

Univerzita Karlova

1. lékařská fakulta

Studijní program: Molekulární a buněčná biologie, genetika a virologie



UNIVERZITA KARLOVA
1. lékařská fakulta

Mgr. Petra Zemánková

**Analýza kvantitativních a kvalitativních genetických znaků
v patogenezi hereditárních forem solidních nádorů**

**Analysis of quantitative and qualitative genetic features in the pathogenesis of
hereditary solid tumors**

Autoreferát disertační práce

Školitel: prof. MUDr. Zdeněk Kleibl, Ph.D.

Školitel konzultant: Mgr. Viktor Stránecký, Ph.D.

Praha, 2019

Doktorské studijní programy v biomedicině
Univerzita Karlova a Akademie věd České republiky

Obor: Molekulární a buněčná biologie, genetik a virologie

Předseda oborové rady: doc. RNDr. Dana Holá, Ph.D.

Školící pracoviště: Ústav biochemie a experimentální onkologie

Školitel: prof. MUDr. Zdeněk Kleibl, Ph.D.

Školitel konzultant: Mgr. Viktor Stránecký, Ph.D.

Disertační práce bude nejméně pět pracovních dnů před konáním obhajoby zveřejněna k nahlížení veřejnosti v tištěné podobě na Oddělení pro vědeckou činnost a zahraniční styky Děkanátu 1. lékařské fakulty.

Abstrakt

Nádorová onemocnění patří mezi druhou nejčastější příčinu úmrtí v ČR. Nosiči mutací v genech predisponujících k dědičné formě onemocnění tvoří malou, ale klinicky velmi významnou skupinu vysoce rizikových osob. V současné době jsou známy desítky predispozičních genů pro vznik hereditárních nádorových syndromů, pro jejichž analýzu se cílené sekvenování nové generace (NGS) stalo metodou první volby. NGS umožňuje rapidní zrychlení určení příčinné mutace v oblasti diagnostiky hereditárních nádorových syndromů. K identifikaci mutací v genech predisponujících ke vzniku dědičných nádorových onemocnění jsme navrhli analýzu pomocí panelového NGS včetně bioinformatického zpracování, které umožňuje spolehlivou identifikaci jednonukleotidových záměn, krátkých inzercí/delecí i rozsáhlých intragenových přestaveb. Bioinformatické postupy, popsáné v této dizertační práci, jsme následně využili k validaci panelového NGS, ale i pro identifikaci alterací v konkrétních genech, která umožnila nalézt jejich doposud nepopsané asociace s dědičnými nádorovými onemocněními. Bioinformatické analýzy se staly základem pro jednotné zpracování rozsáhlých souborů dat z CZEKANCA konsorcia a umožňují konstrukci frekvenční databáze variant, která slouží pro zlepšení klinické diagnostiky nádorové predispozice u pacientů v ČR. Všestrannost NGS umožňuje jeho využití i na analýzu sestřihových variant nádorových predispozičních genů, která je nezbytným předpokladem pro identifikaci patogenních variant způsobujících aberantní sestřih.

Klíčová slova:

Hereditární nádorové syndromy, sekvenování nové generace, bioinformatická analýza

Abstract

Cancer the second most common causes of death in the Czech Republic. Carriers of mutations in genes predisposing to hereditary cancers represent a small but clinically significant group of high risk individuals. Today, dozens of predisposing genes for hereditary tumor syndromes are known and targeted next generation sequencing (NGS) has become a standard approach for their analysis. NGS allows rapid acceleration diagnostics of causal mutation in high-risk individuals. To identify mutations in genes predisposing to hereditary cancers, we designed a panel NGS analysis including subsequent bioinformatics analysis allowing a reliable identification of single nucleotide variants, insertions/deletions, and large intragenic rearrangements. The bioinformatics procedures described in this thesis were used for panel NGS validation, but also for identification of alterations associating with so far undescribed hereditary tumor types. Bioinformatics analyzes have become the basis for the unified processing of large datasets from the CZEKANCA consortium and enable the construction of a population-specific database of genotypes that serve to improve clinical diagnostics of cancer predisposition in Czech patients. The versatility of NGS also allows its use for RNA (cDNA-based) analyzes of splicing variants in the genes of interest, which prerequisite for aberrant splicing identification.

Key words:

Hereditary cancer syndromes, next generation sequencing, bioinformatics analysis

1 ÚVOD

Nádorová onemocnění představují po onemocněních kardiovaskulárního systému druhou nejčastější příčinu úmrtí v naší populaci (28% všech úmrtí v roce 2016; data: www.uzis.cz). Česká republika zaujímá přední místo v celoevropských statistikách incidence nádorů (Arnold M. et al., 2015).

V naší laboratoři se zabýváme především karcinomy prsu, ovaria, pankreatu a tlustého střeva, které zahrnují časté nebo prognosticky nepříznivé nádorové diagnózy.

1.1 Hereditární nádorové syndromy

U všech onkologických diagnóz převládají **sporadická onemocnění** vznikající na základě akumulace somatických mutací genomové DNA. Tyto mutace nejčastěji postihují protoonkogeny, kódující pozitivní regulátory růstu buněk a tkání, nebo tumor supresorové geny, které regulují buněčný růst negativně, indukují apoptózu nebo se podílejí na opravách poruch genomové DNA (Hanahan D., Weinberg R., 2000).

Přibližně 3% všech nádorových onemocnění však vznikají jako **hereditární nádory** (Rahman N., 2014). Hereditární nádorové syndromy mají původ v germinálních mutacích. Dědičnost hereditárních mutací nádorových predispozičních genů je převážně autosomálně dominantní. V některých případech je zastoupení dědičných nádorů podstatně vyšší než zmíněná 3%; > 15% u ovariálních karcinomů, >20% u medulárního karcinomu štítné žlázy a >30% případů u feochromocytomu (Rahman N., 2014).

Identifikace příčinné genetické změny je nezbytným předpokladem prediktivního testování, které v případě přítomnosti patogenní mutace u doposud asymptomatického nosiče, umožňuje jeho zařazení do preventivních programů, jejichž cílem je snížení rizika a v ideálním případě eliminace možnosti vzniku onkologického onemocnění. Přestože podíl hereditárních forem na celkovém počtu nádorových onemocnění je malý, vysoké riziko vzniku onemocnění činí diagnostiku dědičných forem klinicky závažným úkolem, jehož řešení významně přispívá ke zlepšení kvality života nosičů mutací.

Nádorových predispozičních genů bylo doposud popsáno několik set. Největší rozmach zaznamenala charakterizace „hlavních“ predispozičních genů v posledním desetiletí 20. století, kdy byly popsány i mutace v hlavních predispozičních genech pro vznik **hereditární formy karcinomu prsu a ovaria** *BRCA1* (Miky Y. et al., 1994) a *BRCA2* (Wooster R. et al., 1995). Dědičné patogenní mutace zvyšují riziko vzniku karcinomu prsu u nosičů téměř desetinásobně a zvyšují také riziko vzniku karcinomu ovarii a nádorů dalších tkání. Mutace v genu *BRCA1* zvyšují riziko vzniku karcinomu kolorekta, *BRCA2* pak přispívá také ke vzniku karcinomu prsu u mužů, nádorů pankreatu a prostaty (Foulkes W., 2008). Proteiny *BRCA1* i *BRCA2* jsou velké jaderné fosfoproteiny, jejichž dominantní funkcí je podíl na opravách dvouřetězcových zlomů DNA cestou homologní rekombinace (Nielsen F.C. et al., 2016). Kromě *BRCA1* a *BRCA2* jsou s dědičnou predispozicí ke karcinomu prsu a ovaria spojeny i mutace v dalších genech kódujících DNA reparační proteiny (*PALB2* - Rahman N., 2007; *RAD51C* - Pelttari L.M. et al., 2011; *RAD51D* - Loveday C. et al., 2011), nebo proteiny regulující odpověď na poškození DNA (*CHEK2* a *TP53*; Kleibl Z., Kristensen V., 2016).

Asociace s dědičnými poruchami genů kódujících DNA reparační proteiny, které se podílejí na opravách dvouřetězcových zlomů cestou homologní rekombinace, je typickou, i když ne výlučnou, charakteristikou dědičných forem karcinomu prsu a ovaria. Velmi vzácně se vyskytující homozygotní mutace (či přítomnost *trans* patogenní mutací u složených heterozygotů) způsobují vrozená onemocnění provázená závažnými symptomy a časnou letalitou spojenou obvykle se vznikem maligních onemocnění. Typickým příkladem těchto syndromů je Fanconiho anémie, heterogenní onemocnění,

kteřá se projevuje mnohočetnými vrozenými vadami, selháním kostní dřene a vysokou predispozicí k hematologickým i solidním malignitám (Che R. et al., 2017; Nalepa G. a Clapp D.W., 2018). Fanconiho anemii způsobují mutace v některém z 21 FA genů (*FANCA až FANCW*) kódujících proteiny podílející se na opravě zkřížených vazeb a reparaci dvouřetězcových zlomů (Che R. et al., 2017).

Na rozdíl od hereditárních forem karcinomu prsu a ovaria s dědičnými mutacemi v DNA reparačních genech pro proteiny homologní rekombinace, mutace v rodinách s **dědičným nepolypózním karcinomem kolorekta** – Lynchovým syndromem – jsou způsobeny dědičnými mutacemi mismatch-repair (MMR) genů – *MLH1, MSH2, MSH6, PMS2* (Lynch H.T. et al., 2015). Predilekci poruch MMR genů ke karcinomu kolorekta dokumentují i v nedávné době popsané hereditární mutace v genech pro DNA polymerázy (*POLE, POLDI*) účastníci se této reparační cesty, které zvyšují riziko oligopolypózních forem kolorektálního karcinomu (Palles C. et al., 2013).

1.2 Přístupy ke studiu genetické podstaty hereditárních nádorových syndromů

Identifikace osob s dědičně podmíněnými nádorovými syndromy vyžaduje průkazné určení patogenní mutace. S ohledem na skutečnost, že nádorových predispozičních a kandidátních genů bylo identifikováno několik set, jejich mutace se liší **penetrancí** (mírou asociovaného rizika vzniku onemocnění), jedná se o poměrně náročný problém. Podstatnou komplikací je rovněž nedostatek informací, které máme o vztahu genotypu s fenotypem u postižených osob a rodin, protože, s výjimkou několika málo genů (jako je *BRCA1, BRCA2*, či mutátorové geny), jsou dědičné mutace v nádorových predispozičních genech vzácné a často populačně či geograficky specifické.

