

Univerzita Palackého v Olomouci
Lékařská fakulta

Zpracovávání, vyhodnocení a interpretace genomických dat získaných masivním paralelním sekvenováním pro klinickou praxi

Disertační práce

Autor práce: Mgr. Petr Vrtěl
Vedoucí práce: doc. Mgr. Radek Vodička, Ph.D.

2023

Čestné prohlášení

Prohlašuji, že jsem tuto disertační práci vypracoval samostatně pod vedením doc. Mgr. Radka Vodičky, Ph.D. a že jsem všechny použité literární zdroje uvedl v seznamu literatury.

V Olomouci

Mgr. Petr Vrtěl

Poděkování

Děkuji všem, kteří mě během kompletování této práce pomáhali a podporovali. Obzvláště děkuji panu doc. Mgr. Radkovi Vodičkovi, Ph.D. za cenné rady a odborné vedení práce. Dále pak kolegům z Ústavu lékařské genetiky a Hemato-onkologické kliniky FN Olomouc za spolupráci a odborné konzultace v rámci dané problematiky. V neposlední řadě také děkuji své rodině za podporu a hlavně nekonečnou trpělivost.

Abstrakt

Sekvenační techniky druhé generace významně zvyšují testovací kapacitu a stávají se podstatnou součástí celého odvětví klinické genetiky. Díky možnosti současného vyšetření více pacientů na více cílů, umožňují nahrazení dříve užívaných molekulárně diagnostických metod, kterým byla zejména Sangerova sekvenace. Práce popisuje zavedení a zhodnocení optimálního diagnostického přístupu založeného na vysokokapacitním sekvenování v diagnostice vzácných trombofilních stavů – deficitu proteinu S (PS), proteinu C (PC) a antitrombinu (AT). Jedná se pravděpodobně o první aplikaci metodiky vysokokapacitní sekvenace u genetického testování deficitu těchto přirozených antikoagulantů v České republice.

Pro přípravu amplikonových knihoven a cílené sekvenování byla využita platforma Ion Torrent. Na základě definovaných indikačních kritérií byla selektována kohorta 31 pacientů. Pro deficit PS (7), PC (13) a AT (11). V rámci těchto skupin pacientů byla zaznamenána míra detekce mutací 67,7 %, nejvyšší ve skupině pacientů indikovaných pro deficit AT – 76,9 %. Bylo také odhaleno šest dosud nepopsaných poškozujících variant – *PROS1* c.1468delA (p.Ile490fs) a c.1931T>A (p.Ile644Asn); *PROC* c.450C>A (p.Tyr150Ter); c.715G>C (p.Gly239Arg) a c.866C>G (p.Pro289Arg); *SERPINC1* c.398A>T (p.Gln133Leu). Bylo také provedeno porovnání variant ze tří různých softwarů – Torrent Suite, Ion Reporter a NextGene u 10 kontrol. Pro data generovaná z platformy Ion Torrent se zdají být vhodnější software od stejného poskytovatele především z důvodu kvalitního odfiltrování falešně pozitivních variant.

Pozorovaný záchyt poškozujících variant je v souladu se závěry publikovaných studií, které vycházely zejména z časově náročného Sangerova sekvenování. Popisovaný diagnostický postup je tedy vhodný do klinické praxe. Jeho výhodou je také zvýšení testovací kapacity u této pravděpodobně poddiagnostikované skupiny pacientů. Popsaná metodika je navíc vhodná pro využití u dalších oligogenních panelů.

Klíčová slova: sekvenování DNA; vysokokapacitní sekvenování; deficit proteinu S; deficit proteinu C; deficit antitrombinu; míra detekce mutací

Abstract

Second-generation sequencing techniques significantly increase testing capacity within clinical settings and become an essential part of the entire field of clinical genetics. With the ability to test multiple patients on multiple targets simultaneously, they allow the replacement of previously used molecular diagnostic methods, which was primarily Sanger sequencing. This paper describes the introduction and evaluation of an optimal diagnostic approach based on high-throughput sequencing in the diagnosis of rare thrombophilic conditions - deficiency of protein S (PS), protein C (PC) and antithrombin (AT). This is probably the first application of high-throughput sequencing methodology in genetic testing for deficiency of these natural anticoagulants in the Czech Republic.

The Ion Torrent platform was used for the preparation of amplicon libraries and target sequencing. A cohort of 31 patients was selected based on defined indication criteria. For PS deficiency (7), for PC deficiency (13) and for AT (11). Within these patient groups, a mutation detection rate of 67.7% was observed, with the highest rate in the group of patients indicated for AT deficiency - 76.9%. Six previously undescribed damaging variants were also detected - *PROS1* c.1468delA (p.Ile490fs) and c.1931T>A (p.Ile644Asn); *PROC* c.450C>A (p.Tyr150Ter); c.715G>C (p.Gly239Arg) and c.866C>G (p.Pro289Arg); *SERPINC1* c.398A>T (p.Gln133Leu). Variants from three different software, Torrent Suite, Ion Reporter and NextGene, were also compared for 10 controls. For data generated from the Ion Torrent platform, software from the same provider seems to be more suitable, mainly because of the high quality filtering of false positive variants.

The observed detection of deleterious variants is in accordance with the conclusions of published studies, which were mainly based on time-consuming Sanger sequencing. The described diagnostic procedure is therefore suitable for clinical practice. It also allows to increase the testing capacity in this probably underdiagnosed group of patients. The described methodology is also suitable for use in other oligogenic panels.

Keywords: DNA sequencing; high-throughput sequencing; protein S deficiency; protein C deficiency; antithrombin deficiency; mutation detection rate

Obsah

1	Úvod.....	3
2	Teoretická část.....	5
2.1	Sekvenační techniky – historie	5
2.2	První generace sekvenování	6
2.2.1	Sangerova sekvenace	6
2.2.2	Maxam-Gilbert sekvenace.....	6
2.3	Druhá generace sekvenování	7
2.3.1	Masivní paralelní pyrosekvenování – Roche	8
2.3.2	Fluorescenční detekce – Illumina	9
2.3.3	Polovodičová technologie – Ion Torrent	10
2.3.4	Sekvenování po ligaci – SOLiD systém.....	11
2.3.5	Stolní sekvenátory	12
2.4	Třetí generace sekvenování.....	12
2.4.1	Single molecule real-time – Pacific Biosciences	13
2.4.2	Nanopórová technologie – Oxford Nanopore Technology	13
3	Pracovní postup NGS.....	14
3.1	Příprava knihoven.....	15
3.1.1	Metody založené na PCR	15
3.1.2	Metody založené na hybridizačním záchytu	16
3.2	Datová analýza	17
3.2.1	Primární analýza	18
3.2.2	Sekundární analýza.....	20
3.2.3	Terciální analýza	23
4	Aplikace metodiky vysokokapacitního sekvenování do klinické praxe pro vybrané onemocnění	27
4.1	Žilní tromboembolická nemoc.....	28
4.1.1	Faktory zvyšující riziko nástupu VTE.....	28
4.1.2	Genetické predispozice VTE	28
5	Materiál a metodika.....	33
5.1	Indikační kritéria.....	33
5.1.1	Funkční testy.....	34
5.2	Kohorta pacientů indikovaných na genetické testování	34

5.3	Izolace DNA.....	35
5.3.1	Měření koncentrace DNA.....	35
5.3.2	Sekvence na platformě Ion S5™	36
5.3.3	MLPA.....	41
5.3.4	Sangerova sekvence	42
5.4	Porovnání VCF vytvořených z rozdílných softwarů	44
5.5	Statistická analýza	44
6	Výsledky	45
6.1	Souhrnné výsledky u vyšetřených probandů	45
6.2	Výsledky u probandů s deficitem PS	50
6.2.1	Dosud nepopsané varianty u probandů s deficitem PS.....	51
6.3	Výsledky u probandů s deficitem PC	52
6.3.1	Dosud nepopsané varianty u probandů s deficitem PC	54
6.4	Výsledky u probandů s deficitem AT	55
6.4.1	Dosud nepopsané varianty u probandů s deficitem AT	57
6.5	Porovnání VCF vytvořených z rozdílných softwarů	57
6.5.1	NextGene.....	58
6.5.2	Ion Reporter	59
6.5.3	Torrent Suite.....	60
7	Diskuze	61
8	Závěr	67
9	Seznam použitých zkratk	68
10	Seznam publikovaných prací autora	71
11	Použitá literatura	73
12	Přílohy	87

1 Úvod

Rozvoj sekvenačních technik otevřel člověku netušené možnosti v poznávání genetické informace organismů žijících na Zemi. Umožnil podrobné studium jejich genomů a fylogeneze. Těchto dovedností je nyní využíváno od virologie, bakteriologie přes studium genomů vyšších organismů včetně člověka, kde sekvenační techniky umožnily pochopení patogeneze celé řady monogenních a polygenních onemocnění, jejichž genetická příčina zůstávala dlouho záhadou.

Zavedení sekvenačních technik první generace bylo průlomové, teprve však sekvenační techniky generace druhé umožnily, pro svou významně vyšší testovací kapacitu a rychlost, expanzi jak do výzkumu celých genomů, tak do rutinní diagnostiky genových panelů. Značná část geneticky podmíněných onemocnění vyžaduje pro úspěšné určení kauzální genetické varianty přečtení sekvence více genů. To nám tyto sekvenační techniky umožňují, ale musíme si být vědomi jejich limitů a danou metodiku řádně optimalizovat pro naše potřeby.

Téma této práce je spojeno se sekvenačními technikami druhé generace a jejich aplikací do klinické praxe. Práce podává přehled historie sekvenačních technik se zaměřením na sekvenační techniky druhé generace a algoritmy užívané při analýze dat získaných pomocí masivní paralelní sekvenace.

V praktické části práce se zaměřuji na aplikaci metodiky masivního paralelního sekvenování do vyhledávání vzácných mutací u trombofilních stavů, kde byla metoda aplikována a úspěšně zavedena do klinické praxe na našem pracovišti. Pravděpodobně se jedná o první aplikaci metodiky vysokokapacitního sekvenování do vyhledávání vzácných genetických variant u trombofilních stavů v České republice.

Cíle práce:

Poskytnout přehled využívaných sekvenačních technik

Popsat datovou analýzu

Demonstrovat zavedení masivně paralelní sekvenace do klinické praxe pro vrozený deficit proteinu S, C a antitrombinu

Porovnání posuzování variant (variant calling) ze tří běžně užívaných softwarů

2 Teoretická část

2.1 Sekvenační techniky – historie

Základním kamenem, který umožnil studium lidského genomu, byl v roce 1953 objev dvouřetězcové struktury deoxyribonukleové kyseliny (DNA) skládající se ze dvou purinových bází: adenin – A; guanin – G a dvou pyrimidinových bází: thymin – T; cytosin – C (1). Následovaly práce vedoucí k rozluštění genetického kódu organismů, které byly publikovány po roce 1960 (2, 3, 4, 5). V tomto období však nebyly známy techniky, které by určily pořadí bází v polynukleotidové sekvenci řetězce DNA. Rozluštění genetického kódu, znalost DNA struktury organismů a porozumění možné souvislosti změny na úrovni DNA sekvence s určitým fenotypem, akcentovalo důležitost vývoje sekvenačních technik.

První sekvenační pokusy se zaměřovaly na poměrně lehce dostupné ribonukleové kyseliny (RNA), především transferovou RNA, mikrobiální ribozomální RNA nebo jednovláčkovou RNA bakteriofágů. I přes nalezení vhodných RNA organismů a enzymů RNáz, umožňovala dostupná analytická chemie určit pouze složení ne však pořadí zkoumaných nukleotidů (6). První celou nukleotidovou sekvenci vytvořili za využití dostupných technik Holley et al. 1965 (7). Jednalo se o alanin transferovou RNA ze *Saccharomyces cerevisiae*. Ve stejném období Sanger et al. 1965 (8) vyvinuli revoluční techniku založenou na detekci radioaktivně značených částečně naštěpených fragmentů po dvourozměrné frakcionaci. Další adaptace této techniky již umožnily zaměřit pozornost také na DNA sekvenci. Ideální bylo pro tento účel využití bakteriofágů s DNA genomem. První výsledky přináší sekvenace bakteriofágu lambda. Wu and Kaiser 1968 (9) využili jeho 5' kohezních konců a funkce *Escherichia coli* DNA polymerázy k vyplnění konců syntetizovaného vlákna DNA radioaktivně značenými nukleotidy. Tyto byly doplňovány jeden po druhém. Částečně značená DNA mohla být dále štěpena nukleázami, radioaktivně značené nukleotidy izolovány a jejich sekvence tak mohla být stanovena (10). Tyto metody však byly velmi náročné a aplikovatelné jen na krátké úseky DNA. Z tohoto důvodu je masivní využívání DNA sekvenace spojováno až se sekvenačními protokoly A. Coulsona, F. Sangera, M. Maxama a W. Gilberta z druhé poloviny 70. let 20. století.

2.2 První generace sekvenování

Období je spojováno se Sangerovou a Maxam-Gilbertovou sekvenační technologií. Velmi efektivní se v těchto protokolech ukázalo zavedení jediné separace dle délky polynukleotidu pomocí elektroforézy na polyakrylamidovém gelu.

2.2.1 Sangerova sekvenace

Sanger and Coulson 1975 (11) publikovali práci, ve které představili nový „plus a minus“ sekvenační protokol. Revoluční pak byla publikace z roku 1977, kde byla představena tzv. Sangerova řetězec-ukončující nebo také dideoxy technika, která je v modifikované podobě využívána do dnešní doby (12). Metoda využívá jedno vlákno dvouřetězcové DNA jako templát pro nově syntetizované vlákno. Jedná se o modifikovanou replikaci, kde jsou do reakční směsi, kromě běžných deoxyribonukleosidtrifosfátů (dNTP), přidávány také modifikované dideoxyribonukleosidtrifosfáty (ddNTP). Tyto pak při svém zařazení ukončují elongaci nově vznikajícího řetězce DNA pro absenci 3' hydroxylové skupiny. Pokud pak do reakční směsi přidáme kromě běžných dNTP také radioaktivně značené ddNTP, získáme všechny možné délky DNA řetězce. Toho je dosaženo zařazením vhodného ddNTP, který ukončí elongaci nového vlákna DNA. Sekvence se určovala dle čtyř paralelních reakcí, kde každá reakce obsahovala jeden typ ddNTP báze. Následně jsou pak byly tyto čtyři sekvenační produkty separovány a detekovány na polyakrylamidovém gelu. Vylepšení této metody je spojováno především s nahrazením radioaktivního značení fluorescenčním a spojením čtyř paralelních reakcí do jedné. Potenciál pro automatizaci přinesl vývoj separační technologie – přechod od vertikální elektroforézy v polyakrylamidovém gelu ke kapilární elektroforéze v lineárním polyakrylamidu (6). Zavedení kapilárně elektroforetických systémů je spojováno především s firmou Applied Biosystems (ABI), která v roce 1995 přichází s automatickým sekvenátorem ABI Prism 370 (13).

2.2.2 Maxam-Gilbert sekvenace

V roce 1977 uvádí Maxam and Gilbert 1977 (14) metodiku sekvenace DNA založené na chemické degradaci bází. K určení nukleotidové sekvence tato metoda využívá radioaktivní

značení na konci DNA fragmentu a chemická činidla, která řetězec specificky štěpí. Např. puriny jsou štěpeny dimethylsulfátem, pyrimidiny pak hydrazinem. Modifikací využívaných chemikálií pak lze řetězec specificky štěpit: T+C, C, A+G, G. Fragменты ze čtyř reakcí jsou separovány na polyakrylamidovém gelu a detekovány autoradiografií.

Tato metoda byla velmi populární, protože nevyžadovala DNA polymerázu. Postupem času však byla upřednostněna modifikovaná Sangerova sekvenační metoda mimo jiné proto, že Maxam-Gilbertova sekvenace využívá ve svém protokolu toxické a radioaktivní chemikálie (13).

2.3 Druhá generace sekvenování

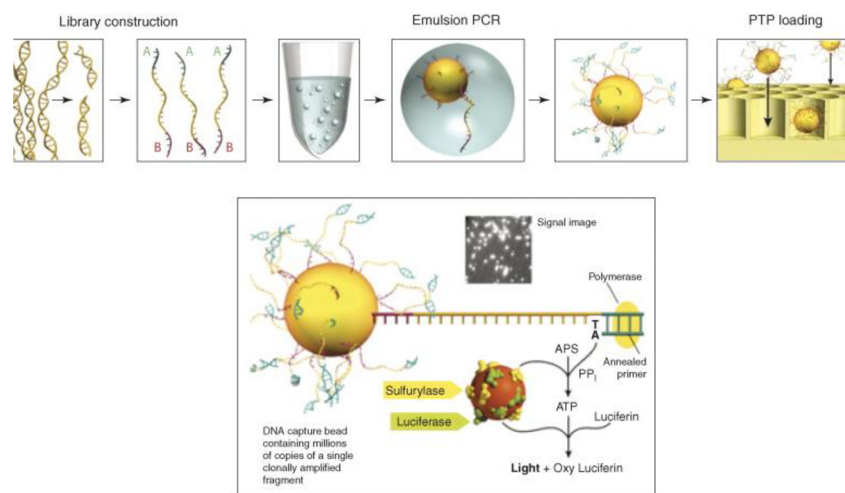
Od počátku 21. století se začínají objevovat vysokokapacitní sekvenační technologie, které změnily od základů přístup k výzkumu a diagnostice na poli genetiky. Pro tyto technologie se zažily termíny: Next generation sequencing (NGS), vysokokapacitní sekvenace, masivně paralelní sekvenování (MPS). Někdy jsou také nazývány sekvenačními technologiemi krátkého čtení, protože zde probíhá paralelně masivní sekvenování krátkých (250–800 bp), klonálně amplifikovaných DNA molekul (15).

Jejich základní charakteristikou je: generování mnoha miliónů krátkých čtení paralelně, přímá detekce sekvenačních produktů bez nutnosti elektroforézy, zrychlení procesu a snížení ceny sekvenace oproti sekvenačním technikám první generace (13). Zavádění těchto technologií, přináší mnohé výzvy stran zpracování sekvenačních dat, ukládání dat, jejich interpretace, etických aspektů a potřeby implementace dalších nezbytných oborů (bioinformatika aj.). Tato práce je zaměřena především na práci s daty získanými pomocí sekvenačních technik druhé generace.

Éra sekvenačních technik druhé generace se začala psát v roce 2000 kdy firma Lynx Therapeutics Company vyvinula technologii: „massively parallel signature sequencing on microbead arrays“, která využívala při čtení bází fluorescenční detekci (16). Jednalo se o průlom v sekvenační technologii, který umožnil nástup komerčně využívaným sekvenačním platformám.

2.3.1 Masivní paralelní pyrosekvenování – Roche

V roce 2004 je uvedena na trh první komerčně dostupná platforma 454 Life Sciences (později Roche). Tato využívá technologii pyrosekvenace, která umožňuje odvození sekvence měřením produkce pyrofosfátu při postupném promývání jednotlivých nukleotidů systémem přes templátovou DNA připevněnou na pevné fázi. Pyrosekvenace je tedy obdobně jako Sangerova metoda technikou tzv. sekvenování po syntéze „sequencing-by-synthesis“. Protože obě metody vyžadují přímou činnost DNA polymerázy k tvorbě pozorovatelného výstupu (6). Prvním krokem je příprava sekvenačních knihoven. Vzorky DNA jsou náhodně fragmentovány a k těmto fragmentům jsou navázány adaptéry. Každý z DNA fragmentů je specificky navázán na syntetické kuličky na jejichž povrchu jsou primery, které jsou specificky komplementární k adapterům. To umožní, že je každá kulička spojena s jedním fragmentem. Polymerázová řetězová reakce (PCR) probíhá na těchto kuličkách v olejové emulzi – emulzní PCR. Tímto způsobem je na povrchu každé kuličky amplifikováno přibližně 1 milion kopií každého fragmentu DNA. Po emulzní PCR jsou kuličky přeneseny na pikotitrační destičku / sklíčko, která obsahuje statisíce až miliony jamek. Každá kulička je umístěna do jedné jamky a následně je omývána směsí jednotlivých nukleotidů. Po každém zařazení nukleotidu je emitováno světlo, které je detekováno a tím je stanovena DNA sekvence. Schéma sekvenačního pracovního postupu „workflow“ viz. Obr. 1. Tento proces probíhá paralelně ve všech jamkách pikotitrační destičky což výrazně zvyšuje sekvenační kapacitu (13, 17, 18, 19). Firma Roche uvedla na trh postupně GS FLX systém a GS Junior systém (www.roche.com cit. 20. 8. 2022).



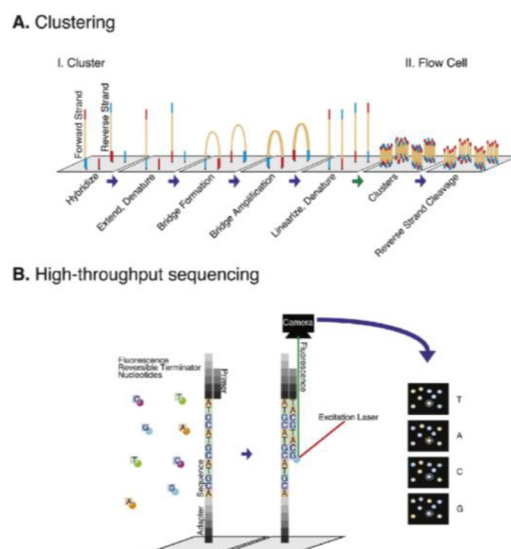
Obr. 1. Princip sekvenace využívané sekvenátorem 454 Roche – převzato z Mardis 2008 (20).

(PTP – pikotitrační destička; sekvenační proces začíná přípravou amplikonové knihovny; přidání specifických adapterů k amplikonům pro vazbu na kuličky; amplifikace navázaných amplikonů pomocí emulzní PCR na kuličkách; nanesení kuliček na

pikotitrační destičku; pyrosekvenční reakce vede k emitaci světla po úspěšném zařazení nukleotidu; tento světelný signál je detekován a tím je stanovena sekvence DNA řetězce)

2.3.2 Fluorescenční detekce – Illumina

V roce 2006 byl uveden na trh tzv. Genomický analyzátor společnosti Solexa, která byla v následujícím roce odkoupena firmou Illumina. DNA knihovny mohou být připraveny jakoukoli metodou, která vede ke vzniku směsi fragmentů oboustranně ohraničených adaptéry (např. amplikonové knihovny, knihovny na principu obohacení cílových sekvencí hybridizací se sondami). Amplifikace těchto DNA fragmentů probíhá pomocí tzv. můstkové PCR (21, 22, 23, 24). Sekvenční technologie Illumina pak vychází z tzv. sekvenování po syntéze „sequencing-by-synthesis“. V prvním kroku jsou připraveny knihovny DNA, bez ohledu na metodu jejich přípravy je společným krokem navázání adapterů (krátkých oligonukleotidových sekvencí) na oba konce DNA fragmentů. Tyto fragmenty jsou poté denaturovány na jednořetězcové úseky DNA, které pomocí adapterů hybridizují s komplementárními sekvencemi na tzv. flow cell, což je transparentní destička. Hybridizace DNA fragmentů k adapterům na flow cello probíhá díky zahřívání a ochlazení. Následuje inkubace s polymerázou a dalšími reakčními složkami, které amplifikují DNA fragmenty pomocí „PCR můstkové amplifikace“ viz. Obr. 2. Takto vznikne mnoho identických kopií každé sekvence, které tvoří klastry. V každém klastru je až milion kopií jedné identické sekvence. Detekce pořadí nukleotidů v DNA řetězci probíhá po vložení flow celly do fluidní kazety sekvenátoru. Jednotlivé klastry jsou zásobeny polymerázou a čtyřmi fluorescenčně značenými nukleotidy s chemicky inaktivovanou 3'OH skupinou. Tato modifikace zaručí v každém cyklu inkorporaci pouze jednoho nukleotidu. Po každém cyklu následuje zobrazovací krok (emitace fluorescence laserem – detekce CCD kamerou) a chemický krok, který odstraní fluorescenční skupinu a odblokuje 3' konec pro možnou inkorporaci dalšího nukleotidu v dalším cyklu (13, 20). Sekvenční workflow viz. Obr. 2. Společnost nabízí sekvenční platformy: iSeq 100, MiniSeq, MiSeq, NextSeq, HiSeq, HiSeq X, NovaSeq (www.illumina.com cit. 24. 8. 2022).

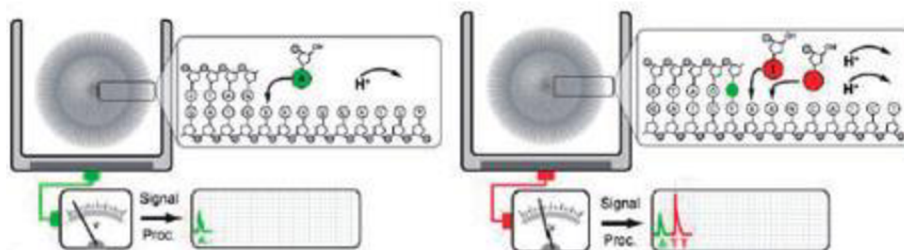


Obr. 2. Princip sekvenace využívané na platformě Illumina – převzato z Chaitankar et al. 2016 (24).

(A – tvorba klastrů → DNA fragmenty jsou denaturovány a nanесeny na flow cellu, kde hybridizují s komplementárními adaptéry; komplementární fragmenty jsou prodlouženy a amplifikovány můstkovou PCR a denaturovány; vzniknou identické jednořetězcové DNA fragmenty – klastry; B – sekvenace → fragmenty jsou sekvenovány za využití reverzibilních terminátorových nukleotidů; nukleotidy jsou detekovány po excitaci laserem a fluorescenční detekci)

2.3.3 Polovodičová technologie – Ion Torrent

Life Technologies uvedla na trh na konci roku 2010 sekvenační platformu Ion Torrent Personal Genome Machine (PGM). Ion Torrent je unikátní díky prvnímu využití tzv. „post-light“ sekvenace, protože nevyužívá fluorescenci ani luminiscenci (25). Prvním krokem je příprava knihovny DNA (zejména metodou AmpliSeq, může být také fragmentace – v závislosti na aplikaci), na vzniklé DNA fragmenty jsou navázány adaptéry, které umožní hybridizaci fragmentů ke kuličkám. Obdobně jako 454 pyrosekvenační technologie využívá Ion Torrent emulzní PCR, pro klonální amplifikaci DNA fragmentu na kuličkách. Kuličky jsou přeneseny do jamek na čipu a postupně omývány jednotlivými nukleotidy. Úspěšná inkorporace je měřena dle změny pH, způsobené uvolněním vodíkového iontu – to je umožněno využitím polovodičové „semikonduktorové“ technologie, kterou jsou vyrobeny mikroprocesorové sekvenační čipy (25, 26). I tato sekvenační technologie tedy využívá sekvenování po syntéze „sequencing-by-synthesis“. Princip sekvenace viz. Obr 3. Technologie Ion Torrent nabízí také stanici Ion Chef™. Tato významně urychluje přípravu knihoven a templátů. Zařízení je kompatibilní s platformou Ion S5™. Posledním sekvenačním systémem je Genexus™ (www.thermofisher.com cit. 25. 8. 2022).

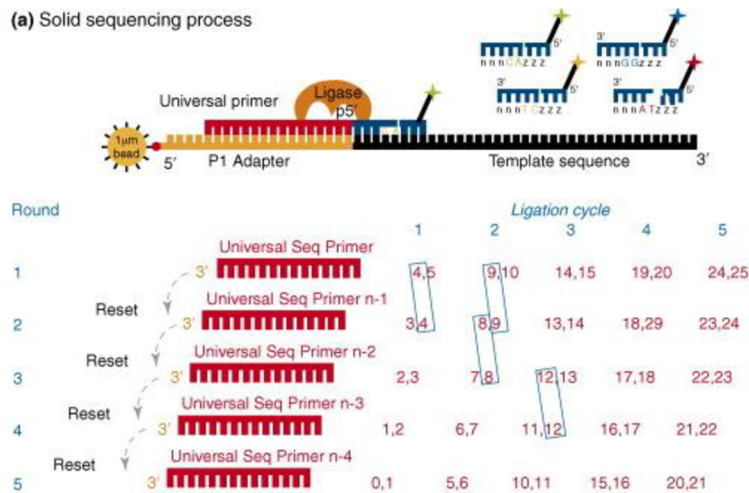


Obr. 3. Princip sekvenační technologie Ion Torrent, detekce správně zařazených nukleotidů – upraveno z Rothberg et al. 2011 (25).

(Kuličky nanesené na čipu jsou postupně omývány jednotlivými nukleotidy; po úspěšném navázání komplementárního nukleotidu je uvolněn vodíkový iont; při každém omytí je měřen elektrický signál v každé jamce na čipu, ten udává počet úspěšně zařazených nukleotidů)

2.3.4 Sekvenování po ligaci – SOLiD systém

V roce 2007 uvedla na trh firma ABI sekvenátor SOLiD (Sequencing by Oligo Ligation and Detection). Tento přístup je unikátní pro využití technologie sekvenování po ligaci „sequencing-by-ligation“, DNA ligáza je využita k přidání sond do rostoucího oligonukleotidového řetězce (27). Prvním krokem je navázání oligo adapterů na DNA fragmenty, tyto se poté specificky váží na magnetické kuličky s komplementárními oligonukleotidy. Tyto kuličky se v následně amplifikují pomocí emulzní PCR. Dalším krokem je nanesení kuliček na sklíčko a vložení do fluidní kazety sekvenátoru. Sekvenační proces pokračuje navázáním univerzálních primerů komplementárních se SOLiD adaptery, ligázy a přidáním oktamerů s fluorescenční značkou. Tyto postupně nasedají na fragmenty DNA, při následné detekci fluorescence. Výstupem je barevný prostor, ve kterém je zakódován konkrétní nukleotid, jsou zde čtyři fluorescenční barvy, které reprezentují 16 možných kombinací dvou bází. Sekvenátor opakuje každý ligační cyklus. Každý cyklus je komplementární vlákno odstraněno a nový cyklus sekvenace začíná na pozici n-1 templátového vlákna. Tento proces se opakuje, dokud není každá báze dvakrát osekvenována. Získaná data jsou pak převedena do konkrétní sekvence bází DNA (13, 20). Sekvenační proces je znázorněn na Obr. 4.



