru) překrývá excitační spektrum druhého (akceptoru) (Kim *et al.*, 2007).

Při detekci prostřednictvím tzv. "molekulárního majáku" (*molecular beacon*) hybridizace k cílové DNA otevírá vlásenkovou strukturu, která drží fluorofor v blízkosti univerzálního zhášedce. Cílová DNA, která je zcela komplementární k sekvenci smyčkové části molekulárního majáku, s ním hybridizuje a otevře vlásenku, což vede ke vzniku fluorescence. Molekulární maják lišící se v jedné bázi nehybridizuje k cílové DNA dostatečně silně, aby narušil vlásenkovou strukturu, a není pozorována žádná fluorescence (Tyagi *et al.*, 1998).

Detekce na bázi fluorescenční polarizace (FP) využívá změnu polarizace v rovině polarizovaného světla molekulou fluorescenčního barviva v důsledku změny jeho molekulové hmotnosti za podmínek konstantní teploty a viskozity rozpouštědla. Polarizace fluorescence se používá v klinické diagnostice a vazebných testech. Výhody detekce založené na polarizaci fluorescence jsou nižší spotřeba fluorescenčně značených nukleotidů ve srovnání s FRET nebo metodami přímé detekce fluorescence (Kwok, 2001; Kim *et al.*, 2007).

Elektrochemická detekce monitoruje změny v elektrochemických vlastnostech produktů reakce. Elektrochemické vlastnosti sondy jsou změněny, když se k ní naváže komplementární DNA. Tento proces je znásoben použitím feromagnetického značení (Kwok, 2001).

Hmotnostní spektrometrie (*MS-mass spectrometry*) měří molekulární hmotnost vytvořených produktů a je proto nejpřímější detekční metodou. MS s vysokým rozlišením může snadno rozpoznat i molekuly DNA, které se liší jen jednou bází.

MALDI-TOF MS je široce používaná metoda pro hmotnostní analýzu různých biomolekul, jako jsou oligonukleotidy, peptidy, proteiny a oligosacharidy. Zahrnuje použití malé organické molekuly zvané matrice, která absorbuje energii z laserového zdroje určité vlnové délky pro ionizaci. Když jsou analyty smíchány a kokrystalizovány s matricí, jsou ionizovány ve formě intaktních molekul v důsledku přenosu energie z molekul matrice. Výsledné ionty jsou urychlovány napěťovým gradientem ve vakuu a odděleny dobou letu každého iontu, která závisí na hmotnosti a náboji iontu. V MALDI-TOF MS může být směs mnoha oligonukleotidů rychle separována a přesně analyzována bez specifických detekčních značek, jako jsou fluorofory. Proto je tato metoda používána v různých genotypizačních metodách. Vhodný návrh sond umožňuje kontrolované multiplexování.

Agena Bioscience MassARRAY<sup>®</sup> Systém umožňuje analýzu až 40 uživatelsky definovaných jednonukleotidových polymorfismů v jednom plexu (Ellis *et al.*, 2017, Petřek *et al.*, 2016). Platforma umožňuje genotypizaci SNP, kvantitativní genovou expresi



a analýzu CNV, methylační analýzu a komparativní sekvenční analýzu. Sequenom iPLEX<sup>®</sup> Gold Assay pro genotypování SNP poskytuje data nejvyšší kvality při nízkých nákladech na genotyp (Cullinan, 2008).

Test TaqMan<sup>®</sup> je jednokrokový test, který využívá detekci založenou na fluorescenci. Je ovšem omezen na genotypizaci jednoho SNP na test. Kombinuje hybridizaci a 5' nukleázovou aktivitu polymerázy spojenou s fluorescenční detekcí. Využívá čtyři oligonukleotidy: dvě alelově specifické oligonukleotidové sondy a pár PCR primerů lemujících oblast obsahující SNP. Alelově specifické sondy nesou fluorescenční barvivo na jednom konci (reportér) a nefluorescenční barvivo na druhém konci (zhášedlo). Neporušené sondy nevykazují žádnou fluorescenci kvůli těsné blízkosti mezi reportérovým a zhášecím barvivem. Pro detekci SNP se provádí PCR reakce s genomovou DNA v přítomnosti všech čtyř sond pomocí polymerázy, která má 5' exonukleázovou aktivitu. Během prodlužování PCR primeru enzym štěpí pouze hybridizovanou sondu, která je dokonale kompatibilní, čímž uvolňuje reportérové barvivo ze zhášedla. Reportérové barvivo generuje fluorescenční signál v nepřítomnosti svého zhášedla, zatímco nesprávně spárovaná sonda zůstává nedotčena a nevykazuje žádnou fluorescenci. Výsledkem je, že obě alely mohou být detekovány během reakce na základě hybridizace příslušných sond. TaqMan<sup>®</sup> vyžaduje pouze jeden enzymatický krok s jednoduchým operačním postupem (Francisco *et al.*, 2005).

Pro aplikaci identifikace jednonukleotidových polymorfismů v medicíně s co nejrychlejším přístupem byly testovány i systémy na principu *point of care*, například Spartan RX Systém založený na testování bukalního stěru, který využívá optickou detekci a následnou automatickou analýzu pro vyhodnocení fluorescenčních signálů z oligonukleotidových sond, které se váží na PCR amplikony a tím určuje genotyp (Petřek *et al.*, 2016).

## 2 Cíle práce

Obecným cílem této dizertační práce bylo přispět k objasnění patofyziologie sarkoidózy studiem polymorfismů v genech, převážně s imunitními funkcemi, které by mohly mít vliv na její vznik a průběh. Replikovat výsledky předchozích studií u populací, u kterých nebyla spojitost polymorfismů se sarkoidózou ještě studována, či se podrobněji zaměřit na souvislost těchto polymorfismů s jednotlivými klinickými stadii a napomoci tak identifikaci možných markerů pro budoucí využití pro diagnostiku, určování prognózy nebo k indikaci léčby u sarkoidózy.

Jednotlivé konkrétní cíle pak byly:

- Výběr SNP spojených se sarkoidózou, návrh primerů pro jejich analýzu pomocí MassArray a optimalizace této metody
- Analýza asociací vybraných SNP u pacientů se sarkoidózou v několika populacích (česká, polská, řecká a korejská populace)
- Analýza asociací vybraných SNP s klinickými fenotypy sarkoidózy (stadia onemocnění, přítomnost Löfgrenova syndromu, vývoj onemocnění po dvou letech sledování, přítomnost mimoplicních projevů onemocnění, indikace k léčbě)
- Analýza HLA lokusů pomocí NGS ve vybraných populacích (česká, korejská a chorvatská populace)
- Asociace HLA variant se sarkoidózou a asociace HLA variant s klinickými fenotypy sarkoidózy.

## 3 Materiál a metody

### 3.1 Biologický materiál

V této *case-control* studii byli analyzováni pacienti se sarkoidózou z České republiky, Řecké republiky, Korejské republiky a Polské republiky. Pacienti byli diagnostikováni na základě *ATS, ERS, and WASOG statement on sarcoidosis (1999)*.

#### 3.1.1 Česká populace

V české populaci diagnostikovala 301 pacientů se sarkoidózou ve *Fakultní nemocnici Brno na klinice nemocí plicních a tuberkulózy* doc. MUDr. Martina Doubková, Ph.D.. Charakteristika této patientské populace pro analýzu 20 SNP a HLA lokusů je uvedena v Tab. 1.

Byly shromážděny vzorky 203 kontrolních zdravých nepříbuzných jedinců pro analýzu 20 jednonukleotidových polymorfismů, a pro analýzu HLA lokusů pak vzorky 309 kontrolních jedinců na pracovištích školitele - *Lékařské fakulty Univerzity Palackého v Olomouci (Ústavy patologické fyziologie, a molekulární a translační medicíny)*, resp. ve *Fakultní nemocnici Olomouc - Laboratoř kardiogenomiky* (Tab. 1).

Z pracovišť byly obdrženy vzorky krve ve zkumavkách K<sub>3</sub>EDTA, ze kterých byla následně v naší laboratoři vyizolována DNA.

Pacienti i kontrolní jedinci byli do studie zařazováni v období od 2.3.2011 do 10.3.2022 a podepsali informovaný souhlas s účastí na studii schválený etickou komisí daných institucí.

Pacienti byli sledováni po dobu minimálně dvou let od stanovení diagnózy pro zjištění vývoje nemoci a reakce na léčbu.

U českých pacientů byly kromě plic postiženy v 91 případech i jiné orgány. U 25 pacientů byla pozorována generalizovaná lymfadenopatie (u 16 pacientů lokalizované postižení lymfatických uzlin), u 25 pacientů postižení kůže, u 16 pacientů postižení sleziny a u 2 pacientů trombocytopenie. Další mimoplicní lokalizace byly: játra (n = 8), srdce (n = 8), oči (n = 4), nervový systém (n = 4), kosti (n = 3), kostní dřeň (n = 1), měkké tkáně (n = 1), nadvarle (n = 1), příušní žláza (n = 1), slinná žláza (n = 1), mandle (n = 1) a střeva (n = 1).

**Tabulka 1: Charakteristika analyzované české populace *n*-počet jedinců**

<b>Česká republika</b>			
	<b>Pacienti <i>n</i> = 301</b>	<b>Kontroly NGS analýza <i>n</i> = 309</b>	<b>Kontroly SNP analýza <i>n</i> = 203</b>
Muži : ženy	149 : 152	143 : 166	94 : 109
Průměrný věk	46.77±11.49 let	32.38±8.63let	31.94±8.89 let
Věkové rozmezí	24 – 72 let	18 – 61 let	18 – 58 let
Löfgrenův syndrom	55	-	-
Mimoplicní postižení	91	-	-
Léčba	194	-	-
Remise : Perzistence	142 : 134	-	-
Rentgenové stadium 1/2/3/4	95/142/58/6	-	-

### 3.1.2 Polská populace

V polské populaci diagnostikovala 104 pacientů na *Department of Pulmonology, Medical University of Gdansk* Prof. Dr. Habil. Anna Dubaniewicz. Jedinci byli do studie zařazováni v období od roku 2010 do 14.11.2016. Pacienti byli sledováni po dobu minimálně dvou let od stanovení diagnózy pro zjištění vývoje nemoci a reakce na léčbu. Charakteristika populace je uvedena v Tab. 2.

Pro analýzu 20 jednonukleotidových polymorfismů byla využita populace 100 kontrolních zdravých nepříbuzných jedinců, kteří byli získáni na *Department of Forensic Medicine, Medical University of Gdansk* (Dr. Habil. Krzysztof Rębała). Charakteristika kontrolních jedinců je uvedena v Tab. 2. Pacienti i kontrolní subjekty podepsali informovaný souhlas s účastí na studii schválený etickou komisí instituce.

Z pracovišť byly obdrženy vzorky DNA.

**Tabulka 2: Charakteristika polské analyzované populace *n*-počet jedinců**

<b>Polská republika</b>		
	<b>Pacienti <i>n</i> = 104</b>	<b>Kontrolní populace <i>n</i> = 100</b>
Muži: ženy	64 : 40	53 : 47
Průměrný věk	41.38±11.50 let	32.99±12.17 let
Věkové rozmezí	21 – 68 let	16 – 79 let
Löfgrenův syndrom	18	-
Léčba	38	-
Remise: Stabilizace: Progrese	31 : 15 : 25	-
Stadium 1/2/3/4	33/48/16/6	-

### 3.1.3 Řecká populace

V Řecké populaci diagnostikovala 103 pacientů v *Sarcoidosis Center, General Hospital for Chest Diseases, Athens* Dr. Angeliki Rapti. Jedinci byli do studie zařazováni v období od 11.10.2017 do 13.11.2019. Pacienti byli sledováni po dobu minimálně dvou let od stanovení diagnózy pro zjištění vývoje nemoci a reakce na léčbu.

Z pracoviště byly obdrženy vzorky krve ve zkumavkách K<sub>3</sub>EDTA, ze kterých byla následně v naší laboratoři vyizolována DNA.

Základní charakteristika řecké populace je uvedena v Tab. 3. V analyzované řecké populaci pacientů se sarkoidózou bylo 57 pacientů léčeno kortikoidy, 5 z nich navíc v kombinaci s methotrexátem. Tato kombinovaná léčba byla použita při výskytu pulmonární hypertenze a u několika případů s kardiopatií. U 2 pacientů byly kortikosteroidy podávány v kombinaci s long-acting  $\beta$ 2-agonist. U 5 pacientů byla léčba pozastavena. U 46 pacientů nebyla indikována žádná léčba.

U řeckých pacientů bylo nejčastějším mimoplicním postižením zasažení kůže (n = 13), kardiopatie zjištěná na základě MRI vyšetření (n = 12), uveitida (n = 8), periferní lymfadenopatie (n = 4), artropatie (n = 3), hyperkalcémie (n = 3), nodózní erytém (n = 2), pulmonární hypertenze (n = 2) a neurosarkoidóza (n = 1).

Pro analýzu 20 jednonukleotidových polymorfismů byla použita populace 100 kontrolních zdravých nepříbuzných jedinců, tito byli získáni mezi zdravými dárci krve v *Department of Immunology and Histocompatibility, Laiko General Hospital, Athens* (Dr. Kalliopi Adam) (Tab. 3). Z pracoviště byly poskytnuty vzorky DNA.

Pacienti i kontrolní jedinci podepsali informovaný souhlas s účastí na studii schválený etickou komisí daných institucí.

**Tabulka 3: Charakteristika řecké analyzované populace n-počet jedinců**

<b>Řecká republika</b>		
	<b>Pacienti n = 103</b>	<b>Kontroly n = 100</b>
Muži: ženy	30 : 73	30 : 70
Průměrný věk	56.55±11.57 let	43.68±12.02 let
Věkové rozmezí	28 – 85 let	21 – 72 let
Löfgrenův syndrom	11	-
Mimoplicní postižení	44	-
Léčba	57	-
Remise: Stabilizace: Progrese	13 :87 :3	-
Stadium 1/2/3/4	29/59/4/5	-

### 3.1.4 Korejská populace

V Korejské populaci diagnostikoval 139 pacientů v *Department of Pulmonary and Critical Care Medicine, Asan Medical Centre, Seoul M.D. Jin Woo Song Ph.D.*. Jedinci byli do studie zařazováni v období od 17.5.1994 do 31.5.2019. Pacienti podepsali informovaný souhlas s účastí na studii schválený etickou komisí instituce. Z pracoviště byly obdrženy vzorky DNA.

Pacienti byli sledováni v rozmezí jednoho a půl až dvacetišesti let pro zjištění vývoje nemoci a reakce na léčbu.

Pro analýzu HLA lokusů byly použity alelické frekvence Korejské populace z databáze allele frequencies.net (*South Korea pop 1* pro lokusy *HLA-DRB1, -DQA1, -DQB1* a *South Korea pop 3* pro lokusy *HLA-A, -B, -C*).

Pro analýzu 20 jednonukleotidových polymorfismů byly použity alelické frekvence SNP v *East Asian population* z databáze ensembl.org.

Základní charakteristika patientské populace je v Tab. 4. V korejské patientské populaci bylo pomocí kortikoidů léčeno 92 pacientů, 17 z nich v kombinaci s Methotrexátem, 7 v kombinaci s Azathioprinem, u 6 pacientů v kombinaci s Mykofenolát mofetilem, u 4 pacientů v kombinaci s Hydroxychloroquinem a u jednoho pacienta v kombinaci s Cyklosporinem. Jeden pacient byl léčen pouze Hydroxychloroquinem a dva pouze Methotrexátem. U 46 pacientů nebyla indikována žádná léčba.

U korejských pacientů patřily mezi mimoplicní lokalizace sarkoidózy oči ( $n = 27$ ) a mimoplicní lymfatické uzliny ( $n = 27$ ), následované zasažením kůže ( $n = 24$ ), srdce ( $n = 7$ ), jater ( $n = 7$ ), sleziny ( $n = 4$ ), ledvin ( $n = 3$ ) a pohybového ústrojí ( $n = 2$ ). Neurosarkoidóza, stejně jako zasažení slinivky, se vyskytovala v 1 případě a u 46 pacientů byla pozorována hyperkalcémie.

**Tabulka 4: Charakteristika analyzované korejské patientské populace  $n$ -počet jedinců**

	Korejská republika	
	Pacienti	
	Analýza SNP $n = 139$	Analýza NGS $n = 103$
Muži : ženy	40 : 99	24 : 79
Průměrný věk	46.9±12.5 let	46.9 ± 12.9 let
Věkové rozmezí	27 – 71 let	27 – 71 let
Mimoplicní postižení	96	73
Löfgrenův syndrom	0	0
Léčba	95	67
Remise : Stabilizace : Progrese	79 : 40 : 9	58 : 32 : 5
Stadium 0/1/2/3/4	4/37/85/6/5	2/28/64/6/3

### 3.1.5 Chorvatská populace

V chorvatské populaci byla provedena NGS analýza HLA lokusů u 120 zdravých nepříbuzných dobrovolných dárců krve. DNA pro NGS analýzu byla izolována na *Department of Medical Chemistry, Biochemistry and Clinical Chemistry, Faculty of Medicine, University of Osijek* Assist. Prof. Stana Tokić Ph.D.. Ve vyšetřované populaci bylo 34 žen a 86 mužů. Medián věku jedinců byl 36 let a věkové rozmezí bylo 20–61 let. Jedinci pocházeli z východní části Chorvatska.

## 3.2 Izolace DNA

### 3.2.1 Přístrojová izolace DNA

DNA byla izolována metodou mikroizolace pomocí přístroje NorDiag Arrow (NorDiag, Oslo, Norway). Izolace byla provedena z 500  $\mu$ l nesrážlivé periferní krve odebrané do zkumavek K<sub>3</sub>EDTA. Izolace probíhala pomocí standardního přednastaveného protokolu Blood DNA 500 v.2.0. Eluční objem, tedy výsledný objem získané DNA byl 400  $\mu$ l (Petřek *et al.*, 2016).

### 3.2.2 Izolace DNA vysolovací metodou

DNA byla zároveň izolována vysolovací metodou, tzv. makroizolací. Pro tuto metodu byly použity 3 ml periferní krve odebrané do zkumavek K3-EDTA, smíchané s 42 ml LB puftru, který obsahoval detergent Triton X100 zajišťující lýzi buněk (0.3M sacharóza; 0.01M Tris (pH 7.5); 5mM MgCl<sub>2</sub>\*6H<sub>2</sub>O; 1% Triton X100). Vzniklá směs byla centrifugována při 4000 ot/min a získaný pelet byl promyt roztokem Tris (pH 7.5). Následně byl pelet rozpuštěn v 950  $\mu$ l roztoku Tris (pH 7.5). K získané směsi bylo přidáno 350  $\mu$ l roztoku P (10mM Tris (pH 8); 40mM EDTA; 4mM CaCl<sub>2</sub>; 300mM NaCl; 4% SDS) a 40  $\mu$ l proteinázy K. Následovala inkubace 24 hodin při teplotě 37 °C. Po inkubaci byla směs vychlazená na pokojovou teplotu a bylo k ní přidáno 400  $\mu$ l 5M NaCl. Důkladným protřepáním došlo k vysrážení proteinů, které byly odděleny centrifugací 15 minut při 4000 ot/min.

DNA byla precipitována v 96% EtOH, promyta v 70% EtOH, a následně rozpuštěna v 350  $\mu$ l Tris-EDTA puftru (Miller *et al.*, 1988).

### 3.2.3 Měření koncentrace DNA

Koncentrace DNA byla stanovena měřením na spektrofotometru Nanodrop 2000 (Thermo Fisher Scientific Inc., Wilmington, Delaware USA). Koncentrace DNA

se pohybovala v rozmezí 30-100 ng/μl u mikroizolací a v rozmezí 150-700 ng/μl u makroizolací. Čistota DNA (poměr absorbancí vlnových délek 260nm/280nm) se pohybovala v rozmezí 1.7-1.9 u obou typů izolace.

### 3.3 Analýza technologií MassArray®

#### 3.3.1 Návrh primerů

Primery byly navrženy na 20 SNP (Tab. 5) pomocí programu *Assay Design Suite 2.0* (Agena Bioscience, Inc., San Diego, CA, USA).

**Tabulka 5: SNP vybrané pro analýzu (Chr. - chromozom)**

Gen	SNP	Minoritní alela	Chr.	Region	Funkce
<i>ANXA11</i>	rs1049550	A	10	Exon (Arg230Cys)	Kalcium dependentní fosfolipidové-vazebné místo
<i>TNFA</i>	rs1800629	A	6	Promotor	Cytokin
<i>TGF-β2</i>	rs1891467	G	1	Intron	Cytokin (aktivace SMAD)
<i>GREM1</i>	rs1919364	C	15	Promotor	Cytokin
<i>(HLA-DRA, HLA-DRB5)</i>	rs1964995	C	6	Intergenní	Imunita, prezentace antigenu
<i>NOD2</i>	rs2066844	T	16	Exon (Arg702Trp)	Imunitní systém
<i>BTNL2</i>	rs2076530	C	6	Exon (Ser360Gly)	Regulace T-lymfocytů (Th1)
<i>HLA-DQA1</i>	rs2187668	T	6	Intron	Imunita, prezentace antigenu
<i>CACFD1</i>	rs3124765	T	9	Exon (Ile58=)	Homeostáza kalcia
<i>C6orf10</i>	rs3129927	C	6	Intron	Riziko pro poruchy imunity
<i>ATF6B</i>	rs3130288	A	6	5'-UTR; vMHC-III	Stres endoplazmatického retikula
<i>HLA-DRA</i>	rs3135394	G	6	Intron	Imunita, prezentace antigenu
<i>TAP2</i>	rs3819717	T	6	Intron	Membránový transport a prezentace antigenu
<i>HLA-DRB1</i>	rs3830135	A	6	Intron	Imunita, prezentace antigenu
<i>TGF-β3</i>	rs3917165	T	14	Intron	Cytokin (diferenciace a vývoj buněk)
<i>TGF-β3</i>	rs3917200	G	14	Intron	Cytokin (diferenciace a vývoj buněk)
<i>HLA-DPB1</i>	rs9277357	G	6	Promotor	Imunita, prezentace antigenu
<i>LRRC16A</i>	rs9295661	C	6	Intron	Produkce megakaryocytů a destiček
<i>HLA-DRB1*03:01</i>	rs2040410	T	6	Intron	Imunita, prezentace antigenu
<i>HLA-DRB1*15:01</i>	rs3135388	A	6	Intergenní	Imunita, prezentace antigenu



Jednotlivé lyofylizované primery pro PCR reakci byly naředy PCR grade vodou na koncentraci 100 pmol/ $\mu$ l. Extenční primery byly naředy PCR grade vodou na koncentraci 400 pmol/ $\mu$ l. Sekvence primerů jsou uvedeny v tabulkách (Tab. 6, Tab. 7).

**Tabulka 6: Nukleotidové sekvence *forward* (F) a *reverse* (R) primerů pro amplifikační reakci *MassArray* analýzy**

Název primeru	Sekvence primerů	
	F	R
rs3130288	ACGTTGGATGATCAGCTCCGCCATCTTTCC	ACGTTGGATGTTACTGGTCGGTAGAGCTTC
rs9295661	ACGTTGGATGTTTCTAGCTCTCTAGTGCTC	ACGTTGGATGACCACACTTTTGGACTGCAC
rs3129927	ACGTTGGATGGAAAATGGTGCCTCATAGCG	ACGTTGGATGTCCTGCCACAACATAAAAGG
rs2187668	ACGTTGGATGTGAGGTGACACATATGAGGC	ACGTTGGATGTGAGGTGACACATATGAGGC
rs3135394	ACGTTGGATGCAACAGCACCAATGTTGCTC	ACGTTGGATGCATCTCCATGGTACAGCAGA
rs1800629	ACGTTGGATGGCATCCTGTCTGGAAGTTAG	ACGTTGGATGCTGATTTGTGTGTAGGACCC
rs1891467	ACGTTGGATGTGCATATAGGGTATAGGCGG	ACGTTGGATGTCAGCAAGCCATTCAAGAC
rs3917165	ACGTTGGATGTATTGAGAGCTGGACTAGGG	ACGTTGGATGTATTGAGAGCTGGACTAGGG
rs3917200	ACGTTGGATGGGAAGGAAAGTGACAATCTC	ACGTTGGATGCCTGGACATCAGAATTAAGC
rs9277357	ACGTTGGATGTCTGGCAGTTGATTACTGTG	ACGTTGGATGTCTGGCAGTTGATTACTGTG
rs3830135	ACGTTGGATGCAGGACATTAGAGCCGTTTG	ACGTTGGATGAGAAAGGTGAGGAGCCTTTG
rs1964995	ACGTTGGATGTCCCTTTAACTGCAGCTCAC	ACGTTGGATGGCAGATGCCTTTTGTGTGAG
rs1049550	ACGTTGGATGGAAAAGTAGGATCTGTGCCG	ACGTTGGATGGAGCATCTCATGACCCCTAC
rs3819717	ACGTTGGATGGTGTCTGAGAAAATGTGTCTG	ACGTTGGATGGGAGCAATAGGCTAAACTAC
rs2076530	ACGTTGGATGAGTGGCAGGAGCAGGTATTG	ACGTTGGATGATACTCAGTGCCAGACCTTC
rs1919364	ACGTTGGATGGAAGCCTTGACTTTTGACGG	ACGTTGGATGGAGCCTGTCGCCAACATGA
rs2066844	ACGTTGGATGTTTCAGATCACAGCAGCCTTC	ACGTTGGATGATGGAGTGGAAAGTGCTTGCG
rs3124765	ACGTTGGATGTTTCAAGCTTGC GCGATCTC	ACGTTGGATGAGATTCCCAGCAGCAAGAC
rs2040410	ACGTTGGATGTGAGGTACCTCATTAGGCAG	ACGTTGGATGGTATATGTAAGAGTCCAGG
rs3135388	ACGTTGGATGCCTCTTGTGTGGTCTAGTC	ACGTTGGATGTTGCTCAGTAGAGATCTCCC

**Tabulka 7: Nukleotidové sekvence primerů pro reakci prodloužení o jednu bázi *MassArray* analýzy**

Název primeru	Sekvence primeru
rs3130288	ACCGTCTCCTGGTTG
rs9295661	AGAAAAACTGAAATTGAAGTGAC
rs3129927	AAAAAGAAGAAGAAATCAAAGTAG
rs2187668	CACCACATGGTCCTCA
rs3135394	CAGATCCCCATTCCTTACT
rs1800629	CCCGGGAGGCTGAACCCCGTCC
rs1891467	TTAGACGTCTCAAGGCCTCA
rs3917200	GGGATTGACAATCTCCTGTCTGA
rs3917165	GCCATTTTCTCTATTTTGTTTAA
rs9277357	AGTTTTTGCCAATTGCGTTA
rs3830135	CTGGAGGAGGAACTATGG
rs1964995	TGCAGCTCACATTATTTT
rs1049550	ATGACTGCCTGGGGAGT
rs3819717	TGTGTCGTTAGATGGTTTCAT
rs2076530	GTGGATCTGAAGGTGGTA
rs2066844	TCTGAGAAGGCCCTGCTC
rs3124765	ATCCACCCTCTGAACAT
rs2040410	CCAAGTGCAGTGTATGTA
rs3135388	ATCTCCCAACAAACCAA

### 3.3.2 Příprava produktů pro genotypizaci

Amplifikace a příprava produktů pro genotypizaci byly provedeny pomocí *iPLEX*<sup>®</sup> *Gold Reagent* Kitu (Agena Bioscience, Inc., San Diego, CA, USA). Nejprve byly vzorky DNA naředěny na koncentraci 10 ng/μl, následně proběhla amplifikace.

PCR reakční směs byla připravena dle rozpisu uvedeného v Tab. 8 a napipetována do 96 jamkové destičky. Poté byl do příslušných jamek napipetován 1 μl DNA o koncentraci 10 ng/μl. Jako negativní kontrola byla použita PCR grade voda. Jako pozitivní kontrola byl použit vždy jeden ze vzorků komerční DNA (NA16689, NA17084, NA17252) získaných z *NIGMS Human Genetic Cell Repository* (Coriell Institute for Medical Research, New Jersey, USA). Amplifikace probíhala v termocykleru dle doporučeného teplotního schématu (Tab. 9).

**Tabulka 8: Složení reakční směsi pro amplifikační reakci *MassArray* analýzy**

Reagencie	Množství pro 1 reakci (μl)
PCR grade voda	2.4
Pufr 10 x	0.5
MgCl <sub>2</sub> 25 mmol/l	0.4
dNTP mix 25 mmol/l	0.2
Primer mix 1 μmol/l	0.5
HotStarTaq 5 U/μl	0.1

**Tabulka 9: Teplotní schéma inkubace vzorků pro amplifikační reakci *MassArray* analýzy**

Počet cyklů	Teplota	Fáze	čas
1x	94 °C	Pre-denaturace	2 min
45 x	94 °C	Denaturace	30 s
	56 °C	Nasedání primerů	30 s
	72 °C	Elongace	60 s
1x	72 °C	Finální extenze	5 min
1x	4 °C	Chlazení	∞

Po prvotní PCR amplifikaci následovala inaktivace volných deoxyribonukleotidů pomocí *shrimp alkaline* fosfatázy (SAP). Byla připravena reakční směs dle Tab. 10. Do každé reakční jamky byly napipetovány 2  $\mu$ l připravené směsi a probíhala inkubace dle teplotního schématu (Tab. 11).

**Tabulka 10: Složení reakční směsi pro SAP reakci *MassArray* analýzy**

Reagencie	Množství pro 1 reakci ( $\mu$ l)
PCR grade voda	1.53
SAP Pufr 10 x	0.17
SAP enzym 1.7 U/ $\mu$ l	0.3

**Tabulka 11: Teplotní schéma inkubace vzorků pro SAP reakci *MassArray* analýzy**

Počet cyklů	Teplota	Fáze	Čas
1 x	37 °C	Inkubace	40 min
1 x	85 °C	Inkubace	5 min
1 x	4 °C	Chlazení	$\infty$

Posledním krokem byla iPLEX reakce. Jedná se o prodloužení získaného amplikonu o jednu bázi, připojením extenčních primerů o známé molekulové hmotnosti. Byla připravena reakční směs (Tab. 12), která byla napipetována ke vzorům v 96 jamkové desce a následovala inkubace v termocykleru dle teplotního schématu (Tab. 13).

**Tabulka 12: Složení reakční směsi pro iPLEX reakci *MassArray* analýzy**

Reagencie	Množství pro 1 reakci ( $\mu$ l)
PCR grade voda	0.739
iPLEX <sup>®</sup> pufrPlus 10x	0.2
iPLEX <sup>®</sup> -Terminační mix 0.5	0.1
Primer mix	0.94
iPLEX <sup>®</sup> enzym	0.021

**Tabulka 13: Teplotní schéma inkubace vzorků při iPLEX reakci *MassArray* analýzy**

Počet cyklů	Teplota	Fáze	Čas	
1x	94 °C	<b>Predenaturace</b>	30 s	
40 x	94 °C	<b>Denaturace</b>	5 s	
	5x	52 °C	<b>Nasedání primerů</b>	5 s
		80 °C	<b>Elongace</b>	5 s
1x	72 °C	<b>Finální extenze</b>	3 min	
1x	4 °C	<b>Chlazení</b>	∞	

Pomocí resinu byly v dalším kroku odstraněny nežádoucí příměsi, především soli, které by mohly negativně ovlivnit následnou analýzu přístrojem Sequenom. Po odsolení byly vzorky přístrojem Nanodispenser nanесeny na čip s maticí, který byl analyzován pomocí platformy *MassARRAY* založené na principu MALDI-TOF MS (Agena Bioscience, San Diego, CA, USA). Výsledky byly získány pomocí programu *MassARRAY Typer 4.2* (Agena Bioscience, San Diego, CA, USA) a následně ručně kontrolovány pomocí *call cluster plot* (Sikorova *et al.*, 2020; Sterclova *et al.*, 2021).

### 3.4 End point genotypizace

U SNP rs3917200 (*TGF-β3*) byly vzorky genotypizovány i metodou *End point genotyping TaqMan assay* (Thermo Fisher Scientific Inc., Waltham, MA USA) pomocí přístroje LightCycler 480 (Roche Diagnostics, Mannheim, Germany). Byla připravena reakční směs (Tab. 14). Následně bylo do příslušných jamek napipetováno 4.5 μl templátové DNA (10 ng/μl). Inkubace probíhala dle schématu (Tab. 15) se zachycením dat v každém cyklu. Poté byla provedena *end point* analýza dat (Sikorova *et al.*, 2020; Sterclova *et al.*, 2021).

**Tabulka 14: Složení reakční směsi pro *TaqMan assay***

Reagencie	Množství pro 1 reakci (μl)
PCR grade voda	0.25
LightCycler 480 Probes Master	5
TaqMan SNP genotyping assay	0.25

**Tabulka 15: Teplotní schéma inkubace vzorků pro *TaqMan assay***

Počet cyklů	Teplota	Fáze	Čas
1x	95 °C	<b>Predenaturace</b>	5 min
45 x	95 °C	<b>Denaturace</b>	10 s
	60 °C	<b>Nasedání primerů</b>	45 s
	72 °C	<b>Elongace</b>	1 s
1x	40 °C	<b>Finální extenze</b>	30 s

### 3.5 NGS

HLA genotypy pro *HLA-A*, *-B*, *-C*, *-DRB1*, *-DQA1* a *-DQB1* lokusy byly určeny pomocí Omixon Holotype HLA 96/7 kitu (Omixon Biocomputing Ltd, Budapest, Hungary) v konfiguraci pro Illumina MiSeq. Tato metoda umožňuje typizaci HLA lokusů pomocí kombinovaného přístupu spočívajícího v prodloužení délky sekvence a pokročilých bioinformatických nástrojů. Má potenciál standardu pro přesnou typizaci variant HLA s rozlišením až na 4 pole a generování HLA plné délky s co nejméně nejednoznačnostmi (Kishore *et Petrek*, 2018).