1.2.1 Vazebné a GWAS analýzy

Charakterizace nádorových predispozičních genů vyžadovala významný technologický pokrok v metodách molekulární biologie a DNA diagnostiky, ke kterému přispěl objev PCR a sekvenování. Zásadní posun v identifikaci nádorových predispozičních genů měly na počátku **vazebné analýzy**, kterými bylo v letech 1980-1990 identifikováno několik predispozičních genů (Foulkes W., 2008). Modifikací vazebných analýz jsou **celogenomové asociační studie** (Genome-Wide Associations studies; GWAS), analyzující frekvenci jednonukleotidových polymorfismů napříč celým genomem v populaci u tisíců zdravých a nemocných jedinců. V současné době dokázaly GWAS studie identifikovat varianty, které jsou asociovány se zvýšeným rizikem vývoje nádorů (Sud A. et al., 2017; Michailidou K. et al., 2017).

1.2.2 Instrumentální přístupy pro analýzu nádorové predispozice

V roce 1977 byl popsán princip tzv. **Sangerova sekvenování**, které výrazně zrychlilo analýzu DNA (Sanger F. et al., 1977). **Sekvenování nové generace** (Next Generation Sequencing - NGS; masivní paralelní sekvenování) je dalším generačním stupněm sekvenování, umožňujícím paralelní analýzu milionů DNA templátů v jediné analýze. Třebaže výkonnost NGS dokáže analyzovat sekvenci celého lidského genomu, většina současných postupů rutinní genetické diagnostiky je založena na použití různě rozsáhlých skupin vyšetřovaných genů - **genových panelů** (Soukupová J., 2016). K jejich selektivnímu **obohacení** (sequence capture) lze využít mnoha technologických a koncepčních přístupů (Ballester L.Y. et al., 2016; Kozarewa I. et al., 2015). Omezení velikosti cílové oblasti sekvenované DNA umožňuje paralelní zpracování vzorků od více vyšetřovaných osob (multiplexing), což zrychluje analýzy v reálném provozu a snižuje jejich finanční náročnost (Shearer A.E. et al., 2012).

Základem každého NGS je **příprava sekvenační knihovny**. V případě analýz predispozice ke vzniku geneticky podmíněných onemocnění slouží jako obvyklý templát genomová DNA. Zásadní limitací současné rutinně používaných technologií NGS je délka čtení omezená na úseky DNA stovek bází. Proto na počátku přípravy knihovny musí být vysokomolekulární genomová DNA štěpena pomocí ultrazvuku nebo enzymaticky. Po štěpení jsou DNA fragmenty upraveny, aby byly připraveny na navázání adaptorů pro namnožení cílových fragmentů a barkódů k rozlišení DNA fragmentů patřících ke konkrétní osobě.

V této práci se budu zabývat dvěma panely, které jsme v naší laboratoři postupně vyvinuli pro analýzu nádorové predispozice a sloužily k sekvenování na technologických platformách SOLiD a Illumina.

SOLiD sekvenování (Sequencing by Oligonucleotide Ligation and Detection) využívá k sekvenačnímu procesu hybridizaci fluorescenčně značených krátkých oligonukleotidů (prob) hybridizujících ke čtenému úseku DNA, které jsou spojovány pomocí DNA ligázy v několika cyklech (Shendure J. et al., 2005). SOLiD sekvenování vynikalo v podobě nízkého počtu sekvenačních chyb. Zásadní omezení této technologie však spočívaly v nemožnosti čtení delších fragmentů DNA, časové náročnosti ligačních cyklů a dosažení omezeného sekvenačního výstupu. Z těchto důvodů nemohlo SOLiD sekvenování čelit nástupu konkurenčních platforem a z dnešní perspektivy se jedná o ukončenou vývojovou větev (Shendure J. et al., 2017).

Dominantní platformou NGS v současnosti je **Illumina**, využívající k amplifikaci tzv. můstkovou PCR, kdy jsou jednořetězcové fragmenty templátů opatřené adaptory hybridizovány na primery imobilizované na dně sekvenační komory (flowcell). Každý fragment během amplifikace kolem sebe vytvoří tzv. cluster (Fedurco M. et al., 2006) - shluk identických kopií daného fragmentu DNA sekvenovaných pomocí DNA polymerázy a značených nukleotidů. Výhodou sekvenačních platforem Illumina je především rychlost a efektivnost, nevýhodou chyby v substitucích, kvůli narůstajícímu šumu pozadí (Ari S. a Arikan M., 2016).

Dalším stupněm technologického rozvoje je tzv. **sekvenování třetí generace**. Výhodou sekvenování třetí generace je výrazně větší délka čtení a vynechání amplifikačního kroku.

2 BIOINFORMATICKÁ ANALÝZA

Sekvenování nové generace produkuje velké množství dat a vyžaduje rozsáhlou bioinformatickou analýzu. Zpracování sekvenačních dat probíhá dominantně na výkonných výpočetních serverech v operačním systému Linux, který je vhodný pro práci s velkými objemy dat (Gb – Tb informací). Analýza sekvenačních dat se skládá z řady na sebe navazujících kroků, které obvykle zahrnují kontrolu kvality, mapování sekvenačních čtení na referenční sekvenci, nalezení odchylek (variant) oproti referenční genomové sekvenci a jejich anotaci.

Primárním výstupem sekvenačních dat jsou soubory v podobě **FASTQ** (FASTA formát spolu s kvalitou každé báze) formátu, což jsou všechny sekvence jednotlivých analyzovaných fragmentů (reads - čtení), ohodnocené kvalitou, vyjadřující pravděpodobnost přítomnosti dané báze. FASTQ soubory představují vstupní formát pro úvodní kontrolu kvality (např. v softwaru FastQC., Andrews S. 2010). V případě přítomnosti nekvalitních bází nebo tzv. pročení se do adaptorů lze sekvenci upravit tzv. trimmingem – odstranění nekvalitních bází na konci jednotlivých čtení nebo sekvencí adaptorů.

FASTQ soubory jsou mapovány na referenční genom. Výsledkem mapování jsou data v **SAM** (sorted alignment map) formátu obsahující sekvenci čtení, její koordinátu (lokalizaci) v referenčním genomu a odchylky oproti referenční sekvenci. Pro další zpracování jsou SAM soubory převedeny do binární podoby – **BAM** formátu, seřazeny a jsou z nich odstraněny PCR duplikáty. Volitelným krokem je

rekalibrace bází, která umožňuje detekci a odstranění systémových chyb sekvenačního procesu. Takto upravené BAM soubory jsou vstupem pro vlastní hodnocení sekvenačních dat, genotypování, CNV analýzu (viz kapitola 2.3), analýza středně velkých inzercí, delecí a pro vizualizaci hodnocené oblasti.

Finálním krokem bioinformatického zpracování je genotypování – nalezení odchylek od referenčního genomu. Výstupem tohoto kroku je soubor ve formátu **VCF** (Variant Call Format), obsahující všechny nalezené varianty, které jsou následně funkčně anotovány.

2.1 Mapování

Základním krokem vyhodnocení sekvenačního výstupu z NGS je mapování získaných sekvencí na referenční genom. S ohledem na současný stav NGS technologií je typickou situací mapování velkého množství krátkých (stovky bp dlouhých) DNA fragmentů. Omezená délka čtení může komplikovat mapování v oblastech repetitivních sekvencí či pseudogenů.

Existuje velké množství mapovacích programů s rozdílnými algoritmy. Některé z nich jsou přednostně určeny pro kratší čtení, jiné zase pro čtení o velikosti nad 100 bp. Úkolem mapovacích algoritmů je lokalizovat pozici sekvenačního čtení v referenčním genomu. Vyhledávání této podobnosti není založeno na přesné shodě sekvence čtení a referenčního genomu, protože v přítomnosti odchylky (varianty) od referenčního genomu přesná shoda nikdy nenastane. Každý algoritmus má proto nastavenou jistou toleranci vůči neshodě se vzorem v genomu. Dva hlavní mapovací přístupy jsou tzv. „**seed and extend**“ algoritmus a „**q-gram filter**“ (Ye H. et al., 2015). Generování základních fragmentů (seedů nebo q-gramů), lze provést pomocí tzv. **hashovací tabulky** nebo s využitím **FM indexu** (Ferragina P. a Manzini G., 2000). Oba procesy jsou založeny na vyhledávání podobnosti v textu.

Výstupem mapovacích programů jsou obvykle soubory ve formátu SAM, které jsou poté převedeny na BAM soubory. Následuje odstranění PCR duplikátů. Takto upravené soubory jsou pak podrobeny dvěma volitelným analýzám – „**realignmentu**“ - znovuzhodnocení inzercí nebo delecí v oblasti, kde se varianta vyskytuje a **rekalibraci** kvality bází – detekování sekvenačních chyb.

2.2 Genotypování

Stanovení genotypu jedinců je obecně sumarizace kvality bází a hloubky čtení z BAM souborů. Dva hlavní způsoby genotypování jsou pomocí bayesiánské pravděpodobnosti, nebo pomocí heuristických faktorů, díky nimž se vytvoří sada pravidel, které dávají vznik genotypům. Nejrozšířenějšími nástroji jsou **SAMTOOLS** (Li H., 2011) a **GATK** (Genome Analysis Toolkit, McKenna A. et al., 2010), pracující na základě bayesiánské pravděpodobnosti. Oba dokáží zajistit i více kroků dříve popsaných úprav (odstranění PCR duplikátů, realignment, recalibrace kvality bází).

2.3 Analýza počtu kopií a velkých přestaveb

Důležitým krokem NGS analýzy dat je analýza velkých přestaveb tzv. „**CNV calling**“. K identifikaci CNV ze sekvenačních dat je možné využít několik přístupů. **Pair-endové mapování** je založeno na sekvenování pomocí pair-endového čtení. Identifikace variant pak spočívá v detekci rozdílné vzdálenosti mezi páry čtení. Další přístup je založen na vyhledání oblastí, ve kterých se vyskytují tzv. rozdělená čtení (**split read-based**). Rozpoznání CNV závisí na namapování částí jednoho čtení na jinou oblast genomu. Nejčastěji používaný přístup je **analýza založená na hloubce pokrytí** daného

segmentu. Další možností je identifikace CNV variant založená na novém sestavení genomové sekvence nebo kombinace předchozích způsobů (Zhao M. et al., 2012).

