



UNIVERZITA KARLOVA
I. lékařská fakulta

Doktorské studijní programy v biomedicíně

Studijní obor: Biochemie a patobiochemie

Mgr. Filip Lhota

**IDENTIFIKACE DĚDIČNÝCH ALTERACÍ PREDISPONUJÍCÍCH KE VZNIKU
KARCINOMU PRSU POMOCÍ „NEXT-GEN“ SEKVENOVÁNÍ**

**IDENTIFICATION OF HEREDITARY ALTERATIONS PREDISPOSING TO BREAST
CANCER DEVELOPMENT USING „NEXT-GEN“ SEQUENCING**

Dizertační práce

Vedoucí závěrečné práce: Doc. MUDr. Zdeněk Kleibl, PhD.

Ústav biochemie a experimentální onkologie, 1. LF UK

Praha, 2018

Prohlášení:

Prohlašuji, že jsem závěrečnou práci zpracoval samostatně a že jsem řádně uvedl a citoval všechny použité prameny a literaturu. Současně prohlašuji, že práce nebyla využita k získání jiného nebo stejného titulu

Souhlasím s trvalým uložením elektronické verze mé práce v databázi systému meziuniverzitního projektu Theses.cz za účelem soustavné kontroly podobnosti kvalifikačních prací.

V Praze, 1.6.2018

.....

FILIP LHOTA

Identifikační záznam:

LHOTA, Filip. *Identifikace dědičných alterací predisponujících ke vzniku karcinomu prsu pomocí „next-gen“ sekvenování. [identification of hereditary alterations predisposing to breast cancer development using „next-gen“ sequencing].* Praha, 2018. 187 s., 4 příl. Dizertační práce (PhD). Univerzita Karlova, 1. lékařská fakulta, Ústav biochemie a experimentální onkologie. Vedoucí práce: Doc. MUDr. Zdeněk Kleibl, PhD.

Abstrakt:

Karcinom prsu je nejčastějším zhoubným nádorovým onemocněním v populaci žen v Evropě. Přibližně 5 – 10% případů připadá na jeho dědičnou formu, která se vyznačuje vysokou penetrancí, časným nástupem, zvýšeným rizikem rekurence a vznikem maligních onemocnění v dalších lokalizacích. Mutační analýzy vysoce rizikových pacientů identifikují kauzální mutaci v některém z nejvíce studovaných predispozičních genů (*BRCA1*, *BRCA2*, *TP53*, *ATM*, *CHEK2*, *NBS1*, *PALB2*) v méně než jedné třetině případů tohoto onemocnění.

V poslední době byla, díky nástupu nových metod masivně paralelního sekvenování, charakterizována řada dalších predispozičních a kandidátních genů, ale výskyt jejich patogenních alterací se v jednotlivých geograficky odlišných populacích často liší. Významná část vysoce rizikových pacientů z rodin s dědičným karcinomem prsu může reprezentovat nosiče populačně specifických nebo i privátních mutací. Většina doposud identifikovaných predispozičních genů, jejichž mutace predisponují ke vzniku karcinomu prsu u vysoce rizikových osob, kóduje proteiny ovlivňující DNA reparační pochody, především pak reparaci dvouřetězcových zlomů DNA. Avšak mutační analýza byla provedena jen ve velmi malé části všech DNA reparačních genů. Lze předpokládat, že v doposud neanalyzovaných genech kódujících DNA reparační proteiny lze identifikovat raritní, ale klinicky významné genetické alterace způsobující nádorovou predispozici v postižených rodinách.

Práce shrnuje analýzu genů predisponujících ke karcinomu prsu v české populaci vysoce rizikových pacientů s karcinomem prsu, kteří byli negativně testováni na přítomnost mutací v genech *BRCA1/BRCA2/PALB2*. Tato analýza identifikovala trunkační mutace u 32 % pacientů. Pouze 9 % pacientů byli nosiči trunkačních variant v genech zahrnutých do současných klinických NGS panelů využívaných v predikci karcinomu prsu. Výsledky práce také ukazují zvýšené zastoupení varianty c.1096_1099dupATTA v genu *FANCL* u vysoce rizikových pacientů. Tato skutečnost ukazuje, že *FANCL* (nebo spíše pouze tato varianta) může obsahovat novou predispoziční alelu pro karcinom prsu, jejíž významnost je nezbytné potvrdit většími studii. V rámci mezinárodní spolupráce byla provedena i analýza potencionálně patogenních variant genu *ERCC2* ve vztahu k hereditární formě karcinomu prsu, která poukázala na důležitý aspekt analýz nádorově-predispozičních genů z hlediska populačně a regionálně specifického hodnocení genetického pozadí. Dále byly identifikovány i trunkační varianty ve skupině genů cytochromu p450, které kódují enzymy metabolismu steroidních hormonů, u 5 % pacientů s karcinomem prsu. Proto může tato funkční skupina genů přispět k vysvětlení doposud neodhalené, chybějící dědičnosti karcinomu prsu u vysoce rizikových pacientů, nicméně pro potvrzení této hypotézy, bude nezbytné provedení dalších studií.

Klíčová slova: karcinom prsu, next-generation sequencing, DNA reparační geny, *FANCL*, *ERCC2*

Summary:

Breast cancer (BC) is the most frequent cancer type in female population of Europe. Approximately 5 – 10 % accounts for its hereditary form which is characterized by high penetrance, early onset, risen recurrence risk and development of other cancers. Mutational analyses of high risk patients identify a predisposing mutation in one of the most studied genes (*BRCA1*, *BRCA2*, *TP53*, *ATM*, *CHEK2*, *NBS1*, *PALB2*) only in less than one third of tested breast cancer patients.

Lately, with the use of new methods of next-generation sequencing, a number of other susceptibility or candidate genes were characterized, but the incidence of their pathogenic alteration is often geographically different. A notable proportion of high risk patients from families with hereditary BC can represent carriers of population-specific, or private mutations. Most of the to date identified BC susceptibility genes codes for proteins involved in DNA repair, especially repair of double strand break DNA repair. Nevertheless the mutation analysis was conducted only on a small fraction of these DNA repair genes. We can expect that in the group of yet nontested genes coding for DNA repair proteins a rare, but clinically important genetic alterations predisposing to BC in affected families can be discovered.

This work describes a comprehensive analysis of BC susceptibility genes in Czech high-risk BC patients negatively tested for mutations in *BRCA1/BRCA2/PALB2* genes and identifies some truncating variant in 32% of analyzed patients and 9% of all patients were carriers of a truncating variant in the genes that are analyzed in the current clinical NGS panel for the BC risk prediction. Results also show an overrepresentation of the *FANCL* variant c.1096_1099dupATTA in high-risk patients, indicating that *FANCL* (or rather this variant alone) may represent a novel BC susceptibility allele, whose importance needs to be further clarified by larger studies. We have also performed an analysis of potentially pathogenic variants for hereditary breast cancer in *ERCC2* gene within an international collaboration study which outpointed an important aspect of analyses of cancer-predisposing genes from the point of population- and region-specific evaluation of genetic background. Interestingly, in other analyzed genes, truncating mutations in the group of cytochrome p450 genes coding the enzymes of steroid hormones metabolism in 5% of BC patients were identified. Therefore, this functional group may contribute to the explanation of so far undisclosed missing heritability in some high-risk BC patients; however, further studies will be necessary to confirm this hypothesis.

Key words: breast cancer, next-generation sequencing, DNA repair genes, *FANCL*, *ERCC2*

OBSAH

1	Karcinom prsu	4
1.1	Epidemiologie karcinomu prsu.....	4
1.2	Patogeneze a klasifikace karcinomu prsu	6
1.3	Mechanismy oprav genomové DNA.....	8
1.3.1	Oprava dvouřetězcových zlomů DNA.....	10
1.3.2	Reparační mechanismy dalších poškození genomové DNA.....	16
2	Geny predisponující ke vzniku dědičných forem karcinomu prsu	20
2.1	Geny s vysokou penetrancí	22
2.1.1	BRCA1.....	22
2.1.2	BRCA2.....	24
2.1.3	Další geny s vysokou penetrancí	26
2.2	Geny se střední penetrancí	28
2.3	Geny s nízkou penetrancí	31
2.4	Kandidátní geny a jejich charakterizace.....	32
3	Next generation sekvenování.....	34
3.1	Vývoj sekvenování DNA.....	34
3.2	Sangerovo sekvenování.....	34
3.3	Sekvenování nové generace, „next-generation sequencing“	35
3.3.1	GS FLX (Roche).....	37
3.3.2	Genome Analyzer (Illumina).....	38
3.3.3	SOLiD (Life technologies).....	40
3.3.4	CGA platforma (BGI).....	40
3.3.5	Ion Torrent (Life technologies).....	42
3.4	Metody single molecule sekvenování.....	43
3.4.1	PacBio.....	43
3.4.2	Nanopore.....	44
4	Pracovní hypotéza a cíle studie	45
5	Metody	46
5.1	Soubor testovaných osob.....	46
5.1.1	Rozšířený soubor pro analýzu variant genu FANCL.....	47
5.1.2	Soubor pro analýzu varianty c.657del5 genu NBN.....	47

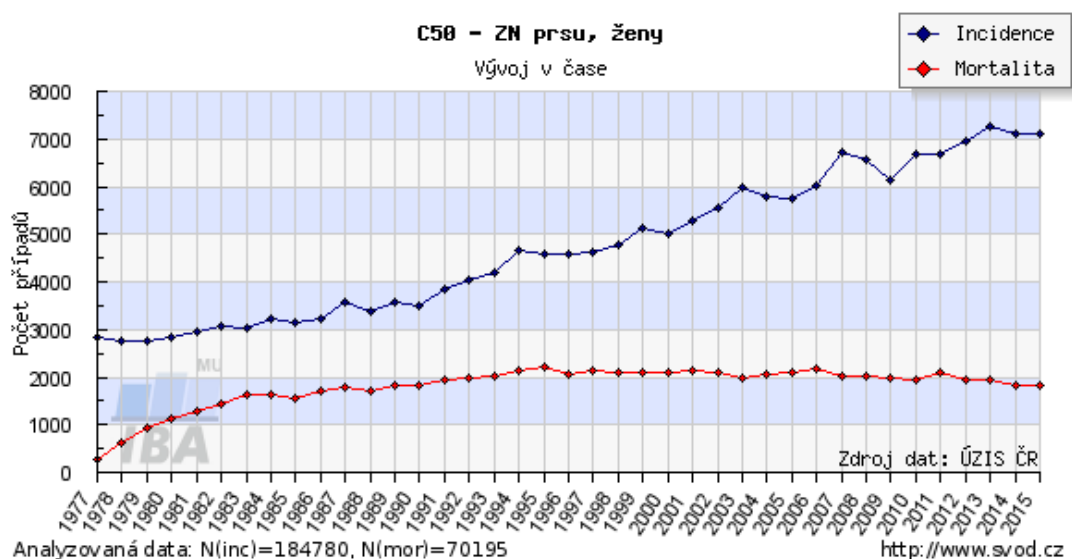
5.2	Izolace genetického materiálu	47
5.2.1	Izolace DNA z leukocytů periferní krve	47
5.2.2	Izolace celkové RNA z periferní krve	48
5.2.3	Kontrola koncentrace a kvality izolovaného genetického materiálu	48
5.3	Příprava sekvenačního panelu	48
5.4	Příprava NGS knihovny	49
5.5	Sekvenování na systému SOLiD4.....	55
5.6	Bioinformatická analýza	55
5.7	Analýza výstupních dat.....	56
5.7.1	Prioritizace missense variant.....	56
5.7.2	Statistická analýza	57
5.8	Následné analýzy potenciálně patogenních variant	57
5.8.1	HRM analýza.....	57
5.8.2	PCR amplifikace	59
5.8.3	Příprava cDNA	60
5.8.4	Izolace DNA amplifikátů z agarózového gelu	60
5.8.5	Sangerovo sekvenování.....	60
5.8.6	Amplifikační a sekvenační primery.....	61
6	Výsledky.....	62
6.1	Identifikace dědičných variant ve vybraných genech u rizikových pacientů s karcinomem prsu negativně testovaných na přítomnost mutací <i>BRCA1/BRCA2/PALB2</i> (Lhota <i>et al.</i> 2016)	62
6.1.1	Trunkační varianty u pacientů negativně testovaných na přítomnost mutací v genech <i>BRCA1</i> , <i>BRCA2</i> a <i>PALB2</i>	62
6.1.2	Hereditární varianty v DNA reparačních genech.....	64
6.1.3	Hereditární varianty v genech neúčastnících se DNA reparace	67
6.1.4	Charakteristiky onemocnění a rodinné anamnézy nosičů trunkačních variant	69
6.1.5	Nalezené varianty ovlivňující sestřih pre-mRNA	70
6.2	Rozšířená analýza výskytu varianty <i>FANCL</i> c.1096_1099dupATTA (Zemankova <i>et al.</i> 2016) 74	
6.3	Identifikace a funkční analýzy <i>ERCC2</i> mutací u pacientů s familiárním karcinomem prsu a ovaria (Rump. <i>et al.</i> 21016).....	75
6.4	Analýza varianty c.657del5 genu <i>NBN</i> ve vztahu ke karcinomu pankreatu (Boreck <i>et al.</i> 2016) 82	
6.5	Systém pro identifikaci sestřihových variant mRNA založený na amplikonovém NGS (Hojny <i>et al.</i> 2017)	84
7	Diskuse	85

8	Závěr.....	93
9	Poděkování.....	94
10	Seznam zkratk	95
11	Seznam příloh.....	97
12	Použitá literatura.....	98
13	Přílohy	109

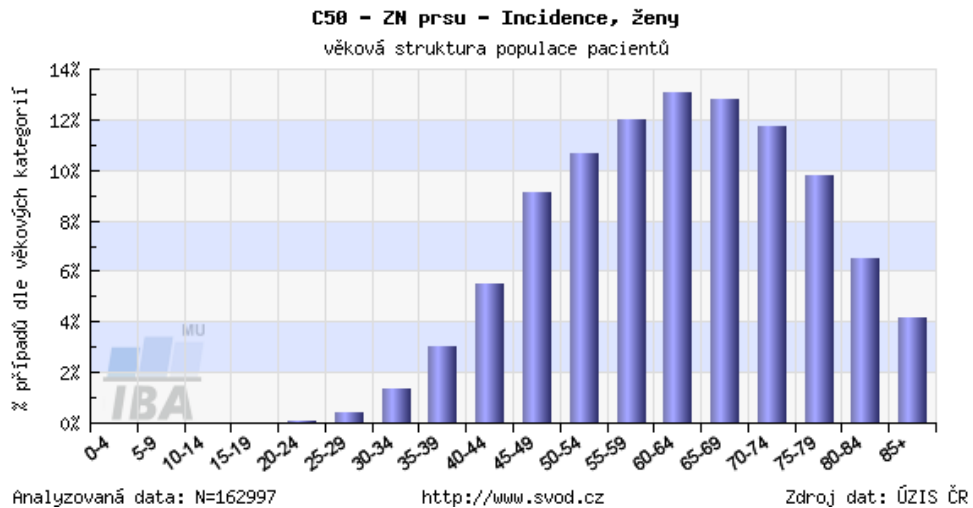
1 KARCINOM PRSU

1.1 Epidemiologie karcinomu prsu

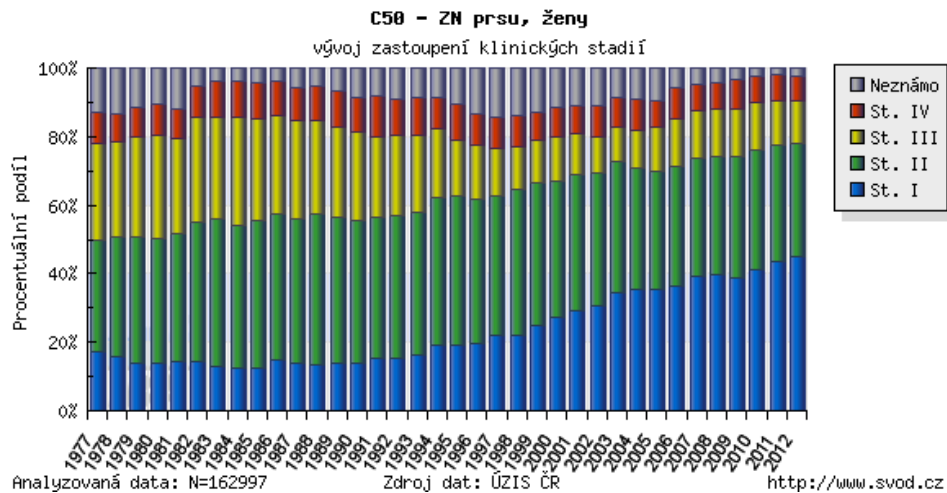
V České republice, jako i v ostatních státech Evropy, je karcinom prsu nejčastějším zhoubným nádorovým onemocněním žen (nebereme-li v potaz diagnózu C44, tedy jiný zhoubný nádor kůže). Incidence tohoto onemocnění v České republice z dlouhodobého hlediska narůstá (Obrázek 1, Novotvary 2014, ÚZIS, <http://www.uzis.cz>). Stejně tak je tomu i v ostatních evropských zemích. V roce 2015 bylo zaznamenáno 7005 nových případů onemocnění. Hrubá incidence se v roce 2015 pohybovala na hodnotě 123 nemocných na 100 000 žen. Nejvyšší výskyt karcinomu prsu u žen zaznamenáváme ve věkové skupině 60-65 let (Obrázek 2). Pozitivní vývoj můžeme pozorovat při pohledu na statistiku zastoupení jednotlivých stádií onemocnění v době diagnózy, neboť se zvyšuje skupina pacientek, které byly diagnostikovány v časném klinickém stádiu nádoru (77,8 % ve stádiu I a II), které se vyznačují poměrně velkou úspěšností léčby (Obrázek 4). Z tohoto pohledu je velice přínosná identifikace rizikových osob a jejich zařazení do preventivního sledovacího režimu, tak aby bylo případné nádorové onemocnění zachyceno již v časném stádiu. Identifikace rizikových pacientů je možná pouze u dědičné formy karcinomu prsu, která tvoří 5-10 % všech případů tohoto onemocnění (Mateju et al. 2010, Pohlreich et al. 2012).



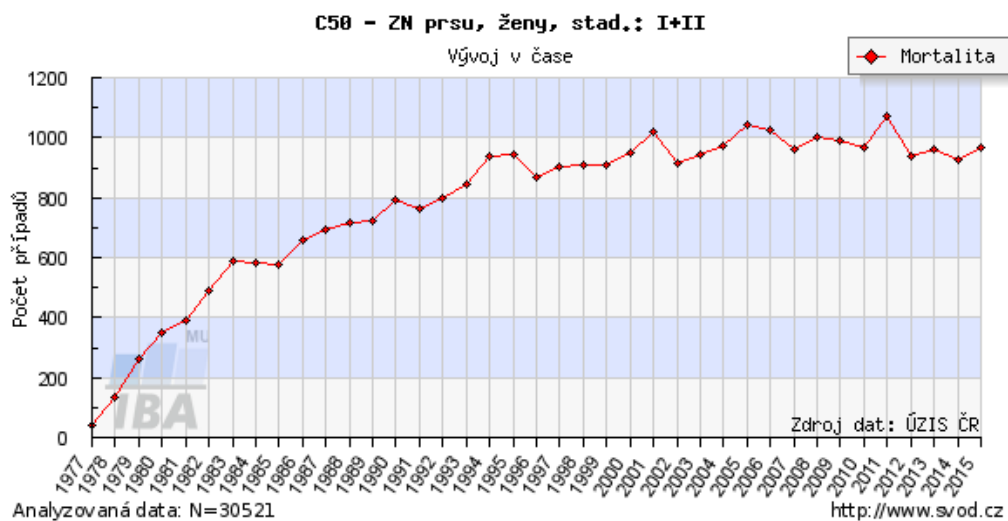
Obrázek 1: Časový vývoj hrubé incidence a hrubé mortality pro diagnózu karcinomu prsu v celé populaci České republiky. ÚZIS, <http://www.uzis.cz>



Obrázek 2: Věková struktura populace pacientek v době diagnózy karcinomu prsu. C50- kód diagnózy, zhoubný novotvar prsu; ZN – zhoubný nádor. ÚZIS, <http://www.uzis.cz>



Obrázek 3: Zastoupení klinických stadií v čase. Graf zobrazuje časový vývoj zastoupení klinických stadií. Klinická stadia jsou určována na základě TNM klasifikace (viz. kapitola 1.2) platné v době stanovení diagnózy. ÚZIS, <http://www.uzis.cz>



Obrázek 4: Mortalita pacientek s diagnostikovaným klinickým stádiem nádoru I, nebo II. Graf zobrazuje časový vývoj od r. 1977 do r. 2015. ÚZIS, <http://www.uzis.cz>

1.2 Patogeneze a klasifikace karcinomu prsu

Riziko vzniku onemocnění v průběhu života roste. Kromě pohlaví, věku a rodinné anamnézy riziko zvyšují některé reprodukční a hormonální faktory, jako časná menarche, pozdní menopauza, pozdní narození prvního dítěte, nuliparita, krátká doba kojení, užívání orální antikoncepce nebo hormonální terapie, které zvyšují vystavení mamární tkáně estrogenů (Barrett 2010). Estrogeny hrají důležitou roli ve vývoji a progresi karcinomu prsu. Přímé i nepřímé mechanismy podporují kontribuci estrogenů ke vzniku onemocnění. U některých metabolitů estrogenů byly navíc prokázány genotoxické vlastnosti. Dalšími rizikovými faktory pro karcinom prsu jsou postmenopauzální obezita, nedostatečná výživa, konzumace alkoholu/tabáku, nedostatek fyzické aktivity, socioekonomické postavení, ionizující záření a předchozí benigní onemocnění mamární tkáně (Downs-Holmes et al. 2011). Tumorigeneze karcinomu prsu je mnohaúrovňový proces, během kterého se normální epitel prsní tkáně vyvíjí přes hyperplázii a karcinom *in situ* do invazivního nádorového onemocnění. Přibližně 95% maligních nádorů prsu jsou karcinomy epitelu mamární tkáně (Berg et al. 1995).

Karcinom prsu vzniká především jako sporadické onemocnění, které je důsledkem kumulace somatických mutací v buňkách duktálního, nebo myoepiteliálního kompartmentu prsní žlázy. Příčinou vzniku těchto mutací jsou vnější (chemické mutageny, záření) i vnitřní vlivy (oxidativní stres, chyby replikace), které neustále poškozují DNA buněk zdravého organismu. Většina těchto poškození je díky reparačním mechanismům opravena, nicméně malá frakce může být zafixována jako mutace a přenášena na dceřiné buňky. Somatické mutace nádorového genomu můžeme dělit podle vlivu na karcinogenezi na „driver“ a „passenger“. Driver mutace propůjčují buňce proliferační výhodu a jsou pozitivně selektovány v rámci mikroevoluce, která probíhá mezi buněčnými populacemi v prekancerózní tkáni. Tyto mutace se vyskytují v tumorsupresorových genech, či protoonkogenech. Ostatní mutace, které nepřispívají k růstovému zvýhodnění daného buněčného klonu, ale vyskytují se ve všech buňkách klonu společně s driver mutacemi, jsou nazývány „passenger“ mutace (Stratton et al. 2009). Tato klasifikace je užitečná pro porozumění genetickým změnám v pokročilých stádiích vývoje nádoru, ale i pro prevenci a časnou detekci nádorového onemocnění, kdy je třeba sledovat vývoj buněk nesoucích takové potencionálně tumorogenní mutace a brát v úvahu i možné jiné vyústění jejich fenotypu než maturaci v maligní nádorovou buňku. Klon buňky zdravého jedince, který nese mutaci, může vyhynout, přetrvávat bez další expanze, vytvořit neinvazivní benigní tumor, nebo progredovat až do nádoru. Pro prevenci a včasnou detekci je důležité tyto různé dráhy rozlišovat a rozlišení pouze na drivers/passengers může být v tomto ohledu příliš zjednodušující. Někteří autoři proto navrhují zavedení i přívlasků zvýhodňujících/znevýhodňujících mutace ve smyslu schopnosti daného klonu růst či přežít v daném prostředí a dále přívlasků predisponujících/nepredisponujících.

Predisponující mutace umožňuje vývoj klonu směrem k plně nádorovému fenotypu buď získáním základních znaků nádoru (Hanahan et al. 2011), nebo destabilizací genomu a následnému hromadění dalších mutací (Kuhner et al. 2016).

Přibližně 5-10 % případů karcinomu prsu vzniká jako dědičné onemocnění na základě germinálních mutací v genech obecně predisponujících k nádorovým onemocněním, či charakteristickým pro vývoj karcinomu prsu, viz kapitola 2.

Histologicky jsou nejběžnějšími epitelovými nádory prsu infiltrující duktální (75 % všech nádorů prsu) a lobulární (15 %) karcinom. Mezi méně časté patří medulární (1-7 %), mucinózní (1-2 %), tubulární (1-2 %), papilární (>1 %) a inflamatorní karcinomy. Neepitelové nádory jsou velmi vzácné. Různé histologické typy mají různou prognózu. Například lobulární karcinom prsu má lepší prognózu, než duktální tumor. Mezi méně častými subtypy je tubulární karcinom dobře diferencovaný karcinom s velice dobrou prognózou, zatímco vzácné inflamatorní karcinomy mají prognózu špatnou (Downs-Holmes et al. 2011). Diferenciace nádorových buněk definuje histologický grading nádoru, který se stanovuje dle metodiky Blooma a Richardsona (Bloom et al. 1957), postupně modifikované Elstonem a Ellisem (Elston et al. 1991). Grading nádoru je založen na třech hlavních znacích: formace tubulů, jaderný pleomorfismus a počet mitóz. Postupně se stala stádia karcinomu prsu důležitým prognostickým faktorem, který je jedním z činitelů, rozhodujících o různých léčebných strategiích. Určení stádia nádoru je založeno na TNM klasifikaci, která popisuje velikost tumoru (T), zasažení lokálních lymfatických uzlin (N) a přítomnost vzdálených metastáz (M). Rutinní patologická klasifikace nádoru prsu také popisuje invazi vaskulárních, nebo lymfatických cév, infiltraci do kůže, míru exprese estrogenního receptoru (ER), progesteronového receptoru (PR) a HER2 (jinak také ERBB2, nebo NEU) receptoru. Pro charakterizaci nádoru se také využívá míra proliferační aktivity a exprese tumorsupresorového genu *TP53* (UICC).

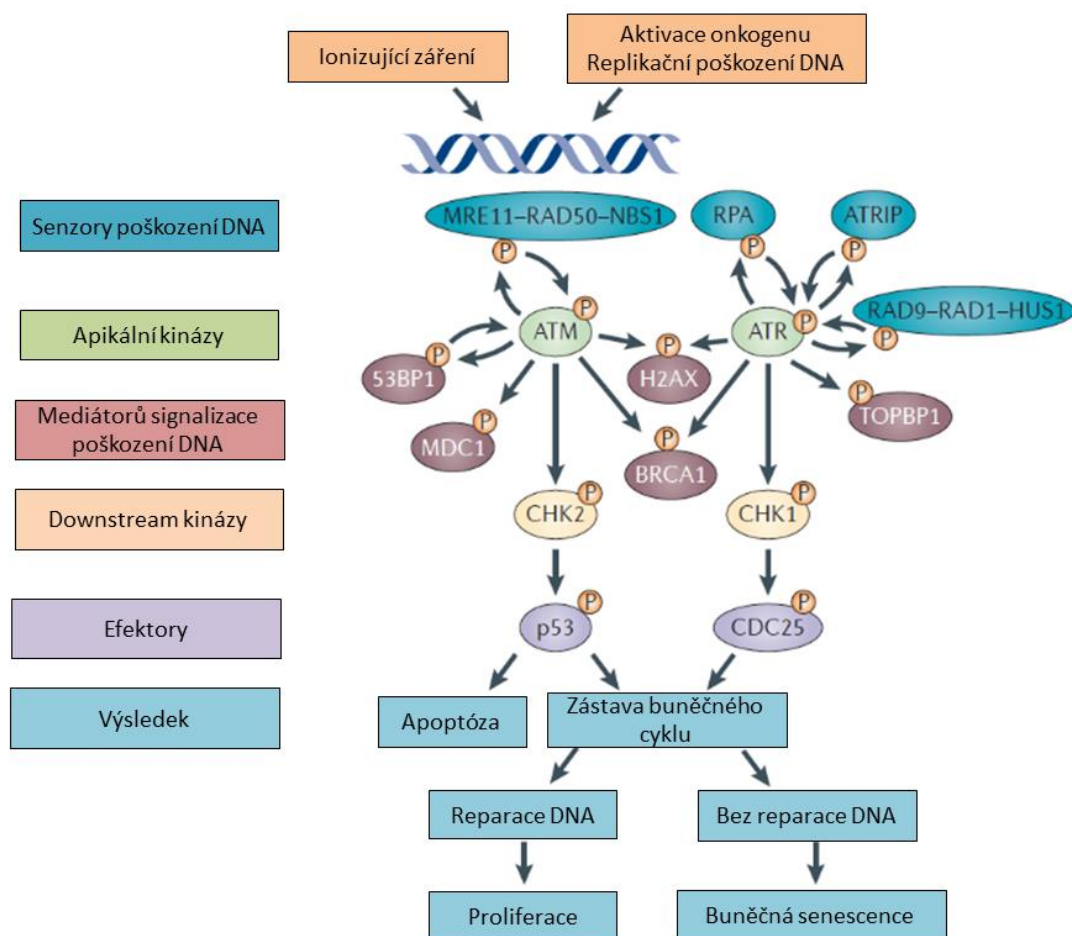
Nový klasifikační systém mamárních nádorů představuje molekulární profilování, které vychází z charakteristik genové exprese v nádorových buňkách, které byly získávány pomocí expresních mRNA čipů. V tomto systému jsou karcinomy prsu rozlišovány na několik molekulárních subtypů, které nejčastěji zahrnují: luminal A, luminal B, HER2+, basal-like a normal-like tumory. Jednotlivé subtypy se liší svým původem, biologickými vlastnostmi i klinickými charakteristikami (Polyak 2007). Basal-like nádory prsu, zahrnující tzv. triple negativní nádory (TNBC; neexprimující ER, PR, ani HER2 receptor) se vyznačují agresivním růstem s vysokou mírou proliferační aktivity, špatnou odpovědí na systémovou léčbu a nepříznivou prognózou. Tento subtyp tvoří 15 % všech karcinomů prsu. Většina nádorů prsu patří do skupiny luminal A (71 %), charakterizovaných ER i PR pozitivitou, nižší proliferační aktivitou a z toho vyplývající příznivější prognózou (Sorlie et al. 2001, Rouzier et al. 2005,

Smid et al. 2008, Hugh et al. 2009, Wiechmann et al. 2009). Přestože se incidence nádorů prsu zvyšuje, prognóza onemocnění se zlepšila. Za hlavní faktory tohoto zlepšení lze považovat časnou diagnózu a také zlepšení systémové léčby onemocnění. Pro léčbu nádorů prsu jsou standardně využívány kombinace chirurgické léčby, radioterapie, chemoterapie a hormonální léčby. Pětileté přežití se dnes pohybuje okolo 82% (Prausová 2010).

Nádory prsu se vzácně vyskytují u mužů, kde představují méně než 1 % zhoubných nádorů. Karcinom prsu je onemocněním starších mužů (po 50. roku života). Nejvýznamnějším rizikovým faktorem karcinomu prsu u muže jsou zárodečné mutace v genu *BRCA2* (Breast cancer type 2 susceptibility protein) a přibližně 13% mužů s nádorem prsu jsou nosiči *BRCA2* mutace. Karcinom prsu u mužů je v 80 % hormonálně závislý. Významným rizikovým faktorem je přísun estrogenů při poruše jaterních funkcí nebo terapii karcinomu prostaty. Příznaky, diagnostika i terapie jsou obdobné jako u karcinomu prsu u žen (Syrjakoski et al. 2004).

1.3 Mechanizmy oprav genomové DNA

Udržení genomové stability je zcela zásadní pro zachování buněčné integrity. Genomová DNA je přitom neustále vystavena poškozujícím vlivům. Každý den dochází k několika tisícům poškození v každé buňce lidského organismu, z nichž většina je způsobena normální metabolickou aktivitou buňky (např. v důsledku vzniku kyslíkových radikálů v průběhu buněčné respirace). Zdrojem závažných poškození genomové DNA v dělicích se buňkách jsou poruchy replikace DNA. Exogenní poškození DNA mohou být způsobena chemickými sloučeninami, ionizujícím nebo UV zářením (Jackson et al. 2009, Ciccia et al. 2010). Poškození DNA může způsobovat vznik mutací a chromozomálních aberací. Z tohoto důvodu se v buňkách vyvinul systém odpovědi na poškození DNA, tzv. DNA Damage Response (DDR). DDR zahrnuje komplexní signalizační síť aktivovanou poškozením DNA, která velice rychle ovlivňuje celou řadu fyziologických procesů (buněčný cyklus, replikace DNA, opravy DNA, senescence a apoptóza), aby bylo zamezeno přenosu poškození na dceřinou buňku (Ciccia et al. 2010). Při závažných poškozeních DNA (např. při vzniku dvouřetězcových zlomů v mitoticky aktivních buňkách) DDR zahrnuje několikastupňovou signální kaskádu, které se účastní desítky až stovky různých proteinů (Jackson et al. 2009).



Obrázek 5. Signální úrovně odpovědi na přítomnost poškození DNA. Iničiální fáze spočívá v aktivaci a translokaci senzorů poškození DNA, které umožňuje aktivaci apikálních kináz, zajišťujících přenos intracelulárního signálu a fosforylaci mediátorů signalizace poškození DNA, které katalyzují tvorbu rozsáhlých komplexů (reparačních ohnisek) pro opravy poškozené DNA. Odpověď na přítomnost závažného poškození DNA však v mitoticky aktivních buňkách vyžaduje harmonizovat procesy oprav genomu s další intracelulární signalizací pomocí efektorových proteinů DDR signalizace aktivovaných prostřednictvím fosforylace specifickými kinázami nebo prostřednictvím aktivace enzymových funkcí mediátorů a kovalentních modifikací (např. ubikvitinylace, ribosylace) jejich substrátů. Výsledkem je za fyziologických okolností reparace DNA a zástava buněčného cyklu. V případě nedostatečné reparace pak indukce senescence či přímo apoptózy. Převzato z (Sulli et al. 2012).

Chybně opravená poškození DNA mohou způsobit zánik buňky či přestavby genomu s významným onkogenním potenciálem. Proto musí být dráhy, které opravují a signalizují poškození DNA velmi citlivé, ale zároveň selektivní a reverzibilní. Integrace pozitivních a především negativních kontrolních bodů zahrnujících posttranslační modifikace, jako je fosforylace, ubikvitinylace a acetylace, je klíčová pro zajištění efektivních odpovědí na poškození DNA a zmírnění dopadů těchto poškození na integritu genomu (Panier et al. 2013).

Procesy oprav DNA reflektují jednotlivé druhy poškození DNA. Oprava dvouřetězcových zlomů (DNA Double Strand Break (DSB) repair) probíhá v lidských buňkách nejčastěji cestou nehomologního spojování konců (Non-Homologous End Joining, NHEJ). V dělicích se buňkách však tato poškození

mohou být reparována pomocí komplexního a vysoce přesného procesu homologní rekombinace (HR).

Specifické reparační mechanismy rozpoznávají a opravují jednořetězcové zlomy (Single Strand Breaks, SSBs) a kovalentní modifikace nukleotidů a jejich aduktů pomocí Nucleotide Excision Repair (NER) a Base Excision Repair (BER). Opravy chybně zařazených bazí při replikaci DNA jsou zajišťovány cestou Mismatch Repair (MMR).

Nejčastějšími poškozeními DNA jsou kovalentní modifikace nukleotidů (např. vznik O⁶-metylguanosinu), které jsou efektivně odstraňovány přímou enzymovou opravou (v uvedeném případě pomocí enzymu MGMT; Methylated-DNA-protein-cysteine methyltransferase). Defekty v DDR přispívají k procesům stárnutí a řady poruch včetně vývojových vad, neurodegenerativních onemocnění a nádorových onemocnění, což podtrhuje zásadní důležitost efektivního DDR pro buňku a viabilitu celého organismu (Hoeijmakers 2001, Jackson et al. 2009).

1.3.1 Oprava dvouřetězcových zlomů DNA

Dvouřetězcové zlomy DNA, při kterých dochází k přerušení fosfodeoxyribonukleotidové kostry obou vláken DNA, jsou nejzávažnější formou poškození DNA v mitoticky aktivních buňkách. Důsledkem nedostatečné či chybějící opravy těchto poškození je vznik chromozomových zlomů a translokací, které se zásadně podílejí na vzniku nádorových onemocnění. DSB vznikají jako důsledek radiačního záření či působením volných radikálů, ale jejich nejčastějším zdrojem jsou poruchy replikace DNA. (Jackson et al. 2009). V odpovědi na přítomnost takového poškození dochází v dělících se buňkách k zástavě buněčného cyklu v kontrolních bodech, nebo v průběhu replikace (G1/S a inter-S kontrolní body) a před rozdělením buňky (G2/M kontrolní bod), čímž předchází duplikaci a segregaci poškozené DNA.

Buňky využívají dvě hlavní cesty pro opravu DSB - homologní rekombinaci (HR) a nehomologní spojování konců (NHEJ). Obě dráhy se funkčně doplňují, avšak působí za různých okolností. Zásadní molekulární regulační krok, který rozhoduje mezi cestou NHEJ a HR, je proces resekce přerušovaných konců DNA za vzniku rozsáhlého úseku ssDNA, který je potřebný pro HR, nikoliv však pro NHEJ.

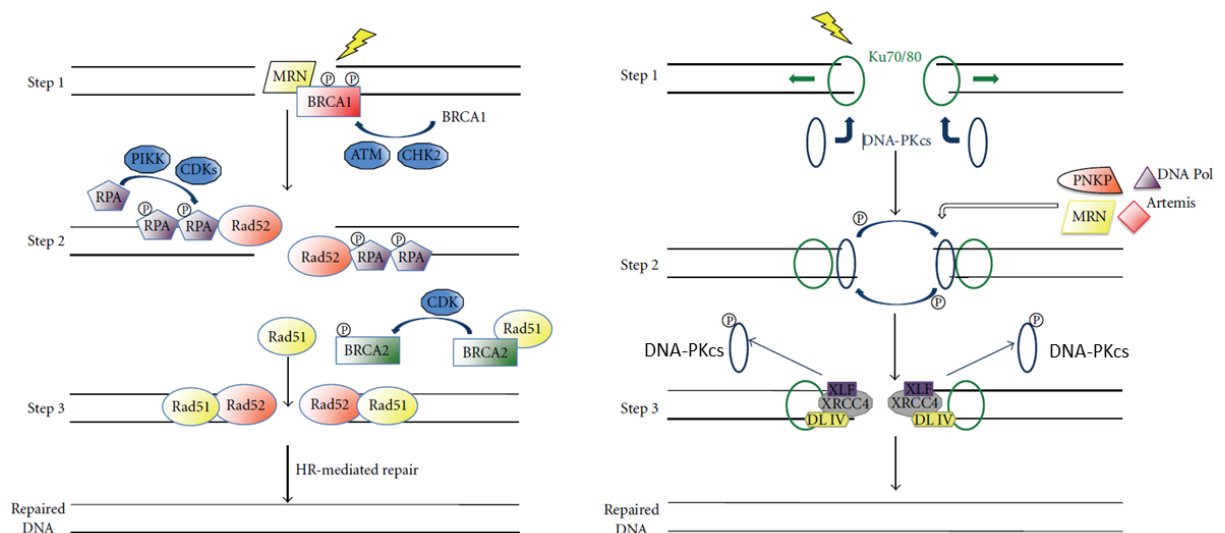
Proteiny zodpovědné za opravy DSB mohou být rozděleny na senzory poškození DNA (např. MRN komplex složený z proteinů Mre11, Rad50 a Nbs1; Ku70/Ku80 (Lupus Ku autoantigen protein p70/XRCC6 a Lupus Ku autoantigen protein p86); RPA (Replication protein A 70 kDa DNA-binding subunit)), přenašeče (apikální kinázy) odpovědi na poškození (např. DNA-PK (DNA-dependent protein

kinase), ATM (Ataxia telangiectasia mutated), ATR (Ataxia telangiectasia and Rad3-related protein)), downstream kinázy (CHEK2 (Checkpoint kinase 2), CHEK1 (Checkpoint kinase 1)) a efektory (např. p53, CDC25 (cell division cycle 25); Obrázek 5). Dvě PIKK kinázy (phosphatidylinositol-3 kinase-like kinase), ATM a ATR, jsou zásadními apikálními regulátory signalizace kontrolních bodů. Zatímco ATM kináza je aktivována primárně dvouřetězcovými zlomy, zablokování replikační vidlice a SSB aktivují ATR kinázu. Třetí zástupce PIKK rodiny, DNA-dependentní proteinkináza (DNA-PK), je také aktivována přítomností DSB v DNA. Oproti ATM a ATR je aktivita DNA-PK omezena pouze na NHEJ opravy.

Proteiny zodpovědné za aktivaci těchto transdukčních kináz a jejich lokalizaci do místa poškození DNA jsou sensorové proteiny – MRN komplex (složený z Mre11, Rad50 a Nbs1) pro ATM, Ku70/Ku80 heterodimer pro DNA-PK a ATRIP (ATR-Interacting Protein) s RPA (Replication Protein A) pro ATR. ATR je cílena do míst DSB pouze v S, nebo G2 fázi (Falck et al. 2005, Jazayeri et al. 2006, Cimprich et al. 2008). MRN komplex a Ku70/80 heterodimer jsou první komplexy, které rozeznávají DSB a spouští signalizaci kontrolního bodu a opravných mechanismů prostřednictvím aktivace PIKK kináz. Primárním cílem PIKK kináz je γ H2AX, varianta histonu H2AX, který je fosforylován na serinu 139. Přítomnost γ H2AX je považována za charakteristický znak odpovědi na přítomnost DSB. Fosforylace γ H2AX působí jako aktivační signál pro interakci s mediátorovým proteinem MDC1 (Mediator of DNA damage checkpoint protein 1) iniciujícím na fosforylaci závislý vznik MRN komplexu a aktivaci ATM. Tyto události působí jako pozitivní regulace a amplifikace odpovědi na poškození DNA DSB (Mirzoeva et al. 2001, Stucki et al. 2005). Fosforylovaný protein MDC1 umožňuje navázání E3 ubikvitin ligáz, jako RNF8 (Ring Finger Protein) a RNF168 do místa DSB, kde ubikvitinylují γ H2AX. Tím je následně katalyzována lokalizace dalších mediátorů, jako 53BP1 (Tumor suppressor p53-binding protein 1) a BRCA1 (Breast cancer 1) pro ATM a TopBP1 (DNA Topoisomerase II-Binding Protein 1) a CLSPN (Claspin) pro ATR, do místa poškození DNA. BRCA1-A komplex je lokalizován do místa DSB interakcí s γ H2AX a podílí se na zástavě buněčného cyklu v kontrolním bodě G2/M. Dalším fosforylačním cílem pro ATM po vzniku DSB je efektorová kináza CHEK2, která se podílí na zástavě buněčného cyklu v kontrolních bodech G1/S a G2/M inhibicí fosfatáz z rodiny Cdc25 a aktivací tumor supresorového transkripčního faktoru p53. ATR s pomocí TopBP1 a Claspinu aktivuje efektorovou kinázu CHEK1, která fosforyluje překrývající se substráty s CHEK2 (např. Cdc25A, Cdc25Cf). Obecně je předpokládáno, že ATM aktivuje CHEK2 a ATR primárně aktivuje CHEK1, nicméně tento koncept byl modifikován studii mapujícími řadu křížení, neboli „crosstalks“, mezi dráhami těchto kináz, především po zjištění, že CHEK1 je aktivován pomocí ATR i ATM (Dai et al. 2010).

1.3.1.1 Opravy DNA homologní rekombinací

HR vyžaduje přítomnost homologního templátu, sesterské chromatidy, která umožní přesnou opravu DSB v S a G2 fázi buněčného cyklu (San Filippo et al. 2008, Moynahan et al. 2010). BRCA1-C komplex umožňuje vznik jednořetězcové DNA v místě DSB tím, že v komplexu s endonukleázou CtIP (CtBP-interacting protein; RBBP8) stimuluje resekci DNA MRN komplexem (Chen et al. 2008). Resekce zahrnuje 5'-3' nukleolytické zpracování konců DNA prostřednictvím MRN komplexu za přispění CtIP, helicáz z rodiny RECQ a nukleáz Exol a Dna2. Takto vzniklé jednořetězcové molekuly DNA jsou rychle obaleny ssDNA vazebným komplexem RPA (Replication protein A), který je následně nahrazen proteinem RAD51 (DNA repair protein RAD51 homolog 1) s pomocí dalších RAD51 paralogů (RAD51C,D) a dalších proteinů reparačního ohniska, které mimo jiné zahrnuje produkty genů ze skupiny Fanconiho Anemie (FA), jako je FANCD1/BRCA2 a FANCN/PALB2. Protein BRCA2 zastává důležitou roli v reparaci pomocí HR, jelikož katalyzuje vytvoření nukleofilament z RAD51 zprostředkovávajících vyhledání homologní sekvence v sesterské chromatidě. Následně homologní řetězec invaduje do struktury templátové DNA. Protein BRCA2 vyžaduje lokalizaci PALB2 (FANCN) v místě poškození k tomu, aby mohl stabilizovat intermediáty HR a tím umožnit vlastní rekombinaci. Po invazi homologního řetězce a zahájení aktivity DNA polymeráz a ligaci konců DNA pomocí Ligázy I, oddělí DNA helicáza intermediáty HR za uvolnění intaktních opravených vláken DNA (Huertas et al. 2008, Bernstein et al. 2009, Mimitou et al. 2009, Rupnik et al. 2010, You et al. 2010, Polo et al. 2011).



Obrázek 6: Zjednodušený model reparace DNA pomocí HR (vlevo) nebo NHEJ (vpravo). Výběr reparační dráhy závisí na uspořádání úvodních komplexů reparace DSB. Interakcí Ku70/Ku80 heterodimeru s místem přerušeni obou DNA řetězců je aktivováno NHEJ, vznikem a aktivací MRN komplexu v místě přerušeni DNA HR. V modelu NHEJ je DLIV názvem pro DNA ligázu IV. V modelu HR, po nasazení RAD51 prostřednictvím BRCA2, vytváří RAD1 nukleoproteinové vlákno, které invaduje homologní sekvenci a aktivuje výměnu vláken sesterské chromatidy. Převzato z: (Summers et al. 2011).

1.3.1.2 Opravy DNA prostřednictvím NHEJ

Oproti opravám homologní rekombinací je reparace DNA pomocí NHEJ aktivní v průběhu celého buněčného cyklu a je jedinou volbou oprav DSB v DNA v průběhu G₀, G₁ a časně S-fáze. Přestože je NHEJ vysoce efektivní, jeho nepřesnost, vycházející z úprav DNA řetězců v místě DSB, jej činí potenciálně náchylným k tvorbě mutací (Lieber et al. 2010)..

NHEJ je nejvyžívanější opravný mechanismus DSB zprostředkovávající přímou ligaci poškozených konců DNA a obvykle vyžaduje minimální úpravu konců DNA (Obrázek 6). Během NHEJ jsou konce DNA nejprve rozeznány a přiblíženy k sobě pomocí Ku70/Ku80 heterodimeru (kódovaného geny *XRCC6* a *XRCC5*) sloužícího pro vazbu a aktivaci holoenzymu DNA-dependentní proteinkinázy (DNA-PK)(Polo et al. 2011, Panier et al. 2013), obsahujícího katalytickou podjednotku DNA-PK (DNA-PKcs), která vůči sobě nastaví konce poškozené DNA (Summers et al. 2011). Před tím, než jsou tyto proti sobě postavené konce DNA zaligovány pomocí komplexu XLF-XRCC4-Ligáza IV, jsou upravovány řadou faktorů, jako je nukleáza Artemis (DCLRE1C), fosfatáza polynukleotidkinázy 3' (PNKP), Aprataxin a APLF (Aprataxin and PNK-like factor).

Alternativou kanonické dráhy NHEJ je alternativní NHEJ. Tyto dráhy se obvykle spoléhají na koncové mikrohomologní sekvence pro spojování a zahrnují některé faktory působící v HR a SSB opravách, jako je MRN komplex (Polo et al. 2011).

1.3.1.3 Reparace kovalentního spojení řetězců DNA

Proteiny rozsáhlé rodiny genů Fanconiho anémie se podílejí na udržování genomové stability reparací kovalentních meziřetězcových spojů (zesíťování DNA tzv. interstrand crosslinks; ICLs).(Alpi et al. 2008)). ICL jsou pro genom velmi závažná poškození, která mohou vznikat při normálním metabolismu buňky, nebo při aplikaci protinádorové terapie (léčba deriváty platiny). ICL brání separaci komplementárních vláken kovalentním spojením obou antiparalelních řetězců dvoušroubovice DNA. Proto ICL inhibují kritické buněčné procesy, jako je transkripce a replikace. Důsledkem ICL může být tvorba dvouřetězcových zlomů v DNA (Bhattacharjee et al. 2017).

Reparace ICL se účastní řada opravných mechanismů, jako je dále popsáný nucleotide excision repair (NER), homologní rekombinace (HR) a translesion synthesis (TLS) (Nojima et al. 2005) (McHugh et al. 2001). Zatímco u neproliferujících buněk hraje základní úlohu v rozpoznání a odstranění ICL mechanismus NER (McHugh et al. 2001), u buněk procházejících duplikací genomu (S-fáze) funguje replikační aparát jako senzor ICL. V posledních dvou desetiletích byla popsána komplexní reparační

dráha zahrnující skupinu proteinů kódovaných geny Fanconio Anemie (FA; OMIM 227650). Tyto proteiny detekují a opravují ICL v průběhu replikace prostřednictvím aktivace a úzké kooperace s reparačními mechanizmy homologní rekombinace (Bhattacharjee et al. 2017). (Nojima et al. 2005)

Reparační dráha Fanconio anemie je tvořena jadernou multiproteinovou sítí, v současnosti čítající 22 komplementačních skupin (A až W) a další asociované proteiny (Che et al. 2018). Dvacet genů skupiny je lokalizováno autosomálně, pouze *FANCB* leží na chromosomu X (Meetei et al. 2004).

Proteiny této dráhy byly rozděleny do tří skupin, podle fáze procesu opravy ICL, které se účastní (Wang 2007). Skupina I, neboli takzvaný "FA nuclear core complex", je tvořena proteiny *FANCA*, *FANCB*, *FANCC*, *FANCE*, *FANCF*, *FANCG*, *FANCL*, *FANCM* a *FANCT* (*UBET2*) spolu s dalšími pěti proteiny, které s core komplexem interagují (*FAAP100*, *FAAP24*, *FAAP20*, *MHF1* a *MHF2*). Do skupiny II, která je také označována jako ID2 komplex patří *FANCD2* a *FANCI*. Skupina III zahrnuje DNA reparační faktory, včetně proteinů HR dráhy: *BRCA2* (*FANCD1*), *BRIP1* (*FANCI*), *PALB2* (*FANCI*), *RAD51C* (*FANCI*), *RAD51* (*FANCI*), *SLX4* (*FANCP*), *BRCA1* (*FANCI*), *XRCC2* (*FANCI*), *REV7* (*FANCI*) a DNA endonukleázu *XPF* (*FANCI*) (Ceccaldi et al. 2016). Core komplex je zodpovědný za monoubikvitinylaci dimeru *FANCD2-FANCI* v reakci na poškození DNA, která je zásadním aktivačním krokem v procesu reparace ICL (Wang 2007). Huang a kolektiv identifikovali tři funkční moduly tohoto komplexu: katalytický modul (*FANCB*, *FANCL*, *FAAP100*) zajišťující E3 ligázovou aktivitu prostřednictvím RING domény proteinu *FANCL*, dále *FANCA-FANCG-FAAP20* modul a *FANCC-FANCE-FANCF* modul, které umožňují lokalizaci core komplexu do místa poškození DNA (Huang et al. 2014), a to především díky interakci s komplexem *FANCM/FAAP24* (Deans et al. 2009). Ubikvitinylovaný *FANCD2* (*FANCD2-Ub*) přivádí do místa poškození podpůrný protein pro reparační nukleázy – *SLX4* (*FANCP*) interakcí prostřednictvím ubikvitin rozeznávajícího motivu *UBZ4*. *SLX4* (*FANCP*) působí jako molekulární platforma pro koordinování a aktivaci specifických endonukleáz rozlišujících ICL strukturu, jako jsou *XPF-ERCC1*, *MUS81-EME1* a *SLX1*, které rozpojí kovalentně spojená vlákna DNA (Bhattacharjee et al. 2017). Následně jsou do místa poškození navázány polymerázy *REV7* (*FANCI*), polymeráza ζ a polymeráza η , které doplňují jednořetězcové mezery, které vznikly při rozpojení ICL. Tyto polymerázy patří do skupiny tzv. translézních polymeráz, které se vyznačují větší vazebnou kapsou pro DNA ve srovnání s replikačními polymerázami a dovolují tak procesovat větší DNA adukty, které vznikají při ICL, či v případě thyminových dimerů (de Groote et al. 2011). Kromě jednořetězcové mezery na jednom vlákně vzniká při rozpojení i dvouřetězcový zlom, jehož oprava je závislá na homologní rekombinaci. Této skutečnosti odpovídá i hypersenzitivita buněk deficientní na HR k DNA zesilujícím látkám působícím vznik ICL (např. mitomycin C). FA proteiny účastníci se HR nejsou potřebné pro *FANCD2-I* monoubikvitinylaci, z čehož vyplývá, že k jejich aktivaci dochází až po ubikvitinylaci *FANCD2-I* komplexu. Konečné vyštěpení (resekce) části DNA je klíčový krok v opravě

dvouřetězcového zlomu, který iniciuje HR. FANCD2 a BRCA1 (FANCS) zahajují navázání resekcčního faktoru CtIP do místa dvouřetězcového zlomu k zahájení HR. V této chvíli zajišťuje FANCC inhibici opravy cestou NHEJ tak, aby proběhla HR (Bunting et al. 2010).

1.3.1.4 Význam vrozených poruch repaře dvouřetězcových zlomů

Heterozygotní dědičné mutace v genech účastnících se oprav dvouřetězcových zlomů DNA (naprosto převážně v genech kódujících proteiny homologní rekombinace) způsobují často autosomálně dědičnou predispozici k některým nádorovým onemocněním, jako například karcinom prsu a ovaria (*BRCA1*, *BRCA2*, *RAD51C*, *RAD51D*, *PALB2/FANCN*, *CHEK2*, *ATM*, *NBN*).

Homozygotní mutace v těchto genech pak způsobují vzácná autosomálně recesivní, závažná syndromická onemocnění, která jsou obvykle doprovázena zvýšenou predispozicí k nádorovým onemocněním, jako například ataxia telangiectasia (OMIM 208900, postižení nervového a imunitního systému, leukémie, lymfomy; gen *ATM*) nebo Nijmegen breakage syndrom (OMIM 251260, postižení imunitního systému, opožděný růst, karcinom prsu, ovaria, prostaty; gen *NBN*).

Vrozené poruchy obou alel genů ze skupiny Fanconiho anemie způsobují vznik vzácných (1 na 1 000 000 narozených), autosomálně recesivních (nebo X-vázaných) syndromů charakterizovaný selháním kostní dřeně s k progresivní aplastickou anemií, genomovou nestabilitou, chromosomální fragilitou a zvýšenou predispozicí ke vzniku nádorových onemocnění, především k akutní myeloidní leukemii (Hu et al.). Řada pacientů FA také trpí vrozenými vývojovými vadami, jako jsou orgánové malformace, hypogonadismus, abnormální pigmentace kůže, či nízký vzrůst. Často je u těchto pacientů diagnostikováno nádorové onemocnění v ranném věku (spinocelulární karcinom hlavy a krku, karcinom děložního hrdla a nádory jater).

Heterozygotní vrozené mutace v FA genech jsou popisovány jako predisponující k familiárním nádorům prsu, ovaria, prostaty, pankreatu, či žaludku ((Friedenson, 2005 #643), viz též kapitola 2).

Vztah poruch FA genů k tumorigenezi je velmi komplexní. Zatímco predispozice k nádorovým onemocněním, především pak k AML (akutní myeloidní leukemie) a spinocelulárnímu karcinomu je společná pro všechny pacienty nesoucí homozygotní mutaci některého z FA genů, predispozice ke karcinomu prsu a ovaria na základě heterozygotní mutace byla popsána pouze u genů skupiny III – tedy genů v pozdní fázi dráhy zprostředkující reparaci pomocí HR (6 z 10 genů skupiny III: *FANCD1/BRCA2*, *FANCI/BRIP1*, *FANCN/PALB2*, *FANCO/RAD51C*, *RAD51 (FANCR)*, *BRCA1 (FANCS)*(Kottemann et al. 2013). U FA pacientů způsobují mutace v genech pozdní fáze dráhy

závažnější fenotypy s vyšší penetrancí a dřívějším nástupem nádorů. Byly zaznamenány případy pacientů s mutacemi *BRCA2* a *PALB2* s embryonálními tumory a tumory v extrémně nízkém věku (pod 2 roky) (Kottemann et al. 2013). Korelace mezi geny pozdní fáze a tumorigenezí je intuitivní, uvědomíme-li si mechaniku jednotlivých částí FA dráhy: zatímco geny časných fází core a ID komplexů jsou více samostatně fungující a spíše koordinující aktivaci dráhy, geny pozdní fáze jsou úzce spjaty s opravami homologní rekombinací, které jsou hojněji a nezávisle využívány mimo FA dráhu (Kottemann et al. 2013).

1.3.2 Reparační mechanizmy dalších poškození genomové DNA

Vznik dvouřetězcových zlomů je mimořádně závažným postižením DNA, avšak jeho důležitost s ohledem na vysoký onkogenní potenciál, je omezena na populace mitoticky aktivních buněk. Poškození jednotlivých nukleotidů, případně modifikace jejich bazí jsou řádově častějšími alteracemi DNA, které vyžadují účelnou opravu prakticky ve všech jaderných buňkách organismu. Selhání jejich reparačních procesů je rovněž zdrojem poškození DNA vedoucích k nádorové transformaci.

Base excision repair (BER) je mechanismus, který odstraňuje kovalentní modifikace bazí ve struktuře DNA, jako jsou oxidované báze, deaminace, alkylace, abazická místa (AP) a jednořetězcové zlomy. Úvodní krok BER zahrnuje rozeznání poškozené báze a její odstranění pomocí DNA glykosylázy (např. OGG1 - 8-oxoguanin DNA glycosylase 1). V případě využití monofunkční glykosylázy je následně rozštěpena fosfodiesterová vazba v AP místě pomocí AP-nukleázy (např. APE1 – apurinic/apyrimidinic endonuclease 1). Vzniklá mezera s nově vzniklým 3'-OH a 5'-deoxyribosylfosfátovým koncem je vyplněna dosyntetizováním přerušenoého řetězce na základě komplementarity bazí DNA polymerázou (nejčastěji POL β - polymerase β) (Ide et al. 2004). Fosfodiesterová vazba v místě přerušenoení je nakonec obnovena prostřednictvím ligázy LIG3 a XRCC1 (DNA ligase 3, X-ray cross complementing 1).

Bifunkční glykosylázy mají navíc AP-lyázovou aktivitu štěpící abazické místo DNA za vzniku jednořetězcového zlomu, který na svém 3'-konci obsahuje buďto fosfátovou skupinu (β , δ -eliminace), nebo α,β -nenasycený aldehyd (PUA, polyunsaturated aldehyde). 3'-PUA i 3'-fosfát jsou dále odstraněny fosfodiesterázovou aktivitou APE1 a fosfatázovou aktivitou polynukleotid kinázy (PNK) za vzniku přerušenoého řetězce DNA (podobně jako v případě monofunkční glykosylázy). Jednonukleotidová mezera je vyplněna Pol β a zlom je zacelen pomocí DNA ligázy LIG3/XRCC1 jako v případě monofunkčních glykosiláz (Obrázek 7).

BER se účastní i řada dalších enzymů společných s jinými reparačními drahami jako je PARP1 (poly (ADP-ribosyl) polymeráza 1) katalyzující syntézu adenosindifosfátribózylového polymeru a jeho

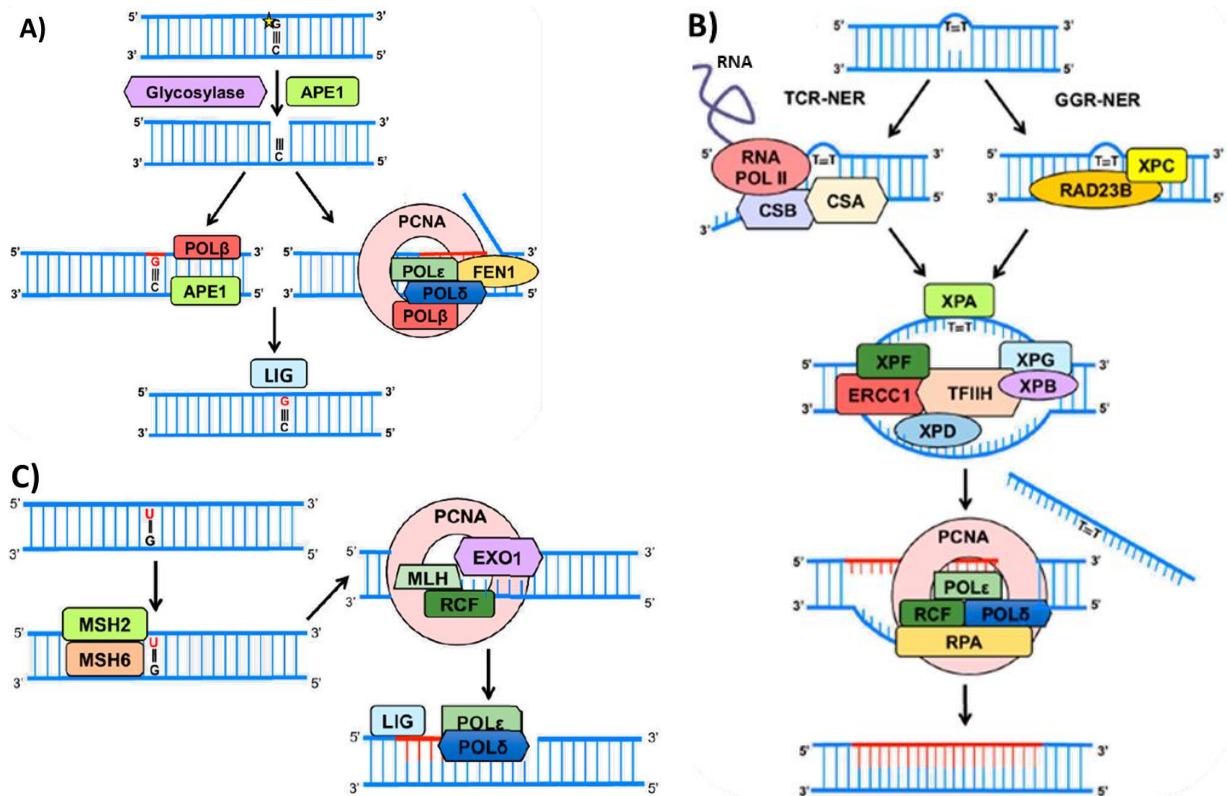
kovalentní navázání na řadu proteinů nukleoproteinového komplexu a proteinů podílejících se na reparaci DNA, včetně ligázy III, LIG1 (ligase 1), PCNA (proliferating cell nuclear antigen) a FEN1 (flap endonuclease 1) (Gomes et al. 2017). Inhibitory PARP1 (PARPi) jsou využívány při léčbě nádorů vykazujících poruchy v reparační dráze HR. Principem použití PARPi při indukci tzv. syntetické letality je inhibice komplementárních drah reparace zlomů DNA – BER (pomocí PARPi) a HR (v nádorové tkáni v případě inaktivace obou alel některého z kritických genů – např. *BRCA1* nebo *BRCA2* v důsledku kombinace vrozené a získané inaktivace/mutace). Inhibice BER pomocí PARPi způsobuje progresi jednořetězcových zlomů do dvouřetězcových, jejichž oprava je v nádorových (ale nikoliv somatických) buňkách blokována. Porucha reparace dvouřetězcových zlomů DNA způsobuje zvýšení genomové nestability v nádorových buňkách, které není slučitelné s jejich přežíváním (Rouleau et al. 2010).

Nucleotide excision repair (NER) je mechanismem opravujícím poškození DNA, která působí konformační deformaci struktury dvoušroubovice blokující postup replikační vidlice a transkribujících RNA polymeráz. Taková poškození vznikají například působením UV záření za vzniku pyrimidinových dimerů a 6-4 pyrimidin-pyrimidonu (Menck et al. 2014). Oprava pomocí NER počíná rozpoznáním poškození a následným rozvolněním struktury dvoušroubovice. Přerušení řetězce na obou stranách od místa kovalentně spojených bazí umožní jejich vyštěpení v podobě oligonukleotidové sekvence. Vzniklá mezera je vyplněna dosyntetizováním komplementárního řetězce s následnou ligací obnovující integritu řetězce DNA.

NER lze rozdělit na dvě odlišné dráhy v závislosti na místě vzniku poškození: tzv. global genome repair (GGR; globální opravy genomu) a transcription-coupled repair (TCR; opravy spřažené s transkripcí) (Gomes et al. 2017).

GGR odstraňuje poškození napříč genomem komplexem XPC-RAD23B (xeroderma pigmentosum complementation group C—human homolog B of *S. cerevisiae* RAD23). Tento komplex se váže na komplementární řetězec řetězci poškozenému a zprostředkovává vazbu komplexu transkripčního faktoru IIIH (TFIIH), který rozplétá dvoušroubovici DNA v délce přibližně 30 nukleoidů. TFIIH komplex je složen z deseti podjednotek. Sedm z nich tvoří tzv. core komplex: XPD, XPB, p62, p52, p44, p34 a TTDA. Cyklin aktivační kinázový subkomplex je tvořen cyklinem H, CDK7 a MAT1 proteiny a je propojen s core komplexem prostřednictvím XPD proteinu. Dvě podjednotky: ERCC2/XPD a ERCC3/XPB mají helikázovou aktivitu a rozplétající dvoušroubovici reparované DNA. Rozvolnění úseku DNA umožňuje nasednutí faktoru XPA (xeroderma pigmentosum complementation group A) a RPA komplexu. Endonukleázy XPF/ERCC1 (xeroderma pigmentosum complementation group F/excision repair, cross-complementation group 1) a XPG (xeroderma pigmentosum complementation group G) vyštěpí 24-30 nukleotidů, které obsahují poškozené místo. DNA polymerázy δ a ϵ asociované s RFC

(replication factor C) a PCNA následně vyplní vzniklou mezeru. Ligázy LIG1 nebo LIG3 nakonec zacelí přerušení nově dosyntetizovaného oligonukleotidového řetězce (Obrázek 7).



Obrázek 7. Schematické znázornění vybraných protein účastnících se drah reparace jednořetězcových poškození DNA. A) Base excision repair (BER); B) Nucleotide excision repair (NER); Mismatch repair (MMR). Upraveno podle (Gomes et al. 2017).

TCR je využíván u transkripčně aktivních genů, kdy poškození DNA zablokuje postup RNA polymerázy II a je rozeznáváno jak vlastní RNA polymerázou, tak i proteiny CSA (Cockayne syndrome A, jinak také ERCC8) a CSB (Cockayne syndrome B, neboli ERCC6). Tyto faktory zprostředkují opět vazbu TFIIH a další průběh reparace je společný s GGR (Shafirovich et al. 2017).

Mismatch repair (MMR) je reparačním mechanismem, který je zodpovědný za detekci i opravu nepárujících bází a inserčních/delečních smyček (insertion/deletion loops, IDL), které vznikají při replikaci, nebo homologní rekombinaci. Celá reparační dráha je zahájena evolučně vysoce konzervovanými proteiny MSH2 a MSH6 (MutS homolog 2 a 6), které spolu tvoří heterodimer MutSα a rozeznávají nepárující báze a jedno-, nebo dvoubázové IDL. V případě IDL delších, než 2 báze je za detekci zodpovědný MSH2 spolu s MSH3 (MutSβ) (Larrea et al. 2010). Po rozeznání poškození nastupuje jeden z heterodimerů MutLα (MLH1/PMS2), MutLβ (MLH1/MLH3), MutLγ (MLH1/MLH2),

které po aktivaci PCNA a replikačním faktorem C svojí endonukleázovou aktivitou (kterou zajišťuje PMS2 a MLH3) vytvoří jednořetězcové přerušení struktury DNA vedle poškození (na 3' stranu od daného nukleotidu). RPA faktor vázající jednořetězcovou DNA následně rekrutuje polymerázu δ (Pol δ), nebo Exonukleázu 1 (Exo1), které dosyntetizují opravený polynukleotidový řetězec. Jednořetězcové přerušení je na závěr spojeno ligázou 1 (Lig1; Obrázek 7).

1.3.2.1 Vrozené poruchy BER a NER

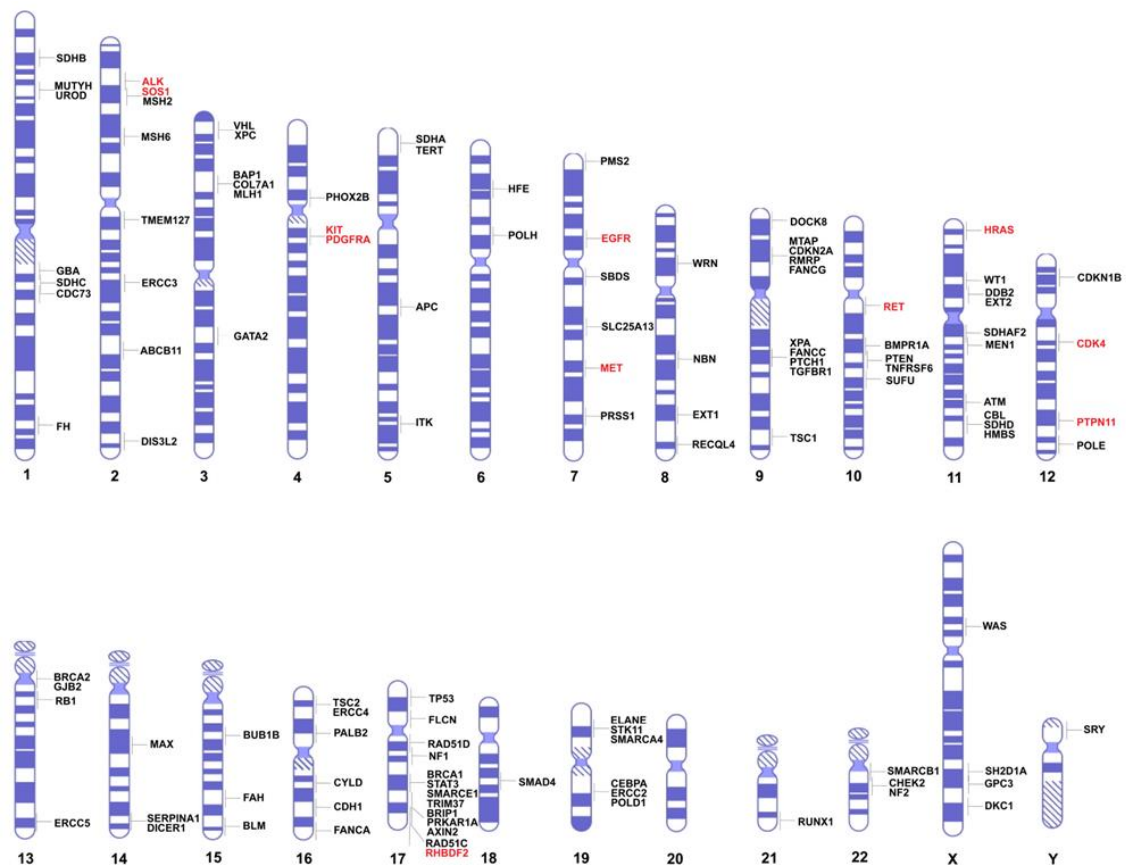
Dědičné mutace v genech účastnících se NER oprav způsobují autosomálně recesivní onemocnění xeroderma pigmentosum (XP, OMIM 278730; geny *XPA*, *XPC*, *XRCC2/XPD*, *XRCC3*, *ERCC4*, *ERCC5/XPG*, *XPE*), které se vyznačuje extrémní citlivostí na UV záření, časným vznikem kožních nádorů a neurologickými poruchami (Lehmann et al. 2014). Poškození genů účastnících se převážně TCR oprav způsobuje další závažná onemocnění, jako je Cockaynův syndrom (CS, OMIM 216400; geny *ERCC6* a *ERCC8*) způsobující ichtyosu, mentální retardaci a u některých jedinců fotosenzitivitu a trichothiodystrofií (TTD, OMIM 601675; geny *ERCC2*, *ERCC3*), onemocnění charakterizované mikrocefalií s fenotypem podobným progerii a mentální retardací. Dalším takovým onemocněním je Cerebro-oculo-facio-skeletální syndrom 2 (COFS2, OMIM 610756; geny *ERCC1*, *ERCC2*, *ERCC5*, *ERCC6*) způsobující degenerativní změny mozku, očí a míchy s častým snížením intelektu, těžkou svalovou hypotoníí a charakteristickými bezděčnými pohyby očí.

Byla identifikována řada vrozených mutací v genech *MSH2*, *MSH6*, *MLH1* i *PMS2* a stejně tak i epigenetické umlčení exprese *MSH2*, nebo *MLH1*, které způsobují defektní reparaci cestou MMR, která vede ke kumulaci spontánních genomických mutací a tumorogenezi. Zatímco studium sporadických alterací ovlivňujících geny MMR dráhy je teprve v počátcích, patogenní varianty predisponující k dědičným nádorovým onemocněním, jako je především hereditární nepolypózní kolorektální karcinom (HNPCC), či karcinom endometria jsou intenzivně studovány již delší dobu (Pena-Diaz et al. 2016). Existuje i řada studií vyšetřujících možnou predispozici mutací v MMR genech i k dalším nádorovým onemocněním, jako je např. karcinom ovaria (Pal et al. 2012), či popisujících vrozené vady a rozvoj nádorových onemocnění v dětském věku u nosičů homozygotních mutací (Baas et al. 2013).

2 GENY PREDISPOJUJÍCÍ KE VZNIKU DĚDIČNÝCH FOREM KARCINOMU PRSU

Jako nádorové predispoziční geny jsou označovány geny, jejichž vrozené patogenní mutace významně zvyšují riziko vzniku nádorového onemocnění. Míra tohoto rizika není pevně stanovena. Podle některých autorů je to větší relativní riziko, nežli 2x, přičemž alespoň u 5 % nosičů patogenních mutací dojde k rozvoji nádorového onemocnění (Rahman 2014). Pro většinu predispozičních genů ale platí výrazně vyšší míra penetrance, než je tato minimální hranice (viz kapitoly 2.1, 2.2).

Během posledních třiceti let byly pomocí řady různých přístupů identifikovány stovky nádorových predispozičních genů (nejvýznamnější zobrazeny na Obrázku 8). Predispoziční geny pro karcinom prsu byly identifikovány zpočátku pomocí genomových vazebných analýz a pozičního klonování, s příchodem nových metod pomocí NGS (panelového, exomového a celogenomového sekvenování) a celogenomovými asociačními studiemi (Genome Wide Association Studies, GWAS) (Turnbull et al. 2008).

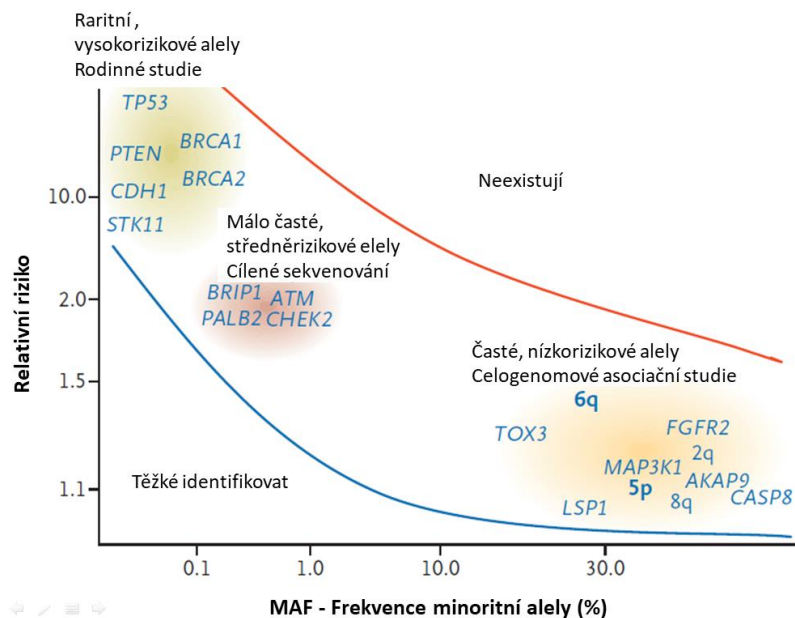


Obrázek 8. Chromozomální lokalizace hlavních nádorově predispozičních genů. Geny, u nichž způsobuje predispozici ztráta funkce, jsou vyobrazeny černě. Geny, u nichž způsobuje nádorovou predispozici získání funkce, jsou vyobrazeny červeně. Převzato z (Rahman 2014)

Naprostá většina predispozičních genů jsou tumorsupresory kódující proteiny účastníci se základních procesů ovlivňujících tkáňovou homeostázu, jako je reparace DNA, regulace buněčného cyklu nebo apoptózy. Dědičnost nádorové predispozice může být různá: u velké části těchto genů je autosomálně dominantní, ale pro řadu z nich můžeme zaznamenat autozomálně recesivní dědičnost pro jiné závažné nenádorové onemocnění mnohdy s výskytem nádorů v raném věku a zároveň autosomálně dominantní pro predispozici k nádorovému onemocnění v pozdějším věku života.

Jak již bylo uvedeno (Kap. 1.2), karcinom prsu vzniká především jako sporadické onemocnění. Přibližně 5-10 % případů karcinomu prsu je však způsobeno dědičnou predispozicí (Parkin 2004, Pohlreich et al. 2012). Podle některých studií, jako třeba projekt zpracovaný Skandinávským registrem dvojčat (Lichtenstein et al. 2000), se mohou genetické faktory podílet až na 27 % případech nádoru prsu (0,04-0,41 CI). Příbuzní prvního stupně pacientů s nádory prsu mají přibližně dvojnásobné riziko vzniku onemocnění ve srovnání s obecnou populací a pozitivní rodinná anamnéza je považována za jeden z nejsilnějších predispozičních faktorů korelujícím s počtem postižených příbuzných (Thompson et al. 2004).

Geny predisponující ke karcinomu prsu jsou rozdělovány do tří kategorií: geny s vysokou penetrancí, ale s nízkou frekvencí vrozených mutací; geny se střední penetrancí, jejichž alterace se vyskytují častěji; geny s nízkou penetrancí, jejichž poruchy jsou populačně časté (Foulkes 2008)(Obrázek 9).



Obrázek 9. Geny a lokusy predisponující ke karcinomu prsu. Znárodně predispoziční geny se vyskytují pouze mezi červenou a modrou křivkou. Vysokopropenetrantní syndromické geny jsou zvýrazněny zeleně. Celoživotní riziko vzniku karcinomu prsu může být u nosičů mutací v těchto genech 30-80 % a relativní riziko více než desetinásobné proti populaci bez mutace v těchto genech. Relativní riziko u středně propenetrantních genů (zvýrazněny červeně) se pohybuje okolo hodnoty 2. Tato skupina se s testováním osob z postižených rodin pomocí NGS sekvenování (viz kapitola 3) rozšiřuje. Časté, nízkopropenetrantní geny jsou znázorněny oranžově. Převzato z (Foulkes 2008).

Tato klasifikace byla učiněna s ohledem na míru rizika onemocnění karcinomem prsu (penetrance popisuje pravděpodobnost vzniku onemocnění u nosiče příčinné mutace) a prevalenci patogenních mutací v obecné populaci. Podle klasifikace navržené Strattonem a Rahmannovou (Stratton et al. 2008) způsobují vysokopenetrantní geny až 10x zvýšené relativní riziko vzniku nádoru prsu, avšak jejich mutační frekvence je v populaci obvykle velice nízká. Geny se střední penetrancí zvyšují riziko 2-4x a frekvence jejich mutací je nižší. Geny s nízkou penetrancí zvyšují riziko onemocnění přibližně 1,5x i méně a frekvence jejich mutací je v obecné populaci vysoká. Mutace ve všech doposud identifikovaných predispozičních genech jsou zodpovědné za pouze přibližně 25 % familiárních nádorů prsu (Turnbull et al. 2008, Ellsworth et al. 2010). Genetická predispozice tří čtvrtin vysoce rizikových rodin tak zůstává bez vysvětlení.

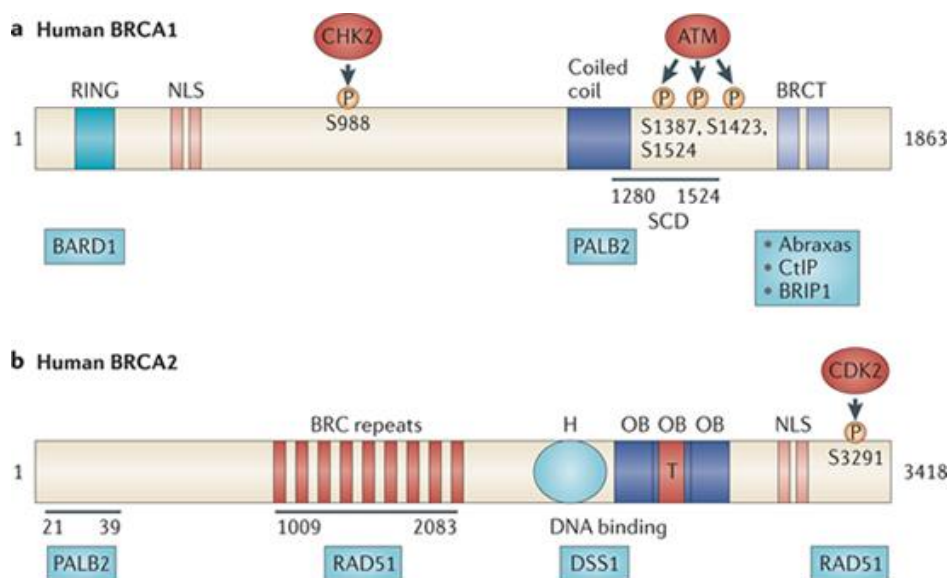
2.1 Geny s vysokou penetrancí

Mutace v genech s vysokou penetrancí jsou v panmiktické populaci velmi vzácné (<0,1 %) a jsou asociovány s vysokým rizikem vzniku nádorů prsu (až 10x vyšší). Jak již bylo uvedeno, zatímco přibližně 10-30 % všech nádorů prsu je připisováno genetickým faktorům, pouze 5-10 % případů vykazuje silnou dědičnou komponentu a méně, než jedna čtvrtina dědičných případů může být vysvětlena mutacemi v genech s vysokou penetrancí, přenášenými autozomálně dominantním způsobem (Ellsworth et al. 2010, Apostolou et al. 2013). Dva nejvýznamnější predispoziční geny, *BRCA1* a *BRCA2*, vysvětlují dohromady kolem 3-6% nádorů prsu (Kleibl et al. 2016) v různých populacích, avšak jejich vzájemný podíl vykazuje značný populačně-specifický charakter. Rodokmen, vykazující evidentní autozomálně dominantní predispozici ke karcinomu prsu, poprvé systematicky popsal francouzský lékař Pierre-Paul Broca v roce 1866. V rodině jeho ženy se vyskytlo devět žen ve čtyřech generacích, které zemřely na toto onemocnění (Broca 1866).

2.1.1 *BRCA1*

Pomocí genetických vazebných analýz ve 23 rodinách s výskytem karcinomu prsu v mladém věku, byl gen *BRCA1* lokalizován na chromozomu 17q21 (Hall et al. 1990) a následně klonován pro sekvenační a funkční analýzy (Miki et al. 1994). Patogenní mutace v *BRCA1* byly nalezeny u přibližně 7-10 % pacientů s karcinomem prsu s familiárním výskytem (Peto et al. 1999). Gen *BRCA1* obsahuje 24 kódujících exonů a jeho 7,4 kb dlouhý transkript kóduje protein o 1863 aminokyselinách. *BRCA1* nemá v lidském genomu žádný známý homolog, ani *BRCA2* jím není, přestože spolu oba proteiny sdílejí některé strukturní rysy (Thompson et al. 2004). Vysoce konzervované funkční motivy *BRCA1* zahrnují důležité domény: N-terminální RING finger doménu, jaderné lokalizační signály (Nuclear

Localization Signal, NLS), dvě protein-protein interakční BRCA1 C-terminal (BRCT) domény na C konci, kterým předchází coiled-coil doména (Caestecker et al. 2013). Protein BRCA1 vytváří řadu proteinových komplexů a to především vazbami s RING a BRCT doménami. Struktura a vazebná místa s interagujícími proteiny jsou naznačena na Obrázku 10.



Obrázek 10: Schematická struktura a interagující proteiny BRCA1 (a) a BRCA2 (b). Interagující proteiny jsou vyznačeny pod oblastmi BRCA1/2, které jsou nezbytné k asociaci. Převzato z (Roy et al. 2012).

Protein BRCA1 se účastní řady důležitých buněčných pochodů zajišťujících genomovou stabilitu, včetně jeho zásadní úlohy spočívající v regulaci oprav dvouřetězcových zlomů DNA (DSB) pomocí HR. Protein BRCA1 se však podílí i na dalších pochodech, jako je regulace kontrolních bodů buněčného cyklu, DNA replikace, regulace transkripce, NER, ubikvitinylace proteinů, a deacetylace histonů při remodelaci chromatinu (Wang 2012). BRCA1 se účastní časných pochodů regulace a aktivace HR, která opravuje DSB a meziřetězcové spoje (interstrand crosslinks, ICL) v DNA.

BRCA1 je fosforylován řadou kináz v odpovědi na přítomnost DSB v DNA, např. ATM, CHK2, a/nebo ATR. Poškození DNA spouští fosforylaci H2AX (γ H2AX) závislou na ATM a ATR. Tato fosforylace je nezbytná pro kumulaci řady DNA reparačních proteinů, cílených do místa reparačního ohniska pro zajištění optimální opravy DNA.

BRCA1 vytváří různé multiproteinové komplexy v místě poškození. Popsáno je pět z těchto komplexů. **Core komplex** je tvořen proteiny BRCA1 a BARD1 spojenými prostřednictvím BRCT domén. Tento stabilní heterodimer zajišťující E3 ubikvitinligázovou aktivitu je přítomen ve všech BRCA1 komplexech. BRCA1/PALB2/BRCA2 superkomplex je sestavován vazbou PALB2 na coiled-coil doménu

BRCA1 a hraje zásadní roli v opravách DNA zprostředkovaných HR. Další **BRCA1-A komplex** obsahuje proteiny ABRAXAS (FAM157B), RAP80/UIMC1, NBA1/MERIT40/BABAM1, BRE/BRCC45 a BRCC36/BRCC3 a je asociován s ubikvitinylovanými histony v blízkosti míst poškození DNA. **BRCA1-B komplex** obsahuje BRIP1 a TOPBP1 a je nezbytný pro regulaci kontrolního bodu indukovaného replikačním stresem a reparace meziřetězcových spojů (ICL repair). **BRCA1-C komplex** se skládá z CtIP a MRN (MRE11/RAD50/NBS1) a detekuje DSB. Tento komplex je zodpovědný za vystřížení úseků DNA v průběhu reparace. Všechny tyto komplexy (A, B, C) jsou formovány prostřednictvím interakcí BRCT domén BRCA1 proteinu (Welsh et al. 2000, Boulton 2006, Huen et al. 2010, Wang 2012, Rosen 2013).

