

UNIVERZITA KARLOVA
3. LÉKAŘSKÁ FAKULTA



Autoreferát disertační práce

Integrace molekulární onkologie a *in silico* metod pro návrh preventivních opatření v chemoterapii nádorů prsu

Integration of molecular oncology and *in silico* methods for formulation of preventive measures in chemotherapy of breast cancer

PharmDr. Mária Kováčová

Praha, 2024

Doktorské studijní programy v biomedicině

Univerzita Karlova a Akademie věd České republiky

Studijní obor: Preventivní medicína a epidemiologie

Předseda oborové rady: doc. MUDr. Jan Polák, Ph.D.

Školící pracoviště: Státní zdravotní ústav, Oddělení toxikogenomiky

Autor: PharmDr. Mária Kováčová

Školitel: prof. RNDr. Pavel Souček, CSc

Autoreferát byl rozeslán dne 26.7.2024

S disertací je možno se seznámit na děkanátě 3. lékařské fakulty Univerzity Karlovy v Praze.

OBSAH

1. Souhrn.....	4
1.1. Summary.....	5
2. Úvod.....	6
2.1. Epidemiologie a etiologie karcinomu prsu	6
2.2. Prevence karcinomu prsu.....	8
2.3. Bioinformatika, studie <i>in silico</i>	10
2.3.1. Bioinformatické databáze	10
2.3.2. Bioinformatické nástroje.....	12
2.4. Hypotéza: Cíle řešení práce	15
3. Materiál, zdrojová data a metody.....	16
3.1.1. Materiál a pacienti.....	16
3.1.2. Bioinformatické algoritmy a metody pro funkční predikci častých variant	16
3.1.3. Bioinformatické algoritmy a metody pro funkční predikci vzácných variant	16
3.1.4. Farmakogenomická data – časté varianty	17
3.1.5. Farmakogenomická data – vzácné varianty	17
3.1.6. Bioinformatické algoritmy pro funkční predikci genu <i>KIF14</i>	17
3.1.7. Statistické metody	18
4. Výsledky	18
4.1. Výsledky pilotní studie častých variant – bioinformatická analýza	18
4.1.1. Datamining - farmakogenomické databáze	18
4.1.2. <i>In silico</i> predikce.....	18
4.2. Výsledky validační studie častých variant.....	19
4.3. Výsledky studie proveditelnosti – vzácné varianty	21
4.4. Výsledky pilotní studie – varianty genu <i>KIF14</i>	24
5. Diskuse metodických postupů a výsledků	25
5.1.1. Pilotní a validační studie – bioinformatická analýza	25
5.1.2. Predikce vzácných variant	28
5.1.3. Predikce variant genu <i>KIF14</i>	30
5.2. Závěry	31
5.2.1. Validační studie	31
5.2.2. Studie vzácných variant	32
5.2.3. Studie variant genu <i>KIF14</i>	32
Seznam použité literatury	33
Seznam publikací.....	43

1. Souhrn

Karcinom prsu, geneticky heterogenní onemocnění, zůstává u žen hlavní příčinou úmrtí, a přestože mortalita postupně ale pomalu klesá, incidence opětovně roste. Terciární a kvartérní prevence se zabývá studiem genetických rozdílů u pacientů a účinků chemoterapie. Základní hypotéza pro tuto práci vychází z předpokladu evolučně, populačně a funkčně odlišných genetických variant v lidském genomu, které jsou veřejně dostupné v celosvětových databázích a využitelné pro další studie. Hypotéza je již využívána v řadě bioinformatických nástrojů a hlavním cílem této disertační práce bylo studium dostupných databází a bioinformatických nástrojů k usnadnění zpracování velkého množství farmakogenomických dat. Prostřednictvím masivního paralelního sekvenování jsme u českých pacientek s karcinomem prsu vyhodnotili časté zárodečné kódující a nekódující varianty panelu složeného z 509 genů, které mají vztah k metabolismu a eliminaci léčiv, k funkci buněčné smrti či signálních drah v klinické onkologii, za pomoci nových *in silico* nástrojů a zavedených farmakogenomických databází. Prioritizované varianty byly technicky a klinicky validovány na souboru 805 pacientek. Asociace rs2227291 v *ATP7A*, rs2293194 v *KCNAB1* (u pacientek v časném stadiu) a rs4376673 v *DFFB* s odpovědí na neoadjuvantní cytotoxickou léčbu poskytují nové cíle pro následné funkční studie s potenciálem přispět k terciární prevenci. U pacientek s lumenálními B nebo trojitě negativními nádory, často studovaná varianta rs1801160 v *DPYD*, významně souvisela s přežitím bez nemoci a představuje další provokační cíl s prognostickým potenciálem.

Při studiu vzácných kódujících variant těchto pacientek, námi nově navržená strategie s cílem vyvážit některé z nedostatků metod strojového učení, vyhodnotila pro geny *ABCC1*, *ABCC4*, *ABCB6*, *ATP7B*, *CYP2D6*, *CYP4F3* a *CFTR* významnou souvislost s odpovědí na léčbu, menopauzálním obdobím, dále s histologickým typem, molekulárním podtypem a diferencovaností (grade) karcinomu prsu. Prokázali jsme, že skupina pacientek s patogenními variantami v *CFTR* měla významně zkrácené přežití bez nemoci (Log Rank, $p = 0,002$) a celkové přežití (Log Rank, $p = 0,006$). Naše studie poskytuje další důkaz, že *CFTR* se potenciálně podílí na procesech ovlivňujících účinnost onkoterapie nebo na samotné maligní progresi onemocnění, jak dokládají recentní publikace. Terciární prevence v kontextu odpovědi na léčbu se teoreticky nabízí možným sledováním funkčnosti *CFTR* nebo přímým zavedením léčivých látek, jako jsou již dostupné modulatory *CFTR*, které do určité míry řeší některé defekty funkčních variant.

Při zaměření na všechny varianty v genu *KIF14*, *in silico* identifikovány funkční nekódující varianty (rs17448931 a rs3806362) byly klinicky ověřeny a představují další zpřesnění cílů v genu *KIF14* s potenciálním využitím v terciární prevenci.

1.1. Summary

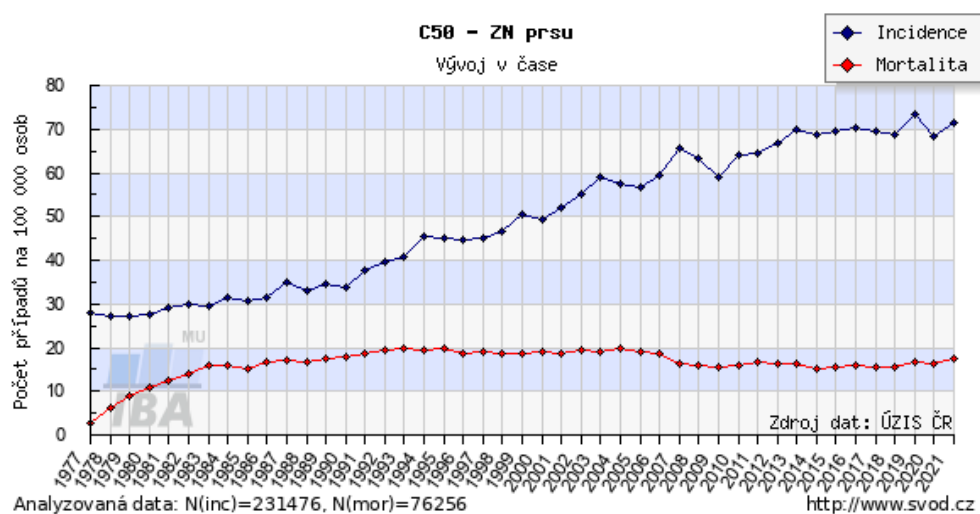
Breast cancer, a genetically heterogeneous disease, continues to be the leading cause of death among women. Although the mortality rate is gradually declining, the incidence is once again on the rise. Tertiary and quaternary prevention focus on studying genetic differences among patients and the effects of chemotherapy. The primary hypothesis for this research is based on the evolutionary, population, and functional differences in genetic variants within the human genome. This hypothesis is already employed in numerous bioinformatics tools. The main objective of this dissertation work was to study available databases and bioinformatics tools to facilitate the processing of extensive pharmacogenomic data. Through massive parallel sequencing, we evaluated frequent germline coding and non-coding variants in a panel of 509 genes related to drug metabolism and elimination, cell death functions, or signaling pathways in oncology among Czech breast cancer patients using novel *in silico* tools and established pharmacogenomic databases. Prioritized variants were technically and clinically validated in a cohort of 805 patients. The association of rs2227291 in *ATP7A*, rs2293194 in *KCNAB1* (in early-stage patients), and rs4376673 in *DFFB* with response to neoadjuvant cytotoxic therapy provides new targets for subsequent functional studies, potentially contributing to tertiary prevention. In patients with luminal B or triple-negative tumors, the frequently studied variant rs1801160 in *DPYD* was significantly associated with disease-free survival, representing an additional predictive target with prognostic potential. For rare coding variants in these patients, we employed a newly proposed strategy to address some limitations of machine learning methods. This approach revealed a significant association of variants in the *ABCC1*, *ABCC4*, *ABCB6*, *ATP7B*, *CYP2D6*, *CYP4F3*, and *CFTR* genes with treatment response, menopausal status, histological type, molecular subtype, and differentiation grade of breast cancer. Patients with pathogenic variants in *CFTR* had significantly reduced disease-free survival (Log Rank, $p = 0.002$) and overall survival (Log Rank, $p = 0.006$). Our study provides further evidence that *CFTR* may be involved in processes affecting the efficacy of cancer therapy or the progression of malignant disease, as supported by recent publications. Tertiary prevention in the context of treatment response could theoretically be achieved through the monitoring of *CFTR* functionality or the administration of therapeutic agents, such as currently available *CFTR* modulators, which partially address defects in functional variants.

By targeting all variants in the *KIF14* gene, functional non-coding variants (rs17448931 and rs3806362) identified through *in silico* analysis have been clinically validated, representing refined targets within the *KIF14* gene with potential applications in tertiary prevention.

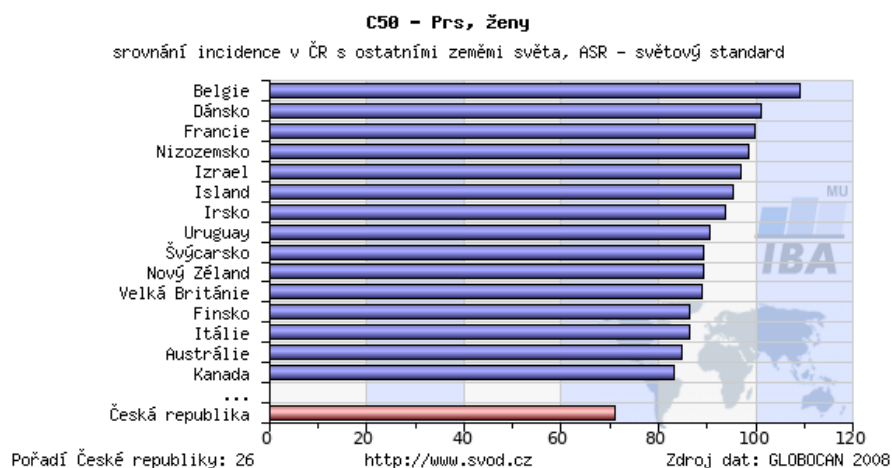
2. Úvod

2.1. Epidemiologie a etiologie karcinomu prsu

Karcinom prsu, geneticky heterogenní onemocnění, zůstává u žen hlavní příčinou úmrtí, a přestože mortalita postupně ale pomalu klesá, incidence od roku 2011 v ČR opětovně roste (viz Graf 2—1). Z celkového počtu 183 zemí se Česká republika nachází na 26. pořadí, tj. v první čtvrtině světové incidence (Graf 2—2). Navzdory všem dostupným preventivním opatřením, celosvětově nádor prsu tvoří 12,5% podíl (Malhotra et al. 2010), (Siegel et al. 2022) a v České republice, obdobně 11,7% podíl ze všech nových případů nádorů.



Graf 2—1 Časový vývoj hrubé incidence (počet nových případů na 100000 osob) a hrubé mortality (počet úmrtí na diagnózu na 100000 osob) pro C50 v celé populaci (www.svod.cz)



Graf 2—2 incidence diagnózy karcinomu prsu (C50) v ČR a ostatních zemí světa (www.svod.cz)

V ženské populaci je karcinom prsu druhým nejčastějším diagnostikovaným nádorem. Z nádorových onemocnění právě zhoubný novotvar prsu (C50) patřil za rok 2018 v počtu 1621 případů k nejčastější příčině úmrtí na nádorové onemocnění u žen (Český statistický úřad, zveřejněno 25.11.2019). Poslední statistické údaje z roku 2020 poukazují pouze na pomalu stabilizující tendenci v počtu 1710 případů. (Český statistický úřad, ÚZIS ČR 2021, ISSN 1210-9967). Podle molekulární klasifikace, karcinom prsu rozlišujeme na tři hlavní nádorové molekulární podtypy: hormonální receptor (ER/PR) pozitivní a zároveň ErbB2/HER2 negativní (70 % pacientek), ErbB2/HER2 pozitivní (15-20 %) a trojitě negativní (TNBC) tj., bez exprese všech tří receptorů (15 %) (Waks a Winer 2019). Molekulární podtypy vykazují významné rozdíly v progresi onemocnění, přičemž nejkratší přežití představuje bazální/triple negativní (ER-/PR-/HER2-) podtyp. Někteří autoři dále rozdělují hormonálně pozitivní podtyp (luminální) podle hodnot proliferativního markeru Ki67 na luminální A a luminální B, které rovněž vykazují rozdílnou prognózu (Malhotra et al. 2010). V předkládané práci studujeme genetický profil vrozených variant u pacientek s hormonálně pozitivním/HER2 negativním (luminální A, B) a TNBC podtypem nádorů.

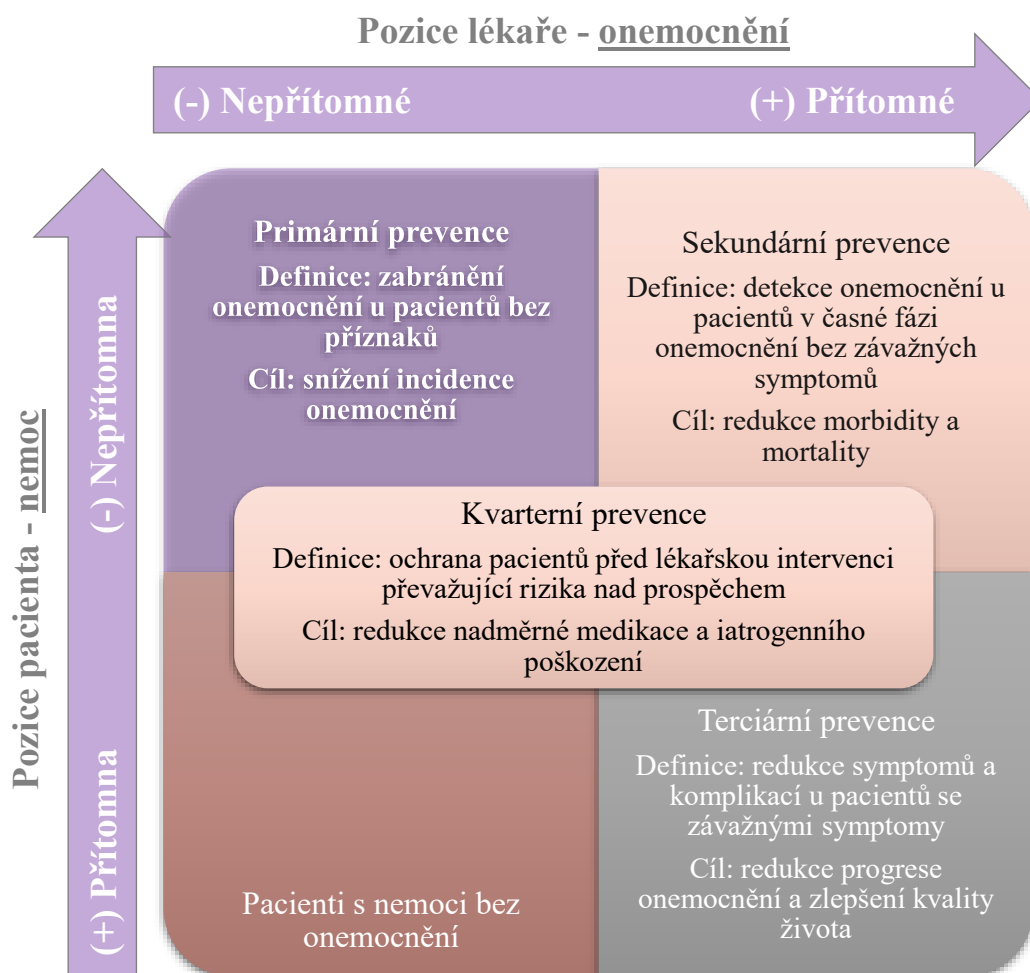
Etiologie

Etiologie karcinomu prsu je dána komplexní interakcí různých modifikovatelných i nemodifikovatelných faktorů, tj. genetickými, environmentálními, nutričními, hormonálními a dědičnými prvky. Mezi rizikové faktory patří předchozí výskyt karcinomu prsu, pozitivní rodinná anamnéza, časná menarche, pozdní menopauza, neaktivní způsob života, hormonální substituční terapie a expozice toxickým látkám (kouření, konzumace alkoholu). Expozice endokrinním disruptorům (xenoestrogeny a syntetické estrogény) obsaženým v chemických látkách v bezprostředním styku s okolím (Admoun a Mayrovitz 2022). Z pohledu rizika genetického portfolia se odhaduje, že mutace, přestavby nebo delece například v již dobře studovaném genu *BRCA1* a *BRCA2*, vysvětlují pouze 15 % familiárních karcinomů s přihlédnutím na rozdílnosti v zastoupení dokumentované pro etnické skupiny. V předkládané práci se celkově zaměřujeme na panel 509 genů důležitých v procesu farmakologie onkologických léčiv a progresu karcinogeneze.