2.4 Anotace

Posledním krokem bioinformatického zpracování NGS dat je anotace variant – pojmenování biologické funkce a významu variant. V současnosti existuje několik databází poskytující referenční genomové sekvence a definice genů včetně jejich transkripčních variant. Nejpoužívanější databáze jsou **UCSC** (Fujita P. et al., 2011), **RefSeq** (Pruitt K. et al., 2012) a **ENSEMBL** (Flicek P. et al., 2012).

Nalezené varianty jsou dále anotovány frekvencemi z populačních databází např. ESP6500, 1000g (Genomes Project C. et al., 2012), ExAC a gNOMAD (Das R. a Ghosh S.K., 2017). K zachyceným variantám je s výhodou přiřadit i informace o klinické závažnosti, které jsou uvedeny např. v databázi ClinVar (Landrum M. J. et al., 2014).

Pro analýzu funkčního dopadu doposud nepopsaných variant nebo variant nejasného významu (VUS) byla vytvořena rozsáhlá kolekce predikčních programů, založených na různých algoritmech jako je **SIFT** (Kumar P. et al., 2009), **PolyPhen-2** (Adzhubei A. et al., 2010), **LRT** (Chun S. a Fay J.C., 2009), **MutationTaster** (Schwarz J.M. et al., 2010), **PhyloP**, (Pollard K.S. et al., 2010), **CADD** (Kircher M. et al., 2014) nebo **GERP** (Cooper G.M. et al., 2005). Pro *in silico* predikci poruch sestřihových signálů lze využít specializované predikční algoritmy (např. **spidex**, **dbNSFP**; Liu X., 2011).

3 VÝCHODISKA A CÍLE PRÁCE

V roce 2010 proto byly v naší laboratoři zahájeny první pokusy se sekvenováním pomocí NGS. Výsledky komerčního sekvenování na panelu 80 genů u 60 nosičů známých mutací však především ukázaly, že pro další rozvoj NGS v laboratoři bude nezbytné zajistit laboratorní i bioinformatické analýzy vlastními silami. Zatímco změna laboratorních postupů, jakkoliv zásadní, umožňovala využít stávajících zkušeností molekulárně biologických technik, které byly v laboratoři dostupné, množství dat generovaných NGS vyžadovalo koncepční změnu interpretace vyšetření a především hodnocení získaných nálezů.

Cílem dizertační práce bylo:

- **vytvoření robustního bioinformatického postupu pro hodnocení NGS dat,**
- **vytvoření postupů pro kontrolu spolehlivosti analýz,**
- **zavedení postupů pro prioritizaci variant, které umožní charakterizovat kandidátní genetické prognostické a prediktivní faktory ovlivňující vznik a vývoj dědičných nádorových onemocnění analyzovaných v naší laboratoři,**
- **vytvoření databáze genotypů s fenotypovými charakteristikami pacientů (histopatologická a klinická data o onemocnění) umožňující jejich statistické zpracování.**

4 SEZNAM PRACÍ, SLOUŽÍCÍCH JAKO PODKLAD DIZERTAČNÍ PRÁCE

V časopisech s IF (bez IF – **označeny**)

1. Lhota F, **Zemankova P**, Kleiblova P, Soukupova J, Vocka M, Stranecky V, Janatova M, Hartmannova H, Hodanova K, Kmoch S, Kleibl Z. Hereditary truncating mutations of DNA repair and other genes in *BRCA1/BRCA2/PALB2*-negatively tested breast cancer patients. *Clin Genet.* 2016;90(4):324-33, (IF₂₀₁₆= 3.326).
2. **Zemankova P**, Lhota F, Kleiblova P, Soukupova J, Vocka M, Janatova M, Kleibl Z. RE: Frameshift variant FANCL*c.1095_1099dupATTA is not associated with high breast cancer risk. *Clin Genet.* 2016;90(4):387-9, (IF₂₀₁₆=3.326).
3. Rump A, Benet-Pages A, Schubert S, Kuhlmann JD, Janavičius R, Macháčková E, Foretová L, Kleibl Z, Lhota F, **Zemankova P**, Betcheva-Krajcir E, Mackenroth L, Hackmann K, Lehmann J, Nissen A, DiDonato N, Opitz R, Thiele H, Kast K, Wimberger P, Holinski-Feder E, Emmert S, Schröck E, Klink B. Identification and Functional Testing of ERCC2 Mutations in a Multi-national Cohort of Patients with Familial Breast- and Ovarian Cancer. *PLoS Genet.* 2016;12(8):e1006248, (IF₂₀₁₆=6.100).
4. Borecka M, **Zemankova P**, Lhota F, Soukupova J, Kleiblova P, Vocka M, Soucek P, Ticha I, Kleibl Z, Janatova M. The c.657del5 variant in NBN gene predisposes to pancreatic cancer. *Gene.* 2016;587(2):169-72, (IF₂₀₁₆=2.415).
5. Soukupová J, **Zemánková P**, Kleiblová P, Janatová M, Kleibl Z. CZE CANCA: CZEch CAncer paNel for Clinical. Application – návrh a příprava cíleného sekvenčního panelu pro identifikaci nádorové predispozice u rizikových osob v České Republice. *Klin Onkol.* 2016;29 Suppl 1:S46-54. (bez IF)
6. Soukupova J, **Zemankova P**, Lhotova K, Janatova M, Borecka M, Stolarova L, Lhota F, Foretova L, Machackova E, Stranecky V, Tavandzis S, Kleiblova P, Vocka M, Hartmannova H, Hodanova K, Kmoch S, Kleibl Z. Validation of CZE CANCA (Czech CAncer paNel for Clinical Application) for targeted NGS – based analysis of hereditary cancer syndromes. *PLoS One.*;13(4):e0195761, (IF₂₀₁₈=2.766).
7. Kleiblova P, Stolarova L, Krizova K, Lhota F, Hojny J, **Zemankova P**, Havranek O, Vocka M, Cerna M, Lhotova K, Borecka M, Janatova M, Soukupova J, Sevcik J, Zimovjanova M, Kotlas J, Panczak A, Vesela K, Cervenкова J, Schneiderova M, Burocziova M, Burdova K, Stranecky V, Foretova L, Machackova E, Tavandzis S, Kmoch S, Macurek L, Kleibl Z. Identification of deleterious germline CHEK2 mutations and their association with breast and ovarian cancer. *Int J Cancer.* 2019 May 3. (IF₂₀₁₈= 7.360)
8. Hojny J, **Zemankova P**, Lhota F, Sevcik J, Stranecky V, Hartmannova H, Hodanova K, Mestak O, Pavlista D, Janatova M, Soukupova J, Vocka M, Kleibl Z, Kleiblova P. Multiplex PCR and NGS-based identification of mRNA splicing variants: Analysis of BRCA1 splicing pattern as a model. *Gene.* 2017; 30(637):41-49, (IF₂₀₁₇=2.498).

Seznam dalších spoluautorských prací v časopisech s IF (bez IF – **označeny**)

9. Vocka M, Zimovjanova M, Bielcikova Z, Tesarova P, Petruzelka L, Mateju M, Krizova L, Kotlas J, Soukupova J, Janatova M, **Zemankova P**, Kleiblova P, Novotny J, Konopasek B, Chodacka M, Brychta

- M, Sochor M, Smejkalova-Musilova D, Cmejlova V, Kozevnikovova R, Miskarova L, Argalacsova S, Stolarova L, Lhotova K, Borecka M, Kleibl Z. Estrogen Receptor Status Oppositely Modifies Breast Cancer Prognosis in BRCA1/BRCA2 Mutation Carriers Versus Non-Carriers. *Cancers* (Basel). 2019 May 28;11(6).
10. Pejsova H, Hubacek JA, **Zemankova P**, Zlatohlavek L. Baseline Leptin/Adiponectin Ratio is a Significant Predictor of BMI Changes in Children/Adolescents after Intensive Lifestyle Intervention. *Exp Clin Endocrinol Diabetes*. 2019 Mar 6. doi: 10.1055/a-0859-7041.
 11. Lovecek M, Janatova M, Skalicky P, Zemanek T, Havlik R, Ehrmann J, Strouhal O, **Zemankova P**, Lhotova K, Borecka M, Soukupova J, Svebisova H, Soucek P, Hlavac V, Kleibl Z, Neoral C, Melichar B, Mohelnikova-Duchonova B. Genetic analysis of subsequent second primary malignant neoplasms in long-term pancreatic cancer survivors suggests new potential hereditary genetic alterations. *Cancer Manag Res*. 2019;11:599-609.
 12. Burke LJ, Sevcik J, Gambino G, Tudini E, Mucaki EJ, Shirley BC, Whiley P, Parsons MT, De Leeneer K, Gutiérrez-Enríquez S, Santamariña M, Caputo SM, Santana Dos Santos E, Soukupova J, Janatova M, **Zemankova P**, Lhotova K, Stolarova L, Borecka M, Moles-Fernández A, Manoukian S, Bonanni B; ENIGMA Consortium, Edwards SL, Blok MJ, van Overeem Hansen T, Rossing M, Diez O, Vega A, Claes KBM, Goldgar DE, Rouleau E, Radice P, Peterlongo P, Rogan PK, Caligo M, Spurdle AB, Brown MA. BRCA1 and BRCA2 5' noncoding region variants identified in breast cancer patients alter promoter activity and protein binding. *Hum Mutat*. 2018;39(12):2025-2039.
 13. Stránecký V, Neřoldová M, Hodaňová K, Hartmannová H, Piherová L, **Zemánková P**, Přistoupilová A, Vrablík M, Adámková V, Kmoch S, Jirsa M. Large copy-number variations in patients with statin-associated myopathy affecting statin myopathy-related loci. *Physiol Res*. 2016;65(6):1005-1011.
 14. Borecka M, **Zemankova P**, Vocka M, Soucek P, Soukupova J, Kleiblova P, Sevcik J, Kleibl Z, Janatova M. Mutation analysis of the PALB2 gene in unselected pancreatic cancer patients in the Czech Republic. *Cancer Genet*. 2016;209(5):199-204.
 15. Janatová M, Borecká M, Soukupová J, Kleiblová P, Stříbrná J, Vočka M, **Zemánková P**, Panczak A, Veselá K, Souček P, Foretová L, Kleibl Z. PALB2 jako další kandidátní gen pro genetické testování u pacientů s hereditárním karcinomem prsu v České republice. *Klin Onkol*. 2016;29 Suppl 1:S31-4.
 16. Janatova M, Soukupova J, Stribrna J, Kleiblova P, Vocka M, **Boudova P**, Kleibl Z, Pohlreich P. Mutation Analysis of the RAD51C and RAD51D Genes in High-Risk Ovarian Cancer Patients and Families from the Czech Republic. *PLoS One*. 2015 Jun 9;10(6):e0127711.