Obr. 4. Schéma sekvenačního procesu založeném na „sekvenování po ligaci“ využívané sekvenátorem SOLiD – převzato z Mardis 2008 (20).

(Sekvence je zahájena nasednutím univerzálního sekvenačního primeru, následuje ligace příslušného fluorescenčně značeného oktameru a detekce v každém cyklu)

2.3.5 Stolní sekvenátory

Důležitým milníkem v klinické praxi bylo úspěšné zavedení „stolních“ NGS sekvenátorů do malých laboratoří. Podmínkou byla malá velikost přístroje a rychlá odezva. Tyto podmínky byly v letech 2010–2011 splněny sekvenátory MiSeq (Illumina), Ion PGM (Ion Torrent) a 454 GS Junior (Roche) (28, 29). Výsadní postavení má společnost Illumina – více než 90 % sekvenačních dat je generováno na jejich přístrojích (30).

2.4 Třetí generace sekvenování

Sekvenační technologie třetí generace byly vyvinuty s cílem eliminovat nedostatky sekvenačních technologií druhé generace. Zejména pak redukce časově a finančně náročné PCR amplifikace a nutnosti co nejpřesněji přečíst repetitivní genomické oblasti, které jsou pro technologie krátkého čtení obtížné. (13, 31) Tyto technologie bývají nazývány také sekvenování s dlouhým čtením nebo sekvenování jednotlivých molekul. Největším přínosem těchto metod je možnost dlouhých sekvenčních čtení >10 kb přímo z nativní DNA (13, 15, 31, 32).

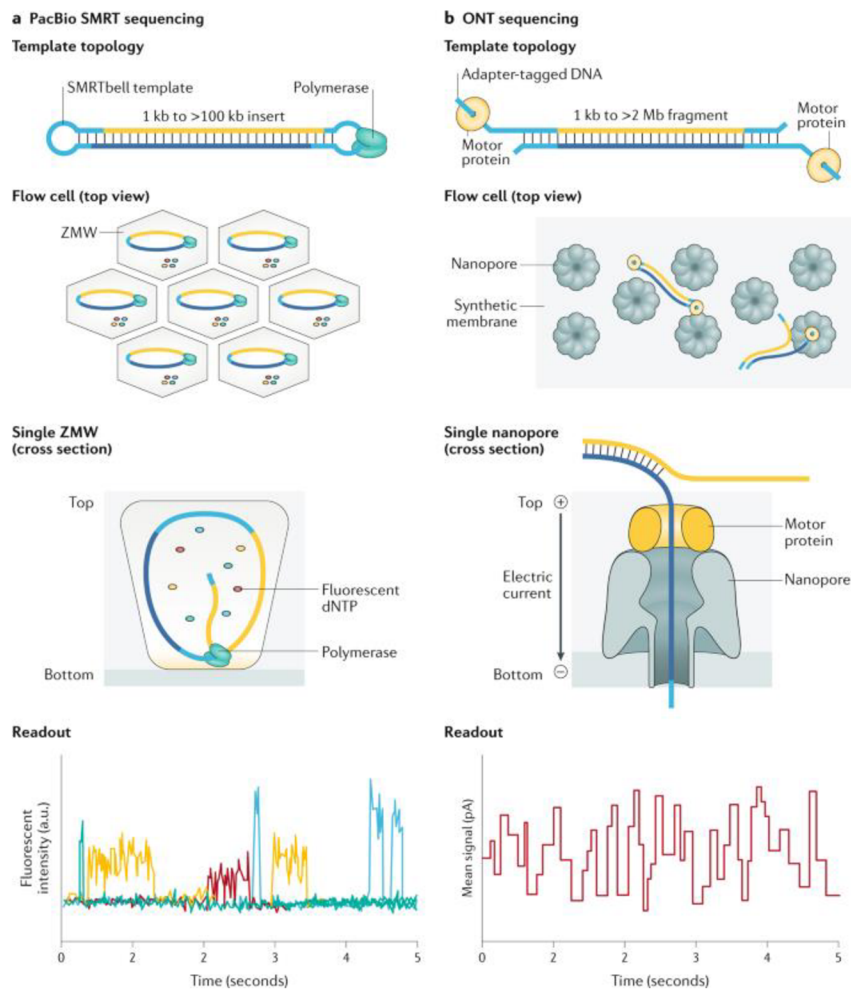
2.4.1 Single molecule real-time – Pacific Biosciences

Společností Pacific Bioscience (PacBio) byla vyvinuta „single molecule real-time“ (SMRT) sekvenační metoda. Metoda využívá templát ve formě jednovláknové kruhové DNA molekuly – SMRTbell. Templát vzniká ligací vlásenkových adaptérů na oba konce dvouřetězce DNA. Sekvence probíhá na čipu – SMRT cell – s malými póry zvanými zero-mode waveguides (ZMW). DNA polymeráza se naváže na adaptér SMRTbell a začíná replikace. Polymeráza obtéká templát a zařazuje fluorescenčně značené dNTP do syntetizovaného vlákna. Po každé inkorporaci je laserem excitován fluorofor za současného zaznamenání excitace. Před další inkorporací dNTP je fluorofor odštěpen. Tento proces se opakuje, dokud není osekvenován celý SMRTbell templát. Tato platforma generuje čtení o délce okolo 10 kb (15, 32, 33). Vhodnou aplikaci PacBio SMRT Rhoads and Au 2015 (33) spatřují ve výzkumu genomu, transkriptomu a epigenetiky. Princip této metody je znázorněn na Obr. 5.

2.4.2 Nanopórová technologie – Oxford Nanopore Technology

Oxford Nanopore Technologies (ONT) v roce 2014 uvedli na trh sekvenátor MinION (34). Tato sekvenační platforma využívá spíše lineární DNA molekuly délky až několika kb. Sekvenátor může generovat čtení i delší než 1Mb (15, 32). Úvodním krokem je připojení dvouřetězcové molekuly DNA k sekvenačnímu adaptéru, který je již naplněn motorickým proteinem. Tato směs je nanášena na flow cellu (nanopórový čip), která obsahuje tisíce nanopórů v syntetické membráně. Motorický protein společně s elektrickým proudem rozvine a protlačí záporně nabitou DNA skrz póry. Při průchodu DNA pórem jsou zaznamenávány charakteristické odchylky proudu, které se analyzují v reálném čase. Tyto pak stanoví sekvenci analyzovaného vlákna DNA (13, 32). Princip této metody je znázorněn na Obr. 5.

Hlavní benefity této metody jsou shodné jako u technologie PacBio pramenící z generování dlouhých čtení. ONT přístroje mají výhodu nižší ceny, rozměrů a přesnosti – tyto vlastnosti lze využít například při terénních aplikacích, jak dokumentují Quick et al. 2016 (35) u sekvenace viru Eboly. Obdobné jsou také hlavní nevýhody, které jsou spojovány především s vysokou úrovní chyb (2–20 %) ve srovnání se sekvenačními technikami 2. generace (13, 15, 34, 35, 36). Příslibem pro zvýšení přesnosti jsou vylepšení v oblasti užívané chemie a algoritmů pro detekci bází „base calling“ (23).



Obr. 5. Schéma sekvenačního procesu vybraných sekvenačních technologií třetí generace a) SMRT; ONT – převzato z Logsdon et al. 2020 (32).

(A – SMRT → žlutě – forward vlákno; modře – reverse vlákno; DNA vlákno je fragmentováno a navázáno k vláskovkým adapterům (světle modrá) – vytvoří SMRTbell; na SMRTbell se váže DNA polymeráza a nanese se na čip – SMRT cell s malými póry zero-mode waveguides (ZMW) zde probíhá sekvenace; při sekvenaci se přidávají fluorescenčně značené dNTP; po úspěšném zařazení značeného dNTP je fluorofoer excitován a detekován kamerou, která zaznamená vlnovou délku a relativní polohu inkorporované báze v nově vznikajícím řetězci. **B – ONT** → žlutě – forward vlákno; modře – reverse vlákno; vlákno DNA je označeno adaptéry (světle modrá) opatřenými motorickým proteinem a vloženo do flow celly, která obsahuje nanopóry v syntetické membráně; adapter je vložen do nanopóru a motorický protein odvíjí DNA; elektrický proud spolu s motorickým proteinem protlačuje DNA přes nanopór; pohyb DNA pórem způsobuje typické odchylky proudu na základě, kterých se sestavuje sekvence DNA).

3 Pracovní postup NGS

Následující kapitoly budou zaměřeny výlučně na sekvenační techniky 2. generace se zaměřením na platformu Ion Torrent z důvodu jejího využití v praktické části této práce.

Vzhledem k velmi odborným výrazům v této oblasti budou některé ustálené termíny ponechány v původním jazyce a vysvětleny přímo v textu.

Proces sekvenace a práce s daty získaných vysokokapacitním sekvenováním lze rozdělit do tří základních kroků: Příprava knihoven, sekvenace a datová analýza. Jednotlivé kroky se mohou značně odlišovat v závislosti na vstupním vzorku: DNA / RNA, případně aplikaci: cílená sekvenace (TS), celoxomová sekvenace (WES), celogenomová sekvenace (WGS). V rámci této závěrečné práce a aplikace metodiky do klinické praxe je předmětem výhradně analýza variant z vybraných oblastí lidského genomu – možné aplikace analýzy RNA nejsou předmětem této práce.

3.1 Příprava knihoven

Základním předpokladem pro úspěšné sekvenování je kvalitní knihovna. Příprava knihovny se skládá ze získání templátů, odpovídajících molekulám, které mají být osekvenovány. Poté přípravou fragmentů, tak aby byly kompatibilní se zvolenou sekvenační platformou (37).

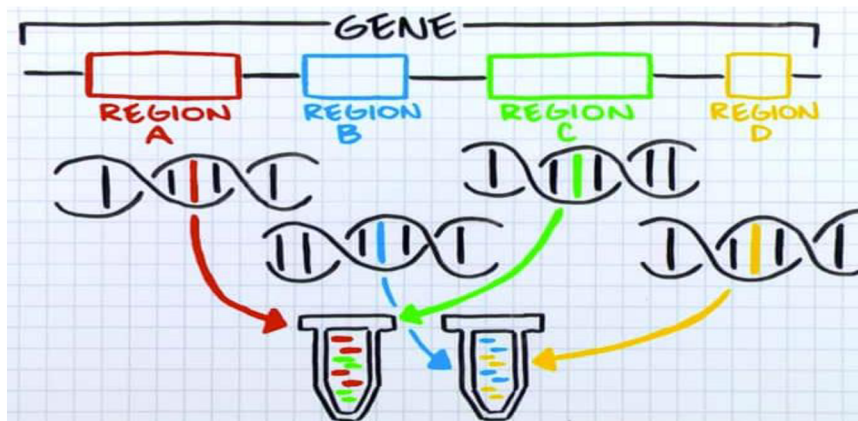
DNA sekvenace může v závislosti na sekvenačním templátu zahrnovat TS, WES, WGS a sekvenování epigenomu (38). Pro přípravu templátu jsou běžně využívány dva přístupy: metody založené na PCR a metody založené na hybridizačním záchytu „capture-based“.

3.1.1 Metody založené na PCR

Amplikonové sekvenování bylo dlouhou dobu kapacitně omezeno na omezený počet genů, exonů. Změnu přinesl nástup „long-range PCR“. Dominantní se stala sekvenace kratších, náhodně vytvořených DNA fragmentů, které byly kratší než amplikon (tzv. „shotgun sekvenace“). Tímto přístupem je možné osekvenovat celé geny, introny a regulační oblasti. Metody založené na PCR jsou nejčastěji využívány pro přípravu templátů u TS (15, 37). Hu et al. 2021 (15) uvádí, že nedostatkem této techniky je občasný výpadek alel.

Společnost Thermo Fisher Scientific vyvinula Ion AmpliSeq™ technologii přípravy knihoven pro TS založené na amplikonovém obohacení. Takto navržený genový panel se skládá z tzv. poolů, ve kterých jsou obsaženy páry primerů, kde každý amplifikuje specifickou genomickou oblast. Jedinečná je možnost multiplexu těchto primerů až do 24 000 párů v jedné PCR. To umožňuje v jednom sekvenačním běhu cílit na jednotky až stovky genů, popřípadě celé exomy. Po amplifikaci se primery rozštěpí a knihovna obsahující syntetizované amplikony je

připravena. Pro naše potřeby je vhodná především AmpliSeq technologie genového designu. Ta vychází ze skládání překrývajících se ampliconů za účelem stanovení souvislé genové sekvence. Z toho důvodu musí být překrývající se páry primerů rozděleny do dvou poolů. Multiplex PCR tedy proběhne pro tyto pooly zvlášť, za účelem zajištění dostatečného překryvu (eliminace případného ampliconového výpadku) a pokrytí viz. Obr. 6 (39).



Obr. 6. Design genového panelu pomocí Ion AmpliSeq™ – převzato z Kim 2017 (39).

3.1.2 Metody založené na hybridizačním záchytu

Principem je využití biotinylovaných sond, které podle komplementarity hybridizují s vybranými oblastmi genomu. SONDY JSOU PAK „VYCHYTÁNY“ POMOCÍ MAGNETICKÝCH KULIČEK, které jsou pokryty streptavidinem. Podle Hu et al. 2021 (15) jsou tyto techniky ve srovnání s těmi založenými na PCR nákladově výhodnější pro TS, zejména v případě větších genomických oblastí a většího počtu genů, uvádějí autoři snížení výpadku alel. Tyto techniky jsou tedy využívány pro všechny zmiňované aplikace – TS, WES, WGS (15, 37, 38).

Příprava knihoven zahrnuje také fragmentaci, opravy konců, ligaci adapterů a výběr velikosti. Pro fragmentaci DNA jsou užívány fyzikální (akustické stříhání, sonikace), chemické a enzymatické metody (nespecifická směs endonukleáz, transpoziční značení) (40). Úprava konců je přípravným krokem pro ligaci adapterů. Cílem je, aby konce DNA fragmentů neobsahovaly „převisy“ s 5' fosfátovou anebo 3' hydroxylovou skupinou, ale pouze tupé zakončení a netemplátový deoxyadenosin 5' monofosfátem. Následuje velikostní selekce, která umožní obohacení DNA fragmentů (v definovaných velikostech) a odstranění kontaminantů. Nejčastěji tak probíhá hybridizací na kuličkách (např. Agencourt AMPure XP) případně

elektroforetickou selekcí (15). Obohacení cílové oblasti bylo vyvinuto za účelem optimalizace sekvenačního běhu, zvýšení časové a nákladové efektivity a snadnější analýzy výsledných dat (41). Technik k obohacení cílových oblastí bylo vyvinuto více např. hybridizační (na kuličkách), metody založené na PCR, molekulárně inverzní sondy, array capture, hybridní aj. (28, 41).

Důležitým krokem v přípravě knihoven je využití krátkých oligonukleotidů „čárového kódu“ k označení DNA fragmentů pocházejících od určitého pacienta. Tento krok umožní smíchání a sekvenaci více vzorků DNA během sekvenačního běhu. Bylo popsáno více přístupů pro označení vzorků čárovými kódy (41, 42).

3.2 Datová analýza

Díky využití vysokokapacitních sekvenátorů získáváme značné množství hrubých dat, které je třeba dále zpracovat, abychom mohli odhalit genomické varianty a posoudit jejich možný klinický dopad. Bioinformatický proces zpracování hrubých dat „bioinformatickou pipeline“ lze rozdělit na primární, sekundární a terciální bioinformatickou analýzu (15, 43). V rámci datové analýzy byl v této práci kladen důraz především na platformu Ion Torrent, která byla využita v praktické části této disertační práce.

V následujících oddílech práce jsou užívány rozdílné typy souborů, které uchovávají sekvenační data v rozdílných formách. Pro přehlednost nejdůležitější / nejčastěji užívané viz. Tab. 1.

Tab. 1. Vybrané typy souborů uchovávající sekvenační data.

Soubor	Popis souboru
FASTQ	Na textu založený formát pro uložení nukleotidové sekvence s odpovídajícím skóre kvality
uBAM	binární soubor unaligned or unmapped – forma BAM formátu, který neobsahuje informace o mapování
BAM	Binary Alignment Map – soubor se užívá k ukládání textových informací o čtení přiřazených k referenční sekvenci. Binární verze SAM formátu
BAI	Binary Alignment index – soubor urychlující přístup pro BAM, který řazen dle koordinát
BED	Browser Extensible Data – definuje pozici na chromozomu anebo určitou oblast
SAM	Sequence Alignment Map – textový formát ukládající čtení přiřazené k referenční sekvenci
TSV	Tab-separated values – soubor oddělený tabulátory, určený pro tabulkový software. Hrubá data v textových prohlížečích
CSV	Comma-separated values – oddělený textový soubor. Každý řádek představuje datový záznam s poli oddělenými čárkou. Tento soubor ukládá tabulková data v prostém textu
VCF	Variant Call Format – textový soubor ukládající sekvenční varianty
XLS	Microsoft Excel formát VCF souboru

3.2.1 Primární analýza

Analýza začíná již samotnou detekcí či analýzou sekvenačního signálu a následnou konverzí hrubých sekvenačních dat do nukleotidových sekvencí – jednotlivých čtení (base calling). Výstupy primární analýzy se z pohledu běžně užívaných platform (Illumina vs Ion Torrent) odlišují. Výstupním formátem primární analýzy u Illuminy je FASTQ u Ion Torrentu uBAM (43).

Primární analýza u platformy Ion Torrent probíhá pomocí softwaru Ion Torrent Suite Software. Tento nástroj nabízí uživateli automatickou, případně přizpůsobitelnou datovou analýzu (44). Analýza začíná detekcí zařazení nukleotidu (senzor v každé jamce na čipu), signál je konvertován v napětí a na server je odeslán jako DAT soubor (nezpracované stopy elektrických signálů). Následuje krok zpracování signálu, konverzí nezpracovaných stop elektrických signálů na jedno číslo pro daný tok nukleotidů a danou jamku na čipu ve WELLS soubor. Primární analýza je zakončena detekcí bází „base callingem“ – pomocí modulu BaseCaller je určena co nejpravděpodobnější sekvence bází na základě dat z WELLS souborů. V těchto krocích dochází k normalizaci signálu – proces je důležitý k vyřešení určitých chyb vznikajících během sequencing-by-synthesis techniky jako je pokles signálu anebo fázování (44).

Sekvenační proces generuje chyby, je tedy nezbytná určitá kontrola kvality base callingu. Využívá se Phred skóre, které přiřazuje logaritmickou pravděpodobnost chyby každé detekované báze. Phred skóre 10 (Q10) značí pravděpodobnost, že 1 z 10 bází je zařazena chybně, tedy přesnost 90 % (45). Důležité pro kontrolu kvality jsou data ve FASTQ formátu. Obsahují hrubé sekvenační čtení, názvy souborů a hodnoty kvality – tento formát lze vygenerovat také u platformy Ion Torrent pomocí modulu FileExporter (44). Hodnota Phred skóre je důležitá pro filtrování (např. nekvalitní čtení) a ořezávání „trimming“ sekvencí. Ořezávání zahrnuje odstranění určitých bází s nižší kvalitou, adaptéry, molekulární značku, případně čárovým kódem (43, 44).

Využití čárových kódů (unikátní krátké sekvence) k označení jednotlivých DNA fragmentů umožňuje smíchat a osekvenovat více vzorků DNA během jednoho sekvenačního běhu a má za následek snížení nákladů na sekvenaci. Z toho vyplývá nutnost demultiplexování čtení před další analýzou – roztřídění všech čtení na základě jejich specifického čárového kódu (46). Tento proces je pro každou platformu specifický – pro Ion Torrent probíhá pomocí Torrent Suite Software. Posledním krokem primární analýzy je odstranění adaptérových sekvencí, které byly navázány během přípravy knihoven, z konců (obvykle 3' konců) jednotlivých sekvenačních čtení. Jejich odstranění je nutné kvůli možné interferenci při mapování a sestavování „assembly“. Tento krok zvyšuje kvalitu a spolehlivost analýzy, je důležitý pro identifikaci jednonukleotidových polymorfismů (SNP) a sestavení genomu „genome assembly“. Zároveň také snižuje čas provádění analýzy (47). Pro Ion Torrent je trimming automaticky prováděn pomocí Torrent Suite Software.

3.2.2 Sekundární analýza

V tomto kroku bioinformatická analýza NGS dat zahrnuje přiřazení jednotlivých sekvenačních čtení k lidskému referenčnímu genomu „read alignment“ a posuzování variant „variant calling“ (VC). Běžně se využívá jeden ze dvou referenčních lidských genomů GRCh37 / hg19 nebo GRCh38 / hg38 (43, 44). Mapování sekvenačních čtení má dva rozdílné přístupy. 1) Read alignment – přiřazování jednotlivých čtení k referenčnímu genomu. 2) *De novo* assembly – spojování čtení bez využití referenčních dat. V klinických aplikacích je využíváno přiřazování jednotlivých čtení proti danému referenčnímu genomu, tedy hg19 nebo hg38. *De novo* assembly se využívá specificky, zejména pokud není referenční genom k dispozici nebo jak popisuje Ameer et al. 2018 (48) k doplnění chybějících sekvencí lidských referenčních genomů. Přístup sestavení *de novo* není běžně aplikován v klinické praxi, proto se ním nebude tato práce podrobně zabývat.

Správné přiřazení k referenční sekvenci je velmi komplikovaný proces, který si musí poradit s krátkým čtením, repetitivními oblastmi v genomu aj. (49). Výstupním formátem sekundární analýzy je BAM a SAM – tyto formáty obsahují sekvenační čtení, umístění alignmentu, skóre kvality zařazení bází, rozdíly oproti referenční sekvenci a skóre kvality mapování. SAM je textový soubor vytvořen pro snadné zpracování, BAM je binární verze stejných dat. Výstupy z těchto formátů mohou být nahrány a fyzicky kontrolovány pomocí prohlížeče Integrative Genomics Viewer (IGV).

3.2.2.1 Sequence Alignment

Je využíván algoritmus, který přiřazuje přečtenou sekvenci k místu v referenčním genomu na základě jejich shody – za současné tolerance určitého počtu neshod, které umožní následnou detekci variant (43).

V této oblasti byly vyvinuty přístupy jako jsou Hash-based alignment metody. Programy, které vychází z této metodiky jsou založeny na hash-tabulce – datové struktuře pro indexování a skenování / rychlé vyhledávání sekvenačních dat (49). Mezi programy pracující na této bázi se řadí např. SOAP (50) a SHRiMP (51). Dále mohou být užívány Burrowsovy-Wheelerovy transformační metody, které jsou využívány programy, jako jsou Bowtie 2 (52), BWA (53).

Tyto využívají datovou strukturu FM indexu. Burrows-Wheelerova transformace umožňuje významnou datovou kompresi například pomocí techniky bzip2. (31, 43, 49).

Proces alignmentu v Torrent Suite Softwaru provádí TMAP algoritmus, který má dvě základní fáze. **1) Počáteční mapování** – kdy jsou sekvence čtení lokalizovány do referenčního genomu. Pro uživatele jsou k dispozici moduly map1, map2, map3, map4 a mapvsw. Základní nastavení je map4 modul – BWA rychlé mapování, kdy vyhledává maximální přesné shody mezi čtením a referencí. **2) Zpřesnění alignmentu** – kdy je každá konkrétní pozice čtení přiřazena k odpovídající pozici v referenčním genomu. Jsou k dispozici také volitelné možnosti zpřesnění alignmentu ke kompenzaci pro Ion Torrent specifických sekvenačních odchylek jako jsou: opětovné přiřazování „re-alignment“ – pro lepší přiřazení v homopolymerních oblastech, re-alignment oblastí s pravděpodobnými chybami ve fázování s nízkým skóre inzercí, delecí (INDEL), aj. Umožňuje také odfiltrování duplicitních čtení (44). TMAP software tedy zpřesňuje proces následného VC.

3.2.2.2 Variant calling

Proces VC je spojován s generováním variant call formátu (VCF). Jedná se o záznam / hledání rozdílů mezi vybraným referenčním genomem a daným sekvenačním čtením. Rozdíl je detekován na úrovni jednotlivých nukleotidů. Vstupním formátem je obvykle BAM soubor.

K dispozici jsou volně dostupné nástroje, které mají algoritmy umožňující detekci jednonukleotidových variant „single nucleotide variants“ (SNV) a krátkých INDELS (obvykle do 30bp) Například GATK (54), FreeBayes (55), SAMtools (56) – které jsou založeny na Bayesovském přístupu (57). Tyto nástroje jsou hojně využívány pro data z Illuminy (43). Běžně jsou VCF generovány pomocí komerčních software – NextGENe (58), Torrent Suite – pomocí vlastního Torrent Variant Caller (TVC) modulu (44). Také na této úrovni lze nalezené varianty filtrovat dle zvolených kritérií.

TVC detekuje SNP, vícenukleotidové polymorfismy a INDEL. Software umožňuje spustit analýzu automaticky, bezprostředně po sekvenačním běhu. Nebo manuálně kde je možné ovlivnit přísnost „stringency“, specifičnost a citlivost. Dále se také vybírá referenční genom, cílové oblasti „target regions“ případně hot spots. Je možné opět využít předdefinované nastavení, jako je třeba u Ion AmpliSeq. Pro amplikonovou sekvenaci je užíváno nastavení

target regions – kde vložíme adekvátní soubor BED, který určuje zájmovou genomickou pozici pro analýzu (44).

3.2.2.3 Copy number variant calling

Kauzální fenotyp pacienta mohou vyvolat kromě menších genomických změn jako jsou SNV a INDELS, také varianty počtu kopií „copy number variants“ (CNV) zahrnující velké delece a duplikace. Kauzální CNV byly popsány např. u geneticky podmíněného deficitu antikoagulačních proteinů (59, 60). V citované literatuře byla diagnostika sekvenační technikou doplněna metodou „Multiplex Ligation-Dependent Probe Amplification“ (MLPA), která detekci CNV umožňuje. Vysokokapacitní sekvenační technologie jsou nyní navíc schopny odhalit balancované i nebalancované strukturální varianty (61).

Analýza metodami NGS byla běžně doplňována metodou pro detekci CNV, jako je již zmiňovaná MLPA – nicméně právě technologie NGS mají potenciál CNV spolehlivě detekovat. Byly vyvinuty nástroje k jejich detekci založené na: hloubce čtení, mapování párových konců, rozdělených čteních, *de novo* assembly a kombinatorickém přístupu. Z běžně užívaných softwarů CNV hodnotí např. NextGENe a Ion Reporter (62).

Ion Torrent umožňuje CNV analýzu pomocí softwaru Ion Reporter – pro knihovny připravené metodou amplikonového sekvenování. Detekce je prováděna pomocí Hidden Markova Modelu – který využívá normalizované pokrytí jednotlivých amplikonů k určení počtu kopií nebo ploidie (63). Při využívání AmpliSeq On-Demand Panelů lze CNV spolehlivě detekovat, pokud CNV pokrývá alespoň 10 optimálně 20 amplikonů. I s menším počtem amplikonů je CNV detekce možná, ale zejména při méně kvalitních datech nebo nižším pokrytí může být CNV přehlédnuta (64). Stanovení CNV v programu Ion Reporter viz. Obr. 7.

Copy Number	CNV Confidence	CNV P-Value	CNV Precision	CytoBand
2	0		487.8	3p25.2q26.33(12625929-181431173)x2

Obr. 7. Detekce CNV v genu *PROS1* pomocí software Ion Reporter.

(Copy Number – počet kopií v detekované oblasti genu; CNV Confidence – odhadovaná procentuální pravděpodobnost, že počet kopií je menší než daný počet kopií; CNV P-Value – významnost měření poměru CNV; CNV Precision – čím je hodnota vyšší tím je určení CNV spolehlivější; CytoBand – oblast na chromozomu, případně oblast, ke které je vztažen konkrétní výsledek CNV analýzy)

3.2.3 Terciální analýza

Závěrečný krok analýzy dat získaných z MPS je spojen s anotací a interpretací nalezených genomických variant z pohledu klinické genetiky, tedy v kontextu s určitým genetickým onemocněním. Na této úrovni nám kromě *in-silico* hodnotících algoritmů a databází může pomoci také prostá korelace genotyp versus fenotyp pacienta. Posledním krokem je tedy zařazení odhalených variant do pěti skupin: patogenní, pravděpodobně patogenní, nejasný význam, pravděpodobně benigní a benigní.

3.2.3.1 Anotace variant

Vstupním formátem pro anotaci je VCF formát, který neobsahuje žádné informace o možném biologickém dopadu varianty na protein / organizmus. Vlastní anotací jsou tyto dodatečné informace o biologickém účinku odhalených variant doplněny a varianty mohou být na jejich podkladě interpretovány. Je nezbytný také komplexní pohled na odhalenou variantu tzn. v jakém genu a na jaké pozici se varianta nachází, jaký má dopad na transkripci / translaci (missense, nonsense, synonymní, stop-loss, frameshift) (43). Komplexní nástroje pro anotaci variant využívají data z mnoha softwarů a databází, které budou dále jmenovány. Mezi tyto komerční nástroje se řadí např. NextGENe, Geneticist Assistant (65), VariantStudio (66), aj. Pro anotaci dat získaných z platformy Ion Torrent je využíván především software Ion Reporter, který je určený a přizpůsobený pro analýzy dat získaných sekundární analýzou pomocí Torrent Suite softwaru.

3.2.3.1.1 Populační databáze

Tyto databáze obsahují osekvenované exomy a genomy statisíců dobrovolníků z různých populací. V různých populacích a etnických skupinách existují genomické odchylky, z toho důvodu jsou tyto soubory osekvenovaných dobrovolníků pro správné určení významu nalezených variant nezbytné (67).

Nejnámější jsou databáze 1000 Genomes Project (68) a Exome Aggregation Consortium (ExAC), jehož data jsou k dispozici v rámci The Genome Aggregation Database (gnomAD). Tato databáze se skládá z 125 748 exomů a 15 708 genomů pro hg19 (69). Varianty vyplývající

z těchto databází jsou vloženy do databáze jednonukleotidových polymorfismů (dbSNP) (70), volně dostupné z National Center for Biotechnology Information (NCBI) (71). Databáze dbSNP shrnuje dostupné informace o dané variantě – populační frekvenci alternativní alely, vliv na zařazenou aminokyselinu, klinickou signifikanci, publikace, aj.