2.2. Prevence karcinomu prsu

Podstatu úspěšnosti terapie karcinomu prsu lze posuzovat i z pohledu prevence. Plně funkční a neinvasivní primární prevence (tj., mimo profylaktické mastektomie) cílená na zabránění vzniku onemocnění zatím není k dispozici (Alaofi et al. 2018). Jediným účinným prostředkem je včasná diagnóza, v době, kdy nejsou plně rozvinuté symptomy onemocnění a který se řadí již k sekundární prevenci. Zde značně pomáhá indikované genetické vyšetření u žen s velmi vysokým rizikem, preventivní screeningová mamografie u asymptomatických žen v dispenzární péči a u žen od 45 let věku (případně od 40 let věku pro samoplátkyně), částečně i samovyšetření prsu. Sekundární prevence, zvyšuje šanci na zpomalení růstu nebo úplné zastavení růstu karcinomu včasnou terapií. Celkovým výsledkem takového preventivního opatření by mělo být snížení úmrtnosti na nádorové onemocnění (Kolak et al. 2017).

K příkladu invazivní primární prevence a sekundární prevence cílenou terapií řadíme odhalení predispozičních genů dědičné formy karcinomu prsu, jako jsou např. autozomálně dominantně děděné mutace v genech *BRCA1/2*. Efektivní detekce potenciálních nositelů mutace zde vede k profylakticko-preventivním opatřením, v případě již zmíněné mastektomie, a významnému zvýšení kvality a délky života díky včas podané terapii (Belizário et al. 2016). Cílem terciární prevence je tak zabránit progresi onemocnění, ztrátě soběstačnosti a s ní spojené snížení kvality života. Důležité postavení zastává i kvartérní prevence s cílem identifikovat pacienta ohroženého nadměrnou medikací, chránit ho před lékařskými invazivními metodami a navrhnout mu intervenci, která je eticky přijatelná (Kisling a M Das 2023; Martins et al. 2018) viz Obrázek 2—1. Terciární a kvartérní prevenci lze docílit studiem mechanismu karcinogeneze a účinků chemoterapie, včetně těch nežádoucích. Obecně, studium chemoterapie na úrovni farmakokinetického a farmakogenetického profilování se zaměřuje na distribuci, biotransformaci a eliminaci cytotoxického léčiva s důrazem na chemorezistenci na molekulární úrovni (O'Donnell a Ratain 2012). Na úrovni farmakodynamického a farmakogenomického studia chemoterapie se jedná o komplexní genetické profilování pacienta s onkologickým onemocněním, ideálně jak na úrovni vrozené variability, tak somatických mutací nádorových buněk.

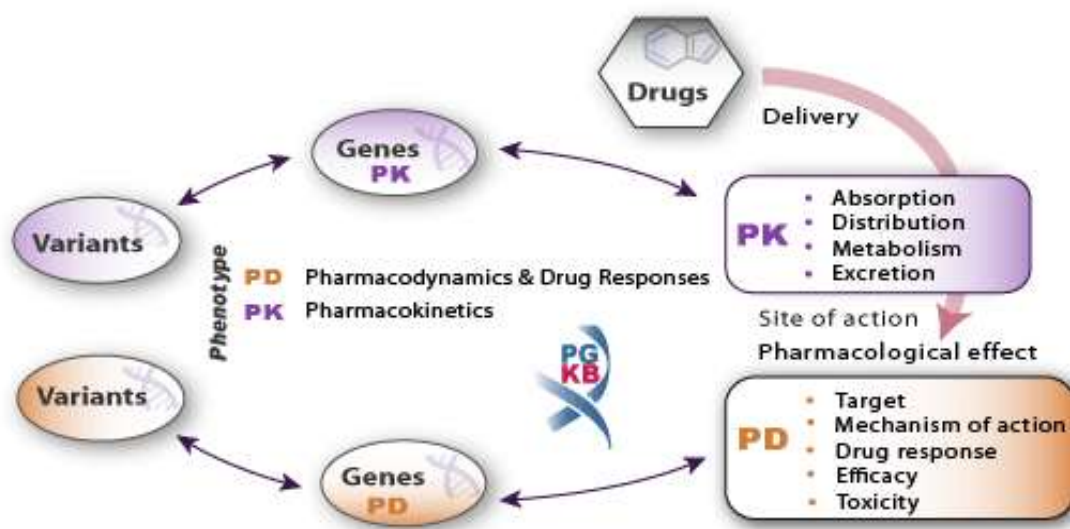


Obrázek 2—1 Nemoc ve vztahu k prevenci (převzato a upraveno (Martins et al. 2018))

Farmakogenetické přístupy jsou užitečné v případech, kdy léčivo interaguje s receptory nebo jinými "lékovými" cíli nebo podléhá metabolismu nebo transportu proteinu, zejména když jsou známy funkční polymorfismy v genech kódujících kandidátní proteiny. Tyto přístupy vedly v posledních letech k identifikaci několika klinicky použitelných farmakokinetických párů typu gen-léčivo, např. mutací *CYP2D6* (Alshabeeb et al. 2022; Carr et al. 2021; Zhang et al. 2015), které metabolizují tamoxifen, a mutací *SLCO1B1*, které ovlivňují clearance metotrexátu u vzácných variant s klinickým dopadem větším než u častých variant (Ramsey et al. 2012; Alshabeeb et al. 2022).

Farmakogenomika se zabývá studiem vlivu na osud léčiva genomu. Studuje rozdílné účinky potenciálních léčiv *in vitro* a *in vivo* na genovou expresi jako celek. Hodnotí expresní profily, variabilitu v mechanismu účinku potenciálních léčiv ve fázi screeningu. Oba termíny – farmakogenetika a farmakogenomika jsou často v publikacích zaměňovány a nejčastěji

adresovány termínem „farmakogenomika“ čehož využíváme i v této práci, kde se zaměřujeme na studium fenotypu a odpovědi organismu na léčivo. Fenotyp, jak farmakodynamický (PK), tak farmakokinetický (PD) se ve výsledku mohou projevit jak zvýšením/snížením účinnosti anebo zvýšenou toxicitou a vznikem nežádoucích účinků viz **Obrázek 2—2** (převzato z www.pharmgkb.org).



Obrázek 2—2 Ovlivnění mechanismu anebo cíle účinku léčiva vrozenou variantou genu

Náplní předkládané práce je studium dostupných bioinformatických databází a nástrojů k usnadnění zpracování velkého množství farmakogenomických dat, a tak k zautomatizování procesů predikce genetického profilu pacienta ve vztahu k terciární a kvartérní prevenci.

2.3. Bioinformatika, studie *in silico*

Biomedicínský výzkum je řízen neustálým navyšováním v schopnosti generovat data (Cabanillas et al. 2017). Na základě těchto dat bylo, a je možné objasnit spojení genetické variability s nemocí nebo s odpovědí pacienta na léčbu. Tato skutečnost se stala hlavním důvodem pro vznik nového vědného odboru – bioinformatiky, která na základě matematického modelování, statistiky a komputace predikuje funkční biologické konsekvence (Afgan et al. 2002).

2.3.1. Bioinformatické databáze

Od zavedení projektu lidského genomu v 90. letech 20. století byly akademickými institucemi, organizacemi, konsorcií a komunitami vytvořeny a udržovány databáze a znalostní databáze (tzv.

Knowledge based database (KBDB)), které shromažďují, uchovávají a umožňují vyhledávání záznamů týkající se genetických a klinických informací. Poskytují dostatek dostupných důkazů a faktů, které spolehlivě prokazují genotypovou a fenotypovou asociaci, jež vysvětluje funkční a patogenní význam genetických variací. Databázové zdroje využívá většina prediktivních algoritmů (v celé práci je také označujeme jako prediktory neboli nástroje) a jejich základní přehled je uveden níže.

2.3.1.1. Primární bioinformatické databáze

Nejčastěji používané databáze obsahují sekvenční informace, údaje o populačním měřítku, fenotypovou ontologii, klinické a experimentální důkazy. Sekvenční primární zdroje GenBank (Sayers et al. 2022), jehož hostitelem je National Institutes of Health (NIH), European Nucleotide Archive (Baker et al. 2000), Evropský institut pro bioinformatiku (European Molecular Biology Laboratory's European Bioinformatics Institute, EMBL-EBI), a DNA Data Bank of Japan (DDBJ) (Okido et al. 2022) jsou nejpoužívanější primární databáze, které uchovávají více než 2,5 miliard nukleotidových sekvencí pro více než 504 000 formálně popsanych druhů (Liu et al. 2022).

2.3.1.2. Populační zdroje

Populační zdroje poskytují informace umožňující zkoumat varianty mezi subpopulacemi s různým etnickým původem nebo stavem onemocnění. Kromě toho je frekvence minoritních alel (MAF) z těchto databází obvykle užitečným ukazatelem pro stanovení priorit kandidátních variant a jedná se taky o důležitý parametr využívající při vytváření predikčních modelů. V předkládané práci využíváme těchto projektů k anotaci frekvence variant pro soubor pacientek v pilotní studii ve srovnání s frekvenčními daty specifickými pro Českou populaci (NCMG) integrovanou do SQL databáze.

2.3.1.3. Fenotypové databáze

Fenotypové databáze popisují fenotypy a nemoci ve spojení s genetickými informacemi. Jsou navrženy tak, aby anotovaly klinické fenotypy a geny pomocí dobře strukturované, počítačově přívětivé, přesné a precizní terminologie. Celkově tyto databáze poskytují významné poznatky pro stanovení priorit a interpretaci genetických dat fenotypech s cílem komplexně přezkoumat vazbu genotyp-fenotyp. Příkladem je OMIM (Online Mendelian Inheritance in Man) (Amberger et al. 2019) a Orphanet (Aymé et al. 1998). Hlavní faktor, který komplikuje adekvátní predikci

je, že v různých databázích se překrývají varianty i s protichůdnou klasifikací klinického dopadu v důsledku nejednotných pravidel a subjektivního názoru různých kurátorů což značně komplikuje přístup prioritizace kandidátních variant.

2.3.1.4. Farmakogenomické databáze a nástroje

Podporu využití farmakogenomických dat doplňují specifická konsorcia a projekty. Konsorcium pro implementaci klinické farmakogenetiky (CPIC) vzniklo v roce 2009 jako společný projekt již zmíněné PharmGKB (<https://www.pharmgkb.org>) a síť Pharmacogenomics Research Network (PGRN) (<http://www.pgrn.org>). CPIC poskytuje klinické pokyny, které umožňují převést výsledky farmakogenetických laboratorních testů do použitelných rozhodnutí o předepisování konkrétních léčivých přípravků. K datu psaní této práce bylo již vytvořeno 30 klinických pokynů (<https://cpicpgx.org/genes-drugs>). Dalším projektem je Pharmacogenetics Working Group z Nizozemska (Royal Dutch Association for Advancement of Pharmacy) a The Pharmacogene Variation (PharmVar) Consortium (<https://www.pharmvar.org/>) centrální úložiště variability farmakogenů (PGx), které se zaměřuje na strukturu haplotypů a alelickou variabilitu (Gaedigk et al. 2021). Databáze PharmGKB a následně i PharmVar byly hlavní zdroj PGx dat předkládané práce. Dále byly využité data z fenotypových databází pro nežádoucí účinky v ADReCS-Target a PheWAS catalog s obsahem 3 144 jednonukleotidových polymorfismů (SNP) z katalogu GWAS NHGRI u 13 835 jedinců Evropského původu z pěti míst sítě Electronic Medical Records and Genomics (eMERGE) (Denny et al. 2013). Za zmínku také stojí DGIdb databáze shromažďující interakce mezi léky a geny tzv. "druggable genom" získané z více než třiceti důvěryhodných zdrojů (Freshour et al. 2021).

2.3.2. Bioinformatické nástroje

Prediktory vlivu variant lze rozdělit několika způsoby, například do kategorií sekvenčních a strukturních nástrojů a metaprediktorů, kterého pro zjednodušení vyžíváme i v této předkládané práci (Hassan et al. 2019a; Yazar a Ozbek 2022; Hassan et al. 2019b). Dalším často používaným rozdělením je na modely strojového učení (Machine Learning, ML) tj., modely s dohledem a bez dohledu a modely bez ML. Pro zpracování hlavního cíle bylo sestaven set z sekvenčních algoritmů (Align-GVGD, LRT, MutationAssessor, PolyPhen2, PROVEAN a SIFT) a strojového učení (CADD, MetaSVM, REVEL a VEST3).

2.3.2.1. Sekvenční algoritmy

Tato třída prediktorů je odvozena ze srovnávací genomiky, kdy u přirozeného výběru jsou změny aminokyselin v konzervativních sekvencích více „škodlivé“, což se určuje vyhledáváním homologních sekvencí napříč druhy oproti nehomologním pozicím, které jsou považovány za „tolerantní“ (Cooper a Shendure 2011).

Sekvenční sada

Ke stanovení prvního cíle předkládané práce (Hlavac et al. 2018) byla zvolena metoda interpretace predikce tzv. striktním konsenzem několika vybraných nástrojů. Pro zhodnocení nesynonymních variant (nsSNV) byla vytvořena sada 6 sekvenčních nástrojů, jejichž modely byly sestaveny právě pro hodnocení nsSNVs. Tento postup byl inovativní oproti již publikovaným studiím.

2.3.2.2. Strukturní algoritmy

Kromě primární struktury proteinu je pro jeho normální funkci zásadní také skládání a stabilita. První poznatky o variantách, které ovlivňují strukturu bílkovin a vedou k aberantním fenotypům, lze datovat do 50. let 20. století, kdy byla objevena záměna aminokyselin u molekuly hemoglobinu, která způsobuje srpkovitou anémii (Ingram 1957). Od té doby byly popsány tisíce mutací (Giardine et al. 2014), které ovlivňují zvýšení funkce (Jones et al. 1979) nebo snížení afinity ke kyslíku (Bonaventura a Riggs 1968), stabilitu (Martínez et al. 1977) a konformaci (Moo-Penn et al. 1988) hemoglobinu.

2.3.2.3. Metody strojového učení (Machine Learning, ML)

Přístupy této kategorie zohledňují jak dříve popsanou homologní sekvenci, tak informace o struktuře proteinu. Predikce využívají kombinaci informací o homologické sekvenci (např. konzervativní skóre) a strukturních vlastností, jako jsou hydropatie, polarita, torsní úhly a elektrostatické interakce. Navíc doplněné o energetické vlastnosti a biochemické vlastnosti, jako je plocha rozhraní přístupná rozpouštědлу (Kulshreshtha et al. 2016). Tyto rysy se někdy transformují nebo vybírají pro trénování modelu a poté jsou vyškoleny k rozlišování mezi známými funkčně škodlivými variantami (pozitivní soubor) a pravděpodobně neutrálními variantami (negativní kontrolní soubor). Často aplikovaným metaprediktorem je Meta-SVM jehož praktická využitelnost byla testována i na omických datech karcinomu prsu. Výsledkem byla identifikace souborů genů konzistentně spojených s onemocněními napříč studiemi. Zejména bylo zjištěno, že zjištěný soubor genů z omických dat TCGA je významně obohacen o

dráhy ABC transportérů, které jsou dobře známy jako kritické pro mechanismus vzniku karcinomu prsu mnoha ohledech. Tohoto faktu dokládá i řada studií z laboratoře školitele předkládané práce, kdy např. *ABCC8* a *ABCC11* korelovali s progresí a odpovědí na léčbu u pacientek s nádorem prsu (Kim et al. 2017; Hlaváč et al. 2013). MetaSVM/LR prediktor byl rovněž aplikován u obou našich studií jak pro časté, tak vzácné varianty (Hlavac et al. 2018; Kováčová et al. 2024).