5 KOMENTÁŘ K VYBRANÝM PUBLIKOVANÝM PRACÍM

5.1 Článek 1: Hereditary truncating mutations of DNA repair and other genes in BRCA1/BRCA2/PALB2-negatively tested breast cancer patients.

Lhota F, **Zemankova P**, Kleiblova P, Soukupova J, Vocka M, Stranecky V, Janatova M, Hartmannova H, Hodanova K, Kmoch S, Kleibl Z. *Clin Genet*. 2016;90(4):324-33, (IF₂₀₁₆= 3.326).

Uvedená práce shrnuje naše počáteční úsilí o analýzu nádorové predispozice pomocí NGS. V práci jsme navrhli sekvenční panel pro cílené obohacení vzorků zaměřené na analýzu 581 genů, pomocí kterého jsme analyzovali soubor 325 pacientů s karcinomem prsu negativně testovaných na hlavní predispoziční geny *BRCA1/BRCA2/PALB2* a 105 nenádorových kontrol na platformě SOLiD.

Panel 581 genů jsme navrhli tak, že 141 genů se účastní přímo DNA reparačních procesů a 449 genů bylo vybráno z databáze Phenopedia (Yu W. et al., 2010).

V rámci analýz jsme kodifikovali postup pro bioinformatické analýzy sekvenčních dat. Tento postup zahrnoval mapování pomocí softwaru Novoalign (CS1.01.08), odstranění duplikátů pomocí softwaru

Picard tools a SAMtools (0.1.8). Varianty byly anotovány pomocí ANNOVARu. Filtrace variant pak probíhala na základě kvality >150 a pokrytí >10x. Mezní hodnotu kvality jsme určili na základě validace různých SNP. Díky frekvenčním databázím ESP6500 a 1000g jsme byli schopni vyloučit varianty s frekvencí >0.01. Dále byly odstraněny populačně časté varianty, u kterých není pravděpodobné, že by se podílely na podstatném zvýšení rizika vzniku nádorů. V posledním kroku byla provedena klasifikace variant a analyzovány varianty nejasného významu (VUS). Za patogenní jsme považovali varianty vedoucí ke zkrácení proteinového produktu z důvodu porušení čtecího rámce (pokud nebyly klasifikovány jinak v databázi ClinVar – např. rekurentní benigní varianta p.K3326* v genu *BRC A2*). Dále byly jako patogenní varianty označeny změny kanonických sestřihových míst (c.+/- 1 a 2) a intronové varianty způsobující aberantní sestřih při analýze z RNA. Jako patogenní byly rovněž hodnoceny varianty klasifikované jako patogenní a potenciálně patogenní v databázi ClinVar, které neměly konfliktní hodnocení. Nesynonymní nové varianty, které byly identifikovány predikčními programy jako škodlivé (SIFT, PolyPhen-2, LRT, MutationTaster a PhyloP) byly označeny jako potenciálně patogenní. Za nevýznamné jsme považovali varianty, které byly jako benigní či pravděpodobně benigní uvedeny v databázích ClinVar a HGMD.

Za použití výše popsané filtrace jsme z počátečního počtu 491,385 variant získali 4,540 vzácných variant (2,647 unikátních variant). V souboru pacientů jsme identifikovali 127 truncačních variant, 34 variant in-frame delecí/inzercí, z 1,599 nesynonymních unikátních variant jich 356 bylo identifikováno potenciálně patogenních. Varianty zkracující proteinový produkt jsme našli celkem u 32% pacientů, u 9% pacientů byly zaznamenány germinální varianty v klinicky významných predispozičních genech.

Největší pozornost jsme věnovali variantám vedoucím ke zkrácení proteinového produktu s posunem čtecího rámce. Celkem 36 těchto variant jsme našli u 25 genů kódujících proteiny zúčastněné v DNA reparačních pochodech. Mezi nejčastěji postižené geny patřily geny kódující proteiny komplexu Fanconiho anemie. V analyzovaném souboru jsme u šesti pacientek s karcinomem prsu a (žádné kontroly) zachytili variantu c.1096_1099dupATTA genu *FANCL*. Proteinový produkt genu *FANCL* je klíčová ubikvitin ligáza FA core komplexu. Varianta c.1096_1099dupATTA přiléhá k PHD/RING finger doméně katalyzující ubikvitinylaci a již dříve byla identifikována Alim A. M. et al., 2009 u pacienta s mírnými projevy Fanconiho anemie komplementační skupiny L. Proto jsme provedli dodatečnou genotypizaci této varianty na souboru 337 vysoce rizikových pacientů s karcinomem prsu, 673 pacientů se sporadickým karcinomem prsu a u 686 nenádorových kontrol pomocí High-Resolution Melting Analysis (HRMA). Varianta c.1096_1099dupATTA po dodatečné analýze identifikována u 9/662 (1,3%) vysoce rizikových pacientů, 3/673 (0,4%) pacientek se sporadickým karcinomem prsu a u 3/791 (0,4%) kontrol. Statisticky významný rozdíl ve frekvenci výskytu byl zachycen pouze mezi vysoce rizikovými pacienty a kontrolami ($p_{Fisher} = 0.04$).

Funkční vztahy mezi ostatními 48 geny postiženými 53 germinálními truncačními mutacemi v ostatních („nereparačních“) genech u 74/325 (22,8%) pacientů a 10/105 (9,5%) kontrol jsme analyzovali pomocí funkčních anotací v nástrojích STRING a KEGG. Tyto geny jsme zařadili do devíti funkčních skupin a naznačili některé další možné funkční skupiny genů ovlivňující nádorovou predispozici ke vzniku karcinomu prsu.

Výsledky této studie nám umožnili vyhodnotit proveditelnost panelového NGS pro analýzu nádorové predispozice od návrhu panelu, přes optimalizace fragmentace DNA, obohacení cílových oblastí, konstrukci sekvenačních knihoven, bioinformatickou analýzu a interpretaci nálezů. Studie zároveň umožnila identifikovat některá slabá místa analýz, jako značnou velikost cílového panelu s obtížnou interpretací variant v genech s málo známým funkčním vztahem ke karcinomu prsu a časově i ekonomicky náročné sekvenování na platformě SOLiD.

5.2 Článek 2: RE: Frameshift variant FANCL*c.1095_1099dupATTA is not associated with high breast cancer risk.

Zemankova P, Lhota F, Kleiblova P, Soukupova J, Vocka M, Janatova M, Kleibl Z. *Clin Genet.* 2016;90(4):387-9, (IF₂₀₁₆=3.326).

V reakci na předchozí článek Lhota et al., 2016 (Kapitola 5.1), ve kterém jsme identifikovali zvýšený výskyt varianty c.1096_1099dupATTA v genu *FANCL* u pacientek s dědičnou formou karcinomu prsu, provedli Pfeifer et al., 2016 genotypizaci této varianty u pacientů a kontrol z Německa a Makedonie. V analyzovaných souborech identifikovali variantu c.1096_1099ATTA u 3/887 pacientů a 5/976 kontrol z Německa a u 1/278 pacientů a 1/229 kontrol z Makedonie. Autoři tak zpochybňují klinický význam našich zjištění.

V naší odpovědi jsme k dříve publikovaným výsledkům rozšířené analýzy Lhota et al., 2016 (Kapitola 5.1), která identifikovala 15 nosičů c.1096_1099dupATTA varianty doplnili klinické a histopatologické charakteristiky nosičů mutací a identifikovali jsme dalších osm nosičů této varianty identifikovaných v rámci sekvenování panelem CZECANCA (viz Kapitola 5.5). Analýza klinicko-patologických dat ukázala, že na rozdíl od Pfeifer et al., kteří našli pozitivní rodinnou anamnézu s přítomností karcinomu prsu pouze u jednoho z 10 (10%) identifikovaných nosičů c.1096_1099dupATTA, v našem souboru byla tato zachycena u 9/23 (39%) nosičů a přítomnost jakéhokoli nádoru jsme zaznamenali u 15/23 (65%) c.1096_1099dupATTA. Nápadná asociace s familiárním výskytem nádorových onemocnění u nosičů varianty tak naznačuje, že varianta může modifikovat riziko vzniku nádorového onemocnění u nosičů mutací, avšak ke zhodnocení tohoto účinku bude nezbytné provedení analýzy velmi rozsáhlých souborů pacientů a kontrol.

5.3 Článek 3: Identification and Functional Testing of ERCC2 Mutations in a Multi-national Cohort of Patients with Familial Breast- and Ovarian Cancer

Rump A, Benet-Pages A, Schubert S, Kuhlmann JD, Janavičius R, Macháčková E, Foretová L, Kleibl Z, Lhota F, **Zemankova P**, Betcheva-Krajcir E, Mackenroth L, Hackmann K, Lehmann J, Nissen A, DiDonato N, Opitz R, Thiele H, Kast K, Wimberger P, Holinski-Feder E, Emmert S, Schröck E, Klink B., *PLoS Genet.* 2016;12(8):e1006248, (IF₂₀₁₆=6.1).

Po publikování našeho prvního článku z panelového sekvenování (Článek 1), ve kterém jsme identifikovali dvě posunové mutace postihující gen *ERCC2*, jsme byli přizváni ke spolupráci na mezinárodní studii zaměřené na význam *ERCC2* mutací u pacientek s dědičným karcinomem prsu a ovarií. Gen *ERCC2* kóduje DNA helikázovou podjednotku (*ERCC2/XPD*) transkripčního faktoru IIIH, která se podílí na nukleotidové excizní reparaci (NER) DNA. Dědičné bi-alelické mutace *ERCC2* genu se manifestují jako tři zcela rozdílná onemocnění: cerebro-okulo-facio-skeletální syndrom 2, fotosenzitivní trichothiodystrofie 1, nebo xeroderma pigmentosum D (XPD). Vznik XPD je spojen se zvýšeným výskytem kožních tumorů, avšak přítomnost monoalelických mutací *ERCC2* genu nebyla u pacientek s familiárním výskytem karcinomu prsu a ovaria studována.

Ve studii se podařilo analyzovat 1,345 pacientů (včetně 325 z naší laboratoře) a 2,400 kontrol (včetně 345 z naší laboratoře) z Německa, ČR a Litvy analyzovaných pomocí panelového NGS. U pacientů bylo identifikováno celkem 5 trunkačních a raritních 20 missense variant, které kolegové z Německa podrobili funkčním in vitro analýzám.