Prvním krokem pracovního postupu byla *long-range PCR*, kdy byly amplifikovány jednotlivé HLA lokusy I a II třídy.

Byla připravena reakční směs (Tab. 16, Tab. 17). Následně bylo do jamek napipetováno 5  $\mu$ l genomické DNA ( $\approx 30$  ng/ $\mu$ l). Vzorky byly inkubovány v termocykleru Mastercycler nexus (Eppendorf, Hamburg, Germany) dle teplotního schématu (Tab. 18, Tab. 19).

**Tabulka 16: Složení reakční směsi pro amplifikaci lokusů *HLA-A*, *-B*, *-C*, *-DRB1* a *-DQA1***

Reagencie	Množství pro 1 reakci ( $\mu$ l)
PCR grade voda	13.85
PCR pufr 10x	2.5
dNTP mix 10 mM	1.25
Primer mix	2
LR PCR enzym	0.4

**Tabulka 17: Složení reakční směsi pro amplifikaci lokusů *HLA-DQB1***

Reagencie	Množství pro 1 reakci ( $\mu$ l)
PCR grade voda	7.85

<b>PCR pufr 10x</b>	2.5
<b>dNTP mix 10 mM</b>	1.25
<b>Primer mix</b>	2
<b>LR PCR enzym</b>	0.8
<b>Kombinovaný enhancer</b>	5.6

**Tabulka 18: Teplotní schéma inkubace vzorků pro amplifikaci lokusů HLA I. třídy**

Počet cyklů	Fáze	Teplota	Čas
1x	<b>Predenaturace</b>	95 °C	3 min
35x	<b>Denaturace</b>	95 °C	15 sec
	<b>Nasedání primerů</b>	65 °C	35 sec
	<b>Elongace</b>	68 °C	5 min
1x	<b>Finální extenze</b>	68 °C	10 min
1x	<b>Chlazení</b>	4 °C	∞

**Tabulka 19: Teplotní schéma inkubace vzorků pro amplifikaci lokusů HLA II. třídy**

Počet cyklů	Fáze	Teplota	Čas
1x	<b>Predenaturace</b>	95 °C	3 min
35x	<b>Denaturace</b>	93 °C	15 sec
	<b>Nasedání primerů</b>	60 °C	35 sec
	<b>Elongace</b>	68 °C	9 min
1x	<b>Finální extenze</b>	68 °C	10 min
1x	<b>Chlazení</b>	4 °C	∞

Následovala kvantifikace DNA amplikonu pomocí fluorometru EnSpire Multimode plate reader (PerkinElmer, Waltham, MA, USA) za použití *QuantiFluor fluorescent dsDNA staining system* (Promega, Madison, Wisconsin, USA). Následně byly všechny vzorky naředěny tak, aby jejich finální koncentrace byla v rozsahu 50–120 ng/μl. Všech 6 produktů pro jeden konkrétní vzorek bylo následně smícháno a finální objem pro jeden vzorek byl 35 μl. Poté byla provedena purifikace produktů za použití *ExoSAP-IT enzyme mix* (Affymetrix Inc., Santa Clara, CA, USA).

V dalších krocích byla připravena knihovna pro sekvenování. Nejprve byla provedena enzymatická fragmentace vzorků, fragmentované konce byly opraveny, adenylvány a v posledním kroku byly na konce ligovány specifické adaptérové indexy. Takto upravené vzorky byly následně smíchány do výsledné knihovny o konečném objemu 900 μl.

Následně byly z knihovny vyselektovány fragmenty odpovídající délky (650 -1300 bp) pomocí magnetických kuliček *AMPure XP beads* (Beckman Coulter, Beverly, Massachusetts, USA).

Koncentrace takto získané knihovny byla změřena qPCR za použití *KAPA Sybr Fast qPCR Master Mix* (KAPA Biosystems, Boston, Massachusetts, USA) kitu, pomocí přístroje LightCycler 480 II (Roche Diagnostics, Mannheim, Germany) analýzou *fitting points*, za použití standardů v rozmezí 0.02 pM - 20 pM obsažených v *KAPA* kitu.

Následně byla knihovna naředěna na koncentraci 2 nM, denaturována NaOH a smíchána s hybridizačním pufrem. Knihovna o finální koncentraci 9 pM byla následně nanášena na kazetu s reagensii pro Illumina MiSeq (Illumina, San Diego, CA, USA) a sekvenována (Tokic *et al.*, 2020; Sikorova *et al.*, 2022). Výstupní data ve fastq formátu byly analyzovány pomocí *Omixon Twin software* v4.1.0. – v4.4.1. a *IPD-IMGT/HLA database Release 3.34-3.45*.

### 3.6 Statistické zpracování dat

Při analýze výsledku *MassArray* bylo nejprve zjištěno, zda výsledky, zejména u kontrolních populací, odpovídají Hardy-Weinbergově rovnováze  $p > 0.05$ .

Hardy-Weinbergova rovnováha (HWE) je nulový model vztahu mezi frekvencemi alel a genotypů v rámci generací i mezi nimi, za předpokladu, že nedochází k mutacím, migraci, selekci, náhodnému páření a nekonečné velikosti populace. Pokud se frekvence genotypů nebo alel významně odchyľují od HWE, může to znamenat systematické chyby v genotypizaci, neočekávanou strukturu populace či přítomnost homologních oblastí v genomu.

Pro jednotlivé SNP byl vypočítán *call rate*, která je definována jako podíl jedinců ve studii, u nichž nechybí informace o příslušném SNP.

Následně byly stanoveny frekvence minoritních alel (MAF) u jednotlivých SNP. Frekvence minoritních alel je frekvence, s níž se v dané populaci vyskytuje druhá nejčastější alela.

$$MAF = \frac{\text{alely pozitivní pro danou variantu}}{\text{všechny identifikované alely}}$$

Asociace polymorfismů se sledovaným znakem byla určována na základě alelického modelu a oboustranného Fisherova testu. Fisherův přesný test je test statistické významnosti používaný při analýze kontingenčních tabulek. Je platný pro všechny velikosti vzorků a v praxi se používá při malých velikostech souborů. Je jedním ze třídy exaktních testů, které

se tak nazývají proto, že významnost odchylky od nulové hypotézy (např.  $p$ -hodnota) může být vypočtena přesně, není tedy závislá na aproximaci. Analýza SNP dat probíhala především pomocí programu SHEsisPlus (bio-x.cn)(10.10.2022) (Shi *et al.*, 2005) a *2x2 Contingency Table* (graphpad.com; vassarstats.net) (9.9.2022);  $p < 0.05$ .

Bylo stanoveno *odds ratio* (OR) a 95% interval spolehlivosti (CI-*confidence interval*). OR je míra souvislosti mezi expozicí a výsledkem, vyjadřuje pravděpodobnost výskytu výsledku při určité expozici v porovnání s pravděpodobností výskytu výsledku bez této expozice.

OR = 1 genotyp neovlivňuje pravděpodobnost následku

OR > 1 genotyp je spojen s vyšší pravděpodobností následku

OR < 1 genotyp spojen s nižší pravděpodobností následku

K odhadu přesnosti OR se používá 95% CI. Velký CI znamená nízkou úroveň přesnosti OR, zatímco malý CI znamená vyšší přesnost OR.

Následně byla provedena korekce pro mnohočetná srovnání pomocí Bonferroniho korekce. Bonferroniho korekce je úprava hodnot  $p$  a používá se ke snížení pravděpodobnosti získání falešně pozitivních výsledků v případě, že se na jednom souboru dat provádí několik závislých nebo nezávislých statistických testů současně.

$$p_{corr} = 1 - (1 - p)^n$$

$n$ -počet zkoumaných SNP v daném plexu

$p$ -původní  $p$ -hodnota před korekcí

Pro analýzu HLA byly použity podobné statistické testy jako v případě analýzy SNP. Byla také stanovena  $p$ -hodnota na základě Fisherova testu, OR a 95% CI. Dále byla stanovena vazebná nerovnováha (LD *linkage disequilibrium*).

Pro analýzy byl v tomto případě použit především software “*HLA analysis tools*” ([https://www.hiv.lanl.gov/content/immunology/hla/hla\\_compare.html](https://www.hiv.lanl.gov/content/immunology/hla/hla_compare.html)) (10.10.2022) a *2x2 Contingency Table* (vassarstats.net) (9.9.2022);  $p < 0.05$ .

Následně byla provedena korekce pro mnohočetná srovnání pomocí Bonferroniho korekce (viz výše).



#### 4.4 HLA analýza v korejské populaci

Bylo analyzováno 6 HLA lokusů (*HLA-A*, *-B*, *-C*, *-DRB1*, *-DQA1*, *-DQB1*) u 103 pacientů se sarkoidózou z Korejské republiky. Frekvence alel HLA I. a II. třídy u pacientů byly porovnány s údaji o frekvenci alel z kontrolní korejské populace z databáze (HLA I. třídy  $n = 485$ ; HLA II. třídy  $n = 324$ ). V lokusu *HLA-A* bylo genotypizací HLA stanoveno 21 různých variant (Přílohy Tab. 21). Jedna z těchto variant, *HLA-A\*33:03*, se vyskytovala častěji u kontrol než u pacientů, což ukazuje na protektivní asociaci, která byla signifikantní na primární úrovni (Tab. 30). V lokusu *HLA-B* bylo rozlišeno 45 různých variant (Přílohy Tab. 21) a dvě z nich, *HLA-B\*15:01* a *HLA-B\*58:01*, byly na primární úrovni asociovány jako protektivní pro sarkoidózu (Tab. 30). Dále byla nalezena *HLA-B\*15:07* jako riziková varianta pro sarkoidózu signifikantní i po korekci pro mnohočetná srovnání ( $p_{corr} = 0.041$ ). V lokusu *HLA-C* bylo rozlišeno 22 různých variant. Varianta *HLA-C\*05:01* byla spojena se sarkoidózou na primární úrovni jako riziková varianta. Ochranná varianta *HLA-C\*03:02* ( $p_{corr} = 1.84 \cdot 10^{-04}$ ; OR: 0.163 (95 % CI: 0.06-0.449)) a rizikové varianty *HLA-C\*03:04* ( $p_{corr} = 0.007$ ; OR: 2.949 (95 % CI: 1.703-5.105)) a *HLA-C\*08:22* ( $p_{corr} = 0.02$ ) byly asociovány i po korekci na mnohočetná srovnání (Tab. 30).

**Tabulka 30: Souhrn variant HLA I. třídy asociovaných s pacienty se sarkoidózou ( $n = 103$ ) při porovnání s kontrolní populací ( $n = 485$ ) v korejské populaci (jsou zobrazeny jen alely asociovány alespoň na primární úrovni) (OR-odds ratio, CI-confidence interval,  $p_{corr}$ - $p$ -hodnota po korekci pro mnohočetná srovnání (signifikantní zvýrazněna tučně), NA-neanalyzováno)**

HLA	OR (95% CI)	$p$	$p_{corr}$
A*33:03	0.492 (0.294-0.821)	0.005	0.100
B*15:01	0.389 (0.193-0.783)	0.005	0.218
<b>B*15:07</b>	<b>NA</b>	<b>0.001</b>	<b>0.041</b>
B*58:01	0.285 (0.102-0.791)	0.008	0.290
<b>C*03:02</b>	<b>0.163 (0.060-0.449)</b>	<b><math>8.38 \cdot 10^{-06}</math></b>	<b><math>1.84 \cdot 10^{-04}</math></b>
<b>C*03:04</b>	<b>2.949 (1.703-5.105)</b>	<b><math>3.16 \cdot 10^{-04}</math></b>	<b>0.007</b>
C*05:01	2.970 (1.061-8.314)	0.046	0.643
<b>C*08:22</b>	<b>NA</b>	<b>0.001</b>	<b>0.020</b>

*HLA-DRB1* je HLA lokus, který je nejčastěji zmiňován v souvislosti se sarkoidózou; v tomto souboru pacientů bylo genotypizací HLA stanoveno 41 různých variant v tomto

lokusu. Varianty HLA-DRB1\*04:06, HLA-DRB1\*13:02, HLA-DRB1\*14:01 byly protektivní a varianta HLA-DRB1\*14:03 působila jako riziková pro sarkoidózu, přičemž asociace byly signifikantní pouze na primární úrovni. Po korekci na mnohočetná srovnání byly varianty HLA-DRB1\*14:54 ( $p_{corr} = 2.31 \cdot 10^{-05}$ ) a HLA-DRB1\*12:01 ( $p_{corr} = 9.12 \cdot 10^{-05}$ ; OR: 3.754 (95 % CI: 2.204–6.394)) identifikovány jako rizikové faktory pro sarkoidózu (Tab. 31).

Pokud jde o lokus *HLA-DQA1*, z genotypizace HLA vyplynulo 17 variant. Varianty HLA-DQA1\*01:01, HLA-DQA1\*03:01, HLA-DQA1\*05:01 a HLA-DQA1\*05:03 byly na primární úrovni určeny jako protektivní pro sarkoidózu. Varianta HLA-DQA1\*05:06 byla na primární úrovni spojena s rizikem vzniku sarkoidózy. Po korekci na mnohočetná srovnání byla HLA-DQA1\*01:02 ( $p_{corr} = 3.64 \cdot 10^{-05}$ ; OR: 0.288 (95 % CI: 0.162–0.512)) shledána jako protektivní a HLA-DQA1\*05:07 ( $p_{corr} = 0.003$ ) a HLA-DQA1\*05:08 ( $p_{corr} = 4.11 \cdot 10^{-05}$ ) jako rizikový faktor pro vznik sarkoidózy (Tab. 31).

V lokusu *HLA-DQB1* bylo stanoveno 19 variant. Varianty HLA-DQB1\*03:03, HLA-DQB1\*04:01 a HLA-DQB1\*04:02 byly na primární úrovni asociovány jako protektivní a varianta HLA-DQB1\*05:03 jako rizikový faktor vzniku sarkoidózy. Po korekci pro mnohočetná srovnání byly jako protektivní pro sarkoidózu zjištěny varianty HLA-DQB1\*03:02 ( $p_{corr} = 2.25 \cdot 10^{-05}$ ; OR: 0.086 (95 % CI: 0.021–0.356)) a HLA-DQB1\*06:05 ( $p_{corr} = 0.029$ ), zatímco rizikovou variantou byla HLA-DQB1\*06:09 ( $p_{corr} = 0.015$ ); viz Tab. 31.

**Tabulka 31: Souhrn variant HLA II. třídy asociovaných s pacienty se sarkoidózou ( $n = 103$ ) při porovnání s kontrolní populací ( $n = 324$ ) v korejské populaci (jsou zobrazeny jen alely asociovány alespoň na primární úrovni) (OR-odds ratio, CI-confidence interval,  $p_{corr}$ - $p$ -hodnota po korekci pro mnohočetná srovnání (signifikantní zvýrazněna tučně), NA-neanalyzováno)**

HLA	OR (95% CI)	$p$	$p_{corr}$
DRB1*04:06	0.157 (0.037–0.654)	0.002	0.080
<b>DRB1*12:01</b>	<b>3.754 (2.204–6.394)</b>	<b>2.22*10<sup>-06</sup></b>	<b>9.12*10<sup>-05</sup></b>
DRB1*13:02	0.445 (0.224–0.881)	0.017	0.506
DRB1*14:03	4.955 (1.371–17.904)	0.016	0.483
<b>DRB1*14:54</b>	<b>NA</b>	<b>5.63*10<sup>-07</sup></b>	<b>2.31*10<sup>-05</sup></b>
DRB1*14:01	NA	0.001	0.056
DQA1*01:01	0.365 (0.154–0.865)	0.015	0.223
<b>DQA1*01:02</b>	<b>0.288 (0.162–0.512)</b>	<b>2.14*10<sup>-06</sup></b>	<b>3.64*10<sup>-05</sup></b>

HLA	OR (95% CI)	<i>p</i>	<i>p<sub>corr</sub></i>
DQA1*03:01	0.425 (0.207–0.870)	0.014	0.217
DQA1*05:01	0.152 (0.021–142)	0.037	0.474
DQA1*05:06	NA	0.014	0.211
<b>DQA1*05:07</b>	<b>NA</b>	<b>1.86*10<sup>-04</sup></b>	<b>0.003</b>
<b>DQA1*05:08</b>	<b>NA</b>	<b>2.42*10<sup>-06</sup></b>	<b>4.11*10<sup>-05</sup></b>
DQA1*05:03	NA	0.010	0.161
<b>DQB1*03:02</b>	<b>0.086 (0.021–0.356)</b>	<b>1.18*10<sup>-06</sup></b>	<b>2.25*10<sup>-05</sup></b>
DQB1*03:03	0.481 (0.249–0.928)	0.027	0.404
DQB1*04:01	0.360 (0.152–0.851)	0.015	0.244
DQB1*04:02	0.120 (0.016–0.891)	0.010	0.169
DQB1*05:03	2.278 (1.130–4.592)	0.024	0.369
<b>DQB1*06:09</b>	<b>NA</b>	<b>0.001</b>	<b>0.015</b>
<b>DQB1*06:05</b>	<b>NA</b>	<b>0.002</b>	<b>0.029</b>

Byla také vypočtena LD mezi variantami asociovanými se sarkoidózou u korejských pacientů. Ve vazbě byly nalezeny varianty HLA II. třídy, HLA-DQA1\*05:08-HLA-DRB1\*12:01 ( $p = 3.33*10^{-06}$ ) a HLA-DQA1\*05:07-HLA-DRB1\*14:03 ( $p = 6.21*10^{-10}$ ). V rámci HLA I. třídy byly nalezeny ve vazbě varianty HLA-B\*58:01-HLA-C\*03:02 ( $p = 2.09*10^{-07}$ ) (Tab. 32).

**Tabulka 32: Vazebná nerovnováha alel nalezených jako asociovaných se sarkoidózou v korejské populaci** (počet provedených dvoucestných testů  $N_2 = 9636$ ,  $p$  hodnota pro dvoucestné srovnání  $0.05/N_2 = 5.189*10^{-06}$  pro 95% CI)

HLA1	HLA2	<i>p</i>
B*58:01	C*03:02	2.09*10 <sup>-07</sup>
DQA1*05:07	DRB1*14:03	6.21*10 <sup>-10</sup>
DQA1*05:08	DRB1*12:01	3.33*10 <sup>-06</sup>

#### 4.4.1 Asociace HLA variant s klinickými fenotypy sarkoidózy

##### 4.4.1.1 Srovnání distribuce variant HLA v kontextu různých stadií na základě rtg. nálezu

Pacienti byli rozděleni do dvou skupin – první skupina zahrnovala 30 pacientů se stadii 0-1 a druhá skupina zahrnovala 73 pacientů s pokročilejšími stadii 2-4. Frekvence

HLA zjištěné v těchto skupinách byly porovnány s frekvencemi v kontrolní populaci. Na primární úrovni bylo 8 variant asociovaných se stadii 0-1, a 2 z nich zůstaly asociovány i po korekci pro mnohočetná srovnání (HLA-DRB1\*12:01  $p_{corr} = 0.033$ ; HLA-DQA1\*05:08,  $p_{corr} = 0.001$ ) (Přílohy Tab. 22). Srovnání stadií 2-4 s kontrolními jedinci vedlo k 28 variantám signifikantně asociovaným s těmito stadii na primární úrovni. Po korekci pro mnohočetná srovnání zůstalo 8 variant asociovaných s pokročilejšími formami onemocnění (HLA-B\*15:07  $p_{corr} = 0.013$ , HLA-C\*03:04  $p_{corr} = 0.007$ , HLA-C\*08:22  $p_{corr} = 0.006$ , HLA-DRB1\*12:01  $p_{corr} = 0.003$ , HLA-DRB1\*14:54  $p_{corr} = 4.57 \cdot 10^{-05}$ , HLA-DQA1\*05:07  $p_{corr} = 0.001$ , HLA-DQA1\*05:08  $p_{corr} = 0.003$ , HLA-DQB1\*06:09  $p_{corr} = 0.004$ ) a 4 byly považovány za protektivní (HLA-A\*33:03  $p_{corr} = 0.038$ , HLA-C\*03:02  $p_{corr} = 0.005$ , HLA-DQA1\*01:02  $p_{corr} = 0.001$ , HLA-DQB1\*03:02  $p_{corr} = 3.86 \cdot 10^{-04}$ ) (Přílohy Tab. 23). Vzájemné porovnání pacientů ve stadiích 0-1 ( $n = 30$ ) a pacientů ve stadiích 2-4 ( $n = 73$ ) ukázalo, že alela HLA-B\*51:01 je na primární úrovni spojena se stadii 2-4 ( $p = 0.042$ ); po korekci pro mnohočetná srovnání však nebyla signifikantní.

#### 4.4.1.2 Srovnání HLA variant u pacientů s mimoplicní manifestací sarkoidózy.

Distribuce polymorfismů HLA u pacientů s mimoplicní manifestací sarkoidózy byla porovnána s údaji kontrolních jedinců. Při tomto srovnání bylo 23 variant asociováno s mimoplicní manifestací na primární úrovni (Přílohy Tab. 24). Z nich 9 bylo signifikantních i po korekci na mnohočetná srovnání, 4 jako protektivní vůči mimoplicní manifestaci (HLA-A\*33:03  $p_{corr} = 0.038$ , HLA-C\*03:02  $p_{corr} = 0.019$ , HLA-DQA1\*01:02  $p_{corr} = 0.003$ , HLA-DQB1\*03:02  $p_{corr} = 3.86 \cdot 10^{-04}$ ) a 5 bylo rizikových pro vznik mimoplicních projevů (HLA-B\*15:07  $p_{corr} = 0.013$ , HLA-C\*08:22  $p_{corr} = 0.047$ , HLA-DRB1\*14:54  $p_{corr} = 0.008$ , HLA-DQA1\*05:08  $p_{corr} = 0.003$ , HLA-DQB1\*06:09  $p_{corr} = 0.004$ ) (Přílohy Tab. 24). Při vzájemném porovnání frekvencí HLA u pacientů s mimoplicním postižením a bez něj byly na primární úrovni odhaleny čtyři asociace (Přílohy Tab. 25), které však po korekci nedosáhly signifikance.

#### 4.4.1.3 Varianty HLA v kontextu klinického průběhu onemocnění (2 roky sledování).

Pacienti byli sledováni nejméně 2 roky po stanovení diagnózy. Byli rozděleni do tří skupin v závislosti na průběhu onemocnění takto: pacienti, jejichž onemocnění přešlo do remise ( $n = 58$ ), pacienti se stabilním onemocněním a pacienti ( $n = 32$ ), jejichž onemocnění progredovalo do závažnějšího stavu ( $n = 5$ ). Dále byla distribuce HLA v těchto třech skupinách porovnána s údaji kontrolních jedinců. S progresivním, zhoršujícím

se průběhem byla spojena 1 varianta pouze na primární úrovni (Přílohy Tab. 26). Důležité je, že 32 variant bylo asociováno s remisí onemocnění; 8 z nich zůstalo signifikantních i po korekci pro mnohočetná srovnání (Přílohy Tab. 27). Z těchto 8 asociovaných variant jich bylo 5 asociováno s pacienty v remisi (HLA-DRB1\*12:01  $p_{corr} = 0.006$ , HLA-DRB1\*14:54  $p_{corr} = 0.003$ , HLA-DQA1\*05:07  $p_{corr} = 1.86 \cdot 10^{-04}$ , HLA-DQA1\*05:08  $p_{corr} = 0.001$ , HLA-DQB1\*06:09  $p_{corr} = 0.010$ ); zbývající 3 varianty byly častější v kontrolní skupině populace. Ve skupině pacientů se stabilním onemocněním bylo na primární úrovni asociováno 12 variant a 2 z nich dosáhly signifikance po korekci pro mnohočetná srovnání (Přílohy Tab. 28). Tyto varianty byly spojeny s pacienty ve stabilizovaném stavu (HLA-DRB1\*08:03  $p_{corr} = 0.002$  a HLA-DRB1\*14:54  $p_{corr} = 2.08 \cdot 10^{-04}$ ).

[REDACTED]

## 4.6 SNP analýza v řecké populaci

V řecké populaci bylo pomocí metody *MassArray* analyzováno 20 SNP u pacientů se sarkoidózou ( $n = 103$ ) a kontrolních jedinců ( $n = 100$ ). Výsledky pro polymorfismus rs3819717 v genu *TAP2* neodpovídaly HWE, a to především ve skupině kontrolních subjektů. Proto byly tyto polymorfismy vyloučeny z hodnocení výsledků u řecké populace.

Analýzou byla identifikována asociace polymorfismu rs3917200 v genu pro *TGF-β3* jako riziková pro vznik sarkoidózy, a to i po korekci pro mnohočetná srovnání ( $p_{corr} = 4.97 \cdot 10^{-6}$ ; OR: 3.04 (95% CI: 1.98-4.69)). Na primární úrovni byl jako asociovaný se sarkoidózou určen i polymorfismus rs1049550 v genu *ANXA11* ( $p = 0.013$ ; OR: 0.59 (95% CI: 0.39-0.89)) jako ochranný faktor pro vznik sarkoidózy (Tab. 35, Obr. 4).

**Tabulka 35: Analýza asociací SNP se sarkoidózou na alelické úrovni pro srovnání pacientů ( $n = 103$ ) a kontrolních jedinců ( $n = 100$ ) v řecké populaci (OR-odds ratio, CI-confidence interval,  $p_{corr}$ - $p$ -hodnota po korekci pro mnohočetná srovnání (signifikantní výsledky zvýrazněny tučně), kurzíva-výsledky, které nespĺňovaly kontrolu kvality)**

Gen	SNP	OR (95% CI)	$p$	$p_{corr}$
<i>ANXA11</i>	rs1049550	0.590 (0.391-0.888)	<b>0.013</b>	0.210
<i>TNF-α</i>	rs1800629	1.987 (0.985-4.006)	0.060	0.672
<i>TGF-β2</i>	rs1891467	1.003 (0.657-1.532)	1.000	1.000
<i>GREM1</i>	rs1919364	1.041 (0.703-1.541)	0.842	1.000
<i>HLA-DRA, HLA-</i>	rs1964995	0.642 (0.344-1.197)	0.209	0.985
<i>NOD2</i>	rs2066844	1.476 (0.515-4.229)	0.600	1.000
<i>BTNL2</i>	rs2076530	0.977 (0.645-1.479)	0.916	1.000
<i>HLA-DQA1</i>	rs2187668	0.968 (0.470-1.993)	1.000	1.000
<i>CACFD1</i>	rs3124765	NA	0.142	0.937
<i>C6orf10</i>	rs3129927	3.000 (0.800-11.248)	0.140	0.934
<i>ATF6B</i>	rs3130288	2.680 (0.700-10.254)	0.220	0.989
<i>HLA-DRA</i>	rs3135394	3.000 (0.800-11.248)	0.140	0.934
<i>TAP2</i>	rs3819717	1.100 (0.697-1.737)	0.727	1.000
<i>HLA-DRB1</i>	rs3830135	0.707 (0.291-1.718)	0.505	1.000
<i>TGF-β3</i>	rs3917165	2.308 (0.868-6.136)	0.108	0.872
<b><i>TGF-β3</i></b>	<b>rs3917200</b>	<b>3.043 (1.976-4.686)</b>	<b>2.76*10<sup>-07</sup></b>	<b>4.97*10<sup>-06</sup></b>
<i>HLA-DPB1</i>	rs9277357	1.040 (0.662-1.632)	0.908	1.000
<i>LRRC16A</i>	rs9295661	1.447 (0.239-8.759)	0.999	1.000
<i>HLA-DRB1*03:01</i>	rs2040410	1.174 (0.657-2.097)	0.658	0.883
<i>HLA-DRB1*15:01</i>	rs3135388	0.606 (0.256-1.435)	0.287	0.492

#### 4.6.1 Asociace SNP s klinickými fenotypy sarkoidózy

Na základě hodnocení stadia podle rtg. plic byli pacienti rozděleni do dvou skupin, na pacienty ve stadiu 1 ( $n = 29$ ) a na pacienty ve stadiu 2 a vyšším ( $n = 68$ ). Při vzájemném porovnání těchto dvou skupin nebyla nalezena asociace s žádným z analyzovaných polymorfismů (Přílohy Tab. 33). Tyto dvě skupiny byly porovnávány i s kontrolními jedinci. V případě polymorfismu rs3917200 v genu *TGF-β3* byla nalezena asociace signifikantní i po korekci pro mnohočetná srovnání, a to jak při porovnání pacientů ve stadiu 1 s kontrolními jedinci ( $p_{corr} = 3.13 \cdot 10^{-03}$ ), tak u pacientů ve stadiu 2 a výše s kontrolními jedinci ( $p_{corr} = 1.71 \cdot 10^{-04}$ ).

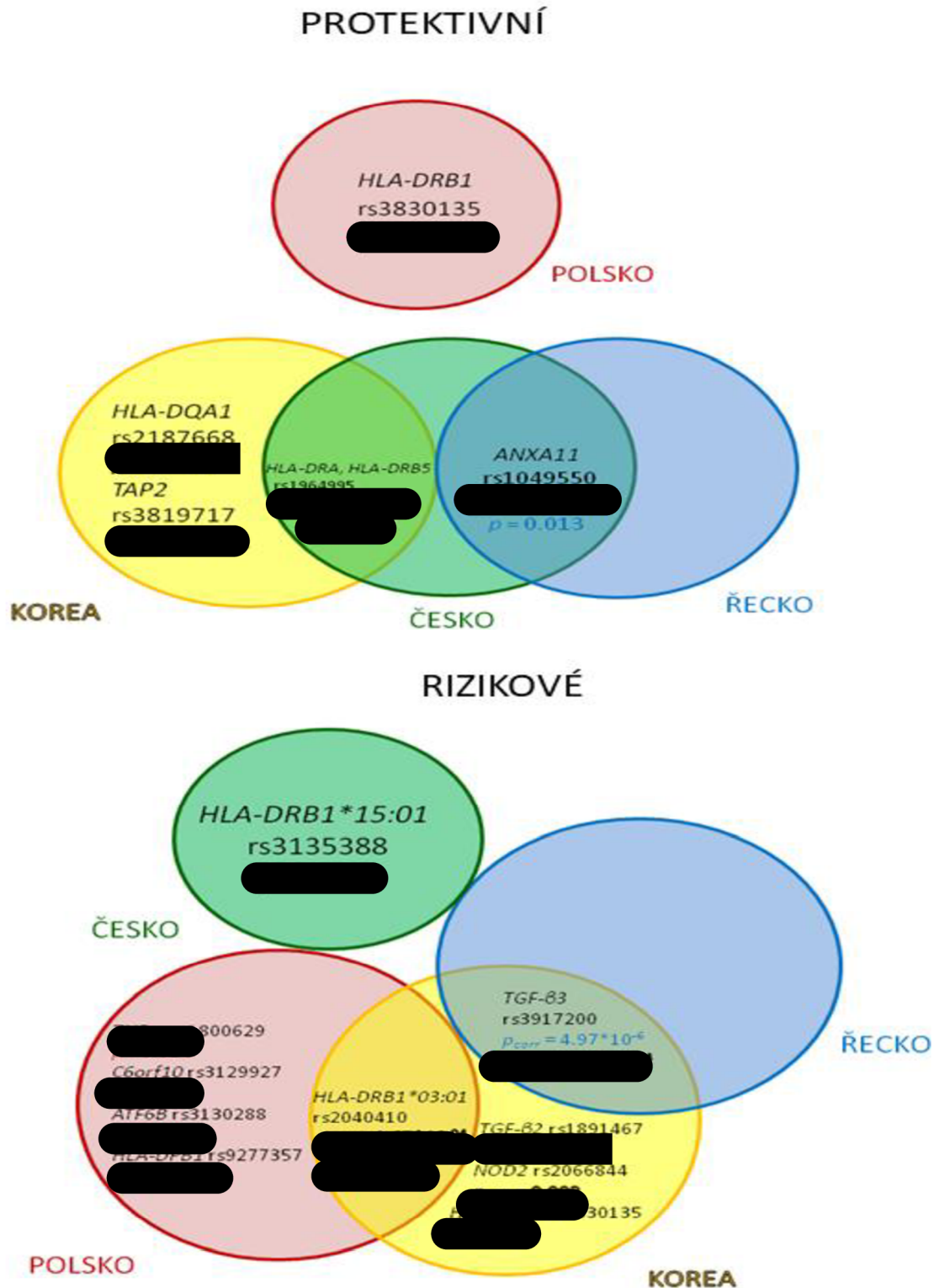
Dále byl nalezen polymorfismus rs3917165 v genu *TGF-β3* asociován s pacienty ve stadiu 1 sarkoidózy v porovnání s kontrolními jedinci na primární úrovni ( $p = 0.005$ ) (Přílohy Tab. 34, Obr. 5).

Jako protektivní byl detekován polymorfismus rs1049550 v genu *ANXA11*, který se signifikantně více vyskytoval u kontrolních jedinců, než u pacientů ve stadiu 2 a vyšším ( $p = 0.007$ ) (Přílohy Tab. 35).

U pacientů ve stadiu 2 a výše byl jako asociovaný nalezen také polymorfismus rs1800629 v genu *TNF-α* ( $p = 0.015$ ) (Přílohy Tab. 35, Obr. 6).