Pro nádorovou predispozici jsou klinicky významné především dědičné nonsense mutace a malé inserce/delece způsobující posun čtecího rámce a následný vznik zkráceného proteinového produktu genu *BRCA1*. Významná je i řada popsaných missense mutací, především v důležitých RING a BRCT doménách. Popisovány jsou i patogenní mutace zahrnující větší část genu, kódující jeden i více exonů (Narod et al. 2004, Thompson et al. 2004, Caestecker et al. 2013).

Průměrné kumulativní riziko karcinomu prsu u nosičů mutací v genu *BRCA1* v 70 letech věku se pohybuje kolem 65 % (95%; CI 44-78 %), karcinomu ovaria 39 % (95% CI 18-54 %). Relativní riziko vzniku karcinomu prsu má vrchol ve věkové skupině 30-49 let a následně klesá, zatímco u karcinomu ovaria není u relativního rizika patrný výrazný trend. U nosičů mutací v *BRCA1* genu stoupá incidence až do věku 45-49 let a pak zůstává v podstatě konstantní (Antoniou et al. 2003). U nosičů mutací v *BRCA1* je také poměrně časté bilaterální onemocnění. *BRCA1*-asociované nádory prsu se často vyznačují vyšším stupněm a vykazují větší mitotickou aktivitu. Nádory jsou často ER-, PR- a HER2-, p53+, aneuploidní a histologicky vykazují medulární charakter. Nádory *BRCA1* nosiček mají podobný imunohistochemický profil jako sporadické bazální karcinomy (Lalloo et al. 2012, Mavaddat et al. 2012).

Četnost vrozených mutací genu *BRCA1* v ČR významně dominuje nad četností mutací v *BRCA2* (Machackova et al. 2008) (Pohlreich et al. 2005). Přibližně 11% germinálních mutací *BRCA1* tvoří intragenové přestavby postihující jeden a nejčastěji několik exonů (Ticha et al. 2010).

2.1.2 *BRCA2*

Gen *BRCA2* byl poprvé lokalizován analýzou patnácti vysoce rizikových rodin s karcinomem prsu, u kterých se nevyskytovala vazba onemocnění k *BRCA1* lokusu. Tato celogenomová vazebná analýza

mapovala *BRCA2* na chromozom 13q12-q13 v roce 1994 a samotný gen byl klonován v následujícím roce (Wooster et al. 1994, Wooster et al. 1995).

Patogenní mutace v *BRCA2* jsou zodpovědné za přibližně 10% případů dědičného karcinomu prsu (Walsh et al. 2007). *BRCA2* je velký gen s 27 exony o délce okolo 70 kb genomové DNA. Kóduje protein o délce 3418 aminokyselin. *BRCA2* obsahuje osm BRC repetic mezi aminokyselinovými zbytky 1009 a 2083, jednu DNA vazebnou doménu, tři oligonukleotid-vázající smyčky a tower doménu na C konci (Roy et al. 2012). Struktura proteinu *BRCA2*, znázorňující také vazebná místa interagujících proteinů je prezentována na Obrázku 10.

Podobně jako *BRCA1*, zásadní úlohou *BRCA2* je účast v reparaci DNA pomocí HR, kde *BRCA2* interaguje s *RAD51* (Turner et al. 2005). *RAD51* je klíčovým proteinem, který obaluje zpracovávanou jednořetězcovou DNA v místě DSB a iniciuje homologní párování a vzájemnou invazi těchto homologních oblastí v průběhu rekombinace (viz kapitola 1.3.1.1) (West 2003). *BRCA2* váže *RAD51* prostřednictvím osmi evolučně konzervovaných BRC repetic a samostatným motivem v C-terminální doméně. *BRCA2* reguluje tvorbu *RAD51* nukleoproteinových filament (Sharan et al. 1997, Davies et al. 2001, Pellegrini et al. 2002). *BRCA2* se také může vázat na jednořetězcovou DNA prostřednictvím C-terminální domény umožňující *BRCA2* zahájit rekombinaci. *BRCA2* působí přednostně na rozhraní mezi dvouřetězcovou DNA jednořetězcovým 3' přesahem zpracovaným pomocí MRN komplexu na odstranění replikačních proteinů z jednořetězcových přesahů a umožňuje nasednutí proteinu *RAD51* (Yang et al. 2002, Yang et al. 2005). Vazba *RAD51* a *BRCA2* je regulována fosforylací závislou na cyklin-dependentních kinázách a zdá se, že působí jako „vypínač“ kontrolující rekombinační opravnou aktivitu v průběhu přechodu z S/G2 do M fáze buněčného cyklu (Esashi et al. 2005).

Průměrné kumulativní riziko vzniku karcinomu prsu u nosičky mutací *BRCA2* ve věku 70 let je 45 % (95 % CI 31-56 %), pro karcinom ovaria 11 % (2,4-19 %). Stanovené relativní riziko karcinomu prsu mezi nosičkami mutací je vyšší v mladší věkové skupině (20-29 let) a s věkem se snižuje, naopak relativní riziko karcinomu ovaria vzrůstá od čtyřicátého roku života do 59 let a následně klesá (Antoniou et al. 2003).

Kromě toho, že je *BRCA2* predispozičním genem pro karcinom prsu a ovaria, je i jedním z genů Fanconiho anemie (FA). Bialelické mutace *BRCA2* způsobují FA, subtyp D1 (OMIM 605724) (Howlett et al. 2002). Dříve bylo předpokládáno, že bialelické mutace jsou embryonálně letální (Evers et al. 2006).

2.1.3 Další geny s vysokou penetrancí

Dalšími čtyřmi geny s vysokým rizikem karcinomu prsu jsou *TP53*, *PTEN*, *STK11*, a *CDH1*. *TP53* vykazuje nejvyšší rizikovost, neboť mutace v tomto genu zvyšují riziko 18-60x ve 45. roce života ve srovnání s obecnou populací. Více, než 50 % nosiček mutací v *TP53* vyvine onemocnění před 50. rokem života (Garber et al. 1991). Mutace v genech *PTEN*, *STK11* a *CDH1* v součtu přispívají k celoživotnímu riziku karcinomu prsu přibližně 40-60 % (Lalloo et al. 2012).

Gen ***TP53*** byl poprvé identifikován v roce 1979 a je dnes velmi dobře znám pro častou přítomnost jeho somatických mutací u nádorových onemocnění. Dědičné zárodečné mutace jsou málo časté, ale způsobují Li-Fraumeni syndrom, který je příčinou vzniku nádorových onemocnění v dětském věku (sarkomy měkkých tkání, osteosarkom, leukemie, nádory mozku, gliomy a adrenokortikální karcinom), ale i karcinomu prsu v raném věku. Více, než 70 % rodin klasifikovaných jako Li-Fraumeni vykazuje mutace v *TP53* a přibližně 30 % žen s mutací vyvine karcinom prsu před 30. rokem života (Malkin et al. 1990). Nicméně zárodečné mutace v *TP53* jsou výrazně méně časté oproti mutacím v *BRCA1* a *BRCA2*. Li-Fraumeni syndrom je zodpovědný za méně, než 0,1 % všech případů karcinomu prsu (Lalloo et al. 2012).

Protein p53 je důležitým transkripčním faktorem ovlivňujícím expresi genů, jejichž proteinové produkty regulují kontrolní body G1/S a G2/M buněčného cyklu, opravy DNA, buněčnou senescenci, apoptózu, autofagii a angiogenezi. Mezi zásadní funkce p53 patří aktivace apoptózy. Jelikož apoptóza nádorových buněk po radioterapii je zprostředkována p53 dráhou, mutační inaktivace p53 má za následek vznik radiorezistentních tumorů. Údaje z *in vitro* studií ukazují, že pacienti s Li-Fraumeni syndromem vykazují abnormální odpověď na nízké dávky radiace a defektní apoptózu (Lowe et al. 1994, Lalloo et al. 2012).

Mutace v genech *PTEN*, *STK11* a *CDH1* predisponují k různým nádorovým syndromům. Proteinový produkt genu ***PTEN*** je tumorsupresorová fosfatáza negativně regulující proproliferativní signální dráhu aktivace proteinkinázy B (PKB/AKT). Navíc bylo dokázáno, že *PTEN* se účastní udržování chromozomální integrity (Shen et al. 2007). Mutace v genu *PTEN* způsobují vzácný autozomálně dominantní dědičný Cowdenův syndrom (OMIM 158350), který zvyšuje riziko vzniku benigních i maligních tumorů prsu, štítné žlázy a endometria (Nelen et al. 1996). Nosiči mutací v *PTEN* mají 67-85% životní riziko vzniku karcinomu prsu (Ngeow et al. 2017) .

Gen ***STK11*** kóduje serin-threoninovou proteinkinázu podílející se na kontrole buněčné proliferace a polarity buňky. Dědičné mutace genu *STK11* způsobují autozomálně dominantní Peutz-Jeghersův syndrom (PJS, OMIM 175200) charakterizovaný hamartomatózními intestinálními polypy,

mukokutánní pigmentací a zvýšenou incidencí některých malignit, včetně karcinomu tlustého střeva, karcinomu prsu a karcinomu ovaria (převážně mucinózní forma) (Walsh et al. 2016). Celoživotní riziko vzniku karcinomu prsu se u nosičů mutací *STK11* pohybuje mezi 45-50% (Rosenthal et al. 2017).

Gen ***CDH1*** kóduje transmembránový E-cadherin, který je důležitý protein pro buněčnou polaritu, motilitu a mezibuněčnou komunikaci. Mutace v *CDH1* jsou zodpovědné za hereditární difúzní karcinom žaludku a autozomálně dominantní nádorový syndrom, který je asociovaný se zvýšeným rizikem vzniku lobulárního karcinomu prsu. Nicméně byly reportovány i rodiny s *CDH1* mutacemi a výskytem karcinomu prsu bez výskytu karcinomu žaludku (Hemminki et al. 1997, Hemminki et al. 1998, Pharoah et al. 2001, Masciari et al. 2007). Nosiči mutací v genu *CDH1* mají přibližně 39-50% riziko vzniku karcinomu prsu (van der Groep et al. 2011, Rosenthal et al. 2017).

Syndromy způsobené mutacemi v genech *PTEN*, *STK11*, nebo *CDH1* jsou velmi vzácné a proto je četnost vzniku karcinomu prsu způsobené mutacemi v těchto genech malá. V současnosti neexistuje důkaz, že by mutace v těchto genech byly zodpovědné za významnou část hereditárních, či sporadických karcinomů prsu bez přítomnosti dalších fenotypových znaků spojených s příslušnými syndromy (Turnbull et al. 2008).

RAD51C je kandidátním genem pro zařazení do skupiny genů s vysokou penetrancí, potvrzeno zatím ale bylo jen riziko vzniku karcinomu ovaria a význam tohoto genu k predispozici karcinomu prsu je kontroverzní. V klinických doporučeních genetického testování americké onkologické společnosti National Comprehensive Cancer Network (NCCN) je tento gen zatím uveden pouze pro karcinom ovaria (Daly et al. 2017). *RAD51C* je součástí komplexu složeného z pěti paralogů *RAD51*, které se účastní v procesu HR. Za tímto účelem vytváří *RAD51C* dva různé subkomplexy v místech poškození DNA díky asociaci se svými paralogy. Tyto subkomplexy následně umožňují reparaci DNA prostřednictvím HR (Masson et al. 2001).

RAD51C byl zařazen mezi predispoziční faktory pro Fanconiho anemii (FA). Bialelické mutace v *RAD51C* byly identifikovány u FA pacientů (FANCO podskupina) a následně byly i monoalelické alterace označeny jako varianty způsobující vysoké riziko nádorů prsu/ovarů ve srovnání s rizikem u nosičů mutací v genech *BRCA1/2* (Meindl et al. 2010, Vaz et al. 2010). Nicméně řada následujících studií neposkytla důkaz, že by *RAD51C* predisponoval specificky ke karcinomu prsu, neboť monoalelické mutace byly nalezeny u rodin s výskytem karcinomů prsu i ovaria. Frekvence mutací v *RAD51C* u rodin s výskytem karcinomu prsu a ovarů se pohybuje kolem 1,3-4 %. Tyto mutace vykazují vysokou, nebo střední penetranci dle (Antoniou et al. 2003, Akbari et al. 2010, Levy-Lahad 2010, Meindl et al. 2010, Clague et al. 2011, Pang et al. 2011, Pelttari et al. 2011, Romero et al. 2011, Silvestri et al. 2011, Vuorela et al. 2011, Thompson et al. 2012),(Zheng et al. 2010, Vuorela et al.

2011, Wong et al. 2011). Relativní riziko karcinomu ovaria pro nosičky mutací v *RAD51C* ve studii Loveday a kol. vycházelo 5,88 což odpovídá > 9% kumulativnímu riziku ve věku 80 let (Pelttari et al. 2011, Loveday et al. 2012). V české studii se vyskytovala mutace v genu *RAD51C* u 1,2 % vysoce rizikových pacientek s karcinomem ovaria (Janatova et al. 2015). V rámci velkých studií byly nalezeny nosičky mutací genu *RAD51C* i v rodinách s výskytem pouze karcinomu prsu. Jednalo se především o pacientky s triple negativním karcinomem. Frekvence nosiček mutací *RAD51C* v těchto studiích jsou nízké (0,4 %) (Buys et al. 2017, Golmard et al. 2017).

2.2 Geny se střední penetrancí

Oproti genům s vysokou penetrancí, u kterých se vyskytují patogenní varianty s velice nízkou populační frekvencí, se u variant genů se střední a nízkou penetrancí setkáváme s vyšší frekvencí (přibližně 1-40%; viz Obrázek 9), avšak prevalence genetických alterací v těchto genech vykazuje výrazné populační rozdíly (Foulkes 2008, Stratton et al. 2008). Odhaduje se, že varianty se střední penetrancí jsou zodpovědné za 5 % případů familiárního karcinomu prsu. Nosiči těchto mutací mají riziko vzniku karcinomu prsu do 60 let na úrovni přibližně 6-10 % ve srovnání se 3 % u běžné populace (Stratton et al. 2008, Hollestelle et al. 2010). Mutace ve všech těchto genech zvyšují 2-3 x riziko vzniku karcinomu prsu.

CHEK2, *ATM*, *BRIP1* a *PALB2* jsou nejvíce studované geny se střední penetrancí, neboť jejich proteinové produkty se účastní reparace DNA a jejich funkce je úzce spjata s proteiny *BRCA1* a *BRCA2* (diskutováno v kapitolách 1.3.1.1 a 2.1.1)(Meijers-Heijboer et al. 2002, Renwick et al. 2006, Seal et al. 2006, Erkko et al. 2007, Southey et al. 2010).

Biochemické analýzy a asociační studie kandidátních genů v řadě populací dokazují, že ***PALB2*** patří spolu s *BRCA1* a *BRCA2* mezi klíčové predispoziční geny karcinomu prsu. Nejvíce patogenní mutace doposud identifikované v genu *PALB2* jsou trunkační frame-shift, nebo nonsense mutace, které jsou nalézány napříč celým genem, nikoliv pouze v konkrétních oblastech. V průměru přispívají všechny mutace v *PALB2* ke střednímu celoživotnímu riziku vzniku onemocnění, ale některé mutace v tomto genu jsou asociovány s vysokým rizikem. Tyto mutace jsou srovnatelné s průměrným rizikem mutací v genu *BRCA2* (Erkko et al. 2007, Foulkes et al. 2007, Janatova et al. 2013). Recentní data z rozsáhlé studie 65057 žen ukazují, že riziko karcinomu prsu spojené s nosičstvím patogenních mutací *PALB2* je skutečně vysoké (OR=7,46; 95%CI 5,12-11,19) a prakticky tak umožňuje zařadit patogenní mutace v *PALB2* mezi varianty s vysokou penetrancí (Couch et al. 2017).

Gen **CHEK2** kóduje kinázu CHK2 (Obr. 5) kontrolující buněčný cyklus především v odpovědi na přítomnost poškození DNA prostřednictvím fosforylace řady substrátů (včetně p53 a BRCA1). Varianta c.1100delC v genu **CHEK2** vytváří zkrácený protein se ztrátou jeho kinázové aktivity. V Severní Evropě je přítomna až u 1,1 % zdravých jedinců a u 5,5 % postižených familiární formou karcinomu prsu. Relativní riziko u nosičů této varianty bylo stanoveno na 2,2 násobek (Meijers-Heijboer et al. 2002). Frekvence nosičů se značně liší mezi populacemi. V Nizozemí dosahuje 0,8 %, ve Finsku a Švédsku 0,7 % a 0,5 %. Nicméně v ostatních evropských zemích, Severní Americe a Austrálii málokde překračuje hodnotu 0,2 %. V České republice dosahuje frekvence c.1100delC 0,27 % (Kleibl et al. 2005, Nevanlinna et al. 2006). Nosiči mutace CHEK2*1100delC mají navíc zvýšené riziko vzniku bilaterálního onemocnění. Byly popsány i rodiny s homozygotní mutací CHEK2*1100delC. U takových nosičů bylo stanoveno zvýšené riziko karcinomu prsu 6 x a navíc je u takových nosičů patrně zvýšené riziko dalších malignit, jako je kolorektální karcinom, či karcinom prostaty (Adank et al. 2011). Nicméně tato mutace nevykazuje úplnou segregaci s nádorovým fenotypem karcinomu prsu v rodinách se znaky vysoké penetrance a nelze tedy vyloučit, že CHEK2*1100delC se podílí jako modifikační faktor na nádorové predispozici (Hollestelle et al. 2010).

ATM je apikální kinázou (Obrázek 5), která ovlivňuje řadu procesů při opravách DSB aktivací kaskády zahrnující p53, BRCA1 a CHK2 (diskutováno v kapitole 1.3.1). Bialelické mutace v genu **ATM** způsobují vzácný syndrom Ataxia-teleangiectasa (A-T, OMIM 208900), charakterizovaný progresivní mozečkovou ataxií, imunodeficiencí a nádorovou predispozicí, včetně predispozice ke karcinomu prsu (Easton 1994, Savitsky et al. 1995). Ženy, nosičky heterozygotních mutací v **ATM**, nevykazují fenotypové znaky A-T, ale jsou ohroženy více než dvojnásobně zvýšeným rizikem vzniku karcinomu prsu (Thompson et al. 2005, Renwick et al. 2006). Doposud byla identifikována řada predisponujících variant v genu **ATM**, včetně missense variant. Prevalence těchto variant se značně liší mezi populacemi z různých geografických oblastí. Penetrance **ATM** je přibližně 15 % (shrnutí v (Hollestelle et al. 2010, Laloo et al. 2012).

Gen **BRIP1** (znám také pod označením **BACH1**) kóduje DEAH helikázu, která interaguje s BRCA1 a má na BRCA1 závislou roli v opravách DNA a kontrole buněčného cyklu (Peng et al. 2006). Výskyt truncačních mutací v genu **BRIP1** je méně častý, než u **CHEK2** a **ATM**. Přibližně polovina nalezených mutací v **BRIP1** tvoří nonsense mutace 2392C>T (R798X). Ostatní patogenní mutace představují malé inserce a delece. Segregační analýzy ukázaly, že nosiči monoalelických mutací dosahují dvojnásobného relativního rizika vzniku karcinomu prsu (Thompson et al. 2005). Dle novějších studií je význam mutací **BRIP1** v predispozici ke karcinomu prsu (OR=1,28) minoritní oproti predispozici ke karcinomu ovaria (OR=8,13) (Rafnar et al. 2011). Jelikož se jedná o gen se střední penetrancí, kosegregace mutací v **BRIP1** s nádorovým fenotypem je neúplná. **BRIP1** byl zahrnut mezi FA geny

v komplementační skupině J (FANCI) u bialelických nosičů mutací. Fenotyp u FANCI skupiny je odlišný od fenotypu s bialelickými mutacemi v *BRCA2* (*FANCD1*) s výskytem významně nižšího počtu nádorů dětského věku (Levitus et al. 2005, Levrán et al. 2005, Litman et al. 2005, Seal et al. 2006).

Heteromerní proteinový MRN komplex (složený z proteinů **MRE11**, **RAD50** a **NBS1**; Obrázek 6) sehrává zásadní roli v buněčné odpovědi aktivované poškozením DNA, popsané v kapitole 1.3.1. Bialelické zárodečné mutace genu **NBN** (*NBS1*) zvyšují chromosomální instabilitu genomu a jsou původcem Nijmegen Breakage syndromu (NBS, OMIM 251260). Nosiči mutací v *NBN* vykazují zvýšené riziko vzniku nádorových onemocnění, včetně karcinomu prsu (Seemanova 1990, Seemanova et al. 2007). Varianta c.657del5 (c.657_661delACAAA) v genu *NBN* je nejčastější patogenní variantou, nalezenou u 90 % pacientů s NBS. Výskyt této varianty byl studován v řadě populací, ale její významná asociace s onemocněním byla prokázána především ve slovanských populacích, kde bylo zjištěno 3,1-násobně zvýšené riziko karcinomu prsu (Steffen et al. 2006, Bogdanova et al. 2008, di Masi et al. 2008). Novější publikace na populacích Německých a Severoamerických ukazují na velice nízkou asociaci s karcinomem prsu (OR=1,39) (Hauke et al. 2018), OR=1,27 (Couch et al. 2017).

Mutace v genu **RAD50** kódujícím další protein MRN komplexu jsou rovněž asociovány se zvýšeným rizikem karcinomem prsu. Ve finské populaci byla jako „founder“ (neboli zakladatelská) mutace, identifikována varianta c.687delT, která zvyšuje riziko karcinomu prsu čtyřikrát (Heikkinen et al. 2006). Další vzácná varianta, Q350X, byla identifikována v jedné rodině ve Velké Británii, nicméně asociace mezi touto mutací a onemocněním zůstává nepotvrzena (Tommiska et al. 2006).

Role germinálních mutací **MRE11** v patogenezi karcinomu prsu zůstává nejasná. Potenciální klinická relevance **MRE11** a celého MRN komplexu v tomto onemocnění je naznačena pozorováními, kdy basal-like nádory vykazují významné snížení exprese, či ztrátu proteinů MRN komplexu (Bartkova et al. 2008). Novější studie asociaci mutací v **MRE11** s predispozicí ke karcinomu prsu spíše potvrzují (Damiola et al. 2014).

Riziko vzniku karcinomu prsu spojené s neurofibromatózou typu 1 (**NF1**, OMIM 162200) může být také středně zvýšeno (celoživotní riziko přibližně 17 %) (Wang et al. 2018). **NF1** je relativně frekventní autosomálně dominantní onemocnění s úplnou penetrancí v rodině způsobenou mutacemi v genu **NF1**. Tento tumorsupresorový gen kóduje neurofibromin, působící jako negativní regulátor aktivovaného Ras proteinu, zvýšením procesivity intrinziční GTPázové aktivity Ras. Pacienti s vrozenými mutacemi *NF1* vykazují významně zvýšené relativní riziko vzniku gliomu, maligních nádorů periferních nervů, juvenilní myelomonocytové leukémie, rhabdomyosarkomu a feochromocytomu (Sorensen et al. 1986, Zoller et al. 1997, Airewele et al. 2001).

2.3 Geny s nízkou penetrancí

Do současné doby byla identifikována řada častých variant, které jsou spojeny s malým zvýšením rizika vzniku karcinomu prsu. Jejich polygenní působení s aditivními účinky může ozřejmit vznik významného počtu hereditárně podmíněných případů karcinomu prsu bez přítomnosti zjevných kauzálních mutací ve známých predispozičních genech. Geny s nízkou penetrancí zvyšují riziko vzniku karcinomu prsu méně než dvojnásobně, ale frekvence jejich heterozygotních variant v obecné populaci je vysoká, mezi 10 a 90 % (Pharoah et al. 2008, Turnbull et al. 2008). Validované geny s nízkou penetrancí jsou uvedeny v Tabulce 1. Pomocí GWAS (Genome Wide Association Study) projektů bylo validováno přibližně 20 SNP na úrovni významnosti, která dosahovala prahu odpovídajícímu velikosti genomu ($p < 5 \times 10^{-7}$) (Trainer et al. 2011). Dnes známé varianty s nízkým rizikem odpovídají za přibližně 10 % rizika karcinomu prsu v populaci, přestože mohou vysvětlovat větší podíl na dědičném onemocnění v rámci rodiny s historií karcinomu prsu (Latif et al. 2010, Mavaddat et al. 2010, Turnbull et al. 2010). Na rozdíl od mutací s vysokou penetrancí, většina z popsaných nízkopenetrantních SNP se nevyskytuje v kódujících oblastech genů. Řada SNP se zdá být asociována s geny se známou funkcí v procesu karcinogeneze nádorů prsu, jako je regulace růstu (*FGFR2*, *MAP3K1*), opravy DNA (*RAD51L*), apoptóza (*CASP8*), nebo hormonální signalizace (*ESR1*). Přibližně polovina SNP asociovaných s rizikem onemocnění se vyskytuje daleko od jakéhokoliv z kandidátních predispozičních genů, v oblastech označovaných jako genové pouště („gene deserts“). Důkladnější analýzy prokázaly, že tyto vzdálené varianty mohou mít vliv na distanční chromatinové interakce, které ovlivňují expresi genů vzdálených několik kilobází. Taková závislost byla identifikována např. pro nekódující oblast 8q24, která je zodpovědná za expresi protoonkogenu *MYC*, ale i pro další kandidátní geny, jako je regulátor buněčného cyklu cyklin D1 (gen *CCND1*) v oblasti 11q13, nebo fibroblastový růstový faktor *FGF10* v oblasti 5q12. Pro další varianty, jako je SNP rs13387042 v oblasti 2q35 zůstává mechanismus jejich účinku nejasný, přestože spojení s rizikem onemocnění bylo prokázáno (Penney et al. 2009, McClellan et al. 2010, Trainer et al. 2011). Doposud byl největší příspěvek k riziku karcinomu prsu připisován SNP alele v lokusu *FGFR2*. Tato varianta je přítomna u přibližně 1 % případů karcinomu prsu a je spojena s přibližně 1,1% relativním rizikem malignity. Zároveň byly nalezeny somatické mutace zvyšující aktivitu výsledného proteinového produktu u řady dalších nádorů (Easton et al. 2007). Naopak u kódující varianty D302H v genu *CASP8* byl prokázán protektivní účinek proti karcinomu prsu. Tato varianta se vyskytuje u 13 % žen evropské nenádorové populace a je spojena s přibližně 12% snížením rizika vzniku karcinomu prsu. Funkční dopady varianty D302H jsou neznámé a je možné, že popsané pozorování může být způsobeno jinými alelami ve vazebné nerovnováze. Nicméně efekt způsobený variantou D302H naznačuje významnost dědičné variability genů kódující efektoři v apoptotické dráze v nádorové predispozici ke karcinomu

prsu (Cox et al. 2007, Mavaddat et al. 2010). Zajímavé je také, že některé SNP byly identifikovány jako modifikátory BRCA1 a BRCA2, neboť u několika frekventních SNP (v lokusech *TOX3*, *FGFR2*, *MAP3K*, *LSP1*, 2q35, *SLC4A7*, 1p11.2, 5p12 and 6q25.1) bylo dokázáno, že zvyšují riziko karcinomu prsu u *BRCA2* mutovaných nosičů. Nicméně pouze SNP v *TOX3*, 2q35 and 6q25.1 vykazují zvýšené riziko u nosičů mutace v *BRCA1* (Antoniou et al. 2008, Antoniou et al. 2011).

Tabulka 1 Predispoziční geny pro karcinom prsu podle Lalloo & Evans 2012, Shuen & Foulkes 2011, Turnbull & Rahman 2008. *podle Southey et al. 2010.

Klasifikace	Gen	Relativní riziko
Vysoká penetrance	<i>BRCA1, BRCA2, TP53</i>	>10
	<i>PTEN, STK11, CDH1</i>	2–10
Střední penetrance	<i>ATM, CHEK2, BRIP1, NF1, RAD50, NBN</i>	2–3
	<i>PALB2</i>	2–6 (19*)
Nízká penetrance	<i>FGFR2, TOX3/TNRC9, MRPS30, FAM84B,</i>	
	<i>TNP1/IGFBP5/IGFBP2/TNS1, NOTCH2, ESR1, CDKN2A</i>	1.08–1.26
	<i>LSP1, MAP3K1, LSP1, NEK10</i>	1,07–1.13
	<i>CASP8, COX11, RAD51L1</i>	0.85–0.97

2.4 Kandidátní geny a jejich charakterizace

Přes recentní objevy genetických predispozic ke karcinomu prsu zůstává stále přibližně 70 % dědičného rizika nevysvětleno známými predispozičními faktory. Objev dalších predispozičních genů s podobným významem jako *BRCA1*, nebo *BRCA2*, není pravděpodobný, avšak lze předpokládat existenci, patrně mnoha, nových predispozičních genů s velmi nízkou frekvencí (<1:10-50 000) patogenních variant v populaci. Jelikož většina z doposud identifikovaných genů se střední penetrancí, jejichž mutace zvyšují riziko vzniku karcinomu prsu, se účastní oprav DNA, je pravděpodobné, že v této funkční skupině mohou být objeveny nové predispoziční geny. (Meijers-Heijboer et al. 2002, Seal et al. 2006, Erkko et al. 2007). Některé nádorové predispoziční geny (*PTEN*, *ATM*, *STK11*, *NBN*) byly identifikovány díky faktu, že jejich bialelické mutace způsobují autosomálně recesivní syndromická onemocnění, která mimo jiné zapříčiňují i predispozici k nádorovým onemocněním.

První predispoziční geny pro karcinom prsu byly nalezeny pomocí vazebných studií a pozičního klonování. Některé geny se střední penetrancí, jako je *CHEK2*, *PALB2* a *BRIP1*, pak byly nalezeny díky hledání v rámci omezeného množství vytipovaných kandidátních genů (Meijers-Heijboer et al. 2002, Seal et al. 2006, Rahman et al. 2007). V poslední době identifikované lokusy s nízkou penetrancí byly identifikovány pomocí celogenomových asociačních SNP studií. Extenzivní mezinárodní spolupráce s využitím nových technologií, jako je sekvenování nové generace, SNP a CNV (Copy Number Variation) ereje, již objevily řadu nových genů s nízkou penetrancí a zřejmě ještě v budoucnu objeví.

Nicméně riziko způsobené těmito novými faktory může být ještě nižší, než u již identifikovaných. Pomocí celogenomových asociačních studií (GWAS) bylo identifikováno více, než 100 jednonukleotidových polymorfismů (Single-Nucleotide Polymorphism – SNP), které mírně zvyšují riziko karcinomu prsu (Easton et al. 2007, Michailidou et al. 2013, Couch et al. 2015). Téměř všechny tyto SNP jsou nekódující a publikovaná data ukazují, že se některé z nich nacházejí v regulačních oblastech sousedních cílových genů a způsobují lehké změny v jejich expresi, jako je tomu např. u *CCND1* (French et al. 2013). Nebo ovlivňují posttranskripční regulaci, jako je třeba změna sestřihu u transkriptu genu *TERT* (Bojesen et al. 2013).

Přestože pro většinu ostatních lokusů nebyl doposud mechanismus vlivu na riziko karcinomu prsu objasněn, lze se domnívat, že pro velkou část z nich to bude právě modifikace exprese cílových genů v jejich blízkosti. Autoři práce Li a kolektivu navíc chtěli prokázat významnější vliv trunkačních variant v kódujících oblastech genů, které obsahují tyto nízkopenetrantní SNP, na expresi což se jim nepodařilo a zdá se tedy, že takovéto varianty mohou nízkou měrou přispívat ke zvýšenému riziku karcinomu prsu, ale patrně nebudou hlavními faktory pro vznik dědičného karcinomu prsu (Couch et al. 2015, Li et al. 2018).

Prohloubení znalostí o genetice karcinomu prsu a predispozici k tomuto onemocnění může vést k získání cílených terapií, založených na individuálním genetickém profilu pacienta. Na příklad bylo zjištěno, že inhibitory PARP (poly ADP-ribóza polymeráza) indukují selektivní nádorovou cytotoxicitu u *BRCA* mutovaných nádorů. Toto specifické působení umožňuje použití nové terapie pro nosiče *BRCA* mutací (viz kapitola 1.3.2) (Farmer et al. 2005).

3 NEXT GENERATION SEKVENOVÁNÍ

3.1 Vývoj sekvenování DNA

Vývoj technologie sekvenování DNA má bohatou historii, nicméně drtivá většina doposud vyprodukovaných sekvenačních dat byla vytvořena na některé z obměn biochemie Sangerovy metody (Hutchison 2007), (Sanger 1988). Nové milénium přineslo revoluční posun na poli DNA sekvenačních strategií, který byl umožněn několika zásadními předpoklady. Nejprve bylo dokončeno sekvenování lidského genomu v rámci The Human Genome Project, které poskytlo volně dostupnou referenční sekvenci lidského genomu a zároveň nastavilo limity v optimalizaci a redukci nákladů klasického Sangerova sekvenování (Venter et al. 2001). Se znalostí sekvence lidského genomu, ale i sekvencí základních modelových organizmů bylo možné rozvíjet metody sekvenování s krátkými čteními, které tak mohly být mapovány na příslušné oblasti genomu. Došlo tak k expanzi řady molekulárně genetických metod sekvenování, které mohou být souhrnně nazývány jako high-throughput DNA sequencing (Hutchison 2007), sekvenování nové generace (next-generation sequencing, NGS), nebo masivní paralelní sekvenování, které jsou v dnešní době hojně využívány nejen k sekvenování genomové DNA, ale i ke studiu genové exprese. Rozvoj těchto metod byl umožněn obecným technologickým pokrokem v oblastech povrchové chemie, polymerázového inženýrství, mikroskopie, bioinformatiky a informatiky skladování a manipulace s mnohem většími objemy dat, než bylo do nedávna možné.

3.2 Sangerovo sekvenování

Od začátku devadesátých let dvacátého století bylo sekvenování DNA prováděno téměř výhradně pomocí kapilárních poloautomatických sekvenátorů, využívajících Sangerovo sekvenování (Hunkapiller et al. 1991). Sekvenování DNA v rámci velkých sekvenačních projektů probíhalo především dvěma přístupy:

- i) pomocí tzv. shotgun *de novo* sekvenování, při kterém je náhodně fragmentovaná DNA klonována do klonovacích plazmidů, které jsou následně použity pro transformaci *Escherichia coli*, nebo
- ii) pomocí cíleného resekvenování, při kterém je využito PCR amplifikace s primery, které ohraničují cílovou oblast genomu.

Výsledkem obou přístupů je amplifikovaný templát, ať už klonální kopie jednoho plazmidového inzertu v izolované bakteriální kolonii, nebo příslušný počet PCR amplikonů obsažených v jednom reakčním objemu.

Biochemická část sekvenování probíhá prostřednictvím cyklické sekvenační reakce, ve které dochází ke střídání denaturace templátu, nasedání primeru a prodlužování DNA produktu. Primer je komplementární ke známé sekvenci ohraničující cílovou oblast sekvenování. Jednotlivé cykly prodlužování primerů jsou náhodně ukončeny fluorescenčně značenými 2',3'-dideoxynukleotidy (ddNTP). Ve výsledné směsi produktů s označenými 3' konci odpovídá fluorescenční značka příslušnému terminálnímu nukleotidu. Separace produktů sekvenační reakce, lišících se o jednu bázi a zakončených fluorescenčně značeným terminálním dideoxynukleotidem probíhá pomocí vysokorozlišovací kapilární gelové elektroforézy. Laserová excitace značených konců zaznamenaná detektorem pak poskytuje informaci produktu s příslušným terminálním nukleotidem při průchodu kapiláry čtecím okénkem. Software následně derivuje intenzitu a pořadí fluorescenčních signálů do křivky elektroforeogramu a příslušné nukleotidové sekvence a zároveň vypočítává chybovou pravděpodobnost detekce jednotlivé báze (Ewing et al. 1998, Ewing et al. 1998).

Automatizovaná Sangerova metoda je považována za technologii „první generace sekvenování“. Souběžná elektroforetická separace v 96, či 384 kapilárách je limitním množstvím pro tuto metodu. Po třiceti letech postupného zdokonalování může být pomocí Sangerova sekvenování dosaženo přesnosti na jednu bázi s hodnotou 99,999% a maximální délky přečteného fragmentu okolo 1000 bp. Náklady na jednu osekvenovanou kilobázi genetické informace se pohybují okolo \$2.

3.3 Sekvenování nové generace, „next-generation sequencing“

Metody nové generace, „next-generation sequencing“ (NGS) využívají různé strategie kombinující přípravu templátu, sekvenování, zobrazování a následných mapovacích bioinformatických metod. Jejich společným jmenovatelem je však především schopnost paralelního sekvenování tisíců až milionů sekvenačních templátů. Hlavní výhodou NGS metod je schopnost poskytnout enormní množství dat za nízké finanční náklady ve velmi rychlém čase. V některých případech objem dat přesahuje 10^{10} jednotlivých čtení („reads“) na jeden sekvenační běh.

Využití metod NGS se neomezuje pouze na sekvenování genomové DNA. Například microarray studie genových expresí jsou dnes nahrazovány metodami NGS, které mohou identifikovat a kvantifikovat

méně časté transkripty bez předchozí znalosti konkrétního genu, nebo mohou poskytnout informace o produktech alternativního sestřihu pre-mRNA (Wang et al. 2009, Hojny et al. 2017). Možnost sekvenovat celé genomy řady příbuzných organismů umožnila provedení velkých komparativních evolučních studií, jejichž realizace byla ještě před několika lety nemyslitelná. Nejrozšířenější využití NGS ovšem zůstává resekvenování lidských genomů pro porozumění genetických rozdílů mezi zdravím a nemocí.

Od doby vzniku NGS metod, se etablovalo několik sekvenačních technologických přístupů - platforem, přičemž některé z nich mají výhody pouze pro určité aplikace (Wold et al. 2008). Sekvenační technologie zahrnují posloupnost technologických postupů zahrnujících:

- i) příprava templátu / knihovny,
- ii) imobilizace a klonální amplifikace templátu,
- iii) sekvenování pořizováním optických/elektrochemických záznamů
- iv) bioinformatická analýza výstupních dat a jejich interpretace

Příprava sekvenační DNA knihovny zahrnuje přidání definované sekvence, zvané adaptor, na konce náhodně fragmentované, či pomocí primerů amplifikované DNA (Obrázek 17 v kapitole 5). Opatření cílové DNA adaptorovými sekvencemi je nezbytné pro ukotvení molekul DNA templátů při NGS a pro definování počátku sekvenační reakce. Všechny současné platformy, s výjimkou nastupující generace single molecule sekvenátorů diskutovaných v kapitole 4., vyžadují místně specifickou klonální amplifikaci molekul sekvenační knihovny pro zajištění dostatečné intenzity signálu pro detekci jednotlivých bází. Amplifikace templátů probíhá buď pomocí emulzní PCR, nebo pomocí bridge-PCR za tvorby klastrů přímo na vlastním sekvenačním skle. Vlastní postup sekvenování probíhá obvykle pomocí syntézy DNA polymerázou, alternativním přístupem je ligace fluorescenčně značených oligonukleotidů (Rodríguez-Ezpeleta 2012).

NGS metody byly komercializovány řadou společností (Tabulka 2). První komerčně dostupnou NGS platformou byl Sekvenační systém GS FLX (Margulies et al. 2005), který byl původně vyvinut společností 454 Life Sciences a později převzat společností Roche. Další technologie, produkující kratší čtení, nežli 454, byl Genome Analyzer společnosti Solexa (Bentley 2006). Illumina se postupně vyvinula do platformy v současné době rozšířené nejvíce. Později vyvinutá technologie SOLiD, společnosti Life Technologies (původně Applied Biosystems), využívá sadu fluorescenčně značených oligonukleotidů, které jsou ligovány na základě komplementarity s templátem (Smith et al. 2008). V roce 2010 byla na trh uvedena technologie polovodičového sekvenování, společnosti IonTorrent Systems, která využívá detekci změny pH při uvolnění H^+ po navázání nukleotidu v průběhu syntézy DNA (Rothberg et al. 2011). Tato platforma nahradila platformu SOLiD ve společnosti Life

Technologies pod označením IonTorrent. Další technologií je GCA systém společnosti Complete Genomics, dnes již BGI Company, který je založen na imobilizaci tzv. DNA nanoballs (klonálních jednomolekulových amplifikátů) na pevném podkladu v kombinaci s ligačním sekvenováním (Drmanac et al. 2010). Tato technologie je oproti výše zmíněným rozšířena jen málo.

Tabulka 2: Přehled DNA sekvenačních platform nové generace

Platforma	Společnost	Sekvenační knihovna	Podklad	Amplifikace	Sekvenační reakce	Detekční metoda
454	Roche	lineární adaptory	pikotitrační destička	emulzní PCR	syntézou	pyrosekvenování
Genome analyzer	Illumina	lineární adaptory	flow cell	bridge PCR	syntézou	reverzně terminátorové nukleotidy s fluorescenční značkou
SOLiD	Life Technologies	lineární adaptory	flow cell	emulzní PCR	ligací	oligonukleotidové próby s fluorescenční značkou
Ion Torrent	Life Technologies	lineární adaptory	polovodičový čip	emulzní PCR	syntézou	změna pH
CGA	Complete Genomics (BGI)	cirkulární adaptory	DNA nanoball	Rolling circle	ligací	oligonukleotidové próby s fluorescenční značkou
PacBio RS	Pacific Biosciences	vlásečkové adaptory	Zero mode waveguide	jediná molekula	syntézou, real-time	nukleotidy s fluorescenční značkou
Nanopore	Oxford Nanopore	vlásečkové adaptory	Nanopore	jediná molekula	syntézou, real-time	změna napětí

3.3.1 GS FLX (Roche)

Sekvenační proces systému GS FLX/Junior (Leguen et al.), zahrnuje přípravu knihovny s upravenými konci fragmentovaných DNA molekul, imobilizaci vzorku na streptavidinových kuličkách a pyrosekvenování.

Prvním krokem přípravy knihovny je náhodná fragmentace DNA na 300-800 bp dlouhé fragmenty. Následně je fragmentovaná dvouřetězcová DNA opracována tzv. end-repair enzymy, které postupně vytvoří tupé konce a jednonukleotidový přesah adeninu na 3' koncích molekul knihovny. Adaptorové oligonukleotidy, označované jako „A“ a „B“, jsou následně naligovány na konce fragmentované DNA. Jednořetězcový zlom mezi inzertem a adaptorem vzniklý po ligaci je odstraněn pomocí Bst DNA polymerázy, která nahradí celý řetězec, ve kterém se zlom nachází. Adaptor „B“ nese navázanou molekulu biotinu, která umožňuje purifikaci fragmentů se správnou kombinací adaptorů. Sekvenační knihovna je izolována pomocí streptavidinových magnetických kuliček. Fragменты obsahující biotinem značený adaptor B jsou navázány na streptavidinové kuličky, zatímco homozygotní, nebiotinylované A/A adaptory jsou odmyty. Imobilizované fragmenty jsou denaturovány, přičemž oba řetězce B/B adaptorových fragmentů zůstanou imobilizovány prostřednictvím vazby biotin-streptavidin a jednořetězcové templáty A/B fragmentů jsou uvolněny a použity k sekvenování.

Knihovna vázaná na kuličky je v systému GS FLX amplifikována pomocí emulzní PCR, při které dochází k paralelní amplifikaci fragmentů v prostředí směsi olejové a vodní fáze. Kapénky vody v olejové emulzi vytváří mikroskopické reaktory, ve kterých se ideálně nachází pouze právě jedna kulička s jediným DNA templátem. Na každé z kuliček mohou být po proběhnutí amplifikace imobilizovány stovky tisíc kopií jediného templátu. Takto vzniklé kuličky s navázanými DNA molekulami knihovny jsou poté deponovány do jamek pikotitrační destičky, která obsahuje více než jeden milion jamek o pikolitrovém objemu (Margulies et al. 2005). Rozměry těchto jamek jsou tak malé, že pouze jediná kulička může obsadit jednu jamku. Sekvenační chemie je rozlévána po celé ploše destičky a k jednotlivým sekvenačním reakcím dochází uvnitř jamek. Vlastní proces detekce inkorporované báze v systému GS FLX je založen na pyrosekvenační reakci (Margulies et al. 2005). V průběhu syntézy DNA je využívána detekce uvolňovaného pyrofosfátu při každé inkorporaci báze DNA polymerázou. Enzymy ATP-sulfuryláza a luciferáza konvertují pyrofosfát na záblesk viditelného světla, který je detekován pomocí CCD kamery. Každý ze čtyř nukleotidtrifosfátů je aplikován na pikolitrovou destičku jednotlivě pro každý sekvenační cyklus. Tato technologie nepoužívá žádné terminátory polymerace, které by zabránily přidání většího počtu stejného typu báze v rámci jednoho cyklu. Tato vlastnost je pro GS FLX systém limitující pro přesné sekvenování v homopolymerních oblastech, protože přidání většího počtu stejných nukleotidů je detekováno pouze na základě intenzity pyrofosfatázové reakce, která nemá zcela lineární průběh.

3.3.2 Genome Analyzer (Illumina)

Tento systém je založen na imobilizaci lineárních molekul DNA knihovny na pevné podložce, na které zároveň probíhá amplifikace. K vlastnímu sekvenování využívá fluorescenčně značené reverzně terminátorové nukleotidy.

Příprava vzorku pro systém Genome analyzer (nyní v podobě platforem iSeq, MiniSeq, Miseq, NextSeq a HiSeq/NovaSeq) společnosti Illumina zahrnuje, podobně jako u platformy GS FLX, přidání specifických adaptorových sekvencí na konce fragmentovaných DNA molekul (Bentley 2006). Fragmentace definuje počáteční pozice sekvenačního čtení v rámci DNA molekuly. V následujícím kroku jsou pomocí enzymatické směsi zarovnány konce DNA fragmentů a dále jsou přidány jednonukleotidové přesahy adeninu na 3' konce. Na takto upravené molekuly jsou naligovány dvouřetězcové adaptorové molekuly, které jsou jen částečně komplementární. Adaptory opatřené DNA je amplifikována pomocí PCR s primery, které do molekuly inkorporují sekvenci specifickou pro příslušný vzorek (tzv. index) a zároveň sekvenci nezbytnou pro imobilizaci molekuly v průběhu vlastního sekvenování.

Tato platforma využívá skleněnou planární komůrku (tzv. flow cell), na jejímž povrchu jsou imobilizovány molekuly DNA knihovny a proudí jí sekvenační reagentie. Vnitřní povrch komůrky je opatřen oligonukleotidy, které jsou na sklo navázány ve vysoké hustotě. Sekvence těchto oligonukleotidů je komplementární k sekvenci adaptorové části molekuly DNA knihovny. Po hybridizaci molekul knihovny k imobilizovaným oligonukleotidům, sloužícím jako primery, dochází pomocí DNA polymerázy k prodlužování dle templátové molekuly DNA knihovny, přičemž vzniká imobilizovaná kopie původní molekuly knihovny. Následně musí dojít ke klonální amplifikaci jednotlivých molekul sekvenační knihovny a tvorbě imobilizovaných, koncentrovaných a mikroskopicky detekovatelných shluků (klastřů) identických DNA molekul. Povrch komůrky opatřený primery slouží jako platforma pro amplifikaci imobilizované sekvenační knihovny v procesu zvaném bridge-PCR. Oproti klasické PCR reakci, která probíhá v roztoku za cyklického střídání kroků denaturace, anelace a extenze, bridge-PCR technika probíhá na pevném podkladu sekvenační komůrky, který je omýván jednotlivými reagentii. V průběhu reakce dochází, stejně jako u klasické PCR, ke střídání zmíněných tří teplot. Bridge-PCR je iniciována hybridizací imobilizované molekuly DNA knihovny (jak bylo popsáno výše) a amplifikačním primerem za tvorby struktury molekulárního mostu, ve které jsou oba konce molekuly připojeny k povrchu komůrky. Takto imobilizovaná DNA molekula je templátem pro elongační reakci zprostředkovanou DNA polymerázou. Výsledná dvouřetězcová DNA molekula je na jednom konci uvolněna denaturačním činidlem. Cyklické změny teplot a omývání amplifikačními reagentii generuje ve výsledku shluky čítající přibližně tisíc kopií příslušné molekuly DNA knihovny. Nakonec jsou odstraněny komplementární vlákna a na povrchu sekvenační komůrky zůstávají ukotveny jednořetězcové molekuly, na jejichž 3' konec je umístěn terminátor polymerizace a následně je k nim hybridizován sekvenační primer (Kircher et al. 2009).

Připravená sekvenační komůrka je napojena na vysoce výkonný zobrazovací systém, který se skládá z mikroskopické CCD kamery, excitačních laserů, či LED zdrojů světla a fluorescenčních filtrů. Tato metoda využívá čtyři různé fluorofory a reverzibilně terminovatelne nukleotidy. Sekvenační reakce započiná prodlužováním hybridizovaného primeru terminačními nukleotidy prostřednictvím DNA polymerázy. Přidaná báze nese specifický fluorofor a terminační skupinu, která zabraňuje inkorporaci dalších nukleotidů. Po inkorporaci jedné báze je pořízen snímek imobilizovaných clusterů. Pozice zaznamenaného nukleotidu definuje první bázi sekvenačního čtení. Před dalším cyklem jsou vratný terminátor polymerace a fluorofor chemicky odstraněny, tak, aby bylo umožněno přidání dalšího nukleotidu. Synchronní extenze sekvenovaného řetězce DNA vždy pouze o jeden nukleotid za jeden cyklus umožňuje přesné čtení i v případě homopolymerních sekvencí. Nicméně s jistou pravděpodobností dochází k selhání inkorporace báze v průběhu jednoho cyklu, což má za následek o jednu bázi posunutě čtení, neboli tzv. off-phasing effect. V průběhu narůstání počtu cyklů se takto

vznikající chyby stávají četnější a přesnost sekvenování klesá se stoupající délkou fragmentů. Z tohoto důvodu je Illumina technologie limitována na kratší sekvenační čtení (do 300bp, nejčastěji však 75-150bp) (Kircher et al. 2009).

3.3.3 SOLiD (Life Technologies)

Sekvenační platforma SOLiD společnosti Applied Biosystems (Valouev et al. 2008, Smith et al. 2010) je založena na technologii Polonator (Shendure et al. 2005), která využívá amplifikaci pomocí emulzní PCR, následnou imobilizaci DNA knihovny na pevný nosič a cyklické ligační sekvenování.

Příprava sekvenační knihovny na systému SOLiD zahrnuje fragmentaci DNA na úseky o délce cca 100-200 bp, úpravu konců a ligaci tzv. „P1“ a „P2“ adaptorových DNA molekul (Valouev et al. 2008). Emulzní PCR je zde využita ke klonální amplifikaci a imobilizaci jednotlivých molekul DNA knihovny na paramagnetické kuličky obalené P1 adaptorovými oligonukleotidy. Pole imobilizovaných P1 kuliček o vysoké hustotě obsahující sekvenačními templáty je vytvořeno díky modifikaci 3' konců DNA molekul knihovny, jejichž prostřednictvím jsou jednotlivé kuličky přichyceny na sekvenační sklo.

Sekvenační chemie systému SOLiD je založena na ligaci. Nejprve dochází k hybridizaci sekvenačního primeru na P1 adaptor imobilizované kuličky. Následně je přidána směs fluorescenčně značených oligonukleotidů všech možných variant v pořadí bází komplementárních k sekvenovanému templátu. Tyto osminukleotidové oligonukleotidy jsou hybridizovány k templátu a pouze přesně nasedající molekuly jsou ligovány k primeru. Každý z oligonukleotidů obsahuje specifickou dinukleotidovou sekvenci, které odpovídá konkrétní fluorescenční značce. Po pořízení obrázku CCD kamerou jsou odstraněny fluorofory. Další cyklus začíná o 5 bází dál od místa počátku čtení. Po deseti sekvenačních cyklech je odstraněn primer a na pozici n-1 od původního startu čtení je hybridizován druhý primer. Dohromady je využito všech sekvenačních primerů (n, n1, n-2, n-3 a n-4). Ve výsledku je padesátinukleotidový fragment čten dvakrát pro zvýšení přesnosti. Jelikož sekvenování platformy SOLiD využívá značené oligonukleotidy s kombinací různých dinukleotidů, vyžaduje interpretace hrubých dat poměrně složitý algoritmus (Valouev et al. 2008).

3.3.4 CGA platforma (BGI)

CGA technologie je založena na přípravě cirkulárních DNA knihoven a následné tzv. „rolling circle“ amplifikaci (RCA) pro tvorbu nanokuliček (tzv. nanoballs), které se ukotvují na pevný nosič (Drmanac et al. 2010).

Při přípravě knihovny je DNA fragmentována na 400-500 bp dlouhé fragmenty, jejichž konce jsou enzymaticky zatupeny a defosforylovány. Univerzální adaptorové molekuly jsou naligovány na konce fragmentů pomocí nick translace. Pouze molekuly opatřené adaptory jsou v následujícím kroku amplifikovány a to pomocí PCR využívající primery, které obsahují uracil. Uracily jsou z výsledných amplikonů odstraněny za vzniku jednořetězcových přesahů. Molekuly knihovny jsou následně štěpeny pomocí restričního enzymu *AclI* a cirkularizovány za použití T4 DNA ligázy za přítomnosti spojovacího oligonukleotidu. Cirkularizované produkty jsou purifikovány pomocí exonukleázy, která degraduje pouze lineární DNA molekuly. Linearizace, ligace adapteru, PCR amplifikace, restrikce a cirkularizace jsou opakovány, dokud nejsou do cirkulárních molekul knihovny inkorporovány čtyři různé unikátní adaptorové sekvence. Před finálním cirkularizačním krokem je purifikován jednořetězcový templát pomocí separace na paramagnetických kuličkách a ošetření exonukleázou. Výsledný produkt obsahuje dva třináctibázové inzerty genomové DNA a dva dvacetišestibázové inzerty genomové DNA přilehlé k adaptorovým sekvencím. K amplifikaci sekvenační knihovny na platformě CGA dochází prostřednictvím RCA procesu s vysoce procesivní polymerázou Phi29, která nahrazuje část sekvence templátu. RCA vytvoří dlouhé řetězce DNA z cirkulární knihovny, které obsahují krátké palindromické sekvence, které uvnitř dlouhého lineárního DNA produktu zapříčiní intramolekulární vytvoření šroubovicové struktury a tvorbu tzv. DNA nanoballs (DNBS), neboli DNA nanokuliček. DNA nanokulička je dlouhý řetězec repetitivních fragmentů amplifikované DNA, která vytváří detekovatelné, trojrozměrné, kondenzované, kulovité objekty, které jsou využívány při vlastním sekvenování.

Sekvenační povrch GCA čipu je pokryt hexamethylsilylanem (HMDS), který inhibuje vazbu DNA. Na povrchu čipu jsou body aminosoli o průměru 300nm, vytvořené pomocí fotolitografie. Tyto body pokrývají 95% povrchu sekvenační platformy. Na rozdíl od HMDS, váží pozitivně nabitá aminosolová struktura DNA nanokuliček. Vzniká tak sekvenační plocha s přibližně 350 miliony imobilizovaných DNA nanokuliček, které jsou od sebe navzájem vzdáleny 1,29 μm . CGA využívá strategii nazvanou „Combinatorial probe anchor ligation“ (cPAL), neboli kombinatorní ligation kotvení sekvence. Celý proces začíná hybridizací kotveního oligonukleotidu k jednomu z unikátních adaptorů na molekule DNA nanokuličky. Čtyři degenerované nonamerní oligonukleotidy značené fluorescenčními značkami odpovídají jednotlivým nukleotidům (A, T, C, nebo G) na první pozici oligonukleotidové próby. Určení sekvence nukleotidů probíhá reakcí, při níž správně párující hybridizační próba nasedá na templát a je ligována pomocí T4 DNA ligázy. Po vytvoření snímku ligovaných produktů CCD kamerou, jsou molekuly prób odmyty. Proces hybridizace, ligace, focení a denaturace je opakován pětikrát, vždy s použitím nového setu fluorescenčně značených oligonukleotidových prób, které obsahují známé báze na pozicích n+1, n+2, n+3 a n+4. Po pěti cyklech

přesnost ligační reakce poklesá a sekvenování pokračuje znovunastavením reakce s použitím kotevní sekvence s degenerovanou oblastí pěti bází. Následně dochází k dalším pěti cyklům sekvenování ligací s využitím nonamerních, fluorescenčně značených práb. Cyklické sekvenování deseti bází může být opakováno až osmkrát, vždy začínající na pozici každé unikátní kotevní sekvence. Výsledkem je 62-70 bází dlouhé čtení z každé DNA nanokuličky. Právě krátká délka jednotlivých čtení je hlavním omezením této metody. V roce 2013 byla společnost Complete genomics spolu se všemi patenty pro technologii GCA koupena společností BGI-Shenzen, která v roce 2016 spustila produkci systému BGI seq-1000, který je na této technologii založen a slibuje délku čtení až 2 x 100bp.

3.3.5 Ion Torrent (Life Technologies)

Ion Torrent technologie, neboli metoda polovodičového sekvenování, je založena na detekci protonů, které jsou uvolňovány během polymerace DNA. Tato platforma je jednou z metod NGS, které využívají sekvenování syntézou, při které je nově syntetizovaný komplementární řetězec syntetizován podle sekvence templátové molekuly DNA (Rothberg et al. 2011).

Příprava sekvenační knihovny z genomové DNA fragmentované na 300-400 bp dlouhé úseky zahrnuje opravu 3' a 5' konců a ligaci adaptorových DNA molekul, obdobně, jako u výše popsaných platform SOLiD, či Illumina. Po ligaci adaptorů jsou produkty sekvenační knihovny purifikovány a zároveň jsou izolovány pouze fragmenty o požadované délce pomocí paramagnetických, DNA vázajících kuliček. K následné amplifikaci je využita emulzní PCR, podobně, jako u systému SOLiD. Amplifikace probíhá na kuličkách opatřených oligonukleotidy, které jsou komplementární k sekvenci adaptoru. V rámci jednotlivých kapének pak dochází ke klonální amplifikaci unikátních molekul DNA knihovny.

Vytvořené kuličky s navázanými DNA molekulami jsou následně umístěny po jedné do jamek na sekvenačním čipu. Jeden čip pojme přibližně půl miliardy sekvenačních kuliček. Sekvenační chemií je v průběhu sekvenování čip omýván. Vlastní sekvenování vychází z principu inkorporace deoxyribonukleotid trifosfátu (dNTP) do molekuly syntetizované DNA. Při něm dochází ke tvorbě kovalentní vazby, uvolnění pyrofosfátu a H^+ . Detekce změny konduktance H^+ na polovodičovém rozhraní je ukazatelem inkorporace příslušného dNTP.

Jamky čipu, které obsahují kuličku s amplifikovaným templátem a DNA polymerázu jsou v průběhu sekvenování postupně zaplavovány roztoky jednotlivých nukleotidů (A, C, T, nebo G dNTP). Pokud se do jamky dostane dNTP, který je komplementární k sekvenovanému templátu, je zainkorporován do syntetizovaného řetězce pomocí DNA polymerázy a je uvolněn proton. Uvolnění protonu je sledováno jako změna pH, které je detekováno pro každou jamku zvlášť. Nepřipojené nukleotidy jsou

odmyty z čipu a v dalším cyklu je jamka zaplavena dalším druhem dNTP. Série elektrických pulzů generovaná pH senzitivní vrstvou umístěnou pod čipem, je následně přeložena do sekvence DNA. Zařazení jednotlivých nukleotidů je měřeno na základě změny elektrického potenciálu a není tedy potřeba fluorescenčních značek a převodu z optického signálu. Slabinou této metody jsou homopolymerní oblasti v sekvenování DNA, neboť v takovém případě je inkorporováno více dNTP molekul v jednom cyklu a odečtení snížení pH není tak přesné, aby rozlíšilo rozdíl mezi např. šesti a sedmi A v homopolymerní sekvenci za sebou. Limit spolehlivosti u takových úseků se pohybuje kolem 4-5 za sebou následujících stejných bází.

3.4 Metody single molecule sekvenování

Přesto, že NGS přineslo revoluční změny v možnostech komplexního pohledu na genomy organismů, jedná se stále o poměrně nákladnou a náročnou technologii. Základním technologickým limitem dnešních přístupů NGS je neschopnost sekvenování dlouhých (>1000 b) úseků DNA. Toto omezení překonává následná generace technologií sekvenujících jednotlivé molekuly DNA, která generuje čtení o délce přesahující 1 Mb genetické informace.

3.4.1 PacBio

Sekvenační technologie PacBio RS využívá DNA polymerázu, fluorescenčně značené nukleotidy a vysokorozlišovací zobrazovací zařízení pro detekci procesu DNA syntézy jediné molekuly v reálném čase (Eid et al. 2009).

SMRTbell je základní metodou pro přípravu sekvenačních knihoven pro systém PacBio RS, díky které dosahuje tato platforma vysoké přesnosti detekce variant (Flusberg et al. 2010). Pro sekvenování genomu je DNA náhodně fragmentována a její konce jsou opraveny, podobně, jako u výše popsaných NGS metod. Dále jsou přidány na 3' konce fragmentované DNA jednonukleotidový přesahy adeninu pro usnadnění ligace adaptoru s jednonukleotidovým přesahem thyminu. Jako adapteru se zde využívá jediného DNA oligonukleotidu, který vytváří intramolekulární vlásenkovou strukturu. SMRTbell templátová DNA je strukturně lineární molekula, ale vlásenkové adaptory z ní činí topologicky cirkulární molekulu. Sekvenační komůrka SMRT technologie obsahuje desetitisíce tzv. zero-mode waveguides (ZMW), (Levene et al. 2003)). ZMW jsou nanotechnologicky vytvořené jamky ve skleněném povrchu, potažené vrstvou hliníku. Objem těchto jamek se pohybuje v zeptolitrových jednotkách. SMRT komůrka je připravena pro imobilizaci DNA polymerázy potažením povrchu streptavidinem. Příprava sekvenační reakce vyžaduje inkubaci biotinylované Phi29 DNA polymerázy

s vytvořenými SMRTbell DNA templáty. Po započetí polymerázové reakce inkorporuje DNA polymeráza navléknutá na vlákno templátu jednotlivé fluorescenčně značené nukleotidy. Každá fluorescenční barva odpovídá jednomu ze čtyř možných nukleotidů A,T,C, nebo G (Korlach et al. 2008). V průběhu iniciace inkorporace nové báze se dostává fluorescenčně značený nukleotid do aktivního místa enzymu a do těsné blízkosti skleněného dna ZMW. Na dně každé ZMW jamky je vysokorozlišovací kamera, která zaznamenává fluorescenci právě inkorporovaného nukleotidu. Během inkorporační reakce je z nukleotidu uvolněn fluorescenčně značený fosfát, čímž fluorescenční signál opět vymizí. Spolu s tím, jak polymeráza syntetizuje kopii templátové molekuly, jsou události inkorporace úspěšných nukleotidů zaznamenávány ve formátu podobném videu. Použití SMRTbell knihoven s krátkými inzerty umožňuje sekvenovat v režimu nazývaném „circular consensus sequencing“, při kterém je stejný templát čten na sense i antisense vlákně několikrát. Množství čtení je odvislé od délky insertu. Tento vysoce redundantní přístup sekvenování zvyšuje přesnost určování pořadí bázi a překonává tak vysoké frekvence chyb, které jsou obecně spojeny s real-time sekvenováním, čímž umožňuje přesnější detekci variant.

3.4.2 Nanopore

Nanopore sekvenování je založeno na zaznamenávání změn elektrického proudění při průchodu syntetizované molekuly DNA pórem umístěným v nepropustné membráně.

Příprava sekvenační knihovny pro Nanopore technologii zahrnuje fragmentaci DNA na velmi dlouhé úseky (až desítky kb), opravu konců a ligace adaptorových molekul na konce DNA fragmentů. Jeden z těchto dvou adaptorových oligonukleotidů vytváří vlásenkovou strukturu, která umožní sekvenování templátu i z druhého řetězce za účelem zvýšení přesnosti sekvenování, podobně, jako metoda circular consensus sequencing u PacBio platformy. Při vlastním sekvenování prochází nově syntetizovaný řetězec 1 nanometr širokým transmembránovým pórem v syntetické membráně uložené v sekvenační komůrce. Nanopór je ponořen ve vodivém roztoku, čímž je vytvořen potenciál napříč membránou. Pórem tak prochází slabý elektrický proud, který je kvantifikován. Velikost proudu procházejícího nanopórem se snižuje při průchodu molekul. Snižování procházejícího proudu je velmi senzitivní na velikost a tvar molekuly pórem procházející. Při průchodu řetězce DNA jsou tak rozeznávány různé výchylky v měřeném napětí, odpovídající jednotlivým bázím v pořadí sekvence čtené DNA molekuly. Výhodou této technologie je možnost přímého real-time sekvenování bez potřeby PCR amplifikace, nebo chemického značení nukleotidů s následnou nezbytnou optickou instrumentací.

4 PRACOVNÍ HYPOTÉZA A CÍLE STUDIE

- V nedávné době byla charakterizována řada dalších predispozičních a kandidátních genů, ale výskyt jejich patogenních alterací se v jednotlivých geograficky odlišných populacích často liší. Významná část vysoce rizikových pacientů z rodin s dědičným karcinomem prsu může reprezentovat nosiče populačně specifických nebo i privátních mutací. Většina doposud identifikovaných predispozičních genů, jejichž mutace predisponují ke vzniku karcinomu prsu u vysoce rizikových osob, kóduje proteiny ovlivňující DNA reparační pochody, především pak reparaci dvouřetězcových zlomů DNA. Avšak mutační analýza byla provedena jen ve velmi malé části všech DNA reparačních genů. Lze předpokládat, že v doposud neanalyzovaných genech kódujících DNA reparační proteiny můžeme identifikovat raritní, ale klinicky významné genetické alterace způsobující nádorovou predispozici v postižených rodinách.
- Cílem dizertační práce byla identifikace populačně specifických alterací genů predisponujících ke vzniku dědičné formy karcinomu prsu v populaci patientek s mnohočetným výskytem tohoto onemocnění v rodině. Analýza byla zaměřena také na otázku významu alterací v genech kódujících proteiny podílejících se na opravách genomové DNA.
- Současně bylo cílem ověřit vhodnost zvolené metodiky sekvenování nové generace (NGS) pro identifikaci obecných i populačně specifických variant genů predisponujících k nádorovému onemocnění, v tomto případě hereditárnímu karcinomu prsu. Navržen byl postup vlastního návrhu NGS panelu, který obsahuje jak známé predispoziční geny pro identifikaci nových, či populačně specifických variant, tak i geny kandidátní zvolené na základě jejich možné funkční spojitosti s tumorigenezí karcinomu prsu. Především byl kladen důraz na geny účastnící se reparace genomové DNA.
- Vybrané nově identifikované kandidátní varianty s možným významem pro predispozici ke karcinomu prsu nebo dalším nádorovým onemocněním byly analyzovány v populacích sporadických nebo vysoce rizikových osob.

5 METODY

5.1 Soubor testovaných osob

Analyzované vzorky byly vybrány z kolekce vysoce rizikových pacientů s karcinomem prsu z Prahy a okolí, kteří byli mezi listopadem 2001 a květnem 2012 v naší laboratoři negativně testováni na přítomnost patogenních mutací v genech *BRCA1* a *BRCA2*, nebo *PALB2*. Pacienti splňovali kritéria pro genetické testování (Bartoňková et al. 2003). Celkem bylo do studie zahrnuto 325 pacientů, kteří poskytli informovaný souhlas s analýzou. Charakteristiky pacientů, včetně nádorových subtypů zohledňujících expresi estrogenního receptoru (ER), progesteronového receptoru (PR), Her2 receptoru a Ki-67, jak je popsáno v (Goldhirsch et al. 2011), jsou uvedeny v Tabulce 3. Pro identifikaci populačně specifických variant s ohledem na možné sekvenční varianty specifické pro dědičné onemocnění karcinom prsu jsme analyzovali 105 kontrolních vzorků, které pocházely od českých žen bez nádorového onemocnění, které byly vybrány na základě věku (medián věku 71 let; 95% CI 34 – 95 let), pocházející ze stejné geografické oblasti (Kleibl et al. 2008).

Tabulka 3: Klinické a histopatologické charakteristiky 325 vyšetřených vysoce rizikových pacientů s karcinomem prsu negativně testovaných na přítomnost mutací v genech *BRCA1*, *BRCA2* a *PALB2*.

	N	procento	Věk v době diagnózy ¹	
			medián	(rozpětí)
Všichni pacienti s karcinomem prsu	325	100.0	43.1	(24.4-81.7)
Ženy	314	96.6	42.8	(24.4-78.0)
<i>Menoaktivita v době diagnózy BC (N = 314; ~100%)</i>				
- Premenopauzální	235	74.8	36.9	(24.4-56.5)
- Postmenopauzální	69	22.0	60.0	(48.0-78.0)
- Neznámý status menoaktivity	10	3.2	51.7	(47.2-58.5)
<i>Lokalizace primárního(-ch) nádoru(-ů) (N = 314; ~100%)</i>				
- Unilaterální BC	281	89.5	41.3	(24.4-78.0)
- - a další primární nádor	22	-	48.6	(27.4-70.4)
- Bilaterální BC	25	8.0	46.0	(33.9-76.1)
- - a další primární nádor	1	-	52.2	-
- BC a karcinom ovaria	8	2.5	49.9	(21.5-60.9)
- - a další primární nádor	2	-	44.3	(30.0; 58.5)
Muži (všichni unilaterální BC)	11	3.4	69.6	(54.7-81.7)
- - a další primární nádor	2	-	64.1	(55.4; 72.8)
<i>Histologie karcinomu prsu (N = 325; ~100%)</i>				
- duktální	264	81.2	41.4	(24.4-81.7)
- lobulární	25	7.7	48.0	(31.2-69.9)
- medulární	13	4.0	50.5	(28.0-77.0)
- ostatní	7	2.2	47.0	(27.5-54.5)
- neznáma	16	4.9	45.0	(34.9-69.3)
<i>Subtypy karcinomu prsu (N = 325; ~100%)</i>				
- luminal A	90	27.7	48.8	(27.9-76.1)
- luminal B	117	36.0	37.8	(25.6-81.7)
- basal like	51	15.7	37.5	(26.5-77.0)
- HER-2	15	4.6	34.6	(24.4-69.5)
- neznámý	52	16.0	41.3	(25.5-78.8)

Poznámka: Charakteristiky karcinomu prsu odrážejí pozitivitu ER, PR, Her2 receptorů a Ki-67 následovně: Luminal A – pozitivní na ER a PR, negativní na HER-2 a Ki-67 <14%; Luminal B – pozitivní na ER a PR, pozitivní na HER-2 nebo/a Ki-67 >14%; Basal like – negativní na ER a PR, negativní na HER-2; HER-2 – negativní na estrogení a progesteronový receptor, pozitivní na HER-2.

¹Pro všechna data byl brán v potaz věk v době diagnózy prvního karcinomu prsu.

5.1.1 Rozšířený soubor pro analýzu variant genu FANCL

V rámci analýz jsme našli rekurentní variantu genu *FANCL*, způsobující posun čtecího rámce. Její další analýzu jsme provedli v rozšířeném souboru dalších 502 vysoce rizikových pacientů s karcinomem prsu, 693 pacientů se sporadickým karcinomem prsu a 664 nenádorových kontrol, které byly již dříve popsány (Mateju et al. 2010, Janatova et al. 2015).

5.1.2 Soubor pro analýzu varianty c.657del5 genu NBN

Delece c.657_661delACAAA v genu *NBN* (viz Kap. 2.2) byla identifikována u pacientky s karcinomem pankreatu, která byla chybně zařazena do NGS studie (viz kapitola 6.5). K ověření doposud nejasné asociace mezi nosičstvím této varianty a vznikem karcinomu pankreatu byl vyšetřen soubor pacientů s karcinomem pankreatu a kontrolních osob bez nádorového onemocnění. Analýza byla provedena u 241 neselektovaných, histopatologicky verifikovaných pacientů s ductálním adenokarcinomem pankreatu. Vzorky byly získány ve spolupráci se Státním zdravotním ústavem (152 vzorků – 59 žen a 93 mužů; medián věku při diagnóze: 63 let; rozmezí 40-80 let) a s Onkologickou klinikou Všeobecné fakultní nemocnice v Praze (89 vzorků – 49 žen a 40 mužů; medián věku při diagnóze: 64 let; rozmezí 38-84 let). Pokud byly k dispozici, byly shromážděny informace o rodinné anamnéze u nosičů hledané varianty. Rozšířený soubor zahrnoval i 915 osob kontrolní populace bez záznamů o nádorovém onemocnění. Všichni pacienti i kontroly byly českého původu.

5.2 Izolace genetického materiálu

Pro izolaci DNA a RNA bylo od každého jedince odebráno 10 ml periferní krve do zkumavek s EDTA.

5.2.1 Izolace DNA z leukocytů periferní krve

Genomová DNA byla izolována pomocí izolačního kitu Wizard Genomic DNA Purification System (Promega) podle návodu výrobce.

5.2.2 Izolace celkové RNA z periferní krve

Čerstvá nesrážlivá krev (3-5 ml) byla navrstvena na Histopaque 1070 (Sigma), pomocí centrifugace byly separovány leukocyty a následně byly promyty v chlazeném roztoku PBS. Celková buněčná RNA byla izolována pomocí fenol-chloroformové extrakce v přítomnosti guanidiniumizothiokyanátu (Chomczynski et al. 1987). Pokud nebylo možné provést izolaci RNA ihned po odběru krve, byl použit systém PAX gene (Qiagen) pro odběr a uchování stabilizované RNA. Do příslušného roztoku kitu byly odebrány 2 ml krve. Izolace RNA byla dále provedena izolačním kitem podle instrukcí výrobce.

5.2.3 Kontrola koncentrace a kvality izolovaného genetického materiálu

Koncentrace a čistota izolovaného genetického materiálu (DNA, RNA) byla stanovena na základě měření absorpance při vlnových délkách 260 a 280 nm na spektrofotometru NanoDrop 1000 (Thermo Fisher scientific). Integrita izolované DNA i RNA byla ověřena elektroforeticky (SubCell; BioRad) na 1% agarózovém gelu (UltraPure Agarose, Gibco) barveném pomocí GelRed (Biotium) ve standardním TBE pufru po dobu 50 min při 80 V. Před elektroforetickou separací RNA byla provedena její denaturace při 65°C. DNA byla následně uchovávána v 1,5ml zkumavkách při 4°C a skladována při -20°C, RNA byla skladována při teplotě -80°C.

5.3 Příprava sekvenačního panelu

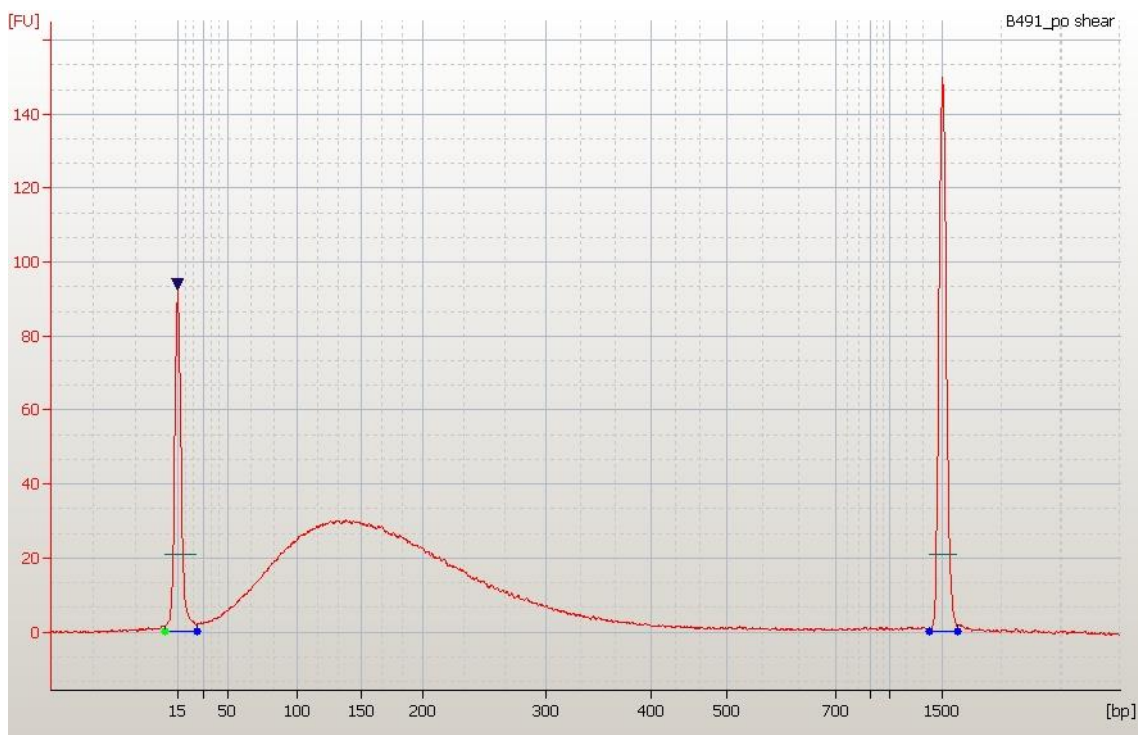
Vybrané geny pro sekvenování se sestávaly ze dvou překrývajících se skupin. První byla zastoupena 141 geny kódujícími proteiny, které se účastní reparace DNA a buněčné odpovědi na poškození DNA. Výběr těchto genů byl proveden na základě studia primárních literárních pramenů z databáze PubMed s využitím klíčových výrazů: DNA repair; („DNA repair“ AND „cancer“); „homologous recombination“ AND „DNA“ AND „repair“. Druhá skupina čítala 448 genů, které byly získány z databáze Phenopedia na webu HuGE Navigator za použití termínu „breast neoplasm“. Do výběru byly zařazeny geny s minimálně dvěma záznamy ((Yu et al. 2010); verze z února 2012)). Průnikem obou skupin vznikl konečný seznam 590 genů. Použitá SeqCap EZ Choice hybridizační knihovna (Leguen et al.) pro vycytání exonů a exon-intronových hranic cílových genů, pokrývala 96,8 % cílových bází. Tři geny nebyly pokryty próbami. Šest genů s nízkým sekvenačním pokrytím bylo vyloučeno z analýz. Sekvenování bylo tedy úspěšně provedeno na 581 genech (Příloha I – Seznam cílových genů).

5.4 Příprava NGS knihovny

Sekvenování probíhalo na platformě SOLiD (Kap. 3.3.3). Pro přípravu sekvenační knihovny byl použit 1 µg genomové DNA. Fragmentace DNA byla provedena ultrazvukem (Covaris S2, Covaris) s nastavením přístroje optimálním pro získání fragmentů genomové DNA o délkách v rozmezí 100-200bp (Tabulka 4, Obrázek 11). Byla provedena optimalizace parametrů Intenzita, Duty cycle a čas sonikace.

Tabulka 4. Nastavení parametrů sonikátoru Covaris S2 pro fragmentaci gDNA

Parametr	hodnota	jednotky
Počet cyklů	6	
Teplota lázně	5	°C
Limit teploty vodní lázně	8	°C
Režim	Frequency sweeping	
Funkce Water Quality Testing	off	
Duty Cycle	10	%
Intenzita	5	
Cycles per Burst	200	
Čas	60	s
Objem vzorku	50	µl



Obrázek 11: Elektroforeogram kapilární elektroforézy systému Bioanalyzer. Výstup popisuje délkovou distribuci fragmentů genomové DNA po ultrazvukové fragmentaci sonikátorem Covaris S2.

Fragmentovaná DNA byla následně podrobena třem editačním krokům:

- (i) Nejprve byly vytvořeny tupé konce dvouřetězcových DNA fragmentů. Takto upravená DNA byla fosforylována na 5' koncích s využitím Fast End Repair Enzyme mix kitu (Thermo Fisher Scientific). Vzorek byl purifikován pomocí SPRI (Solid Phase Reversible Immobilization) magnetických kuliček AMPure XP (Beckmann Coulter) v poměru 1,4x vůči objemu vzorku. (Kiiski et al.)
- (ii) Na konce upravených DNA fragmentů byly naligovány dvouřetězcové adaptorové DNA molekuly pro sekvenování na systému SOLiD (Obrázek 12). Tato úprava byla provedena pomocí Rapid DNA Ligation Kit (Thermo Fisher Scientific).



Obrázek 12: Sekvence adaptorů pro SOLiD sekvenování ligovaných na fragmentovanou gDNA při přípravě knihovny.

- (iii) Následná LM-PCR (Ligation-Mediated PCR) reakce amplifikující všechny molekuly DNA s naligovanými adaptory zajistí inkorporaci specifické desetinukleotidové sekvence (tzv. barcode) do molekuly vzniklého PCR produktu (Harakalova et al. 2011). Tato amplifikace probíhá pomocí primerů (Obrázek 13), jejichž část nasedá na adaptor a zbylá část přidává barcode sekvenci a dále sekvenci rozeznávanou při vlastním masivně paralelním sekvenování.

Pro amplifikaci byl použit Platinum PCR SuperMix (Thermo Fisher Scientific). LM-PCR probíhala za následujících podmínek:

Krok	Teplota	Čas	
Nick translace	72°C	20 min	8 cyklů
Denaturace	95°C	5 min	
Anelace	54°C	15 s	
Elongace	70°C	1 min	
Terminální elongace	70°C	4 min	
Chlazení	4°C	∞	

Solid_Ada_1: 5' - CTATGGGCAGTCGGTGAT - 3'

Adaptor P1

Solid_2_BCxx: 5' - CTGCCCCGGGTTCTCATTCTCT BBBBBBBBBB CTGCTGTACGGCCAAGGCG - 3'

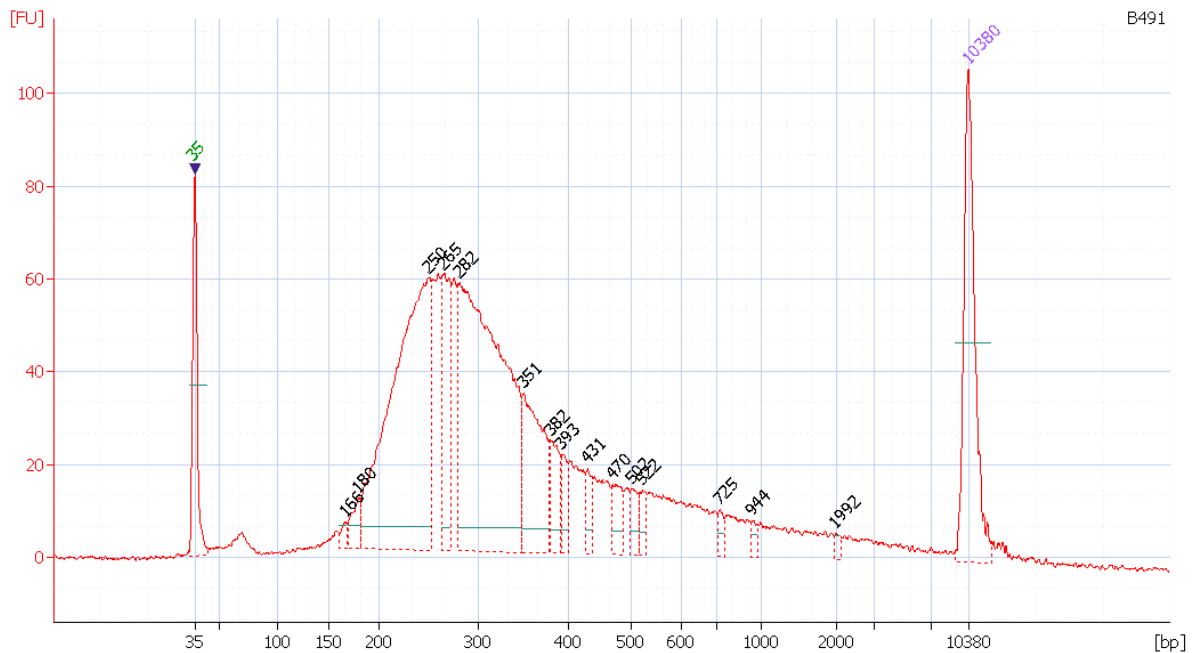
Adaptor P2

Barcode

Interní adaptor

Obrázek 13: Primery pro amplifikaci fragmentů opatřených adaptory. Ada2 primer obsahuje tzv. barcode sekvenci deseti nukleotidů unikátní pro daný vzorek (BBBBBBBBB, konkrétní pořadí nukleotidů lze nalézt v dokumentu SOLiD Fragment Library Barcoding Kit 1-96 Protocol na adrese www.lifetechnologies.com). Lze použít až 96 barcode primerů (v projektu bylo pro sekvenování v jednom běhu použito maximálně 60 vzorků. Ada1 primer je univerzální.

Vzniklé DNA knihovny byly purifikovány na kolonkách PureLink PCR Purification Kit (Thermo Fisher Scientific), kvantifikovány za použití fluorimetru Qubit (Life Technologies) a distribuce délek fragmentů byla ověřena na kapilární elektroforéze Bioanalyzer 2100 s kitem DNA 1000 (Agilent Technologies, Obrázek 14).



Obrázek 14: Elektroforeogram jedné z knihoven po přidání adaptorových sekvencí a LM-PCR, pomocí které byla přidána specifická barcode sekvence.

K obohacení o cílovou DNA (tzv. DNA enrichment, či sequence capture) byl využit na zakázku vytvořený set biotinylovaných hybridizačních prób (SeqCap EZ Choice Library, Roche). Jednalo se o hybridizaci probíhající v roztoku.

Pro enrichment bylo smícháno vždy 12 vzorků v ekvimolárním množství o celkové hmotnosti 1 μg , dále 5 μg lidské Cot-1 DNA pro zamezení nespecifické hybridizace. Do hybridizační reakce byly přidány jednořetězcové oligonukleotidy nasedající na koncové části vytvořených DNA fragmentů knihovny (tzv. barblocks), které obsahují adaptorové sekvence a sekvence barcodů (Obrázek 15). Pro tento účel bylo smícháno 12 příslušných P2 barblocků o molárním množství 1000 pmol a jeden univerzální P1 barblock taktéž v množství 1000 pmol. Takto připravená směs byla vysušena ve vakuovém rotačním koncentrátoru SpeedVac (Eppendorf), následně byla DNA rozpuštěna v hybridizačním pufru (SeqCap Hybridization and Wash Kits, Roche) a zdenaturována při teplotě 95°C po dobu 10 min.

Solid_P1_Block : 5' - CTATGGGCAGTCGGTGAT /3ddC/-3'
Adaptor P1

Solid_BarBlock_xx: 5' - CTGCCCCGGGTTTCCTCATTCTCT BBBBBBBBBBB CTGCTGTACGGCCAAGGCG /3ddC/-3'
Adaptor P2 Barcode Interní adaptor

Obrázek 15: Sekvence oligonukleotidů použitých při hybridizační reakci k zakrytí koncových sekvencí DNA molekul knihovny. Barbločky obsahují na 3' konci terminální dideoxycytydin.

