

UNIVERZITA KARLOVA
3. LÉKAŘSKÁ FAKULTA

Disertační práce

Praha, červenec 2024

PharmDr. Mária Kováčová

UNIVERZITA KARLOVA
3. LÉKAŘSKÁ FAKULTA



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

Školitel: prof. RNDr. Pavel Souček, CSc

Praha, červenec 2024

PharmDr. Mária Kováčová

OBSAH

1.	Souhrn.....	10
1.1.	Summary.....	12
2.	Literární přehled, současný stav poznání.....	14
2.1.	Epidemiologie a etiologie karcinomu prsu.....	14
2.2.	Prevence karcinomu prsu.....	16
2.3.	Farmakoterapie karcinomu prsu.....	19
2.4.	Precizní medicína.....	20
2.5.	Bioinformatika, studie <i>in silico</i>	22
2.5.1.	Bioinformatické databáze.....	23
2.5.2.	Bioinformatické nástroje.....	27
2.5.3.	Výběr vhodného(ých) prediktoru(ů) pro koncového uživatele.....	47
2.6.	Hypotéza: Cíle řešení.....	55
3.	Materiál, zdrojová data a metody.....	56
3.1.1.	Materiál a pacienti.....	56
3.1.2.	Populační zdrojová data.....	58
3.1.3.	Bioinformatické algoritmy a metody pro funkční predikci častých variant.....	59
3.1.4.	Bioinformatické algoritmy a metody pro funkční predikci vzácných variant.....	61
3.1.5.	Farmakogenomická data – časté varianty.....	65
3.1.6.	Farmakogenomická data – vzácné varianty.....	66
3.1.7.	Bioinformatické algoritmy pro funkční predikci genu <i>KIF14</i>	66
3.1.8.	Statistické metody.....	66
4.	Výsledky.....	69
4.1.	Výsledky pilotní studie častých variant – bioinformatická analýza.....	69
4.1.1.	Datamining - farmakogenomické databáze.....	69
4.1.2.	<i>In silico</i> predikce.....	80
4.2.	Výsledky validační studie častých variant.....	86
4.3.	Výsledky studie proveditelnosti – vzácné varianty.....	90
4.4.	Výsledky pilotní studie – varianty genu <i>KIF14</i>	99
5.	Diskuse metodických postupů a výsledků.....	100
5.1.1.	Pilotní a validační studie – bioinformatická analýza.....	100
5.1.2.	Predikce vzácných variant.....	105
5.1.3.	Predikce variant genu <i>KIF14</i>	108
5.2.	Závěry.....	109
5.2.1.	Validační studie.....	109
5.2.2.	Studie vzácných variant.....	110
5.2.3.	Studie variant genu <i>KIF14</i>	110

Seznam publikací	112
Seznam příkladových kódů.....	113
Seznam webových zdrojů	116
Seznam použité literatury	117

Seznam použitých zkratk:

ABC	z angl. ATP binding cassette transporters
ACMG/ AMP	z angl. American College of Medical Genetics/ Association for Molecular Pathology
AI	umělá inteligence, z angl. Artificial Intelligence
ALFA	z angl. Allele Frequency Aggregator (ALFA)
C50	zhoubný novotvar prsu
CI	interval spolehlivosti, z angl. Confidential Interval
CPIC	z angl. Clinical Pharmacogenetics Implementation Consortium
CYP	Cytochrom P450
dbGaP	z angl. database of Genotypes and Phenotypes
DFS	přežití bez nemoci, z angl. Disease - Free Survival
EMA	Evropská léková agentura, z angl. European Medicine Agency
ER	Estrogenový Receptor
ExAC	z angl. Exome Aggregation Consortium
FDA	z angl. Food and Drug Administration
FDR	míra falešného objevu, z angl. False Discovery Rate
GCO	z angl. Global Cancer Observatory
GPR	genový patogenní poměr, z angl. Gene Pathogenic Ratio
GRCh37	z angl. Genome Reference Consortium Human Build 37 (hg19)
GRCh38	z angl. Genome Reference Consortium Human Build 38 (hg38)
HCSC	z angl. Health Canada, Santé Canada
HER2/ ERBB2	lidský epidermální růstový faktor receptoru 2/ erb-b2 tyrozinkinázový receptor z angl. Human Epidermal growth factor Receptor 2/ erb-b2 receptor tyrosine kinase
HGMD	z angl. Human Gene Mutation Database
HGVS	z angl. Human Genome Variation Society
HPO	z angl. Human Phenotype Ontology
HR	poměr rizik, z angl. Hazard Ratio
HWE	Hardyho–Weinbergův zákon, z angl. Hardy-Weinberg equilibrium
IDC	invazivní duktální karcinom
KBDB	z angl. Knowledge based Database
kbp	z angl. kilo base pair
Ki-67	marker proliferace Ki-67
LOVD	z angl. Leiden Open-source Variation Database
LSDBs	z angl. Locus Specific Databases
ML	strojové učení, z angl. Machine Learning
NCMG	Národní centrum lékařské genomiky
NGS	Next-Generation Sequencing
nsSNV/s	nesynonymní jednonukleotidová varianta/ varianty
OS	celkové přežití, z angl. Overall Survival
PR	Progesteronový Receptor
rsID	referenční identifikátor varianty

S.D.	směrodatná odchylka, z angl. standard deviation
SNV	jednonukleotidová varianta
SmPC	Souhrn Údajů o Přípravku, z angl. Summary of Product Characteristics
sSNV	synonymní jednonukleotidová varianta
TFBP	profil vazby transkripčního faktoru, z angl. Transcription Factor Binding Profile
TFBS	vazebné místo transkripčního faktoru, z angl. Transcription Factor Binding Site
TNBC	Trojité negativní karcinom prsu, z angl. Triple - Negative Breast Cancer
UTR	nepřekládaná oblast, z angl. UnTranslated Region
ÚZIS	Ústav Zdravotnických Informací a Statistiky ČR
VCF	z angl. Variant Call Format
VDA	z angl. Variant - Drug Associations
VEP	z angl. Variant Effect Predictor

Prohlášení:

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

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

V Praze, 25.7.2024

Mária Kováčová

Identifikační záznam:

Kováčová, Mária. 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]. Praha, 2024. 145 stran. Disertační práce. Univerzita Karlova, 3. lékařská fakulta, Preventivní medicína a epidemiologie. Vedoucí závěrečné práce: prof. RNDr. Pavel Souček, CSc.

Klíčová slova: bioinformatika, nádor prsu, farmakogenomika, prediktivní algoritmy, terciární prevence

Key words: bioinformatics, breast cancer, pharmacogenomics, predictive algorithms, tertiary prevention

Poděkování

Ráda bych poděkovala svému školiteli prof. RNDr. Pavlu Součkovi, CSc. za odborné vedení a konzultace při vyvíjení metod a zpracování dat, při psaní článků a této disertační práce. Dále bych chtěla velice poděkovat kolegům a kolegyním v laboratoři Oddělení Toxikogenomiky ve Státním zdravotním ústavu v Praze. Nejvíce bych chtěla poděkovat Mgr. Viktorovi Hlaváčovi, Ph.D., za předání zkušeností, cenné rady a pomoc se studiem.

Velké poděkování patří všem spoluautorům, kteří jsou uvedeni v publikovaných pracích. Na závěr bych chtěla poděkovat své rodině a přátelům za podporu bez nichž by studium a vypracování této práce bylo mnohem náročnější.

Poděkování patří Národnímu centru lékařské genomiky za poskytnutí alelických frekvencí pro srovnání s etnicky odpovídající populací (projekt CZ.02.1.01/0.0/0.0/16_013/0001634), grantovým agenturám za nezbytnou finanční podporu k publikování studií, které jsou podkladem této disertační práce (GAUK 1776218 a NV22-08-0028).

1. Souhrn

Karcinom prsu, geneticky heterogenní onemocnění, zůstává u žen hlavní příčinou úmrtí, a přestože mortalita postupně ale pomalu klesá, incidence opětovně roste. Plně funkční primární prevence cílená na snížení rizika vzniku onemocnění zatím není k dispozici a proto sekundární, terciární a popřípadě i kvartérní prevence se stávají součástí klinického managementu u onkologických pacientů. Nejúčinnějším prostředkem je včasná diagnóza, sekundární prevence. Terciární prevence s cílem zabránit progresi onemocnění napomáhá ztrátě soběstačnosti a s ní spojené snížení kvality života. Kvartérní prevence s cílem identifikovat pacienta ohroženého nadměrnou medikací se zaměřuje na snížení dodatečných nežádoucích účinků terapie. Terciární a kvartérní prevence vychází ze studia genetických rozdílů u pacientů a jejich následné stratifikace podle genetické predispozice a molekulární povahy karcinomu a účinků chemoterapie. Z důvodu genetických změn lze studovat i sníženou odpověď na léčbu, která může být zapříčiněna variabilitou v metabolismu, či transportu léčiva a následné expozice léčivu.

Od doby, kdy v roce 2003 celosvětově největší biologický projekt – Human Genome Project, vygeneroval data pokrývající přibližně 92 % lidského genomu se dostáváme do doby bioinformatické, kdy po 20ti letech již disponujeme 100% pokrytím lidského genomu. 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 a farmakogenetických dat, a tak k zautomatizování procesů predikce genetického profilu pacientek s nádorem prsu ve vztahu k terciární a popřípadě kvartérní prevenci.

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 několik signálních drah v klinické onkologii, za pomoci nových *in silico* nástrojů a zavedených farmakogenomických databází. *In silico* 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 luminálními 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 bylo cílem práce navrhnout efektivní interpretaci funkční predikce bioinformatickými nástroji založenými na metodách strojového učení a sekvenčních a strukturních algoritmech. Námí nově navržená strategie s cílem vyvážit některé z nedostatků metod strojového učení, vyhodnotila přiměřený počet genů se zátěží funkčně škodlivých nesynonymních variant na souboru pacientek s nádorem prsu. 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. 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 terapeutických 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*, považovaném za jeden ze 70 prognostických biomarkerů u karcinomu prsu, byly bioinformatickým nástrojem RegulomeDB identifikovány funkční nekódující varianty (rs17448931 a rs3806362). Varianty byly následně úspěšně potvrzeny na rozsáhlých klinických datech ($n = 808$) pacientek studie. Celkově tak 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 správně předpověděna zvolenou *in silico* metodou.

Naše práce přinesly řadu prioritních výsledků využitelných v terciární a kvartérní prevenci karcinomu prsu, avšak rovněž demonstrovaly význam důkladného bioinformatického vyhodnocení a následné klinické validace jako nezbytných nástrojů.

1.1. Summary

Breast cancer, a genetically heterogeneous disease, remains the leading cause of death in women, and although mortality is slowly but gradually decreasing, incidence and prevalence is increasing in last years. A fully effective primary prevention to reduce the risk of developing the disease is still not available. An early diagnosis as part of a secondary prevention is the only effective risk minimisation measure. A tertiary prevention with the aim to avoid the disease progression, a loss of self-sufficiency associated with a lower quality of life, and a quaternary prevention identifying patients at risk of over-medication are becoming an integral part of the clinical management of oncologic patients. Tertiary and quaternary prevention are based on the study of genetic differences in patients and their subsequent stratification according to genetic predisposition, the molecular nature of the cancer, and the effects of chemotherapy. Due to genetic variations, a reduced response to treatment can be studied as well, which may be attributed to variability in metabolism or drug transport and subsequent drug exposure.

In 2003, data on approximately 92 % of the human genome had been generated via the world's largest biological project - the Human Genome Project. Twenty years later, we have found ourselves in the age of the bioinformatics, in which 100 % of the human genome has been mapped. The basic hypothesis of this work is that there are evolutionary, population and functionally distinct genetic variants in the human genome and these are publicly accessible in the global databases for further studies. This work aimed to investigate available databases and bioinformatics tools to facilitate processing of large amounts of pharmacogenomic and pharmacogenetic data and thus to systematize the prediction process of genetic profile of breast cancer patients with applicability in tertiary and potentially quaternary cancer prevention.

Through massive parallel sequencing, we have evaluated common germline coding and non-coding variants in a panel of 509 genes related to drug metabolism and elimination, cell death and major oncogenic pathways using emerging *in silico* tools and established pharmacogenomic databases for the first time in Czech breast cancer patients in such an extent. *In silico* 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 novel targets for follow-up functional studies with the potential to contribute to tertiary prevention. In patients with luminal B or triple-negative breast cancer tumors, the frequently studied variant rs1801160 in *DPYD* was significantly associated with disease-free survival and represents another provocative target with prognostic potential.