Nezbytné pro identifikaci potenciálně patogenních variant je hodnota Minor allele frequency (MAF), která představuje frekvenci druhé nejčastější alely v dané populaci. Výpočet MAF odpovídá podílu pozorované alely a všech osekvenovaných alel v rámci vybrané databáze (72). V rámci databáze dbSNP tedy MAF vychází z dat 1000 Genome.

Genetická onemocnění, na jejichž diagnostiku se pomocí NGS technologie zaměřujeme se nejčastěji řadí mezi vzácná onemocnění, které jsou nejčastěji způsobena méně častými (MAF: 1–5%) a vzácnými (MAF: <1%) genetickými variantami (73, 74, 75). Z tohoto důvodu je vhodné využít při hodnocení odhalených variant filtr pro MAF <5 %.

3.2.3.1.2 Databáze variant

Klasifikace nalezených variant se také opírá o databáze, které obsahují informace o interpretaci a klinické signifikanci již popsaných variant. Z neznámějších je již jmenovaná dbSNP, dále také ClinVar (76). Databáze ClinVar je dostupná z NCBI. Obsahuje volně přístupné informace o možné korelaci varianty a daného fenotypu bez ohledu na velikost varianty (SNV, CNV). K daným variantám jsou uvedeny také podpůrné materiály, které umožnily stanovit klinickou signifikanci varianty – tato databáze má pro anotaci a interpretaci variant nezastupitelnou úlohu.

Hojně využívanou, volně dostupnou databází, je Leiden Open Variation Database (LOVD). Databáze se zaměřuje na kombinaci gen / onemocnění / pacient. Databáze tedy shromažďuje sekvenční varianty určitého jedince s informací o možné souvislosti s určitým genetickým onemocněním (77). Právě informace o jedinci, u kterého byla daná varianta odhalena, dělá tuto databázi unikátní.

Mezi databáze variant řadíme také Human Gene Mutation Database (HGMD), která si klade za cíl shromáždit všechny známé genetické poškození zodpovědné za lidské dědičné choroby, společně s polymorfizmy spojenými s onemocněními, které byly publikovány v odborné literatuře (78).

Uvedené databáze vycházejí z Human Genome Variation Society (HGVS) nomenklatury, která standardizovala zápis zaznamenaných variant. Systém nomenklatury variant navržený den Dunnem et al. 2000 (79) byl přijat odbornou veřejností. Poslední update této nomenklatury publikoval den Dunnem et al. 2016 (80). Pravidla HGVS nomenklatury jsou pak přehledně dostupná na webu (<http://varnomen.hgvs.org/> cit. 20. 9. 2022).

3.2.3.1.3 *In-silico* predikční nástroje

Klasifikace odhalených variant se také opírá o data ze softwarových prediktivních algoritmů, které předpovídají funkci proteinů. Tyto nástroje vypočítávají skóre pro každou variantu, vychází z parametrů jako např. stupeň evoluční konzervace zbytku aminokyseliny, statistická predikce na základě známé mutace, homologie dané sekvence, proteinová struktura aj.

Jedny z nejčastěji využívaných jsou predikční nástroje SIFT (Sorting Intolerant From Tolerant) (81) a PolyPhen-2 (Polymorphism Phenotyping v2) (82). SIFT je predikční nástroj umístěný na webovém rozhraní, který určuje, zda je záměna aminokyseliny poškozující. Skóre nabývá hodnot od 0,0 (poškozující) do 1,0 (tolerovaná) (81). PolyPhen-2 je obdobný nástroj, jehož skóre má opačný význam: nabývá hodnot od 0,0 (tolerovaná) do 1,0 (poškozující) (82). K dispozici jsou také další predikční nástroje např. MutationTaster2 (83) a VEP (Ensembl Variant Effect Predictor) (84). Výstupem programu VEP jsou mimo jiné také data z jiných prediktorů a nabízí tak poměrně ucelený pohled na možný dopad dané varianty – není tedy predikčním nástrojem v pravém slova smyslu, ale sdružuje výstupy z jiných nástrojů. Pro posouzení možného dopadu variant na sestřih messenger ribonukleové kyseliny (mRNA) byl navržen Human Splicing Finder (85).

Důležitým nástrojem je také PhyloP (phylogenetic p-values), který posuzuje evoluční konzervaci jednotlivých pozorovaných nukleotidů napříč vybranými zástupci třídy savců, případně vybranými obratlovci. Na základě toho identifikuje kandidátní funkční elementy v genomických sekvencích. Pro nulovou hypotézu neutrální evoluce / nulového driftu představují pozitivní hodnoty PhyloP skóre evolučně konzervovaná místa, záporné hodnoty pak představují oblasti podléhající zrychlené evoluci (86).

3.2.3.1.4 Genomové prohlížeče – vizualizace

Genomové prohlížeče umožňují fyzickou kontrolu určité sekvence, kde jsou k dispozici všechny dostupné informace, o již popsaných variantách. Díky informacím dostupným na jednom místě pak můžeme variantu na základě současných znalostí správně interpretovat.

Většina genomových prohlížečů umožňuje ve svém prostředí zobrazovat četné 1D stopy „track“ obsahující informace o referenčním genomu, transkriptomu, přiřazených čteních, nalezených mutacích, anotacích z veřejných zdrojů a dalších informacích důležitých pro správnou interpretaci nalezených variant (87). Na tomto principu jsou založeny volně dostupné genomické prohlížeče např. **Genome Data Viewer**, který integruje veškeré informace a některé nástroje (např. BLAST, primer BLAST) dostupné z NCBI (88). Dalším využívaným prohlížečem je **University of Santa Cruz (UCSC) Genome Browser**, který funguje od roku 2001 jako grafický prohlížeč pro zkoumání anotací genomu. Obsahuje stovky anotačních stop a užitečných nástrojů, např. in-silico PCR a LiftOver, aj. (89). Obdobně je založen také **Ensembl Genome Browser**. Projekt Ensembl vyvíjí prostředí pro poskytování referenčních dat pro interpretaci genomu jakéhokoli druhu. Samozřejmě jsou také vlastní programy dostupné z prostředí Ensembl jako jsou VEP, BLAST (90).

Velmi přehledný je prohlížeč **VarSome**. Tento prohlížeč umožňuje uživateli vyhledat varianty v genomickém kontextu, shromažďuje data ze třiceti externích databází a využívá četných predikčních nástrojů. Největším přínosem je umožnění komunitě sdílet informace o lidských variantách (91). Velkým benefitem je také implementace automatického klasifikátoru variant, který hodnotí variantu dle doporučení American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG/AMP) (92). Výstup z automatického klasifikátoru variant VarSome viz. Obr. 8. Tato klasifikace nám poskytne základní informace pro následnou interpretaci, což oceníme zejména u doposud nepopsaných tzv. „novel“ variant.

The image shows a list of five classification criteria for a variant in the PROC gene, each with a status icon (PS3, PM1, PM2, PP2, PP5) and a brief description of the evidence. The criteria are:

- PS3 Strong**: A VarSome user has classified this variant as Pathogenic, confirmed by a functional study, mentioned in [25533856](#).
- PM1 Moderate**: UniProt protein PROC_HUMAN domain 'Peptidase S1' has 113 missense/in-frame variants (55 pathogenic variants, 46 uncertain variants and 12 benign variants), pathogenicity = 48.7%, which qualifies as pathogenic.
- PM2 Supporting**: Using strength Supporting because the variant doesn't alter the protein length (missense) and the position isn't moderately conserved (the position is poorly conserved (phyloP100way = 2.39 is greater than 1.4)) GnomAD exomes homozygous allele count = 0 is less than 2 for AD/AR with good gnomAD exomes coverage = 98.7. Variant not found in gnomAD genomes with good gnomAD genomes coverage = 33.8.
- PP2 Supporting**: 120 out of 149 non-VUS missense variants in gene PROC are pathogenic = 80.5% which is more than threshold of 40.2%, and 151 out of 348 clinically reported variants in gene PROC are pathogenic = 43.4% which is more than threshold of 35.2%.
- PP5 Supporting**: UniProt Variants classifies this variant as Pathogenic, citing 2 articles ([7865674](#) and [7792728](#)), associated with Thrombophilia due to protein C deficiency, autosomal dominant. ClinVar classifies this variant as Uncertain Significance, 1 star (criteria provided, 3 submissions), citing 5 articles ([34355501](#), [25533856](#), [23174622](#), [7865674](#) and [7792728](#)), associated with None Provided, Not Provided, Proc Deficiency, Autosomal Dominant and 3 more, with 3 submissions (2 pathogenic, 1 uncertain, 0 benign).

Obr. 8. Kritéria / důkazy (dle ACMG/AMP) pro klasifikaci nalezené varianty v genu *PROC* z VarSome automatického klasifikátoru variant.

3.2.3.2 Interpretace nalezených variant

Správná interpretace nalezené varianty v lidském genomu, je velkou výzvou, jak pro pracovníky v molekulárně genetických laboratořích, tak pro klinické genetiky. I navzdory zavedení obecně uznávaných doporučení pro klasifikaci nalezených variant stále zůstává značná míra subjektivity (46). Správná interpretace varianty pro pacienta znamená benefit z pohledu možné kauzální terapie, případně ochranného režimu v rodině. Na druhou stranu špatná interpretace může pacientovi a jeho rodině výrazně uškodit.

V současné době interpretace sekvenčních variant vychází z doporučení, které publikovali Richards et al. 2015 (92) a jsou známy jako ACMG/AMP doporučení pro interpretaci germinálních sekvenčních variant. Pro popis nalezených variant nalezených v genech způsobujících Mendelovské poruchy doporučují užívat standartní terminologii: **patogenní, pravděpodobně patogenní, nejasný význam, pravděpodobně benigní, benigní**. Toto doporučení popisuje postup při klasifikaci odhalených variant do těchto kategorií za využití typických důkazů např. populační data, funkční data, *in-silico* predikce a segregáční data (92). Konkrétně jsou pro zařazení varianty do uvedených pěti kategorií využívána kritéria pro klasifikaci patogenních variant (Příloha 1), kritéria pro klasifikaci benigních variant (Příloha 2) a vlastní pravidla pro kombinaci kritérií pro klasifikaci variant (Příloha 3). Uvedené postupy slouží pouze jako doporučení, Richards et al. 2015 (92) zdůrazňuje důležitost vlastního odborného úsudku klinického či laboratorního genetika s ohledem na konkrétního pacienta, případně vzorek.

4 Aplikace metodiky vysokokapacitního sekvenování do klinické praxe pro vybrané onemocnění

Využití NGS je vhodné zejména v oblasti klinické diagnostiky, kde je předpoklad většího množství pacientů s určitou indikací a požadavkem na vyšetření předem dané skupiny genů. Tato metodika tak může plně nahradit do té doby využívaný, časově náročnější, molekulárně diagnostický postup kombinující Sangerovu sekvenaci doplněnou metodou MLPA.

V rámci zavedení metodiky NGS do rutinní diagnostiky byla navázána spolupráce s Hemato-onkologickou klinikou Fakultní nemocnice Olomouc, která naše pracoviště oslovila ohledně možné aplikace vysokokapacitní sekvenční technologie pro diagnostiku vrozeného deficitu

přirozených antikoagulantů (proteinu S, C a antitrombinu), které významně predisponují dané jedince k trombóze.

4.1 Žilní tromboembolická nemoc

Žilní tromboembolická nemoc (venous thromboembolism – VTE) zahrnuje žilní trombózu a plicní embolii (PE). Žilní trombóza je formování *trombu* v žilním systému, který může rezultovat v obstrukci *arteria pulmonalis* – kde utržení *trombu* způsobí PE, která může být pro člověka až letální.

Odhadovaná roční incidence VTE, pro populaci Evropského původu, se pohybuje v rozmezí od 104 do 183 případů na 100 000 lidí, což je obdobné jako u cévní mozkové příhody (93). U pacientů s VTE byla pozorována roční míra mortality v rozmezí 16,7–25 %. (94, 95, 96).

4.1.1 Faktory zvyšující riziko nástupu VTE

Manifestace tromboembolického stavu vykazuje multifaktoriální podstatu, kde je však dědičná komponenta stanovena až na 60 % (pro selektované pacienty s první epizodou trombózy do 45 let) (97). Mezi vnější faktory zvyšující riziko manifestace tromboembolické nemoci uvádí Souto et al. 2000 (97) operace, imobilizaci, kouření, věk, nádorová onemocnění a u žen pak zejména užívání orální antikoncepce a těhotenství. Z geneticky podmíněných faktorů to jsou zejména, v populaci poměrně časté SNP – faktor V Leiden a mutace faktoru II protrombinu G20210A. Vzácným, ale velmi závažným predisponujícím faktorem k nástupu VTE je deficit antikoagulačních proteinů S (PS), proteinu C (PC) a antitrombinu (AT).

4.1.2 Genetické predispozice VTE

4.1.2.1 Faktor V Leiden a protrombin G20210A

Příčinou hereditární trombofilie jsou poškozující varianty v genech, jejichž proteinové produkty jsou zapojeny do procesu srážení krve. Rutinní genetická diagnostika je prováděna

pouze u SNP faktoru V Leiden a protrombinu G20210A – u jejich nositelů je zvýšené riziko vzniku hyper koagulačních stavů.

Varianta faktor V Leiden v genu *F5* způsobuje záměnu aminokyseliny argininu za glutamin na pozici 506. Varianta vede ke ztrátě štěpného místa pro aktivovaný PC, tím se zvyšuje riziko nástupu VTE v důsledku zvýšení tvorby trombinu (98, 99). Důvodem provádění rutinní diagnostiky je poměrně vysoká populační frekvence dané alely – u varianty faktor V Leiden literatura uvádí frekvenci 2–10 % (100, 101, 102). Pokud je varianta přítomna v heterozygotním stavu je riziko nástupu VTE oproti běžné populaci zvýšeno 3–8x (98, 101, 103, 104).

Varianta protrombinu G20210A, která se nachází v genu *F2*, představuje tranzici guaninu za adenin v 3' nepřekládané oblasti. Záměna má za následek mírně zvýšenou hladinu protrombinu v plazmě, bez změny vlastní molekuly (98, 99). U varianty protrombinu G20210A byla pozorována populační frekvence 1–4 %, pro Evropskou populaci je uvažována spíše nižší hodnota daného rozmezí (98, 102, 104, 105). Nositelé této varianty v heterozygotním stavu mají oproti běžné populaci 3–4 vyšší riziko vzniku VTE (98, 103, 105).

Laboratorní diagnostika využívala a stále využívá metod navržených pro rychlou a levnou detekci známých SNP: PCR-RFLP, Strip Assay®, alelově-specifickou PCR, TaqMan real-time PCR a SNaPshot® (106, 107, 108, 109, 110, 111).

4.1.2.2 Deficit přirozených antikoagulačních proteinů

Vrozený deficit antikoagulačních proteinů S, C a AT je vzácným, avšak závažným rizikovým faktorem dědičné trombofilie (98, 103, 112). Za poruchu funkce přirozených antikoagulantů u hereditárních forem deficitů těchto proteinů mohou poškozující varianty v genech, které tyto proteiny kódují.

4.1.2.2.1 Deficit proteinu S

PS řadíme mezi vitamín K-dependentní plazmatické proteiny. Protein je kódován genem *PROS1*, který je lokalizován na 3q11.1 a dle transkriptu ENST00000394236.9 (délka transkriptu 3 556 bp) se skládá z 15 exonů a obsahuje 676 aminokyselinových zbytků. Nejčastěji je exprimován v játrech a srdci. U lidí se protein nachází ve volné formě ze 40 %,

tento podíl může vytvořit aktivní komplex s PC. Ze 60 % je vázaný k β -řetězci C4b vazebného proteinu (113, 114). Hlavní funkce PS je kofaktorová, působí jako zesilovač aktivovaného PC, který štěpí faktor V, působí tedy jako inhibitor koagulace. Byla však popsána také přímá antikoagulační aktivita pomocí vyvazování vazebných míst na fosfolipidové membráně určených pro prokoagulační faktory (114).

Incidenci deficitu PS v kohortě pacientů s VTE ve věkové skupině < 45 let stanovil De Stefano et al. 1996 (115) na 1,4–7,5 %. V běžné Evropské populaci se prevalence hereditárního deficitu PS odhaduje v rozmezí 0,1–0,5 % (59, 112, 114). I přes nízkou prevalenci se jedná o závažný rizikový faktor nástupu VTE, nositelé kauzální varianty mají 5–10x vyšší riziko oproti běžné populaci (59, 103, 112, 114). Dědičnost je uváděna ve většině případů jako autozomálně dominantní (AD) s redukovanou penetrancí a variabilní expresivitou. Vzácně je pozorován také autozomálně recesivní (AR) typ dědičnosti – OMIM 176880 (<https://omim.org/> cit. 21. 9. 2022).



Obr. 9. Rozložení poškozujících variant v genu *PROS1* dle prohlížeče VarSome.

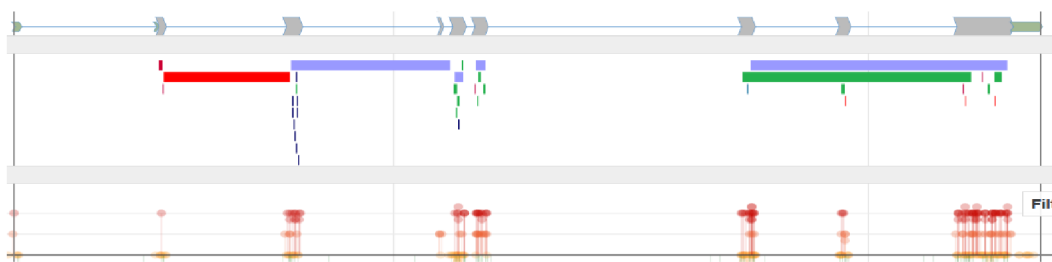
(Tmavě červené tečky označují patogenní varianty; světle červené tečky označují pravděpodobně patogenní varianty; oranžové tečky označují varianty nejasného významu)

4.1.2.2.2 Deficit proteinu C

PC je vitamín-K dependentní plazmatický protein, kódovaný genem *PROC*, který je lokalizován na pozici 2q14.3. Dle transkriptu ENST00000234071.3 (délka transkriptu 1 773 bp) obsahuje 9 exonů, které kódují 461 aminokyselinových zbytků. Exprese probíhá především

v játrech, ale v menší míře také v ledvinách. Protein se v krevní plazmě nachází jako inaktivní serinová proteáza, která je aktivována v reakci na zvýšenou hladinu trombinu. Aktivace probíhá na endotelovém receptoru PC, který je kódován genem **PROCR**. Hlavní biologická funkce spočívá v aktivním štěpení, inaktivaci aktivovaných prokoagulačních faktorů V a VIII. Jeho funkce zesiluje působení PS a faktoru V (98, 116).

V kohortě pacientů s VTE ve věkové skupině < 45 let byla pozorována incidence deficitu PC v rozmezí 1,4–8,6 % (115), což je výrazně více ve srovnání s prevalencí v Evropské populaci 0,15–0,3 % (59, 112). Z uvedeného vyplývá, že se jedná o vzácnou formu dědičného onemocnění, které je však významným rizikovým faktorem VTE. Pacienti s kauzální genetickou variantou mají 5–10x vyšší riziko nástupu VTE oproti běžné populaci (59, 103, 112). U většiny pacientů byla pozorována AD dědičnost s redukovanou penetrancí a variabilní expresivitou. Vzácně je pozorován také AR typ dědičnosti – OMIM 612283 (<https://omim.org/cit.21.9.2022>).



Obr. 10. Rozložení poškozujících variant v genu *PROC* dle prohlížeče VarSome.

(Tmavě červené tečky označují patogenní varianty; světle červené tečky označují pravděpodobně patogenní varianty; oranžové tečky označují varianty nejasného významu)

4.1.2.2.3 Deficit antitrombinu

AT je plazmatický protein kódovaný genem **SERPINC1**. Gen je situován do chromozomové pozice 1q25.1. Podle transkriptu ENST00000367698.3 (délka transkriptu 1 594 bp) obsahuje 7 exonů, které kódují 464 aminokyselinových zbytků. Gen je exprimován výhradně v játrech. AT je hlavním inhibítozem serinových proteáz, především lyzuje trombin aktivovaný faktor X. Dále také blokuje aktivované faktory XII, XI, IX, které se řadí mezi serinové proteázy. Účinek AT se výrazně zvyšuje v přítomnosti heparinu (59, 117).

De Stefano et al. 1996 (115) uvádí incidenci pacientů s deficitem AT v rozmezí 0,5–4,9 %, data se vztahují ke kohortě pacientů s VTE mladších 45 let. V běžné Evropské populaci je pak prevalence nižší 0,02–0,2 % (59, 98, 112, 115). Pacienti s dědičným deficitem AT mají riziko

vzniku VTE 5–16x vyšší oproti zdravé populaci (59, 98, 112). U tohoto geneticky podmíněného onemocnění je uváděn AD typ dědičnosti s redukovanou penetrancí a variabilní expresivitou. Vzácně je pozorován také AR typ dědičnosti – OMIM 613118 (<https://omim.org/> cit. 21. 9. 2022).



Obr. 11. Rozložení poškozujících variant v genu *SERPINC1* dle prohlížeče VarSome.

(Tmavě červené tečky označují patogenní varianty; světle červené tečky označují pravděpodobně patogenní varianty; oranžové tečky označují varianty nejasného významu)

4.1.2.3 Molekulárně genetická diagnostika u deficitů proteinů S, C a antitrombinu

S ohledem na velmi nízkou populační prevalenci hereditárních deficitů PS, PC a AT, jsou tyto řazeny mezi vzácná genetická onemocnění. Z toho vyplývá nutnost preselekcce pacientů vhodných pro genetické testování (112). Genetická diagnostika u těchto deficiencí využívá metody, které jsou vhodné pro detekci očekávaných kauzálních variant. Nejčastěji se jedná o SNV s absencí mutačních hot-spots oblastí. Vzácně jsou popsány také CNV, které jsou detekovány v genech kódujících zájmové proteiny. Na jejich detekci se využívala, a stále hojně užívá, kombinace Sangerovy sekvenace a metody MLPA (59, 114). Pozorované spektrum kauzálních genových variant a časová náročnost těchto molekulárně diagnostických metod je důvodem zavedení metodiky vysokokapacitního sekvenování do diagnostiky těchto geneticky podmíněných deficitů. Pilotní práci v tomto směru publikoval Lotta et al. 2012 (118).

5 Materiál a metodika

V následujících částech práce je prezentována metodika a vyhodnocení diagnostického postupu založeného na vysokokapacitním sekvenování, vhodného pro malý počet genů, pro rutinní diagnostiku pacientů s podezřením na vrozený deficit PS, PC a AT. V České republice se jedná pravděpodobně o první aplikaci metodiky NGS do diagnostiky vzácných genetických variant u trombofilních stavů.

5.1 Indikační kritéria

Úspěšný záchyt kauzálních genových variant vyžaduje precizně preselektovanou skupinu pacientů s podezřením na vrozený deficit jednoho z antikoagulačních proteinů. Byla využita indikační kritéria, která shrnuje Colucci and Tsakiris 2020 (112) viz. Tab 2.

Tab. 2. Navrhovaná indikační kritéria pro testování VTE, přizpůsobeno dle Colucci and Tsakiris 2020 (112).

Muži a ženy
VTE neznámé příčiny < 50 let
VTE na neobvyklém místě
První VTE a pozitivní rodinná historie VTE
Mladí pacienti s arteriální ischemií způsobené embolií
Ženy
VTE během těhotenství nebo po porodu
VTE následkem užívání orální antikoncepce nebo hormonální terapie
VTE před nasazením hormonální terapie
Vícečetné nevysvětlitelné těhotenské ztráty
Mladé pacientky s pozitivní rodinnou historií VTE před předepsáním orální antikoncepce

Pro naše zaměření, výhradně na genetické testování pacientů se suspektním deficitem vybraných antikoagulačních proteinů, daná indikační kritéria rozšiřujeme o:

- 1) Vyloučení běžně testovaných polymorfizmů faktor V Leiden a G20210A**
- 2) U pacienta musí být opakovaně prokázána nízká aktivita daného antikoagulačního proteinu na základě příslušného funkčního testu**

Pokud pacient splnil tato kritéria byl na našem pracovišti otestován na přítomnost zárodečných genových variant na platformě Ion S5™ (Thermo Fisher Scientific).

5.1.1 Funkční testy

Využité funkční analýzy jsou pro každý ze sledovaných proteinů specifické – tyto byly prováděny na externích pracovištích, které se odpovídající diagnostikou zabývají. Cut-off hodnoty jsou stanoveny empiricky, mohou se v rámci různých pracovišť lišit. Vliv hodnot funkčních testů na míru detekce mutací je součástí kapitoly – Diskuze. Uvedené hodnoty jsou stanoveny v rámci Fakultní nemocnice Olomouc. Přehled užitých testů shrnuje Tab. 3.

Tab. 3. Princip testů měřící aktivitu sledovaných proteinů s příslušným cut-off.

Aktivita proteinu	Typ funkčního testu	Cut-off (%)
Volný protein S	Latexová imunoanalýza pro volný protein S	< 53
Protein C	Testy užívající chromogenního substrátu nebo aktivovaného částečného tromboplastinového času	< 72
Antitrombin	Testy založeny na přidání nadbytku aktivovaného faktoru X nebo protrombinu – byly využity chromogenní testy	< 75

5.2 Kohorta pacientů indikovaných na genetické testování

Za využití výše uvedených indikačních kritérií bylo v období let 2016–2021 pro sekvenaci indikováno 31 probandů, z nichž bylo 21 žen a 10 mužů. Nejstarším otestovaným byl muž ročník narození 1932, nejmladší byl muž narozen v roce 2019. Pacienti starší 45 let byli do kohorty zahrnuti z důvodu manifestace tromboembolické nemoci do 45 roků života, pozitivní rodinné anamnézy a nízké aktivity určitého antikoagulačního proteinu. V dané kohortě bylo 7 pacientů testováno na suspektní vrozený deficit PS, 13 pro deficit PC a 11 pro suspektní dědičný deficit AT. Pacienti pro genetické testování byli indikováni z Fakultní nemocnice Olomouc, HEMACENTRUM, spol. s r.o. Plzeň a Fakultní nemocnice Martin.

V závislosti na suspektní deficit přirozeného antikoagulačního proteinu byl u pacientů pomocí NGS vyšetřen odpovídající gen, který daný protein kóduje – *PROS1*, *PROC*, *SERPINC1*. U

pacientů s deficitem PC byl do panelu zařazen kromě genu *PROC*, také gen *PROCR*, který kóduje endotelový receptor PC.

Všichni otestovaní pacienti podepsali informovaný souhlas s genetickým vyšetřením. Společně s probandy byli v některých případech konzultováni klinickým genetikem také rodinní příslušníci, kteří byli v případě nálezu kauzální varianty otestováni na nosičství dané varianty Sangerovou metodou. Tito jsou v práci uvedeni pouze pokud byli využiti v rámci segregací analýzy u novel variant.

5.3 Izolace DNA

Primárním materiálem bylo 9 ml periferní krve odebírané do zkumavek s EDTA. Izolace a purifikace DNA z lymfocytů periferní krve probíhala pomocí QIAcube (QIAGEN) automatického izolátoru, za využití QIAamp DNA Blood Mini Kitu nebo QIAamp DNA Mini QIAcube Kit (oba QIAGEN). Využívané kity vyžadují vstupní množství 200 μ l periferní krve a umožňují volitelný eluční objem 50–200 μ l (www.qiagen.com cit. 30. 9. 2022). Pro izolaci byl na přístroji QIAcube vybrán program QIAamp DNA Blood or body fluid (nastaven eluční objem 100 μ l) a dále se postupovalo dle pokynů na monitoru přístroje.

5.3.1 Měření koncentrace DNA

5.3.1.1 Spektrofotometr

Koncentrace izolované DNA byla primárně měřena pomocí spektrofotometru NanoDrop™ 2000/2000c (Thermo Fisher Scientific), kde výrobce uvádí limit detekce pro dsDNA od 2 ng/ μ l do 15 000 ng/ μ l (www.thermofisher.com cit. 30. 9. 2022). Před měřením je nutné stanovení nulové hodnoty – BLANK. Využíváme vzorek nosného roztoku (eluční pufr z QIAamp izolačního kitu), poté můžeme měřit koncentraci vzorku pacienta – do přístroje pipetujeme 1,5 μ l vzorku.

Popsané rozmezí detekce je dostatečné pro běžné PCR aplikace a metodu MLPA. Vstupní koncentrace DNA pro přípravu knihoven dle manuálu (MAN0013432) pro Ion AmpliSeq Library Preparation na Ion Chef System, je optimálně 0,67 ng/ μ l, což je pod detekčním limitem daného spektrofotometru.

5.3.1.2 Fluorimetr

Pro precizní stanovení koncentrace vstupní DNA pro MPS byl využit Qubit® 2.0 Fluorometer (Thermo Fisher Scientific). Při využití Qubit® dsDNA HS Assay Kitu (Thermo Fisher Scientific) lze kvantifikovat velmi citlivě dsDNA (www.thermofisher.com cit. 30. 9. 2022). V závislosti na objemu vzorku je souprava navržena pro přesné měření koncentrace DNA od 0,005 do 120 ng/μl, což je pro námi požadované rozmezí vyhovující. Pracovní roztok byl připraven smícháním koncentrované assay reagent (barvičky) a pufru do celkového objemu 200 μl v poměru 1:199 pro každý měřený vzorek, případně kalibrační standard. Dále již pipetujeme do 0,5 ml měřících tenkostěnných čirých zkumavek, které jsou s přístrojem kompatibilní.

Příprava kalibračních standardů: 10 μl daného standardu + 190 μl pracovního roztoku

Příprava vzorků: 2 μl vzorku DNA + 198 μl pracovního roztoku

Po napipetování krátce promícháme na vortexu a stočíme. Ponecháme 2 minuty inkubovat při pokojové teplotě. Nejprve kalibrujeme přístroj pomocí standardů a poté kvantifikujeme koncentraci DNA u vzorků.