Přesto, že trunkační a missense mutace klasifikované jako funkčně-defektní varianty byly uvedeny v databázi ExAC jako velmi raritní (s výjimkou jediné všechny se vyskytovaly s frekvencí <0,05%),

analýzy populačně-porovnatelných kontrolních vzorků odhalily přítomnost i některých patogenních variant se značně prevalencí (např. frekvence posunové mutace p.F568fs* v naší populaci dosahovala 0,43% u pacientů, ale také 0,44% v kontrolách). Segregační analýza v dostupných rodinách nosičů mutací neprokázala jasnou asociaci fenotypu s přítomností patogenních variant *ERCC2* genu.

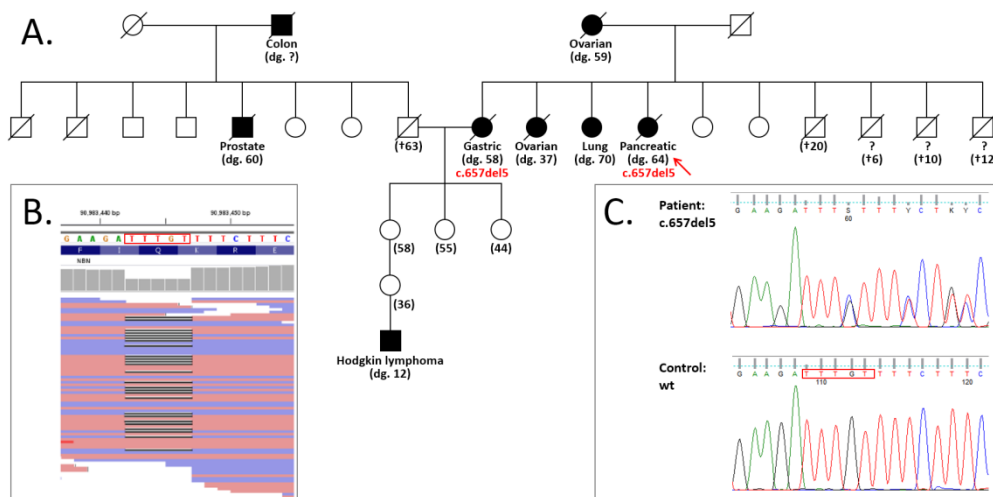
Výsledky studie tak nepotvrdily předpoklad, že nosičství patogenních mutací genu *ERCC2* je spojeno se zvýšeným rizikem vzniku karcinomu prsu a ovaria. Přinesly však cenné pozorování, které ukazuje, že analýza nádorové predispozice u genů s neúplnou penetrancí a s raritními dědičnými variantami bude vyžadovat pro porovnání analýzu dostatečně velkého souboru etnicky a geograficky srovnatelné kontrolní populace. Z důvodu možného „founder“ efektu se pro hodnocení frekvencí výskytu variant v těchto genech nelze spoléhat pouze na veřejné multietnické databáze.

5.4 Článek 4: The c.657del5 variant in NBN gene predisposes to pancreatic cancer

Borecka M, **Zemankova P**, Lhota F, Soukupova J, Kleiblova P, Vocka M, Soucek P, Ticha I, Kleibl Z, Janatova M., *Gene*. 2016;587(2):169-72, (IF₂₀₁₆=2.415).

Při přípravě dat k publikaci článku Lhota F. et al., 2016 (Článek 1) jsme provedli re-evaluaci klinických a histopatologických údajů analyzovaných pacientů, která ukázala, že 13 sekvenovaných vzorků pochází od pacientů s jinou diagnózou, než je karcinom prsu. Tyto vzorky byly z práce vyřazeny. Jedním z vyřazených byl vzorek pacientky s karcinomem pankreatu s mnohočetným nádorovým výskytem v rodině. Při analýze vzorku probandky byla zachycena mutace v genu *NBN* c.657del5, kterou jsme následně potvrdili i v histologickém bločku z nádoru u její sestry s ca žaludku (Obr. 1).

Obr. 1. Rodokmen pacientky s karcinomem pankreatu (A) s mutací c.657del5 v genu *NBN*, identifikované pomocí NGS (B) a konfirmované Sangerovým sekvenováním (C)



Gen *NBN* kóduje nibrin, součást MRN komplexu, katalyzujícího zahájení procesu homologní rekombinace při reparaci dvouřetězcových zlomů DNA v S a G2 fázi buněčného cyklu (Carney J. P. et al., 1998). Bialelické patogenní mutace v tomto genu způsobují vznik Nijmegen-breakage syndromu a heterozygotní mutace zvyšují riziko lymfoidních malignit a jiných nádorů (Varon R. et al., 1998), jako např. karcinomu prsu (Gorski B. et al., 2003), non-Hodgkinova lymfomu (Steffen J. et al., 2006) a karcinomu prostaty (Cybulski C. et al., 2013). Mutace c.657del5 v genu *NBN* byla popsána jako častá varianta v naší populaci (Varon R. et al., 1998).

Protože asociace přítomnosti této varianty s karcinomem pankreatu nebyla známa, provedli jsme genotypizaci c.657del5 varianty pomocí HRMA na souboru 241 neselektovaných pacientů s karcinomem pankreatu. Při této analýze jsme zachytili dalších 5 nosičů mutace c.657del5 v souboru pacientů (2,07%). Výskyt u pacientů s karcinomem pankreatu byl signifikantně vyšší, než výskyt této varianty v kontrolním souboru (2/915; 0,23%), což naznačuje, že mutace v genu *NBN* mohou zvyšovat riziko vzniku karcinomu pankreatu (OR 9.7; 95% CI: 1,9 – 50,2; $p=0,006$) a reprezentují tak další predispoziční gen pro hereditární formu tohoto závažného onemocnění, které je šestým nejčastějším nádorem v České republice a tumorem s velmi nepříznivou prognózou (pětileté přežití je pouze 7%, s mediánem přežití 6 měsíců; Siegel R. L., et al., 2015).

Souběžně s publikací našich výsledků byla publikována studie Polských autorů (Lener M.R. et al 2016), která zaznamenala srovnatelnou frekvenci c.657del5 varianty u pacientů s karcinomem pankreatu v Polsku (8/383; 2.09%), která se rovněž lišila od frekvence v kontrolách (22/4,000; 0,55%; OR 3.80; $p=0,002$).

5.5 Článek 5: CZECANCA: CZEch Cancer paNel for Clinical. Application – návrh a příprava cíleného sekvenačního panelu pro identifikaci nádorové predispozice u rizikových osob v České Republice

Soukupová J, Zemánková P, Kleiblová P, Janatová M, Kleibl Z. *Klin Onkol.* 2016;29 Suppl 1:S46-54., ([časopis bez IF](#)).

Zkušenosti z přípravy sekvenačního panelu, sekvenování a bioinformatického zpracování popisovaném v Článku 1 (Lhota F. et al., 2016) jsme využili při návrhu panelu a testování, určenému ke klinickému využití v diagnostice nosičů patogenních mutací pro vznik dědičných nádorových onemocnění v ČR.

Cílové geny jsme identifikovali na základě důkladné literární rešerše, ze které vyplynul seznam cílových genů. Kromě genů s již stanovenou klinickou relevancí, jsme se rozhodli zahrnout i DNA reparační geny, u nichž je klinický význam zatím nejasný. Panel byl navržen v online NimbleDesign softwaru (NimbleGen, Roche) za přísných podmínek, znemožňujících navrženým próbám nasednout více jak na tři cílové sekvence v genomu. Cílová oblast o velikosti přibližně 600 kb zahrnovala všechny popsané exony, intron-exonové přechody a pro vybrané geny i promotorové oblasti.

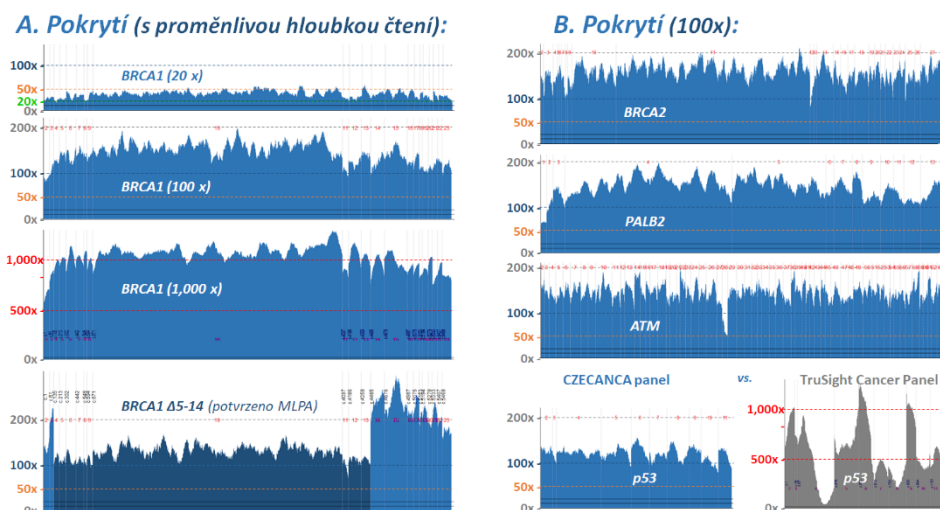
Od počátku jsme plánovali využití panelu i v dalších laboratořích v ČR, což by umožnilo analyzovat velké množství indikovaných osob a získat tak údaje o výskytu i raritních variant ve studovaných genech v naší populaci. Proto byl součástí projektu CZECANCA nejen sekvenační panel, ale zároveň optimalizovaný postup pro přípravu sekvenačních knihoven a jednotný postup pro bioinformatickou analýzu umožňující snazší interpretaci nalezených variant. Příprava knihovny CZECANCA panelu byla optimalizována na platformě Illumina MiSeq, jako nejrozšířenější NGS platformy současnosti.

Analyzované geny byly rozříděny do skupiny A-C podle klesající klinické významnosti. Z analytického hlediska jsme velkou pozornost při přípravě panelu věnovali rovnoměrnosti pokrytí cílových oblastí. Pro jeho rutinní kontrolu jsme vyvinuli skript Boudalyzer (v prostředí R) umožňující rychlou vizuální kontrolu celkového pokrytí hodnocených genů ([Obr. 2](#)).

Koncept CZECANCA projektu umožnil vznik národního konsorcia, ke kterému se do současné doby připojilo devět laboratoří analyzujících vzorky pacientů s podezřením na nádorovou predispozici. Účastníci projektu společně sdílejí hrubá sekvenační data (prostřednictvím portálu BaseSpace), která zpracováváme v naší laboratoři jednotným postupem pro konstrukci frekvenční databáze. Genotypové

údaje jsou účastníky doplněny o klinicko-patologické charakteristiky sekvenovaných pacientů s onkologickými onemocněními.