Through the study of rare coding variants in these patients, the purpose of this work was to establish an effective approach for the adequate prediction and interpretation of functional variants using bioinformatics tools that rely on sequencing and structural algorithms, as well as machine learning techniques. To address some of the limitations inherent to machine learning methodologies, our newly proposed strategy assessed a substantial number of genes harboring functionally deleterious non-synonymous variants in a cohort of breast cancer patients. Among the twelve prioritized genes, we identified significant associations for seven genes—*ABCC1*, *ABCC4*, *ABCB6*, *ATP7B*, *CYP2D6*, *CYP4F3*, and *CFTR*—with response to treatment, menopausal status, histological type, molecular subtype, and differentiation grade of breast cancer. Specifically, we demonstrated that patients with pathogenic variants in *CFTR* exhibited significantly reduced disease-free survival (Log Rank, $p = 0.002$) and overall survival (Log Rank, $p = 0.006$). Our findings further substantiate the involvement of *CFTR* in processes that potentially influence the efficacy of cancer therapy or the progression of malignant disease, corroborating recent publications.

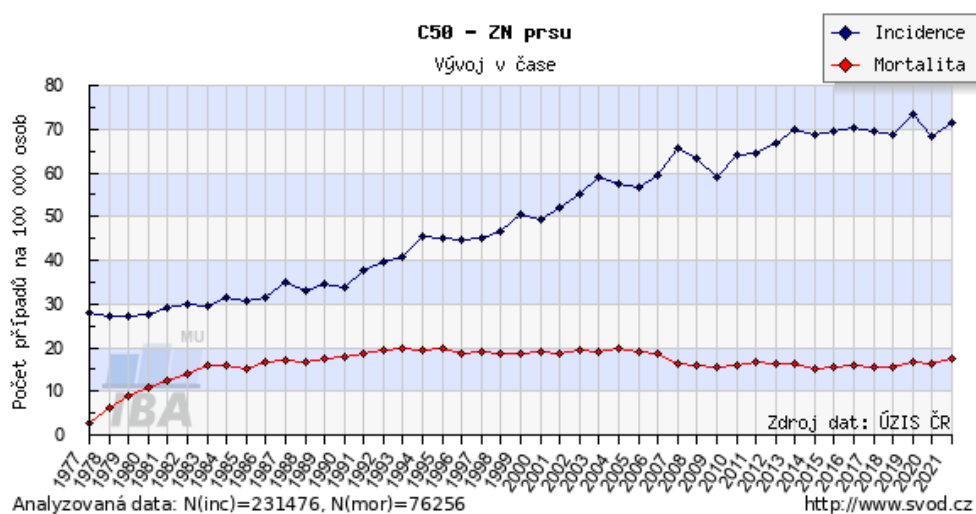
Further focusing on all variants in the *KIF14* gene, known as one of 70 breast cancer prognostic biomarkers, and using *in silico* analysis via RegulomeDB, we have identified functional non-coding variants (rs17448931 and rs3806362). These variants were successfully replicated in a large clinical data from breast cancer patients ($n = 808$) of the study. Overall, of the 23 germline variants with an allele frequency of more than 5%, only two (rs17448931 and rs3806362) were validated. These variants are located in the intronic region of the genome, and their regulatory function was accurately predicted through the chosen *in silico* method. Our results refine the targets within the *KIF14* gene, and potentially allows for use in the tertiary prevention.

Our work provides a range of high-priority outcomes that can be utilised for further research, specifically in the areas of tertiary and quaternary cancer prevention. In addition, we have also emphasised the importance of thorough bioinformatic evaluation and subsequent clinical validation as essential tools.

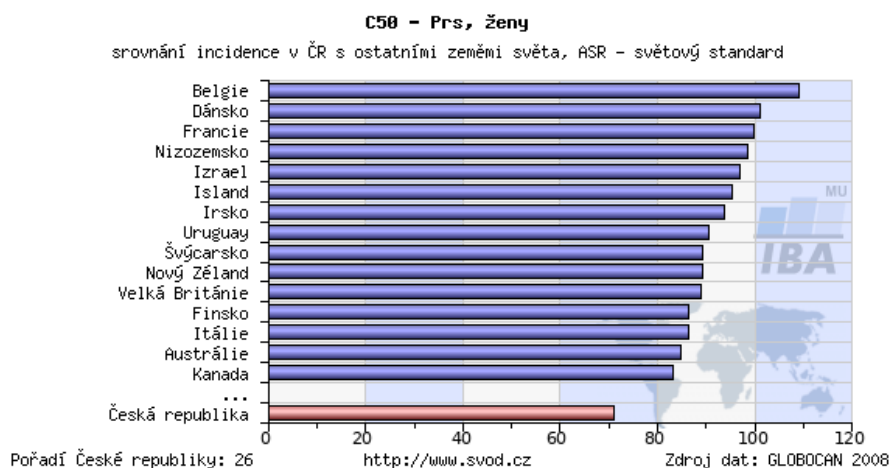
2. Literární přehled, současný stav poznání

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: hormonálně 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 proliferačního 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

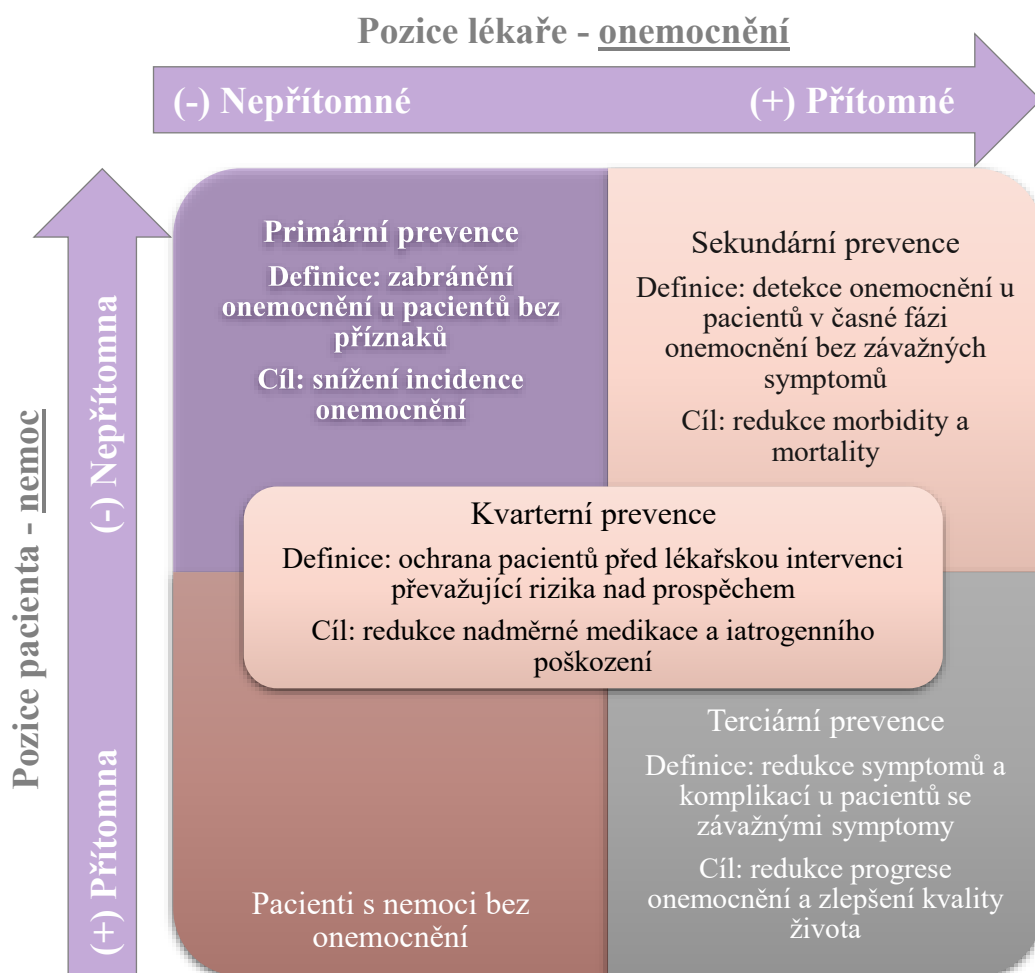
Etiologie karcinomu prsu je dána komplexní interakcí různých modifikovatelných i nemodifikovatelných faktorů, tj. genetickými, environmentálními, nutričními, hormonálními a dědičnými prvky. Mezi rizikové faktory patří předchozí výskyt karcinomu prsu, pozitivní rodinná anamnéza, časná menarche, pozdní 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. Mezi další vzácné, ale vysoce penetrantní geny patří *PTEN* a *TP53* (rovněž zahrnuté v našem panelu genů studií předkládané práce), dále pak *CDHI* a *STK11*, z nichž každý propůjčuje odlišný klinický fenotyp (Shiovitz a Korde 2015). 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 signifikantnímu zvýšení kvality a délky života díky včas podané terapii (Belizário et al. 2016). Seznam dostupných genetických testů rozšířen o další významné geny a mutace jak vrozených, tak somatických pro diagnózu a terapii karcinomu prsu poskytuje a průběžně aktualizuje americký úřad pro kontrolu potravin a léčiv (FDA) <https://www.fda.gov/medical-devices/in-vitro-diagnostics/nucleic-acid-based-tests#human>.

Nedílnou součástí klinického managementu pacientů se stává i terciární prevence, zvolením účinných léčiv s adekvátní odpovědí po celou dobu terapie. 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.

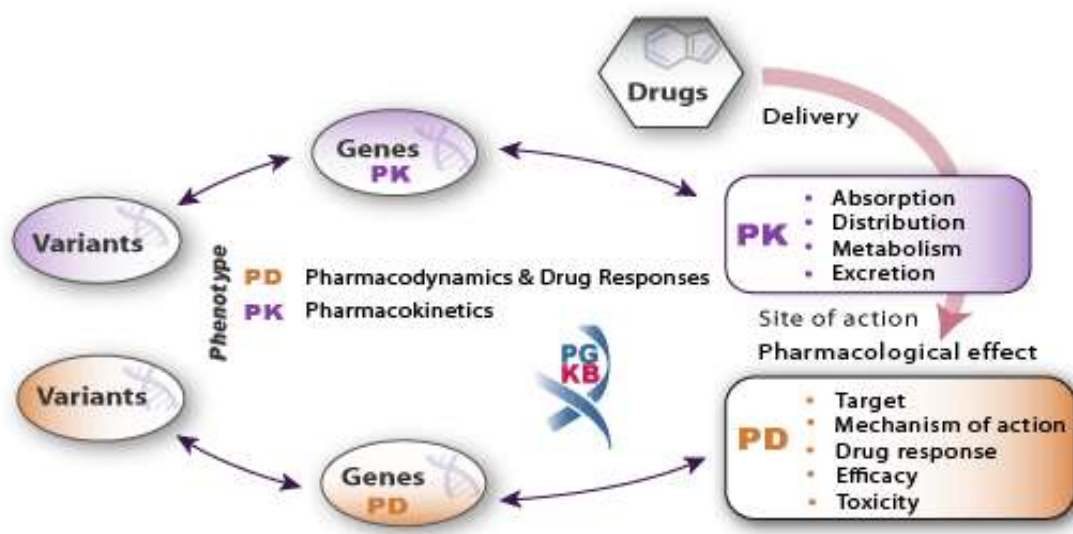


Obrázek 2—1 Nemoc ve vztahu k prevenci (převzato a upraveno (Martins et al. 2018))

Terciární a kvartérní prevenci lze docílit studiem mechanismu karcinogeneze a účinků chemoterapie, včetně těch nežádoucích. Dále pak stratifikací pacientů i podle genetické predispozice a molekulární povahy karcinomu s ohledem na genetické, epigenetické a biochemické změny (Uzilov et al. 2016). 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. Při rozdělení dle vědního oboru, farmakogenetika se zabývá studiem jednotlivých genů zapojených do farmakologických procesů (Kranzler et al. 2017). Studuje rozdílné účinky léčiva u rozdílných pacientů v reálné klinické praxi, které jsou dány přítomností dědičných variant genů, variabilitou pacientů.

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) a **Obrázek 2—3**, který popisuje situaci proléčiva a léčiva pro fenotyp pomalých („poor“), normálních („extensive“) a ultrarychlých metabolizátorů .