5.3.2 Sekvence na platformě Ion S5™

Pro vysokokapacitní sekvenaci byla zvolena Ion AmpliSeq™ technologie pro TS. Tento přístup nám dovoluje designovat pro každý sekvenační běh unikátní Ion AmpliSeq™ On-Demand panel, který je vhodný zejména pro testování menších skupin genů obvykle < 50. V každém sekvenačním běhu byli nakombinováni pacienti s deficitem přirozených antikoagulačních proteinů s pacienty indikovanými na jinou diagnózu dle požadavku lékařů (např. deficit biotinidázy, Wilsonova choroba, von Willebrandova choroba, RASopatie aj.). Tato metoda je vhodná k detekci SNV, malých inzercí, delecí, duplikací do 20 bp. Při dostatečném pokrytí cílových oblastí je provedena také CNV analýza.

Během celé analýzy bylo postupováno podle příslušných manuálů, které jsou k dispozici k daným kitům (Thermo Fisher Scientific) viz. Tab. 4.

Tab. 4. Využité kity a příslušné manuály dostupné z (www.thermofisher.com cit. 1. 10. 2022)

Název kitu	Účel	Název manuálu	Číslo manuálu
Ion AmpliSeq™ Kit for Chef DL8	Reagencie a spotřební materiál pro automatickou přípravu knihoven v přístroji Ion Chef™	Ion AmpliSeq Library Preparation on the Ion Chef System	MAN0013432
Ion 510™ & Ion 520™ & Ion 530™ Kit – Chef	Reagencie a spotřební materiál pro automatickou tvorbu templátu a chip loading v přístroji Ion Chef™	Ion 510™ & Ion 520™ & Ion 530™ Kit – Chef	MAN0016854
	Reagencie a spotřební materiál pro sekvenaci na přístroji Ion S5™		

5.3.2.1 Design amplikonových knihoven

Pro návrh primerů byl využit software Ion AmpliSeq™ Designer (Thermo Fisher Scientific), kde byl zvolen Ion AmpliSeq™ On-Demand panels (Thermo Fisher Scientific). V tomto nastavení byl navržen vlastní / unikátní genový panel, který na základě indikace zahrnoval následující geny: deficit PS – *PROS1*; deficit proteinu C – *PROC* a *PROCR*; deficit AT – *SERPINC1*. Geny obsažené v panelech jsou předem otestovány v laboratoři pro ověření požadovaných vlastností. Amplikony jsou *in-silico* navrženy ke 100% pokrytí kódujících oblastí a rozhraní exonů / intronů. Design vycházel z referenčního lidského genomu hg19, navržené amplikony byly v rozmezí 125–275 bp. Pro tuto aplikaci byly navrženy primery vždy 2-poolové. Ze softwaru Ion AmpliSeq™ Designer je možnost stažení pro každý design unikátního souboru BED. Ten je nezbytný pro analýzu sekvenačních dat. Program je dostupný z www.ampliseq.com.

5.3.2.2 Příprava knihovny

Příprava sekvenační knihovny je automatická pomocí přístroje Ion Chef™ (Thermo Fisher Scientific). Vstupní množství DNA je velmi nízké – 10 ng v objemu 15 µl. Ion Chef™ vyžaduje pro přípravu knihoven kit Ion AmpliSeq™ Kit for Chef DL8 (Thermo Fisher Scientific), který obsahuje materiál pro přípravu knihoven 32 pacientů. Při přípravě knihoven a obsluze přístroje postupujeme dle návodu viz. Tab. 4. Zde je klíčové především nastavení počtu PCR cyklů a času annealingu / extenze v závislosti na počtu párů primerů obsažených v poolech unikátního On-Demand panelu viz. Tab. 5. Tučně jsou vyznačeny využívané hodnoty.

Tab. 5. Doporučený počet amplifikačních cyklů. Upraveno dle manuálu MAN0013432 dostupného z (www.thermofisher.com cit. 1. 10. 2022).

Počet párů primerů na pool	Doporučený počet amplifikačních cyklů	Annealing / extenze (min)
12–24	22	4
25–48	21	4
49–96	20	4
97–192	19	4
193–384	18	4
385–768	17	4
769–1 536	16	8
1 537–3 072	15	8
3 073–6 144	14	16
6 145–24 576	13	16

5.3.2.3 Tvorba templátu, plnění čipu

Tyto kroky jsou automatizovány pomocí přístroje Ion Chef™, který k tomuto kroku využívá kitu Ion 510™/520™/530™ Kit-Chef (Thermo Fisher Scientific). Sekvenační čipy byly zvoleny dle počtu pacientů zařazených do jednoho sekvenačního běhu. Ion 520 pro 8 pacientů. Ion 530 pro 16 pacientů. V tomto kroku byl také v prostředí Torrent Suite softwaru vytvořen pracovní list „sample sheet“, kde byla nastavena reference hg19, nahrán příslušný soubor BED a nastaven templátovací protokol na 400 bp. Dále bylo postupováno dle příslušného návodu viz. Tab. 4.

5.3.2.4 Sekvenace

Sekvenace byla provedena na platformě Ion S5™, za využití sekvenačních reagentů z kitu Ion 510™/520™/530™ Kit-Chef. K zajištění dostatečného „end to end“ pokrytí jednotlivých ampliconů nastavíme 550 cyklů omytí. Při nastavení a obsluze přístroje bylo postupováno dle příslušného manuálu viz. Tab. 4.

5.3.2.5 Analýza sekvenačních dat

5.3.2.5.1 Torrent Suite software

Primární a sekundární analýza byla provedena dle základního nastavení v prostředí Torrent Suite™ software (Thermo Fisher Scientific). Pro každý sekvenační běh software poskytne data kontroly kvality sekvenace a pokrytí jednotlivých vzorků (pro cílové amplikony a báze zvlášť). Klíčovou metrikou analýzy pokrytí je pro nás pokrytí jednotlivých ampliconů: minimální průměrné pokrytí ampliconů ≥ 200 čtení; pokrytí ampliconů nejméně 20 čteními $\geq 98\%$.

Výstupními datovými formáty jsou uBAM, ale také FASTQ – ten je možné stáhnout v rámci modulu – FileExporter. Automaticky probíhá také přiřazování čtení k zadané referenční sekvenci (hg19) a cílovým oblastem (unikátní soubor BED pro každý design On-Demand panelu) – výstupním formátem je BAM a BAI. V rámci analýzy dat v Torrent Suite softwaru je automaticky provedena analýza BAM formátu v integrovaném modulu TVC – parametry pro VC byly ponechány základní, které jsou doporučené. Výstupní formáty daného modulu jsou VCF a XLS. Vybrané parametry základního nastavení variantCalleru viz. Tab. 6.

Tab. 6. Vybrané parametry využitých softwarů pro variant calling – základní nastavení.

	Torrent Suite		Ion Reporter		NextGene	
	INDEL	SNP	INDEL	SNP	INDEL	SNP
MAF	0,25	0,18	0,1	0,1	0,2	0,2
MQ	20	15	10	10	20	20
MC	10	45	10	5	5	5
MCES	3	0	4	0	0	0
MSB	0,95	0,94	0,95	0,98	0,85	0,85

(MAF – minimum allele frequency; MQ – minimum quality; MC – minimum coverage; MCES – minimum coverage of either(each) strand; MSB – maximum strand bias)

5.3.2.5.2 Ion Reporter

Tento software od firmy Thermo Fisher Scientific byl využit k sekundární a terciální analýze sekvenačních dat získaných z Torrent Suite softwaru. Pacientská sekvenační data (formát BAM, VCF) byla do Ion Reporteru nahrána pomocí modulu IonReporterUploader. Ion Reporter je dostupný online na <https://ionreporter.thermofisher.com/>.

5.3.2.5.3 Anotace VCF souborů

V softwaru Ion Reporter byla provedena anotace VCF souborů nahraných ze softwaru Torrent Suite. Parametry byly ponechány v základním nastavení. Tímto způsobem byli posuzováni všichni pacienti. Ze sady anotací využíváme zejména: *in-silico* predikce (PolyPhen, SIFT), PhyloP, dbSNP, ClinVar, MAF a 5000 exomes European MAF.

5.3.2.5.4 Anotace z BAM souborů, CNV analýza

Software provádí anotaci variant a CNV analýzu z nahraných BAM souborů. Pro provedení CNV analýzy je nutné manuálně definovat vzorky (z nahraných souborů BAM), kde u všech musí být nastaveno pohlaví – muž. Následuje tvorba tzv. „CNV baseline“ (referenční rozmezí proti kterému je CNV pacienta stanovena), do které vybereme všechny vzorky ze sekvenačního běhu (minimálně 10), pomocí kterých chceme stanovit CNV.

Dalším krokem je tvorba sekvenačního workflow – jeho nastavení: referenční sekvence hg19, nahrání unikátního souboru BED, výběr unikátní CNV baseline pro danou skupinu pacientů. Parametry pro anotaci: mapování, bamstats, hledání variant – ponechány v základním nastavení. Pro zvýšení senzitivity nastaven parametr CNV Finding – CNV Sensitivity – HIGH. Pomocí takto vytvořeného workflow byly vzorky analyzovány. Ze sady anotací využíváme zejména: *in-silico* predikce (PolyPhen, SIFT), PhyloP, dbSNP, ClinVar, MAF a 5000 exomes European MAF.

5.3.2.5.5 Postup hodnocení / interpretace

Nalezené varianty typu SNV, delece, inserce a duplikace (< 20 bp) byly primárně hodnoceny z dat generovaných programem Ion Reporter z BAM souborů. Hodnocení proběhlo na základě popsaných anotačních nástrojů jako jsou *in-silico* predikční softwary, MAF, efekt mutace na protein, PhyloP skóre, dbSNP a ClinVar. Následovala kontrola přítomnosti varianty pomocí IGV (např. k vyloučení možného sekvenačního artefaktu) a vyhledání dostupných informací a dat z dalších predikčních nástrojů v prohlížeči VarSome případně UCSC. Při klasifikaci vycházíme také z dostupných informací o pacientově fenotypu a rodinné anamnéze – při interpretaci pak zohledňujeme ACMG / AMP doporučení. Postup při posuzování možné

patogenity nalezených variant viz. Tab. 7. Správný zápis a přítomnost varianty je ověřen kontrolou v anotovaném VCF souboru. Pro zápis odhalených variant jsme použili referenční transkripty: *PROS1* (NM_000313.4); *PROC* (NM_000312.4); *SERPINC1* (NM_000488.4) a *PROCR* (NM_006404.5).

CNV je detekováno z BAM souborů pomocí Ion Reporteru. V případě kauzálního nálezu by následovala confirmace metodou MLPA.

Tab. 7. Postup při posuzování možné patogenity nalezených variant.

Krok 1	Krok 2	Krok 3	Krok 4
Suspektní varianta v Ion Reporter	"Fyzická" kontrola přítomnosti varianty	Kontrola varianty v genomickém prohlížeči	Interpretace při zohlednění ACMG / AMP klasifikace*
<i>In-silico</i> (PolyPhen, SIFT)	IGV	VarSome (doplnění informací o možné patogenitě)	
MAF < 0,01			
Efekt varianty na protein		UCSC	
PhyloP			
ClinVar			

*Zejména u doposud nepopsaných variant klademe důraz na fenotyp pacienta a rodinnou anamnézu

5.3.3 MLPA

Metoda má sloužit pouze jako confirmační v případě nálezu rozsáhlé přestavby ve vyšetřovaných genech CNV analýzou pomocí Ion Reporteru. Pro případnou confirmaci byly k dispozici SALSA MLPA Probemixy (MRC Holland) uvedené v Tab. 8. Pro gen *PROCR* není v současné době k dispozici žádný probemix.

Tab. 8. Probemixy pro případnou confirmaci odhalených CNV ve vyšetřovaných genech.

Vyšetřovaný gen	Název probemixu
<i>PROS1</i>	SALSA MLPA P112 PROS1
<i>PROC</i>	SALSA MLPA P265 PROC
<i>SERPINC1</i>	SALSA MLPA P227 SERPINC1

5.3.4 Sangerova sekvenace

Tato metoda byla využita pouze u pacientů s novel variantou, u kterých byli k dispozici rodinní příslušníci ochotní podrobit se segregací analýze.

Primery byly navrženy, tak aby pokrývali oblast výskytu určité kauzální varianty. V Tab. 9. jsou uvedeny sekvence primerů s annealingovou teplotou, délkou PCR produktu a odpovídající variantou, která má být detekována.

Tab. 9. Primery navržené pro segregací analýzu u probandů s doposud nepopsanou poškozující variantou.

Varianta	F sekvence primeru 5' -> 3'	R sekvence primeru 5' -> 3'	Annealing (°C)	Délka (bp)
<i>PROS1</i> c.1931T>A	TGTCTTTTCCAACTGATGGAC	GGTTGGAAACAGGAAGTCTGAATG	59	243
<i>PROS1</i> c.1468delA	AGCATAAGACTAGAGTGGGCA	ACTCATAATCGAGCCACTGTTT	56	358
<i>PROC</i> c.866C>G	TCTCCGGGTGAACCTTCTTC	GACGAAGGTGCGGTTTCTCT	56	410
<i>PROC</i> c.450C>A	AGGTGAGCTTCCTCAATTGCT	ATTCCTGGGCGATGTATTGGG	56	166
<i>SERPINC1</i> c.398A>T	GCATAATGAGCGCCCCTAGT	TCCCATGTGCATTTACCGCT	59	377

Složení PCR master mixů, teplotní profily reakcí a nastavení sekvenační runu pomocí kapilární elektroforézy na přístroji ABI 3130 je zobrazeno v Tab. 10.–Tab. 17.

Sekvenační data byla hodnocena pomocí software SeqScanner v1.0.

Tab. 10. Složení 1. PCR směsi.

Složky 1. PCR směsi (1 reakce)	Objem (μl)
PPP Master Mix (Top-Bio)	7,5
PCR H ₂ O (Top-Bio)	5,5
Primer F (10pmol)	0,5
Primer R (10pmol)	0,5
DNA	1
Celkem	15

Tab. 11. Teplotní profil 1. PCR.

1. PCR teplotní profil (°C)	Čas (min, sec)	Cykly
95	5	1x
94	0,30	34x
59 / 56 dle Tab. 9.	1,30	
72	1	
72	10	1x
6	∞	

Tab. 12. Enzymatické přečištění – složky.

Složky pro enzymatické přečištění (1 reakce)	Objem (μl)
Alkalická fosfatáza (Thermo Fisher Scientific)	1
Exonukleáza (Thermo Fisher Scientific)	0,5
PCR produkt	5
Celkem	6,5

Tab. 13. Teplotní profil enzymatického přečištění.

Enzymatické přečištění teplotní profil (°C)	Čas (min, sec)	Cykly
37	15	1x
85	15	1x
10	∞	

Tab. 14. Složení 2. PCR směsi.

Složky 2. PCR směsi (1 reakce)	Objem (μl)
PCR H ₂ O (Top-Bio)	4,5
BigDye® Terminator v1.1 (ABI)	1,5
BDX64 buffer (MCLAB)	0,94
Big Dye® Terminator v 3.1 kit (ABI)	0,063
Primer F / R (10pmol)	1
sekvenační templát	2
Celkem	10

Tab. 15. Teplotní profil 2. PCR.

2. PCR teplotní profil (°C)	Čas (min, sec)	Cykly
94	2	1x
96	0,10	34x
50	0,05 (Ramp 1°C/sec)	
60	4	
10	∞	

Tab. 16. Přečištění sekvenačních produktů – složky.

Složky pro přečištění sekvenačních produktů (1 reakce)	Objem (μl)
SAM™ Solution (ABI)	22,5
XTerminator™ Solution (ABI)	5
Sekvenační produkt	5
Celkem	32,5

Tab. 17. Nastavení separačních podmínek na kapilární elektroforéze ABI 3130.

Podmínky sekvenace pro ABI 3130	Hodnota
Separační teplota	60 °C
Doba nástřiku vzorku	12 sec
Napětí při nástřiku vzorku	1,2 kV
Napětí při separaci vzorku	8,5 kV
Doba separace vzorku	2780 sec

5.4 Porovnání VCF vytvořených z rozdílných softwarů

V rámci této práce bylo provedeno porovnání variant z pohledu jejich prevalence / absence v daném VCF souboru. Porovnávány byly VCF vytvořené z Torrent Suite softwaru pomocí integrovaného modulu variantCaller; VCF vytvořené z BAM souborů v softwaru Ion Reporter a VCF, které byly generovány pomocí NextGene softwaru.

Bylo vybráno 10 pacientů z jednoho sekvenčního běhu, ve kterém byli pacienti indikováni pro deficit PS, PC a AT, ale také pro RASopatie (tito podepsali souhlas s anonymním využitím výsledků). Pacienti byli anonymizováni a hodnoceny pouze varianty odhalené v genech *PROS1*, *PROC* a *SERPINC1*.

Nastavení variantCalleru v programu Torrent Suite a Ion Reporteru ponecháno základní, vybrané parametry viz. Tab. 6. výše v textu.

Pro zpracování dat programem NextGene byly využity soubory FASTQ, které byly staženy Torrent Suite pomocí modulu – FileExporter.

Nastavení NextGene: V prvním kroku byl zvolen typ přístroje – Ion Torrent, aplikace – SNP / INDEL discovery a sequence alignment. Nastavení referenční sekvence – Human_v37p10dbSNP135 a nahrán příslušný BED soubor, který je pro každý design On-Demand panelu jedinečný. Vybrané parametry nastavení viz. Tab. 6. výše v textu.

5.5 Statistická analýza

V rámci statistické analýzy byl zkoumán možný vztah hladin funkčních testů u pacientů s potvrzenou kauzální variantou a skupinou pacientů bez nalezené patogenní varianty. K statistickému vyhodnocení možného vztahu hladin funkčních testů k těmto skupinám pacientů byl využit nepárový t-test. Pro nulovou hypotézu: “neexistuje rozdíl v hodnotách funkčních testů mezi pacienty s potvrzenou a pacienty bez potvrzené kauzální genetické varianty“.

Z důvodu rozdílného počtu provedených funkčních testů u jednotlivých pacientů ze souboru, byl t-test vypočítán z mediánů funkčních testů.

Tabulky a grafy byly vytvořeny pomocí softwaru Microsoft Excel, Microsoft PowerPoint a GraphPad. Rodokmen byl sestaven pomocí Progeny (Progeny Genetics).

6 Výsledky

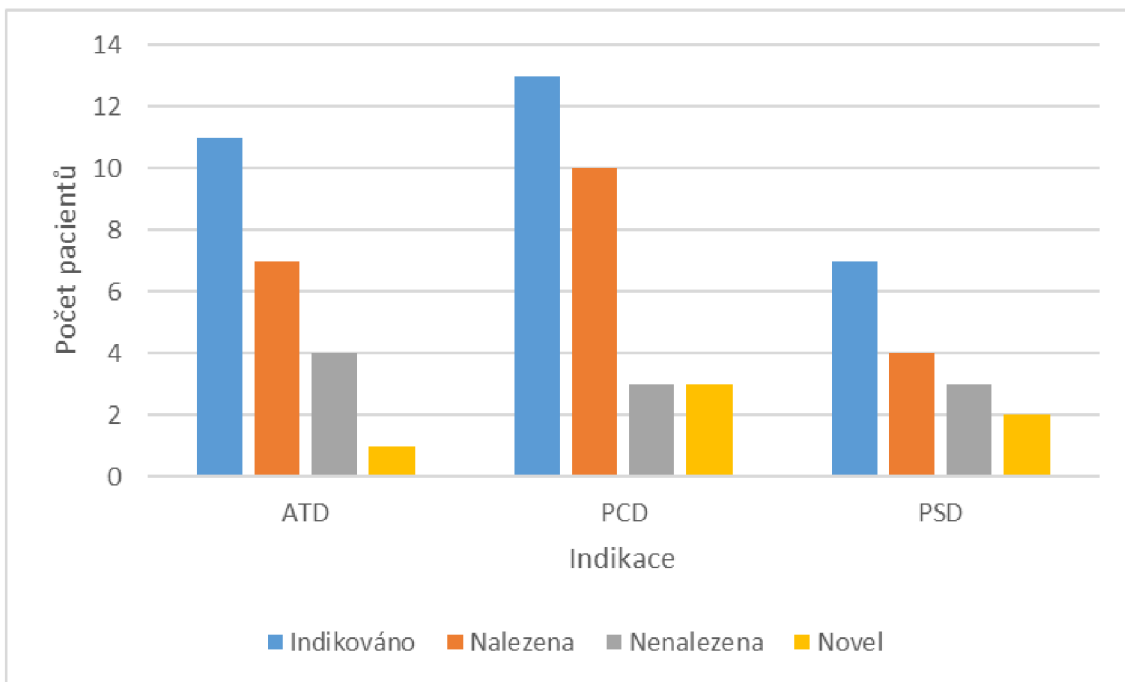
6.1 Souhrnné výsledky u vyšetřených probandů

Do této práce bylo zahrnuto celkem 31 probandů, kteří splnili indikační kritéria pro genetické testování vrozeného deficitu jednoho ze sledovaných antikoagulačních proteinů. Celkem bylo pomocí MPS na platformě Ion S5 odhaleno 21 pravděpodobně patogenních / patogenních variant. To odpovídá míře detekce mutací (MDR) 67,7 %. Kauzální varianty byly zaznamenány u 14 žen a 7 mužů. Z nalezených variant bylo 6 variant dosud nepopsaných tzv. novel. Nejčastějším typem mutací byly missense varianty – 76,2 %. U 10 probandů nebyla odhalena žádná SNV, CNV analýza byla negativní. Souhrnné výsledky jsou shrnuty v Tab. 18. a zobrazeny v Obr. 12. a Obr. 13. Pacienti s nalezenou variantou jsou zobrazeni v Tab. 19. Informace o pacientech bez kauzální varianty shrnuje Tab. 20. Pacienti s detekovanou novel variantou s doplňujícími informacemi pro klasifikaci nalezených variant jsou uvedeni v Tab. 21.

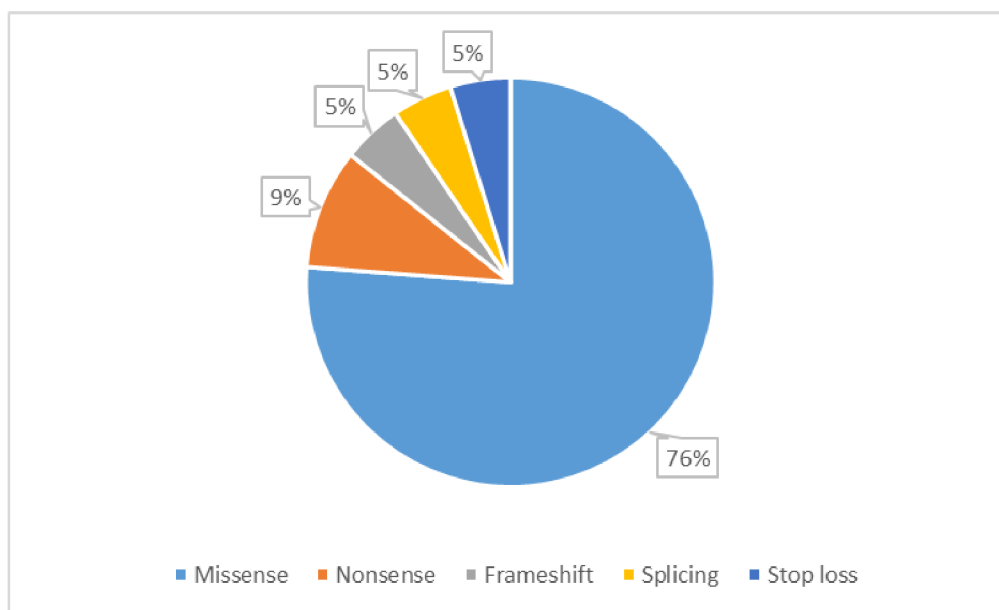
Pomocí t-testu byly porovnány mediány měření funkčních testů u pacientů s nalezenou oproti skupině pacientů bez nalezené kauzální genetické varianty. Signifikantní rozdíl byl zaznamenán pouze u pacientů indikovaných pro deficit PS ($p = 0,0047$) – zde můžeme zamítnout nulovou hypotézu viz. Tab. 18. Distribuce mediánů funkčních testů je zobrazen na Obr. 14.

Tab. 18. Soubor pacientů rozdělený dle příslušné indikace s odpovídající MDR.

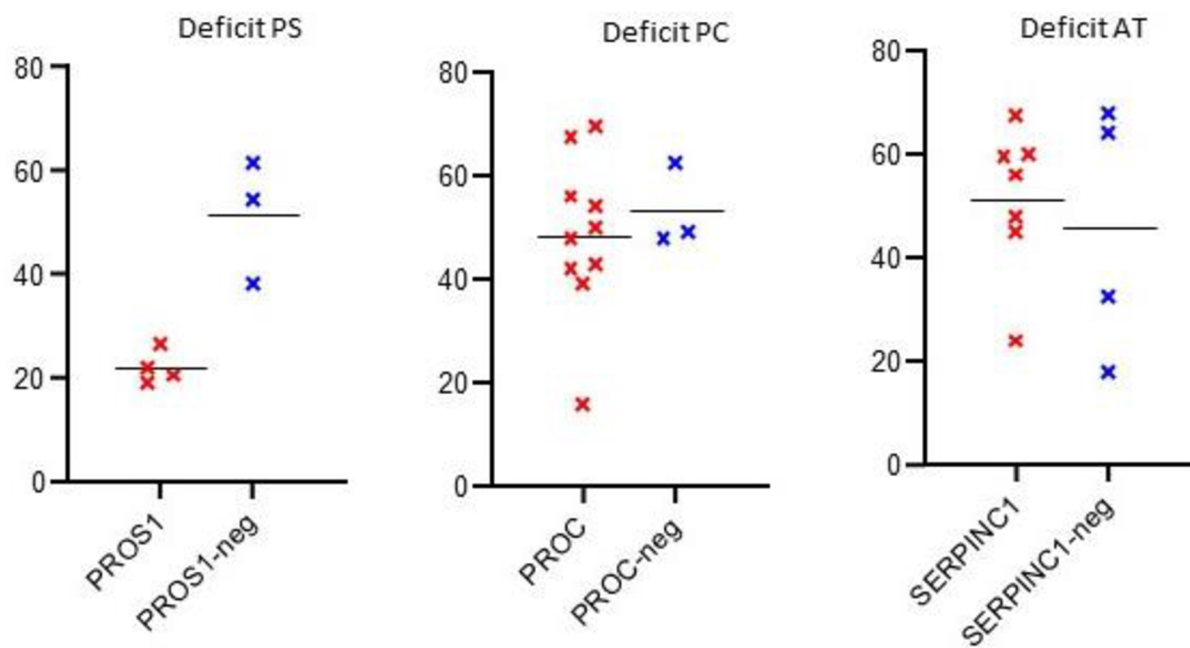
Indikace	Pozitivní vs negativní (T-test)	Testováno / (varianta nalezena)	Missense	Nonsense	Frameshift	Splicing	Stop loss	Novel varianta
ATD	0,6241	11 (7) 63,6%	6	1	–	–	–	1
PCD	0,6286	13 (10) 76,9%	8	1	–	–	1	3
PSD	0,0047	7 (4) 57,1%	2	–	1	1	–	2
Celkem	–	31 (21) 67,7%	16 (76,2%)	2 (9,5%)	1 (4,8%)	1 (4,8%)	1 (4,8%)	6 (28,6%)



Obr. 12. Soubor pacientů rozdělený dle příslušné indikace a nalezených / nenalezených / novel variant.



Obr. 13. Spektrum nalezených kauzálních variant (zaokrouhleno).



Obr. 14. Distribuce mediánů funkčních testů u pacientů dle příslušných indikací – nalezena vs nenalezena varianta.

(červeně – mediány funkčních testů pro pacienty s nalezenou variantou; modře – mediány funkčních testů pro pacienty bez nalezené varianty; osa y – aktivita proteinu v %; osa x – pacienti s nalezenou vs pacienti bez nalezené varianty)

Tab. 19. Kohorta pacientů s nalezenou kauzální variantou.

Ročník narození (Pohlaví)	Funkční test (%)	Funkční test medián (%)	Gen	Varianta	Exon/ Intron	Koodináta hg19	Efekt varianty	Zygozita /novel *	In-silico predikce	Rodinná historie	Klasifikace
1981 (Ž)	60;61;63	60	SERPINC1	c.79T>C p.Trp27Arg	2	1:173884020 rs116581658	missense	het	damaging	Neznámá	likely pathogenic
1983 (Ž)	67;68	67,5	SERPINC1	c.133C>T p.Arg45Trp	2	1:173883966 rs768704768	missense	het	damaging	Ne	likely pathogenic
1993 (Ž)	56	56	SERPINC1	c.391C>T p.Leu131Phe	2	1:173883708 rs121909567	missense	het	damaging	Ano	pathogenic
1973 (M)	54;65	59,5	SERPINC1	c.236G>A p.Arg79His	2	1:173883863 rs121909552	missense	het	damaging	Ano	likely pathogenic
1971 (Ž)	48	48	SERPINC1	c.236G>A p.Arg79His	2	1:173883863 rs121909552	missense	het	damaging	Ano	likely pathogenic
2004 (M)	42;45;47	45	SERPINC1	c.398A>T p.Gln133Leu	2	1:173883701	missense	het *	damaging	Ne	likely pathogenic
1932 (M)	24	24	SERPINC1	c.685C>T p.Arg229Ter	4	1:173879969	nonsense	het	_	Ano	likely pathogenic
1980 (Ž)	50	50	PROC	c.450C>A p.Tyr150Ter	6	2:128180899	nonsense	het *	_	Ano	likely pathogenic
1971 (M)	39;45	42	PROC	c.677A>T p.Gln226Leu	7	2:128183802	missense	het	damaging	Ano	pathogenic
1976 (Ž)	41;45	43	PROC	c.677A>T p.Gln226Leu	7	2:128183802	missense	het	damaging	Ano	pathogenic
1973 (Ž)	49;56;57	56	PROC	c.715G>C p.Gly239Arg	8	2:128184717	missense	het *	damaging	Ano	pathogenic
1988 (Ž)	54	54	PROC	c.759C>A p.His253Gln	8	2:128184761 rs145866973	missense	het	damaging	Ano	likely pathogenic
1971 (Ž)	67;68;71;71	69,5	PROC	c.1301T>C p.Val434Ala	9	2:128186437	missense	het	damaging	Ano	pathogenic
1988 (Ž)	65;70	67,5	PROC	c.1384T>C p.Ter462Gln	9	2:128186520 rs370298954	stop loss	het	_	Ano	likely pathogenic
1978 (Ž)	47;48;59	48	PROC	c.1106C>T p.Pro369Leu	9	2:128186242 rs121109869	missense	het	damaging / tolerated	Ne	pathogenic
1959 (M)	4,4;27,1	15,75	PROC	c.866C>G p.Pro289Arg	9	2:128186002	missense	hom *	damaging	Ano	pathogenic
1994 (Ž)	39	39	PROC	c.1019C>T p.Thr340Met	9	2:128186155 rs766261022	missense	het	damaging	Ano	pathogenic
1963 (M)	22	22	PROS1	c.1155+5G>A	10	3:93611772	splicing	het	_	Ano	pathogenic
1981 (Ž)	9;9;11;12;19; ;20;23;27;29	19	PROS1	c.1468delA p.Ile490fs	12	3:93603596	frameshift delece (INDEL)	het *	_	Ano	likely pathogenic
1953 (Ž)	18;23	20,5	PROS1	c.1916G>A p.Cys639Tyr	15	3:93593204	missense	het	damaging	Ano	likely pathogenic
2002 (M)	22,1;26,5;27, 9	26,5	PROS1	c.1931T>A p.Ile644Asn	15	3:93593189	missense	het *	possibly damaging / damaging	Ano	likely pathogenic

(Ž – žena; M – muž; het – heterozygot; hom – homozygot; * – novel)

Tab. 20. Kohorta pacientů bez nalezené kauzální varianty.