Vlastní hybridizační reakce byla inkubována 64-72 hodin při teplotě 42°C podle protokolu výrobce (NimbleGen SeqCap EZ Library SR User's Guide, Version 4.2, Roche). DNA hybridizovaná s biotinylovanými próbami byla izolována pomocí magnetických kuliček s navázaným streptavidinem (M-270 Streptavidin Dynabeads, Thermo Fisher Scientific). Po eluci byla takto obohacená DNA použita jako templát pro osmicyklovou PCR pomocí primerů nasedajících na adaptorové sekvence (Obrázek 16) Pro amplifikaci byl použit Platinum PCR Amplification Mix (Life Technologies). PCR reakce probíhala za následujících podmínek:

Krok	Teplota	Čas	
Denaturace	95°C	5 min	8 cyklů
Denaturace	95°C	15 s	
Anelace	60°C	15 s	
Elongace	70°C	1 min	
Terminální elongace	70°C	5 min	
Chlazení	4°C	∞	

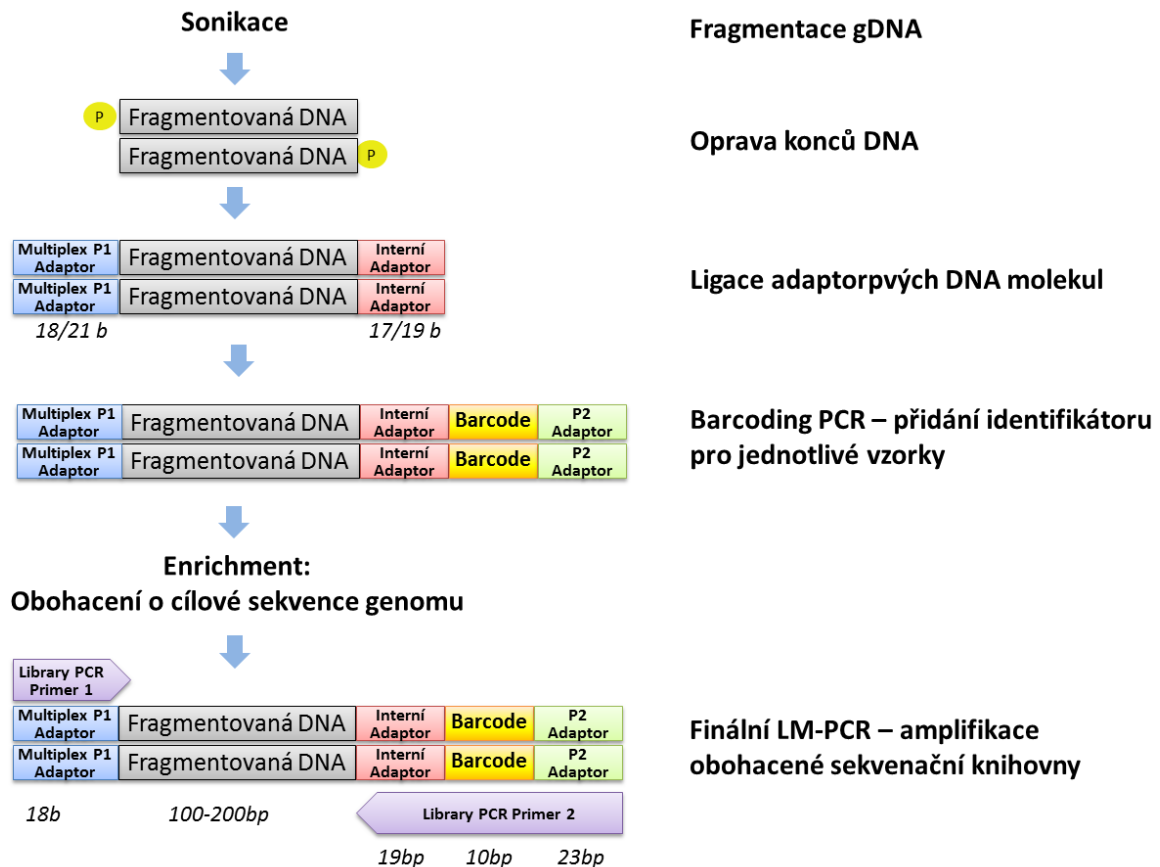
Library PCR primer_2: 5- CTGCCCCGGGTTTCCTCATTCT -3
Adaptor P2

Library PCR primer_1: 5- CCACTACGCCTCCGCTTTCCTCT CTATGGGCAGTCGGTGAT -3
Adaptor P1

Obrázek 16: Sekvence primerů použitých pro finální amplifikaci obohacené sekvenační DNA knihovny. Library PCR primer_1 přidává k DNA ampikonu koncovou sekvenci P1 adaptoru nezbytnou pro následnou emulzní PCR.

Výsledná DNA knihovna byla purifikována pomocí PureLink PCR Purification Kit (Thermo Fisher Scientific) s elucí do low TE pufru (10mM Tris , 0.1mM EDTA, pH 8.0) a kvantifikována pomocí fluorimetru Qubit (Life Technologies). Distribuce délek fragmentů v rámci vytvořené DNA knihovny byla vyšetřena na systému Bioanalyzer 2100 instrument, s čipem DNA high sensitivity (Agilent Technologies).

Proces úprav DNA při přípravě sekvenační knihovny je znázorněn na Obrázku 17.

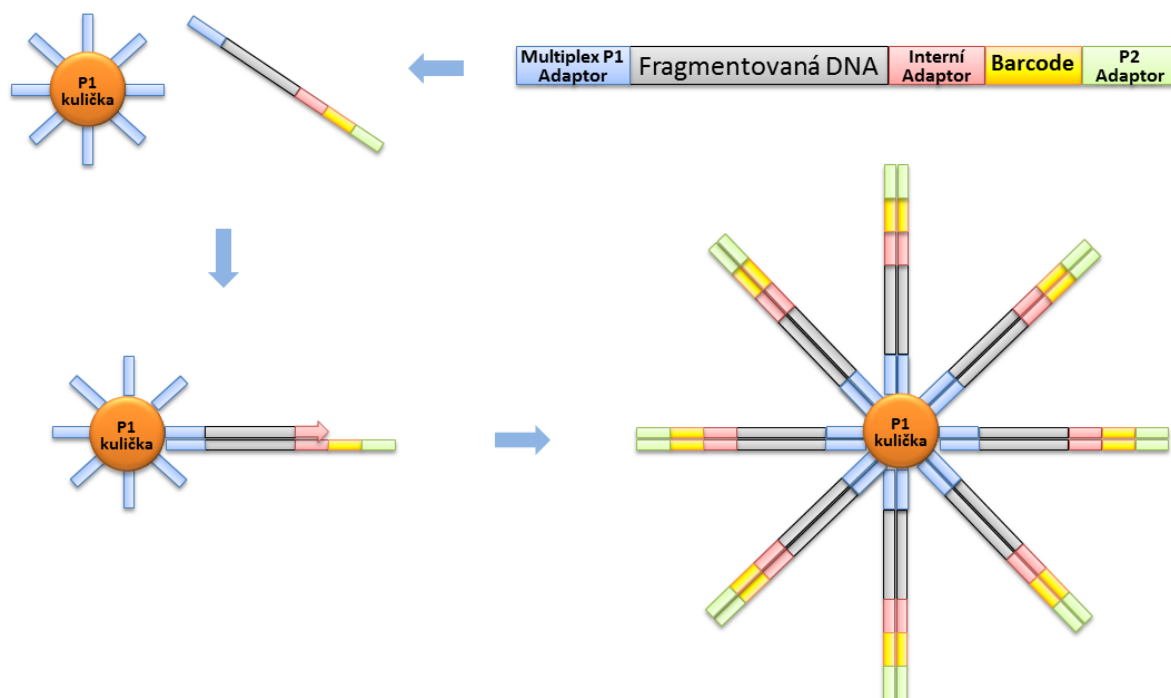


Obrázek 17: Schematické znázornění úprav DNA molekul připravovaných pro sekvenování na systému SOLiD4.

V dalším kroku byly jednotlivé DNA molekuly knihovny klonálně amplifikovány pomocí emulzní PCR (ePCR) na kuličkách pokrytých oligonukleotidy komplementárními k P1 adaptorové části DNA fragmentů (P1 kuličky, Obrázek 18). Pro emulzní PCR byly vytvořeny mikrokapénky vodné fáze ve fázi olejové pomocí přístroje SOLiD EZ Bead Emulsifier (Life Technologies). Do roztoku vodné fáze byly přidány paramagnetické P1 kuličky a DNA knihovny, které byly velmi naředěny, tak, aby jedna mikrokapénka obsahovala ideálně právě jeden DNA fragment. Statisticky se v 10-15% mikrokapének vyskytuje jedna templátová molekula, jedna P1 kulička a amplifikační mix. V průběhu ePCR byly DNA molekuly amplifikovány pomocí P1 adaptoru navázaného na kuličku a tím bylo dosaženo jejich fixaci na povrchu P1 kuličky. Mikrokapénky, které obsahují více, než jeden typ DNA molekuly knihovny dají vzniknout polyklonálním kuličkám, které produkují multifuorescenční signál v průběhu sekvenování a nemohou být analyzovány. Stejně tak jako kapénky bez templátu s P1 kuličkami bez navázané DNA jsou v následujícím kroku obohacování kuliček eliminovány. Knihovny amplifikované při ePCR byly naředěny na finální koncentraci 600 pM a všechny použité roztoky byly součástí kitu EZ Bead™ Templated Bead Preparation (Life Technologies). P1 kuličky bylo nutné před přidáním do vodné fáze promýt v Bead Block Solution za použití magnetického stojánku a sonikace na systému Covaris S2 s následujícím nastavením:

Parametr	Cyklus 1	Cyklus 2	jednotky
Počet cyklů	1	1	
Teplota lázně	5	5	°C
Limit teploty vodní lázně	8	8	°C
Režim	Frequency sweeping		
Funkce Water Quality Testing	off	off	
Duty Cycle	1	5	%
Intenzita	5	5	
Cycles per Burst	50	100	
Čas	5	60	s
Objem vzorku	200	200	μl

Emulzní PCR proběhla podle protokolu výrobce Templated bead preparation protocol for SOLiD V4 na cykleru SOLiD EZ Bead Amplifier (Life Technologies) v měřítku E20, či E80 (20, nebo 80 ml reakce) dle počtu zpracovávaných knihoven. Amplifikační protokol výrobce nezveřejňuje a je součástí operačního softwaru cykleru.



Obrázek 18: Schematické znázornění procesu klonální amplifikace DNA molekul sekvenční knihovny pomocí emulzní PCR.

Po proběhnutí reakce byla emulze rozbita pomocí butan-2-olu a P1 kuličky byly promyty a naředěny na koncentraci 750 000 až 1 250 000 kuliček/μl. Koncentrace byla měřena na spektrofotometru NanoDrop 1000 (Thermo Fischer scientific) po oddělení shluknutých kuliček pomocí sonikátoru Covaris S2 s následujícím programem:

Parametr	Cyklus 1	Cyklus 2	jednotky
Počet cyklů	1	1	
Teplota lázně	5	5	°C
Limit teploty vodní lázně	8	8	°C
Režim	Frequency sweeping		
Funkce Water Quality Testing	off	off	
Duty Cycle	2	5	%
Intenzita	6	9	
Cycles per Burst	100	100	
Čas	4	15	s
Objem vzorku	200	200	μl

80-90% takto zpracovaných kuliček nemá navázaný žádný DNA amplikon a vzorky byly proto následně obohaceny o P1 kuličky s úspěšně navázanými, klonálně naamplifikovanými molekulami knihovny pomocí systému SOLiD EZ Bead Enricher (Life Technologies) podle návodu výrobce Templated Bead Preparation Guide.

5.5 Sekvenování na systému SOLiD4

Pro imobilizaci obohacených kuliček na povrchu sekvenačního skla SOLiD, byly na P2 adaptorové konce DNA molekul knihovny připojeny „Bead linker“ molekuly (výrobce nezveřejňuje chemickou strukturu) podle protokolu výrobce Templated Bead Preparation Guide. Pro dosažení počtu přibližně 500 milionů obohacených kuliček na sekvenačním skle, je nezbytné deponovat na sklo nanášku o cca 700 milionech kuliček. Po změření koncentrace kuliček na spektrofotometru NanoDrop byla provedena inkubace v „Bead deposition buffer“ po dobu 2 hodin při teplotě 37°C pro ukotvení obohacených kuliček knihovny na sekvenační sklo. Na Sekvenačním systému SOLiD4 byl nastaven multiplexní běh, který přiřazuje jednotlivé barcodes příslušným vzorkům. Pro sekvenování byla použita chemie SOLiD Opti Fragment Library Sequencing Kit podle protokolu výrobce. Sekvenování probíhalo po dobu 5 dnů s výtěžkem kolem 400-500 milionů čtení o délce 50bp.

5.6 Bioinformatická analýza

Pro technickou kontrolu enrichmentu knihovny a masivně paralelního sekvenování byla použita v každé dvanáctici vzorků interně kontrolní DNA jedné osoby, která byla analýzována celkem třicet sedmkrát. Varianty, které se v těchto kontrolních vzorcích vyskytovaly mezi jednotlivými sekvenačními běhy s větší variabilitou (6,5 – 65 %) byly hodnoceny jako náhodné sekvenační chyby a byly vyřazeny z výsledného data setu. Zároveň bylo v prvním běhu sekvenováno 6 kontrolních vzorků

se známými mutacemi v predispozičních genech *BRCA1*, *BRCA2*, nebo *PALB2*, již dříve vyšetřovaných v naší laboratoři. Všechny kontrolní mutace u těchto vzorků byly nalezeny i pomocí NGS.

Výstupem ze sekvenátoru SOLiD jsou sekvence jednotlivých čtení („readů“) v základním formátu FASTQ. Sekvenační data byla zpracována pomocí Genome Analysis Toolkit (GATK (McKenna et al. 2010)) a Samtools. Tato struktura obsahuje několik samostatných kroků. Používá standardní sequence alignment/map (SAM) formát (Venter et al.), který reprezentuje jednotlivé čtení použitím production-quality SAM knihovny. Binární verze tohoto formátu je pak binary alignment/map (Hu et al.) formát (Hu et al.), který byl vytvořen pomocí SAMtools (0.1.8). Data byla namapována na referenční genomovou sekvenci (GRCh37/hg19) pomocí Novoalign (CS 1.01.08) s ohledem na nejlepší pozice „mapping quality“. Po tomto kroku, byly odstraněny sekvenční duplikáty (pomocí aplikace Picard). Následovala úprava/filtrace kvalitativně odpovídajících čtení (Indel – based reallignment, rekaliibrace kvality bazí, rekaliibrace kvality variant v databázi SNP, Genotype Phasing and Refinement). V posledním kroku byla provedena funkční anotace sekvenčních dat umožňující přiřazení základní biologické resp. fenotypové informace (ANNOVAR (Wang et al. 2010)).

Vzniklý data set obsahoval 2 855 706 variant, z nichž 561 434 se vyskytuje v exonech a přilehlých exon-intronových spojeních (+/- 2bp) cílových genů v rámci vyšetřené souboru 325 vysoce rizikových pacientů s karcinomem prsu a 105 kontrolních osob bez nádorového onemocnění. Pro stanovení hranice spolehlivosti dat na základě hodnoty sekvenační kvality bylo pomocí Sangerova sekvenování potvrzeno 221 variant u 155 pacientů a kontrol (popsáno v kapitole 5.9.2 a 5.9.5). Jako spolehlivá limitní hodnota sekvenační kvality byla určena hodnota 150× a hloubka čtení s minimální hodnotou 10×. Toto kritérium splňuje 12 221 variant lokalizovaných v cílové sekvenci. K další analýze byly vybrány varianty, které se vyskytují s frekvencí ≤1% v databázích ESP6500 (<http://evs.gs.washington.edu>) a G1000 (<http://www.1000genomes.org>) a zároveň se vyskytují v rámci studovaného souboru maximálně ve dvou vzorcích kontrolní populace (vyloučení častých populačně-specifických variant). Z další analýzy byly vyřazeny varianty, které se vyskytovaly pouze u kontrol. Uvedeným způsobem filtrování zůstalo k finálním analýzám výsledných 4 540 variant přítomných u 325 pacientů a 105 kontrol.

5.7 Analýza výstupních dat

5.7.1 *Prioritizace missense variant*

Missense varianty tvořily 2347 z celkových 4540 variant, tedy největší skupinu identifikovaných zárodečných alterací. Pro posouzení možného negativního vlivu varianty na strukturu a funkci

výsledného proteinového produktu přispívajícího k predispozici ke karcinomu prsu byla učiněna prioritizace za použití pěti predikčních algoritmů - SIFT (Kumar et al. 2009), PolyPhen (Adzhubei et al. 2010), LRT (Doniger et al. 2008), Mutation taster (Schwarz et al. 2010), PhyloP (Pollard et al. 2010) a dvou databází sdružujících genotypová data a odpovídající klinické charakteristiky (ClinVar, <http://www.ncbi.nlm.nih.gov/clinvar/> a HGMD, <http://www.hgmd.cf.ac.uk/>). Za varianty, které by mohly být asociovány s predispozicí ke karcinomu prsu, byly považovány ty, které byly i) označeny všemi predikčními algoritmy jako patogenní (nebo bez určení), nebo ii) označeny za asociované s onemocněním v databázi ClinVar nebo HGMD. Číselně hodnotící predikční algoritmy CADD (Kircher et al. 2014) score a GERP (Cooper et al. 2005) byly využity ke kontrole výše popsané prioritizace.

5.7.2 Statistická analýza

Rozdíly mezi analyzovanými skupinami a podskupinami byly evaluovány pomocí Pearson chi-kvadrátového testu, nebo Fisherovým testem, pokud byl počet očekávaných událostí nižší, než šest.

5.8 Následné analýzy potenciálně patogenních variant

Pro vyšetření nalezené varianty c.1094_1095insAATT (p.P365fs*) genu *FANCL*, způsobující posun čtecího rámce byl vyšetřen rozšířený soubor pacientů a kontrol popsáný v kapitole 5.1.1. DNA osob tohoto souboru byla nejprve vyšetřena pomocí HRM (High-Resolution Melting) analýzy a vzorky vykazující pozitivitu pomocí této metody byly dále confirmovány Sangerovým sekvenováním z nezávislé PCR amplifikace.

5.8.1 HRM analýza

HRM (High Resolution Melting) analýza variant c.1094_1095insAATT genu *FANCL* a c.657del5 genu *NBN* na rozšířených souborech osob byla prováděna s použitím Fast EvaGreen HRM Master Mix (Biotium). Pro amplifikaci a barvení produktu bylo použito následujícího složení reakční směsi:

Složka	Objem
EvaGreen HRM Master Mix	0,6 µl
25 ng gDNA	0,24 µl
Forward primer (40µM)	0,0175 µl
Reverse primer (40µM)	0,0175 µl
PCR grade H ₂ O	2,265 µl
Σ	3,14 µl

Pro PCR amplifikaci byly použity následující reakční podmínky:

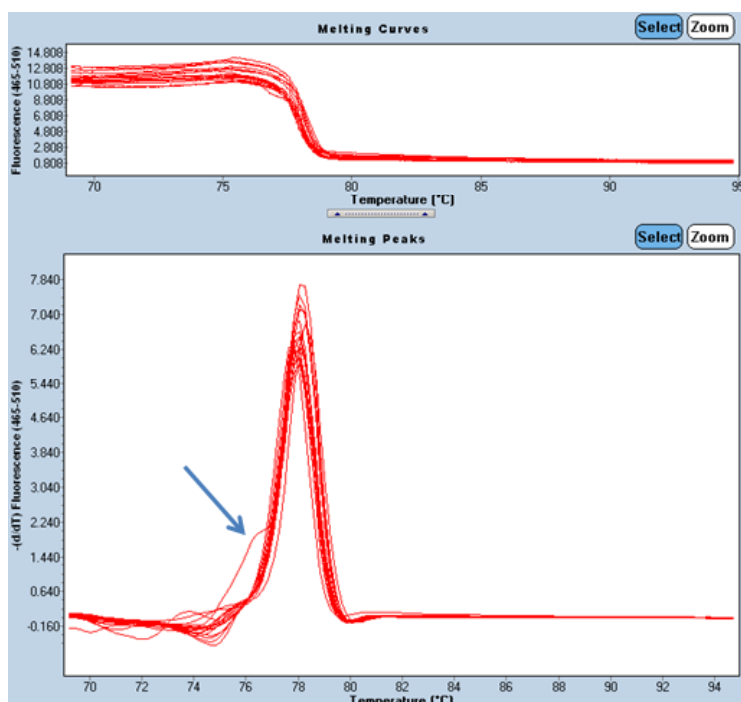
Krok	Teplota	Čas	
Denaturace	95°C	15 min	55 cyklů
Denaturace	95°C	25 s	
Anelace	60°C	30 s	
Elongace	72	20 s	

Křivka tání	Ramp rate °C/s			25°C celkem
	95°C	1 min	4,8	
	40°C	1min	2,5	
	70°C	1 s	1	
	95°C→70°C	kontinuálně	0,02	

Pro amplifikaci části genu *FANCL*, resp. *NBN* obsahující oblast s variantou byly použity primery o sekvenci a délce zobrazené níže:

Název	Sekvence 5'-3'	T _m (°C)	Délka amplikonu
FANCL_14f2	GAGTGTCCACAACTTGTAAAAATATG	54,0	112 bp
FANCL_14r2	TTCCAGCTCTTCACCGAAATG	54,9	
NBS_6f	CAGATAGTCACTCCGTTTACAA	58,2	293bp
NBS_6r	GTTGTTAACGTATTTTCATTTGGG	58,6	

PCR amplifikace s následnou denaturací amplifikátů byla prováděna na přístroji Light Cycler 480 (Leguen et al.). Pro identifikaci amplikonů obsahujících mutaci pomocí teplotního profilu křivky tání, byla provedena analýza pomocí Roche Light cycler 480 software release 1.5.0 SP4 (Obrázek 19).



Obrázek 19: Příklad analýzy křivek tání pomocí HRM. Šipkou je označena křivka mutovaného vzorku, ostatní křivky přísluší wild type amplikonům.

5.8.2 PCR amplifikace

PCR amplifikace s následným Sangerovým sekvenováním byla využita pro:

- (i) Stanovení hodnoty spolehlivosti sekvenační kvality výstupů ze sekvenování analyzátozem SOLiD, která je jedním z výsledných parametrů bioinformaticky zanalyzovaných dat z next-gen sekvenování. Pomocí konfirmací nalezených variant s různou hodnotou sekvenační kvality byla nalezena hranice (kvalita 150), nad kterou byly všechny varianty spolehlivě zjištěny i pomocí sangerova sekvenování.
- (ii) Konfirmaci nalezených variant, způsobujících posun čtecího rámce, nonsense variant a variant vyskytujících se v sestřihovém místě (+/- 1-2bp od exonu) pomocí Sangerova sekvenování
- (iii) Pro amplifikace cDNA odpovídající kódujících částí genů, u nichž byly nalezeny varianty potenciálně ovlivňující sestřih mRNA a následné zjištění vlivu na výsledný transkript.

Pro všechny tyto aplikace bylo použito stejné složení reakční směsi s použitím FastStart Taq DNA Polymerase (Roche):

10x PCR Reaction Buffer s 20 mM MgCl ₂	1 μl
100 ng DNA (gDNA, nebo cDNA)	0,5 μl
Forward primer (40μM)	0,1 μl
Reverse primer (40μM)	0,1 μl
PCR Grade Nucleotide Mix (každý 10mM)	0,2 μl
FastStart Taq DNA Polymerase (5 U/μl)	0,08 μl
PCR grade H ₂ O	8,02 μl
Σ	10 μl

Pro PCR amplifikaci s „touch-down“ byly použity následující reakční podmínky:

Krok	Teplota	Čas	
Denaturace	95°C	6 min	
Denaturace	95°C	30 s	15 cyklů
Anelace („touch down“)	64-57°C	1 min	
Elongace	72°C	1 min 45 s	
Denaturace	95°C	30 s	27 cyklů
Anelace	57°C	45 s	
Elongace	72°C	1 min 30 s	
Terminální elongace	72°C	7 min	
Chlazení	4°C	∞	

Pro vyšetření nalezených intronových variant v oblastech +/- 1-2 bp od konců exonů, které mohou mít potenciální vliv na sestřih výsledné mRNA, byla provedena analýza RNA, zahrnující amplifikaci

příslušného úseku cDNA, elektroforetickou separaci PCR produktu, v případě potřeby izolaci vybraného fragmentu DNA z gelu s PCR reamplifikací a Sangerovo sekvenování.

5.8.3 Příprava cDNA

Komplementární DNA (cDNA) byla připravena pomocí reverzní transkripce z celkové buněčné RNA s použitím enzymu SuperScript® III Reverse Transcriptase (Life Technologies). Před vlastní syntézou cDNA byly vzorky inkubovány 30min v 37°C s DNázou I (DNase I, RNase-free, Life Technologies) pro odstranění možné kontaminace DNA. Pro syntézu 20 µl cDNA byly použity 2 µg celkové RNA, 0,05 µl náhodných hexanukleotidů (50 ng/ µl; Roche) a 0,8 µl dNTP (každý 10 mM). Po hybridizaci hexanukleotidů (65°C - 10 min) a rychlém zchlazení (na ledu), bylo do reakční směsi přidáno 100U reverzní transkriptázy SuperScript III, 4 µl 5x RT pufru (Invitrogen) a 1 µl DTT (0,1 M). Reverzní transkripce byla provedena dle protokolu výrobce (25°C - 5 min; 50°C - 50 min; 70°C - 15 min). Získaná cDNA sloužila jako templát pro amplifikaci fragmentů oblastí genů, u kterých byly pomocí NGS nalezeny varianty s potenciálním dopadem na sestřih mRNA.

5.8.4 Izolace DNA amplifikátů z agarózového gelu

Elektroforetická separace byla prováděna na 1-3 % (dle délky analyzovaného fragmentu) agarózovém gelu v 1xTBE pufru při napětí 10V/cm gelu. Pro přibližné určení délky fragmentů byl použit velikostní standard GeneRuler DNA Ladder Mix (100-10000 bp, Life Technologies), pro vizualizaci na UV transluminátoru Gel Red (Biotium).

U vzorků, které po PCR amplifikaci obsahovaly více než dva amplifikáty a nebylo možné pomocí Sangerova sekvenování spolehlivě určit sekvenci obou, bylo nejprve přistoupeno k izolaci jednotlivých ampliconů z gelu a následné reamplifikaci, či přímému sekvenování vyizolované DNA příslušného elektroforetického pruhu. Fragment o požadované velikosti byl vyříznut z gelu skalpelem a DNA byla purifikována pomocí Zymoclean Gel DNA Recovery Kit (Zymo Research) podle originálního protokolu výrobce. Výsledná DNA kvantifikována pomocí spektrofotometru NanoDrop.

5.8.5 Sangerovo sekvenování

Před vlastním sekvenováním byly z proběhlých PCR reakcí odstraněny oligonukleotidy a deoxyribonukleotidtrifosfáty pomocí 15min inkubace s enzymovou směsí obsahující DNA

exonukleázu a alkalickou fosfatázu EXOSAP-IT (USB Corp.). Z každé PCR reakce byl použit 1 µl, ke kterému bylo přidáno 0,8 µl purifikační směsi a 1 µl ddH₂O. Po purifikaci byly vzorky inkubovány při 80°C pro inaktivaci enzymů. Pro následnou sekvenační reakci bylo ke vzorkům přidáno 0,2 µl příslušného sekvenačního primeru o koncentraci 40 µM. Dále byl přidán 1 µl 5x Big Dye Sequencing buffer a 1 µl Big Dye Terminator Ready Reaction Mix v.3.1 (Applied Biosystems) obsahující všechny komponenty potřebné pro sekvenační reakci (včetně fluorescenčně značených ddNTPs). Inkubace sekvenační reakce probíhala za následujících podmínek:

Krok	Teplota	Čas	
Denaturace	95°C	2 min	20 cyklů
Denaturace	95°C	5 s	
Anelace	60°C	10 s	
Elongace	72°C	4 min	
Chlazení	4°C	∞	

Výsledný produkt sekvenační reakce byl precipitován a rozpuštěn ve formamidu dle následujícího postupu. Do zkumavky s produkty sekvenační reakce bylo přidáno 1,3 µl Na-acetátu (3M), 30 µl 100% etanolu a 1,3 µl EDTA (0,125 M), směs byla krátce vortexována a centrifugována po dobu 20 min, při teplotě 4°C a 14 000 rpm. Supernatant byl odstraněn a k precipitátu bylo přidáno 55 µl 70% etanolu. Vzorek byl centrifugován za stejných podmínek. Po opětovném odstranění supernatantu byl precipitát sušen v otevřené zkumavce 2 min při 94°C a následně 15 min při pokojové teplotě. Precipitovaná DNA byla rozpuštěna v HiDi formamidu (Applied Biosystems), denaturována 2 min při 94°C a ihned umístěna do bločku vychlazeného na 4 °C. Takto připravené produkty sekvenační reakce byly analyzovány v automatickém sekvenátoru ABI Prism 3130 (Applied Biosystems) v kapiláře o délce 36 cm s polymerem POP-7 za denaturujících podmínek. Výsledky analýzy byly zpracovány pomocí programu Sequencing Analysis v.2.5.5 (Applied Biosystems) a v následném zpracování vizualizovány pomocí FinchTV programu (Geospiza).

5.8.6 Amplifikační a sekvenační primery

Primery pro confirmaci nalezených variant, způsobujících posun čtecího rámce, nonsense variant a variant vyskytujících se v sestřihovém místě (+/- 1-2bp od exonu) pomocí Sangerova sekvenování a stejně tak i primery pro amplifikace cDNA odpovídající kódujících částí genů, u nichž byly nalezeny varianty potenciálně ovlivňující sestřih mRNA, byly navrženy pomocí webové aplikace Primer Blast (<http://www.ncbi.nlm.nih.gov/tools/primer-blast/>; Příloha II)

6 VÝSLEDKY

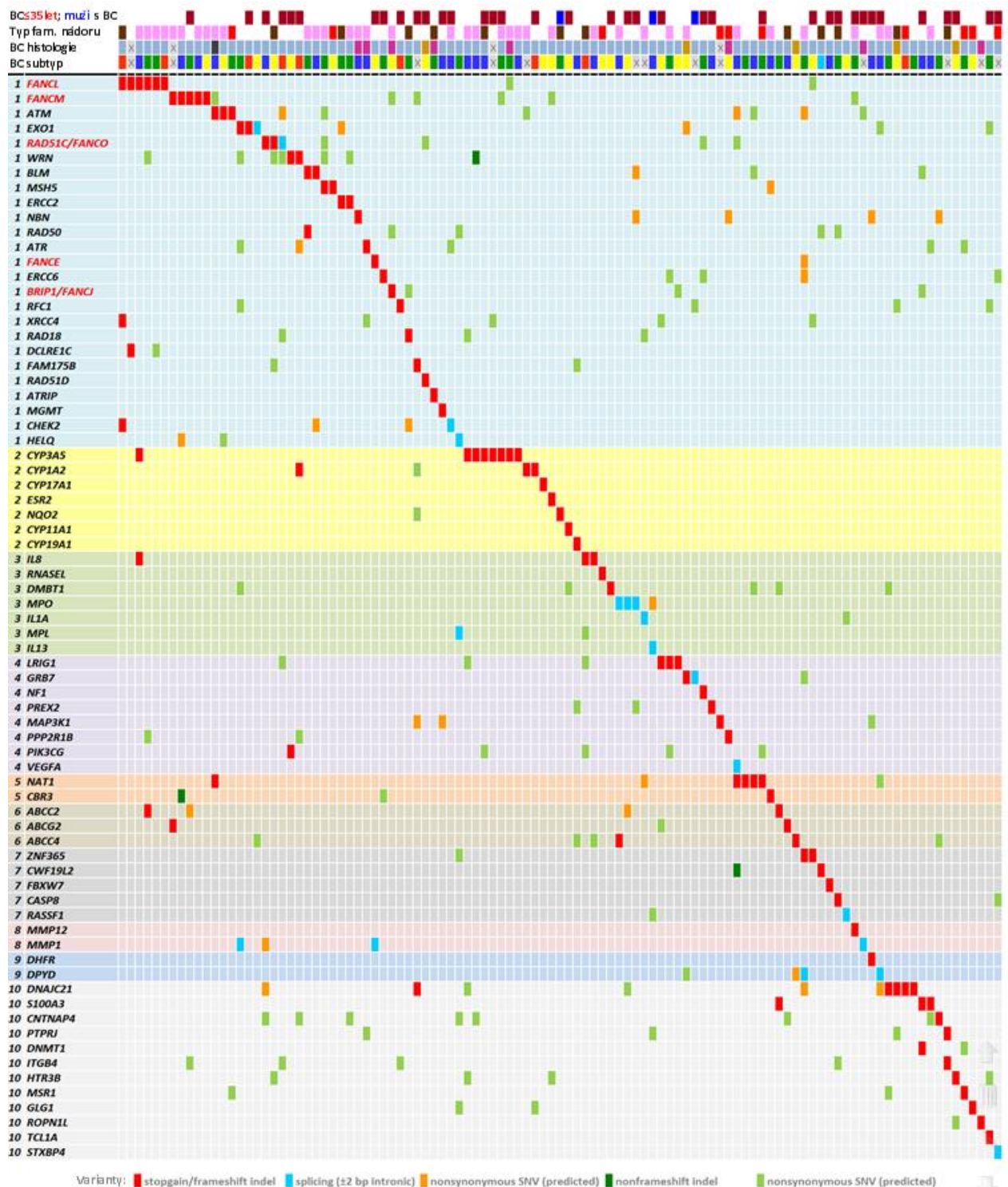
6.1 Identifikace dědičných variant ve vybraných genech u rizikových pacientů s karcinomem prsu negativně testovaných na přítomnost mutací *BRCA1/BRCA2/PALB2* (Lhota *et al.* 2016)

Hlavní náplní studie byla identifikace vrozených genetických variant u 325 vysoce rizikových nemocných s karcinomem prsu, kteří byli v předchozích letech negativně testováni na přítomnost mutací v genech *BRCA1*, *BRCA2*, nebo *PALB2*. K NGS analýze byl použit panel 581 genů popsáný v kapitole 5.3.

6.1.1 Trunkační varianty u pacientů negativně testovaných na přítomnost mutací v genech *BRCA1*, *BRCA2* a *PALB2*

Za použití filtračního postupu popsáného v kapitole 5.6 bylo identifikováno 4 540 variant s nízkou populační frekvencí z celkového počtu 2 855 706 variant, které byly výstupem sekvenování 430 analyzovaných vzorků (325 vysoce rizikových pacientů (Tabulka 3) s karcinomem prsu a 105 kontrolních nenádorových vzorků). Bylo dosaženo průměrného sekvenačního pokrytí 56,5 a 93 % sekvenované cílové DNA bylo pokryto minimálně 10x. Vyfiltrované varianty reprezentovaly 2 647 unikátních změn v sekvenci DNA, lokalizovaných v exonických oblastech a přilehlých +/-2 bázích a to ve 496/581 analyzovaných genech (85,4 %). Celkem bylo nalezeno 144 trunkačních mutací, variant způsobujících posun čtecího rámce a následné zkrácení proteinového produktu (nonsense, frame-shift inserce/delece, nebo alterace donorových a akceptorových sestřihových míst), které představovaly 89 unikátních variant, v 73 z 581 cílových genů (12,6 %). Trunkační varianty byly dvojnásobně častější u pacientů s karcinomem prsu v porovnání s kontrolní populací (16/105; 15.2%; $p = 0.009$).

V souboru 325 pacientů s karcinomem prsu bylo nalezeno 4 053 variant (2 647 unikátních událostí), včetně 127 trunkačních variant (87 unikátních), 34 inzercí/delecí neměnicích čtecí rámec (22 unikátních), 2 347 missense SNV (Single Nucleotide Variant, 1 599 unikátních) a 1 545 synonymních SNV (937 unikátních). Primárně byla pozornost zaměřena na trunkační mutace, které byly identifikovány u 105/325 (32,3 %) pacientů s karcinomem prsu (Tabulky 3 a 5). Devatenáct pacientů byli nosiči více, než jedné trunkační varianty (Obrázek 20).



Obrázek 20: Přehled variant v 73 genech (řádky) ovlivněných nejméně jednou trunkační variantou (nonsense, frameshift, nebo sestřihová variant), které byly nalezeny u 105 pacientů (sloupce). Patologické charakteristiky nádorů (histologie a subtypy) a vybrané klinické charakteristiky (pacientky do 35 let, karcinom prsu u muže, výskyt familiárního nádorového onemocnění) jsou znázorněny v prvních čtyřech řádcích (legenda barevného označení je znázorněna na Obrázku 21). Pacienti a geny jsou řazeni/y dle celkového množství nalezených variant, geny (geny pro Fanconiho anemii jsou vyznačeny červeně) jsou řazeny do skupin dle funkčního vztahu kódovaných proteinů následovně: skupina genů kódující protein účastníci se reparace DNA (1); syntéza steroidních hormonů a jejich signalizace; imunitní odpověď (3); signalizace membránovými receptory (4); metabolismus xenobiotik (5); membránový transport molekul (6); regulace buněčného cyklu a apoptózy (7); mezibuněčná komunikace (8); metabolismus nukleotidů (9); ostatní procesy (10).

Prioritizační *in silico* analýza, která hodnotila missense SNV, naznačila, že 365 z 1 599 unikátních jednonukleotidových záměn (22 %) může být potenciálně patogenní, z důvodu jejich asociace se známými fenotypy v databázích HGMD, ClinVar, nebo na základě softwarové predikce (Příloha III). Numerické hodnoty CADD skóre a GERP algoritmu potenciálně patogenních variant [23.7 (95%CI 23.0 – 24.5) respektive 4.6 (95%CI 4.4 – 4.7)] se značně lišily od hodnot predikovaných jako nepatogenní [17.0 (95%CI 16.5 – 17.4) respektive 2.1 (95%CI 1.9 – 2.3)].

6.1.2 Hereditární varianty v DNA reparačních genech

Celkem 36 z 89 (40 %) nalezených unikátních trunkačních variant, které ovlivňují 25 DNA reparačních genů, bylo nalezeno u 41 (12,6 %) pacientů s karcinomem prsu (Tabulka 5). Nejčastější alterace byly nalezeny u genů, které kódují proteiny účastníci se oprav dvouřetězcových zlomů DNA (DDSB) a meziřetězcových spojů (ICL). Mezi tyto pacienty patří 16 nosičů devíti unikátních trunkačních variant v pěti genech Fanconiho anemie (FA; *FANCL*, *FANCM*, *BRIP1/FANCI*, *RAD51C/FANCO*, a *FANCE*). Některé z nich mohou mít vliv na predispozici k dědičným formám karcinomu prsu a ovaria.

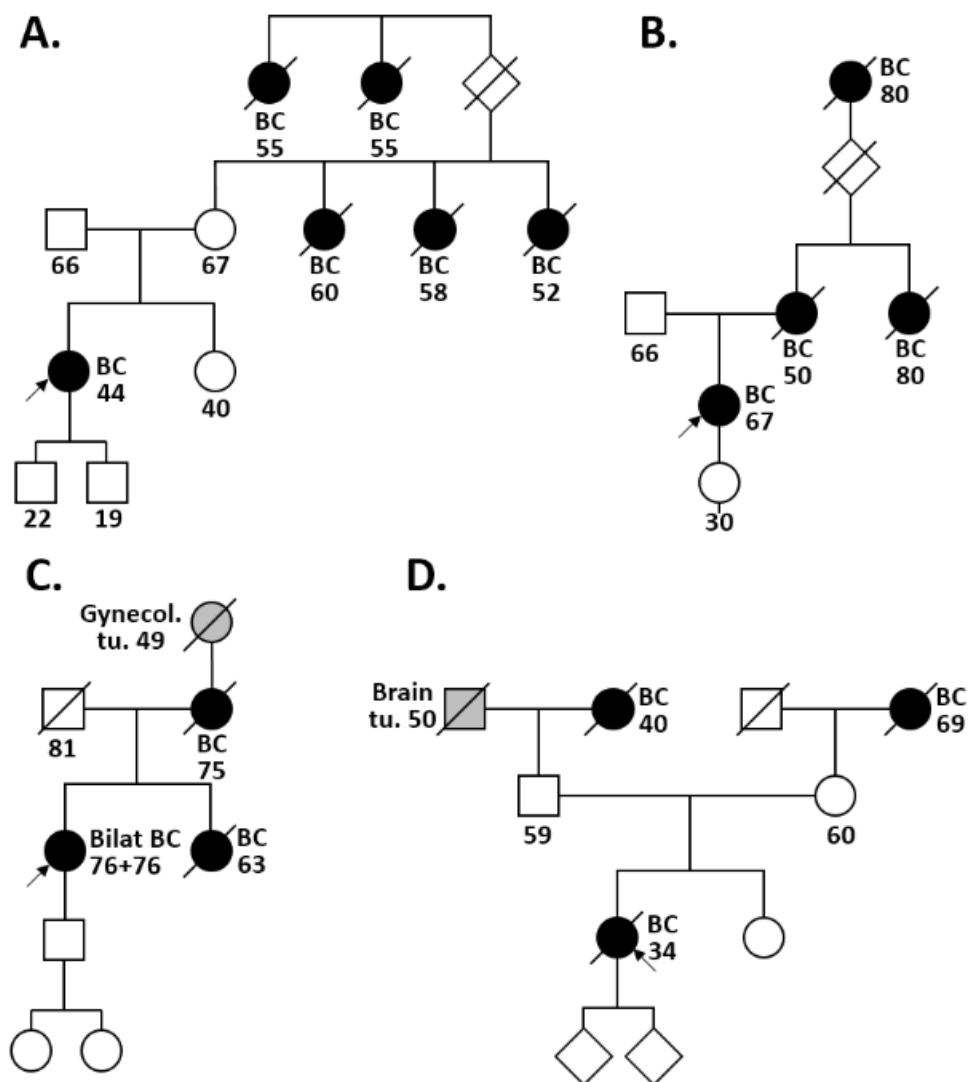
Inserční varianta genu *FANCE*, c.929dupC (p.P310fs), byla ve studovaném souboru nalezena jedenkrát u pacienta a jedenkrát u kontroly. Tato variant byla popsána skupinou (Akbari et al. 2011) jako kauzální pro karcinom jícnu. Dále byla nalezena varianta téhož genu, c.736_738del (p.246_246del) u jednoho analyzovaného pacienta. U této varianty nebyl popsán známý klinický význam a zároveň nezpůsobuje posun čtecího rámce.

V genu *FANCM* byly nalezeny čtyři varianty. Tři následující non-sense mutace: c.C5101T (p.Q1701X) u dvou pacientů, která byla v minulosti popsána jako pravděpodobně predispoziční varianta pro triple-negativní karcinom prsu (Kiiski et al. 2014); dále varianta c.C5791T (p.R1931X) u jednoho pacienta a varianta c.C1972T (p.R658X), která byla nalezena u jednoho pacienta a jedné kontrolní osoby. Zároveň jsme identifikovali jednu doposud nepopsanou variantu genu *FANCM* způsobující posun čtecího rámce, c.3979_3980del (p.1327_1327del), která se nachází v ERCC4-like C-terminální doméně, zodpovědné za vazbu větvených struktur DNA (Ciccia et al. 2008).

Nejčastější varianta identifikovaná pouze u pacientů (6/325) s karcinomem prsu (věk 36 – 76 let), byla čtyřnukleotidová inserce c.1096_1099dupATTA (p.P365fs*) v genu *FANCL*. Tato varianta je lokalizována v těsném sousedství RING-finger domény zodpovědné za autoubikvitinylaci (Gurtan et al. 2006) a způsobuje posun čtecího rámce, při němž vzniká proteinový produkt zakončený aberantní C-koncovou částí s dvanácti aminokyselinami, který je o tři aminokyselinové zbytky delší, než je wt

forma proteinu FANCL. Tato varianta byla dříve asociována s Fanconiho anémií typu L (Ali et al. 2009) a zvýšeným rizikem karcinomu jícnu (Akbari et al. 2011).

Čtyři z nosičů této *FANCL* varianty pocházeli z rodin s dědičným karcinomem prsu (Obrázek 24), další pacientka (která je současně nosičkou trunkační mutace v genu *CHEK2*) pocházela z rodiny s mnohočetným výskytem různých nádorových onemocnění a u jedné pacientky nebyla známa rodinná historie nádorových onemocnění. Za povšimnutí také stojí skutečnost, že čtyři z těchto šesti pacientek byly nosičkami další trunkační varianty zjištěné při NGS analýze.



Obrázek 24: Rodokmeny čtyř nosičů mutace *FANCL* c.1096_1099dupATTA (p.T367Nfs*13) z HBC rodin. Probandi s *FANCL* mutací jsou vyznačeni šipkou. BC – karcinom prsu, Brain tu. – tumor mozku, Gynecol. Tu. – gynekologický tumor, Bilat. - bilaterální

Tabulka 5: Seznam 36 truncačních variant (nonsense, frame-shift inserce/delece, nebo alterace sestřihových míst) v genech kódujících proteiny účastníci se oprav DNA a buněčné odpovědi na poškození DNA, které byly nalezeny u pacientů s karcinomem prsu (Pac) a u kontrolních vzorků (Ctrl). Geny jsou seřazeny podle celkové četnosti truncačních variant u příslušného genu.

Gen	HGVS značení	HGVS protein	Klasifikace	HGMD/ ClinVar	Pac (N)	Ctrl (N)	ExAC (mut/celkem) ^a	p ^b
FANCL	c.1096_1099dupATTA	p.T367Nfs*13	indel		6	0	232/65648	0.01
FANCM	c.1972C>T	p.R658X	nonsense	DM	1	1	7/66502	<0.001
	c.3979_3980delCA	p.Q1327Vfs*16	indel		1	0	0/66498	<0.001
	c.5101C>T	p.Q1701X	nonsense	DM?	2	0	95/66562	0.3
	c.5791C>T	p.R1931X	nonsense	DM	1	0	63/66622	0.6
ATM	c.3850delA	p.T1284Qfs*9	indel	DM	1	0	-	-
	c.7327C>T	p.R2443X	nonsense	DM/P	2	0	-	-
EXO1	c.1522dupT	p.C508Lfs*7	indel		1	0	-	-
	c.2358delG	p.L787Yfs*37	indel		1	0	-	-
	c.2212-1G>C	p.V738_K743del	splicing	DM	1	1	-	-
CHEK2	c.277delT	p.W93Gfs*17	indel		1	0	-	-
	c.444+1G>A	p.R148Vfs*6	splicing	DM	2	0	11/66720	<0.001
RAD51C	c.502A>T	p.R168X	nonsense		2	0	-	-
	c.905-2_1delAG	p.L301Gfs*42	splicing	DM	1	0	-	-
BLM	c.1642C>T	p.Q548X	nonsense	DM	2	0	21/66322	<0.001
ERCC2	c.230_231delTG	p.V77Afs*4	indel		1	0	-	-
	c.1703_1704delTT	p.F568Yfs*2	indel	DM	1	2	11/65444	0.01
MSH5	c.541C>T	p.R181X	nonsense		1	0	13/65882	0.02
	c.1900C>T	p.R634X	nonsense		1	0	-	-
WRN	c.604A>T	p.K202X	nonsense		1	0	-	-
	c.4216C>T	p.R1406X	nonsense		1	0	87/65788	0.9
ATR	c.5342T>A	p.L1781X	nonsense		1	0	-	-
ATRIP	c.827_828delAG	p.E276Gfs*2	indel		1	0	-	-
BRIP1	c.2392C>T	p.R798X	nonsense	DM/P	1	0	16/65688	0.04
DCLRE1C	c.1903dupA	p.S635Kfs*6	indel		1	0	-	-
ERCC6	c.3693C>G	p.Y1231X	nonsense		1	0	-	-
FAM175B	c.1084delC	p.Q362Kfs*19	indel		1	0	-	-
FANCE	c.929dupC	p.V311Sfs*2	indel	DM	1	1	-	-
HELQ	c.2677-1G>A	p.Q348Pfs*17	splicing		1	1	27/66528	0.15
MGMT	c.207_210dupACGT	p.S70Yfs*5	indel		1	0	-	-
NBN	c.657_661delACAAA	p.K219Nfs*16	indel		1	1	21/65324	0.09
RAD18	c.1430_1431insGCGG	p.T478Rfs*6	indel		1	0	-	-
RAD50	c.1093C>T	p.R365X	nonsense		1	0	-	-
RAD51D	c.355_358deldeITGTA	p.C119Wfs*16	indel		1	0	-	-
RFC1	c.2191delA	p.R731Gfs*7	indel		1	0	-	-
XRCC4	c.25delC	p.H9Tfs*8	indel		1	0	42/66632	0.4
Celkem					46	7		

Poznámka: Všechny vyobrazené varianty byly potvrzeny pomocí Sangerova sekvenování. Pro popis variant byla použita nomenklatura podle pravidel Human Genome Variation Society (HGVS). Varianty obsažené v databázích HGMD nebo ClinVar: onemocnění způsobující (patologické) mutace (DM), mutace pravděpodobně způsobující onemocnění (pravděpodobně patogenní; DM?), nebo patogenní (P). Obsáhlejší verze této tabulky (včetně missense variant predikovaných jako patogenní, referenčních transkriptů a frekvencí v databázích ExAC, ESP6500 a G1000) je přiložena jako Příloha III.

^a alelické frekvence z databáze ExAC v evropské populaci (mimo finskou), (mutované alely/wt alely)

^b Rozdíly mezi alelickou frekvencí v evropské (mimo finskou) populaci (ExAC) a v populaci studovaného souboru pacientů (Fisherův test).

Dalších 19 pacientů bylo nosiči 19 unikátních variant ovlivňujících geny, které se účastní drah reparace DDSB, včetně reparace homologní rekombinací (HR; *ATM*, *EXO1*, *WRN*, *BLM*, *DCLRE1C*, *FAM175B/ABRO1*, *HELQ*, *NBN*, *RAD18*, *RAD50*, *RAD51D*, *CHEK2*, a *RFC1*), ale i nehomologním spojováním konců (NHEJ; *XRCC4*). Osm truncačních variant, každá u jednoho pacienta, bylo identifikováno v genech, jejichž proteinové produkty se účastní DNA reparace jednořetězcových zlomů (*ATR*, *ATRIP*), mechanismu nucleotide excision repair (NER; *ERCC2*, *ERCC6*), mechanismu mismatch repair (MMR; *MSH5*) a přímého odstraňování alkylovaného guaninu (MGMT). Toto zařazení postižených genů nemusí popisovat úplnou funkci příslušných proteinových produktů, neboť se některé z těchto genů účastní několika mechanismů reparace DNA.

Čtyři varianty ovlivňující konsenzuální sestřihové místo byly identifikovány u pěti pacientů s karcinomem prsu. Dva z nich byli nosiči známé mutace v *CHEK2*. Současně byly nalezeny dvě dosud nepopsané varianty v genech kódujících DNA reparační protein *RAD51C* a helikázu *HELQ*, které způsobují posun čtecího rámce ve výsledném mRNA transkriptu (podrobněji popsáno v kapitole 6.4.1).

Celkem bylo nalezeno 10 unikátních missense variant v 59 DNA reparačních genech u 133 pacientů (34 z těchto variant u 56 pacientů se vyskytovalo v 15 genech, ve kterých byla zároveň nalezena alespoň jedna truncační varianta). Nejčastější a potenciálně nejvíce patogenní varianty u pacientů byly nalezeny v genech *ATM* (12 variant u 17 pacientů) a *CHEK2* (4 varianty u 13 pacientů). Prioritizované varianty v genu *CHEK2* zahrnovaly variantu c.470T>C (p.I157T), což byla jediná rekurentní varianta *CHEK2* identifikovaná u devíti pacientů a jedné kontroly. Mezi prioritizovanými variantami byla také identifikována jedna patogenní varianta v genu *BRCA1* (p.C39R; IARC třída 5), *TP53* (p.G245S; Align GVGD C55), a *CDH1* (p.T340A) u tří nepříbuzných, mladých pacientek s karcinomem prsu (s věkem v době diagnózy 28, 35 respektive 33 let) bez přítomnosti jiných truncačních variant.

V rámci vyšetřovaného souboru byly nalezeny dvě truncační (p.V77Afs*4 u 1 pacienta, p.F568Yfs*2 u 1 pacienta a 2 kontrolních osob) a 2 missense varianty (u dvou pacientů: p.Asp423Asn, p.Val611Ala) v genu *ERCC2*. Rozšířená analýza variant *ERCC2* genu je popsána v kapitole 6.5.

6.1.3 Hereditární varianty v genech neúčastnících se DNA reparace

Zbývajících 81 truncačních variant (reprezentujících 53 unikátních mutací) ve 48 genech, které se neúčastní reparace DNA, bylo nalezeno u 74 pacientů s karcinomem prsu (Tabulka 6).

Tabulka 6: Seznam 53 trunkačních variant (nonsense, frame-shift inzerce/delece, nebo alterace sestřihových míst) v genech kódujících proteiny, které se neúčastní oprav DNA, jež byly nalezeny u pacientů s karcinomem prsu (Pac) a u kontrolních vzorků (Ctrl). Geny jsou seřazeny podle celkové četnosti trunkačních variant u příslušného genu. Příslušnost genu do skupin popsanych v Obrázku 20 je vyznačena ve sloupci Sk.

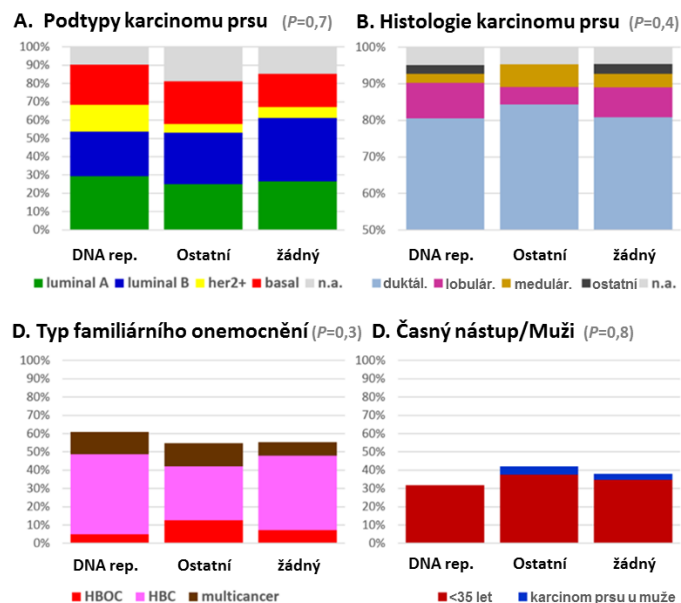
Gen	Sk.	HGVS značení	HGVS protein	Klasifikace	HGMD/ ClinVar	Pac (N)	Ctrl (N)	ExAC (mut/celkem) ^a	P ^b
CYP3A5	2	c.92dupG	p.L32Tfs*3	indel		7	1	732/66688	0.9
	2	c.246dupG	p.A83Gfs*40	indel		1	0	21/66728	0.09
CYP1A2	5	c.816T>A	p.Y272X	nonsense	FTV	3	0	16/65958	<0.001
CYP11A1	2	c.835delA	p.I279Yfs*10	indel	DM	1	0	2/66694	<0.001
CYP17A1	2	c.1072C>T	p.R358X	nonsense	DM	1	0	-	-
CYP19A1	2	c.1058dupT	p.L353Ffs*10	indel		1	0	1/60606	<0.001
ESR2	2	c.76G>T	p.E26X	nonsense		1	0	1/66734	<0.001
NQO2	2	c.628C>T	p.Q210X	nonsense		1	0	1/66386	<0.001
IL8	3	c.91G>T	p.E31X	nonsense		3	2	104/66426	0.05
DMBT1	3	c.2227delC	p.Q743Rfs*4	indel		1	0	-	-
IL13	3	c.174+2delT	p.(?)	splicing		1	0	-	-
IL1A	3	c.319+2T>C	p.(?)	splicing		1	0	-	-
MPL	3	c.79+2T>A	p.(?)	splicing	DM	1	0	114/66230	0.9
MPO	3	c.2031-2A>C	p.R677Wfs*73	splicing	DM	3	1	470/66434	0.5
RNASEL	3	c.793G>T	p.E265X	nonsense	DM/P	1	1	381/66212	0.2
LRIG1	4	c.3149_3150delCG	p.A1050Gfs*17	indel		3	0	102/66704	0.04
GRB7	4	c.862C>T	p.Q288X	nonsense		1	0	2/48666	<0.001
	4	c.801+1G>C	p.(?)	splicing		1	0	-	-
MAP3K1	4	c.4151dupT	p.L1384Lfs*36	indel		1	0	-	-
NF1	4	c.5690delG	p.G1897Vfs*28	indel		1	0	-	-
PIK3CG	4	c.41_42delAG	p.E14Gfs*147	indel		1	0	5/62474	<0.001
PPP2R1B	4	c.342_343delTG	p.V115Cfs*3	indel		1	0	81/66082	0.8
PREX2	4	c.3210_3213delAACA	p.D1072Vfs*17	indel		1	0	-	-
VEGFA	4	c.1085+2T>C	p.(?)	splicing		1	0	15/66648	0.03
NAT1	5	c.559C>T	p.R187X	nonsense	FP	5	1	252/66632	0.1
CBR3	5	c.533delA	p.D178Afs*46	indel		1	0	102/66716	1
ABCC2	6	c.3196C>T	p.R1066X	nonsense	DM/P	2	0	35/66738	<0.001
ABCC4	6	c.2468dupA	p.N823Kfs*12	indel		1	1	26/66634	
	6	c.1150C>T	p.R384X	nonsense		1	0	-	-
ABCG2	6	c.706C>T	p.R236X	nonsense	FP	1	0	24/66634	0.1
	6	c.736C>T	p.R246X	nonsense	FP/P	1	0	5/66692	<0.001
ZNF365	7	c.1065G>A	p.W355X	nonsense		2	0	4/66740	<0.001
CASP8	7	c.106delG	p.E36Nfs*7	indel		1	0	-	-
CWF19L2	7	c.1605delA	p.K535Nfs*4	indel		1	0	-	-
FBXW7	7	c.310delC	p.H104Mfs*389	indel		1	0	-	-
RASSF1	7	c.888+1G>A	p.V258Gfs*7	splicing		1	0	1/64856	<0.001
MMP1	8	c.105+2T>C	p.Q35Vfs*11	splicing	FTV	3	0	114/66230	0.08
MMP12	8	c.327C>T	p.W109X	nonsense		1	0	0/65722	<0.001
DPYD	9	c.1905+1G>A	p.D581_N635del	splicing	DM	2	1	389/66688	0.36
DHFR	9	c.95delT	p.F32Sfs*7	indel		1	0	-	-
DNAJC21	10	c.1503delA	p.K501Nfs*10	indel		3	0	33/66560	<0.001
	10	c.1629delT	p.F543Lfs*4	indel		2	0	5/11578	0.01
S100A3	10	c.208delG	p.V70Wfs*83	indel		3	1	291/66718	0.9
CNTNAP4	10	c.3913G>T	p.E1305X	nonsense		1	0	3/56224	<0.001
DNMT1	10	c.1035dupC	p.K346Qfs*35	indel		1	0	-	-
GLG1	10	c.3520C>T	p.R1174X	nonsense		1	0	-	-
HTR3B	10	c.871C>T	p.Q291X	nonsense		1	0	-	-
ITGB4	10	c.665delG	p.G222Efs*60	indel		1	0	-	-
MSR1	10	c.569delT	p.L190Cfs*5	indel		1	0	-	-
PTPRJ	10	c.1191T>A	p.Y397X	nonsense		1	0	-	-
ROPN1L	10	c.135T>A	p.Y45X	nonsense		1	1	126/66680	0.8
STXBP4	10	c.181-1G>A	p.K60Vfs*28	splicing		1	0	3/66388	<0.001
TCL1A	10	c.253C>T	p.R85X	nonsense		1	0	-	-
Celkem						81	10		

Za účelem identifikování možných poškození drah, které mohou přispívat k nádorové predispozici, jsme rozdělili postižené geny do devíti skupin (skupiny 2-9 v Tabulce 6 a Obrázku 20) sdružujících funkčně příbuzné proteiny. Dvanáct genů (skupina 10) nelze přiřadit do společné funkční skupiny. Šestnáct nosičů (5 % všech pacientů) devíti různých trunkačních alterací bylo identifikováno ve skupině 2, která obsahuje geny účastnící se metabolismu steroidních hormonů a jejich.

Dále bylo detekováno 250 unikátních, prioritizovaných, potenciálně patogenních missense variant ve 150 genech u 213 pacientů. Nejčastější prioritizované SNV v genech neúčastnících se DNA reparace postihovaly APC. Devět missense variant bylo nalezeno u osmi nosičů s časným nástupem onemocnění v podobě duktálního karcinomu prsu (medián věku: 29 let, rozpětí 27 – 48 let). Pozitivní rodinná historie nádorového onemocnění tlustého střeva byla dokumentována u 4 z 8 nosičů.

6.1.4 Charakteristiky onemocnění a rodinné anamnézy nosičů trunkačních variant

Nebyly nalezeny významné rozdíly v charakteristice pacientů (rodinná historie nádorových onemocnění a nástup nemoci) a nádorů (podtypy karcinomu prsu a histologické vlastnosti) mezi nosiči trunkačních variant v DNA reparačních genech a ostatních genech a pacienty bez trunkační varianty v analyzovaných genech (Obrázek 21).



Obrázek 21: Patologické charakteristiky nádorů a klinické charakteristiky 325 analyzovaných pacientů s karcinomem prsu, zařazených do skupin podle přítomnosti trunkační varianty v některém z DNA reparačních genů (41 pacientů), variant v pouze některém z ostatních genů (64 pacientů) a nepřítomnosti žádné trunkační varianty (220 pacientů). Znázorněné p hodnoty vyznačují nesignifikantní rozdíly mezi vyobrazenými charakteristikami mezi analyzovanými podskupinami. Podtypy a histologické vlastnosti karcinomu prsu dle popisu v kapitole 1.2. HBOC: pacienti z rodin s výskytem hereditárního karcinomu prsu a ovaria; HBC: pacienti z rodin s výskytem hereditárního karcinomu prsu; multicancer: pacienti z rodin s mnohočetným výskytem nádorových onemocnění.

Trunkační varianty byly nejčastější u Her-2 pozitivních nádorů (9/15; 60 %) a méně časté byly u subtypu luminal B (28/110; 25 %). Byl zaznamenán mírně zvýšený podíl nosičů trunkačních variant u pacientů s podtypy karcinomu prsu spojenými s horší prognózou (bazální a Her-2 pozitivní subtypy; 33/68 pacientů – 49 %) oproti subtypům s lepší prognózou (podtypy luminal A a luminal B; 57/207 pacientů – 28 %; p=0,1).

6.1.5 Nalezené varianty ovlivňující sestřih pre-mRNA

Skupina trunkačních variant zahrnovala 20 variant ovlivňujících sestřih pre-mRNA (sestřihové varianty; 14 unikátních, z nichž každá ovlivňovala jeden gen) v pozicích +/- 2 bp okolo exonů. Vliv těchto sestřihových variant byl zkoumán na úrovni mRNA (dostupné z periferní krve osmi pacientů). Sedm z osmi analyzovaných vzorků se sestřihovou variantou vykazovalo posun čtecího rámce a aberantní terminaci translace proteinového produktu. Dvě varianty v genech *DPYD* a *CHEK2* byly již dříve popsány jako sestřih alterující varianty.

6.1.5.1 Sestřihové varianty v DNA reparačních genech

Mutace c.444+1G>A (r.[444+1_444+4ins; 444+1g>a], p.R148Vfs*6, NM_001005735) v sestřihovém místě genu **CHEK2** má za následek 4-bp inzerci z důvodu využití alternativního donorového sestřihového místa v intronu 2. Tato, v minulosti popsaná (Dong et al. 2003), mutace vytváří předčasný terminační kodon v exonu 3 a eliminuje část FHA domény a celou kinázovou aktivační doménu výsledného CHEK2 proteinu. Popsaná sestřihová varianta se v rámci vyšetřovaného souboru vyskytovala u dvou pacientů.

V exonu 11 genu **EXO1** byla nalezena záměna G>C na pozici -1 a způsobovala 18 nukleotidů dlouhou in-frame delecí ve výsledném transkriptu (c.2212-1G>C, r.2212_2229del, p.V738_K743del, NM_130398.3, Obrázek 22), neboť jako nový AG motiv 3' sestřihového místa jsou rozeznávány nukleotidy 17 a 18 z exonu 12. Tato varianta byla nalezena u jednoho pacienta a jedné kontrolní osoby.

V genu **HELQ**, jednonukleotidová záměna G>A na 3' konci intronu, způsobuje frame-shift delecí sekvence TTTAG na 5' konci exonu 7 a rozpoznání nového sestřihového motivu AG. Výsledný transkript obsahuje sekvenci 16 odlišných aminokyselin, zakončenou předčasným terminačním kodonem (c.2677-1G>A, r.2677_2681del, p.Q348Pfs*17, NM_133636.3, Obrázek 22). Varianta byla nalezena u jedné kontroly a jednoho pacienta v rámci vyšetřovaného souboru.

Varianta IVS6-2_1delAG genu **RAD51C** způsobuje frame-shift výpadek celého exonu 7 ve výsledném mRNA transkriptu (c.905-2_1delAG, r.905_965del, p.L301Gfs*42, NM_058216.1). Transkript potom obsahuje nový terminační kodon po 41 tripletech kódujících odlišné aminokyseliny oproti divoké formě RAD51C. Tato variant byla nalezena u jednoho pacienta v rámci vyšetřované populace.



Obrázek 22: Charakterizace sestřihových variant nalezených v DNA reparačních genech (EXO1, HELQ a RAD51C). Schémata ukazují sekvenční chromatogramy z gDNA (A); elektroforeogramy PCR produktů amplifikovaných z cDNA (B); porovnání sekvenčních chromatogramů wild-type (C) a mutantní (D) cDNA; schematické zobrazení poškozené části aberantního transkriptu (červeným písmem, E), deletované části transkriptu maturované mRNA (žlutě), a oligonukleotidové primery použité pro analýzu cDNA (černé šipky). Schéma F ukazuje pozici deletovaných/zkrácených částí výsledného proteinu.

6.1.5.2 Sestřihové varianty v genech neúčastnících se reparace DNA

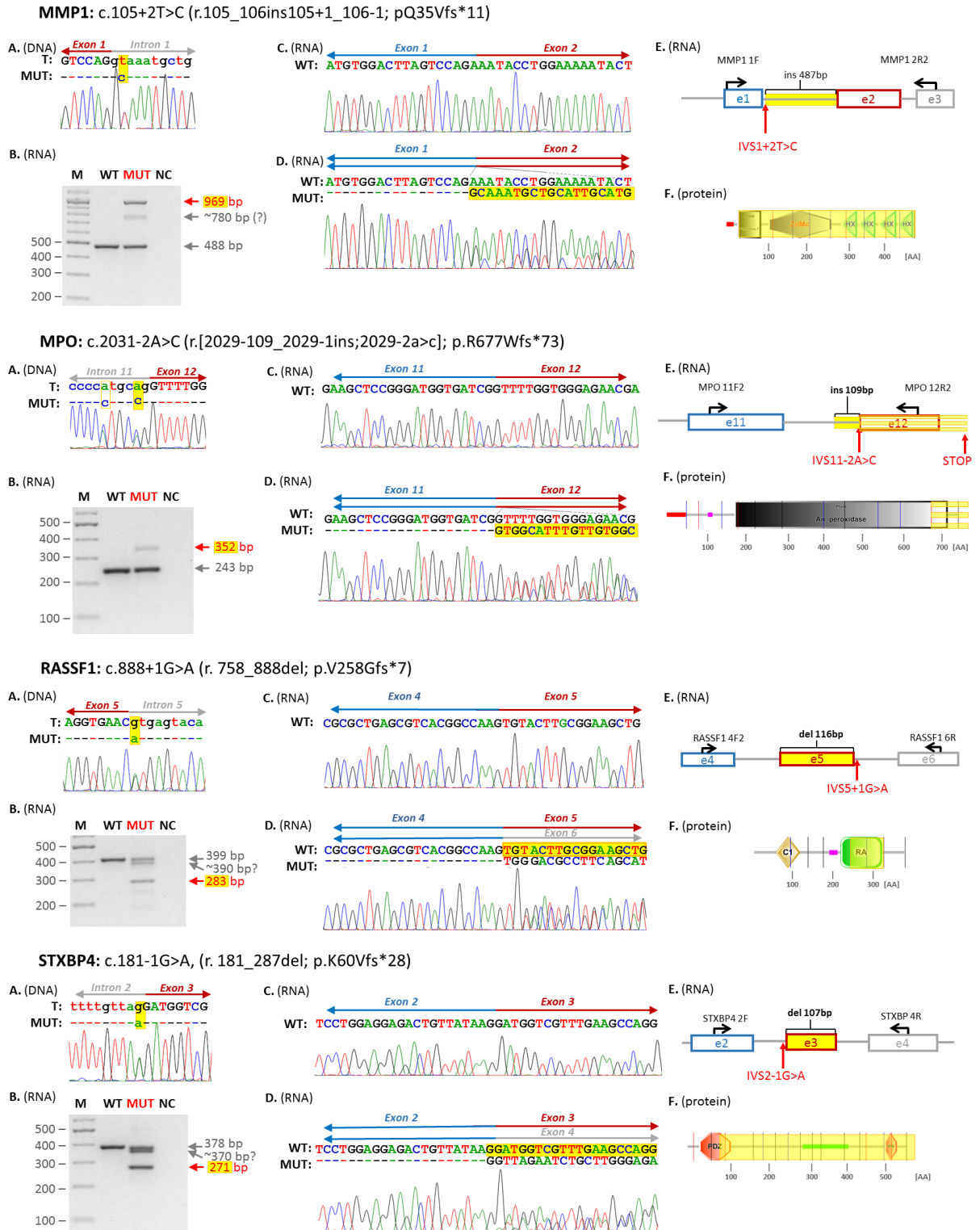
Jedna z námi nalezených variant genu **DPYD** byla dobře popsána záměna G>A v akceptorovém místě sestřihu pro exon 14, způsobující in-frame delecí 55 aminokyselin (c.1905+1G>A, r.1741_1905del, p.D581_N635del, NM_000110.3). Tato varianta vykazuje výrazně sníženou enzymatickou aktivitu výsledného translatovaného proteinu a prodlouženou clearance 5-fluorouracilu, která vede k těžké toxicitě a případné smrti (Meinsma et al. 1995) (van Kuilenburg et al. 2012). Tato varianta se vyskytovala u dvou pacientů a jedné kontrolní osoby v rámci vyšetřovaného souboru.

Záměna c.105+2T>C v genu **MMP1** způsobuje zachování celého intronu 1 ve výsledném sestřiženém mRNA transkriptu. Může tak docházet k posunu čtecího rámce a vzniku předčasného terminačního kodonu po 10 odlišných aminokyselinách oproti divoké formě MMP1 (r.105_106ins105+1_106-1, p.Q35Vfs*11, NM_002421.3, Obrázek 23). Tato varianta byla identifikována ve vzorcích dvou pacientů vyšetřovaného souboru.

Jednonukleotidová záměna A>C na pozici -2 exonu 11 v genu **MPO** způsobuje vznik transkriptu s inzercí 109 nukleotidů sekvence intronu (c.2031-2A>C, r.[2029-109_2029-1ins;2029-2a>c], p.R677Wfs*73, NM_000250.1, Obrázek 23). Tímto může vzniknout proteinový produkt delší o 4 aminokyseliny a rozdílnými 72 aminokyselinami C-koncové části oproti divoké formě MPO. Tato sestřihová varianta byla nalezena u dvou pacientů a jedné kontrolní osoby.

V genu **RASSF1** způsobuje záměna G>A na 5' konci intronu 5 frame-shift delecí celého exonu 5 ve výsledném mRNA produktu (c.888+1G>A, r. 758_888del, p.V258Gfs*7, NM_170714.1, Obrázek 23). Příslušný transkript má po 6 aminokyselinách, které jsou odlišné od divoké formy, zařazen terminační kodon. Tato varianta byla nalezena u jednoho pacienta.

Varianta IVS-1G>A, ovlivňující sestřihové místo, genu **STXBP4**, způsobuje frame-shift delecí celého exonu 3 a do potenciálního výsledného proteinu zavádí 28 nových aminokyselin, následovaných terminačním kodonem (c.181-1G>A, r. 181_287del, p.K60Vfs*28, NM_178509.5, Obrázek 23). Výsledný transkript kóduje 218 aminokyselin krátký protein, ve srovnání s 553 aminokyselin dlouhou divokou formou STXBP4. Tato varianta byla identifikována u jednoho pacienta.



Obrázek 23: Charakterizace sestřihových variant nalezených v genech neúčastnících se reparace DNA (MMP1, MPO, RASSF1 a STXBP4). Schémata ukazují sekvenční chromatogramy z gDNA (A); elektroforeogramy PCR produktů amplifikovaných z cDNA (B); porovnání sekvenčních chromatogramů wild-type (C) a mutantní (D) cDNA; schematické zobrazení postižené části aberantního transkriptu (červeným písmem, E), deletované části transkriptu maturované mRNA (žlutě), a oligonukleotidové primery použité pro analýzu cDNA (černé šipky). Schéma F ukazuje pozici deletovaných/zkrácených částí výsledného proteinu.

6.2 Rozšířená analýza výskytu varianty *FANCL* c.1096_1099dupATTA (Zemankova et al. 2016)

V předchozí práci (Kap. 6.1) jsme identifikovali rekurentní variantu v genu *FANCL* u vysoce rizikových pacientek s karcinomem prsu (Obrázek 24). Nález významně zvýšeného výskytu varianty c.1096_1099dupATTA u vysoce rizikových pacientů byl konfrontován v práci Pfeifer a kol, s nevýznamným nálezem výskytu této varianty a její asociaci s nádorovým onemocněním v německém a makedonském souboru testovaných pacientů s BC a kontrol (Pfeifer et al. 2016). V reakci na tuto skutečnost jsme provedli rozšířenou analýzu *FANCL* c.1096_1099dupATTA v populaci 502 vysoce rizikových osob, 693 pacientů se sporadickým onemocněním a 686 kontrol.

Z důvodu nedostatečného množství vzorků kontrolní populace, ve které při NGS analýze nebyla varianta c.1096_1099dupATTA zachycena, byla nejprve porovnána frekvence této varianty ve studovaném souboru s frekvencí v databázi ExAc (<http://exac.broadinstitute.org>; data z května 2015) uvádějící 342 nosičů c.1096_1099dupATTA z celkových 59 537 testovaných exomů (0,57 %), což naznačuje vyšší výskyt varianty ve studovaném NGS souboru vysoce rizikových pacientů s karcinomem prsu ($p=0,01$; Tabulka 5).

Možný vztah této varianty ke zvýšení rizika vzniku karcinomu prsu jsme dále analyzovali vyšetřením 502 vysoce rizikových pacientů s karcinomem prsu nebo ovaria (všechny *BRCA1/BRCA2/PALB2* negativní). Mezi těmito pacientkami bylo nalezeno dalších pět nosiček mutace c.1096_1099dupATTA – tři s duktálním karcinomem prsu (s diagnózou v 27, 34 a 38 letech) a dvě s karcinomem ovaria (26 a 39 let). Celkově byla tato varianta identifikována u 11/827 vysoce rizikových jedinců (1,3 %), představujících 11/805 vysoce rizikových žen (1,4 %).

Pro identifikaci nosičů mutace c.1096_1099dupATTA u sporadických pacientů s karcinomem prsu a dalších geograficky odpovídajících nenádorových kontrolních jedinců, bylo dále genotypováno 693 neselektovaných případů s karcinomem prsu a 686 nenádorových kontrol (313 žen a 373 mužů). Tato analýza odhalila tři nosiče v každé ze studovaných skupin s frekvencí 0,4 % jak u případů se sporadickým karcinomem prsu (3/693), tak u kontrol (3/791 – včetně 105 NGS kontrol). Frekvence této varianty oproti kontrolnímu souboru je tedy silně zvýšena u vysoce rizikových jedinců ($p=0,039$), nikoliv však u pacientů se sporadickým onemocněním ($p=0,9$). Všech 14 nosičů (11 ze skupiny vysoce rizikových pacientů a 3 z neselektované skupiny pacientů s karcinomem prsu) mezi pacienty byly ženy, zatímco všechny tři případy nalezené u kontrol byli muži bez nádorového onemocnění. Přehled frekvencí této varianty v rámci studovaných souborů osob je uveden v Tabulce 7.

Tabulka 7. Přehled testovaných osob na přítomnost varianty c.1094_1095insAATT (p.P365fs) genu FANCL

Soubor	Počet jedinců v souboru	Počet jedinců s mutací	% zastoupení v rámci souboru
Původní NGS soubor pacientů	325	6	1,85
Pacienti BRCA1/2/PALB2 negativní	502	5	1
Pacienti se sporadickým karcinomem prsu	693	3	0,43
Celkem v souborech pacientů	1520	14	0,92
Původní NGS soubor kontrol	105	0	0
Rozšířená kontrolní populace	686	3 (muži bez onemocnění)	0,43
Celkem v kontrolních souborech	791	3 (muži bez onemocnění)	0,38

Od doby publikování hlavní publikace této dizertační práce (Lhota et al. 2016)(Příloha IV-1) jsme mezitím v rámci aplikace sekvenačního projektu CZEKANCA (Soukupova et al. 2016) identifikovali dalších 8 nosičů této varianty (Zemankova et al. 2016) (Příloha IV-2). Ze všech identifikovaných nosičů této varianty s karcinomem prsu, bylo 14 s poměrně nízkým věkem v době diagnózy (průměr 47.2 let (rozmezí 28–76 let). Zároveň jsme recentně identifikovali tři nosiče c.1096_1099dupATTA u pacientek s karcinomem ovaria v nízkém věku. Oproti pozorování skupiny Pfeifer a kol., která našla pozitivní rodinnou anamnézu karcinomu prsu pouze u jednoho nosiče (příbuzní prvního a druhého stupně), v rámci všech námi vyšetřených osob byla rodinná anamnéza karcinomu prsu nalezena u 23 nosičů této varianty (39 %) a rodinná anamnéza jakéhokoliv nádorového onemocnění u dalších 15 nosičů (celkem 65 %). Ze 14 nosičů varianty c.1096_1099dupATTA vyšetřených pomocí NGS byla nalezena další trunkační varianta v dalším genu.

6.3 Identifikace a funkční analýzy ERCC2 mutací u pacientů s familiárním karcinomem prsu a ovaria (Rump. et al. 21016)

Analýzou hereditárních variant u pacientů s karcinomem prsu (Kap. 6.1) byly nalezeny dvě trunkační (p.V77Afs*4 u 1 pacienta, p.F568Yfs*2 u 1 pacienta a 2 kontrolních osob) a 2 missense varianty (u dvou pacientů: p. Asp423Asn, p.Val611Ala) v genu ERCC2. V rámci mezinárodní spolupráce s laboratořemi v Technische Universität Dresden (Německo), Medical Genetics Center (MGZ) Munich (Německo), Cologne Center for Genomics (CCG, Německo), Vilnius University Hospital Santariskiu Clinics (Lotyšsko) a Masarykovým onkologickým ústavem Brno, byly tyto a další varianty ERCC2 hodnoceny z hlediska geograficko-populační významnosti a vlivu na funkci výsledného proteinového produktu (Rump et al. 2016)(Příloha IV-3). Ve společné studii bylo celkem analyzováno 1345 osob s karcinomem prsu nebo/a ovaria ze zmíněných středo- a východoevropských zemí pomocí multigenového panelového sekvenování, při němž byly nalezeny rekurentní zakladatelské mutace v ERCC2, které byly funkčně validovány esejemi na buněčných liniích v německé skupině.

V rámci celého souboru 1345 pacientů s karcinomem prsu nebo/a ovaria byly nalezeny tři varianty posunující čtecí rámec (p.Val77fs, p.Phe568fs a p.Ser746fs), dvě varianty ovlivňující sestřih pre-mRNA (c.1903-2A>G, c.2150C>G) a 20 missense variant s nízkou populační frekvencí (Tabulka 8).

Popis varianty					Predikovaný efekt	Funkční efekt		BC/OC případů				
Police	Exon	Nukleotidová změna	AK změna	rs-ID	Kumulativní hodnocení	Komplementace NER-deficientních buněk	Negativní modulace transkripce	GE N = 587	CZ N = 353	LT N = 405	celkem N = 1345	BC/ OC
hg19	(23)	NM_000400.3	max = 760 aa									
19:45873459	2	c.37C>T	p.(Pro13Ser)	-	pathogenic	n.t.	n.t.	1	0	0	1	BC
19:45872203	4	c.230_231delTG	p.(Val77Alafs)	-	n.a.	n.t.	n.t.	0	1	0	1	BC+OC
19:45868194	7	c.496C>T	p.(Arg166Cys)	-	pathogenic	n.t.	n.t.	0	0	2	2	BC
19:45868191	7	c.499G>C	p.(Glu167Gln)	rs367829012	benign	n.t.	n.t.	1	0	0	1	BC
19:45868127	7	c.563G>C	p.(Gly188Ala)	-	benign	n.t.	n.t.	1	0	0	1	BC
19:45867756	8	c.644C>T	p.(Pro215Leu)	-	pathogenic	n.t.	n.t.	0	0	1	1	BC
19:45867354	10	c.839G>A	p.(Arg280His)	-	pathogenic	n.t.	n.t.	0	0	1	1	BC
19:45867247	10	c.946C>G	p.(Gln316Glu)	-	benign	n.t.	n.t.	1	0	0	1	BC
19:45860928	13	c.1267G>A	p.(Asp423Asn)	rs143710107	benign	no	yes	0	1	8	9	4xBC, 5xOC
19:45860760	14	c.1349G>A	p.(Arg450His)	rs146632315	pathogenic	yes	no	2	0	0	2	BC
19:45860626	15	c.1381C>G	p.(Leu461Val)	rs121913016	benign	yes	yes	3	0	0	3	2xBC, 1xOC
19:45860548	15	c.1459C>T	p.(Arg487Trp)	rs562132292	pathogenic	no	yes	0	0	4	4	2xBC, 2xOC
19:45858929	16	c.1537G>T	p.(Asp513Tyr)	-	pathogenic	yes	yes	1	0	0	1	BC
19:45858047	17	c.1606G>A	p.(Val536Met)	rs142568756	pathogenic	yes	yes	2	0	0	2	BC
19:45856554	18	c.1703_1704delTT	p.(Phe568fs)	-	pathogenic	no	no	1	3	1	5	BC
19:45856532	18	c.1726G>A	p.(Glu576Lys)	rs201165309	pathogenic	n.t.	n.t.	1	0	0	1	BC
19:45856397	19	c.1775G>A	p.(Arg592His)	rs147224585	pathogenic	yes	no	1	0	7	8	BC
19:45856370	19	c.1802G>A	p.(Arg601Gln)	rs140522180	pathogenic	yes	yes	2	1	0	3	BC
19:45856074	20	c.1832T>C	p.(Val611Ala)	-	benign	n.t.	n.t.	0	1	0	1	BC
19:45856015	20	c.1891C>T	p.(Arg631Cys)	rs144511865	pathogenic	no	no	1	0	1	2	1xBC, 1xOC
19:45855909	IVS 20	c.1903-2A>G	splice site	-	n.a.	n.t.	n.t.	1	0	0	1	BC+OC
19:45855778	21	c.2032G>C	p.(Val678Leu)	-	benign	n.t.	n.t.	0	0	1	1	BC
19:45855507	22	c.2150C>G	splice effect	rs144564120	pathogenic	n.t.	n.t.	3	0	0	3	2xBC, 1xOC
19:4585492	22	c.2165G>A	p.(Arg722Gln)	rs138569838	pathogenic	n.t.	n.t.	0	1	0	1	BC
19:45854932	23	c.2238delA	p.(Ser746fs)	-	n.a.	yes	yes	1	0	0	1	OC

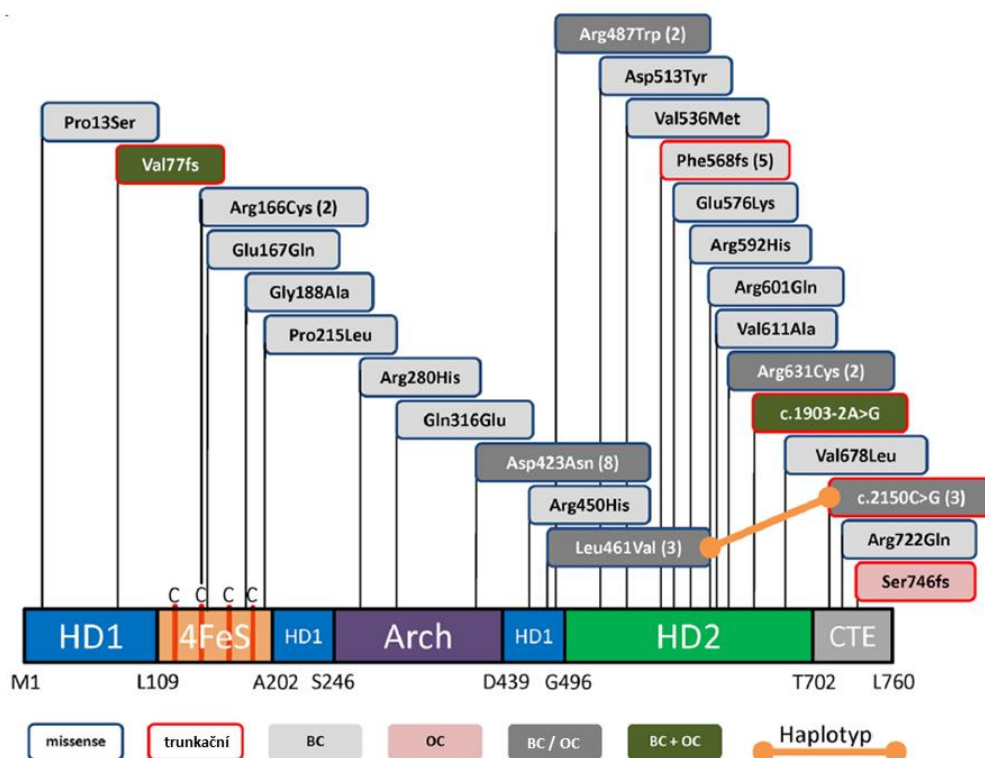
Tabulka 8. Mutace a raritní varianty genu ERCC2 identifikované panelovým sekvenováním jedinců s dědičným onemocněním karcinomem prsu a/nebo ovaria (HBOC). AK = aminokyselina; N = velikost souboru; n.a. = není dostupné; n.t. = netestováno; CZ = ČR, GE = Německo, LT = Litva. Kumulativní hodnocení je založeno na výsledcích několika predikčních algoritmů. (Rump et al. 2016)

Zatímco 14 mutací bylo v rámci souboru unikátních (2 frame-shift, 1 sestřihová varianta, 11 missense), 11 mutací (1 frame-shift, 1 sestřihová varianta, 9 missense) bylo nalezeno ve 43 nezávislých rodinách. Nejčastěji se vyskytovala varianta p.Asp423Asn, identifikovaná u 8 nosičů v Litevském souboru a u jednoho z České republiky. Časté polymorfismy p.Lys751Gln a p.Asp312Asn byly nalezeny u 64 % vyšetřených případů. Jelikož byly tyto varianty označeny za funkčně irelevantní (Pabalan et al. 2010), nebyly zahrnuty do funkčního testování v rámci projektu. Mezi 20 raritními missense variantami reportovanými v Tabulce 9 je 13 variant predikovaných dle různých algoritmů jako patogenní. Další *in silico* evaluace konzervace (PhyloP) a kombinované predikce patogenity (CADD) silně podpořily závažnost těchto variant.