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

V neposlední řadě c.) alkylační činidla, cytostatika jako je pro léčivo cyklofosfamid, které je jaterním systémem cytochromu P450 metabolizováno na aktivní metabolity. Další léčiva s hyperstabilizací mikrotubulů tj., inhibitory polymerace mikrotubulů vinca alkaloidy (vincristin, vinblastin, vinorelbin) nebyly do naší studie zahrnuty. Farmakokinetické vlastnosti těchto léčiv jsou blíže popsány v jednotlivých souhrnech údajů o přípravku (SmPC), ze kterých byly převzaty i výše uvedené informace, dále pak v databázích DrugBank a PharmGKB. Detailní seznam genů studovaného panelu a jejich rozdělení ve vztahu k fázím metabolismu léčiv, hlavně metabolismu jaterního je uveden v příloze (Supplementary Table S2) publikované studie (Hlavac et al. 2018). Adjuvantní hormonální léčba u postmenopauzálních pacientek zahrnovala inhibitory aromatázy (anastrozol, letrozol) a selektivní modulátory estrogenových receptorů (SERM), jako je syntetický nesteroidní antiestrogen tamoxifen. U premenopauzálních pacientek byl v rámci adjuvantní léčby podáván agonista gonadotropin uvolňujícího hormonu (GnRH) goserelin. Celkově se naše studie omezují pouze na vyhodnocení cytotoxické léčby, a proto hormonální terapie není dále popisována blíže.

2.4. Precizní medicína

"Precizní medicína (obecně považovaná za obdobu personalizované medicíny nebo individualizované medicíny) je inovativní přístup, který využívá informace o genomu, životním prostředí a životním stylu jednotlivce k rozhodování o jeho léčbě. Cílem precizní medicíny je poskytnout přesnější přístup k prevenci, diagnostice a léčbě onemocnění" (Delpierre a Lefèvre 2023).

Precizní medicína byla za posledních 20 let nejúspěšnější v léčbě karcinomu, kde analýza genetických změn zárodečných anebo v nádorových (somatických) buňkách přinesla biomarkery, které se v klinické praxi již rutinně používají. Biomarkery jsou podle definice objektivní, kvantifikovatelné charakteristiky biologických procesů (Strimbu a Tavel 2010). Ačkoli většina farmakogenomických možností v onkologii zahrnuje získané somatické mutace, které často předpovídají farmakodynamické účinky onkologických léčiv, např. somatické mutace genu pro receptor epidermálního růstového faktoru (EGFR) (Wheeler et al. 2013; Ikediobi 2008), zárodečné variace, které ovlivňují vývoj karcinomu, mohou rovněž pomoci předvídat terapeutické a nežádoucí účinky protinádorových léčiv (Evans a McLeod 2003; Filipski et al. 2014; Hertz 2013).

Dobrým příkladem personalizované medicíny již implementované do klinické praxe je publikovaná analýza vycházející ze souhrnů údajů o přípravku (SmPC), které jsou mistry

výstupních dat studií léčiv při a po registraci léčivých přípravků. Od roku 1995 do srpna 2014, kdy bylo centralizovaně registrováno (v EU přes Evropskou Lékovou Agenturu) 517 léčivých přípravků, téměř 15 % z nich obsahovalo farmakogenomickou informaci v různých bodech SPC. V případě vrozené mutace, za zmínku stojí příklad kapecitabinu (perorálně podávané proléčivo 5-FU), kde je přímo kontraindikováno jeho použití u pacientů s genetickou změnou ovlivňující aktivitu dihydropyrimidindehydrogenázy (DPD). Ačkoli DPD deficit nemůže být přesně stanoven, je známo, že pacienti s určitými homozygotními nebo složenými heterozygotními mutacemi v místě *DPYD* genu, které mohou způsobit úplnou nebo téměř úplně chybějící DPD aktivitu enzymu, mají vyšší riziko život ohrožující toxicity (Meulendijks et al. 2015).

Realitou ale je, že chybějící informace týkající se stanovení přesné dávky, stěžují rutinní terapeutické využití genomických dat (Crona a Innocenti 2012). Tato omezení se snaží Evropská Léková Agentura (EMA), ve spojení s ostatními lékovými agenturami, kompenzovat podporou relevantních farmakogenomických dat, která jsou požadována u registrovaných léčivých přípravků (Chin et al. 2017) https://www.ema.europa.eu/en/documents/report/genomics-genetics-transcriptomics-epigenetics-subgroup-report_en.pdf. Příkladem je upozornění pro držitele rozhodnutí o registraci u přípravků obsahující tamoxifen <https://www.sukl.cz/tamoxifen-a-cytochrom-p450-cyp2d6?highlightWords=tamoxifen>. Pracovní skupina pro farmakovigilanci (PhVWP) Evropské lékové agentury ve spolupráci s pracovní skupinou pro farmakogenetiku (CHMP Pharmacogenomics Working Party, PGWP) a odborným poradním sborem pro onkologii (Scientific Advice Group on Oncology, SAG O) uvedla, že polymorfismus genů pro *CYP2D6* může být spojen s různou odpovědí na léčbu tamoxifenem, kdy pomalí metabolizátoři *CYP2D6* mají snížené plazmatické hladiny endoxifenu, jednoho z nejdůležitějších aktivních metabolitů tamoxifenu a mohou vykazovat sníženou odpověď na léčbu a doporučila přidání těchto informací do SmPC.

V případě somatických mutací u nádorů prsu, kdy je zvýšená exprese proteinu HER2 nebo jejichž nádory vykazují amplifikaci genu HER2, byly na podkladě farmakogenomických dat registrované léčivé přípravky jako Herceptin (od roku 2000), Tyverb (od roku 2008) a Perjeta (od roku 2013). V precizní medicíně se využívá AI a strojové učení k individualizaci léčby s ohledem na faktory, jako jsou charakteristiky onemocnění, genotyp pacienta, panely biomarkerů a klinické parametry k upřesnění výběru pacientů a dávkování léčiva do klinických studií dle předem připravené knihovny variant. V roce 2020, byla navíc provedena studie za využití strojového učení a dat fenotypové manifestace u DPD-deficientních pacientů z farmakovigilanci databáze EMA. Výsledky studie ukazují, že 7 až 13 % pacientů má profil podobný DPD-deficientním pacientům

ve vztahu k projevům toxicity na léčivo (fluorouracil, kapecitabin, tegafur, a flucytosin). Studie také poukázala na další využitelnost umělé inteligence v regulačním prostředí.

V červenci 2023, navíc EMA vydala dokument, aktuálně k veřejnému připomínkování, týkající se využití umělé inteligence (artificial intelligence, AI) v průběhu života léčivého přípravku a lze tedy očekávat další využití těchto metod i v regulaci léčivých přípravků. <https://www.ema.europa.eu/en/news/reflection-paper-use-artificial-intelligence-lifecycle-medicines>.

2.5. 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 vědního odboru – bioinformatiky, která na základě matematického modelování, statistiky a komputace predikuje funkční biologické konsekvence (Afgan et al. 2002). Vše započalo ojedinelým, celosvětově největším společným biologickým projektem – Human Genome Project, který v roce 2003 vygeneroval data pokrývající až 92 % lidského genomu a veřejně je zpřístupnil (Weng et al. 2013). Byly vyvinuty efektivní „high throughput“ metody, které snížily čas i finance potřebné například ke zjištění etiologie onemocnění (Iorio et al. 2016). Celkově technologie sekvenování nové generace (NGS) dosáhla stonásobného snížení nákladů ve srovnání se Sangerovým sekvenováním a její cena je aktuálně nižší než 20 000 Kč za jeden lidský genom. V současné době mohou platformy NGS produkovat miliardy čtení až pro 48 vzorků v jednom runu (Hu et al. 2021; Liu et al. 2022). Ze surových dat NGS lze pomocí analytického souboru nástrojů (pipeline) vytvořit standardní a dobře rozpoznatelné soubory ve formátu VCF (Variant Call Format), které uchovávají identifikované varianty a poskytují genomickou pozici variant, typ substituce nukleotidů, vyhodnocené skóre kvality, genotyp a další relevantní informace podle zarovnání sekvencí. Na základě uvedených informací lze varianty přiřadit ke konkrétním genům nebo transkriptům, klasifikovat je do různých typů a vyvozovat závěry o jejich důsledcích (Wang et al. 2010; Cingolani et al. 2012; McLaren et al. 2016).

Nejnáročnější částí je následující krok, tj. funkční anotace neboli, v našem případě, přiřazení škodlivosti varianty způsobující onemocnění či odpověď na léčivo (Araujo et al. 2018) z přibližně 50 000 variant celoxomového nebo dokonce milionů variant z celogenomového sekvenování (Liu et al. 2022). Existuje spousta datových zdrojů uchovávajících evidované informace o vztahu genotyp-fenotyp, kdy informace o volání variant jsou uchovávány v tzv. Locus Specific

DataBases (LSDBs). Jedna z nejznámějších je Leiden Open-source Variation Database (LOVD) a ClinVar (viz 2.5.1.3.).

Do jisté míry jsme schopni tyto záznamy využít k interpretaci vzniku, průběhu, diagnostiky a léčby onemocnění z genetického hlediska. Avšak i ty nejuznávanější databáze, jako je například ClinVar (Landrum et al. 2018), obsahují pouze 14 000 vysoce spolehlivých variant s důkazy vyhodnocenými genetickými experty, což je malý zlomek ve srovnání s obrovským počtem variant identifikovaných z NGS. Tato situace dramaticky snižuje klinickou využitelnost genetiky. Kromě toho také představuje velkou výzvu pro pochopení rozdílného působení genů mezi jednotlivci, populacemi a druhy a také pro dešifrování vztahu genotyp-fenotyp (Orgogozo et al. 2015). K řešení těchto problémů se objevily výpočetní nástroje tzv. prediktory (případně obecně, nazývané jako algoritmy) pro předpovídání vlivu neboli konsekvence variant. Prediktory mohou pomoci překlenout propast mezi obrovským množstvím generovaných genomických dat a omezenými známými genetickými důkazy, a nakonec sestavit potenciální vztah genotyp-fenotyp pro nově identifikované varianty (Liu et al. 2022) (viz 2.5.2.). Obecně tak aplikace *in silico* přístupů kombinujících matematické strategie s odborným názorem umožňuje výzkumníkům efektivně a ekonomicky analyzovat biologický význam genetických dat. Kategorizace založená na predikci poskytuje vhled do charakterizace variant a stanovení priorit.

2.5.1. Bioinformatické databáze

Od zavedení projektu lidského genomu v 90. letech 20. století byly akademickými institucemi, organizacemi, konsorciemi 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. Množství bioinformatických nástrojů a databází, které se generují za posledních 20 let, představuje přibližně 54 nových databází a až kolem stovky aktualizací již používaných databází každý rok (Galperin et al. 2017).

Databázové zdroje využívá většina prediktivních algoritmů (v celé práci je také označujeme jako prediktory neboli nástroje) a jejich základní přehled je uveden níže.

2.5.1.1. 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). Představují základ pro genetickou analýzu, protože zarovnání čistých čtení k referenčnímu genomu je nezbytným krokem při analýze NGS jak již bylo zmíněno výše. Proteinové sekvence se 100 %, 90 % a 50 % identitou k jednotlivým druhům následně vytvářejí klastry, které se ukládají do informativních databází, jako je UniProt Reference Clusters (UniRef) (Suzek et al. 2007), odnož Universal Protein Resource (UniProt) (UniProt, 2021). Tyto klastry se využívají k vytváření souborů vícenásobného zarovnání sekvencí (Multiple Sequence Alignment, MSA), které představují základ přístupu založeného na homologii sekvencí a také základní algoritmus některých z prediktorů jmenovaných v kapitolách níže.

2.5.1.2. 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.

Existuje několik celosvětových populačních projektů, včetně NCBI dbSNP (Smigielski et al. 2000), 1000 genome project (1KGP) (Sudmant et al. 2015), HapMap (Gibbs et al. 2003), UK10K (Walter et al. 2015), Genome Aggregation Database (gnomAD) (Karczewski et al. 2019) a NHLBI GO Exome Sequencing Project (ESP) (Fu et al. 2013). 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.

2.5.1.3. 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). Podle statistik OMIM je k dnešnímu dni na molekulární

úrovni možné diagnostikovat přibližně 6315 tzv. jedno-genových poruch způsobených zhruba 4441 geny (<https://omim.org/statistics/geneMap>). Orphanet ke dni (07.07.2023) zahrnuje 5835, 6172 nemocí a 45734 diagnostických testů <https://www.orpha.net/consor/cgi-bin/index.php>).

Pro popis nemocí ve struktuře dat vhodné pro algoritmy byly vyvinuty ontologické databáze, jako je Gene Ontology (GO) (Ashburner et al. 2000), Mammalian Phenotype Ontology (Smith et al. 2004) a Human Phenotype Ontology (HPO) (Robinson et al. 2008). Tyto databáze jsou také známé jako databáze specifických lokusů (LSDB). Data a záznamy jsou často kurátorsky zpracovávány z literatury a klinických studií. Rozsah LSDB se pohybuje od jednoho genu se zhruba 4000 variantami (Findlay et al. 2018) až po stovky milionů variant. Cílem LSDB je jednoznačně a přesně definovat a kategorizovat korelaci mezi genotypem a fenotypem, pochopit funkce a účinky genů, poskytnout mapu genetické distribuce napříč populacemi a nemocemi a pomoci klinikům/diagnostickým laboratořím při provádění dalších validačních testů tím, že poskytnou detailní informace o molekule, patogenitě a účincích variant. Mezi specifické LSDB zřízené pro jednotlivé fenotypy/nemoci, jako je například Fabryho nemoc, fabry-database.org (Saito et al. 2011) a *CFTR2* pro cystickou fibrózu jejíž dat jsme využili v průběhu studia vzácných variant pro dílčí cíl této předkládané práce (The Clinical and Functional TRanslation of *CFTR* (*CFTR2*); <http://CFTR2.org>).