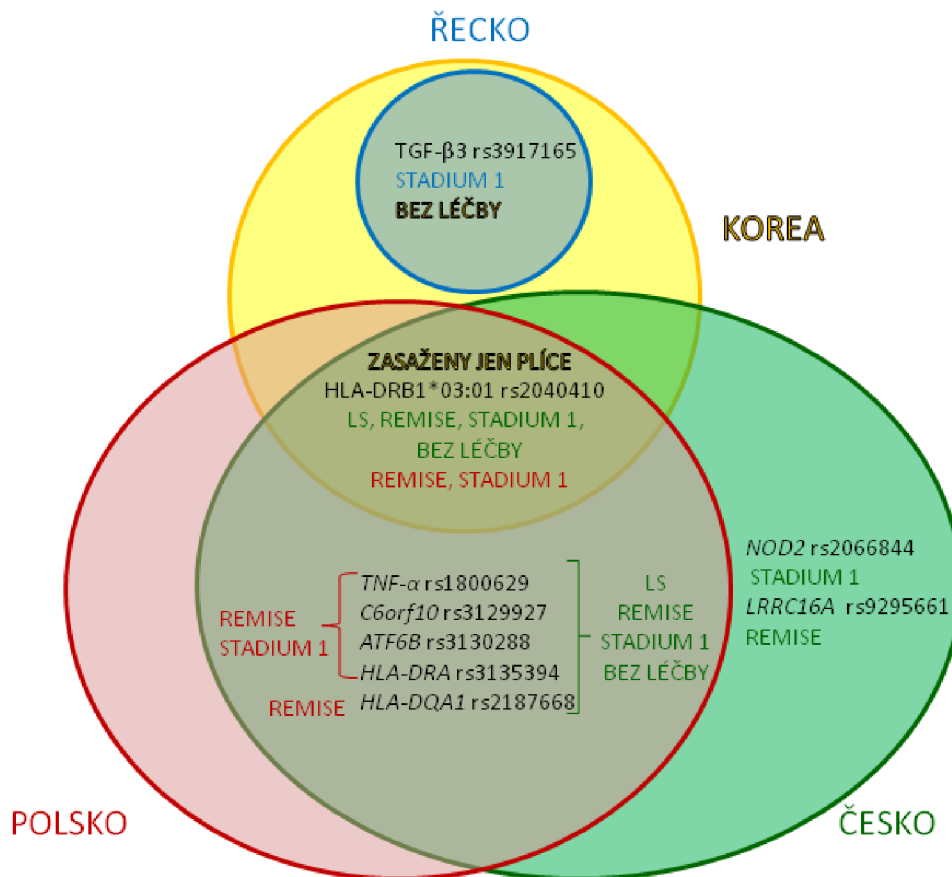
Porovnáním pacientů s mimoplicními projevy sarkoidózy s pacienty, u kterých sarkoidóza zasáhla jen plíce, nebyl identifikován žádný polymorfismus spojený s jednou či druhou skupinou. Stejně tomu bylo i u porovnání pacientů, kteří podstupovali léčbu, s pacienty, kteří nebyli indikováni k léčbě.

**Obrázek 4: Varianty identifikované jako protektivní a rizikové při vzniku sarkoidózy v české, polské, řecké a korejské populaci** zelené písmo – česká populace, červené písmo-polská populace, modré písmo – řecká populace, žluté písmo s černým orámováním-korejská populace,  $p$  – varianta signifikantní jen na primární úrovni,  $p_{corr}$ -varianta signifikantní po korekci pro mnohočetná srovnání.

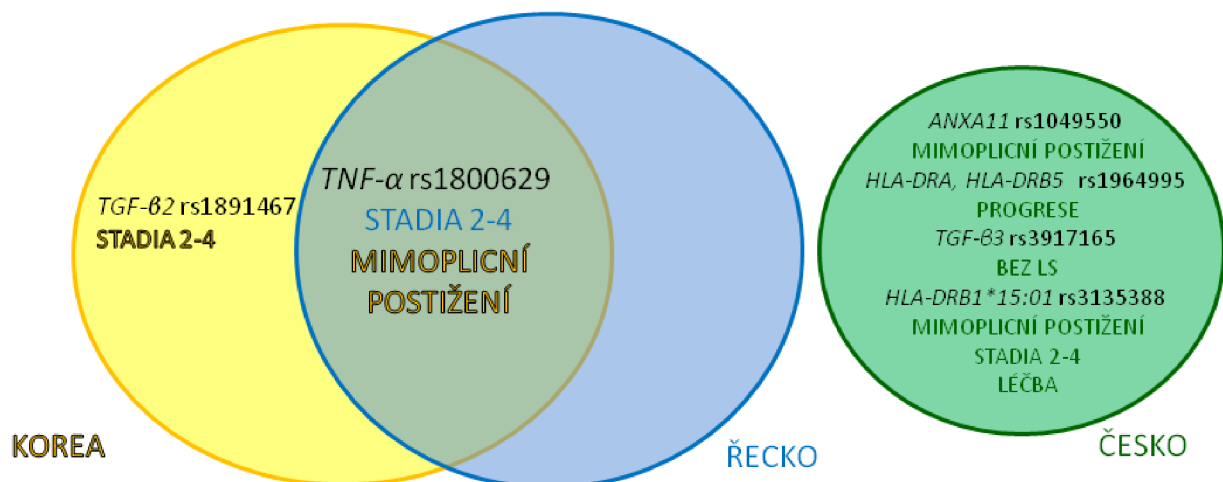




**Obrázek 5: Varianty spojené s příznivým průběhem sarkoidózy v české, polské, řecké a korejské populaci** zelené písmo – česká populace, červené písmo – polská populace, modré písmo – řecká populace, žluté písmo s černým orámováním – korejská populace



**Obrázek 6: Varianty spojené se závažnějším průběhem sarkoidózy v české, řecké a korejské populaci** zelené písmo – česká populace, modré písmo – řecká populace, žluté písmo s černým orámováním – korejská populace



#### 4.6.2 Analýza HLA u chorvatských zdravých jedinců

Bylo hodnoceno 120 jedinců chorvatské zdravé populace v šesti HLA lokusech (*HLA-A*, *-B*, *-C*, *-DRB1*, *-DQA1*, *-DQB1*) s rozlišením na 4 pole. Z důvodu nízké kvality bylo z hodnocení vyloučeno 9 vzorků.

V lokusu *HLA-A* bylo identifikováno 25 různých variant. Nejčastěji byl pozorován výskyt varianty *HLA-A\*02:01:01* (32.43 %), druhou nejčastější byla varianta *HLA-A\*01:01:01:01* (12.16 %) následována variantami *HLA-A\*03:01:01:01* (11.71 %), *HLA-A\*24:02:01:01* (11.71 %) a *HLA-A\*11:01:01:01* (9.91 %). Nejčastěji se vyskytující varianta *HLA-A\*02:01:01* byla přítomna v analyzované populaci u 59 jedinců, u 13 z nich v homozygotní formě (Přílohy Tab. 36).

Největší počet rozlišených variant v rámci HLA I. třídy byl v lokusu *HLA-B*, kde bylo rozlišeno 39 různých variant. Nejčastěji se vyskytovala varianta *HLA-B\*51:01:01* (8.56 %) následovaná variantami *HLA-B\*35:01:01* (8.11 %), *HLA-B\*08:01:01* (8.11 %), *HLA-B\*07:02:01* (6.76 %), *HLA-B\*35:03:01* (6.76 %) a *HLA-B\*18:01:01* (5.41 %) (Přílohy Tab. 36).

V lokusu *HLA-C* bylo typizací zjištěno 26 různých HLA variant. Nejčastěji se v této populaci vyskytovala varianta *HLA-C\*04:01:01:01* (13.96 %) následovaná *HLA-C\*02:02:02* (9.46 %), *HLA-C\*12:03:01* (7.21 %) a *HLA-C\*03:03:01* (5.41 %). Nejrozličnější skupinou byla *HLA-C\*07*, která byla zastoupena v populaci v 27.48 % a skládala se z 6 variant, nejčastější z nich byly *HLA-C\*07:01:01G* (14.41 %) a *HLA-C\*07:02:01G* (8.11 %) (Přílohy Tab. 36).

Největší variabilita v rámci HLA II. třídy byla pozorována v rámci lokusu *HLA-DRB1*, kde bylo identifikováno 40 variant. Nejčastější variantou zde byla *HLA-DRB1\*16:01:01* (13.51 %) následována *HLA-DRB1\*01:01:01* (10.81 %), *HLA-DRB1\*07:01:01* (9.91 %), *HLA-DRB1\*15:01:01* (6.76 %), *HLA-DRB1\*03:01:01:01* (5.41 %), *HLA-DRB1\*11:01:01:01* (4.5 %). Nejvyšší variabilita zde byla sledována ve skupině *DRB1\*04*, ve které bylo rozlišeno 9 variant (Přílohy Tab. 37).

Analýza lokusu *HLA-DQA1* odhalila 37 různých variant, z nichž nejčastější byla *HLA-DQA1\*01:02:02* (15.77 %). Následovaly varianty *HLA-DQA1\*03:01:01* (7.21 %) a *HLA-DQA1\*02:01:01:01* (6.76 %) (Přílohy Tab. 37).

V lokusu *HLA-DQB1* bylo zjištěno 25 různých variant. Nejčastěji se v této populaci vyskytovaly varianty *DQB1\*05:02:01:01* (15.77 %), *DQB1\*03:01:01:03* (14.86 %), *DQB1\*05:01:01:03* (10.36 %), *HLA-DQB1\*02:01:01* (7.66 %), *HLA-DQB1\*03:02:01:01* (7.21 %) a *HLA-DQB1\*02:02:01:01* (6.76 %) (Přílohy Tab. 37).

Byla provedena i haplotypová analýza. V rámci HLA I. třídy se čteněji ve zkoumané populaci vyskytoval jen ancestrální evropský haplotyp A\*01:01:01:01-C\*07:01:01-B\*08:01:01 (Přílohy Tab. 38).

Mezi HLA II. třídy byl nejčtenější haplotyp DRB1\*16:01:01-DQA1\*01:02:02-DQB1\*05:02:01:01 (13.51 %). Mezi další často se vyskytující haplotypy patřily 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 %) a DRB1\*01:01:01-DQA1\*01:01:01:01-DQB1\*05:01:01:03 (4.05 %) (Přílohy Tab. 38). Mezi haplotypy zahrnujícími všech 6 lokusů byl nejčastějším ancestrální haplotyp 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 %).

## 5 Diskuze

### 5.1 HLA varianty asociované se sarkoidózou v české populaci

Tato studie provedená na české populaci zkoumala souvislost HLA variant se sarkoidózou. Zaměřila se na souvislosti mezi HLA variantami a klinickými fenotypy sarkoidózy. Jedná se o první podrobnou analýzu HLA variant v české populaci v lokusech *HLA-A*, *-B*, *-C*, *-DRB1*, *-DQA1* a *-DQB1* s rozlišením na 4 pole. Nově byla objevena asociace mezi sarkoidózou a HLA-DQB1\*06:04 jako rizikovým faktorem pro sarkoidózu a asociace mezi sarkoidózou a HLA-DQA1\*03:01 jako protektivním faktorem v české populaci. Byla potvrzena také asociace mezi HLA-DQB1\*06:02 (riziková) a HLA-DRB1\*01:01, HLA-DQB1\*03:02 (protektivní), které byly dříve popsány v menších studiích u jiných populací (Sikorova *et al.*, 2022; Berlin *et al.*, 1997; Voorter *et al.*, 2005; Sato *et al.*, 2010a; Wennerstrom *et al.*, 2013). Co se týká klinických fenotypů sarkoidózy byly identifikovány varianty HLA-A\*01:01, HLA-B\*08:01, HLA-C\*07:01, HLA-DRB1\*03:01, HLA-DQA1\*05:01 a HLA-DQB1\*02:01 asociovány s LS, varianty HLA-DRB1\*03:01 a HLA-DQA1\*05:01 spojené s remisí onemocnění a stadiem 1 a varianty HLA-DRB1\*11:01 a HLA-DQA1\*05:05 asociovány s onemocněním ve stadiu 2-4. S mimoplicními projevy sarkoidózy byla asociovaná varianta HLA-DQB1\*05:03. Tato varianta byla již dříve nalezena v populaci Velké Británie, v níž byla spojena s výskytem uveitidy u pacientů se sarkoidózou (Sato *et al.*, 2010a).

U sarkoidózy, multisystémového granulomatózního onemocnění postihujícího především plíce, prezentují alveolární CD4<sup>+</sup> T buňky neznámý antigen interakcí s molekulami HLA II. třídy. Tato úloha HLA II. třídy může být jedním z důvodů, proč většina variant HLA asociovaných se sarkoidózou i u našich korejských pacientů patří do HLA II. třídy (Judson, 2019).

Při srovnání pacientů se sarkoidózou s kontrolní populací byla jako riziková varianta pro sarkoidózu nalezena HLA-DQB1\*06:02. Tato varianta byla v souvislosti se sarkoidózou zmiňována již dříve, především se závažnější prognózou u sarkoidózy (Sato *et al.*, 2002), a souvisela s variantami HLA-DRB1\*15:01 a HLA-B\*07 (Voorter *et al.*, 2005). HLA-DRB1\*15:01 a HLA-B\*07:02 byly rovněž pozorovány jako rizikové faktory sarkoidózy, ale pouze na primární úrovni. Tyto dvě varianty byly rovněž zaznamenány v LD s variantou HLA-DQB1\*06:02 (Tab. 21).

Nově byla identifikována varianta HLA-DQB1\*06:04 jako rizikový faktor pro rozvoj sarkoidózy, což je v rozporu s předchozí studií u populace z Nizozemí, kde nebyl zjištěn

významný rozdíl ve výskytu této varianty u pacientů ve srovnání s kontrolní populací (Voorter *et al.*, 2005). Tento rozdíl může odrážet rozdílnou velikost souboru a odlišný etnický původ vyšetřovaných pacientů (van Moorsel *et al.*, 2022).

Varianta HLA-DRB1\*01:01 byla dříve popsána jako protektivní faktor při vzniku sarkoidózy v populaci Spojeného království, Nizozemska, Finska a Japonska, kde byla identifikována jako součást haplotypu HLA-DQA1\*01:01 -DQB1\*05:01 -DRB1\*01:01. Všechny tyto varianty jsou podle zde prezentovaných údajů v LD a působí protektivně proti rozvoji sarkoidózy (Wennerstrom *et al.*, 2013; Grutters *et al.*, 2003; Naruse *et al.*, 2000). Kromě HLA-DRB1\*01:01, která se ve zde prezentované české kontrolní populaci vyskytovala častěji než u pacientů, byly dalšími dvěma protektivními variantami HLA-DQA1\*03:01 a HLA-DQB1\*03:02. Varianta HLA-DQB1\*03:02 byla nedávno popsána jako protektivní faktor vzniku sarkoidózy také u korejské populace (Sikorova *et al.*, 2022). HLA-DQA1\*03:01 nebyla dosud v souvislosti se sarkoidózou nikdy zmíněna. Tato pozorovaná asociace však může být výsledkem LD mezi HLA-DQA1\*03:01 a HLA-DQB1\*03:02 a další protektivní variantou – HLA-DRB1\*04:01, která je zde uváděna jako protektivní na primární úrovni. Tyto tři varianty jsou součástí jednoho z nejběžnějších haplotypů v kavkazské populaci (Kockum *et al.*, 1999). Další běžný haplotyp, HLA-DQB1\*06:02 -DQA1\*01:02, se v této studii vztahuje k riziku sarkoidózy.

Při porovnání klinických fenotypů sarkoidózy byla jako rizikový faktor pro pokročilejší stadia sarkoidózy stanovena varianta HLA-DRB1\*11:01. Tato varianta byla již dříve popsána ve vztahu k sarkoidóze obecně (Rossman *et al.*, 2003). V této práci byla varianta poprvé identifikována ve spojitosti s vyššími stadii onemocnění.

V této práci bylo zaznamenáno 6 HLA variant (HLA-A\*01:01, HLA-B\*08:01, HLA-C\*07:01, HLA-DRB1\*03:01, HLA-DQA1\*05:01 a HLA-DQB1\*02:01) tvořících 8.1 haplotyp, asociovaných s LS. Haplotyp 8.1 se nejčastěji vyskytuje v severní Evropě a je spojen s rizikem autoimunitních onemocnění (Gambino *et al.*, 2018). Byla popsána i jeho souvislost se sarkoidózou (Price *et al.*, 1999; Grunewald, 2010). Vazba variant HLA-B\*08, HLA-DRB1\*03:01 a HLA-DQB1\*02:01 s příznivou prognózou sarkoidózy byla dříve pozorována v chorvatské populaci (Grubic *et al.*, 2007). Příznivý průběh onemocnění byl dříve spojen také s variantou HLA-DQB1\*02:01 (Sato *et al.*, 2002; Grunewald *et al.*, 2004).

V této práci byla zjištěna spojitost variant HLA-DRB1\*03:01 a HLA-DQA1\*05:01 se stadiem 1 a remisí onemocnění, spolu s variantou HLA-A\*01:01 se častěji vyskytovaly u pacientů, u kterých nebyla indikována léčba kortikosteroidy. Většina asociačních studií provedených v souvislosti se sarkoidózou uvádí spojitost varianty HLA-DRB1\*03:01

se skupinou pacientů, zejména těch s příznivější prognózou (Mrazek *et al.*, 2005; Grunewald *et al.*, 2004; Yanardag *et al.*, 2017). Byla také popsána asociace varianty HLA-B\*08:01 s mírným průběhem onemocnění, zejména s LS (Grunewald *et al.*, 2004, Fischer *et al.*, 2015). V této práci byla popsána asociace varianty HLA-DQA1\*05:01 se skupinami s příznivější prognózou. Již dříve byl zaznamenán významně vyšší výskyt této varianty u pacientů než u kontrolní populace (Ishihara *et al.*, 1994). Také byla zaznamenána souvislost varianty HLA-A\*01:01 s pacienty v remisi onemocnění (Grunewald *et al.*, 2004).

S mimoplicním postižením byla v této práci asociovaná varianta HLA-DQB1\*05:03, dříve zmiňovaná v souvislosti se sarkoidózní uveitidou u pacientů ve Velké Británii (Sato *et al.*, 2010a). Dále byla HLA-DQB1\*05:03 spojena s persistentním onemocněním na primární úrovni, stejně jako HLA-DRB1\*14:54 a HLA-DQA1\*01:04. Tyto tři varianty se v české populaci vyskytují v LD. V indické populaci byly tyto varianty spojeny s rizikem sarkoidózy (Sharma *et al.*, 2003).

O vztahu mezi imunogenetickými markery a indikací k léčbě kortikosteroidy je zatím poměrně málo informací. V předchozí studii byl u českých pacientů se sarkoidózou indikovaných k léčbě kortikosteroidy zjištěn zvýšený výskyt mutantní alely CCR5Delta32 (Petrek *et al.*, 2000). Bylo zde pozorováno, že nositelé HLA-A\*02:01 potřebují léčbu dvakrát častěji než pacienti, kteří tuto variantu nenesou. Vzhledem k tomu, že léčba kortikosteroidy je indikována u pacientů s významným postižením plic sarkoidózou, u nichž se předpokládá vyšší riziko budoucí mortality nebo trvalé invalidity (Baughman *et al.*, 2021), může být užitečná dostupnost dalšího laboratorního markeru, který bude zahrnut do komplexního rozhodování o zahájení léčby. Před potenciálním klinickým využitím je však třeba toto zjištění replikovat. Na druhou stranu nebyla potřeba léčby spojena se třemi alelami haplotypu 8.1 a toto pozorování dobře zapadá do tímto uváděné souvislosti tohoto haplotypu s mírnější formou sarkoidózy.

V této části práce byla provedena typizace HLA lokusů pomocí metody NGS s vysokým rozlišením na 4 pole. Pro statistickou analýzu bylo použito rozlišení na dvě pole, protože je to nejpoužívanější formát mezi statistickými programy. Hlavní význam této části práce spočívá ve velkém počtu podrobně charakterizovaných pacientů z jednoho zdravotnického centra, což umožňuje porovnávat i adekvátní podskupiny na základě klinických diagnóz. Byla analyzována data 301 přesně charakterizovaných pacientů se sarkoidózou s reprezentativními fenotypy. Jediným omezením byl nízký počet pacientů ve 4. stadiu.

## 5.2 SNP asociované se sarkoidózou v české populaci

V české populaci bylo analyzováno 20 polymorfismů spojených především s imunitním systémem u 301 pacientů se sarkoidózou. Výsledky byly porovnány s výsledky genotypizace 203 kontrolních jedinců. Cílem této části práce byla replikace již dříve zjištěných asociací v české populaci na poměrně velkém souboru z konkrétního zdravotnického centra a vyhodnocení jejich přítomnosti u jednotlivých klinických fenotypů sarkoidózy.

Jako rizikový pro vznik sarkoidózy byl v této práci odhalen polymorfismus rs3135388, který je spojen s HLA variantou HLA-DRB1\*15:01 (Karakaya *et al.*, 2019). Tato varianta byla popsána již dříve asociovaná s rizikem vzniku onemocnění, a především s jeho progresí (Grutters *et al.*, 2003; Karakaya *et al.*, 2019; Grunewald, 2010). Polymorfismus rs3135388 byl v této práci identifikován i jako rizikový faktor pro vznik mimoplicních projevů onemocnění a spojený s pacienty indikovanými k léčbě kortikoidy, což potvrzují dřívější výsledky (Grutters *et al.*, 2003; Karakaya *et al.*, 2019; Grunewald, 2010).

Se závažnějším průběhem onemocnění byl v této práci asociován polymorfismus rs3917165 v genu pro *TGF-β3*, jehož minoritní alela A se častěji vyskytovala u pacientů, u kterých nebyl diagnostikován LS, v porovnání s pacienty s LS. Izoformy *TGF-β* jsou obecně spojeny s fibrotickými onemocněními. V holandské populaci byl tento polymorfismus asociován s pacienty, u kterých se sarkoidóza projevovala přítomností plicní fibrózy (Kruit *et al.*, 2006).

Polymorfismus rs1964995 spojený s genem *HLA-DRA*, *HLA-DRB5* byl dříve zmiňován především jako asociovaný s pacienty, u kterých nebyl diagnostikován LS, tedy obecně se závažnější prognózou onemocnění (Xiong *et al.*, 2021). V této práci byl polymorfismus rs1964995 nalezen asociovaný s pacienty, u kterých došlo v průběhu sledování k progresi onemocnění.

V rámci porovnání pacientů s mimoplicními projevy onemocnění a pacientů, u kterých onemocnění postihovalo pouze plíce, byl jako spojený s mimoplicní formou onemocnění identifikován i polymorfismus rs1049550 v genu *ANXA11*.

Jako protektivní proti vzniku sarkoidózy byl nalezen polymorfismus rs1964995 v genu *HLA-DRA*, *HLA-DRB5*, který se vyskytoval u kontrolní populace dvakrát častěji než u pacientů.

S kontrolní populací byl asociován i polymorfismus rs1049550 v genu *ANXA11*. Tento gen má vliv na tvorbu granulomu a bylo prokázáno, že má roli v udržování granulomatózního zánětu (Mirsaeidi *et al.*, 2016). Nesynonymní mutace rs1049550 nacházející se v exonu 6 genu *ANXA11* vede k záměně aminokyseliny z argininu na cytosin lokalizované na evolučně

konzervované pozici 230 první annexinové domény. Tyto sekvenční změny mohou způsobit narušení struktury proteinu nebo interferovat s proteinovými interakcemi (Hofmann *et al.*, 2008). Zvýšený výskyt rs1049550\*T varianty v kontrolní populaci při porovnání s pacienty se sarkoidózou byl nalezen v mnoha dřívějších studiích (Mrazek *et al.*, 2011; Karakaya *et al.*, 2022; Zhou *et al.*, 2016).

Další polymorfismy byly spojeny s příznivým průběhem onemocnění. Polymorfismus rs9295661 v genu *LRRC16A*, který je exprimován i v CD4<sup>+</sup> a CD8<sup>+</sup> T-lymfocytech (Kim *et al.*, 2014), byl dříve spojován s pacienty, u kterých byl diagnostikován LS (Rivera *et al.*, 2016). Výše zmíněná asociace byla potvrzena i v této práci.

Dále byla v této práci nalezena souvislost polymorfismu rs2066844, který je zodpovědný za nesynonymní mutaci v genu *NOD2/CARD15*, kdy dochází k záměně argininu za trypsin na pozici 702. Tato varianta byla již dříve identifikována jako spojená se sarkoidózou. Na rozdíl od této práce, kde byla identifikována asociovaná s pacienty ve stadiu 1, byla ovšem dříve nalezena spojená s vyššími stadii sarkoidózy a jejím závažnějším průběhem (Sato *et al.*, 2010b).

Ve spojitosti s příznivým průběhem sarkoidózy bylo v této práci identifikováno 6 polymorfismů v genech nacházejících se na chromozomu 6 (rs1800629 v genu *TNF-α*; rs2187668 v genu *HLA-DQA1*; rs3129927 v lokusu *C6orf10*; rs3130288 v genu *ATF6B*; rs315394 v genu *HLA-DRA* a rs2040410 spojený s variantou HLA-DRB1\*03:01). Tyto polymorfismy byly asociovány s výskytem LS, s pacienty v remisi onemocnění, s pacienty ve stadiu 1 a s pacienty, u kterých nebyla indikována léčba kortikosteroidy. Ve studii zabývající se autoimunitním onemocněním – celiakií, byla identifikována spojitost výskytu polymorfismu rs2187668 v genu *HLA-DQA1* s výskytem HLA variant HLA-DQA1\*05:01 a HLA-DQB1\*02:01 (Monsoon *et al.*, 2008). Tyto varianty jsou spolu s variantou HLA-DRB1\*03:01 (spjatou s polymorfismem rs2040410, který byl v této práci také analyzován) a genem *TNF-α* (ve kterém byl v této práci analyzován polymorfismus rs1800629) součástí 8.1 ancestrálního haplotypu. Ten byl v této práci spojen s příznivějším průběhem sarkoidózy i v analýze HLA variant pomocí NGS. Polymorfismus rs1800629 v genu *TNF-α* byl asociován s příznivější prognózou, konkrétně výskytem LS i v analýze provedené dříve u české populace (Mrazek *et al.*, 2005).

S příznivějším průběhem onemocnění byl spojen i polymorfismus rs3130288 v genu *ATF6B*. Protein kódovaný tímto genem je transkripčním faktorem v signální dráze odpovědi na stres ER (Park *et al.*, 2014). *ATF6B* je lokalizován v oblasti MHC na chromozomu 6p21 (Moudi *et al.*, 2020) a je to pravděpodobně jeho pozice v této oblasti, co určuje jeho spojitost



s geny asociovanými s lepším průběhem sarkoidózy. Polymorfismy rs3129927 v lokusu *C6orf10* a rs3135394 v genu *HLA-DRA* byly v této práci také asociovány s příznivějším průběhem onemocnění a nachází se v oblasti chromozomu 6. Tyto tři polymorfismy (rs3130288; rs3129927; rs3135394) byly již dříve nalezeny spojeny s pacienty s LS (Rivera *et al.*, 2016). Jejich spojitost s příznivým průběhem onemocnění a společný výskyt s polymorfismy v genech 8.1 ancestrálního haplotypu by z nich mohl činit potenciální kandidáty tohoto haplotypu. Polymorfismy ve výše zmíněných genech by do budoucna mohly sloužit jako markery příznivé prognózy u pacientů se sarkoidózou.

### 5.3 SNP asociované se sarkoidózou v polské populaci

V polské populaci bylo vyšetřováno 20 jednonukleotidových polymorfismů u 104 pacientů se sarkoidózou. Data byla porovnána s výsledky genotypizace 100 kontrolních jedinců. Výsledky v této populaci byly podobné těm získaným analýzou polymorfismů v české populaci, což odpovídá podobnému geografickému umístění i etnickému původu těchto dvou populací.

Ve spojitosti s příznivým průběhem sarkoidózy bylo v této práci identifikováno 6 polymorfismů v genech nacházejících se na chromozomu 6 (rs1800629 v genu *TNF- $\alpha$* ; rs2187668 v genu *HLA-DQA1*; rs3129927 v lokusu *C6orf10*; rs3130288 v genu *ATF6B*; rs315394 v genu *HLA-DRA* a rs2040410 spojený s variantou HLA-DRB1\*03:01). Polymorfismus rs2187668 v genu *HLA-DQA1* byl identifikován jako predispozice k remisi onemocnění. Jak je již výše zmíněno, tento polymorfismus je asociován s výskytem variant HLA-DQA1\*05:01 a HLA-DQB1\*02:01 (Monsuur *et al.*, 2008). Tyto dvě varianty byly spojeny s příznivějším průběhem sarkoidózy v polské populaci v dřívější studii (Dubaniewicz *et al.*, 2007). Stejně tak byla v této studii (Dubaniewicz *et al.*, 2007) jako riziková pro LS a pacienty ve stadiu 1, v porovnání s kontrolními jedinci, objevena varianta HLA-DRB1\*03:01. Tato varianta je asociovaná s polymorfismem rs2040410, který byl identifikován jako rizikový pro vznik sarkoidózy u polských pacientů a zároveň spojený s pacienty v remisi a pacienty ve stadiu 1. Stejně tak byly v této práci v souvislosti s remisí onemocnění a stadiem 1 identifikovány polymorfismy rs1800629 v genu *TNF- $\alpha$* ; rs2187668 v genu *HLA-DQA1*, rs3129927 v lokusu *C6orf10*; rs3130288 v genu *ATF6B* a rs315394 v genu *HLA-DRA*. Polymorfismy rs3129927 v lokusu *C6orf10* a rs3130288 v genu *ATF6B* byly zároveň v této práci identifikovány jako rizikové pro vznik sarkoidózy obecně. Jejich spojitost se sarkoidózou a jejím příznivějším průběhem byla popsána již dříve (Xiong *et al.*, 2021; Rivera *et al.*, 2016).

Polymorfismus rs1800629 v genu *TNF- $\alpha$*  byl v této práci u polské populace nalezen jako rizikový faktor pro vznik sarkoidózy. V rámci ní je ovšem spjat s remisí onemocnění a pacienti ve stadiu 1. Tato spojitost s příznivou prognózou onemocnění je v souladu s předchozími studiemi (Seitzer *et al.*, 1997; Mrazek *et al.*, 2005).

Polymorfismus rs9277357 v genu *HLA-DPB1* byl se sarkoidózou spojený již dříve (Rivera *et al.*, 2016). V této studii byla identifikována jeho funkce jako rizikového faktoru pro vznik sarkoidózy v polské populaci.

Jako protektivní před vznikem sarkoidózy byl u polské populace identifikovaný polymorfismus rs3830135 v genu *HLA-DRB1*. Asociován s kontrolními jedinci v porovnání s pacienti se sarkoidózou byl identifikován již dříve ve švédské populaci (Rivera *et al.*, 2016). Tento polymorfismus se v populaci českých pacientů se sarkoidózou vyskytoval vždy spolu s variantou HLA-DRB1\*01 (Tab. 27). Tato varianta byla nalezena jako protektivní před vznikem sarkoidózy i v polské a dalších populacích (Foley *et al.*, 2001).

Analýzou polymorfismů v polské populaci byly identifikovány možné markery vzniku sarkoidózy i jejího průběhu. Výsledky analýz provedených v polské populaci odpovídají výsledkům získaným analýzou české populace. Také v polské populaci byla objevena spojitost polymorfismů v genech na 6. chromozomu se vznikem i průběhem sarkoidózy. Limitací podrobnější analýzy u polské populace byla především velikost souboru pacientů i kontrolních jedinců. Pro podrobnější analýzu by bylo potřeba rozšířit soubor pacientů, a to zejména pro analýzu asociací u jednotlivých klinických fenotypů sarkoidózy.

#### **5.4 HLA varianty asociované se sarkoidózou v korejské populaci**

V rámci práce byly vyšetřovány HLA alely u korejských pacientů se sarkoidózou. Byla zjišťována jejich spojitost se samotným onemocněním i klinickými fenotypy sarkoidózy. Byly identifikovány varianty asociované s rizikem sarkoidózy, jako jsou HLA-C\*03:04, HLA-DRB1\*12:01 a HLA-DRB1\*14:54, dříve popsané v jiných populacích (Rossman *et al.*, 2003; Levin *et al.*, 2015; Ishihara *et al.*, 1994). Dále byly identifikovány nové varianty HLA asociované s rizikem sarkoidózy (např. HLA-DQA1\*05:08) a nové ochranné varianty HLA-DQB1\*03:02 a HLA-DQA1\*01:02 v korejské populaci. V neposlední řadě potvrzujeme výsledky předchozích studií, kdy varianta HLA-DRB1\*12:01 byla nalezena ve spojitosti s příznivější prognózou (Levin *et al.*, 2015), což bylo zde potvrzeno i v korejské populaci.

Lokus *HLA-DRB1* je často spojován s autoimunitními onemocněními, včetně sarkoidózy (Rossman *et al.*, 2003; Grunewald *et al.*, 2016). V této práci byla identifikována varianta HLA-DRB1\*12:01 rizikový faktor pro vznik sarkoidózy. V souladu s tímto zjištěním

byla tato varianta popsána jako riziková u afroamerické populace (Rossman *et al.*, 2003; Levin *et al.*, 2015). Tato varianta byla v korejské populaci také více zastoupena u pacientů, s remisí onemocnění. HLA-DRB1\*12:01 tedy může být spojena s lepší prognózou sarkoidózy v korejské populaci. Dále byla v této populaci stanovena varianta HLA-DRB1\*14:54 jako rizikový faktor sarkoidózy. Je zajímavé, že z omezeného počtu studií o HLA u sarkoidózy ve východoasijské populaci Ishihara *et al.*, 1994 uvádí, že sarkoidóza je spojena s HLA-DRB1\*12:01 a také s alelami skupiny HLA-DRB1\*14 u japonských pacientů. Dále byly alely skupiny HLA-DRB1\*14, do které patří HLA-DRB1\*14:54, dříve asociovány s rizikem onemocnění, rozsáhlým mimoplicním postižením a chronickým průběhem onemocnění u turecké populace (Yanardag *et al.*, 2017; Esendagli *et al.*, 2018).