2.3.2.4. Prediktory nekódující a regulační části genů

Převážná většina zárodečných variant se vyskytuje v nekódující části genomu a poměrově jen pouze malá část z nich je považována za funkční. Více důkazů však naznačuje, že varianty DNA mimo oblasti genomu kódující proteiny (zde označované jako nekódující varianty) ovlivňují lidské fenotypy, včetně onkologických onemocnění či odpovědí na léčbu (Schipper a Posthuma 2022). Pro lepší pochopení architektury nekódujícího genomu je k dispozici několik konsorcií obdobně jako u kódujících variant popsanych výše. Encyclopedia of DNA Elements (ENCODE Project Consortium 2012) a NIH Roadmap Epigenomics Mapping Consortium (Bernstein et al. 2010) jsou zaměřené na charakterizaci histonových modifikací a přístupnosti DNA na základě široké škály testů u 127 typů buněk. Ke stanovení prvního cíle předkládané práce (Hlavac et al. 2018) byly aplikovány následující in silico nástroje: dbNSFP (dbSNV), VEP (TFBP), MaxEntScan, IW scoring framework, RegulomeDB, TargetScan a PINES.

Závěrem literárního přehledu lze shrnout jedny z významných příkladů úspěšně aplikované prevence u pacientek s karcinomem prsu postavené na základu farmakogenomických dat a metod. Odhalení *BRCA* genu představuje signifikantní úspěch v sekundární prevenci, kdy včasná a náležitá diagnóza zvyšuje šanci na zajištění dostatečné odpovědi na terapii. Definování molekulárních subtypů nádorů a účinné moderní terapie léčivými přípravky (Herceptin, Tyverb a Perjeta) omezuje progresi onemocnění, ztrátě soběstačnosti a s ní spojené snížení kvality života v terciární prevenci. Úprava farmakoterapie v systému kvartérní prevence u pacientů s alterací v *DPYD* genu ohrožených nadměrnou medikací a následnou toxicitou. V předkládané práci představujeme nový koncept aplikace bioinformatických nástrojů u českých pacientek s molekulárně definovaným karcinomem prsu, k odhalení alterací genů vrozenými mutacemi souvisejícími s odpovědí na léčbu, relapsem či progresi onemocnění s cílem možného využití v prevenci.

2.4. Hypotéza: Cíle řešení práce

Základní hypotéza pro tuto práci vychází z předpokladu evolučně, populačně a funkčně odlišných genetických variant v lidském genomu a veřejně dostupných v celosvětových databázích. Hypotéza je využívána v řadě bioinformatických nástrojů a hlavním cílem práce bylo nalezení vhodného výběru počítačových algoritmů s možností predikce funkčního dopadu genetických variant v odpovědi na terapii karcinomu prsu. Základní bioinformatické nástroje jako je PolyPhen a SIFT již byly úspěšně aplikovány ve studiích na pracovišti školitele. Publikovaná literatura navíc dokládá téměř 20leté zkušenosti integrované bioinformatické analýzy ke prioritizaci genetických variant a genů. V této práci jsem se zaměřila na novější a složitější algoritmy a jejich kombinace vyžadující aplikaci programovacích jazyků a vyhodnocování výsledků i za pomoci strojového učení čili podoblasti umělé inteligence. Hlavním cílem bylo nalezení zárodečných variant ve vztahu k odpovědi na léčbu, klinicko-patologickým faktorům a progresi onemocnění. Následná identifikace cílené léčby na molekulární úrovni nese potenciál redukce relapsů a progresu onemocnění u terciární, potažmo kvartérní prevence. Na oddělení Toxikogenomiky SZÚ byl sestaven panel 509 farmakogenomicky významných genů a bylo provedeno cílené sekvenování všech kódujících a regulačních oblastí těchto genů u 105 pacientek s karcinomem prsu léčených klasickou chemoterapií založenou na antracyklinech, taxanech a nukleosidových analogích. Funkční hodnocení genetické variace u pacientek s nádorem prsu proběhlo za použití výše zmíněných algoritmů s korelací k farmakoterapii a celkové prognózy onemocnění. Téměř 20 000 variant bylo vyhodnoceno z funkčního i klinického hlediska za pomoci dostupných bioinformatických metod a veřejných farmakogenomických a genomických databází. Potenciálně škodlivé varianty genů predikovaných *in silico* a vyhodnocených na 105 pacientek byly následně ověřeny ve validační studii na souboru 805 pacientek.

Dílčí cíle:

1. Výběr algoritmů pro zpracování a zhodnocení kandidátních variant z testovací studie do validační studie *in silico* bioinformatickou analýzou, umožňující rychlejší a komplexnější predikci funkční konsekvence vrozených variant všech typů, tj. v kódující i nekódující části genomu.
2. Validace kandidátních variant z bioinformatické analýzy k potenciálnímu využití v terciární a kvartérní prevenci karcinomu prsu.
3. Výběr algoritmů pro hodnocení a selekci kandidátních genů vzácných variant za pomoci funkčních prediktorů do testovací studie (studie proveditelnosti).

3. Materiál, zdrojová data a metody

3.1.1. Materiál a pacienti

Materiál použitý v *in vitro* metodách k extrakci DNA a následného sekvenování nové generace (NGS) včetně vybraného panelu genů je popsán v publikaci, která je součástí disertační práce (Hlavac et al. 2018).

Pacientky, jejichž vzorky a data byly studovány, prostudovaly a podepsaly před vstupem do studie Informovaný souhlas pacienta a studie byla schválena etickou komisí Státního zdravotního ústavu v Praze a spolupracujících poskytovatelů zdravotní péče (FN Motol, Medicon a.s. a Nemocnice Atlas ve Zlíně).

3.1.2. Bioinformatické algoritmy a metody pro funkční predikci častých variant

Pro funkční predikci byly zárodečné varianty anotovány pomocí platform Annovar a VEP poskytující výsledky predikce pro zvolené nástroje. Software Annovar verze 2018 Apr 16 (obsahujícího databázi dbNSFP verze 3.2), pro lidský genom Genome Reference Consortium Human Build 37 (GRCh37/hg19) s preferencí „Filter-based annotation“ byl stažen a integrován systémem Unix případně Linux do interní databáze provozované přes relační databázové systémy MySQL a DBeaver. Software VEP (Variant Effect Predictor) verze v92.2, pro lidský genom (GRCh37/hg19) byl stažen a využíván ke zpracování dat v systému Unix případně Linux. Celkově k filtraci a anotaci variant byly použity programovací jazyky SQL v operačním systému Microsoft případně Bash (Unix shell) a AWK v operačním systému Unix. Použité příkladové kódy jsou uvedeny v seznamu disertační práce. Ke statistické analýze metody míry falešného objevu (false discovery rate, FDR) a grafickému zobrazení KM plotů byl použit programovací jazyk R verze 4.2.3 (survfit, ggsurvplot).

Získané výsledky tvořily základ funkční evidence seřazené dle váhy námi vytvořených pravidel pro prioritizaci variant následně určených k ověření ve validační fázi studie.

3.1.3. Bioinformatické algoritmy a metody pro funkční predikci vzácných variant

Bioinformatická funkční predikce (důkazy na úrovni vzácných variant a genů)

Pro predikci patogenity jsme vybrali tři nástroje založené na strojovém učení (REVEL, MetaSVM a CADD) a jeden nástroj založený na zachování sekvence (MutationAssessor) s prahem optimalizovaným pro PGx varianty (Kováčová et al. 2024). Výsledky predikce byly integrovány do interní relační databáze pomocí MySQL a/nebo DBeaver. U všech metod byl

výběr omezen na nesynonymní varianty s nízkou frekvencí (MAF mezi 0,05 a 0,01) a na vzácné varianty ($MAF \leq 0,01$) kromě PGx variant, kde byla filtrace rozšířena na všechny typy kódujících variant v genech s publikovanými klinickými důkazy. Prahové hodnoty pro predikci (binární nebo číselné) se řídily doporučenými hodnotami vyváženými pro optimální specifickou a citlivost, jak je definovali vývojáři nástroje.

Metoda in silico predikce a interpretace predikce

Výpočetní předpověď byla odvozena pro všechny nsSNV a rozdělena podle počtu nsSNVs na úroveň genu. Geny s nejvyšším počtem domněle patogenních variant byly dále zkoumány z hlediska klinické asociace u pacientek s karcinomem prsu ve studii. Se záměrem snížit počet falešně pozitivních předpovědí zvolenými *in silico* nástroji provozovanými samostatně byla do procesu filtrace integrována specifická sada tzv., Benign set nástrojů řízených umělou inteligencí (REVEL, VEST3 a MetaSVM) se shodnou předpovědí benigního následku. Absence předpovědi některým z nástrojů v kombinaci vyřadila varianty z tohoto souhlasného hodnocení (podrobně viz *Proces filtrace*) v disertační práci.

3.1.4. Farmakogenomická data – časté varianty

Integrace dat dostupných na stránkách PharmGKB (TSV soubory s hodnotami oddělenými tabulátory) do interní databáze v MySQL/DBeaver podléhala pravidlům hlavních vztahových identifikátorů (primárních a cizích klíčů) jako je „rsID“ tj., označení využívané ve veřejných databázích specifických pro lokus varianty např. Clinvar a Human Gene Mutation Database (HGMD) a „PA(ID)“ označení v PharmGKB databázi. Bližší popis obsahu souborů a příkladový kód k vytvoření interní databáze je uveden v disertační práci.

3.1.5. Farmakogenomická data – vzácné varianty

Pro získání farmakogenomických dat bylo provedeno vyhledávání variant souvisejících s PGx v databázích PharmVar a ClinVar.

3.1.6. Bioinformatické algoritmy pro funkční predikci genu KIF14

Bioinformatická analýza pro gen KIF14 byla provedena platformou SNPnexus (Dayem Ullah et al. 2018) s cílem predikce funkčního dopadu aminokyselinové záměny v proteinu kombinací nástrojů SIFT (Ensembl Variation 95) a PolyPhen2 (Ensembl Variation 95). Prahové hodnoty byly pro vybrané algoritmy nastaveny v souladu se studiemi výše tj., častých a vzácných variant.

3.1.7. Statistické metody

K realizaci všech cílů předkládané práce byly vrozené varianty získané z cíleného sekvenování rekalibrovány pomocí GATK 3.7. Hardy-Weinbergův test byl vypočten pomocí nástroje Bcftools 1.5 (Cambridge, UK). Pro statistické a funkční hodnocení byly vzaty v úvahu pouze varianty v Hardy-Weinbergově rovnováze ($p > 0,01$), méně než 50 % chybějících dat a s FILTER = PASS (VCF). Hladina významnosti $p < 0,05$ byla považována za statisticky významnou. Analýzy byly provedeny pomocí statistického programu SPSS v16.0 (SPSS, Chicago, IL, USA) a s použitím skriptu v R.

4. Výsledky

4.1. Výsledky pilotní studie častých variant – bioinformatická analýza

4.1.1. Datamining - farmakogenomické databáze

PharmGKB a PharmVar

Získané informace z PharmGKB, poskytly míru funkční významnosti klinické a veřejně dostupné evidence pro prioritizaci variant začleněných do validačního souboru. Dotazovaná repositoria zahrnovaly databáze PharmGKB (McDonagh et al. 2011) a PharmVar (ke dni analýzy k dispozici pouze pro geny CYP2C9, 2C19, 2D6) se souhrnnými výsledky statisticky významných variant uvedených v tabulkách disertační práce. Další využitou farmakogenomickou databází byla, an Adverse Drug Reaction Classification System-Target, ADReCS-Target (<http://bioinf.xmu.edu.cn/ADReCS-Target>) (Huang et al. 2018), která poskytuje komplexní informace o nežádoucích reakcích léčiv způsobených interakcí léčiv s proteinovými, genovými a genetickými variacemi. Data z databáze PheWas Resources (phenome-wide association studies) byla filtrována opětovně v souladu s farmakoterapií námi sledovaných pacientek, a tudíž pro antineoplastická léčiva a statisticky významné varianty uvedeny v disertační práci. Pro vybrané varianty databáze neuváděla fenotypovou asociaci ve statistické významnosti na hladině alfa 5%.

4.1.2. *In silico* predikce

Sekvenční sada nástrojů

Pro funkční predikci nesynonymních variant byla sestavena samostatná sada nástrojů využívající algoritmy sekvenčních srovnání. Pro interpretaci predikce byl použit přísný konsenzus ve výsledcích všech nástrojů za vyloučení všech variant s chybějící predikcí minimálně v jednom z nástrojů. Ze statisticky významných nekódujících variant ve vztahu k odpovědi na léčbu,

molekulárnímu subtypu či DFS bylo predikováno 6 variant nástroji MaxEntScan anebo TFBP (VEP). Pro tyto varianty však heterozygotní genotyp prezentoval statisticky výraznější vliv ve srovnání s oběma homozygoty, protože byly vyloučeny z konečného souboru statisticky významných variant určených k validaci. PINES poskytl seřazený seznam významnosti predikce nekódujících variant (nevážená hodnota) s funkční charakteristikou ovlivněnou daty pro jaterní tkáň (vážená hodnota) a játra jakožto hlavní cílový orgán biotransformace léčiv (Bodea et al. 2018).

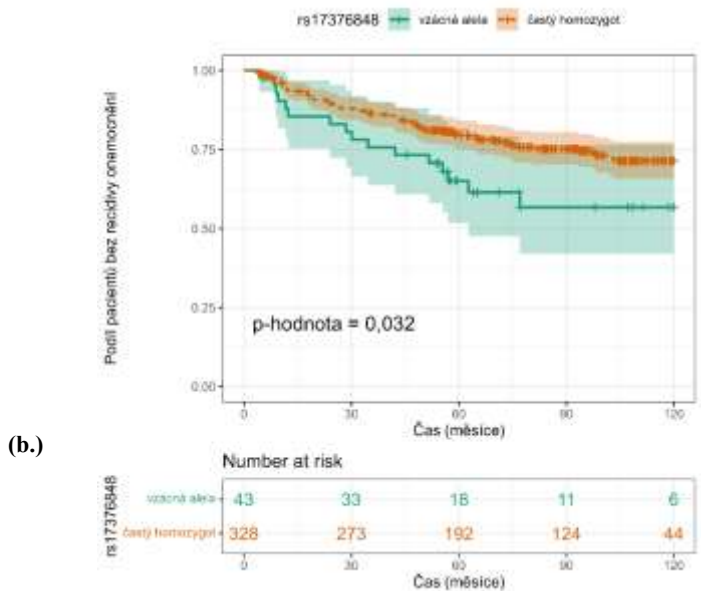
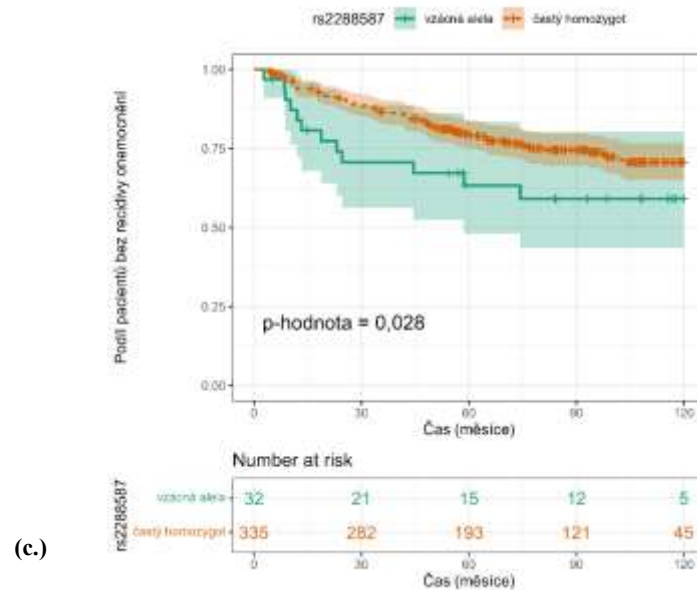
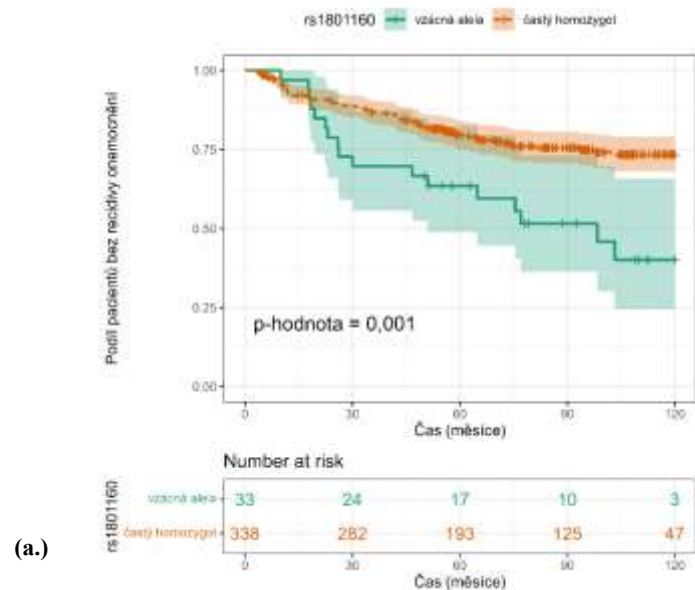
Klinicky významné varianty v odpovědi na léčbu a délce přežití

Do sady prioritizovaných variant byly přidány 4 intronové varianty vykazující silnou statistickou a klinickou korelaci, jak v nedostatečné odpovědi na léčbu, tak zkrácením časového období bez nemoci. Tyto varianty nezískaly žádnou *in silico* evidenci v predikci škodlivosti dle vybraných nástrojů.