Většina nástrojů založených na fenotypu vyžaduje jako vstup soubory variant (VCF) a termíny HPO, zatímco některé nástroje vyžadují pouze termíny HPO. Prioritizace genů u Mendelovských nemocí za pomoci srovnání 10 *in silico* metod, jak s HPO terminologií, tak VCF formátu jako vstupních dat byla nedávno publikována. Výsledky ukazují, že metody využívající jako vstupní údaje termíny HPO a soubory VCF dosahují lepšího celkového výkonu než metody využívající pouze fenotypová data (Yuan et al. 2022).

Hlavní faktor, který komplikuje adekvátní predikci je, že v různých LSDB 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.

2.5.1.4. 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). Všechny tyto konsorcia poskytují recenzní řízení (peer-review), aktualizované farmakogenetické pokyny pro chemoterapeutika, připravené na základě „evidence-based“ znalostí také z laboratorních podmínek farmakogenetického a farmakogenomického výzkumu. Nezávislá pracovní skupina pro hodnocení genomických aplikací v praxi a prevenci (Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Working Group (EWG)) z USA navíc vyvinula kritéria pro stanovení klinické validity a použitelnosti dat.

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).

PharmGKB navíc shromažďuje a dále spravuje preskripční údaje léků obsahující farmakogenetické informace schválené FDA, Evropskou agenturou pro léčivé přípravky (EMA), Švýcarskou agenturou pro léčivé přípravky (Swissmedic), Japonskou agenturou pro léčivé přípravky a zdravotnické prostředky (PMDA) a Kanadskou agenturou pro zdraví (Santé Canada, HCSC) k plnému využití pro personalizovanou medicínu. PharmGKB (pharmacogenetics and pharmacogenomics knowledge base, www.pharmgkb.org).

2.5.1.5. Databáze somatických mutací nádorů

K podobnému účelu systematického kurátorství byla rovněž vytvořena i řada znalostních databází somatických variant. V předkládané práci využíváme těchto zdrojů pouze okrajově. Stejně jako ClinVar i níže zmíněné znalostní databáze mají stanovenou úroveň použitelnosti a obohacují komunitu přispíváním variant s odbornou kurátorskou podporou.

Tyto zdroje jsou také důležité pro pochopení vývoje bioinformatických nástrojů využívající somatické mutace jako podklad trénování a validování.

Mezi nejznámější znalostní databáze zahrnujeme OncoKB (<https://www.oncokb.org/>) s 743 geny a 126 léčivy (ke dni 07.07.2023) a CIViC (civic.genome.wustl.edu) s 496 geny a 183 varianty při

filtraci dat pro nádor prsu (ke dni 07.07.2023). Mezi další patří MyCancerGenome (<https://www.mycancergenome.org/>), CanDL (<https://candl.osu.edu/>). Příkladem nového bioinformatického prediktoru ke klinické interpretaci somatických variant nádorů založenou na důkazech je CancerVar (<https://cancervar.wglab.org/about.php>). Nástroj využívá umělou inteligenci ke zhodnocení a poskytuje kategorizaci predikce v souladu s pokyny AMP/ASCO/CAP/CGC pro období 2017-2019.

V rámci postgraduálního studia a předkládané práce byly využity i další veřejné databáze a nástroje např. Catalogue Of Somatic Mutations In Cancer (<https://cancer.sanger.ac.uk/cosmic>), cBioportal (<https://www.cbioportal.org/>) využívající data z velkých projektů ICGC (International Cancer Genome Consortium, <https://dcc.icgc.org/>) a TCGA (The Cancer Genome Atlas Program, <https://www.cancer.gov/ccg/research/genome-sequencing/tcga>) a další, které již implementovaly algoritmy bioinformatické analýzy pro funkční hodnocení genetického profilu (Jalali Sefid Dashti a Gamielien 2017). Dále projekty jako např., CloudBioLinux, CloudMan, Galaxy a Gemstone, které navíc poskytují virtuální výzkumní prostředí (virtual research environment) umožňující zpracování objemných dat a řadí se tak mezi další novodobé atributy na poli bioinformatiky (Afgan et al. 2002). Pro tvorbu funkčních genových interakcí jsou k dispozici webové prohlížeče, aplikace či pluginy jako např. Cytoscape, GeneMANIA a BioGRID) (Stark et al. 2006). Tato prostředí implementují populární nástroje analýzy a redukuje zdroje výpočetní techniky potřebné pro ukládání, analýzu či personální obsazení laboratoria bioinformatikem.

2.5.2. Bioinformatické nástroje

In silico predikční algoritmy představují užitečný nástroj pro předběžné zhodnocení škodlivosti vrozených variant, například dopadu aminokyselinových záměn na funkci proteinu, a především pro výběr kandidátních variant, pro funkční analýzu rozsáhlých dat metodou sekvenování. Volba těchto nástrojů je kritická pro spolehlivou interpretaci dopadu variant a měla by být založena na systematickém hodnocení jejich výkonu, a to jak z pohledu specifity, tak senzitivity (viz 2.5.3). Základním, ale velice zjednodušeným biologickým principem je, že jediná záměna aminokyseliny v proteinu může ovlivnit jeho strukturu a funkci. Analýza mutací v proteinu je klíčová pro zjištění jejich dopadu a pochopení mechanismu onemocnění. Bioinformatické nástroje poskytují pro danou mutaci výsledky založené na skóre, které udává pravděpodobnost poškození/škodlivosti/patogenity. V posledních letech, hlavně zavedením umělé inteligence,

vývoj stávajících programů a zavedení nových predikčních algoritmů změnil současný stav analýzy proteinových mutací (Choudhury et al. 2022).

Publikované studie však dokládají značné rozdíly v interpretaci predikce při použití několika nástrojů, pravděpodobně důsledkem různé váhy hodnocených parametrů u jednotlivých predikčních algoritmů, kvůli obsahu a velikosti testovacích datových souborů ale taky z důvodu neadekvátního zvolení algoritmu k zamýšlenému cíli studie (Ghosh et al. 2017; Niroula a Vihinen 2019).

Vedlejší cíl této práce zahrnuje i zhodnocení rozdílností ve výsledku predikce u zvolených nástrojů a metod pro funkční interpretaci genetické variace, v našem případě vzácných variant u vybraného panelu 509 farmakogenů a onkologicky významných genů.

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) a tyto prediktory jsou blíže popsány.

Základní výběr nástrojů lze také rozdělit na část genu, kterého se predikce týká, zjednodušeně na kódující a nekódující. Pro kódující část (exon) je dále dělíme dle typu varianty (mutace). Varianty způsobující sekvenční změny se dělí především na čtyři typy: inserce, delece, jednonukleotidová varianta (SNV) a ostatní substituce (Eilbeck et al. 2005). SNV se skládají z nesynonymních SNV (nsSNV) a synonymních SNV (sSNV). Mezi nimi jsou nejčastěji identifikovány nsSNV, stejně jako v naší studii. Proto se také v předkládané práci zaměřujeme hlavně na výpočetní přístupy, které jsou vyvinuty k odvození dopadu nsSNVs v kódujících oblastech. Minoritně bylo využito i nástrojů hodnotící nekódující část genů (dbscSNV, TFBP, MaxEntScan, IW scoring framework, RegulomeDB, TargetScan a PINES).

2.5.2.1. 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). Metodologicky, prediktory nejprve konstruují mnohočetné přiřazení neboli zarovnání sekvencí (MSA) buď seskupením více proteinových

sekvencí s danou podobností ze zarovnáním BLAST (Altschul et al. 1990) nebo jen načtením vlastních selektivních sekvencí z výše jmenovaných genomických databází pro mnohonásobné zarovnání pomocí MULTIZ (Blanchette et al. 2004), nebo MUSCLE (Edgar 2004). Na základě MSA je vypočtena pozičně specifická skórovací matice (PSSM) (Gribskov et al. 1987), která generuje výsledek predikce pomocí pravděpodobnostního skóre (Ng a Henikoff 2003), poměru pravděpodobnosti (Chun a Fay 2009), průměrné vzdálenosti mezi cílovými druhy a ostatními v podčeledi (Choi et al. 2012) nebo rozdílu entropie (Reva et al. 2007; Hopf et al. 2017). Prediktivní výsledky jsou obvykle spojité hodnoty s mezní, také nazývanou prahovou hodnotou, doporučenou dle vývojáře metody a ověřenou na souboru mutačních dat.

Příkladem jsou varianty stop-gain a frameshift (posunové) mutace v anotačních nástrojích, např. ve VEP (McLaren et al. 2016) a SnpEff (Cingolani et al. 2012) obecně hodnoceny jako "HIGH" tzn. s vysokým dopadem na biologickou sekvenci. Vliv na protein však není vždy jednoznačný a zdá se, že změny aminokyselin zejména ty, které se nacházejí v blízkosti C-konce proteinu, jsou tolerantní (MacArthur et al. 2012) a některé varianty frameshiftů, dokonce i v homozygotním stavu, jsou často pozorovány mezi populacemi, což naznačuje možný snížený dopad na jejich patogenní charakter (Eilbeck et al. 2017).

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. Striktní shoda ve výsledcích predikce znamenala vyloučení všech variant pro některou z chybějících predikcí nástroje v "sadě sekvenčních nástrojů *in silico*". Sada zahrnuje nástroje Align-GVGD, LRT, MutationAssessor, PolyPhen2, PROVEAN a SIFT podrobněji popsane níže. Metody sekvenční sady nejsou založené na metodách strojového učení, a překonávají tak obavy z kruhovosti typu 2 v důsledku nedostatečného rozlišení škodlivých variant od neutrálních v rámci daného proteinu v trénovacím souboru dat (viz 2.5.3). Tento postup byl inovativní oproti již publikovaným studiím (viz 5.1.1).

Align-GVGD

Align-GVGD akronym pro Grantham Variation, Grantham Deviation, patřící pod Huntsman Cancer Institute v Utahu, pracuje se sekvenční homologií a záměnu ve vysoce konzervovaných místech hodnotí dle fyzikálně-chemických vlastností aminokyselin tj., polarita, náboj atd.

(Tavtigian et al. 2008). Pravděpodobnost, že substituce je škodlivá vyjadřuje škála od „C0“ do „C65“, kde třída „C65“ označuje největší pravděpodobnost. Nástroj je dostupný z <http://agvgd.iarc.fr>, kde má uživatel možnost srovnání vložené sekvence (ve FASTA formátu) s některým z důležitých predispozičních genů (např. BRCA1). Align-GVGD je rozšíření původního Granthamova rozdílu ve vícenásobném zarovnání sekvencí a skutečného simultánního vícenásobného porovnání. Ve verzi 2012 (říjen) programu ANNOVAR, využívaného v našich studiích, je k dispozici argument --aamatrixfile, k vyvolání skóre GRANTHAM, nebo jakoukoli jinou matici substituce aminokyselin pro nesynonymní varianty v anotaci založené na genu. Např. do výstupu je přidán zápis AAMatrix=43, který označuje, že změna R->Q má GRANTHAM skóre 43 a vyjadřuje tak škodlivost záměny (Garcia et al. 2022).