*HLA-DQB1* byl dalším lokusem HLA II. třídy obsahujícím variantu asociovanou se sarkoidózou v korejské populaci, konkrétně HLA-DQB1\*03:02 fungující jako protektivní faktor. Tato varianta nebyla dosud v souvislosti se sarkoidózou zmiňována.

Lokus *HLA-DQA1* nebyl v souvislosti se sarkoidózou příliš často zmiňován. V této práci byla zjištěna LD mezi alelami *HLA-DQA1* (např. HLA-DQA1\*05) a souvisejícími variantami *HLA-DRB1*, což by mohlo být důvodem zjištěných asociací. V souvislosti s průběhem onemocnění byla identifikována varianta HLA-DQA1\*01:02 jako protektivní v korejské populaci.

Lokusy HLA I. třídy (*HLA-A*, *-B*, *-C*) byly dříve v souvislosti se sarkoidózou a autoimunitními onemocněními zmiňovány jen zřídka. V případech, kdy byla zaznamenána asociace HLA I. třídy se sarkoidózou, byla jako vysvětlení navrhována buď vazebná nerovnováha mezi HLA I. a II. třídy, nebo nezávislé působení HLA I. třídy v souvislosti s imunitní odpovědí na intracelulární mykobakteriální infekci (Rybicki *et Iannuzzi*, 2007; Tarasidis *et Arce*, 2020). U korejské populace byly se sarkoidózou asociovány varianty HLA-C\*03:02 a HLA-C\*03:04, resp. HLA-C\*03:04, která navíc souvisela s pokročilejšími stadii a závažnější prognózou.

Další varianta HLA I. třídy, HLA-A\*33:03, byla spojena s ochranou před mimoplicním postižením. Tato varianta byla dříve pozorována s frekvencí vyšší než 10 % v celé korejské populaci (Lee *et al.*, 2005).

Zde prezentovaná analýza HLA u korejské populace má svá omezení. Hlavní limitací jsou menší počty pacientů zejména při porovnávání klinických fenotypů. Optimální by byla možnost typizace vlastní kontrolní korejské populace namísto použití údajů z populačních databází. Pozorované asociace je proto třeba interpretovat s určitou rezervou, dokud nebudou data ověřena v další skupině korejských pacientů a kontrolních jedinců.

Naproti tomu existují důležité pozitivní aspekty tohoto srovnání. Dosud bylo publikováno pouze několik studií zkoumajících roli variability v HLA lokusech u sarkoidózy, kde byly použity metody genotypizace s vysokým rozlišením (Karakaya *et al.*, 2019). Největší přínosem této části práce však je, že HLA u sarkoidózy nebyla v korejské populaci vůbec zkoumána, a to bez ohledu na úroveň přesnosti.

## 5.5 SNP asociované se sarkoidózou v korejské populaci

V korejské populaci bylo analyzováno 20 SNP u 139 pacientů se sarkoidózou. Byla zjišťována jejich spojitost se samotným onemocněním i klinickými fenotypy sarkoidózy. Byly identifikovány polymorfismy asociované se vznikem onemocnění obecně (rs1891467 v genu *TGF-β2*; rs3917200 v genu *TGF-β3*; rs2066844 v genu *NOD2*; rs3830135 v genu *HLA-DRB1* a rs2040410 spojený s variantou *HLA-DRB1\*03:01*) i polymorfismy protektivní proti vzniku sarkoidózy (rs1964995 v lokusu *HLA-DRA*, *HLA-DRB5*; rs2187668 v genu *HLA-DQA1* a rs3819717 v genu *TAP2*). U pacientů s mimoplicní lokalizací sarkoidózy byl identifikován polymorfismus rs1800629 v genu *TNF-α*, naopak polymorfismus rs2040410 spojený s variantou *HLA-DRB1\*03:01* se vyskytoval více u pacientů, u kterých jsou zasaženy jen plíce. S pacienty, u nichž nebyla indikována léčba, byl asociován polymorfismus rs3917165 v genu *TGF-β3* se závažnějším průběhem onemocnění. Se stadii 2-4 byl naopak asociován polymorfismus rs1891467 v genu *TGF-β2*.

Jako rizikový pro vznik sarkoidózy byl identifikován polymorfismus rs2066844 v genu *NOD2*, který byl již dříve asociován se sarkoidózou především s její progresivní formou (Sato *et al.*, 2010b). Tento polymorfismus vede k záměně aminokyselin v *C-terminal leucine-rich repeat* (LRR) a ta má regulační i rozpoznávací funkci pro PAMPs. LRR jsou spojeny s časným nástupem sarkoidózy. Mutace *NOD2* jsou spojeny s multisystémovým granulomatózním onemocněním charakterizovaným artritidou, uveitidou a kožními chorobami (Negroni *et al.*, 2018).

Jako další rizikový polymorfismus byl nalezen rs3830135 v genu *HLA-DRB1*. Tento polymorfismus zde byl v rámci polské populace i v předešlých studiích evropských populací (Rivera *et al.*, 2016) nalezen jako ochranný proti vzniku sarkoidózy. V asijské populaci tento polymorfismus zatím nebyl zkoumán.

Polymorfismus rs2040410 asociovaný s variantou *HLA-DRB1\*03:01* byl nalezen v této práci jak rizikový pro sarkoidózu obecně, tak i spojený s pacienty, kteří mají sarkoidózu zasaženy jen plíce. Toto odpovídá předchozím studiím na evropské a americké populaci (Levin *et al.*, 2015; Lahtela *et al.*, 2019). Varianta *HLA-DRB1\*03:01* má zároveň vysokou

vazebnou afinitu k peptidům na povrchu bakterií dříve již spojených s patogenezí sarkoidózy (*Mycobacterium tuberculosis* a *Propionibacterium acnes*) (Levin *et al.*, 2015).

Jako rizikové pro vznik sarkoidózy byly v této práci nalezeny polymorfismy rs1891467 a rs3917200 v genech *TGF-β2* a *TGF-β3*. Polymorfismy v izoformách genu *TGF-β* byly nalezeny asociovány se sarkoidózou již dříve (Sikorova *et al.*, 2020; Pabst *et al.*, 2011; Kruit *et al.*, 2006). *TGF-β* vede k aktivaci a proliferaci fibroblastů a stimuluje je k produkci kolagenu, fibronektinu a proteoglykanů; inhibuje degradaci kolagenu deaktivací matrixových metaloproteináz a aktivací inhibitorů proteináz (Kato *et al.*, 2022). Rovnováha mezi izoformami *TGF-β* může určovat povahu hojení. Isoformy 1 a 2 rodiny *TGF-β* jsou obecně popisovány jako profibrotické, zatímco *TGF-β3* působí antifibroticky (Kruit *et al.*, 2006). Polymorfismus rs3917165 v genu *TGF-β3*, nalezen v této práci především u pacientů, kterým nebyla indikována léčba, byl popsán jako spojený se vznikem fibrózy v holandské populaci (Kruit *et al.*, 2006). Polymorfismus rs1891467 v genu *TGF-β2* byl nalezen asociován s vyššími stadii u korejských pacientů, což odpovídá i výsledkům v holandské populaci (Kruit *et al.*, 2006), nikoliv však výsledkům v německé populaci (Pabs *et al.*, 2011).

Polymorfismus rs1800629 v genu *TNF-α* byl v této práci asociován s mimoplicními projevy sarkoidózy. V souvislosti s postižením srdce sarkoidózou byl nalezen již dříve v japonské populaci (Takashige *et al.*, 1999).

Polymorfismus rs1964995 v lokusu *HLA-DRA*, *HLA-DRB5* byl nalezen v této práci jako protektivní, což odpovídá výsledkům zde prezentovaným u české populace i předchozí studii provedené u evropské populace (Rivera *et al.*, 2016). Jako protektivní byl v této práci identifikován i polymorfismus rs2187668 v genu *HLA-DQA1*, spojený s variantou *HLA-DQA1\*05:01*, která byla v této práci nalezena asociována s kontrolní populací při porovnání s korejskými pacienty se sarkoidózou. Jako protektivní byla v této práci identifikována také přítomnost alely T v polymorfismu rs3819717 v genu *TAP2*. V evropské populaci byla alela G v polymorfismu rs3819717 v genu *TAP2* identifikována jako riziková pro vznik sarkoidózy (Rivera *et al.*, 2016).

## 5.6 SNP asociované se sarkoidózou v řecké populaci

Bylo analyzováno 20 jednonukleotidových polymorfismů, spojených s imunitní reakcí či metabolizací vápníku. V této souvislosti se jedná o první studii genetické variability genů souvisejících s imunitou provedenou u řeckých pacientů se sarkoidózou. Hlavním zjištěním byla souvislost mezi sarkoidózou a variantou rs3917200 v genu *TGF-β3*. Dále tato práce

potvrdila, že varianta *ANXA11* rs1049550\*A byla asociována se sarkoidózou jako celkem i u řeckých pacientů.

Gen *ANXA11* se podílí na apoptóze a má vliv na tvorbu granulomu (Mirsaeidi *et al.*, 2016). Polymorfismus rs1049550 vede k aminokyselinové substituci arginin/cytosin na pozici 230 ve vazebné doméně zodpovědné za vazbu vápenatých iontů. Asociace tohoto polymorfismu se sarkoidózou byla poprvé zaznamenána v rámci celogenomové asociační studie provedené v německé populaci (Hofmann *et al.*, 2008) a následně byla replikována v české (Mrazek *et al.*, 2011), portugalské (Morais *et al.*, 2013) a americké (Levin *et al.*, 2013) populaci s protektivním účinkem. Protektivní asociace tohoto polymorfismu s pokročilejšími stadii 2-4 u sarkoidózy byla zatím zaznamenána pouze u české populace (Mrazek *et al.*, 2011). Asociace s perzistencí onemocnění byla zaznamenána u americké populace (Levin *et al.*, 2013).

U genu *TGF-β3* se předpokládá antifibrotický účinek při hojení ran (Gilbert *et al.*, 2016; Shah *et al.*, 1995). Polymorfismus rs3917200 v genu *TGF-β3* byl již dříve popsán v souvislosti s fibrotickými procesy (Pabst *et al.*, 2011). V předchozí studii tohoto polymorfismu u německé populace byl pozorován trend naznačující mutaci alely C tohoto SNP jako faktor přispívající ke vzniku plicní fibrózy u pacientů se sarkoidózou (Pabst *et al.*, 2011). Podobná pozorování týkající se tohoto SNP v *TGF-β3* byla zaznamenána u nizozemské populace (Kruit *et al.*, 2006). Je tedy možné předpokládat, že SNP rs3917200 v *TGF-β3* by mohl být potenciálním markerem progresu onemocnění u pacientů se sarkoidózou.

Byla pozorována také asociace polymorfismu rs1800629 v genu *TNF-α* s pokročilejšími stadii 2-4 u sarkoidózy ve srovnání s kontrolními jedinci. Asociace mezi tímto polymorfismem a sarkoidózou obecně byla již dříve zaznamenána u jiné populace (Mrazek *et al.*, 2005), ne však asociace s vyššími stadii 2-4 u sarkoidózy. *TNF-α* hraje důležitou roli v patogenezi sarkoidózy při tvorbě granulomů a polymorfismus rs1800629 může tedy ovlivňovat transkripci tohoto genu, a tím i jeho produkci.

Provedená analýza polymorfismů u řecké populace měla určitá omezení. Byl vyšetřen relativně nízký počet pacientů, zejména v podskupinách se závažnější prognózou a vyššími stadii. K ověření těchto výsledků a prozkoumání jejich možného vztahu ke klinickému průběhu onemocnění by bylo zapotřebí replikovat výsledky na větší kohortě pacientů.

## 5.7 Analýza zdravé chorvatské populace

Jedná se o první analýzu diverzity HLA lokusů v populaci východního Chorvatska provedenou u zdravých dárců krve. Analýza byla provedena u 6 HLA lokusů pomocí NGS umožňujícího rozsáhlou analýzu s pokrytím exonů/intronů s minimem nejednoznačností. Poprvé v chorvatské populaci byla provedena analýza 6 lokusů s rozlišením na 4 pole. Výsledky byly porovnány s chorvatským registrem dárců kostní dřeně.

Porovnání výsledků v lokusu *HLA-A* neodhalilo žádné signifikantní rozdíly. Shodovalo se i pořadí nejčastějších alel A\*02:01:01, A\*01:01:01:01, A\*03:01:01:01, a A\*24:01:01:01. Rozdíly mezi údaji prezentovanými v této práci a údaji z registru dárců kostní dřeně byly pozorovány v lokusu *HLA-B*. Varianta HLA-B\*18:0, která zde byla uvedena jako 6. nejčastější, byla v registru druhou nejčastější (8.16 %) (Grubic *et al.*, 2014). Toto spolu s výskytem varianty HLA-B\*35:01, která byla v této práci pozorována u chorvatské populace jako druhá nejčastější, poukazuje na silný vliv jihovýchodoevropských populací na složení HLA u východních Chorvatů, jelikož tyto se shodují se zastoupením v turecké (Pingel *et al.* 2013) či italské populaci (Rendine *et al.*, 2012).

V pořadí variant v lokusu *HLA-C* bylo pozorováno také několik odlišností. V registru dárců kostní dřeně byla na prvním místě varianta HLA-C\*07:01 (21.77 %) a na druhém varianta HLA-C\*04:01 (15.59 %), ve zde prezentované populaci byla HLA-C\*07:01 (14.41 %) na druhém místě a HLA-C\*04:01 (18.01 %) na prvním místě. Obecně bylo pořadí alel shodné spíše s výsledky v turecké populaci (Pingel *et al.*, 2013).

Odlišnosti byly i ve výsledcích lokusu *HLA-DRB1*. Varianty HLA – DRB1\*16:01 (13.96 %) a DRB1\*03:01 (7.66 %) byly ve zde prezentovaných výsledcích první a pátá nejčastější, zatímco mezi chorvatskými dárci kostní dřeně byly v tomto pořadí na čtvrtém a prvním místě. Zde prezentované výsledky tak odpovídají spíše východoevropské, bulharské, populaci (Ivanova *et al.*, 2002). Naopak poměr výskytu variant HLA-DRB1\*11:01:01:01 a HLA-DRB1\*11:04:01 byl v souladu s výsledky chorvatského registru dárců kostní dřeně.

Haplotypovou analýzou byly zjištěny a potvrzeny haplotypy pozorované i v rámci registru dárců kostní dřeně. Četnost výskytu haplotypu A\*11:01:01:01-C\*04:01:01:01-B\*35:01:01-DRB1\*01:01:01-DQA1\*01:01:01:01-DQB1\*05:01:01:03 (1.35 %) se shodovala s výsledky polské populace (Nowak *et al.*, 2008). Byl pozorován i výskyt haplotypů DRB1\*03:01:01:01- DQA1\*05:01:01:02-DQB1\*02:01:01, DRB1\*15:01:01-DQA1\*01:02:01:01- DQB1\*06:02:01:01, DRB1\*01:01:01-DQA1\*01:01:01:01-DQB1\*05:01:01:03 a 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 v této práci dříve pozorovaných ve spojitosti s výskytem či průběhem sarkoidózy v české populaci.

## 5.8 Spojitost výskytu polymorfismů s geografickou polohou

V české a polské populaci, kdy se jedná o populace sousedních středoevropských států, se výsledky analýzy spojitosti polymorfismů se sarkoidózou téměř shodovaly. V obou případech byla identifikována spojitost skupiny polymorfismů na 6. chromozomu (rs1800629, rs2187668, rs3129927, 3130288, rs3135394 a rs2040410) s příznivější prognózou sarkoidózy (Obr. 5). U polské populace byla identifikována nejen protektivní asociace polymorfismu rs3830135 před vznikem sarkoidózy, ale i ukazatel rizika vzniku sarkoidózy polymorfismus rs9277357. U české populace byly navíc identifikovány polymorfismy rs3917165 a rs3135388 spojené se závažnějším průběhem onemocnění. Dále byla u české populace potvrzena ochranná funkce polymorfismu rs1049550 před vznikem sarkoidózy, což bylo možné pozorovat i v další evropské populaci, a to sice v řecké populaci (Obr. 4). Zde byly již pozorovány jisté odlišnosti ve výskytu asociací se sarkoidózou mezi touto a středoevropskými populacemi. Například polymorfismus rs1800629 zde byl asociován s vyššími stadii onemocnění. V této populaci byl identifikován nový pravděpodobný rizikový faktor onemocnění polymorfismus rs3917200 v genu *TGF- $\beta$ 3*, který byl rizikovým pro vznik sarkoidózy i v korejské populaci (Obr. 4). V těchto dvou populacích, řecké a korejské, byla společným znakem i spojitost polymorfismu rs1800629 se závažnějším průběhem sarkoidózy (Obr. 6).

Ve východoasijské populaci z Korejské republiky byly pozorovány další odlišnosti oproti populacím evropským. Se sarkoidózou zde byly asociovány především polymorfismy v lokusu *TGF- $\beta$* . Jedná se o cytokin vyskytující se ve více izoformách a důsledkem narušení rovnováhy (poměru) těchto izoform může být vznik fibrózy. Polymorfismy rs1891467, rs2066844, rs3830135 a rs3917200 byly v korejské populaci identifikovány jako rizikové pro vznik sarkoidózy, naopak polymorfismy rs1964995, rs2187668 a rs3819717 měly v rámci sarkoidózy ochranný charakter. Polymorfismus rs2040410 byl v korejské populaci pozorován jako rizikový pro vznik sarkoidózy. V rámci pacientů se sarkoidózou se ovšem více vyskytoval ve skupině, kdy nemoc zasahovala pouze plíce, což odpovídá spojitosti s příznivější prognózou ve středoevropských populacích (Obr. 5).

V rámci analýzy HLA variant je v této práci také popsána podrobná charakteristika diverzity HLA u východochorvatské populace, která poukazuje na charakteristické znaky alel a haplotypů, které odpovídají složité populační historii studované skupiny Chorvatů a této



geografické oblasti. Tato data zvyšují populační a geografické pokrytí daty NGS a dávají základ klinickým a genetickým údajům. Tato studie představuje užitečnou referenci pro populační studie a studie asociací HLA s nemocemi.

V analýze jednonukleotidových polymorfismů jednotlivých populací České republiky, Polské republiky, Řecké republiky a Korejské republiky byly nalezeny polymorfismy, které pro danou skupinu neodpovídaly Hardy-Weinbergově rovnováze. Tyto polymorfismy se nacházely především na chromozomu 6 v oblasti MHC. Jedná se o oblast s velmi vysokou variabilitou a výskytem paralogií úseků, ovlivňujících HWE a komplikujících tak genotypizační analýzu (Kennedy *et al.*, 2017).

## 6 Závěr

V této práci bylo vybráno a analyzováno 20 jednonukleotidových polymorfismů v genech spojených s patofyziologií sarkoidózy pomocí metody *MassArray* v české, polské, řecké a korejské populaci. Dále byly analyzovány HLA varianty u české a korejské populace pomocí NGS metody s vysokým rozlišením. Následně byla analyzována spojitost těchto polymorfismů a HLA variant se sarkoidózou. Byla analyzována i spojitost polymorfismů a HLA variant s jednotlivými klinickými fenotypy, a to především v detailně popsané populaci 301 českých pacientů se sarkoidózou, kdy byla k dispozici i početná kontrolní skupina.

V této práci se podařilo replikovat výsledky předchozích studií a objevit i některé nové souvislosti mezi zkoumanými polymorfismy HLA variantami a sarkoidózou. Byla také zjevná závislost přítomnosti polymorfismů a různých HLA variant na geografickém umístění dané populace. V české a polské populaci byla identifikována spojitost skupiny polymorfismů na 6. chromozomu s příznivější prognózou sarkoidózy. Spojitost byla pozorovatelná i u české a řecké populace. Zde byla potvrzena ochranná funkce polymorfismu rs1049550 před vznikem sarkoidózy.

V korejské populaci byly se sarkoidózou asociovány polymorfismy v lokusu *TGF- $\beta$* , polymorfismus rs3917200 byl asociován i s řeckými pacienty se sarkoidózou. Polymorfismus rs2040410 byl v korejské populaci pozorován jako rizikový pro vznik sarkoidózy a s pacienty, u kterých nemoc zasahovala pouze plíce. To odpovídá výsledkům ve střeoevropských populacích.

Byly také analyzovány HLA varianty. Je zde popsána podrobná charakteristika diverzity HLA u východochorvatské populace. Oblast MHC je dlouhodobě uváděna ve spojitosti se sarkoidózou. V české populaci byla jako rizikový faktor pro vznik sarkoidózy potvrzena varianta HLA-DQB1\*06:02 a nově byla jako riziková pozorována varianta HLA-DQB1\*06:04. Byla navržena také souvislost varianty HLA-DRB1\*11:01, již dříve spojené s rizikem sarkoidózy, jako rizikového faktoru pro progredující průběh onemocnění. Naopak byly pozorovány některé varianty HLA spojené s mimoplicním postižením a chronickým průběhem onemocnění. Nositelé těchto HLA variant by měli být pečlivě sledováni, vyšetřováni na mimoplicní postižení a měla by být zvažována jejich léčba. V této práci, jako v jedné z prvních, byla potvrzena role 8.1 ancestrálního haplotypu u sarkoidózy, jenž byl často zmiňován u jiných autoimunitních onemocnění.

Přínos interakce 8.1 ancestralního haplotypu s fenotypem autoimunitního onemocnění, zejména s lepší prognózou sarkoidózy, je nejdůležitějším zjištěním. Tuto souvislost je třeba zvážit s ohledem na přístup personalizované medicíny a může být užitečná, pokud jde o indikaci léčby kortikosteroidy.

V korejské populaci, byly potvrzeny asociace mezi některými variantami HLA a sarkoidózou, které byly dříve popsány v jiných populacích. Byly identifikovány i některé, dříve nepopsané, spojení HLA variant se sarkoidózou, zejména v lokusech HLA II. třídy, definované pomocí přesné genotypizace NGS ve zde použitém souboru korejských pacientů. Než však bude možné dále zkoumat pravděpodobný klinický význam nebo funkční význam zde uvedených asociací, je nutné provést potvrzující studie v jiné korejské kohortě pacientů se sarkoidózou.

Zde uváděné výsledky imunogenetických analýz, získané studiem čtyř různých populací, potvrzují heterogenitu sarkoidózy na základě etnického původu pacientů. Toto zjištění by mělo být, jak z této studie vyplývá, zohledněno i v případě personalizace přístupu k diagnostice a určování prognózy či indikaci léčby tohoto onemocnění.

## 7 Seznam publikací a konferenčních příspěvků

### 7.1 Seznam publikací spojených s dizertační prací

**Sikorová K**, Moon SJ, Yoon HY, Strnad A, Song JW, Petrek M. HLA class II variants defined by next generation sequencing are associated with sarcoidosis in Korean patients. *Scientific Reports* 2022; 12(1): 1-9. IF (2021): 4.997, Q2

**Sikorova K**, Kishore A, Rapti A, Adam K, Kocourkova L, Zizkova V, Charikiopoulou M, Kalianos A, Bouros E, Bouros D, Petrek M. Association of TGF- $\beta$ 3 and ANXA11 with pulmonary sarcoidosis in Greek population. *Expert Rev Respir Med.* 2020;14(10):1065-1069. IF (2019): 2.725, Q2

Tokić S, Žižkova V, Štefanić M, Glavaš-Obrovac L, Marci S, Samardžija M, **Sikorova K**, Petrek M. 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. *Sci Rep.* 2020;10(1):5513. IF (2019): 3.998, Q1

### 7.2 Seznam ostatních publikací

Sterclova M, Kishore A, **Sikorova K**, Skibova J, Petrek M, Vasakova M. Effect of genotype on the disease course in idiopathic pulmonary fibrosis despite antifibrotic treatment. *Biomed Rep.* 2021 Nov;15(5):87. doi: 10.3892/br.2021.1463. Epub 2021 Aug 23. PMID: 34589215; PMCID: PMC8444193. AIS (2021): 0.422, Q3

Navratilova Z, Losse S, Petrova P, **Sikorova K**, Chabronova A, Petrek M. The Effect of Tobacco Smoking and Smoking Cessation on Urinal miRNAs in a Pilot Study. *Life (Basel).* 2020;10(9):191. doi: 10.3390/life10090191. IF (2019): 2.991, Q2

### 7.3 Kapitola v monografii

Petřek M, Kishore A, **Sikorová K**: Moderní koncepce genetiky sarkoidózy z pohledu genetika, str. 79-86. In: Sarkoidóza - nemoc mnoha tváří. Kolek V (Ed.). Maxdorf Praha 2019, 438 stran.

### 7.4 Konferenční příspěvky

**Sikorová K**, Yoon H, Strnad A, Kocourková L, Moon S, Song J, Petrek M. Polymorphisms in immune related genes are associated with Sarcoidosis in Korean Population. (ERS 2022 – European Respiratory Society (ERS) International Congress, September 4 – 6, 2022, Barcelona, Spain, poster)

**Sikorova K**, Gavronova A, Petrková J, Kocourkova L, Osoegawa K, Fernandez-Vina MA, Petrek M. HLA variants and severe COVID-19: Preliminary report from Czechia. *HLA* 2022; 527-527. (Abstracts for the 35th European Immunogenetics and Histocompatibility Conference, May 17 – 20, 2022, Amsterdam, the Netherlands, poster)

**Sikorová K**, Žižková V, Kocourková L, Strnad A, Doubková M, Petřek M. HLA associations in Czech Patients with Sarcoidosis. *European Respiratory Journal* 58 (suppl 65) PA702 (ERS 2021 – European Respiratory Society (ERS) International Congress, virtual, September 4 – 8, 2021)

**Sikorová K**, Kim M, Žižková V, Moon Su-jin, Woo Song J, Petřek M. Association of HLA variants with sarcoidosis in Korean population- pilot study. *Abstract Book*, page 36. (14th East West Immunogenetics Conference (EWIC), March 5-7, 2020, Budapest, Hungary)

**Sikorová K**, Kishore A, Rapti A, Kalliopi A, Zizkova V, Charikiopoulou M, Kalianos A, Petrek M. Association of TGF- $\beta$ 3 and ANXA11 with Pulmonary Sarcoidosis in Greek Population. *Abstract Book*, page 151. (International Conference on Sarcoidosis and Interstitial Lung Diseases - WASOG 2018, June 7-9, 2018, Heraklion Crete, Greece)

**Sikorová K**, Kishore A, Doubková M, Rębał K, Dubaniewicz A, Kolek V, Petřek M. Evaluation of sarcoidosis genetic risk based on 18 susceptibility markers in a west-slavonic population. *Abstract Book*, page 10. (IMTM reactor PhD. conference, June 14 – 16, 2017, Pastviny) - ústní sdělení

**Sikorova K**, Kishore A, Doubkova M, Rębała K, Dubaniewicz A, Kolek V, Petrek M. Evaluation of sarcoidosis genetic risk based on 18 susceptibility markers in a West-Slavonic population: subanalysis in context of Löfgren syndrome. *HLA* 2017;89(6):463. (31st European Immunogenetics and Histocompatibility Conference (EFI) & 25th Annual Meeting of the German Society for Immunogenetics (DGI), May 30 – June 2, 2017, Mannheim/Heidelberg, Germany) - poster

## 8 Seznam zkratek

°C	stupně celsia
μl	mikrolitr
A, T, C, G	adenin, thymin, cytosin, guanin
ACE	angiotensin-konvertující enzym
Ala	Alanin
ANXA11	annexin 11
APC	<i>antigen presenting cell</i> (antigen prezentující buňky)
Arg	Arginin
ATF6B	<i>activating transcription factor 6 beta</i>
ATP	<i>adenosine triphosphate</i> adenosintrifosfát
ATS	<i>american thoracic society</i>
BAFF	<i>B-cell activating factor</i>
BAL	bronchoalveolární laváž
BMPR2	<i>bone morphogenetic protein receptor type 2</i>
bp	pár bází
BTNL2	<i>butyrophilin like 2</i> (butyrofilin 2)
C1orf141	<i>chromosome 1 open reading frame 141</i>
C6orf10	<i>chromosome 6 open reading frame 10</i>
CACFD1	<i>calcium channel flower domain containing 1</i>
CaCl <sub>2</sub>	<i>chlorid vápenatý</i>
CCL	<i>C-C motive chemokine ligand</i> C-C chemokinový ligand
CD	<i>cluster of diferenciation</i> diferenciační skupina
CI	<i>conficence interval</i> interval spolehlivosti
CNV	copy number variation variace počtu kopií
CT	<i>computed tomography</i> výpočetní tomografie
CXCL	<i>C-X-C motif chemokine ligand</i> C-X-C chemokinový ligand
Cys	Cystein
DNA	<i>deoxyribonucleic acid</i> deoxyribonukleová kyselina
dNTP	<i>deoxynucleotide triphosphate</i> deoxynukleotid trifosfát
dsDNA	double-stranded DNA dvouvláknová DNA
EDTA	<i>ethylen diaminotetraacetic acid</i> ethylen diaminotetraoctová kyselina
EKG	elektrokardiografie

ERS	<i>European Respiratory Society</i>
EtOH	ethanol
F	<i>forward</i>
<i>f</i>	frekvence
FEV1	<i>forced expired volume in 1 second</i> usilovně vydechnutý objem za 1 sekundu
FP	fluorescenční polarizace
FRET	<i>fluorescence resonance energy transfer</i>
FVC	<i>forced vital capacity</i> usilovná vitální kapacita,
Gly	Glycin
GREM1	<i>gremlin 1</i>
GWAS	<i>genome wide asociation study</i> celogenomová asociční studie
HLA	<i>human leukocyte antigen</i> lidský leukocytový antigen
HWE	Hardy–Weinberg equilibrium Hardy-Weinbergova rovnováha
Chr.	chromozóm
Ig	Immunoglobulin
IL	interleukin
Ile	Isoleucin
Inc	<i>Incorporated</i>
INF- $\gamma$	interferon- $\gamma$
IPD-IMGT	<i>The Immuno Polymorphism Database- Immunogenetics database</i>
JAK	<i>Janus kinase</i> Janusova kináza
JAK-STAT	<i>Janus kinase/signal transducers and activators of transcription</i>
LD	<i>linkage disequilibrium</i> vazebná nerovnováha
LRR	<i>leucine-rich repeat</i>
LRRC16A	Leucine Rich Repeat Containing Protein 16A
LS	Löfregnúv syndrom
M	molární
MAF	<i>minor allele frequency</i> frekvence minoritních alel
MALDI	<i>matrix assisted laser desorption/ionization</i>
MgCl <sub>2</sub>	chlorid hořečnatý
MgCl <sub>2</sub> *6H <sub>2</sub> O	chlorid hořečnatý hexahydrát
MHC	<i>major histocompatibility complex</i> hlavní histokompatibilní komplex
min	minuta
ml	mililitr

MS	<i>mass spektrometry</i> hmotnostní spektrometrie
mTORC1	<i>mammalian target of rapamycin complex 1</i>
n	<i>number</i> počet
NA	neanalyzováno
NaCl	chlorid sodný
NaOH	hydroxid sodný
NFκB	nuclear factor-kappa B
ng	nanogram
NGS	<i>next generation sequencing</i> sekvenování příští generace
NHGRI	<i>National Human Genome Research Institute</i>
NIGMS	<i>Human Genetic Cell Repository</i>
NLR	NOD-like receptory
nm	nanometr
NOD2	<i>Nucleotide Binding Oligomerization Domain Containing 2</i>
NOTCH4	Notch receptor 4
OR	<i>odds ratio</i> poměr šancí
OS9	osteosarcoma 9
Ot	otáčky
PAMP	<i>pathogen-associated molecular patterns</i>
p <sub>corr</sub>	p-hodnota po korekci pro mnohočetná srovnání
PCR	<i>polymerase chain reaction</i> polymerázová řetězová reakce
PET	pozitronová emisní tomografie
pH	potential of hydrogen
pM	pikomolární
Pro	Prolin
PRS	<i>polygenic risk score</i>
qPCR	<i>quantitative PCR</i> kvantitativní polymerázová řetězová reakce
R	<i>reverse</i>
RNA	<i>ribonucleic acid</i> ribonukleová kyselina
RTG	rentgenové záření
s	sekunda
SAA	sérový amyloid A
SAP	<i>shrimp alkaline phosphatase</i>
SDS	dodecylsírán sodný



Ser	Serin
SLCO1B1	<i>solute carrier organic anion transporter family member 1B1</i>
SMAD	<i>suppressor of mothers against decapentaplegic</i>
SMRT	<i>single molekule real time</i>
SNP	<i>single nucleotide polymorphism</i> jednonukleotidový polymorfismus
SRRM3	<i>serine/arginine repetitive matrix 3</i>
SSOP	<i>sequence specific oligonucleotide probes</i> sekvenčně specifické oligonuklotidové sondy
SSP	sekvenčně specifické primery
STAT	<i>signal transducer and activator of transcription</i>
STYXL1	<i>serine/threonine/tyrosine-interacting-like protein 1</i>
TAP2	<i>antigen peptide transporter 2</i>
Tc	<i>T-cytotoxic cell</i> cytotoxický T lymfocyt
TCR	<i>T cell receptor</i> T lymfocytový receptor
TE	Tris- EDTA
TGF- $\beta$	<i>transforming growth factor <math>\beta</math></i> transformující růstový faktor $\beta$
Th	<i>T helper cell</i> pomocný T lymfocyt
TLR	<i>Toll Like Receptor</i>
TNF	<i>tumor necrosis factor</i> tumor nekrotizující faktor
TOF	<i>time of fly</i> analyzátor doby letu
Treg	regulačními T-lymfocyty
Tris	tris(hydroxymethyl)aminomethane
Trp	Tryptofan
TYK2	<i>tyrosine kinase 2</i>
USA	<i>United States of America</i>
UTR	<i>untranslated region</i> nepřekládaná oblast
WASOG	<i>World Association of Sarcoidosis and other Granulomatous Disorders</i>

## 9 Seznam literatury

1. Ahmadzai, H., Loke, W. S. J., Huang, S., Herbert, C., Wakefield, D., & Thomas, P. (2014). Biomarkers in sarcoidosis: a review. *Curr Biomark Find*, 4, 93-106.
2. Arkema, E. V., & Cozier, Y. C. (2018). Epidemiology of sarcoidosis: current findings and future directions. *Therapeutic Advances in Chronic Disease*, 9(11), 227-240.
3. Arkema, E. V., & Cozier, Y. C. (2020). Sarcoidosis epidemiology: recent estimates of incidence, prevalence and risk factors. *Current opinion in pulmonary medicine*, 26(5), 527-534.
4. ATS, ERS (1999). WASOG. Statement on sarcoidosis. Joint Statement of the American Thoracic Society (ATS), the European Respiratory Society (ERS) and the World Association of Sarcoidosis and Other Granulomatous Disorders (WASOG) adopted by the ATS Board of Directors and by the ERS Executive Committee, February 1999. *Am J Respir Crit Care Med*, 160, 736-755.
5. Backman, J. D., Li, A. H., Marcketta, A., Sun, D., Mbatchou, J., Kessler, M. D., ... & Ferreira, M. A. (2021). Exome sequencing and analysis of 454,787 UK Biobank participants. *Nature*, 599(7886), 628-634.
6. Barton, A., Thomson, W., Ke, X., Eyre, S., Hinks, A., Bowes, J., ... & Worthington, J. (2008). Rheumatoid arthritis susceptibility loci at chromosomes 10p15, 12q13 and 22q13. *Nature genetics*, 40(10), 1156-1159.
7. Baughman, R. P., Valeyre, D., Korsten, P., Mathioudakis, A. G., Wuyts, W. A., Wells, A., ... & Ouellette, D. R. (2021). ERS clinical practice guidelines on treatment of sarcoidosis. *European respiratory journal*, 58(6).
8. Bergantini, L., Bianchi, F., Cameli, P., Mazzei, M. A., Fui, A., Sestini, P., ... & Bargagli, E. (2019). Prognostic biomarkers of sarcoidosis: a comparative study of serum chitotriosidase, ACE, lysozyme, and KL-6. *Disease Markers*, 2019:8565423.
9. Berlin, M., Fogdell-Hahn, A., Olerup, O., Eklund, A., & Grunewald, J. (1997). HLA-DR predicts the prognosis in Scandinavian patients with pulmonary sarcoidosis. *American journal of respiratory and critical care medicine*, 156(5), 1601-1605.
10. Bogdan, M., Nitsch-Osuch, A., Kanecki, K., Gorynski, P., Tyszko, P., Baranska, A., & Samel-Kowalik, P. (2019). Sarcoidosis among hospitalized patients in Poland: a study based on a National hospital registry. *Pol Arch Intern Med*, 129(9), 580-5.
11. Bravo-Egana, V., Sanders, H., & Chitnis, N. (2021). New challenges, new opportunities: Next generation sequencing and its place in the advancement of HLA typing. *Human Immunology*, 82(7), 478-487.
12. Brito-Zerón, P., Kostov, B., Superville, D., Baughman, R. P., & Ramos-Casals, M. (2019). Geoepidemiological big data approach to sarcoidosis: geographical and ethnic determinants. *Clin Exp Rheumatol*, 37(6), 1052-64.