Ročník narození (Pohlaví)	Funkční test (%)	Funkční test medián (%)	Gen	CNV
1982 (Ž)	24;30;30;31; 31;32;33;35; 35;36;37;38	32,5	<i>SERPINC1</i>	neg
1984 (Ž)	58;60;64;66; 68;68;69;70; 70	68	<i>SERPINC1</i>	neg
2019 (M)	18	18	<i>SERPINC1</i>	neg
1974 (Ž)	64	64	<i>SERPINC1</i>	neg
1999 (Ž)	44;54	49	<i>PROC</i>	neg
2003 (M)	62;63	62,5	<i>PROC</i>	neg
1993 (Ž)	41;45;48;48; 49;49	48	<i>PROC</i>	neg
1983 (Ž)	31;45,6	38,3	<i>PROS1</i>	neg
1952 (M)	61,2;61,9	61,55	<i>PROS1</i>	neg
1990 (Ž)	45;64	54,5	<i>PROS1</i>	neg

(Ž – žena; M – muž; neg – negativní)

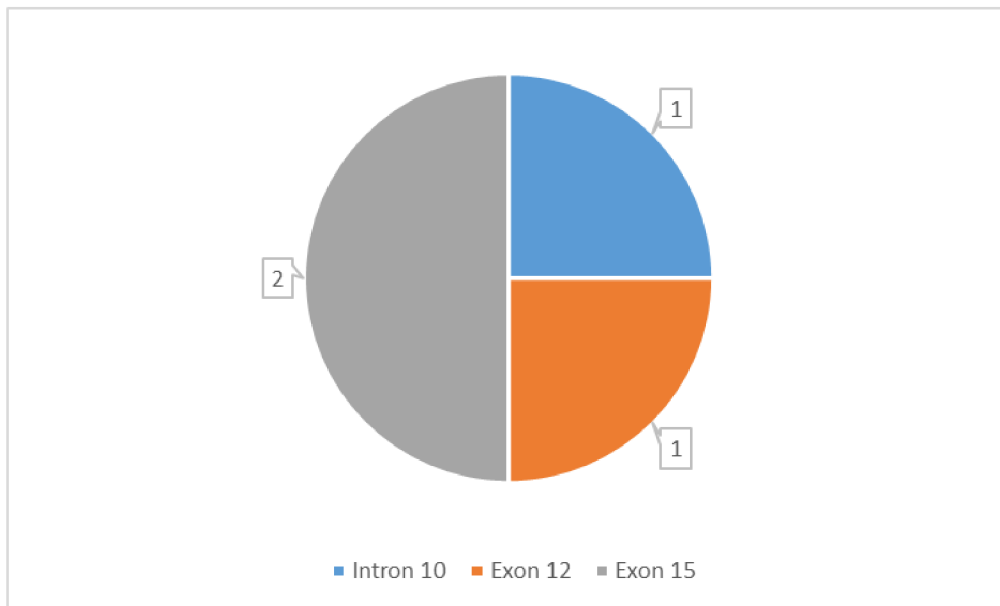
Tab. 21. Pacienti s dosud nepopsanou kauzální variantou.

Ročník narození (pohlaví)	Gen	Varianta	Efekt varianty	Zygotita	<i>In-silico</i> predikce	Rodinná historie	ACMG/AMP
2004 (M)	<i>SERPINC1</i>	c.398A>T p.Gln133Leu	missense	het	damaging	Ne	PM2, PM1, PM5, PP2, PP3
1980 (Ž)	<i>PROC</i>	c.450C>A p.Tyr150Ter	nonsense	het	–	Ano	PVS1, PM2
1973 (Ž)	<i>PROC</i>	c.715G>C p.Gly239Arg	missense	het	damaging	Ano	PM1, PM2, PS1, PP2, PP3
1959 (M)	<i>PROC</i>	c.866C>G p.Pro289Arg	missense	hom	damaging	Ano	PM1, PM5, PM2, PP2, PP3
1981 (Ž)	<i>PROS1</i>	c.1468delA p.Ile490fs	Frameshift Deletion (INDEL)	het	–	Ano	PVS1, PM2
2002 (M)	<i>PROS1</i>	c.1931T>A p.Ile644Asn	missense	het	possibly damaging/ damaging	Ano	PM1, PM2, PP2, PP3

(Ž – žena; M – muž; het – heterozygot; hom – homozygot; ACMG/AMP – důkazy o patogenitě varianty)

6.2 Výsledky u probandů s deficitem PS

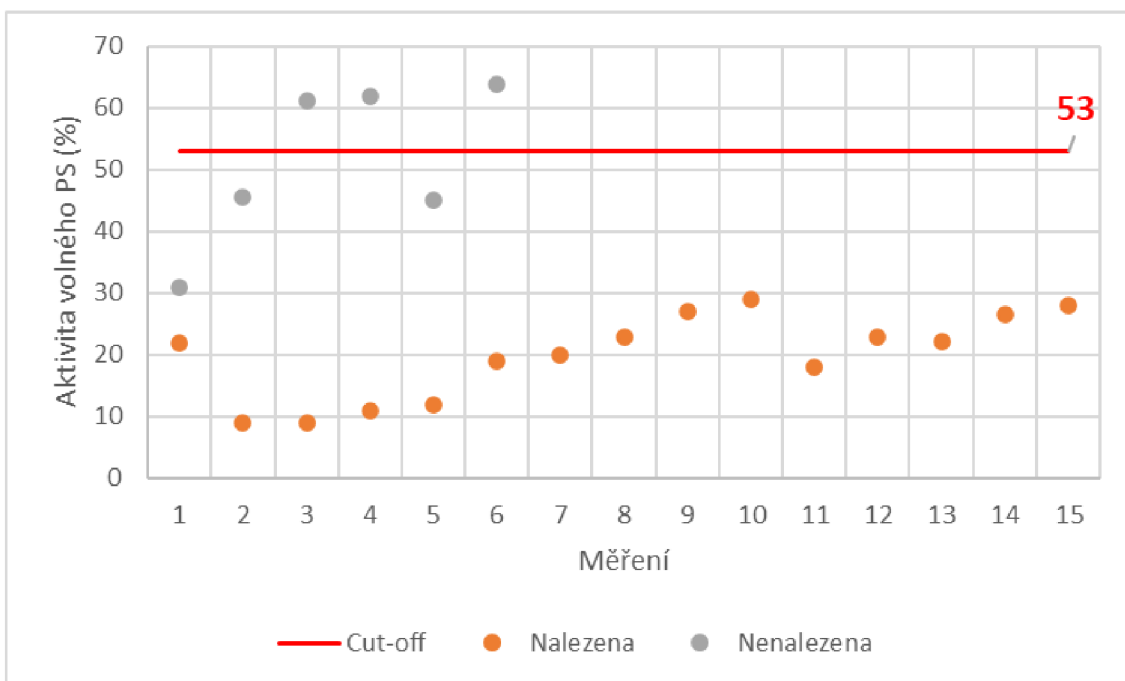
Pro suspektní dědičný deficit PS bylo na MPS genu *PROS1* indikováno 7 probandů (4 ženy a 3 muži). Kauzální varianta byla odhalena u 4 probandů viz. Tab. 19., kdy všichni měli pozitivní rodinou historií tromboembolických stavů. MDR bylo u této kohorty pacientů 57,1 %. Nejvyšší četnost (2) měly varianty v 15. exonu genu *PROS1* zobrazeno na Obr. 15. Nejčastěji byly zjištěny missense varianty (2) dále pak frameshift (1) a splicing (1). Všechny uvedené byly nalezeny v heterozygotním stavu.



Obr. 15. Četnost nalezených kauzálních variant v exonech / intronech genu *PROS1*

V této patientské kohortě byl zjištěn signifikantní rozdíl ($p = 0,0047$) mediánů měření funkčních testů mezi pacienty se zjištěnou a pacienty bez odhalené kauzální varianty viz. Obr. 14. a Tab. 18.

Všechna dostupná měření aktivity volného PS s příslušným cut-off testu jsou vizualizovány na Obr. 16.

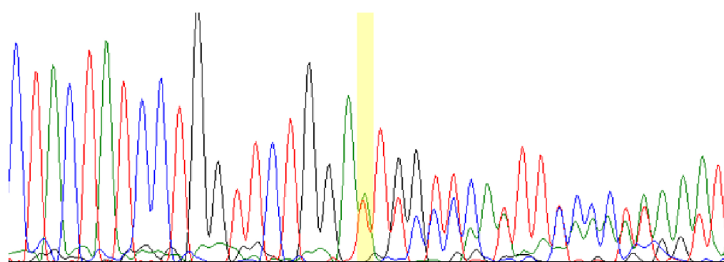


Obr. 16. Hodnoty všech provedených funkčních testů u pacientů indikovaných pro deficit PS s vyznačeným cut-off.

6.2.1 Dosud nepopsané varianty u probandů s deficitem PS

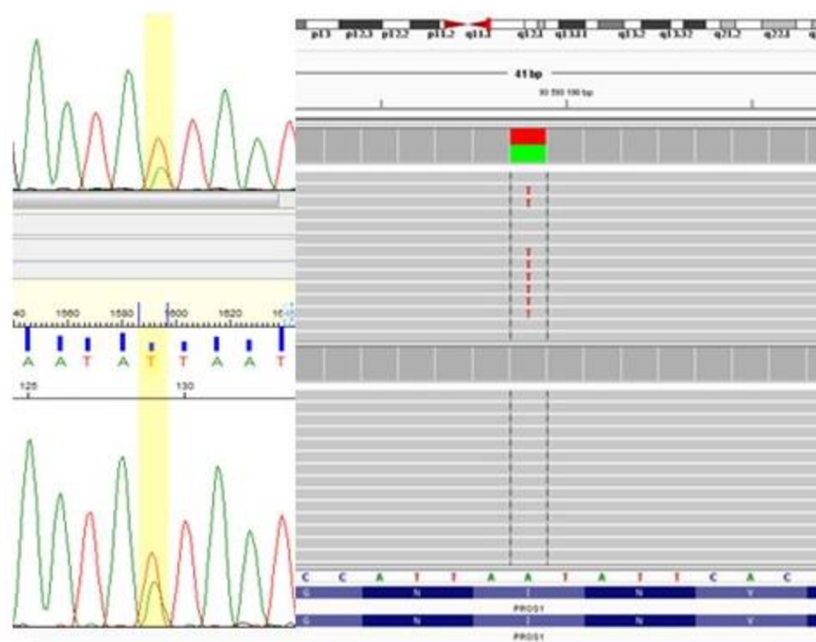
V rámci kohorty pacientů s deficitem PS byly popsány dvě novel varianty - c.1468delA (p.Ile490fs) a c.1931T>A (p.Ile644Asn). Informace o variantách uvedeny v Tab. 21.

Varianta **c.1468delA (p.Ile490fs)** byla nalezena u probandky narozené v roce 1981, která byla indikována pro opakovaně nízkou aktivitu volného PS (9–29 %). Tato varianta byla Sangerovou sekvenací potvrzena také u matky probandky a klasifikována jako pravděpodobně patogenní. Klasifikace varianty proběhla na základě absence alternativní alely v populačních databázích, dopadu na protein – frameshift varianta a pozitivní rodinné anamnézy. Detekce této kauzální varianty u matky probandky pomocí Sangerovy sekvenace viz. Obr. 17.



Obr. 17. Varianta *PROS1*:c.1468delA (p.Ile490fs) detekována Sangerovou sekvenací u matky probandky.

Druhá poprvé popsaná byla missense varianta **c.1931T>A (p.Ile644Asn)**. Varianta byla prokázána u probanda narozeného v roce 2002. Pacient byl indikován na základě prodělané hluboké žilní trombózy pravé dolní končetiny a následně provedených funkčních testů – aktivita volného PS (22,1–27,9 %). Pacient má pozitivní rodinnou anamnézu VTE z matčiny strany rodokmenu. Varianta následně potvrzena Sangerovým sekvenováním u matky a maternální tety probanda (obě prodělaly trombózu dolní končetiny) a klasifikována jako pravděpodobně patogenní. Klasifikace varianty proběhla na základě absence alternativní alely v populačních databázích, PhyloP skóre – 7,5, *in-silico* predikčních nástrojů a pozitivní rodinné anamnézy. Varianta je zobrazena na Obr. 18.



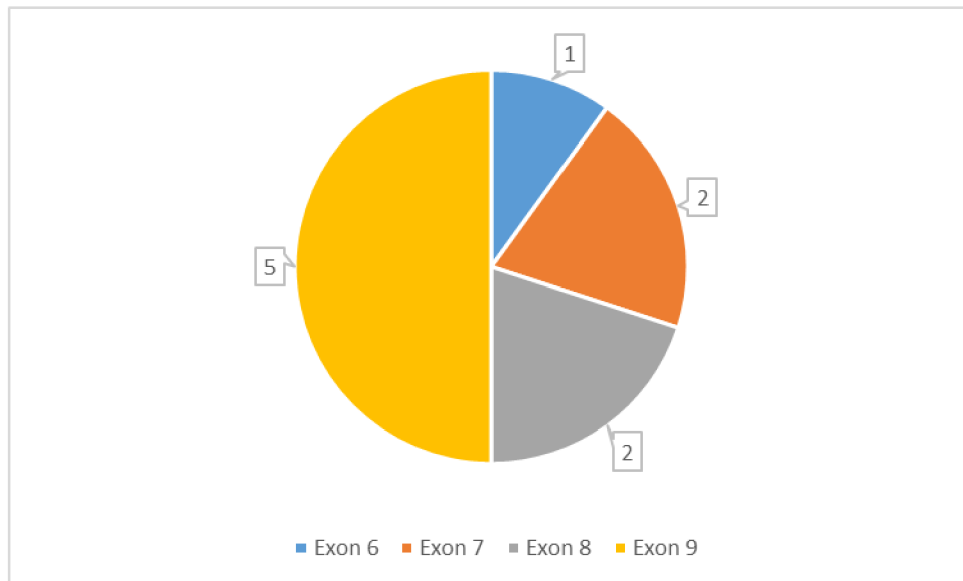
Obr. 18. Vizualizace varianty *PROS1*:c.1931T>A (p.Ile644Asn).

(levá část – Sangerovo sekvenování proband/matka R řetězec; pravá část – vizualizace varianty u probanda pomocí IGV)

6.3 Výsledky u probandů s deficitem PC

Pro podezření na dědičný deficit PC bylo na MPS genů *PROC* a *PROCR* indikováno 13 pacientů (10 žen a 3 muži). Patogenní / pravděpodobně patogenní varianty byly nalezeny pouze v genu *PROC* u 10 probandů, což představuje MDR 76,9 % viz. Tab. 18. U skupiny pacientů s deficitem PC jsme tak zaznamenali nejvyšší MDR z námi vyšetřovaných deficitů antikoagulačních proteinů. Pozitivní rodinnou anamnézu tromboembolických stavů uvedlo 9 z 10 pozitivně testovaných probandů. Nejvyšší četnost variant (5) byla zjištěna v 9. exonu genu

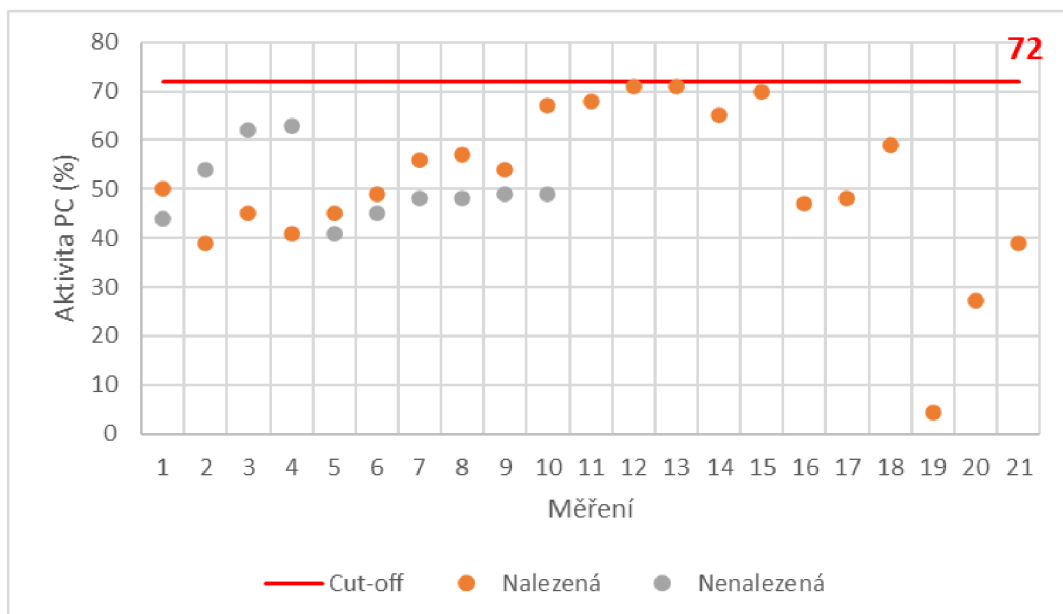
PROC viz. Obr. 19. Nejčastější varianty byly missense (8), nonsense (1) a stop loss (1). Z odhalených variant bylo 9 v heterozygotním a jedna varianta v homozygotním stavu.



Obr. 19. Četnost nalezených kauzálních variant v exonech genu *PROC*.

V této kohortě pacientů nebyl zjištěn signifikantní rozdíl ($p = 0,6286$) mediánů měření funkčních testů mezi pacienty se zjištěnou a pacienty bez odhalené kauzální varianty viz. Obr. 14. a Tab. 18.

Všechna dostupná měření aktivity PC s příslušným cut-off testu jsou zobrazeny na Obr. 20.



Obr. 20. Hodnoty všech provedených funkčních testů u pacientů indikovaných pro deficit PC s vyznačeným cut-off.

6.3.1 Dosud nepopsané varianty u probandů s deficitem PC

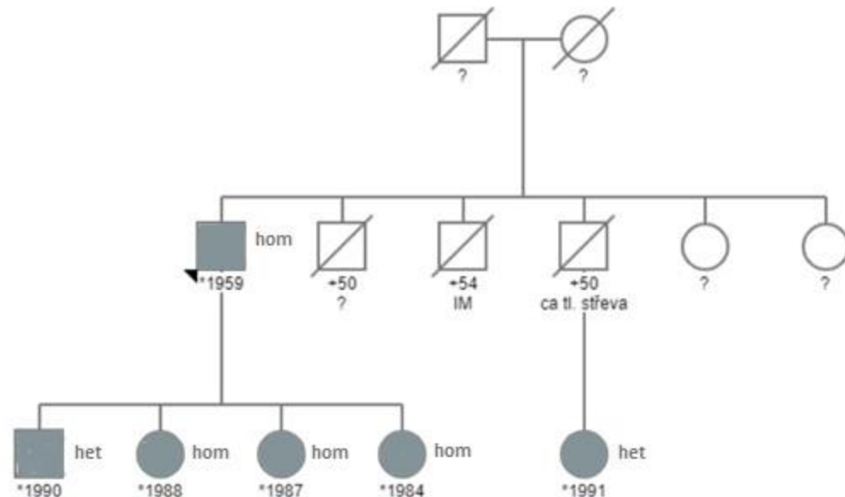
U skupiny pacientů indikovaných pro deficit PC byly odhaleny tři novel varianty - c.450C>A (p.Tyr150Ter); c.715G>C (p.Gly239Arg) a c.866C>G (p.Pro289Arg). Informace o probandech / variantách sumarizuje Tab. 21.

Varianta **c.450C>A (p.Tyr150Ter)** v heterozygotním stavu byla odhalena u probandky s rokem narození 1980. U pacientky laboratorně stanovena aktivita PC 50 %, doposud z pohledu embolizace asymptomatická. U matky probandky zjištěn deficit PC, otec opakovaně hluboké žilní trombózy a prodělal PE. Paternální dědeček na PE umřel. Segregační analýzu v tomto případě nebylo možné provést, rodinní příslušníci v riziku se genetickému testování nepodrobili. Nalezená varianta byla klasifikována jako pravděpodobně patogenní na základě absence alternativní alely v populačních databázích, dopadu na protein – nonsense a pozitivní rodinné anamnézy.

U pacientky narozené v roce 1973 byla nalezena varianta **c.715G>C (p.Gly239Arg)** v heterozygotním stavu. U této probandy prokázaly funkční testy opakovaně nízkou aktivitu PC (49–57 %). Pozitivní rodinná anamnéza – otec opakovaně trombózy a PE, u bratra hluboká žilní trombóza. U bratra byla potvrzena stejná varianta jako u probandky. Tato varianta byla klasifikována jako patogenní. Klasifikace varianty proběhla na základě absence alternativní alely v populačních databázích, PhyloP skóre – 9,29, *in-silico* predikčních nástrojů a pozitivní rodinné anamnézy. Stejná varianta na proteinové úrovni, ale rozdílná záměna nukleotidů (c.715G>A) byla korelována s očekávaným fenotypem.

Zajímavá je zejména varianta **c.866C>G (p.Pro289Arg)**, která byla přítomna v homozygotním stavu. Tato varianta byla odhalena u probanda ročník narození 1959 s pozitivní osobní a rodinnou anamnézou VTE, pozorovaná aktivita PC v rozmezí 4,4–27,1 %. Pacient prodělal opakovaně epizody hluboké žilní trombózy a PE, ale první zaznamenaná ataka VTE až v roce 1983! To může naznačovat AR formu onemocnění. S testováním přítomnosti této familiární varianty souhlasilo šest členů dané rodiny. U třech žen ročník narození 1984, 1987 a 1988 byla varianta v genu *PROC*:c.866C>G (p.Pro289Arg) potvrzena v homozygotním stavu. Poslední anamnestické údaje ze strany VTE z roku 2005 – negativní. Stejná varianta v heterozygotním stavu odhalena u dvou žen (1943 a 1991) a jednoho muže 1990. Z těchto rodinných příslušníků prodělala epizodu trombózy pouze pacientka ročník narození 1943 v roce 2006. Na základě dostupných informací se přesný příbuzenský vztah této ženy k probandovi nepodařilo stanovit. Varianta byla klasifikována jako patogenní. Klasifikace varianty proběhla na základě absence

alternativní alely v populačních databázích, *in-silico* predikčních nástrojů a pozitivní rodinné anamnézy. Varianta na stejné pozici c.866C>T(p.Pro289Leu) byla popsána s očekávaným fenotypem. Rodokmen známých členů této rodiny zobrazen v Obr. 21.

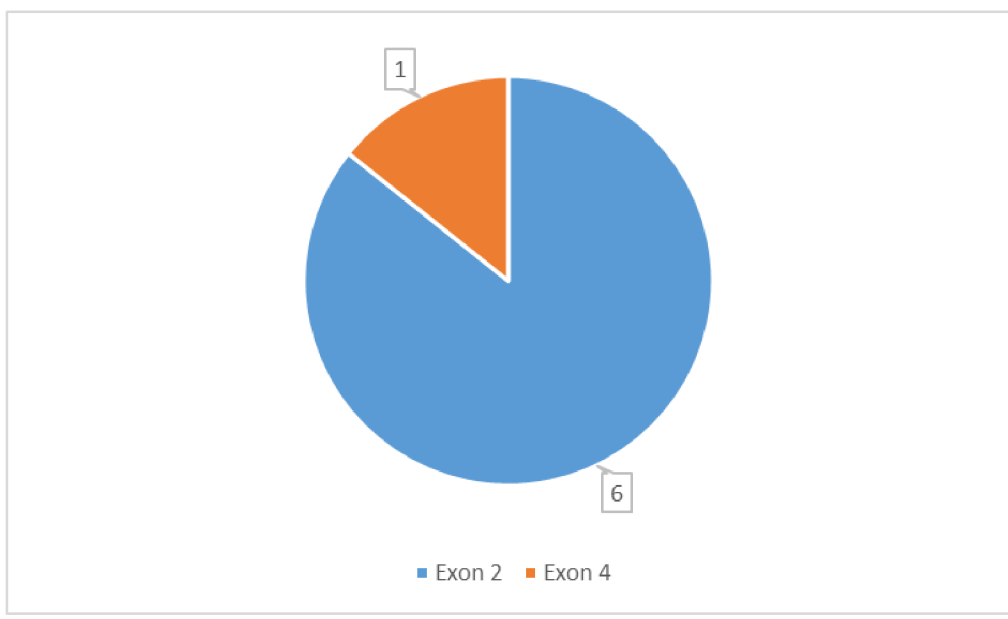


Obr. 21. Segregace varianty *PROC:c.866C>G* (p.Pro289Arg) v rodině probanda.

(* – narození; + – umrtí; IM – infarkt myokardu; ca – karcinom; hom – homozygot; het – heterozygot)

6.4 Výsledky u probandů s deficitem AT

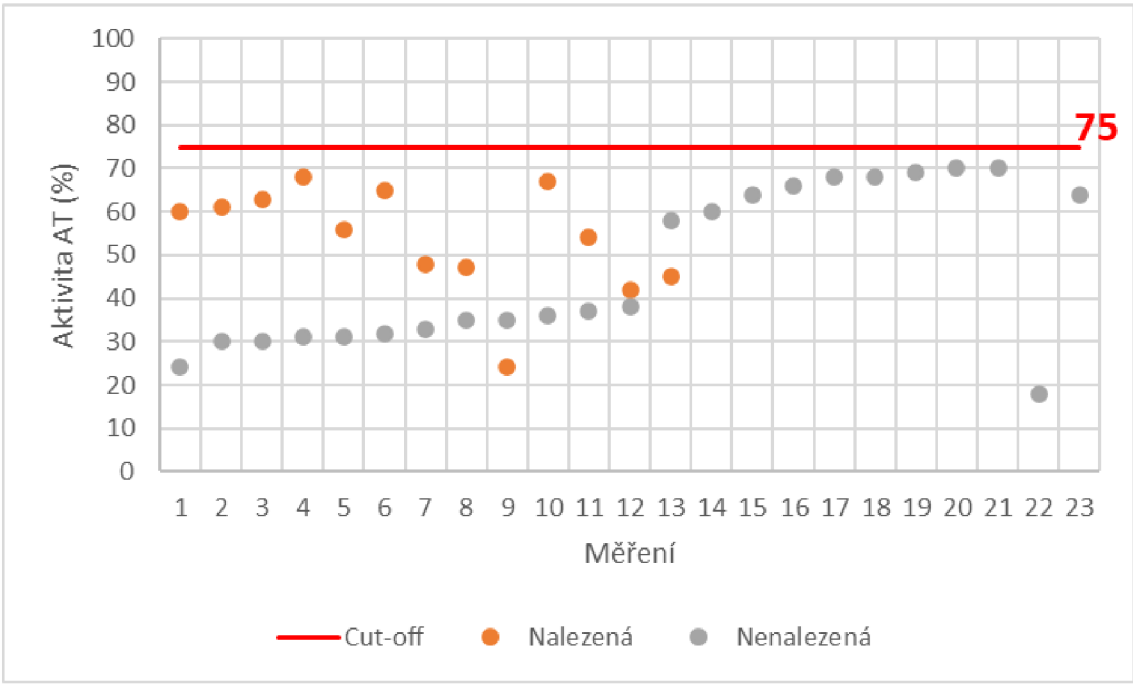
Pro genetické testování genu *SERPINC1* bylo indikováno 11 probandů (7 žen a 4 muži). Patogenní / pravděpodobně patogenní varianta byla odhalena u 7 pacientů viz. Tab. 19. U dvou pacientů z této skupiny nebyla uvedena pozitivní rodinná anamnéza VTE. V této kohortě pacientů s deficitem AT bylo zaznamenáno MDR 63,6 %. Nejvyšší četnost kauzálních variant byla zaznamenána v 2. exonu (6) genu *SERPINC1* vyneseno na Obr. 22. Byly popsány missense (6) a nonsense (1) varianty. Všechny varianty byly nalezeny v heterozygotním stavu.



Obr. 22. Četnost nalezených kauzálních variant v exonech genu *SERPINC1*.

V této kohortě pacientů nebyl zjištěn signifikantní rozdíl ($p = 0,6241$) mediánů měření funkčních testů mezi pacienty se zjištěnou a pacienty bez odhalené kauzální varianty viz. Obr. 14. a Tab. 18.

Všechna dostupná měření aktivity AT s příslušným cut-off testu jsou zobrazeny na Obr. 23.



Obr. 23. Hodnoty všech provedených funkčních testů u pacientů indikovaných pro deficit AT s vyznačeným cut-off.

6.4.1 Dosud nepopsané varianty u probandů s deficitem AT

U skupiny pacientů indikovaných pro deficit AT byla odhalena jedna novel varianta – c.398A>T (p.Gln133Leu). Informace o probandovi / variantě sumarizuje Tab. 21.