4.2. Výsledky validační studie častých variant

Za účelem ověření klinických souvislostí zjištěných v testovací fázi byly varianty hodnoceny na základě odpovědi a přežití pacientek ve validačním souboru. Všechny homozygotní genotypy pozorované u méně, než pěti pacientek byly pro zvýšení statistické síly srovnání seskupeny s odpovídajícím heterozygotním genotypem. V panelu 509 farmakogenů a genů souvisejících s léčivou u pacientek s karcinomem prsu, pouze minoritní část, důležitá z hlediska individualizované terapie, byla potvrzena ve validační studii. Všechny validované varianty byly rovněž predikované za škodlivé aspoň jedním ze zvolených nástrojů.

Další tři varianty v genech *DPYD* a *IRSI*, z nichž pouze jedna byla významná, lze považovat za prognostické z hlediska DFS po cytotoxické léčbě. V multivariační analýze Coxovou regresí upravenou podle grade karcinomu, stádia onemocnění a stavu exprese hormonálních receptorů byla vzácná alela rs1801160 v *DPYD* spojena s významným nárůstem rizika snížení období bez recidivy onemocnění (HR = 2,58; 95 % CI = 1,48-4,50; p = 0,001), ale další dvě varianty (rs17376848 rovněž v *DPYD* a rs2288587 v *IRSI*) byly již nevýznamné (p = 0,071 a p = 0,115).

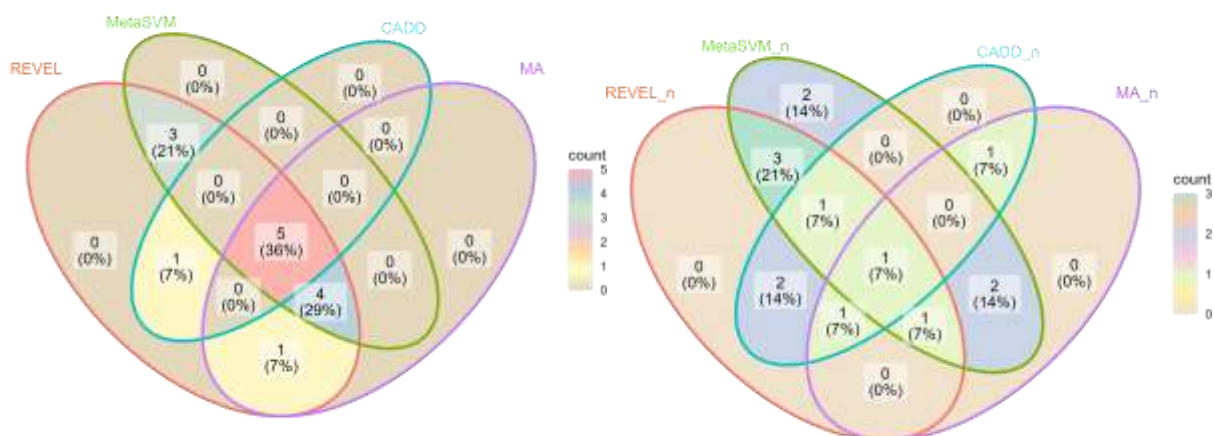


Graf 4—1 Kaplan-Meierovy grafy validovaných variant ve vztahu k DFS patientek léčených cytotoxickou terapií

Čárkovaná čára (oranžová) představuje častý homozygotní genotyp a plná čára (zelená) vzácnou alelu. Významnost byla hodnocena pomocí log-rank testu. Stínované oblasti označují 95% intervaly spolehlivosti. Number at risk = počet patientek v riziku. Při absenci vzácných homozygotů nebo nízkých počtech ve srovnávaných skupinách byl hodnocen vliv vzácné alely. (a): rs1801160; (b): rs17376848; (c): rs228858.

4.3. Výsledky studie proveditelnosti – vzácné varianty

Bioinformatická funkční predikce odhalila 1080 vzácných nsSNV u 365 z 509 sekvenovaných genů. Frekvence alel pro 13,1 % ze zjištěných nsSNV ($n = 141$) chyběla v etnicky blízké populaci (databáze ExAC – NFE) a pro 12,3 % variant ($n = 133$) v České zdravé populaci (databáze NCMG). Nejvyšší počet variant na chromozom a gen byl zjištěn u chromozomu 13 s 25 nsSNVs ve čtyřech genech.



Graf 4—2 Vennův diagram shodně predikovaných genů po filtraci úrovně 1 a 2, (A) prediktory řízené strojovým učením, tj. REVEL, MetaSVM a CADD ve srovnání s MutationAssessor (MA); (B) prediktory řízené umělou inteligencí a MutationAssessor seřazené dle znormalizované hodnoty poměru patogenních genů (GPR). Legenda „count“ určuje zbarvení diagramu dle počtu sdílených genů (R script, knihovna: ggVennDiagram, ggplot2).

B.) Farmakogenomická analýza

Varianty související s PGx byly identifikovány u více než 26 % pacientek s karcinomem prsu ve studii. PharmVar v současné době obsahuje 25 genů (květen 2023), všechny kromě jednoho (*NUDT15*) jsou sdílené s genovým panelem studie. Exonové varianty určené pro sníženou funkci *DPYD* ($n=2$) a *CYP2D6* ($n=3$) byly identifikovány u 16 % pacientek v našem souboru. Varianty výrazně snižující nebo zvyšující funkci genu nebyly identifikovány.

C.) Statistická asociace s klinickými parametry

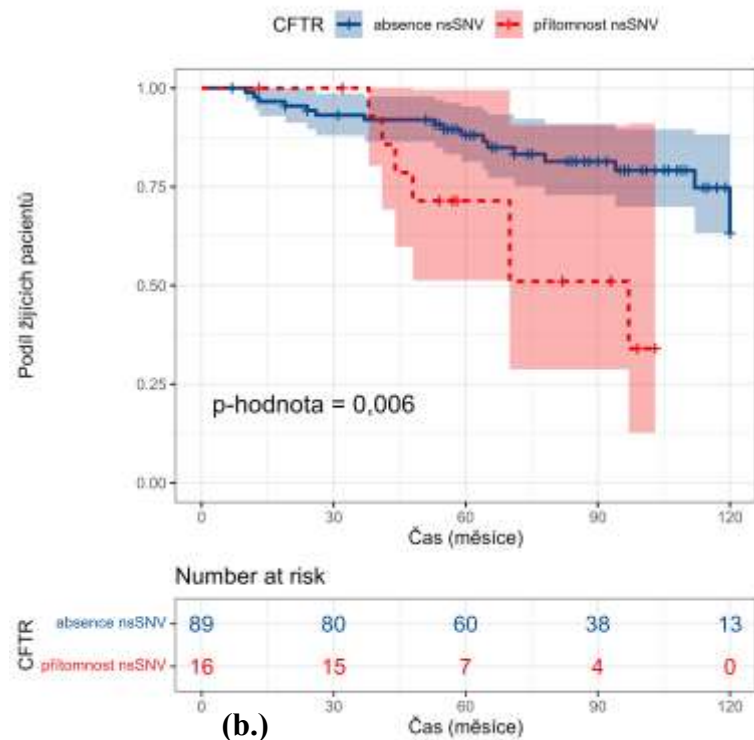
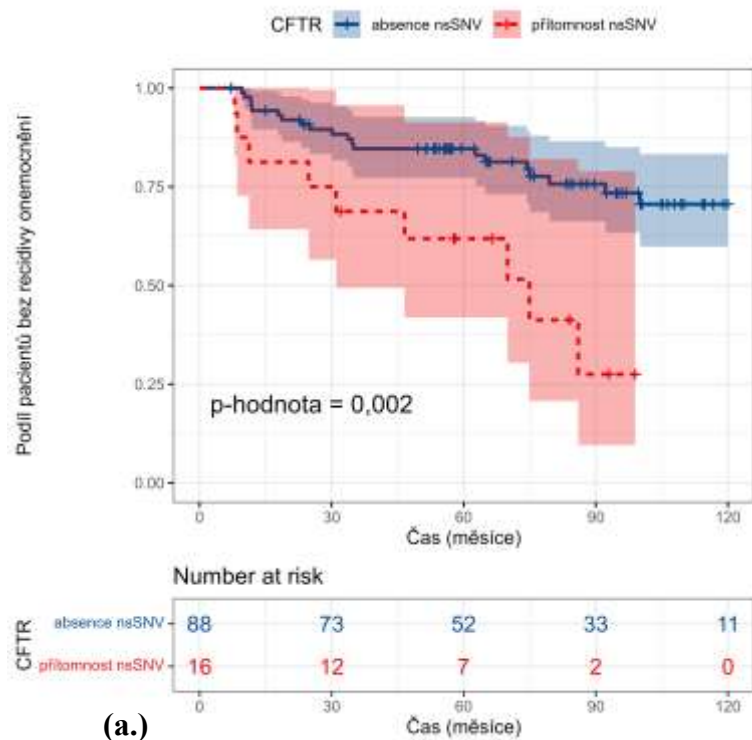
Pacientky s přítomností patogenní varianty v genu *CFTR* měli významně kratší přežití bez onemocnění (Log Rank, Mantel-Cox, $p = 0,002$) a celkové přežití (Log Rank, Mantel-Cox, $p = 0,006$) ve srovnání s pacientkami bez nsSNV (viz Graf 4—3). Studie odhalila *CFTR* jako potenciálně důležitý gen korelující s prognózou karcinomu prsu.

Multivariační analýza pomocí Coxovy regrese upravené podle stupně nádoru a stadia onemocnění potvrdila, že přítomnost vzácných nsSNVs v genu CFTR poskytuje nezávislou prognostickou informaci pro DFS (HR = 3,03 [95% CI: 1,32 - 7,00], p = 0,009), a pro OS (HR=3,38 [95% CI: 1,37 - 8,37], p = 0,008). Pacientky, které byly nositelky některé ze vzácných variant v genu ABCC1, měly nejméně třikrát vyšší pravděpodobnost vzniku nádorů 3. stupně (OR = 3,1 [95% CI: 1,2-8,3], p = 0,021, Pearsonův chí-kvadrát).

Naopak varianty v ABCC4 vykazovaly asociaci naznačující jejich protektivní roli, pokud jde o nízký stupeň nádoru (G1 nebo G2) a jiný než invazivní ductální karcinom (IDC) histologický typ nádoru ve srovnání s pacientkami bez vzácných variant v ABCC4 (p = 0,046, Pearsonův chí-kvadrát a p=0,045, Fisherův přesný test). U pacientek s nsSNV v ABCC4 bylo pozorováno více než devětkrát vyšší riziko nedostatečné odpovědi na léčbu (OR = 9,64 [95%CI: 1,75 - 53,26] p = 0,007, Fisherův přesný test). Přítomnost nsSNVs se vzácnou frekvencí ve farmakogenu *CYP2D6* naznačovala ochrannou roli, protože u pacientek s nádory vysokého stupně (G3) nebyly zjištěny žádné varianty (p = 0,023, Fisherův přesný test).

Pacientky s divokým fenotypem ATP7B měli významně častěji luminální (LA nebo LB) podtyp, a navíc také nižší pravděpodobnost nádorů vysokého stupně G3 než nositelky nsSNV v ATP7B (p = 0,040 a p = 0,013, Fisherův přesný test). Pro gen *CYP4F3* byly nositelky vzácných variant ze skupiny s LB podtypem nádoru ve srovnání s TNBC (p = 0,042, Fisherův přesný test). Pacientky v premenopauzálním období měly častěji nsSNV v ABCB6 (p = 0,03, Pearsonův chí-kvadrát) než pacientky v postmenopauzálním období. Kromě toho bylo nosičství nsSNV v ABCB6 významně spojeno s téměř pětkrát vyšším rizikem nedostatečné odpovědi na neoadjuvantní cytotoxickou chemoterapii (OR=4,89 [95% CI:1,04-22,96], p=0,046, Fisherův přesný test). U vyjmenovaných genů nebyla pozorována žádná souvislost s DFS nebo OS. Coxova regresní analýza pro DFS a OS upravená podle stupně nádoru (G3) a stadia onemocnění proto neprokázala statistickou významnost. U žádné z exonických variant souvisejících s PGx v *DPYD* a *CYP2D6* nebyla zjištěna žádná asociace.

V analýze zaměřené na všechny exonické varianty (tj. nesynonymní, synonymní, frameshift a inframe indels) u *DPYD* (n=5) a *CYP2D6* (n=18) nebyla zjištěna žádná asociace pro žádnou z testovaných proměnných ve srovnání s pacienty bez těchto variant.



Graf 4—3 Kaplan-Meierovy grafy pro asociaci nsSNV s DFS (a.) a OS (b.) patientek

Kaplan-Meierovy grafy pro asociaci nsSNV s pacienty bez recidivy onemocnění, DFS (a.) a celkového přežití, OS (b.) patientek. Červená přerušovaná čára představuje pacientky s přítomností nesynonymní vzácné varianty v genu *CFTR*. Modrá čára představuje pacientky bez nesynonymní varianty v genu *CFTR*. Stínované oblasti označují 95% intervaly spolehlivosti. Number at risk = počet patientek v riziku. Významnost byla hodnocena pomocí Log-rank testu na 5% hladině.

4.4. Výsledky pilotní studie – varianty genu *KIF14*

Bioinformatická analýza pro gen *KIF14* byla provedena platformou SNPnexus (Dayem Ullah et al. 2018) s cílem predikce funkčního dopadu aminokyselinové záměny v proteinu kombinací nástrojů SIFT (Ensembl Variation 95) a PolyPhen2 (Ensembl Variation 95). Celkově byly rozpoznány 4 nsSNVs s dostatečnou datovou podporou, tj. označené v predikci s vysokou důvěrou (rs141768593, rs145426227, rs75449932 a rs77157287). Pouze varianta (rs75449932, T>G) vedoucí k záměně aminokyseliny serinu za arginin byla definována jako škodlivá, zatímco ostatní varianty byly benigní. Predikce dle nástroje PolyPhen2 byla shodná s výsledkem testu SIFT, kdy mezní hodnota pro predikci benigních variant byla navýšená o skupinu tzv. pravděpodobně škodlivých variant za účelem snížení falešně pozitivních výsledků. Interpretace funkční predikce dvěma bioinformatickými nástroji následovala metodu plného konsenzu. Populační data projektu 1000 Genome Project (fáze 3) poskytly údaje o frekvenci minoritní alely G=0,0026, která je v rozmezí pro vzácné varianty (MAF \leq 001). Pro nekódující varianty v *KIF14* by zvolen nástroj RegulomeDB, který celkově predikoval dvě škodlivé intronové varianty rs17448931, A>G kategorie (3a) a rs3806362, C>T vyšší kategorie (2a).

Následující část studie byla provedena ve spolupráci s kolegy Ing. Ivonou Krusovou Ph.D., a Viktorem Hlaváčem, Ph.D., a doplněna o další sadu anotací nástrojem Annovar (Krus et al. 2022).

Z první sady *in silico* predikovaných variant byla statistická souvislost klinických parametrů testovacího souboru pacientek (n = 105) prokázána u nekódující varianty rs17448931 a rs3806362. Celkové přežití u pacientek s minoritní alelou G (rs17448931) bylo signifikantně sníženo (p = 0,036) v porovnání s běžným genotypem AA. Pro nositelky alely T (rs3806362) bylo pozorováno významně kratší celkové přežití (p = 0,023) jak v testujícím, tak ve validačním setu. Ve validačním setu (n = 808) měly nositelky minoritní alely G (rs17448931) rovněž významně kratší celkové přežití než pacientky bez variace (p = 0,010), čímž se nám podařilo úspěšně reprodukovat výsledky testovacího souboru založené na predikci nástrojem RegulomeDB. Asociace zůstala významná pro obě varianty i v multivariační Coxově regresní analýze, která zahrnovala kovariáty jako molekulární podtyp tumoru a stadium onemocnění v době diagnózy (poměr rizik, HR = 1,7 a 95% interval spolehlivosti (1,1 - 2,8) pro rs17448931 a HR = 1,9 a 95% interval spolehlivosti (1,2 - 3,0) pro rs3806362.