Tabulka 9. Predikce efektu missense variant ERCC2. Pravděpodobnost vlivu nesynonymních variant ERCC2 byla predikována počítačovými algoritmy: SIFT, Sorting Invariant from Tolerated (skóre pod 0,05: not tolerated; rozmezí 0–1); PolyPhen-2, Klasifikace podle PSIC skóre (HumVar, "benign"- "possibly damaging"—"probably damaging ", rozmezí: 0–1); Provean, Protein Variation Effect Analyze; MAPP, Multivariate Analysis of Protein Polymorphism; Align-GVGD, Skóre (C0, C15, C25, C35, C45, C55, C65) od C0 (likely benign) po C65 (likely pathogenic); CADD, Combined Annotation Dependent Depletion. Dlt = deleterious, PrD = probably damaging, PsD = possibly damaging, Bgn = benign, Ntr = neutral. Konzervovanost byla analyzována pomocí PhyloP (Rozmezí skóre od -14.1 to 6.4). Grantham distance score (Rozmezí 0–215).

AK záměna (N = 20)	SIFT	PPH2	Pro-vean	MAPP	Align GVGD	CADD	PhyloP	Grantham Distance	Souhrn
Pro13Ser	Dlt	PrD	Dlt	bad	C0	19.70	5.21	74	patogenní
Arg166Cys	Dlt	PrD	Dlt	bad	C0	29.80	3.35	180	patogenní
Glu167Gln	Tol	Bgn	Ntr	good	C0	11.52	1.5	29	benigní
Gly188Ala	Tol	Bgn	Ntr	good	C0	24.10	3.44	60	benigní
Pro215Leu	Dlt	PrD	Dlt	bad	C65	29.40	5.13	98	patogenní
Arg280His	Dlt	Bgn	Dlt	bad	C0	24.40	2.30	29	patogenní
Gln316Glu	Tol	Bgn	Ntr	good	C0	14.16	1.98	29	benigní
Asp423Asn	Tol	Bgn	Ntr	good	C0	21.00	5.53	23	benigní
Arg450His	Dlt	Bgn	Dlt	bad	C0	27.60	4.75	29	patogenní
Leu461Val	Dlt	PsD	Ntr	good	C0	25.60	4.64	32	benigní
Arg487Trp	Dlt	PrD	Dlt	bad	C0	26.00	0.29	101	patogenní
Asp513Tyr	Dlt	PrD	Dlt	good	C25	22.60	5.05	160	patogenní
Val536Met	Dlt	PrD	Dlt	bad	C15	22.60	5.45	21	patogenní
Glu576Lys	Dlt	PrD	Dlt	bad	C55	22.50	5.61	56	patogenní
Arg592His	Dlt	PrD	Dlt	bad	C25	16.23	5.61	29	patogenní
Arg601Gln	Dlt	PsD	Dlt	bad	C35	32.00	5.61	43	patogenní
Val611Ala	Tol	Bgn	Ntr	good	C0	18.83	4.48	64	benigní
Arg631Cys	Dlt	PrD	Dlt	bad	C65	22.00	3.54	180	patogenní
Val678Leu	Tol	Bgn	Ntr	good	C0	17.23	0.77	32	benigní
Arg722Gln	Dlt	PsD	Dlt	bad	C0	36.00	5.21	43	patogenní

Studované varianty byly rozprostřeny po celé délce kódující sekvence ERCC2 (Obrázek 25). Aminokyselinové zbytky 13, 450, 461, 513, 536, 576, 592, 601, 611, 631, 678 se vyskytují v klastru helikázových motivů HD1 a HD2 katalytických domén a zbytky 166, 167, 188, 215, 280, 316, 423, 487, 722 lokalizují do domén vážících komplex transkripčních faktorů TFIIH (domény Arch, FES, a C-terminální doména). Varianty popsané jako patogenní změny vyvolávající xeroderma pigmentosum (XPD) jsou lokalizovány v HD2 doméně a inaktivují helikázovou reparační aktivitu bez narušení struktury proteinu. Naopak mutace, které způsobují trichothiodystrofii (TTD), jsou lokalizovány daleko od katalytického místa enzymu a destabilizují strukturu XPD (XPD je proteinovým produktem genu ERCC2) a interakce s TFIIH proteiny (Fan et al. 2008). Dle našich dat předpokládáme, že BC / OC relevantní varianty postihují katalytickou aktivitu i stabilitu proteinu.



Obrázek 25. Doménová struktura a poloha mutací ERCC2. Znáznorněn je ERCC2 protein se čtyřmi XPD doménami: HD1 (modře), HD2 (zeleně), FeS (oranžově) a Arch (fialově). C-terminální extenze (šedě) patrně vytváří interakční plochu pro protein p44. Potencionálně patogenní varianty ERCC2 jsou znázorněny v rámečcích (modré, nebo červené: missense nebo truncační varianty; výplně: světle šedá – případy s karcinomem prsu - BC, růžová: případy s pouze karcinomem ovaria - OC, tmavě šedá: případy s karcinomem prsu nebo ovaria – BC/OC, tmavě zelená: pacientky s karcinomem prsu i ovaria (BC + OC). Čísla v závorkách vyznačují rekurentní mutace.

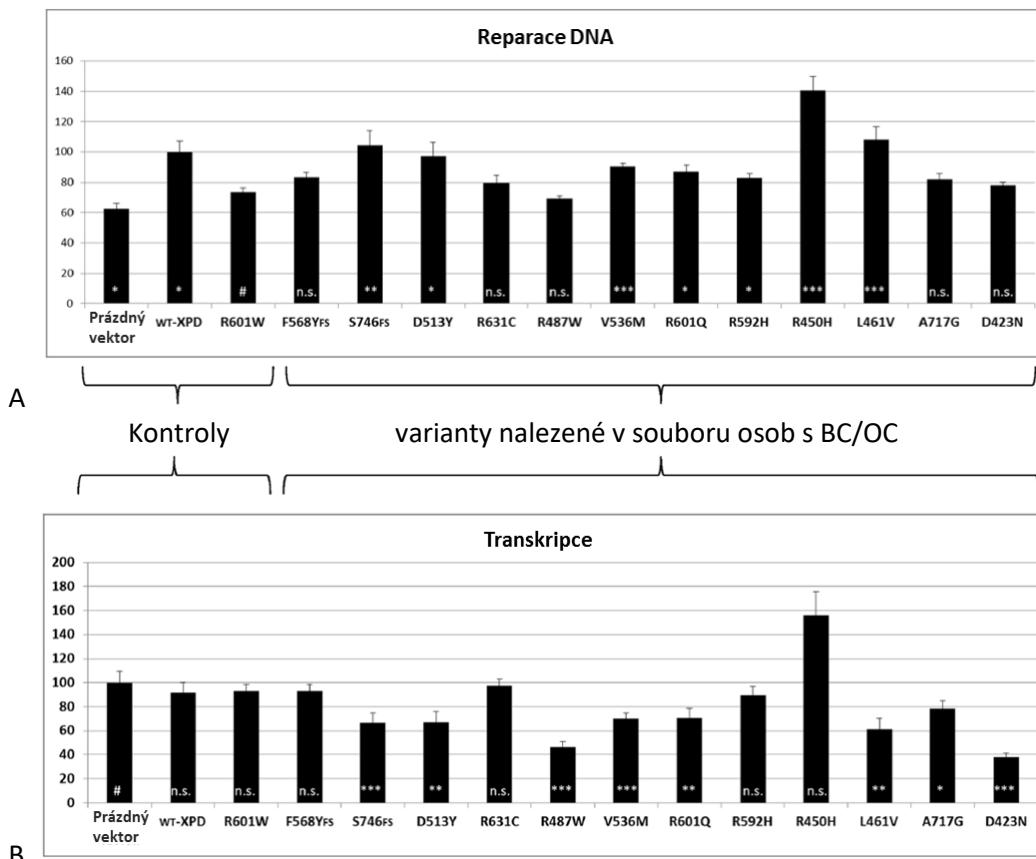
Ve spolupracující laboratoři v Technische Universität Dresden bylo v rámci projektu testováno zatím 11 variant (9 rekurentních a 2 unikátní varianty; Tabulka 10) pomocí funkčních analýz hodnotících reparační kapacitu NER (Obrázek 22 A) a regulaci transkripce (Obrázek 22 B).

Tabulka 10: Charakteristiky mutací testovaných na účinnost reparační a transkripční aktivitu.

Mutace / charakteristiky	Asp 423 Asn	Arg 450 His	Leu 461 Val	Arg 487 Trp	Asp 513 Tyr	Val 563 Met	Phe 568 fs	Arg 592 His	Arg 601 Gln	Arg 631 Cys	Ser 746 fs
Kodon spojený s onemocněním	TTD	---	TTD	TTD	---	---	TTD	---	XPD	---	---
In-silico predikce vlivu	ben	pat	ben	pat	pat	pat	fs	pat	pat	pat	fs
Opakovaně se vyskytující	ano	ano	ano	ano	---	ano	ano	ano	ano	ano	---

Zatímco 6 z 11 BC/OC asociovaných variant ERCC2 testovaných v této studii zatím nebylo popsáno ve spojitosti s onemocněním (AK zbytky 423, 450, 513, 536, 631, 746), 5 aminokyselinových zbytků bylo popsáno jako mutovaných u TTD: AK 461 (Botta et al. 1998), AK 487 (Viprakasit et al. 2001), AK 568

(Zhou et al. 2010), AK 592 (Nishiwaki et al. 2004) nebo XPD AK 601 (Cleaver et al. 1999) (Obrázek 22). Podle funkčních analýz nedokázaly čtyři varianty proteinu *ERCC2* (p.Asp423Asn, p.Arg487Trp, p.Phe568Tyrfs a p.Arg631Cys) zvýšit funkční NER reportérového genu v plazmidu po UV ozáření, což naznačuje poškození reparační kapacity XPD/*ERCC2*. Zbývajících 7 testovaných variant si zachovávalo mírnou schopnost NER (Obrázek 26 A). Při analýze vlivu na transkripci byl detekován dominantně negativní vliv sedmi variant *ERCC2* (p.Asp423Asn, p.Leu461Val, p.Arg487Trp, p.Asp513Tyr, p.Val536Met, p.Arg601Gln, p.Ser746fs) na expresi reportérového genu (Obrázek 26 B) naznačující blokování transkripce. Celkem tedy 10 z 11 mutací vykazovalo snížení účinnosti reparace excízi a/nebo sníženou schopnost transkripční aktivity. Nejvyšší dopad na funkci proteinu v tomto ohledu vykazovaly varianty p.Asp423Asn a p.Arg487Trp.



Obrázek 26: Kapacita nucleotide excision repair (NER) a transkripční aktivita PD/*ERCC2* variant asociovaných s karcinomem prsu. (A) Několik XPD/*ERCC2* variant zaklonovaných do expresního vektoru bylo analyzováno z hlediska komplementace v *ERCC2*-defektní buněčné linii XP6BE, které nadměrně exprimují NER-deficientní R601W XPD mutantní protein (Liu et al. 2008). Černé sloupce ukazují střední relativní reparační kapacitu (v %, WT-XPD byl vzat jako 100 %) plazmidu s reporterovým luciferázovým genem po ozáření UV (UVC 1000J/m²) po reaktivaci v transfekované buňce (n>6 v triplikátech). (B) Dominantní modulace exprese reporterového luciferázového genu (bez ozáření) pomocí overexprese XPD/*ERCC2* variant asociovaných s karcinomem prsu/ovaria v buňkách schopných transkripce ale s XPD/*ERCC2*-závislým deficitem reparace – linie XP6BE. Černé sloupce vyznačují střední relativní expresi reporterového genu (v %, prázdný vektor = 100 %), zajištěnou bazální transkripcí prostřednictvím CMV-promotoru (n>6 v triplikátech). Startovní odchylky jsou vyznačeny. Byly vypočteny míry významnosti testováním normální distribuce hodnot pomocí příslušných statistických nástrojů pro porovnávání dvou skupin (T-Test, nebo U-Test, # = referenční skupina, *** = p<0.001, ** = p<0.01, * = p<0.05, n.s. = not significant - nevýznamné). (Rump et al. 2016)

Většina nalezených germinálních *ERCC2* mutací jsou zakladatelské mutace. Mezi jejich typické znaky patří opakovaný výskyt, populační specificita a sdílení haplotypu. Mutace v *ERCC2* jsou rekurentní, neboť 11 z 25 bylo pozorováno alespoň dvakrát ve sledované společné skupině osob s karcinomem prsu/ovarií (poslední sloupec v Tabulce 6). Mezi 11 rekurentními variantami bylo 5 identifikováno pouze v jedné ze tří populací (např. p.Asp423Asn: 8 x L, 1 x CZ, 0 x GE). Pro dvě další mutace specificky nabohacené v některých populacích můžeme demonstrovat sdílení haplotypu: (i) mutace c.1381C>G (rs121913016) se vždy vyskytovala v kosegregaci s c.2150C>G (rs144564120), což je haplotyp, který byl opakovaně pozorován u TTD/XPD pacientů (Takayama et al. 1995). (ii) Téměř u všech případů (10/11) se frame-shift mutace c.1703_1704delTT vyskytovala spolu s polymorfizmem c.1758+32C>G (rs238417). Navíc jsou tyto dvě varianty vzdáleny pouze 84 nukleotidů a všechny NGS čtení zahrnující tyto pozice obsahovaly obě varianty, což ukazuje jejich vzájemnou lokalizaci v *cis* pozici.

Frekvence *ERCC2* variant charakterizovaných v rámci studie byly porovnány s frekvencemi veřejných databází ESP (NHLBI Exome Sequencing Project; (Venter et al. 2001) a ExAC (Exome Aggregation Consortium). Některé z nalezených mutací, jako je p.Phe568fs* a Asp423Asn byly zachyceny u analyzovaných pacientek s karcinomem prsu a ovaria významně častěji, než u osob z populací v ExAC databázi, což naznačuje asociaci těchto variant s výskytem onemocnění. Pro potvrzení tohoto výsledku jsme učinili segreganční analýzy, nicméně žádná ze studovaných mutací *ERCC2* nevykazovala přesvědčující kosegregaci s výskytem karcinomu prsu/ovaria. Navíc následná analýza populačně specifických kontrolních souborů ukázala, že téměř všechny zakladatelské mutace se v souborech BC/OC pacientů vyskytovaly s podobnou frekvencí jako u etnicky odpovídajících kontrol. Jedinou výjimkou je doposud litevská mutace p.Arg487Trp, nalezená u čtyř vyšetřovaných v souboru BC/OC pacientek, a nepřítomná v kontrolní populaci (Tabulka 11). Je však nutné upozornit, že populační kontrolní soubor zahrnující pouze 100 jedinců, je nedostatečný na průkaz statistické významnosti tohoto nálezu. Jedním z výstupů provedené analýzy je upozornění na skutečnost, že stanovení frekvencí především zakladatelských variant v populačně specifických souborech kontrol, přestože mohou být malé, je přesnější, než stanovení frekvence těchto variant v rozsáhlých souborech globálních databází.

Tabulka 11. Alelické frekvence ERCC2 variant u pacientů s karcinomem prsu/ovaria (BC/OC) a u příslušných kontrolních kohort. Alelická frekvence je vypočtena na základě velikosti souboru (v závorce) a počtu pozorovaných případů (viz Tabulka 6) se zohledněním hetero-, homozygotity.

AK změna (N = 25)	CZ	CZ	LT	LT	GE	GE	ExAc
	BC/OC	Kontroly	BC/OC	Kontroly	BC/OC	Kontroly	vers. 0.2
	[353] ^a	[453] ^b	[405]	[103]	[587] ^c	[1844] ^d	[variabilní] ^e
Pro13Ser	0	0	0	0	0.0851	0	0
Val77Alafs	0.1416	0	0	0	0	0	0
Arg166Cys	0	0	0.2469	0	0	0	0
Glu167Gln	0	0	0	0	0.0851	0	0.0033
Gly188Ala	0	0	0	0	0.0851	0	0
Pro215Leu	0	0	0.1234	0	0	0	0
Arg280His	0	0	0.1234	0	0	0	0.0072
Gln316Glu	0	0	0	0	0.0851	0	0.0152
Asp423Asn	0.1416	0.1104	0.9876	1.456	0	0.0542	0.0248
Arg450His	0	0	0	0	0.1704	0.0813	0.0214
Leu461Val	0	0	0	0	0.2553	0.1356	0.1345
Arg487Trp	0	0	0.4938	0	0	0	0.0034
Asp513Tyr	0	0	0	0	0.0851	0	0
Val536Met	0	0	0	0	0.1704	0	0.0231
p.Phe568fs	0.4249	0.4415	0.1234	0	0.0851	0	0.0093
Glu576Lys	0	0	0	0	0.0851	0.0542	0.0008
Arg592His	0	0	0.8642	0	0.0851	0	0.0332
Arg601Gln	0	0.1104	0	0	0.1704	0.0542	0.0175
Val611Ala	0.1416	0	0	0	0	0	0.0042
Arg631Cys	0	0	0.1234	0	0.0851	0	0.0025
c.1903-2A>G	0	0	0	0	0.0851	0	0
Val678Leu	0	0	0.1234	0	0	0	0
c.2150C>G	0	0	0	0	0.2553	0.0813	0.0349
Arg722Gln	0.1416	0	0	0	0	0	0.0067
p.Ser746fs	0	0	0	0	0.0851	0	0

BC/OC = jedinci s karcinomem prsu a/nebo ovaria; Kontroly = zdravé osoby bez nádorového onemocnění; CZ = Czech Republic, GE = Germany, LT = Lithuania; AK AK = aminokyselina; N = velikost souboru; nt = nukleotid; ExAC = Exome Aggregation Consortium, Cambridge, MA (URL: <http://exac.broadinstitute.org>);

^a 28 vzorků z MOU Brno (TruSight-Cancer) + 325 z našeho pracoviště (custom panel 580 genů; Nimblegen, ÚBEO, 1.LF UK)

^b 108 nenádorových kontrolních osob ženského pohlaví (MOU, Brno) + 105 nenádorových kontrolních osob ženského pohlaví (ÚBEO, 1.LF UK) + 240 nenádorových kontrol sekvenovaných hromadně pomocí Trusight-Cancer (ÚBEO, 1.LF UK)

^c 271 vzorků z TU Dresden + 316 vzorků z MGZ Mnichov (TruSight-Cancer)

^d 1629 vzorků exomového sekvenování z CCG Kolín n. Rýnem + 79 osob bez BC/OC anamnézy z TU Dresden (Trusight-One) + 136 osob bez BC/OC anamnézy z TU Dresden a MGZ Mnichov (TruSight-Cancer)

^e Neboť byla exomová data získána z různých zdrojů s rozdílnými strategiemi obohacení o cílové oblasti, liší se velikosti souborů pro jednotlivé varianty. Alelické frekvence jednotlivých variant byly vypočteny z příslušných velikostí souborů.

Výskyt onemocnění spojených s přítomností patogenních germinálních mutací ERCC2 neodpovídá frekvenci pravděpodobně patogenních zakladatelských mutací v příslušných populacích. Přestože zakladatelské mutace testované v této studii nemusí predisponovat k BC/OC, stále mohou propůjčovat status nosičství onemocnění s recesivní dědičností, kterými jsou XPD (OMIM 278730), TTD (OMIM 601675) a COFS2 (OMIM 610756). Dokonce i samotná mutace p.Phe568fs byla

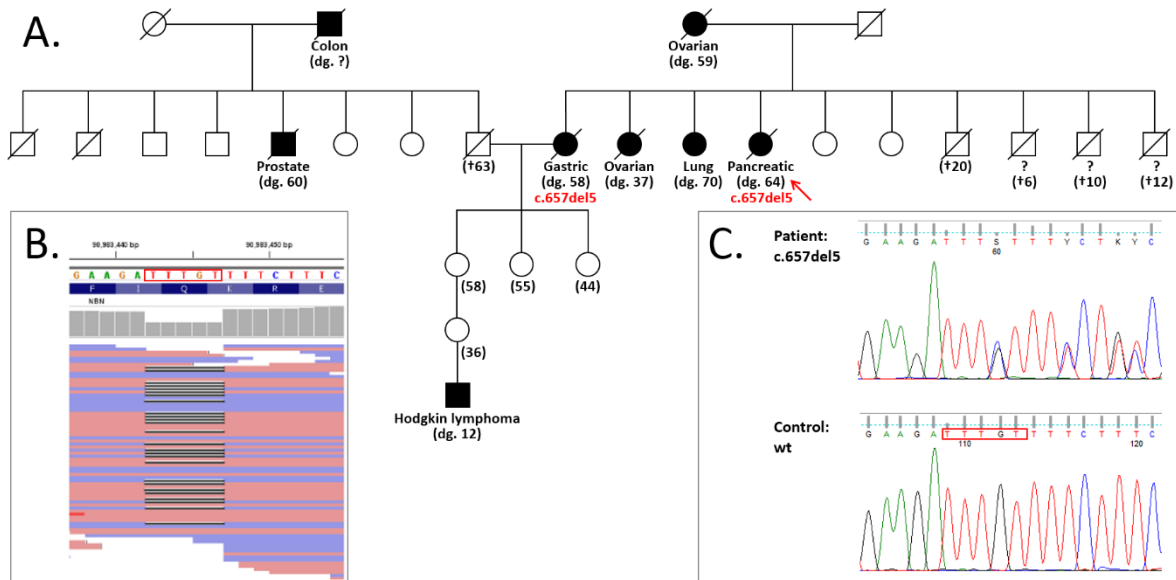
detekována u 7 z 806 vzorků z České republiky, což znamená, že frekvence heterozygotních nosičů této varianty je přibližně 0,86 %. Podle Hardy-Weinbergova rovnovážného modelu by incidence TTD byla 1/30 000. Na základě kombinovaných dat z diagnostických center zabývajících se poruchami reparace DNA ve Francii, Německu, Itálii, Holandsku a Velké Británii je reálná incidence TTD 1,2 na milion osob (Kleijer et al. 2008). Jelikož lze předpokládat, že (i) incidence 1/30 000 by nebyla přehlédnuta českými genetiky a (ii) publikace charakterizující variantu p.Phe568fs jako kauzální pro TTD se nemýlí (Takayama et al. 1995, Broughton et al. 2001), diskrepance mezi alelickou frekvencí a incidencí onemocnění naznačuje, že homozygotita varianty p.Phe568fs je patrně embryonálně letální. Toto tvrzení souhlasí s pozorováním, že úplná ztráta aktivity *ERCC2* je letální u homozygotních knock-out myši (de Boer et al. 1998) a zároveň je konzistentní s pozorováním, že všichni XPD a TTD doposud testovaní pacienti vykazovali reziduální aktivitu *ERCC2* (Lehmann et al. 2014). Jelikož v Litvě nebyla popsána zvýšená incidence TD/XPD, lze předpokládat, že frekventní litevská zakladatelská mutace p.Asp423Asn (Tabulka 11), která jasně vykazuje funkční deficienci, je také v homozygotní podobě embryonálně letální.

6.4 Analýza varianty c.657del5 genu *NBN* ve vztahu ke karcinomu pankreatu (Borecka et al. 2016)

Kontrolní analýzou klinických a histopatologických dat pacientů s karcinomem prsu analyzovaných panelovým NGS (Lhota et al. 2016; Kap. 6.1.) jsme zjistili, že se u 22 osob vyšetřovaného souboru pacientů vyskytovalo nádorové onemocnění jiné diagnózy než karcinom prsu. Tito jedinci byli ze souboru 325 pacientů pro analýzu predispozice ke karcinomu prsu vyřazeni. Protože se však jednalo o kvalitní sekvenční výstupy získané u nemocných s onkologickými diagnózami, byly výsledky sekvenování podrobeny další kontrole. U jednoho ze sekvenovaných vzorků souboru nemocných bylo zjištěno, že se jedná o vzorek pacientky s diagnózou duktálního adenokarcinomu pankreatu, nikoliv karcinomu prsu (Obrázek 27).

U této pacientky byla zjištěna varianta c.657delACAAA v genu *NBN*, což je nejčastější, patogenní, středoevropská mutace u pacientů s Nijmegen-breakage syndromem (NBS), autosomálně recesivním onemocněním charakterizovaným chromozomální instabilitou a zvýšeným rizikem lymfoidních malignit a dalších nádorových onemocnění (Varon et al. 1998). Bylo ukázáno, že heterozygotní mutace *NBN* predisponují ke karcinomu prsu (Gorski et al. 2003), non-Hodgkinovu lymfomu, a karcinomu prostaty, nicméně role *NBN* v predispozici a karcinogenezi karcinomu pankreatu nebyla doposud studována. Námi nalezená varianta c.657delACAAA je nejfrekventnější patogenní variantou u pacientů s *NBN* asociovanými nádorovými onemocněními ve slovanské populaci (Varon et al. 2000). U

této pacientky (diagnostikované v 64 letech) nebyla nalezena žádná další trunkační varianta v dalších predispozičních genech pro karcinom pankreatu (*BRCA1*, *BRCA2*, *PALB2*, *MLH1*, *MSH2*, *MSH6*, *PMS2*, *STK11*, *APC*, *CDKN2A*). Přítomnost této mutace byla potvrzena i u zemřelé sestry probandky s karcinomem žaludku (Obrázek 27).



Obrázek 27: Náhodný záchyt mutace *c.657del5* v genu *NBN* u pacientky s karcinomem pankreatu. A. rodokmen pacientky (šipka), B. záchyt varianty pomocí NGS (zobrazeno v IGV viewer) a C. potvrzení mutace přímým sekvenováním.

Pro analýzu frekvence této varianty v populaci pacientů s karcinomem pankreatu a kontrolní populaci byla varianta *c.657del5* genotypována u 241 neselektovaných pacientů s duktálním adenokarcinomem pankreatu. Mezi těmito pacienty bylo identifikováno 5 nosičů vyšetřované varianty (2,07 %). Frekvence této varianty v kontrolní populaci byla významně nižší (2/915), což naznačuje, že nosiči varianty *c.657del5* mají vyšší riziko vzniku duktálního adenokarcinomu pankreatu (OR = 9.7; 95%CI: 1.9–50.2; PFET = 0.006). Rodinná historie karcinomu pankreatu byla zdokumentována pouze u jednoho z pěti nosičů mutace, nicméně druhý pacient měl nádorovou historii v rodině (sestru s karcinomem žaludku) a třetí pacient ženského pohlaví pak spolu s karcinomem pankreatu (ve 46 letech) zároveň i karcinom prsu (v 64 letech). Průměrný věk v době diagnózy u nosičů varianty *c.657del5* byl 65,8 roku (v rozmezí 59-73) (Borecka et al. 2016) (Příloha IV-4).

6.5 Systém pro identifikaci sestřihových variant mRNA založený na amplikonovém NGS (Hojny et al. 2017)

Recentní přístup identifikace predispozičních variant pomocí NGS generuje i velké množství variant nejasného významu, které mohou ovlivňovat sestřih pre-mRNA za vzniku aberantních sestřihových isoform. Proto je nezbytným předpokladem pro klinickou interpretaci významu takových variant znát i paletu všech přirozeně se vyskytujících alternativních sestřihových variant daného genu v dané tkáni. Na našem pracovišti jsme vyvinuli komplexní přístup k takovéto katalogizaci sestřihových variant a jeho aplikaci jsme ukázali na příkladu analýz sestřihových variant genu *BRCA1* (Hojny et al. 2017) (Příloha IV-5).

Pro identifikaci alternativních sestřihových variant byl navržen přístup založený na multiplexní PCR (mPCR) amplifikaci všech teoreticky možných exon-exonových spojení s následnou selekcí dle velikosti amplikonu a charakterizací pomocí NGS. Účinnost této metody byla prokázána na komplexním modelu analýzy alternativních sestřihových variant genu *BRCA1* v lidských leukocytech, zdravé mamární tkáni, tukové tkáni a stabilních buněčných liniích. Klíčovou částí této metody je navržení všech primerových párů pro jednotlivé přechody exonů tak, aby forward primer cílil na 3' oblast exonu a reverse primer na 5' oblast exonu následujícího. Pro takovouto analýzu *BRCA1* bylo navrženo 45 primerů cílících 22 kódujících exonů kanonického transkriptu NM_007294 a dále i alternativní exony 11q a 13A. Primery byly použity pro amplifikaci cDNA v 31 mPCR reakcích. Produkty těchto amplifikací byly smíchány a následně byl proveden výběr krátkých amplikonů, odpovídajících zkráceným (alternativně sestřiženým) transkriptům, pomocí reverzibilní imobilizace na paramagnetických kuličkách. Z těchto PCR produktů byla připravena standardní NGS knihovna pro sekvenování na platformě Illumina. Bioinformatická analýza identifikovala fragmenty s příslušnými exon-exonovými spojeními a byl vytvořen katalog všech nalezených variant pro příslušnou tkáň/buněčnou linii. Celkem bylo tímto přístupem identifikováno 94 alternativních sestřihových variant *BRCA1*. Pouze 29 variant bylo přítomno ve všech testovaných tkáních, což potvrzuje rozdílnou expresi sestřihových variant *BRCA1* v jednotlivých tkáních (Orban et al. 2003, Colombo et al. 2014, Romero et al. 2015). Čtyřicet dva z celkových 94 nalezených variant nebylo doposud popsáno v předchozích studiích (Colombo, Romero, Orban). Tato metoda umožňuje vysoce citlivou a rychlou detekci alternativních sestřihových variant pro teoreticky libovolný gen.

7 DISKUSE

Panelové sekvenování nové generace představuje spolehlivý nástroj pro analýzu nádorových predispozičních genů v klinické praxi, ale i způsob identifikace nových kandidátních genů u vysoce rizikových jedinců. Oproti exomovému, či celogenomovému sekvenování umožňuje identifikaci nosičů patogenních variant s nižší nákladností, větší flexibilitou ve výběru cílových genů, senzitivity a méně náročné bioinformatické zpracování pro rutinní praxi (Schroeder et al. 2015). Analýza představená v této práci odhalila truncační varianty u téměř jedné třetiny vyšetřených pacientů; 30 pacientů (9 %) byli nosiči truncační varianty v jednom z 15 genů (*ATM, ATR, BLM, BRIP1, ERCC2, FANCE, FANCL, FANCM, CHEK2, NBN, NF1, RAD50, RAD51C, RAD51D, WRN*), v současnosti testovaných komerčně dostupnými klinickými NGS panely (Easton et al. 2015). Ze 73 genů s truncačními mutacemi, byla pouze u 51 z nich nalezena jen jedna truncační varianta. Tato skutečnost naznačuje, že by mohly být extrémně vzácné varianty identifikovány u značné části vysoce rizikových jedinců. Klinická interpretace takových variant a jejich odlišení od náhodných nálezů, které nejsou spojeny s predispozicí ke karcinomu prsu, může být velice obtížné.

I přes omezené množství analyzovaných vzorků, nebylo 22/36 (61 %) unikátních truncačních variant v **DNA reparačních genech** přítomno v databázi ExAc a frekvence pouze šesti variant (17 %) se nelišila mezi analyzovaným souborem a databází ExAc (Tabulka 5). Zároveň pouze 20/53 (38 %) unikátních truncačních variant v genech neúčastnících se reparační DNA bylo reportováno v databázi ExAc a frekvence 13 z nich (25 %) se nelišila mezi analyzovaným souborem a databází (Tabulka 10). Tyto nálezy naznačují, že i) alespoň některé truncační varianty nemusí zásadně zvyšovat riziko karcinomu prsu u svých nosičů a ii) že tento jev je pravděpodobněji u genů neúčastnících se reparační DNA.

Zajímavým nálezem této práce je vysoká frekvence potenciálně patogenních variant v pěti FA genech (*FANCL, FANCM, RAD51C/FANCO, FANCE, a BRIP1/FANCJ*) v 16/325 (4,9 %) vysoce rizikových pacientech analyzovaných pomocí NGS. FA geny kódují DNA reparační proteiny přispívající ke genomové stabilitě prostřednictvím oprav ICL (Kottemann et al. 2013). FA a FA asociované proteiny vytvářejí řadu protein-proteinových komplexů (Walden et al. 2014). Místo poškození DNA je rozeznáno kotevním komplexem (*FANCM, FAAP24, MHF1, MHF2*), který zahajuje sestavování multiproteinového Core komplexu (*FANCA, FANCB, FANCC, FANCE, FANCF, FANCG, FANCL, FAAP20, FAAP100*), který je nezbytný k monoubikvitinylaci heterodimeru *FANCI/FANCD2 (ID2)*. Monoubikvitinovaný komplex ID2 následně aktivuje FA nukleázy (*SLX4/FANCP, XPF/FANCO*) a další DNA reparační faktory (*BRIP1/FANCJ, PALB2/FANCN, BRCA2/FANCD1, BRCA1/FANCS, RAD51C/FANCO*) mechanismem, který zatím není plně objasněn. Dědičné bílelické mutace v FA genech jsou zodpovědné za vývoj FA, charakterizované kongenitálními vadami, selháním kostní

dřeně, buněčné přecitlivělosti k látkám způsobujícím meziřetězcové spoje DNA a predispozici k nádorovým onemocněním.

Nejčastěji identifikovaná truncační varianta v naší studii byla varianta c.1096_1099dupATTA v genu *FANCL*, který kóduje ubikvitin ligázu zajišťující monoubikvitinylaci ID2 (Kap. 1.3.1.3), což je klíčovým krokem v aktivaci FA dráhy (Meetei et al. 2003). Varianta c.1096_1099dupATTA byla popsána jako jedna ze dvou patogenních mutací *FANCL* u pacienta bez evidentních hematologických abnormalit, který patřil do FA-L komplementační skupiny (Ali et al. 2009). Mutovaný protein *FANCL* (p.T367Nfs*13) obsahuje aberantní řetězec dvanácti aminokyselin (který je o tři aminokyseliny delší, než wild type *FANCL*), který zasahuje do PHD/RING finger domény katalyzující ubikvitin ligázovou aktivitu *FANCL*. Ali a kol. zjistili pomocí funkční studie, že *In vitro* tato hypomorfní varianta snižuje vazbu *FANCL* s ostatními FA vazebnými partnery Core komplexu a významně snižuje schopnost monoubikvitinylace D2I komplexu. Dále bylo dokázáno, že tato varianta vykazuje navození G2/M zástavy buněčného cyklu po vystavení melphalanu (látka vytvářející meziřetězcové spoje DNA), dále zvýšenou senzitivitu k mitomycinu c a zvýšenou frekvenci tvorby chromozomálních zlomů oproti buňkám s divokou formou genu *FANCL* (Ali et al. 2009). Stejná varianta byla také identifikována u turkmenského pacienta s familiálním skvamózním karcinomem jícnu (Akbari et al. 2011). Nedávno byly identifikovány další, bialelické mutace, *FANCL* u tří případů VACTERL syndromu (OMIM 192350), které ukazují těžké, nenáhodné malformace (Vetro et al. 2015). Zajímavé je, že bylo zatím popsáno pouze šest FA-L pacientů, což činí tuto komplementační skupinu velmi vzácnou.

Naše výsledky ukazující vyšší zastoupení varianty c.1096_1099dupATTA mezi vysoce rizikovými pacienty s karcinomem prsu, ale zároveň nízkou frekvencí mezi neselektovanými případy karcinomu prsu a mezi nenádorovými kontrolami, což naznačuje, že tato varianta může představovat novou alelu predisponující ke karcinomu prsu. Jelikož frekvence této varianty v databázi ExAc (s nejvyšší frekvencí v latinskoamerické populaci: 1.2 % a nejnižší v jihoasijské: 0.02 %) se celosvětově významně liší, bude pro potvrzení vztahu této varianty k predispozici ke karcinomu prsu a jejímu možnému klinickému využití, potvrzení rozsáhlejšími sekvenčními studii. Jelikož bylo šest nosičů varianty c.1096_1099dupATTA, zároveň nosiči jiné truncační varianty v DNA reparačních genech (*XRCC4*, *CHEK2*), nebo v dalších analyzovaných genech (*CYP3A5+IL8*, *ABCC2*), nemůžeme vyloučit ani možnost, že tato varianta může fungovat jako modifikující, spíše než jako samostatně predisponující. Zajímavé je, že byl v našem souboru kontrol nalezen homozygotní jedinec (muž, 66 let) bez nádorového onemocnění a příznaků FA. Všichni Jedinci kontrolní populace s touto variantou byli mužského pohlaví. Absence jakýchkoliv znaků FA byla také pozorována u homozygotních nosičů nonsense variant asociovaných s karcinomem prsu – *FANCM* c.5101C>T a c.5791C>T (diskutovány dále), které zkracují C-terminální část *FANCM* proteinu (Lim et al. 2014). Časté mutace ovlivňující *FANCL* a *FANCM*

C-terminální části proteinu naznačují, že ne všechny truncační varianty FA genů, a obzvláště ty, které jsou lokalizované na C-konci, úplně znemožňují funkci kódovaného proteinu, ale mohou narušit FA dráhu za specifických, doposud neobjasněných okolností. Podobně jako popsané varianty *FANCL* a *FANCM* může v nádorové predispozici fungovat popisovaná nonsense varianta v *BRCA2/FANCD1* c.9976A>T (p.K3326X), zkracující výsledný proteinový produkt o 93 aminokyselin. V kontrastu s většinou truncačních, vysoce penetrantních variant genu *BRCA2*, byla varianta K3326X označena za pouze mírně zvyšující riziko (OR=1,26). Tato alela zvyšuje riziko vzniku karcinomu jícnu, pankreatu, plic a nádorů ORL oblasti (Delahaye-Sourdeix et al. 2015).

U genu *FANCM*, kódujícího DNA helikázu, která přispívá k formování kotevního FA komplexu, jsme našli čtyři truncační varianty u pěti pacientů s karcinomem prsu (věk diagnózy 35-57 let). Truncační varianty *FANCM* byly nedávno popsány ve spojení s triple negativním karcinomem prsu (Kiiski et al. 2014). U tří pacientů (žádní z nich s triple negativním nádorem) jsme identifikovali dříve popsanou nonsense variantu (c.5101C>T), nebo „exon skipping“ variantu c.5791C>T. U obou těchto variant bylo dříve prokázáno zvýšení rizika vzniku karcinomu prsu (OR=1,86-3,93) (Kiiski et al. 2014, Peterlongo et al. 2015). Zbývající dvě námi identifikované *FANCM* mutace zahrnovaly nonsense variantu c.1972C>T (u pacientky s luminal B subtypem karcinomu prsu, jejíž matka a sestra trpěly bilaterálním karcinomem) a dosud nepopsanou variantu c.3979_3980delCA (u pacientky s mnohočetným výskytem karcinomu prsu a kolorektálního karcinomu v rodině). Asociace mezi kolorektálním karcinomem a germinálními mutacemi v genu *FANCM* byla nedávno popsána v nádorové tkáni (včetně delece somatické wt alely), získané ze dvou nosičů varianty c.5791C>T (Smith et al. 2013). S ohledem na nejnovější studie a výsledky naší analýzy se zdá, že vztah truncačních variant *FANCM* k predispozici ke karcinomem prsu, může být obdobný, jako je u truncačních variant genu *CHEK2* (Kiiski et al. 2014).

Identifikovali jsme také tři nosiče mutací v genu *RAD51C/FANCO* (0,9 % pacientů, věk 32, 35 a 41 let). Mutace *RAD51C* byly původně popsány jako varianty predisponující ke karcinomem ovaria (Loveday et al. 2012), nicméně, poslední data poukazují také na predispozici ke karcinomem prsu (Couch et al. 2015). Dva ze čtyř pacientů s karcinomem prsu byli nosiči nonsense mutace c.502A>T (p.R168X) a jeden sestřihové varianty c.905-2_1delAG, která způsobuje vynechání exonu 7 a posun čtecího rámce. Nedávno byly popsány dvě další patogenní varianty u dvou pacientek s karcinomem ovaria (Janatova et al. 2015). Tato data naznačují, že mutace v genu *RAD51C*, mohou ovlivňovat až 1 % vysoce rizikových, *BRCA1/BRCA2* negativních pacientů s karcinomem prsu, nebo ovaria v české populaci.

V naší studii byli též identifikováni dva nosiči mutací ve *FANCE* a *BRIP1/FANCI* s Her2+ karcinomem prsu. Obě mutace byly dříve reportovány ve spojitosti s karcinomem jícnu (Akbari et al. 2011) a triplenegativním karcinomem prsu (Couch et al. 2015).

Dále byly v rámci projektu nalezeny vzácné varianty v řadě dalších genů, kódujících DNA reparační proteiny (včetně *ATM*, *CHEK2*, *BLM*, *WRN*, *RAD51D* a *XRCC4*), účastníci se opravy dvouřetězcových zlomů DNA a které byly již dříve popsány ve spojitosti s karcinomem prsu. Zároveň byla také identifikována řada variant v genech kódujících proteiny účastnících se ostatních DNA reparačních drah. Mimo jiné, zajímavým kandidátním genem je *EXO1*, který kóduje exonukleázu zúčastněnou v množství různých procesů reparace DNA. Kromě dvou indel variant (u pacientek s karcinomem prsu diagnostikovaných ve věku 29 a 44 let), byla identifikována popsána sestřihová varianta c.2212-1G>C (u pacientky s bazálním karcinomem prsu s diagnózou ve 37 letech), která způsobuje in-frame delecii šesti aminokyselin účastnících se interakce *EXO1* s *MSH2* v průběhu procesu DNA mismatch repair (Rasmussen et al. 2000). V rozporu s naší analýzou RNA, popsali Wu a kolektiv tutéž variantu jako variantu posunující čtecí rámeček u pacienta s hereditárním nonpolypózním karcinomem (Wu et al. 2001). Přestože jsme nenalezli další PCR fragmenty při analýze RNA (resp. cDNA) z lymfocytů nosičky této mutace (Obrázek 23), nelze vyloučit přítomnost alternativních procesů sestřihu mRNA, jejichž produkty mohou být degradovány prostřednictvím nonsense mediated decay (NMD). V genu *EXO1* jsme navíc identifikovali dvě prioritizované missense varianty (p.E109K a p.S369R) u pěti pacientů s karcinomem prsu a pozitivní rodinou historií. U tří z nich došlo zároveň k rozvoji druhotného nádorového onemocnění (Hodgkinův lymfom, medulární karcinom štítné žlázy a karcinom ovaria). Kombinační analýza přítomnosti mutací v *EXO1* a zároveň v dalších genech, zúčastněných např. v NER (nucleotide excision repair; *ERCC2*, *ERCC6*), naznačuje, že narušení těchto reparačních procesů může zvyšovat nádorovou predispozici, včetně predispozice ke vzniku karcinomu prsu. Míra s jakou mohou tyto vzácné varianty přispívat k rozvoji karcinomu prsu a dalších nádorů, bude nepochybně předmětem dalších studií. Takové bádání je nezbytné především pro geny, u kterých nebyly doposud nalezeny žádné dědičné varianty (např. *MGMT*).

V rámci mezinárodní spolupráce byly nalezeny trunkační, potenciálně kauzální varianty výše zmíněného genu ***ERCC2***. Krom významné funkce *ERCC2* v NER opravách genomové DNA (Takayama et al. 1995), hraje tento protein důležitou roli v bazální buněčné transkripci jako součást transkripčního faktoru IIH (Singh et al. 2015). Nejlépe popsané dědičné onemocnění spojené s bialelickými mutacemi v *ERCC2* je Xeroderma Pigmentosum, typ D (XPD, OMIM278730), syndrom charakterizovaný extrémní fotosenzitivitou kůže a časným vznikem kožních tumorů (Lehmann et al.

2014). Proto je *ERCC2* nadějným kandidátním nádorově predispozičním genem. Na druhé straně mohou bialelické mutace *ERCC2* také vést k syndromům bez zvýšené predispozice k vývoji nádorů, například Trichothiodystrofie 1 (TTD; OMIM 601675) a cerebrookulofascioskeletální syndrom (COFS2; OMIM610756). Tyto skutečnosti naznačují, že ne všechny funkčně relevantní mutace *ERCC2* zvyšují nádorovou predispozici u nosičů. Z výsledků tohoto projektu také vyplývá, že územně specifické soubory kontrol, přestože mohou být malé, jsou významnější a užitečnější, než globální multietnické soubory kontrol. Jelikož genotypová data umožňují lokalizovat geografický původ daného jedince v rozmezí stovek kilometrů (Novembre et al. 2008), měl by být termín „regionálně specifický“ definovaný jako „méně, než 300 km od místa náběru“. Proto mají regionálně specifické kontroly ještě větší významnost, než populačně specifické kontroly, neboť populace se mezi sebou mísí, obzvláště v regionech blízkých k hranicím. Například varianta p.Phe568fs byla zachycena pouze jednou v německém souboru pacientů s BC/OC a ani jednou mezi 1844 německými kontrolami. Na základě populačně specifických dat bychom byli nadšeni z takového nálezu. Nicméně tento německý pacient pocházel z Drážďan, vzdálených 118km od Prahy a stejná mutace byla nalezena dvakrát v malém souboru 105 nenádorových kontrol našeho Pražského souboru. Tato zjištění vyzdvihují význam regionálních kontrol a multinárodních studií pro spolehlivou evaluaci variant.

Za účelem identifikace genetických variant i u genů, jejichž proteinové produkty se **neúčastní reparační DNA**, bylo sekvenováno 448 genů, které byly zmíněny ve studiích, zabývajících se karcinomem prsu. Podobně jako u DNA reparačních genů, byla pozornost zaměřena na přítomnost potencionálně škodlivých dědičných variant, které zahrnovaly truncační mutace a sestřihové alterace. Takové varianty byly nalezeny u 48 genů. Dotčené geny byly ručně rozděleny do osmi funkčně příbuzných skupin.

Nejčastější varianty (u 16 pacientů) byly nalezeny v genech, které kódují enzymy zúčastněné v syntéze steroidních hormonů, jejich obratu a v ER β (*ESR2*), který zprostředkovává intracelulární signalizaci závislou na estrogenu. Tato skupina obsahovala především geny z rodiny cytochromu p450, jejichž produkty přispívají k biosyntéze estrogenu (*CYP11A1*, *CYP17A1*, *CYP19A1*), nebo jeho metabolismu (*CYP3A5*, *CYP1A2*), jak bylo shrnuto v (Blackburn et al. 2015). Jelikož estrogen a jeho metabolity ovlivňují vznik karcinomu prsu prostřednictvím buněčné proliferace, mohou CYP geny účastníci se metabolismu estrogenu přispívat k riziku vzniku karcinomu prsu. Je s podivem, že přestože řada studií analyzovala polymorfizmy, nebo expresi CYP genů u pacientů s karcinomem prsu, nebyla doposud učiněna žádná systematická mutační analýza těchto genů u nádorových pacientů. Varianty v *CYP11A1* a *CYP17A1* nalezené u pacientů s Her2+ karcinomem prsu byly dříve popsány u

homozygotů, nebo složených heterozygotů s adrenální insuficiencí (Hiort et al. 2005) a kongenitální adrenální hyperplazií (OMIM 201910) (Hwang et al. 2011). Hopper a kol. popsali heterozygotní mutaci p.R239* v genu *CYP17A1* u tří *BRCA1/2* negativních mladých sester s karcinomem prsu a hypotetizují, že tato varianta je zodpovědná za dominantně dědičný a pravděpodobně vysoce rizikový karcinom prsu (Hopper et al. 2005). Nedávno popsali Yang a kol. mutaci p.Y329* v genu *CYP17A1* u pacienta z rodiny s hereditárním výskytem karcinomu prsu a ovarií (Yang et al. 2015). V našem souboru byla nalezena doposud nepopsaná varianta c.1058dupT v genu *CYP19A1* u pacientky s duplicitou karcinomu prsu a Nonhodgkinským lymfomem (diagnózy ve 44 a 40 letech), jejíž matka vyvinula bilaterální karcinom prsu (v 64 a 66 letech). Bialelické mutace na podobné pozici v *CYP19A1* způsobují deficit aromatázy (OMIM #613546). Defekty v enzymech katabolizujících estrogen naznačují mechanisticky jasnější patofyziologické spojení s vznikem karcinomu prsu. Jelikož jsou estrogény známými substráty *CYP35A* a *CYP1A2* (Blackburn et al. 2015), může mít 11 nosičů nalezených trunkačních variant v těchto genech potencionálně sníženou clearance estrogenu. V našem souboru byla také nalezena varianta v genu *NQO2*, kódující chinon reduktázu, která eliminuje chinony estrogenu zodpovědné za karcinogenezi způsobenou estrogenem (Gaikwad et al. 2009). U jednoho pacienta byla nalezena mutace v genu *ESR2*, který kóduje Erβ s antiproliferativní signalizací (Caiazza et al. 2015). Velké zastoupení pacientů s konstitutivními mutacemi v genech metabolismu steroidních hormonů podporuje hypotéza Hopper a kol. (Hopper et al. 2005), který navrhuje, že kauzální mutace v genech metabolismu steroidních hormonů, mohou představovat nový patofyziologický mechanismus spojující genetické a environmentální interakce s nádorovou predispozicí.

Další funkční skupiny spojující pacienty s trunkačními variantami ve funkčně relevantních genech byly menší, s výjimkou genů kódujících proteiny imunitní odpovědi a membránové signalizace (obě obsahovaly 11 nosičů trunkačních variant). Nicméně je evidentní, že přinejmenším některé nalezené varianty v genech neúčastnících se DNA reparace mají velice omezený (zdali vůbec nějaký) dopad na predispozici ke karcinomu prsu a spíše představují náhodné nálezy raritních variant přispívajících k jiným onemocněním. Příkladem mohou být mutace v genu *ABCC2*, identifikované již dříve u pacientů s Dublin-Johnsonovým syndromem (Pacifico et al. 2010), které způsobují zvýšené hladiny bilirubinu v séru s protektivním účinkem proti karcinogenezi (Sticova et al. 2013). Podobně známá mutace v genu *DPYD* se vztahem k toxicitě fluoropyrimidinů (Kleibl et al. 2009). Reportování těchto náhodných nálezů je značně diskutabilní a přemětem intenzivní debaty (Blackburn et al. 2015).

Při posouzení pacientů a charakteristiky onemocnění u nosičů patogenních mutací v hlavních predispozičních genech *BRCA1* a *BRCA2*, které se vyskytují u pacientů s časným nástupem onemocnění, u agresivnějších subtypů karcinomu prsu, či zvýšeného výskytu familiárních nádorových onemocnění, by se dalo očekávat totéž u nosičů mutací v dalších predispozičních genech. Ve skutečnosti jsme nezaznamenali významné rozdíly v klinických a histopatologických charakteristikách mezi nosiči trunkačních mutací v DNA reparačních genech, genech účastnících se reparační DNA a u pacientů bez trunkačních variant. Tento výsledek se nezměnil, ani když byly do porovnávání přidány prioritizované varianty (data nejsou součástí práce). Toto pozorování může být ovlivněno několika faktory. Kromě toho, že analyzovaný soubor obsahoval limitované množství jedinců vysoce preselektované populace pacientů, mají ostatní predispoziční, nebo kandidátní geny, nižší penetranci než *BRCA1* a *BRCA2*. Ve velké studii čítající 1 824 triple negativních pacientů s karcinomem prsu (Couch et al. 2015), která sledovala dědičné mutace v 17 genech, byly významné rozdíly v obohacení o rodinnou historii karcinomu prsu a ovarií a v charakteristikách nádoru, nalezeny pouze u nosičů mutací v genech *BRCA1* a *BRCA2*, nikoliv však u nosičů mutací v ostatních predispozičních genech. Oproti hlavním predispozičním genům karcinomu prsu, které silně ovlivňují primárně riziko vzniku karcinomu prsu, mohou ostatní geny, či dokonce jejich konkrétní varianty, ovlivňovat (se stejnou mírou a s omezenou penetrancí) mnohem větší spektrum nádorových diagnóz. Tento jev byl zdokumentován například u několika germinálních mutací genu *CHEK2*, které u nosičů zvyšují riziko karcinomu prsu, prostaty, kolorekta, štítné žlázy, nebo ledvin, ale snižují riziko vzniku některých nádorů dýchacího ústrojí (Narod 2010). Z tohoto hlediska bude nutná obezřetná evaluace rodinných výskytů nádorových onemocnění a segregčních analýz u studovaných variant, pro správné určení klinické relevance nových genů predisponujících ke karcinomu prsu. Vzhledem k tomu, že frekvence variant všech kandidátních predispozičních genů (s minimálně střední penetrancí) jsou o řád nižší, než frekvence v genech *BRCA1/2*, bude pro evaluaci klinické aplikovatelnosti nových predispozičních genů zapotřebí úsilí mezinárodních konsorcií.

Jedním z vedlejších poznatků této práce a podnět k dalšímu studiu je náhodný nález **varianty c.657del5 v genu *NBN*** u pacienta s karcinomem pankreatu a následná analýza na rozšířeném souboru pacientů s tímto onemocněním a na kontrolní nenádorové populaci. Doposud nejvyšší frekvence nosičů mutace v genu *NBN* (3,7 % pacientů) byla nalezena u pacientů s karcinomem prsu ze střední a východní Evropy (Gorski et al. 2003). Pozdější metaanalýzy ukázaly, že c.657del5 je středně penetrantní alela pro karcinom prsu (OR = 2.51; 95%CI: 1.68–3.73) a lymfom (OR = 2.93; 95%CI: 1.62–5.29) a zároveň silně zvyšuje riziko karcinomu prostaty (OR=5.87; 95%CI: 2.51–13.75) (Gao et al. 2013). Asociace germinálních mutací v *NBN* s predispozicí ke karcinomu prsu vedla

k zahrnutí *NBN* do NGS analýz pomocí multinádorových sekvenačních panelů (Couch et al. 2015). Zatímco ve dvou studiích analyzujících varianty pomocí panelového testování u pacientů s karcinomem pankreatu nebyly nalezeny žádné trunkační mutace v genu *NBN* (Grant et al. 2015, Roberts et al. 2016), v práci Hu a kolektivu byla nalezena jednou varianta c.657del5 mezi 96 pacienty (Hu et al. 2016). Recentně provedla skupina Lener a kol. analýzu deseti prevalentních zakladatelských mutací v genech *BRCA1*, *CHEK2*, *PALB2* a *NBN* (vč. c.657del5) u 383 pacientů s karcinomem pankreatu a našli 8 nosičů této mutace (2,09 %), což naznačuje zvýšené riziko karcinomu u nosičů varianty c.657del5 v polské populaci (OR = 3.8; 95%CI: 1.68–8.60) (Lener et al. 2016). Naše výsledky a výsledky práce Lener a kol. naznačují, že varianta c.657del5 může být nová predisponující alela významně zvyšující riziko karcinomu pankreatu. Kombinované OR vypočtené ze souhrnného počtu vzorků naší a Lener a kol. práce (624 pacientů s karcinomem pankreatu – 13 nosičů varianty c.657del5 a 4915 kontrol – 24 nosičů) činí 4,33 (95%CI 2.2–8.56; $p < 0.001$) (Borecka et al. 2016). Nicméně bude zapotřebí dalších studií s většími soubory testovaných osob včetně segregáčnických analýz pro potvrzení tohoto pozorování. Jejich důležitým hlediskem bude cílený a především vyrovnaný výběr pacientů a etnicky rovnocenného souboru kontrol, protože analýza bude jistě významně ovlivněna efektem zakladatelské mutace c.657del5 nabohacené v prostoru Střední Evropy.

Spojujícím tématem předkládané dizertační práce je použití NGS metod v diagnostice a aplikovaném výzkumu při identifikaci dědičných nádorových onemocnění. Výsledky této práce byly východiskem pro vývoj klinicky aplikovaného panelu CZE CANCA (CZEch CAncer panel for Clinical Application) (Soukupova et al. 2018) sloužícího pro identifikaci kauzálních variant u jedinců z rodin s výskytem hereditárních nádorových onemocnění. Při jeho vývoji se plně uplatnily zkušenosti a optimalizované postupy designu, laboratorní metodiky i bioinformatické analýzy panelového sekvenování, které jsou předmětem předkládané práce. CZE CANCA panel je od roku 2015 využíván klinickými genetickými laboratořemi v České republice, které se podílejí na tvorbě konsensu zpracování a interpretace sekvenačních dat prostřednictvím pravidelných pracovních setkání. Do současné doby se do projektu CZE CANCA zapojilo 8 laboratoří a bylo vyšetřeno více, než 5 tisíc klinických a výzkumných vzorků. Podařilo se tak posunout genetické vyšetřování hereditárních nádorů od časově nesmírně náročného testování jednotlivých genů k rychlé multigenové analýze zajišťující identifikaci kauzální predispozice u vysoce rizikových osob a rodin, kterým lze nabídnout terapeutickou a preventivní péči. Díky projektu CZE CANCA se daří unifikovat poskytovanou péči genetického testování z hlediska kvality, rozsahu vyšetřovaných genů a především klinické interpretace.

8 ZÁVĚR

Incidence karcinomu prsu má v České republice trvale rostoucí tendenci. Vznik tohoto nádorového onemocnění je následkem hromadění alterací genů, které jsou z většiny účastny v reparaci genomové DNA a regulaci buněčného cyklu. Za významnou část hereditárních případů karcinomu prsu jsou zodpovědné dědičné alterace genů *BRCA1* a *BRCA2* a několika málo dalších genů s vysokou, či střední penetrancí. Značná část případů tohoto onemocnění však zůstává bez objasněné genetické příčiny.

Cílem práce byla identifikace alterací genů predisponujících ke vzniku dědičné formy karcinomu prsu v české populaci pacientek s mnohočetným výskytem tohoto onemocnění v rodině. Analýza byla zaměřena také na otázku významu alterací v genech kódujících proteiny podílejících se na opravách genomové DNA.

Předkládaná práce shrnuje obsáhlou analýzu genů predisponujících ke karcinomu prsu v české populaci vysoce rizikových pacientů s karcinomem prsu, kteří byli negativně testováni na přítomnost mutací v genech *BRCA1/BRCA2/PALB2*. Tato analýza identifikovala trunkační mutace u 32 % pacientů. Pouze 9 % pacientů byli nosiči trunkačních variant v genech zahrnutých do současných klinických NGS panelů využívaných v predikci karcinomu prsu. Výsledky práce také ukazují zvýšené zastoupení varianty c.1096_1099dupATTA v genu *FANCL* u vysoce rizikových pacientů. Tato skutečnost ukazuje, že *FANCL* (nebo spíše pouze tato varianta) může obsahovat novou predispoziční alelu pro karcinom prsu, jejíž významnost je nezbytné potvrdit většími studii. Dále je zajímavé, že byly identifikovány i trunkační varianty ve skupině genů cytochromu p450, které kódují enzymy metabolismu steroidních hormonů, u 5 % pacientů s karcinomem prsu. Proto může tato funkční skupina genů přispět k vysvětlení doposud neodhalené, chybějící dědičnosti karcinomu prsu u vysoce rizikových pacientů. Z práce vyplynul i vedlejší projekt, při kterém byla po vyšetření rozšířeného souboru identifikována varianta c.657del5 genu *NBN* jako potencionálně predispoziční pro karcinom pankreatu. Nicméně pro potvrzení výše vyřčených hypotéz, bude nezbytné provedení dalších studií, převážně z důvodu omezených počtů testovaných osob. Výsledky projektu byly také součástí mezinárodní spolupráce evaluující varianty genu *ERCC2*, ze které vyplynula jasná potřeba vyšetřování genetického pozadí pomocí sekvenčních projektů kontrolních osob a multinárodních studií pro spolehlivou evaluaci variant. Práce potvrzuje vhodnost zvoleného přístupu panelového NGS pro identifikaci obecných i populačně specifických variant genů predisponujících k nádorovému onemocnění, v tomto případě hereditárnímu karcinomu prsu a zároveň identifikaci možných nových kandidátních genů. Zkušenosti nabyté při vývoji a analýze dat tohoto projektu byly úspěšně implementovány v rámci klinické aplikace panelu CZEKANCA, který je v současnosti významným diagnostickým nástrojem pro vyšetření hereditárních nádorových syndromů v České republice.

9 PODĚKOVÁNÍ

Na tomto místě bych rád poděkoval především svému školiteli Doc. MUDr. Zdeňku Kleiblovi, PhD., který mě uvedl do problematiky analýzy genů přispívajících ke vzniku nádorových onemocnění prsu a ovarií. Velmi si cením jeho rad, podpory a taktního směřování k cíli.

Za cenné rady a inspirující debaty, které významně ovlivnily mé studium a práci bych dále chtěl poděkovat i ostatním spolupracovníkům a kamarádům z naší laboratoře a stejně tak ze spolupracující Laboratoře lékařské genomiky a bioinformatiky z Ústavu dědičných metabolických poruch.

V neposlední řadě děkuji svým nejbližším za vytvoření příjemného zázemí, bez kterého by výše popsané stránky nikdy nevznikly.

V Modřanech dne 1. května 2018.

10 SEZNAM ZKRATEK

Zkratka	Plný název
53BP1	Tumor suppressor p53-binding protein 1
APE1	apurinic/apyrimidinic endonuclease 1
APLF	Aprataxin and PNK-like factor
ATM	Ataxia telangiectasia mutated
ATR	Ataxia telangiectasia and Rad3-related protein
ATRIP	ATR-Interacting Protein
BC	breast cancer (karcinom prsu)
BER	Base Excision Repair
BRCA1	Breast cancer 1
CDC25	Ras-specific guanine nucleotide-releasing factor 1
cDNA	komplementární DNA (complementary DNA)
CLSPN	Claspin
cPAL	Combinatorial probe anchor ligation
CSA/ERCC8	Cockayne syndrome A
CSB/ERCC6	Cockayne syndrome B
CSR	Class Switch Recombination
ddNTP	dideoxyribonukleotidtrifosfát
DDR	DNA Damage Response
DNA-PK	DNA-dependent protein kinase
ER	estrogenní receptor
ESP	Exome Sequencing Project
ExAC	Exome Aggregation Consortium
FA	Fanconiho anemie
FEN1	flap endonuclease 1
gDNA	Genomová DNA
GGR	global genome repair, globální opravy genomu
GWAS	Genome Wide Association Studies
GWAS	Genome Wide Association Studies
HER2/neu	Human epidermal growth factor receptor 2; ErbB-2
HGVS	Human Genome Variation Society
HNPCC	hereditární nepolypózní kolorektální karcinom
HR	homologní rekombinace
HRM	High-Resolution Melting
HTS	high-throughput DNA sequencing
CHEK1	Checkpoint kinase 1
ICL	interstrand crosslinks, meziřetězcové spoje, zesíťování DNA
Ku70	Lupus Ku autoantigen protein p70
Ku80	Lupus Ku autoantigen protein p86
LM-PCR	ligation-mediated PCR
MDC1	Mediator of DNA damage checkpoint protein 1
MGMT	Methylated-DNA-protein-cysteine methyltransferase
MLPA	Multiple ligation-dependent probe amplification

MMR	Mismatch Repair
mRNA	messenger RNA
MSH2	MutS homolog 2
MSH6	MutS homolog 6
NER	Nucleotide excision repair
NES	Nuclear Export Signal
NGS	next-generation sequencing
NHEJ	Non-Homologous End Joining
NLS	Nuclear Localization Signal
PARP	poly ADP-ribóza polymeráza
PARP	poly ADP-ribóza polymeráza
PBS	phosphate buffered saline
PCNA	proliferating cell nuclear antigen
PCR	polymerázová řetězová reakce (polymerase chain reaction)
PIKK	phosphatidylinositol-3 kinase-like kinase
RAD51	DNA repair protein RAD51 homolog 1
RBBP8	CtBP-interacting protein
RCA	"rolling circle" amplifikace
RFC	replication factor C
RPA	Replication protein A 70 kDa DNA-binding subunit
RPA	Replication Protein A
RT-PCR	Zpětná PCR (Reverse transcription PCR)
SMRT	Single-molecule real-time
SNP	single nucleotide polymorphism
SNP	Single-Nucleotide Polymorphism
SSB	Single Strand Breaks
TBE	Tris/Borate/EDTA pufr
TCR	transcription-coupled repair, opravy spřažené s transkripcí
TE	Tris/EDTA pufr
TLS	translesion synthesis
TopBP1	DNA Topoisomerase II-Binding Protein 1
TTD	trichothiodystrofie
UICC	Union for International Cancer Control
ÚZIS	Ústav zdravotnických informací a statistiky ČR
wt	normální alela (wild type)
XP	xeroderma pigmentosum
XPA	xeroderma pigmentosum complementation group A
XPF/ERCC1	xeroderma pigmentosum complementation group F/excision repair, cross-complementation group 1
XPG	xeroderma pigmentosum complementation group G
ZMW	zero-mode waveguides

11 SEZNAM PŘÍLOH

Příloha	Obsah
Příloha I	Seznam cílových genů
Příloha II	Seznam a sekvence primerů použitých pro confirmaci a charakterizaci nalezených zárodečných variant
Příloha III	Identifikované truncační varianty (nonsense, inzerce/delece, a ovlivňující splicing) a prioritizované missense varianty
Příloha IV	Seznam publikací
Příloha IV-1	Lhota <i>et al.</i> ; Hereditary truncating mutations of DNA repair and other genes in <i>BRCA1/BRCA2/PALB2</i> -negatively tested breast cancer patients
Příloha IV-2	Zemankova <i>et al.</i> ; RE: frameshift variant <i>FANCL</i> *c.1096_1099dupATTA is not associated with high breast cancer risk
Příloha IV-3	Rump <i>et al.</i> ; Identification and Functional Testing of <i>ERCC2</i> Mutations in a Multi-national Cohort of Patients with Familial Breast- and Ovarian Cancer
Příloha IV-4	Borecka <i>et al.</i> ; The c.657del5 variant in the <i>NBN</i> gene predisposes to pancreatic cancer
Příloha IV-5	Hojny <i>et al.</i> ; Multiplex PCR and NGS-based identification of mRNA splicing variants: Analysis of <i>BRCA1</i> splicing pattern as a model

12 POUŽITÁ LITERATURA

- Adank, M. A., M. A. Jonker, I. Kluijt, S. E. van Mil, R. A. Oldenburg, et al. (2011). CHEK2*1100delC homozygosity is associated with a high breast cancer risk in women. *J Med Genet* **48**(12): 860-863.
- Adzhubei, I. A., S. Schmidt, L. Peshkin, V. E. Ramensky, A. Gerasimova, et al. (2010). A method and server for predicting damaging missense mutations. *Nat Methods* **7**(4): 248-249.
- Airewele, G. E., A. J. Sigurdson, K. J. Wiley, B. E. Frieden, L. W. Caldarera, et al. (2001). Neoplasms in neurofibromatosis 1 are related to gender but not to family history of cancer. *Genet Epidemiol* **20**(1): 75-86.
- Akbari, M. R., R. Malekzadeh, P. Lepage, D. Roquis, A. R. Sadjadi, et al. (2011). Mutations in Fanconi anemia genes and the risk of esophageal cancer. *Hum Genet* **129**(5): 573-582.
- Akbari, M. R., P. Tonin, W. D. Foulkes, P. Ghadirian, M. Tischkowitz, et al. (2010). RAD51C germline mutations in breast and ovarian cancer patients. *Breast Cancer Res* **12**(4): 404.
- Ali, A. M., M. Kirby, M. Jansen, F. P. Lach, J. Schulte, 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): E761-770.
- Alpi, A. F., P. E. Pace, M. M. Babu and K. J. Patel (2008). Mechanistic insight into site-restricted monoubiquitination of FANCD2 by Ube2t, FANCL, and FANCI. *Mol Cell* **32**(6): 767-777.
- Antoniou, A., P. D. Pharoah, S. Narod, H. A. Risch, J. E. Eyfjord, et al. (2003). Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case Series unselected for family history: a combined analysis of 22 studies. *Am J Hum Genet* **72**(5): 1117-1130.
- Antoniou, A. C., C. Kartsonaki, O. M. Sinilnikova, P. Soucy, L. McGuffog, et al. (2011). Common alleles at 6q25.1 and 1p11.2 are associated with breast cancer risk for BRCA1 and BRCA2 mutation carriers. *Hum Mol Genet* **20**(16): 3304-3321.
- Antoniou, A. C., A. B. Spurdle, O. M. Sinilnikova, S. Healey, K. A. Pooley, et al. (2008). Common breast cancer-predisposition alleles are associated with breast cancer risk in BRCA1 and BRCA2 mutation carriers. *Am J Hum Genet* **82**(4): 937-948.
- Apostolou, P. and F. Fostira (2013). Hereditary breast cancer: the era of new susceptibility genes. *Biomed Res Int* **2013**: 747318.
- Baas, A. F., M. Gabbett, M. Rimac, M. Kansikas, M. Raphael, et al. (2013). Agenesis of the corpus callosum and gray matter heterotopia in three patients with constitutional mismatch repair deficiency syndrome. *Eur J Hum Genet* **21**(1): 55-61.
- Barrett, S. V. (2010). Breast cancer. *J R Coll Physicians Edinb* **40**(4): 335-338; quiz 339.
- Bartkova, J., J. Tommiska, L. Oplustilova, K. Aaltonen, A. Tamminen, et al. (2008). Aberrations of the MRE11-RAD50-NBS1 DNA damage sensor complex in human breast cancer: MRE11 as a candidate familial cancer-predisposing gene. *Mol Oncol* **2**(4): 296-316.
- Bartoňková, H., L. Foretová, E. Helmichová, R. Kalábová, Z. Kleibl, et al. (2003). Doporučené zásady péče o nemocné s nádory prsu a vaječníků a zdravé osoby se zárodečnými mutacemi genů BRCA1 nebo BRCA2. *Klinická onkologie*(16): 28-34.
- Bentley, D. R. (2006). Whole-genome re-sequencing. *Curr Opin Genet Dev* **16**(6): 545-552.
- Berg, J. W. and R. V. Hutter (1995). Breast cancer. *Cancer* **75**(1 Suppl): 257-269.
- Bernstein, K. A. and R. Rothstein (2009). At loose ends: resecting a double-strand break. *Cell* **137**(5): 807-810.
- Bhattacharjee, S. and S. Nandi (2017). DNA damage response and cancer therapeutics through the lens of the Fanconi Anemia DNA repair pathway. *Cell Commun Signal* **15**(1): 41.
- Blackburn, H. L., D. L. Ellsworth, C. D. Shriver and R. E. Ellsworth (2015). Role of cytochrome P450 genes in breast cancer etiology and treatment: effects on estrogen biosynthesis, metabolism, and response to endocrine therapy. *Cancer Causes Control* **26**(3): 319-332.
- Blackburn, H. L., B. Schroeder, C. Turner, C. D. Shriver, D. L. Ellsworth, et al. (2015). Management of Incidental Findings in the Era of Next-generation Sequencing. *Curr Genomics* **16**(3): 159-174.
- Bloom, H. J. and W. W. Richardson (1957). Histological grading and prognosis in breast cancer; a study of 1409 cases of which 359 have been followed for 15 years. *Br J Cancer* **11**(3): 359-377.
- Bogdanova, N., S. Feshchenko, P. Schurmann, R. Waltes, B. Wieland, et al. (2008). Nijmegen Breakage Syndrome mutations and risk of breast cancer. *Int J Cancer* **122**(4): 802-806.
- Bojesen, S. E., K. A. Pooley, S. E. Johnatty, J. Beesley, K. Michailidou, et al. (2013). Multiple independent variants at the TERT locus are associated with telomere length and risks of breast and ovarian cancer. *Nat Genet* **45**(4): 371-384, 384e371-372.
- Borecka, M., P. Zemankova, F. Lhota, J. Soukupova, P. Kleiblova, et al. (2016). The c.657del5 variant in the NBN gene predisposes to pancreatic cancer. *Gene* **587**(2): 169-172.

- Botta, E., T. Nardo, B. C. Broughton, S. Marinoni, A. R. Lehmann, et al. (1998). Analysis of mutations in the XPD gene in Italian patients with trichothiodystrophy: site of mutation correlates with repair deficiency, but gene dosage appears to determine clinical severity. *Am J Hum Genet* **63**(4): 1036-1048.
- Boulton, S. J. (2006). BRCA1-mediated ubiquitylation. *Cell Cycle* **5**(14): 1481-1486.
- Broca (1866). *Traité des Tumeurs*. Paris, P. Asselin.
- Broughton, B. C., M. Berneburg, H. Fawcett, E. M. Taylor, C. F. Arlett, et al. (2001). Two individuals with features of both xeroderma pigmentosum and trichothiodystrophy highlight the complexity of the clinical outcomes of mutations in the XPD gene. *Hum Mol Genet* **10**(22): 2539-2547.
- Bunting, S. F. and A. Nussenzweig (2010). Dangerous liaisons: Fanconi anemia and toxic nonhomologous end joining in DNA crosslink repair. *Mol Cell* **39**(2): 164-166.
- Buys, S. S., J. F. Sandbach, A. Gammon, G. Patel, J. Kidd, et al. (2017). A study of over 35,000 women with breast cancer tested with a 25-gene panel of hereditary cancer genes. *Cancer* **123**(10): 1721-1730.
- Caestecker, K. W. and G. R. Van de Walle (2013). The role of BRCA1 in DNA double-strand repair: past and present. *Exp Cell Res* **319**(5): 575-587.
- Caiazza, F., E. J. Ryan, G. Doherty, D. C. Winter and K. Sheahan (2015). Estrogen receptors and their implications in colorectal carcinogenesis. *Front Oncol* **5**: 19.
- Ceccaldi, R., P. Sarangi and A. D. D'Andrea (2016). The Fanconi anaemia pathway: new players and new functions. *Nat Rev Mol Cell Biol* **17**(6): 337-349.
- Ciccia, A. and S. J. Elledge (2010). The DNA damage response: making it safe to play with knives. *Mol Cell* **40**(2): 179-204.
- Ciccia, A., N. McDonald and S. C. West (2008). Structural and functional relationships of the XPF/MUS81 family of proteins. *Annu Rev Biochem* **77**: 259-287.
- Cimprich, K. A. and D. Cortez (2008). ATR: an essential regulator of genome integrity. *Nat Rev Mol Cell Biol* **9**(8): 616-627.
- Clague, J., G. Wilhoite, A. Adamson, A. Bailis, J. N. Weitzel, et al. (2011). RAD51C germline mutations in breast and ovarian cancer cases from high-risk families. *PLoS One* **6**(9): e25632.
- Cleaver, J. E., L. H. Thompson, A. S. Richardson and J. C. States (1999). A summary of mutations in the UV-sensitive disorders: xeroderma pigmentosum, Cockayne syndrome, and trichothiodystrophy. *Hum Mutat* **14**(1): 9-22.
- Colombo, M., M. J. Blok, P. Whiley, M. Santamarina, S. Gutierrez-Enriquez, 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): 3666-3680.
- Cooper, G. M., E. A. Stone, G. Asimenos, N. C. S. Program, E. D. Green, et al. (2005). Distribution and intensity of constraint in mammalian genomic sequence. *Genome Res* **15**(7): 901-913.
- Couch, F. J., S. N. Hart, P. Sharma, A. E. Toland, X. Wang, et al. (2015). Inherited mutations in 17 breast cancer susceptibility genes among a large triple-negative breast cancer cohort unselected for family history of breast cancer. *J Clin Oncol* **33**(4): 304-311.
- Couch, F. J., H. Shimelis, C. Hu, S. N. Hart, E. C. Polley, et al. (2017). Associations Between Cancer Predisposition Testing Panel Genes and Breast Cancer. *JAMA Oncol* **3**(9): 1190-1196.
- Cox, A., A. M. Dunning, M. Garcia-Closas, S. Balasubramanian, M. W. Reed, et al. (2007). A common coding variant in CASP8 is associated with breast cancer risk. *Nat Genet* **39**(3): 352-358.
- Dai, Y. and S. Grant (2010). New insights into checkpoint kinase 1 in the DNA damage response signaling network. *Clin Cancer Res* **16**(2): 376-383.
- Daly, M. B., R. Pilarski, M. Berry, S. S. Buys, M. Farmer, et al. (2017). NCCN Guidelines Insights: Genetic/Familial High-Risk Assessment: Breast and Ovarian, Version 2.2017. *J Natl Compr Canc Netw* **15**(1): 9-20.
- Damiola, F., M. Pertesi, J. Oliver, F. Le Calvez-Kelm, C. Voegelé, et al. (2014). Rare key functional domain missense substitutions in MRE11A, RAD50, and NBN contribute to breast cancer susceptibility: results from a Breast Cancer Family Registry case-control mutation-screening study. *Breast Cancer Res* **16**(3): R58.
- Davies, A. A., J. Y. Masson, M. J. McIlwraith, A. Z. Stasiak, A. Stasiak, et al. (2001). Role of BRCA2 in control of the RAD51 recombination and DNA repair protein. *Mol Cell* **7**(2): 273-282.
- de Boer, J., I. Donker, J. de Wit, J. H. Hoeijmakers and G. Weeda (1998). Disruption of the mouse xeroderma pigmentosum group D DNA repair/basal transcription gene results in preimplantation lethality. *Cancer Res* **58**(1): 89-94.
- de Groote, F. H., J. G. Jansen, Y. Masuda, D. M. Shah, K. Kamiya, et al. (2011). The Rev1 translesion synthesis polymerase has multiple distinct DNA binding modes. *DNA Repair (Amst)* **10**(9): 915-925.
- Deans, A. J. and S. C. West (2009). FANCM connects the genome instability disorders Bloom's Syndrome and Fanconi Anemia. *Mol Cell* **36**(6): 943-953.
- Delahaye-Sourdeix, M., D. Anantharaman, M. N. Timofeeva, V. Gaborieau, A. Chabrier, et al. (2015). A rare truncating BRCA2 variant and genetic susceptibility to upper aerodigestive tract cancer. *J Natl Cancer Inst* **107**(5).
- di Masi, A. and A. Antoccia (2008). NBS1 Heterozygosity and Cancer Risk. *Curr Genomics* **9**(4): 275-281.

- Dong, X., L. Wang, K. Taniguchi, X. Wang, J. M. Cunningham, et al. (2003). Mutations in CHEK2 associated with prostate cancer risk. *Am J Hum Genet* **72**(2): 270-280.
- Doniger, S. W., H. S. Kim, D. Swain, D. Corcuera, M. Williams, et al. (2008). A catalog of neutral and deleterious polymorphism in yeast. *PLoS Genet* **4**(8): e1000183.
- Downs-Holmes, C. and P. Silverman (2011). Breast cancer: overview & updates. *Nurse Pract* **36**(12): 20-26; quiz 27.
- Drmanac, R., A. B. Sparks, M. J. Callow, A. L. Halpern, N. L. Burns, et al. (2010). Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. *Science* **327**(5961): 78-81.
- Easton, D. F. (1994). Cancer risks in A-T heterozygotes. *Int J Radiat Biol* **66**(6 Suppl): S177-182.
- Easton, D. F., P. D. Pharoah, A. C. Antoniou, M. Tischkowitz, S. V. Tavtigian, et al. (2015). Gene-panel sequencing and the prediction of breast-cancer risk. *N Engl J Med* **372**(23): 2243-2257.
- Easton, D. F., K. A. Pooley, A. M. Dunning, P. D. Pharoah, D. Thompson, et al. (2007). Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* **447**(7148): 1087-1093.
- Eid, J., A. Fehr, J. Gray, K. Luong, J. Lyle, et al. (2009). Real-time DNA sequencing from single polymerase molecules. *Science* **323**(5910): 133-138.
- Ellsworth, R. E., D. J. Decewicz, C. D. Shriver and D. L. Ellsworth (2010). Breast cancer in the personal genomics era. *Curr Genomics* **11**(3): 146-161.
- Elston, C. W. and I. O. Ellis (1991). Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. *Histopathology* **19**(5): 403-410.
- Erkko, H., B. Xia, J. Nikkila, J. Schleutker, K. Syrjakoski, et al. (2007). A recurrent mutation in PALB2 in Finnish cancer families. *Nature* **446**(7133): 316-319.
- Esashi, F., N. Christ, J. Gannon, Y. Liu, T. Hunt, et al. (2005). CDK-dependent phosphorylation of BRCA2 as a regulatory mechanism for recombinational repair. *Nature* **434**(7033): 598-604.
- Evers, B. and J. Jonkers (2006). Mouse models of BRCA1 and BRCA2 deficiency: past lessons, current understanding and future prospects. *Oncogene* **25**(43): 5885-5897.
- Ewing, B. and P. Green (1998). Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res* **8**(3): 186-194.
- Ewing, B., L. Hillier, M. C. Wendl and P. Green (1998). Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* **8**(3): 175-185.
- Falck, J., J. Coates and S. P. Jackson (2005). Conserved modes of recruitment of ATM, ATR and DNA-PKcs to sites of DNA damage. *Nature* **434**(7033): 605-611.
- Fan, L., J. O. Fuss, Q. J. Cheng, A. S. Arvai, M. Hammel, et al. (2008). XPD helicase structures and activities: insights into the cancer and aging phenotypes from XPD mutations. *Cell* **133**(5): 789-800.
- Farmer, H., N. McCabe, C. J. Lord, A. N. Tutt, D. A. Johnson, et al. (2005). Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature* **434**(7035): 917-921.
- Flusberg, B. A., D. R. Webster, J. H. Lee, K. J. Travers, E. C. Olivares, et al. (2010). Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat Methods* **7**(6): 461-465.
- Foulkes, W. D. (2008). Inherited susceptibility to common cancers. *N Engl J Med* **359**(20): 2143-2153.
- Foulkes, W. D., P. Ghadirian, M. R. Akbari, N. Hamel, S. Giroux, et al. (2007). Identification of a novel truncating PALB2 mutation and analysis of its contribution to early-onset breast cancer in French-Canadian women. *Breast Cancer Res* **9**(6): R83.
- French, J. D., M. Ghossaini, S. L. Edwards, K. B. Meyer, K. Michailidou, et al. (2013). Functional variants at the 11q13 risk locus for breast cancer regulate cyclin D1 expression through long-range enhancers. *Am J Hum Genet* **92**(4): 489-503.
- Gaikwad, N. W., L. Yang, E. G. Rogan and E. L. Cavalieri (2009). Evidence for NQO2-mediated reduction of the carcinogenic estrogen ortho-quinones. *Free Radic Biol Med* **46**(2): 253-262.
- Gao, P., N. Ma, M. Li, Q. B. Tian and D. W. Liu (2013). Functional variants in NBS1 and cancer risk: evidence from a meta-analysis of 60 publications with 111 individual studies. *Mutagenesis* **28**(6): 683-697.
- Garber, J. E., A. M. Goldstein, A. F. Kantor, M. G. Dreyfus, J. F. Fraumeni, Jr., et al. (1991). Follow-up study of twenty-four families with Li-Fraumeni syndrome. *Cancer Res* **51**(22): 6094-6097.
- Goldhirsch, A., W. C. Wood, A. S. Coates, R. D. Gelber, B. Thurlimann, et al. (2011). Strategies for subtypes--dealing with the diversity of breast cancer: highlights of the St. Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2011. *Ann Oncol* **22**(8): 1736-1747.
- Golmard, L., L. Castera, S. Krieger, V. Moncoutier, K. Abidallah, et al. (2017). Contribution of germline deleterious variants in the RAD51 paralogs to breast and ovarian cancers. *Eur J Hum Genet* **25**(12): 1345-1353.
- Gomes, L. R., C. F. M. Menck and G. S. Leandro (2017). Autophagy Roles in the Modulation of DNA Repair Pathways. *Int J Mol Sci* **18**(11).
- Gorski, B., T. Debniak, B. Masojc, M. Mierzejewski, K. Medrek, et al. (2003). Germline 657del5 mutation in the NBS1 gene in breast cancer patients. *Int J Cancer* **106**(3): 379-381.
- Grant, R. C., I. Selander, A. A. Connor, S. Selvarajah, A. Borgida, et al. (2015). Prevalence of germline mutations in cancer predisposition genes in patients with pancreatic cancer. *Gastroenterology* **148**(3): 556-564.