LRT

LRT (Test (Test) Poměru (Ratio) Pravděpodobnosti (Likelihood)) se koncepčně odlišuje od ostatních srovnávacích genomických metod. Obecně LRT představuje statistický test shody dvou modelů. Relativně složitější model se porovnává s jednodušším modelem, aby se zjistilo, zda se na určitý soubor dat hodí výrazně lépe. Všechny předchozí metody určené k identifikaci škodlivých mutací se spoléhají na heuristické postupy k rozlišení míst v proteinu, která jsou konzervovaná, od těch, která konzervovaná nejsou. Toho se dosáhne výběrem sekvencí, které nejsou příliš blízce nebo příliš vzdáleně příbuzné se sekvencí, která je předmětem zájmu, a porovnáním stupně zachování v místě zájmu s jinými místy v proteinu. Výhodou tohoto přístupu je, že se nevyžaduje fylogenetický vztah a evoluční vzdálenost mezi sekvencemi. Tento přístup má však i svá omezení, protože nerozlišuje mezi vzdáleně příbuznými proteiny, které jsou vysoce konzervované, a blízce příbuznými proteiny, které se rychle vyvíjely. Ačkoli lze vždy vybrat proteiny, které vykazují požadovaný stupeň podobnosti, je pravděpodobnější, že u vzdáleně příbuzných proteinů došlo ke kompenzačním změnám (Ng a Henikoff 2006). Oproti tomu LRT byl navržen tak, aby explicitně modeloval fylogenetické vztahy pomocí pravděpodobnostního rámce a minimalizoval kompenzační změny, které jsou považovány za běžné a mohou způsobit falešně negativní výsledky, a to použitím blízce příbuzných druhů obratlovců: savců, kuřat, žab a ryb. LRT se také odlišuje od jiných srovnávacích genomických metod tím, že všechny změny aminokyselin jsou považovány za stejné, místo aby se radikální a konzervativní změny aminokyselin vážily rozdílně. Ačkoli se očekává, snížení síly predikce LRT, empiricky je míra falešně pozitivních i falešně negativních výsledků LRT nižší u radikálních změn oproti

konzervativním změnám aminokyselin. To může být důsledkem toho, že genetický kód a další mutační parametry korelují s poměrem radikálních a konzervativních změn aminokyselin (Chun a Fay 2009). Když vývojáři LRT porovnávali SIFT a PolyPhen s jejich testem poměru pravděpodobnosti (LRT) při hodnocení 3 lidských genomů zjistili, že 76 % variant bylo předpovězeno jako škodlivé pouze jedním nástrojem, a jenom 5 % variant bylo označeno za škodlivé všemi třemi nástroji. V případě striktního konsensu pro interpretaci predikce za použití těchto nástrojů, tak dochází k vysoké eliminaci předpovězených variant považovaných za škodlivé.

MutationAssessor

Vývoj této metody byl založen na fenomenologické analýze, která získává informace ze zarovnání proteinových rodin velkého počtu homologických sekvencí seskupených do zarovnaných souborů rodin a pod-rodin a využívá 3D struktury sekvenčních homologů. Evoluční analýza v tomto případě představuje mocný nástroj, kdy přirozený výběr určité sekvenční varianty odráží souhrnný účinek molekulárních změn na fyziologii buněk, tkání a orgánů k předpovědi funkčního dopadu mutací obecně a zejména u nádorových onemocnění např. dodatečným využitím dat z databáze COSMIC (Reva et al. 2011). Nedostatek dat (3D struktury a komplexy) a přesných přístupů pro de novo modelování struktury a funkce proteinů na molekulární úrovni jsou známou limitací, která je zde kompenzována již zmíněných srovnání sekvencí (Liu et al. 2022).

PROVEAN

Analyzátor vlivu variací na proteiny (PROVEAN) vypočítává funkční důsledky záměny jedné aminokyseliny na protein (Choi et al. 2012). PROVEAN kategorizuje mutace jako škodlivé nebo neutrální. Mutace ze skóre PROVEAN $< -2,5$ je škodlivé, zatímco mutace se skóre $> -2,5$ jsou považovány za neutrální. Webový server PROVEAN se skládá ze tří nástrojů: PROVEAN Protein (zahrnuje všechny druhy), PROVEAN Protein Batch a PROVEAN Genome Variants (konkrétně pro myš a člověka). Nástroj PROVEAN Protein Batch vrací také výsledek nástroje SIFT a může zpracovávat velké množství proteinových variant. Vstup pro tento program přebírá záměnu aminokyselin a podporuje identifikátory proteinů z veřejné domény z NCBI RefSeq, UniProt a Ensembl.

SIFT

SIFT akronym pro Sorting Intolerant from Tolerant, nástroj dostupný na <http://sift.jcvi.org/> pracuje na základě sekvenční homologie a podobnosti fyzikálně-chemických vlastností aminokyselin. V případech, kde je aminokyselina v určité pozici vysoce konzervovaná, je její záměna vyhodnocena jako intolerantní, naopak při nižším stupni je záměna spíše tolerována (Kumar et al. 2009). Výsledkem je skóre, které určuje pravděpodobnost, že je změna je tolerována, tudíž výsledky blíže k nule představují negativní vliv dané záměny. Varianty s hodnotami $\geq 0,05$ jsou označovány za patogenní, ostatní za tolerované.

2.5.2.2. 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.

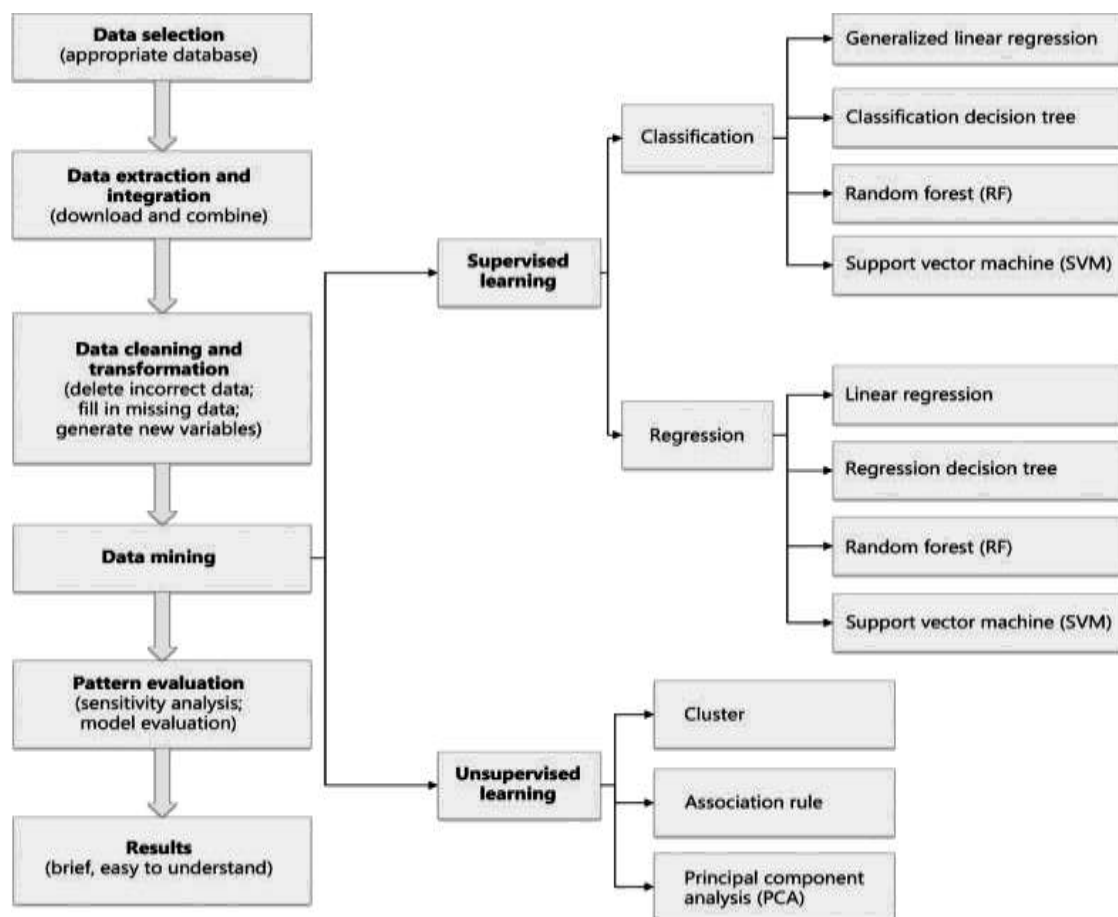
Obecně nesynonymní varianty ovlivňují expresi proteinu (Haraksingh a Snyder 2013), posttranslační modifikace (Kim et al. 2015) nebo vazebnou afinitu včetně farmakogenomických cílů (Morningstar-Kywi et al. 2021) a lékové rezistence (Pires et al. 2017). Odhadem přibližně 75 % variant způsobujících onemocnění vede přímo k destabilizaci proteinu, čímž se stabilita proteinu stává hlavním faktorem přispívajícím k patologii onemocnění, zatímco asi 7 % variant má také funkční úlohu (Yue et al. 2005; 2014). Ve srovnání s polymorfismy ovlivňují mutace způsobující onemocnění převážně centrální část proteinu, zatímco ~70 % se nachází ve strukturních a funkčně nezbytných oblastech (Sunyaev et al. 1999; Beer et al. 2013). Protein-proteinová rozhraní jsou horkými místy tzv. hot spots pro nsSNVs způsobující onemocnění (David et al. 2012; Petukh et al. 2015). Vliv nsSNV na stabilitu proteinu se obvykle odhaduje výpočtem změny Gibbsovy volné energie ($\Delta\Delta G_f$), která je výsledkem záměny aminokyseliny. Fyzikální energetická funkce, statistická potenciální funkce a empiricky definovaná potenciální funkce jsou tři typy metodik výpočtu energie. Vzhledem k tomu, že první funkce je výpočetně náročná, jsou častěji využívány poslední dvě. Predikce stability proteinů založené na struktuře se přisuzují především empirickým potenciálům, které integrují s fyzikální a statistickou složkou energie související se strukturou (Guerois et al. 2002). Příkladem nástroje využívající techniky ML je

MAESTRO (Laimer et al. 2015) a PoPMuSiC-2.0 k předpovědi vlivu mutace/variace na stabilitu proteinu (Dehouck et al. 2009).

2.5.2.3. 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).

Do konečného modelu jsou někdy zahrnuty i stovky rysů, což dále komplikuje srovnání napříč prediktory. Z algoritmického hlediska se v predikcích s více rysy běžně používají přístupy ML pod dohledem, včetně SVM (Calabrese et al. 2009; Li et al. 2009), naivního Bayesova klasifikátoru (Adzhubei et al. 2010), neuronové sítě (NN) (Hecht et al. 2015), Random Forest (Carter et al. 2013; Niroula a Vihinen 2017) a regrese „boosted tree regression“ například pro nástroj ENTPIRE (Zhou et al. 2016). Výhody a nevýhody prediktorů se odvíjejí od zvolených modelů a následné kvality primárních dat použitých k trénování a validaci modelů. Obrázek 2—4 znázorňuje řadu výzkumných koncepcí. „Proces získávání dat je rozdělen do několika kroků: (1) výběr databáze podle účelu výzkumu; (2) extrakce a integrace dat, včetně stažení požadovaných dat a kombinace dat z více zdrojů; (3) čištění a transformace dat, včetně odstranění nesprávných údajů, doplnění chybějících údajů, vytvoření nových proměnných, konverze formátu dat a zajištění konzistence dat; (4) dolování dat, zahrnující extrakci implicitních vztahových modelů prostřednictvím tradiční statistiky nebo ML; (5) vyhodnocení vzorů, které se zaměřuje na parametry platnosti a hodnoty vztahových modelů extrahovaných dat; a (6) vyhodnocení výsledků, zahrnující převod extrahovaného modelu vztahů mezi daty a daty do srozumitelných znalostí zpřístupněných veřejnosti“ (Wu et al. 2021). Učení bez dozoru, známé také jako strojové učení bez dozoru, využívá algoritmy strojového učení k analýze a shlukování souborů neoznačených dat. Tyto algoritmy odhalují skryté vzory nebo seskupení dat bez nutnosti lidského zásahu. Díky své schopnosti odhalovat podobnosti a rozdíly v informacích jsou ideálním řešením pro průzkumnou analýzu objemných dat.



Obrázek 2—4 Přehled základního rozdělení modelů strojového učení (Wu et al. 2021)

Metaprediktory

Metaprediktory jsou nástroje, které provádějí predikci integrací výsledků již existujících prediktorů. Základní myšlenkou metaprediktorů je využít potenciální komplementární výkonnosti vybraných prediktorů při klasifikaci variant. Každý prediktor má svou vlastní metriku a stupnici, což ztěžuje porovnávání více prediktorů a brání jejich současnému použití (Tian et al. 2019). Pokud jde o metaprediktory, existují v porovnání s výše uvedenými sekvenčními a strukturálními algoritmy dvě vylepšení. Za prvé, poskytují komplexní hodnocení vybraných již existujících nástrojů (Liu et al. 2022). Za druhé, mají vlastní způsob interpretace skóre z vybraných nástrojů, a to transformací na srovnatelnou škálu jako je normalizované skóre (Bendl et al. 2016) nebo binární hodnoty (González-Pérez a López-Bigas 2011). Kromě toho jsou metaprediktory schopny celkově zlepšit predikční výkon což umožňuje vyhnout se zkreslení a generalizaci jednotlivými prediktory (Kircher et al. 2014). Pokud jde o chybějící hodnoty predikce, kde částečně již některé existující nástroje selhávají, metaprediktory pomocí imputace poskytují škodlivé/neutrální mezní hodnoty jako např. u nástroje Meta-SNP (Capriotti et al. 2013), průměrného skóre v CADD

(Kircher et al. 2014) a DANN (Quang et al. 2015), fixního skóre (Quinodoz et al. 2022), maximálního patogenního skóre v prediktoru M-CAP (Jagadeesh et al. 2016) nebo flexibilní imputace pomocí průměrné hodnoty k-nejbližších sousedů (Ioannidis et al. 2016) v nástroji REVEL. V případě nástroje M-CAP, ale naše preliminární studie vzácných variant odhalila významné množství chybějících hodnot predikce pro nsSNVs oproti jiným nástrojům. M-CAP nebyl proto použitý do studie proveditelnosti.