5. Diskuse metodických postupů a výsledků

5.1.1. Pilotní a validační studie – bioinformatická analýza

Veřejné databáze jako je např. PharmGKB jsou bohatým zdrojem informací o zárodečných variantách. Databáze vznikly na základě pracného kurátorství publikovaných studií a silné potřeby systematického využívání získaných znalostí v personalizované medicíně (McDonagh et al. 2011; Smirnov et al. 2018). K datu 6. 7. 2023 je v databázi PharmGKB (<https://www.pharmgkb.org/>) přístupných 26 402 anotací pro 762 léčiv souvisejících s odpovědí na léčbu na farmakodynamické a/nebo farmakokinetické úrovni. Navzdory značnému počtu dostupných anotací, není k dispozici webový nástroj ani nástroj příkazového řádku, který by usnadnil další použití těchto variant v laboratořích a v klinické praxi. Uvedené anotace na stránkách PharmGKB nebo PharmVar jsou přístupné pouze v TSV formátech a vyžadují další pracné zpracování. Publikováno bylo několik málo pokusů o vytvoření nástrojů designovaných pro farmakogeny a pouze jedna studie s optimalizací současných modelů pro farmakogenomické hodnocení (Gerek et al. 2015; Zhou et al. 2018).

Tři varianty ve vztahu k odpovědi na léčbu byly objeveny pro genotyp v rs2227291 genu *ATP7A* a varianty rs4376673 genu *DFFB* u častého (MAF > 5 %) homozygota anebo pro častou alelu v rs2293194 genu *KCNABI*. Dále pacienti s četným homozygotním genotypem v rs1801160 *DPYD* přežívali déle bez relapsu po cytotoxické léčbě než nositelé vzácné alely. Tento efekt byl zvláště výrazný u pacientů s lumenálním B a triple negativním molekulárním podtypem.

Zajímavého výsledku jsme dosáhli pro následující dvě intronové varianty. Pro gen *DFFB* (rs4376673) a *KCNABI* (rs2293194), byly obě varianty prioritizované z důvodu silné klinické evidence s odpovědí na léčbu a zároveň s délkou přežití bez nemoci. Navíc tyto varianty získaly podporu predikce *in silico* dle nového nástroje PINES s výpočtem váženým vzhledem na fenotypová data pro jaterní tkáň jako hlavního místa metabolismu léčiv. Z celkového počtu 6 prioritizovaných variant se silnou klinickou evidencí odpovědi na léčbu a DFS, pouze právě tyto dvě varianty, predikované nástrojem PINES, byly validovány. Varianta rs2293194 byla významně spojena s odpovědí u pacientů s časným stadiem onemocnění 0 (in situ) nebo I ($p < 0,001$). V genetice nádorových onemocnění může gen *KCNABI* hrát roli v patogenezi karcinomu prsu, protože jeho nadměrná exprese byla zjištěna v nádorech prsu ve srovnání s nenádorovými tkáněmi (Hlaváč et al. 2014, s. 450).

Gen *DFFB* je podjednotkou Beta a aktivní složkou proteinu DNA Fragmentation Factor (DFF). Například zvýšená exprese *DFFB* v kombinaci s léčbou doxorubicinem nebo sulfonamidy zvýšila mortalitu nádorových buněk karcinomu prsu T47-D prostřednictvím apoptózy *in vitro*

(Bagheri et al. 2015; 2014). Variace a deregulace genu *DFFB* v přítomnosti léků indukujících apoptózu by tedy mohla mít vliv na jejich účinnost v nádorových buňkách. Dalším genem relevantním pro farmakogenomiku nukleosidových analog je dihydropyrimidin dehydrogenáza (*DPYD*). DPD je aktivní v metabolické dráze 5-fluorouracilu a mutace v jeho genu vedou ke zvýšenému riziku toxicity u onkologických pacientů, kterým je administrována chemoterapie 5-fluorouracilem (van Kuilenburg 2004). Polymorfismus *DPYD* rs1801160, spojený s přežitím pacientek s karcinomem prsu, je dokumentován ve PharmGKB pro jiné než v naší studii detekované alely (CC, CT, TT). Avšak v nedávné studii u pacientů s karcinomem tlustého střeva léčených režimy sestávajícími z 5-fluorouracilu nebo kapecitabinu v kombinaci s oxaliplatinou bylo nosičství rs1801160 (G>A) v *DPYD* spojeno s nežádoucími rizikovými účinky, např. neutropenie třetího až čtvrtého stupně (Ruzzo et al. 2017). Další recentní studie u onkologických pacientů s farmakoterapií fluoropyrimidiny navíc dokládá významnost objevu rs1801160 (c.2194G>A) související s odpovědí na léčbu, kdy autoři přímo doporučují zvážení přidání této varianty do klasického *DPYD* panelu testování u pacientů (Božina et al. 2022). Přestože bylo nedávno navrženo individualizované dávkování řízené genotypem *DPYD* pro lepší bezpečnost léčby fluoropyrimidiny jako nový standard péče (Henricks et al. 2018), naše studie nebyla navržena tak, aby se zabývala nežádoucími účinky anebo úpravou terapie. Nicméně ověřená asociace varianty *DPYD* s DFS přidává do znalostní báze nové a důležité pozorování.

Z dalších validovaných variant, spojených s odpovědí na léčbu, vzbuzuje zvláštní pozornost rs2227291 v *ATP7A*, protože je nesynonymní (V767L), a tudíž pravděpodobně bezprostředněji funkční. Pozoruhodné je, že rs2227291 je jedinou asociací s odpovědí ve validačním souboru, která prošla mírou falešného objevu ($p = 0,011$). Transportér mědi *ATP7A* kóduje transmembránový protein, který funguje při transportu mědi přes membrány a je často studován v souvislosti s citlivostí na léky na bázi platiny, např. cisplatinu. Poměrně nedávno byla rs2227291 spojena s rezistencí na cisplatinu u pacientek s epiteliálním karcinomem vaječnicků léčených kombinací platiny a taxanu (Li et al. 2017). Pro gen *SLC28A3* je rs10868138 méně studovanou variantou, dle publikovaných dat, byla spojena s vyšší koncentrací azathioprinu v erytrocytech pacientů s neuromyelitis optica (Mei et al. 2017), což naznačuje, že může být funkční *in vivo*.

Souhrnně, farmakogenomicky již známé varianty uvedené v repositoriu PharmGKB a prioritizované našeho výběru významně souvisely s odpovědí na léčbu. Pacientky s dobrou odpovědí na neoadjuvantní cytotoxickou léčbu měly vzácnou alelu rs10868138 (PA166157820)

v genu *SLC28A3* významně častěji než pacientky bez odpovědi. Varianta byla predikovaná nástrojem CADD za škodlivou. ClinVar neuváděl v době analýzy význam této varianty. Genetická variabilita *SLC28A3* byla již dříve spojována se změnou farmakokinetických parametrů nukleosidových analogů (Khatri et al. 2014) a s kardiotoxicitou antracyklinů (Visscher et al. 2013). Ukázalo se, že variace rs7867504 v *SLC28A3* se podílí na toxicitě gemcitabinu u pacientů s metastazujícím karcinomem prsu, kteří dostávají udržovací léčbu (Park et al. 2013). Asociace této varianty však nebyla po úpravě na stadium onemocnění v multivariační analýze významná.

Obecně překvapivého výsledku jsme dosáhli při celkovém srovnání škodlivých variant predikovaných nástrojem CADD, které se ve 28,5 % shodovalo s již kurátory ošetřenými farmakogenomickými variantami v PharmGKB a ve 100 % u tří nesynonymních variant ověřených ve validační fázi. CADD nebyl trénován na sadě těchto farmakogenomických variant, ale naše, i když počtem omezené výsledky, poukazují na vyšší specifitu nástroje, i vzhledem k PGx datům. Algoritmus predikce je u CADD založen na strojovém učení.

Navzdory výběru nástrojů do sekvenční sady nejlépe odpovídající povaze farmakogenomickým variantám a obcházející výše zmíněné známé nedostatky, např. kruhovost (Grimm et al. 2015) u ostatních kategorií (tj. metody strukturální, strojové učení), nebyla ani jedna ze tří prioritizovaných variant validována. Nedávná studie z roku 2022 demonstrovala sníženou výkonnost prediktorů strojového učení právě z důvodu nadměrného přizpůsobení („overfitting“) trénovaných dat naproti nástrojům jako je SIFT a PROVEAN využitě v naší studii (Quinodoz et al. 2022). Nicméně, při dodatečném porovnání s nástrojem CADD, všechny tři nsSNVs byly považované rovněž za škodlivé, což naznačuje, že chybovost predikce nebyla daná výběrem sekvenčních modelů bez strojového učení.

V rámci nově navržené strategie prioritizace variant jsme implementovali jiný postup, než byl používán na pracovišti školitele, kdy vzhledem k zjištěným omezením *in silico* analýz jsme nejdříve aplikovali statistickou analýzu na menším vzorku pacientek v korelaci ke klinickým datům, která vyfiltrovala potenciálně důležité varianty a zároveň zvýšila efektivitu postupu.

Podporu tohoto postupu dokládají i výsledky studie (Ernst et al. 2018), kde testovali výkonnost čtyř predikčních nástrojů (Align-GVGD, SIFT, PolyPhen-2, MutationTaster2) na souboru 236 nesynonymních vrozených variant *BRCAl/2*, které byly již dříve klasifikovány odbornými komisemi. Identifikovali 161 variant, jejichž klasifikace je nezávislá na jakékoli předchozí *in silico* predikci a prokázali, že vzhledem k nízké specifitě nebyly predikční nástroje *in silico* vhodné k předpovědi patogenity variant s nejistým významem tzv. VUS u *BRCAl/2*. Klinické

důsledky by tedy nikdy neměly být založeny pouze na předpovědích *in silico*. Celkově tak naše statistická evidence asociace s klinickými daty na malém vzorku pacientů (n = 105) předcházela pouze doplňující, avšak komplexní *in silico* prioritizaci doplněnou o data z databází s již ověřenou funkční predikcí. Rozdílnost ve funkčních aspektech již klinicky ověřených variant je obecným problémem, a proto jsme krok opětovného ověřování již publikovaných variant považovali za zásadní. Zmíněná studie dále naznačila, že nástroje SIFT a MutationTaster2 by mohly být vhodné k předpovídání benignity, protože oba nástroje nevedly k falešně negativním předpovědím (Ernst et al. 2018). Strategii, která rozděluje výběr specifických nástrojů pro benigní varianty a samostatně pro patogenní neboli škodlivé, rovněž v případě studie (Ghosh et al. 2017), jsme použili v další studii pro *in silico* predikci vzácných variant dílčího cíle.

Z celkového počtu 18245 variant identifikovaných u našich pacientek se nacházelo 3256 v exonech, 9458 v intronu a 3872 v oblastech 3'UTR nebo 5'UTR při anotaci podle databáze NCBI Reference Sequence Database, RefSeq (RefSeq; <https://www.ncbi.nlm.nih.gov/refseq/>) v Annovaru. Celkem 2565 (14 %) variant bylo nových (tj. nenalezených v dbSNP Build 150). Identifikovali jsme 327 variant (dvě nové) souvisejících s odpovědí na neoadjuvantní cytotoxickou léčbu a 418 variant (tři nové) souvisejících s DFS. V pilotní fázi projektu bylo vybráno 58 variant na základě nové strategie a 5 variant, z toho všechny predikované za škodlivé *in silico*, byly klinicky ověřeny.

5.1.2. Predikce vzácných variant

Úprava *in silico* predikce podle délky genu dále zpřesnila proces filtrace a umožnila zařazení *DPYD* - VIP farmakogenu do konečného souboru genů. Dolování dat v databázích PGx také odhalilo významný počet pacientek naší studie, které měly relevantní PGx exonické varianty, a to konkrétně v *DPYD*, *CYP2D6* a *ABCC1*. Funkční předpovědi *in silico* pomocí čtyř nástrojů byly k dispozici pouze pro dvě nsSNV (rs67376798 v *DPYD* a rs45511401 v *ABCC1*) z osmi identifikovaných PGx variant. Gen *DPYD* kóduje enzym dihydropyrimidin dehydrogenázu (DPD), který je zodpovědný za farmakokinetiku a související toxicitu fluoropyrimidinů. U heterozygotních pacientů byla pro rs67376798 nebo rs56038477 hlášena toxicita související s fluoropyrimidiny 3. nebo 4. stupně vyžadující 50% snížení počáteční dávky (Knikman et al. 2021; Carr et al. 2021). V naší studii však nebyla u pacientek s výskytem vzácných nsSNVs nebo exonických variant u *DPYD* zjištěna žádná souvislost se žádným z testovaných klinických parametrů. Naopak, zatímco ani u exonických variant v *CYP2D6* nebyla nalezena žádná

asociace, u skupiny pacientek s přítomností nsSNV byla prokázána asociace se stupněm nádoru, tj., 1 a 2 ($p = 0,023$). Studie rovněž ukázala, že u pacientek s karcinomem prsu, které měly některou ze vzácných variant v *ABCC1*, byla nejméně třikrát vyšší pravděpodobnost vzniku vysoce agresivních nádorů 3. stupně. Publikované údaje dokládají, že rs45511401 v *ABCC1* byla významně ($p = 0,012$) spojena se zvýšeným rizikem kardiotoxicity u pacientů s Non-Hodgkinovým lymfomem léčených adriamycinem (Wojnowski et al. 2005). Antracykliny, jako je adriamycin, řadí k běžné onkologické léčbě pacientek s karcinomem prsu a výskyt závažných nežádoucích vedlejších účinků může vést k přerušení a selhání léčby.

Mutace v *CFTR* byly v nedávné době spojeny se zvýšenou náchylností k různým formám malignit, což naznačuje, že funkce *CFTR* je ještě širší než jen funkce iontového kanálu a poukazuje na schopnost regulovat četné signální dráhy. Již několik autorů (Zhang et al. 2018; FitzMaurice a Nazareth 2021; Dvorak et al. 2017) popsalo souvislost úrovně exprese *CFTR* s přežitím pacientů včetně možného využití *CFTR* jako prognostického biomarkeru u lidských malignit, jako je například karcinom prsu. Obecně platí, že zárodečné mutace v *CFTR* způsobují cystickou fibrózu (CF) - nejčastější život limitující recesivní genetické onemocnění u bělochů a *CFTR* je známý především pro svou PGx hodnotu u CF. Z celkem 12 variant v genu *CFTR* pozorovaných u našich pacientek s karcinomem prsu je osm variant uloženo také v databázi The Clinical and Functional TRanslation of CFTR (CFTR2). Tyto komplexní varianty jsou považovány buď za nezpůsobující CF, nebo mají různé klinické důsledky v závislosti na spoluobjevující se variantě (variantách) umístěné v pozici cis/trans, která ovlivňuje produkci proteinu CFTR (např. pankreatická insuficience). Pro naši studii to naznačuje potenciál pro funkční aktivitu zjištěných variant. Z hlediska poměru patogenních variant na gen se *CFTR* nacházel v rozmezí 10 nejvíce patogenních genů. Několik dalších studií ukázalo, že délka transkriptu koreluje s funkční aktivitou ko-exprimovaných genů nebo proteinů souvisejících s rakovinou (Lopes et al. 2021; Sahakyan a Balasubramanian 2016). *CFTR* s délkou transkriptu 6132 bp a velikostí proteinu 1480 aminokyselin lze považovat za delší gen, což značí, že mechanismus působení v rezistenci na léky nebo v maligním procesu může být závislý i na jiných genech. Již dříve jsme prokázali, že *CFTR*, dříve známý jako *ABCC7/MRP7*, převažoval mezi nejvíce dysregulovanými geny na úrovni transkriptu u karcinomu prsu, kolorekta a pankreatu (Dvorak et al. 2017). ABC transportéry představují velkou rodinu aktivních přenašečů, které transportují látky, včetně protinádorových léčiv, přes extracelulární a intracelulární membrány a ovlivňují tak eflux léčiv související s rozvojem chemorezistence nádorových buněk (Fletcher et al. 2016; Chen et al. 2016). Zde jsme identifikovali čtyři ze sedmi klinicky asociovaných genů,

keré se řadí do rodiny ABC. Jako příklad lze uvést transportéry podrodiny *ABCC*, které snižují intracelulární koncentraci 5-fluorouracilu (Fukuda a Schuetz 2012). U *ABCC4*, a navíc u *ABCB6* jsme prokázali významně zvýšené riziko nedostatečné odpovědi na neoadjuvantní cytotoxickou léčbu u pacientek s výskytem vzácných nsSNVs.

O dalších prioritizovaných a klinicky souvisejících genech, *ATP7B* a později *CYP4F3* nebo *CYP2D6*, je známo, že se podílejí na metabolismu protinádorových léčiv a na rezistenci k chemoterapii založené na cisplatině u některých druhů rakoviny (Kanzaki et al. 2002; Nakayama et al. 2002). Naše zjištění naznačují, že tato rezistence nemusí být omezena pouze na platinové deriváty, ale také na další strukturně a mechanisticky různorodá léčiva. Souvislost mezi genetickou variabilitou těchto genů a rezistencí však čeká na potvrzení v nezávislých studiích na větších populacích a následných funkčních analýzách. Jsme si vědomi toho, že naše současné výsledky nemusí být nezávislé na dalších neznámých klinických nebo molekulárních faktorech, které ještě výrazněji ovlivňují prognózu pacientů. Kromě skromné velikosti vzorku je dalším omezením naší studie panel genů s obohacením o cílové ABC transportéry, což mohlo ovlivnit výhodu vypočteného poměru patogenních variant na gen. Na druhou stranu má naše studie jasný přínos v zaměření na nejvýznamnější farmakogeny a onkogeny a hodnocení etnicky homogenní populace, což přispívá ke konceptu diverzity v medicíně a konkrétně v precizní onkologii a možné prevenci.