- Gurtan, A. M., P. Stuckert and A. D. D'Andrea (2006). The WD40 repeats of FANCL are required for Fanconi anemia core complex assembly. *J Biol Chem* **281**(16): 10896-10905.
- Hall, J. M., M. K. Lee, B. Newman, J. E. Morrow, L. A. Anderson, et al. (1990). Linkage of early-onset familial breast cancer to chromosome 17q21. *Science* **250**(4988): 1684-1689.
- Hanahan, D. and R. A. Weinberg (2011). Hallmarks of cancer: the next generation. *Cell* **144**(5): 646-674.
- Harakalova, M., M. Mokry, B. Hrdlickova, I. Renkens, K. Duran, et al. (2011). Multiplexed array-based and in-solution genomic enrichment for flexible and cost-effective targeted next-generation sequencing. *Nat Protoc* **6**(12): 1870-1886.
- Hauke, J., J. Horvath, E. Gross, A. Gehrig, E. Honisch, et al. (2018). Gene panel testing of 5589 BRCA1/2-negative index patients with breast cancer in a routine diagnostic setting: results of the German Consortium for Hereditary Breast and Ovarian Cancer. *Cancer Med*.
- Heikkinen, K., K. Rapakko, S. M. Karppinen, H. Erkkö, S. Knuutila, et al. (2006). RAD50 and NBS1 are breast cancer susceptibility genes associated with genomic instability. *Carcinogenesis* **27**(8): 1593-1599.
- Hemminki, A., D. Markie, I. Tomlinson, E. Avizienyte, S. Roth, et al. (1998). A serine/threonine kinase gene defective in Peutz-Jeghers syndrome. *Nature* **391**(6663): 184-187.
- Hemminki, A., I. Tomlinson, D. Markie, H. Jarvinen, P. Sistonen, et al. (1997). Localization of a susceptibility locus for Peutz-Jeghers syndrome to 19p using comparative genomic hybridization and targeted linkage analysis. *Nat Genet* **15**(1): 87-90.
- Hiort, O., P. M. Holterhus, R. Werner, C. Marschke, U. Hoppe, et al. (2005). Homozygous disruption of P450 side-chain cleavage (CYP11A1) is associated with prematurity, complete 46,XY sex reversal, and severe adrenal failure. *J Clin Endocrinol Metab* **90**(1): 538-541.
- Hoeymakers, J. H. (2001). Genome maintenance mechanisms for preventing cancer. *Nature* **411**(6835): 366-374.
- Hojny, J., P. Zemankova, F. Lhota, J. Sevcik, V. Stranecky, et al. (2017). Multiplex PCR and NGS-based identification of mRNA splicing variants: Analysis of BRCA1 splicing pattern as a model. *Gene* **637**: 41-49.
- Hollestelle, A., M. Wasielewski, J. W. Martens and M. Schutte (2010). Discovering moderate-risk breast cancer susceptibility genes. *Curr Opin Genet Dev* **20**(3): 268-276.
- Hopper, J. L., V. M. Hayes, A. B. Spurdle, G. Chenevix-Trench, M. A. Jenkins, et al. (2005). A protein-truncating mutation in CYP17A1 in three sisters with early-onset breast cancer. *Hum Mutat* **26**(4): 298-302.
- Howlett, N. G., T. Taniguchi, S. Olson, B. Cox, Q. Waisfisz, et al. (2002). Biallelic inactivation of BRCA2 in Fanconi anemia. *Science* **297**(5581): 606-609.
- Hu, C., S. N. Hart, W. R. Bamlet, R. M. Moore, K. Nandakumar, et al. (2016). Prevalence of Pathogenic Mutations in Cancer Predisposition Genes among Pancreatic Cancer Patients. *Cancer Epidemiol Biomarkers Prev* **25**(1): 207-211.
- Huang, Y., J. W. Leung, M. Lowery, N. Matsushita, Y. Wang, et al. (2014). Modularized functions of the Fanconi anemia core complex. *Cell Rep* **7**(6): 1849-1857.
- Huen, M. S., S. M. Sy and J. Chen (2010). BRCA1 and its toolbox for the maintenance of genome integrity. *Nat Rev Mol Cell Biol* **11**(2): 138-148.
- Huertas, P., F. Cortes-Ledesma, A. A. Sartori, A. Aguilera and S. P. Jackson (2008). CDK targets Sae2 to control DNA-end resection and homologous recombination. *Nature* **455**(7213): 689-692.
- Hugh, J., J. Hanson, M. C. Cheang, T. O. Nielsen, C. M. Perou, et al. (2009). Breast cancer subtypes and response to docetaxel in node-positive breast cancer: use of an immunohistochemical definition in the BCIRG 001 trial. *J Clin Oncol* **27**(8): 1168-1176.
- Hunkapiller, T., R. J. Kaiser, B. F. Koop and L. Hood (1991). Large-scale and automated DNA sequence determination. *Science* **254**(5028): 59-67.
- Hutchison, C. A., 3rd (2007). DNA sequencing: bench to bedside and beyond. *Nucleic Acids Res* **35**(18): 6227-6237.
- Hwang, D. Y., C. C. Hung, F. G. Riepe, R. J. Auchus, A. E. Kulle, et al. (2011). CYP17A1 intron mutation causing cryptic splicing in 17alpha-hydroxylase deficiency. *PLoS One* **6**(9): e25492.
- Che, R., J. Zhang, M. Nepal, B. Han and P. Fei (2018). Multifaceted Fanconi Anemia Signaling. *Trends Genet* **34**(3): 171-183.
- Chen, L., C. J. Nievera, A. Y. Lee and X. Wu (2008). Cell cycle-dependent complex formation of BRCA1.CtIP.MRN is important for DNA double-strand break repair. *J Biol Chem* **283**(12): 7713-7720.
- Chomczynski, P. and N. Sacchi (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem* **162**(1): 156-159.
- Ide, H. and M. Kotera (2004). Human DNA glycosylases involved in the repair of oxidatively damaged DNA. *Biol Pharm Bull* **27**(4): 480-485.
- Jackson, S. P. and J. Bartek (2009). The DNA-damage response in human biology and disease. *Nature* **461**(7267): 1071-1078.
- Janatova, M., Z. Kleibl, J. Stribrna, A. Panczak, K. Vesela, et al. (2013). The PALB2 gene is a strong candidate for clinical testing in BRCA1- and BRCA2-negative hereditary breast cancer. *Cancer Epidemiol Biomarkers Prev* **22**(12): 2323-2332.

- Janatova, M., J. Soukupova, J. Stribrna, P. Kleiblova, M. Vocka, et al. (2015). Mutation Analysis of the RAD51C and RAD51D Genes in High-Risk Ovarian Cancer Patients and Families from the Czech Republic. *PLoS One* **10**(6): e0127711.
- Jazayeri, A., J. Falck, C. Lukas, J. Bartek, G. C. Smith, et al. (2006). ATM- and cell cycle-dependent regulation of ATR in response to DNA double-strand breaks. *Nat Cell Biol* **8**(1): 37-45.
- Kiiski, J. I., L. M. Pelttari, S. Khan, E. S. Freysteinsdottir, I. Reynisdottir, et al. (2014). Exome sequencing identifies FANCM as a susceptibility gene for triple-negative breast cancer. *Proc Natl Acad Sci U S A* **111**(42): 15172-15177.
- Kircher, M., U. Stenzel and J. Kelso (2009). Improved base calling for the Illumina Genome Analyzer using machine learning strategies. *Genome Biol* **10**(8): R83.
- Kircher, M., D. M. Witten, P. Jain, B. J. O'Roak, G. M. Cooper, et al. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* **46**(3): 310-315.
- Kleibl, Z., J. Fidlerova, P. Kleiblova, S. Kormunda, M. Bilek, et al. (2009). Influence of dihydropyrimidine dehydrogenase gene (DPYD) coding sequence variants on the development of fluoropyrimidine-related toxicity in patients with high-grade toxicity and patients with excellent tolerance of fluoropyrimidine-based chemotherapy. *Neoplasma* **56**(4): 303-316.
- Kleibl, Z., O. Havranek, J. Novotny, P. Kleiblova, P. Soucek, 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**(1): 159-164.
- Kleibl, Z. and V. N. Kristensen (2016). Women at high risk of breast cancer: Molecular characteristics, clinical presentation and management. *Breast* **28**: 136-144.
- Kleibl, Z., J. Novotny, D. Bezdickova, R. Malik, P. Kleiblova, et al. (2005). The CHEK2 c.1100delC germline mutation rarely contributes to breast cancer development in the Czech Republic. *Breast Cancer Res Treat* **90**(2): 165-167.
- Kleijer, W. J., V. Laugel, M. Berneburg, T. Nardo, H. Fawcett, et al. (2008). Incidence of DNA repair deficiency disorders in western Europe: Xeroderma pigmentosum, Cockayne syndrome and trichothiodystrophy. *DNA Repair (Amst)* **7**(5): 744-750.
- Korlach, J., A. Bibillo, J. Wegener, P. Peluso, T. T. Pham, et al. (2008). Long, processive enzymatic DNA synthesis using 100% dye-labeled terminal phosphate-linked nucleotides. *Nucleosides Nucleotides Nucleic Acids* **27**(9): 1072-1083.
- Kottemann, M. C. and A. Smogorzewska (2013). Fanconi anaemia and the repair of Watson and Crick DNA crosslinks. *Nature* **493**(7432): 356-363.
- Kuhner, M. K., R. Kostadinov and B. J. Reid (2016). Limitations of the Driver/Passenger Model in Cancer Prevention. *Cancer Prev Res (Phila)* **9**(5): 335-338.
- Kumar, P., S. Henikoff and P. C. Ng (2009). Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* **4**(7): 1073-1081.
- Laloo, F. and D. G. Evans (2012). Familial breast cancer. *Clin Genet* **82**(2): 105-114.
- Larrea, A. A., S. A. Lujan and T. A. Kunkel (2010). SnapShot: DNA mismatch repair. *Cell* **141**(4): 730 e731.
- Latif, A., K. D. Hadfield, S. A. Roberts, A. Shenton, F. Laloo, et al. (2010). Breast cancer susceptibility variants alter risks in familial disease. *J Med Genet* **47**(2): 126-131.
- Leguen, Y., J. M. Brechot, E. Varon and J. Rochemaure (1998). [Unusual cause of necrotized subacute pneumopathy]. *Rev Pneumol Clin* **54**(3): 161-163.
- Lehmann, J., S. Schubert and S. Emmert (2014). Xeroderma pigmentosum: diagnostic procedures, interdisciplinary patient care, and novel therapeutic approaches. *J Dtsch Dermatol Ges* **12**(10): 867-872.
- Lener, M. R., R. J. Scott, W. Kluzniak, P. Baszuk, C. Cybulski, et al. (2016). Do founder mutations characteristic of some cancer sites also predispose to pancreatic cancer? *Int J Cancer* **139**(3): 601-606.
- Levene, M. J., J. Korlach, S. W. Turner, M. Foquet, H. G. Craighead, et al. (2003). Zero-mode waveguides for single-molecule analysis at high concentrations. *Science* **299**(5607): 682-686.
- Levitus, M., Q. Waisfisz, B. C. Godthelp, Y. de Vries, S. Hussain, et al. (2005). The DNA helicase BRIP1 is defective in Fanconi anemia complementation group J. *Nat Genet* **37**(9): 934-935.
- Levrán, O., C. Attwooll, R. T. Henry, K. L. Milton, K. Neveling, et al. (2005). The BRCA1-interacting helicase BRIP1 is deficient in Fanconi anemia. *Nat Genet* **37**(9): 931-933.
- Levy-Lahad, E. (2010). Fanconi anemia and breast cancer susceptibility meet again. *Nat Genet* **42**(5): 368-369.
- Lhota, F., P. Zemankova, P. Kleiblova, J. Soukupova, M. Vocka, et al. (2016). Hereditary truncating mutations of DNA repair and other genes in BRCA1/BRCA2/PALB2-negatively tested breast cancer patients. *Clin Genet* **90**(4): 324-333.
- Li, N., S. M. Rowley, E. R. Thompson, S. McInerney, L. Devereux, et al. (2018). Evaluating the breast cancer predisposition role of rare variants in genes associated with low-penetrance breast cancer risk SNPs. *Breast Cancer Res* **20**(1): 3.
- Lieber, M. R. and T. E. Wilson (2010). SnapShot: Nonhomologous DNA end joining (NHEJ). *Cell* **142**(3): 496-496 e491.

- Lichtenstein, P., N. V. Holm, P. K. Verkasalo, A. Iliadou, J. Kaprio, et al. (2000). Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland. *N Engl J Med* **343**(2): 78-85.
- Lim, E. T., P. Wurtz, A. S. Havulinna, P. Palta, T. Tukiainen, et al. (2014). Distribution and medical impact of loss-of-function variants in the Finnish founder population. *PLoS Genet* **10**(7): e1004494.
- Litman, R., M. Peng, Z. Jin, F. Zhang, J. Zhang, et al. (2005). BACH1 is critical for homologous recombination and appears to be the Fanconi anemia gene product FANCI. *Cancer Cell* **8**(3): 255-265.
- Liu, H., J. Rudolf, K. A. Johnson, S. A. McMahon, M. Oke, et al. (2008). Structure of the DNA repair helicase XPD. *Cell* **133**(5): 801-812.
- Loveday, C., C. Turnbull, E. Ruark, R. M. Xicola, E. Ramsay, et al. (2012). Germline RAD51C mutations confer susceptibility to ovarian cancer. *Nat Genet* **44**(5): 475-476; author reply 476.
- Lowe, S. W., S. Bodis, A. McClatchey, L. Remington, H. E. Ruley, et al. (1994). p53 status and the efficacy of cancer therapy in vivo. *Science* **266**(5186): 807-810.
- Machackova, E., L. Foretova, M. Lukesova, P. Vasickova, M. Navratilova, et al. (2008). Spectrum and characterisation of BRCA1 and BRCA2 deleterious mutations in high-risk Czech patients with breast and/or ovarian cancer. *BMC Cancer* **8**: 140.
- Malkin, D., F. P. Li, L. C. Strong, J. F. Fraumeni, Jr., C. E. Nelson, et al. (1990). Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms. *Science* **250**(4985): 1233-1238.
- Margulies, M., M. Egholm, W. E. Altman, S. Attiya, J. S. Bader, et al. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**(7057): 376-380.
- Masciari, S., N. Larsson, J. Senz, N. Boyd, P. Kaurah, et al. (2007). Germline E-cadherin mutations in familial lobular breast cancer. *J Med Genet* **44**(11): 726-731.
- Masson, J. Y., M. C. Tarsounas, A. Z. Stasiak, A. Stasiak, R. Shah, et al. (2001). Identification and purification of two distinct complexes containing the five RAD51 paralogs. *Genes Dev* **15**(24): 3296-3307.
- Mateju, M., J. Stribrna, M. Zikan, Z. Kleibl, M. Janatova, et al. (2010). Population-based study of BRCA1/2 mutations: family history based criteria identify minority of mutation carriers. *Neoplasma* **57**(3): 280-285.
- Mavaddat, N., A. C. Antoniou, D. F. Easton and M. Garcia-Closas (2010). Genetic susceptibility to breast cancer. *Mol Oncol* **4**(3): 174-191.
- Mavaddat, N., D. Barrowdale, I. L. Andrulis, S. M. Domchek, D. Eccles, et al. (2012). Pathology of breast and ovarian cancers among BRCA1 and BRCA2 mutation carriers: results from the Consortium of Investigators of Modifiers of BRCA1/2 (CIMBA). *Cancer Epidemiol Biomarkers Prev* **21**(1): 134-147.
- McClellan, J. and M. C. King (2010). Genetic heterogeneity in human disease. *Cell* **141**(2): 210-217.
- McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, et al. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**(9): 1297-1303.
- Meetei, A. R., J. P. de Winter, A. L. Medhurst, M. Wallisch, Q. Waisfisz, et al. (2003). A novel ubiquitin ligase is deficient in Fanconi anemia. *Nat Genet* **35**(2): 165-170.
- Meetei, A. R., M. Levitus, Y. Xue, A. L. Medhurst, M. Zwaan, et al. (2004). X-linked inheritance of Fanconi anemia complementation group B. *Nat Genet* **36**(11): 1219-1224.
- Meijers-Heijboer, H., A. van den Ouweland, J. Klijn, M. Wasielewski, A. de Snoo, et al. (2002). Low-penetrance susceptibility to breast cancer due to CHEK2(*)1100delC in noncarriers of BRCA1 or BRCA2 mutations. *Nat Genet* **31**(1): 55-59.
- Meindl, A., H. Hellebrand, C. Wiek, V. Erven, B. Wappenschmidt, et al. (2010). Germline mutations in breast and ovarian cancer pedigrees establish RAD51C as a human cancer susceptibility gene. *Nat Genet* **42**(5): 410-414.
- Meisma, R., P. Fernandez-Salguero, A. B. Van Kuilenburg, A. H. Van Gennip and F. J. Gonzalez (1995). Human polymorphism in drug metabolism: mutation in the dihydropyrimidine dehydrogenase gene results in exon skipping and thymine uraciluria. *DNA Cell Biol* **14**(1): 1-6.
- Menck, C. F. and V. Munford (2014). DNA repair diseases: What do they tell us about cancer and aging? *Genet Mol Biol* **37**(1 Suppl): 220-233.
- McHugh, P. J., V. J. Spanswick and J. A. Hartley (2001). Repair of DNA interstrand crosslinks: molecular mechanisms and clinical relevance. *Lancet Oncol* **2**(8): 483-490.
- Michailidou, K., P. Hall, A. Gonzalez-Neira, M. Ghoussaini, J. Dennis, et al. (2013). Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat Genet* **45**(4): 353-361, 361e351-352.
- Miki, Y., J. Swensen, D. Shattuck-Eidens, P. A. Futreal, K. Harshman, et al. (1994). A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science* **266**(5182): 66-71.
- Mimitou, E. P. and L. S. Symington (2009). Nucleases and helicases take center stage in homologous recombination. *Trends Biochem Sci* **34**(5): 264-272.
- Mirzoeva, O. K. and J. H. Petrini (2001). DNA damage-dependent nuclear dynamics of the Mre11 complex. *Mol Cell Biol* **21**(1): 281-288.
- Moynahan, M. E. and M. Jasin (2010). Mitotic homologous recombination maintains genomic stability and suppresses tumorigenesis. *Nat Rev Mol Cell Biol* **11**(3): 196-207.

- Narod, S. A. (2010). Testing for CHEK2 in the cancer genetics clinic: ready for prime time? *Clin Genet* **78**(1): 1-7.
- Narod, S. A. and W. D. Foulkes (2004). BRCA1 and BRCA2: 1994 and beyond. *Nat Rev Cancer* **4**(9): 665-676.
- Nelen, M. R., G. W. Padberg, E. A. Peeters, A. Y. Lin, B. van den Helm, et al. (1996). Localization of the gene for Cowden disease to chromosome 10q22-23. *Nat Genet* **13**(1): 114-116.
- Nevanlinna, H. and J. Bartek (2006). The CHEK2 gene and inherited breast cancer susceptibility. *Oncogene* **25**(43): 5912-5919.
- Ngeow, J., K. Sesock and C. Eng (2017). Breast cancer risk and clinical implications for germline PTEN mutation carriers. *Breast Cancer Res Treat* **165**(1): 1-8.
- Nishiwaki, Y., N. Kobayashi, K. Imoto, T. A. Iwamoto, A. Yamamoto, et al. (2004). Trichothiodystrophy fibroblasts are deficient in the repair of ultraviolet-induced cyclobutane pyrimidine dimers and (6-4)photoproducts. *J Invest Dermatol* **122**(2): 526-532.
- Nojima, K., H. Hochegger, A. Saberi, T. Fukushima, K. Kikuchi, et al. (2005). Multiple repair pathways mediate tolerance to chemotherapeutic cross-linking agents in vertebrate cells. *Cancer Res* **65**(24): 11704-11711.
- Novembre, J., T. Johnson, K. Bryc, Z. Kutalik, A. R. Boyko, et al. (2008). Genes mirror geography within Europe. *Nature* **456**(7218): 98-101.
- Orban, T. I. and E. Olah (2003). Emerging roles of BRCA1 alternative splicing. *Mol Pathol* **56**(4): 191-197.
- Pabalan, N., O. Francisco-Pabalan, L. Sung, H. Jarjanazi and H. Ozcelik (2010). Meta-analysis of two ERCC2 (XPD) polymorphisms, Asp312Asn and Lys751Gln, in breast cancer. *Breast Cancer Res Treat* **124**(2): 531-541.
- Pacifico, L., C. Carducci, E. Poggiogalle, F. Caravona, I. Antonozzi, et al. (2010). Mutational analysis of ABCC2 gene in two siblings with neonatal-onset Dubin Johnson syndrome. *Clin Genet* **78**(6): 598-600.
- Pal, T., M. R. Akbari, P. Sun, J. H. Lee, J. Fulp, et al. (2012). Frequency of mutations in mismatch repair genes in a population-based study of women with ovarian cancer. *Br J Cancer* **107**(10): 1783-1790.
- Pang, Z., L. Yao, J. Zhang, T. Ouyang, J. Li, et al. (2011). RAD51C germline mutations in Chinese women with familial breast cancer. *Breast Cancer Res Treat* **129**(3): 1019-1020.
- Panier, S. and D. Durocher (2013). Push back to respond better: regulatory inhibition of the DNA double-strand break response. *Nat Rev Mol Cell Biol* **14**(10): 661-672.
- Parkin, D. M. (2004). International variation. *Oncogene* **23**(38): 6329-6340.
- Pellegrini, L., D. S. Yu, T. Lo, S. Anand, M. Lee, et al. (2002). Insights into DNA recombination from the structure of a RAD51-BRCA2 complex. *Nature* **420**(6913): 287-293.
- Pelttari, L. M., T. Heikkinen, D. Thompson, A. Kallioniemi, J. Schleutker, et al. (2011). RAD51C is a susceptibility gene for ovarian cancer. *Hum Mol Genet* **20**(16): 3278-3288.
- Pena-Diaz, J. and L. J. Rasmussen (2016). Approaches to diagnose DNA mismatch repair gene defects in cancer. *DNA Repair (Amst)* **38**: 147-154.
- Peng, M., R. Litman, Z. Jin, G. Fong and S. B. Cantor (2006). BACH1 is a DNA repair protein supporting BRCA1 damage response. *Oncogene* **25**(15): 2245-2253.
- Penney, K. L., C. A. Salinas, M. Pomerantz, F. R. Schumacher, C. A. Beckwith, et al. (2009). Evaluation of 8q24 and 17q risk loci and prostate cancer mortality. *Clin Cancer Res* **15**(9): 3223-3230.
- Peterlongo, P., I. Catucci, M. Colombo, L. Caleca, E. Mucaki, et al. (2015). FANCM c.5791C>T nonsense mutation (rs144567652) induces exon skipping, affects DNA repair activity and is a familial breast cancer risk factor. *Hum Mol Genet* **24**(18): 5345-5355.
- Peto, J., N. Collins, R. Barfoot, S. Seal, W. Warren, et al. (1999). Prevalence of BRCA1 and BRCA2 gene mutations in patients with early-onset breast cancer. *J Natl Cancer Inst* **91**(11): 943-949.
- Pfeifer, K., P. Schurmann, N. Bogdanova, K. Neuhauser, I. M. Kostovska, et al. (2016). Frameshift variant FANCL*c.1096_1099dupATTA is not associated with high breast cancer risk. *Clin Genet* **90**(4): 385-386.
- Pharoah, P. D., A. C. Antoniou, D. F. Easton and B. A. Ponder (2008). Polygenes, risk prediction, and targeted prevention of breast cancer. *N Engl J Med* **358**(26): 2796-2803.
- Pharoah, P. D., P. Guilford, C. Caldas and C. International Gastric Cancer Linkage (2001). Incidence of gastric cancer and breast cancer in CDH1 (E-cadherin) mutation carriers from hereditary diffuse gastric cancer families. *Gastroenterology* **121**(6): 1348-1353.
- Pohlreich, P., Z. Kleibl, P. Kleiblova, M. Janatova, J. Soukupova, et al. (2012). [The clinical importance of a genetic analysis of moderate-risk cancer susceptibility genes in breast and other cancer patients from the Czech Republic]. *Klin Onkol* **25** Suppl: S59-66.
- Pohlreich, P., M. Zikan, J. Stribrna, Z. Kleibl, M. Janatova, et al. (2005). High proportion of recurrent germline mutations in the BRCA1 gene in breast and ovarian cancer patients from the Prague area. *Breast Cancer Res* **7**(5): R728-736.
- Pollard, K. S., M. J. Hubisz, K. R. Rosenbloom and A. Siepel (2010). Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res* **20**(1): 110-121.
- Polo, S. E. and S. P. Jackson (2011). Dynamics of DNA damage response proteins at DNA breaks: a focus on protein modifications. *Genes Dev* **25**(5): 409-433.
- Polyak, K. (2007). Breast cancer: origins and evolution. *J Clin Invest* **117**(11): 3155-3163.
- Prusová, J. (2010). Karcinom prsu – problém i v 21. století. *Interní Med.* **1**(12): 26-32.
- Rafnar, T., D. F. Gudbjartsson, P. Sulem, A. Jonasdottir, A. Sigurdsson, et al. (2011). Mutations in BRIP1 confer high risk of ovarian cancer. *Nat Genet* **43**(11): 1104-1107.

- Rahman, N. (2014). Realizing the promise of cancer predisposition genes. *Nature* **505**(7483): 302-308.
- Rahman, N., S. Seal, D. Thompson, P. Kelly, A. Renwick, et al. (2007). PALB2, which encodes a BRCA2-interacting protein, is a breast cancer susceptibility gene. *Nat Genet* **39**(2): 165-167.
- Rasmussen, L. J., M. Rasmussen, B. Lee, A. K. Rasmussen, D. M. Wilson, 3rd, et al. (2000). Identification of factors interacting with hMSH2 in the fetal liver utilizing the yeast two-hybrid system. In vivo interaction through the C-terminal domains of hEXO1 and hMSH2 and comparative expression analysis. *Mutat Res* **460**(1): 41-52.
- Renwick, A., D. Thompson, S. Seal, P. Kelly, T. Chagtai, et al. (2006). ATM mutations that cause ataxia-telangiectasia are breast cancer susceptibility alleles. *Nat Genet* **38**(8): 873-875.
- Roberts, N. J., A. L. Norris, G. M. Petersen, M. L. Bondy, R. Brand, et al. (2016). Whole Genome Sequencing Defines the Genetic Heterogeneity of Familial Pancreatic Cancer. *Cancer Discov* **6**(2): 166-175.
- Rodríguez-Ezpeleta (2012). Overview of Sequencing Technology Platforms. *Bioinformatics for High Throughput Sequencing*.
- Romero, A., F. Garcia-Garcia, I. Lopez-Perolio, G. Ruiz de Garibay, J. A. Garcia-Saenz, et al. (2015). BRCA1 Alternative splicing landscape in breast tissue samples. *BMC Cancer* **15**: 219.
- Romero, A., P. Perez-Segura, A. Tosar, J. A. Garcia-Saenz, E. Diaz-Rubio, et al. (2011). A HRM-based screening method detects RAD51C germ-line deleterious mutations in Spanish breast and ovarian cancer families. *Breast Cancer Res Treat* **129**(3): 939-946.
- Rosen, E. M. (2013). BRCA1 in the DNA damage response and at telomeres. *Front Genet* **4**: 85.
- Rosenthal, E. T., B. Evans, J. Kidd, K. Brown, H. Goringe, et al. (2017). Increased Identification of Candidates for High-Risk Breast Cancer Screening Through Expanded Genetic Testing. *J Am Coll Radiol* **14**(4): 561-568.
- Rothberg, J. M., W. Hinz, T. M. Rearick, J. Schultz, W. Mileski, et al. (2011). An integrated semiconductor device enabling non-optical genome sequencing. *Nature* **475**(7356): 348-352.
- Rouleau, M., A. Patel, M. J. Hendzel, S. H. Kaufmann and G. G. Poirier (2010). PARP inhibition: PARP1 and beyond. *Nat Rev Cancer* **10**(4): 293-301.
- Rouzier, R., C. M. Perou, W. F. Symmans, N. Ibrahim, M. Cristofanilli, et al. (2005). Breast cancer molecular subtypes respond differently to preoperative chemotherapy. *Clin Cancer Res* **11**(16): 5678-5685.
- Roy, R., J. Chun and S. N. Powell (2012). BRCA1 and BRCA2: different roles in a common pathway of genome protection. *Nat Rev Cancer* **12**(1): 68-78.
- Rump, A., A. Benet-Pages, S. Schubert, J. D. Kuhlmann, R. Janavicius, et al. (2016). Identification and Functional Testing of ERCC2 Mutations in a Multi-national Cohort of Patients with Familial Breast- and Ovarian Cancer. *PLoS Genet* **12**(8): e1006248.
- Rupnik, A., N. F. Lowndes and M. Grenon (2010). MRN and the race to the break. *Chromosoma* **119**(2): 115-135.
- San Filippo, J., P. Sung and H. Klein (2008). Mechanism of eukaryotic homologous recombination. *Annu Rev Biochem* **77**: 229-257.
- Sanger, F. (1988). Sequences, sequences, and sequences. *Annu Rev Biochem* **57**: 1-28.
- Savitsky, K., A. Bar-Shira, S. Gilad, G. Rotman, Y. Ziv, et al. (1995). A single ataxia telangiectasia gene with a product similar to PI-3 kinase. *Science* **268**(5218): 1749-1753.
- Seal, S., D. Thompson, A. Renwick, A. Elliott, P. Kelly, et al. (2006). Truncating mutations in the Fanconi anemia J gene BRIP1 are low-penetrance breast cancer susceptibility alleles. *Nat Genet* **38**(11): 1239-1241.
- Seemanova, E. (1990). An increased risk for malignant neoplasms in heterozygotes for a syndrome of microcephaly, normal intelligence, growth retardation, remarkable facies, immunodeficiency and chromosomal instability. *Mutat Res* **238**(3): 321-324.
- Seemanova, E., P. Jarolim, P. Seeman, R. Varon, M. Digweed, et al. (2007). Cancer risk of heterozygotes with the NBN founder mutation. *J Natl Cancer Inst* **99**(24): 1875-1880.
- Shafirovich, V. and N. E. Geacintov (2017). Removal of oxidatively generated DNA damage by overlapping repair pathways. *Free Radic Biol Med* **107**: 53-61.
- Sharan, S. K., M. Morimatsu, U. Albrecht, D. S. Lim, E. Regel, et al. (1997). Embryonic lethality and radiation hypersensitivity mediated by Rad51 in mice lacking Brca2. *Nature* **386**(6627): 804-810.
- Shen, W. H., A. S. Balajee, J. Wang, H. Wu, C. Eng, et al. (2007). Essential role for nuclear PTEN in maintaining chromosomal integrity. *Cell* **128**(1): 157-170.
- Shendure, J., G. J. Porreca, N. B. Reppas, X. Lin, J. P. McCutcheon, et al. (2005). Accurate multiplex polony sequencing of an evolved bacterial genome. *Science* **309**(5741): 1728-1732.
- Schroeder, C., U. Faust, M. Sturm, K. Hackmann, K. Grundmann, et al. (2015). HBOC multi-gene panel testing: comparison of two sequencing centers. *Breast Cancer Res Treat* **152**(1): 129-136.
- Schwarz, J. M., C. Rodelsperger, M. Schuelke and D. Seelow (2010). MutationTaster evaluates disease-causing potential of sequence alterations. *Nat Methods* **7**(8): 575-576.
- Silvestri, V., P. Rizzolo, M. Falchetti, I. Zanna, G. Masala, et al. (2011). Mutation screening of RAD51C in male breast cancer patients. *Breast Cancer Res* **13**(1): 404.
- Singh, A., E. Compe, N. Le May and J. M. Egly (2015). TFIIH subunit alterations causing xeroderma pigmentosum and trichothiodystrophy specifically disturb several steps during transcription. *Am J Hum Genet* **96**(2): 194-207.

- Smid, M., Y. Wang, Y. Zhang, A. M. Sieuwerts, J. Yu, et al. (2008). Subtypes of breast cancer show preferential site of relapse. *Cancer Res* **68**(9): 3108-3114.
- Smith, A. M., L. E. Heisler, R. P. St Onge, E. Farias-Hesson, I. M. Wallace, et al. (2010). Highly-multiplexed barcode sequencing: an efficient method for parallel analysis of pooled samples. *Nucleic Acids Res* **38**(13): e142.
- Smith, C. G., M. Naven, R. Harris, J. Colley, H. West, et al. (2013). Exome resequencing identifies potential tumor-suppressor genes that predispose to colorectal cancer. *Hum Mutat* **34**(7): 1026-1034.
- Smith, D. R., A. R. Quinlan, H. E. Peckham, K. Makowsky, W. Tao, et al. (2008). Rapid whole-genome mutational profiling using next-generation sequencing technologies. *Genome Res* **18**(10): 1638-1642.
- Sorensen, S. A., J. J. Mulvihill and A. Nielsen (1986). Long-term follow-up of von Recklinghausen neurofibromatosis. Survival and malignant neoplasms. *N Engl J Med* **314**(16): 1010-1015.
- Sorlie, T., C. M. Perou, R. Tibshirani, T. Aas, S. Geisler, et al. (2001). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* **98**(19): 10869-10874.
- Soukupova, J., P. Zemankova, P. Kleiblova, M. Janatova and Z. Kleibl (2016). [CZECANCA: CZEch CANcer paNel for Clinical Application-- Design and Optimization of the Targeted Sequencing Panel for the Identification of Cancer Susceptibility in High-risk Individuals from the Czech Republic]. *Klin Onkol* **29 Suppl 1**: S46-54.
- Soukupova, J., P. Zemankova, K. Lhotova, M. Janatova, M. Borecka, et al. (2018). Validation of CZECANCA (CZEch CANcer paNel for Clinical Application) for targeted NGS-based analysis of hereditary cancer syndromes. *PLoS One* **13**(4): e0195761.
- Southey, M. C., Z. L. Teo, J. G. Dowty, F. A. Odefrey, D. J. Park, et al. (2010). A PALB2 mutation associated with high risk of breast cancer. *Breast Cancer Res* **12**(6): R109.
- Steffen, J., D. Nowakowska, A. Niwinska, D. Czapczak, A. Kluska, et al. (2006). Germline mutations 657del5 of the NBS1 gene contribute significantly to the incidence of breast cancer in Central Poland. *Int J Cancer* **119**(2): 472-475.
- Sticova, E. and M. Jirsa (2013). New insights in bilirubin metabolism and their clinical implications. *World J Gastroenterol* **19**(38): 6398-6407.
- Stratton, M. R., P. J. Campbell and P. A. Futreal (2009). The cancer genome. *Nature* **458**(7239): 719-724.
- Stratton, M. R. and N. Rahman (2008). The emerging landscape of breast cancer susceptibility. *Nat Genet* **40**(1): 17-22.
- Stucki, M., J. A. Clapperton, D. Mohammad, M. B. Yaffe, S. J. Smerdon, et al. (2005). MDC1 directly binds phosphorylated histone H2AX to regulate cellular responses to DNA double-strand breaks. *Cell* **123**(7): 1213-1226.
- Sulli, G., R. Di Micco and F. d'Adda di Fagagna (2012). Crosstalk between chromatin state and DNA damage response in cellular senescence and cancer. *Nat Rev Cancer* **12**(10): 709-720.
- Summers, K. C., F. Shen, E. A. Sierra Potchanant, E. A. Phipps, R. J. Hickey, et al. (2011). Phosphorylation: the molecular switch of double-strand break repair. *Int J Proteomics* **2011**: 373816.
- Syrjakoski, K., T. Kuukasjarvi, K. Waltering, K. Haraldsson, A. Auvinen, et al. (2004). BRCA2 mutations in 154 finnish male breast cancer patients. *Neoplasia* **6**(5): 541-545.
- Takayama, K., E. P. Salazar, A. Lehmann, M. Stefanini, L. H. Thompson, et al. (1995). Defects in the DNA repair and transcription gene ERCC2 in the cancer-prone disorder xeroderma pigmentosum group D. *Cancer Res* **55**(23): 5656-5663.
- Thompson, D., S. Duedal, J. Kirner, L. McGuffog, J. Last, et al. (2005). Cancer risks and mortality in heterozygous ATM mutation carriers. *J Natl Cancer Inst* **97**(11): 813-822.
- Thompson, D. and D. Easton (2004). The genetic epidemiology of breast cancer genes. *J Mammary Gland Biol Neoplasia* **9**(3): 221-236.
- Thompson, E. R., S. E. Boyle, J. Johnson, G. L. Ryland, S. Sawyer, et al. (2012). Analysis of RAD51C germline mutations in high-risk breast and ovarian cancer families and ovarian cancer patients. *Hum Mutat* **33**(1): 95-99.
- Ticha, I., Z. Kleibl, J. Stribrna, J. Kotlas, M. Zimovjanova, et al. (2010). Screening for genomic rearrangements in BRCA1 and BRCA2 genes in Czech high-risk breast/ovarian cancer patients: high proportion of population specific alterations in BRCA1 gene. *Breast Cancer Res Treat* **124**(2): 337-347.
- Tommiska, J., S. Seal, A. Renwick, R. Barfoot, L. Baskcomb, et al. (2006). Evaluation of RAD50 in familial breast cancer predisposition. *Int J Cancer* **118**(11): 2911-2916.
- Trainer, A. H., E. Thompson and P. A. James (2011). BRCA and beyond: a genome-first approach to familial breast cancer risk assessment. *Discov Med* **12**(66): 433-443.
- Turnbull, C., S. Ahmed, J. Morrison, D. Pernet, A. Renwick, et al. (2010). Genome-wide association study identifies five new breast cancer susceptibility loci. *Nat Genet* **42**(6): 504-507.
- Turnbull, C. and N. Rahman (2008). Genetic predisposition to breast cancer: past, present, and future. *Annu Rev Genomics Hum Genet* **9**: 321-345.
- Turner, N., A. Tutt and A. Ashworth (2005). Targeting the DNA repair defect of BRCA tumours. *Curr Opin Pharmacol* **5**(4): 388-393.
- UICC Internationa Union Against Cancer. URI: <http://www.uicc.org/resources/tnm>.

- Valouev, A., J. Ichikawa, T. Tonthat, J. Stuart, S. Ranade, et al. (2008). A high-resolution, nucleosome position map of *C. elegans* reveals a lack of universal sequence-dictated positioning. *Genome Res* **18**(7): 1051-1063.
- van der Groep, P., E. van der Wall and P. J. van Diest (2011). Pathology of hereditary breast cancer. *Cell Oncol (Dordr)* **34**(2): 71-88.
- van Kuilenburg, A. B., P. Hausler, A. Schalhorn, M. W. Tanck, J. H. Proost, et al. (2012). Evaluation of 5-fluorouracil pharmacokinetics in cancer patients with a c.1905+1G>A mutation in DPYD by means of a Bayesian limited sampling strategy. *Clin Pharmacokinet* **51**(3): 163-174.
- Varon, R., E. Seemanova, K. Chrzanowska, O. Hnateyko, D. Piekutowska-Abramczuk, et al. (2000). Clinical ascertainment of Nijmegen breakage syndrome (NBS) and prevalence of the major mutation, 657del5, in three Slav populations. *Eur J Hum Genet* **8**(11): 900-902.
- Varon, R., C. Vissinga, M. Platzer, K. M. Cerosaletti, K. H. Chrzanowska, et al. (1998). Nibrin, a novel DNA double-strand break repair protein, is mutated in Nijmegen breakage syndrome. *Cell* **93**(3): 467-476.
- Vaz, F., H. Hanenberg, B. Schuster, K. Barker, C. Wiek, et al. (2010). Mutation of the RAD51C gene in a Fanconi anemia-like disorder. *Nat Genet* **42**(5): 406-409.
- Venter, J. C., M. D. Adams, E. W. Myers, P. W. Li, R. J. Mural, et al. (2001). The sequence of the human genome. *Science* **291**(5507): 1304-1351.
- Vetro, A., M. Iascone, I. Limongelli, N. Ameziane, S. Gana, et al. (2015). Loss-of-Function FANCL Mutations Associate with Severe Fanconi Anemia Overlapping the VACTERL Association. *Hum Mutat* **36**(5): 562-568.
- Viprakasit, V., R. J. Gibbons, B. C. Broughton, J. L. Tolmie, D. Brown, et al. (2001). Mutations in the general transcription factor TFIIH result in beta-thalassaemia in individuals with trichothiodystrophy. *Hum Mol Genet* **10**(24): 2797-2802.
- Vuorela, M., K. Pylkas, J. M. Hartikainen, K. Sundfeldt, A. Lindblom, et al. (2011). Further evidence for the contribution of the RAD51C gene in hereditary breast and ovarian cancer susceptibility. *Breast Cancer Res Treat* **130**(3): 1003-1010.
- Walden, H. and A. J. Deans (2014). The Fanconi anemia DNA repair pathway: structural and functional insights into a complex disorder. *Annu Rev Biophys* **43**: 257-278.
- Walsh, M. F., K. L. Nathanson, F. J. Couch and K. Offit (2016). Genomic Biomarkers for Breast Cancer Risk. *Adv Exp Med Biol* **882**: 1-32.
- Walsh, T. and M. C. King (2007). Ten genes for inherited breast cancer. *Cancer Cell* **11**(2): 103-105.
- Wang, B. (2012). BRCA1 tumor suppressor network: focusing on its tail. *Cell Biosci* **2**(1): 6.
- Wang, K., M. Li and H. Hakonarson (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* **38**(16): e164.
- Wang, W. (2007). Emergence of a DNA-damage response network consisting of Fanconi anaemia and BRCA proteins. *Nat Rev Genet* **8**(10): 735-748.
- Wang, X., J. K. Teer, R. N. Tounsi, A. M. Levin, D. Boulware, et al. (2018). Breast cancer risk and germline genomic profiling of women with neurofibromatosis type 1 who developed breast cancer. *Genes Chromosomes Cancer* **57**(1): 19-27.
- Wang, Z., M. Gerstein and M. Snyder (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* **10**(1): 57-63.
- Welch, P. L., K. N. Owens and M. C. King (2000). Insights into the functions of BRCA1 and BRCA2. *Trends Genet* **16**(2): 69-74.
- West, S. C. (2003). Molecular views of recombination proteins and their control. *Nat Rev Mol Cell Biol* **4**(6): 435-445.
- Wiechmann, L., M. Sampson, M. Stempel, L. M. Jaks, S. M. Patil, et al. (2009). Presenting features of breast cancer differ by molecular subtype. *Ann Surg Oncol* **16**(10): 2705-2710.
- Wold, B. and R. M. Myers (2008). Sequence census methods for functional genomics. *Nat Methods* **5**(1): 19-21.
- Wong, M. W., C. Nordfors, D. Mossman, G. Pecanpetelovska, K. A. Avery-Kiejda, et al. (2011). BRIP1, PALB2, and RAD51C mutation analysis reveals their relative importance as genetic susceptibility factors for breast cancer. *Breast Cancer Res Treat* **127**(3): 853-859.
- Wooster, R., G. Bignell, J. Lancaster, S. Swift, S. Seal, et al. (1995). Identification of the breast cancer susceptibility gene BRCA2. *Nature* **378**(6559): 789-792.
- Wooster, R., S. L. Neuhausen, J. Mangion, Y. Quirk, D. Ford, et al. (1994). Localization of a breast cancer susceptibility gene, BRCA2, to chromosome 13q12-13. *Science* **265**(5181): 2088-2090.
- Wu, Y., M. J. Berends, J. G. Post, R. G. Mensink, E. Verlind, et al. (2001). Germline mutations of EXO1 gene in patients with hereditary nonpolyposis colorectal cancer (HNPCC) and atypical HNPCC forms. *Gastroenterology* **120**(7): 1580-1587.
- Yang, H., P. D. Jeffrey, J. Miller, E. Kinnucan, Y. Sun, et al. (2002). BRCA2 function in DNA binding and recombination from a BRCA2-DSS1-ssDNA structure. *Science* **297**(5588): 1837-1848.
- Yang, H., Q. Li, J. Fan, W. K. Holloman and N. P. Pavletich (2005). The BRCA2 homologue Brh2 nucleates RAD51 filament formation at a dsDNA-ssDNA junction. *Nature* **433**(7026): 653-657.

- Yang, X., J. Wu, J. Lu, G. Liu, G. Di, et al. (2015). Identification of a comprehensive spectrum of genetic factors for hereditary breast cancer in a Chinese population by next-generation sequencing. PLoS One **10**(4): e0125571.
- You, Z. and J. M. Bailis (2010). DNA damage and decisions: CtIP coordinates DNA repair and cell cycle checkpoints. Trends Cell Biol **20**(7): 402-409.
- Yu, W., M. Clyne, M. J. Khoury and M. Gwinn (2010). Phenopedia and Genopedia: disease-centered and gene-centered views of the evolving knowledge of human genetic associations. Bioinformatics **26**(1): 145-146.
- Zemankova, P., F. Lhota, P. Kleiblova, J. Soukupova, M. Vocka, et al. (2016). RE: frameshift variant FANCL*c.1096_1099dupATTA is not associated with high breast cancer risk. Clin Genet **90**(4): 387-389.
- Zheng, Y., J. Zhang, K. Hope, Q. Niu, D. Huo, et al. (2010). Screening RAD51C nucleotide alterations in patients with a family history of breast and ovarian cancer. Breast Cancer Res Treat **124**(3): 857-861.
- Zhou, X., S. G. Khan, D. Tamura, N. J. Patronas, W. M. Zein, et al. (2010). Brittle hair, developmental delay, neurologic abnormalities, and photosensitivity in a 4-year-old girl. J Am Acad Dermatol **63**(2): 323-328.
- Zoller, M. E., B. Rembeck, A. Oden, M. Samuelsson and L. Angervall (1997). Malignant and benign tumors in patients with neurofibromatosis type 1 in a defined Swedish population. Cancer **79**(11): 2125-2131.

13 PŘÍLOHY

PŘÍLOHA I:

Seznam cílových genů

Gen	Gen (plné jméno)	Referenční transkript
RBFOX1	RNA binding protein, fox-1 homolog 1	NM_018723.3
ABCB1	ATP-binding cassette, sub-family B (MDR/TAP), member 1	NM_000927.4
ABCC2	ATP-binding cassette, sub-family C (CFTR/MRP), member 2	NM_000392.4
ABCC4	ATP-binding cassette, sub-family C (CFTR/MRP), member 4	NM_005845.4
ABCG2	ATP-binding cassette, sub-family G (WHITE), member 2	NM_004827.2
ABL1	ABL proto-oncogene 1, non-receptor tyrosine kinase	NM_005157.4
ACACA	acetyl-CoA carboxylase alpha	NM_198834.2
ACE	angiotensin I converting enzyme	NM_000789.3
ADAMTS18	ADAM metalloproteinase with thrombospondin	NM_199355.2
ADH1B	alcohol dehydrogenase 1B (class I), beta polypeptide	NM_000668.5
ADH1C	alcohol dehydrogenase 1C (class I), gamma polypeptide	NM_000669.4
ADIPOQ	adiponectin, C1Q and collagen domain containing	NM_001177800.1
ADIPOR1	adiponectin receptor 1	NM_015999.5
ADRB2	adrenoceptor beta 2, surface	NM_000024.5
ADRB3	adrenoceptor beta 3	NM_000025.2
AGTR1	angiotensin II receptor, type 1	NM_031850.3
AHR	aryl hydrocarbon receptor	NM_001621.4
AKAP13	A kinase (PRKA) anchor protein 13	NM_006738.5
AKR1C2	aldo-keto reductase family 1, member C2	NM_001354.5
AKR1C3	aldo-keto reductase family 1, member C3	NM_003739.5
AKR1C4	aldo-keto reductase family 1, member C4	NM_001818.3
AKT1	v-akt murine thymoma viral oncogene homolog 1	NM_005163.2
ALDH1A1	aldehyde dehydrogenase 1 family, member A1	NM_000689.4
ALDH2	aldehyde dehydrogenase 2 family (mitochondrial)	NM_000690.3
ALDH3A1	aldehyde dehydrogenase 3 family, member A1	NM_000691.4
ANKLE1	ankyrin repeat and LEM domain containing 1	NM_152363.5
ANKRD16	ankyrin repeat domain 16	NM_019046.2
APC	adenomatous polyposis coli	NM_000038.4
APEX1	APEX nuclease 1	NM_001641.3
APOE	apolipoprotein E	NM_001302688.1
AR	androgen receptor	NM_000044.3
ARHGAP10	Rho GTPase activating protein 10	NM_024605.3
ARHGEF10L	Rho guanine nucleotide exchange factor 10-like protein	NM_018125.3
ARL11	ADP-ribosylation factor-like 11	NM_138450.5
ARNT	aryl hydrocarbon receptor nuclear translocator	NM_001668.3
ATF1	activating transcription factor 1	NM_005171.4
ATM	ATM serine/threonine kinase	NM_000051.3
ATMIN	ATM interactor	NM_015251.2
ATR	ATR serine/threonine kinase	NM_001184.3
ATRIP	ATR interacting protein	NM_130384.2
AURKA	aurora kinase A	NM_198433.1
AXIN2	axin 2	NM_004655.3
BAP1	BRCA1 associated protein-1	NM_004656.2
BARD1	BRCA1 associated RING domain 1	NM_000465.2
BAG6	BCL2-associated athanogene 6	NM_004639.3
BCL2	B-cell CLL/lymphoma 2	NM_000633.2
BCL6	B-cell CLL/lymphoma 6	NM_001706.4
BCR	breakpoint cluster region	NM_004327.3
BID	BH3 interacting domain death agonist	NM_197966.2
BLM	Bloom syndrome, RecQ helicase-like	NM_000057.2
BRAP	BRCA1 associated protein	NM_006768
BRCA1	breast cancer 1, early onset	NM_007294.3
BRCA2	breast cancer 2, early onset	NM_000059.3
BRCC3	BRCA1/BRCA2-containing complex, subunit 3	NM_024332.3
BRE	brain and reproductive organ-expressed	NM_004899.4
BRIP1	BRCA1 interacting protein C-terminal helicase 1	NM_032043.2
BRSK1	BR serine/threonine kinase 1	NM_032430.1
BTBD3	BTB (POZ) domain containing 3	NM_181443.3

BTLA	B and T lymphocyte associated	NM_181780.3
BUB1B	BUB1 mitotic checkpoint serine/threonine kinase B	NM_001211.5
C11orf30	chromosome 11 open reading frame 30	NM_001300942.1
MIEN1	migration and invasion enhancer 1	NM_032339.3
C19ORF40	chromosome 19 open reading frame 40	NM_152266.4
BABAM1	BRISC and BRCA1 A complex member 1	NM_001033549.2
C1QA	complement component 1, q subcomponent, A chain	NM_015991.2
AP5S1	adaptor-related protein complex 5, sigma 1 subunit	NM_001204446.1
SLC52A3	solute carrier family 52 (riboflavin transporter)	NM_033409.3
C2orf43	lipid droplet associated hydrolase	NM_021925.3
CCDC170	coiled-coil domain containing 170	NM_025059.3
CAMK1D	calcium/calmodulin-dependent protein kinase ID	NM_153498.2
CARM1	coactivator-associated arginine methyltransferase 1	NM_199141.1
CASP10	caspase 10, apoptosis-related cysteine peptidase	NM_032977.3
CASP8	caspase 8, apoptosis-related cysteine peptidase	NM_001228.4
CAT	catalase	NM_001752.3
CBR1	carbonyl reductase 1	NM_001757.3
CBR3	carbonyl reductase 3	NM_001236.3
CBS	cystathionine-beta-synthase	NM_000071.2
CCND1	cyclin D1	NM_053056.2
CCND2	cyclin D2	NM_001759.3
CCND3	cyclin D3	NM_001760.4
CCNE1	cyclin E1	NM_001238.2
CCR5	chemokine (C-C motif) receptor 5 (gene/pseudogene)	NM_000579.3
CD44	CD44 molecule (Indian blood group)	NM_001001391.1
CDH1	cadherin 1, type 1, E-cadherin (epithelial)	NM_004360.3
CDK2	cyclin-dependent kinase 2	NM_001798.4
CDK4	cyclin-dependent kinase 4	NM_000075.3
CDK6	cyclin-dependent kinase 6	NM_001145306.1
CDK8	cyclin-dependent kinase 8	NM_001260.1
CDKN1A	cyclin-dependent kinase inhibitor 1A (p21, Cip1)	NM_078467.2
CDKN1B	cyclin-dependent kinase inhibitor 1B (p27, Kip1)	NM_004064.4
CDKN2A	cyclin-dependent kinase inhibitor 2A	NM_000077.4
CDKN2C	cyclin-dependent kinase inhibitor 2C (p18)	NM_001262.2
CDKN2D	cyclin-dependent kinase inhibitor 2D (p19)	NM_079421.2
CLPTM1L	CLPTM1-like	NM_030782.3
CLSPN	claspin	NM_022111.3
CNTNAP4	contactin associated protein-like 4	NM_033401.3
NELFB	negative elongation factor complex member B	NM_015456.3
COL11A1	collagen, type XI, alpha 1	NM_080629.2
COL18A1	collagen, type XVIII, alpha 1	NM_030582.3
COL1A1	collagen, type I, alpha 1	NM_000088.3
COL4A1	collagen, type IV, alpha 1	NM_001845.4
COMT	catechol-O-methyltransferase	NM_000754.3
CREBBP	CREB binding protein	NM_004380.2
CRY2	cryptochrome circadian clock 2	NM_021117.3
CSNK1D	casein kinase 1, delta	NM_001893.4
CSNK1E	casein kinase 1, epsilon	NM_001894.4
CTCF	CCCTC-binding factor	NM_006565.3
CTLA4	cytotoxic T-lymphocyte-associated protein 4	NM_005214.4
CTNNB1	catenin	NM_001904.3
CWF19L2	CWF19-like 2	NM_152434.2
CXCL12	chemokine (C-X-C motif) ligand 12	NM_000609.6
CYP11A1	cytochrome P450, family 11, subfamily A, polypeptide 1	NM_000781.2
CYP11B2	cytochrome P450, family 11, subfamily B, polypeptide 2	NM_000498.3
CYP17A1	cytochrome P450, family 17, subfamily A, polypeptide 1	NM_000102.3
CYP19A1	cytochrome P450, family 19, subfamily A, polypeptide 1	NM_031226.2
CYP1A1	cytochrome P450, family 1, subfamily A, polypeptide 1	NM_000499.3
CYP1A2	cytochrome P450, family 1, subfamily A, polypeptide 2	NM_000761.3

CYP1B1	cytochrome P450, family 1, subfamily B, polypeptide 1	NM_000104.3
CYP2B6	cytochrome P450, family 2, subfamily B, polypeptide 6	NM_000767.4
CYP2C19	cytochrome P450, family 2, subfamily C, polypeptide 19	NM_000769.1
CYP2C8	cytochrome P450, family 2, subfamily C, polypeptide 8	NM_000770.3
CYP2C9	cytochrome P450, family 2, subfamily C, polypeptide 9	NM_000771.3
CYP2D6	cytochrome P450, family 2, subfamily D, polypeptide 6	NM_000106.5
CYP2E1	cytochrome P450, family 2, subfamily E, polypeptide 1	NM_000773.3
CYP3A4	cytochrome P450, family 3, subfamily A, polypeptide 4	NM_017460.5
CYP3A5	cytochrome P450, family 3, subfamily A, polypeptide 5	NM_000777.4
DCC	DCC netrin 1 receptor	NM_005215.3
DCLRE1B	DNA cross-link repair 1B	NM_022836.3
DCLRE1C	DNA cross-link repair 1C	NM_001033855.1
DHFR	dihydrofolate reductase	NM_000791.3
DLC1	DLC1 Rho GTPase activating protein	NM_182643.2
DMBT1	deleted in malignant brain tumors 1	NM_007329.2
DMC1	DNA meiotic recombinase 1	NM_007068.3
DNMT1	DNA (cytosine-5-)-methyltransferase 1	NM_001130823.1
DNMT3B	DNA (cytosine-5-)-methyltransferase 3 beta	NM_006892.3
DPYD	dihydropyrimidine dehydrogenase	NM_000110.3
DRD2	dopamine receptor D2	NM_000795.3
DRD3	dopamine receptor D3	NM_000796.5
DYNC2H1	dynein, cytoplasmic 2, heavy chain 1	NM_001080463.1
EDN1	endothelin 1	NM_001955.4
EGF	epidermal growth factor	NM_001963.4
EGFR	epidermal growth factor receptor	NM_005228.3
EHMT1	euchromatic histone-lysine N-methyltransferase 1	NM_024757.4
EHMT2	euchromatic histone-lysine N-methyltransferase 2	NM_001289413.1
ECHDC1	ethylmalonyl-CoA decarboxylase 1	NM_001002030.1
EIF3A	eukaryotic translation initiation factor 3, subunit A	NM_003750.2
ENAH	enabled homolog (Drosophila)	NM_001008493.1
EP300	E1A binding protein p300	NM_001429.3
EPAS1	endothelial PAS domain protein 1	NM_001430.4
EPHX1	epoxide hydrolase 1, microsomal	NM_001136018.3
ERAP1	endoplasmic reticulum aminopeptidase 1	NM_016442.3
ERBB2	erb-b2 receptor tyrosine kinase 2	NM_004448.2
ERBB4	erb-b2 receptor tyrosine kinase 4	NM_005235.2
ERCC1	excision repair cross-complementation group 1	NM_001983.3
ERCC2	excision repair cross-complementation group 2	NM_000400.3
ERCC3	excision repair cross-complementation group 3	NM_000122.1
ERCC4	excision repair cross-complementation group 4	NM_005236.2
ERCC5	excision repair cross-complementation group 5	NM_000123.3
ERCC6	excision repair cross-complementation group 6	NM_000124.2
ESPL1	extra spindle pole bodies homolog 1	NM_012291.4
ESR1	estrogen receptor 1	NM_001122742.1
ESR2	estrogen receptor 2	NM_001040275.1
EXO1	exonuclease 1	NM_130398.3
EYA2	EYA transcriptional coactivator and phosphatase 2	NM_005244.4
F2	coagulation factor II (thrombin)	NM_000506.3
F5	coagulation factor V (proaccelerin, labile factor)	NM_000130.4
FAM175A	family with sequence similarity 175, member A	NM_139076.2
FAM175B	family with sequence similarity 175, member B	NM_032182.3
FANCA	Fanconi anemia, complementation group A	NM_000135.2
FANCB	Fanconi anemia, complementation group B	NM_001018113.1
FANCC	Fanconi anemia, complementation group C	NM_000136.2
FANCD2	Fanconi anemia, complementation group D2	NM_033084.3
FANCE	Fanconi anemia, complementation group E	NM_021922.2
FANCF	Fanconi anemia, complementation group F	NM_022725.3
FANCG	Fanconi anemia, complementation group G	NM_004629.1
FANCI	Fanconi anemia, complementation group I	NM_001113378.1

FANCL	Fanconi anemia, complementation group L	NM_018062.3
FANCM	Fanconi anemia, complementation group M	NM_020937.2
FAS	Fas cell surface death receptor	NM_000043.4
FASLG	Fas ligand (TNF superfamily, member 6)	NM_000639.1
FBLN2	fibulin 2	NM_001165035.1
FBN1	fibrillin 1	NM_000138.4
FBXL7	F-box and leucine-rich repeat protein 7	NM_012304.4
FBXO10	F-box protein 10	NM_012166.2
FBXW7	F-box and WD repeat domain containing 7	NM_033632.3
FCRL3	Fc receptor-like 3	NM_052939.3
FEN1	flap structure-specific endonuclease 1	NM_004111.5
FGF10	fibroblast growth factor 10	NM_004465.1
FGF12	fibroblast growth factor 12	NM_021032.4
FGFR1	fibroblast growth factor receptor 1	NM_023110.2
FGFR2	fibroblast growth factor receptor 2	NM_000141.4
FGFR4	fibroblast growth factor receptor 4	NM_213647.2
FLT1	fms-related tyrosine kinase 1	NM_002019.4
FOXE1	forkhead box E1	NM_004473.3
FOXP4	forkhead box P4	NM_001012426.1
GABBR1	gamma-aminobutyric acid (GABA) B receptor, 1	NM_001470.2
GADD45A	growth arrest and DNA-damage-inducible, alpha	NM_001924.3
GATA3	GATA binding protein 3	NM_001002295.1
GC	group-specific component (vitamin D binding protein)	NM_000583.3
GH1	growth hormone 1	NM_000515.4
GHR	growth hormone receptor	NM_000163.4
GHRH	growth hormone releasing hormone	NM_021081.4
GHRHR	growth hormone releasing hormone receptor	NM_000823.3
GHRL	ghrelin/obestatin prepropeptide	NM_001134944.1
GHSR	growth hormone secretagogue receptor	NM_198407.2
GLG1	golgi glycoprotein 1	NM_012201.5
GLO1	glyoxalase I	NM_006708.2
GNB3	guanine nucleotide binding protein (G protein)	NM_002075.3
GNRH1	gonadotropin-releasing hormone 1	NM_001083111.1
GNRHR	gonadotropin-releasing hormone receptor	NM_000406.2
GPER1	G protein-coupled estrogen receptor 1	NM_001039966.1
GPRC6A	G protein-coupled receptor, class C, group 6, member A	NM_148963.3
GPX4	glutathione peroxidase 4	NM_001039847.2
GRB7	growth factor receptor-bound protein 7	NM_005310.3
GRIK1	glutamate receptor, ionotropic, kainate 1	NM_000830.3
GSK3B	glycogen synthase kinase 3 beta	NM_002093.3
GSR	glutathione reductase	NM_000637.3
GSTA1	glutathione S-transferase alpha 1	NM_145740.3
GSTA2	glutathione S-transferase alpha 2	NM_000846.4
GSTM1	glutathione S-transferase mu 1	NM_000561.3
GSTM2	glutathione S-transferase mu 2	NM_000848.3
GSTM3	glutathione S-transferase mu 3	NM_000849.4
GSTO1	glutathione S-transferase omega 1	NM_001191003.1
GSTO2	glutathione S-transferase omega 2	NM_183239.1
GSTP1	glutathione S-transferase pi 1	NM_000852.3
GSTT1	glutathione S-transferase theta 1	NM_000853.3
GSTZ1	glutathione S-transferase zeta 1	NM_145870.2
HCN1	hyperpolarization activated cyclic nucleotide gated pota	NM_021072.3
HDAC2	histone deacetylase 2	NM_001527.3
HELQ	helicase, POLQ-like	NM_133636.3
HFE	hemochromatosis	NM_000410.3
HIF1A	hypoxia inducible factor 1, alpha subunit	NM_001530.3
HLA-A	major histocompatibility complex, class I, A	NM_002116.7
HLA-B	major histocompatibility complex, class I, B	NM_005514.6
HLA-C	major histocompatibility complex, class I, C	NM_001243042.1

HLA-DPA1	major histocompatibility complex, class	NM_001242525.1
HLA-DPB1	major histocompatibility complex, class	NM_002121.5
HLA-DQA1	major histocompatibility complex, class	NM_002122.3
HLA-DQB1	major histocompatibility complex, class	NM_002123.4
HLA-DRB1	major histocompatibility complex, class	NM_001243965.1
HLA-F	major histocompatibility complex, class I, F	NM_001098479.1
HMOX1	heme oxygenase 1	NM_002133.2
HNF4A	hepatocyte nuclear factor 4, alpha	NM_000457.4
HSD11B1	hydroxysteroid (11-beta) dehydrogenase 1	NM_181755.2
HSD17B1	hydroxysteroid (17-beta) dehydrogenase 1	NM_000413.2
HSD17B2	hydroxysteroid (17-beta) dehydrogenase 2	NM_002153.2
HSD17B3	hydroxysteroid (17-beta) dehydrogenase 3	NM_000197.1
HSD17B4	hydroxysteroid (17-beta) dehydrogenase 4	NM_000414.3
HSD17B7	hydroxysteroid (17-beta) dehydrogenase 7	NM_016371.3
HSD3B1	hydroxy-delta-5-steroid dehydrogenase	NM_000862.2
HTR3B	5-hydroxytryptamine (serotonin) receptor 3B	NM_006028.4
HUS1	HUS1 checkpoint homolog	NM_004507.3
CHEK1	checkpoint kinase 1	NM_001114121.2
CHEK2	checkpoint kinase 2	NM_007194.3
CHST3	carbohydrate (chondroitin 6) sulfotransferase 3	NM_004273.4
ICAM5	intercellular adhesion molecule 5, telencephalin	NM_003259.3
IDH1	isocitrate dehydrogenase 1 (NADP+)	NM_005896.3
IDH2	isocitrate dehydrogenase 2 (NADP+)	NM_002168.3
IFNG	interferon, gamma	NM_000619.2
IGF1	insulin-like growth factor 1 (somatomedin C)	NM_001111283.1
IGF1R	insulin-like growth factor 1 receptor	NM_000875.4
IGF2	insulin-like growth factor 2	NM_000612.5
IGF2R	insulin-like growth factor 2 receptor	NM_000876.2
IGFALS	insulin-like growth factor binding protein	NM_001146006.1
IGFBP1	insulin-like growth factor binding protein 1	NM_000596.2
IGFBP2	insulin-like growth factor binding protein 2	NM_000597.2
IGFBP3	insulin-like growth factor binding protein 3	NM_001013398.1
IGFBP4	insulin-like growth factor binding protein 4	NM_001552.2
IGFBP5	insulin-like growth factor binding protein 5	NM_000599.3
IGFBP6	insulin-like growth factor binding protein 6	NM_002178.2
IKZF3	IKAROS family zinc finger 3	NM_012481.4
IL10	interleukin 10	NM_000572.2
IL13	interleukin 13	NM_002188.2
IL1A	interleukin 1, alpha	NM_000575.3
IL1B	interleukin 1, beta	NM_000576.2
IL1RN	interleukin 1 receptor antagonist	NM_173841.2
IL23R	interleukin 23 receptor	NM_144701.2
IL4	interleukin 4	NM_000589.3
IL4R	interleukin 4 receptor	NM_000418.3
IL6	interleukin 6	NM_000600.3
IL8	chemokine (C-X-C motif) ligand 8	NM_000584.3
INS	insulin	NM_001185097.1
INSIG2	insulin induced gene 2	NM_016133.2
INSR	insulin receptor	NM_000208.2
IRS1	insulin receptor substrate 1	NM_005544.2
IRS2	insulin receptor substrate 2	NM_003749.2
ITGA2	integrin, alpha 2 (CD49B)	NM_002203.3
ITGA9	integrin, alpha 9	NM_002207.2
ITGB3	integrin, beta 3	NM_000212.2
ITGB4	integrin, beta 4	NM_000213.3
JAK2	Janus kinase 2	NM_004972.3
JMY	junction mediating and regulatory protein, p53 cofactor	NM_152405.4
KAT5	K(lysine) acetyltransferase 5	NM_182710.2
KCNH1	potassium channel, voltage gated eag related subfamily H	NM_172362.2

KCNK2	potassium channel, two pore domain subfamily K, member 2	NM_001017424.2
KDR	kinase insert domain receptor	NM_002253.2
AP5Z1	adaptor-related protein complex 5, zeta 1 subunit	NM_014855.2
KIF1B	kinesin family member 1B	NM_015074.3
KLF6	Kruppel-like factor 6	NM_001300.5
KLK3	kallikrein-related peptidase 3	NM_001648.2
KPNB1	karyopherin (importin) beta 1	NM_002265.5
KRAS	Kirsten rat sarcoma viral oncogene homolog	NM_004985.4
LEP	leptin	NM_000230.2
LEPR	leptin receptor	NM_002303.5
LHCGR	luteinizing hormone/choriogonadotropin receptor	NM_000233.3
LIG1	ligase I, DNA, ATP-dependent	NM_000234.1
LIG3	ligase III, DNA, ATP-dependent	NM_013975.3
LIG4	ligase IV, DNA, ATP-dependent	NM_002312.3
LSP1	lymphocyte-specific protein 1	NM_002339.2
LTA	lymphotoxin alpha	NM_000595.3
LUM	lumican	NM_002345.3
LZTS1	leucine zipper, putative tumor suppressor 1	NM_021020.3
MAD2L1	MAD2 mitotic arrest deficient-like 1	NM_002358.3
MAP3K1	mitogen-activated protein kinase kinase kinase 1	NM_005921.1
MBD2	methyl-CpG binding domain protein 2	NM_003927.4
MCM8	minichromosome maintenance complex component 8	NM_032485.5
MCPH1	microcephalin 1	NM_024596.3
MDC1	mediator of DNA-damage checkpoint 1	NM_014641.2
MDM2	MDM2 proto-oncogene, E3 ubiquitin protein ligase	NM_002392.5
MDM4	MDM4, p53 regulator	NM_002393.4
MED1	mediator complex subunit 1	NM_004774.3
MFAP1	microfibrillar-associated protein 1	NM_005926.2
MGMT	O-6-methylguanine-DNA methyltransferase	NM_002412.3
MLH1	mutL homolog 1	NM_000249.3
MMP1	matrix metalloproteinase 1	NM_002421.3
MMP12	matrix metalloproteinase 12	NM_002426.4
MMP13	matrix metalloproteinase 13	NM_002427.3
MMP2	matrix metalloproteinase 2	NM_004530.4
MMP3	matrix metalloproteinase 3	NM_002422.3
MMP7	matrix metalloproteinase 7	NM_002423.3
MMP8	matrix metalloproteinase 8	NM_002424.2
MMP9	matrix metalloproteinase 9	NM_004994.2
MPL	MPL proto-oncogene, thrombopoietin receptor	NM_005373.2
MPO	myeloperoxidase	NM_000250.1
MRE11A	MRE11 meiotic recombination 11 homolog A	NM_005591.3
MRPS30	mitochondrial ribosomal protein S30	NM_016640.3
MSH2	mutS homolog 2	NM_000251.2
MSH3	mutS homolog 3	NM_002439.4
MSH5	mutS homolog 5	NM_172165.3
MSH6	mutS homolog 6	NM_000179.2
MSR1	macrophage scavenger receptor 1	NM_138715.2
MTHFD1	methylenetetrahydrofolate dehydrogenase	NM_005956.3
MTHFR	methylenetetrahydrofolate reductase	NM_005957.4
MTR	5-methyltetrahydrofolate-homocysteine methyltransferase	NM_000254.2
MTRR	5-methyltetrahydrofolate-homocysteine methyltransferase	NM_002454.2
MUS81	MUS81 structure-specific endonuclease subunit	NM_025128.4
MUTYH	mutY homolog	NM_001128425.1
MYC	v-myc avian myelocytomatosis viral oncogene homolog	NM_002467.4
NAT1	N-acetyltransferase 1 (arylamine N-acetyltransferase)	NM_001160179.2
NAT2	N-acetyltransferase 2 (arylamine N-acetyltransferase)	NM_000015.2
NBN	nibrin	NM_002485.4
NCAM1	neural cell adhesion molecule 1	NM_181351.4
NCOA1	nuclear receptor coactivator 1	NM_003743.4

NCOA3	nuclear receptor coactivator 3	NM_181659.2
NCOR1	nuclear receptor corepressor 1	NM_006311.3
NCOR2	nuclear receptor corepressor 2	NM_006312.5
NEK10	NIMA-related kinase 10	NM_199347.2
NEK2	NIMA-related kinase 2	NM_001204182.1
NEUROD2	neuronal differentiation 2	NM_006160.3
NFKB1	nuclear factor of kappa light polypeptide gene enhancer	NM_003998.3
NFKBIA	nuclear factor of kappa light polypeptide gene enhancer	NM_020529.2
NHEJ1	nonhomologous end-joining factor 1	NM_024782.2
NME1	NME/NM23 nucleoside diphosphate kinase 1	NM_198175.1
NOD2	nucleotide-binding oligomerization domain containing 2	NM_022162.1
NOS3	nitric oxide synthase 3 (endothelial cell)	NM_000603.4
NOTCH2	notch 2	NM_024408.3
NPAS2	neuronal PAS domain protein 2	NM_002518.3
NPM1	nucleophosmin (nucleolar phosphoprotein B23, numatrin)	NM_002520.6
NQO1	NAD(P)H dehydrogenase, quinone 1	NM_000903.2
NQO2	NAD(P)H dehydrogenase, quinone 2	NM_000904.4
NR1I2	nuclear receptor subfamily 1, group I, member 2	NM_003889.3
NR1I3	nuclear receptor subfamily 1, group I, member 3	NM_001077482.2
NR5A2	nuclear receptor subfamily 5, group A, member 2	NM_205860.2
NRP1	neuropilin 1	NM_003873.5
NRP2	neuropilin 2	NM_201266.1
NTRK2	neurotrophic tyrosine kinase, receptor, type 2	NM_006180.4
NUMA1	nuclear mitotic apparatus protein 1	NM_006185.3
OGG1	8-oxoguanine DNA glycosylase	NM_016828.2
OPN5	opsin 5	NM_181744.3
PADI4	peptidyl arginine deiminase, type IV	NM_012387.2
PADI6	peptidyl arginine deiminase, type VI	NM_207421.4
PAK1	p21 protein (Cdc42/Rac)-activated kinase 1	NM_001128620.1
PALB2	partner and localizer of BRCA2	NM_024675.3
PARP1	poly (ADP-ribose) polymerase 1	NM_001618.3
PBOV1	prostate and breast cancer overexpressed 1	NM_021635.2
PCDH7	protocadherin 7	NM_001173523.1
PCNA	proliferating cell nuclear antigen	NM_002592.2
PGAP3	post-GPI attachment to proteins 3	NM_033419.4
PGD	phosphogluconate dehydrogenase	NM_002631.2
PGR	progesterone receptor	NM_000926.4
PHB	prohibitin	NM_002634.3
PIK3CA	phosphatidylinositol-4,5-bisphosphate 3-kinase	NM_006218.2
PIK3CB	phosphatidylinositol-4,5-bisphosphate 3-kinase	NM_006219.2
PLA2G5	phospholipase A2, group V	NM_000929.2
PLAU	plasminogen activator, urokinase	NM_002658.3
PLCE1	phospholipase C, epsilon 1	NM_016341.3
PLK1	polo-like kinase 1	NM_005030.3
POLB	polymerase (DNA directed), beta	NM_002690.2
POLD1	polymerase (DNA directed), delta 1, catalytic subunit	NM_001256849.1
PON1	paraoxonase 1	NM_000446.5
POT1	protection of telomeres 1	NM_015450.2
POU1F1	POU class 1 homeobox 1	NM_001122757.2
PPARA	peroxisome proliferator-activated receptor alpha	NM_005036.4
PPARG	peroxisome proliferator-activated receptor gamma	NM_138712.3
PPARGC1A	peroxisome proliferator-activated recept	NM_013261.3
PPM1D	protein phosphatase, Mg ²⁺ /Mn ²⁺ dependent, 1D	NM_003620.3
PPP1R1B	protein phosphatase 1, regulatory (inhibitor) subunit 1B	NM_032192.3
PPP2R1B	protein phosphatase 2, regulatory subunit A, beta	NM_181699.2
PPP2R5B	protein phosphatase 2, regulatory subunit B', beta	NM_006244.3
PRDM2	PR domain containing 2, with ZNF domain	NM_012231.4
PREX2	phosphatidylinositol-3,4,5-trisphosphate-dependent Race	NM_024870.2
PRKCA	protein kinase C, alpha	NM_002737.2

PRKDC	protein kinase, DNA-activated, catalytic polypeptide	NM_006904.6
PRL	prolactin	NM_001163558.2
PRLR	prolactin receptor	NM_000949.6
PSCA	prostate stem cell antigen	NM_005672.4
PTEN	phosphatase and tensin homolog	NM_000314.4
PTGDS	prostaglandin D2 synthase 21kDa (brain)	NM_000954.5
PTGES	prostaglandin E synthase	NM_004878.4
PTGIS	prostaglandin I2 (prostacyclin) synthase	NM_000961.3
PTGS1	prostaglandin-endoperoxide synthase 1	NM_000962.2
PTGS2	prostaglandin-endoperoxide synthase 2	NM_000963.3
PTPRJ	protein tyrosine phosphatase, receptor type, J	NM_002843.3
PTPRN2	protein tyrosine phosphatase, receptor type	NM_002847.3
PTTG1	pituitary tumor-transforming 1	NM_001282382.1
RAB27A	RAB27A, member RAS oncogene family	NM_183235.1
RAD1	RAD1 checkpoint DNA exonuclease	NM_002853.3
RAD17	RAD17 homolog	NM_133338.2
RAD18	RAD18 E3 ubiquitin protein ligase	NM_020165.3
RAD23B	RAD23 homolog B	NM_002874.4
RAD50	RAD50 homolog	NM_005732.3
RAD51	RAD51 recombinase	NM_133487.3
RAD51AP1	RAD51 associated protein 1	NM_001130862.1
RAD51C	RAD51 paralog C	NM_058216.1
RAD51B	RAD51 paralog B	NM_133509.3
RAD51D	RAD51 paralog D	NM_002878.3
RAD52	RAD52 homolog	NM_134424.3
RAD54B	RAD54 homolog B	NM_012415.3
RAD54L	RAD54-like	NM_001142548.1
RAD9A	RAD9 homolog A	NM_004584.2
RASSF1	Ras association (RalGDS/AF-6) domain family member 1	NM_170714.1
RB1	retinoblastoma 1	NM_000321.2
RBBP8	retinoblastoma binding protein 8	NM_002894.2
RCC2	regulator of chromosome condensation 2	NM_018715.3
RECQL	RecQ helicase-like	NM_002907.3
RECQL4	RecQ protein-like 4	NM_004260.3
RECQL5	RecQ protein-like 5	NM_004259.6
RFC1	replication factor C (activator 1)	NM_002913.4
RFC2	replication factor C (activator 1) 2	NM_181471.2
RFC4	replication factor C (activator 1) 4	NM_002916.3
RFX6	regulatory factor X, 6	NM_173560.3
RGS1	regulator of G-protein signaling 1	NM_002922.3
RHOU	ras homolog family member U	NM_021205.5
RNASEL	ribonuclease L	NM_021133.3
DROSHA	drosha, ribonuclease type III	NM_013235.4
RNF146	ring finger protein 146	NM_001242844
RNF168	ring finger protein 168, E3 ubiquitin protein ligase	NM_152617.3
RNF8	ring finger protein 8, E3 ubiquitin protein ligase	NM_003958.3
ROPN1L	rhophilin associated tail protein 1-like	NM_031916.4
RPA1	replication protein A1, 70kDa	NM_002945.3
RRP1B	ribosomal RNA processing 1B	NM_015056.2
S100A3	S100 calcium binding protein A3	NM_002960.1
SART1	squamous cell carcinoma antigen recognized by T cells	NM_005146.4
SCARB1	scavenger receptor class B, member 1	NM_005505.4
SERPINE1	serpin peptidase inhibitor, clade E	NM_000602.4
SETX	senataxin	NM_015046.5
SHBG	sex hormone-binding globulin	NM_001040.4
SHC1	SHC (Src homology 2 domain containing)	NM_003029.4
SHFM1	split hand/foot malformation (ectrodactyly) type 1	NM_006304.1
SHMT1	serine hydroxymethyltransferase 1 (soluble)	NM_004169.4
SHPRH	SNF2 histone linker PHD RING helicase	NM_001042683.2

SIPA1	signal-induced proliferation-associated 1	NM_153253.29
SIRT6	sirtuin 6	NM_016539.2
SKP2	S-phase kinase-associated protein 2	NM_005983.3
SLC22A16	solute carrier family 22	NM_033125.3
SLC4A7	solute carrier family 4	NM_003615.4
SLC6A4	solute carrier family 6	NM_001045.5
SNRNPB	small nuclear ribonucleoprotein polypeptides B and B1	NM_198216.1
SNW1	SNW domain containing 1	NM_012245.2
SOD1	superoxide dismutase 1, soluble	NM_000454.4
SOD2	superoxide dismutase 2, mitochondrial	NM_000636.2
SPOCK2	sparc/osteonectin, cwcv and kazal-like domains proteogly	NM_014767.2
SRD5A1	steroid-5-alpha-reductase, alpha polypeptide 1	NM_001047.2
SRD5A2	steroid-5-alpha-reductase, alpha polypeptide 2	NM_000348.3
SST	somatostatin	NM_001048.3
SSTR1	somatostatin receptor 1	NM_001049.2
SSTR2	somatostatin receptor 2	NM_001050.2
SSTR3	somatostatin receptor 3	NM_001051.4
SSTR4	somatostatin receptor 4	NM_001052.2
SSTR5	somatostatin receptor 5	NM_001053.3
ST14	suppression of tumorigenicity 14	NM_021978.3
STARD3	StAR-related lipid transfer (START) domain containing 3	NM_006804.3
STK11	serine/threonine kinase 11	NM_000455.4
STK39	serine threonine kinase 39	NM_013233.2
STS	steroid sulfatase (microsomal), isozyme S	NM_000351.4
STXBP4	syntaxin binding protein 4	NM_178509.5
STXBP5	syntaxin binding protein 5 (tomosyn)	NM_001127715.2
SULF1	sulfatase 1	NM_001128205.1
SULT1A1	sulfotransferase family, cytosolic, 1A, phenol-preferrin	NM_001055.3
SULT1A2	sulfotransferase family, cytosolic, 1A, phenol-preferrin	NM_001054.3
SULT1E1	sulfotransferase family 1E, estrogen-preferring, member	NM_005420.2
TBXAS1	thromboxane A synthase 1 (platelet)	NM_030984.3
TCF7L2	transcription factor 7-like 2 (T-cell specific, HMG-box)	NM_030756.4
TCL1A	T-cell leukemia/lymphoma 1A	NM_021966.2
TELO2	telomere maintenance 2	NM_016111.3
TEP1	telomerase-associated protein 1	NM_007110.4
TERF1	telomeric repeat binding factor (NIMA-interacting) 1	NM_017489.2
TERF2	telomeric repeat binding factor 2	NM_005652.4
TERF2IP	telomeric repeat binding factor 2, interacting protein	NM_018975.3
TERT	telomerase reverse transcriptase	NM_198253.2
TGFB1	transforming growth factor, beta 1	NM_000660.5
TGFBR1	transforming growth factor, beta receptor 1	NM_004612.2
TGFBR2	transforming growth factor, beta receptor II	NM_001024847.2
TIMP2	TIMP metalloproteinase inhibitor 2	NM_003255.4
TIMP3	TIMP metalloproteinase inhibitor 3	NM_000362.4
TLK2	tousled-like kinase 2	NM_001284333.1
TLR4	toll-like receptor 4	NM_138554.3
TNC	tenascin C	NM_002160.3
TNF	tumor necrosis factor	NM_000594.3
TNFRSF19	tumor necrosis factor receptor superfamily, member 19	NM_018647.3
TNFSF10	tumor necrosis factor (ligand) superfamily, member 10	NM_003810.3
TNKS	tankyrase, TRF1-interacting ankyrin-related ADP-ribose	NM_003747.2
TOP2A	topoisomerase (DNA) II alpha 170kDa	NM_001067.3
TOPBP1	topoisomerase (DNA) II binding protein 1	NM_007027.3
TOX	thymocyte selection-associated high mobility group box	NM_014729.2
TOX3	high mobility group box family member 3	NM_001146188.1
TP53	tumor protein p53	NM_000546.5
TP53BP1	tumor protein p53 binding protein 1	NM_001141980.1
TP63	tumor protein p63	NM_003722.4
TP73	tumor protein p73	NM_005427.3

TPTE2	transmembrane phosphoinositide 3-phosphatase and tensin	NM_199254.2
TRIM29	tripartite motif containing 29	NM_012101.3
TRIO	trio Rho guanine nucleotide exchange factor	NM_007118.2
TSHR	thyroid stimulating hormone receptor	NM_000369.2
TTK	TTK protein kinase	NM_003318.4
TUBD1	tubulin, delta 1	NM_016261.3
TXN	thioredoxin	NM_003329.3
TXN2	thioredoxin 2	NM_012473.3
TXNRD1	thioredoxin reductase 1	NM_001093771.2
TXNRD2	thioredoxin reductase 2	NM_006440.3
TYMS	thymidylate synthetase	NM_001071.2
UBE2A	ubiquitin-conjugating enzyme E2A	NM_003336.3
UBE2B	ubiquitin-conjugating enzyme E2B	NM_003337.3
UBE2I	ubiquitin-conjugating enzyme E2I	NM_003345.4
UBE2N	ubiquitin-conjugating enzyme E2N	NM_003348.3
UBE2V2	ubiquitin-conjugating enzyme E2 variant 2	NM_003350.2
UBE4B	ubiquitination factor E4B	NM_001105562.2
UGT1A6	UDP glucuronosyltransferase 1 family, polypeptide A6	NM_001072.3
UGT2B15	UDP glucuronosyltransferase 2 family, polypeptide B15	NM_001076.3
UGT2B4	UDP glucuronosyltransferase 2 family, polypeptide B4	NM_021139.2
UGT2B7	UDP glucuronosyltransferase 2 family, polypeptide B7	NM_001074.2
UIMC1	ubiquitin interaction motif containing 1	NM_001199297.1
VDR	vitamin D (1,25- dihydroxyvitamin D3) receptor	NM_001017535.1
VEGFA	vascular endothelial growth factor A	NM_001025366.2
WEE1	WEE1 G2 checkpoint kinase	NM_003390.3
WNT7A	wingless-type MMTV integration site family, member 7A	NM_004625.3
WRN	Werner syndrome, RecQ helicase-like	NM_000553.4
XPA	xeroderma pigmentosum, complementation group A	NM_000380.3
XPC	xeroderma pigmentosum, complementation group C	NM_004628.4
XRCC1	X-ray repair complementing defective repair in Chinese h	NM_006297.2
XRCC2	X-ray repair complementing defective repair in Chinese h	NM_005431.1
XRCC3	X-ray repair complementing defective repair in Chinese h	NM_001100119.1
XRCC4	X-ray repair complementing defective repair in Chinese h	NM_022406.2
XRCC5	X-ray repair complementing defective repair in Chinese h	NM_021141.3
XRCC6	X-ray repair complementing defective repair in Chinese h	NM_001469.4
YAP1	Yes-associated protein 1	NM_001130145.2
YPEL2	yippee-like 2 (Drosophila)	NM_001005404.3
ZMIZ1	zinc finger, MIZ-type containing 1	NM_020338.3
ZNF350	zinc finger protein 350	NM_021632.3
ZNF365	zinc finger protein 365	NM_199451.2
ZNF577	zinc finger protein 577	NM_032679.2
SLX4	SLX4 structure-specific endonuclease subunit	NM_032444.2
DNAJC21	DnaJ (Hsp40) homolog, subfamily C, member 21	NM_194283.3
MLH3	mutL homolog 3	NM_001040108.1
ESRRG	estrogen-related receptor gamma	NM_001438.3
LRIG1	leucine-rich repeats and immunoglobulin-like domains 1	NM_015541.2
NF1	neurofibromin 1	NM_001042492.2
PIK3CG	phosphatidylinositol-4,5-bisphosphate 3-kinase, catalyti	NM_002649.3
PTTG2	pituitary tumor-transforming 2	NM_006607.2
PMS2	PMS2 postmeiotic segregation increased 2	NM_000535.5
UBE2NL	ubiquitin-conjugating enzyme E2N-like	NM_001012989.2

PŘÍLOHA II:

**Seznam a sekvencí primerů použitých pro confirmaci a charakterizaci
nalezených zárodečných variant**

Gen	Skupina genů	typ varianty	kódující s.	protein s.	RNA	Genomická koordináta	forward primer 5' - 3'	reverse primer 5' - 3'
ABCC2	6	nonsense	c.3196C>T	p.R1066X		chr10:101591826	AGTCTACGGAGCTCTGGGAT	TGAGATACTTACGCCGGCAA
ABCC4	6	indel	c.2468dupA	p.N823Kfs*12		chr13:95768245	ACATTTCAACTCCCTGCCCG	GTCAGCGGCAGCAAATCATC
ABCC4	6	nonsense	c.1150C>T	p.R384X		chr13:95858797	CGTGTTGTGACCTTCACCAC	GGCATCTGACACCAACCTGG
ABCG2	6	nonsense	c.706C>T	p.R236X		chr4:89039396	TCAGGCTGAACTAGAGCAAACA	AGACCAAACAGCACTCCTGC
ABCG2	6	nonsense	c.736C>T	p.R246X		chr4:89039366	TCAGGCTGAACTAGAGCAAACA	AGACCAAACAGCACTCCTGC
ATM	1	indel	c.3850delA	p.T1284Qfs*9		chr11:108155057	TGCTGATGGTATTAACAGTTT	GGTTGGCTATGCTAGATAATGAT
ATM	1	nonsense	c.7327C>T	p.R2443X		chr11:108200960	TCCTTAGAAGTTTGCTTTTTTC	CTGGACCAAGTGCTAGGAATA
ATR	1	nonsense	c.5342T>A	p.L1781X		chr3:142218507	TGGCACACTTTAATTTGTTGAAGA	ATGCCAGCCTGGTTTCATT
ATRIP	1	indel	c.827_828delAG	p.E276Gfs*2		chr3:48498814	TTTTGCACTGCTCCTCTGACA	GCCCATGTGGCAAGAAATTACA
BLM	1	nonsense	c.1642C>T	p.Q548X		chr15:91304245	TTGCTTTTGTGGCCTACCAG	AGGCAATGATGATTTGCTATGGT
BRIP1	1	nonsense	c.2392C>T	p.R798X		chr17:59793412	CACACAGCTGATTAGTTATATGC	CTGACATTTGGGGGAGGTG
CASP8	7	indel	c.106delG	p.E36Nfs*7		chr2:202131315	CCTTTGATGAACAAGCCAGCA	AGGCTGTCCATTCAACCCAC
CBR3	5	indel	c.533delA	p.D178Afs*46		chr21:37518509	TGGTTGACTCTGTGATATGCTT	CTTGTGGCTCAGTGGCATCT
CNTNAP4	10	nonsense	c.3913G>T	p.E1305X		chr16:76592566	TGTAGTGGAACCCACAAGGC	TTCCATTGAGAACCCCTGGC
CWF19L2	7	indel	c.1605delA	p.K535Nfs*4		chr11:107286953	TGAAGGCTACATGAGTGTGGT	GCCCGTGGTAGTTTACACA
CYP11A1	2	indel	c.835delA	p.I279Yfs*10		chr15:74635473	CGTGCTCAGTAAACGGCAG	GGAACGCCTGAACAACAGG
CYP17A1	2	nonsense	c.1072C>T	p.R358X		chr10:104592335	GAGCCACTACTGGAAGGGA	GTTGAATGCATCATGGGGCT
CYP19A1	2	indel	c.1058dupT	p.L353Ffs*10		chr15:51504723	TGGTTGCTGGTGTGCATTAGA	CGTGTGTGCTCCTGGTGAG
CYP11A2	5	nonsense	c.816T>A	p.Y272X		chr15:75042895	TCCTTCGCTACCTGCCTAAC	ACTGTTCTGAGGGGCACAA
CYP3A5	2	indel	c.92dupG	p.L32Tfs*3		chr7:99273812	CTCCCTCTCTTGACCATTCCAG	AACCTCAGAACTCCCTCCA
CYP3A5	2	indel	c.246dupG	p.A83Gfs*40		chr7:99270275	AAAGTCTGGCTTCTGGGTG	GCAAAAGTTATTGCAGTTGCCG
DCLRE1C	1	indel	c.1903dupA	p.S635Kfs*6		chr10:14950588	AGGAATCCAGCTTCACTGC	TCACCAGTTGCCAGCTTCTC
DHFR	9	indel	c.95delT	p.F32Sfs*7		chr5:79949869	CACCGCTCAGGTATCTGCC	AATTTGCTCGTGCGTTGACA
DMBT1	3	indel	c.2227delC	p.Q743Rfs*4		chr10:124351838	GAACCTCTGGTTGCAGTCGT	CCAGTCTAGGAGAGGGAGCC
DNAJC21	10	indel	c.1503delA	p.K501Nfs*10		chr5:34954040	AGAGAGCACTCAAATAATGATGGT	TGCAGATCACTGAAATTTAACTCT
DNAJC21	10	indel	c.1629delT	p.F543Lfs*4		chr5:34954717	AGAGAGCACTCAAATAATGATGGT	TGCAGATCACTGAAATTTAACTCT
DNMT1	10	indel	c.1035dupC	p.K346Qfs*35		chr19:10271072	CCGAAGCACTATCTGGGTGT	CTGACGCACCTTGGTCTAT
DPYD	9	splicing	c.1905+1G>A	p.D581_N635del	r.1741_1905del	chr1:97915614	ACGGCTGCATATTGGTGTCAAAG	TTCAGCAGAGTCAATCCACCAG
ERCC2	1	indel	c.230_231delTG	p.V77Afs*4		chr19:45872205	TCTGGGGAGGTGAAGTACT	AGGACTTGTGGTTGGACGAAA