Dalším z nástrojů, který integruje skóre několika prediktorů najednou je např. Imhotep (Knecht et al. 2017). Zde bylo sjednoceno devět populárních predikčních nástrojů (PolyPhen2, SNPs&GO, MutPred, SIFT, MutationTaster2, MutationAssessor a FATHMM, jakož i Granthamovo skóre a PhyloP) do jediného prediktoru. Optimální kombinace těchto nástrojů byla vybrána pomocí široké škály statistických modelovacích technik, přičemž se vycházelo z 10029 jednonukleotidových variant (SNV) z Human Gene Mutation Database ve vztahu k nemocím a 10002 domněle "benigních" nesynonymních SNV z UCSC (The University of California Santa Cruz, Genome Browser database). Bylo zjištěno, že prediktivní výkonnost se výrazně zlepšila integrací modelů, zatímco maximální prediktivní schopnosti bylo dosaženo pomocí algoritmu Random Forest, rozhodovacího stromu (Decision Tree) nebo logistické regresní analýzy. Dále bylo zjištěno, že kombinace nástrojů PolyPhen-2, SNPs&GO, MutPred, MutationTaster2 a FATHMM má stejnou výkonnost jako všechny nástroje dohromady. Další zajímavé zjištění důležité pro zvolení adekvátních prediktorů v naší studii, učinili autoři tohoto nástroje při testování výkonnosti jednotlivých integrovaných nástrojů, a to konkrétně zmíněného Granthamovo skóre, využitého i pro účely predikce vzácných farmakogenomických variant jiné studie (Kozyra et al. 2017). Granthamovo skóre prokázalo horší výkonnost než ostatní nástroje. Toto skóre bylo totiž vyvinuto téměř před 50 lety, kdy původně nebylo určeno ke klasifikaci patogenních variant, ale ke zkoumání evoluce proteinů (Grantham 1974). Nicméně, vzhledem k tomu, že Granthamovo skóre kvantifikuje biochemický rozdíl mezi dvěma aminokyselinami z hlediska jejich složení, polarity a molekulárního objemu, lze předpokládat, že skóre odráží i funkční dopad nsSNVs. Všechny ostatní nástroje uvažované i v naší studii pro predikci navíc využívají další informace, jako je sekundární struktura nebo evoluční zachování, přičemž hlavně ty druhé jmenované jsou prediktivní (Ng a Henikoff 2003). Zdá se pravděpodobné, že nedostatek těchto informací je hlavním důvodem slabé prediktivní výkonnosti Granthamova skóre. Navíc autoři Granthamova skóre použili mezní hodnotu pro binární klasifikaci, která byl doporučen již před několika desetiletími (Li et al. 1984). Granthamovo skóre je ve výzkumu nadále aplikováno i přes jeho

poměrnou jednoduchost. Recentní studie z roku 2022 (González-del Pozo et al. 2022) využila tohoto nástroje v kombinaci s ostatními 13 často užívanými nástroji za současné optimalizace prahové hodnoty k určení predikce patogenity. Tréninková kohorta panelového sekvenování obsahovala kurátorská data od 209 geneticky diagnostikovaných jedinců s IRD (inherited retinal dystrophies). Optimální kombinace nástrojů, které byly dříve ověřeny u dalších 50 jedinců s IRD, byly testovány také u pacientů s dědičným nádorovým onemocněním ($n = 109$) a s neurologickým onemocněním ($n = 47$). Statistická analýza ukázala, že optimální kombinace filtrování zahrnovala nástroje CADDv1.6, MAPP, Grantham a SIFT a umožnila identifikovat jednu homozygotní variantu v kandidátním genu CFAP20 (c.337 C > T; p.Arg113Trp) konzervovaném ciliárním genu, který byl abundantně exprimován v lidské sítnici a nacházel se ve vrstvě fotoreceptorů (González-del Pozo et al. 2022).

Dalším 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 studii. 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).

CADD

CADD akronym pro Combined Annotation Dependent Depletion, je schopen vyhodnotit riziko substituce i v nekódujících oblastech, jelikož netolerantní záměny nejsou omezeny pouze na exomové sekvence genů (Kircher et al. 2014). Přestože existuje mnoho nástrojů pro anotaci a skórování variant, většina anotací využívá jeden typ informace (např. zachování) a/nebo má omezený rozsah (např. na nesynonymní změny). CADD je rámec, který integruje více anotací do jedné metriky tím, že porovnává varianty, které prošli přirozeným výběrem se simulovanými mutacemi. Skóre CADD je založeno na různých genomických vlastnostech odvozených z kontextu okolních sekvencí, anotací genových modelů, evolučních omezení, epigenetických měření a funkčních předpovědí. Pro každou variantu jsou všechny tyto anotace integrovány do jediného skóre CADD pomocí modelu strojového učení s dohledem a algoritmem logistické regrese. Pro lepší interpretaci jsou transformovány do pořadí tzv. PHRED skóre (tj. logaritmická

škála) pro přibližně všech 9 miliard potenciálních SNV celého genomu, tedy souboru všech tří nereferenčních alel na každé pozici referenční sestavy (Rentzsch et al. 2019). Skóre „CADD_phred“ bylo využito k predikci častých a vzácných variant studií předkládané práce. Na rozdíl jiných prediktorů CADD nebyl trénován na relativně omezeném počtu genomových variant, u nichž je "znám" patogenní nebo benigní stav. CADD byl postaven na větší trénovací sadě variant, které vznikly a přetrvaly napříč vývoje od posledního předka člověka a opice, a protože přetrvaly miliony let navzdory purifikační selekce, lze je většinou považovat za benigní nebo neutrální. CADD prediktor byl integrován mezi predikce strojového učení naší studie pro vzácné varianty (Kováčová et al. 2024).

REVEL

REVEL je ansámblová metoda vyvinuta pro hodnocení vzácných nesynonymních variant. Patogenita je vyhodnocena na základě kombinace skóre z 13 jednotlivých nástrojů: MutPred, FATHMM v2.3, VEST 3.0, PolyPhen-2, SIFT, PROVEAN, MutationAssessor, MutationTaster, LRT, GERP++, SiPhy, phyloP a phastCons. REVEL byl vycvičen pomocí nedávno objevených patogenních a vzácných neutrálních nesynonymních variant, s výjimkou těch, které byly dříve použity k tréninku jeho složek. Skóre REVEL pro jednotlivé nesynonymní varianty se může pohybovat v rozmezí od 0 do 1, přičemž vyšší skóre odráží větší pravděpodobnost, že varianta způsobuje onemocnění. REVEL poskytuje předem vypočtené skóre REVEL pro všechny možné lidské nesynonymní varianty. Autoři demonstrovali, že REVEL má ve srovnání se stávajícími metodami trvale nejlepší celkový výkon, zejména při rozlišování mutací spojených s onemocnění od neobvyklých neutrálních nesynonymních variant s AF pod 3 % (Ioannidis et al. 2016). Skóre nástroje REVEL bylo aplikováno k predikci vzácných variant studie předkládané práce.

VEST

Nástroj VEST akronym pro Variant Effect Scoring Tool je metoda strojového učení, která předpovídá funkční význam nesynonymních mutací na základě pravděpodobnosti, že jsou škodlivé (Carter et al. 2013; Douville et al. 2016). Metoda vyžaduje značné znalosti bioinformatiky, systému Linux a přístupu k linuxovému serveru s velkou pamětí. Nejaktuálnější verzí je VEST 4.0 (ke dni 2.7.2023). VEST je poměrně nová metoda pro stanovení priorit nsSNVs mutací, které mění aktivitu proteinů. VEST používá algoritmus strojového učení pod dohledem (Random Forest) k identifikaci pravděpodobných funkčních nsSNVs mutací. Tréninkovou sadou je pozitivní třída nsSNVs z databáze mutací lidských genů a negativní třída častých nsSNVs

detekovaných v databázi v populaci projektu sekvenování exomu, Exome Sequencing Project (ESP). Frekvence alel 1 % nebo vyšší byla považována za častou (běžnou) pro vývoj této metody. Všechny mutace jsou popsány souborem 86 kvantitativních znaků. Ačkoli je tréninková sada VEST určena k identifikaci nsSNVs, které mění aktivitu proteinu, jeho konstrukce zahrnuje dodatečný kontext selektivního rozdílu mezi nemocemi a častými variantami. Nástroj VEST byl vytvořen za účelem identifikace kauzálních genů, kdy se zkombinovali p-hodnoty pro všechny mutace podle genů a napříč všemi geny exomů všech onemocnění, k vytvoření statistiky na úrovni genů. Dle vývojářů by mělo být možné identifikovat tzv. kauzální geny na základě obohacení o funkční geny, mutací napříč exony onemocnění, aniž by bylo nutné odfiltrovat běžné varianty. Tento přístup k prioritizaci genů se liší od současných statistických metod pro testování genové zátěže v tom, že nevyžaduje kontrolní populaci, ačkoli pokud je k dispozici je užitečná pro eliminaci genů, které jsou významné jako artefakty výsledku sekvenování (Luo et al. 2011; Zeggini 2011). Skóre nástroje VEST3 bylo využito k predikci vzácných variant studie předkládané práce.

2.5.2.4. 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). Sekvenování celého genomu (WGS) výrazně zvýšilo schopnost odhalit genetickou variabilitu i v nekódujících oblastech, která se často podílí na regulaci genů. Aby bylo možné tyto nekódující varianty systematicky studovat a posoudit jejich funkční/patogenní potenciál, je nutná důkladná anotace a určení regulačních oblastí. Některé příklady anotace nekódujících oblastí zahrnují promotory, enhancery, dlouhé nekódující RNA, místa startu transkripce, vazebná místa transkripčních faktorů, regulační sekvence, vlastnosti přístupnosti chromatinu a vzorce modifikace histonů a místa citlivá k hydrolytickému štěpení enzymem deoxyribonukleázou (DNáza) I. Pro lepší pochopení architektury nekódujícího genomu je k dispozici několik konsorcií obdobně jako u kódujících variant popsaných 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. Vliv variant na předpokládané motivy vazebných míst transkripčních faktorů (TFBS) lze odhadnout

pomocí pozičně váhových matic (PWM) z dostupných databází, jako jsou například TRANSFAC (Wingender et al. 1996) a JASPAR (Mathelier et al. 2014). Dále, projekt dlouhých nekódujících RNA (lncRNA) FANTOM, 6. edice (Functional ANnotation Of the Mammalian genome (Abugessaisa et al. 2020) a 4DGenome databáze chromatinových interakcí (Schaid et al. 2018).

Význam nekódujících mutací zůstává celkově náročnou oblastí genomiky nádorových onemocnění. Data WGS jsou stále dostupnější, zejména prostřednictvím TCGA (The Cancer Genome Atlas) a ICGC (International Cancer Genome Consortium) konsorcií zmíněných výše, což podněcuje více studií ve stylu pan-karcinomu, které hledají významně mutované regulační prvky.