5.1.3. Predikce variant genu *KIF14*

Celkově z 23 zárodečných variant s alelickou frekvencí větší než 5%, byly pouze 2 (rs17448931 a rs3806362) klinicky ověřeny a představují další zpřesnění cílů v genu *KIF14* s potenciálním využitím v terciární prevenci. Varianty se nacházejí v intronové části genomu a jejich regulační funkce byla demonstrována a správně předpověděna zvolenou *in silico* predikcí.

Funkční aspekt intronové části těchto variant podporuje aktualizovaná verze databáze Regulome (2.2) (Dong et al. 2023), která disponuje novými daty a poskytuje anotaci, kdy kategorizuje variantu rs17448931 ze skupiny (3a), v době naší studie, do (1f) tj., významných funkčně regulačních elementů. Podkategorie 1f indikuje dostupnost důkazů o vazbě varianty na transkripční faktor nebo místa citlivého ke štěpení chromatinu DNázou. Druhá varianta rs3806362 byla rovněž zařazena do vyšší kategorie z (2a) do (1a), tzn. že databáze obsahuje důkazy o vazbě transkripčního faktoru (TF), motivu pro TF a místa citlivého ke štěpení chromatinu DNázou. Zařazení těchto variant do jedné z nejvyšších funkčních kategorií nezávisle na našich výsledcích nepřímo poukazuje na významnost našeho výsledku.

Dostupné metody predikce, zejména pokud jsou adekvátně kombinovány, mají potenciál pro správnou predikci škodlivosti. Zachycení diagnostických požadavků, zlepšení prevence a klinické léčby však vyžaduje další optimalizaci. Statistické nebo strukturální metody a metody strojového učení, založené na pravidlech pro identifikaci a analýzu závislostí nsSNVs nesou potenciál ke zvýšení přesnosti diagnostických a prognostických přístupů k odhadu odpovědi na léčbu a zůstanou proto součástí evidence ve filtraci a prioritizaci variant na pracovišti školitele.

5.2. Závěry

5.2.1. Validáční studie

Prostřednictvím masivního paralelního sekvenování byla u českých pacientek s karcinomem prsu poprvé vyhodnocena zárodečná variabilita panelu genů ve vztahu k metabolismu a eliminaci léčiv, k funkci buněčné smrti či signálních drah markerů v klinické onkologii. Pomocí nových nástrojů *in silico* a zavedených farmakogenomických databází byly v testovací fázi získány prioritizované varianty u 105 pacientek a dále technicky a klinicky ověřeny na souboru dalších 805 pacientek. Asociace rs2227291 v *ATP7A*, rs2293194 v *KCNAB1* (u pacientek v časném stadiu) a rs4376673 v *DPYD* s odpovědí na neoadjuvantní cytotoxickou léčbu poskytují nové cíle pro následné funkční studie s potenciálem přispět k terciární prevenci selhání léčby. U pacientek s lumenálními B nebo trojitě negativními nádory, často studovaná varianta rs1801160 v *DPYD* významně souvisela s přežitím bez nemoci a představuje další provokační cíl s prognostickým potenciálem. Dva roky po zveřejnění naší studie, EMA doporučila testování snížené aktivity enzymu DPD, s častým výskytem v kavkazské populaci (9 %), až úplné deficience DPD se vzácnou frekvencí variant 0.5% pro predikci rizika toxicity (neutropenie, neurotoxicita atd.) u léčby fluoropyrimidiny (de With et al. 2023) https://www.ema.europa.eu/en/documents/press-release/ema-recommendations-dpd-testing-prior-treatment-fluorouracil-capecitabine-tegafur-flucytosine_en.pdf. Recentní studie u onkologických pacientů s farmakoterapií fluoropyrimidiny navíc dokládá významnost objevu rs1801160 (c.2194G>A) související s odpovědí na léčbu, kdy autoři přímo doporučují zvážení přidání této varianty do klasického *DPYD* panelu testování u pacientů (Božina et al. 2022). Studie tak přispívá k dalšímu posílení klinické evidence oproti vydanému pokynu CPIC z roku 2017 pro danou variantu (<https://cpicpgx.org/guidelines/guideline-for-fluoropyrimidines-and-dpyd/>). Jedná se také o první studii, která v české populaci poskytuje genetická data v takovém

rozsahu a s vazbou na chemorezistenci hodnocenou ve vztahu k prognóze a výsledku léčby pacientek s karcinomem prsu.

5.2.2. Studie vzácných variant

Naše studie proveditelnosti přinesla několik inovativních postupů prioritizace domněle funkčních vzácných variant. Pokud je nám známo, je to poprvé, kdy byl jako výchozí bod při určování priorit patogenních genů u českých pacientek s karcinomem prsu zaveden poměr patogenních genů (GPR) a vylučovací kritéria pro benigní geny s kombinací nástrojů, kterým dominuje predikce řízená umělou inteligencí. Pro sedm z dvanácti prioritizovaných genů, tj. *ABCC1*, *ABCC4*, *ABCB6*, *ATP7B*, *CYP2D6*, *CYP4F3* a *CFTR* jsme našli významnou souvislost s odpovědí na léčbu, menopauzálním obdobím, dále s histologickým typem, molekulárním podtypem a diferencovaností (grade) karcinomu prsu. Naše studie poskytuje další důkaz, že *CFTR* se potenciálně podílí na procesech ovlivňujících účinnost onkologické léčby nebo na samotné maligní progresi onemocnění.

Terciární prevence v kontextu odpovědi na léčbu se teoreticky nabízí možným sledováním funkčnosti *CFTR* nebo přímým zavedením terapeutických léčivých látek, jako jsou již dostupné modulátory *CFTR*, které do určité míry řeší některé defekty funkčních variant.

5.2.3. Studie variant genu *KIF14*

In silico analýza nekódujících variant genu *KIF14* nástrojem RegulomeDB odhalila funkční nekódující variantu (rs17448931), která byla úspěšně reprodukována na klinických datech pacientek ve validačním setu (n = 808) České populace. Celkově z 23 zárodečných variant s alelickou frekvencí větší než 5%, pouze dvě (rs17448931 a rs3806362) byly klinicky ověřeny. Varianty se nacházejí v intronové části genu a jejich regulační funkce byla správně předpověděna zvolenou *in silico* metodou.

Tato práce splnila všechny vytýčené cíle a přispěla k výzkumu v oblasti terciární prevence selhání léčby v onkologii, zejména u karcinomu prsu. Ukazuje nejen klinický význam jednotlivých genů a variant, ale i cesty k bioinformatickému vyhodnocení a získání dalších úrovní důkazů pro další využití v moderní precizní medicíně.

Seznam použité literatury

- ADMOUN, Claudia a Harvey N. MAYROVITZ, 2022. The Etiology of Breast Cancer. In: Harvey N. MAYROVITZ, ed. *Breast Cancer* [online]. Brisbane (AU): Exon Publications [vid. 2023-07-29]. ISBN 978-0-645-33203-2. Dostupné z: <http://www.ncbi.nlm.nih.gov/books/NBK583809/>
- AFGAN, Enis, Brad CHAPMAN, Margita JADAN, Vedran FRANKE a James TAYLOR, 2002. Using Cloud Computing Infrastructure with CloudBioLinux, CloudMan, and Galaxy. In: *Current Protocols in Bioinformatics* [online]. B.m.: John Wiley & Sons, Inc. [vid. 2017-10-27]. ISBN 978-0-471-25095-1. Dostupné z: doi:10.1002/0471250953.bi1109s38
- ALAOFI, Rawan K., Mohammed O. NASSIF a Marwan R. AL-HAJEILI, 2018. Prophylactic mastectomy for the prevention of breast cancer: Review of the literature. *Avicenna Journal of Medicine* [online]. **8**(3), 67–77 [vid. 2023-11-09]. ISSN 2231-0770. Dostupné z: doi:10.4103/ajm.AJM_21_18
- ALSHABEEB, Mohammad A., Mesnad ALYABSI, Mohammad A. AZIZ a Salah ABOHELAIKA, 2022. Pharmacogenes that demonstrate high association evidence according to CPIC, DPWG, and PharmGKB. *Frontiers in Medicine* [online]. **9** [vid. 2023-02-06]. ISSN 2296-858X. Dostupné z: <https://www.frontiersin.org/articles/10.3389/fmed.2022.1001876>
- AMBERGER, Joanna S, Carol A BOCCHINI, Alan F SCOTT a Ada HAMOSH, 2019. OMIM.org: leveraging knowledge across phenotype–gene relationships. *Nucleic Acids Research* [online]. **47**(D1), D1038–D1043 [vid. 2023-07-07]. ISSN 0305-1048. Dostupné z: doi:10.1093/nar/gky1151
- AYMÉ, S, B URBERO, D OZIEL, E LECOUTURIER a AC BISCARAT, 1998. Information sur les maladies rares : le projet Orphanet. *La Revue de Médecine Interne* [online]. **19**, 376–377 [vid. 2023-07-07]. ISSN 0248-8663. Dostupné z: doi:10.1016/S0248-8663(98)90021-2
- BAGHERI, Fatemeh, Shahrokh SAFARIAN, Mohamadreza Baghaban ESLAMINEJAD a Nader SHEIBANI, 2014. Stable overexpression of DNA fragmentation factor in T-47D cells: sensitization of breast cancer cells to apoptosis in response to acetazolamide and sulfabenzamide. *Molecular Biology Reports* [online]. **41**(11), 7387–7394. ISSN 1573-4978. Dostupné z: doi:10.1007/s11033-014-3626-3
- BAGHERI, Fatemeh, Shahrokh SAFARIAN, Mohamadreza Baghaban ESLAMINEJAD a Nader SHEIBANI, 2015. Sensitization of breast cancer cells to doxorubicin via stable cell line generation and overexpression of DFF40. *Biochemistry and Cell Biology = Biochimie Et Biologie Cellulaire* [online]. **93**(6), 604–610. ISSN 1208-6002. Dostupné z: doi:10.1139/bcb-2015-0007
- BAKER, Wendy, Alexandra VAN DEN BROEK, Evelyn CAMON, Pascal HINGAMP, Peter STERK, Guenter STOESSER a Mary Ann TULI, 2000. The EMBL Nucleotide Sequence Database. *Nucleic Acids Research* [online]. **28**(1), 19–23 [vid. 2023-08-11]. ISSN 0305-1048. Dostupné z: doi:10.1093/nar/28.1.19

- BELIZÁRIO, José E., Beatriz A. SANGIULIANO, Marcela PEREZ-SOSA, Jennifer M. NEYRA a Dayson F. MOREIRA, 2016. Using Pharmacogenomic Databases for Discovering Patient-Target Genes and Small Molecule Candidates to Cancer Therapy. *Frontiers in Pharmacology* [online]. 7 [vid. 2017-10-29]. ISSN 1663-9812. Dostupné z: doi:10.3389/fphar.2016.00312
- BERNSTEIN, Bradley E, John A STAMATOYANNOPOULOS, Joseph F COSTELLO, Bing REN, Aleksandar MILOSAVLJEVIC, Alexander MEISSNER, Manolis KELLIS, Marco A MARRA, Arthur L BEAUDET, Joseph R ECKER, Peggy J FARNHAM, Martin HIRST, Eric S LANDER, Tarjei S MIKKELSEN a James A THOMSON, 2010. The NIH Roadmap Epigenomics Mapping Consortium. *Nature biotechnology* [online]. 28(10), 1045–1048 [vid. 2023-07-26]. ISSN 1087-0156. Dostupné z: doi:10.1038/nbt1010-1045
- BODEA, Corneliu A., Adele A. MITCHELL, Aaron G. DAY-WILLIAMS, Heiko RUNZ a Shamil R. SUNYAEV, 2018. Phenotype-specific information improves prediction of functional impact for noncoding variants. *bioRxiv* [online]. 083642 [vid. 2018-07-22]. Dostupné z: doi:10.1101/083642
- BONAVENTURA, J a A RIGGS, 1968. Hemoglobin Kansas, a Human Hemoglobin with a Neutral Amino Acid Substitution and an Abnormal Oxygen Equilibrium. *Journal of Biological Chemistry* [online]. 243(5), 980–991 [vid. 2023-07-07]. ISSN 0021-9258. Dostupné z: doi:10.1016/S0021-9258(18)93612-4
- BOŽINA, Nada, Ivan BILIĆ, Lana GANOCI, Livija ŠIMIČEVIĆ, Stjepko PLEŠTINA, Lucija LEŠNJAKOVIĆ a Vladimir TRKULJA, 2022. DPYD polymorphisms c.496A>G, c.2194G>A and c.85T>C and risk of severe adverse drug reactions in patients treated with fluoropyrimidine-based protocols. *British Journal of Clinical Pharmacology* [online]. 88(5), 2190–2202. ISSN 1365-2125. Dostupné z: doi:10.1111/bcp.15144
- CABANILLAS, Rubén, Marta DIÑEIRO, David CASTILLO, Patricia C. PRUNEDA, Cristina PENAS, Guadalupe A. CIFUENTES, Álvaro DE VICENTE, Noelia S. DURÁN, Rebeca ÁLVAREZ, Gonzalo R. ORDÓÑEZ a Juan CADIÑANOS, 2017. A novel molecular diagnostics platform for somatic and germline precision oncology. *Molecular Genetics & Genomic Medicine* [online]. 5(4), 336–359 [vid. 2017-11-13]. ISSN 2324-9269. Dostupné z: doi:10.1002/mgg3.291
- CARR, Daniel F., Richard M. TURNER a Munir PIRMOHAMED, 2021. Pharmacogenomics of anticancer drugs: Personalising the choice and dose to manage drug response. *British Journal of Clinical Pharmacology* [online]. 87(2), 237–255 [vid. 2023-04-07]. ISSN 1365-2125. Dostupné z: doi:10.1111/bcp.14407
- COOPER, Gregory M. a Jay SHENDURE, 2011. Needles in stacks of needles: finding disease-causal variants in a wealth of genomic data. *Nature Reviews Genetics* [online]. 12(9), 628–640 [vid. 2023-07-07]. ISSN 1471-0064. Dostupné z: doi:10.1038/nrg3046
- DAYEM ULLAH, Abu Z., Jorge OSCANOVA, Jun WANG, Ai NAGANO, Nicholas R. LEMOINE a Claude CHELALA, 2018. SNPnexus: assessing the functional relevance of genetic variation to facilitate the promise of precision medicine. *Nucleic Acids Research* [online]. 46(W1), W109–W113 [vid. 2018-07-20]. ISSN 0305-1048. Dostupné z: doi:10.1093/nar/gky399

- DENNY, Joshua C, Lisa BASTARACHE, Marylyn D RITCHIE, Robert J CARROLL, Raquel ZINK, Jonathan D MOSLEY, Julie R FIELD, Jill M PULLEY, Andrea H RAMIREZ, Erica BOWTON, Melissa A BASFORD, David S CARRELL, Peggy L PEISSIG, Abel N KHO, Jennifer A PACHECO, Luke V RASMUSSEN, David R CROSSLIN, Paul K CRANE, Jyotishman PATHAK, Suzette J BIELINSKI, Sarah A PENDERGRASS, Hua XU, Lucia A HINDORFF, Rongling LI, Teri A MANOLIO, Christopher G CHUTE, Rex L CHISHOLM, Eric B LARSON, Gail P JARVIK, Murray H BRILLIANT, Catherine A MCCARTY, Iftikhar J KULLO, Jonathan L HAINES, Dana C CRAWFORD, Daniel R MASYS a Dan M RODEN, 2013. Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. *Nature biotechnology* [online]. **31**(12), 1102–1110 [vid. 2018-01-30]. ISSN 1087-0156. Dostupné z: doi:10.1038/nbt.2749
- DE WITH, M., A. SADLON, E. CECCHIN, V. HAUFROID, F. THOMAS, M. JOERGER, R. H. N. VAN SCHAIK, R. H. J. MATHIJSEN, C. R. LARGIADÈR, a ‘THE WORKING GROUP ON THE IMPLEMENTATION OF DPD-DEFICIENCY TESTING IN EUROPE’, 2023. Implementation of dihydropyrimidine dehydrogenase deficiency testing in Europe. *ESMO open* [online]. **8**(2), 101197. ISSN 2059-7029. Dostupné z: doi:10.1016/j.esmoop.2023.101197
- DONG, Shengcheng, Nanxiang ZHAO, Emma SPRAGINS, Meenakshi S. KAGDA, Mingjie LI, Pedro ASSIS, Otto JOLANKI, Yunhai LUO, J. Michael CHERRY, Alan P. BOYLE a Benjamin C. HITZ, 2023. Annotating and prioritizing human non-coding variants with RegulomeDB v.2. *Nature Genetics* [online]. **55**(5), 724–726 [vid. 2023-08-27]. ISSN 1546-1718. Dostupné z: doi:10.1038/s41588-023-01365-3
- DVORAK, Pavel, Martin PESTA a Pavel SOUCEK, 2017. ABC gene expression profiles have clinical importance and possibly form a new hallmark of cancer. *Tumor Biology* [online]. **39**(5), 1010428317699800 [vid. 2022-08-05]. ISSN 1010-4283. Dostupné z: doi:10.1177/1010428317699800
- ENCODE PROJECT CONSORTIUM, 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* [online]. **489**(7414), 57–74. ISSN 1476-4687. Dostupné z: doi:10.1038/nature11247
- ERNST, Corinna, Eric HAHNEN, Christoph ENGEL, Michael NOTHNAGEL, Jonas WEBER, Rita K. SCHMUTZLER a Jan HAUKE, 2018. Performance of in silico prediction tools for the classification of rare BRCA1/2 missense variants in clinical diagnostics. *BMC Medical Genomics* [online]. **11**(1), 35 [vid. 2023-04-09]. ISSN 1755-8794. Dostupné z: doi:10.1186/s12920-018-0353-y
- FITZMAURICE, Thomas Simon a Dilip S. NAZARETH, 2021. Incidence of breast cancer in people with cystic fibrosis: A cause for concern? *Journal of Cystic Fibrosis* [online]. **0**(0) [vid. 2022-08-06]. ISSN 1569-1993, 1873-5010. Dostupné z: doi:10.1016/j.jcf.2021.11.017
- FLETCHER, Jamie I., Rebekka T. WILLIAMS, Michelle J. HENDERSON, Murray D. NORRIS a Michelle HABER, 2016. ABC transporters as mediators of drug resistance and contributors to cancer cell biology. *Drug Resistance Updates* [online]. **26**, 1–9 [vid. 2023-05-11]. ISSN 1368-7646. Dostupné z: doi:10.1016/j.drug.2016.03.001
- FRESHOUR, Sharon L., Susanna KIWALA, Kelsy C. COTTO, Adam C. COFFMAN, Joshua F. MCMICHAEL, Jonathan J. SONG, Malachi GRIFFITH, Obi L. GRIFFITH a Alex H.