ERCC2	1	indel	c.1703_1704delTT	p.F568Yfs*2		chr19:45856555	TGCCCCACTTATCCAGCAAT	AGAGAGCTCTGGGAAGACACC
ERCC6	1	nonsense	c.3693C>G	p.Y1231X		chr10:50678313	CAAGGCTGAAAGTTTGGGGGA	TCCCTGTGGCAAACGTATCA
ESR2	2	nonsense	c.76G>T	p.E26X		chr14:64749628	AGCCATTATACTTGCCACGA	CAGGGCATCCTGTGTTTTGG
EXO1	1	indel	c.1522dupT	p.C508Lfs*7		chr1:242042058	ATGTGTGAATGTTGGCATAATCA	CCAGTCTCTTGCCCTCTTGGT
EXO1	1	indel	c.2358delG	p.L787Yfs*37		chr1:242048762	ACCATTGCTCATGGGTTATTTTGT	GGAAGAGTTGGGAGAAAGGGA
EXO1	1	splicing	c.2212-1G>C	p.V738_K743del	r.2212_2229del	chr1:242048615	ACCATTGCTCATGGGTTATTTTGT	GGAAGAGTTGGGAGAAAGGGA
					r.2212_2229del		GACCAGACCTCCAAGCTACG	GAACACGGCCACATTCAGGT
FAM175B	1	indel	c.1084delC	p.Q362Kfs*19		chr10:126523376	ACTTCAGAAATTTGTCATGGCCT	AACAGTGCTAGGTGGAGAGT
FANCE	1	indel	c.929dupC	p.V311Sfs*2		chr6:35425721	GGGAGGGGGACTGGAGTAAA	GGTGACCCCTATCGCTGAGA
FANCL	1	indel	c.1096_1099dupATTA	p.T367Nfs*13		chr2:58386933	TGTGCCTAGCTTTCAGTAATATGT	ACACTTCCACAGTCAGCACG
FANCM	1	indel	c.3979_3980delCA	p.Q1327Vfs*16		chr14:45645936	TGTTTCGAATCAAGCACTAATACCA	TGCTTCTTTGAATCTTTGTAGGTT
FANCM	1	nonsense	c.1972C>T	p.R658X		chr14:45636336	AGTATTTAAGTGGATCGGGGTTT	CCAGTGTCTCTCCATTTGCCA
FANCM	1	nonsense	c.5101C>T	p.Q1701X		chr14:45658326	AATTGCCATGTTTCAGCTTGGT	CTAAGCACCCCTGTGAATTGCC
FANCM	1	nonsense	c.5791C>T	p.R1931X		chr14:45667921	TGGCACCAGTTTGCTCAATG	AAACTTCTATGGAATGCTTTTACGA
FBXW7	7	indel	c.310delC	p.H104Mfs*389		chr4:153332646	CCCCTGCAGAATGTGAAAACC	ACTTTCATTACCTTCTGTAAGACA
GLG1	10	nonsense	c.3520C>T	p.R1174X		chr16:74487085	TTTTCTTATCTCCTGCCCTCCC	TGCTAATGCTCTAACACGGGG
GRB7	4	nonsense	c.862C>T	p.Q288X		chr17:37900866	CCTGTACCAAGTCTGCTCA	TGTCTGAATTGGGGCTCAGA
GRB7	4	splicing	c.801+1G>C	p.(?)	r.(spl?)	chr17:37900461	CATCCTGGTGGTAGAGGGGAA	ATGGTAAACTGGAGGTGGCTG
HELQ	1	splicing	c.2677-1G>A	p.Q348Pfs*17	r.2677_2681del	chr4:84347246	TCCCCTCAGCCATCTGGTAT	TTGGGTGCTCAGGTATGTTT
					r.2677_2681del		AGGACTTGTGCTTGAAGCCT	AGCCTGTTGACAACGTTCTTG
HTR3B	10	nonsense	c.871C>T	p.Q291X		chr11:113813878	AAAAACGTTGGGTCTTCGGG	AGATGAGTGTCCAGTGACCCC
CHEK2	1	indel	c.277delT	p.W93Gfs*17		chr22:29130433	ATATCCAGCTCCTCTACCAGC	CAGAACCTTCCACCTGGTAATAC
CHEK2	1	splicing	c.444+1G>A	p.R148Vfs*6	r.[444+1_444+4ins; 444+1g>a]	chr22:29121230	TCAACAGCCCTCTGATGCATG	GCCGCTTCCAGTAACCATAAGATAATA
IL13	3	splicing	c.174+2delT	p.(?)	r.(spl?)	chr5:131994054	CATCCGCTCCTCAATCCTCTC	CATCTTAGGACCAGGCCAG
IL1A	3	splicing	c.319+2T>C	p.(?)	r.(spl?)	chr2:113539179	TACCTGCCTAGTGAGTGTGGA	CTAGGCACGCCTTGGAATC
IL8	3	nonsense	c.91G>T	p.E31X		chr4:74607285	TGATGCCTTCCATAGTCTCCA	CCACCTTCTTAATTTTAAAGTTTGC
ITGB4	10	indel	c.665delG	p.G222Efs*60		chr17:73725444	CCCCTCCTCTTTCACCCAC	CCTTGCCAGAAGTGCTCA
LRIG1	4	indel	c.3149_3150delCG	p.A1050Gfs*17		chr3:66430824	GGGAGCTGCTTAAACGGTCA	TGAGTGACGCTTGAACCCAA
MAP3K1	4	indel	c.4151dupT	p.L1384Lfs*36		chr5:56183241	AGAGCTGTGTGCCTTtagCA	GCCACGGACTTCTAGTTGGT
MGMT	1	indel	c.207_210dupACGT	p.S70Yfs*5		chr10:131334631	CTTCGGTGTTCGCTTGGTTA	ATGCGTTACCCTGCCATCAA
MMP1	8	splicing	c.105+2T>C	p.Q35Vfs*11	r.105_106ins105+1 _106-1	chr11:102668717	GCTTCTAGCTGGGATATTGGAG	GAGAGCGTATGCATGGCTAGA

					r.105_106ins105+1_106-1		CTGTTCTGGGGTGTGGTGTC	GCTCTTGGCAAATCTGGCGT
MMP12	8	nonsense	c.327C>T	p.W109X		chr11:102743618	CCTGGTCCAAAAGCATCAGTGT	GGGCTGTCTAACTGGTTCAGG
MPL	3	splicing	c.79+2T>A	p.(?)	r.(spl?)	chr1:43803600	CTGAAGGGAGGATGGGCTAAG	AGCAGGAGAGCCTCTTACTC
MPO	3	splicing	c.2031-2A>C	p.R677Wfs*73	r.[2029-109_2029-1ins;2029-2a>c]	chr17:56348226	GCATCACTTGTGTGAAAGCCC	CTCATTTTCTCAGCTGCACCC
					r.[2029-109_2029-1ins;2029-2a>c]		AGTATGGCACGCCCAACAAC	AGACACGGTGGTGTGCCT
MSH5	1	nonsense	c.541C>T	p.R181X		chr6:31711970	GCACAAGTGCTAGGGCTGAA	ACCCACCCCCACTTTACTT
MSH5	1	nonsense	c.1900C>T	p.R634X		chr6:31728554	GGGCTATGGTCAGGATTCGG	TAAAGGGCATGCAGCAGAGG
MSR1	10	indel	c.569delT	p.L190Cfs*5		chr8:16026030	AGCGTAGTTCACACGTTGTAG	ACTGCCTAAGGAGACGAGACT
NAT1	5	nonsense	c.559C>T	p.R187X		chr8:18080115	AGCCTACTCAAATCCAAGTGT	CCACAGGCCATCTTTAAAATAC
NBN	1	indel	c.657_661delACAAA	p.K219Nfs*16		chr8:90983445	CAGATAGTCACTCCGTTTACAA	CCCAAAATGAAATACGTTAACAAC
NF1	4	indel	c.5690delG	p.G1897Vfs*28		chr17:29657394	TGGTTATATCAAGTGTGCCCTT	GTTCCCTCAGAGTTCCTTCACT
NQO2	2	nonsense	c.628C>T	p.Q210X		chr6:3019821	TGTTTCACACCATTTCCTCC	GCCTCCTAGTGTGCTGCTTA
PIK3CG	4	indel	c.41_42delAG	p.E14Gfs*147		chr7:106508047	ACATGTACGCCGCTATACC	CTGCACCACCTGGTACTTGT
PPP2R1B	4	indel	c.342_343delTG	p.V115Cfs*3		chr11:111631739	GGGAATTGCCATCTGGGACA	GTCAGCTACAAGTCCCCAGT
PREX2	4	indel	c.3210_3213delAACA	p.D1072Vfs*17		chr8:69028051	ACAAGAGAAGTTTGAATGACATGCT	TCTTTGTGGAAATCCAGCATCA
PTPRJ	10	nonsense	c.1191T>A	p.Y397X		chr11:48149429	ACCGTCATGCTTTCCTAGCC	TGTGGTCAGCCAACAGTCAC
RAD18	1	indel	c.1430_1431insGCGG	p.T478Rfs*6		chr3:8923099	AAGGGGTTTCAAGCTCATTACT	TAGAGGCAGGAGGCAACTGA
RAD50	1	nonsense	c.1093C>T	p.R365X		chr5:131924420	ATCATTGGGAGAACTGGGCA	TGCCAAAATGGAGTCCAACCA
RAD51C	1	nonsense	c.502A>T	p.R168X		chr17:56774151	CCTTAGATCATCATCATGATTTGG	GGTCTCAGATGGGCACAAATGC
RAD51C	1	splicing	c.905-2_1delAG	p.L301Gfs*42	r.905_965del	chr17:56801399	ACCAAGTCAGTAAGGCCATATAC	CACAGGACTAGCTTAAGAAACC
					r.905_965del		CGTCATGACCTAGATGACCTG	GGTAACAAGTCCACTTGTACAC
RAD51D	1	indel	c.355_358deldeITGTA	p.C119Wfs*16		chr17:33434129	GACTCAGCCCATTTGTGTTG	AGCAAGTTTGAAGGCAAGGA
RASSF1	7	splicing	c.888+1G>A	p.V258Gfs*7	r.758_888del	chr3:50368764	GTGGGCGCAAGCAGTAATTT	TGGAAGTGTTCAGGGCT
					r.758_888del		CGCCGCACTTCTTTTACCT	CTCTGGATCTTCTGGCGGC
RFC1	1	indel	c.2191delA	p.R731Gfs*7		chr4:39304694	GGAAGGGAATTGGGGTCTGC	CAGCAACATGCCTTAACTGCC
RNASEL	3	nonsense	c.793G>T	p.E265X		chr1:182555149	CTCCTTGATGAGATGGGGGC	GCAGTGTCTTTTGCATCA
ROPN1L	10	nonsense	c.135T>A	p.Y45X		chr5:10448375	CATAAGCCCGGATGTTGGT	TCTGTTTTCTGGGTTGCCGT
S100A3	10	indel	c.208delG	p.V70Wfs*83		chr1:153520257	GTTGAAGTCGGCCCTCAGAA	AGGAAGGAGCCCTCATCT
STXBP4	10	splicing	c.181-1G>A	p.K60Vfs*28	r.181_287del	chr17:53076705	TGAAGGCATTCTTACGAAAAC	ACTGCTTCAAGGAGACATTCTA
					r.181_287del		GGCCTTGGCCTGAAGGACTA	GGGAGTGGATGAGGTCTTTGG

TCL1A	10	nonsense	c.253C>T	p.R85X		chr14:96178601	TGCATGGCCTAAGGAAGTGA	GCAAACCAAGATCACCCGA
VEGFA	4	splicing	c.1085+2T>C	p.(?)	r.(spl?)	chr6:43748593	GCTTTGCTTTGGTCGTTCCC	CAAATCTACCCGTTGGTGCC
WRN	1	nonsense	c.604A>T	p.K202X		chr8:30924648	TGTGGACTGCATATGAGGCTT	TGGTAGAGTTTCTGCCACTACT
WRN	1	nonsense	c.4216C>T	p.R1406X		chr8:31030535	ACTACATGATGAGATGGCAAGT	ACATGAATTGCGGAAGGCTG
XRCC4	1	indel	c.25delC	p.H9Tfs*8		chr5:82400763	TTTGTGTAGCTGAGAGGCCAG	TCCCTGAGGACTGAAGAATAGCA
ZNF365	7	nonsense	c.1065G>A	p.W355X		chr10:64382946	GCTATGTGCTATAGGTTCAAATGGT	TGTGGAACCTCTGCCTGGAA

PŘÍLOHA III:

Identifikované truncační varianty (nonsense, inzerce/delece, a ovlivňující splicing) a prioritizované missense varianty

Gen	Skupina genů	Refseq	rs number	Varianta	Kódující	Protein	RNA	HGMD	Clinvar	EXAC celý	EXAC Evropská mimo finskou	frekvence v 1000 genomes	frekvence v ESP 6500
ABCC2	6	NM_000392.4	rs72558199	nonsense	c.3196C>T	p.R1066X		DM	pathogenic	0.0004	0.0005		0.001
ABCC2	6	NM_000392.5	rs17222561	missense	c.1483A>G	p.K495E				0.004	0.0044	0.0023	0.002307
ABCC4	6	NM_005845.4		indel	c.2468dupA	p.N823Kfs*12				0.0002228	0.0003902		
ABCC4	6			nonsense	c.1150C>T	p.R384X							
ABCG2	6	NM_004827.2	rs140207606	nonsense	c.706C>T	p.R236X		FP		0.0003	0.0004	0.0005	0.000384
ABCG2	6	NM_004827.2	rs200190472	nonsense	c.736C>T	p.R246X		FP	pathogenic	0.00006596	0.00007497		0.000077
ABCG2	6			missense	c.350T>A	p.F117Y							
ATM	1	NM_000051.3		indel	c.3850delA	p.T1284Qfs*9		DM					
ATM	1		rs121434220	nonsense	c.7327C>T	p.R2443X		DM	pathogenic				
ATM	1		rs148590073	missense	c.370A>G	p.I124V		DM		0.0008	0.0001	0.0014	0.002693
ATM	1		rs56128736	missense	c.1229T>C	p.V410A		DM		0.0022	0.0033	0.0005	0.00177
ATM	1		rs2235006	missense	c.1744T>C	p.F582L		DM		0.001	0.001	0.0018	0.000615
ATM	1			missense	c.3059C>T	p.T1020I		DM		0.000008289			
ATM	1		rs1800059	missense	c.5071A>C	p.S1691R		DM		0.002	0.0031	0.0005	0.002154
ATM	1		rs11212587	missense	c.6067G>A	p.G2023R		DM?		0.0016	0.0023	0.0014	0.002308
ATM	1			missense	c.6068G>A	p.G2023E							
ATM	1		rs139770721	missense	c.6095G>A	p.R2032K		DM		0.00004124	0.00007501		0.000077
ATM	1			missense	c.7463G>A	p.C2488Y		DM		0.00001648	0.00002997		
ATM	1			missense	c.8567T>A	p.V2856E							
ATM	1			missense	c.8581A>G	p.I2861V							
ATM	1			missense	c.8734A>G	p.R2912G		DM		0.0003	0.0005		0.000231
ATR	1	NM_001184.3		nonsense	c.5342T>A	p.L1781X							
ATR	1		rs148064542	missense	c.4351C>T	p.R1451W				0.0003	0.0004	0.0005	0.000384
ATRIP	1	NM_130384.2		indel	c.827_828delAG	p.E276Gfs*2							
BLM	1	NM_000057.2	rs200389141	nonsense	c.1642C>T	p.Q548X		DM		0.0002	0.0003		0.000154
BLM	1			missense	c.2254C>A	p.Q752K		DM					
BRIP1	1	NM_032043.2	rs137852986	nonsense	c.2392C>T	p.R798X		DM	pathogenic	0.0002	0.0002		0.000154
BRIP1	1		rs28903098	missense	c.139C>G	p.P47A		DM	pathogenic	0.0002	0.0004		0.000231

EXO1	1	NM_130398.3		indel	c.1522dupT	p.C508Lfs*7							
EXO1	1			indel	c.2358delG	p.L787Yfs*37							
EXO1	1		rs4150000	splicing	c.2212-1G>C	p.V738_K743del	r.2212_2229del	DM		0.0019	0.0027	0.0014	0.002153
EXO1	1		rs143546023	missense	c.325G>A	p.E109K		DM		0.0002	0.0004		0.000231
EXO1	1			missense	c.1105A>C	p.S369R							
FAM175B	1	NM_032182.3		indel	c.1084delC	p.Q362Kfs*19							
FANCE	1	NM_021922.2		indel	c.929dupC	p.V311Sfs*2		DM					
FANCE	1		rs145068586	missense	c.253C>T	p.P85S				0.0017	0.0012	0.0018	0.000308
FANCL	1	NM_018062.3		indel	c.1096_1099dupATTA	p.T367Nfs*13				0.003534	0.002881		
FANCM	1	NM_020937.2		indel	c.3979_3980delCA	p.Q1327Vfs*16				8.269e-06			
FANCM	1			nonsense	c.1972C>T	p.R658X		DM		0.00006622	0.0001		0.000077
FANCM	1		rs147021911	nonsense	c.5101C>T	p.Q1701X		DM?		0.0013	0.0014	0.0018	0.000384
FANCM	1		rs144567652	nonsense	c.5791C>T	p.R1931X		DM		0.0009	0.0009	0.0014	0.000846
FBXW7	7	NM_033632.3		indel	c.310delC	p.H104Mfs*389							
GLG1	10	NM_012201.5		nonsense	c.3520C>T	p.R1174X							
GRB7	4	NM_005310.3		nonsense	c.862C>T	p.Q288X				0.00002323	0.0000411		
GRB7	4			splicing	c.801+1G>C	p.(?)	r.(spl?)						
HELQ	1	NM_133636.3	rs200992133	splicing	c.2677-1G>A	p.Q348Pfs*17	r.2677_2681del			0.0002	0.0004		0.000154
HELQ	1			missense	c.1418G>A	p.R473H				0.00002475	0.000015		
HTR3B	10	NM_006028.4		nonsense	c.871C>T	p.Q291X							
CHEK2	1	NM_007194.3		indel	c.277delT	p.W93Gfs*17							0.00008
CHEK2	1			splicing	c.444+1G>A	p.R148Vfs*6	r.[444+1_444+4in	DM		0.0001	0.0002		
CHEK2	1		rs17879961	missense	c.470T>C	p.I157T		DFP	pathogenic	0.0041	0.0048	0.0023	0.001615
CHEK2	1			missense	c.917G>C	p.G306A		DM		0.00005768	0.0001		
CHEK2	1			missense	c.980A>G	p.Y327C		DM		0.00002471	0.00002997		
CHEK2	1			missense	c.1037G>A	p.R346H				0.00001656	0.0000301		
IL13	3	NM_002188.2		splicing	c.174+2delT	p.(?)	r.(spl?)						
IL1A	3	NM_000575.3		splicing	c.319+2T>C	p.(?)	r.(spl?)						
IL8	3	NM_000584.3	rs188378669	nonsense	c.91G>T	p.E31X				0.0012	0.0016	0.0005	0.000692
ITGB4	10	NM_000213.3		indel	c.665delG	p.G222Efs*60							
ITGB4	10		rs150668075	missense	c.2833C>T	p.R945W				0.0002	0.00003037	0	0.000384
LRIG1	4	NM_015541.2		indel	c.3149_3150delCG	p.A1050Gfs*17				0.001063	0.001529		
LRIG1	4		rs138265012	missense	c.698G>A	p.G233E				0.0003	0.0003	0.0005	0.000308

MAP3K1	4	NM_005921.1		indel	c.4151dupT	p.L1384Lfs*36							
MAP3K1	4			missense	c.1510G>A	p.E504K				0.00004976	0.00009002		
MAP3K1	4			missense	c.3242T>A	p.M1081K				0.00002489	0.00003001		
MGMT	1	NM_002412.3		indel	c.207_210dupACGT	p.S70Yfs*5							
MMP1	8	NM_002421.3	rs139018071	splicing	c.105+2T>C	p.Q35Vfs*11	r.105_106ins105+	FTV		0.0012	0.0017	0.0023	0.001999
MMP1	8		rs143788145	missense	c.248C>A	p.T83N				0.0022	0.0034	0.0005	0.002384
MMP12	8	NM_002426.4		nonsense	c.327C>T	p.W109X				0.000008461	0		
MPL	3	NM_005373.2	rs146249964	splicing	c.79+2T>A	p.(?)	r.(spl?)	DM		0.0003	0.0006		0.000308
MPO	3	NM_000250.1	rs35897051	splicing	c.2031-2A>C	p.R677Wfs*73	r.[2029-109_2029	DM		0.0043	0.0071	0.0014	0.005228
MPO	3		rs119468010	missense	c.1705C>T	p.R569W		DM	pathogenic	0.0015	0.0026		0.00223
MPO	3			missense	c.1900T>A	p.Y634N							
MSH5	1	NM_172165.3	rs147515280	nonsense	c.541C>T	p.R181X				0.0001	0.0002		0.000231
MSH5	1	NM_172165.3		nonsense	c.1900C>T	p.R634X							
MSH5	1		rs61748589	missense	c.1808A>G	p.K603R				0.006	0.0094	0.0023	0.006305
MSR1	10	NM_138715.2		indel	c.569delT	p.L190Cfs*5							
MSR1	10		rs138749399	missense	c.1322A>G	p.H441R		DM?		0.0013	0.002	0.0018	0.001538
NAT1	5	NM_001160179	rs5030839	nonsense	c.559C>T	p.R187X		FP		0.0027	0.0038	0.0027	0.002768
NAT1	5		rs56172717	missense	c.752A>T	p.D251V		FP		0.0024	0.0033	0.0018	0.002384
NBN	1	NM_002485.4		indel	c.657_661delACAAA	p.K219Nfs*16				0.0001928	0.0003215		
NBN	1		rs61753720	missense	c.283G>A	p.D95N		DM		0.0019	0.003	0.0018	0.00223
NBN	1		rs61754966	missense	c.511A>G	p.I171V		DM	pathogenic	0.0014	0.002	0.0009	0.00123
NBN	1		rs34767364	missense	c.643C>T	p.R215W		DM	pathogenic	0.003	0.0047	0.0023	0.002463
NF1	4	NM_001042492.2		indel	c.5690delG	p.G1897Vfs*28							
NF1	4			missense	c.2739A>G	p.I913M							
NF1	4			missense	c.7354C>T	p.R2452C				0.000008238	0.00001499		0.000231
NQO2	2	NM_000904.4		nonsense	c.628C>T	p.Q210X				0.000008277	0.00001506		
PIK3CG	4	NM_002649.3		indel	c.41_42delAG	p.E14Gfs*147				5.38e-05	8.003e-05		
PPP2R1B	4	NM_181699.2		indel	c.342_343delTG	p.V115Cfs*3				0.0007002	0.001226		
PREX2	4	NM_024870.2		indel	c.3210_3213delAACA	p.D1072Vfs*17							
PREX2	4		rs11784582	missense	c.934G>A	p.D312N				0.0016	0.0026	0.0014	0.001768
PREX2	4			missense	c.4730C>T	p.A1577V				0.000008275			
PTPRJ	10	NM_002843.3		nonsense	c.1191T>A	p.Y397X							
PTPRJ	1		rs140874449	missense	c.3254C>T	p.S1085L				0.00007428	0.00004503		0.000154

RAD18	1	NM_020165.3		indel	c.1430_1431insGCGG	p.T478Rfs*6							
RAD50	1	NM_005732.3		nonsense	c.1093C>T	p.R365X							
RAD50	1		rs28903088	missense	c.671G>A	p.R224H		DM?		0.0006	0.0008		0.000615
RAD50	1		rs61749630	missense	c.2397G>C	p.Q799H				0.0003	0.0005		0.000384
RAD50	1			missense	c.3487A>T	p.I1163L							
RAD51C	1	NM_058216.1		nonsense	c.502A>T	p.R168X							
RAD51C	1			splicing	c.905-2_1delAG	p.L301Gfs*42	r.905_965del	DM					
RAD51C	1			missense	c.335G>C	p.G112A							0.000077
RAD51D	1	NM_002878.3		indel	c.355_358deldeITGTA	p.C119Wfs*16							
RASSF1	7	NM_170714.1		splicing	c.888+1G>A	p.V258Gfs*7	r.758_888del			0.000009005	0		
RFC1	1	NM_002913.4		indel	c.2191delA	p.R731Gfs*7							
RNASEL	3	NM_021133.3	rs74315364	nonsense	c.793G>T	p.E265X		DM	pathogenic	0.004	0.0058	0.0023	0.003229
RNASEL	3		rs114166108	missense	c.184C>T	p.P62S				0.00005812	0.0001	0.0005	0.000231
RNASEL	3			missense	c.875T>C	p.L292P				0.00001655	0.00003011		
ROPN1L	10	NM_031916.4	rs41280363	nonsense	c.135T>A	p.Y45X				0.0012	0.0019	0.0009	0.001384
S100A3	10	NM_002960.1		indel	c.208delG	p.V70Wfs*83				0.00271	0.004362		
STXBP4	10	NM_178509.5		splicing	c.181-1G>A	p.K60Vfs*28	r.181_287del			0.00002489	0.00004519		0.000077
TCL1A	10	NM_021966.2		nonsense	c.253C>T	p.R85X							
VEGFA	4	NM_001025366	rs149528656	splicing	c.1085+2T>C	p.(?)	r.(spl?)			0.0001	0.0002		0.000308
WRN	1	NM_000553.4		nonsense	c.604A>T	p.K202X							
WRN	1	NM_000553.4	rs11574410	nonsense	c.4216C>T	p.R1406X				0.0035	0.0013	0.0018	0.000769
WRN	1		rs150148567	missense	c.1717A>G	p.T573A				0.001	0.0016	0.0014	0.00208
XRCC4	1	NM_022406.2		indel	c.25delC	p.H9Tfs*8				0.0004377	0.0006303		
ZNF365	7	NM_199451.2	rs142406094	nonsense	c.1065G>A	p.W355X				0.00003295	0.00005993		0.000231

PŘÍLOHA IV:

Seznam publikací souvisejících s předkládanou prací:

1. 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 Dec 30;637:41-49. doi: 10.1016/j.gene.2017.09.025. Epub 2017 Sep 14. (v příloze IV-1)
IF=2.4
2. Zemankova P, **Lhota F**, Kleiblova P, Soukupova J, Vocka M, Janatova M, Kleibl Z.; RE: frameshift variant FANCL*c.1096_1099dupATTA is not associated with high breast cancer risk.; *Clin Genet*. 2016 Oct;90(4):387-9. doi: 10.1111/cge.12842. (v příloze IV-2)
IF= 3.9
3. Rump A, Benet-Pages A, Schubert S, Kuhlmann JD, Janavičius R, Macháčková E, Foretová L, Kleibl Z, **Lhota F**, Zemankova P, Betscheva-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 Aug 9;12(8):e1006248. doi: 10.1371/journal.pgen.1006248. eColl 2016 Aug 9. (v příloze IV-3)
IF=6.7
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 the NBN gene predisposes to pancreatic cancer.; *Gene*. 2016 Aug 10;587(2):169-72. doi: 10.1016/j.gene.2016.04.056. Epub 2016 May (v příloze IV-4)
IF=2.3
5. **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 Oct;90(4):324-33. doi: 10.1111/cge.12748. Epub 2016 Mar 4. (v příloze IV-5)
IF=3.9

Seznam ostatních publikací s IF:

1. 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*. 2018 Apr 12;13(4):e0195761. doi: 10.1371/journal.pone.0195761. eCollection 2018.
IF=3.5
2. Havranek O, Kleiblova P, Hojny J, **Lhota F**, Soucek P, Trneny M, Kleibl Z.; Association of Germline CHEK2 Gene Variants with Risk and Prognosis of Non-Hodgkin Lymphoma.; *PLoS One*. 2015 Oct 27;10(10):e0140819. doi: 10.1371/journal.pone.0140819. eColl. 2015 Oct 27.
IF=3.1

3. Sevcik J, Falk M, Macurek L, Kleiblova P, **Lhota F**, Hojny J, Stefancikova L, Janatova M, Bartek J, Stribrna J, Hodny Z, Jezkova L, Pohlreich P, Kleibl Z.; Expression of human BRCA1 Δ 17-19 alternative splicing variant with a truncated BRCT domain in MCF-7 cells results in impaired assembly of DNA repair complexes and aberrant DNA damage response.; Cell Signal. 2013 May;25(5):1186-93. doi: 10.1016/j.cellsig.2013.02.008. Epub 2013 Feb 14.
IF=4.2
4. Sevcik J, Falk M, Kleiblova P, **Lhota F**, Stefancikova L, Janatova M, Weiterova L, Lukasova E, Kozubek S, Pohlreich P, Kleibl Z.; The BRCA1 alternative splicing variant Δ 14-15 with an in-frame deletion of part of the regulatory serine-containing domain (SCD) impairs the DNA repair capacity in MCF-7 cells.; Cell Signal. 2012 May;24(5):1023-30. doi: 10.1016/j.cellsig.2011.12.023. Epub 2012 Jan 3.
IF=4.2
5. Kantola AK, Ryyänänen MJ, **Lhota F**, Keski-Oja J, Koli K.; Independent regulation of short and long forms of latent TGF-beta binding protein (LTBP)-4 in cultured fibroblasts and human tissues.; J Cell Physiol. 2010 Jun;223(3):727-36. doi: 10.1002/jcp.22082.
IF=4.1

Ostatní publikace bez IF:

Přednáška: Identifikace dědičných alterací predisponujících ke vzniku karcinomu prsu pomocí „next-gen“ sekvenování.

Filip **Lhota**, Petra Boudová, Petra Kleiblová, Jana Soukupová, Viktor Stránecký, Hana Hartmanová, Kateřina Hodaňová, Zdeněk Kleibl

Sborník: 15. Studentská vědecká konference 1.LF UK v Praze; 2014; Galén; ISBN 978-80-7492-150-6

Poster: Identification and quantification of BRCA1 splicing variants

Filip **Lhota**, Jan Hojny, Petra Kleiblova, Jan Sevcik, Jana Soukupova, Marketa Janatova, Petra Boudova, Marianna Borecka, P. Pohlreich, Z. Kleibl

Eur J Cancer. 2014; 50, Suppl. 5; EACR-23, Scientific Programme

Poster: Targeted Next-Gen Sequencing in High-Risk BRCA1-and BRCA2-Negative Breast Cancer Patients

Filip **Lhota**, Viktor Stranecky, Petra Boudova, Jana Soukupova, Katerina Hodanova, Hana Hartmannova, Marketa Janatova, Petra Kleiblova, Petr Pohlreich, Stanislav Kmoch, Zdenek Kleibl

Curr Oncol. 2014 Apr; 21(2): e358–e391. doi: 10.3747/co.21.2077

Překlad učebnice: Harperova ilustrovaná biochemie

Robert K. Murray, David A. Bender, Kathleen M. Botham, Peter J. Kennelly, Victor W. Rodwell, P. Anthony Weil

Kapitoly: 40: Membrány: Struktura a funkce; 41: Rozmanitost endokrinního systému; 42: Působení hormonů a přenos signálu

Galén 2013; EAN: 9788072629077

Prezentace: Identification and characterization of BRCA1 alternative splicing variants in breast cancer cell line.

Filip **Lhota**, Jan Hojny, Petra Kleiblova, Jan Sevcik and Zdenek Kleibl

Biomania Student Scientific Meeting 2013. 2013. ISBN 978-80-210-7933-5

Poster: USER-based Approach for Identification of BRCA1 Alternative Splicing Variants

Filip **Lhota**, Petra Kleiblova, Jan Hojny, Jan Sevcik, Marketa Janatova, Petr Pohlreich, Zdenek Kleibl

Eur J Cancer. 2013; 48, Suppl. 5; EACR-22, Scientific Programme

Přednáška: Identifikace alternativních sestřihových variant genu BRCA1 v buněčné linii MCF-7

Filip **Lhota**, Jan Hojný, Jan Ševčík, Petra Kleiblová

Sborník: 13. Studentská vědecká konference 1.LF UK v Praze; 2012; Galén; ISBN 978-80-7262-943-5

PŘÍLOHA IV-1

Lhota *et al.*; Hereditary truncating mutations of DNA repair and other genes in *BRCA1/BRCA2/PALB2*-negatively tested breast cancer patients



Original Article

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. Hereditary truncating mutations of DNA repair and other genes in *BRCA1/BRCA2/PALB2*-negatively tested breast cancer patients. Clin Genet 2016; 90: 324–333. © John Wiley & Sons A/S. Published by John Wiley & Sons Ltd, 2016

Hereditary breast cancer comprises a minor but clinically meaningful breast cancer (BC) subgroup. Mutations in the major BC-susceptibility genes are important prognostic and predictive markers; however, their carriers represent only 25% of high-risk BC patients. To further characterize variants influencing BC risk, we performed SOLiD sequencing of 581 genes in 325 BC patients (negatively tested in previous *BRCA1/BRCA2/PALB2* analyses). In 105 (32%) patients, we identified and confirmed 127 truncating variants (89 unique; nonsense, frameshift indels, and splice site), 19 patients harbored more than one truncation. Forty-six (36 unique) truncating variants in 25 DNA repair genes were found in 41 (12%) patients, including 16 variants in the Fanconi anemia (FA) genes. The most frequent variant in FA genes was c.1096_1099dupATTA in *FANCL* that also show a borderline association with increased BC risk in subsequent analysis of enlarged groups of BC patients and controls. Another 81 (53 unique) truncating variants were identified in 48 non-DNA repair genes in 74 patients (23%) including 16 patients carrying variants in genes coding proteins of estrogen metabolism/signaling. Our results highlight the importance of mutations in the FA genes' family, and indicate that estrogen metabolism genes may reveal a novel candidate genetic component for BC susceptibility.

Conflict of interest

All authors declare no conflict of interest.

**F. Lhota^{a,†}, P. Zemankova^{a,†},
P. Kleiblova^{a,b}, J. Soukupova^a,
M. Vocka^c, V. Stranecky^d,
M. Janatova^a, H.
Hartmannova^d, K. Hodanova^d,
S. Kmoch^d
and Z. Kleibl^a**

^aInstitute of Biochemistry and Experimental Oncology, First Faculty of Medicine, Charles University in Prague, Prague, Czech Republic, ^bInstitute of Biology and Medical Genetics, First Faculty of Medicine, Charles University in Prague and General University Hospital in Prague, Prague, Czech Republic, ^cDepartment of Oncology, First Faculty of Medicine, Charles University in Prague and General University Hospital in Prague, Prague, Czech Republic, and ^dInstitute of Inherited Metabolic Disorders, First Faculty of Medicine, Charles University in Prague and General University Hospital in Prague, Prague, Czech Republic

[†]These authors contributed equally to this work.

Key words: DNA repair – *FANCL* – genetic predisposition testing – hereditary breast cancer – high-throughput nucleotide sequencing – sequencing panel – SOLiD next-generation sequencing

Corresponding author: Zdenek Kleibl, MD, PhD, Institute of Biochemistry and Experimental Oncology, First Faculty of Medicine, Charles University in Prague, U Nemocnice 5, 128 53 Prague 2, Czech Republic.
Tel.: +420 22 496 5745;
fax.: +420 22 496 5732;
e-mail: zdekleje@lf1.cuni.cz

Received 12 December 2015, revised and accepted for publication 25 January 2016

Hereditary truncating mutations of DNA repair and other genes

Breast cancer (BC; OMIM#114480) emerges as a leading cause of cancer death in female population worldwide. Hereditary breast cancer (HBC) accounts approximately for 5–10% of cases. Clinical importance of HBC results from the high lifetime risk of BC development, increased risk of other associated cancers, early disease onset, and 50% probability of the mutant allele's transmission to the offspring (1). Hence, the identification of germline mutations that confer BC susceptibility is an important task of clinical oncogenetics with considerable clinical utility, including tailored healthcare focused on early cancer identification, preventive surgical strategies decreasing cancer risk, and specific therapeutic strategies (2). The most frequently mutated genes in HBC patients are *BRCA1* and *BRCA2*; however, mutations in these genes account for less than 25% of cases in HBC patients. Since the identification of major BC-susceptibility genes, numerous other predisposition genes have been identified. Their characterization has been strongly accelerated with the availability of next-generation sequencing (NGS) technologies (3). Mutational analyses of recently identified BC-susceptibility genes indicate that frequencies of their mutations are substantially lower than that in *BRCA1* and *BRCA2*, besides being highly variable among populations worldwide (4). However, mutations in these newly established BC-susceptibility genes could collectively epitomize another 25% of ascertained genetic risk in HBC patients and thus their analyses are gradually introduced into the clinical analyses (5).

A striking characteristic of the majority of known BC-susceptibility genes is the contribution of their protein products in DNA damage repair (6). On the other hand, the existence of known BC-susceptibility genes that code for proteins not directly involved in these processes (e.g. *PTEN*, *CDH1*, or *NF1*) indicates that non-DNA repair genes could also contribute to BC susceptibility (7).

Our previous gene-by-gene mutational analyses revealed that the most frequent mutations in Czech HBC patients are found in the *BRCA1* gene (8–10). Less frequently, we identified pathogenic variants in *BRCA2* or *PALB2* (11). In this study, we aimed to describe the presence of potentially pathogenic hereditary variants in other known BC-susceptibility genes using the targeted NGS and to identify further variants that may contribute to BC susceptibility in high-risk Czech BC patients.

Materials and methods

Detailed methods are available in Supporting information methods.

Patients and samples

The 325 successfully sequenced patients' samples were selected from a sample collection of high-risk Czech BC patients that fulfilled testing criteria described previously (8, 9, 11), were negatively tested for the presence of mutations in *BRCA1/BRCA2/PALB2*, and gave their

informed consent approved by local ethical committee. As controls, we analyzed 105 samples obtained from Czech non-cancer elder females selected according to their age (>50 years; median age 71 years; ranged 54–95 years) from non-cancer controls described previously (12). The genotyping of the c.1096_1099dupATTA variant in *FANCL* was performed on additional sample sets of 337 high-risk BC patients, 673 sporadic BC patients and 686 non-cancer controls (13, 14) using high-resolution melting analysis and confirmed by Sanger sequencing (Fig. S1, Supporting information). Clinical and histopathological characteristics of analyzed high-risk BC patients are available in Tables S1 and S2.

Sequencing gene panel

The targeted genes comprised two groups consisting of 141 DNA repair genes and 449 genes retrieved from Phenopedia database (15) using the disease term 'breast neoplasms' with at least two entries (assessed February 2012). Finally, 581 targeted genes (listed in Table S3) were sequenced successfully.

Library construction, sequence capture and sequencing

Fragmented DNA was subjected to ligation of SOLiD sequencing adaptors and polymerase chain reaction (PCR)-based incorporation of bar codes, as described previously (16). The target DNA enrichment was performed by a custom SeqCap EZ Choice Library (Roche), and SOLiD sequencing primers were introduced by PCR. The final libraries were amplified by an emulsion PCR and sequenced on a SOLiD 4 System (Thermo Fisher, Waltham, MA, USA).

Bioinformatics pipeline, variant filtration, and prioritization of missense variants

Sequencing reads were aligned to the human genome reference (GRCh37/hg19) using Novoalign (CS1.01.08). Picard was used for duplicate removal and SAMtools (0.1.8) for SAM-to-BAM conversion and calling of single nucleotide variants (SNVs) and small insertions and deletions (indels). Variant annotation was performed with ANNOVAR (17).

Variant filtration excluded off-target sequences and low confidence variants (sequence quality <150; sequencing coverage <10). We also excluded common variants with allelic frequencies in ESP6500 and 1000 Genomes databases >0.01. To reflect the population-specific variants and variants influencing cancer susceptibility, we excluded variants presented in no patient or in more than two controls.

To identify missense variants with a putative contribution to BC susceptibility, we performed a prioritization that considered five prediction algorithms (SIFT, PolyPhen-2, LRT, MutationTaster, and PhyloP) and two databases (ClinVar and HGMD) aggregating data about genotypes and corresponding clinical characteristics. Prioritized variants were considered those that were

called by each prediction software as deleterious (or unknown) or considered as disease-associated in ClinVar and HGMD databases.

Confirmation of truncating variants

All truncating variants were confirmed by conventional Sanger sequencing. The variants affecting a conservative splice site were analyzed from the blood-isolated patient's RNA, when available, using RT-PCR and sequencing as described previously (18). All primers are listed in Table S4.

Statistical analysis

The differences among analyzed groups and subgroups were calculated by the chi-square test or Fischer exact test if the expected number of events was lower than six.

Results

In the set of 325 patients' samples and 105 controls, we obtained 491,385 variants in exome and adjacent intronic sequences of 581 targeted genes. The mean sequencing coverage was 56.5 and 93% of the captured sequence was covered by >10 reads. Using the variant filtration, we identified 4540 rare variants in the final dataset representing 2647 unique variants of 496/581 targeted genes (85.4%). We found 144 truncating variants (either nonsense, frameshift indels, or splice site alterations), representing 89 unique variants, in 73/581 targeted genes (12.6%).

The set of 325 BC patients harbored 4053 rare variants (2647 unique) including 127 truncating variants (89 unique), 34 in-frame indels (22 unique), 2347 missense SNVs (1599 unique), and 1545 synonymous SNVs (937 unique). We primarily focused on the truncating variants that were identified in 105/325 (32.3%) BC patients (Fig. 1) and were all confirmed by Sanger sequencing. Nineteen patients carried more than one truncating variant (1 patient carried four, 1 patient carried three, and 17 patients carried two truncations), 86 patients carried one truncating variant. The group of truncating variants included 20 splicing variants (14 unique, each affecting one particular gene) flanking to intronic (± 2 bp) sequences. Their impact on splicing was studied at the mRNA level (available from eight patients). Seven out of eight analyzed splicing variants showed frameshift (Figs S2 and S3). The prioritization analysis revealed 356 potentially pathogenic variants out of 1599 rare unique missense variants (22%).

Hereditary variants in DNA repair genes

In 25 DNA repair genes, we identified 46/127 truncations (36/89 unique) in 41 (12.6%) BC patients (Table 1). The most frequent alterations affected genes that code for DNA double-strand break (DDSB)/interstrand crosslink (ICL) repair proteins. These included 16 patients carrying nine unique truncating variants in five Fanconi anemia (FA) genes (Fig. 1).

Another 19 patients harbored 19 unique variants affecting other genes coding for proteins involved in the DDSB repair pathways, including homologous recombination (HR; *ATM*, *EXO1*, *WRN*, *BLM*, *DCLRE1C*, *FAM175B/ABRO1*, *HELQ*, *NBN*, *RAD18*, *RAD50*, *RAD51D*, *CHEK2*, and *RFC1*) but also non-homologous end joining (NHEJ; *XRCC4*) repair. Finally, eight truncating variants, each in one patient, were identified in the genes that code for proteins involved in other DNA repair processes including single-strand DNA repair (*ATR*, *ATRIP*), nucleotide excision repair (NER; *ERCC2*, *ERCC6*), mismatch repair (MMR; *MSH5*), and direct removal of alkylated guanine (*MGMT*). Two patients carried truncating variants in more than one gene involved in different DNA repair pathways.

Altogether, 106 unique prioritized missense variants in 59 DNA repair genes were identified in 133 patients (34 of these variants, in 56 patients, were found in 15 genes in which at least 1 truncating variant was also detected; Table S5). The most frequent potentially pathogenic missense variants were found in *ATM* (12 variants in 17 patients) and *CHEK2* (4 variants in 13 patients). Among prioritized variants, we also identified pathogenic missense variants in *BRCA1* (c.115T>C; p.C39R), *TP53* (c.733G>A; p.G245S), and *CDH1* (c.1018A>G; p.T340A) in three young BC patients.

Extended analysis of *FANCL* c.1096_1099dupATTA

The most frequent frameshift variant found in six BC patients and none NGS control was c.1096_1099dupATTA (p.T367Nfs*13) in *FANCL* (previously described in an FA patient belonging to the FA-L complementation group; OMIM#614083) (19). Because of the insufficient number of NGS controls, we first compared the frequency of this *FANCL* variant among our patients with data from the Exome Aggregation Consortium (ExAC) database (<http://exac.broadinstitute.org>; accessed May 2015) indicating an overrepresentation of this variant among our high-risk BC patients (Table 1). Therefore, we further analyzed another 337 high-risk BC Czech patients (329 females, 8 males; all *BRCA1/BRCA2/PALB2*-negative). Among these, we identified another three c.1096_1099dupATTA carriers with BC (all diagnosed before age of 38 years). Overall, the c.1096_1099dupATTA was identified in 9/662 high-risk BC individuals (1.3%).

To identify the carriers of c.1096_1099dupATTA in sporadic BC patients and other controls, we genotyped 673 unselected BC cases and 686 non-cancer controls (313 females and 373 males). This analysis revealed three carriers in each analyzed group, showing its frequency as 0.4% in both BC cases (3/693) and controls (3/791; including 105 NGS controls and 686 genotyped controls), respectively. Thus, the frequency of c.1096_1099dupATTA was significantly (Fisher exact test) overrepresented only among high-risk individuals ($p = 0.04$) but not in sporadic BC patients ($p = 0.9$). All 14 carriers among patients were females, while all three carriers in controls were males.

Hereditary truncating mutations of DNA repair and other genes

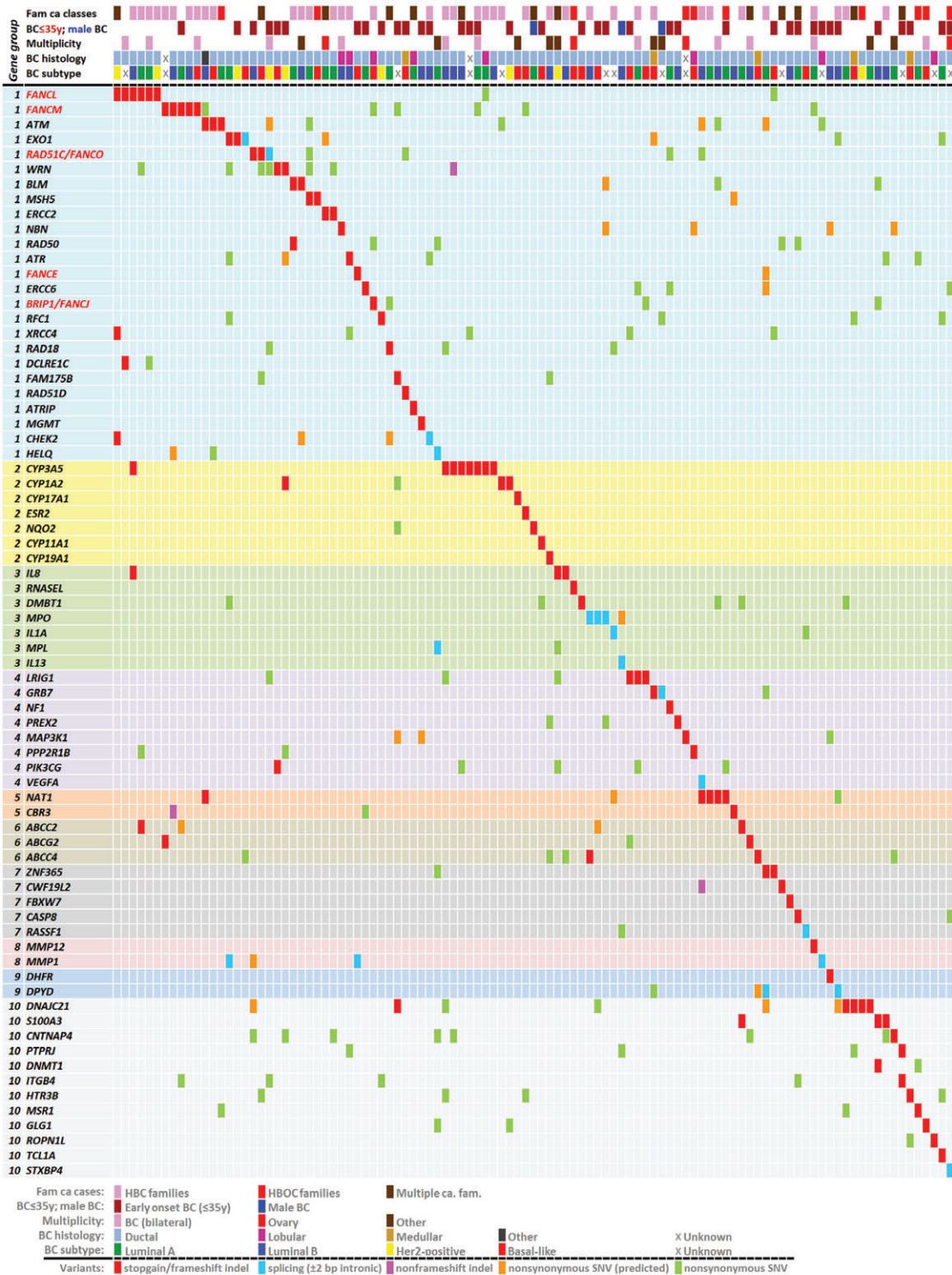


Fig. 1. Overview of variants in 73 genes (rows) affected by at least one truncating (nonsense, frameshift, or splicing) variant, that were identified in 105 BC patients (columns). Pathological characteristics of BC tumors (histology and subtypes) and selected clinical characteristics (BC in females at the age of <35 years or male BC, and the presence of familial cancer) are shown in five upper lines (color markings are displayed in Fig. 2; X denotes a missing information). The patients and genes are ordered according to the overall number of found variants, genes (with the Fanconia anemia gene members highlighted in red letters) are grouped by functional relationship of coded proteins (see note). Note: Genes in gene groups (1–10; number (N) patients with at least one truncating variant) were ascertained as follows: the genes coding proteins involved in DNA repair (1; N=41); steroid hormones synthesis, turnover or signaling (2; N=16); immune response (3; N=11); membrane receptor signaling (4; N=11); metabolism of xenobiotics (5; N=6); membrane transport of molecules (6; N=6); cell cycle/apoptosis regulation (7; N=6); cell-to-cell communication (8; N=4); nucleotide metabolism (9; N=3); or other (unsorted) processes (10; N=16). Color markings used for pathological and clinical characteristics (shown in legend) are identical to that presented in Fig. 2. Clinical and histopathological characteristics of truncating mutation carriers are shown in Table S6.

Table 1. List of 36 unique truncating variants (nonsense, frameshift indels, or splicing) that were found in 25 genes coding for proteins involved in DNA repair and DNA damage response identified in 41/325 BC patients (Pts) and 7/105 non-cancer controls (Ctrls)^a

Gene	HGVS coding	HGVS protein ^b	Classification	Rs number	HGMD/ ClinVar			ExAC (mut/all) ^{b, c}
					Pts (N)	Ctrls (N)		
<i>FANCL</i>	c.1096_1099dupATTA	p.T367Nfs*13	Indel			6	0	232/65648*
<i>FANCM</i>	c.1972C>T	p.R658*	Nonsense		DM	1	1	7/66502*
	c.3979_3980delCA	p.Q1327Vfs*16	Indel			1	0	0/66498*
	c.5101C>T	p.Q1701*	Nonsense	rs147021911	DM?	2	0	95/66562
	c.5791C>T	p.R1931*	Nonsense	rs144567652	DM	1	0	63/66622
<i>ATM</i>	c.3850delA	p.T1284Qfs*9	Indel		DM	1	0	n.r.
	c.7327C>T	p.R2443*	Nonsense	rs121434220	DM/P	2	0	n.r.
<i>EXO1</i>	c.1522dupT	p.C508Lfs*7	Indel			1	0	n.r.
	c.2358delG	p.L787Yfs*37	Indel			1	0	n.r.
	c.2212-1G>C	p.V738_K743del	Splicing	rs4150000	DM	1	1	172/63478
<i>CHEK2</i>	c.277delT	p.W93Gfs*17	Indel			1	0	n.r.
	c.444+1G>A	p.R148Vfs*6	Splicing		DM	2	0	11/66720*
<i>RAD51C</i>	c.502A>T	p.R168*	Nonsense			2	0	n.r.
	c.905-2_1delAG	p.L301Gfs*42	Splicing		DM	1	0	n.r.
<i>BLM</i>	c.1642C>T	p.Q548*	Nonsense	rs200389141	DM	2	0	21/66322*
<i>ERCC2</i>	c.230_231delTG	p.V77Afs*4	Indel			1	0	n.r.
	c.1703_1704delTT	p.F568Yfs*2	Indel		DM	1	2	11/65444*
<i>MSH5</i>	c.541C>T	p.R181*	Nonsense	rs147515280		1	0	13/65882*
	c.1900C>T	p.R634*	Nonsense			1	0	n.r.
<i>WRN</i>	c.604A>T	p.K202*	Nonsense			1	0	n.r.
	c.4216C>T	p.R1406*	Nonsense	rs11574410		1	0	87/65788
<i>ATR</i>	c.5342T>A	p.L1781*	Nonsense			1	0	n.r.
<i>ATRIP</i>	c.827_828delAG	p.E276Gfs*2	Indel			1	0	n.r.
<i>BRIP1</i>	c.2392C>T	p.R798*	Nonsense	rs137852986	DM/P	1	0	16/65688*
<i>DCLRE1C</i>	c.1903dupA	p.S635Kfs*6	Indel			1	0	27/66734*
<i>ERCC6</i>	c.3693C>G	p.Y1231*	Nonsense			1	0	n.r.
<i>FAM175B</i>	c.1084delC	p.Q362Kfs*19	Indel			1	0	n.r.
<i>FANCE</i>	c.929dupC	p.V311Sfs*2	Indel		DM	1	1	n.r.
<i>HELQ</i>	c.2677-1G>A	p.Q348Pfs*17	Splicing	rs200992133		1	1	27/66528
<i>MGMT</i>	c.207_210dupACGT	p.S70Yfs*5	Indel			1	0	n.r.
<i>NBN</i>	c.657_661delACAAA	p.K219Nfs*16	Indel			1	1	21/65324
<i>RAD18</i>	c.1430_1431insGCGG	p.T478Rfs*6	Indel			1	0	n.r.
<i>RAD50</i>	c.1093C>T	p.R365*	Nonsense			1	0	n.r.
<i>RAD51D</i>	c.355_358delTGTGTA	p.C119Wfs*16	Indel			1	0	n.r.
<i>RFC1</i>	c.2191delA	p.R731Gfs*7	Indel			1	0	n.r.
<i>XRCC4</i>	c.25delC	p.H9Tfs*8	Indel			1	0	42/66632
Total variants						46	7	

Variants listed in HGMD or ClinVar databases: DM, disease-causing (pathological) mutations; DM?, likely disease-causing (likely pathological) mutation; P, pathogenic. ExAC, Exome Aggregation Consortium; n.r., variant not reported in ExAC.

^aThe enhanced version of the table (including missense variant predicted as pathogenic, numbers of reference transcripts, and frequencies in ExAC, ESP6500, and 1000 genomes databases) is available as Table S5.

^bExAC allelic frequency in European non-Finnish population (mutated alleles/wt alleles).

^cAsterisk (*) indicates significant differences ($p < 0.05$) between allelic frequencies in European non-Finnish population (ExAC) and in analyzed population of patients (Fisher exact test).

Hereditary variants in non-DNA repair genes

The remaining 81/127 truncations (53/89 unique) in 48 non-DNA repair genes were identified in 74 (22.8%) BC patients (Table 2). We found variants in only non-DNA repair genes in 64 of them, while in 10 patients we also detected some truncating variants in DNA repair genes. To identify possible defects in pathways that may contribute to BC susceptibility, we sorted the affected genes into nine groups (Group 2–9 in Table 2 and Fig. 1) clustering functionally related proteins. Twelve genes (Group 10) comprised proteins with unrelated or unclear function. Sixteen carriers (5% of all patients) of eight different truncating variants have been identified in

the ‘Group 2’ associating genes that code for proteins involved in steroid hormones metabolism or signaling.

Further, we detected 250 unique, prioritized, potentially pathogenic missense variants in 150 genes in 213 patients. The most frequent prioritized SNVs in non-DNA repair genes affected the *APC* gene (in eight patients).

Individual and disease characteristics in carriers of truncating variants

We found no significant differences in the characteristics of patients and tumors between the carriers of truncating

Hereditary truncating mutations of DNA repair and other genes

Table 2. List of 53 unique truncating variants (nonsense, frameshift indels, or splicing) in 48 non-DNA repair genes identified in 74/325 BC patients and 10/105 non-cancer controls^a

Gene	Gr	HGVS coding	HGVS protein ^b	Classification	Rs number	HGMD/ClinVar	Pts (N)	Ctrs (N)	ExAC (mut/all) ^{b, c}
<i>CYP3A5</i>	2	c.92dupG	p.L32Tfs*3	Indel			7	1	732/66688
	2	c.246dupG	p.A83Gfs*40	Indel			1	0	21/66728
<i>CYP1A2</i>	2	c.816T>A	p.Y272*	Nonsense	rs140421378	FTV	3	0	16/65958*
<i>CYP11A1</i>	2	c.835delA	p.I279Yfs*10	Indel		DM	1	0	2/66694*
<i>CYP17A1</i>	2	c.1072C>T	p.R358*	Nonsense		DM	1	0	n.r.
<i>CYP19A1</i>	2	c.1058dupT	p.L353Ffs*10	Indel			1	0	1/60606*
<i>ESR2</i>	2	c.76G>T	p.E26*	Nonsense			1	0	1/66734*
<i>NQO2</i>	2	c.628C>T	p.Q210*	Nonsense			1	0	1/66386*
<i>IL8</i>	3	c.91G>T	p.E31*	Nonsense	rs188378669		3	2	104/66426
<i>DMBT1</i>	3	c.2227delC	p.Q743Rfs*4	Indel			1	0	n.r.
<i>IL13</i>	3	c.174+2delTT	p.(?)	Splicing			1	0	n.r.
<i>IL1A</i>	3	c.319+2T>C	p.(?)	Splicing			1	0	n.r.
<i>MPL</i>	3	c.79+2T>A	p.(?)	Splicing	rs146249964	DM	1	0	114/66230
<i>MPO</i>	3	c.2031-2A>C	p.R677Wfs*73	Splicing	rs35897051	DM	3	1	470/66434
<i>RNASEL</i>	3	c.793G>T	p.E265*	Nonsense	rs74315364	DM/P	1	1	381/66212
<i>LRIG1</i>	4	c.3149_3150delCG	p.A1050Gfs*17	Indel			3	0	102/66704*
<i>GRB7</i>	4	c.862C>T	p.Q288*	Nonsense			1	0	2/48666*
	4	c.801+1G>C	p.(?)	Splicing			1	0	n.r.
<i>MAP3K1</i>	4	c.4151dupT	p.L1384Lfs*36	Indel			1	0	n.r.
<i>NF1</i>	4	c.5690delG	p.G1897Vfs*28	Indel			1	0	n.r.
<i>PIK3CG</i>	4	c.41_42delAG	p.E14Gfs*147	Indel			1	0	5/62474*
<i>PPP2R1B</i>	4	c.342_343delITG	p.V115Cfs*3	Indel			1	0	81/66082
<i>PREX2</i>	4	c.3210_3213delAACA	p.D1072Vfs*17	Indel			1	0	n.r.
<i>VEGFA</i>	4	c.1085+2T>C	p.(?)	Splicing	rs149528656		1	0	15/66648*
<i>NAT1</i>	5	c.559C>T	p.R187*	Nonsense	rs5030839	FP	5	1	252/66632
<i>CBR3</i>	5	c.533delA	p.D178Afs*46	Indel			1	0	102/66716
<i>ABCC2</i>	6	c.3196C>T	p.R1066*	Nonsense	rs72558199	DM/P	2	0	35/66738*
<i>ABCC4</i>	6	c.2468dupA	p.N823Kfs*12	Indel			1	1	26/66634
	6	c.1150C>T	p.R384*	Nonsense			1	0	n.r.
<i>ABCG2</i>	6	c.706C>T	p.R236*	Nonsense	rs140207606	FP	1	0	24/66634
	6	c.736C>T	p.R246*	Nonsense	rs200190472	FP/P	1	0	5/66692*
<i>ZNF365</i>	7	c.1065G>A	p.W355*	Nonsense	rs142406094		2	0	4/66740*
<i>CASP8</i>	7	c.106delG	p.E36Nfs*7	Indel			1	0	n.r.
<i>CWF19L2</i>	7	c.1605delA	p.K535Nfs*4	Indel			1	0	n.r.
<i>FBXW7</i>	7	c.310delC	p.H104Mfs*389	Indel			1	0	n.r.
<i>RASSF1</i>	7	c.888+1G>A	p.V258Gfs*7	Splicing			1	0	1/64856*
<i>MMP1</i>	8	c.105+2T>C	p.Q35Vfs*11	Splicing	rs139018071	FTV	3	0	114/66230
<i>MMP12</i>	8	c.327C>T	p.W109*	Nonsense			1	0	0/65722*
<i>DPYD</i>	9	c.1905+1G>A	p.D581_N635del	Splicing	rs3918290	DM	2	1	389/66688
<i>DHFR</i>	9	c.95delT	p.F32Sfs*7	Indel			1	0	n.r.
<i>DNAJC21</i>	10	c.1503delA	p.K501Nfs*10	Indel			3	0	33/66560*
	10	c.1629delT	p.F543Lfs*4	Indel			2	0	5/11578*
<i>S100A3</i>	10	c.208delG	p.V70Wfs*83	Indel			3	1	291/66718
<i>CNTNAP4</i>	10	c.3913G>T	p.E1305*	Nonsense			1	0	3/56224*
<i>DNMT1</i>	10	c.1035dupC	p.K346Qfs*35	Indel			1	0	n.r.
<i>GLG1</i>	10	c.3520C>T	p.R1174*	Nonsense			1	0	n.r.
<i>HTR3B</i>	10	c.871C>T	p.Q291*	Nonsense			1	0	n.r.
<i>ITGB4</i>	10	c.665delG	p.G222Efs*60	Indel			1	0	n.r.
<i>MSR1</i>	10	c.569delT	p.L190Cfs*5	Indel			1	0	n.r.
<i>PTPRJ</i>	10	c.1191T>A	p.Y397*	Nonsense			1	0	n.r.
<i>ROPN1L</i>	10	c.135T>A	p.Y45*	Nonsense	rs41280363		1	1	126/66680
<i>STXBP4</i>	10	c.181-1G>A	p.K60Vfs*28	Splicing			1	0	3/66388*
<i>TCL1A</i>	10	c.253C>T	p.R85*	Nonsense			1	0	n.r.
Total variants							81	10	

Variants listed in HGMD or ClinVar databases: DM, disease-causing (pathological) mutations; DM?, likely disease-causing (likely pathological) mutation; P, pathogenic. ExAC, Exome Aggregation Consortium; n.r., variant not reported in ExAC.

^aGenes are grouped (Gr 2–10) according to the functional relationship of coded proteins, as described in the legend of Fig. 1. The enhanced version of the table (including missense variant predicted as pathogenic, numbers of reference transcripts, and frequencies in ExAC, ESP6500, and 1000 genomes databases) is available as Table S5.

^bExAC allelic frequency in European non-Finnish population (mutated alleles/wt alleles).

^cAsterisk (*) indicates Significant differences ($p < 0.05$) between allelic frequencies in European non-Finnish population (ExAC) and in analyzed population of patients (Fisher exact test).

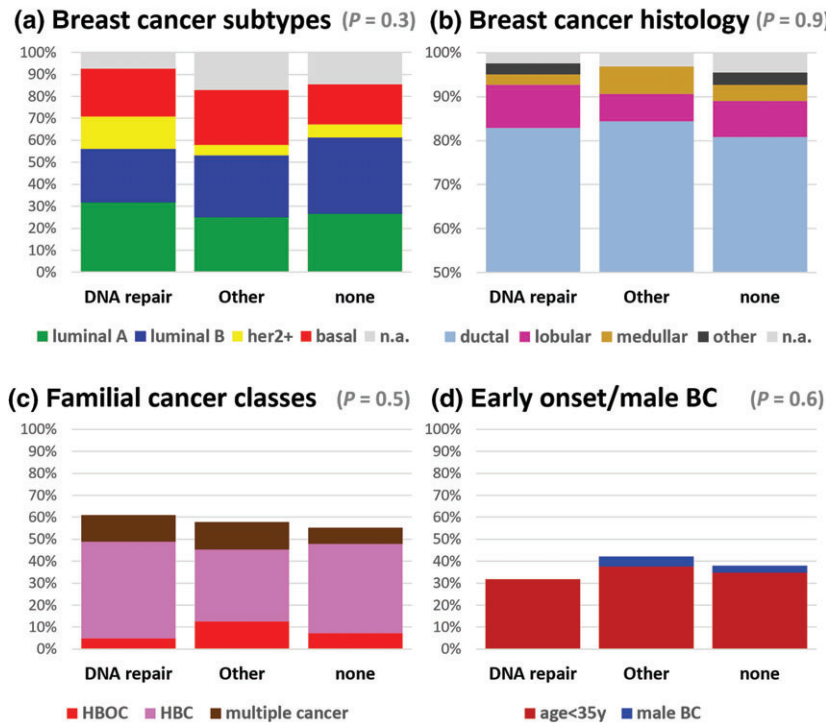


Fig. 2. Pathological characteristics of tumors and clinical characteristics of 325 analyzed BC patients grouped according to the presence of truncating variant in any DNA repair gene (DNA repair; 41 patients), variant in only other genes (other; 64 patients), and no truncating variant (none; 220 patients). The p-values (chi-square test) indicate insignificant differences in displayed characteristics among the analyzed subgroups.

variants in the DNA repair genes, carriers of truncating variants in other genes, and patients not carrying truncating variants (Fig. 2).

Discussion

Panel NGS represents a reliable approach for the analysis of cancer susceptibility in clinical settings but also in identification of candidate genes in high-risk individuals. In contrast to exome or even genome NGS, it allows the identification of the carriers of pathogenic variants in a cost-effective manner, with flexibility in the selection of gene targets, sensitivity, and manageable bioinformatics load for routine practice (20). Our analysis revealed the presence of truncating variants in nearly one third of analyzed patients and 30 patients (9%) carried truncating variants in some of 15 genes (*ATM*, *ATR*, *BLM*, *BRIP1*, *ERCC2*, *FANCE*, *FANCL*, *FANCM*, *CHEK2*, *NBN*, *NF1*, *RAD50*, *RAD51C*, *RAD51D*, *WRN*) analyzed by currently clinically used NGS panels (5). Out of 73 genes with truncating variants, in 51 genes we found only a single truncation. This indicates that rare variants could be identified in a substantial proportion of high-risk individuals; however, their clinical interpretation and differentiation from incidental findings not associating with BC susceptibility would be difficult.

Characterization of variants in DNA repair genes

The interesting result of our study is the high frequency of potentially pathogenic variants in five FA genes in

4.9% high-risk patients. FA genes code for DNA repair proteins contributing to genome stability maintenance by the ICL repair [reviewed in (21, 22)]. FA proteins form several protein–protein complexes (22). Hereditary bi-allelic mutations of FA genes are responsible for the development of FA characterized by congenital abnormalities, bone marrow failure, cellular hypersensitivity to DNA crosslinking agents and cancer susceptibility. The most frequent truncating variant was c.1096_1099dupATTA in *FANCL* that codes an ubiquitin ligase catalyzing the monoubiquitination of FANCI/FANCD2 (ID2) complex – the key step in FA pathway activation (23, 24). The c.1096_1099dupATTA variant was described by Ali et al. (19) in a patient that belonged to the FA-L complementation group. The mutated FANCL protein (p.T367Nfs*13) contains an aberrant chain of 12 amino acid residues that flanks to the PHD/RING finger domain catalyzing ubiquitin ligase activity. Ali et al. (19) performed its functional characterization revealing that the c.1096_1099dupATTA is a hypomorphic mutation resulting in the formation of altered protein with reduced binding to FA core complex and reduced FANCD2 monoubiquitination. Same variant was also identified by Akbari et al. (25) in a patient with familial esophageal squamous cell carcinoma. The results of our study, showing the overrepresentation of c.1096_1099dupATTA among high-risk BC patients, indicate that this variant may represent a novel BC-susceptibility allele. However, further studies, including segregation analyses providing information

about the association of c.1096_1099dupATTA with cancer phenotype in affected families and analyses of the variant in other populations, will be necessary to evaluate its potential clinical utility. As four out of six c.1096_1099dupATTA carriers identified by our panel NGS carried also other truncating variants (Fig. 1), we could not rule out the possibility that c.1096_1099dupATTA could act as a rather modifying variant. The recurrent mutations affecting *FANCL* and *FANCM* at their C-termini indicate that truncating variants of FA genes located in far C-terminal regions may impair the FA pathway under specific and so far uncomprehended circumstances. Such phenomenon has been proposed also for the nonsense c.9976A>T (p.K3326*) variant in *BRCA2/FANCD1* truncating the last 93 amino acids. In contrast to the majority of *BRCA2* pathogenic variants, the p.K3326* has been recognized as only a modest BC-susceptibility allele (OR = 1.26) increasing a risk of other cancers (26).

In *FANCM*, coding a helicase contributing to the formation of the FA anchor complex, we identified four truncating variants in five patients. Truncations in *FANCM* were recently associated with susceptibility for triple-negative BC (27). In three patients (none of them triple-negative), we identified previously characterized nonsense or exon skipping mutations that were shown to increase BC risk (27, 28). The remaining two *FANCM* variants included the rare nonsense mutation c.1972C>T (p.R658*; in a luminal BC patient whose mother and her sister suffered from bilateral BC) and the novel mutation c.3979_3980delCA (p.Q1327Vfs*16; in a BC patient with multiple breast and colorectal cancer (CRC) cases in the family). The association between CRC and germinal *FANCM* mutation has recently been identified in CRC tumor samples obtained from two c.5791C>T (p.R1931*) carriers (29).

We have also identified three *RAD51C/FANCO* mutation carriers (0.9% of patients). The *RAD51C* was originally identified as OC-susceptibility gene (30); however, later data conferred also increased BC susceptibility (31). Recently, we described two other pathogenic *RAD51C* variants in two OC patients (13). These data indicate that mutations in *RAD51C* may affect ~1% of Czech high-risk BC or OC patients.

Finally, the carriers of variants in FA genes comprised also two basal-like BC patients carrying pathogenic variants in *FANCE* and *BRIPI/FANCI*, respectively. Both variants were reported in association with esophageal cancer (25) and triple-negative BC (31), respectively.

We found rare truncating variants in several other genes associated with hereditary BC that code for DDSB repair proteins; however, we also identified several truncating variants in the genes coding proteins engaged in other DNA repair pathways. Among others, an interesting candidate is *EXO1*, which codes for exonuclease involved in numerous DNA repair pathways. Besides two indels, we identified and characterized the c.2212-1G>C splicing mutation resulting in six amino acids in-frame deletion (p.V738_K743del), involving the interaction of *EXO1* with *MSH2* during MMR (32). Contrary to our analysis, Wu et al. (33) characterized the identical

variant as a frameshift in a patient with hereditary non-polyposis CRC. Moreover, we further identified also two rare *EXO1* prioritized missense variants [c.325G>A (p.E109K) and c.1105A>C (p.S369R)] in five patients. Clustering of mutations in *EXO1* and presence of mutations in other genes involved for example in NER (*ERCC2*, *ERCC6*) suggest that an impairment of these repair processes by hereditary alterations could increase BC susceptibility. The degree to which these variants may influence BC susceptibility remains to be investigated by further studies.

Variants in non-DNA repair genes

The potentially deleterious hereditary variants were identified in 48/448 non-DNA repair genes, most frequently (in 16 BC patients) in the genes that code for the enzymes of steroid hormone metabolism and signaling. The group primarily included members of the cytochrome p450 superfamily contributing to the estrogen biosynthesis (*CYP11A1*, *CYP17A1*, *CYP19A1*) or catabolism (*CYP3A5*, *CYP1A2*) [reviewed in (34)]. Given that estrogens may affect BC etiology, variants in *CYP* genes may influence BC risk.

Variants in *CYP11A1* and *CYP17A1* identified in basal-like patients were previously described in patients suffering from severe congenital adrenal insufficiency (35) (OMIM#613743) and congenital adrenal hyperplasia (36) (OMIM#202110), respectively. Interestingly, Hopper et al. (37) reported p.R239* (c.775C>T) variant in *CYP17A1* in three *BRCA1/2*-negative young sisters with BC and hypothesized that this variant is responsible for dominantly inherited and possibly high-risk BC. Recently, Yang et al. (38) identified c.987delC (p.Y329*) variant in *CYP17A1* in a patient from an HBOC family. We found a novel variant, c.1058dupT (p.L353Ffs*10) in *CYP19A1*, in a patient with a BC and non-Hodgkin lymphoma duplicity whose mother developed bilateral BC. Mutations in similar positions cause aromatase deficiency (OMIM#613546).

Defects in estrogen-catabolizing enzymes suggest a mechanistically more obvious pathophysiological link to BC promotion. As estrogens are known substrates of *CYP3A5* and *CYP1A2* (34), 11 identified carriers of truncating variants in these genes could potentially have reduced estrogen clearance. We also found a nonsense variant in *NQO2* coding a quinone reductase eliminating estrogen quinones responsible for estrogen-initiated carcinogenesis (39) and one truncating variant in *ESR2* that codes for ER β with anti-proliferative signaling (40). The high proportion of patients carrying constitutive truncating variants in steroid hormone metabolism genes supports the hypothesis of Hopper et al. (37) suggesting that cancer-causing mutations in these genes may represent a new pathophysiological mechanism linking genetic and environmental interactions in BC susceptibility.

The other functional groups associating the patients with truncating variants in functionally relevant genes were smaller. It is obvious that at least some variants in non-DNA repair genes have very limited (if any) impact on BC susceptibility and they rather represent incidental

findings [e.g. mutations in *ABCC2* was identified in a patient with the Dubin–Johnson syndrome (41) or known *DPYD* mutation related to the fluoropyrimidines toxicity (42)]. Reporting of incidental findings is highly questionable and a matter of debate (43, 44).

Disease and individual characteristics in carriers of truncating variants

Considering the patients and disease characteristics in the carriers of mutations in the major BC predisposing genes, the earlier age at BC diagnosis or more aggressive form of BC subtypes would be expected also in the carriers of mutations in other BC-susceptibility genes. In fact, we did not identify significant differences in clinical and histopathological characteristics between the carriers and non-carriers of truncating variants. This result did not change even when prioritized variants were added into the comparison (data not shown). Similar behavior was also documented recently in a large study of 1824 triple-negative BC patients analyzed by Couch et al. (31) for hereditary mutations in 17 genes, where significant differences in enrichment for family BC/OC history and tumor characteristics were identified only in the carriers of *BRCA1/2* mutations but not in carriers of non-*BRCA1/2* mutations. We suggest that some principal changes in the evaluation of clinical and histopathological characteristics will be required to assess the clinical importance of non-*BRCA1/2* BC-susceptibility genes. Since the frequencies of mutations in these genes are lower by order than that in *BRCA1/2*, an international and consortia effort will be required for such analyses.

Conclusions

Our study identified truncating variants in 32% of patients, and 9% of patients were carriers of a truncating variant in the genes currently analyzed in clinical NGS panels for the cancer risk prediction. The most frequent truncating variants affected FA genes that, together with *BRCA1*, *BRCA2*, and *PALB2*, make this group the most important for cancer susceptibility in BC patients. Our results also show an overrepresentation of the *FANCL* variant c.1096_1099dupATTA in high-risk patients, indicating that this variant may represent a novel BC-susceptibility allele. Moreover, we identified potentially pathogenic variants in several rarely mutated DNA repair genes indicating that despite its low frequency, variants in these genes may influence the development of HBC in Czech patients. We believe that it is important to analyze such genes and in international co-operation to evaluate their contribution to the BC development because they may represent clinically valuable predictors of cancer risk in families of mutation carriers. Interestingly, in other analyzed genes, we found truncating variants in the genes coding the P450 enzymes of steroid hormones metabolism in 5% of BC patients. Therefore, this functional group may contribute to the explanation of so far undisclosed missing heritability in some high-risk BC patients. We are aware that exact role of

both c.1096_1099dupATTA and *CYP* genes in BC susceptibility needs to be further clarified by independent and larger studies.

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's web-site.

Acknowledgements

The authors wish to thank all individuals who participated in this study. This study was supported by the grant of Ministry of Health of the Czech Republic NT14054 and by BIO-CEV CZ.1.05/1.1.00/02.0109 and Charles University projects PVOUK-P27/LF1/1, UK 260148/2015, and UNCE 204011.

References

1. Robson M, Offit K. Clinical practice. Management of an inherited predisposition to breast cancer. *N Engl J Med* 2007; 357 (2): 154–162.
2. Couch FJ, Nathanson KL, Offit K. Two decades after BRCA: setting paradigms in personalized cancer care and prevention. *Science* 2014; 343 (6178): 1466–1470.
3. Rahman N. Realizing the promise of cancer predisposition genes. *Nature* 2014; 505 (7483): 302–308.
4. Karami F, Mehdipour P. A comprehensive focus on global spectrum of BRCA1 and BRCA2 mutations in breast cancer. *Biomed Res Int* 2013; 2013: 928562.
5. Easton DF, Pharoah PD, Antoniou AC et al. Gene-panel sequencing and the prediction of breast-cancer risk. *N Engl J Med* 2015; 372 (23): 2243–2257.
6. Kean S. Breast cancer. The 'other' breast cancer genes. *Science* 2014; 343 (6178): 1457–1459.
7. Ripperger T, Gadzicki D, Meindl A, Schlegelberger B. Breast cancer susceptibility: current knowledge and implications for genetic counselling. *Eur J Hum Genet* 2009; 17 (6): 722–731.
8. Pohlreich P, Stribrna J, Kleibl Z et al. Mutations of the BRCA1 gene in hereditary breast and ovarian cancer in the Czech Republic. *Med Princ Pract* 2003; 12 (1): 23–29.
9. Pohlreich P, Zikan M, Stribrna J et al. High proportion of recurrent germline mutations in the BRCA1 gene in breast and ovarian cancer patients from the Prague area. *Breast Cancer Res* 2005; 7 (5): R728–R736.
10. Ticha I, Kleibl Z, Stribrna J et al. Screening for genomic rearrangements in BRCA1 and BRCA2 genes in Czech high-risk breast/ovarian cancer patients: high proportion of population specific alterations in BRCA1 gene. *Breast Cancer Res Treat* 2010; 124 (2): 337–347.
11. Janatova M, Kleibl Z, Stribrna J et al. The PALB2 gene is a strong candidate for clinical testing in BRCA1- and BRCA2-negative hereditary breast cancer. *Cancer Epidemiol Biomarkers Prev* 2013; 22 (12): 2323–2332.
12. Kleibl Z, Havranek O, Hlavata I et al. The CHEK2 gene I157T mutation and other alterations in its proximity increase the risk of sporadic colorectal cancer in the Czech population. *Eur J Cancer* 2009; 45 (4): 618–624.
13. Janatova M, Soukupova J, Stribrna J et al. Mutation analysis of the RAD51C and RAD51D genes in high-risk ovarian cancer patients and families from the Czech Republic. *PLoS One* 2015; 10 (6): e0127711.
14. Mateju M, Stribrna J, Zikan M et al. Population-based study of BRCA1/2 mutations: family history based criteria identify minority of mutation carriers. *Neoplasma* 2010; 57 (3): 280–285.
15. Yu W, Clyne M, Khoury MJ, Gwinn M. Phenopedia and Genopedia: disease-centered and gene-centered views of the evolving knowledge of human genetic associations. *Bioinformatics* 2010; 26 (1): 145–146.
16. Harakalova M, Mokry M, Hrdlickova B et al. Multiplexed array-based and in-solution genomic enrichment for flexible and cost-effective targeted next-generation sequencing. *Nat Protoc* 2011; 6 (12): 1870–1886.
17. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 2010; 38 (16): e164.

Hereditary truncating mutations of DNA repair and other genes

18. Sevcik J, Falk M, Macurek L et al. Expression of human BRCA1 Delta 17–19 alternative splicing variant with a truncated BRCT domain in MCF-7 cells results in impaired assembly of DNA repair complexes and aberrant DNA damage response. *Cell Signal* 2013; 25 (5): 1186–1193.
19. Ali AM, Kirby M, Jansen M et al. Identification and characterization of mutations in FANCL gene: a second case of Fanconi anemia belonging to FA-L complementation group. *Hum Mutat* 2009; 30 (7): E761–E770.
20. Schroeder C, Faust U, Sturm M et al. HBOC multi-gene panel testing: comparison of two sequencing centers. *Breast Cancer Res Treat* 2015; 152 (1): 129–136.
21. Kottemann MC, Smogorzewska A. Fanconi anaemia and the repair of Watson and Crick DNA crosslinks. *Nature* 2013; 493 (7432): 356–363.
22. Walden H, Deans AJ. The Fanconi anemia DNA repair pathway: structural and functional insights into a complex disorder. *Annu Rev Biophys* 2014; 43 (1): 257–278.
23. Meetei AR, de Winter JP, Medhurst AL et al. A novel ubiquitin ligase is deficient in Fanconi anemia. *Nat Genet* 2003; 35 (2): 165–170.
24. Meetei AR, Yan Z, Wang W. FANCL replaces BRCA1 as the likely ubiquitin ligase responsible for FANCD2 monoubiquitination. *Cell Cycle* 2004; 3 (2): 179–181.
25. Akbari MR, Malekzadeh R, Lepage P et al. Mutations in Fanconi anemia genes and the risk of esophageal cancer. *Hum Genet* 2011; 129 (5): 573–582.
26. Delahaye-Sourdeix M, Anantharaman D, Timofeeva MN et al. A rare truncating BRCA2 variant and genetic susceptibility to upper aerodigestive tract cancer. *J Natl Cancer Inst* 2015; 107 (5): djv037.
27. Kiiski JI, Peltari LM, Khan S et al. Exome sequencing identifies FANCM as a susceptibility gene for triple-negative breast cancer. *Proc Natl Acad Sci U S A* 2014; 111 (42): 15172–15177.
28. Peterlongo P, Catucci I, Colombo M et al. FANCM c.5791C>T nonsense mutation (rs144567652) induces exon skipping, affects DNA repair activity and is a familial breast cancer risk factor. *Hum Mol Genet* 2015; 24 (18): 5345–5355.
29. Smith CG, Naven M, Harris R et al. Exome resequencing identifies potential tumor-suppressor genes that predispose to colorectal cancer. *Hum Mutat* 2013; 34 (7): 1026–1034.
30. Loveday C, Turnbull C, Ruark E et al. Germline RAD51C mutations confer susceptibility to ovarian cancer. *Nat Genet* 2012; 44 (5): 475–476.
31. Couch FJ, Hart SN, Sharma P et al. Inherited mutations in 17 breast cancer susceptibility genes among a large triple-negative breast cancer cohort unselected for family history of breast cancer. *J Clin Oncol* 2015; 33 (4): 304–311.
32. Rasmussen LJ, Rasmussen M, Lee B et al. Identification of factors interacting with hMSH2 in the fetal liver utilizing the yeast two-hybrid system. *In vivo* interaction through the C-terminal domains of hEXO1 and hMSH2 and comparative expression analysis. *Mutat Res* 2000; 460 (1): 41–52.
33. Wu Y, Berends MJ, Post JG et al. Germline mutations of EXO1 gene in patients with hereditary nonpolyposis colorectal cancer (HNPCC) and atypical HNPCC forms. *Gastroenterology* 2001; 120 (7): 1580–1587.
34. Blackburn HL, Ellsworth DL, Shriver CD, Ellsworth RE. Role of cytochrome P450 genes in breast cancer etiology and treatment: effects on estrogen biosynthesis, metabolism, and response to endocrine therapy. *Cancer Causes Control* 2015; 26 (3): 319–332.
35. Hiort O, Holterhus PM, Werner R et al. Homozygous disruption of P450 side-chain cleavage (CYP11A1) is associated with prematurity, complete 46, XY sex reversal, and severe adrenal failure. *J Clin Endocrinol Metab* 2005; 90 (1): 538–541.
36. Hwang DY, Hung CC, Riepe FG et al. CYP17A1 intron mutation causing cryptic splicing in 17alpha-hydroxylase deficiency. *PLoS One* 2011; 6 (9): e25492.
37. Hopper JL, Hayes VM, Spurdle AB et al. A protein-truncating mutation in CYP17A1 in three sisters with early-onset breast cancer. *Hum Mutat* 2005; 26 (4): 298–302.
38. Yang X, Wu J, Lu J et al. Identification of a comprehensive spectrum of genetic factors for hereditary breast cancer in a Chinese population by next-generation sequencing. *PLoS One* 2015; 10 (4): e0125571.
39. Gaikwad NW, Yang L, Rogan EG, Cavalieri EL. Evidence for NQO2-mediated reduction of the carcinogenic estrogen ortho-quinones. *Free Radic Biol Med* 2009; 46 (2): 253–262.
40. Caiazza F, Ryan EJ, Doherty G, Winter DC, Sheahan K. Estrogen receptors and their implications in colorectal carcinogenesis. *Front Oncol* 2015; 5: 19.
41. Pacifico L, Carducci C, Poggiogalle E et al. Mutational analysis of ABCC2 gene in two siblings with neonatal-onset Dubin Johnson syndrome. *Clin Genet* 2010; 78 (6): 598–600.
42. Kleibl Z, Fidlerova J, Kleiblova P et al. Influence of dihydropyrimidine dehydrogenase gene (DPYD) coding sequence variants on the development of fluoropyrimidine-related toxicity in patients with high-grade toxicity and patients with excellent tolerance of fluoropyrimidine-based chemotherapy. *Neoplasma* 2009; 56 (4): 303–316.
43. Blackburn HL, Schroeder B, Turner C, Shriver CD, Ellsworth DL, Ellsworth RE. Management of incidental findings in the era of next-generation sequencing. *Curr Genomics* 2015; 16 (3): 159–174.
44. Christenhusz GM, Devriendt K, Dierickx K. Disclosing incidental findings in genetics contexts: a review of the empirical ethical research. *Eur J Med Genet* 2013; 56 (10): 529–540.