13. Calender, A., Weichhart, T., Valeyre, D., & Pacheco, Y. (2020). Current insights in genetics of sarcoidosis: functional and clinical impacts. *Journal of Clinical Medicine*, 9(8), 2633.
14. Candore, G., Lio, D., Romano, G. C., & Caruso, C. (2002). Pathogenesis of autoimmune diseases associated with 8.1 ancestral haplotype: effect of multiple gene interactions. *Autoimmunity reviews*, 1(1-2), 29-35.
15. Cornaby, C., Schmitz, J. L., & Weimer, E. T. (2021). Next-generation sequencing and clinical histocompatibility testing. *Human Immunology*, 82(11), 829-837.
16. Cornaby, C., Montgomery, M. C., Liu, C., & Weimer, E. T. (2022). Unique Molecular Identifier-Based High-Resolution HLA Typing and Transcript Quantitation Using Long-Read Sequencing. *Frontiers in genetics*, 13.
17. Crouser, E. D., Maier, L. A., Wilson, K. C., Bonham, C. A., Morgenthau, A. S., Patterson, K. C., ... & Baughman, R. P. (2020). Diagnosis and detection of sarcoidosis. An official American Thoracic Society clinical practice guideline. *American journal of respiratory and critical care medicine*, 201(8), e26-e51.
18. Cullinan, A., & Cantor, C. (2008). Sequenom, Inc. *Pharmacogenomics*, 9(9), 1211-5.
19. Drent, M., Crouser, E. D., & Grunewald, J. (2021). Challenges of sarcoidosis and its management. *New England Journal of Medicine*, 385(11), 1018-1032.
20. Dubaniewicz, A., & Moszkowska, G. (2007). Analysis of occurrence of DRB and DQ alleles in sarcoidosis and tuberculosis from Northern Poland. *Advances in Respiratory Medicine*, 75(1), 13-21.
21. Ellis, J. A., & Ong, B. (2017). The MassARRAY<sup>®</sup> system for targeted SNP genotyping. In *Genotyping* (pp. 77-94).
22. Esendagli, D., Ozmen, F., Koksall, D., Onder, S., & Emri, S. (2018). Association of class II human leukocyte antigen (HLA) alleles with pulmonary sarcoidosis. *Sarcoidosis, Vasculitis, and Diffuse Lung Diseases*, 35(2), 143.
23. Fischer, A., Schmid, B., Ellinghaus, D., Nothnagel, M., Gaede, K. I., Schürmann, M., ... & Schreiber, S. (2012). A novel sarcoidosis risk locus for Europeans on chromosome 11q13. 1. *American journal of respiratory and critical care medicine*, 186(9), 877-885.
24. Fischer, A., Ellinghaus, D., Nutsua, M., Hofmann, S., Montgomery, C. G., Iannuzzi, M. C., ... & Schreiber, S. (2015). Identification of immune-relevant factors conferring sarcoidosis genetic risk. *American journal of respiratory and critical care medicine*, 192(6), 727-736.
25. Foley, P. J., McGrath, D. S., Puscinska, E., Petrek, M., Kolek, V., Drabek, J., ... & du Bois, R. M. (2001). Human leukocyte antigen-DRB1 position 11 residues are a common protective marker for sarcoidosis. *American journal of respiratory cell and molecular biology*, 25(3), 272-277.
26. Francisco, M. D. L. V., Lazaruk, K. D., Rhodes, M. D., & Wenz, M. H. (2005). Assessment of two flexible and compatible SNP genotyping platforms: TaqMan<sup>®</sup> SNP Genotyping Assays and the SNPlex<sup>™</sup> Genotyping System. *Mutation*

- Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 573(1-2), 111-135.
27. Franke, A., Fischer, A., Nothnagel, M., Becker, C., Grabe, N., Till, A., ... & Schreiber, S. (2008). Genome-wide association analysis in sarcoidosis and Crohn's disease unravels a common susceptibility locus on 10p12. *Gastroenterology*, 135(4), 1207-1215.
  28. Gambino, C. M., Aiello, A., Accardi, G., Caruso, C., & Candore, G. (2018). Autoimmune diseases and 8.1 ancestral haplotype: an update. *HLA*, 92(3), 137-143.
  29. Garman, L., Pezant, N., Pastori, A., Savoy, K. A., Li, C., Levin, A. M., ... & Montgomery, C. G. (2021). Genome-wide association study of ocular sarcoidosis confirms HLA associations and implicates barrier function and autoimmunity in African Americans. *Ocular immunology and inflammation*, 29(2), 244-249.
  30. Gilbert, R. W., Vickaryous, M. K., & Vilorio-Petit, A. M. (2016). Signalling by transforming growth factor beta isoforms in wound healing and tissue regeneration. *Journal of developmental biology*, 4(2), 21.
  31. Grubić, Z., Žunec, R., Peroš-Golubičić, T., Tekavec-Trkanjec, J., Martinez, N., Alilović, M., ... & Kerhin-Brkljačić, V. (2007). HLA class I and class II frequencies in patients with sarcoidosis from Croatia: role of HLA-B8,- DRB1\* 0301, and- DQB1\* 0201 haplotype in clinical variations of the disease. *Tissue antigens*, 70(4), 301-306.
  32. Grubic, Z., Burek Kamenaric, M., Mikulic, M., Stingl Jankovic, K., Maskalan, M., & Zunec, R. (2014). HLA-A, HLA-B and HLA-DRB 1 allele and haplotype diversity among volunteer bone marrow donors from Croatia. *International journal of immunogenetics*, 41(3), 211-221.
  33. Grunewald, J., Eklund, A., & Olerup, O. (2004). Human leukocyte antigen class I alleles and the disease course in sarcoidosis patients. *American journal of respiratory and critical care medicine*, 169(6), 696-702.
  34. Grunewald, J. (2010). Role of genetics in susceptibility and outcome of sarcoidosis. *Seminars in respiratory and critical care medicine*, 31(04), 380-389.
  35. Grunewald, J., Kaiser, Y., Ostadkarampour, M., Rivera, N. V., Vezzi, F., Lötstedt, B., ... & Eklund, A. (2016). T-cell receptor-HLA-DRB1 associations suggest specific antigens in pulmonary sarcoidosis. *European Respiratory Journal*, 47(3), 898-909.
  36. Grutters, J. C., Sato, H., Welsh, K. I., & du Bois, R. M. (2003). The importance of sarcoidosis genotype to lung phenotype. *American journal of respiratory cell and molecular biology*, 29(3), S59.
  37. Gwadera, Ł., Białas, A. J., Iwański, M. A., Górski, P., & Piotrowski, W. J. (2019). Sarcoidosis and calcium homeostasis disturbances—Do we know where we stand?. *Chronic respiratory disease*, 16, 1479973119878713.
  38. Hena, K. M. (2020). Sarcoidosis epidemiology: race matters. *Frontiers in immunology*, 11, 537382.
  39. Hilberg, O., Hoffmann-Vold, A. M., Smith, V., Bouros, D., Kilpeläinen, M., Guiot, J., ... & Wuyts, W. (2022). Epidemiology of interstitial lung diseases and

- their progressive-fibrosing behaviour in six European countries. *European Respiratory Journal Open Research*, 8(1), 00597-2021.
40. Hofmann, S., Franke, A., Fischer, A., Jacobs, G., Nothnagel, M., Gaede, K. I., ... & Schreiber, S. (2008). Genome-wide association study identifies ANXA11 as a new susceptibility locus for sarcoidosis. *Nature genetics*, 40(9), 1103-1106.
  41. Hofmann, S., Fischer, A., Till, A., Müller-Quernheim, J., Häslner, R., Franke, A., ... & GenPhenReSa Consortium. (2011). A genome-wide association study reveals evidence of association with sarcoidosis at 6p12. 1. *European Respiratory Journal*, 38(5), 1127-1135.
  42. Hofmann, S., Fischer, A., Nothnagel, M., Jacobs, G., Schmid, B., Wittig, M., ... & Schreiber, S. (2013). Genome-wide association analysis reveals 12q13. 3–q14. 1 as new risk locus for sarcoidosis. *European Respiratory Journal*, 41(4), 888-900.
  43. Howell, W. M., Carter, V., & Clark, B. (2010). The HLA system: immunobiology, HLA typing, antibody screening and crossmatching techniques. *Journal of clinical pathology*, 63(5), 387-390.
  44. Chung, J. H., Cox, C. W., Forssen, A. V., Biederer, J., Puderbach, M., & Lynch, D. A. (2014). The dark lymph node sign on magnetic resonance imaging: a novel finding in patients with sarcoidosis. *Journal of Thoracic Imaging*, 29(2), 125-129.
  45. Iannuzzi, M. C., Rybicki, B. A., & Teirstein, A. S. (2007). Sarcoidosis. *The New England journal of medicine*, 357(21), 2153–2165.
  46. Ishihara, M., Ohno, S., Ishida, T., Ando, H., Naruse, T., Nose, Y., & Inoko, H. (1994). Molecular genetic studies of HLA class II alleles in sarcoidosis. *Tissue Antigens*, 43(4), 238-241.
  47. Ivanova, M., Rozemuller, E., Tyufekchiev, N., Michailova, A., Tilanus, M., & Naumova, E. (2002). HLA polymorphism in Bulgarians defined by high-resolution typing methods in comparison with other populations. *Tissue antigens*, 60(6), 496-504.
  48. Jain, R., Yadav, D., Puranik, N., Guleria, R., & Jin, J. O. (2020). Sarcoidosis: causes, diagnosis, clinical features, and treatments. *Journal of Clinical Medicine*, 9(4), 1081.
  49. Jeny, F., Bernaudin, J. F., Aubart, F. C., Brillet, P. Y., Bouvry, D., Nunes, H., & Valeyre, D. (2020). Diagnosis issues in sarcoidosis. *Respiratory Medicine and Research*, 77, 37-45.
  50. Jeon, M. H., Kang, T., Yoo, S. H., Swan, H. S., Kim, H. J., & Ahn, H. S. (2020). The incidence, comorbidity and mortality of sarcoidosis in Korea, 2008-2015: a nationwide population-based study. *Sarcoidosis, Vasculitis, and Diffuse Lung Diseases*, 37(1), 24.
  51. Jiang, L., Zheng, Z., Fang, H., & Yang, J. (2021). A generalized linear mixed model association tool for biobank-scale data. *Nature genetics*, 53(11), 1616-1621.
  52. Judson, M. A. (2019). A sarcoidosis clinician's perspective of MHC functional elements outside the antigen binding site. *Human Immunology*, 80(1), 85-89.
  53. Judson, M. A. (2020). Environmental risk factors for sarcoidosis. *Frontiers in Immunology*, 11, 1340.

54. Karakatsani, A., Papakosta, D., Rapti, A., Antoniou, K. M., Dimadi, M., Markopoulou, A., ... & Bouros, D. (2009). Epidemiology of interstitial lung diseases in Greece. *Respiratory medicine*, 103(8), 1122-1129.
55. Karakaya, B., Schimmelpennink, M. C., Kocourkova, L., van der Vis, J. J., Meek, B., Grutters, J. C., ... & van Moorsel, C. H. M. (2019). Bronchoalveolar lavage characteristics correlate with HLA tag SNPs in patients with Löfgren's syndrome and other sarcoidosis. *Clinical & Experimental Immunology*, 196(2), 249-258.
56. Karakaya, B., van der Vis, J. J., Veltkamp, M., Biesma, D. H., Grutters, J. C., & van Moorsel, C. H. (2022). ANXA11 rs1049550 Associates with Löfgren's Syndrome and Chronic Sarcoidosis. *Cells*, 11(9), 1557.
57. Katchar, K., Eklund, A., & Grunewald, J. (2003). Expression of Th1 markers by lung accumulated T cells in pulmonary sarcoidosis. *Journal of internal medicine*, 254(6), 564-571.
58. Kato, S., Sakai, Y., Okabe, A., Kawashima, Y., Kuwahara, K., Shiogama, K., ... & Morimoto, S. I. (2022). Histology of Cardiac Sarcoidosis with Novel Considerations Arranged upon a Pathologic Basis. *Journal of Clinical Medicine*, 11(1), 251.
59. Kennedy, A. E., Ozbek, U., & Dorak, M. T. (2017). What has GWAS done for HLA and disease associations?. *International journal of immunogenetics*, 44(5), 195-211.
60. Kim, H. J., Kim, H. M., & Song, J. W. (2021). Clinical characteristics and outcomes of Korean patients with sarcoidosis. *Scientific reports*, 11(1), 1-8.
61. Kim, M. S., Pinto, S. M., Getnet, D., Nirujogi, R. S., Manda, S. S., Chaerkady, R., ... & Pandey, A. (2014). A draft map of the human proteome. *Nature*, 509(7502), 575-581.
62. Kim, S., & Misra, A. (2007). SNP genotyping: technologies and biomedical applications. *Annu. Rev. Biomed. Eng.*, 9, 289-320.
63. Kishore, A., & Petrek, M. (2013). Immunogenetics of sarcoidosis. *J Int Trends Immun*, 1, 43-53.
64. Kishore, A., & Petrek, M. (2018). Next-generation sequencing based HLA typing: deciphering immunogenetic aspects of sarcoidosis. *Frontiers in genetics*, 9, 503.
65. Kockum, I., Sanjeevi, C. B., Eastman, S., Landin-Olsson, M., Dahlquist, G., & Lernmark, Å. (1999). Complex interaction between HLA DR and DQ in conferring risk for childhood type 1 diabetes. *European journal of immunogenetics*, 26(5), 361-372.
66. Kolek, V. (1994). Epidemiological study on sarcoidosis in Moravia and Silesia. *Sarcoidosis*, 11(2), 110-112.
67. Kolek, V., Žurková, M., & Lošťáková, V. (2017). Sarkoidóza–nadále záhadné onemocnění. *Vnitř Lék*; 63(11): 807-814
68. Kowalska, M., Niewiadomska, E., & Zejda, J. E. (2014). Epidemiology of sarcoidosis recorded in 2006-2010 in the Silesian voivodeship on the basis of routine medical reporting. *Annals of Agricultural and Environmental Medicine*, 21(1), 55-8.

69. Kruit, A., Grutters, J. C., Ruven, H. J., van Moorsel, C. H., Weiskirchen, R., Mengsteab, S., & van den Bosch, J. M. (2006). Transforming growth factor- $\beta$  gene polymorphisms in sarcoidosis patients with and without fibrosis. *Chest*, *129*(6), 1584-1591.
70. Kwok, P. Y. (2001). Methods for genotyping single nucleotide polymorphisms. *Annual review of genomics and human genetics*, *2*(1), 235-258.
71. Lahtela, E., Kankainen, M., Sinisalo, J., Selroos, O., & Lokki, M. L. (2019). Exome sequencing identifies susceptibility loci for sarcoidosis prognosis. *Frontiers in immunology*, *10*, 2964.
72. Lee, K. W., Oh, D. H., Lee, C., & Yang, S. Y. (2005). Allelic and haplotypic diversity of HLA-A,-B,-C,-DRB1, and-DQB1 genes in the Korean population. *Tissue antigens*, *65*(5), 437-447.
73. Levin, A. M., Iannuzzi, M. C., Montgomery, C. G., Trudeau, S., Datta, I., McKeigue, P., ... & Rybicki, B. A. (2013). Association of ANXA11 genetic variation with sarcoidosis in African Americans and European Americans. *Genes & Immunity*, *14*(1), 13-18.
74. Levin, A. M., Adrianto, I., Datta, I., Iannuzzi, M. C., Trudeau, S., Li, J., ... & Rybicki, B. A. (2015). Association of HLA-DRB1 with sarcoidosis susceptibility and progression in African Americans. *American journal of respiratory cell and molecular biology*, *53*(2), 206-216.
75. Liu, J. Z., Hov, J. R., Folseraas, T., Ellinghaus, E., Rushbrook, S. M., Doncheva, N. T., ... & Karlsen, T. H. (2013). Dense genotyping of immune-related disease regions identifies nine new risk loci for primary sclerosing cholangitis. *Nature genetics*, *45*(6), 670-675.
76. Madden, K., & Chabot-Richards, D. (2019). HLA testing in the molecular diagnostic laboratory. *Virchows Archiv*, *474*(2), 139-147.
77. Meguro, A., Ishihara, M., Petrek, M., Yamamoto, K., Takeuchi, M., Mrazek, F., ... & Mizuki, N. (2020). Genetic control of CCL24, POR, and IL23R contributes to the pathogenesis of sarcoidosis. *Communications biology*, *3*(1), 1-10.
78. Miller, S. (1988). A simple salting-out procedure tissue for extracting DNA from human nucleated cells. *Nucleic Acids Res.*, *16*, 221.
79. Mirsaeidi, M., Machado, R. F., Schraufnagel, D., Sweiss, N. J., & Baughman, R. P. (2015). Racial difference in sarcoidosis mortality in the United States. *Chest*, *147*(2), 438-449.
80. Mirsaeidi, M., Gidfar, S., Vu, A., & Schraufnagel, D. (2016). Annexins family: insights into their functions and potential role in pathogenesis of sarcoidosis. *Journal of translational medicine*, *14*(1), 1-9.
81. Monsuur, A. J., de Bakker, P. I., Zhernakova, A., Pinto, D., Verduijn, W., Romanos, J., ... & Wijmenga, C. (2008). Effective detection of human leukocyte antigen risk alleles in celiac disease using tag single nucleotide polymorphisms. *PloS one*, *3*(5), e2270.
82. Morais, A., Lima, B., Peixoto, M., Melo, N., Alves, H., Marques, J. A., & Delgado, L. (2013). Annexin A11 gene polymorphism (R230C variant) and sarcoidosis in a Portuguese population. *Tissue Antigens*, *82*(3), 186-191.

83. Morimoto, T., Azuma, A., Abe, S., Usuki, J., Kudoh, S., Sugisaki, K., ... & Nukiwa, T. (2008). Epidemiology of sarcoidosis in Japan. *European Respiratory Journal*, *31*(2), 372-379.
84. Moudi, M., Sargazi, S., Heidari Nia, M., Saravani, R., Shirvaliloo, M., & Shakiba, M. (2020). Polymorphism in the 3'-UTR of LIF but not in the ATF6B gene associates with schizophrenia susceptibility: a case-control study and in silico analyses. *Journal of Molecular Neuroscience*, *70*(12), 2093-2101.
85. Mrazek, F., Holla, L. I., HutYROVÁ, B., Znojil, V., Vasku, A., Kolek, V., ... & Petrek, M. (2005). Association of tumour necrosis factor- $\alpha$ , lymphotoxin- $\alpha$  and HLA-DRB1 gene polymorphisms with Löfgren's syndrome in Czech patients with sarcoidosis. *Tissue antigens*, *65*(2), 163-171.
86. Mrazek, F., Stahelova, A., Kriegova, E., Fillerova, R., Zurkova, M., Kolek, V., & Petrek, M. (2011). Functional variant ANXA11 R230C: true marker of protection and candidate disease modifier in sarcoidosis. *Genes & Immunity*, *12*(6), 490-494.
87. Müller-Quernheim, J., Prasse, A., & Zissel, G. (2012). Pathogenesis of sarcoidosis. *La Presse Médicale*, *41*(6), e275-e287.
88. Naruse, T. K., Matsuzawa, Y., Ota, M., Katsuyama, Y., Matsumori, A., Hara, M., ... & Inoko, H. (2000). HLA-DQB1\* 0601 is primarily associated with the susceptibility to cardiac sarcoidosis. *Tissue antigens*, *56*(1), 52-57.
89. Negroni, A., Pierdomenico, M., Cucchiara, S., & Stronati, L. (2018). NOD2 and inflammation: current insights. *Journal of inflammation research*, *11*, 49.
90. Nowak, J., Mika-Witkowska, R., Polak, M., Zajko, M., Rogatko-Koroś, M., Graczyk-Pol, E., & Lange, A. (2008). 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*(3), 193-205.
91. Pabst, S., Fränken, T., Schönau, J., Stier, S., Nickenig, G., Meyer, R., ... & Grohé, C. (2011). Transforming growth factor- $\beta$  gene polymorphisms in different phenotypes of sarcoidosis. *European Respiratory Journal*, *38*(1), 169-175.
92. Park, T. J., Kim, J. H., Pasaje, C. F., Park, B. L., Bae, J. S., Uh, S. T., ... & Shin, H. D. (2014). Polymorphisms of ATF6B are potentially associated with fev1 decline by aspirin provocation in asthmatics. *Allergy, asthma & immunology research*, *6*(2), 142-148.
93. Petrek, M., Kocourkova, L., Zizkova, V., Nosek, Z., Taborsky, M., & Petrkova, J. (2016). Characterization of three CYP2C19 gene variants by MassARRAY and point of care techniques: Experience from a Czech Centre. *Archivum immunologiae et therapeuticae experimentalis*, *64*(1), 99-107.
94. Petrek, M., Drabek, J., Kolek, V., Zlamal, J., Welsh, K. I., Bunce, M., ... & du Bois, R. (2000). CC chemokine receptor gene polymorphisms in Czech patients with pulmonary sarcoidosis. *American journal of respiratory and critical care medicine*, *162*(3), 1000-1003.
95. Petrkova, J., Taborsky, M., & Petrek, M. (2018). Pharmacogenetics of Cardiovascular Disease: Genetic Variation and Statin Intolerance. *Genetic Diversity and Disease Susceptibility*.



96. Pingel, J., Solloch, U. V., Hofmann, J. A., Lange, V., Ehninger, G., & Schmidt, A. H. (2013). High-resolution HLA haplotype frequencies of stem cell donors in Germany with foreign parentage: how can they be used to improve unrelated donor searches?. *Human immunology*, 74(3), 330-340.
97. Polverino, F., Balestro, E., & Spagnolo, P. (2020). Clinical presentations, pathogenesis, and therapy of sarcoidosis: state of the art. *Journal of clinical medicine*, 9(8), 2363.
98. Price, P., Witt, C., Allock, R., Sayer, D., Garlepp, M., Kok, C. C., ... & Christiansen, F. (1999). The genetic basis for the association of the 8.1 ancestral haplotype (A1, B8, DR3) with multiple immunopathological diseases. *Immunological reviews*, 167(1), 257-274.
99. Profaizer, T., Lázár-Molnár, E., Close, D. W., Delgado, J. C., & Kumánovics, A. (2016). HLA genotyping in the clinical laboratory: comparison of next-generation sequencing methods. *HLA*, 88(1-2), 14-24.
100. Rahaghi, F. F., Baughman, R. P., Saketkoo, L. A., Sweiss, N. J., Barney, J. B., Birring, S. S., ... & Scholand, M. B. (2020). Delphi consensus recommendations for a treatment algorithm in pulmonary sarcoidosis. *European Respiratory Review*, 29(155).
101. Rendine, S., Ferrero, N. M., Sacchi, N., Costa, C., Pollichieni, S., & Amoroso, A. (2012). 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. *Human immunology*, 73(4), 399-404.
102. Rivera, N. V., Ronninger, M., Shchetynsky, K., Franke, A., Nöthen, M. M., Müller-Quernheim, J., ... & Grunewald, J. (2016). High-density genetic mapping identifies new susceptibility variants in sarcoidosis phenotypes and shows genomic-driven phenotypic differences. *American journal of respiratory and critical care medicine*, 193(9), 1008-1022.
103. Rivera, N. V., Patasova, K., Kullberg, S., Diaz-Gallo, L. M., Iseda, T., Bengtsson, C., ... & Padyukov, L. (2019). A gene–environment interaction between smoking and gene polymorphisms provides a high risk of two subgroups of sarcoidosis. *Scientific reports*, 9(1), 1-11.
104. Rossman, M. D., Thompson, B., Frederick, M., Maliarik, M., Iannuzzi, M. C., Rybicki, B. A., ... & ACCESS Group. (2003). HLA-DRB1\* 1101: a significant risk factor for sarcoidosis in blacks and whites. *The American Journal of Human Genetics*, 73(4), 720-735.
105. Rubio-Rivas, M., Franco, J., & Corbella, X. (2020). Sarcoidosis presenting with and without Löfgren's syndrome: clinical, radiological and behavioral differences observed in a group of 691 patients. *Joint Bone Spine*, 87(2), 141-147.
106. Rybicki, B. A., Iannuzzi, M. C., Frederick, M. M., Thompson, B. W., Rossman, M. D., Bresnitz, E. A., ... & ACCESS Research Group. (2001). Familial aggregation of sarcoidosis: a case–control etiologic study of sarcoidosis (ACCESS). *American journal of respiratory and critical care medicine*, 164(11), 2085-2091.

107. Rybicki, B. A., & Iannuzzi, M. C. (2007). Epidemiology of sarcoidosis: recent advances and future prospects. *Seminars in respiratory and critical care medicine* 28 (01), 022-035.
108. Sakaue, S., Kanai, M., Tanigawa, Y., Karjalainen, J., Kurki, M., Koshiba, S., ... & Okada, Y. (2021). A cross-population atlas of genetic associations for 220 human phenotypes. *Nature genetics*, 53(10), 1415-1424.
109. Sato, H., Grutters, J. C., Pantelidis, P., Mizzon, A. N., Ahmad, T., Van Houte, A. J., ... & Du Bois, R. M. (2002). HLA-DQB1\* 0201: a marker for good prognosis in British and Dutch patients with sarcoidosis. *American journal of respiratory cell and molecular biology*, 27(4), 406-412.
110. Sato, H., Woodhead, F. A., Ahmad, T., Grutters, J. C., Spagnolo, P., van den Bosch, J. M., ... & Welsh, K. I. (2010a). Sarcoidosis HLA class II genotyping distinguishes differences of clinical phenotype across ethnic groups. *Human molecular genetics*, 19(20), 4100-4111.
111. Sato, H., Williams, H. R. T., Spagnolo, P., Abdallah, A., Ahmad, T., Orchard, T. R., ... & Welsh, K. I. (2010b). CARD15/NOD2 polymorphisms are associated with severe pulmonary sarcoidosis. *European Respiratory Journal*, 35(2), 324-330.
112. Seitzer, U., Swider, C., Stüber, F., Suchnicki, K., Lange, A., Richter, E., ... & Gerdes, J. (1997). Tumour necrosis factor alpha promoter gene polymorphism in sarcoidosis. *Cytokine*, 9(10), 787-790.
113. Sève, P., Pacheco, Y., Durupt, F., Jamilloux, Y., Gerfaud-Valentin, M., Isaac, S., ... & El Jammal, T. (2021). Sarcoidosis: a clinical overview from symptoms to diagnosis. *Cells*, 10(4), 766.
114. Shah, M., Foreman, D. M., & Ferguson, M. W. (1995). Neutralisation of TGF-beta 1 and TGF-beta 2 or exogenous addition of TGF-beta 3 to cutaneous rat wounds reduces scarring. *Journal of cell science*, 108(3), 985-1002.
115. Sharma, S. K., Balamurugan, A., Pandey, R. M., Saha, P. K., & Mehra, N. K. (2003). Human leukocyte antigen-DR alleles influence the clinical course of pulmonary sarcoidosis in Asian Indians. *American journal of respiratory cell and molecular biology*, 29(2), 225-231.
116. Shen, Y., Pang, C., Wu, Y., Li, D., Wan, C., Liao, Z., ... & Wen, F. (2016). Diagnostic performance of bronchoalveolar lavage fluid CD4/CD8 ratio for sarcoidosis: a meta-analysis. *EBioMedicine*, 8, 302-308.
117. Shi, Y. Y., & He, L. (2005). SHEsis, a powerful software platform for analyses of linkage disequilibrium, haplotype construction, and genetic association at polymorphism loci. *Cell research*, 15(2), 97.
118. Schurmann, M., Reichel, P., Muller-Myhsok, B., Schlaak, M. A. X., Muller-Quernheim, J., & Schwinger, E. (2001). Results from a genome-wide search for predisposing genes in sarcoidosis. *American journal of respiratory and critical care medicine*, 164(5), 840-846.
119. Sikorova, K., Kishore, A., Rapti, A., Adam, K., Kocourkova, L., Zizkova, V., ... & Petrek, M. (2020). Association of TGF-β3 and ANXA11 with pulmonary

- sarcoidosis in Greek population. *Expert Review of Respiratory Medicine*, 14(10), 1065-1069.
120. Sikorova, K., Moon, S. J., Yoon, H. Y., Strnad, A., Song, J. W., & Petrek, M. (2022). HLA class II variants defined by next generation sequencing are associated with sarcoidosis in Korean patients. *Scientific Reports*, 12(1), 1-9.
  121. Silva, M., Nunes, H., Valeyre, D., & Sverzellati, N. (2015). Imaging of sarcoidosis. *Clinical reviews in allergy & immunology*, 49(1), 45-53.
  122. Starshinova, A. A., Malkova, A. M., Basantsova, N. Y., Zinchenko, Y. S., Kudryavtsev, I. V., Ershov, G. A., ... & Yablonskiy, P. K. (2020). Sarcoidosis as an autoimmune disease. *Frontiers in Immunology*, 10, 2933.
  123. Sterclova, M., Kishore, A., Sikorova, K., Skibova, J., Petrek, M., & Vasakova, M. (2021). Effect of genotype on the disease course in idiopathic pulmonary fibrosis despite antifibrotic treatment. *Biomedical reports*, 15(5), 1-7.
  124. Takashige, N., Naruse, T. K., Matsumori, A., Hara, M., Nagai, S., Morimoto, S., ... & Inoko, H. (1999). Genetic polymorphisms at the tumour necrosis factor loci (TNFA and TNFB) in cardiac sarcoidosis. *Tissue antigens*, 54(2), 191-193.
  125. Tarasidis, A., & Arce, S. (2020). Immune response biomarkers as indicators of sarcoidosis presence, prognosis, and possible treatment: An Immunopathogenic perspective. *Autoimmunity Reviews*, 19(3), 102462.
  126. Terrington, D. L., Hayton, C., Peel, A., Fowler, S. J., Fraser, W., & Wilson, A. M. (2019). The role of measuring exhaled breath biomarkers in sarcoidosis: a systematic review. *Journal of breath research*, 13(3), 036015.
  127. Tokić, S., Žiškova, V., Štefanić, M., Glavaš-Obrovac, L., Marczy, S., Samardžija, M., ... & 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*, 10(1), 1-13.
  128. Tripipitsiriwat, A., Komoltri, C., Ruangchira-Urai, R., & Ungprasert, P. (2020). Clinical Characteristics of Sarcoidosis in Asian Population: A 14-year Single Center Retrospective Cohort Study from Thailand. *Sarcoidosis, Vasculitis, and Diffuse Lung Diseases*, 37(4).
  129. Tyagi, S., Bratu, D. P., & Kramer, F. R. (1998). Multicolor molecular beacons for allele discrimination. *Nature biotechnology*, 16(1), 49-53.
  130. Uffelmann, E., Huang, Q. Q., Munung, N. S., De Vries, J., Okada, Y., Martin, A. R., ... & Posthuma, D. (2021). Genome-wide association studies. *Nature Reviews Methods Primers*, 1(1), 1-21.
  131. Ungprasert, P., Ryu, J. H., & Matteson, E. L. (2019). Clinical manifestations, diagnosis, and treatment of sarcoidosis. *Mayo Clinic Proceedings: Innovations, Quality & Outcomes*, 3(3), 358-375.
  132. Valeyre, D., Bernaudin, J. F., Uzunhan, Y., Kambouchner, M., Brillet, P. Y., Soussan, M., & Nunes, H. (2014). Clinical presentation of sarcoidosis and diagnostic work-up. *Seminars in respiratory and critical care medicine*, 35(03), 336-351.