Varianta v genu *SERPINC1* **c.398A>T (p.Gln133Leu)** byla odhalena u probanda ročník narození 2004 v heterozygotním stavu. Pacient byl indikován na genetické testování z důvodů nástupu VTE v roce 2019, která nastala zřejmě jako následek předchozí hospitalizace pacienta. Prokázaná nízká aktivita AT v rozmezí 42–47 %. V paternální ani maternální straně rodiny nejsou uváděny epizody VTE. V rámci segreganční analýzy testována matka na nosičství dané varianty s negativním výsledkem. Fenotyp probanda a rodinná anamnéza naznačuje, že by se mohlo jednat o *de novo* kauzální variantu, nicméně pro definitivní potvrzení dané teze by bylo vhodné doplnit vyšetření u otce. Nalezená varianta byla klasifikována jako pravděpodobně patogenní. Klasifikace varianty proběhla na základě absence alternativní alely v populačních databázích, PhyloP skóre – 7,69, *in-silico* predikčních nástrojů. Varianta ve stejném kodonu c.397C>A (p.Gln133Lys) je klasifikována jako pravděpodobně patogenní.

6.5 Porovnání VCF vytvořených z rozdílných softwarů

U 10 pacientů sekvenovaných v rámci jednoho běhu společně byly detekovány varianty v genech *PROS1*, *PROC* a *SERPINC1* pomocí tří různých softwarů: Torrent Suite variantCalleru, Ion Reporteru a NextGene. U těchto pacientů detekovaly tyto softwary při svém základním nastavení v genu *PROS1* – 16, 19 a 27 variant; v genu *PROC* – 17, 17, 19 variant; pro gen *SERPINC1* – 15, 15, 16 variant. Nalezené varianty pro jednotlivé pacienty jsou shrnuty v Příloze 4, počet variant ve vybraných genech zobrazen v Tab. 22.

Tab. 22. Počty variant detekovaných vybranými softwary u jednotlivých pacientů.

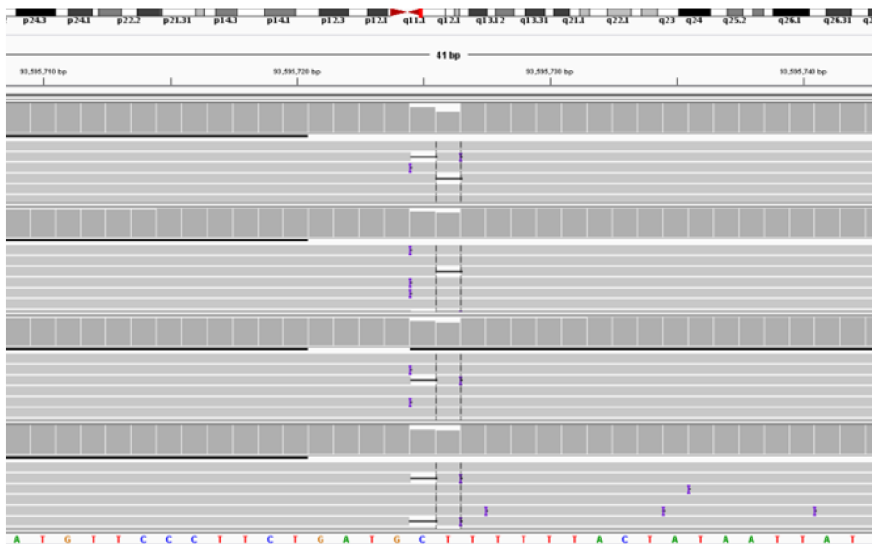
Pacient	Torrent Suite			Ion Reporter			NextGene			Celkem
	<i>PROS1</i>	<i>PROC</i>	<i>SERPINC1</i>	<i>PROS1</i>	<i>PROC</i>	<i>SERPINC1</i>	<i>PROS1</i>	<i>PROC</i>	<i>SERPINC1</i>	
1	2	2	0	3	2	0	2	2	0	13
2	1	2	0	1	2	0	1	2	0	9
3	3	2	0	4	2	0	4	3	0	18
4	0	2	1	0	2	1	3	2	1	12
5	2	1	3	3	1	3	3	1	3	20
6	2	2	3	2	2	3	4	2	3	23
7	2	1	3	3	1	3	2	1	3	19
8	2	2	3	2	2	3	5	2	3	24
9	2	2	0	1	2	0	1	3	1	12
10	0	1	2	0	1	2	2	1	2	11
Celkem	16	17	15	19	17	15	27	19	16	161

Všechny uvedené softwary odhalily jednu kauzální variantu v genu *SERPINC1:c.79T>C* (p.Trp27Arg), která byla prokázána u pacienta číslo 4.

6.5.1 NextGene

Program NextGene oproti TVC a Ion Reporteru detekoval INDEL varianty přítomné v homopolymerních oblastech: v genu *PROS1* c.1870+84delA (u pacientů 3, 4, 6, 8), c.602-23delT (u pacientů 4, 5, 6, 8, 10), c.469+25delC (u pacientů 4 a 8), c.1156-9dupT (u pacienta 10). V genu *PROC* c.1212delG (p.Met406TrpfsTer15) (u pacientů 3 a 9). V genu *SERPINC1* detekoval u pacienta 9 variantu c.409-13_409-12delTT. Frekvence alternativních alel u těchto variant se pohybovala v rozmezí 20,6–36,9 %.

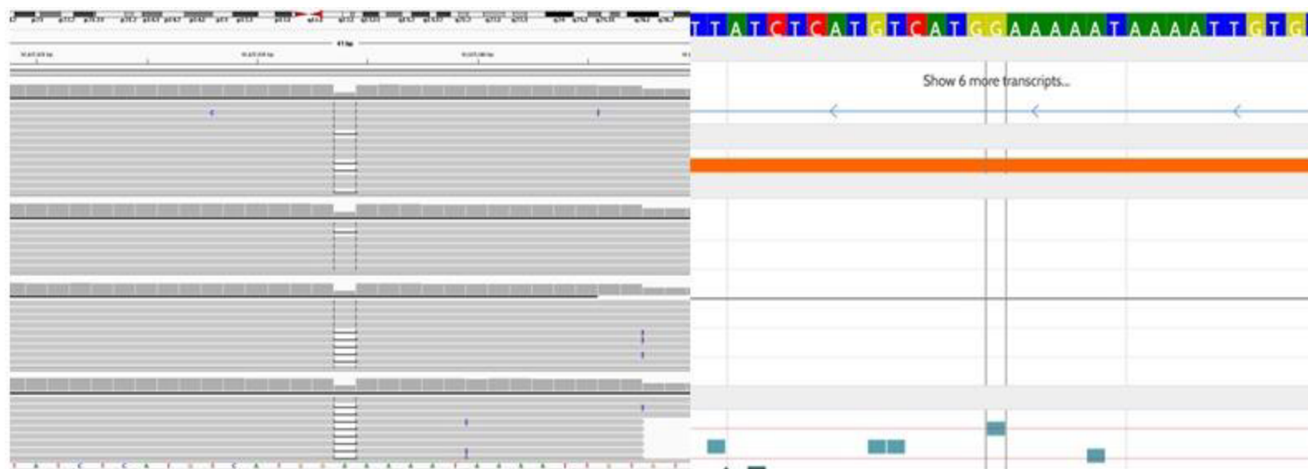
Prezence INDEL variant v homopolymerních oblastech u více pacientů ve stejném sekvenačním běhu s relativně nízkou frekvencí naznačuje, že se jedná o sekvenační artefakty viz. Obr. 24.



Obr. 24. Vizualizace chyby sekvenace – varianty *PROS1*:c.1870+84delA pomocí IGV.

6.5.2 Ion Reporter

Analýza v softwaru Ion Reporter se od ostatních porovnávaných softwarů odlišovala detekcí variant v genu *PROS1* c.347-47delT u pacientů 1 a 7 (frekvence alternativní alely 50,98 a 55,49 %) a c.939A>T (p.Leu313Phe) u pacientů 3 a 5 (frekvence alternativní alely 20,91 a 21,97%). Zajímavá je především prevalence varianty c.347-47delT, která vykazuje frekvenci alternativní alely >50 %. Dle kontroly v IGV byla varianta přítomna u všech sledovaných pacientů v rámci daného sekvenčního běhu, v genomické prohlídce VarSome není uvedena. Tento INDEL se nachází v oblasti bohaté na A/T má nízkou kvalitu, kterou má pro VC nejnižší posazenou právě Ion Reporter, pravděpodobně se jedná o sekvenční artefakt. Varianta je zobrazena na Obr. 25.

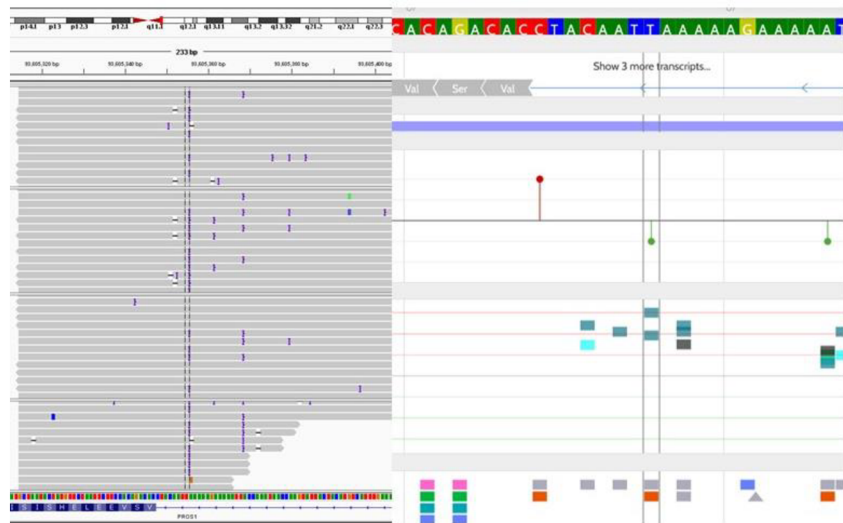


Obr. 25. Vizualizace chyby sekvenace – varianty *PROS1*:c.347-47delT.

(levá část – IGV; pravá část oblast očekávaného výskytu varianty vizualizovaná pomocí VarSome)

6.5.3 Torrent Suite

Pomocí softwaru Torrent Suite byla oproti porovnávaným softwarům nalezena u pacienta 9 varianta v genu *PROS1* c.1156-9_1156 s frekvencí alternativní alely 43,96 %. Stejná varianta (zapsána c.1156-9dupT), byla detekována u pacienta 10 softwarem NextGene s frekvencí alternativní alely 34,7 %. Jedná se o INDEL v oblasti bohaté na A/T – dle IGV přítomna u všech pacientů v daném sekvenačním běhu, není uvedena ve VarSome – s velkou pravděpodobností se jedná o sekvenační artefakt. Zobrazeno na Obr. 26.



Obr. 26. Vizualizace chyby sekvenace – varianty *PROS1*:c.1156-9dupT.

(levá část – IGV; pravá část oblast očekávaného výskytu varianty vizualizovaná pomocí VarSome)

7 Diskuze

Sekvenační techniky druhé generace umožnily výrazně zvýšit testovací kapacitu, za současného snížení sekvenačních nákladů v přepočtu na osekvenovanou bázi. Tyto technologie umožňují vyšetření více genů v rámci jedné reakce, což nabízí vhodnou aplikaci do diagnostiky monogenně děděných chorob (67). Aplikace těchto sekvenačních technologií byly již dříve popsány v rámci výzkumu. Ng et al. 2009 (119) osekvenovali dvanáct lidských exomů, kde u čtyřech nepříbuzných jedinců s Freeman-Sheldon syndromem prokázali, že kandidátní geny monogenně děděné poruchy lze detekovat pomocí exomového sekvenování. Přehled aplikací exomového sekvenování pro vyhledávání kandidátních genů, kauzálních variant u monogenních poruch a *de novo* variant u sporadických forem onemocnění přináší Ku et al. 2011 (120). Autoři také upozorňují na limitaci tohoto přístupu, zejména pak v nepokrytí regulačních genových oblastí, evolučně konzervované oblasti v nekódujících oblastech a neúplné pokrytí všech exonů. Důležitou úlohu hraje MPS (WES, WGS) pro klinickou praxi v případě nediodagnostikovaných pacientů, kde lékař očekává na základě fenotypu pacienta genetickou příčinu onemocnění. Literatura uvádí MDR do 26 % v závislosti na indikaci pacientů (121, 122, 123).

Tato práce byla zaměřena na aplikaci vhodné metodiky MPS do rutinní klinické diagnostiky vybraných deficitů antikoagulačních proteinů. Fidalgo and Ribeiro 2017 (124) poukazují na vhodnost aplikace vysokokapacitního sekvenování do genetické diagnostiky pacientů s deficitem PS, PC a AT z důvodu navýšení testovací kapacity této pravděpodobně poddiagnostikované skupiny pacientů. Pilotní práce na aplikaci NGS do problematiky VTE byla zaměřena na sekvenaci 186 vybraných genů, kde byly úspěšně detekovány kauzální varianty (118). Následovaly práce, které se zaměřovaly na velké genové panely zahrnující geny pro dědičné krvácivé poruchy, trombotické poruchy a poruchy funkce krevních destiček. Většina prací vychází z metodiky WES. Simeoni et al. 2016 (125) publikovali výsledky u 300 pacientů, u kterých odhalili 204 kauzálních variant včetně 8 CNV. Autoři se zaměřili na 63 genů. Lee et al. 2017 (126) u 64 pacientů s VTE uvádí 60,9% MDR. Zaměřili se na hledání vzácných trombofilních variant v 55 genech asociovaných s trombofilií. Data s doposud největšího patientského souboru publikovali Downes et al. 2019 (127), když pomocí ThromboGenomics HTS panelu (96 genů) osekvenovali 2396 pacientů, indikovaných pro krvácivé poruchy, trombotické poruchy a poruchy krevních destiček a zaznamenali celkové

MDR 37,3 %. Z této patientské kohorty bylo 284 pacientů indikováno pro trombózu – zjištěné MDR v této skupině pacientů bylo 48,9 %.

Výše uvedené studie nejsou primárně zaměřeny na vyhledávání kauzálních genových variant u pacientů indikovaných pro deficit PS, PC a AT. V této problematice je stále nejvíce rozšířenou metodou molekulární diagnostiky Sangerovo sekvenování (59, 114, 128–143). Tato sekvenační metoda byla považována za „zlatý standard“ pro detekci poškozujících variant. Naráží však na limity, kterými jsou nižší senzitivita zachycení CNV, neschopnost vyšetřovat více cílů paralelně a s tím spojená časová náročnost. Tyto nedostatky jsou odstraněny vysokokapacitními sekvenačními technikami, které tak zvyšují testovací kapacitu, jak bylo popsáno například v oboru molekulární patologie (144, 145). Sangerova sekvenace není vhodná pro detekci velkých delecí a duplikací, běžně je tedy doplňována metodou MLPA (146). CNV mohou být kauzální příčinou deficitu přirozených antikoagulačních proteinů, z toho důvodu je kombinace Sangerovy sekvenace a metody MLPA využívána také v této problematice (59, 114). MPS techniky jsou schopny detekovat široké spektrum mutací včetně CNV (144, 145). Patel et al. 2016 (147) popisují na příkladu sekvenace HIV viru klesající náklady na MPS, které jsou již obdobné jako u Sangerovy metody. Z uvedeného vyplývá že NGS technologie mají pro diagnostické laboratoře větší informační potenciál a pro lékaře větší rychlost odezvy než Sangerova sekvenace.

V této práci byla využita AmpliSeq technologie založena na sekvenaci designovaných amplikonů pro své hlavní výhody, kterými jsou: malé množství vstupní DNA (10 ng), vysoké pokrytí amplikonů a lepší zacílení na gen *PROS1*, který je vysoce homologní s pseudogenem *PROSP*. Výsledky uváděné v této práci pochází především z námi publikovaných dat v publikaci Vrtel et al. 2022 (148).

V rámci této práce bylo pomocí metody MPS osekvenováno 41 pacientů, ne všichni však splnili definovaná indikační kritéria. Tři pacienti byli rodinní příslušníci probandů, tři pacienti byli indikováni pro detekci variant FV Leiden / protrombin G20210A. Tři pacienti vykazovali kombinovaný deficit antikoagulačních proteinů a jedna probandka byla bez anamnestických údajů. Tito pak byly ze studie vyloučeny. Do studie bylo tedy zahrnuto 31 probandů.

Pro zavedení nové metody do klinické praxe, respektive nahrazení starší techniky novou, je důležitým ukazatelem záchyt kauzálních variant u indikovaných pacientů. Napříč indikacemi – deficit PS, PC a AT jsme zaznamenali celkovou MDR 67,7 %. Publikace vycházejí ze Sangerova sekvenování a uvádí podobné hodnoty MDR: Wypasek et al. 2017 (59) – 72 %;

Caspers et al. 2012 (129) – 65,8 %; Kim et al. 2014 (131) – 55,9 %. U pacientů s deficitem PS jsme zjistili MDR 57,1 %, což je nejvyšší hodnota z publikovaných prací 37,8–53 % (59, 128, 129, 131, 143). U kohorty pacientů indikovaných pro deficit PC jsme pozorovali MDR 76,9 %, srovnatelná data uvádí také literatura 58,1–89,3 % (59, 128, 129, 131, 136, 142). Nejvyšší MDR je popisován u pacientů s deficitem AT 74–83,5 %, my jsme zaznamenali MDR 63,6 % (59, 128, 129, 130, 131). Pozorovaná data naznačují, že námi prezentovaná metodika automatické přípravy knihoven, templátování a sekvenace na platformě Ion S5TM je vhodná pro molekulární diagnostiku pacientů s podezřením na dědičný deficit antikoagulačních proteinů.

Sekvenací genu *PROSI* jsme u pacientů s laboratorně potvrzených deficitem PS odhalili čtyři patogenní / pravděpodobně patogenní varianty. Varianty c.1155+5G>A a c.1916G>A (p.Cys639Tyr) byly již dříve popsány v literatuře (114, 127, 141). Popsali jsme také dvě novel varianty c.1468delA (p.Ile490fs) a c.1931T>A (p.Ile644Asn). V publikaci Li and Long 1996 (140) popsali jinou variantu ve stejné pozici c.1931T>G (p.Ile644Ser) u rodiny s deficitem PS a epizodami trombóz.

V genu *PROC* jsme u indikovaných probandů zachytili deset kauzálních variant. Sedm bylo již dříve popsáno v literatuře – c.677A>T (p.Gln226Leu), c.759C>A (p.His253Gln), c.1301T>C (p.Val434Ala), c.1384T>C (p.Ter462Gln), c.1106C>T (p.Pro369Leu), c.1019C>T (p.Thr340Met) (59, 129, 136, 139). Variantu c.677A>T jsme detekovali u dvou nepříbuzných probandů. V našem souboru pacientů jsme odhalili tři novel varianty c.450C>A (p.Tyr150Ter), c.715G>C (p.Gly239Arg) a c.866C>G (p.Pro289Arg). Varianta c.866C>G byla u probanda ročník narození 1959 odhalena v homozygotním stavu a sledována v rámci segregace analýzy v rodině probanda (3x homozygot, 2x heterozygot). Zajímavé je především, že první zaznamenaný nástup VTE u probanda je až v jeho 24 letech. Obvyklá dědičnost deficitu PC je AD, literatura popisuje také případy, kdy se deficit PC dědí jako AR znak. Grundy et al. 1992 (137) a Grundy et al. 1991 (138) popisují variantu na stejné nukleotidové pozici c.866C>T (p.Pro289Leu) nalezenou u probanda, který pochází z příbuzenského svazku. V této rodině pak autoři popisují dědičnost jako neúplnou AR s pozdním nástupem. V literatuře je pozdní nástup VTE u pacientů s homozygotní mutací v genu *PROC* již evidován (149, 150, 151). V literatuře byla již dříve popsána také varianta c.715G>A (p.Gly239Arg), tato je ve shodě s námi popsanou variantou c.715G>C (p.Gly239Arg) na proteinové úrovni, ale odlišuje se záměnou nukleotidu (127, 131).

Vysokokapacitní sekvenací genu *SERPINC1* jsme odhalili sedm kauzálních variant. Šest bylo již dříve publikováno – c.79T>C (p.Trp27Arg), c.133C>T (p.Arg45Trp), c.391C>T

(p.Leu131Phe), c.236G>A (p.Arg79His), c.685C>T (p.Arg229Ter) (59, 127, 129, 132, 135). Variantu c.236G>A jsme detekovali u dvou nepříbuzných probandů s pozitivní rodinnou anamnézou trombóz. Popsali jsme také jednu novel variantu c.398A>T (p.Gln133Leu) u pacienta se suspektním fenotypem. Trait et al. 1994 (134) uvádí variantu ve stejném kodonu c.397C>A (p.Gln133Lys) u probanda s pozitivní rodinnou anamnézou trombóz. Zajímavá je také varianta c.79T>C (p.Trp27Arg), kterou jsme zaznamenali u pacientky s opakovaně nízkou aktivitou AT (60–63%). Varianta je v dbSNP uvedena jako rs1165816584, není však v literatuře spojována s námi pozorovaným fenotypem. Ve stejném kodonu popsali Provanikova et al. 2020 (133) nonsense variantu c.80G>A (p.Trp27Ter).

U probandů, kde nebyla odhalena kauzální varianta by se mohlo jednat o sekundární příčinu deficitu antikoagulačního proteinu. Nedostatek proteinů C a S může být také získaným stavem např. při nedostatku vitamínu K, těžké jaterní dysfunkci, infekci nebo dlouhodobé terapii warfarinem (152, 153), U proteinu S je uváděna také orální antikoncepce (152). U pacientů, kde nebyla nalezena genetická příčina onemocnění, můžeme také zvažovat jiné molekulární mechanismy pro objasnění chybějící dědičnosti – epigenetiku. U okluzivních žilních příhod byla popsána modifikace histonů (citrulinace) a možný vliv určitých mikro RNA, jejich vyšetření by mohlo zlepšit přesnost diagnostiky VTE. Klinické studie na vliv DNA metylace u VTE nejsou v současné době k dispozici (154).

Zvážili jsme také možný posun cut-off funkčních testů. Za tímto účelem jsme pomocí t-testu porovnali mediány funkčních testů u pacientů s nalezenou kauzální variantou a pacienty bez nalezené varianty. Signifikantní rozdíl byl prokázán pouze u kohorty pacientů s deficitem PS ($p = 0,0047$). Zde by se dalo uvažovat o změně cut-off. Bohužel pro rozsáhlejší závěry nemáme dostatečný soubor pacientů. Korelace hodnot funkčních testů a MDR byla již dříve hodnocena. Například bylo prokázáno vyšší MDR u pacientů s aktivitou AT <75 %, výrazně redukované MDR u pacientů s aktivitou PC >60 % a žádné kauzální varianty nebyly odhaleny u pacientů s volným PS nad 53 a 55 % (59, 129, 143). Další zpřísňování cut-off není žádoucí z důvodu úspěšného zavedení metodiky NGS do dané problematiky, která významně zvyšuje testovací kapacitu oproti časově náročné Sangerově sekvenaci.

Varianty, které jsme označili v rámci této práce jako kauzální byly klasifikovány do skupin patogenní nebo pravděpodobně patogenní. Při interpretaci jsme zohlednili doporučení, které stanovili Richards et al. 2015 (92) a jsou známé jako ACMG/AMP doporučení pro interpretaci germinálních sekvenčních variant. Samotná interpretace variant je stále významně ovlivněna lidským faktorem a značnou mírou subjektivity. I přes využití ACMG/AMP doporučení je

udávána značná míra diskrepance v interpretaci genetických variant mezi laboratořemi i v rámci jednotlivých laboratoří (46, 155). Amendola et al. 2016 (155) uvádí 71% konkordanci u interpretací vybraných variant mezi laboratořemi. U diskordantních stanovení byly zaznamenány také rozdíly v kategoriích, které by mohly ovlivnit rozhodnutí lékaře (patogenní/pravděpodobně patogenní versus nejasného významu, pravděpodobně benigní, benigní). Správná interpretace nalezené varianty může pacientovi poskytnout benefit z pohledu léčby, prevence, ochranného režimu v rodině a případné preimplantační genetické diagnostiky. Chybná interpretace však může vézt k závažným dopadům jak pro probanda samotného, tak pro jeho rodinu (156, 157, 158). Evans et al. 2001 (159) upozorňují, že chybná interpretace může negativně ovlivnit nejen diagnostické testy, ale také prediktivní testy používané u asymptomatických osob k předpovědi budoucího rizika onemocnění. Jako absurdní příklad mohou být na podkladu prediktivního testování chybně indikovány profylaktické operační zákroky u nositelek mutací *BRCA1/BRCA2* (160).

Genetické testování pacientů s deficitem proteinu S, C nebo AT má význam především pro definitivní potvrzení dědičné komponenty a umožní detekovat familiární variantu u příbuzných probanda (tj. doposud klinicky asymptomatických nositelů patogenní varianty ve smyslu nástupu VTE). Tato identifikace umožňuje zavedení profylaktické léčby ať už permanentně nebo během vysoce rizikových situací (např. velký chirurgický zákrok, úraz, imobilizace, těhotenství a šestinedělí). Neexistují však žádná data, která by odůvodňovala rizika dlouhodobé nebo trvalé profylaktické antikoagulace u asymptomatického jedince (tj. jedince, který neprodělal tromboembolickou příhodu) (161). Genetické poradenství je přínosem pro postižené jedince a jejich rodiny (162).

V rámci této práce bylo provedeno také porovnání variant detekovaných třemi softwary – Torrent Suite (plugin TVC); Ion Reporteru a NextGene. Všechny úspěšně detekovaly poškozující variantu *SERPINC1:c.79T>C* (p.Trp27Arg) a všechny SNV. U INDEL variant zejména v homopolymerních oblastech jsme zaznamenali nejvíce chybně detekovaných variant u softwaru NextGene. Shin et al. 2017 (163) provedli porovnání dvou bioinformatických pipeline pro optimální zpracování dat získaných ze sekvenátoru Ion Torrent S5 XL pomocí Torrent Suite (modul TVC) a NextGene. V rámci studie byly správně nalezeny všechny očekávané varianty (681 SNV, 15 malých INDEL a 3 CNV). Falešně pozitivní byla pouze jedna SNV detekovaná pomocí NextGene. Platforma Ion Torrent má vyšší míru chybovosti u INDEL ve srovnání s technologií sekvenování po syntéze, kterou využívá Illumina. Literatura uvádí chybovost rozmezí 0,46–2,4 % (29, 36, 163, 164, 165). Tyto nedostatky jsou odstraňovány

optimalizací bioinformatické pipeline, zvýšením hloubky čtení a rovnoměrnosti pokrytí (163, 165, 166). Z uvedeného vyplývá, že pro optimální VC je vhodné užívat především přístupy navržené výrobcem platformy (TVC a Ion Reporter). Shin et al. 2017 (163) optimalizovali bioinformatické nastavení jehož součástí byl TVC.

8 Závěr

Tato práce podává ucelený přehled sekvenačních technik se zaměřením na techniky druhé generace. Byl navržen a optimalizován diagnostický přístup založený na vysokokapacitním sekvenování v diagnostice vzácných trombofilních stavů. Jedná se pravděpodobně o první aplikaci metodiky vysokokapacitní sekvenace u genetického testování deficitu přirozených antikoagulačních proteinů v České republice. Popisovaná metodika v klinické praxi nahrazuje doposud běžně užívanou časově náročnou Sangerovu sekvenaci. Námi popisovaný přístup je vhodný také u dalších oligogenních panelů.

V návaznosti na tuto práci by bylo vhodné nadále rozšiřovat kohortu pacientů indikovanou pro deficit přirozených antikoagulantů z důvodu provedení genotyp-fenotypové korelace. Dále pak vyhodnotit MDR u jiných oligogenních onemocnění, které jsou prezentovanou metodikou vysokokapacitního sekvenování na našem pracovišti vyšetřováni.

9 Seznam použitých zkratek

DNA – deoxyribonukleová kyselina

A – adenin

G – guanin

T – tymin

C – cytosin

RNA – ribonukleová kyselina

dNTP – deoxyribonukleotidtrifosfát

ddNTP – dideoxyribonukleotidtrifosfát

ABI – Applied Biosystems

NGS – next generation sequencing

MPS – masivně paralelní sekvenování

PCR – polymerázová řetězová reakce

PacBio – Pacific Bioscience

SMRT – single molecule real-time

ZMW – zero-mode waveguides

ONT – Oxford Nanopore Technologies

PGM – Ion Torrent Personal Genome Machine

TS – cílená sekvenace

WGS – celogenomová sekvenace

WES – celoexomová sekvenace

SNP – jednonukleotidový polymorfismus

IGV – Integrative Genomics Viewer

VC – variant calling

INDEL – inzerce, delece

VCF – variant call format

SNV – single nucleotide variant

TVC – Torrent Variant caller

CNV – copy number variants

MLPA – Multiplex Ligation-Dependent Probe Amplification

ExAC – Exome Aggregation Consortium

gnomAD – The Genome Aggregation Database

dbSNP – databáze jednonukleotidových polymorfismů

NCBI – National Center for Biotechnology Information

MAF – Minor allele frequency

LOVD – Leiden Open Variation Database

HGMD – Human Gene Mutation Database

HGVS – Human Genome Variation Society

SIFT – Sorting Intolerant From Tolerant

PolyPhen-2 – Polymorphism Phenotyping v2

VEP – Ensembl Variant Effect Predictor

PhyloP – phylogenetic p-values

mRNA – messenger ribonukleová kyselina

UCSC – University of Santa Cruz

ACMG / AMP – American College of Medical Genetics and Genomics and the Association for Molecular Pathology

VTE – žilní tromboembolická nemoc

PE – plicní embolie

PS – protein S

PC – protein C

AT – antitrombin

AD – autozomálně dominantní

AR – autozomálně recesivní

MDR – míra detekce mutací

DOAC – direct oral anticoagulant

10 Seznam publikovaných prací autora

Publikace

Vrtel P., Slavik L., Vodicka R., Stellmachova J., Prochazka M., Prochazkova J., Ulehlova J., Rohon P., Simurda T., Stasko J., Martinkova I. and Vrtel R. Detection of Unknown and Rare Pathogenic Variants in Antithrombin, Protein C and Protein S Deficiency Using High-Throughput Targeted Sequencing. *Diagnostics*. 2022; 12:1060.