Překvapivě vysoké zastoupení nekódujících variant se nachází i ve farmakogenomické databázi PharmGKB tvořících 88 % ze známých vztahů mezi varianty a léčivy (tzv. variant-drug-association, „VDA“). Mnoho nekódujících variant se nachází v regulačních oblastech, jako je promotor, 5' nepřekládané oblasti (5'-UTR), intronové oblasti, 3' nepřekládané oblasti (3'-UTR) a mezigenové oblasti, které řídí sestřih genů, transkripci a translaci. Není známo, zda nekódující varianty přímo ovlivňují funkci proteinu, expresi nebo jsou přítomny v haplotypu, který zahrnuje funkčně relevantní genetické varianty tzv. syntetickou asociaci (Park et al. 2020). Dickson (Dickson et al. 2010) a kolegové definovali syntetickou asociaci jako asociaci genotypizovaného společného markeru, který je výsledkem více nepozorovaných nízkofrekvenčních kauzálních variant. Z pohledu farmakogenomiky, lze účinky nekódujících variant na reakce na léčiva tedy interpretovat jako potenciální dopad nekódujících variant na funkční aspekty kódujícího genu. Ku příkladu nekódující, zde ale časté varianty v genu pro komplex epoxidreduktázu 1 vitamínu K (VKORC1) ovlivňují transkripci genu přes vazbu na transkripční místo a způsobují inhibici aktivity VKORC1 vyžadující snížené dávkování warfarinu (Rieder et al. 2005; Flockhart et al. 2008). Dále pak nekódující RNA (ncRNA) jako regulátory na post-transkripční úrovni představují nové zdroje poznatků s potenciálem překonání chemorezistence u terapie FOLFOX tzn., kombinace kyseliny folinové (FOL), 5-fluorouracilu (F) a oxaliplatinu (OX)) (Hon et al. 2018). Studie sledující výše zmíněnou syntetickou asociaci vyhodnotila zátěž nekódujících VDA v databázi PharmGKB pomocí šesti různých metod *in silico* (tj. SIFT, MutationTaster, PolyPhen-2 HDIV, PolyPhen-2 HVAR, phyloP a GERP++) u různých etnických skupin a dle klasifikačního systému léčiv dle terapeutické oblasti ATC (The Anatomical Therapeutic Chemical). Studie prokázala významnou korelaci mezi známými nekódujícím VDA (n = 2011) v PharmGKB s funkčně relevantním účinkem vzácných a běžných kódujících (nesynonymních) variant a 3668

farmakogenů pro 5099 léčiv z databáze DrugBank (Park et al. 2020). Použité *in silico* metody však byly vyvinuty pro hodnocení hlavně kódujících variant (nsSNVs), což částečně zpochybňuje nalezené korelace či následování postupu zvolených metod těchto autorů ke splnění prvního cíle předkládané práce, který zahrnuje funkční predikci nekódujících variant.

Nedávná studie demonstrovala několik sestřihových variant s dopadem na prognózu přežití bez onemocnění u trojitě negativního karcinomu prsu (Lianto et al. 2021). *In silico* nástroje pro predikci variant ovlivňujících sestřih zohledňují různé aspekty mechanismu sestřihu, jež se skládají ze (i) sestřihových signálů, včetně 5'/3' místa sestřihu a bodu rozvětvení; ii) regulační prvky sestřihu, tj. exonické/intronické zesilovače/silencery sestřihu; a iii) spliceozom a další trans-aktivní prvky, které se vážou na cis-aktivní prvky.

Obecně, dokončení a regulace sestřihu se opírá o složité biochemické reakce mezi sekvencemi nukleotidů tzv. cis-aktivními elementy a různými vzájemnými bílkovinami tzv. trans-aktivní elementy. Cis-aktivní elementy obsahují 5' místo sestřihu (spojnice mezi exonem a intronem), 3' sestřihové místo (spojnice mezi intronem a exonem), 3' sestřihové místo (spojnice mezi intronem a exonem), bod větvení (desítky nukleotidů před místem 3' sestřihu), exonické zesilovače sestřihu (ESE), intronické zesilovače sestřihu, exonické sestřihové silencery, intronické sestřihové silencery. Trans-aktivní prvky zahrnují spliceozom. Klíčovým krokem sestřihu je lokalizace exon-intronových hranic zachycením sestřihových signálů obsažených v pre-mRNA pomocí spliceozomu. Rozsáhlá srovnání sekvencí na hranicích exonů a intronů odhalily klíčové sestřihové signály, které jsou tvořeny jak přítomnosti téměř neměnných míst GT-AG (první a poslední dvě místa intronu), tak méně sekvenčně konzervovaných míst v blízkosti těchto hranic nazvaných 5' a 3' konsensní sekvence sestřihu. V jednotlivých genech se konečný produkt sestřihu může lišit v důsledku alternativního sestřihu s výsledkem různých sekvencí proteinů bez škodlivých účinků na jeho funkce. Důsledkem alternativního sestřihu může být vynechání exonu (tzv. exon skipping), použití odlišných 5' nebo 3' sestřihových míst (alternativní 5' nebo 3' sestřihová místa), zachování jednoho ze dvou exonů, nebo zachování intronu. U lidských genů je častý alternativní sestřih, který diverzifikuje genovou expresi (např. v různých tkáních). Mutace v místech sestřihu mohou také modifikovat vzorce sestřihu škodlivým způsobem. Například, mutované místo sestřihu může narušit autentický exon-intron a tím změnit vazebné místo spliceozomu, což vede k aberantnímu sestřihu (Jian et al. 2014).

Většina nástrojů vyžaduje zadání jedné nebo více sekvencí s určením nebo bez určení exon-intronových hranic. Společným rysem obou formátů je poskytnout nástroji tzv. sekvence kolem místa sestřihu (buď ručně, nebo automaticky, což naznačuje, že předpověď dané mutace závisí pouze na samotném kontextu sekvence, bez ohledu na to, který nástroj je zvolen. Ve skutečnosti hlavní rozdíly mezi nástroji jsou konsenzuální sekvence, které používají pro porovnávání se vstupními sekvencemi, použité statistické modely pro toto porovnávání, a metody trénování implementované v přístupech strojového učení. Většina těchto metod, včetně CADD, GWAVA a FATHMM-MKL používá algoritmy strojového učení. Přestože přístup strojového učení je vysoce automatický, jeho nevýhoda je stejně zřejmá jako jeho výhoda, a tou je nadměrné přizpůsobení (tzv. overfitting) popsaná pro nástroje kódujících oblastí níže. Pokud je pro klasifikátor zvoleno příliš mnoho parametrů v trénovacím souboru vzhledem k počtu příkladů, zobecnění bude pravděpodobně nízké. To znamená, že při použití nadměrně přizpůsobené neuronové sítě nebo stroje s podpůrnými vektory k předpovídání neznámých míst sestřihu, nemusí být optimalizované kritérium vhodné. Příkladem jsou GWAVA a deltaSVM, které vyžadují trénovací soubor funkčních i nefunkčních příkladů. Kvalitních a zejména experimentálně ověřených trénovacích datových sad pro nekódující varianty je však stále velmi málo a jsou neúplné, což omezuje přesnost predikce. Druhou nevýhodou je, že mnoho existujících skórovacích metod globálně slučuje anotace napříč mnoha různými typy buněk, což ignoruje pozorování, že regulační prvky mohou fungovat velmi specifickým způsobem pro daný typ buněk. Navíc většina metod, které umožňují analýzu specifickou pro daný buněčný typ, jako je deltaSVM, tak činí omezením svých anotačních zdrojů nebo trénovacího souboru pouze na zájmový buněčný typ. Tento přístup má několik nevýhod. Zaprvé jsou ignorovány potenciálně informativní anotace z podobných typů buněk, což může snížit statistickou sílu. Za druhé, pokud není znám správný typ buňky, může zaměření analýzy na předem definovaný typ buňky vést k výsledkům, které nemají vypovídací hodnotu o základní biologii. V neposlední řadě je na uživatele kladena velká výpočetní zátěž, protože kromě spuštění skórovací metody musí poskytnout anotace a trénovací data z buněčných typů, které je zajímaví (Bodea et al. 2018). Přestože většina genomických oblastí je již nyní anotována množstvím epigenetických prvků, problémem zůstává vyvodit z těchto anotací smysluplné závěry, zejména proto, že data jsou vysoce dimenzionální a mnoho epigenetických prvků je korelovaných.

Pro bioinformatiky, kteří mají obvykle k dispozici velké množství variant k anotaci a predikci, je vhodný nástroj, který dokáže provádět "batch" analýzu tzv. analýzu v šaržích (např. MaxEntScan

a SplicePredictor mají tuto možnost). Ideálním přístupem je platforma s obsahem několika nástrojů a populačních dat, jak je tomu u dbNSFP a VEP.

Při vybírání nástrojů do našich studií sehrálo toto kritérium jednu z nevyšších priorit. dbNSFP v3.0. obsahuje 82,832,027 nsSNVs a ssSNVs, 24 nástrojů pro funkční predikci a skóre konzervace a řadí se mezi jedny z nejpoužívanějších platforem.

dbNSFP (dbscSNV), VEP (TFBP), MaxEntScan, IW scoring framework, RegulomeDB a TargetScan, PINES

dbscSNV

Pro vývoj prediktoru dbscSNV bylo celkově zkoumáno osm nástrojů, poskytujících rozhraní příkazového řádku s možností tzv. analýzy v šaržích. Pozitivní skupina byla sestavena z unikátních scSNV měnících sestřih 408 genů a ze tří databází, např. HGMD (již zmíněná výše), SpliceDisease a použitím dvou ansámblových metod učení, AdaBoost a Random Forest. Pro další zlepšení predikční výkonnosti byly k oběma modelům přidány dvě skóre zachování (phyloP46way_placental a phyloP46way_primate a dvě skóre funkční predikce celého genomu (CADD_raw a CADD_phred. Přidání těchto čtyř skóre dále zlepšilo oba modely s optimální senzitivitou a specificitou (nad 90%), což vedlo významnému zvýšení ve srovnání s jakoukoli jinou hodnocenou metodou.

Výsledné skóre předpovědí (ada_score a rf_score) byly předem vypočteny pro všechny potenciální scSNV v lidském genomu na základě RefSeq verze 62 a Ensembl verze 73. Obě skóre i odpovídající anotační informace jsou navíc automaticky začleněné jako připojená databáze s názvem dbscSNV do dbNSFP.

Spolehlivost predikce dbscSNV (Jian et al. 2014) autoři ověřili na nezávislém testovacím souboru a souboru onkologických dat somatických mutací z databáze COSMIC. MutPred Splice (Mort et al. 2014) poskytoval podobné skóre pravděpodobnosti jako dbscSNV, ale pouze pro kódující varianty a jeho výkonnost na uvedené datové sadě nebyla uspokojivá (AUC = 0,715).

Platforma dbNSFP byla použita ve všech našich studiích vzhledem její vysoké efektivnosti a neustále probíhajícím aktualizacím vylepšujícím platformu.

VEP, Transcription Factor Binding Profile (TFBP)

Webová platforma poskytující anotaci i nekódujících variant obsahuje několik atributů hodnocení vazebného profilu transkripčních faktorů (TFBP) (McLaren et al. 2016). Konkrétně se jedná o

relativní polohu odchylky v zarovnaném TFBP tzv. „MOTIF_POS“, dále příznak označující, zda varianta spadá do vysoce informační pozice vazebného profilu transkripčních faktorů (HIGH_INF_POS) a rozdíl ve skóre motivu referenční sekvence a sekvence varianty pro TFBP (MOTIF_SCORE_CHANGE). Hodnocení atributů vazebného profilu transkripčních faktorů v platformě VEP nemá velkou oporu v publikovaných datech. Vzhledem však jeho snadnému využití v již integrovaných výsledcích společně s ostatními nástroji a populačními daty, byl použit pro svou informační povahu v naší studii častých variant předkládané práce jako podpůrný prediktor.

Nexus IW, IW scoring framework

Nexus IW skóre začleňuje 11 nejčastěji používaných funkčních anotačních metod a skórovacích systémů pro genetické variace v nekódujícím genomu. Výhodou těchto vážených skórovacích metod je, že se nespolehlají na tzv. zlatou standardní tréninkovou sadu variant spojených s onemocněním a domněle benigních variant, která jsou často nepřesná a nepraktická pro rozsah variant spojených se studii k predikci synonymních variant, které nemění sekvenci aminokyselin, ale ovlivňují funkčnost produktů genu (přerušením transkripčních procesů, modulací sestřihu, změnou stability struktury messengerové RNA a ovlivněním rychlosti translace). Například nástroj SILVA k predikci synonymních variant, byl vyškolen pouze na 33 experimentálně určených škodlivých variantách a 785 domněle neutrálních variantách, což je příliš málo pro přesné vyškolení modelu (Zhou a Lauschke 2021). Pomocí neřízeného spektrálního přístupu založeného na kovarianci rysů, IW-Scoring selektivně spojuje některé z těchto jedenácti metod (tj. celý soubor nebo podmnožiny) do jediného "ansámblového" algoritmu a poskytuje dvě samostatná, lineární, vážená, funkční skórovací schémata pro známé a nové varianty. V nezávislých testovacích souborech dat poskytuje IW-scoring konzistentně vysoce kvalitní a stabilní výkon při rozlišování funkčně významných variací od ostatních, zatímco metody, z nichž se skládá, vykazují nekonzistentní výkon. Vývojáři nástroje následně demonstrovali vysokou výkonnost a užitečnost IW-skórování při identifikaci a upřednostňování funkčních variant ve studiích GWAS, kvantitativních lokusů exprese (eQTL) a karcinomu (Wang et al. 2018a).