133. van Moorsel, C. H., Petrek, M., & Rivera, N. (2022). Unravelling the genetic basis of sarcoidosis. In: Bonnella F., Culver D.A., Israël-Biet D. eds. *Sarcoidosis (ERS Monograph)*. Sheffield, *European Respiratory Society* 41-56.
134. Voorter, C. E., Drent, M., & van den Berg-Loonen, E. M. (2005). Severe pulmonary sarcoidosis is strongly associated with the haplotype HLA-DQB1\* 0602–DRB1\* 150101. *Human immunology*, 66(7), 826-835.
135. Wennerström, A., Pietinalho, A., Lasota, J., Salli, K., Surakka, I., Seppänen, M., ... & Lokki, M. L. (2013). Major histocompatibility complex class II and BTNL2 associations in sarcoidosis. *European Respiratory Journal*, 42(2), 550-553.
136. Wojtan, P., Mierzejewski, M., Osieńska, I., & Domagała-Kulawik, J. (2016). Macrophage polarization in interstitial lung diseases. *Central-european Journal of Immunology*, 41(2), 159.
137. Xiong, Y., Kullberg, S., Garman, L., Pezant, N., Ellinghaus, D., Vasila, V., ... & Rivera, N. V. (2021). Sex Differences in the Genetics of Sarcoidosis Across European and African Ancestry Populations. <https://doi.org/10.21203/rs.3.rs-382787/v1>
138. Yanardag, H., Tetikkurt, C., Bilir, M., & Yilmaz, E. (2017). Association of HLA antigens with the clinical course of sarcoidosis and familial disease. *Monaldi Archives for Chest Disease*, 87(3).
139. Yoon, H. Y., Kim, H. M., Kim, Y. J., & Song, J. W. (2018). Prevalence and incidence of sarcoidosis in Korea: a nationwide population-based study. *Respiratory research*, 19(1), 1-8.
140. Zhou, E. R., & Arce, S. (2020). Key players and biomarkers of the adaptive immune system in the pathogenesis of sarcoidosis. *International Journal of Molecular Sciences*, 21(19), 7398.
141. Zhou, H., Diao, M., & Zhang, M. (2016). The Association between ANXA11 Gene Polymorphisms and Sarcoidosis: a Meta-Analysis and systematic review. *Sarcoidosis, Vasculitis, and Diffuse Lung Diseases: Official Journal of WASOG*, 33(2), 102-111.
142. Zurkova, M., Kolek, V., Tomankova, T., & Kriegova, E. (2014). Extrapulmonary involvement in patients with sarcoidosis and comparison of routine laboratory and clinical data to pulmonary involvement. *Biomedical Papers*, 158(4), 613-620.

## 10 Internetové odkazy k statistickým nástrojům a databázím

- [GWAS Catalog \(ebi.ac.uk\)](https://www.ebi.ac.uk/gwas/)
- [The Allele Frequency Net Database - Allele, haplotype and genotype frequencies in Worldwide Populations \(allele-frequencies.net\)](http://www.allele-frequencies.net/)
- [Ensembl genome browser 108](https://www.ncbi.nlm.nih.gov/genome/browser108/)
- [SHEsisPlus \(bio-x.cn\)](https://shesisplus.bio-x.cn/)
- [Analyze a 2x2 contingency table \(graphpad.com\)](https://www.graphpad.com/quickcalcs/contingency2.php)
- [2x2 Contingency Table with Odds Ratios, etc. \(vassarstats.net\)](http://vassarstats.net/contingency2.html)
- [HLA Analysis Tools \(lanl.gov\)](https://www.lanl.gov/HLA/)

**Přílohy Tabulka 21: Rozlišené varianty a varianty související s onemocněním v šesti zkoumaných lokusech HLA (úroveň rozlišení na dvě pole)** (tabulka uvádí počet variant, které se nám podařilo rozlišit pomocí NGS genotypizace HLA, a počet variant asociovaných se sarkoidózou ve zkoumaných HLA lokusech pozorovaných u českých pacientů. Významná asociace: na "primární úrovni":  $p < 0.05$ , před Bonferroniho korekcí na mnohočetná srovnání (\*\*))

<b>Locus</b>	<b>DETERMINED VARIANTS</b>	<b>ASSOCIATED ON PRIMARY LEVEL</b>	<b>ASSOCIATED AFTER CORRECTION*</b>
<i>HLA-A</i>	21	1	0
<i>HLA-B</i>	45	3	1
<i>HLA-C</i>	22	4	3
<i>HLA-DRB1</i>	41	6	2
<i>HLA-DQA1</i>	17	8	3
<i>HLA-DQB1</i>	19	7	3

**Přílohy Tabulka 22: Souhrn HLA variant asociovaných s pacienty ve stadiích 0-1 ( $n = 30$ ) při porovnání s kontrolními jedinci (HLA I. třídy  $n = 485$ ; HLA II. třídy  $n = 324$ ) v korejské populaci (jsou zobrazeny jen alely asociovány alespoň na primární úrovni) ( $f$ -frekvence, OR-odds ratio, CI-confidence interval,  $p_{corr}$ - $p$ -hodnota po korekci pro mnohočetná srovnání (signifikantní zvýrazněna tučně), NA-neanalyzováno)**

HLA	$f$ stadia 0-1	$f$ kontroly	OR (95% CI)	$p$	$p_{corr}$
HLA-A*02:07	0.08	0.03	2.94 (1.10-7.89)	0.043	0.601
HLA-C*03:02	0.02	0.11	0.14 (0.02-1.02)	0.016	0.294
<b>HLA-DRB1*12:01</b>	<b>0.17</b>	<b>0.05</b>	<b>4.24 (1.95-9.19)</b>	<b>0.001</b>	<b>0.033</b>
HLA-DRB1*14:54	0.03	0.00	NA	0.007	0.252
HLA-DQA1*01:02	0.07	0.20	0.28 (0.10-0.79)	0.009	0.145
HLA-DQA1*03:01	0.02	0.10	0.16 (0.02-1.16)	0.033	0.436
<b>HLA-DQA1*05:08</b>	<b>0.07</b>	<b>0.00</b>	<b>NA</b>	<b><math>4.7 \cdot 10^{-05}</math></b>	<b>0.001</b>
HLA-DQB1*03:02	0.02	0.10	0.15 (0.02-1.09)	0.034	0.485

**Přílohy Tabulka 23: Souhrn HLA variant asociovaných s pacienty ve stadiu 2-4 ( $n = 73$ ) při porovnání s kontrolními jedinci (HLA I. třídy  $n = 485$ ; HLA II. třídy  $n = 324$ ) v korejské populaci (jsou zobrazeny jen alely asociovány alespoň na primární úrovni) ( $f$ -frekvence, OR-odds ratio, CI-confidence interval,  $p_{corr}$ - $p$ -hodnota po korekci pro mnohočetná srovnání (signifikantní zvýrazněna tučně), NA-neanalyzováno)**

HLA	$f$ stadia 2-4	$f$ kontroly	OR (95% CI)	$p$	$p_{corr}$
<b>HLA-A*33:03</b>	<b>0.07</b>	<b>0.16</b>	<b>0.38 (0.19-0.73)</b>	<b>0.002</b>	<b>0.038</b>
HLA-B*15:01	0.05	0.11	0.43 (0.2-0.94)	0.035	0.798
<b>HLA-B*15:07</b>	<b>0.03</b>	<b>0.00</b>	NA	<b><math>2.83 \cdot 10^{-04}</math></b>	<b>0.013</b>
HLA-B*51:01	0.14	0.08	1.73 (1.03-2.92)	0.046	0.877
HLA-B*58:01	0.02	0.07	0.30 (0.09-0.97)	0.036	0.812
<b>HLA-C*03:02</b>	<b>0.02</b>	<b>0.11</b>	<b>0.17 (0.05-0.55)</b>	<b><math>2.41 \cdot 10^{-04}</math></b>	<b>0.005</b>
<b>HLA-C*03:04</b>	<b>0.12</b>	<b>0.04</b>	<b>3.25 (1.78-5.93)</b>	<b><math>3.13 \cdot 10^{-04}</math></b>	<b>0.007</b>
HLA-C*05:01	0.03	0.01	3.51 (1.18-10.48)	0.036	0.553
<b>HLA-C*08:22</b>	<b>0.03</b>	<b>0.00</b>	NA	<b><math>2.83 \cdot 10^{-04}</math></b>	<b>0.006</b>
HLA-DRB1*01:01	0.02	0.06	0.31 (0.1-1.02)	0.044	0.842
HLA-DRB1*03:01	0.00	0.03	NA	0.033	0.750
HLA-DRB1*04:06	0.01	0.06	0.22 (0.05-0.93)	0.021	0.576
<b>HLA-DRB1*12:01</b>	<b>0.14</b>	<b>0.05</b>	<b>3.56 (1.97-6.44)</b>	<b><math>7.83 \cdot 10^{-05}</math></b>	<b>0.003</b>
HLA-DRB1*12:10	0.01	0.00	NA	0.034	0.754
HLA-DRB1*13:02	0.04	0.1	0.37 (0.16-0.88)	0.017	0.503
HLA-DRB1*14:03	0.04	0.01	7.08 (1.95-25.66)	0.004	0.147
<b>HLA-DRB1*14:54</b>	<b>0.05</b>	<b>0.00</b>	NA	<b><math>1.11 \cdot 10^{-06}</math></b>	<b><math>4.57 \cdot 10^{-05}</math></b>
HLA-DRB1*14:01	0.00	0.04	NA	0.008	0.283
HLA-DQA1*01:01	0.03	0.08	0.34 (0.12-0.97)	0.042	0.514
<b>HLA-DQA1*01:02</b>	<b>0.07</b>	<b>0.20</b>	<b>0.29 (0.15-0.57)</b>	<b><math>6.13 \cdot 10^{-05}</math></b>	<b>0.001</b>
HLA-DQA1*05:01	0.00	0.03	NA	0.035	0.454
HLA-DQA1*05:06	0.01	0.00	NA	0.034	0.441
<b>HLA-DQA1*05:07</b>	<b>0.04</b>	<b>0.00</b>	NA	<b><math>3.55 \cdot 10^{-05}</math></b>	<b>0.001</b>
<b>HLA-DQA1*05:08</b>	<b>0.03</b>	<b>0.00</b>	NA	<b><math>1.99 \cdot 10^{-04}</math></b>	<b>0.003</b>
HLA-DQB1*02:01	0.00	0.03	NA	0.033	0.474
<b>HLA-DQB1*03:02</b>	<b>0.01</b>	<b>0.10</b>	<b>0.06 (0.01-0.44)</b>	<b><math>2.03 \cdot 10^{-05}</math></b>	<b><math>3.86 \cdot 10^{-04}</math></b>
<b>HLA-DQB1*06:09</b>	<b>0.03</b>	<b>0.00</b>	NA	<b><math>1.98 \cdot 10^{-04}</math></b>	<b>0.004</b>
HLA-DQB1*06:05	0.00	0.04	NA	0.008	0.143

**Přílohy Tabulka 24: Souhrn HLA variant asociovaných s pacienty s mimoplicním postižením ( $n = 73$ ) při porovnání s kontrolními jedinci (HLA I. třídy  $n = 485$ ; HLA II. třídy  $n = 324$ ) v korejské populaci (jsou zobrazeny jen alely asociovány alespoň na primární úrovni) ( $f$ -frekvence, OR-odds ratio, CI-confidence interval,  $p_{corr}$ - $p$ -hodnota po korekci pro mnohočetná srovnání (signifikantní zvýrazněna tučně), NA-neanalyzováno)**

HLA	$f$ mimoplicní	$f$ kontroly	OR (95% CI)	$p$	$p_{corr}$
<b>HLA-A*33:03</b>	<b>0.07</b>	<b>0.16</b>	<b>0.38 (0.19-0.73)</b>	<b>0.002</b>	<b>0.038</b>
<b>HLA-B*15:07</b>	<b>0.03</b>	<b>0.00</b>	NA	<b><math>2.82 \cdot 10^{-04}</math></b>	<b>0.013</b>
HLA-B*44:02	0.04	0.01	3.54 (1.3-9.64)	0.022	0.631
HLA-B*55:02	0.00	0.03	NA	0.039	0.832
<b>HLA-C*03:02</b>	<b>0.03</b>	<b>0.11</b>	<b>0.23 (0.08-0.64)</b>	<b>0.001</b>	<b>0.019</b>
HLA-C*03:04	0.10	0.04	2.82 (1.51-5.28)	0.002	0.054
HLA-C*05:01	0.04	0.01	4.24 (1.51-11.92)	0.012	0.230
<b>HLA-C*08:22</b>	<b>0.02</b>	<b>0.00</b>	NA	<b>0.002</b>	<b>0.047</b>
HLA-DRB1*04:06	0.01	0.06	0.11 (0.01-0.81)	0.005	0.187
HLA-DRB1*12:01	0.12	0.05	2.79 (1.49-5.23)	0.002	0.097
HLA-DRB1*12:10	0.01	0.00	NA	0.034	0.754
<b>HLA-DRB1*14:54</b>	<b>0.03</b>	<b>0.00</b>	NA	<b><math>1.99 \cdot 10^{-04}</math></b>	<b>0.008</b>
HLA-DRB1*14:01	0.00	0.04	NA	0.008	0.283
<b>HLA-DQA1*01:02</b>	<b>0.08</b>	<b>0.20</b>	<b>0.32 (0.17-0.61)</b>	<b><math>1.78 \cdot 10^{-04}</math></b>	<b>0.003</b>
HLA-DQA1*05:07	0.02	0.00	NA	0.006	0.099
<b>HLA-DQA1*05:08</b>	<b>0.03</b>	<b>0.00</b>	NA	<b><math>1.99 \cdot 10^{-04}</math></b>	<b>0.003</b>
<b>HLA-DQB1*03:02</b>	<b>0.01</b>	<b>0.10</b>	<b>0.06 (0.01-0.44)</b>	<b><math>2.03 \cdot 10^{-05}</math></b>	<b><math>3.86 \cdot 10^{-04}</math></b>
HLA-DQB1*04:01	0.03	0.08	0.34 (0.12-0.95)	0.029	0.426
HLA-DQB1*04:02	0.00	0.04	NA	0.008	0.143
HLA-DQB1*05:03	0.09	0.03	3.05 (1.48-6.29)	0.004	0.078
HLA-DQB1*06:03	0.04	0.01	3.87 (1.28-11.65)	0.019	0.312
<b>HLA-DQB1*06:09</b>	<b>0.03</b>	<b>0.00</b>	NA	<b><math>1.99 \cdot 10^{-04}</math></b>	<b>0.004</b>
HLA-DQB1*06:05	0.00	0.04	NA	0.008	0.143
<b>HLA-DPB1*02:01</b>	<b>0.07</b>	<b>0.25</b>	<b>0.23 (0.12-0.44)</b>	<b><math>3.32 \cdot 10^{-07}</math></b>	<b><math>5.98 \cdot 10^{-06}</math></b>
HLA-DPB1*02:02	0.01	0.05	0.13 (0.02-0.93)	0.012	0.199
<b>HLA-DPB1*04:01</b>	<b>0.01</b>	<b>0.09</b>	<b>0.07 (0.01-0.50)</b>	<b><math>8.40 \cdot 10^{-05}</math></b>	<b>0.002</b>
HLA-DPB1*04:02	0.01	0.07	0.19 (0.04-0.78)	0.006	0.105
HLA-DPB1*13:01	0.01	0.07	0.19 (0.04-0.78)	0.006	0.105
HLA-DPB1*135:01	0.02	0.00	NA	0.006	0.105
HLA-DPB1*15:01	0.01	0.00	NA	0.034	0.460



**Přílohy Tabulka 25: Souhrn HLA variant asociovaných s pacienty s mimoplicním postižením ( $n = 73$ ) při porovnání s pacienty, u kterých byly zasaženy jen plíce ( $n = 30$ ) v korejské populaci (jsou zobrazeny jen alely asociovány alespoň na primární úrovni) ( $f$ -frekvence,  $p_{corr}$ - $p$ -hodnota po korekci pro mnohočetná srovnání)**

<b>HLA</b>	<b><math>f</math> jen plíce</b>	<b><math>f</math> mimoplicní</b>	<b><math>p</math></b>	<b><math>p_{corr}</math></b>
HLA-A*02:06	0.143	0.046	0.031	0.483
HLA-C*15:02	0.080	0.007	0.018	0.334
HLA-DRB1*12:01	0.259	0.125	0.030	0.714
HLA-DQA1*05:06	0.068	0.000	0.021	0.297

**Přílohy Tabulka 26: Souhrn HLA variant asociovaných s pacienty s progresí onemocnění ( $n = 5$ ) při porovnání s kontrolními jedinci (HLA I. třídy  $n = 485$ ; HLA II. třídy  $n = 324$ ) v korejské populaci (jsou zobrazeny jen alely asociovány alespoň na primární úrovni) ( $f$ - frekvence, OR-*odds ratio*, CI-*confidence interval*,  $p_{corr}$ - $p$ -hodnota po korekci pro mnohočetná srovnání)**

<b>HLA</b>	<b><i>f</i> progresse</b>	<b><i>f</i> kontroly</b>	<b>OR (95% CI)</b>	<b><i>p</i></b>	<b><i>p</i><sub>corr</sub></b>
HLA-DQB1*05:03	0.20	0.03	7.81 (1.56-39.15)	0.041	0.548

**Přílohy Tabulka 27: Souhrn HLA variant asociovaných s pacienty s remisí onemocnění ( $n = 58$ ) při porovnání s kontrolními jedinci (HLA I. třídy  $n = 485$ ; HLA II. třídy  $n = 324$ ) v korejské populaci (jsou zobrazeny jen alely asociovány alespoň na primární úrovni) ( $f$ -frekvence, OR-odds ratio, CI-confidence interval,  $p_{corr}$ - $p$ -hodnota po korekci pro mnohočetná srovnání (signifikantní zvýrazněna tučně), NA-neanalyzováno)**

HLA	$f$ remise	$f$ kontroly	OR (95% CI)	$p$	$p_{corr}$
<b>HLA-A*33:03</b>	<b>0.06</b>	<b>0.16</b>	<b>0.33 (0.15-0.72)</b>	<b>0.002</b>	<b>0.049</b>
HLA-B*15:07	0.02	0.00	NA	0.011	0.401
HLA-B*44:02	0.05	0.01	4.51 (1.65-12.31)	0.008	0.312
HLA-B*44:03	0.03	0.09	0.29 (0.09-0.92)	0.026	0.699
HLA-C*03:02	0.03	0.11	0.22 (0.07-0.70)	0.003	0.059
HLA-C*03:04	0.10	0.04	2.85 (1.44-5.62)	0.004	0.094
HLA-C*05:01	0.05	0.01	5.40 (1.91-15.23)	0.004	0.090
HLA-C*08:22	0.02	0.00	NA	0.011	0.222
HLA-DRB1*01:01	0.02	0.06	0.26 (0.06-1.09)	0.048	0.869
HLA-DRB1*04:06	0.00	0.06	NA	0.002	0.088
<b>HLA-DRB1*12:01</b>	<b>0.15</b>	<b>0.05</b>	<b>3.64 (1.93-6.86)</b>	<b>1.49*10<sup>-04</sup></b>	<b>0.006</b>
HLA-DRB1*12:10	0.02	0.00	NA	0.023	0.613
HLA-DRB1*13:01	0.04	0.01	4.07 (1.27-13.00)	0.024	0.634
HLA-DRB1*13:02	0.03	0.10	0.31 (0.11-0.87)	0.015	0.459
HLA-DRB1*14:03	0.05	0.01	9.01 (2.48-32.75)	0.001	0.054
<b>HLA-DRB1*14:54</b>	<b>0.04</b>	<b>0.00</b>	<b>NA</b>	<b>7.49*10<sup>-05</sup></b>	<b>0.003</b>
HLA-DRB1*14:01	0.00	0.04	NA	0.023	0.610
HLA-DQA1*01:01	0.03	0.08	0.32 (0.10-1.05)	0.047	0.557
<b>HLA-DQA1*01:02</b>	<b>0.06</b>	<b>0.20</b>	<b>0.25 (0.12-0.56)</b>	<b>1.13*10<sup>-04</sup></b>	<b>0.002</b>
HLA-DQA1*02:01	0.01	0.06	0.15 (0.02-1.08)	0.031	0.418
HLA-DQA1*03:01	0.03	0.10	0.25 (0.08-0.80)	0.011	0.168
HLA-DQA1*05:06	0.03	0.00	NA	0.003	0.057
<b>HLA-DQA1*05:07</b>	<b>0.05</b>	<b>0.00</b>	<b>NA</b>	<b>1.10*10<sup>-05</sup></b>	<b>1.86*10<sup>-04</sup></b>
<b>HLA-DQA1*05:08</b>	<b>0.04</b>	<b>0.00</b>	<b>NA</b>	<b>7.49*10<sup>-05</sup></b>	<b>0.001</b>
HLA-DQB1*02:02	0.01	0.05	0.15 (0.02-1.12)	0.031	0.447
<b>HLA-DQB1*03:02</b>	<b>0.01</b>	<b>0.10</b>	<b>0.08 (0.01-0.56)</b>	<b>2.46*10<sup>-04</sup></b>	<b>0.005</b>
HLA-DQB1*04:01	0.03	0.08	0.32 (0.10-1.04)	0.047	0.596
HLA-DQB1*04:02	0.00	0.04	NA	0.023	0.354
HLA-DQB1*06:03	0.04	0.01	4.07 (1.27-13.00)	0.024	0.373
HLA-DQB1*06:04	0.01	0.06	0.13 (0.02-0.95)	0.013	0.225
<b>HLA-DQB1*06:09</b>	<b>0.03</b>	<b>0.00</b>	<b>NA</b>	<b>0.001</b>	<b>0.010</b>
HLA-DQB1*06:05	0.00	0.04	NA	0.023	0.356

**Přílohy Tabulka 28: Souhrn HLA variant asociovaných s pacienty se stabilním onemocněním ( $n = 32$ ) při porovnání s kontrolními jedinci (HLA I. třídy  $n = 485$ ; HLA II. třídy  $n = 324$ ) v korejské populaci (jsou zobrazeny jen alely asociovány alespoň na primární úrovni) ( $f$ -frekvence, OR-odds ratio, CI-confidence interval,  $p_{corr}$ - $p$ -hodnota po korekci pro mnohočetná srovnání (signifikantní zvýrazněna tučně), NA-neanalyzováno)**

HLA	$f$ stabilizace	$f$ kontroly	OR (95% CI)	$p$	$p_{corr}$
HLA-B*15:07	0.03	0.00	NA	0.004	0.156
HLA-B*40:02	0.09	0.04	2.62 (1.06-6.45)	0.044	0.869
HLA-C*03:02	0.02	0.11	0.13 (0.02-0.95)	0.017	0.308
HLA-C*03:04	0.09	0.04	2.55 (1.04-6.28)	0.049	0.667
HLA-C*08:22	0.03	0.00	NA	0.004	0.080
<b>HLA-DRB1*08:03</b>	<b>0.25</b>	<b>0.07</b>	<b>4.23 (2.24-8.02)</b>	<b><math>4.07 \cdot 10^{-5}</math></b>	<b>0.002</b>
HLA-DRB1*12:01	0.13	0.05	3.03 (1.32-6.93)	0.013	0.415
<b>HLA-DRB1*14:54</b>	<b>0.08</b>	<b>0.00</b>	<b>NA</b>	<b><math>5.07 \cdot 10^{-6}</math></b>	<b><math>2.08 \cdot 10^{-4}</math></b>
HLA-DQA1*01:02	0.09	0.20	0.41 (0.17-0.97)	0.044	0.538
HLA-DQA1*05:08	0.03	0.00	NA	0.008	0.127
HLA-DQB1*03:01	0.03	0.13	0.22 (0.05-0.92)	0.024	0.368
HLA-DQB1*03:02	0.02	0.10	0.14 (0.02-1.02)	0.022	0.348

**Přílohy Tabulka 33: Analýza asociací SNP v porovnání pacientů se stadiem 1 ( $n = 29$ ) s pacienty se stadii 2-4 ( $n = 68$ ) na alelické úrovni v řecké populaci (OR-odds ratio, CI-confidence interval,  $p_{corr}$ - $p$ -hodnota po korekci pro mnohočetná srovnání, signifikantní výsledky zvýrazněny tučně, kurzíva-výsledky, které nesplňovaly kontrolu kvality, NA-neanalyzováno)**

<b>Gen</b>	<b>SNP</b>	<b>OR (95% CI)</b>	<b><i>p</i></b>	<b><i>p<sub>corr</sub></i></b>
<i>ANXA11</i>	rs1049550	1.374 (0.741-2.547)	0.340	0.999
<i>TNF-<math>\alpha</math></i>	rs1800629	0.438 (0.157-1.223)	0.120	0.900
<i>TGF-<math>\beta</math>2</i>	rs1891467	1.403 (0.757-2.600)	0.338	0.999
<i>GREM1</i>	rs1919364	1.551 (0.865-2.783)	0.183	0.974
<i>HLA-DRA, HLA-DRB5</i>	rs1964995	0.472 (0.150-1.483)	0.217	0.988
<i>NOD2</i>	rs2066844	1.563 (0.406-6.018)	0.496	1.000
<i>BTNL2</i>	rs2076530	1.533 (0.837-2.808)	0.210	0.986
<i>HLA-DQA1</i>	rs2187668	0.616 (0.191-1.986)	0.585	1.000
<i>CACFD1</i>	rs3124765	NA	0.999	1.000
<i>C6orf10</i>	rs3129927	0.533 (0.107-2.639)	0.721	1.000
<i>ATF6B</i>	rs3130288	0.617 (0.121-3.143)	0.716	1.000
<i>HLA-DRA</i>	rs3135394	0.533 (0.107-2.639)	0.721	1.000
<i>TAP2</i>	<i>rs3819717</i>	<i>0.806 (0.410-1.585)</i>	<i>0.613</i>	<i>1.000</i>
<i>HLA-DRB1</i>	rs3830135	0.940 (0.227-3.879)	1.000	1.000
<i>TGF-<math>\beta</math>3</i>	rs3917165	2.709 (0.900-8.151)	0.082	0.786
<i>TGF-<math>\beta</math>3</i>	rs3917200	0.916 (0.514-1.634)	0.882	1.000
<i>HLA-DPB1</i>	rs9277357	0.631 (0.314-1.266)	0.237	0.992
<i>LRRC16A</i>	rs9295661	NA	0.552	1.000
HLA-DRB1*03:01	rs2040410	0.460 (0.178-1.190)	0.138	0.257
HLA-DRB1*15:01	rs3135388	0.970 (0.235-4.001)	1.000	1.000

**Přílohy Tabulka 34: Analýza asociací SNP v porovnání pacientů se stadiem 1 ( $n = 29$ ) s kontrolními jedinci ( $n = 100$ ) na alelické úrovni v řecké populaci (OR-odds ratio, CI-confidence interval,  $p_{corr}$ - $p$ -hodnota po korekci pro mnohočetná srovnání, signifikantní výsledky zvýrazněny tučně, kurzíva-výsledky, které nesplňovaly kontrolu kvality, NA-neanalyzováno)**

Gen	SNP	OR (95% CI)	$p$	$p_{corr}$
<i>ANXA11</i>	rs1049550	0.685 (0.372-1.261)	0.288	0.998
<i>TNF-<math>\alpha</math></i>	rs1800629	1.342 (0.457-3.936)	0.566	1.000
<i>TGF-<math>\beta</math>2</i>	rs1891467	1.182 (0.635-2.196)	0.631	1.000
<i>GREM1</i>	rs1919364	1.374 (0.762-2.478)	0.300	0.998
<i>HLA-DRA, HLA-DRB5</i>	rs1964995	0.337 (0.098-1.155)	0.103	0.859
<i>NOD2</i>	rs2066844	2.370 (0.645-8.703)	0.240	0.993
<i>BTNL2</i>	rs2076530	1.176 (0.641-2.158)	0.639	1.000
<i>HLA-DQA1</i>	rs2187668	0.842 (0.270-2.626)	1.000	1.000
<i>CACFD1</i>	rs3124765	1.380 (0.657-2.900)	0.426	1.000
<i>C6orf10</i>	rs3129927	2.321 (0.378-14.238)	0.317	0.999
<i>ATF6B</i>	rs3130288	2.321 (0.378-14.238)	0.317	0.999
<i>HLA-DRA</i>	rs3135394	2.321 (0.378-14.238)	0.317	0.999
<i>TAP2</i>	<i>rs3819717</i>	0.839 (0.410-1.717)	0.723	1.000
<i>HLA-DRB1</i>	rs3830135	0.541 (0.117-2.493)	0.532	1.000
<b><i>TGF-<math>\beta</math>3</i></b>	rs3917165	4.960 (1.645-14.954)	<b>0.005</b>	0.086
<b><i>TGF-<math>\beta</math>3</i></b>	<b>rs3917200</b>	<b>3.260 (1.765-6.021)</b>	<b>1.74*10<sup>-04</sup></b>	<b>0.003</b>
<i>HLA-DPBI</i>	rs9277357	0.793 (0.388-1.617)	0.601	1.000
<i>LRRC16A</i>	rs9295661	NA	0.999	1.000
<i>HLA-DRB1*03:01</i>	rs2040410	0.826 (0.320-2.131)	0.819	0.967
<i>HLA-DRB1*15:01</i>	rs3135388	0.724 (0.200-2.613)	0.770	0.947

**Přílohy Tabulka 35: Analýza asociací SNP v porovnání pacientů se stadii 2-4 ( $n = 68$ ) s kontrolními jedinci ( $n = 100$ ) na alelické úrovni v řecké populaci (OR-ratio, CI-confidence interval,  $p_{corr}$ -p-hodnota po korekci pro mnohočetná srovnání, signifikantní výsledky zvýrazněny tučně, kurzíva-výsledky, které nesplňovaly kontrolu kvality, NA-neanalyzováno)**

Gen	SNP	OR (95% CI)	$p$	$p_{corr}$
<i>ANXA11</i>	rs1049550	0.526 (0.329-0.842)	<b>0.007</b>	0.119
<i>TNF-<math>\alpha</math></i>	rs1800629	2.496 (1.195-5.213)	<b>0.015</b>	0.238
<i>TGF-<math>\beta</math>2</i>	rs1891467	0.889 (0.549-1.439)	0.713	1.000
<i>GREM1</i>	rs1919364	0.894 (0.575-1.390)	0.653	1.000
<i>HLA-DRA, HLA-DRB5</i>	rs1964995	0.792 (0.404-1.555)	0.613	1.000
<i>NOD2</i>	rs2066844	1.240 (0.37-4.149)	0.761	1.000
<i>BTNL2</i>	rs2076530	0.837 (0.52-1.345)	0.474	1.000
<i>HLA-DQA1</i>	rs2187668	1.118 (0.511-2.447)	0.841	1.000
<i>CACFD1</i>	rs3124765	NA	0.126	0.911
<i>C6orf10</i>	rs3129927	3.582 (0.909-14.111)	0.096	0.837
<i>ATF6B</i>	rs3130288	3.095 (0.760-12.601)	0.164	0.960
<i>HLA-DRA</i>	rs3135394	3.582 (0.909-14.111)	0.096	0.837
<i>TAP2</i>	<i>rs3819717</i>	<i>1.181 (0.713-1.959)</i>	<i>0.519</i>	<i>1.000</i>
<i>HLA-DRB1</i>	rs3830135	0.722 (0.264-1.974)	0.625	1.000
<i>TGF-<math>\beta</math>3</i>	rs3917165	1.476 (0.465-4.680)	0.557	1.000
<b><i>TGF-<math>\beta</math>3</i></b>	<b>rs3917200</b>	<b>2.953 (1.836-4.750)</b>	<b>9.52*10<sup>-06</sup></b>	<b>1.71*10<sup>-04</sup></b>
<i>HLA-DPB1</i>	rs9277357	1.203 (0.733-1.975)	0.525	1.000
<i>LRRC16A</i>	rs9295661	2.221 (0.366-13.477)	0.399	1.000
HLA-DRB1*03:01	rs2040410	1.458 (0.785-2.709)	0.263	0.457
HLA-DRB1*15:01	rs3135388	0.613 (0.229-1.637)	0.358	0.588

**Přílohy Tabulka 36: Alelické frekvence HLA-A, HLA-B a HLA-C u 111 nepříbuzných dárců krve z východního Chorvatska (*n*-počet jedinců, *f* – frekvence)**