Vrtěl P., Slavík L., Vodička R., Procházka M., Procházková J., Vrtěl R., Úlehlová J., Rohoň P., Štellmachová J. Vyhledávání genetických variant u trombofilních stavů. *Časopis Lékařů Českých*, 2019; 158(1): 28–32.

Vrtěl P., Vrtěl R., Klaskova E., Vrbicka D., Adamova K., Pavlicek J., Hana V., Hana V. jr., Soucek O., Stara V., Lebl J., Snajdrova M., Zapletalova J., Fürst T., Prochazka M., Kapralova S., Tauber Z., Krejcirikova E., Routilova E., Stellmachova J., Vodicka R. Haplotype analysis of the X chromosome in patients with Turner syndrome in order to verify the possible effect of imprinting on selected symptoms. *Biomedical papers*. 2021; 165:XXX.

Hůrková V., Šantavý J., Vlk R., Procházka M., **Vrtěl P.**, Vrtěl R. Pohled genetika na rozšířené záhlaví u plodu mezi 11. – 14. týdnem gravidity. *Gynekologie a porodnictví*. 2017; 1(1): 40–44.

Adamová K., Vrbická D., Vrtěl R., **Vrtěl P.**, Klásková E., Procházka M. Současný pohled na Turnerův syndrom. *Gynekologie a porodnictví*. 2020; 4(5): 279–281.

Štellmachová J., **Vrtěl P.**, Vrtěl R., Janíková M., Kolaříková K., Procházka M., Vodička R. Nádory ovaria a genetická dispozice. *Česká gynekologie* 2022; 87(3): 211–216.

Pavlicek J., Soucek O., Vrtel R., Klaskova E., Hana V., Stara V., Adamova K., Fürst T., Hana V Jr., Kapralova S., Prochazka M., Snajderova M., Tomaskova H., Tüdös Z., Vrbicka D., **Vrtel P.**, Zapletalova J., Tauber Z., Lebl J. Karyotyping of Lymphocytes and Epithelial Cells of Distinct Embryonic Origin Does Not Help to Predict the Turner Syndrome Features. *Horm Res Paediatr* . 2022; 1–11.

Přednášky a postery

Vrtěl P., Vodička R., Procházka M., Vrtěl R., Slavík L., Procházková J. Vyhledávání vzácných genetických variant u trombofilních stavů – zavedení do klinické praxe. Dolní Morava 2-4.11.2016 (XX. Celostátní konference DNA diagnostiky).

Vrtěl P., Slavík L., Vodička R., Procházka M., Vrtěl R., Úlehlová J. Aplikace metodologie masivního paralelního sekvenování do vyhledávání vzácných genetických variant u trombofilních stavů. Hradec Králové 18-20. 5. 2017 (XXIV. Česko-slovenská konference o trombóze a hemostáze).

Vrtěl P., Slavík L., Vodička R., Vrtěl R., Úlehlová J., Procházková J., Hluší A., Palova M., Procházka M. Application of massive parallel sequencing searching for rare genetic variants associated with thrombophilia – implementation to clinical practice. Basilej 15-18. 2. 2017 (61st Annual Meeting of the Society of Thrombosis and Hemostasis Research).

Vrtěl P., Slavík L., Vodička R., Procházka M., Vrtěl R., Procházková J., Úlehlová J. Methodology of massive parallel sequencing on Ion Torrent platform – searching for rare genetic variants associated with thrombophilia. Vienna 20-23.2.2018 (62nd Annual Meeting of the Society of Thrombosis and Hemostasis Research).

Vrtěl P., Vrtěl R., Adamova K., Vrbicka D., Stellmachova J., Klaskova E., Pavlicek J., Hana V., Stara V., Soucek O., Lebl J., Prochazka M. X chromosome haplotypization in the group of Turner syndrome patients. Houston 15-19.10.2019 (ASHG 2019 annual meeting).

Vrtěl P., Vrtěl R., Adamova K., Vrbicka D., Prochazka M., Stellmachova J., Klaskova E., Zapletalova J., Pavlicek J., Hana V., Stara V., Soucek O., Lebl J. Possible effect of the imprinting on the Turner's phenotype. Online 6-9.6. 2020 (European Human Genetics Virtual Conference 2020).

Vrtěl P., Šantavá A., Vrtěl R., Slavík L., Vodička R., Štellmachová J. Detekce CNV pomocí kombinace masivní paralelní sekvence s metodou MLPA – kazuistika. Smokovec 30. 9.–1. 10. 2021 (XXXI. Izakovičův memoriál).

Vrtěl P., Vrtěl R., Vodička R., Slavík L., Procházková J., Úlehlová J., Štellmachová J. Hledání vzácných genetických variant asociovaných s trombózou pomocí vysokokapacitní sekvenční technologie. Košice 21. 9.–23. 9. 2022 (XXXII. Izakovičův memoriál).

11 Použitá literatura

- 1) Watson JD, Crick FHC. Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature*, 1953; 171: 737–738.
- 2) Crick FHC, Barnett L, Brenner S, et al. General nature of the genetic code for proteins. *Nature*, 1961; 192: 1227–1232.
- 3) Nirenberg M, Leder P. RNA codewords and protein synthesis: The effect of trinucleotides upon the binding of sRNA to ribosomes. *Science*, 1964; 145: 1399–1407.
- 4) Khorana HG, Büuchi H, Ghosh H, et al. Polynucleotide synthesis and the genetic code. *Cold Spring Harb Symp Quant Biol* 1966; 31:39–49.
- 5) Crick FHC. The origin of the genetic code. *J Mol Biol* 1968; 38: 367–379.
- 6) Heather JM, Chain B. The sequence of sequencers: The history of sequencing DNA. *Genomics* 2016; 107(1): 1–8.
- 7) Holley RW, Apgar J, Everett GA, et al. Structure of a ribonucleic acid. *Science* 1965; 147: 1462–1465.
- 8) Sanger F, Brownlee G, Barrell B. A two-dimensional fractionation procedure for radioactive nucleotides. *J Mol Biol* 1965; 13: 373–398.
- 9) Wu R, Kaiser A. Structure and base sequence in the cohesive ends of bacteriophage lambda DNA. *J Mol Biol* 1968; 35(3): 523–537.
- 10) Wu R. Nucleotide sequence analysis of DNA. *Nature New Biol* 1972; 236: 198–200.
- 11) Sanger F, Coulson AR. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J Mol Biol* 1975; 94: 441–448.
- 12) Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci* 1977; 74(12): 5463–5467.
- 13) Kchouk M, Gibrat JF, Elloumi M. Generations of Sequencing Technologies: From First to Next Generation. *Biol Med* 2017; 9(3): 1–8.
- 14) Maxam AM, Gilbert W. A new method for sequencing DNA. *Proc Natl Acad Sci USA* 1977; 74(2): 560–564.

- 15) Hu T, Chitnis N, Monos D, et al. Next-generation sequencing technologies: An overview. *Human Immunol* 2021; 82(11): 801–811.
- 16) Brenner S, Johnson M, Bridgham J, et al. Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nature Biotechnol* 2000; 18(6): 630–634.
- 17) Patrick KL. 454 Life Sciences: Illuminating the future of genome sequencing and personalized medicine. *Yale J Biol Med* 2007; 80(4): 191–194.
- 18) Mardis ER. Next-generation DNA sequencing methods. *Annu Rev Genomics Hum Genet* 2008; 9: 387–402.
- 19) Berka J, Chen YJ, Leamon JH, et al. “Bead emulsion nucleic acid amplification”. U.S. Patent Application, 2005.
- 20) Mardis ER. The impact of next-generation sequencing technology on genetics. *Trends Genet* 2008; 24(3): 133–141.
- 21) Adessi C, Matton G, Ayala G, et al. Solid phase DNA amplification: characterisation of primer attachment and amplification mechanisms. *Nucleic Acids Res* 2000; 28(20): e87.
- 22) Fedurco M, Romieu A, Williams S, et al. BTA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies. *Nucleic Acids Res* 2006; 34(3): e22.
- 23) Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet* 2016; 17: 333–351.
- 24) Chaitankar V, Karakulah G, Ratnapriya R, et al. Next generation sequencing technology and genomewide data analysis: Perspectives for retinal research. *Prog Retin Eye Res* 2016; 55: 1–31.
- 25) Rothberg JM, Hinz W, Rearick TM, et al. An integrated semiconductor device enabling non-optical genome sequencing. *Nature* 2011; 475: 348–352.
- 26) Golan D, Medvedev P. Using state machines to model the Ion Torrent sequencing process and to improve read error rates. *Bioinformatics* 2013; 29(13): i344–i351.

- 27) McKernan KJ, Peckham HE, Costa GL, et al. Sequence and structural variation in a human genome uncovered by short-read, massively parallel ligation sequencing using two-base encoding. *Genome Res* 2009; 19: 1527–1541.
- 28) Liu L, Li H, Li S, et al. Comparison of Next-Generation Sequencing Systems. *J Biomed Biotechnol* 2012; 2012.
- 29) Loman NJ, Misra R, Dallman T, et al. Performance comparison of bench-top high-throughput sequencing platforms. *Nature Biotechnol* 2012; 30: 434–439.
- 30) Illumina. Illumina sequencing platforms. 2022 [online]. [cit. 2022-10-11] Available from: <https://www.illumina.com/systems/sequencing-platforms.html>.
- 31) Kanzi AM, San JE, Chimukangara B, et al. Next Generation Sequencing and Bioinformatics Analysis of Family Genetic Inheritance. *Front Genet* 2020; 11: 544162.
- 32) Logsdon GA, Vollger MR, Eichler EE. Long-read human genome sequencing and its applications. *Nat Rev Genet* 2020; 21: 597–614.
- 33) Rhoads A, Au KF. PacBio Sequencing and Its Applications. *Gen Prot Bioinf* 2015; 13(5): 278–289.
- 34) Mikheyev AS, Tin MM. A first look at the Oxford Nanopore MinION sequencer. *Mol Ecol Resour* 2014; 14: 1097–1102.
- 35) Quick J, Loman NL, Duraffour S, et al. Real-time, portable genome sequencing for Ebola surveillance. *Nature* 2016; 530: 228–232.
- 36) Quail MA, Smith M, Coupland P, et al. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics* 2012; 13: 1–13.
- 37) Head SR, Komori HK, LaMere SA, et al. Library construction for next-generation sequencing: overviews and challenges. *Biotechniques* 2014; 56(2): 61–77.
- 38) Rizzo JM, Buck MJ. Key Principles and Clinical Applications of “Next-Generation” DNA Sequencing. *Cancer Prev Res* 2012; 5: 887–900.
- 39) Kim MJ. Complete Guide to Ion Ampliseq Technology – How Does Ion Ampliseq Technology Work? 2017 [online]. [cit. 2022-10-12] Available from:

<https://www.thermofisher.com/blog/behindthebench/complete-guide-to-ion-ampliseq-technology-how-does-ion-ampliseq-technology-work/>.

- 40) Marine R, Polson SW, Ravel J, et al. Evaluation of a Transposase Protocol for Rapid Generation of Shotgun High-Throughput Sequencing Libraries from Nanogram Quantities of DNA. *Appl Environ Microbiol* 2011; 77(22): 8071–8079.
- 41) Mamanova L, Coffey AJ, Scott CE, et al. Target-enrichment strategies for next-generation sequencing. *Nat Methods* 2010; 7(2): 111–118.
- 42) Craig DW, Pearson JW, Szelinger S, et al. Identification of genetic variants using bar-coded multiplexed sequencing. *Nat Methods* 2008; 5(10): 887–893.
- 43) Pereira R, Oliveira J, Sousa M. Bioinformatics and Computational Tools for Next-Generation Sequencing Analysis in Clinical Genetics. *J Clin Med* 2020; 9: 132.
- 44) Thermo Fisher Scientific. *Torrent Suite™ Software 5.12 USER GUIDE (MAN0017972)*. 2019 [online]. [cit. 2022-9-14] Available from: https://assets.thermofisher.com/TFS-Assets/LSG/manuals/MAN0017972_031419_TorrentSuite_5_12_UG_.pdf.
- 45) Ewing B, Hillier L, Wendl MC, et al. Base-calling of automated sequencer traces using Phred. I. Accuracy assessment. *Genome Res* 1998; 8: 175–185.
- 46) Yohe S, Thyagarajan B. Review of clinical next-generation sequencing. *Arch Pathol Lab Med* 2017; 141(11): 1544–1557.
- 47) Del Fabbro Ch, Scalabrin S, Morgante M, et al. An Extensive Evaluation of Read Trimming Effects on Illumina NGS Data Analysis. *PLOS ONE* 2013; 8: e85024.
- 48) Ameer A, Che H, Martin M, et al. De Novo Assembly of Two Swedish Genomes Reveals Missing Segments from the Human GRCh38 Reference and Improves Variant Calling of Population-Scale Sequencing Data. *Genes* 2018; 9(10): 486.
- 49) Flicek P, Birney E. Sense from sequence reads: methods for alignment and assembly. *Nat Methods* 2009; 6: S6–S12.
- 50) Li R, Li Y, Kristiansen K, et al. SOAP: short oligonucleotide alignment program. *Bioinformatics* 2008; 24(5): 713–714.

- 51) Rumble SM, Lacroute P, Dalca AV, et al. SHRiMP: Accurate Mapping of Short Color-space Reads. *PLOS Comput Biol* 2009; 5: e1000386.
- 52) Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012; 9(4): 357–359.
- 53) Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 2009; 25(14): 1754–1760.
- 54) DePristo MA, Banks E, Poplin RE, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 2011; 43(5): 491–498.
- 55) Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. *rXiv Prepr arXiv12073907* 2012; 1–9.
- 56) Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009; 25(16): 2078–2079.
- 57) Sandmann S, de Graaf AO, Karimi M, et al. Evaluating Variant Calling Tools for Non-Matched Next-Generation Sequencing Data. *Sci Rep* 2017; 7: 1–12.
- 58) SoftGenetics. NextGENe. 2022 [Software]. [cit. 2022-10-11] Available from: <https://softgenetics.com/products/nextgene/>.
- 59) Wypasek E, Corral J, Alhenc-Gelas M, et al. Genetic characterization of antithrombin, protein C, and protein S deficiencies in Polish patients. *Pol Arch Int Med* 2017; 127: 512–523.
- 60) Pintao MC, Garcia AA, Borgel D et al. Gross deletions/duplications in PROS1 are relatively common in point mutation-negative hereditary protein S deficiency. *Hum Genet* 2009; 126(3): 449–456.
- 61) Chen W, Kalscheuer V, Tzschach A, et al. Mapping translocation breakpoints by next-generation sequencing. *Genome Res* 2008; 8: 1143–1149.
- 62) Thermo Fisher Scientific. Ion Reporter. 2022 [Software]. [cit. 2022-9-11] Available from: <https://ionreporter.thermofisher.com/ir/>.
- 63) Thermo Fisher Scientific. CNV detection by Ion semiconductor sequencing. 2014 [online]. [cit. 2022-9-14] Available from: <https://tools.thermofisher.com/content/sfs/brochures/CNV-Detection-by-Ion.pdf>.

- 64) Thermo Fisher Scientific. Ion AmpliSeq Designer. 2022 [Software]. [cit. 2022-9-12] Available from: <https://ampliseq.com/>.
- 65) SoftGenetics. Geneticist Assistant. 2022 [Software]. [cit. 2022-10-11] Available from: <https://softgenetics.com/products/geneticist-assistant/>.
- 66) Illumina. VariantStudio Software. 2017 [Software]. [cit. 2022-9-10] Available from: <https://www.illumina.com/products/by-type/informatics-products/variantstudio.html>.
- 67) Jamuar SS, Tan EC. Clinical application of next-generation sequencing for Mendelian diseases. *Hum Genomics* 2015; 9: 1–10.
- 68) 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* 2015; 526: 68–74.
- 69) Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* 2020; 581: 434–443.
- 70) Sherry ST, Ward MH, Kholodov M, et al. dbSNP: the NCBI database of genetic variation. *Nucl Acids Res* 2001; 29(1): 308–311.
- 71) National Library of Medicine. National Center for Biotechnology Information. 2022 [online]. [cit. 2022-8-10] Available from: <https://www.ncbi.nlm.nih.gov/>.
- 72) Charames GS, Sabatini P, Watkins N. Clinical and genetic evidence and population evidence. In: Lázaro C, Lerner-Ellis J, Spurdle A. *Translational and Applied Genomics*. Academic Press 2021; 59–87.
- 73) Chouraki V, Seshadri S. Genetics of Alzheimer’s Disease. *Adv Genet* 2014; 87: 245–294.
- 74) Hernandez RD, Uricchio LH, Hartman K, et al. Ultra-rare variants drive substantial cis-heritability of human gene expression. *Nat Genet* 2019; 51(9): 1349–1355.
- 75) Lasky-Su J. Statistical Techniques for Genetic Analysis. In: Robertson D, Williams GH. *Clinical and Translational Science (Second Edition)*. Academic Press 2017; 347–362.
- 76) Landrum MJ, Chitipiralla S, Brown GR, et al. ClinVar: improvements to accessing data. *Nucl Acids Res* 2020; 48(D1): D835–D844.
- 77) Fokkema IF, Taschner PE, Schaafsma GC, et al. LOVD v.2.0: the next generation in gene variant databases. *Hum Mutat* 2011; 32(5):557–63.

- 78) Stenson PD, Mort M, Ball EV, et al. The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. *Hum Genet* 2014; 133: 1–9.
- 79) den Dunnen JT, Antonarakis SE. Mutation nomenclature extensions and suggestions to describe complex mutations: A discussion. *Hum Mutat* 2000; 15(1):7–12.
- 80) den Dunnen JT, Dalgleish R, Maglott DR, et al. HGVS recommendations for the description of sequence variants: 2016 update. *Hum Mutat* 2016; 37(6):564–569.
- 81) Sim NL, Kumar P, Hu J, et al. SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucl Acids Res* 2012; 40(W1): W452–W457.
- 82) Adzhubei I, Jordan DM, Sunyaev SR. Predicting Functional Effect of Human Missense Mutations Using PolyPhen-2. *Curr Protoc Hum Genet* 2013; Chapter 7.
- 83) Schwarz JM, Cooper DN, Schuelke M, et al. MutationTaster2: mutation prediction for the deep-sequencing age. *Nat Methods* 2014; 11: 361–362.
- 84) McLaren W, Gil L, Hunt SE, et al. The Ensembl Variant Effect Predictor. *Genome Biol* 2016; 17: 1–14.
- 85) Desmet FO, Hamroun D, Lalande M, et al. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res* 2009; 37(9): e67.
- 86) Pollard KS, Hubisz MJ, Rosenbloom KR, et al. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res* 2010; 20:110–21.
- 87) Pabinger S, Dander A, Fischer M, et al. A survey of tools for variant analysis of next-generation genome sequencing data. *Brief Bioinform* 2014; 15(2): 256–278.
- 88) Rangwala SH, Kuznetsov A, Ananiev V, et al. Accessing NCBI data using the NCBI Sequence Viewer and Genome Data Viewer (GDV). *Genome Res* 2021; 31:159–169.
- 89) Haeussler M, Zweig AS, Tyner C, et al. The UCSC Genome Browser database: 2019 update. *Nucleic Acids Res* 2019; 47(D1): D853–D858.
- 90) Cunningham F, Allen JE, Allen J, et al. Ensembl 2022. *Nucleic Acids Res* 2022; 50(1): D988–D995.

- 91) Kopanos Ch, Tsiolkas V, Kouris A, et al. VarSome: the human genomic variant search engine. *Bioinformatics* 2019; 35(11): 1978–1980.
- 92) Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the american college of medical genetics and genomics and the association for molecular pathology. *Genet Med* 2015; 17: 405–423.
- 93) Heit JA. Epidemiology of venous thromboembolism. *Nat Rev Cardiol* 2015; 12: 464–474.
- 94) Prandoni P, Lensing AWA, Cogo A, et al. The long-term Clinical course of acute deep venous thrombosis. *Ann Intern Med* 1996; 125: 1–7.
- 95) Tagalakis V, Patenaude V, Kahn SR, et al. Incidence of and mortality from venous thromboembolism in a real-world population: The Q-VTE Study Cohort. *Am. J. Med.* 2013, 126, 832.e13–832.e21.
- 96) Anderson FA, Wheeler HB, Goldberg RJ, et al. A population-based perspective of the hospital incidence and case-fatality rates of deep vein thrombosis and pulmonary embolism: The Worcester DVT Study. *Arch Intern Med* 1991; 151: 933–938.
- 97) Souto JC, Almasy L, Borell M, et al. Genetic susceptibility to thrombosis and its relationship to physiological risk factors: The GAIT study. *Genetic Analysis of Idiopathic Thrombophilia*. *Am J Hum Genet* 2000; 67: 1452–1459.
- 98) Dahlbäck B. Advances in understanding pathogenic mechanisms of thrombophilic disorders. *Blood* 2008; 112: 19–27.
- 99) Ekim M, Yrlmaz YK. The prevalence of Factor V Leiden, prothrombin G20210A, MTHFR C677T and MTHFR A1298C mutations in healthy Turkish population. *Hippokratia* 2015; 19(4): 309–313.
- 100) Rees DC, Cox M, Clegg MB. World distribution of factor V Leiden. *The Lancet* 1995; 346: 1133–1134.
- 101) Segers K, Dahlbäck B, Nicolaes GA. Coagulation factor V and thrombophilia: background and mechanisms. *Thromb Haemost* 2007; 98: 530–542.

- 102) Kvasnička J, Hájková J, Bobčíková P, et al. Prevalence trombofilních mutací FV Leiden, protrombinu G20210A a PAI-1 4G/5G a jejich vzájemných kombinací v souboru 1450 zdravých osob středního věku v regionu Praha a střední Čechy (výsledky real-time PCR analýzy FRET). *Čas Lék čes* 2012; 151: 76–82.
- 103) Kessler P. Trombofilní stavy. *Interní medicína pro praxi*, 2007; 8(9): 374–379.
- 104) Emmerich J, Rosendaal FR, Cattaneo M, et al. Combined effect of factor V Leiden and prothrombin 20210A on the risk of venous thromboembolism. *ThrombHaemost* 2001; 86: 809–816.
- 105) Rosendaal FR, Doggen CJM, Zivelin A et al. Geographic distribution of the 20210 G to A prothrombin variant. *Thromb Haemost* 1998; 79(4): 706–708.
- 106) Bertina RM, Koelman BP, Koster T et al. Mutation in blood coagulation factor V associated with resistance to activated protein C. *Nature* 1994; 369: 64–67.
- 107) Ekim M, Ekim H, Yılmaz YK. The prevalence of Factor V Leiden, prothrombin G20210A, MTHFR C677T and MTHFR A1298C mutations in healthy Turkish population. *Hippokratia* 2015; 19(4): 309–313.
- 108) Blasczyk R, Ritter M, Thiede M et al. Simple and rapid detection of factor V Leiden by allele-specific PCR amplification. *Thromb Haemost* 1996; 75(5): 757–759.
- 109) Yousefian E, Kardi MT, Allahveisi A. Methylenetetrahydrofolate reductase C677T and A1298C polymorphism in Iranian women with idiopathic recurrent pregnancy losses. *Iran Red Crescent Med J* 2014; 16(7): e16763.
- 110) Awad NS, Almaki TA, Sabry AM et al. Screening of factor VG1691A (Leiden) and factor II/prothrombin G20210A polymorphisms among apparently healthy Taif-Saudi Arabia population using a reverse hybridization strip assay approach. *World J Med Sci* 2013; 9(4): 202–207.
- 111) van den Bergh FA, van Oeveren-Dybicz AM, Bon MA. Rapid single-tube genotyping of the factor V Leiden and prothrombin mutations by real-time PCR using dual-color detection. *Clin Chem* 2000; 46(8): 1191–1195.
- 112) Colucci G, Tsakiris DA. Thrombophilia screening revisited: An issue of personalized medicine. *J Thromb Thrombol* 2020; 49: 618–629.

- 113) Dahlbäck B. Protein S and C4b-binding protein: components involved in the regulation of the protein C anticoagulant system. *Thromb Haemost* 1991; 66(1): 49–61.
- 114) García de Frutos P, Fuentes-Prior P, Hurtado B, et al. Molecular basis of protein S deficiency. *Thromb Haemost* 2007; 98(3): 543–556.
- 115) De Stefano V, Finazzi G, Mannucci PM Inherited thrombophilia: Pathogenesis, clinical syndromes, and management. *Blood* 1996; 87: 3531–3544.
- 116) Goldenber NA, Manco-Johnson MJ. Protein C deficiency. *Haemophilia* 2008; 14(6): 1214–1221.
- 117) Lu Z, Wang F, Liang M. SerpinC1/Antithrombin III in kidney-related diseases. *Clin Sci* 2017; 131(9): 823–831.
- 118) Lotta LA, Wang M, Yu J, et al. Identification of genetic risk variants for deep vein thrombosis by multiplexed next-generation sequencing of 186 hemostatic/pro-inflammatory genes. *BMC Med Genom* 2012; 5: 7.
- 119) Ng SB, Turner EH, Robertson PD, et al. Targeted Capture and Massively Parallel Sequencing of Twelve Human Exomes. *Nature* 2009; 461(7261): 272–276.
- 120) Ku CH-S, Naidoo N, Pawitan Y. Revisiting Mendelian disorders through exome sequencing. *Hum Genet* 2011; 129: 351–370.
- 121) Shashi V, McConkie-Rosell A, Rosell B, et al. The utility of the traditional medical genetics diagnostic evaluation in the context of next-generation sequencing for undiagnosed genetic disorders. *Genet Med* 2014; 16 (2): 76–182.
- 122) Yang Y, Muzny DM, Reid JG, et al. Clinical Whole-Exome Sequencing for the Diagnosis of Mendelian Disorders. *N Engl J Med* 2013; 369:1502–1511.
- 123) Lee H, Deignan JL, Dorrani N, et al. Clinical Exome Sequencing for Genetic Identification of Rare Mendelian Disorders. *JAMA* 2014; 312(18): 1880–1887.
- 124) Fidalgo T, Ribeiro ML. Added value of next-generation sequencing for haemostasis diagnosis. *Thromb Haemost Res* 2017; 1: 1007.
- 125) Simeoni I, Stephens, JC, Hu F, et al. A high-throughput sequencing test for diagnosing inherited bleeding, thrombotic, and platelet disorders. *Blood* 2016; 127 (23): 2791–2803.

- 126) Lee EJ, Dykas DJ, Leavitt, AD, et al. Whole-exome sequencing in evaluation of patients with venous thromboembolism. *Blood Adv* 2017; 1 (16): 1224–1237.
- 127) Downes K, Megy K, Duarte D, et al. Diagnostic high-throughput sequencing of 2396 patients with bleeding, thrombotic, and platelet disorders. *Blood* 2019; 134 (23): 2082–2091.
- 128) Berezky Z, Gindele R, Speker M, et al. Deficiencies of the Natural Anticoagulants – Novel Clinical Laboratory Aspects of Thrombophilia Testing. *EJIFCC* 2016; 27: 130–146.
- 129) Caspers M, Pavlova A, Driesen J, et al. Deficiencies of antithrombin, protein C and protein S – Practical experience in genetic analysis of a large patient cohort. *Thromb Haemost* 2012; 108: 247–257.
- 130) Castaldo G, Cerbone AM, Guida A, et al. Molecular analysis and genotype-phenotype correlation in patients with antithrombin deficiency from Southern Italy. *Thromb Haemost* 2012; 107: 673–680.
- 131) Kim HJ, Seo JY, Lee KO, et al. Distinct frequencies and mutation spectrums of genetic thrombophilia in Korea in comparison with other Asian countries both in patients with thromboembolism and in the general population. *Haematologica* 2014; 99: 561–569.
- 132) Luxembourg B, Delev D, Geisen C, et al. Molecular basis of antithrombin deficiency. *Thromb Haemost* 2011; 105: 635–646.
- 133) Provaznikova D, Matyskova M, Capova I, et al. Seventeen novel SERPINC1 variants causing hereditary antithrombin deficiency in a Czech population. *Thromb Res* 2020; 189: 39–41.
- 134) Trait RC, Walker ID, Perry DJ, et al. Prevalence of antithrombin deficiency in the healthy population. *Br J Haematol* 1994; 87: 106–112.
- 135) Owen MC, Borg JY, Soria C, et al. Heparin binding defect in a new antithrombin III variant: Rouen, 47 Arg to His. *Blood* 1987; 69: 1275–1279.
- 136) Gandrille S, Aiach M. Identification of Mutations in 90 of 121 Consecutive Symptomatic French Patients With a Type I Protein C Deficiency. *Blood* 1995; 86: 2598–2605.
- 137) Grundy CB, Chisholm M, Kakkar VV, Cooper DN. A novel homozygous missense mutation in the protein C (PROC) gene causing recurrent venous thrombosis. *Hum Genet* 1992; 89: 683–684.

- 138) Grundy CB, Melissari E, Lindo V, et al. Late-onset homozygous protein C deficiency. *Lancet* 1991; 338: 575–576.
- 139) Fidalgo T, Martinho P, Salvado R, et al. Familial thrombotic risk based on the genetic background of Protein C Deficiency in a Portuguese Study. *Eur J Hum Genet* 2015; 95: 294–307.
- 140) Li M, Long GL. Identification of two novel point mutations in the human protein S gene associated with familial protein S deficiency and thrombosis. *Arterioscler Thromb Vasc Biol* 1996; 16: 1407–1415.
- 141) Choi J, Kim HJ, Chang MH, et al. A Rare Splicing Mutation in the PROS1 Gene of a Korean Patient with Type I Hereditary Protein S Deficiency. *Ann Clin Lab Sci* 2011; 41: 397–400.
- 142) Winther-Larsen A, Kjaergaard AD, Larsen OH, et al. Protein C deficiency; PROC gene variants in a Danish population. *Thromb Res* 2020; 185: 153–159.
- 143) Larsen OH, Kjaergaard AD, Hvas AM, et al. Genetic Variants in the Protein S (PROS1) Gene and Protein S Deficiency in a Danish Population. *TH Open* 2021; 5(4): e479–e488.
- 144) Arsenic R, Treue D, Lehmann A, et al. Comparison of targeted next-generation sequencing and Sanger sequencing for the detection of PIK3CA mutations in breast cancer. *BMC Clin Pathol* 2015; 15: 1–9.
- 145) Walsh T, Casadei S, Lee MK, et al. Mutations in 12 genes for inherited ovarian, fallopian tube, and peritoneal carcinoma identified by massively parallel sequencing. *PNAS* 2011; 108(44): 18032–18037.
- 146) Herman S, Varga D, Deissler HL, et al. Medium-sized deletion in the BRCA1 gene: limitations of Sanger sequencing and MLPA analyses. *Genet Mol Biol* 2012; 35(1): 53–56.
- 147) Patel N, Ferns BR, Nastouli E, et al. Cost analysis of standard Sanger sequencing versus next generation sequencing in the ICONIC study. *The Lancet* 2016; 388: s86.
- 148) Vrtel P, Slavik L, Vodicka R, et al. Detection of Unknown and Rare Pathogenic Variants in Antithrombin, Protein C and Protein S Deficiency Using High-Throughput Targeted Sequencing. *Diagnostics* 2022; 12(5): 1060.