Obr. 2. Zobrazení kódujících oblastí libovolných genů umožňuje vlastní skript Boudalyzer. V panelu A ukazuje, že rovnoměrnost pokrytí při analýze BRCA1 není závislá na hloubce čtení a umožňuje dobré rozpoznání CNV. V panelu B ukazuje, že rovnoměrné pokrytí při analýze cílené na pokrytí 100x je dosahováno s panelem CZEKANCA i u jiných genů. V posledním obrázku je porovnání s panelem TruSight.



5.6 Článek 6: Validation of CZEKANCA (Czech CAncer paNel for Clinical Application) for targeted NGS – base analysis of hereditary cancer syndromes

Soukupova J, **Zemankova P**, Lhotova K, Janatova M, Borecka M, Stolarova L, Lhota F, Foretova L, Machackova E, Stranecky V, Tavandzis S, Kleiblova P, Vocka M, Hartmannova H, Hodanova K, Kmoch S, Kleibl Z. *PLoS One*;13(4):e0195761, (IF₂₀₁₈=2.766).

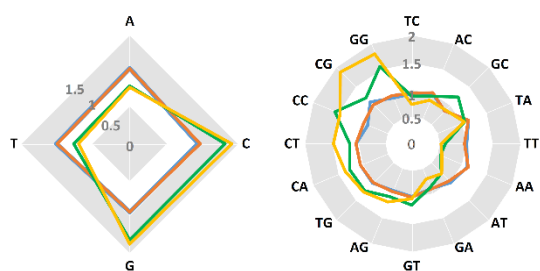
Před zavedením sekvenačního panelu CZEKANCA do rutinní klinické diagnostiky hereditárních nádorových syndromů jsme provedli důkladné testování parametrů panelu.

Testování ukázalo, že v rutinních podmínkách se sekvenováním cíleným na průměrné pokrytí 100x dosahuje sekvenování panelem CZEKANCA 100x pokrytí v 85,8% a pokrytí 50x v 98,4% cílové oblasti. Pouze 0,3% cílové oblasti je pokryto <20x, což je hodnota spolehlivosti při analýze heterozygotních germinálních variant.

Uniformita pokrytí v datech je velmi vyhovující a umožňuje snadnější a spolehlivější detekci CNV variant. Uniformita panelu vzhledem k reprodukovatelnosti výsledků, byla ověřena několika způsoby. Do jednoho runu jsme zahrnuli DNA z jednoho vzorku, která byla zpracována v odlišných koncentracích (33 ng – 100%, 24.75 ng – 75% a 16.5 ng – 50%). Výsledky ukazují, že 289/293 (98,6%) variant bylo nalezeno ve všech třech replikátech.

Další analýza funkčnosti panelu proběhla testováním stejné DNA v rozdílných sekvenačních analýzách. Z 356 unikátních variant se jich v obou nezávislých sekvenačních analýzách nacházelo 354 (99,4%). Dodatečná analýza ukazuje na uniformitu zpracování ve čtyřech laboratořích, kde je CZEKANCA rutinně používána. Bylo detekováno 332 unikátních variant. Jeden vzorek v rámci čtyřech pracovišť tedy vykazuje shodu 331/332 (99,7%), 327/332 (98,5%), 329/332 (99,1%) a 329/332 (99,1%).

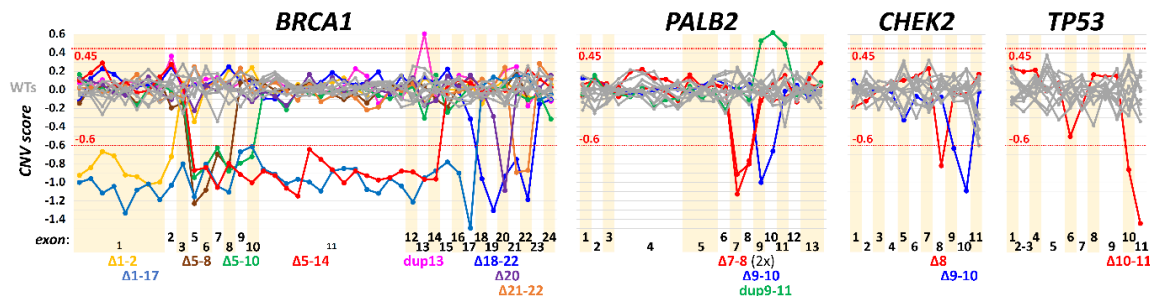
Mírné rozdíly mezi jednotlivými laboratořemi lze částečně přisoudit odlišnému způsobu fragmentace genomové DNA. Analýzou terminálních nukleotidů a dinukleotidů sekvenovaných fragmentů DNA po štěpení pomocí fragmentázy a ultrazvukové fragmentaci jsme prokázali, že ultrazvukovým štěpením lze dosáhnout rovnoměrnější fragmentace DNA (Obr. 3).



Obr. 3. Normalizovaný poměr výskytu terminálních nukleotidů a dinukleotidů sekvenovaných fragmentů štěpených ultrazvukem (červená a modrá) a enzymaticky (žlutá a zelená) provedená ve čtyřech nezávislých laboratořích. Při zcela náhodném štěpení by poměr terminálních nukleotidů a dinukleotidů měl být identický (blízký 1). Tomu se blíží fragmentace pomocí ultrazvuku, zatímco enzymatické štěpení preferenčně lépe fragmentuje CG-bohaté oblasti.

CNV analýzu jsme optimalizovali na panelu CZEKANCA analýzou více než 300 vzorků souběžně analyzovaných pomocí MLPA. Po úpravě protokolu k dosažení uniformního a dostatečného pokrytí, což byl základ pro tuto analýzu, jsme dospěli k softwaru CNVkit (Talevich E., 2016), fungujícího na bázi Linuxu. Díky pozitivním kontrolám, jsme provedli dodatečnou analýzu pomocí vlastního skriptu v prostředí R. Teprve po sloučení obou postupů jsme dosáhli 100% záchytu všech sekvenovaných pozitivních kontrol s velkými přestavbami souběžně analyzovanými pomocí MLPA (Obr. 4).

Obr. 4. Výsledek kalibrace CNV analýzy na našem pracovišti. Hodnota < -0.6 znamená přítomnost delecce, hodnota > 0.5 znamená přítomnost duplikace. Panel BRCA1 obsahuje devět pozitivních kontrol – osm delecí a jednu duplikaci, panel PALB2 obsahuje čtyři pozitivní kontroly – tři delecce a jednu duplikaci, panel CHEK2 obsahuje dvě delecce a panel TP53 jednu delecí.



Kvalitativní parametry a výsledky interních kontrol kvality detekce SNV a CNV jsme doplnili o nezávislé hodnocení European Molecular Genetics Quality Network (www.emqng.org), kdy jsme dosáhli dle externího hodnocení 100% senzitivity v detekci variant. Další kontrola kvality zahrnovala sekvenování pěti vzorků od Coriell Institute for Medical Research, pro které jsou dostupné vysoce kvalitní a verifikované BAM a VCF soubory umožňující vyčíslení falešné pozitivivity/negativity.

Výsledky validace sekvenačního panelu CZEKANCA prokázaly, že analýza pomocí tohoto systému je funkční a vyhovující pro současnou oblast diagnostiky v detekci variant s vysokou specificitou, senzitivitou a robustností z hlediska klinické praxe. Umožňuje analyzovat a spolehlivě detekovat nejen SNV a malé inserce/delece, ale i CNV u většiny z vyšetřovaných genů. Analýzy rovněž prokázaly, že bioinformatické zpracování je plně srovnatelné se zavedenými standardy a bioinformatický postup analýz všech vzorků členů CZEKANCA konsorcia povede ke konstrukci hodnotné frekvenční databáze germinálních variant u vysoce rizikových osob.

5.7 Článek 7: Identification of deleterious germline CHEK2 mutations and their association with breast and ovarian cancer.

Kleiblova P, Stolarova L, Krizova K, Lhota F, Hojny J, **Zemankova P**, Havranek O, Vocka M, Cerna M, Lhotova K, Borecka M, Janatova M, Soukupova J, Sevcik J, Zimovjanova M, Kotlas J, Panczak A, Vesela K, Cervenкова J, Schneiderova M, Burocziowa M, Burdova K, Stranecky V, Foretova L, Machackova E, Tavandzis S, Kmoch S, Macurek L, Kleibl Z. *Int J Cancer*. 2019 May 3, doi: 10.1002/ijc.32385. (IF₂₀₁₇=7.360).

Frekvenční databáze CZECANCA obsahuje v současnosti genotypová data od více než 6000 pacientů. Mezi nimi výrazně převažují pacientky s diagnózou karcinomu prsu a ovarií. Mutace u těchto pacientek dominantně postihují geny *BRCA1* a *BRCA2*, třetí nejčastější jsou mutace v genu *CHEK2*.

Gen *CHEK2* kóduje kinázu *CHK2* aktivovanou proteinem *ATM* v odpovědi na přítomnost dvouřetězcových zlomů v DNA, která následně fosforyluje řadu proteinů ovlivňujících DNA reparaci (např. *BRCA1*, *KAP1*) a zástavu buněčného cyklu (např. *p53*). Vrozené mutace v genu *CHEK2* byly asociovány se vznikem různých nádorových onemocnění, včetně karcinomu prsu, kolorekta, štítné žlázy, prostaty, ledviny, nebo non-Hodgkinova lymfomu (Cybulski C. et al., 2004; Kleibl Z. et al., 2008; Havranek O. et al., 2015). Významné rozdíly v prevalenci alterací *CHEK2* genu v jednotlivých populacích vedou k diskrepantním odhadům rizika asociovaného s nosičstvím mutací *CHEK2* genu.

Cílem práce byla identifikace dědičných variant *CHEK2* genu v naší populaci u 1,928 vysoce rizikových pacientů analyzovaných na nádorovou predispozici ke karcinomu prsu a ovaria a u 3,360 nenádorových kontrol, funkční charakterizace významu nalezených missense variant, vyčíslení rizika vzniku karcinomu prsu a ovaria u nosičů trunkačních a funkčně-defektních missense variant a analýza klinických a histopatologických charakteristik nádorů u nosičů těchto alterací.