	<b>HLA-A</b>	<i>n</i>	<i>f</i>	<b>HLA-B</b>	<i>n</i>	<i>f</i>	<b>HLA-C</b>	<i>n</i>	<i>f</i>
1	A*02:01:01	72	32.43%	B*51:01:01	19	8.56%	C*04:01:01:01	31	13.96%
2	A*01:01:01:01	27	12.16%	B*08:01:01	18	8.11%	C*07:01:01	25	11.26%
3	A*03:01:01:01	26	11.71%	B*35:01:01	18	8.11%	C*02:02:02	21	9.46%
4	A*24:02:01:01	26	11.71%	B*07:02:01	15	6.76%	C*07:02:01	18	8.11%
5	A*11:01:01:01	22	9.91%	B*35:03:01	15	6.76%	C*12:03:01	16	7.21%
6	A*32:01:01:01	8	3.60%	B*18:01:01	12	5.41%	C*06:02:01:01	15	6.76%
7	A*23:01:01:01	5	2.25%	B*15:01:01:01	9	4.05%	C*03:03:01	12	5.41%
8	A*25:01:01:01	5	2.25%	B*27:05:02	9	4.05%	C*01:02:01	11	4.96%
9	A*31:01:02:01	5	2.25%	B*44:02:01:01	9	4.05%	C*05:01:01	11	4.96%
10	A*26:01:01:01	4	1.80%	B*27:02:01:01	8	3.60%	C*07:04:01	11	4.96%
11	A*30:01:01	3	1.35%	B*52:01:01	8	3.60%	C*04:01:01:06	9	4.05%
12	A*68:01:02:02	3	1.35%	B*55:01:01	8	3.60%	C*12:02:02	8	3.60%
13	A*24:02:01:05	2	0.90%	B*57:01:01:01	7	3.15%	C*15:02:01:01	7	3.15%
14	A*29:02:01:01	2	0.90%	B*13:02:01:01	6	2.70%	C*03:04:01	5	2.25%
15	A*66:01:01:01	2	0.90%	B*44:27:01	6	2.70%	C*07:01:02	4	1.80%
16	A*03:01:01:03	1	0.45%	B*44:03:01	5	2.25%	C*08:02:01:01	3	1.35%
17	A*03:01:01:05	1	0.45%	B*15:17:01:01	4	1.80%	C*14:02:01	3	1.35%
18	A*26:01:01:06	1	0.45%	B*35:02:01	4	1.80%	C*07:18	2	0.90%
19	A*26:08	1	0.45%	B*38:01:01	4	1.80%	C*15:05:02	2	0.90%
20	A*29:01:01:01	1	0.45%	B*39:01:01	4	1.80%	C*17:03:01	2	0.90%
21	A*31:01:02:04	1	0.45%	B*14:02:01:01	3	1.35%	C*03:02:02:01	1	0.45%
22	A*33:01:01:01	1	0.45%	B*40:01:02	3	1.35%	C*03:04:02	1	0.45%
23	A*33:03:01:02	1	0.45%	B*40:02:01	3	1.35%	C*06:02:01:02	1	0.45%
24	A*68:01:01:02	1	0.45%	B*41:02:01	3	1.35%	C*07:01:09	1	0.45%
25	A*68:02:01:01	1	0.45%	B*58:01:01	3	1.35%	C*16:02:01	1	0.45%
26				B*18:03	2	0.90%		1	0.45%
27				B*35:08:01:01	2	0.90%			
28				B*37:01:01:01	2	0.90%			
29				B*49:01:01	2	0.90%			
30				B*56:01:01	2	0.90%			
31				B*07:05:01	1	0.45%			
32				B*07:06:01	1	0.45%			
33				B*15:01:06	1	0.45%			
34				B*15:10:01	1	0.45%			
35				B*27:02:01:04	1	0.45%			
36				B*44:02:01:03	1	0.45%			
37				B*44:05:01	1	0.45%			
38				B*50:01:01:01	1	0.45%			
39				B*54:01:01	1	0.45%			



**Přílohy Tabulka 37: Alelické frekvence HLA-DRB1, HLA-DQA1 a HLA-DQB1 u 111 nepříbuzných dárců krve z východního Chorvatska (*n* – počet jedinců, *f* – frekvence)**

	HLA-DRB1	<i>n</i>	<i>f</i>	HLA-DQA1	<i>n</i>	<i>f</i>	HLA-DQB1	<i>n</i>	<i>f</i>
1	DRB1*16:01:01	30	13.51%	DQA1*01:02:02	35	15.77%	DQB1*05:02:01:01	35	15.77%
2	DRB1*01:01:01	24	10.81%	DQA1*03:01:01	16	7.21%	DQB1*03:01:01:03	33	14.86%
3	DRB1*07:01:01	22	9.91%	DQA1*02:01:01:01	15	6.76%	DQB1*05:01:01:03	23	10.36%
4	DRB1*15:01:01	15	6.76%	DQA1*01:03:01:02	13	5.86%	DQB1*02:01:01	17	7.66%
5	DRB1*03:01:01:01	12	5.41%	DQA1*05:01:01:02	12	5.41%	DQB1*03:02:01:01	16	7.21%
6	DRB1*11:01:01:01	10	4.50%	DQA1*01:02:01:01	10	4.50%	DQB1*02:02:01:01	15	6.76%
7	DRB1*11:04:01	10	4.50%	DQA1*05:05:01:09	10	4.50%	DQB1*05:03:01	13	5.86%
8	DRB1*14:54:01:02	9	4.05%	DQA1*01:01:01:01	9	4.05%	DQB1*06:02:01:01	13	5.86%
9	DRB1*08:01:01	8	3.60%	DQA1*05:05:01:04	9	4.05%	DQB1*06:03:01:01	13	5.86%
10	DRB1*13:01:01:02	8	3.60%	DQA1*01:01:01:03	7	3.15%	DQB1*03:03:02:01	7	3.15%
11	DRB1*13:01:01:01	7	3.15%	DQA1*01:03:01:01	7	3.15%	DQB1*04:02:01:01	7	3.15%
12	DRB1*15:02:01:01	7	3.15%	DQA1*02:01:01:02	7	3.15%	DQB1*06:01:01/15	7	3.15%
13	DRB1*11:01:01:03	6	2.70%	DQA1*01:04:01:01	6	2.70%	DQB1*06:04:01	4	1.80%
14	DRB1*13:02:01:02	5	2.25%	DQA1*05:05:01:01	6	2.70%	DQB1*03:01:01:05	3	1.35%
15	DRB1*03:01:01:03	4	1.80%	DQA1*01:02:01:04	5	2.25%	DQB1*05:01:01:05	3	1.35%
16	DRB1*04:01:01:03	3	1.35%	DQA1*01:04:01:02	5	2.25%	DQB1*06:03:01:02	3	1.35%
17	DRB1*04:02:01	3	1.35%	DQA1*05:01:01:03	5	2.25%	DQB1*03:03:02:02	2	0.90%
18	DRB1*04:03:01:01	3	1.35%	DQA1*01:01:01:02	4	1.80%	DQB1*03:01:01:01	1	0.45%
19	DRB1*10:01:01:03	3	1.35%	DQA1*01:05:01	4	1.80%	DQB1*03:01:01:02	1	0.45%
20	DRB1*11:03:01	3	1.35%	DQA1*04:01:01:03	4	1.80%	DQB1*03:01:01:10	1	0.45%
21	DRB1*12:01:01/12:10	3	1.35%	DQA1*05:05:01:02	4	1.80%	DQB1*03:01:01:19	1	0.45%
22	DRB1*13:03:01	3	1.35%	DQA1*05:05:01:05	4	1.80%	DQB1*03:02:01:02	1	0.45%
23	DRB1*04:01:01:01	2	0.90%	DQA1*01:01:01:05	3	1.35%	DQB1*05:01:01:01	1	0.45%
24	DRB1*04:05:01:03	2	0.90%	DQA1*01:02:01:05	3	1.35%	DQB1*05:01:01:02	1	0.45%
25	DRB1*04:08:01	2	0.90%	DQA1*03:03:01:01	3	1.35%	DQB1*05:04	1	0.45%
26	DRB1*09:01:02	2	0.90%	DQA1*03:02:01:01	2	0.90%			
27	DRB1*14:01:01	2	0.90%	DQA1*04:02	2	0.90%			
28	DRB1*16:02:01:02	2	0.90%	DQA1*05:05:01:06	2	0.90%			
29	DRB1*01:02:01	1	0.45%	DQA1*05:05:01:08	2	0.90%			
30	DRB1*03:01:01:02	1	0.45%	DQA1*01:01:02	1	0.45%			
31	DRB1*04:01:01:02	1	0.45%	DQA1*01:02:01:03	1	0.45%			
32	DRB1*04:04:01	1	0.45%	DQA1*01:03:01:06	1	0.45%			
33	DRB1*04:15	1	0.45%	DQA1*01:04:01:03	1	0.45%			
34	DRB1*08:04:01	1	0.45%	DQA1*01:04:01:04	1	0.45%			
35	DRB1*10:01:01:01	1	0.45%	DQA1*01:10	1	0.45%			
36	DRB1*11:01:08	1	0.45%	DQA1*04:01:02:01	1	0.45%			
37	DRB1*13:05:01	1	0.45%	DQA1*05:05:01:03	1	0.45%			
38	DRB1*14:05:01:02	1	0.45%						
39	DRB1*14:54:01:01	1	0.45%						
40	DRB1*16:01:02	1	0.45%						

**Přílohy Tabulka 38: Nejčastější ( $f > 1$ ) haplotypy v chorvatské populaci zdravých dárců krve ( $n = 111$ ) ( $n$ -počet,  $f$ -frekvence)**

HLA-A-B-C			$n$	$f$	
1	A*01:01:01:01	B*08:01:01	C*07:01:01	15	6.76%
2	A*11:01:01:01	B*35:01:01	C*04:01:01:01	8	3.60%
3	A*03:01:01:01	B*07:02:01	C*07:02:01	6.89	3.11%
4	A*02:01:01	B*57:01:01:01	C*06:02:01:01	6	2.70%
5	A*02:01:01	B*27:02:01:01	C*02:02:02	5	2.25%
6	A*02:01:01	B*35:01:01	C*04:01:01:06	5	2.25%
7	A*02:01:01	B*35:03:01	C*04:01:01:01	5	2.25%
8	A*02:01:01	B*44:02:01:01	C*05:01:01	5	2.25%
9	A*03:01:01:01	B*35:03:01	C*04:01:01:01	5	2.25%
10	A*02:01:01	B*44:27:01	C*07:04:01	4	1.80%
11	A*11:01:01:01	B*52:01:01	C*12:02:02	4	1.80%
12	A*02:01:01	B*27:05:02	C*02:02:02	3.89	1.75%
13	A*02:01:01	B*18:01:01	C*07:01:01	3	1.35%
14	A*02:01:01	B*52:01:01	C*12:02:02	3	1.35%
15	A*03:01:01:01	B*15:01:01:01	C*07:04:01	3	1.35%
16	A*23:01:01:01	B*44:03:01	C*04:01:01:01	3	1.35%
17	A*24:02:01:01	B*13:02:01:01	C*06:02:01:01	3	1.35%
18	A*24:02:01:01	B*35:02:01	C*04:01:01:06	3	1.35%
19	A*24:02:01:01	B*51:01:01	C*15:02:01:01	3	1.35%
20	A*02:01:01	B*07:02:01	C*07:02:01	2.88	1.30%
HLA-DQA1-DQB1-DRB1			$n$	$f$	
1	DQA1*01:02:02	DQB1*05:02:01:01	DRB1*16:01:01	30	13.51%
2	DQA1*02:01:01:01	DQB1*02:02:01:01	DRB1*07:01:01	11.82	5.32%
3	DQA1*05:01:01:02	DQB1*02:01:01	DRB1*03:01:01:01	11	4.96%
4	DQA1*01:02:01:01	DQB1*06:02:01:01	DRB1*15:01:01	10	4.50%
5	DQA1*01:01:01:01	DQB1*05:01:01:03	DRB1*01:01:01	9	4.05%
6	DQA1*01:01:01:03	DQB1*05:01:01:03	DRB1*01:01:01	7	3.15%
7	DQA1*03:01:01	DQB1*06:01:01/15	DRB1*15:02:01:01	7	3.15%
8	DQA1*01:03:01:01	DQB1*06:03:01:01	DRB1*13:01:01:01	7	3.15%
9	DQA1*03:01:01	DQB1*05:03:01	DRB1*14:54:01:02	6	2.70%
10	DQA1*01:03:01:02	DQB1*03:01:01:03	DRB1*11:04:01	5	2.25%
11	DQA1*01:04:01:01	DQB1*05:01:01:03	DRB1*01:01:01	4	1.80%
12	DQA1*05:05:01:09	DQB1*06:04:01	DRB1*13:02:01:02	4	1.80%
13	DQA1*01:01:01:02	DQB1*06:03:01:01	DRB1*13:01:01:02	4	1.80%
14	DQA1*01:02:01:04	DQB1*04:02:01:01	DRB1*08:01:01	4	1.80%
15	DQA1*01:03:01:02	DQB1*02:01:01	DRB1*03:01:01:03	4	1.80%
16	DQA1*04:01:01:03	DQB1*03:03:02:01	DRB1*07:01:01	3.82	1.72%
17	DQA1*05:01:01:03	DQB1*03:03:02:01	DRB1*07:01:01	3.18	1.43%
18	DQA1*02:01:01:02	DQB1*02:02:01:01	DRB1*07:01:01	3.18	1.43%
19	DQA1*02:01:01:01	DQB1*05:01:01:03	DRB1*01:01:01	3	1.35%
20	DQA1*02:01:01:02	DQB1*06:02:01:01	DRB1*15:01:01	3	1.35%
21	DQA1*01:01:01:05	DQB1*03:02:01:01	DRB1*04:01:01:03	3	1.35%
22	DQA1*01:02:01:05	DQB1*03:02:01:01	DRB1*04:02:01	3	1.35%
23	DQA1*03:01:01	DQB1*03:02:01:01	DRB1*04:03:01:01	3	1.35%
24	DQA1*05:05:01:02	DQB1*03:01:01:03	DRB1*11:01:01:01	3	1.35%
25	DQA1*05:05:01:04	DQB1*03:01:01:03	DRB1*11:01:01:01	3	1.35%
26	DQA1*05:05:01:05	DQB1*03:01:01:03	DRB1*13:03:01	3	1.35%
27	DQA1*05:05:01:09	DQB1*03:01:01:03	DRB1*11:01:01:03	3	1.35%

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-β3* 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-β3* 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-β2</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-β3</i>	rs3917165	<b>5.00*10<sup>-3</sup></b>	4.96	0.56	1.48
<i>TGF-β3</i>	rs3917200	<b>1.74*10<sup>-4</sup> §</b>	3.26	<b>9.52*10<sup>-6</sup> †</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-β3* 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.

Gene	SNP	Sarcoidosis (n = 103) vs. controls, (n = 100)	
		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-β2</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-β3</i>	rs3917165	0.11	2.31 [0.87–6.14]
<b><i>TGF-β3</i></b>	rs3917200	<b>2.76*10<sup>-7</sup> §</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-β3* 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-β3* 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

1. Dubrey S, Shah S, Hardman T, et al. Sarcoidosis: the links between epidemiology and aetiology. *Postgrad Med J.* 2014;90(1068):582–589.
2. 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.
3. 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.**
4. Mortaz E, Masjedi MR, Tabarsi P, et al. Immunopathology of sarcoidosis. *Iran J Allergy Asthma Immunol.* 2014;13(5):300–306.
5. Karakatsani A, Papakosta D, Rapti A, et al. Epidemiology of interstitial lung diseases in Greece. *Respir Med.* 2009;103(8):1122–1129.
6. Hunninghake GW, Costabel U, Ando M, et al. Statement on sarcoidosis. *Am J Respir Crit Care Med.* 1999;160(2):736–755.
7. 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.
8. 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.**
9. 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.**
10. 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.
11. 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.
12. 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.
13. 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.
14. Mrazek 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.
15. 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.**
16. Gilbert RWD, Vickaryous MK, Viloria-Petit AM. Signalling by transforming growth factor beta isoforms in wound healing and tissue regeneration. *J Dev Biol.* 2016;4(2):21.
17. 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.
18. 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.
19. Mrazek 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.
20. 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.
21. 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.
22. 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.
23. 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.
24. Ziegenhagen MW, Müller-Quernheim J. The cytokine network in sarcoidosis and its clinical relevance. *J Intern Med.* 2003;253(1):18–30.
25. 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)
26. 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.**
27. 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 class II variants defined by next generation sequencing are associated with sarcoidosis in Korean patients

Kateřina Sikorová<sup>1</sup>, Su-Jin Moon<sup>2</sup>, Hee-Young Yoon<sup>3</sup>, Adam Strnad<sup>4</sup>, Jin Woo Song<sup>2</sup>✉ & Martin Petrek<sup>1</sup>✉

Polymorphic genes with immune functions, namely those of the human leukocyte antigen (HLA) system, have been implicated in sarcoidosis pathogenesis. As HLA polymorphisms in sarcoidosis have not been yet investigated in the Korean population, we used next-generation sequencing (NGS), allowing detailed characterization of HLA alleles to investigate the role of HLA variation in Korean sarcoidosis patients. We enrolled 103 patients diagnosed by the ATS/ERS/WASOG guidelines at Asan Medical Centre, Seoul, Korea. Among those, genotyping of 7 HLA loci (HLA-A, -B, -C, -DQA1, -DQB1, -DRB1, -DPB1) was performed using Omixon Holotype™ kit and HLA-Twin software™. HLA allele frequencies were compared with frequency data on healthy Koreans from the allelic frequency databases, and 4-digit characteristics of HLA genotyping were used. Associations were assessed by two-tailed Fischer's exact test with correction for multiple comparisons. Variants previously associated with sarcoidosis risk (HLA-C\*03:04, HLA-DRB1\*12:01, HLA-DRB1\*14:54) and a known protective variant HLA-DPB1\*04:01, were associated with sarcoidosis in Koreans. Further, we suggest new HLA variants associated with sarcoidosis risk (e.g., HLA-DQA1\*05:08) and novel protective variants HLA-DQB1\*03:02 and HLA-DQA1\*01:02 in Koreans. This first study of HLA variation in Korean patients with sarcoidosis by precise genotyping methodology reports data that could serve future meta-analyses on HLA variation's role in sarcoidosis.

Altered immune response to the yet unknown antigen(s), influenced by genetic and environmental factors, is the hallmark of sarcoidosis pathogenesis. The presentation of sarcoidosis is remarkably diverse as it is the course of the disease. The disease presents in mild asymptomatic to progressive forms, which occur in 15–20% of patients<sup>1</sup>.

The occurrence and clinical course of sarcoidosis depend on ethnicity; it is higher and more severe in African Americans than in patients of European descent<sup>2,3</sup>. In the Asian population, a generally lower incidence of sarcoidosis was reported<sup>4</sup>; however, new cases have increased in the last few years—for example, in South Korea, with an incidence rate of 0.85 per 100,000<sup>5</sup>. Therefore, it is beneficial to distinguish between sarcoidosis phenotypes in their early stages for diagnostics, subsequent treatment, and general patient management. To accomplish this, it can be helpful to investigate the genetic profile of the patients in the context of their immunogenetic backgrounds, such as human leukocyte antigen (HLA) allele variants, HLA coding region genes, and other genes related to immune response and granuloma formation.

The key role of HLA, the most polymorphic human genetic system, in susceptibility to sarcoidosis and in modifying its clinical course has already been shown. However, most studies have been performed on populations of European or Afro-American descent<sup>6–8</sup>. However, in the East Asian region, only a limited number of studies on Japanese populations have been conducted<sup>9,10</sup>. To our knowledge, HLA polymorphisms have not yet been investigated in Korean sarcoidosis patients. Therefore, we focused this study on investigating polymorphic HLA alleles in sarcoidosis and its clinical phenotypes in Koreans.

<sup>1</sup>Department of Pathological Physiology & Institute of Molecular and Translational Medicine, Faculty of Medicine and Dentistry, Palacky University Olomouc, Olomouc, Czech Republic. <sup>2</sup>Department of Pulmonary and Critical Care Medicine, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Republic of Korea. <sup>3</sup>Division of Allergy and Respiratory Diseases, Soonchunhyang University Seoul Hospital, Seoul, Republic of Korea. <sup>4</sup>Department of Pathological Physiology, Faculty of Medicine and Dentistry, Palacky University Olomouc, Olomouc, Czech Republic. ✉email: jwsongasan@gmail.com; martin.petrek2@fnol.cz



Locus	Determined variants	Associated on primary level	Associated after correction
HLA-A	21	1	0
HLA-B	45	3	1
HLA-C	22	4	3
HLA-DRB1	41	6	2
HLA-DQA1	17	8	3
HLA-DQB1	19	7	3
HLA-DPB1	18	6	4

**Table 1.** Determined and associated alleles in seven HLA loci. Numbers of alleles that we were able to distinguish using NGS HLA genotyping and numbers of sarcoidosis associated alleles in seven HLA loci (primary level = before correction for multiple comparisons). *HLA* human leukocyte antigen.

HLA	OR [95% CI]	<i>p</i> -value	<i>p</i> <sub>corr</sub>
A*33:03	0.492 [0.294–0.821]	0.005	0.100
B*15:01	0.389 [0.193–0.783]	0.005	0.218
<b>B*15:07</b>	NA	0.001	<b>0.041</b>
B*58:01	0.285 [0.102–0.791]	0.008	0.290
<b>C*03:02</b>	0.163 [0.06–0.449]	$8.38 \times 10^{-06}$	<b><math>1.84 \times 10^{-04}</math></b>
<b>C*03:04</b>	2.949 [1.703–5.105]	$3.16 \times 10^{-04}$	<b>0.007</b>
C*05:01	2.97 [1.061–8.314]	0.046	0.643
<b>C*08:22</b>	NA	0.001	<b>0.02</b>

**Table 2.** HLA class I alleles associated with sarcoidosis in Korean patients (only the alleles associated at least on a primary level are shown). NA not analyzed (no variants were present in one of the analyzed groups), OR odds ratio, CI confidence interval, *HLA* human leukocyte antigen. Significant *p*<sub>corr</sub> (*p*-value after correction for multiple comparisons) in bold.

## Results

**Comparison of the distribution of HLA variants between patients and healthy population.** This study analyzed seven HLA loci at allele level in 103 patients with sarcoidosis from the Republic of Korea. HLA class I and class II allele frequencies in patients were compared with allele frequency data from healthy Korean population.

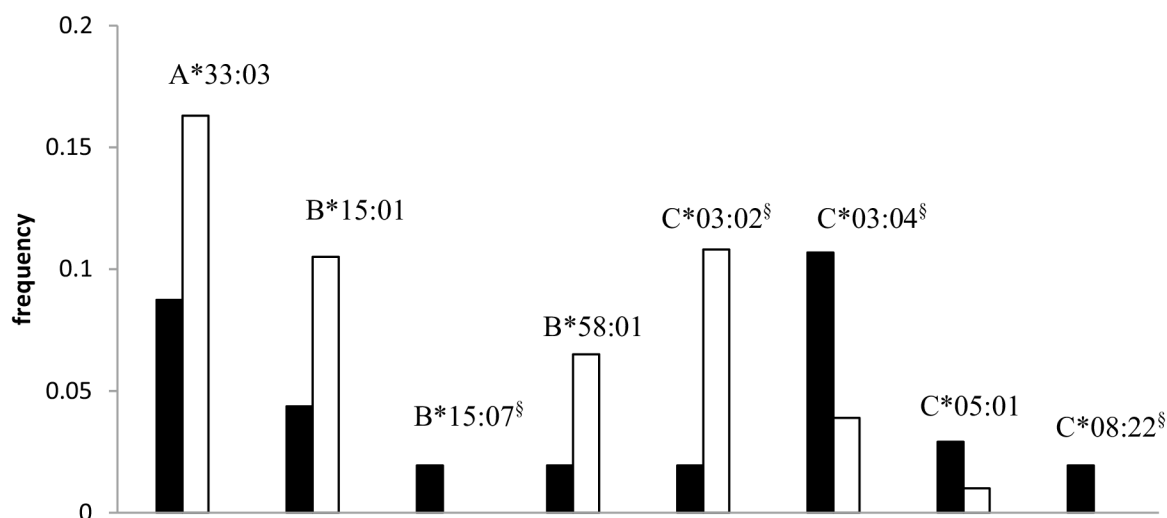
In the *HLA-A* locus, 21 different variants were determined by HLA genotyping (Table 1). One of these variants, HLA-A\*33:03, occurred more often in the controls than in the patients, showing a protective association significant on the primary level. In the *HLA-B* locus, 45 different variants were distinguished (Table 1), and two of them, HLA-B\*15:01 and B\*58:01, were associated at the primary level as protective for sarcoidosis (Table 2). Further, HLA-B\*15:07 was associated as a risk variant for sarcoidosis after correction for multiple comparisons (*p*<sub>corr</sub> = 0.041). In the *HLA-C* locus, 22 different variants were distinguished (Table 1). HLA-C\*05:01 was associated with sarcoidosis at a primary level as a risk variant. Protective variant HLA-C\*03:02 (*p*<sub>corr</sub> =  $1.84 \times 10^{-04}$ ) and risk variants HLA-C\*03:04 (*p*<sub>corr</sub> = 0.007) and HLA-C\*08:22 (*p*<sub>corr</sub> = 0.02) were associated after correction for multiple comparisons (Table 2, Fig. 1).

*HLA-DRB1* is the HLA locus most often mentioned in connection with sarcoidosis; 41 different variants in this locus were determined by HLA genotyping in our patient group (Table 1). Variants HLA-DRB1\*04:06, HLA-DRB1\*13:02, HLA-DRB1\*14:01 were protective and variant HLA-DRB1\*14:03 acted as a risk allele for sarcoidosis, associations being significant on a primary level. After correction for multiple comparisons, alleles HLA-DRB1\*14:54 (*p*<sub>corr</sub> =  $2.31 \times 10^{-05}$ ) and HLA-DRB1\*12:01 (*p*<sub>corr</sub> =  $9.12 \times 10^{-05}$ ) were identified as risk factors for sarcoidosis (Table 3, Fig. 2).

Regarding the *HLA-DQA1* locus, 17 variants resulted from HLA genotyping (Table 1). HLA-DQA1\*05:03, HLA-DQA1\*03:01, HLA-DQA1\*05:01, and HLA-DQA1\*01:01 were associated on the primary level as protective for sarcoidosis. Variant HLA-DQA1\*05:06 was associated with the risk for sarcoidosis on the primary level. After correction for multiple comparisons, HLA-DQA1\*01:02 (*p*<sub>corr</sub> =  $3.64 \times 10^{-05}$ ) was found as protective and HLA-DQA1\*05:07 (*p*<sub>corr</sub> = 0.003) and HLA-DQA1\*05:08 (*p*<sub>corr</sub> =  $4.11 \times 10^{-05}$ ) as risk factors (Table 3).

In the *HLA-DQB1* locus, 19 variants were determined (Table 1). Variants HLA-DQB1\*03:03, HLA-DQB1\*04:01, and HLA-DQB1\*04:02 were associated as protective on the primary level, and variant HLA-DQB1\*05:03, a risk factor for sarcoidosis. After correction for multiple comparisons, HLA-DQB1\*03:02 (*p*<sub>corr</sub> =  $2.25 \times 10^{-05}$ ) and HLA-DQB1\*06:05 (*p*<sub>corr</sub> = 0.029) were found as protective for sarcoidosis while risk variant was HLA-DQB1\*06:09 (*p*<sub>corr</sub> = 0.015); see Table 3.

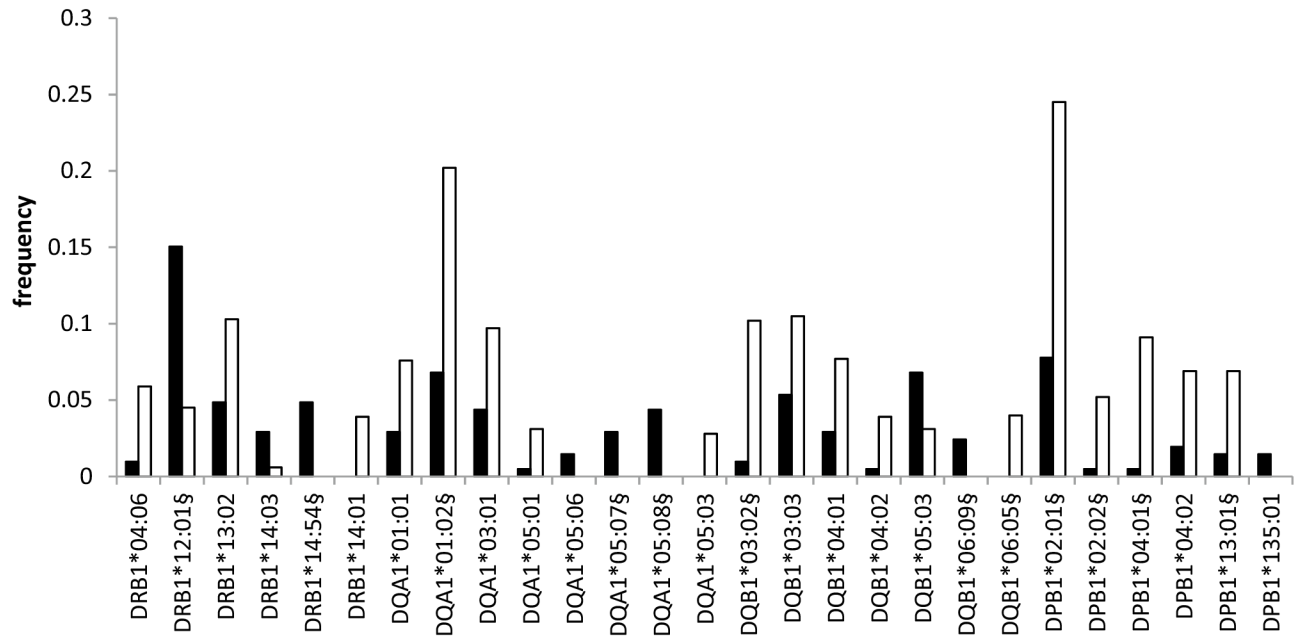
In the *HLA-DPB1* locus, 18 different variants were characterized (Table 1). Variant HLA-DPB1\*04:02 was considered protective on primary level and variant HLA-DPB1\*135:01 associated with risk. After correction for multiple comparisons, variants HLA-DPB1\*02:01 (*p*<sub>corr</sub> =  $5.92 \times 10^{-07}$ ), HLA-DPB1\*02:02 (*p*<sub>corr</sub> = 0.017),



**Figure 1.** Frequencies of HLA class I alleles associated with sarcoidosis in Koreans. Patients—black columns, Control population—white columns. §Significant  $p_{corr}$ .

HLA	OR [95% CI]	<i>p</i> -value	<i>p</i> <sub>corr</sub>
DRB1*04:06	0.157 [0.037–0.654]	0.002	0.080
<b>DRB1*12:01</b>	3.754 [2.204–6.394]	$2.22 \times 10^{-06}$	<b><math>9.12 \times 10^{-05}</math></b>
DRB1*13:02	0.445 [0.224–0.881]	0.017	0.506
DRB1*14:03	4.955 [1.371–17.904]	0.016	0.483
<b>DRB1*14:54</b>	NA	$5.63 \times 10^{-07}$	<b><math>2.31 \times 10^{-05}</math></b>
DRB1*14:01	NA	0.001	0.056
DQA1*01:01	0.365 [0.154–0.865]	0.015	0.223
<b>DQA1*01:02</b>	0.288 [0.162–0.512]	$2.14 \times 10^{-06}$	<b><math>3.64 \times 10^{-05}</math></b>
DQA1*03:01	0.425 [0.207–0.87]	0.014	0.217
DQA1*05:01	0.152 [0.021–1.42]	0.037	0.474
DQA1*05:06	NA	0.014	0.211
<b>DQA1*05:07</b>	NA	$1.86 \times 10^{-04}$	<b>0.003</b>
<b>DQA1*05:08</b>	NA	$2.42 \times 10^{-06}$	<b><math>4.11 \times 10^{-05}</math></b>
DQA1*05:03	NA	0.01	0.161
<b>DQB1*03:02</b>	0.086 [0.021–0.356]	$1.18 \times 10^{-06}$	<b><math>2.25 \times 10^{-05}</math></b>
DQB1*03:03	0.481 [0.249–0.928]	0.027	0.404
DQB1*04:01	0.36 [0.152–0.851]	0.015	0.244
DQB1*04:02	0.12 [0.016–0.891]	0.01	0.169
DQB1*05:03	2.278 [1.13–4.592]	0.024	0.369
<b>DQB1*06:09</b>	NA	0.001	<b>0.015</b>
<b>DQB1*06:05</b>	NA	0.002	<b>0.029</b>
<b>DPB1*02:01</b>	0.259 [0.151–0.445]	$3.28 \times 10^{-08}$	<b><math>5.92 \times 10^{-07}</math></b>
<b>DPB1*02:02</b>	0.089 [0.012–0.654]	0.001	<b>0.017</b>
<b>DPB1*04:01</b>	0.049 [0.007–0.354]	$1.08 \times 10^{-06}$	<b><math>1.94 \times 10^{-05}</math></b>
DPB1*04:02	0.267 [0.095–0.753]	0.005	0.093
<b>DPB1*13:01</b>	0.199 [0.061–0.649]	0.002	<b>0.027</b>
DPB1*135:01	NA	0.014	0.222

**Table 3.** HLA class II alleles associated with sarcoidosis in Korean patients (only the alleles associated at least on a primary level are shown). NA not analyzed (no variants were present in one of the analyzed groups), OR odds ratio, CI confidence interval, HLA human leukocyte antigen. Significant  $p_{corr}$  (*p*-value after correction for multiple comparisons) in bold.



**Figure 2.** HLA Frequencies of HLA class II alleles associated with sarcoidosis in Koreans. Patients—black columns, Control population—white columns. \$Significant  $p_{corr}$ .