- WAGNER, 2021. Integration of the Drug-Gene Interaction Database (DGIdb 4.0) with open crowdsource efforts. *Nucleic Acids Research* [online]. **49**(D1), D1144–D1151. ISSN 1362-4962. Dostupné z: doi:10.1093/nar/gkaa1084
- FUKUDA, Yu a John D. SCHUETZ, 2012. ABC Transporters and their Role in Nucleoside and Nucleotide Drug Resistance. *Biochemical Pharmacology* [online]. **83**(8), 1073–1083 [vid. 2022-12-16]. ISSN 0006-2952. Dostupné z: doi:10.1016/j.bcp.2011.12.042
- GAEDIGK, Andrea, Scott T CASEY, Michelle WHIRL-CARRILLO, Neil A MILLER a Teri E KLEIN, 2021. Pharmacogene Variation Consortium: A Global Resource and Repository for Pharmacogene Variation. *Clinical pharmacology and therapeutics* [online]. **110**(3), 542–545 [vid. 2023-07-08]. ISSN 1532-6535. Dostupné z: doi:10.1002/cpt.2321
- GEREK, Nevin Z, Li LIU, Kristyn GEROLD, Pegah BIPARVA, Eric D THOMAS a Sudhir KUMAR, 2015. Evolutionary Diagnosis of non-synonymous variants involved in differential drug response. *BMC Medical Genomics* [online]. **8**(S1) [vid. 2018-08-13]. ISSN 1755-8794. Dostupné z: doi:10.1186/1755-8794-8-S1-S6
- GHOSH, Rajarshi, Ninad OAK a Sharon E. PLON, 2017. Evaluation of in silico algorithms for use with ACMG/AMP clinical variant interpretation guidelines. *Genome Biology* [online]. **18**, 225 [vid. 2018-07-19]. ISSN 1474-760X. Dostupné z: doi:10.1186/s13059-017-1353-5
- GIARDINE, Belinda, Joseph BORG, Emmanouil VIENNAS, Cristiana PAVLIDIS, Kamran MORADKHANI, Philippe JOLY, Marina BARTSAKOULIA, Cathy RIEMER, Webb MILLER, Giannis TZIMAS, Henri WAJCMAN, Ross C. HARDISON a George P. PATRINOS, 2014. Updates of the HbVar database of human hemoglobin variants and thalassemia mutations. *Nucleic Acids Research* [online]. **42**(D1), D1063–D1069 [vid. 2023-07-07]. ISSN 0305-1048. Dostupné z: doi:10.1093/nar/gkt911
- GRIMM, Dominik G., Chloé-Agathe AZENCOTT, Fabian AICHELER, Udo GIERATHS, Daniel G. MACARTHUR, Kaitlin E. SAMOCHA, David N. COOPER, Peter D. STENSON, Mark J. DALY, Jordan W. SMOLLER, Laramie E. DUNCAN a Karsten M. BORGWARDT, 2015. The Evaluation of Tools Used to Predict the Impact of Missense Variants Is Hindered by Two Types of Circularity. *Human Mutation* [online]. **36**(5), 513–523 [vid. 2018-02-25]. ISSN 1098-1004. Dostupné z: doi:10.1002/humu.22768
- HASSAN, Marwa S., A. A. SHAALAN, M. I. DESSOUKY, Abdelaziz E. ABDELNAIEM a Mahmoud ELHEFNAWI, 2019a. A review study: Computational techniques for expecting the impact of non-synonymous single nucleotide variants in human diseases. *Gene* [online]. **680**, 20–33 [vid. 2023-07-07]. ISSN 0378-1119. Dostupné z: doi:10.1016/j.gene.2018.09.028
- HASSAN, Marwa S., A. A. SHAALAN, M. I. DESSOUKY, Abdelaziz E. ABDELNAIEM a Mahmoud ELHEFNAWI, 2019b. Evaluation of computational techniques for predicting non-synonymous single nucleotide variants pathogenicity. *Genomics* [online]. **111**(4), 869–882 [vid. 2023-07-07]. ISSN 0888-7543. Dostupné z: doi:10.1016/j.ygeno.2018.05.013
- HENRICKS, Linda M, Carin A T C LUNENBURG, Femke M DE MAN, Didier MEULENDIJKS, Geert W J FREDERIX, Emma KIENHUIS, Geert-Jan CREEMERS, Arnold BAARS, Vincent O DEZENTJÉ, Alexander L T IMHOLZ, Frank J F JEURISSEN, Johanna E A PORTIELJE, Rob L H JANSEN, Paul HAMBERG, Albert J TEN TIJE, Helga

- J DROOGENDIJK, Miriam KOOPMAN, Peter NIEBOER, Marlène H W VAN DE POEL, Caroline M P W MANDIGERS, Hilde ROSING, Jos H BEIJNEN, Erik van WERKHOVEN, André B P VAN KUILENBURG, Ron H N VAN SCHAİK, Ron H J MATHIJSSSEN, Jesse J SWEN, Hans GELDERBLUM, Annemieke CATS, Henk-Jan GUCHELAAR a Jan H M SCHELLENS, 2018. DPYD genotype-guided dose individualisation of fluoropyrimidine therapy in patients with cancer: a prospective safety analysis. *The Lancet Oncology* [online]. **19**(11), 1459–1467 [vid. 2023-07-07]. ISSN 1470-2045. Dostupné z: doi:10.1016/S1470-2045(18)30686-7
- HLAVAC, Viktor, Maria KOVACOVA, Katerina ELSNEROVA, Veronika BRYNYCHOVA, Renata KOZEVNIKOVOVA, Karel RAUS, Katerina KOPECKOVA, Sona MESTAKOVA, David VRANA, Jiri GATEK, Pavel OSTASOV, Radka VACLAVIKOVA a Pavel SOUCEK, 2018. Use of Germline Genetic Variability for Prediction of Chemoresistance and Prognosis of Breast Cancer Patients. *Cancers* [online]. **10**(12), 511 [vid. 2019-01-14]. Dostupné z: doi:10.3390/cancers10120511
- HLAVÁČ, Viktor, Veronika BRYNYCHOVÁ, Radka VÁCLAVÍKOVÁ, Marie EHRLICHOVÁ, David VRÁNA, Václav PECHA, Renata KOŽEVNIKOVOVÁ, Markéta TRNKOVÁ, Jiří GATĚK, Dana KOPPEROVÁ, Ivan GUT a Pavel SOUČEK, 2013. The expression profile of ATP-binding cassette transporter genes in breast carcinoma. *Pharmacogenomics* [online]. **14**(5), 515–529 [vid. 2023-06-16]. ISSN 1462-2416. Dostupné z: doi:10.2217/pgs.13.26
- HLAVÁČ, Viktor, Veronika BRYNYCHOVÁ, Radka VÁCLAVÍKOVÁ, Marie EHRLICHOVÁ, David VRÁNA, Václav PECHA, Markéta TRNKOVÁ, Roman KODET, Marcela MRHALOVÁ, Kateřina KUBÁČKOVÁ, Jiří GATĚK, Petr VÁŽAN a Pavel SOUČEK, 2014. The role of cytochromes p450 and aldo-keto reductases in prognosis of breast carcinoma patients. *Medicine* [online]. **93**(28), e255. ISSN 1536-5964. Dostupné z: doi:10.1097/MD.0000000000000255
- HUANG, Li-Hong, Qiu-Shun HE, Ke LIU, Jiao CHENG, Min-Dong ZHONG, Lin-Shan CHEN, Li-Xia YAO a Zhi-Liang JI, 2018. ADReCS-Target: target profiles for aiding drug safety research and application. *Nucleic Acids Research* [online]. **46**(D1), D911–D917 [vid. 2018-01-25]. ISSN 0305-1048, 1362-4962. Dostupné z: doi:10.1093/nar/gkx899
- CHEN, Zhaolin, Tianlu SHI, Lei ZHANG, Pengli ZHU, Mingying DENG, Cheng HUANG, Tingting HU, Ling JIANG a Jun LI, 2016. Mammalian drug efflux transporters of the ATP binding cassette (ABC) family in multidrug resistance: A review of the past decade. *Cancer Letters* [online]. **370**(1), 153–164 [vid. 2023-05-11]. ISSN 0304-3835. Dostupné z: doi:10.1016/j.canlet.2015.10.010
- INGRAM, V. M., 1957. Gene Mutations in Human Hæmoglobin: the Chemical Difference Between Normal and Sickle Cell Hæmoglobin. *Nature* [online]. **180**(4581), 326–328 [vid. 2023-07-07]. ISSN 1476-4687. Dostupné z: doi:10.1038/180326a0
- JONES, Clyde M., Samuel CHARACHE a Patricia J. HATHAWAY, 1979. The Effect of Hemoglobin F-Chesapeake (α 292 Arg. \rightarrow Leu γ 2) on Fetal Oxygen Affinity and Erythropoiesis. *Pediatric Research* [online]. **13**(7), 851–853 [vid. 2023-07-07]. ISSN 1530-0447. Dostupné z: doi:10.1203/00006450-197907000-00011

- KANZAKI, Atsuko, Masakazu TOI, Nouri NEAMATI, Hitoshi MIYASHITA, Masahiro OUBU, Kentaro NAKAYAMA, Hiroko BANDO, Kenji OGAWA, Masato MUTOH, Shiro MORI, Kunihiko TERADA, Toshihiro SUGIYAMA, Manabu FUKUMOTO a Yuji TAKEBAYASHI, 2002. Copper-transporting P-Type Adenosine Triphosphatase (ATP7B) Is Expressed in Human Breast Carcinoma. *Japanese Journal of Cancer Research* [online]. **93**(1), 70–77 [vid. 2023-05-11]. ISSN 1349-7006. Dostupné z: doi:10.1111/j.1349-7006.2002.tb01202.x
- KHATRI, A., B. W. WILLIAMS, J. FISHER, R. C. BRUNDAGE, V. J. GURVICH, L. G. LIS, K. M. SKUBITZ, A. Z. DUDEK, E. W. GREENO, R. A. KRATZKE, J. K. LAMBA a M. N. KIRSTEIN, 2014. SLC28A3 genotype and gemcitabine rate of infusion affect dFdCTP metabolite disposition in patients with solid tumours. *British Journal of Cancer* [online]. **110**(2), 304–312. ISSN 1532-1827. Dostupné z: doi:10.1038/bjc.2013.738
- KIM, SungHwan, Jae-Hwan JHONG, JungJun LEE a Ja-Yong KOO, 2017. Meta-analytic support vector machine for integrating multiple omics data. *BioData Mining* [online]. **10**, 2 [vid. 2023-02-10]. ISSN 1756-0381. Dostupné z: doi:10.1186/s13040-017-0126-8
- KISLING, Lisa A. a Joe M DAS, 2023. Prevention Strategies. In: *StatPearls* [online]. Treasure Island (FL): StatPearls Publishing [vid. 2023-06-19]. Dostupné z: <http://www.ncbi.nlm.nih.gov/books/NBK537222/>
- KNIKMAN, Jonathan E., Hans GELDERBLOM, Jos H. BEIJNEN, Annemieke CATS, Henk-Jan GUCHELAAR a Linda M. HENRICKS, 2021. Individualized Dosing of Fluoropyrimidine-Based Chemotherapy to Prevent Severe Fluoropyrimidine-Related Toxicity: What Are the Options? *Clinical Pharmacology and Therapeutics* [online]. **109**(3), 591–604 [vid. 2023-04-09]. ISSN 0009-9236. Dostupné z: doi:10.1002/cpt.2069
- KOLAK, Agnieszka, Marzena KAMIŃSKA, Katarzyna SYGIT, Agnieszka BUDNY, Dariusz SURDYKA, Bożena KUKIEŁKA-BUDNY a Franciszek BURDAN, 2017. Primary and secondary prevention of breast cancer. *Annals of Agricultural and Environmental Medicine* [online]. **24**(4), 549–553 [vid. 2023-03-25]. ISSN 1232-1966, 1898-2263. Dostupné z: doi:10.26444/aaem/75943
- KOVÁČOVÁ, Mária, Viktor HLAVÁČ, Renata KOŽEVNIKOVOVÁ, Karel RAUŠ, Jiří GATĚK a Pavel SOUČEK, 2024. Artificial Intelligence-driven Prediction Revealed CFTR Associated With Therapy Outcome Of Breast Cancer: A Feasibility Study. *Oncology* [online]. ISSN 1423-0232. Dostupné z: doi:10.1159/000540395
- KRUS, Ivona, Veronika BRYNYCHOVÁ, Viktor HLAVÁČ, Radka VÁCLAVÍKOVÁ, Maria KOVÁČOVÁ, Renata KOŽEVNIKOVOVÁ, Katerina KOPEČKOVÁ, Jannis TORNIKIDIS, David VRÁNA, Jiří GATĚK a Pavel SOUČEK, 2022. Single Nucleotide Variants in KIF14 Gene May Have Prognostic Value in Breast Cancer. *Molecular Diagnosis & Therapy* [online]. **26**(6), 665–678 [vid. 2023-02-11]. ISSN 1179-2000. Dostupné z: doi:10.1007/s40291-022-00616-z
- KULSHRESHTHA, Shweta, Vigi CHAUDHARY, Girish K. GOSWAMI a Nidhi MATHUR, 2016. Computational approaches for predicting mutant protein stability. *Journal of Computer-Aided Molecular Design* [online]. **30**(5), 401–412 [vid. 2023-07-07]. ISSN 1573-4951. Dostupné z: doi:10.1007/s10822-016-9914-3