PŘÍLOHA IV-2:

Zemankova *et al.*; RE: frameshift variant *FANCLc.1096_1099dupATTA is not associated with high breast cancer risk**



Correspondence

RE: frameshift variant *FANCL**c.1096_1099dupATTA is not associated with high breast cancer risk

To the Editor,

In our recent publication, we noted a borderline association of the truncating variant c.1096_1099dupATTA in the *FANCL* gene with increased breast cancer (BC) risk in high-risk *BRCA1/BRCA2/PALB2*-negative BC patients (1). However, the subsequent analysis by Pfeifer et al. (2) genotyping the c.1096_1099dupATTA variant in 2370 samples from German and Macedonian BC patients and controls failed to confirm association of this variant with BC risk.

We agree with Pfeifer et al. that *FANCL**c.1096_1099dupATTA is unlikely a high-risk BC susceptibility allele with an immediate clinical utility. There are several lines of evidence that do not support strong involvement of this variant in BC susceptibility including: (i) the relative high frequency of this variant especially in European populations, (ii) functional characteristics demonstrating that cells expressing *FANCL* isoform coded by c.1096_1099dupATTA variant retain the residual *FANCL* functional capacity *in vitro* (3), and (iii) phenotypic characteristics that show only a mild Fanconi anemia (FA) complementation group L phenotype in the compound heterozygote carrying *FANCL**c.1096_1099dupATTA (alongside an another truncating *FANCL* variant) (3). Moreover, we identified a male c.1096_1099dupATTA homozygote (in controls) who had no signs of FA at his age of 57 years (Table 1).

We also agree that *FANCL**c.1096_1099dupATTA may confer a low (or lower) risk variant. We hypothesized that this variant may represent a modifying factor because its carriers were overrepresented only

in a subgroup of high-risk BC patients in our study and also four out of six c.1096_1099dupATTA carriers analyzed by a panel next-generation sequencing (NGS) carried truncating variant(s) in other known or candidate cancer-susceptibility gene(s). After publication of our study reporting 15 carriers of c.1096_1099dupATTA in 2126 analyzed samples of Czech BC patients and controls, we identified another eight carriers using the CZECANCA multicancer panel NGS (4). The individual characteristics of c.1096_1099dupATTA carriers (Table 1) indicate a relatively low mean age at diagnose [47.2 years (range 28–76 years)] in 14 carriers with BC. We also recently identified three c.1096_1099dupATTA carriers with ovarian cancer diagnosed at early age. Contrary to Pfeifer et al. who reported that only one out of 10 identified c.1096_1099dupATTA carriers had a family BC history, we have noticed a known familial history of BC (in a first or second degree relative) in nine out of 23 carriers (39%) and a familial history of some cancer in 15 carriers (65%). The c.1096_1099dupATTA variant was accompanied by another truncating variant(s) in nine out of 14 cancer patients analyzed by a panel NGS (Table 1). We suppose that these characteristics indicate that c.1096_1099dupATTA may (perhaps mildly) modify the breast (or other) cancer risk or cancer onset. However, further studies are required to estimate the risk of cancer development in c.1096_1099dupATTA carriers precisely. The segregation analyses and NGS analyses in families of carriers would be also required to evaluate the involvement of this hypomorphic variant in the risk of other cancer development or in modification of cancer onset.

Correspondence

Table 1. The clinical and genetic characteristics of c.1096_1099dupATTA carriers identified in the Czech population^a

Proband no.	Cancer diagnosis	Age at dg (*at analysis)	Family cancer history (date at diagnosis, or + death)	NGS panel	Truncating variant in cancer susceptibility or candidate genes
The carriers of c.1096_1099dupATTA in FANCL referred in Lhota et al. (1)					
1249 ^b	BC	36	FM-BC (+40); FF-Brain tumor(+73); MM-BC (+70).	SP	None.
1252 ^b	BC	41	F-RC(62); FM-RC(75); FB-PrC(65); MM-BC .	SP	<i>CHEK2</i> : c.444+1G>A (p.R148Vfs*6), c.277delT (p.W93Gfs*17); <i>XRCC4</i> : c.25delC (p.H9Tfs*8).
748 ^b	BC	44	MS1-BC (52); MS2-BC (58); MS3-BC (60).	SP	<i>ABCC2</i> : c.C3196T (p.R1066*).
C0211 ^b	BC&BC	42&43	Unknown.	SP	<i>DCLRE1C</i> : c.1903dupA (p.S635Kfs*6).
1316 ^b	BC	67	M-BC (+50); MS-BC (80).	SP	<i>CYP3A5</i> : c.92dupG (p.L32Tfs*3); <i>IL8</i> : c.91G > T (p.E31*).
1331 ^b	BC	76	M-BC (75); S-BC (63).	SP	None.
960 ^c	BC	33	None.	CZ	None.
1908 ^c	BC	38	S-melanoma(48); FM-RC.	CZ	<i>MSR1</i> : c.877C>T (p.R293*).
1782 ^c	BC	28	None.	CZ	None.
A546 ^d	BC	47	None.	n.d.	–
A626 ^d	BC	73	Unknown.	n.d.	–
A032 ^d	BC	54	None.	n.d.	–
K101 ^e	None	*80	B-myeloma(+65).	n.d.	–
C015 ^{e,f}	None	*57	F-PrC.	n.d.	–
C308 ^e	None	*50	None.	n.d.	–
The carriers of c.1096_1099dupATTA in FANCL identified after publication of Lhota et al. (1) study					
3100	BC	29	None.	CZ	None.
2946a16	BC	57	M-BC (51); D-Hodgkin (24); F-LC(69); FM-BC (50).	CZ	None.
2885a16	None.	*44	M-BC (69).	CZ	None.
1846a15	None	*37	MM-BC (40)& BC (70); FB-CRC.	CZ	<i>DNAJC21</i> : c.1503delA (p.K501Nfs*10).
3524a15	None	*31	M-CRC(48); MM-UBC; two cousins-BC.	CZ	None.
868	OC	26	None.	CZ	<i>BRCA2</i> : c.A9976T (p.K3326*).
120	OC	39	F-UBC(58).	CZ	<i>TSHR</i> : c.2102dupG (p.Q702Pfs*17).
2864	OC	48	F-LC(+74); FS2-GBC; FS3-BC(+82) ; FB-LC(+74).	CZ	<i>RAD51C</i> : c.502A>T (p.R168*).

B, brother; CZ, CZECA NCA (CZEch CAncer paNel for Clinical Application) cancer panel (219 genes), described in (4); CRC, colorectal cancer; D, daughter; GBC, gallbladder cancer; FB, father's brother; FM, father's mother; LC, lung cancer; MM, mother's mother; MS, mother's sister; n.d., not done; PrC, prostate cancer; RC, renal cancer; S, sister; SP, SOLiD gene panel (581 genes) used in Lhota et al. (1); UBC, urinary bladder cancer.

^a20/23 carriers (except K101, C015, and C308) were females. All BC histologies were ductal BCs (except the second medullary BC in patient C0211).

^bSix carriers identified by NGS analysis in 325 high-risk BC patients.

^cThree carriers identified in additional sample set of 337 high-risk BC patients by genotyping (analyzed further by CZECA NCA panel NGS).

^dThree carriers identified in 673 sporadic BC patients.

^eThree carriers (males) identified in 686 controls.

^fHomozygote for the c.1096_1099dupATTA variant in the *FANCL* gene.

Acknowledgements

This study was supported by the Ministry of Health of the Czech Republic, grant nr. 15-27695A and Charles University in Prague projects PRVOUK P27 and SVV UK 260256/2016

P. Zemankova^a

F. Lhota^{a,b}

P. Kleiblova^{a,c}

J. Soukupova^a

M. Vocka^d

M. Janatova^a

Z. Kleibl^a

^a*Institute of Biochemistry and Experimental Oncology,
First Faculty of Medicine, Charles University in Prague,
Prague, Czech Republic*

^b*Centre for Medical Genetics and Reproductive Medicine
GENNET,
Prague, Czech Republic*

^c*Institute of Biology and Medical Genetics,
First Faculty of Medicine, Charles University in Prague
and General University Hospital in Prague,
Prague, Czech Republic*

^d*Department of Oncology,
First Faculty of Medicine, Charles University in Prague
and General University Hospital in Prague,
Prague, Czech Republic
e-mail: zdekleje@lf1.cuni.cz*

References

1. Lhota F, Zemankova P, Kleiblova P et al. Hereditary truncating mutations of DNA repair and other genes in BRCA1/BRCA2/PALB2-negatively tested breast cancer patients. *Clin Genet* 2016; 90 (4): 324–333.
2. Pfeifer K, Schürmann P, Bogdanova N et al. Frameshift variant FANCL*c.1096_1099dupATTA is not associated with high breast cancer risk. *Clin Genet* 2016; 90 (4): 386–387.
3. Ali AM, Kirby M, Jansen M et al. Identification and characterization of mutations in FANCL gene: a second case of Fanconi anemia belonging to FA-L complementation group. *Hum Mutat* 2009; 30 (7): E761–E770.
4. Soukupová J, Zemánková P, Kleiblová P, Janatová M, Kleibl Z. [CZECANCA: CZEch CAncer paNel for Clinical Application- design and optimization of the targeted sequencing panel for the identification of cancer susceptibility in high-risk individuals from the Czech Republic]. *Klin Onkol* 2016; 29 (Suppl 1): S46–S54.




Correspondence: Zdenek Kleibl, MD, PhD,
Institute of Biochemistry and Experimental Oncology, First Faculty of
Medicine, Charles University in Prague, U Nemocnice 5, 128 53
Prague 2, Czech Republic.
Tel.: +420 224965745;
fax: +420 224965732;
e-mail: zdekleje@lf1.cuni.cz

PŘÍLOHA IV-3:

Rump *et al.*; Identification and Functional Testing of *ERCC2* Mutations in a Multi-national Cohort of Patients with Familial Breast- and Ovarian Cancer

RESEARCH ARTICLE

Identification and Functional Testing of *ERCC2* Mutations in a Multi-national Cohort of Patients with Familial Breast- and Ovarian Cancer

Andreas Rump^{1,2,3}, Anna Benet-Pages⁴, Steffen Schubert⁵, Jan Dominik Kuhlmann^{2,3,6,7*}, Ramūnas Janavičius^{8,9}, Eva Macháčková¹⁰, Lenka Foretová¹⁰, Zdenek Kleibl¹¹, Filip Lhota¹¹, Petra Zemankova¹¹, Elitza Betcheva-Krajcir^{1,2,3}, Luisa Mackenroth^{1,2,3}, Karl Hackmann^{1,2,3,6}, Janin Lehmann⁵, Anke Nissen⁴, Nataliya DiDonato^{1,2,3}, Romy Opitz^{2,3,6,7}, Holger Thiele¹², Karin Kast^{2,3,6,7}, Pauline Wimberger^{2,3,6,7}, Elke Holinski-Feder⁴, Steffen Emmert^{5,13}, Evelin Schröck^{1,2,3,6}, Barbara Klink^{1,2,3,6}



CrossMark
click for updates

 OPEN ACCESS

Citation: Rump A, Benet-Pages A, Schubert S, Kuhlmann JD, Janavičius R, Macháčková E, et al. (2016) Identification and Functional Testing of *ERCC2* Mutations in a Multi-national Cohort of Patients with Familial Breast- and Ovarian Cancer. *PLoS Genet* 12(8): e1006248. doi:10.1371/journal.pgen.1006248

Editor: Charis Eng, Cleveland Clinic Genomic Medicine Institute, UNITED STATES

Received: June 8, 2016

Accepted: July 18, 2016

Published: August 9, 2016

Copyright: © 2016 Rump et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: The Czech-Brno study was supported by MH CZ - DRO (MMCI, 00209805) and RECAMO, CZ.1.05/2.1.00/03.0101. The Czech-Prague study was supported by Ministry of Health of the Czech Republic, grant nr. 15-28830A and projects of the Charles University in Prague PRVOUK P27 and SVV UK 260256/2016. Functional analyses were funded by the Claudia von Schilling Foundation for breast cancer research (to SS and SE). The Lithuanian

1 Institute for Clinical Genetics, Faculty of Medicine Carl Gustav Carus, Technische Universität Dresden, Dresden, Germany, 2 German Cancer Consortium (DKTK), Dresden, Germany, 3 German Cancer Research Center (DKFZ), Heidelberg, Germany, 4 MGZ—Medical Genetics Center, Munich, Germany, 5 Clinic for Dermatology Venerology and Allergology, Göttingen, Germany, 6 National Center for Tumor Diseases (NCT), Partner Site Dresden, Germany, 7 Department of Gynecology and Obstetrics, Medical Faculty and University Hospital Carl Gustav Carus, Technische Universität Dresden, Germany, 8 Vilnius University Hospital Santariskiu Clinics, Hematology, Oncology and Transfusion Medicine Center, Vilnius, Lithuania, 9 State Research Institute Innovative Medicine Center, Vilnius, Lithuania, 10 Masaryk Memorial Cancer Institute, Brno, Czech Republic, 11 Institute of Biochemistry and Experimental Oncology, First Faculty of Medicine, Charles University in Prague, Prague, Czech Republic, 12 Cologne Center for Genomics, Cologne, Germany, 13 Clinic of Dermatology, Rostock, Germany

 These authors contributed equally to this work.

* jan.kuhlmann@uniklinikum-dresden.de

Abstract

The increasing application of gene panels for familial cancer susceptibility disorders will probably lead to an increased proposal of susceptibility gene candidates. Using *ERCC2* DNA repair gene as an example, we show that proof of a possible role in cancer susceptibility requires a detailed dissection and characterization of the underlying mutations for genes with diverse cellular functions (in this case mainly DNA repair and basic cellular transcription). In case of *ERCC2*, panel sequencing of 1345 index cases from 587 German, 405 Lithuanian and 353 Czech families with breast and ovarian cancer (BC/OC) predisposition revealed 25 mutations (3 frameshift, 2 splice-affecting, 20 missense), all absent or very rare in the ExAC database. While 16 mutations were unique, 9 mutations showed up repeatedly with population-specific appearance. Ten out of eleven mutations that were tested exemplarily in cell-based functional assays exert diminished excision repair efficiency and/or decreased transcriptional activation capability. In order to provide evidence for BC/OC predisposition, we performed familial segregation analyses and screened ethnically matching controls. However, unlike the recently published *RECQL* example, none of our recurrent *ERCC2* mutations showed convincing co-segregation with BC/OC or significant

study was supported by the Research Council of Lithuania (Project SEN18/2015). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Contributions:

- Conceptualization:** AR ABP SS BK.
- Data curation:** AR ABP SS.
- Formal analysis:** AR ABP SS RJ ZK EM.
- Funding acquisition:** PW EHF SE ES BK.
- Investigation:** AR ABP SS RJ ZK EM.
- Methodology:** AR ABP SS RJ ZK EM BK.
- Project administration:** BK.
- Resources:** AR ABP SS JDK RJ LF ZK FL PZ EBK LM KH JL AN ND RO HT KK.
- Supervision:** AR.
- Validation:** AR ABP SS.
- Visualization:** AR ABP SS AN.
- Writing - original draft:** AR ABP SS JDK SE BK.
- Writing - review & editing:** AR JDK.

overrepresentation in the BC/OC cohort. Interestingly, we detected that some deleterious founder mutations had an unexpectedly high frequency of > 1% in the corresponding populations, suggesting that either homozygous carriers are not clinically recognized or homozygosity for these mutations is embryonically lethal. In conclusion, we provide a useful resource on the mutational landscape of ERCC2 mutations in hereditary BC/OC patients and, as our key finding, we demonstrate the complexity of correct interpretation for the discovery of “bonafide” breast cancer susceptibility genes.

Author Summary

Approximately 5–10% of breast/ovarian cancer (BC/OC) cases have inherited an increased risk of developing this malignancy. However, mutations in the two major breast cancer susceptibility genes BRCA1 and BRCA2 explain only 15–20% of all familial BC/OC cases. With the emergence of the high throughput NGS-technology, the number of proposed novel candidate genes for breast cancer predisposition continuously increases. However, a “bonafide” proof of cancer susceptibility requires a detailed characterization of candidate mutations, which we addressed in the current study. Using the DNA repair gene ERCC2 as an example, we performed a comprehensive multi-center approach, analyzing ERCC2 mutations in 1000+ patients with hereditary BC/OC. We identified 25 potential candidate mutations for cancer breast cancer susceptibility, some of them affecting ERCC2 functional activity in appropriate cell-culture based assays. However, a more dissected analysis showed no convincing co-segregation with BC/OC and there was no longer a significant overrepresentation in BC/OC when compared to regionally matched controls instead of the global ExAc variant data base, pointing to the relevance of founder-mutations. In conclusion, we provide a useful resource on the mutational landscape of ERCC2 mutations in hereditary BC/OC patients and, as our key finding, we highlight the complexity of correct interpretation for the discovery of “bonafide” breast cancer susceptibility genes.

Introduction

Since it became evident that only 15%-20% of the familial risk for BC/OC can be explained by mutations in the major breast cancer-susceptibility genes BRCA1 and BRCA2 [1], the search for additional BC/OC susceptibility loci has been pursued. In times of limited sequencing power this pursuit was based on carefully selected candidate genes which typically came from (i) cancer-associated syndromes (ii) linkage screens in large BRCA1/2-negative families and (iii) case-control association studies using single-nucleotide polymorphisms [2,3]. Since sequencing power is no longer an issue, the candidate approach is on its decline and about to be replaced by next generation sequencing (NGS) of large gene panels which, taken together, cover a total of more than 100 genes, only 21 of which have been associated with breast cancer so far [4]. This offers amazing opportunities for detection of novel susceptibility loci but also bears the danger of substantial misuse [4], because variants picked up by these panels are not clinically validated. Therefore, post-marketing data validation is absolutely essential [5]. Rare variants, however, need huge case-control datasets in order to reach the requested statistical significance of $P < 0.0001$ [4]. Until such large datasets become available, variant validation needs to focus on mutations that are clearly deleterious on functional level but still frequent enough to be validated by a few thousand controls. Such recurrent yet harmful variants are best

identified by screening various populations for founder mutations. In *NBN*, for example, a protein-truncating variant (c.657del5) has been identified in Eastern Europe, which is sufficiently common to allow its evaluation in a BC/OC case-control study [6]. Also the successful validation of deleterious Polish and Canadian founder mutations in *RECQL* [7] underlines the huge potential of multi-national BC/OC cohorts.

In this study we sequenced 1345 BC/OC cases from 3 different Central- and East European countries with multi-gene panels and identified recurrent founder mutations in *ERCC2*, which were functionally validated in cell-culture based assays. As essential component of transcription factor IIH, the ERCC2 protein is involved in basal cellular transcription [8] and nucleotide excision repair (NER) of DNA lesions [9]. The most known inherited disease associated with bi-allelic mutations in *ERCC2* is Xeroderma pigmentosum type D (XPD, OMIM 278730), a hereditary cancer-prone syndrome characterized by extreme skin photosensitivity and early development of multiple skin tumors [10]. Therefore, *ERCC2* is a plausible candidate gene for cancer susceptibility. On the other hand, bi-allelic mutations in *ERCC2* can also lead to syndromes without increased propensity to tumor development, namely Trichothiodystrophy 1 (TTD; OMIM 601675) and cerebrooculofacioskeletal syndrome (COFS2; OMIM 610756). This indicates that not all functionally relevant *ERCC2* mutations increase cancer susceptibility in their carriers.

Results and Discussion

Panel sequencing identifies a broad spectrum of rare variants as well as recurrent founder mutations in *ERCC2*

Within the entire set of 1345 BC/OC index cases, we have detected three different frame-shift (fs) mutations [p.(Val77fs), p.(Phe568fs) and p.(Ser746fs)], one splice-acceptor site mutation (c.1903-2A>G), one nucleotide exchange that activates a cryptic splice site (c.2150C>G) and 20 rare missense mutations (Table 1, Fig 1). Whereas 14 mutations were unique (2 fs, 1 splice-site, 11 missense), 11 mutations (1 fs, 1 splice-affecting, 9 missense) have been found in 43 independent families. The most frequent mutation was p.(Asp423Asn) identified in 8 carriers from Lithuania and one from the Czech Republic. The common polymorphisms p.(Lys751Gln) and p.(Asp312Asn) have each been encountered in approximately 64% of our cases; since these variants have been considered to be functionally irrelevant [11], we did not include them in our functional study. Among the 20 rare missense variants reported in Table 1, thirteen are predicted by various computer algorithms to be pathogenic (Table 1 and S4 Table). Further computational analysis of the conservation (PhyloP) and depletion (CADD) scores [12] for the mutated nucleotides strongly supported pathogenicity for these variants (S2 Fig). Mapping the mutated AA positions onto the ERCC2 protein structure revealed a widespread distribution pattern (Fig 1). Residues 13, 450, 461, 513, 536, 576, 592, 601, 611, 631, 678 cluster at the helicase motifs of the HD1 and HD2 catalytic domains and residues 166, 167, 188, 215, 280, 316, 423, 487, 722 locate at the TFIID transcription factor complex binding domains (Arch, FES, and C-terminal). XPD-causing mutations located at the HD2 domain have been shown to inactivate helicase repair capability without disrupting protein structure. Mutations causing trichothiodystrophy (TTD, OMIM 601675), on the other hand, are located well away from the catalytic site of the enzyme and destabilize ERCC2 structure and TFIID protein interactions [13–15]. We suggest that BC/OC relevant mutations might affect both—catalytic activity as well as protein stability.

Functional testing identifies *ERCC2* mutations with deleterious effects on protein level

So far, 11 variants (9 recurrent founder mutations and 2 unique variants; Fig 2C) were tested in functional assays for nucleotide excision repair (NER) capability (Fig 2A) as well as

Table 1. Mutations and rare variants in ERCC2 identified through panel sequencing of individuals with familial breast and/or ovarian cancer. AA = amino acid; N = sample size; n.a. = not applicable; n.t. = not tested; CZ = Czech Republic, GE = Germany, LT = Lithuania. The cumulative assessment is based on the results of various effect prediction algorithms; details see [S4 Table](#).

Variant description				Predicted effect			Functional effect			BC/OC cases			
Position	Exon	Nucleotide change	AA change	rs-ID	Cumulative assessment	Complementation of NER-deficient cells	Negative modulation of transcription	GE	CZ	LT	total	Tumor type	
hg19	(23)	NM_000400.3	max = 760 aa					N = 587	N = 353	N = 405	N = 1345		
19:45873459	2	c.37C>T	p.(Pro13Ser)	-	pathogenic	n.t.	n.t.	1	0	0	1	BC	
19:45872203	4	c.230_231delITG	p.(Val77Alafs)	-	n.a.	n.t.	n.t.	0	1	0	1	BC+OC	
19:45868194	7	c.496C>T	p.(Arg166Cys)	-	pathogenic	n.t.	n.t.	0	0	2	2	BC	
19:45868191	7	c.499G>C	p.(Glu167Gln)	rs3678229012	benign	n.t.	n.t.	1	0	0	1	BC	
19:45868127	7	c.563G>C	p.(Gly188Ala)	-	benign	n.t.	n.t.	1	0	0	1	BC	
19:45867756	8	c.644C>T	p.(Pro215Leu)	-	pathogenic	n.t.	n.t.	0	0	1	1	BC	
19:45867354	10	c.839G>A	p.(Arg280His)	-	pathogenic	n.t.	n.t.	0	0	1	1	BC	
19:45867247	10	c.946C>G	p.(Gln316Glu)	-	benign	n.t.	n.t.	1	0	0	1	BC	
19:458660928	13	c.1267G>A	p.(Asp423Asn)	rs143710107	benign	no	yes	0	1	8	9	4xBC, 5xOC	
19:45860760	14	c.1349G>A	p.(Arg450His)	rs146632315	pathogenic	yes	no	2	0	0	2	BC	
19:45860626	15	c.1381C>G	p.(Leu461Val)	rs121913016	benign	yes	yes	3	0	0	3	2xBC, 1xOC	
19:45860548	15	c.1459C>T	p.(Arg487Trp)	rs562132292	pathogenic	no	yes	0	0	4	4	2xBC, 2xOC	
19:45858929	16	c.1537G>T	p.(Asp513Tyr)	-	pathogenic	yes	yes	1	0	0	1	BC	
19:45858047	17	c.1606G>A	p.(Val536Met)	rs142568756	pathogenic	yes	yes	2	0	0	2	BC	
19:45856554	18	c.1703_1704delIT	p.(Phe568fs)	-	pathogenic	no	no	1	3	1	5	BC	
19:45856532	18	c.1726G>A	p.(Glu576Lys)	rs201165309	pathogenic	n.t.	n.t.	1	0	0	1	BC	
19:45856397	19	c.1775G>A	p.(Arg592His)	rs147224585	pathogenic	yes	no	1	0	7	8	BC	
19:45856370	19	c.1802G>A	p.(Arg601Gln)	rs140522180	pathogenic	yes	yes	2	1	0	3	BC	
19:45856074	20	c.1832T>C	p.(Val611Ala)	-	benign	n.t.	n.t.	0	1	0	1	BC	
19:45856015	20	c.1891C>T	p.(Arg631Cys)	rs1444511865	pathogenic	no	no	1	0	1	2	1xBC, 1xOC	
19:45855909	IVS 20	c.1903-2A>G	splice site	-	n.a.	n.t.	n.t.	1	0	0	1	BC+OC	
19:45855778	21	c.2032G>C	p.(Val678Leu)	-	benign	n.t.	n.t.	0	0	1	1	BC	
19:45855507	22	c.2150C>G	splice effect	rs1444564120	pathogenic	n.t.	n.t.	3	0	0	3	2xBC, 1xOC	
19:45855492	22	c.2165G>A	p.(Arg722Gln)	rs138569838	pathogenic	n.t.	n.t.	0	1	0	1	BC	
19:45854932	23	c.2238delA	p.(Ser746fs)	-	n.a.	yes	yes	1	0	0	1	OC	

doi:10.1371/journal.pgen.1006248.t001

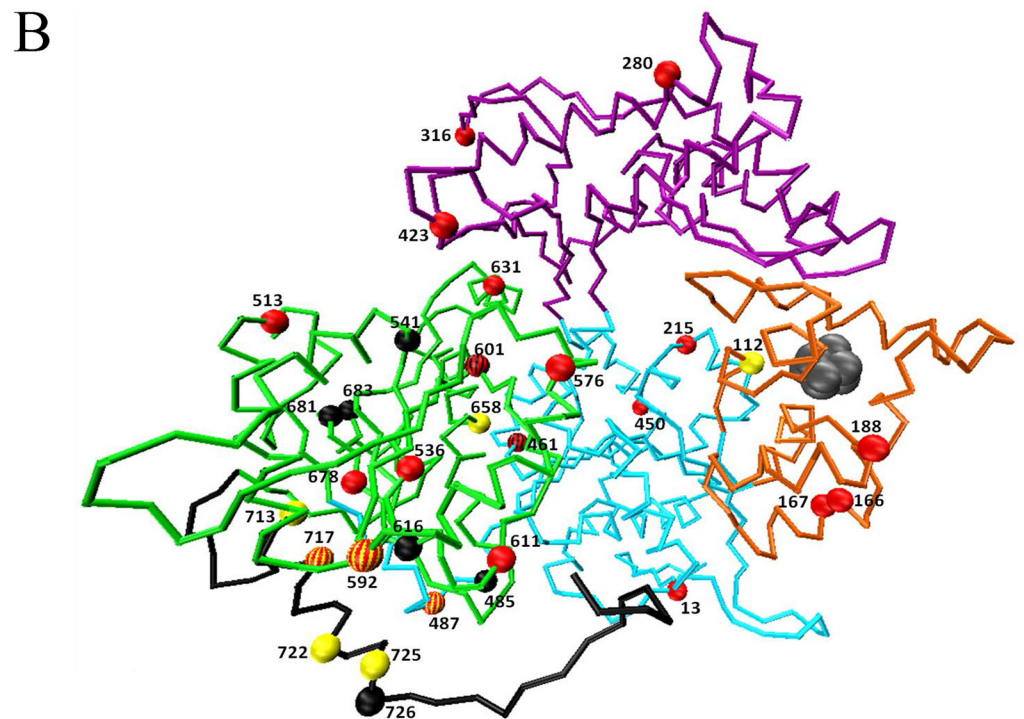
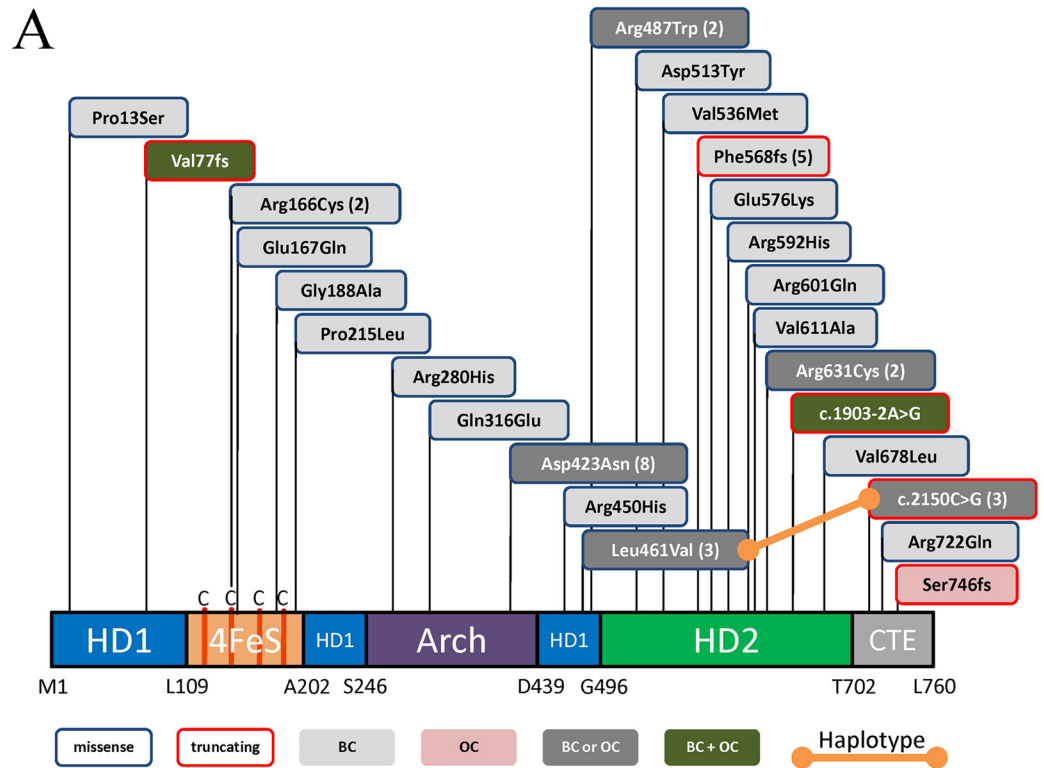


Fig 1. Domain structure and modeling of the ERCC2 mutations. (A) Mutations in the XPD/ERCC2 protein domains. The diagram shows the ERCC2 protein with the four XPD domains shown as HD1 (blue), HD2 (green), FeS (Orange) and Arch (purple). The human enzyme has a C-terminal (grey) extension (CTE) that probably forms an interaction surface with the p44 protein. Disease-relevant ERCC2 mutation sites are indicated in boxes (blue or red frame: missense or truncating mutation, respectively; fillings: light-gray, cases with breast cancer

(BC); pink, case with ovarian cancer only (OC); dark-gray: cases with either breast- or ovarian cancer (BC or OC); dark-green, patients with both breast- and ovarian cancer (BC + OC)). Numbers in brackets indicate recurrent mutations. (B) Structural placement of mutations on a C-alpha trace model of human ERCC2. The residues targeted by HBOC-causing mutations are represented as space-filled red spheres. Xeroderma pigmentosum (XP) and trichothiodystrophy (TTD) disease causing mutations sites as reported in ClinVar are shown in yellow and black spheres. Missense variants at residue position 423, 461, 487, 568, 461 and 722 have been found in both BC/OC as well as XP (red-yellow spheres) and TTD (red-black spheres) patients.

doi:10.1371/journal.pgen.1006248.g001

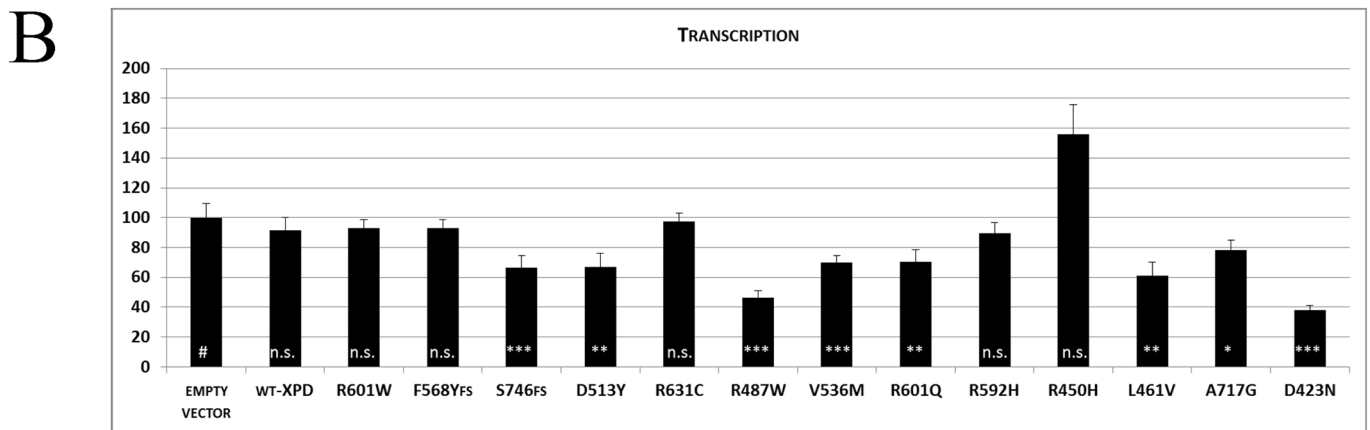
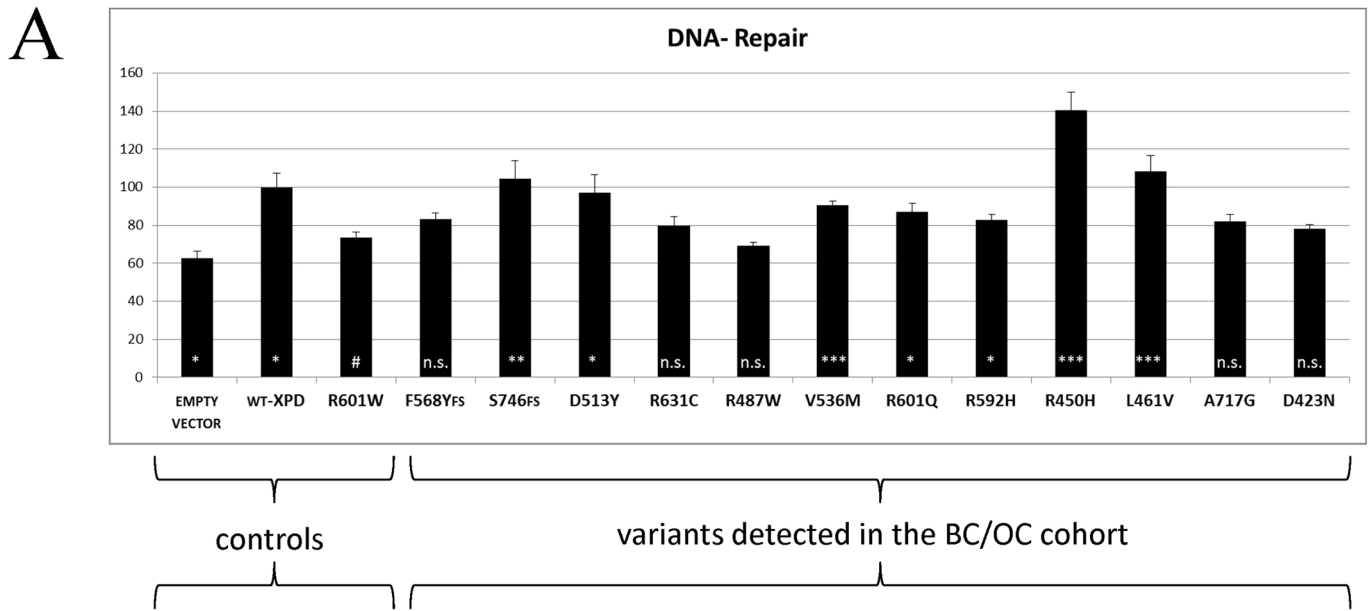
transcription (Fig 2B). Whereas six out of the 11 BC/OC-associated ERCC2 variants tested in this study, have not yet been linked to any disease [AA positions 423, 450, 513, 536, 631, 746], five AA positions have already been found to be mutated in either TTD [AA 461 [16], 487 [17], 568 [18,19], 592 [20]] or XPD [AA 601 [21]] (Figs 1B and 2C). According to our functional assays, four ERCC2 protein variants [p.(Asp423Asn), p.(Arg487Trp), p.(Phe568Tyrfs) and p.(Arg631Cys)] failed to enhance functional NER of an UV-treated reporter gene plasmid indicating the impairment of ERCC2 repair capacity. The remaining seven tested variants retained some NER capability (Fig 2A). Concerning transcription, we detected a dominant negative influence of seven ERCC2 protein variants [p.(Asp423Asn), p.(Leu461Val), p.(Arg487Trp), p.(Asp513Tyr), p.(Val536Met), p.(Arg601Gln), p.(Ser746fs)] on reporter gene expression (Fig 2B) indicating transcription blocking. In summary, 10 of 11 mutations display diminished excision repair efficiency and/or decreased transcriptional activation capability, with p.(Asp423Asn) and p.(Arg487Trp) being the variants with the highest impact on protein function.

The majority of the ERCC2 mutations are founder mutations

The hallmarks of a founder mutation are recurrent appearance, population specificity and haplotype sharing. As to recurrent appearance, 11 out of 25 ERCC2 mutations were seen at least twice in our BC/OC cohort (last column in Table 1). Among the 11 recurrent variants, 5 were identified exclusively in one of the three populations tested in this study (e.g. p.(Arg487Trp): 4x LT only) and another 5 were significantly overrepresented in one of the 3 populations (e.g. p.(Asp423Asn): 8x LT, 1x CZ, 0x GE). For two of the population-enriched recurrent founder mutations, we could also demonstrate haplotype sharing: (i) the mutation c.1381C>G (rs121913016) always co-occurred and co-segregated with mutation c.2150C>G (rs144564120), a haplotype which has been observed repeatedly in TTD/XPD patients [9,16,22]. (ii) In almost all cases (10/11) the frame-shift mutation c.1703_1704delTT co-occurred with the c.1758+32C>G polymorphism (rs238417). Furthermore, these two variants are only 84 nt apart from each other and all NGS-reads covering both variants showed these variants simultaneously, i.e. these variants are definitely localized in *cis* on the same DNA molecule.

Even small region-specific control cohorts outnumber huge public variant databases

In the variant discovery phase of this project, the frequencies of ERCC2 mutations found in the BC/OC cohort were compared to the corresponding frequencies in public databases provided by the NHLBI Exome Sequencing Project (ESP) and the Exome Aggregation Consortium (ExAC). As shown in Table 2, some intriguing mutations, like p.(Phe568fs) and p.(Asp423Asn), have very low frequencies according to ExAC, suggesting significant odds ratios (OR). As a first proof of principle measure, we performed segregation analysis. However, none of our recurrent ERCC2 mutations showed convincing co-segregation with BC/OC (Fig 3). Moreover, as soon as a small number of population-specific control probands has been sequenced, it became clear that almost all founder mutations in the BC/OC cohort showed



C

Mutation / characteristics	Asp 423 Asn	Arg 450 His	Leu 461 Val	Arg 487 Trp	Asp 513 Tyr	Val536Met	Phe 568 fs	Arg 592 His	Arg 601 Gln	Arg 631 Cys	Ser 746 fs
Codon linked to disease	TTD	---	TTD	TTD	---	---	TTD	---	XPD	---	---
In-silico effect prediction	Bgn	Pat	Bgn	Pat	Pat	Pat	fs	Pat	Pat	Pat	fs
Recurrent mutation	Yes	Yes	Yes	Yes	---	Yes	Yes	Yes	Yes	Yes	---

Fig 2. Nucleotide excision repair (NER) capacity and Transcriptional activity of breast cancer associated XPD/ERCC2 variants. (A) Several XPD/ERCC2 variants cloned into an expression vector were analyzed regarding to complementation of ERCC2-defective XP6BE cells overexpressing the NER-deficient R601W XPD mutant [15] (normalization for overexpression artifacts). Black bars indicate the mean relative repair capacity (in %, WT-XPD was set to 100%) of an UV irradiated firefly luciferase reporter gene plasmid (UVC 1000 J/m²) obtained by host cell reactivation (n>6 in triplicates). Red lines mark the range between DNA-repair levels of empty vector, i.e. residual repair activity of the cells, and

WT-XPD, i.e. 100% repair capacity. (B) Dominant modulation of firefly luciferase reporter gene expression (without irradiation) via overexpression of XPD/ERCC2 BC/OC-associated variants was estimated in the transcriptionally-proficient but repair-deficient XPD/ERCC2-defective XP6BE cells. Black bars indicate the mean relative reporter gene expression (in %, empty vector control was set to 100%), obtained by CMV-promotor driven basal transcription ($n > 6$ in triplicates). Error bars indicate the standard error of the mean. Significance levels were calculated, after pairwise testing for normal distribution of the values, using appropriate statistical tests for comparison of two groups (T-Test or U-Test, # = reference group, *** = $p < 0.001$, ** = $p < 0.01$, * = $p < 0.05$, n.s. = not significant). (C) Additional characteristics of the mutations tested for repair efficiency and transcriptional activity.

doi:10.1371/journal.pgen.1006248.g002

similar frequencies in the ethnically matching control cohorts. The only exception so far is the Lithuanian mutation p.(Arg487Trp), which was found 4 times in the Lithuanian BC/OC cohort and not (yet) in the corresponding control cohort (Table 2). With just above 100 individuals this cohort is way too small to be of any statistical relevance. Therefore, the acquisition of additional samples is mandatory. But even in this very early phase of variant (de-)validation it becomes evident that regionally matching control cohorts—as small as they may be—are superior to any huge global cohort. Since genotypic data allow to locate the geographic origin of a given individual within a few hundred kilometers [23], the term “regionally matching” should be defined as “less than ca. 300 km distance from the recruitment center”. As a consequence, regionally matching controls are even superior to population-specific controls, because populations do mix, especially in regions close to national borders. The p.Phe568fs mutation, for example, has been seen only once in a German BC/OC index case and never in the 1844 German controls. Based on population-specific data we would have been very excited about this finding. But the German case was recruited in Dresden, close to the Czech border, and in Prague, 118 km away, the same mutation has been found twice in a small control cohort of only 105 non-cancer females. This underlines the importance of regional controls and multi-national studies for reliable variant validation.

ERCC2 mutations with tumorigenic relevance are probably located in very small and scattered areas of the protein

Due to its involvement in DNA repair and due to encoding a helicase like *RECQL* [7], *ERCC2* is a plausible gene candidate for familial cancer susceptibility. Bi-allelic mutations in *ERCC2*, however, can cause the cancer-prone disease XPD as well as the “non-cancer”-disease TTD [27] and there is no evident genotype-phenotype correlation [19]. The pathogenic p.(Arg112His) mutation, for example, has been identified in TTD patients as well as in a patient with major features of XPD [19]. Furthermore, impairment of DNA repair capacity is not correlated with tumor burden: the mutation p.(Phe568Tyrfs), for example, has been identified in non-cancer TTD patients twice, but not once in cancer-prone XPD patients, although this study (Fig 2) as well as a previous study [19] clearly show diminished repair capability of this frameshift variant. From these observations we have to conclude that a limited subset of mutations in *ERCC2* might predispose to cancer but these mutations are not likely to cluster in a defined area of the gene nor do they necessarily affect a specific sub-function of the ERCC2 protein. Therefore, cancer predisposing *ERCC2* mutations are very likely to be discovered only on the basis of familial co-segregation with cancer and overrepresentation in cancer cohorts vs. region-specific controls.

The incidence of ERCC2-related diseases is not in line with the frequency of deleterious founder mutations in the corresponding populations

Although the founder mutations tested in this study may not predispose to BC/OC they still confer carrier status for the recessive disorders XPD (OMIM 278730), TTD (OMIM 601675)

Table 2. ERCC2 allele frequencies (%) in BC/OC patients and corresponding control cohorts. The allele frequency is counted on the basis of sample size (in brackets) and number of observed cases (see Table 1) with hetero- and homozygosity.

AA / nt change (N = 25)	CZ BC/OC [353] ^a	CZ Ctrl [453] ^b	LT BC/OC [405]	LT Ctrl [103]	GE BC/OC [587] ^c	GE Ctrl [1844] ^d	ExAc vers. 0.2 [variable] ^e
Pro13Ser	0	0	0	0	0.0851	0	0
Val77Alafs	0.1416	0	0	0	0	0	0
Arg166Cys	0	0	0.2469	0	0	0	0
Glu167Gln	0	0	0	0	0.0851	0	0.0033
Gly188Ala	0	0	0	0	0.0851	0	0
Pro215Leu	0	0	0.1234	0	0	0	0
Arg280His	0	0	0.1234	0	0	0	0.0072
Gln316Glu	0	0	0	0	0.0851	0	0.0152
Asp423Asn	0.1416	0.1104	0.9876	1.456	0	0.0542	0.0248
Arg450His	0	0	0	0	0.1704	0.0813	0.0214
Leu461Val	0	0	0	0	0.2553	0.1356	0.1345
Arg487Trp	0	0	0.4938	0	0	0	0.0034
Asp513Tyr	0	0	0	0	0.0851	0	0
Val536Met	0	0	0	0	0.1704	0	0.0231
p.Phe568fs	0.4249	0.4415	0.1234	0	0.0851	0	0.0093
Glu576Lys	0	0	0	0	0.0851	0.0542	0.0008
Arg592His	0	0	0.8642	0	0.0851	0	0.0332
Arg601Gln	0	0.1104	0	0	0.1704	0.0542	0.0175
Val611Ala	0.1416	0	0	0	0	0	0.0042
Arg631Cys	0	0	0.1234	0	0.0851	0	0.0025
c.1903-2A>G	0	0	0	0	0.0851	0	0
Val678Leu	0	0	0.1234	0	0	0	0
c.2150C>G	0	0	0	0	0.2553	0.0813	0.0349
Arg722Gln	0.1416	0	0	0	0	0	0.0067
p.Ser746fs	0	0	0	0	0.0851	0	0

BC/OC = index cases with breast- and/or Ovarian cancer; Ctrl = healthy or non-cancer related individuals; CZ = Czech Republic, GE = Germany, LT = Lithuania; AA = Amino acid; nt = nucleotide; ExAC = Exome Aggregation Consortium, Cambridge, MA (URL: <http://exac.broadinstitute.org>) [accessed May 2015];

^a 28 samples from Brno (TruSight-Cancer) + 325 samples from Prague [24,25] (custom panel with 581 genes);

^b 105 female non-cancer samples from Prague [25,26] (custom panel with 581 genes) + 108 female non-cancer samples from Brno, sequenced in pools with the TruSight-Cancer panel + 240 non-cancer samples from Prague, sequenced in pools with the TruSight-Cancer panel;

^c 271 samples from Dresden + 316 samples from Munich (MGZ), all sequenced with the TruSight-Cancer panel;

^d 1629 individual exome samples from the Cologne Center for Genomics (CCG) + 79 individual non-BC/OC TruSight-One samples from Dresden + 136 individual non-BC/OC TruSight-Cancer samples from Dresden and Munich (MGZ);

^e Since the exome data have been collected from various sources with various enrichment strategies, the sample size varies for each variant. Each allele frequency has been calculated with the corresponding sample size for that allele.

doi:10.1371/journal.pgen.1006248.t002

and COFS2 (OMIM 610756). Even the TTD-causing mutation p.(Phe568fs) alone has been detected in 7 of 806 samples from the Czech Republic (CZ), i.e. the frequency of heterozygous carriers of this mutation is approx. 0.86%. According to Hardy-Weinberg equilibrium model, this would result in a TTD incidence of 1/30.000. Based on combined data from the DNA repair diagnostic centers in France, West-Germany, Italy, the Netherlands and the United Kingdom the actual incidence for TTD is 1.2 per million [28]. Since it is reasonable to assume

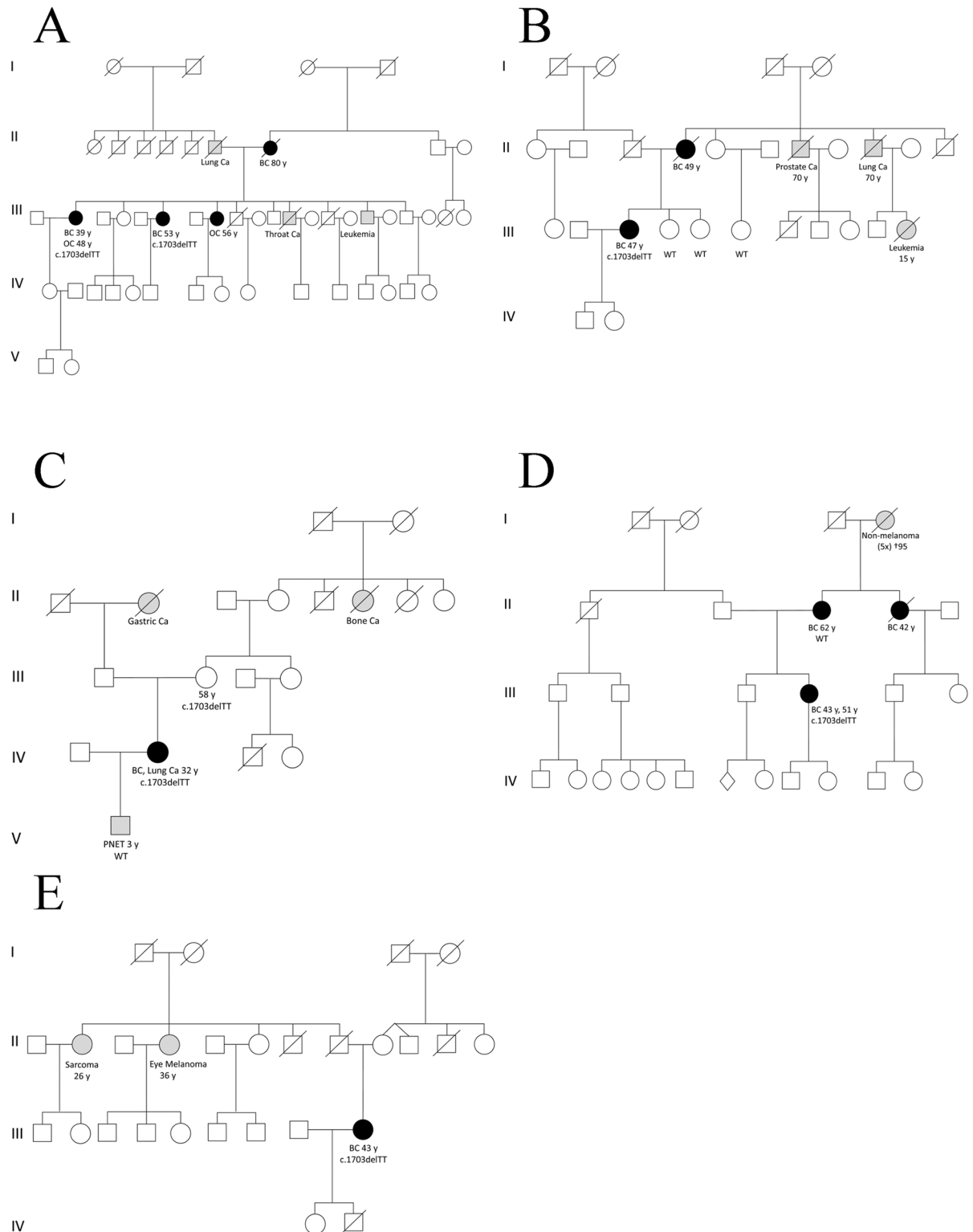


Fig 3. ERCC2 frameshift mutation c.1703_1704delTT (p.Phe568fs) in familial breast and ovarian cancer pedigrees. Individuals with breast cancer (BC), ovarian cancer (OC) or both (BC, OC) are shown as circles filled in black. Individuals tested positive for the familial mutation are indicated in detail; those with WT (wild-type) have been tested negative. All affected individuals with BC or OC not tested for germline mutations in ERCC2 were either deceased or refused testing. (A) German, (B) Lithuanian and (C-E) Czech pedigrees.

doi:10.1371/journal.pgen.1006248.g003

that (i) a TTD incidence of 1/30.000 would not be missed by the clinical geneticists in CZ and (ii) the publications reporting p.(Phe568fs) as TTD-causing [9,19] are not wrong, there is one logical explanation for the discrepancy between allele frequency and disease incidence: homozygosity for p.Phe568fs is embryonic lethal. This is in-line with the observation that complete loss of ERCC2 activity is not compatible with life in homozygous knock-out mice [29] and it is also consistent with the observation that all XPD and TTD patients tested so far have residual ERCC2 activity [30]. Since an elevated TTD/XPD incidence has not been reported in Lithuania either, we can assume that homozygosity of the frequent Lithuanian founder mutation p.(Asp423Asn) (Table 2), which clearly displayed functional deficiency in our experiments (Fig 2), is embryonic lethal as well.

In conclusion, this multi-national study of *ERCC2* mutations in patients with familial BC/OC and regionally matching controls identified and functionally verified a broad spectrum of unique and recurrent *ERCC2* mutations. Although the frequent founder mutations are not very likely to predispose to BC/OC, some mutations, like p.(Val77Alafs), that are unique to the BC/OC cohort are worth to be considered in future large-scale association studies.

Materials and Methods

Ethics statement

Informed written consent was obtained from all patients and the study was approved by the Local Research Ethics Committee (EK 162072007).

Subjects, families and pedigrees

We enrolled affected individuals from 587 German BC and BC/OC pedigrees with hereditary gynecological malignancies through a genetic counseling program at two centers (Dresden, Munich) from the “German Consortium for hereditary breast- and ovarian cancer” (GC-HBOC) and at the Medical Genetics Center (MGZ) in Munich. Additional 131 BC- and 136 BC/OC families were collected at the Vilnius University Hospital Santariskiu Klinikos in Vilnius, Lithuania and 28 BC/OC families were gathered in the Czech Republic at Brno. The Czech Prague subgroup involved 325 BC patients negatively tested for presence of pathogenic *BRCA1* and *BRCA2* variants [24] and 105 non-cancer controls analyzed as described recently [25,26], and additional 240 controls [26] sequenced in pools. The BC pedigrees fulfilled the criterion that at least three affected females with breast cancer but no ovarian cancers were present (breast cancer pedigrees). In the BC/OC pedigrees, at least one case of breast and one ovarian cancer had occurred. All individuals with variant *ERCC2* alleles were checked for mutations in 10 BC/OC core genes defined by GC-HBOC (*ATM*, *BRCA1*, *BRCA2*, *CDH1*, *CHEK2*, *NBN*, *PALB2*, *RAD51C*, *RAD51D* and *TP53*). Informed consent was obtained from all people participating in the study, and the experiments were approved by the ethics committees of the institutions contributing to this project.

TruSight-Cancer panel sequencing

DNA was obtained from peripheral blood of all patients. For panel enrichment approximately 85 ng genomic DNA was required. We used the TruSight Cancer Illumina kit (Illumina), which targets the coding sequences of 94 genes associated with a predisposition towards cancer (S1 Table), following the manufacturer's instructions. Sequencing was carried out on an Illumina MiSeq instrument as 150 bp paired-end runs with V2 chemistry. Reads were aligned to the human reference genome (GRCh37/hg19) using BWA (v 0.7.8-r455) with standard parameters. Duplicate reads and reads that did not map unambiguously were removed. The

percentage of reads overlapping targeted regions and coverage statistics of targeted regions were calculated using Shell scripts. Single-nucleotide variants and small insertions and deletions (INDELs) were called using SAMtools (v1.1). We used the following parameters: a maximum read depth of 10000 (parameter -d), a maximum per sample depth of 10000 for INDEL calling (parameter -L), adjustment of mapping quality (parameter -C) and recalculation of per-Base Alignment Quality (parameter -E). Additionally, we required putative SNVs to fulfill the following criteria: a minimum of 20% of reads showing the variant base and the variant base is indicated by reads coming from different strands. For INDELs we required that at least 15% of reads covering this position indicate the INDEL. Variant annotation was performed with snpEff (v 4.0e) and Alamut-Batch (v 1.3.1) based on the RefSeq database. Only variants (SNVs/small INDELs) in the coding region and the flanking intronic regions (± 15 bp) were evaluated.

Custom breast cancer panel sequencing

The data related to the ERCC2 gene in this study were retrieved from the custom-made gene panel sequencing analysis described recently [25]. Briefly, genomic DNA was obtained from a peripheral blood of 325 BC Czech patients from the Prague area that were negatively tested for a presence of pathogenic variants in the *BRCA1* or *BRCA2* gene previously [24]. The frequency of population-specific variants was assessed by a concurrent analysis of 105 control DNAs obtained from non-cancer individuals [26]. One μg of genomic DNA was used for library construction. The DNA was fragmented by ultrasonication and edited for SOLiD sequencing. Target DNA enrichment was performed by a custom solution-based sequence capture (SeqCap EZ Choice Library, Roche) according to the NimbleGenSeqCap EZ Library SR User's Guide (Version 4.2, Roche). Five hundred and ninety targeted genes include 141 genes that code for known proteins involved in DNA repair and DNA damage response pathways, and an additional set of genes retrieved from Phenopedia at HuGE Navigator16 web site associated with "breast neoplasms" (assessed February 2012). Captured libraries were sequenced on SOLiD4 system. Finally, exonic regions of 581 genes were captured successfully with sufficient coverage. Reads were aligned to the human reference genome (GRCh37/hg19) using Novoalign (CS 1.01.08) with standard parameters. Conversion of SAM to BAM format was performed with SAMtools (0.1.8). Single-nucleotide variants and small insertions and deletions (INDELs) were called using SAMtools (0.1.8). Variant annotation was performed with ANNOVAR [31]. For final evaluations, small INDELs, intronic variants flanking ± 2 bp to exon borders, and rare SNPs (presented in 1000 genome or exome sequencing (ESP) projects with frequency $< 1\%$) were considered.

Sanger sequencing

Validation of *ERCC2* variants in probands and family members was performed by classical Sanger sequencing. Additional DNAs from 8 HBOC patients affected by malignant melanoma (5 cases) or presence of melanoma in other family members (3 cases) were analyzed for the complete *ERCC2* coding region. *ERCC2* exons were amplified with intronic primers (S2 Table) and sequenced using the ABI Prism Terminator Cycle Sequencing Ready Reaction Kit (Applied Biosystems). Genomic DNA (50 ng) containing 1x PCR Master Mix (Qiagen) and 0.25 μM of each forward and reverse primers in 15 μl reaction volume was subjected to PCR amplification for 25 cycles (30 sec at 95°C, 30 sec at 64°C and 30 sec at 72°C).

Functional validation of *ERCC2* variants

Variant cloning. Wild type *ERCC2* cDNA was amplified from reverse transcribed mRNA isolated from fibroblasts derived from healthy donors (RevertAid H Minus First strand cDNA

synthesis kit; Thermo scientific, Waltham, MA, USA) using forward (5'TTAGGTACCATGA AGCTCAACGTGGACG) and reverse (5' TTATCTAGATCAGAGCTGCTGAGCAATCT) primers and cloned into the pJET1.2/blunt vector (CloneJET PCR Cloning Kit; Life technologies, Waltham, MA, USA). These primers carry *KpnI* and *XbaI* restriction sites to release *ERCC2* cDNA by double restriction enzyme digestion (Life technologies). The *ERCC2* cDNA was purified from agarose gels using the Wizard SV Gel and PCR Clean-Up System (Promega, Klaus, Austria) and cloned into the pcDNA3.1(+) mammalian expression vector (Life technologies) and subsequently transformed into DH5 α *E.coli* cells. Colony PCR (using T7 and M13 primers) and Sanger sequencing of the entire gene was performed using the BigDye Terminator v3.1 Cycle Sequencing Kit (Life technologies, for primers see [S3 Table](#)).

For generation of the *ERCC2* variants, site directed mutagenesis was applied using Phusion High-Fidelity DNA Polymerase (Life technologies) and specific primer pairs in either the classical protocol (for variants Ser746FS and D513Y, Stratagene) or an optimized site-directed method (all other variants, for primers see [S3 Table](#)). For the latter, template (100 ng *ERCC2* in pcDNA3.1(+)) was first subjected to dam methylation using dam methyl transferase (NEB, Frankfurt a. M., Germany). Afterwards, a first PCR was conducted with the forward-primer using Phusion polymerase (Life technologies) in a 2-Step PCR protocol with 5 minutes of annealing and elongation at 72°C for 18 cycles. Then over-night enzyme digestion with *DpnI* (Life technologies) was followed by ethanol precipitation. A second PCR using reverse primers (same conditions) was performed with this template and ethanol precipitated. The final reaction product was subject to transformation of DH5 α *E.coli* cells. Positive clones were verified by Sanger sequencing as described above.

Assay set-up. The host cell reactivation (HCR) assay measures the amount of nucleotide excision repair (NER) in actively transcribed genes. This dual reporter gene assay deploys the turnover rate of firefly luciferase substrate as readout for the NER capacity of host cells transfected with the (UV-) damaged reporter gene plasmid encoding for firefly luciferase [32]. HCR can be used for DNA repair capacity assessment of NER deficient host cells transfected with DNA repair gene variants as well as for measuring in situ transcription using non-irradiated firefly luciferase reporter gene plasmids [33,34].

ERCC2-deficient XP6BE-SV-immortalized fibroblasts were a generous gift of K.H. Kraemer (NIH, Bethesda, MD, USA) and harbor two differently mutated *ERCC2* alleles [p.Arg683Trp and an in-frame deletion of amino acids (AA) 36–61] [9]. XP6BE cells were transfected using Attractene Transfection Reagent (Qiagen, Hilden, Germany) according to the manufacture's advice, with plasmids coding for firefly luciferase (100 ng), renilla-luciferase (50 ng) and an empty pcDNA3.1(+) vector or XPD-variants cloned into the pcDNA3.1(+) expression vector (100 ng) (for cloning see above). The plasmid coding for firefly luciferase was divided into two fractions prior to transfection. One fraction was irradiated with 1000 J/m² of UVC light, a second fraction stayed untreated. The non-irradiated renilla-luciferase plasmid serves as an internal control for normalization of transfection efficacy.

After incubation of transfected XP6BE cells for one day (37°C, 5% CO₂), which allows sufficient repair of the UV-photoproducts and protein expression of the luciferases, cells were lysed and analyzed using Dual-Luciferase Reporter Assay System (Promega, Klaus, Austria). The luminescence measurements were performed in a white Glomax 96 microplate using the Glomax luminometer (Promega, Klaus, Austria).

The relative repair capacity is estimated using this formula:

$$repair (\%) = \frac{\text{mean (irradiated firefly/renilla per well)}}{\text{mean (unirradiated firefly/renilla per well)}} \times 100$$

The repair capacity of XP6BE cells transfected with the wild type *ERCC2* cDNA containing expression vector was set to 100%.

Transcriptional activity was calculated as the amount of firefly luciferase expression from non-irradiated plasmids in XP6BE cells transfected either with wild type *ERCC2* or breast cancer associated *ERCC2* variants containing expression vectors relative to the amount of firefly luciferase expression in XP6BE cells transfected with the empty expression vector. The latter was set to 100%. Every experiment (NER capacity as well as transcription) was conducted at least six times in triplicates.

Modeling of ERCC2 protein structure

Structural modeling of the ERCC2 variants. Homology modeling of the human ERCC2 protein was performed with SWISS-MODEL (ExPASy). The crystal structure of the ATP-dependent DNA helicase Ta0057 from *Thermoplasma acidophilum* (RCSB:4A15, UniProt: Q9HM14) was used as template structure for modeling. Predicted models for the residue changes of the detected missense mutations in *ERCC2* were displayed and analyzed using Visual Molecular Dynamics (VMD) (S1 Fig). The predicted models were superimposed onto the Ta0057 structure with the Multiseq tool integrated in VMD.

In-silico interpretation of missense variants. The probability of effect of non-synonymous mutations in *ERCC2* was calculated by the amino acid (AA) substitution prediction methods SIFT, PolyPhen2, Provean, Mutation Taster, MAPP, and AGVD (S4 Table). Based on these data, a summarizing rating was assessed (last column in S4 Table and Table 1). Distribution of PhyloP and Grantham scores [35] for dbSNP, ClinVar and all variants identified in *ERCC2* were analyzed. Statistical probability scores of PhyloP and Grantham scores and analysis of distribution plots are provided (S4 Table and S2 Fig).

AA conservation alignment

A multiple alignment of ERCC2 AA sequences was done according to HomoloGene (NCBI) in order to assess the AA conservation of the detected variants in 20 species with homologous proteins (S3 Fig).

Supporting Information

S1 Fig. ERCC2 domain structure and overlay of ERCC2 missense mutations Arg478Trp and Asp423Asn.

A) Schematic showing the domain structure and canonical motifs of human ERCC2. Helicase motor domains HD1 (blue) and HD2 (green) form the DNA ATP-binding interface. The FeS (orange) and the Arch (purple) domains are inserted into HD1. The boundaries of the FeS cluster binding domain are indicated by red spheres. The human enzyme C-terminal (grey) extension (CTE) is indicated in grey. Domain boundaries are indicated by residue numbers. B,C) 3D representation of the native (cyan) and mutant (pink) overlaid ERCC2 protein structures show a detailed structural environment of the wild-type (green), Arg487 and Asp423 residues in comparison to the Arg487Trp and Asp423Asn mutants (red). Surrounding amino acids (AAs) are indicated as licorice. (B) Note the significant changes in the AA constellations Arg424, Thr425 induced by the by the Asp423Asn replacement. (C) The Arg487Trp AA replacement introduces a tryptophan residue which protrudes beyond the protein surface and might destabilize the interactions with the surrounding AAs His700, Glu690 and Leu701 within the protein loop.

(TIF)

S2 Fig. Distribution of PhyloP and CADD scores for 1000G, ClinVar and the mutations identified in this study in the ERCC2 gene. A) Evolutionary conservations (PhyloP) and Combined Annotation Dependent Depletion (CADD) scores are represented for all non-synonymous ERCC2 variants found in BC/OC patients. Blue: Variants with no significant functional effect; Red: variants which showed a deleterious functional effect by no complementation of NER-deficient cells and/or negative modulation of transcription; Green: variants not tested. B) This analysis was further extended to analyze these combined scores for all non-synonymous variants reported in 1000G and ClinVar with no reported clinical significance (Class 1–3), or ClinVar reported pathogenic variants (Class 4–5) to visualize the probability for the ERCC2 variants which have not been functionally tested to be pathogenic or benign. Heat maps show the distribution and frequency for the combined PhyloP and CADD scores in 1000G and ClinVar. Red colors indicate a low frequency and green colors a high frequency. ERCC2 variants showing no functional pathogenic effect (circle), pathogenic variants with NER complementation failure and/or negative modulation of transcription (triangles), and variants not tested in our functional studies (black square) are represented. ERCC2 variants with deleterious functional effects show a better overlap with ClinVar pathogenic variants (Class 4–5) by their location mostly restricted to dark green and yellow as indicated. In contrast, location of variants shows within the dark red plot region when compared to 1000G and ClinVar (Class 1–3). Variants not included in our functional studies show a similar distribution pattern as functional deleterious variants which overlaps with ClinVar pathogenic variants (Class 4–5). In total, most of the ERCC2 variants are located in areas of high conservation and high deleteriousness. Statistical probability scores for these analyses are provided in [S4 Table](#). PhyloP and CADD scores for 1000G and ClinVar variants were obtained from the annotation browser SNIIPA [\[36\]](#).

(TIF)

S3 Fig. ERCC2 amino acid (AA) sequence alignment. Multiple sequence alignment of protein regions from various species surrounding the identified human ERCC2 missense variants ([S4 Table](#)). Affected residues are indicated in red letters. The dotted lines correspond to sequence gaps or sequence regions not yet available. Except Glu167, all affected residues showed strong conservation across vertebrates (Arg166, Gly188, Arg280, Gln316, Asp423, Leu461, Arg487, Val611, Val678, Ala717, Arg722) or even across all species (Pro13, Pro215, Arg450, D513, Val536, Glu576, Arg592, Arg601, Arg631). The AA variability at codon 167 is in line with the results of the effect prediction algorithms which predict the Glu167Gln replacement as benign ([S4 Table](#)). Accession number of the ERCC2 protein sequences used for AA sequence comparison are as follows: Homo sapiens (NP_000391.1); Pan troglodytes (NP_001233519.1); Macaca mulatta (XP_002808245.1); Canis lupus (XP_541562.3); Bos taurus (NP_001096787.1); Mus musculus (NP_031975.2); Rattus norvegicus (NP_001166280.1); Xenopus tropicalis (NP_001008131.1); Danio rerio (NP_957220.1); Drosophila melanogaster (NP_726036.2); Anopheles gambiae (XP_311900.4); Caenorhabditis elegans (NP_497182.2); Saccharomyces cerevisiae (NP_011098.3); Kluyveromyces lactis (XP_452994.1); Eremothecium gossypii (NP_986780.1); Schizosaccharomyces pombe (NP_593025.1); Magnaporthe oryzae (XP_003716866.1); Neurospora crassa (XP_956536.2); Arabidopsis thaliana (NP_171818.1); and Oryza sativa (NP_001054627.1).

(TIF)

S1 Table. Genes covered by the TruSight-Cancer gene panel.

(DOCX)

S2 Table. ERCC2/XPD primers for Sanger-validation of NGS derived mutations and analysis of familial segregation. All sequences shown in 5' → 3' direction. XPD = alternate name of ERCC2.

(DOCX)

S3 Table. Primer pairs used for PCR amplification, Sanger sequencing, and site directed mutagenesis of ERCC2 variants. All primers (de-salted and deprotected) were synthesized by Sigma-Aldrich (Taufkirchen, Germany).

(DOCX)

S4 Table. Effect prediction of ERCC2 missense variants. The probability of effect of non-synonymous mutations in ERCC2 was predicted by the computer programs: SIFT, Sorting Invariant from Tolerated (Score under 0,05: not tolerated; Range 0–1); PolyPhen-2, Classification following PSIC scores (HumVar, "benign"- "possibly damaging"—"probably damaging", Range: 0–1); Provean, Protein Variation Effect Analyze; MAPP, Multivariate Analysis of Protein Polymorphism; Align-GVGD, Scores (C0, C15, C25, C35, C45, C55, C65) from C0 (likely benign) to C65 (likely pathogenic); CADD, Combined Annotation Dependent Depletion [12]. Dlt = deleterious, PrD = probably damaging, PsD = possibly damaging, Bgn = benign, Ntr = neutral. Conservation was calculated with PhyloP (Score range from -14.1 to 6.4). Grantham [35] distance scores (Range 0–215). AA exchanges in gray background are located in cis and form a haplotype.

(DOCX)

Acknowledgments

We thank all the families for providing samples for this study. We are deeply indebted to the German Cancer Aid for establishing the German Consortium for Hereditary Breast- and Ovarian Cancer (GC-HBOC).

References

1. Couch FJ, Nathanson KL, Offit K (2014) Two decades after BRCA: setting paradigms in personalized cancer care and prevention. *Science* 343: 1466–1470. doi: [10.1126/science.1251827](https://doi.org/10.1126/science.1251827) PMID: [24675953](https://pubmed.ncbi.nlm.nih.gov/24675953/)
2. Turnbull C, Rahman N (2008) Genetic predisposition to breast cancer: past, present, and future. *Annu Rev Genomics Hum Genet* 9: 321–345. doi: [10.1146/annurev.genom.9.081307.164339](https://doi.org/10.1146/annurev.genom.9.081307.164339) PMID: [18544032](https://pubmed.ncbi.nlm.nih.gov/18544032/)
3. Antoniou AC, Easton DF (2006) Models of genetic susceptibility to breast cancer. *Oncogene* 25: 5898–5905. PMID: [16998504](https://pubmed.ncbi.nlm.nih.gov/16998504/)
4. Easton DF, Pharoah PD, Antoniou AC, Tischkowitz M, Tavtigian SV, et al. (2015) Gene-panel sequencing and the prediction of breast-cancer risk. *N Engl J Med* 372: 2243–2257. doi: [10.1056/NEJMs1501341](https://doi.org/10.1056/NEJMs1501341) PMID: [26014596](https://pubmed.ncbi.nlm.nih.gov/26014596/)
5. Evans BJ, Burke W, Jarvik GP (2015) The FDA and genomic tests—getting regulation right. *N Engl J Med* 372: 2258–2264. doi: [10.1056/NEJMs1501194](https://doi.org/10.1056/NEJMs1501194) PMID: [26014592](https://pubmed.ncbi.nlm.nih.gov/26014592/)
6. Zhang G, Zeng Y, Liu Z, Wei W (2013) Significant association between Nijmegen breakage syndrome 1 657del5 polymorphism and breast cancer risk. *Tumour Biol* 34: 2753–2757. doi: [10.1007/s13277-013-0830-z](https://doi.org/10.1007/s13277-013-0830-z) PMID: [23765759](https://pubmed.ncbi.nlm.nih.gov/23765759/)
7. Cybulski C, Carrot-Zhang J, Kluzniak W, Rivera B, Kashyap A, et al. (2015) Germline RECQL mutations are associated with breast cancer susceptibility. *Nat Genet*.
8. Singh A, Compe E, Le May N, Egly JM (2015) TFIIF Subunit Alterations Causing Xeroderma Pigmentosum and Trichothiodystrophy Specifically Disturb Several Steps during Transcription. *Am J Hum Genet* 96: 194–207. doi: [10.1016/j.ajhg.2014.12.012](https://doi.org/10.1016/j.ajhg.2014.12.012) PMID: [25620205](https://pubmed.ncbi.nlm.nih.gov/25620205/)
9. Takayama K, Salazar EP, Lehmann A, Stefanini M, Thompson LH, et al. (1995) Defects in the DNA repair and transcription gene ERCC2 in the cancer-prone disorder xeroderma pigmentosum group D. *Cancer Res* 55: 5656–5663. PMID: [7585650](https://pubmed.ncbi.nlm.nih.gov/7585650/)

10. Lehmann J, Schubert S, Emmert S (2014) Xeroderma pigmentosum: diagnostic procedures, interdisciplinary patient care, and novel therapeutic approaches. *J Dtsch Dermatol Ges* 12: 867–872. doi: [10.1111/ddg.12419](https://doi.org/10.1111/ddg.12419) PMID: [25262888](https://pubmed.ncbi.nlm.nih.gov/25262888/)
11. Pabalan N, Francisco-Pabalan O, Sung L, Jarjanazi H, Ozcelik H (2010) Meta-analysis of two ERCC2 (XPD) polymorphisms, Asp312Asn and Lys751Gln, in breast cancer. *Breast Cancer Res Treat* 124: 531–541. doi: [10.1007/s10549-010-0863-6](https://doi.org/10.1007/s10549-010-0863-6) PMID: [20379847](https://pubmed.ncbi.nlm.nih.gov/20379847/)
12. Kircher M, Witten DM, Jain P, O’Roak BJ, Cooper GM, et al. (2014) A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* 46: 310–315. doi: [10.1038/ng.2892](https://doi.org/10.1038/ng.2892) PMID: [24487276](https://pubmed.ncbi.nlm.nih.gov/24487276/)
13. Fan L, Fuss JO, Cheng QJ, Arvai AS, Hammel M, et al. (2008) XPD helicase structures and activities: insights into the cancer and aging phenotypes from XPD mutations. *Cell* 133: 789–800. doi: [10.1016/j.cell.2008.04.030](https://doi.org/10.1016/j.cell.2008.04.030) PMID: [18510924](https://pubmed.ncbi.nlm.nih.gov/18510924/)
14. Kuper J, Braun C, Elias A, Michels G, Sauer F, et al. (2014) In TFIIH, XPD helicase is exclusively devoted to DNA repair. *PLoS Biol* 12: e1001954. doi: [10.1371/journal.pbio.1001954](https://doi.org/10.1371/journal.pbio.1001954) PMID: [25268380](https://pubmed.ncbi.nlm.nih.gov/25268380/)
15. Liu H, Rudolf J, Johnson KA, McMahon SA, Oke M, et al. (2008) Structure of the DNA repair helicase XPD. *Cell* 133: 801–812. doi: [10.1016/j.cell.2008.04.029](https://doi.org/10.1016/j.cell.2008.04.029) PMID: [18510925](https://pubmed.ncbi.nlm.nih.gov/18510925/)
16. Botta E, Nardo T, Broughton BC, Marinoni S, Lehmann AR, et al. (1998) Analysis of mutations in the XPD gene in Italian patients with trichothiodystrophy: site of mutation correlates with repair deficiency, but gene dosage appears to determine clinical severity. *Am J Hum Genet* 63: 1036–1048. PMID: [9758621](https://pubmed.ncbi.nlm.nih.gov/9758621/)
17. Viprakasit V, Gibbons RJ, Broughton BC, Tolmie JL, Brown D, et al. (2001) Mutations in the general transcription factor TFIIH result in beta-thalassaemia in individuals with trichothiodystrophy. *Hum Mol Genet* 10: 2797–2802. PMID: [11734544](https://pubmed.ncbi.nlm.nih.gov/11734544/)
18. Zhou X, Khan SG, Tamura D, Patronas NJ, Zein WM, et al. (2010) Brittle hair, developmental delay, neurologic abnormalities, and photosensitivity in a 4-year-old girl. *J Am Acad Dermatol* 63: 323–328. doi: [10.1016/j.jaad.2010.03.041](https://doi.org/10.1016/j.jaad.2010.03.041) PMID: [20633800](https://pubmed.ncbi.nlm.nih.gov/20633800/)
19. Broughton BC, Berneburg M, Fawcett H, Taylor EM, Arlett CF, et al. (2001) Two individuals with features of both xeroderma pigmentosum and trichothiodystrophy highlight the complexity of the clinical outcomes of mutations in the XPD gene. *Hum Mol Genet* 10: 2539–2547. PMID: [11709541](https://pubmed.ncbi.nlm.nih.gov/11709541/)
20. Nishiwaki Y, Kobayashi N, Imoto K, Iwamoto TA, Yamamoto A, et al. (2004) Trichothiodystrophy fibroblasts are deficient in the repair of ultraviolet-induced cyclobutane pyrimidine dimers and (6–4)photo-products. *J Invest Dermatol* 122: 526–532. PMID: [15009740](https://pubmed.ncbi.nlm.nih.gov/15009740/)
21. Cleaver JE, Thompson LH, Richardson AS, States JC (1999) A summary of mutations in the UV-sensitive disorders: xeroderma pigmentosum, Cockayne syndrome, and trichothiodystrophy. *Hum Mutat* 14: 9–22. PMID: [10447254](https://pubmed.ncbi.nlm.nih.gov/10447254/)
22. Taylor EM, Broughton BC, Botta E, Stefanini M, Sarasin A, et al. (1997) Xeroderma pigmentosum and trichothiodystrophy are associated with different mutations in the XPD (ERCC2) repair/transcription gene. *Proc Natl Acad Sci U S A* 94: 8658–8663. PMID: [9238033](https://pubmed.ncbi.nlm.nih.gov/9238033/)
23. Novembre J, Johnson T, Bryc K, Kutalik Z, Boyko AR, et al. (2008) Genes mirror geography within Europe. *Nature* 456: 98–101. doi: [10.1038/nature07331](https://doi.org/10.1038/nature07331) PMID: [18758442](https://pubmed.ncbi.nlm.nih.gov/18758442/)
24. Pohlreich P, Zikan M, Stribrna J, Kleibl Z, Janatova M, et al. (2005) High proportion of recurrent germline mutations in the BRCA1 gene in breast and ovarian cancer patients from the Prague area. *Breast Cancer Res* 7: R728–736. PMID: [16168118](https://pubmed.ncbi.nlm.nih.gov/16168118/)
25. Lhota F, Zemankova P, Kleiblova P, Soukupova J, Vocka M, et al. (2016) Hereditary truncating mutations of DNA repair and other genes in BRCA1/BRCA2/PALB2-negatively tested breast cancer patients. *Clin Genet*.
26. Kleibl Z, Havranek O, Novotny J, Kleiblova P, Soucek P, 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: 159–164. PMID: [18058223](https://pubmed.ncbi.nlm.nih.gov/18058223/)
27. Berneburg M, Clingen PH, Harcourt SA, Lowe JE, Taylor EM, et al. (2000) The cancer-free phenotype in trichothiodystrophy is unrelated to its repair defect. *Cancer Res* 60: 431–438. PMID: [10667598](https://pubmed.ncbi.nlm.nih.gov/10667598/)
28. Kleijer WJ, Laugel V, Berneburg M, Nardo T, Fawcett H, et al. (2008) Incidence of DNA repair deficiency disorders in western Europe: Xeroderma pigmentosum, Cockayne syndrome and trichothiodystrophy. *DNA Repair (Amst)* 7: 744–750.
29. de Boer J, Donker I, de Wit J, Hoeijmakers JH, Weeda G (1998) Disruption of the mouse xeroderma pigmentosum group D DNA repair/basal transcription gene results in preimplantation lethality. *Cancer Res* 58: 89–94. PMID: [9426063](https://pubmed.ncbi.nlm.nih.gov/9426063/)
30. Lehmann AR (2001) The xeroderma pigmentosum group D (XPD) gene: one gene, two functions, three diseases. *Genes Dev* 15: 15–23. PMID: [11156600](https://pubmed.ncbi.nlm.nih.gov/11156600/)

31. Wang K, Li M, Hakonarson H (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 38: e164. doi: [10.1093/nar/gkq603](https://doi.org/10.1093/nar/gkq603) PMID: [20601685](https://pubmed.ncbi.nlm.nih.gov/20601685/)
32. Protic-Sabljić M, Kraemer KH (1985) One pyrimidine dimer inactivates expression of a transfected gene in xeroderma pigmentosum cells. *Proc Natl Acad Sci U S A* 82: 6622–6626. PMID: [2995975](https://pubmed.ncbi.nlm.nih.gov/2995975/)
33. Schafer A, Schubert S, Gratchev A, Seebode C, Apel A, et al. (2013) Characterization of three XPG-defective patients identifies three missense mutations that impair repair and transcription. *J Invest Dermatol* 133: 1841–1849. doi: [10.1038/jid.2013.54](https://doi.org/10.1038/jid.2013.54) PMID: [23370536](https://pubmed.ncbi.nlm.nih.gov/23370536/)
34. Khobta A, Anderhub S, Kitsera N, Epe B (2010) Gene silencing induced by oxidative DNA base damage: association with local decrease of histone H4 acetylation in the promoter region. *Nucleic Acids Res* 38: 4285–4295. doi: [10.1093/nar/gkq170](https://doi.org/10.1093/nar/gkq170) PMID: [20338881](https://pubmed.ncbi.nlm.nih.gov/20338881/)
35. Grantham R (1974) Amino acid difference formula to help explain protein evolution. *Science* 185: 862–864. PMID: [4843792](https://pubmed.ncbi.nlm.nih.gov/4843792/)
36. Arnold M, Raffler J, Pfeufer A, Suhre K, Kastenmuller G (2015) SNIIPA: an interactive, genetic variant-centered annotation browser. *Bioinformatics* 31: 1334–1336. doi: [10.1093/bioinformatics/btu779](https://doi.org/10.1093/bioinformatics/btu779) PMID: [25431330](https://pubmed.ncbi.nlm.nih.gov/25431330/)

PŘÍLOHA IV-4:

Borecka *et al.*; The c.657del5 variant in the *NBN* gene predisposes to pancreatic cancer



Research paper

The c.657del5 variant in the *NBN* gene predisposes to pancreatic cancer

Marianna Borecka^a, Petra Zemankova^a, Filip Lhota^a, Jana Soukupova^a, Petra Kleiblova^a, Michal Vocka^b, Pavel Soucek^c, Ivana Ticha^d, Zdenek Kleibl^a, Marketa Janatova^{a,*}

^a Institute of Biochemistry and Experimental Oncology, First Faculty of Medicine, Charles University in Prague, Prague, Czech Republic

^b Department of Oncology, First Faculty of Medicine, Charles University in Prague and General University Hospital in Prague, Prague, Czech Republic

^c Department of Toxicogenomics, National Institute of Public Health, Prague, Czech Republic

^d Department of Pathology, First Faculty of Medicine, Charles University in Prague and General University Hospital in Prague, Prague, Czech Republic

ARTICLE INFO

Article history:

Received 10 March 2016

Received in revised form 22 April 2016

Accepted 29 April 2016

Available online 2 May 2016

Keywords:

Pancreatic ductal adenocarcinoma

The *NBN* gene

Inherited risk

ABSTRACT

Pancreatic ductal adenocarcinoma (PDAC) is the sixth most frequent cancer type in the Czech Republic with a poor prognosis that could be improved by an early detection and subsequent surgical treatment combined with chemotherapy. Genetic factors play an important role in PDAC risk. We previously identified one PDAC patient harboring the Slavic founder deleterious mutation c.657del5 in the *NBN* gene, using a panel next-generation sequencing (NGS). A subsequent analysis of 241 unselected PDAC patients revealed other mutation carriers. The overall frequency of c.657del5 in unselected PDAC patients (5/241; 2.07%) significantly differed from that in non-cancer controls (2/915; 0.2%; $P = 0.006$). The result indicates that the *NBN* c.657del5 variant represents a novel PDAC-susceptibility allele increasing PDAC risk (OR = 9.7; 95% CI: 1.9 to 50.2). The increased risk of PDAC in follow-up recommendations for *NBN* mutation carriers should be considered if other studies also confirm an increased frequency of c.657del5 carriers in PDAC patients from other populations.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Pancreatic ductal adenocarcinoma (PDAC) is the sixth most frequent cancer type (with an incidence of 19.6/100,000 persons in 2013) and the fifth most frequent cause of cancer death in the Czech Republic (www.svod.cz). The prognosis of PDAC is poor with a 5-year survival of 7% and a median survival of 6 months (Siegel et al., 2015). Early detection and subsequent surgical treatment combined with chemotherapy can improve the 5-year survival up to 40% (Nakao et al., 2006). While population screening is not rational due to the low PDAC incidence, the identification of high-risk individuals, who may benefit from the available screening methods, is desirable.