- 149) Lind B, Gedde-Dahl T, Tjønnfjord G, et al. Protein C Deficiency caused by Homozygosity for a Novel PROC D180G Mutation – in vitro Expression and Structural Analysis of the Mutation. *Thromb Haemost* 2002; 88: 632–638.
- 150) Iijima K, Nakamura A, Kurokawa H, et al. A homozygous protein C deficiency (Lys 192 del) who developed venous thrombosis for the first time at adulthood. *Thromb Res* 2010; 125:100–101.
- 151) Yu T, Dai J, Liu H, et al. Homozygous protein C deficiency with late onset venous thrombosis: identification and in vitro expression study of a novel Pro275Ser mutation. *Pathology* 2012; 44(4): 348–353.
- 152) Gupta A, Tun AM, Gupta K, et al. Protein S Deficiency [online]. StatPearls 2022 [cit. 2022-11-12] Available from: https://www.ncbi.nlm.nih.gov/books/NBK544344/?fbclid=IwAR3yczuXfvMMGJ__nsjgVDWi1-wf2ebYAl7QFkXJ9UIWLcXKUh8kttMRbEo.
- 153) Gupta A, Patibandla S. Protein C Deficiency. [online]. StatPearls 2022 [cit. 2022-11-12] Available from: <https://www.ncbi.nlm.nih.gov/books/NBK542222/>.
- 154) Benincasa G, Costa D, Infante T, et al. Interplay between genetics and epigenetics in modulating the risk of venous thromboembolism: A new challenge for personalized therapy. *Thromb Res* 2019; 177: 145–153.
- 155) Amendola LM, Jarvik GP, Leo MC, et al. Performance of ACMG-AMP Variant-Interpretation Guidelines among Nine Laboratories in the Clinical Sequencing Exploratory Research Consortium. *Am J Hum Genet* 2016; 98(6): 1067–1076.
- 156) Quintáns B, Ordóñez-Ugalde A, Cacheiro P, et al. Medical genomics: The intricate path from genetic variant identification to clinical interpretation. *Appl Transl Genomics* 2014; 3(3): 60–67.
- 157) Shestak AG, Makarov LM, Komoliatova VN, et al. Coexistence of Two Rare Genetic Variants in Canonical and Non-canonical Exons of SCN5A: A Potential Source of Misinterpretation. *Front Genet* 2021; 12: 722291.
- 158) Arbustini E, Behr ER, Carrier L, et al. Interpretation and actionability of genetic variants in cardiomyopathies: a position statement from the European Society of Cardiology Council on cardiovascular genomics. *Eur Heart J* 2022; 43(20): 1901–1916.

- 159) Evans JP, Skrzynia C, Burke W. The complexities of predictive genetic testing. *BMJ* 2001; 322: 1052–1056.
- 160) Novotny J, Zimovjanova M. Možnosti preventivních opatření u nosičů mutací genů BRCA1 a BRCA2. *Onkologie* 2011; 5(1): 30–33.
- 161) UpToDate. Screening for inherited thrombophilia in asymptomatic adults. 2021 [online]. [cit. 2022-12-12] Available from: https://www.uptodate.com/contents/screening-for-inherited-thrombophilia-in-asymptomatic-adults?search=enetics%20protein%20C%20deficiency&source=search_result&selectedTitle=2~98&usage_type=default&display_rank=2.
- 162) National Organization for Rare Disorders. Protein S Deficiency. 2019 [online]. [cit. 2022-12-12] Available from: <https://rarediseases.org/rare-diseases/protein-s-deficiency/>.
- 163) Shin S, Kim Y, Oh SCH, et al. Validation and optimization of the Ion Torrent S5 XL sequencer and OncoPrint workflow for BRCA1 and BRCA2 genetic testing. *Oncotarget* 2017; 8(21): 34858–34866.
- 164) Glenn TC. Field guide to next-generation DNA sequencers. *Mol Ecol Resour* 2011; 11: 759–769.
- 165) Laehnemann D, Borkhardt A, McHardy AC. Denoising DNA deep sequencing data—high-throughput sequencing errors and their correction. *Brief Bioinform* 2015; 17(1): 154–179.
- 166) Tarabeux J, Zeitouni B, Moncoutier V, et al. Streamlined ion torrent PGM-based diagnostics: BRCA1 and BRCA2 genes as a model. *Eur J Hum Genet* 2014; 22: 535–541.

12 Přílohy

Příloha 1. Kritéria pro klasifikaci patogenních variant – převzato z Richards et al. 2015 (92).

Evidence of pathogenicity	Category
Very strong	<p>PVS1 null variant (nonsense, frameshift, canonical ± 1 or 2 splice sites, initiation codon, single or multiexon deletion) in a gene where LOF is a known mechanism of disease</p> <p>Caveats:</p> <ul style="list-style-type: none"> • Beware of genes where LOF is not a known disease mechanism (e.g., <i>GFAP</i>, <i>MYH7</i>) • Use caution interpreting LOF variants at the extreme 3' end of a gene • Use caution with splice variants that are predicted to lead to exon skipping but leave the remainder of the protein intact • Use caution in the presence of multiple transcripts
Strong	<p>PS1 Same amino acid change as a previously established pathogenic variant regardless of nucleotide change</p> <p>Example: Val→Leu caused by either G>C or G>T in the same codon</p> <p>Caveat: Beware of changes that impact splicing rather than at the amino acid/protein level</p> <p>PS2 De novo (<u>both</u> maternity and paternity confirmed) in a patient with the disease and no family history</p> <p>Note: Confirmation of paternity only is insufficient. Egg donation, surrogate motherhood, errors in embryo transfer, and so on, can contribute to nonmaternity.</p> <p>PS3 Well-established in vitro or in vivo functional studies supportive of a damaging effect on the gene or gene product</p> <p>Note: Functional studies that have been validated and shown to be reproducible and robust in a clinical diagnostic laboratory setting are considered the most well established.</p> <p>PS4 The prevalence of the variant in affected individuals is significantly increased compared with the prevalence in controls</p> <p>Note 1: Relative risk or OR, as obtained from case–control studies, is >5.0, and the confidence interval around the estimate of relative risk or OR does not include 1.0. See the article for detailed guidance.</p> <p>Note 2: In instances of very rare variants where case–control studies may not reach statistical significance, the prior observation of the variant in multiple unrelated patients with the same phenotype, and its absence in controls, may be used as moderate level of evidence.</p>
Moderate	<p>PM1 Located in a mutational hot spot and/or critical and well-established functional domain (e.g., active site of an enzyme) without benign variation</p> <p>PM2 Absent from controls (or at extremely low frequency if recessive) (Table 6) in Exome Sequencing Project, 1000 Genomes Project, or Exome Aggregation Consortium</p> <p>Caveat: Population data for insertions/deletions may be poorly called by next-generation sequencing.</p> <p>PM3 For recessive disorders, detected in <i>trans</i> with a pathogenic variant</p> <p>Note: This requires testing of parents (or offspring) to determine phase.</p> <p>PM4 Protein length changes as a result of in-frame deletions/insertions in a nonrepeat region or stop-loss variants</p> <p>PM5 Novel missense change at an amino acid residue where a different missense change determined to be pathogenic has been seen before</p> <p>Example: Arg156His is pathogenic; now you observe Arg156Cys</p> <p>Caveat: Beware of changes that impact splicing rather than at the amino acid/protein level.</p> <p>PM6 Assumed de novo, but without confirmation of paternity and maternity</p>
Supporting	<p>PP1 Cosegregation with disease in multiple affected family members in a gene definitively known to cause the disease</p> <p>Note: May be used as stronger evidence with increasing segregation data</p> <p>PP2 Missense variant in a gene that has a low rate of benign missense variation and in which missense variants are a common mechanism of disease</p> <p>PP3 Multiple lines of computational evidence support a deleterious effect on the gene or gene product (conservation, evolutionary, splicing impact, etc.)</p> <p>Caveat: Because many in silico algorithms use the same or very similar input for their predictions, each algorithm should not be counted as an independent criterion. PP3 can be used only once in any evaluation of a variant.</p> <p>PP4 Patient's phenotype or family history is highly specific for a disease with a single genetic etiology</p> <p>PP5 Reputable source recently reports variant as pathogenic, but the evidence is not available to the laboratory to perform an independent evaluation</p>

LOF, loss of function; OR, odds ratio.

Příloha 2. Kritéria pro klasifikaci benigních variant – převzato z Richards et al. 2015 (92).

Evidence of benign impact	Category
Stand-alone	BA1 Allele frequency is >5% in Exome Sequencing Project, 1000 Genomes Project, or Exome Aggregation Consortium
Strong	<p>BS1 Allele frequency is greater than expected for disorder (see Table 6)</p> <p>BS2 Observed in a healthy adult individual for a recessive (homozygous), dominant (heterozygous), or X-linked (hemizygous) disorder, with full penetrance expected at an early age</p> <p>BS3 Well-established in vitro or in vivo functional studies show no damaging effect on protein function or splicing</p> <p>BS4 Lack of segregation in affected members of a family</p> <p>Caveat: The presence of phenocopies for common phenotypes (i.e., cancer, epilepsy) can mimic lack of segregation among affected individuals. Also, families may have more than one pathogenic variant contributing to an autosomal dominant disorder, further confounding an apparent lack of segregation.</p>
Supporting	<p>BP1 Missense variant in a gene for which primarily truncating variants are known to cause disease</p> <p>BP2 Observed in <i>trans</i> with a pathogenic variant for a fully penetrant dominant gene/disorder or observed in <i>cis</i> with a pathogenic variant in any inheritance pattern</p> <p>BP3 In-frame deletions/insertions in a repetitive region without a known function</p> <p>BP4 Multiple lines of computational evidence suggest no impact on gene or gene product (conservation, evolutionary, splicing impact, etc.)</p> <p>Caveat: Because many in silico algorithms use the same or very similar input for their predictions, each algorithm cannot be counted as an independent criterion. BP4 can be used only once in any evaluation of a variant.</p> <p>BP5 Variant found in a case with an alternate molecular basis for disease</p> <p>BP6 Reputable source recently reports variant as benign, but the evidence is not available to the laboratory to perform an independent evaluation</p> <p>BP7 A synonymous (silent) variant for which splicing prediction algorithms predict no impact to the splice consensus sequence nor the creation of a new splice site AND the nucleotide is not highly conserved</p>

Příloha 3. Pravidla pro kombinaci kritérií pro klasifikaci variant – převzato z Richards et al. 2015 (92).

Pathogenic	<ul style="list-style-type: none"> (i) 1 Very strong (PVS1) <i>AND</i> <ul style="list-style-type: none"> (a) ≥ 1 Strong (PS1–PS4) <i>OR</i> (b) ≥ 2 Moderate (PM1–PM6) <i>OR</i> (c) 1 Moderate (PM1–PM6) and 1 supporting (PP1–PP5) <i>OR</i> (d) ≥ 2 Supporting (PP1–PP5) (ii) ≥ 2 Strong (PS1–PS4) <i>OR</i> (iii) 1 Strong (PS1–PS4) <i>AND</i> <ul style="list-style-type: none"> (a) ≥ 3 Moderate (PM1–PM6) <i>OR</i> (b) 2 Moderate (PM1–PM6) <i>AND</i> ≥ 2 Supporting (PP1–PP5) <i>OR</i> (c) 1 Moderate (PM1–PM6) <i>AND</i> ≥ 4 supporting (PP1–PP5)
Likely pathogenic	<ul style="list-style-type: none"> (i) 1 Very strong (PVS1) <i>AND</i> 1 moderate (PM1–PM6) <i>OR</i> (ii) 1 Strong (PS1–PS4) <i>AND</i> 1–2 moderate (PM1–PM6) <i>OR</i> (iii) 1 Strong (PS1–PS4) <i>AND</i> ≥ 2 supporting (PP1–PP5) <i>OR</i> (iv) ≥ 3 Moderate (PM1–PM6) <i>OR</i> (v) 2 Moderate (PM1–PM6) <i>AND</i> ≥ 2 supporting (PP1–PP5) <i>OR</i> (vi) 1 Moderate (PM1–PM6) <i>AND</i> ≥ 4 supporting (PP1–PP5)
Benign	<ul style="list-style-type: none"> (i) 1 Stand-alone (BA1) <i>OR</i> (ii) ≥ 2 Strong (BS1–BS4)
Likely benign	<ul style="list-style-type: none"> (i) 1 Strong (BS1–BS4) and 1 supporting (BP1–BP7) <i>OR</i> (ii) ≥ 2 Supporting (BP1–BP7)
Uncertain significance	<ul style="list-style-type: none"> (i) Other criteria shown above are not met <i>OR</i> (ii) the criteria for benign and pathogenic are contradictory

Příloha 4. Nalezené varianty pro jednotlivé pacienty – variant calling ze tří různých softwarů.

Číslo pacienta								
VCF	Gen	Koordináta	Genotyp	Reference	% frekvence alternativní alely	kódující změna	změna aminokyseliny	Pokrytí
Pacient 1								
Torrent Suite	PROS1	chr3:93593119	C/C	T	100	c.2001A>G	p.Pro667=	208
		chr3:93646396	T/T	C	100	c.77-145G>A	p.?	400
	PROC	chr2:128180872	T/T	G	100	c.423G>T	p.Ser141=	162
		chr2:128184770	C/C	T	99,5	c.768T>C	p.Asp256=	399
SERPINC1								
Ion Reporter	PROS1	chr3:93593119	C/C	T	100	c.2001A>G	p.Pro667=	208
		chr3:93625033	GA/G	GA	55,49	c.347-47delT	p.?	337
		chr3:93646396	T/T	C	100	c.77-145G>A	p.?	1326
	PROC	chr2:128180872	T/T	G	100	c.423G>T	p.Ser141=	163
		chr2:128184770	C/C	T	99,76	c.768T>C	p.Asp256=	2974
SERPINC1								
NextGene	PROS1	chr3:93593119	C/C	T	100	c.2001A>G	p.Pro667=	209
		chr3:93646396	T/T	C	100	c.77-145G>A	p.?	1462
	PROC	chr2:128180872	T/T	G	100	c.423G>T	p.Ser141=	162
		chr2:128184770	C/C	T	99,7	c.768T>C	p.Asp256=	2960
SERPINC1								
Pacient 2								
Torrent Suite	PROS1	chr3:93646396	C/T	C	41,25	c.77-145G>A	p.?	400
		chr2:128177424	C/G	C	39,87	c.-21-74C>G	p.?	306
	PROC	chr2:128180872	T/T	G	100	c.423G>T	p.Ser141=	204
		SERPINC1						
Ion Reporter	PROS1	chr3:93646396	C/T	C	45,52	c.77-145G>A	p.?	1564
		chr2:128177424	C/G	C	53,54	c.-21-74C>G	p.?	396
	PROC	chr2:128180872	T/T	G	100	c.423G>T	p.Ser141=	204
		SERPINC1						
NextGene	PROS1	chr3:93646396	C/T	C	46,8	c.77-145G>A	p.?	1608
		chr2:128177424	C/G	C	53,5	c.-21-74C>G	p.?	396
	PROC	chr2:128180872	T/T	G	100	c.423G>T	p.Ser141=	205
		SERPINC1						
Pacient 3								
Torrent Suite	PROS1	chr3:93593119	C/C	T	100	c.2001A>G	p.Pro667=	226
		chr3:93619677	C/T	C	51	c.698G>A	p.Arg233Lys	400
		chr3:93646396	T/T	C	100	c.77-145G>A	p.?	400
	PROC	chr2:128180872	T/T	G	100	c.423G>T	p.Ser141=	238
		chr2:128184770	T/C	T	46,62	c.768T>C	p.Asp256=	399
SERPINC1								
Ion Reporter	PROS1	chr3:93593119	C/C	T	100	c.2001A>G	p.Pro667=	226
		chr3:93615446	T/A	T	21,97	c.939A>T	p.Leu313Phe	924
		chr3:93619677	C/T	C	52,48	c.698G>A	p.Arg233Lys	1008
		chr3:93646396	T/T	C	100	c.77-145G>A	p.?	1296
	PROC	chr2:128180872	T/T	G	100	c.423G>T	p.Ser141=	238
		chr2:128184770	T/C	T	48,99	c.768T>C	p.Asp256=	2611
SERPINC1								
NextGene	PROS1	chr3:93593119	C/C	T	100	c.2001A>G	p.Pro667=	223
		chr3:93595725	CTTTTT> CTTTTT	CTTTTT	34,6	c.1870+84del A	p.?	1162
		chr3:93619677	C/T	C	52,6	c.698G>A	p.Arg233Lys	1012
		chr3:93646396	T/T	C	100	c.77-145G>A	p.?	1328
	PROC	chr2:128180872	T/T	G	100	c.423G>T	p.Ser141=	236
		chr2:128184770	T/C	T	49	c.768T>C	p.Asp256=	2596
		chr2:128186342	TGGGGG G>TGGGGG GG	TGGGGG	26,9	c.1212delG	p.Met406Trpfs Ter15	2472
		SERPINC1						

Pacient 4										
Torrent Suite	PROS1									
	PROC	chr2:128180872	G/T	G		57,89	c.423G>T	p.Ser141=	114	
		chr2:128184770	T/C	T			51,25	c.768T>C	p.Asp256=	400
SERPINC1	chr1:173884020	A/G	A			59,4	c.79T>C	p.Trp27Arg	399	
Ion Reporter	PROS1									
	PROC	chr2:128180872	G/T	G		57,89	c.423G>T	p.Ser141=	114	
		chr2:128184770	T/C	T			51,05	c.768T>C	p.Asp256=	2762
SERPINC1	chr1:173884020	A/G	A			58,89	c.79T>C	p.Trp27Arg	2615	
NextGene	PROS1	chr3:93595725	CTTTTT> CTTTTT	CTTTTT		35,9	c.1870+84delA	p.?	1163	
		chr3:93619795	TAAAAA >TAAAAA A	TAAAAA		23,2	c.602-23delT	p.?	1263	
		chr3:93624839	TGGGGG G>TGGGG GG	TGGGGG		25,9	c.469+25delC	p.?	1075	
	PROC	chr2:128180872	G/T	G			56,6	c.423G>T	p.Ser141=	113
		chr2:128184770	T/C	T			50,8	c.768T>C	p.Asp256=	2744
	SERPINC1	chr1:173884020	A/G	A			58,6	c.79T>C	p.Trp27Arg	2600
	Pacient 5									
Torrent Suite	PROS1	chr3:93593119	T/C	T		48,97	c.2001A>G	p.Pro667=	243	
		chr3:93646396	C/T	C			46,25	c.77-145G>A	p.?	400
	SERPINC1	chr2:128180872	T/T	G			100	c.423G>T	p.Ser141=	127
		chr1:173876705	C/T	C			46,75	c.1154-53G>A	p.?	400
		chr1:173878832	T/C	T			44,12	c.1011A>G	p.Gln337=	102
chr1:173878862	T/C	T			44,23	c.981A>G	p.Val327=	104		
Ion Reporter	PROS1	chr3:93593119	T/C	T		48,97	c.2001A>G	p.Pro667=	243	
		chr3:93615446	T/A	T			20,91	c.939A>T	p.Leu313Phe	1081
		chr3:93646396	C/T	C			45,08	c.77-145G>A	p.?	1544
	SERPINC1	chr2:128180872	T/T	G			100	c.423G>T	p.Ser141=	128
		chr1:173876705	C/T	C			48,14	c.1154-53G>A	p.?	2206
chr1:173878832	T/C	T			44,66	c.1011A>G	p.Gln337=	103		
chr1:173878862	T/C	T			44,23	c.981A>G	p.Val327=	104		
NextGene	PROS1	chr3:93593119	T/C	T		48,5	c.2001A>G	p.Pro667=	241	
		chr3:93619795	TAAAAA >TAAAAA A	TAAAAA		25,8	c.602-23delT	p.?	1211	
		chr3:93646396	C/T	C			46,6	c.77-145G>A	p.?	1600
	SERPINC1	chr2:128180872	T/T	G			100	c.423G>T	p.Ser141=	128
		chr1:173876705	C/T	C			47,9	c.1154-53G>A	p.?	2196
	chr1:173878832	T/C	T			44,2	c.1011A>G	p.Gln337=	104	
chr1:173878862	T/C	T			44,2	c.981A>G	p.Val327=	104		
Pacient 6										
Torrent Suite	PROS1	chr3:93593119	T/C	T		52,94	c.2001A>G	p.Pro667=	221	
		chr3:93646396	C/T	C			43,5	c.77-145G>A	p.?	400
	PROC	chr2:128177461	G/A	G			52,25	c.-21-37G>A	p.?	400
		chr2:128180872	T/T	G			99,17	c.423G>T	p.Ser141=	121
	SERPINC1	chr1:173876561	C/G	C			47,25	c.1218+27G>C	p.?	400
		chr1:173878832	T/C	T			47,73	c.1011A>G	p.Gln337=	132
chr1:173878862	T/C	T			47,69	c.981A>G	p.Val327=	130		
Ion Reporter	PROS1	chr3:93593119	T/C	T		52,94	c.2001A>G	p.Pro667=	221	
		chr3:93646396	C/T	C			43,41	c.77-145G>A	p.?	1327
	PROC	chr2:128177461	G/A	G			52,48	c.-21-37G>A	p.?	505
		chr2:128180872	T/T	G			99,17	c.423G>T	p.Ser141=	121
	SERPINC1	chr1:173876561	C/G	C			46,9	c.1218+27G>C	p.?	2017
chr1:173878832	T/C	T			47,73	c.1011A>G	p.Gln337=	132		
chr1:173878862	T/C	T			47,69	c.981A>G	p.Val327=	130		
NextGene	PROS1	chr3:93593119	T/C	T		52,9	c.2001A>G	p.Pro667=	221	
		chr3:93595725	CTTTTT> CTTTTT	CTTTTT		25,2	c.1870+84delA	p.?	1083	
		chr3:93619795	TAAAAA >TAAAAA A	TAAAAA		21,9	c.602-23delT	p.?	1009	
	SERPINC1	chr3:93646396	C/T	C			44,5	c.77-145G>A	p.?	1349
		chr2:128177461	G/A	G			52,5	c.-21-37G>A	p.?	505
	PROC	chr2:128180872	T/T	G			99,2	c.423G>T	p.Ser141=	120
		chr1:173876561	C/G	C			46,9	c.1218+27G>C	p.?	2014
chr1:173878832	T/C	T			47,7	c.1011A>G	p.Gln337=	130		
chr1:173878862	T/C	T			48,1	c.981A>G	p.Val327=	129		

Pacient 7								
Torrent Suite	PROS1	chr3:93593119	C/C	T	100	c.2001A>G	p.Pro667=	173
		chr3:93646396	T/T	C	100	c.77-145G>A	p.?	400
	PROC	chr2:128180872	G/T	G	42,86	c.423G>T	p.Ser141=	105
		chr1:173876561	C/G	C	45,25	c.1218+27G>C	p.?	400
	SERPINC1	chr1:173878832	T/C	T	42,65	c.1011A>G	p.Gln337=	136
chr1:173878862		T/C	T	43,38	c.981A>G	p.Val327=	136	
Ion Reporter	PROS1	chr3:93593119	C/C	T	100	c.2001A>G	p.Pro667=	173
		chr3:93625033	GA/G	GA	50,98	c.347-47delT	p.?	357
	chr3:93646396	T/T	C	100	c.77-145G>A	p.?	1093	
	PROC	chr2:128180872	G/T	G	42,86	c.423G>T	p.Ser141=	105
		chr1:173876561	C/G	C	48,27	c.1218+27G>C	p.?	1966
SERPINC1	chr1:173878832	T/C	T	42,65	c.1011A>G	p.Gln337=	136	
	chr1:173878862	T/C	T	43,38	c.981A>G	p.Val327=	136	
NextGene	PROS1	chr3:93593119	C/C	T	100	c.2001A>G	p.Pro667=	167
		chr3:93646396	T/T	C	100	c.77-145G>A	p.?	1138
	PROC	chr2:128180872	G/T	G	42,5	c.423G>T	p.Ser141=	106
		chr1:173876561	C/G	C	48,1	c.1218+27G>C	p.?	1956
	SERPINC1	chr1:173878832	T/C	T	42,5	c.1011A>G	p.Gln337=	134
chr1:173878862		T/C	T	42,5	c.981A>G	p.Val327=	134	
Pacient 8								
Torrent Suite	PROS1	chr3:93593119	T/C	T	52,17	c.2001A>G	p.Pro667=	207
		chr3:93646396	C/T	C	46,5	c.77-145G>A	p.?	400
	PROC	chr2:128177461	G/A	G	45,75	c.-21-37G>A	p.?	400
		chr2:128180872	T/T	G	99,38	c.423G>T	p.Ser141=	162
	SERPINC1	chr1:173876705	C/T	C	50,5	c.1154-53G>A	p.?	400
chr1:173878832		T/C	T	49,67	c.1011A>G	p.Gln337=	153	
chr1:173878862		T/C	T	49,67	c.981A>G	p.Val327=	151	
Ion Reporter	PROS1	chr3:93593119	T/C	T	52,17	c.2001A>G	p.Pro667=	207
		chr3:93646396	C/T	C	45,18	c.77-145G>A	p.?	1452
	PROC	chr2:128177461	G/A	G	44,32	c.-21-37G>A	p.?	440
		chr2:128180872	T/T	G	99,39	c.423G>T	p.Ser141=	163
	SERPINC1	chr1:173876705	C/T	C	50,55	c.1154-53G>A	p.?	1996
chr1:173878832		T/C	T	49,67	c.1011A>G	p.Gln337=	153	
chr1:173878862		T/C	T	49,67	c.981A>G	p.Val327=	151	
NextGene	PROS1	chr3:93593119	T/C	T	51,7	c.2001A>G	p.Pro667=	209
		chr3:93595725	CTTTTT> CTTTTT	CTTTTT	36,9	c.1870+84delA	p.?	1126
		chr3:93619795	TAAAAA >TAAAAA A	TAAAAA	22,1	c.602-23delT	p.?	1105
		chr3:93624839	TGGGGG G>TGGGG	TGGGGG	20,6	c.469+25delC	p.?	964
		chr3:93646396	C/T	C	48	c.77-145G>A	p.?	1525
	PROC	chr2:128177461	G/A	G	44,4	c.-21-37G>A	p.?	439
		chr2:128180872	T/T	G	99,4	c.423G>T	p.Ser141=	163
	SERPINC1	chr1:173876705	C/T	C	50,6	c.1154-53G>A	p.?	1983
		chr1:173878832	T/C	T	50	c.1011A>G	p.Gln337=	152
		chr1:173878862	T/C	T	50	c.981A>G	p.Val327=	150
Pacient 9								
Torrent Suite	PROS1	chr3:93605355	T/TA	T	43,96	c.1156-9_1156	p.?	389
		chr3:93646396	C/T	C	45,25	c.77-145G>A	p.?	400
	PROC	chr2:128180872	T/T	G	99,75	c.423G>T	p.Ser141=	398
		chr2:128184770	C/C	T	100	c.768T>C	p.Asp256=	400
SERPINC1								
Ion Reporter	PROS1	chr3:93646396	C/T	C	46,46	c.77-145G>A	p.?	1244
		chr2:128180872	T/T	G	99,78	c.423G>T	p.Ser141=	920
	chr2:128184770	C/C	T	99,96	c.768T>C	p.Asp256=	2829	
SERPINC1								
NextGene	PROS1	chr3:93646396	C/T	C	47,9	c.77-145G>A	p.?	1287
		chr2:128180872	T/T	G	99,7	c.423G>T	p.Ser141=	924
		chr2:128184770	C/C	T	99,9	c.768T>C	p.Asp256=	2816
	PROC		TGGGGG G>TGGGG	TGGGGG			p.Met406Trpfs Ter15	
		chr2:128186342	GG	TGGGGG	21,8	c.1212delG		2347
SERPINC1	chr1:173881163	CAAAAA AAAT>CA AAAAAA AT	CAAAAAAT	22,7	c.409-13_409- 12delTT	p.?	1156	

Pacient 10								
Torrent Suite	<i>PROS1</i>							
	<i>PROC</i>	chr2:128180872	T/T	G	100	c.423G>T	p.Ser141=	400
	<i>SERPINC1</i>	chr1:173878832	T/C	T	49,45	c.1011A>G	p.Gln337=	91
chr1:173878862		T/C	T	49,45	c.981A>G	p.Val327=	91	
Ion Reporter	<i>PROS1</i>							
	<i>PROC</i>	chr2:128180872	T/T	G	99,77	c.423G>T	p.Ser141=	885
	<i>SERPINC1</i>	chr1:173878832	T/C	T	49,45	c.1011A>G	p.Gln337=	91
chr1:173878862		T/C	T	49,45	c.981A>G	p.Val327=	91	
NextGene	<i>PROS1</i>	chr3:93605355	TAAAAA> TAAAAA	TAAAAAA	34,7	c.1156-9dupT	p.?	1027
		chr3:93619795	TAAAAAA >TAAAAA A	TAAAAA	24,2	c.602-23delT	p.?	952
	<i>PROC</i>	chr2:128180872	T/T	G	99,6	c.423G>T	p.Ser141=	889
	<i>SERPINC1</i>	chr1:173878832	T/C	T	50	c.1011A>G	p.Gln337=	90
		chr1:173878862	T/C	T	50	c.981A>G	p.Val327=	90

(Červeně – patogenní varianta, nalezená u pacienta všemi testovanými softwarey; šedě – pravděpodobné sekvenační artefakty detekovány napříč pacienty ve vybraných genech)