- LI, Tailin, Jingbo PENG, Feiyue ZENG, Keqiang ZHANG, Jinyang LIU, Xi LI, Qianying OUYANG, Guo WANG, Liansheng WANG, Zhaoqian LIU a Yingzi LIU, 2017. Association between polymorphisms in CTR1, CTR2, ATP7A, and ATP7B and platinum resistance in epithelial ovarian cancer. *International Journal of Clinical Pharmacology and Therapeutics* [online]. **55**(10), 774–780. ISSN 0946-1965. Dostupné z: doi:10.5414/CP202907
- LIU, Ye, William S. B. YEUNG, Philip C. N. CHIU a Dandan CAO, 2022. Computational approaches for predicting variant impact: An overview from resources, principles to applications. *Frontiers in Genetics* [online]. **13**, 981005 [vid. 2023-03-12]. ISSN 1664-8021. Dostupné z: doi:10.3389/fgene.2022.981005
- LOPES, Inês, Gulam ALTAB, Priyanka RAINA a João Pedro DE MAGALHÃES, 2021. Gene Size Matters: An Analysis of Gene Length in the Human Genome. *Frontiers in Genetics* [online]. **12** [vid. 2022-08-06]. ISSN 1664-8021. Dostupné z: <https://www.frontiersin.org/articles/10.3389/fgene.2021.559998>
- MALHOTRA, Gautam K, Xiangshan ZHAO, Hamid BAND a Vimla BAND, 2010. Histological, molecular and functional subtypes of breast cancers. *Cancer Biology & Therapy* [online]. **10**(10), 955–960 [vid. 2023-03-19]. ISSN 1538-4047. Dostupné z: doi:10.4161/cbt.10.10.13879
- MARTÍNEZ, G., F. LIMA a B. COLOMBO, 1977. Haemoglobin J Guantanamo ($\alpha 2\beta 2$ (H6) Ala^{Asp}) A new fast unstable haemoglobin found in a Cuban family. *Biochimica et Biophysica Acta (BBA) - Protein Structure* [online]. **491**(1), 1–6 [vid. 2023-07-07]. ISSN 0005-2795. Dostupné z: doi:10.1016/0005-2795(77)90034-4
- MARTINS, Carlos, Maciek GODYCKI-CWIRKO, Bruno HELENO a John BRODERSEN, 2018. Quaternary prevention: reviewing the concept: Quaternary prevention aims to protect patients from medical harm. *European Journal of General Practice* [online]. **24**(1), 106–111 [vid. 2023-06-19]. ISSN 1381-4788, 1751-1402. Dostupné z: doi:10.1080/13814788.2017.1422177
- MCDONAGH, Ellen M, Michelle WHIRL-CARRILLO, Yael GARTEN, Russ B ALTMAN a Teri E KLEIN, 2011. From pharmacogenomic knowledge acquisition to clinical applications: the PharmGKB as a clinical pharmacogenomic biomarker resource. *Biomarkers in Medicine* [online]. **5**(6), 795–806 [vid. 2018-06-30]. ISSN 1752-0363. Dostupné z: doi:10.2217/bmm.11.94
- MEI, Shenghui, Xindi LI, Xiaoqing GONG, Xingang LI, Li YANG, Heng ZHOU, Yonghong LIU, Anna ZHOU, Leting ZHU, Xinghu ZHANG a Zhigang ZHAO, 2017. LC-MS/MS Analysis of Erythrocyte Thiopurine Nucleotides and Their Association With Genetic Variants in Patients With Neuromyelitis Optica Spectrum Disorders Taking Azathioprine. *Therapeutic Drug Monitoring* [online]. **39**(1), 5 [vid. 2023-07-07]. ISSN 0163-4356. Dostupné z: doi:10.1097/FTD.0000000000000362
- MOO-PENN, Winston F., Danny L. JUE, Mary H. JOHNSON, Kenneth W. OLSEN, Daniel SHIH, Richard T. JONES, Sam E. LUX, Paul RODGERS a Arthur ARNONE, 1988. Hemoglobin Brockton [β .138 (H16) Ala .fwdarw. Pro]: an unstable variant near the C-terminus of the β -subunits with normal oxygen-binding properties. *Biochemistry*

[online]. **27**(20), 7614–7619 [vid. 2023-07-07]. ISSN 0006-2960. Dostupné z: doi:10.1021/bi00420a007

NAKAYAMA, Kentaro, Atsuko KANZAKI, Kenji OGAWA, Kohji MIYAZAKI, Nouri NEAMATI a Yuji TAKEBAYASHI, 2002. Copper-transporting P-type adenosine triphosphatase (ATP7B) as a cisplatin based chemoresistance marker in ovarian carcinoma: Comparative analysis with expression of MDR1, MRP1, MRP2, LRP and BCRP. *International Journal of Cancer* [online]. **101**(5), 488–495 [vid. 2023-05-11]. ISSN 1097-0215. Dostupné z: doi:10.1002/ijc.10608

O'DONNELL, Peter H. a Mark J. RATAIN, 2012. Germline pharmacogenomics in oncology: Decoding the patient for targeting therapy. *Molecular Oncology* [online]. **6**(2), 251–259 [vid. 2017-11-13]. ISSN 1574-7891. Dostupné z: doi:10.1016/j.molonc.2012.01.005

OKIDO, Toshihisa, Yuichi KODAMA, Jun MASHIMA, Takehide KOSUGE, Takatomo FUJISAWA a Osamu OGASAWARA, 2022. DNA Data Bank of Japan (DDBJ) update report 2021. *Nucleic Acids Research* [online]. **50**(D1), D102–D105 [vid. 2023-08-11]. ISSN 0305-1048. Dostupné z: doi:10.1093/nar/gkab995

PARK, Yeon Hee, Kyung Hae JUNG, Seock-Ah IM, Joo Hyuk SOHN, Jungsil RO, Jin-Hee AHN, Sung-Bae KIM, Byung-Ho NAM, Do Youn OH, Sae-Won HAN, Soohyeon LEE, In Hae PARK, Keun Seok LEE, Jee Hyun KIM, Seok Yun KANG, Moon Hee LEE, Hee Sook PARK, Jin Seok AHN a Young-Hyuck IM, 2013. Phase III, Multicenter, Randomized Trial of Maintenance Chemotherapy Versus Observation in Patients With Metastatic Breast Cancer After Achieving Disease Control With Six Cycles of Gemcitabine Plus Paclitaxel As First-Line Chemotherapy: KCSG-BR07-02. *Journal of Clinical Oncology* [online]. **31**(14), 1732–1739 [vid. 2023-07-07]. ISSN 0732-183X. Dostupné z: doi:10.1200/JCO.2012.45.2490

QUINODOZ, Mathieu, Virginie G. PETER, Katarina CISAROVA, Beryl ROYER-BERTRAND, Peter D. STENSON, David N. COOPER, Sheila UNGER, Andrea SUPERTI-FURGA a Carlo RIVOLTA, 2022. Analysis of missense variants in the human genome reveals widespread gene-specific clustering and improves prediction of pathogenicity. *The American Journal of Human Genetics* [online]. **109**(3), 457–470 [vid. 2023-05-29]. ISSN 0002-9297, 1537-6605. Dostupné z: doi:10.1016/j.ajhg.2022.01.006

RAMSEY, Laura B., Gitte H. BRUUN, Wenjian YANG, Lisa R. TREVIÑO, Selina VATTATHIL, Paul SCHEET, Cheng CHENG, Gary L. ROSNER, Kathleen M. GIACOMINI, Yiping FAN, Alex SPARREBOOM, Torben S. MIKKELSEN, Thomas J. CORYDON, Ching-Hon PUI, William E. EVANS a Mary V. RELLING, 2012. Rare versus common variants in pharmacogenetics: SLCO1B1 variation and methotrexate disposition. *Genome Research* [online]. **22**(1), 1–8 [vid. 2023-06-24]. ISSN 1088-9051, 1549-5469. Dostupné z: doi:10.1101/gr.129668.111

RUZZO, A., F. GRAZIANO, Fabio GALLI, Francesca GALLI, E. RULLI, S. LONARDI, M. RONZONI, B. MASSIDDA, V. ZAGONEL, N. PELLA, C. MUCCIARINI, R. LABIANCA, M. T. IONTA, I. BAGALONI, E. VELTRI, P. SOZZI, S. BARNI, V. RICCI, L. FOLTRAN, M. NICOLINI, E. BIONDI, A. BRAMATI, D. TURCI, S. LAZZARELLI, C. VERUSIO, F. BERGAMO, A. SOBRERO, L. FRONTINI, M. MENGHI a M. MAGNANI, 2017. Dihydropyrimidine dehydrogenase pharmacogenetics for predicting fluoropyrimidine-related toxicity in the randomised, phase III adjuvant TOSCA trial in high-

- risk colon cancer patients. *British Journal of Cancer* [online]. **117**(9), 1269–1277 [vid. 2023-07-07]. ISSN 1532-1827. Dostupné z: doi:10.1038/bjc.2017.289
- SAHAKYAN, Aleksandr B. a Shankar BALASUBRAMANIAN, 2016. Long genes and genes with multiple splice variants are enriched in pathways linked to cancer and other multigenic diseases. *BMC Genomics* [online]. **17**(1), 225 [vid. 2022-11-30]. ISSN 1471-2164. Dostupné z: doi:10.1186/s12864-016-2582-9
- SAYERS, Eric W, Mark CAVANAUGH, Karen CLARK, Kim D PRUITT, Conrad L SCHOCH, Stephen T SHERRY a Ilene KARSCH-MIZRACHI, 2022. GenBank. *Nucleic Acids Research* [online]. **50**(D1), D161–D164 [vid. 2023-08-11]. ISSN 0305-1048. Dostupné z: doi:10.1093/nar/gkab1135
- SCHIPPER, Marijn a Danielle POSTHUMA, 2022. Demystifying non-coding GWAS variants: an overview of computational tools and methods. *Human Molecular Genetics* [online]. **31**(R1), R73–R83 [vid. 2023-10-15]. ISSN 0964-6906. Dostupné z: doi:10.1093/hmg/ddac198
- SIEGEL, Rebecca L., Kimberly D. MILLER, Hannah E. FUCHS a Ahmedin JEMAL, 2022. Cancer statistics, 2022. *CA: A Cancer Journal for Clinicians* [online]. **72**(1), 7–33 [vid. 2023-03-17]. ISSN 1542-4863. Dostupné z: doi:10.3322/caac.21708
- SMIRNOV, Petr, Victor KOFIA, Alexander MARU, Mark FREEMAN, Chantal HO, Nehme EL-HACHEM, George-Alexandru ADAM, Wail BA-ALAWI, Zhaleh SAFIKHANI a Benjamin HAIBE-KAINS, 2018. PharmacoDB: an integrative database for mining in vitro anticancer drug screening studies. *Nucleic Acids Research* [online]. **46**(D1), D994–D1002. ISSN 1362-4962. Dostupné z: doi:10.1093/nar/gkx911
- VAN KUILENBURG, André B. P, 2004. Dihydropyrimidine dehydrogenase and the efficacy and toxicity of 5-fluorouracil. *European Journal of Cancer* [online]. **40**(7), 939–950 [vid. 2023-07-07]. ISSN 0959-8049. Dostupné z: doi:10.1016/j.ejca.2003.12.004
- VISSCHER, H., C. J. D. ROSS, S. R. RASSEKH, G. S. S. SANDOR, H. N. CARON, E. C. VAN DALEN, L. C. KREMER, H. J. VAN DER PAL, P. C. ROGERS, M. J. RIEDER, B. C. CARLETON, M. R. HAYDEN, a CPNDS CONSORTIUM, 2013. Validation of variants in SLC28A3 and UGT1A6 as genetic markers predictive of anthracycline-induced cardiotoxicity in children. *Pediatric Blood & Cancer* [online]. **60**(8), 1375–1381. ISSN 1545-5017. Dostupné z: doi:10.1002/pbc.24505
- WAKS, Adrienne G. a Eric P. WINER, 2019. Breast Cancer Treatment: A Review. *JAMA* [online]. **321**(3), 288–300 [vid. 2023-03-12]. ISSN 0098-7484. Dostupné z: doi:10.1001/jama.2018.19323
- WOJNOWSKI, Leszek, Bettina KULLE, Markus SCHIRMER, Gregor SCHLÜTER, Albrecht SCHMIDT, Albert ROSENBERGER, Stefan VONHOF, Heike BICKEBÖLLER, Mohammad Reza TOLIAT, Eun-Kyung SUK, Mladen TZVETKOV, Anke KRUGER, Silvia SEIFERT, Marita KLOESS, Heidi HAHN, Markus LOEFFLER, Peter NÜRNBERG, Michael PFREUNDSCHUH, Lorenz TRÜMPER, Jürgen BROCKMÖLLER a Gerd HASENFUSS, 2005. NAD(P)H oxidase and multidrug resistance protein genetic polymorphisms are associated with doxorubicin-induced cardiotoxicity. *Circulation* [online].

112(24), 3754–3762. ISSN 1524-4539. Dostupné z: doi:10.1161/CIRCULATIONAHA.105.576850

YAZAR, Metin a Pemra OZBEK, 2022. Assessment of 13 in silico pathogenicity methods on cancer-related variants. *Computers in Biology and Medicine* [online]. **145**, 105434 [vid. 2022-11-19]. ISSN 0010-4825. Dostupné z: doi:10.1016/j.combiomed.2022.105434

ZHANG, Jieting, Yan WANG, Xiaohua JIANG a Hsiao Chang CHAN, 2018. Cystic fibrosis transmembrane conductance regulator—emerging regulator of cancer. *Cellular and Molecular Life Sciences* [online]. **75**(10), 1737–1756 [vid. 2022-08-05]. ISSN 1420-9071. Dostupné z: doi:10.1007/s00018-018-2755-6

ZHANG, Xuefeng, Zhichen PU, Jun GE, Jie SHEN, Xiaolong YUAN a Haitang XIE, 2015. Association of CYP2D6*10, OATP1B1 A388G, and OATP1B1 T521C polymorphisms and overall survival of breast cancer patients after tamoxifen therapy. *Medical Science Monitor: International Medical Journal of Experimental and Clinical Research* [online]. **21**, 563–569. ISSN 1643-3750. Dostupné z: doi:10.12659/MSM.893473

ZHOU, Yitian, Souren MKRTCHIAN, Masaki KUMONDAI, Masahiro HIRATSUKA a Volker M. LAUSCHKE, 2018. An optimized prediction framework to assess the functional impact of pharmacogenetic variants. *The Pharmacogenomics Journal* [online]. 1 [vid. 2018-09-23]. ISSN 1473-1150. Dostupné z: doi:10.1038/s41397-018-0044-2

Seznam publikací

Publikace, které jsou podkladem práce s IF

HLAVAC, Viktor*, Maria KOVACOVA*, Katerina ELSNEROVA, Veronika BRYNYCHOVA, Renata KOZEVNIKOVOVA, Karel RAUS, Katerina KOPECKOVA, Soná MESTAKOVA, David VRANA, Jiri GATEK, Pavel OSTASOV, Radka VACLAVIKOVA a Pavel SOUCEK, 2018. Use of Germline Genetic Variability for Prediction of Chemoresistance and Prognosis of Breast Cancer Patients. *Cancers* [online]. 10(12), 511 [vid. 2019-01-14]. Dostupné z: doi: 10.3390/cancers10120511, IF_[2023] = 4,5

*Oba autoři přispěli k této práci stejným dílem.

Podíl autorky dizertace v % na jednotlivých aspektech díla: přípravě projektu 50%, provádění prací 45%, interpretaci výsledků 50% a přípravě publikace 50%.

KOVÁČOVÁ, Mária, Viktor HLAVÁČ, Renata KOŽEVNIKOVOVÁ, Karel RAUŠ, Jiří GATĚK a Pavel SOUČEK, 2024. Artificial Intelligence-driven Prediction Revealed CFTR Associated With Therapy Outcome Of Breast Cancer: A Feasibility Study. *Oncology* [online]. ISSN 1423-0232. Dostupné z: doi:10.1159/000540395, IF_[2023] = 2,5

Podíl autorky dizertace v % na jednotlivých aspektech díla: přípravě projektu 70%, provádění prací 80%, interpretaci výsledků 100% a přípravě publikace 90%.

KRUS, Ivona, Veronika BRYNYCHOVÁ, Viktor HLAVÁČ, Radka VÁCLAVÍKOVÁ, Maria KOVÁČOVÁ, Renata KOŽEVNIKOVOVÁ, Katerina KOPEČKOVÁ, Jannis TORNIKIDIS, David VRANA, Jiří GATĚK a Pavel SOUČEK, 2022. Single Nucleotide Variants in KIF14 Gene May Have Prognostic Value in Breast Cancer. *Molecular Diagnosis & Therapy* [online]. 26(6), 665–678 [vid. 2023-02-11]. ISSN 1179-2000. Dostupné z: doi:10.1007/s40291-022-00616-z, IF_[2023] = 4,1

Podíl autorky dizertace v % na jednotlivých aspektech díla: přípravě projektu 20%, provádění prací 10%, interpretaci výsledků 40% a přípravě publikace 15%.

Publikace, které nejsou podkladem práce (bez IF)

KOVÁČOVÁ, Mária; HLAVÁČ, Viktor; BRYNYCHOVÁ, Veronika; VÁCLAVÍKOVÁ, Radka; SOUČEK, Pavel. Gene-based evidence for burden of rare pathogenic variants in pharmacogenes and oncogenes of Czech breast cancer patients. In: *Proceedings 2019 IEEE International Conference on Bioinformatics and Biomedicine*. San Diego: Institute of Electrical and Electronics Engineers, 2019, p. 1214-1216. ISBN 978-1-72811-867-3. DOI: [10.1109/BIBM47256.2019.8983216](https://doi.org/10.1109/BIBM47256.2019.8983216). Conference: IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2019, San Diego, 18.11.2019 – 21.11.2019. Typ publikace: C02. Statě ve sbornících; stať v konferenčním sborníku v cizím jazyce

Podíl autorky dizertace v % na jednotlivých aspektech díla: přípravě projektu 70%, provádění prací 80%, interpretaci výsledků 100% a přípravě publikace 90%.