Analýza zahrnovala výsledky testování mutací *CHEK2* genu, které byly prováděny v období více než 10 let. Polovina vzorků však pocházela z dat zpracovaných panelovým NGS, které umožnilo i spolehlivou identifikaci dvou rozsáhlých delecí *CHEK2* genu, které tvořily třetinu všech nalezených trunkačních mutací. Protože v původním návrhu panelu CZECANCA nebyly z důvodů velkého výskytu pseudogenů zahrnuty exony 12-15, byly tyto oblasti analyzovány dodatečně u všech vzorků analyzovaných pomocí NGS.

V analyzovaném souboru jsme identifikovali 10 různých trunkačních mutací a 26 missense variant. Zatímco frekvence trunkačních mutací se významně lišila mezi souborem pacientů (2,39% nosičů) a kontrol (0,33% nosičů; $p=1,1 \times 10^{-14}$), celková frekvence missense alterací byla srovnatelná (4,56% vs. 3,90%; $p=0,42$). Pod vedením Dr. Macůrka z ÚMG AVČR, bylo vyvinuto funkční vyšetření variant *CHEK2* genu, které umožnilo kvantifikovat enzymovou aktivitu *CHK2* kinázy v lidských netransformovaných buňkách. Tato analýza umožnila funkčně klasifikovat 11 VUS jako varianty s podstatnou poruchou kinázové aktivity, 5 variant s částečnou poruchou kinázové aktivity, a 10 variant se zachovanou kinázovou aktivitou.

Na základě funkčních analýz jsme byli schopni vyčísřit rizika spojená s nosičstvím germinálních mutací v *CHEK2* genu u pacientek s karcinomem prsu a ovaria, které ukazují, že riziko je klinicky významné pro trunkační mutace a vznik karcinomu prsu u žen, ale i u mužů, asociace s karcinomem ovaria je méně významná. Funkčně defektní missense varianty jsou spojeny pravděpodobně s nižším rizikem než trunkační mutace.

Práce dokumentuje zásadní význam v detekci CNV pomocí panelového NGS. Analýza funkčního významu umožní klasifikovat všechny nalezené VUS popsané CZECANCA konsorciem, což zvýší klinickou výpovědní hodnotu analýz u vysoce rizikových pacientů.

5.8 Článek 8: Multiplex PCR and NGS-based identification of mRNA splicing variants: Analysis of BRCA1 splicing pattern as a model

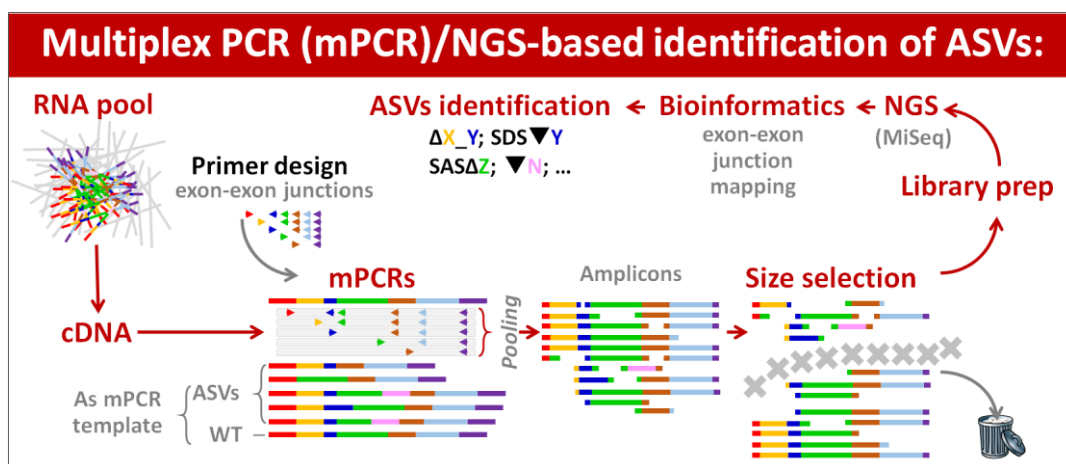
Hojny J, **Zemankova P**, Lhota F, Sevcik J, Stranecky V, Hartmannova H, Hodanova K, Mestak O, Pavlista D, Janatova M, Soukupova J, Vocka M, Kleibl Z, Kleiblova P. *Gene*. 2017; 30(637):41-49, (IF₂₀₁₇=2.498).

V naší laboratoři se dlouhodobě zabýváme analýzou genů predisponujících ke vzniku dědičných nádorových syndromů. Dlouhodobým zájmem jsou i analýzy alternativních a aberantních sestřihových variant. Fyziologicky se v buňkách vyskytují alternativně sestřižené formy proteinových produktů. V odlišných tkáních je zastoupeno různorodé procento konkrétní transkripční varianty. Varianty v DNA mohou ovlivňovat alternativní sestřih a vytvářet sestřih aberantní, který může mít značný vliv na funkci proteinu. Detekovat vliv variant na sestřih primárního transkriptu je tudíž zásadní a vyžaduje znalost fyziologicky se vyskytujících variant alternativního sestřihu.

Současné metody zahrnují různé strategie cílené amplifikace alternativních produktů, které však neumožňují kvantitativní hodnocení alternativních sestřihových forem. Výsledkem pokroku a rozvoje NGS je sekvenování RNA (RNAseq), avšak oblastí našeho zájmu jsou geny s velmi nízkou genovou expresí na úrovni ~1 TPM, a proto analýza jejich, často minoritních, alternativních sestřihových variant pomocí RNAseq je ekonomicky nevýhodná.

V naší práci jsme navrhli a na analýze sestřihových variant BRCA1 mRNA validovali nový systém pro identifikaci alternativních sestřihových variant (Obr. 5).

Obr. 5. Schéma analýzy alternativních sestřihových variant pomocí multiplexního PCR amplifikujících všechny exon-exonová spojení z cDNA a následného NGS velikostně nabohacených alternativně sestřižených (krátkých) izoforem. Nabohacení probíhá pomocí velikostní selekce, po které krátké fragmenty slouží pro přípravu sekvenční knihovny.



Navržením primerů na všechna exon-exonová spojení jsme docílili amplifikace pokrývající všechny možné varianty alternativního sestřihu BRCA1, který jsme použili jako modelový příklad. Vzorke RNA (resp. cDNA) jsme analyzovali od 32 jednotlivců – z toho 16 nenádorových kontrol, osmi pacientů s karcinomem prsu bez přítomnosti mutace v genu *BRCA1* a osmi pacientů s karcinomem prsu s *BRCA1* mutací. RNA byla izolována z leukocytů periferní krve, perimamární tukové a mamární tkáně. Paralelně jsme analyzovali RNA ze stabilních buněčných linií (MCF7, EM-G3, HeLa a MDA-MB-231).

Pro analýzu dat bylo nezbytné vyvinout vlastní postup. Data ze sekvenátoru jsme mapovali pomocí softwaru Novoalign dvojím procesem. V prvním kroku sloužil jako referenční genom fasta soubor, který jsme zkonstruovali ze sekvencí všech kombinatorických možností exon-exon spojení. V druhém kroku

byly data namapovány na DNA sekvenci daného genu, abychom zachytili exonizované introny. Před mapováním vzhledem k nejednotné délce PCR produktů, bylo nutné provést trimming adaptorů a nekvalitních bází na konci všech čtení pomocí softwaru Trimmomatic. Konverze souborů ze SAM formátu do BAM byla provedena softwarem Picard tools. Statistiky o pokrytí jednotlivých oblastí byly spočítány díky softwaru SAMTOOLS a soft-clipové báze ručně zhodnoceny ve vizualizačním softwaru IGV.

Takto navrženým systémem se nám podařilo detekovat celkem 94 variant, tedy více alternativních sestřihových variant než v dříve popsané studii (Colombo M. et al., 2014). Největší množství (72 variant) bylo detekováno v mamární tkáni, dále v leukocytech (67) a perimamární tukové tkáni (57). Ze 76 variant detekovaných v buněčných liniích jich 11 nebylo zachyceno v žádném jiném vzorku.

Postup analýzy je možné aplikovat na vyšetření libovolných transkriptů, včetně velmi málo exprimovaných genů, jako je *BRCA1*. Na rozdíl od RNAseq umožňuje naše analýza přibližně 1000x vyšší detekci kanonických exon-exonových spojení a tím i značně spolehlivější kvantifikaci i minoritních sestřihových variant (tvořících <1% z celkového sestřihu) za nesrovnatelně nižších nákladů, protože analýza *BRCA1* transkriptů všech analyzovaných vzorků spotřebovávala v našem případě méně než 10% sekvenační kapacity při sekvenování na MiSeq.

6 SHRNU TÍ A ZÁVĚR

Nástup sekvenování nové generace v poslední dekádě znamenal revoluční posun v diagnostice genetických onemocnění. NGS se v současnosti stalo rutinní metodou používanou v diagnostických laboratořích. Největší výzva v této oblasti je sestavení bioinformatického pracovního postupu pro analýzu, ale zejména interpretaci nalezených variant.

Panelové sekvenování je v současnosti nejefektivnější formou diagnostiky dle potřeby konkrétního pracoviště. S jeho nástupem se rapidně mění oblast diagnostiky příčiny hereditárních karcinomů. Jedná se o efektivní analýzu, která umožňuje diagnostiku pacientů s různými diagnózami, pro které je dostupná řada cílených panelů.

Panely umožňují výběr genů analyzujících predispozici ke konkrétním nádorovým diagnózám, avšak u pacientů s nejasným fenotypem (např. mnohočetným výskytem rozdílných nádorových diagnóz v rodině probanda) je výběr konkrétního panelu je obtížný. Z těchto důvodů jsme se při návrhu panelu CZECANCA zaměřili na širokou oblast predispozičních genů všech častých nádorových onemocnění.

Sekvenace exomů nebo genomů by nejasných případech byla ideálním postupem, avšak není ekonomicky reálná v současném rutinním provozu.

V naší laboratoři jsme prováděli vyšetření klasickým způsobem do zavedení metody NGS. S mým nástupem a optimalizací zpracování dat jsme byli schopni zpracovat panel 581 genů a vytvořit nový panel CZECANCA, který je součástí diagnostiky již v několika centrech v České republice. Díky panelovému sekvenování jsme byli schopni zpětně detekovat mutace u pacientek, které byly vyšetřeny klasickými postupy.