A genetic predisposition is the major endogenous risk factor of PDAC development, together with chronic pancreatitis and diabetes mellitus

(Becker et al., 2014). It has been estimated that 5–10% of PDAC patients have a positive family PDAC history. The genetic basis of most familial PDAC cases has not been explained yet; however, several PDAC-susceptibility genes have been identified, including genes (*BRCA1*, *BRCA2*, *PALB2*, *MLH1*, *MSH2*, *MSH6*, *PMS2*, *STK11*, *APC*, *CDKN2A*) associated with hereditary cancer syndromes (reviewed in (Becker et al., 2014)). The protein products of numerous PDAC-susceptibility genes are directly involved in DNA repair and the DNA damage response. The most prevalent mutations have been identified in *BRCA2* (up to 6% of patients and increasing PDAC risk 3.5-fold (Couch et al., 2007)) and *PALB2* (3% of patients (Jones et al., 2009)). Their protein products share a common functional role in the DNA double-strand break (DDSB) repair. The *NBN* gene encodes nibrin, a protein participating in the formation of the multiprotein MRN (*MRE11-RAD50-NBN*) complex, an inevitable sensor of DNA damage in the DDSB repair (Carney et al., 1998). Biallelic *NBN* mutations predispose to the autosomal recessive Nijmegen-breakage syndrome (NBS) characterized by chromosomal instability and an increased risk of lymphoid malignancies and other cancers (Varon et al., 1998). Heterozygous *NBN* mutations predispose to breast cancer (BC) (Gorski et al., 2003), non-Hodgkin lymphoma (Steffen et al., 2006), and prostate cancer (Cybulski et al., 2013); however, their role in PDAC predisposition has not been studied yet. The most frequent pathogenic mutation in NBS patients and *NBN*-associated cancers is the recurrent Slavic founder mutation c.657del5 (c.657_661delACAAA) (Varon et al., 2000).

The next-generation sequencing (NGS) technology introduced analyses of large gene collections into genetic analyses in patients

Abbreviation: PDAC, pancreatic ductal adenocarcinoma; *BRCA1*, breast cancer gene 1; *BRCA2*, breast cancer gene 2; *PALB2*, partner and localizer of *BRCA2*; *MLH1*, homolog of *Escherichia coli* MutL 1; *MSH2*, homolog of *E. coli* MutS 2; *MSH6*, homolog of *E. coli* MutS 6; *PMS2*, postmeiotic segregation increased; *STK11*, serine/threonine protein kinase 11; *APC*, adenomatous polyposis coli gene; *CDKN2A*, cyclin-dependent kinase inhibitor 2A; *DDSB*, DNA double-strand break; *NBN*, nibrin; *MRN*, *MRE11/RAD50/NBN* complex; *NBS*, Nijmegen breakage syndrome; *NGS*, next-generation sequencing; *FFPE*, formalin-fixed paraffin-embedded; *HRM*, high resolution melting; *FET*, Fischer's exact test; *OR*, odds ratio; *CI*, confidence interval; *BC*, breast cancer; *CHEK2*, checkpoint kinase 2; *PARP*, poly(ADP-ribose) polymerase.

* Corresponding author at: Institute of Biochemistry and Experimental Oncology, First Faculty of Medicine, Charles University in Prague, U Nemocnice 5, 128 53 Prague 2, Czech Republic.

E-mail address: mjana@f1.cuni.cz (M. Janatova).

with cancer susceptibility. Among others, *NBN* is routinely analyzed in many cancer gene sequencing panels. Recently, we have performed a study of germline variants influencing the breast cancer susceptibility in high-risk breast cancer patients using the custom panel NGS (Lhota et al., 2016). We subsequently used the identical approach for the analysis of pancreatic cancer predisposition in a PDAC patient from multiple cancer family. We identified the c.657del5 germline mutation in the *NBN* gene in this patient. Therefore, we aimed to determine the frequency of c.657del5 in unselected Czech PDAC patients.

2. Materials and methods

2.1. Panel NGS analysis in a patient with pancreatic ductal adenocarcinoma

In order to identify possible germline pathogenic variant in PDAC-susceptibility genes, we performed custom panel NGS targeting 581 genes in a PDAC patient from multiple cancer family (Fig. 1). The NGS and bioinformatics analysis was performed as described previously (Lhota et al., 2016) and revealed germline c.657del5 *NBN* variant. The mutation was confirmed by Sanger sequencing from independent PCR amplified blood DNA sample. The presence of the c.657del5 *NBN* variant in deceased proband's sister with gastric cancer (Fig. 1) was analyzed in DNA isolated from FFPE tumor tissue using the Cobas DNA Sample Preparation Kit (Roche).

2.2. Patients with pancreatic ductal adenocarcinoma

We genotyped c.657del5 *NBN* variant in blood-isolated DNA samples from 241 unselected, histopathologically-verified PDAC patients, which included 152 samples from the National Institute of Public Health [median age at diagnosis: 63 years (ranged 40–82); 59 females] and 89 samples from the Department of Oncology, General University Hospital in Prague [median age at diagnosis: 64 years (ranged 38–84); 49 females]. Information about family history of cancer in c.657del5 carriers was gathered from medical records when available.

The control group included 915 non-cancer individuals and it had been described and genotyped previously. All patients and controls

were of Slavic descent and of Czech origin. The study was approved by the local Ethical Committees and a written informed consent was obtained from all participants.

2.3. The *NBN* c.657del5 genotyping

The exon 6 of the *NBN* gene was analyzed by a high resolution melting (HRM; LightCycler 480; Roche) using HOTT FirePol EvaGreen HRM Mix (Solis BioDyne). The primer sequences had been described previously (Mateju et al., 2012). The presence of c.657del5 was confirmed by sequencing.

2.4. Statistical analysis

The difference between groups was calculated using the Fisher exact test (FET).

3. Results

We analyzed a PDAC patient (diagnosed at 64 years) from multiple-cancer family and identified the c.657del5 *NBN* mutation using the panel NGS (Fig. 1). Except to this germline mutation, we found no other truncating variants in other known PDAC-susceptibility genes (*BRCA1*, *BRCA2*, *PALB2*, *MLH1*, *MSH2*, *MSH6*, *PMS2*, *STK11*, *APC*, *CDKN2A*). The presence of c.657del5 mutation was confirmed also in the proband's sister deceased from gastric cancer (Fig. 1).

In the subsequent analysis, we genotyped c.657del5 in other 241 unselected PDAC patients and found five mutation carriers among them (2.07%). Thus, the frequency of c.657del5 among PDAC patients was significantly higher than that in previously analyzed controls (2/915), suggesting that the carriers of c.657del5 have an increased risk of PDAC development (OR = 9.7; 95%CI: 1.9–50.2; $P_{FET} = 0.006$). A PDAC family history was documented in none of the five c.657del5 carriers from 241 unselected PDAC patients; however, one patient had family cancer history (a sister with gastric cancer), and another female patient suffered from a duplicity of BC (at 46 years) and PDAC

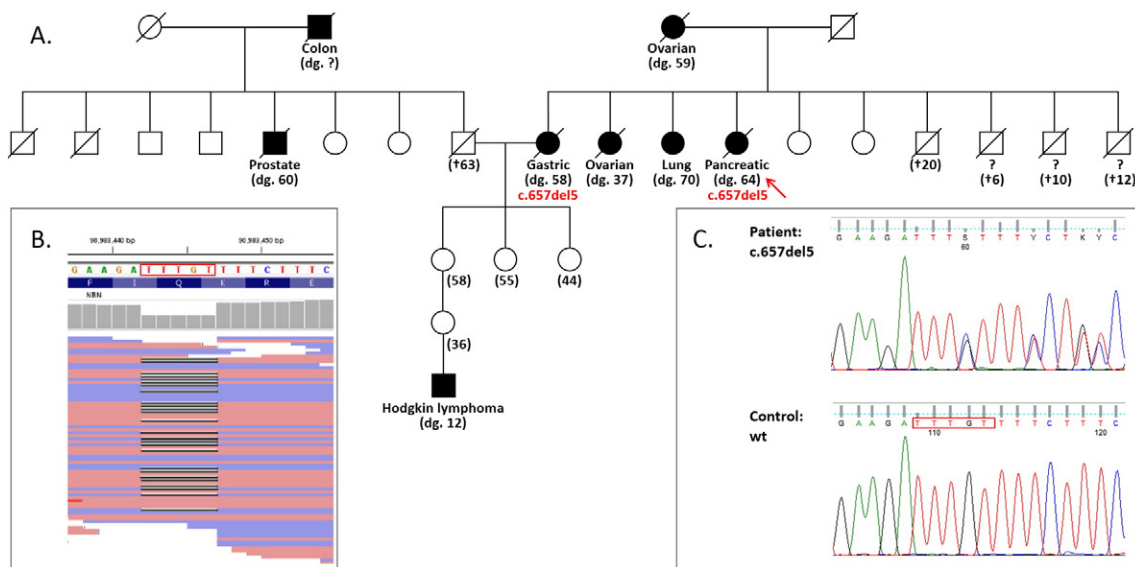


Fig. 1. Pedigree (A) of the multiple cancer family showing the proband with PDAC (indicated by an arrow) and her sister, both carrying c.657del5. DNA samples from other relatives were not available for genotyping. The ages of cancer diagnoses (dg.) or cessation (†) are indicated in the pedigree. The deletion of five nucleotides (TTTGT from reverse strand) is highlighted by a red frame in NGS analysis (B), confirmed by Sanger sequencing (C).

(at 64 years). The mean age at diagnosis for the c.657del5 carriers was 65.8 years (range 59–73).

4. Discussion

The highest frequency of *NBN* mutation carriers (up to 3.7% of patients) was found in BC patients from Central and Eastern Europe (Gorski et al., 2003). Recent meta-analysis indicated that c.657del5 is a moderate BC (OR = 2.51; 95%CI: 1.68–3.73) and lymphoma (OR = 2.93; 95%CI: 1.62–5.29) susceptibility allele, and that it also strongly increases the risk of prostate cancer (OR = 5.87; 95%CI: 2.51–13.75) (Gao et al., 2013). The association of the hereditary *NBN* mutations with BC susceptibility led to the inclusion of *NBN* into multigene cancer panel NGS analyses in high-risk individuals (Couch et al., 2015). Two studies have reported the results of hereditary mutation analysis performed by multi-gene panel testing in PDAC patients. While no truncating *NBN* mutation was identified in two previous studies of 290 and 638 patients, respectively (Grant et al., 2015; Roberts et al., 2016), Hu et al. found one c.657del5 carrier in 96 patients (also carrying the *CHEK2* mutation) (Hu et al., 2016). Recently, Lener et al. performed analysis of 10 prevalent founder mutations in *BRCA1*, *CHEK2*, *PALB2* and *NBN* (incl. c.657del5) in 383 pancreatic cancer patients and detected eight carriers of c.657del5 (2.09%), indicating the increased risk of pancreatic cancer in c.657del5 carriers (OR = 3.8; 95%CI: 1.68–8.60) in Poland (Lener et al., 2016).

The high frequencies of c.657del5 identified in PDAC patients in our and Lener et al. studies indicate that *NBN* is another DNA repair gene involved in PDAC-susceptibility. In comparison with our current study identifying 2.07% of c.657del5 carriers in unselected PDAC patients, earlier analyses found considerably lower frequencies of the mutation in Czech unselected BC (0.3%), colorectal cancer (0.3%), and lymphoma patients (0.8%) (Lhota et al., 2016; Pardini et al., 2009; Soucek et al., 2003). Our results and Lener et al. study (Lener et al., 2016) suggest that c.657del5 may be a novel PDAC-susceptibility allele significantly increasing the risk of PDAC development [combined OR calculated from this and Lener et al. studies comprising 624 pancreatic cancer patients (13 carriers of c.657del5) and 4915 controls (24 carriers) is 4.33; 95%CI 2.2–8.56; $p < 0.001$]. However, further studies in larger populations together with segregation analyses will be necessary to confirm our observation. They also may help specify the PDAC-associated risk more precisely, which is required for clinical management of the carriers and evaluation of c.657del5 as a putative predictive biomarker for therapy using DNA cross-linking agents or PARP inhibitors in carriers with PDAC (Schroder-Heurich et al., 2014).

Only the first patient identified in our preliminary NGS analysis had an indicative family cancer history (Fig. 1) and c.657del5 co-segregated with cancer diagnoses in the family. Its presence in the proband's sister with gastric cancer indicates that the carriers of c.657del5 may develop a broader spectrum of cancers. One in five carriers from the unselected PDAC group had a sister with gastric cancer (unfortunately, no DNA from this patient was available). The other mutation carriers displayed no family cancer history, just like the c.657del5 mutation carrier in the aforementioned report by Hu et al. (Hu et al., 2016). The similar mean age at PDAC diagnosis in carriers and non-carriers in our analysis (65.8 and 63.5 years, respectively) suggests that the c.657del5 mutation is not associated with an earlier disease onset.

In conclusion, our study suggests a novel role of the c.657del5 mutation in PDAC susceptibility. Future analyses of *NBN* in multi-gene cancer panels will help identify the hereditary pathogenic *NBN* mutations throughout the entire gene and enable a more accurate estimation of *NBN*-associated cancer risks.

Conflict of interest

None.

Acknowledgements

We would like to thank all patients who agreed to participate in this study, Marketa Dostalova and Marie Epsteinova for their excellent technical help. This study was supported by grants: IGA MZCR NT-14006-3/2013 and NT-14054-3/2013, PRVOUK-P27/LF1/1, and SVV-UK 260256/2016.

References

- Becker, A.E., Hernandez, Y.G., Frucht, H., Lucas, A.L., 2014. Pancreatic ductal adenocarcinoma: risk factors, screening, and early detection. *World J. Gastroenterol.* 20, 11182–11198.
- Carney, J.P., Maser, R.S., Olivares, H., Davis, E.M., Le Beau, M., Yates 3rd, J.R., Hays, L., Morgan, W.F., Petriani, J.H., 1998. The hMre11/hRad50 protein complex and Nijmegen breakage syndrome: linkage of double-strand break repair to the cellular DNA damage response. *Cell* 93, 477–486.
- Couch, F.J., Johnson, M.R., Rabe, K.G., Brune, K., de Andrade, M., Goggins, M., Rothenmund, H., Gallinger, S., Klein, A., Petersen, G.M., Hruban, R.H., 2007. The prevalence of *BRCA2* mutations in familial pancreatic cancer. *Cancer Epidemiol. Biomark. Prev.* 16, 342–346.
- Couch, F.J., Hart, S.N., Sharma, P., Toland, A.E., Wang, X., Miron, P., Olson, J.E., Godwin, A.K., Pankratz, V.S., Olsowald, C., Slettedahl, S., Hallberg, E., Guidugli, L., Davila, J.I., Beckmann, M.W., Janni, W., Rack, B., Ekici, A.B., Slamon, D.J., Konstantopoulou, I., Fostira, F., Vratimos, A., Fountzilias, G., Pelttari, L.M., Tapper, W.J., Durcan, L., Cross, S.S., Pilarski, R., Shapiro, C.L., Klemp, J., Yao, S., Garber, J., Cox, A., Brauch, H., Ambrosone, C., Nevanlinna, H., Yannoukakos, D., Slager, S.L., Vachon, C.M., Eccles, D.M., Fasching, P.A., 2015. Inherited mutations in 17 breast cancer susceptibility genes among a large triple-negative breast cancer cohort unselected for family history of breast cancer. *J. Clin. Oncol.* 33, 304–311.
- Cybulski, C., Wokolorczyk, D., Kluzniak, W., Jakubowska, A., Górski, B., Gronwald, J., Huzarski, T., Kashyap, A., Byrski, T., Dębniak, T., Gołab, A., Gliniewicz, B., Sikorski, A., Switała, J., Borkowski, T., Borkowski, A., Antczak, A., Wojnar, L., Przybyła, J., Sosnowski, M., Małkiewicz, B., Zdrojowy, R., Sikorska-Radek, P., Matych, J., Wilkosz, J., Różański, W., Kiś, J., Bar, K., Bryniarski, P., Paradyś, A., Jersak, K., Niemirowicz, J., Stupski, P., Jarzemeski, P., Skrzypczyk, M., Dobruch, J., Domagała, P., SA, N., Lubiński, J., 2013. Polish hereditary prostate cancer consortium. An inherited *NBN* mutation is associated with poor prognosis prostate cancer. *Br. J. Cancer* 108, 461–468.
- Gao, P., Ma, N., Li, M., Tian, Q.B., Liu, D.W., 2013. Functional variants in *NBS1* and cancer risk: evidence from a meta-analysis of 60 publications with 111 individual studies. *Mutagenesis* 28, 683–697.
- Gorski, B., Dębniak, T., Masojc, B., Mierzejewski, M., Medrek, K., Cybulski, C., Jakubowska, A., Kurzawski, G., Chosia, M., Scott, R., Lubiński, J., 2003. Germline 657del5 mutation in the *NBS1* gene in breast cancer patients. *Int. J. Cancer* 106, 379–381.
- Grant, R.C., Selander, I., Connor, A.A., Selvarajah, S., Borgida, A., Briollais, L., Petersen, G.M., Lerner-Ellis, J., Holter, S., Gallinger, S., 2015. Prevalence of germline mutations in cancer predisposition genes in patients with pancreatic cancer. *Gastroenterology* 148, 556–564.
- Hu, C., Hart, S.N., Bamlet, W.R., Moore, R.M., Nandakumar, K., Eckloff, B.W., Lee, Y.K., Petersen, G.M., McWilliams, R.R., Couch, F.J., 2016. Prevalence of pathogenic mutations in cancer predisposition genes among pancreatic cancer patients. *Cancer Epidemiol. Biomark. Prev.* 25, 207–211.
- Jones, S., Hruban, R.H., Kamiyama, M., Borges, M., Zhang, X., Parsons, D.W., Lin, J.C., Palmisano, E., Brune, K., Jaffe, E.M., Iacobuzio-Donahue, C.A., Maitra, A., Parmigiani, G., Kern, S.E., Velculescu, V.E., Kinzler, K.W., Vogelstein, B., Eshleman, J.R., Goggins, M., Klein, A.P., 2009. Exomic sequencing identifies *PALB2* as a pancreatic cancer susceptibility gene. *Science* 324, 217.
- Lener, M.R., Scott, R.J., Kluzniak, W., Baszuk, P., Cybulski, C., Wiechowska-Kozłowska, A., Huzarski, T., Byrski, T., Kładny, J., Pietrzak, S., Soluch, A., Jakubowska, A., Lubiński, J., 2016. Do founder mutations characteristic of some cancer sites also predispose to pancreatic cancer? *Int. J. Cancer* <http://dx.doi.org/10.1002/ijc.30116>.
- Lhota, F., Zemanekova, P., Kleiblova, P., Soukupova, J., Vocka, M., Stranecky, V., Janatova, M., Hartmannova, H., Hodanova, K., Kmoch, S., Kleibl, Z., 2016. Hereditary truncating mutations of DNA repair and other genes in *BRCA1/BRCA2/PALB2*-negatively tested breast cancer patients. *Clin. Genet.* <http://dx.doi.org/10.1111/cge.12748>.
- Mateju, M., Kleiblova, P., Kleibl, Z., Janatova, M., Soukupova, J., Ticha, I., Novotny, J., Pohlreich, P., 2012. Germline mutations 657del5 and 643C > T (R215W) in *NBN* are not likely to be associated with increased risk of breast cancer in Czech women. *Breast Cancer Res. Treat.* 133, 809–811.
- Nakao, A., Fujii, T., Sugimoto, H., Kanazumi, N., Nomoto, S., Kodera, Y., Inoue, S., Takeda, S., 2006. Oncological problems in pancreatic cancer surgery. *World J. Gastroenterol.* 12, 4466–4472.
- Pardini, B., Naccarati, A., Polakova, V., Smerhovsky, Z., Hlavata, I., Soucek, P., Novotny, J., Vodickova, L., Tomanova, V., Landi, S., Vodicka, P., 2009. *NBN* 657del5 heterozygous mutations and colorectal cancer risk in the Czech Republic. *Mutat. Res.* 666, 64–67.
- Roberts, N.J., Norris, A.L., Petersen, G.M., Bondy, M.L., Brand, R., Gallinger, S., Kurtz, R.C., Olson, S.H., Rustgi, A.K., Schwartz, A.G., Stoffel, E., Syngal, S., Zogopoulos, G., Ali, S.Z., Axilbund, J., Chaffee, K.G., Chen, Y.C., Cote, M.L., Childs, E.J., Douville, C., Goes, F.S., Herman, J.M., Iacobuzio-Donahue, C., Kramer, M., Makohon-Moore, A., McCombie, R.W., McMahon, K.W., Niknafs, N., Parla, J., Pirooznia, M., Potash, J.B., Rhim, A.D., Smith, A.L., Wang, Y., Wolfgang, C.L., Wood, L.D., Zandi, P.P., Goggins, M., Karchin, R., Eshleman, J.R., Papadopoulos, N., Kinzler, K.W., Vogelstein, B., Hruban, R.H., Klein,

- A.P., 2016. Whole genome sequencing defines the genetic heterogeneity of familial pancreatic cancer. *Cancer Discov.* 6, 166–175.
- Schroder-Heurich, B., Bogdanova, N., Wieland, B., Xie, X., Noskowitz, M., Park-Simon, T.W., Hillemanns, P., Christiansen, H., Dörk, T., 2014. Functional deficiency of NBN, the Nijmegen breakage syndrome protein, in a p.R215W mutant breast cancer cell line. *BMC Cancer* 14, 434.
- Siegel, R.L., Miller, K.D., Jemal, A., 2015. Cancer statistics, 2015. *CA Cancer J. Clin.* 65, 5–29.
- Soucek, P., Gut, I., Trneny, M., Skovlund, E., Grenaker Alnaes, G., Kristensen, T., Børresen-Dale, A.L., Kristensen, V.N., 2003. 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.
- Steffen, J., Maneva, G., Poplawska, L., Varon, R., Mioduszezewska, O., Sperling, K., 2006. Increased risk of gastrointestinal lymphoma in carriers of the 657del5 NBS1 gene mutation. *Int. J. Cancer* 119, 2970–2973.
- Varon, R., Vissinga, C., Platzer, M., Cerosaletti, K.M., Chrzanowska, K.H., Saar, K., Beckmann, G., Seemanova, E., Cooper, P.R., Nowak, N.J., Stumm, M., Weemaes, C.M., Gatti, R.A., Wilson, R.K., Digweed, M., Rosenthal, A., Sperling, K., Concannon, P., Reis, A., 1998. Nibrin, a novel DNA double-strand break repair protein, is mutated in Nijmegen breakage syndrome. *Cell* 93, 467–476.
- Varon, R., Seemanova, E., Chrzanowska, K., Hnateyko, O., Piekutowska-Abramczuk, D., Krajewska-Walasek, M., Sykut-Cegielska, J., Sperling, K., Reis, A., 2000. Clinical ascertainment of Nijmegen breakage syndrome (NBS) and prevalence of the major mutation, 657del5, in three Slav populations. *Eur. J. Hum. Genet.* 8, 900–902.

PŘÍLOHA IV-5:

Hojny *et al.*; Multiplex PCR and NGS-based identification of mRNA splicing variants: Analysis of *BRCA1* splicing pattern as a model



Methodological paper

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



Jan Hojny^a, Petra Zemankova^a, Filip Lhota^a, Jan Sevcik^a, Viktor Stranecky^b, Hana Hartmannova^b, Katerina Hodanova^b, Ondrej Mestak^c, David Pavlista^d, Marketa Janatova^a, Jana Soukupova^a, Michal Vocka^e, Zdenek Kleibl^a, Petra Kleiblova^{a,f,*}

^a Institute of Biochemistry and Experimental Oncology, First Faculty of Medicine, Charles University, Prague 12853, Czech Republic

^b Institute of Inherited Metabolic Disorders, First Faculty of Medicine, Charles University and General University Hospital in Prague, Prague 120 00, Czech Republic

^c Department of Plastic Surgery, First Faculty of Medicine, Charles University and Na Bulovce Hospital, Prague 180 81, Czech Republic

^d Department of Obstetrics and Gynecology, First Faculty of Medicine, Charles University and General University Hospital in Prague, Prague 120 00, Czech Republic

^e Department of Oncology, First Faculty of Medicine, Charles University and General University Hospital in Prague, Prague 120 00, Czech Republic

^f Institute of Biology and Medical Genetics, First Faculty of Medicine, Charles University and General University Hospital in Prague, Prague 120 00, Czech Republic

ARTICLE INFO

Keywords:

Alternative splicing
mRNA splicing variant
BRCA1
Multiplex PCR
NGS

ABSTRACT

Alternative pre-mRNA splicing increases transcriptome plasticity by forming naturally-occurring alternative splicing variants (ASVs). Alterations of splicing processes, caused by DNA mutations, result in aberrant splicing and the formation of aberrant mRNA isoforms. Analyses of hereditary cancer predisposition genes reveal many DNA variants with unknown clinical significance (VUS) that potentially affect pre-mRNA splicing. Therefore, a comprehensive description of ASVs is an essential prerequisite for the interpretation of germline VUS in high-risk individuals.

To identify ASVs in a gene of interest, we have proposed an approach based on multiplex PCR (mPCR) amplification of all theoretically possible exon-exon junctions and subsequent characterization of size-selected and pooled mPCR products by next-generation sequencing (NGS). The efficiency of this method is illustrated by a comprehensive analysis of *BRCA1* ASVs in human leukocytes, normal mammary, and adipose tissues and stable cell lines.

We revealed 94 *BRCA1* ASVs, including 29 variants present in all tested samples. While differences in the qualitative expression of *BRCA1* ASVs among the analyzed human tissues were minor, larger differences were detected between tissue and cell line samples.

Compared with other ASV analysis methods, this approach represents a highly sensitive and rapid alternative for the identification of ASVs in any gene of interest.

1. Introduction

Hereditary mutations in cancer-susceptibility genes are responsible for tumor development in about 5% of all cancer patients. The carriers of hereditary mutations face a high life-time risk of cancer, which often develops at an early age (Rahman, 2014). Tailored care improving life expectancy in these high-risk individuals requires an unequivocal identification of causative mutations in hundreds of known cancer-susceptibility genes. The recent introduction of next-generation sequencing (NGS) into clinical diagnostics enables a simultaneous analysis of multiple genes; however, its clinical utility is hampered by the

presence of many variants with unknown significance (VUS) (Cheon et al., 2014). These genetic changes emerge as rare germline missense, silent or intronic variants with an uncertain biological and functional impact on the resulting protein isoform. The number of identified VUS rises proportionally to the length of the analyzed genomic sequence, and many of them may alter mRNA splicing processes (Tavtigian & Chenevix-Trench, 2014).

Pre-mRNA splicing controls the composition of matured mRNA by regulated intron exclusion and exon linking. A primary wild-type (wt) transcript (pre-mRNA) can be variably processed by alternative splicing into alternative mRNA variants translated into protein isoforms with

Abbreviations: ASV(s), alternative splicing variant(s); mPCR, multiplex PCR; NGS, next generation sequencing; SAS, splice acceptor shift; SDS, splice donor shift; VUS, variant of unknown clinical significance

* Corresponding author at: Institute of Biochemistry and Experimental Oncology, First Faculty of Medicine, Charles University, U Nemocnice 5, Prague 2 128 53, Czech Republic.

E-mail address: petra.kleiblova@lf1.cuni.cz (P. Kleiblova).

<http://dx.doi.org/10.1016/j.gene.2017.09.025>

Received 28 June 2017; Received in revised form 26 July 2017; Accepted 13 September 2017

Available online 14 September 2017

0378-1119/© 2017 Elsevier B.V. All rights reserved.

different biological activities (Bentley, 2014). Alternative splicing must be distinguished from aberrant splicing resulting from a dysregulation of natural splice site recognition caused by DNA mutations. The DNA sequence variants affecting pre-mRNA splicing are more prevalent than estimated up to now, and they account for at least 15% of disease-causing mutations, and for up to 50% of all mutations described in some genes (Caminsky et al., 2015; Soukarieh et al., 2016). Various splicing assays help to disclose the impact of VUS on splicing processes (Whiley et al., 2014), and variants that cause aberrant splicing are considered pathogenic. The evaluation of aberrant splicing requires a precise knowledge of alternative splicing variants (ASVs) for the analyzed primary transcript (Colombo et al., 2014). Despite large-scale RNA sequencing projects (e.g. ENCODE, GTEx), there is no precise catalogue of ASVs or validated RNAseq data for most clinically-relevant genes (Sloan et al., 2016; Baralle & Buratti, 2017).

Recently, two articles have described a comprehensive analysis of naturally occurring splicing variants in *BRCA1* (Colombo et al., 2014; Romero et al., 2015), one of the most studied cancer-susceptibility genes responsible for hereditary breast and ovarian cancer (Kleibl & Kristensen, 2016). In both studies, the RNA-based analysis involved a combination of various techniques including RT-PCR, exon scanning, cloning, sequencing, and relative (semi)quantification. This experimental variability negatively affects the reproducibility of splicing variant analyses from various mRNA sources and makes the methods difficult to use in the analyses of other gene products. The large number of these analytic techniques makes the analysis laborious, may negatively affect its reproducibility, and adaptation to characterize another gene transcripts. Therefore, we aimed to develop a versatile approach suitable for the characterization of ASVs in any gene of interest based on NGS of multiplex PCR-generated amplicons covering all theoretically possible exon-exon junctions.

2. Materials and methods

2.1. An overview of experimental design

We aim to characterize the ASVs of any gene from an RNA sample, dominantly on the qualitative level. The analysis comprises four steps: i) multiplex PCR (mPCR) amplification of all theoretically possible mRNA splicing variants from a cDNA template, ii) pooling of mPCRs, purification and size selection of pooled mPCR products targeting short amplicons, iii) standard NGS library preparation from size-selected mPCR fragments followed by routine Illumina sequencing, and iv) a bioinformatics analysis. We have demonstrated the efficiency of the mPCR/NGS approach by the characterization of *BRCA1* ASVs because i) the *BRCA1* gene is the most frequently altered breast cancer susceptibility gene in many countries including the Czech Republic and many *BRCA1* VUS contribute to aberrant splicing, ii) 63 *BRCA1* mRNA variants were recently described using conventional RT-PCR and capillary-electrophoresis by Colombo et al. (2014) and Romero et al. (2015), indicating that iii) the *BRCA1* mRNA splicing isoform pattern is highly variable.

2.2. Alternative *BRCA1* splice site nomenclature

All alternative splicing events were classified into biotypes based on previously published nomenclature (Colombo et al., 2014; Romero et al., 2015). The insertions (▼) and deletions (Δ) denote splicing events affecting a single exon (cassette) or > 1 consecutive exons (multicassette). The deletions affecting the 5' and 3' ends of an exon were described as an exon number with an added "p" or "q", respectively. The extension of an exon sequence into an adjacent intronic region is described as an exon number with an "a". The splice donor/acceptor shift (SDS/SAS) variants were identified and counted as NGS reads with deletions of nucleotides at the exon-exon junctions or insertions of intronic parts flanking to the 5' or 3' ends of an exon. The

mixed biotypes denoted combinations of the above-mentioned events. The *BRCA1* exons were numbered according to the Breast Cancer Information Core Database (<https://research.nhgri.nih.gov/bic/>) nomenclature.

2.3. Patients and samples

The characterization of *BRCA1* ASVs was performed in 96 RNA samples obtained from 32 individuals (Supplementary Table S1), including 16 non-cancer controls, eight breast-cancer (BC) patients without *BRCA1* mutation, and eight *BRCA1*-mutation carriers. Simultaneously-obtained tissue samples were collected during BC surgery or preventive mastectomy (in BC patients and *BRCA1* mutation carriers) or during cosmetic breast surgery (in controls). All enrolled individuals were Caucasians of a Czech origin who gave a written informed consent approved by ethical committees to participate in the study. RNA samples were isolated from the leukocytes and macroscopically dissected fresh mammary and adipose perimammary tissues of each individual. We further analyzed RNA samples from stable human cell lines (from MCF7 cells, and from pooled EM-G3, HeLa, and MDA-MB-231 cells). The cell lines were maintained as described previously (Brozova et al., 2007; Sevcik et al., 2012; Sevcik et al., 2013; Vondruskova et al., 2008).

2.3.1. Total RNA isolation, quality control and cDNA synthesis

All RNA samples were processed according to MIQE guidelines (Bustin et al., 2009). Peripheral blood samples (2.5 ml) were collected into PAXgene Blood RNA tubes, incubated overnight at room temperature, and stored at -20°C . All stored samples were thawed and stored for 2 h at room temperature before RNA isolation performed with PAXgene Blood RNA Kit (PreAnalytiX). Solid tissue samples (~100 mg/sample) were submerged into 1 ml of RNeasy Lysis Buffer (Qiagen) immediately after surgical excision, processed according to the manufacturer, and after overnight incubation ($2-8^{\circ}\text{C}$) stored at -80°C until RNA isolation. Forty micrograms of thawed, RNeasy-preserved samples were homogenized using MagNA Lyser Green Beads tubes on MagNA Lyser Instrument (Roche) in the presence of 1 ml Qiazol (Qiagen). Total RNAs from the homogenated tissues and cultured cells were isolated with RNeasy Tissue Mini Kit (Qiagen).

All RNA samples were treated by DNase I, quantified on NanoDrop 1000 (Thermo Fisher Scientific) and characterized by the RNA integrity number (RIN) using Bioanalyzer 2100 with RNA 6000 Nano Kit (Agilent Technologies; tissue samples $\text{RIN}_{\text{mean}} = 7.4$; range 6.3–8.9).

Overall, 1.5 μg of RNA was used for cDNA synthesis (in a reaction volume of 20 μl). The cDNA synthesis was performed using SuperScript III Reverse Transcriptase (Thermo Fisher Scientific) and random hexamers (Roche) as described previously (Kleiblova et al., 2010). A routine PCR control of cDNA quality/integrity was performed prior to further analyses (not shown).

2.4. Multiplex PCR (mPCR) amplification and size selection

2.4.1. Primer designing

The primers were designed to specifically cover all possible exon-exon junctions. The resulting PCR amplicons thus enable the identification of all canonical as well as alternative splicing mRNA isoforms. Forward primers targeted the 3' region while reverse primers aimed at the 5' region of an exon (Supplementary Fig. S1). For the analysis of a single gene transcript consisting of N exons, the number of $N-2$ forward and $N-2$ reverse primers is required for the amplification of all theoretically possible exon-exon junctions in at least $N-2$ mPCR reactions that are finally pooled into one mPCR pool.

For the analysis of *BRCA1* ASVs, we designed 45 primers targeting 22 coding exons of the canonical *BRCA1* transcript (NM_007294), and alternative exons 11q and 13A (Fig. 1A, Supplementary Table S2). All individual PCRs were optimized separately (Supplementary Fig. S2)

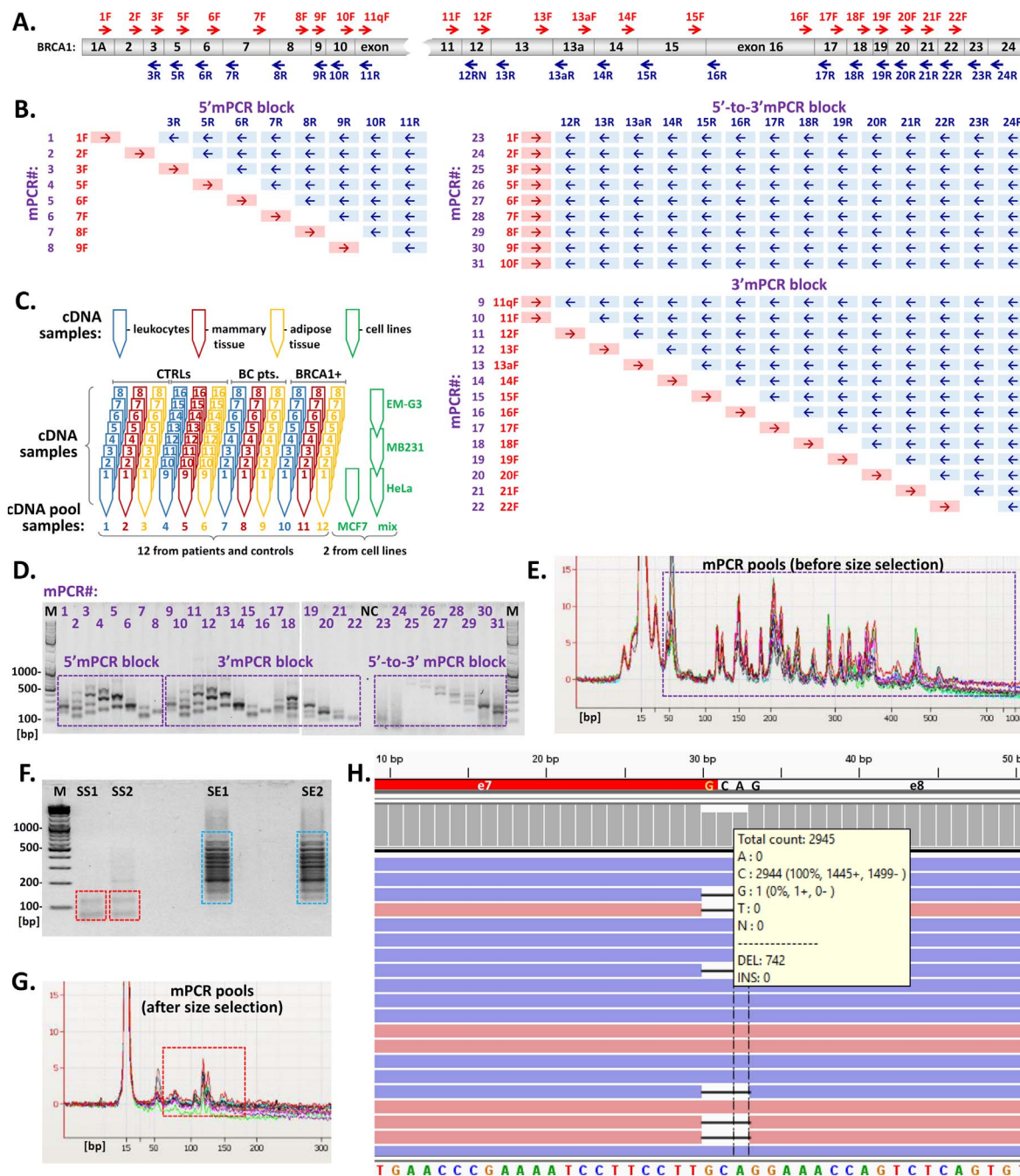


Fig. 1. Method overview. The chart (A) shows primer pairs (forward – red; reverse – blue) used for mPCR amplifications of all theoretically possible BRCA1 ASVs. Due to the presence of the large exon 11, 31 mPCR reactions (violet letters; B) were performed in three ‘blocks’. All 31 mPCR reactions were performed with each of 14 cDNA pools (C). Twelve cDNA pools of human cDNA samples and two cDNA pools from cell line samples served as templates for 31 mPCR reactions. Agarose gel electrophoresis of 31 mPCRs (ranging in size between ~50 and ~700 bp; purple dashed line) from a single cDNA pool is shown in (D). All 31 mPCRs from each cDNA pool sample were further pooled together (to create an mPCR pool) and analyzed on Agilent Bioanalyzer (E; the overlaying electrophoretograms of 12 mPCR pools show good reproducibility). The double-sided size selection was used to enrich the short mPCR amplicons that were subsequently used for NGS library preparation. The agarose gel electrophoresis (F) displays the enrichment of size-selected fragments (SS; red dashed line boxes) while size-excluded fragments (SE; blue dashed line boxes) were discarded. The size-selected (SS) samples (ranging at 50–150 bp in length; red dashed line box in G) were verified by Agilent Bioanalyzer. The MiSeq reads were mapped to the bam files (Supplementary Table S3) containing all theoretically possible BRCA1 splicing cassette and multicassete events. Visual inspection of reads in IGV viewer enabled direct assessment and quantification of SDS/SAS as shown (H) for the Δ CAG at the 5’ end of exon 8 (erroneously mapped as Δ GCA; coverage depth is shown as grey vertical bars; forward (pink) and reverse (blue) reads are shown as vertical bars). In normal mammary tissue of BC patients (shown in H); Δ CAG accounted for 742 reads (20.1%), while 2945 sequencing reads were recorded for wt sequence (Supplementary Table S6). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

and subsequently in mPCR (Supplementary Fig. S3).

2.4.2. mPCR amplification and preparation of mPCR pools

All theoretically possible splicing events of a gene of interest can be amplified in a few mPCRs. The number of mPCRs depends primarily on

the number of exons of the analyzed gene. Every individual mPCR contains a single forward primer and a set of reverse primers targeting all consecutive exons except the exon directly flanking to the exon targeted by the forward primer (Supplementary Fig. S1). The undesired synthesis of the amplicons of several consecutive exons was reduced by

a short elongation time (Supplementary Fig. S3B).

The analysis of BRCA1 ASVs in each cDNA template required 31 mPCRs (Fig. 1B). To overcome the usually limited amount of RNA (and resulting cDNA) available from human tissue samples, the cDNA template for mPCRs was prepared by the pooling of individual cDNAs (from the same tissue and sample subgroup; Fig. 1C). This may be necessary especially with multiexonic and low-expressed genes (including *BRCA1*) which require a larger number of mPCRs consuming an increased amount of cDNA. Although the cDNA pooling resulted in a loss of information about the expression of ASVs in the individual cDNA samples, it increased the chances of detecting low-expressed ASVs.

For the BRCA1 analysis, eight individual cDNA samples (16 μ l each) from the same tissue type and patient group were pooled together to obtain 12 patient cDNA pools (each 128 μ l; Fig. 1C). We also analyzed cDNAs from stable human cell lines (cDNA from MCF7 cells and a pool of cDNAs from EM-G3, HeLa and MD-MB-231). Each of the 12 patient cDNA pools, MCF7 cDNA, and cDNA cell line pools served as a template for 31 mPCRs. Each 40 μ l mPCR contained 4 μ l of pooled cDNA template (equivalent to 300 ng of RNA), a single forward primer (final concentration 225 nM), a variable set of reverse primers (final concentration 75 nM of each), and FastStart Taq DNA Polymerase (Roche) according to the manufacturer's instructions. The mPCR amplifications involved 4-minute incubation at 95 °C followed by 35 cycles (95 °C for 10 s, 62 °C for 20 s, and 72 °C for 15 s) and final extension at 72 °C for 7 min. The individual mPCR products were analyzed electrophoretically (Fig. 1D). After that, 35 μ l of each of the 31 mPCRs from a single cDNA pool were mixed together to provide an mPCR pool. The resulting mPCR pools were characterized by capillary electrophoresis using 2100 Bioanalyzer and DNA 1000 Kit (Agilent Technologies; Fig. 1E).

2.4.3. Size selection and purification of mPCR pools

To reduce the presence of longer mPCR-amplified fragments containing short consecutive exons, the mPCRs pools were subjected to size selection using double-sided solid phase reversible immobilization with magnetic beads. The size-selected and purified amplicons served as templates to prepare a standard NGS library.

As the length of the targeted mPCR amplicons of BRCA1 ASVs was expected to range mostly at 80–90 bp, we performed size selection using magnetic beads to remove undesired amplicons (< 50 bp and > 150 bp; Fig. 1F). First, we used 1.8 \times concentration of Agencourt AMPure XP reagent (Beckman Coulter) to bind and remove amplicons > 150 bp (dominantly containing PCR products amplified from the canonical BRCA1 mRNA). Subsequently, we mixed the supernatant from the first reaction with a reagent to the final 2.5 \times concentration to withdraw DNA fragments > 50 bp in length. The size-selected and purified samples were characterized on Agilent Bioanalyzer with DNA 1000 Kit (Fig. 1G).

2.4.4. NGS library preparation, MiSeq sequencing

First, Dynazyme II DNA Polymerase (Thermo Fisher Scientific) was used to create 3'-dA overhangs in DNA fragments in size-selected and purified mPCR pools. Subsequently, Illumina sequencing adaptors were ligated using Rapid DNA Ligation Kit (Thermo Fisher Scientific). Finally, the seven-cycle PCR reaction (NEBNext High Fidelity PCR Master Mix, NEB) with a universal primer introduced a 6-bp index sequence unique for each mPCR pool. The processed samples were purified using Agencourt AMPure XP Reagent after each step of sequencing library preparation. The quality of the prepared libraries was characterized on 2100 Bioanalyzer and quantified fluorimetrically (Qubit; Thermo Fisher Scientific). To achieve sufficient sequencing coverage and sample diversity, we admixed our mPCR-prepared libraries with panel sequencing libraries of high complexity to standard MiSeq runs (one mPCR pool library represented 1/30 of sequencing capacity). Runs were sequenced with MiSeq Reagent Kit v3 (150 cycle).

2.4.5. Bioinformatics

Bioinformatics requires double-step mapping. First, sequencing data are mapped to a user-designed fasta file containing the exon-exon junctions of all theoretically possible splicing variants of the gene of interest. Thus we can identify all cassette and multicassette splicing deletions and variants resulting from short SDS/SAS. Second, sequencing data are mapped to the genomic sequence of the analyzed gene in order to identify the exonized intronic sequences.

In our BRCA1 analysis, the primary raw data sets (in the fastq.gz format) were processed by a routine bioinformatics pipeline using software tools specified below using the default settings (if not otherwise specified in the list of commands listed in Supplementary Table S7). The remaining sequences were trimmed (to remove adapters and low-quality bases) before mapping by Trimmomatic (ver. 0.32; <http://www.usadellab.org>). First, we mapped raw data sets to the prepared fasta file using Novoalign (ver. 2.08.03; Novocraft). Reads with insufficient sequencing quality were removed. This “BRCA1 splicing” fasta file contained 311 sequences (“exon-exon_computed” in Supplementary Table S3) which considered all possible combinations of known BRCA1 exons (NM_007294), including alternative exon 13A, and known SDS/SAS combinations > 10 bp (11q, and 5q). Output in the SAM format was transformed to BAM by Picard tools (ver. 1.129; <https://broadinstitute.github.io/picard/>) and displayed in IGV (Integrative Genomics Viewer, Broad Institute; Fig. 1H). Coverage statistics were created by SAMtools (ver. 0.1.19; <http://samtools.sourceforge.net/>). Since this approach ignored the presence of exonized intron sequences, we further mapped all raw data to the BRCA1 gene sequence (81,189 bp from NG_005905 spanning sequence 92,500–173,688). The mapping results were analyzed in IGV manually to remove reads with soft-clipped bases. Mapped reads that exceeded from exons into flanking introns or reads mapped to deep intronic sequences, respectively, were recorded including sequences of unmapped nucleotides. The unmapped parts of reads were BLASTed with the BRCA1 gene sequence (<https://blast.ncbi.nlm.nih.gov/>). If the unmapped sequence contained the 5' or 3' parts of a BRCA1 exon, we were able to identify the entire intronic insertion. Sequences comprising the identified intronic insertions with flanking exonic sequences were added to a fasta file (“from_IGV_BLASTed” in Supplementary Table S3) and used for new mapping from the original dataset. Mapped reads were manually inspected in IGV in order to eliminate incorrectly mapped reads with soft-clipped bases.

To compare the numbers of ASV reads among the examined mPCR pools, we expressed the number of sequencing reads of an ASV as value normalized to 10⁶ reads in the given mPCR pool (Supplementary Tables S4 and S5).

3. Results

In order to prepare a versatile method for a direct assessment of splice junction events, we designed the mPCR/NGS-based approach enabling the identification of ASVs in a gene of interest. Using the described method, we identified 94 BRCA1 ASVs (Table 1) comprising all previously described biotypes (Colombo et al., 2014). Our analysis of simultaneously-obtained tissue RNA samples revealed that the highest number of ASVs were expressed in mammary tissue (72 variants), followed by leukocytes and perimammary adipose tissues (67 and 54 variants, respectively).

Forty-eight ASVs were identified in all examined human tissue types, with 29 of them expressed in each cDNA tissue sample (referred here as “ubiquitous”; Fig. 2; Table 1).

Only slight qualitative differences were identified among tissue samples from non-cancer controls, BC patients and BRCA1-mutation carriers. In contrast, the spectrum of ASVs differed between tissue and cell-line samples (Fig. 2). Altogether, 76 ASVs (11 of them exclusively) were expressed in cell lines.

The detected variants included 25 in-frame ASVs that may

Table 1

Description of all BRCA1 ASVs identified in this study, including variant name, systematic description of the variant at the cDNA level, functional annotation, and biotype class. The values for the expression of particular variants in analyzed cell lines and human tissue samples show normalized sequencing reads (reads per 10⁶ reads), the colours indicate normalized sequencing coverage: 0 reads = white; 1–9 reads = green; 10–99 reads = yellow; 100–999 reads = light red; 1000–9999 reads = red; > 9999 reads = dark red. The “ubiquitous” variants (expressed in all analyzed human tissue samples) are highlighted by bold letters and blue lines. The presence of variants in the analyzed human tissue samples and cell lines was compared to that reported as predominant (P), present (1), or absent (0) in studies by Colombo et al. (2014), Romero et al. (2015) and Orban and Olah (2003). The extended version of this table is shown in Supplementary Table S5.

Variant description	HGVS description	Functional annotation	Biot ype	Analyzed cell lines		Analyzed human tissues									References					
				MCF7	mix	Leukocytes			Mammary			Adipose			Leuko.	Mam.	Tu	Orban Olah (2003)		
						Non-cancer controls	BC patients	BRCA1 mut. carriers	Non-cancer controls	BC patients	BRCA1 mut. carriers	Non-cancer controls	BC patients	BRCA1 mut. carriers	Colombo (2014)	Romero (2015)	Colombo (2014)		Romero (2015)	Romero (2015)
1Aq	c.-25_-20del6	UTR	SDSΔ	1185	7854	3632	5045	4210	2981	11439	4010	2508	6035	4955	P	P	P	P	P	1
1Aq, 2a	c.-25_-20del6, c.-19-59_-19-1 ins59	UTR	SDSΔ + SAS ▼	0	0	0	0	0	30	0	0	0	0	0	0	0	0	0	0	0
1Aq, Δ2	c.-25_80del105	n.c.	SDSΔ + CΔ	176	5032	40	101	150	46	191	28	59	82	205	1	1	1	0	0	0
1Aq, Δ2_3	c.-25_134del159	n.c.	SDSΔ + mCΔ	0	1575	0	202	13	14	46	0	0	0	0	1	1	1	0	0	0
1Aq, Δ2_5	c.-25_212del237	n.c.	SDSΔ + mCΔ	0	438	0	0	0	2	0	83	0	0	0	1	1	1	0	0	0
1Aq, Δ2_5, 6p	c.-25_282del307	n.c.	SDSΔ + mCΔ + SASΔ	0	350	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1Aq, Δ2_7	c.-25_441del466	n.c.	SDSΔ + mCΔ	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0
1Aq, Δ2_7, 8p	c.-25_444del469	n.c.	SDSΔ + mCΔ + SASΔ	0	0	0	34	0	0	0	0	0	0	0	0	0	0	0	0	0
1Aq, Δ2_10	c.-25_670del695	n.c.	SDSΔ + mCΔ	0	438	0	0	7	0	0	0	0	0	0	1	1	1	0	0	0
1Aq, Δ2_17	c.-25_5074del5099	n.c.	SDSΔ + mCΔ	88	1225	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1Aq, Δ2_19	c.-25_5193del5218	n.c.	SDSΔ + mCΔ	0	0	0	0	0	0	289	0	0	0	0	0	0	0	0	0	0
1aA	c.-20+1_-20+89ins89	UTR	SDS ▼	6824	12098	321	219	1211	1501	6557	5429	940	427	349	1	1	1	0	P	0
2p	c.-19_-7del13	UTR	SASΔ	88	963	140	118	232	77	214	28	76	238	164	1	P	1	0	P	0
Δ2	c.-19_80del99	n.c.	CΔ	176	3982	482	1063	1987	323	3305	1337	481	1308	462	1	P	1	P	P	0
Δ2_3	c.-19_134del153	n.c.	mCΔ	0	1575	0	0	0	36	364	0	5	0	82	1	P	1	0	P	0
Δ2_3, ▼4	c.-19_134del153 + c.135-4047_135-3932ins116	n.c.	mCΔ + Cq	110	0	0	0	0	0	0	0	5	0	51	1	1	0	0	0	0
Δ2_5	c.-19_212del231	n.c.	mCΔ	88	1203	20	0	0	73	35	0	0	0	0	1	1	1	0	0	0
Δ2_7, 8p	c.-19_444del463	n.c.	mCΔ + SASΔ	0	2253	10	169	0	0	0	0	0	0	0	0	0	0	0	0	0
Δ2_10	c.-19_670del689	n.c.	mCΔ	0	88	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1
Δ2_17	c.-19_5074del5093	n.c.	mCΔ	0	2625	0	0	0	0	0	11	0	0	0	0	0	0	0	0	0
Δ2_19	c.-19_5193del5212	n.c.	mCΔ	0	0	0	101	0	0	976	0	0	22	0	0	0	0	0	0	0
▼ 145 bp int 2	c.81-3486_81-3342ins145	FS	C ▼	636	459	40	34	72	24	87	276	86	100	82	0	0	0	0	0	0
Δ3	c.81_134del54	FS	CΔ	987	21089	3120	10293	8436	4141	9336	9436	2135	5217	6750	1	P	1	0	P	1
Δ3, ▼4	c.81_134del54 + c.135-4047_135-3932ins116	FS	CΔ + C ▼	22	44	0	0	0	44	0	0	0	0	0	0	0	0	0	0	0
Δ3_5	c.81_212del132	IF	mCΔ	0	875	60	405	215	109	474	165	108	0	0	0	0	0	0	0	0
▼ 116bp int 3	c.134+3124_134+3239ins116	FS	C ▼	461	1663	20	0	26	61	23	83	38	212	62	0	0	0	0	0	0
▼ 4	c.135-4047_135-3932ins116	FS	C ▼	7460	2188	281	489	238	367	306	276	470	279	574	1	1	1	0	P	0
Δ5	c.135_212del178	IF	C ▼	395	7351	8016	13043	8569	5524	7476	10169	4778	4013	8823	P	1	P	P	P	1
Δ5_6	c.135_301del167	FS	mCΔ	0	0	0	0	0	0	0	634	0	0	0	0	0	0	0	0	
Δ5_6, 7p	c.135_307del173	FS	mCΔ + SASΔ	0	0	0	0	333	0	127	0	0	0	0	0	0	0	0	0	
Δ5_7	c.135_441del307	FS	mCΔ	0	481	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
Δ5_7, 8p	c.135_444del310	FS	mCΔ + SASΔ	0	0	20	0	0	0	0	0	0	0	0	0	0	0	0	0	
Δ5_9	c.135_593del459	IF	mCΔ	0	44	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
5q	c.191_212del122	FS	SDSΔ	549	1269	3602	2818	2137	8023	4501	3624	5886	9339	4268	P	P	P	P	P	1
5q, Δ6	c.191_301del111	IF	SDSΔ + CΔ	0	0	0	0	0	0	0	0	22	0	0	0	0	0	0	0	1
Δ6	c.213_301del89	FS	CΔ	44	0	40	0	0	65	35	0	0	19	0	0	0	0	0	0	
6q	c.293_301del9	IF	SDSΔ	0	44	431	371	695	202	116	28	124	7	62	0	0	0	0	0	
8p	c.442_444del3	IF	SASΔ	197	7022	15079	10883	8566	7283	4287	4657	6356	5024	4011	P	P	P	P	P	1
Δ8	c.442_547del106	FS	CΔ	0	0	10	0	0	0	0	55	92	0	0	0	0	0	0	0	
Δ8_9	c.442_593del152	FS	mCΔ	0	0	0	0	20	141	0	0	0	0	0	1	1	0	0	P	
Δ8_10	c.442_670del229	FS	mCΔ	0	88	0	0	0	2	0	0	0	0	0	1	1	1	0	P	
Δ8_16	c.442_498del4545	IF	mCΔ	88	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
▼ 94 bp int 8	c.548-297_548-204ins94	FS	C ▼	110	481	251	202	457	71	17	69	238	0	154	0	0	0	0	0	
▼ 97 bp int 8	c.548-300_548-204ins97	FS	C ▼	461	438	371	456	757	244	312	248	286	204	92	0	0	0	0	0	
Δ9	c.548_593del146	FS	CΔ	1295	9888	3642	12334	17110	2295	7303	9301	2092	7042	8689	P	1	P	0	P	1
Δ9_10*	c.548_670del123	IF	mCΔ	4542	19558	562	1063	2261	998	2165	3362	940	1561	5109	P	P	P	P	P	1
Δ9_11	c.548_4096del3549	IF	mCΔ	1009	16079	100	270	454	2	0	413	22	0	923	P	1	P	0	0	1
Δ9_12	c.548_4185del3638	FS	mCΔ	0	0	10	0	0	0	0	165	16	0	0	1	1	1	0	0	

10a	c.594_21_594-1ins21	IF	SAS▼	4827	2056	1575	5113	594	430	4957	1157	665	465	1036	0	0	0	0	0	0
Δ10	c.594_670del77	FS	CΔ	154	1684	30	34	0	12	0	0	27	7	0	1	1	1	0	P	0
Δ10_11	c.594_4096del3503	FS	mCΔ	0	4922	70	101	131	22	0	0	0	0	513	1	1	1	0	0	0
Δ10_12	c.594_4185del3592	FS	mCΔ	0	416	0	0	0	0	0	0	0	89	0	1	1	1	0	0	0
Δ11	c.671_4096del3426	IF	CΔ	439	5163	60	337	104	69	185	303	265	71	246	1	1	1	0	0	1
Δ11_12	c.671_4185del3515	FS	mCΔ	0	0	0	0	0	10	23	0	0	0	0	1	0	1	0	0	0
Δ11_12, 13p	c.671_4188del3518	FS	mCΔ + SASΔ	0	175	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0
11q	c.788_4096del3309	IF	SDSΔ	6012	16451	823	928	1965	2971	1491	5883	1048	2557	7879	P	1	P	0	0	1
11 Δ3094	c.788_3881del3094	FS	Intronization	241	219	0	0	39	6	0	0	0	0	0	0	0	0	0	0	0
11 Δ3110	c.788_3897del3110	FS	Intronization	2041	700	40	67	7	32	0	220	32	223	41	1	1	0	0	0	0
11 Δ3240	c.788_4027del3240	IF	Intronization	614	0	80	34	0	0	0	0	0	0	0	1	1	0	0	0	0
13p	c.4186_4188del3	IF	SASΔ	263	1378	3652	3881	9294	2684	560	3114	3297	847	6217	P	1	P	0	P	0
Δ13	c.4186_4357del172	FS	CΔ	44	1006	70	405	111	36	69	138	86	141	482	1	1	0	0	P	0
Δ13, 14p	c.4186_4360del175	FS	CΔ + SASΔ	44	219	20	0	0	6	12	0	0	0	0	1	1	0	0	0	0
Δ13_14	c.4186_4484del299	FS	mCΔ	0	2231	0	675	333	0	0	0	76	0	0	0	0	0	0	0	0
Δ13_15	c.4186_4675del490	FS	mCΔ	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0
▼13A	c.4358-2785_4358-2720ins66	IF	C▼	6495	24721	3963	1974	7267	873	1196	703	1254	5953	4309	1	1	1	0	P	0
▼13A, 14p	c.4358-2785_4358-2720ins66 + c.4358_4360del3	IF	C▼ + SASΔ	6604	40494	130	169	166	75	12	14	49	123	21	1	1	1	0	0	0
▼13A, Δ14	c.4358-2785_4358-2720ins66 + c.4358_4484del127	FS	C▼ + CΔ	0	0	0	0	819	115	0	0	0	0	0	0	0	0	0	0	0
14p	c.4358_4360del3	IF	SASΔ	4125	3238	542	1282	917	333	335	358	303	892	533	P	P	P	P	P	0
Δ14	c.4358_4484del127	FS	CΔ	417	1488	1304	877	1429	607	1231	1529	773	1442	882	1	0	0	0	0	0
Δ14_15	c.4358_4675del318	IF	mCΔ	241	1663	10	67	52	44	0	55	38	0	144	1	1	0	0	0	0
Δ14_17	c.4358_5074del717	IF	mCΔ	132	919	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1
Δ15	c.4485_4675del191	FS	CΔ	285	766	642	911	1723	220	254	468	168	286	1231	1	1	0	0	P	0
Δ15_16	c.4485_4986del502	FS	mCΔ	0	1619	30	0	274	0	0	0	0	0	0	0	0	0	0	0	0
Δ15_17	c.4485_5074del590	FS	mCΔ	176	372	130	945	1214	8	162	55	103	0	718	1	1	1	0	P	1
Δ15_19	c.4485_5193del709	FS	mCΔ	88	744	0	0	228	4	0	0	0	0	0	1	1	1	0	0	0
Δ15_23	c.4485_5467del983	FS	mCΔ	0	175	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16a	c.4986+1_4986+65ins65	FS	SDS▼	570	44	5568	0	6687	944	2461	3541	740	4292	0	0	0	0	0	0	0
Δ17	c.4987_5074del88	FS	CΔ	88	1466	2358	337	1951	1265	503	7854	1762	1988	6217	1	1	0	0	P	0
Δ17_18	c.4987_5152del166	FS	mCΔ	0	613	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Δ17_19	c.4987_5193del207	IF	mCΔ	88	1006	0	0	75	0	532	262	0	0	0	0	0	0	0	0	0
Δ17_20	c.4987_5277del291	IF	mCΔ	0	0	0	0	0	107	0	0	0	0	0	0	0	0	0	0	0
Δ18	c.5075_5152del78	IF	CΔ	0	219	60	0	0	44	0	0	0	0	0	1	1	0	0	0	0
Δ18_19	c.5075_5193del119	FS	mCΔ	0	0	10	67	0	4	81	28	11	7	0	0	0	0	0	0	0
Δ18_20	c.5075_5277del203	FS	mCΔ	0	0	30	0	0	4	0	0	0	0	0	0	0	0	0	0	0
▼143 bp int 19	c.5194-1231_5194-1089ins143	FS	C▼	132	0	20	0	0	381	324	138	573	0	431	0	0	0	0	0	0
▼146 bp int 19	c.5194-1234_5194-1089ins146	FS	C▼	373	197	80	101	653	40	0	289	32	19	421	0	0	0	0	0	0
Δ20	c.5194_5277del84	IF	CΔ	0	481	191	607	0	60	0	2618	0	0	0	1	1	1	0	0	0
Δ21	c.5278_5332del55	FS	CΔ	1711	8226	492	776	1765	389	376	1226	195	22	1057	1	1	1	0	P	0
Δ21_22	c.5278_5406del129	IF	mCΔ	88	438	0	0	0	0	549	0	0	0	0	1	1	1	0	P	0
▼129 bp int 21	c.5332+873_5332+1001ins129	FS	C▼	1711	263	592	641	1250	230	618	400	432	342	657	0	0	0	0	0	0
▼119 bp int 21	c.5333-706_5333-588ins119	FS	C▼	66	328	90	0	65	28	52	179	11	45	318	0	0	0	0	0	0
Δ22	c.5333_5406del174	FS	CΔ	4739	16451	1585	4590	3064	4405	11919	7372	3464	3872	5160	1	P	1	0	P	0
23a	c.5407-9_5407-1ins9	IF	SAS▼	44	66	10	0	39	35	168	55	68	74	0	0	0	0	0	0	0
Δ23	c.5407_5467del61	FS	CΔ	263	613	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0
Sum of identified variants in each sample (mPCR pool)				54	71	56	46	51	64	49	48	50	40	41						
Number of identified variants in analyzed cell lines / tissue type				76		67		72		54										

potentially result in a translation of BRCA1 protein isoforms altering its biological functions, as demonstrated for Δ14_15 and Δ17_19 previously (Sevcik et al., 2012; Sevcik et al., 2013). Nine out of 11 “ubiquitous” in-frame variants (Δ5; 8p; Δ9_10; Δ11; 11q; 13p; ▼13A; ▼13A, 14p) represented known ASVs while 6q and the highly expressed 10a variant were surprisingly not scored previously. The remaining four of the 25 in-frame variants (Δ3_5; Δ9_11; Δ14_15; 23a) were detected in most of the analyzed samples. Interestingly, Δ17_19 was detected in the cell lines and mammary tissue samples of BRCA1-mutation carriers and BC patients, but in no control sample.

Our approach enabled a direct quantification of 29 SDS/SAS

variants (including 17 mixed biotypes). Twelve SDS/SAS-only biotype variants were identified. Most of them (10 out of 12) were identified as “ubiquitous”, and only the 16a and 23a splicing variants were not present in some analyzed samples. Out of 17 more complex splicing mRNA BRCA1 isoforms (mixed biotypes), only 1Aq, Δ2 and ▼13A, 14p were “ubiquitous”, while the other variants occurred rather rarely (Supplementary Table S6). Besides the 11q splicing variant, which lacks 3,309 bp from exon 11 and its identification was done with a specific forward primer, the other 28 relatively short indels were co-amplified stoichiometrically alongside the corresponding canonical splicing variant (Fig. 1H) and therefore we were able to quantify their relative

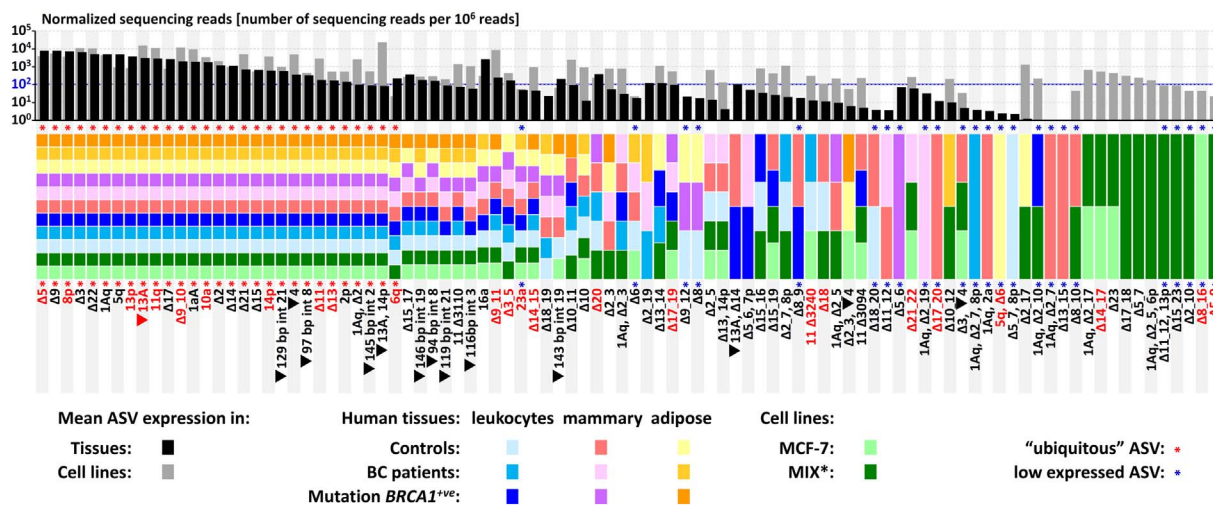


Fig. 2. Qualitative description of the presence of 94 identified BRCA1 ASVs (red letters indicate in-frame variants) in analyzed cDNA sample pools (colour bars). The grey-scale graph (upper part) shows the mean expression (in normalized reads per 10^6 reads) in human tissues (black) and cell line samples (grey). Red asterisks indicate ubiquitous variants (Table 1); blue asterisks indicate variants with low expression ($< 10^2$ normalized reads averaged in both tissue and cell samples).

*Mix of cDNAs from EM-G3, HeLa, and MDA-MB-231 cell lines. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

expression (compared with other existing variants in the particular exon-exon junction including wt transcript) directly from sequencing reads.

To test biological reproducibility, we compared the presence of BRCA1 ASVs in two independently analyzed sets of control samples, each consisting of leukocytes, mammary tissue and adipose tissue cDNA sample pools from eight individuals (Fig. 3A). Altogether, 74/94 BRCA1 ASVs were identified in at least one tissue control sample, while 20/94 were not present in any of them. Thirty-five variants (Fig. 3B) were consistently present (or absent – $\Delta 17_{20}$ in leukocytes and adipose tissue, $11\Delta 3240$ in mammary and adipose tissue, and $\Delta 20$ in adipose tissue) in the analyzed tissue type biological replicate. These 35 variants included the majority of “ubiquitous” ASVs as they were detected with high mean coverage per variant and sample (371 and 1496 in absolute and normalized sequencing reads, respectively). The remaining 39 variants (shown in Fig. 3C), discordantly expressed in paired biological replicates, represented low expressed events with low mean coverage per variant and sample (11 and 41 in absolute and normalized sequencing reads, respectively). The differences in sequencing coverage between 35 consistently present and 39 discordantly expressed variants are shown in Fig. 3D.

4. Discussion

An accurate description of ‘naturally occurring’ ASVs is a prerequisite to understanding their biological significance. RNA-sequencing (RNA-seq) of human RNA samples revealed that 90% of multi-exon genes undergo alternative splicing (Wang et al., 2015). While RNA-seq represents a superior tool for qualitative and quantitative transcriptome analyses, including ASV identification (Byron et al., 2016), it is unsuitable for small-scale projects targeting a few or a single gene. Moreover, RNA-seq analyses of low expressed transcripts require the sequencing of up to 100 million mapped reads and sophisticated bioinformatics instruments (Wang et al., 2015; Conesa et al., 2016).

A pioneering systematic description of BRCA1 ASVs was made by Orban and Olah (2003) who reviewed 23 BRCA1 ASVs known in 2003. Recently, Colombo et al. (2014) identified 63 BRCA1 ASVs by an analysis of 38 blood-derived samples and one healthy breast tissue sample. Subsequently, Romero et al. (2015) revealed 54 BRCA1 ASVs in an analysis of 70 breast tumor samples, four breast samples from healthy individuals and 72 blood-derived samples. Two later studies described the characterization of BRCA1 ASVs by capillary electrophoresis, which

required further cloning or sequencing of fragments containing splice junction events in order to identify the presented peaks. However, only the in silico imputation has been used to explain the peak pattern observed in capillary electrophoresis for a subset of events (Colombo et al., 2014).

Overall, 42 out of 94 BRCA1 splicing events described by our approach had not been identified in previous studies (Supplementary Table S5) which we used to compare the obtained results (Colombo et al., 2014; Romero et al., 2015; Orban & Olah, 2003).

The most common biotypes identified in our study (59/94; 63% variants) were cassette and multicassette ASVs (Supplementary Table S5). We found 27 cassette ASVs that included all 17 variants described in Colombo's study and 10 novel variants (eight intron exonizations, $\Delta 6$ and $\Delta 8$). Of 32 multicassette biotype ASVs ascertained in our study, 16 were described previously. We did not detect four multicassette ASVs ($\Delta 14_{18}$; $\Delta 14_{19}$; $\Delta 21_{23}$; $\Delta 22_{23}$) reported by Colombo's study as minor variants.

The second most frequent biotype variants were SDS/SAS (12/94; 13% variants). Besides nine described by Colombo et al. (2014), we found another four variants containing the exonizations of adjacent intronic regions (“ubiquitous” in-frame 6q and 10a, rare in-frame 23a, and a frameshift 16a) in all analyzed patient tissue types.

We found three large intronizations affecting exon 11, including two described by Colombo et al. (2014) previously, and the sparsely expressed frameshift variant $11\Delta 3094$. We did not target terminal modifications involving the alternative exon 1B and IRIS in our analysis.

Furthermore, we recorded 20 mixed biotype variants including two “ubiquitous” (1Aq, $\Delta 2$ and $\nabla 13A, 14p$). Nine mixed biotype variants were described previously and 11 rare were novel. We did not find six variants detected in Colombo's study, including three variants with untested alternative exon 1B, and another three (1Aq, 2p; 1Aq, $\Delta 2_3$, $\nabla 4$; and $\Delta 10_{13p}$) previously described as minor.

The most complex SAS/SDS events affected the non-coding 5' untranslated region (at the exon 1A-2 junction). The 1Aa variant containing an insertion of 89 nucleotides prevailed in cell line samples. The dominant variant in all analyzed tissue samples was wt exon 1A accompanied by the shortened variant exon 1Aq (in approximately one-third of all mapped reads). The expression of three other SAS variants 8p; 14p; 13p (lacking the CAG nucleotides at the 5'-end of an exon) was ~10% of all sequencing reads in most of the analyzed samples. These variants rank among the NAGNAG tandem acceptors, a common kind of ASVs resulting in single amino acid exclusion (Sinha et al., 2009). The

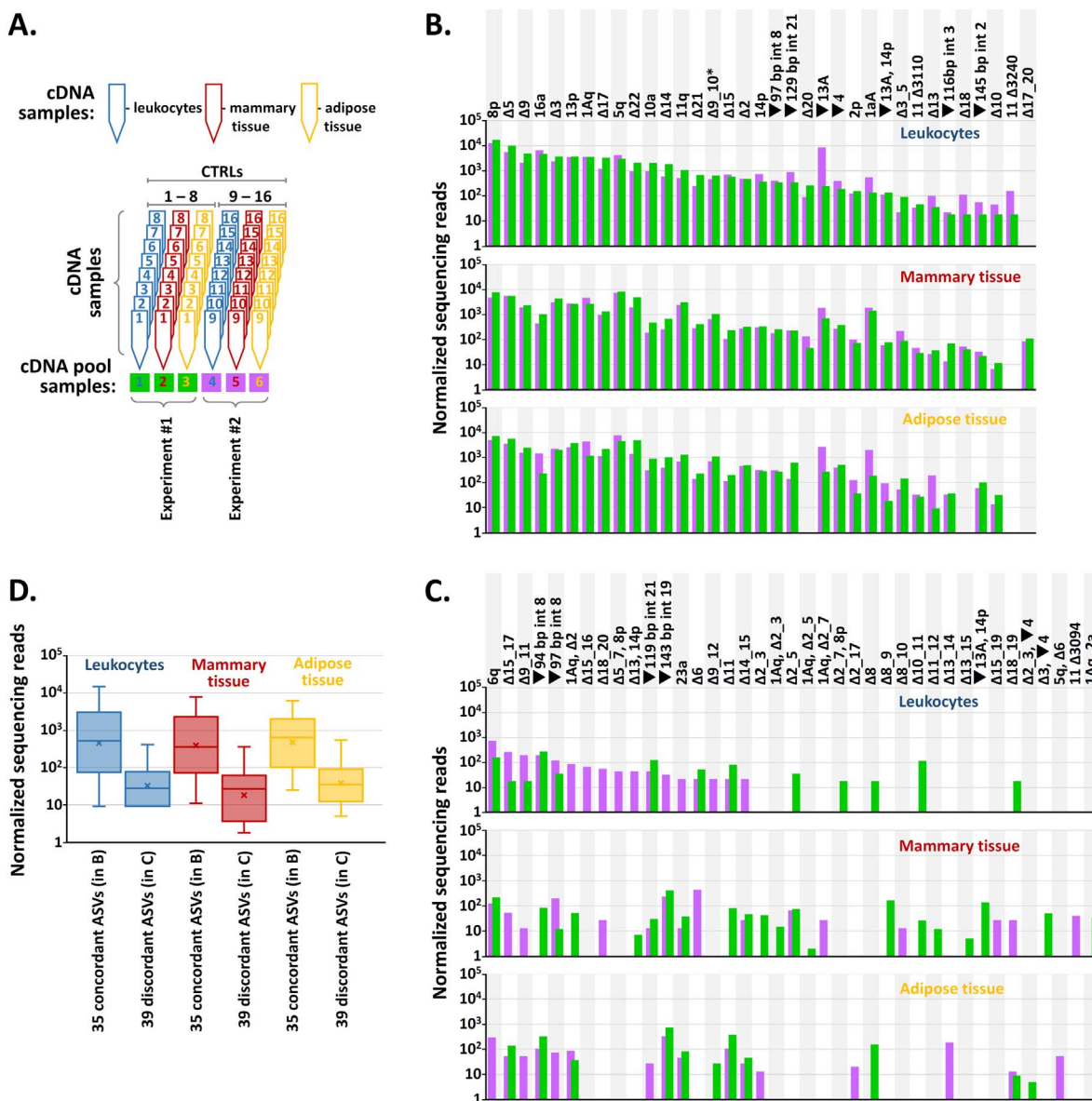


Fig. 3. The reproducibility test of the method in biological replicates involved an analysis of independent cDNAs from three types of tissue obtained from 16 control individuals. Panel A (adjusted from Fig. 1C) shows the arrangement of experiments #1 and #2, each consisting of NGS analyses from leukocytes, mammary tissue, and adipose tissue proceeded and sequenced independently by a pipeline described in the Method section. Graphs B and C compare the numbers of normalized sequencing reads (in log scale) for BRCA1 ASVs (listed in Supplementary Table 4) expressed in leukocytes, mammary and adipose tissues in two independent sets (Experiment #1 and #2) of control samples obtained from 16 individuals. The samples from control individuals 1–8 were analyzed in Experiment #1 (green bars), while the samples from control individuals 9–16 were analyzed in Experiment #2 (violet bars). Panel B shows the expression of 35 fully reproducible ASVs that gave concordant results in all analyzed biological duplicates. Panel C shows the expression of 39 non-fully reproducible ASVs that gave discordant results in at least one biological duplicate. The expressions of 35 fully reproducible ASVs were substantially higher than the expression of 39 non-fully reproducible ASVs as shown in panel D. The box plot charts show the values of sequencing normalized coverage (in log₁₀ scale) for 35 fully reproducible ASVs (shown in B) and 39 non-fully reproducible ASVs (shown in C) in the analyzed tissues. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

other variants were minor or rare, with the exception of five ASVs (10a; 16a; 1Aq, Δ2; Δ13, 14p; and ▼13A, 14p) expressed with a higher proportion in cell line samples.

The presented approach showed satisfactory reproducibility documented at the level of mPCR amplification (Fig. 1E and G) and also in the analysis of biological replicates from two sets of tissue samples from control individuals (Fig. 3). We suppose that discrepancies in the detection of individual ASVs in biological replicates resulted rather from differentially expressed BRCA1 ASVs in the individual RNA samples (mixed in the mPCR pools) than from analytical variability, because reproducibility was strongly positively correlated with the level of variant expression (i.e. coverage; Fig. 3D). Our analysis, performed in native tissues and cells (with unmodified nonsense-mediated decay pathway), revealed 48 frame-shift variants. However, their ratio to in-

frame variants was strongly reduced in the subset of “ubiquitous” variants (48/25 in the entire set of BRCA1 ASVs and 13/11 in “ubiquitous” variants). It has to be noted that BRCA1 mRNA is expressed at low levels, a few tenths of copies per cell in MCF7 cells (Lee et al., 2014). Many of newly identified ASVs were represented by a low number of reads (Fig. 2), indicating that they were probably expressed in a few copies per cell or present only in a subset of cDNA samples in the analyzed cDNA pool. We suppose that at least some of them may represent stochastic noise determining the number of alternative isoforms and their abundance (Melamud & Moul, 2009).

Deletions in isoform 1 (NM_007294.3) are the most frequently described BRCA1 ASVs. The longest well described ASV intron exonization is ▼4 (116 bp from intron 3). The predicted length of PCR amplicon covering ▼4 was 263 bp in our analysis, while the shortest

predicted amplicons (covering variants $\Delta 8_{17}$; 11q, $\Delta 12_{17}$; $\Delta 13_{17}$; $\Delta 17$; $\Delta 19_{21}$) were 60 bp long. Therefore, the setup of mPCR and the size-selection protocol were targeted to enrich amplicons with mean fragment length of 100–150 bp in order to disclose majority of putative ASVs. We were able to identify splicing events covered by PCR fragments ranging between 60 and 287 bp. One of the shortest identified amplicons was “ubiquitous” variant $\Delta 17$, while the variant 11 Δ 3094 was characterized from PCR product of 287 bp (the above mentioned variant $\nabla 4$, identified from 263 bp amplicon, occurred as “ubiquitous”). These findings indicate the range of amplicon lengths (60–263 bp) which can be analyzed under defined conditions.

In conclusion, the mPCR/NGS approach enables direct identification of all biotype classes of splicing events, including mixed biotypes containing exonizations of flanking intronic sequences. The analysis of BRCA1 mRNA revealed the broadest spectrum of its splicing variants, including their distribution in the analyzed human tissue and cell line samples. Similar to most other methods (including recent RNA-seq analyses), the analysis is not able to identify possible combinations of splicing events affecting both 5' and 3' portions of the large BRCA1 transcripts. We are also aware that our approach could miss large deep intronic exonizations (substantially exceeding the targeted PCR amplification and/or range of size selection). We would like to emphasize that the described method can be easily adopted for an analysis of any gene of interest in order to identify its ASVs, not only in human samples. Additionally, we suppose that our approach may represent an interesting option for the functional classification of VUS introducing aberrant splicing with modified protocol using individual (instead of pooled) cDNA template for mPCR step (limited to the region of interest).

Funding

This work was supported by the Grant Agency of the Czech Republic [grant number P301-12-1850], Charles University projects [PROGRES Q28/LF1, and SVV2017/260367], and a Ministry of Education, Youth and Sports project [The National Center for Medical Genomics LM2015091].

Disclosure statement

The authors have no conflicts of interest to disclose.

Acknowledgements

We thank The National Center for Medical Genomics for their instrumental and technical support with NGS sequencing and data analyses.

Authors' contributions

Study design (ZK, PK), experimental procedures (JH, FL, HH, KH, MJ, JSo), bioinformatics (PZ, VS), sample collection and characterization (OM, DP, MV, JSe), data analysis (JH, PK), manuscript preparation (PK, ZK, JH, JSe), manuscript approval (all authors).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.gene.2017.09.025>.

References

- Baralle, D., Buratti, E., 2017. RNA splicing in human disease and in the clinic. *Clin. Sci. (Lond.)* 131, 355–368.
- Bentley, D.L., 2014. Coupling mRNA processing with transcription in time and space. *Nat. Rev. Genet.* 15, 163–175.
- Brozova, M., Kleibl, Z., Netikova, I., Sevcik, J., Scholzova, E., Brezinova, J., Chaloupkova, A., Vesely, P., Dundr, P., Zadinova, M., et al., 2007. Establishment, growth and in vivo differentiation of a new clonal human cell line, EM-G3, derived from breast cancer progenitors. *Breast Cancer Res. Treat.* 103, 247–257.
- Bustin, S.A., Benes, V., Garson, J.A., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M.W., Shipley, G.L., et al., 2009. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Chem. Chem.* 55, 611–622.
- Byron, S.A., Van Keuren-Jensen, K.R., Engelthaler, D.M., Carpten, J.D., Craig, D.W., 2016. Translating RNA sequencing into clinical diagnostics: opportunities and challenges. *Nat. Rev. Genet.* 17, 257–271.
- Caminsky, N., Mucaki, E., Rogan, P., 2015. Interpretation of mRNA splicing mutations in genetic disease: review of the literature and guidelines for information-theoretical analysis. *F1000Res* 18, 282.
- Cheon, J.Y., Mozerker, J., Cook-Deegan, R., 2014. Variants of uncertain significance in BRCA: a harbinger of ethical and policy issues to come? *Genome Med.* 6, 121.
- Colombo, M., Blok, M.J., Whiley, P., Santamarina, M., Gutierrez-Enriquez, S., Romero, A., Garre, P., Becker, A., Smith, L.D., De Vecchi, G., 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, 3666–3680.
- Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szczesniak, M.W., Gaffney, D.J., Elo, L.L., Zhang, X., et al., 2016. A survey of best practices for RNA-seq data analysis. *Genome Biol.* 17, 13.
- Kleibl, Z., Kristensen, V.N., 2016. Women at high risk of breast cancer: molecular characteristics, clinical presentation and management. *Breast* 28, 136–144.
- Kleiblova, P., Dostalova, I., Bartlova, M., Lacinova, Z., Ticha, I., Krejci, V., Springer, D., Kleibl, Z., Haluzik, M., 2010. Expression of adipokines and estrogen receptors in adipose tissue and placenta of patients with gestational diabetes mellitus. *Mol. Cell. Endocrinol.* 314, 150–156.
- Lee, K., Cui, Y., Lee, L.P., Irudayaraj, J., 2014. Quantitative imaging of single mRNA splice variants in living cells. *Nat. Nanotechnol.* 9, 474–480.
- Melamud, E., Moul, J., 2009. Stochastic noise in splicing machinery. *Nucleic Acids Res.* 37, 4873–4886.
- Orban, T.I., Olah, E., 2003. Emerging roles of BRCA1 alternative splicing. *Mol. Pathol.* 56, 191–197.
- Rahman, N., 2014. Realizing the promise of cancer predisposition genes. *Nature* 505, 302–308.
- Romero, A., Garcia-Garcia, F., Lopez-Perolio, I., Ruiz de Garibay, G., Garcia-Saenz, J.A., Garre, P., Ayllon, P., Benito, E., Dopazo, J., Diaz-Rubio, E., et al., 2015. BRCA1 alternative splicing landscape in breast tissue samples. *BMC Cancer* 15, 219.
- Sevcik, J., Falk, M., Kleiblova, P., Lhota, F., Stefancikova, L., Janatova, M., Weiterova, L., Lukasova, E., Kozubek, S., Pohlreich, P., et al., 2012. The BRCA1 alternative splicing variant Delta14-15 with an in-frame deletion of part of the regulatory serine-containing domain (SCD) impairs the DNA repair capacity in MCF-7 cells. *Cell. Signal.* 24, 1023–1030.
- Sevcik, J., Falk, M., Macurek, L., Kleiblova, P., Lhota, F., Hojny, J., Stefancikova, L., Janatova, M., Bartek, J., Stribrna, J., et al., 2013. Expression of human BRCA1Delta17-19 alternative splicing variant with a truncated BRCT domain in MCF-7 cells results in impaired assembly of DNA repair complexes and aberrant DNA damage response. *Cell. Signal.* 25, 1186–1193.
- Sinha, R., Nikolajewa, S., Szafranski, K., Hiller, M., Jahn, N., Huse, K., Platzer, M., Backofen, R., 2009. Accurate prediction of NAGNAG alternative splicing. *Nucleic Acids Res.* 37, 3569–3579.
- Sloan, C.A., Chan, E.T., Davidson, J.M., Malladi, V.S., Strattan, J.S., Hitz, B.C., Gabdank, I., Narayanan, A.K., Ho, M., Lee, B.T., et al., 2016. ENCODE data at the ENCODE portal. *Nucleic Acids Res.* 44, D726–732.
- Soukari, O., Gaildrat, P., Hamieh, M., Drouet, A., Baert-Desurmont, S., Frebourg, T., Tosi, M., Martins, A., 2016. Exonic splicing mutations are more prevalent than currently estimated and can be predicted by using in silico tools. *PLoS Genet.* 12, e1005756.
- Tavtigian, S.V., Chenevix-Trench, G., 2014. Growing recognition of the role for rare missense substitutions in breast cancer susceptibility. *Biomark. Med.* 8, 589–603.
- Vondruskova, E., Malik, R., Sevcik, J., Kleiblova, P., Kleibl, Z., 2008. Long-term BRCA1 down-regulation by small hairpin RNAs targeting the 3' untranslated region. *Neoplasma* 55, 130–137.
- Wang, J., Ye, Z., Huang Tim, H.M., Shi, H., Jin, V., 2015. A survey of computational methods in transcriptome-wide alternative splicing analysis. *Biomol. Concepts* 6, 59–66.
- Whiley, P.J., de la Hoya, M., Thomassen, M., Becker, A., Brandao, R., Pedersen, I.S., Montagna, M., Menendez, M., Quiles, F., Gutierrez-Enriquez, S., et al., 2014. Comparison of mRNA splicing assay protocols across multiple laboratories: recommendations for best practice in standardized clinical testing. *Clin. Chem.* 60, 341–352.