HLA1	HLA2	$p$ -value
B*58:01	C*03:02	$2.09 \times 10^{-07}$
DQA1*05:07	DRB1*14:03	$6.21 \times 10^{-10}$
DQA1*05:08	DRB1*12:01	$3.33 \times 10^{-06}$

**Table 4.** Linkage disequilibrium between HLA alleles observed to be associated with sarcoidosis in Koreans. HLA human leukocyte antigen.

HLA-DPB1\*04:01 ( $p_{corr} = 1.94 \times 10^{-05}$ ), and HLA-DPB1\*13:01 ( $p_{corr} = 0.027$ ) were found protective for sarcoidosis (Table 3).

Linkage disequilibrium between the alleles associated with sarcoidosis in Korean patients was also calculated. Non-randomly associated variants in HLA class II, HLA-DQA1\*05:08–HLA-DRB1\*12:01 ( $p = 3.33 \times 10^{-06}$ ) and HLA-DQA1\*05:07–HLA-DRB1\*14:03 ( $p = 6.21 \times 10^{-10}$ ) were found. The pair of linked variants in HLA class I, HLA-B\*58:01–HLA-C\*03:02 ( $p = 2.09 \times 10^{-07}$ ), was observed (Table 4).

### Comparison of the distribution of HLA variants in the context of different chest radiography stages.

The association of HLA polymorphisms with CXR stage of the disease was analyzed. Patients were divided into two groups—the first group included 30 patients with CXR stage 0–1, and the second group included 73 patients with the more advanced CXR stages 2–4. HLA frequencies observed in the groups were compared with those of the control population. There were 12 variants associated with CXR stage 0–1 on primary level and two of them remained after correction for multiple comparisons (HLA-DRB1\*12:01  $p_{corr} = 0.033$ ; HLA-DQA1\*05:08  $p_{corr} = 0.001$ ) (Supplementary Table S1). The comparison between group CXR stages 2–4 and healthy controls resulted in 33 variants significantly associated with sarcoidosis at the primary level. After correction for multiple comparisons eight variants remained associated with more advanced forms of disease (HLA-B\*15:07  $p_{corr} = 0.013$ , HLA-C\*03:04  $p_{corr} = 0.007$ , HLA-C\*08:22  $p_{corr} = 0.006$ , HLA-DRB1\*12:01  $p_{corr} = 0.003$ , HLA-DRB1\*14:54  $p_{corr} = 4.57 \times 10^{-05}$ , HLA-DQA1\*05:07  $p_{corr} = 0.001$ , HLA-DQA1\*05:08  $p_{corr} = 0.003$ , HLA-DQB1\*06:09  $p_{corr} = 0.004$ ) and seven were considered protective (HLA-A\*33:03  $p_{corr} = 0.038$ , HLA-C\*03:02  $p_{corr} = 0.005$ , HLA-DQA1\*01:02  $p_{corr} = 0.001$ , HLA-DQB1\*03:02  $p_{corr} = 3.86 \times 10^{-04}$ , HLA-DPB1\*02:01  $p_{corr} = 2.12 \times 10^{-05}$ , HLA-DPB1\*02:02  $p_{corr} = 0.021$ , HLA-DPB1\*04:01  $p_{corr} = 0.002$ ) (Supplementary Table S2). Mutual comparison between CXR stage 0–1 patients ( $n = 30$ ) and CXR stages 2–4 patients ( $n = 73$ ) showed HLA-B\*51:01 allele being associated with CXR 2–4 at the primary level ( $p = 0.0421$ ); however, it did not attain significance after correction for multiple comparisons.

**HLA variants in patients with an extrapulmonary manifestation of sarcoidosis.** Distribution of HLA polymorphisms in patients with extrapulmonary manifestation of sarcoidosis were compared with data from healthy controls. In this comparison, 30 variants were associated with extrapulmonary manifestation at

the primary level (Supplementary Table S3). Eleven of them were significant also after correction for multiple comparisons, six as protective to extrapulmonary manifestation (HLA-A\*33:03  $p_{corr} = 0.038$ , HLA-C\*03:02  $p_{corr} = 0.019$ , HLA-DQA1\*01:02  $p_{corr} = 0.003$ , HLA-DQB1\*03:02  $p_{corr} = 3.86 \times 10^{-04}$ , HLA-DPB1\*02:01  $p_{corr} = 5.98 \times 10^{-06}$ , HLA-DPB1\*04:01  $p_{corr} = 0.002$ ) and five as risk factors (HLA-B\*15:07  $p_{corr} = 0.013$ , HLA-C\*08:22  $p_{corr} = 0.047$ , HLA-DRB1\*14:54  $p_{corr} = 0.008$ , HLA-DQA1\*05:08  $p_{corr} = 0.003$ , HLA-DQB1\*06:09  $p_{corr} = 0.004$ ). When mutual comparison of HLA frequencies in patients with and without extrapulmonary involvement was performed, four associations were revealed at the primary level (Supplementary Table S4); however, these did not attain significance after the correction.

**HLA variants in the context of clinical disease course (2 years follow up).** The patients were followed up at least 2 years after diagnosis. They were divided into three groups depending on the disease course as follows: patients whose disease had gone into remission, patients with stable disease, and patients whose disease progressed to a worse state. Next, the distribution of HLA in these three groups was compared with control population data. Two HLA variants are associated with a progressive, worsening course only at the primary level (Supplementary Table S5). Importantly, 39 variants were associated with disease improvement; ten of them remained significant after correction for multiple comparisons (Supplementary Table S6). Out of these ten stringently associated variants, five were associated with disease improvement (HLA-DRB1\*12:01  $p_{corr} = 0.006$ , HLA-DRB1\*14:54  $p_{corr} = 0.003$ , HLA-DQA1\*05:07  $p_{corr} = 1.86 \times 10^{-04}$ , HLA-DQA1\*05:08  $p_{corr} = 0.001$ , HLA-DQB1\*06:09  $p_{corr} = 0.010$ ); remaining five variants were more frequent in population control group. In the group of patients with stable disease, 14 variants were associated on the primary level, and three of them attained significance after correction for multiple comparisons (Supplementary Table S7). Two of these three variants were associated with a stable disease course (HLA-DRB1\*08:03  $p_{corr} = 0.002$  and HLA-DRB1\*14:54  $p_{corr} = 2.08 \times 10^{-04}$ ); HLA-DPB1\*02:01 was more frequent in the control group.

## Discussion

This study focused on immunogenetic profiling of Korean patients with sarcoidosis. It explored if variations in polymorphic HLA genes are related either to the disease itself or to several aspects of its clinical course. We report variants associated with the risk of sarcoidosis such as HLA-C\*03:04, HLA-DRB1\*12:01, HLA-DRB1\*14:54, and a protective variant HLA-DPB1\*04:01, previously described in other populations<sup>6,7,11,12</sup>, associated with sarcoidosis also in Koreans. Further, we suggest new HLA variants associated with sarcoidosis risk (e.g., HLA-DQA1\*05:08) and novel protective variants HLA-DQB1\*03:02 and HLA-DQA1\*01:02 in Koreans. Finally, we confirm the previous reports that variants HLA-DRB1\*12:01 and HLA-DPB1\*04:01, linked to better prognosis<sup>7,13</sup>, are also valid for Koreans. In this context, we report a new association of HLA-DPB1\*02:01 and -DPB1\*02:02 with better disease prognosis in Korean patients.

In sarcoidosis, a multi-systemic granulomatous disease affecting mainly the lungs, alveolar CD4+ T cells present an unknown antigen interacting with HLA class II molecules. This role for HLA II can be one of the reasons for the observation that most of the HLA variants associated with sarcoidosis, also in our Korean patients, belong to HLA class II loci<sup>3</sup>.

*HLA-DRB1* locus has been widely associated with autoimmune diseases, including sarcoidosis<sup>6,14</sup>. We observed that HLA polymorphisms within this class II locus were associated with sarcoidosis also in Koreans. First, the variant HLA-DRB1\*12:01 was detected as a risk factor more than threefold in the patients compared with the control population. In line with our finding, this variant was reported as susceptible in African American population, where it was associated with a doubled risk of disease<sup>6,7</sup>. We also observed that this variant was overrepresented in the patients who improved after 2 years of follow-up. Therefore, HLA-DRB1\*12:01 may be associated with a better prognosis of sarcoidosis in Koreans. Second, we determined HLA-DRB1\*14:54 as a risk factor for sarcoidosis in Koreans. Interestingly, from a limited number of studies on HLA in sarcoidosis in the East Asian population, Ishihara *et al.*<sup>11</sup> reported that sarcoidosis is associated with HLA-DRB1\*12:01 and also with alleles of the group HLA-DRB1\*14 in Japanese patients. Further, alleles of the group HLA-DRB1\*14, to which HLA-DRB1\*14:54 belongs, have been previously associated with the risk of the disease, extensive extrapulmonary involvement, and the prolonged disease course in the Turkish population<sup>15,16</sup>.

*HLA-DQB1* was another HLA class II locus containing sarcoidosis-associated variant in Koreans, specifically HLA-DQB1\*03:02 as a protective factor. This variant has not been mentioned in connection with sarcoidosis yet; nevertheless, it may have an important role in the immune response as a protective impact of this variant was mentioned in the context of HIV infection<sup>17</sup>.

*HLA-DQA1* locus has not been mentioned in connection with sarcoidosis very often. In our study, there was a linkage disequilibrium between *HLA-DQA1* alleles (e.g., HLA-DQA1\*05) and associated *HLA-DRB1* variants, which could be the reason for the observed associations. Therefore, in the context of the disease course, we can nominate HLA-DQA1\*01:02 as a protective factor against sarcoidosis in Koreans.

Regarding the *HLA-DPB1* locus, our observation of the over-representation of HLA-DPB1\*04:01 in the Korean control population agrees with reports from Finnish sarcoidosis patients, where HLA-DPB1\*04:01 was identified as a protective marker<sup>12</sup>. Further, when combined in the haplotype with HLA-DRB1\*04:01, it was characteristic for spontaneous sarcoidosis resolution within 2 years<sup>13</sup>. In a wider context, the HLA-DPB1\*04:01 variant was suggested as protective also in other autoimmune diseases, such as celiac disease<sup>18</sup>. In addition to HLA-DPB1\*04:01, we characterized variants HLA-DPB1\*02:01 and HLA-DPB1\*02:02 as protective factors in Korean sarcoidosis patients.

Finally, regarding the *HLA-DPB1* locus, in our Korean cohort, HLA-DPB1\*02:02 could be associated with a better sarcoidosis prognosis, as it did not occur in patients with CXR stage 2–4. In line with this speculation,

HLA-DPB1\*02:02 was associated with a better prognosis in another autoimmune condition—Graves' disease in the Japanese population<sup>19</sup>.

HLA class I loci (*HLA-A*, *-B*, *-C*) have been previously rarely mentioned in connection with sarcoidosis and autoimmune diseases. In cases where HLA class I associations with sarcoidosis have been reported, either linkage disequilibrium between HLA class I and II or independent action in the context of immune response to intracellular mycobacterial infection were suggested for explanation<sup>20,21</sup>. As a protective factor and a risk factor, HLA-C\*03:02 and HLA-C\*03:04 were associated with sarcoidosis in Koreans, respectively. Moreover, HLA-C\*03:04 was related to more advanced CXR stages and a worse prognosis. Another class I variant, HLA-A\*33:03, was associated with protection from extrapulmonary involvement; this variant was previously observed with a frequency higher than 10% in the overall Korean population<sup>22</sup>.

Our study has limitations. Due to the smaller number of patients, some of the observed associations, especially in comparisons between clinical subgroups, could arise by chance and, therefore, should not be over-interpreted before independent replication. Ideally, the patient data should be compared with the data from a real, well-matched control population instead of the data from population databases. Although we could not use the preferred option due to pandemic limitations, the allelefrequencies.net database, from where we retrieved the control data for this study, has been considered an authoritative and reputable source representative of particular local/national populations<sup>23</sup>. Further, a high degree of ethnic homogeneity in the South Korean population<sup>24</sup> is rather supporting than contradicting, using this kind of population control data for HLA allele frequency comparisons. However, we acknowledge that we could not use an ideally matched control population. Therefore, the observed associations should be interpreted cautiously until their verification in another cohort of Korean patients.

By contrast, there are important positive aspects of our study. So far, there have been only a few reports investigating the role of polymorphisms in HLA loci in sarcoidosis, where high resolution (i.e., very precise) genotyping methods were used<sup>25</sup>. However, the greatest strength of our study is that HLA in sarcoidosis has not been investigated in the Korean population at all, irrespective of precision level. Thus, our report provides the first data on the distribution of HLA variants in patients with sarcoidosis of this particular ethnicity, which is relevant to the importance of the ethnic concept of immunogenetics investigations in sarcoidosis<sup>26</sup>. Despite the above-discussed limitations, hereby reported high-resolution data can serve as a basis for future meta-analyses on the role of HLA variation in sarcoidosis.

## Conclusions

In this study, the first one about HLA polymorphisms and sarcoidosis in the Korean population, we extend associations between some HLA variants and sarcoidosis previously reported in other populations to Koreans. We suggest other new disease-associated variants, especially in the HLA class II loci, defined by NGS precise genotyping in our Korean patient cohort. However, before these associations' plausible clinical significance or functional meaning can be investigated further, confirmatory (i.e., replication) studies in another Korean sarcoidosis patient cohort(s) are necessary. In the meantime, the reported data can be, however, useful for meta-analytical approaches to sarcoidosis genetics across ethnicities.

## Methods

**Patients.** For this case–control study, 103 sarcoidosis cases (mean age  $\pm$  SD, 46.9  $\pm$  12.9 years; age range, 27–71 years; male [M]: female [F] ratio, 24: 79) were enrolled. All subjects were unrelated original inhabitants of South Korea diagnosed according to the ATS, ERS, and WASOG statement on sarcoidosis<sup>27,28</sup> at Asan Medical Centre, Seoul, the Republic of Korea, from 17.5.1994 to 31.5.2019.

Sarcoidosis cases were followed for 1.73–26.76 years after their diagnosis (median 11.95 years). After 2 years from the diagnosis, the patients were evaluated for the disease status as follows: improved (n = 58), stable disease (n = 32), and patients with progression of the disease (n = 5) for the remaining eight patients this information was missing due to loss to follow-up, death of the patients or other reasons. Concerning the treatment, 64 patients were treated with corticosteroids, in some cases in combination with methotrexate (n = 12), azathioprine (n = 4), or hydroxychloroquine sulphate (n = 4), and in one case in combination with mycophenolate mofetil. In two cases, the treatment with methotrexate was used independently without a combination of corticosteroids, and one case was treated with hydroxychloroquine sulphate only. Therefore, 36 patients did not receive any treatment.

Patients were also evaluated based on the Chest radiography stages (thoracic involvement) using Scadding's criteria: stage 1 which involved bilateral hilar lymphadenopathy (BHL) only (n = 28), stage 2 for BHL with pulmonary infiltrates (n = 64), stage 3 with parenchymal infiltrates only (n = 6), stage 4 with pulmonary fibrosis (n = 3); two patients (stage 0) had normal radiograph.

Regarding extrapulmonary involvement, other organs than lungs have also been affected in some patients—eyes (n = 21), extra pulmonary lymph nodes (n = 19), skin (n = 16), spleen (n = 11), heart (n = 7), liver (n = 7), musculoskeletal manifestation (n = 2), renal affection (n = 2), neurosarcoidosis (n = 2), pancreas (n = 1), and hypercalcemia was presented in 34 patients.

**NGS HLA analysis.** HLA genotypes for HLA-A, -B, -C, -DRB1, -DQA1, -DQB1, and -DPB1 loci were determined by high-resolution NGS technique using Omixon Hologate HLA kits configuration 96/7 v2 (Omixon Biocomputing Ltd, Budapest, Hungary) for library preparation and subsequent sequencing on Illumina MiSeq next-generation sequencing platform.

We worked with DNA isolated at Asan Medical Centre, Seoul, Korea. Blood samples were collected in EDTA-coated blood collection tubes (BD Vacutainer No. 367844). DNA in the blood was extracted using QIAamp DNA mini and Blood Kit (Qiagen No. 51306).

The HLA class I and class II loci were amplified by long-range PCR in separate samples in 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 (35 ng/µl). A combined DQB1 enhancer (5.6 µl/sample) was added to the DQB1 master mix. 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, followed by 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 incubation at 68 °C for 10 min. Amplicon presence was validated by 2% agarose gel electrophoresis. DNA was quantified using QuantiFluor fluorescent dsDNA staining system (Promega, Madison, Wisconsin, USA) and EnSpire Multimode plate reader (PerkinElmer, Waltham, MA, USA).

All seven amplicons from each sample were pooled into a final 35 µl volume on a new 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).

The next step was library preparation. The pooled amplicons were fragmented, followed by end repair and ligation with sample-specific indexed adaptors. 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 clean-up. Pooled library fragments ranging between 650 and 1300 bp in size were subsequently selected on Agencourt AMPure XP beads. 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 to 20 pM concentrations. Prior to sequencing, the library was diluted to a 2 nM concentration, denatured by NaOH, diluted with hybridization buffer, and the 9 pM library was loaded on MiSeq (Illumina, San Diego, CA, USA) and sequenced in a single 300-cycle (V2) paired-end sequencing run. Collected reads were exported in .fastq format.

Allele assignment for all seven loci was performed with the Omixon Twin software v4.2.0 and the IPD-IMGT/HLA database Release 3.39.0 (July 2020).

The laboratory where the HLA typing took place has a consistent record of passing international external proficiency testing (quality control) for HLA DNA genotyping since 2014. All experiments, from DNA extraction to the determination of HLA genotypes, were performed in accordance with relevant guidelines and regulations.

**Statistical analysis of the patient genotyping data.** Comparison with a healthy population from Korea in HLA-A, -B, -C (HLA-class I) and HLA-DQA1, -DQB1, -DRB1, -DPB1 (HLA-class II) loci was performed. HLA allele frequencies obtained in patients were compared with frequencies of alleles determined in the corresponding loci in a healthy population from Korea. For HLA class I, population one from South Korea was selected as a healthy population, whereas for HLA class II, population three was used. Data for both populations were retrieved from the database allele frequencies.net ([http://allelefrequencies.net/pop6001c.asp?pop\\_id=1336](http://allelefrequencies.net/pop6001c.asp?pop_id=1336); [http://allelefrequencies.net/pop6001c.asp?pop\\_id=1870](http://allelefrequencies.net/pop6001c.asp?pop_id=1870))<sup>23</sup>.

The association of HLA variants typed by NGS was validated by the two-tailed Fisher exact probability test to estimate the p-value, odds ratio (OR), and 95% confidence interval (CI). Correction for multiple comparisons was performed according to the following formula:  $1 - (1 - p)^n$ , where n = count of distinguished alleles in the loci. The statistical analysis was performed using <http://vassarstats.net/odds2x2.html>, <https://www.graphpad.com/quickcalcs/contingency2/>, and <https://www.hiv.lanl.gov/content/immunology/hla/index.html>.

The linkage disequilibrium calculation was also performed by software <https://www.hiv.lanl.gov/content/immunology/hla/index.html>. The number of two-way tests performed was 9636 ( $N_2 = 9636$ ). The p-value for two-way comparisons as  $0.05/N_2 = 5.189 \times 10^{-6}$  for a 95% confidence level was calculated. As significant, results with p-values less than  $5.189 \times 10^{-6}$  were considered.

**Ethical approval.** Patients provided informed consent with participation in the study, which was approved by the institutional review board (Institutional Review Board of Asan Medical Center IRB No—2018-2170).

### Data availability

The complete data, which served as the basis for analyses presented in the manuscript, including the data on the distribution of all detected HLA alleles in the investigated Korean patients with sarcoidosis (Supplementary Table S8), are provided within the Supplementary Information file.

Received: 27 January 2022; Accepted: 23 May 2022

Published online: 03 June 2022

### References

- Calender, A., Weichhart, T., Valeyre, D. & Pacheco, Y. Current insights in genetics of sarcoidosis: Functional and clinical impacts. *J. Clin. Med.* **9**(8), 2633 (2020).
- Arkema, E. V. & Cozier, Y. C. Epidemiology of sarcoidosis: Current findings and future directions. *Therap. Adv. Chronic Dis.* **9**(11), 227–240 (2018).
- Judson, M. A. A sarcoidosis clinician's perspective of MHC functional elements outside the antigen binding site. *Hum. Immunol.* **80**(1), 85–89 (2019).
- Jeon, M. H. *et al.* The incidence, comorbidity and mortality of sarcoidosis in Korea, 2008–2015: A nationwide population-based study. *Sarcoidosis Vasculitis Diffuse Lung Dis.* **37**(1), 24 (2020).
- Yoon, H. Y., Kim, H. M., Kim, Y. J. & Song, J. W. Prevalence and incidence of sarcoidosis in Korea: A nationwide population-based study. *Respir. Res.* **19**(1), 1–8 (2018).

6. Rossman, M. D. *et al.* HLA-DRB1\*1101: A significant risk factor for sarcoidosis in blacks and whites. *Am. J. Hum. Genet.* **73**(4), 720–735 (2003).
7. Levin, A. M. *et al.* Association of HLA-DRB1 with sarcoidosis susceptibility and progression in African Americans. *Am. J. Respir. Cell Mol. Biol.* **53**(2), 206–216 (2015).
8. Rivera, N. V. *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.* **193**(9), 1008–1022 (2016).
9. Sato, H. *et al.* Sarcoidosis HLA class II genotyping distinguishes differences of clinical phenotype across ethnic groups. *Hum. Mol. Genet.* **19**(20), 4100–4111 (2010).
10. Meguro, A. *et al.* Genetic control of CCL24, POR, and IL23R contributes to the pathogenesis of sarcoidosis. *Commun. Biol.* **3**(1), 1–10 (2020).
11. Ishihara, M. *et al.* Molecular genetic studies of HLA class II alleles in sarcoidosis. *Tissue Antigens* **43**(4), 238–241 (1994).
12. Lahtela, E., Kankainen, M., Sinisalo, J., Selroos, O. & Lokki, M. L. Exome sequencing identifies susceptibility loci for sarcoidosis prognosis. *Front. Immunol.* **10**, 2964 (2019).
13. Lahtela, E., Wolin, A., Pietinalho, A., Lokki, M. L. & Selroos, O. Disease marker combination enhances patient characterization in the Finnish sarcoidosis patients. *Respir. Med.* **132**, 92–94 (2017).
14. Grunewald, J. *et al.* T-cell receptor–HLA-DRB1 associations suggest specific antigens in pulmonary sarcoidosis. *Eur. Respir. J.* **47**(3), 898–909 (2016).
15. Yanardag, H., Tetikkurt, C., Bilir, M. & Yilmaz, E. Association of HLA antigens with the clinical course of sarcoidosis and familial disease. *Monaldi Arch. Chest Dis.* <https://doi.org/10.4081/monaldi.2017.835> (2017).
16. Esendagli, D., Ozmen, F., Koksall, D., Onder, S. & Emri, S. Association of class II human leukocyte antigen (HLA) alleles with pulmonary sarcoidosis. *Sarcoidosis Vasculitis Diffuse Lung Dis.* **35**(2), 143 (2018).
17. Rallon, N. *et al.* Human leucocyte antigen (HLA)-DQB1\*03:02 and HLA-A\*02:01 have opposite patterns in their effects on susceptibility to HIV infection. *HIV Med.* **18**(8), 587–594 (2017).
18. Hadley, D. *et al.* HLA-DPB1\*04:01 protects genetically susceptible children from celiac disease autoimmunity in the TEDDY study. *Am. J. Gastroenterol.* **110**(6), 915 (2015).
19. Takahashi, M., Yasunami, M., Kubota, S., Tamai, H. & Kimura, A. HLA-DPB1\*0202 is associated with a predictor of good prognosis of Graves' disease in the Japanese. *Hum. Immunol.* **67**(1–2), 47–52 (2006).
20. Rybicki, B. A. & Iannuzzi, M. C. Sarcoidosis and human leukocyte antigen class I and II genes: It takes two to tango? *Am. J. Respir. Crit. Care Med.* **169**(6), 665–666 (2004).
21. Tarasidis, A. & Arce, S. Immune response biomarkers as indicators of sarcoidosis presence, prognosis, and possible treatment: An immunopathogenic perspective. *Autoimmun. Rev.* **19**(3), 102462 (2020).
22. Lee, K. W., Oh, D. H., Lee, C. & Yang, S. Y. Allelic and haplotypic diversity of HLA-A,-B,-C,-DRB1, and-DQB1 genes in the Korean population. *Tissue Antigens* **65**(5), 437–447 (2005).
23. Gonzalez-Galarza, F. F. *et al.* Allele frequency net database. *Methods Mol. Biol.* **1802**, 49–62 (2018).
24. New World Encyclopedia Contributors. Demographics of South Korea. *New World Encyclopedia*. [https://www.newworldencyclopedia.org/p/index.php?title=Demographics\\_of\\_South\\_Korea&oldid=1038227](https://www.newworldencyclopedia.org/p/index.php?title=Demographics_of_South_Korea&oldid=1038227) (2020). (accessed date 20.4.2022)
25. Karakaya, B. *et al.* Bronchoalveolar lavage characteristics correlate with HLA tag SNPs in patients with Löfgren's syndrome and other sarcoidosis. *Clin. Exp. Immunol.* **196**(2), 249–258 (2019).
26. van Moersel, C. H., Petrek, M. & Rivera, N. V. Unravelling the genetic basis of sarcoidosis. In *Sarcoidosis (ERS Monograph)* (ed. Bonella, F.) 41–56 (European Respiratory Society, 2022).
27. Statement on Sarcoidosis. Joint Statement of the American Thoracic Society (ATS), the European Respiratory Society (ERS) and the World Association of Sarcoidosis and Other Granulomatous Disorders (WASOG) adopted by the ATS Board of Directors and ERS. *Am. J. Respir. Crit. Care Med.* **160**, 736–755. <https://doi.org/10.1164/ajrccm.160.2.ats4-99> (1999).
28. Kim, H. J., Kim, H. M. & Song, J. W. Clinical characteristics and outcomes of Korean patients with sarcoidosis. *Sci. Rep.* **11**(1), 1–8 (2021).

## Acknowledgements

This research was funded by IGAUP: LF\_2021\_014, IGAUP: LF\_2022\_005, NV18-05-00134, CZ.02.1.01/0.0/0.0/16\_019/0000868 (ENOCH), Basic Science Research Program (NRF-2019R1A2C2008541, NRF-2022R1A2B5B02001602) and the Bio & Medical Technology Development Program (NRF-2022M3A9E4082647) of the National Research Foundation of Korea (NRF) funded by the Ministry of Science & ICT Technology, Republic of Korea. The technical and logistical assistance of Ms. Gabriela Strakova is acknowledged.

## Author contributions

Idea and Study design: M.P. & J.W.S.; Patients recruitment and patient data: J.W.S. and H.Y.Y.; Sample processing for DNA: M.S.J.; HLA genotyping: K.S., A.S.; Statistical analysis: K.S., M.P.; Writing: K.S., M.P., J.W.S.; All authors have read and approved the submitted version of the M.S.; Responsible for the integrity of the study—clinical data: J.W.S., laboratory data: M.P.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-13199-w>.

**Correspondence** and requests for materials should be addressed to J.W.S. 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 licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022



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 ( $\leq 2500$  bp for class I and  $\leq 5000$  bp for class II genes) and/or low key exon coverage depth ( $\leq 30$ )]. 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).

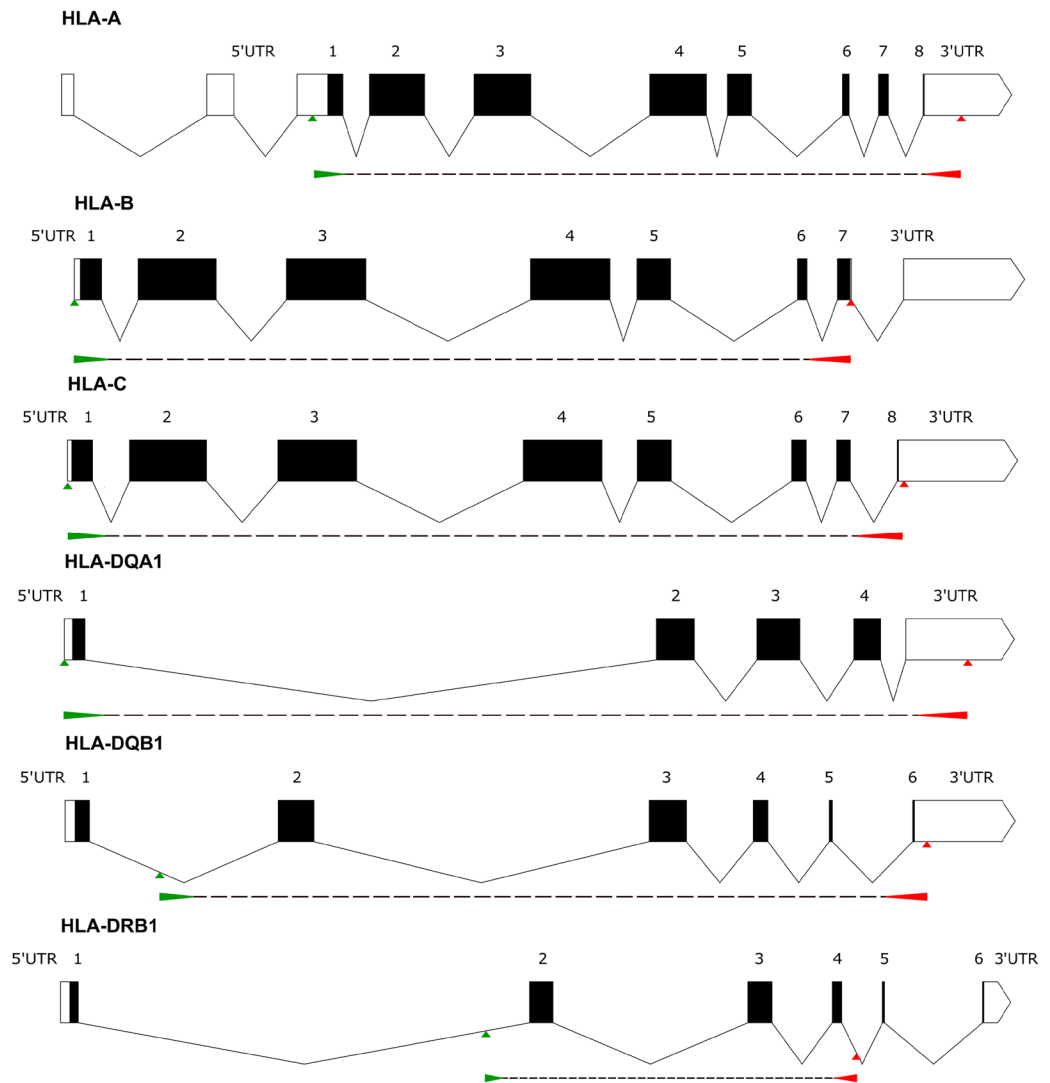
Ambiguous allele calls, which remained unresolvable due to inherent product limitations of the Omixon Holotype Kit (missing data on SNPs or INDEL variations within the unsequenced 3' UTR, 5' UTR and intron 1/exon 1 regions), were reported as ambiguous (i.e. DQB1\*06:01:01/15), or up to the third field 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  $\geq 10\%$ . The most commonly observed *HLA-A\*02:01:01* group alleles (32.43%) were followed by *HLA-A\*01:01:01* (12.16%), *HLA-A\*03:01:01* (11.71%), *HLA-A\*24:02:01* (11.71%) and *HLA-A\*11:01:01* (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.

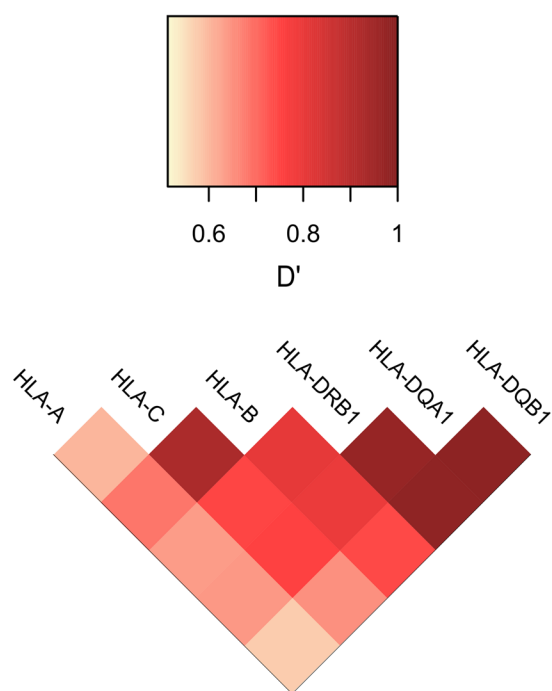


**Figure 2.** Schematic representation of HLA-A, -B, -C, -DQA1, -DQB1 and -DRB1 gene regions covered by Omixon Hotype PCR primers are depicted using Exon-Intron Graphic Maker v.4. (<http://wormweb.org/exonintron>)<sup>57</sup>. Start positions of forward and reverse primers are marked by green and red 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* (5.41%). Together with *HLA-DRB1\*07:01:01* (9.91%) and *HLA-DRB1\*11:01:01* (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* (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 CBMDR. 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\*11:01:01:01~C\*04:01:01:01~B\*35:01:01~DRB1\*01:01:01~DQA1\*01:01:01:01~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  $OD_{260}/OD_{280} > 1.8$  and  $OD_{260}/OD_{230} > 1.6$  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.allelefrequencys.net](http://www.allelefrequencys.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 ( $D^{52,53}$ ,  $W_n^{54}$ ) 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-Martín, 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 Republic 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., Genadieva Stavrik, S., Senev, A., Efinska 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