7 LITERATURA

- Adzhubei I.A., Schmidt S., Peshkin L. et al., 2010, A method and server for predicting damaging missense mutations. *Nat Methods.*;7(4):248-9.
- Andrews S., 2010, FastQC: a quality control tool for high throughput sequence data. Available online at:<http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
- Ali A.M., Kirby M., Jansen M. et al., 2009, Identification and characterization of mutations in FANCL gene: a second case of Fanconi anemia belonging to FA-L complementation group, *Hum Mutat.*;30(7):E761-70.
- Ari S. a Arkan M., 2016, Next Generation Sequencing: Advantages, Disadvantages and Future, In “Plant Omics-Trends and Applications” Springer International Publishing: 109-136.
- Arnold M., Karim-Kos H.E., Coebergh J.W. et al., 2015, Recent trends in incidence of five common cancers in 26 European countries since 1988: Analysis of the European Cancer Observatory, *Eur J Cancer.*;51(9):1164-87.
- Ballester L.Y., Luthra R., Kanagal-Shamanna R. et al., 2016, Advances in clinical next-generation sequencing: target enrichment and sequencing technologies, *Expert Rev Mol Diagn.*;16(3):357-72.
- Carney J.P., Maser R.S., Olivares H. et al., 1998, The hMre11/hRad50 protein complex and Nijmegen breakage syndrome: linkage of double-strand break repair to the cellular DNA damage response, *Cell.*;93(3):477-86.
- Che R., Zhang J., Nepal M. et al., 2017, Multifaceted Fanconi Anemia Signaling, *Trends Genet.*;34(3):171-183.
- Chun S. a Fay J.C., 2009, Identification of deleterious mutations within three human genomes. *Genome Res.*;19(9):1553-61.
- Colombo M., Blok M.J., Whiley P. et al., 2014, Comprehensive annotation of splice junctions supports pervasive alternative splicing at the BRCA1 locus: a report from the ENIGMA consortium, *Hum Mol Genet.*;23(14):3666-80.
- Cooper G.M., Stone E.A., Asimenos G. et al., 2005, Distribution and intensity of constraint in mammalian genomic sequence. *Genome Res.*;15(7):901-13.
- Cybulski C., Gorski B., Huzarski T. et al., 2004, CHEK2 is a multiorgan cancer susceptibility gene. *Am J Hum Genet.*;75:1131–5.
- Cybulski C., Wokołarczyk D., Kluźniak W. et al., 2013, An inherited NBN mutation is associated with poor prognosis prostate cancer, *Br J Cancer.*;108(2):461-8.
- Das R. a Ghosh S.K., 2017, Genetic variants of the DNA repair genes from Exome Aggregation Consortium (EXAC) database: significance in cancer. *DNA Repair (Amst).*;52:92-102.
- Fedurco M., Romieu A., Williams S., et al., 2006, BTA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies. *Nucleic Acids Res.* 34(3):e22.
- Ferragina P. a Manzini G, 2000, Opportunistic data structures with applications, In proceedings of the 41st Annual Symposium on foundations of computer Science (FOCS 2000). IEEE computer society.
- Flicek .P, Amode M., Barrell D. et al., 2012, Ensembl 2012. *Nucleic Acids Res.*;40:D84–D90.
- Foulkes W.D., 2008, Inherited Susceptibility to Common Cancers, *N Engl J Med.*;359(20):2143-53.
- Fujita P., Rhead B., Zweig A. et al., 2011, The UCSC genome browser database: update 2011. *Nucleic Acids Res.*;39:D876–D882.
- Genomes Project C., Abecasis G.R., Auton A. et al., 2012, An integrated map of genetic variation from 1,092 human genomes. *Nature*;491(7422):56-65.
- Górski B., Debniak T., Masojć B. et al., 2003, Germline 657del5 mutation in the NBS1 gene in breast cancer patients, *Int J Cancer.*;106(3):379-81.

- Hanahan D. a Weinberg R.A, 2000, The Hallmarks of Cancer Review, *Cell*;100:57-70.
- Havranek O., Kleiblova P., Hojny J. et al., 2015, Association of Germline CHEK2 gene variants with risk and prognosis of non-Hodgkin lymphoma. *PLoS One*;10:e0140819.
- Kircher M., Witten D.M., Jain P. et al., 2014, A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet.*;46(3):310-5.
- Kleibl Z. a Kristensen V.N., 2016, Women at high risk of breast cancer: Molecular characteristics, clinical presentation and management, *Breast*;28:136-44.
- Kleibl Z., Havranek O., Novotny J. et al., 2008, Analysis of CHEK2 FHA domain in Czech patients with sporadic breast cancer revealed distinct rare genetic alterations. *Breast Cancer Res Treat*;112:159–64.
- Kozarewa I., Armisen J., Gardner A.F. et al., 2015, Overview of Target Enrichment Strategies, *Curr Protoc Mol Biol.*;112:7.21.1-23.
- Kumar P., Henikoff S., Ng P.C., 2009, Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc.*;4(7):1073-81.
- Landrum M.J., Lee J.M., Riley G.R. et al., 2014, ClinVar: public archive of relationships among sequence variation and human phenotype, *Nucleic Acids Res.*;42 (Database issue):D980-5.
- Lener M.R., Scott R.J., Kluźniak W. et al., 2016, Do founder mutations characteristic of some cancer sites also predispose to pancreatic cancer?, *Int J Cancer.*;139(3):601-6.
- Li H., 2011, A statistical Framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data, *Bioinformatics.*;27(21):2987-93.
- Liu X., Jian X., Boerwinkle E., 2011. dbNSFP: a lightweight database of human non-synonymous SNPs and their functional predictions. *Human Mutation.* 32:894-899.
- Loveday C., Turnbull C., Ramsay E. et al., 2011, Germline mutations in RAD51D confer susceptibility to ovarian cancer, *Nat Genet.*;43(9):879-882.
- Lynch H.T., Snyder C.L., Shaw T.G. et al., 2015, Milestones of Lynch syndrome: 1895-2015. *Nat Rev Cancer*;15(3)181-94.
- McKenna A., Hanna M., Banks E. et al., 2010The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data , *Genome Res.*;20:1297-303.
- Michailidou K., Lindström S., Dennis J. et al., 2017, Association analysis identifies 65 new breast cancer risk loci, *Nature*;551(7678):92-94.
- Miki Y., Swensen J., Shattuck-Eidens D. et al., 1994, A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1, *Science*;266(5182):66-71.
- Nalepa G. a Clapp D.W., 2018, Fanconi anaemia and cancer: an intricate relationship, *Nat Rev Cancer.* 2018;18(3):168-185.
- Nielsen F.C., van Overeem Hansen T., Sørensen C.S., 2016, Hereditary breast and ovarian cancer: new genes in confined pathways, *Nat Rev Cancer.*;16(9):599-612.
- Palles C., Cazier J.B., Howarth K.M. et al., 2013, Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas, *Nat Genet.*;45(2):136-44.
- Pelttari L.M., Heikkinen T., Thompson D. et al., 2011, RAD51C is a susceptibility gene for ovarian cancer, *Hum Mol Genet.*;20(16):3278-88.
- Pfeifer K., Schürmann P., Bogdanova N. et al., 2016, Frameshift variant FANCL*c.1096_1099dupATTA is not associated with high breast cancer risk, *Clin Genet.*;90(4):385-6.
- Pollard K.S., Hubisz M.J., Rosenbloom K.R. et al., 2010, Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.*;20(1):110-21.

- Pruitt K., Tatusova T., Brown G. et al., 2012,. NCBI Reference sequences (RefSeq): current status, new features and genome annotation policy. *Nucleic Acids Res.*;40:D130–D135.
- Rahman N., 2014, Mainstreaming genetic testing of cancer predisposition genes, *Clin Med (Lond)*. 14(4):436-9.
- Rahman N., Seal S., Thompson D. et al., 2007, PALB2, which encodes a BRCA2-interacting protein, is a breast cancer susceptibility gene, *Nat Genet.*;39(2):165-7.
- Sanger F., Nicklen S. a Coulson A.R., 1977, DNA sequencing with chain-terminating inhibitors, *Proc. Nati. Acad. Sci. USA.*;74(12):5463-5467.
- Schwarz J.M., Rodelsperger C., Schuelke M. et al., 2010, MutationTaster evaluates disease-causing potential of sequence alterations. *Nat Methods.*;7(8):575-6.
- Shearer A.E., Hildebrand M.S., Ravi H., et al., 2012, Pre-capture multiplexing improves efficiency and cost-effectiveness of targeted genomic enrichment, *BMC Genomics*;13:618.
- Shendure J., Balasubramanian S., Church G.M., et al., 2017, DNA sequencing at 40: past, present and future, *Nature*;550(7676):345-353.
- Shendure J., Porreca G. J., Reppas N. B., et al., 2005, Accurate multiplex polony sequencing of an evolved bacterial genome. *Science*; 309(5741):1728-1732.
- Siegel R.L., Miller K.D., Jernal A., 2015 Multiplex single-tube screening for mutations in the Nijmegen breakage syndrome (NBS1) gene in Hodgkin's and non- Hodgkin's lymphoma patients of Slavic origin, *Eur. J. Hum. Genet.*;11:416-419.
- Soukupova J., 2016, Úskalí interpretace sekvenčních dat v diagnostice dědičných nádorových syndromů, *Labor Aktuell*; 04(16):23-26.
- Steffen J., Maneva G., Popławska L. et al., 2006, Increased risk of gastrointestinal lymphoma in carriers of the 657del5 NBS1 gene mutation, *Int J Cancer.*;119(12):2970-3.
- Sud A., Kinnersley B. a Houlston R.S., 2017, Genome-wide association studies of cancer: current insights and future perspectives, *Nat Rev Cancer.*;17(11):692-704.
- Varon R., Vissinga C., Platzer M. et al., 1998, Nibrin, a novel DNA double-strand break repair protein, is mutated in Nijmegen breakage syndrome, *Cell.*;93(3):467-76.
- Wooster R1, Bignell G, Lancaster J et al., 1995, Identification of the breast cancer susceptibility gene BRCA2, *Nature*;378(6559):789-92.
- Ye H., Meehan J., Tong W. et al., 2015, Alignment of Short Reads: A Crucial Step for Application of Next-Generation Sequencing Data in Precision Medicine, *Pharmaceutics*, 7, p. 523-541.
- Yu W., Clyne M., Khoury M.J. et al., 2010 Phenopedia and Genopedia: disease-centered and gene-centered views of the evolving knowledge of human genetic associations. *Bioinformatics*;26(1):145-6.
- Zhao M., Wang Q., Wang Q. et al., 2013, Computational tools for copy number variation (CNV) detection using next-generation sequencing data: features and perspectives, *BMC Bioinformatics.*;14 Suppl 11:S1.