MaxEntScan

Dosud nejpresnější aproximací pro modelování krátkých sekvenčních motivů je použití distribuce maximální entropie (MED). Ve srovnání s jinými metodami je jediným předpokladem MED konzistence s vlastnostmi empirického rozdělení odhadnutého z dostupných dat. MED také zohledňuje závislosti mezi nesousedními i sousedními pozicemi. Namísto jediného modelu je

MED rámcem s velkou flexibilitou pro generování různých modelů pouhou změnou sad omezení. Tento přístup byl využit u nástroje MaxEntScan (<http://genes.mit.edu/burgelab/>) (Jian et al. 2014). Ve studii (Houdayer et al. 2012) vyhodnotili 6 *in silico* predikčních nástrojů s použitím 272 variant neznámého významu (tzv. VUS) v genech BRCA1 a BRCA2. Tyto varianty byly analyzovány *in silico* a *in vitro*. Výsledkem byla křivka operační charakteristiky přijímače dále použita k identifikaci optimalizované mezní hodnoty pro každý nástroj a k porovnání jejich prediktivní výkonnost pro varianty v 5' a 3' konsenzuálních oblastech (s výjimkou míst GT-AG, protože všechny mutace v místech GT-AG ovlivňují sestřih a byly úspěšně předpovězeny všemi nástroji). Nástroj MaxEntScan, jehož hlavním cílem je předpovídat dopad mutací v 5' a 3' sestřihových konsenzuálních oblastech byl využit i v naší studii častých variant (MAF > 5%).

Dalším příkladem nástroje, který kombinuje *in silico* předpovědi z programů SpliceSiteFinder-like a MaxEntScan a využívá algoritmu logistické regrese je Splicing Prediction in Consensus Elements (SPiCE) (Leman et al. 2018). Predikční protokol SPiCE však využil data pouze pro 11 různých genů, včetně genů *BRCA1*, *BRCA2*, *CFTR* a *RHD* k tréninku a ověření a nebyl do naší bioinformatické analýzy zařazen.

RegulomeDB

Heuristický bodovací systém založený na funkční důvěryhodnosti varianty. Kvantitativní skórovací systém, který určuje, zda se varianta nachází ve funkčním místě a pravděpodobně vede k funkčnímu důsledku, tj. ke změně vazby transkripčního faktoru a regulačnímu účinku genu. Varianty se známými eQTL (tzv. eQTL-SNPs) tj., místy, které jsou asociovány s úrovní transkripce čili ty, u nichž byla prokázána souvislost s expresí (Breitling et al. 2008), nástroj považuje za nejpravděpodobněji významné a označuje je jako varianty kategorie 1. Podíl vzácných (20 %), méně častých (30 %) a běžných (50 %) variant se v jednotlivých kategoriích kromě kategorie 1 příliš neodlišuje. Kategorie 1, která vyžaduje informace eQTL, má velmi málo vzácných variant v důsledku zkrácení při výběru kandidátů za použití matic. V rámci kategorie 1 jsou v podkategoriích uvedeny další anotace od nejjistější (1a, která obsahuje vazbu TF, motiv pro TF a stopu DNázy) po nejméně jistou (1f, která obsahuje pouze vazbu TF nebo vrchol DNázy). Další kategorie představují analogické anotace jako u kategorie 1, ale bez údajů eQTL, a tedy bez známého přímého vlivu na vazbu. Kategorie 2 (a-c) vykazuje přímý důkaz vazby prostřednictvím ChIP-seq a DNázy buď s odpovídající PWM k faktoru ChIP-seq, nebo DNázovou stopou. Kategorie 3(a-b) je považována za méně jistou v ovlivnění vazby vzhledem k neúplnému souboru důkazů. Tato místa mají důkaz ChIP-seq případně motiv, který odpovídá údajům ChIP-seq, ale

žádný důkaz DNázy, nebo důkaz DNázy a jakýkoli jiný motiv. Poslední kategorie 4-6 postrádají důkazy o tom, že varianta skutečně narušuje místo vazby. Patří sem důkaz DNázy a ChIP-seq (kategorie 4), důkaz DNázy nebo ChIP-seq (kategorie 5) nebo jakákoli jednotlivá anotace, která nepatří do výše uvedených kategorií (kategorie 6) (Boyle et al. 2012). Prediktor RegulomeDB byl využitý pro analýzu běžných nekódujících variant a nekódujících variant genu *KIF14* předkládané práce.

TargetScan, https://www.targetscan.org/vert_80/

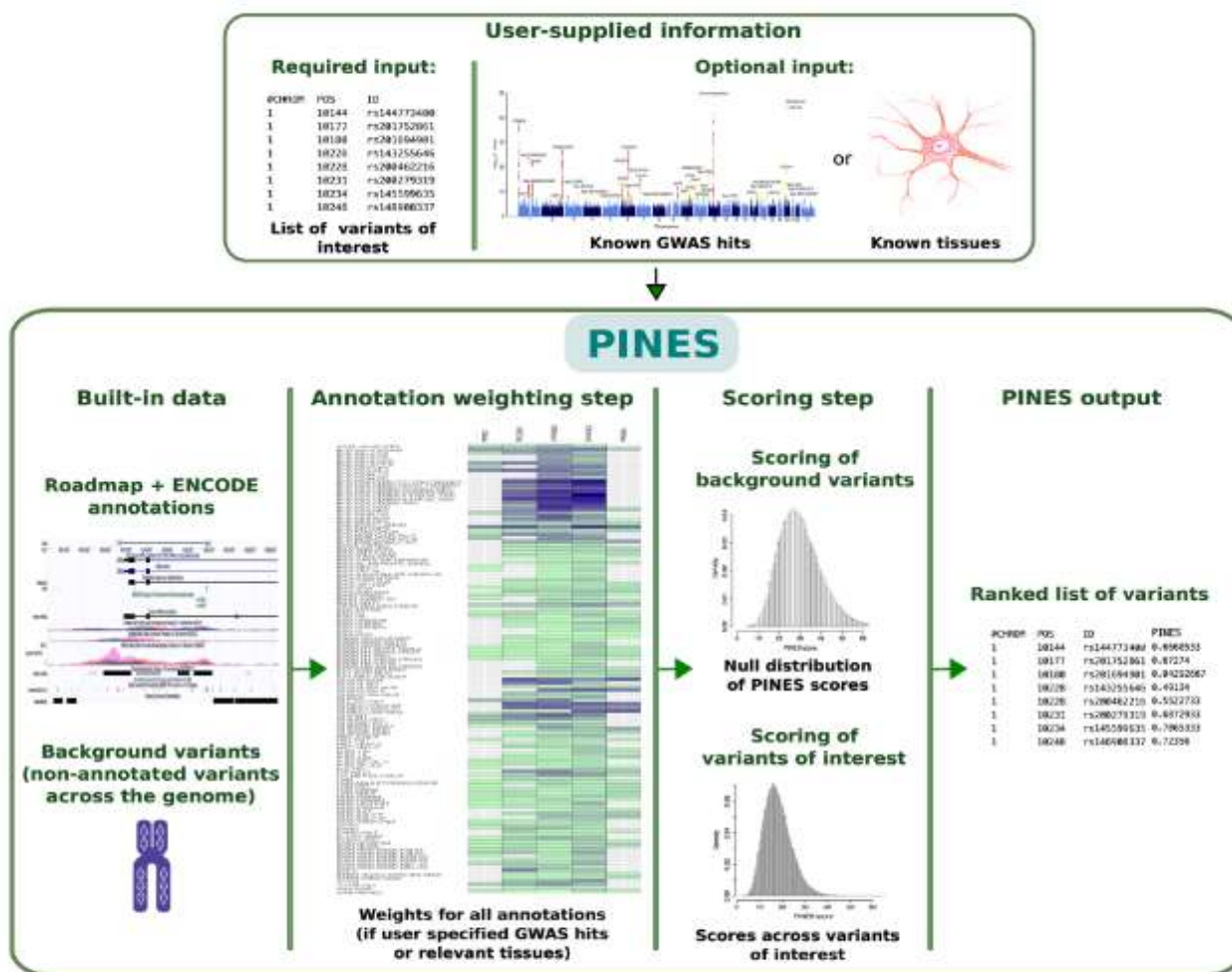
Nástroj využívající analýzu přednostně konzervovaných motivů párování miRNA v rámci 3'UTR s následnou identifikací několika tříd pro cílová místa. Nejúčinnějšími kanonickými typy míst, uvedenými v pořadí podle klesající preferenční zachovalosti a účinnosti, jsou 8merové místo (Watson-Crickova shoda s pozicemi miRNA 2-8 s naproti pozici 1, 7merové místo-m8 (shoda s pozicemi 2-8) a 7merové místo-A1 (shoda s pozicemi 2-7 s A opozitní pozicí 1) (Lewis et al. 2005; Brennecke et al. 2005; Krek et al. 2005; Agarwal et al. 2015). K řízení produkce proteinů se miRNA vážou primárně na kanonická místa. Na základě těchto výsledků Agarwal a kolegové (Agarwal et al. 2015), vypracovali statistický model, který předpovídá účinky vazby miRNA na kanonická místa. Model zohledňuje 14 různých vlastností miRNA, místa miRNA nebo mRNA – včetně sekvence mRNA a byl vycvičen na datech získaných pouze z jednoho typu buněk. Prediktor TargetScan byl aplikován k předpovědi funkčnosti mRNA častých variant předkládané práce.

PINES

PINES akronym pro Phenotype-Informed Noncoding Element Scoring. PINES je čistě epigenetický přístup založený na prioritizaci nekódujících variant a nezahrnuje souhrnné statistiky GWAS přímo do svého skórovacího postupu. Hypotéza vývoje PINES byla založena na upřednostnění nekódujících variant, které mají význam pro konkrétní onemocnění nebo znak, kdy by měla zásadní integrace předchozích genomických informací o fenotypu zvýšit spolehlivost předpovědi. Tyto předchozí informace mohou zahrnovat typy buněk, tkání, genů nebo drah relevantních pro fenotyp. Dále vývojáři předpokládali, že pro předpovídání funkční relevance nekódujících variant by mohlo být přínosné zohlednit obrovské množství variant, u nichž dosud nebyla anotována žádná funkce (varianty v pozadí), a použít je jako výchozí bod pro hledání variant, které se od tohoto pozadí významně odchyľují. Ačkoli absence nahlášených funkčních údajů nemusí nutně znamenat, že všechny varianty v pozadí pocházejí z nefunkčních míst, může

tato sbírka variant přesto sloužit jako užitečná reference, s níž lze porovnávat jednotlivé varianty. PINES se tak od většiny předchozích skórovacích nástrojů odlišuje tím, že k systematickému hodnocení funkčních vlastností používá neřízený přístup. Konkrétně se k trénování modelu PINES nepoužívají žádné sady variant v nastaveních bez vážení a s ručním vážením. Pokud nejsou k dispozici žádné informace relevantní pro fenotyp, PINES hodnotí jednotlivé varianty výhradně porovnáním s úrovní pozadí (zmiňného výše) anotací v celém genomu, přičemž bere v úvahu korelační strukturu mezi anotacemi. Toto tzv. semi-supervisované nastavení zadané varianty používají pouze jako pomůcka při výběru anotačních vah. Tento nástroj umožňuje fenotypově závislé bodování nekódujících variant, které lze přizpůsobit anotacím považovaným za nejvýznamnější pro zájmový fenotyp. Během procesu bodování jsou těmto anotacím přiřazeny nejvyšší váhy, zatímco anotacím s nižší relevancí pro daný fenotyp jsou váhy sníženy viz Obrázek 2—5 (Bodea et al. 2018).

Vývojáři nástroje PINES představili inovativní přístup, postaven na semi-supervizovaném algoritmu, který obecně vyvažuje nedostatky metod strojového učení bez a pod dozorem. Dodatečná možnost specifikace predikce pro jaterní tkán, jako hlavního místa metabolismu léčiv byla dalším důležitým faktorem pro zařazení tohoto nástroje k anotaci častých ($MAF > 5\%$) nekódujících variant předkládané práce (Hlavac et al. 2018).



Obrázek 2—5 schéma modelu predikce nekódujících variant nástroje PINES

2.5.3. Výběr vhodného(ých) prediktoru(ů) pro koncového uživatele

Vzhledem k tomu, že existují desítky prediktorů, stává se výběr vhodného(ých) prediktoru(ů) pro koncové uživatele výzvou. Hlavním praktickým faktorem je často aplikace daného nástroje ve smyslu „web-based“ přístupu, kdy není zapotřebí bioinformatického vzdělání a kódování naproti již náročnému rozhraní příkazového řádku tzv. „command-based“ přístupu využívajícího základní anebo rozšířené znalosti kódovacích jazyků např. Python, R anebo C++.

K prokázání výkonnosti modelu se ve studiích běžně používá několik hodnotících kritérií, jako je citlivost, specifčnost, pozitivní prediktivní hodnota (PPV), negativní prediktivní hodnota (NPV), přesnost a Matthewsův korelační koeficient (MCC) (Thusberg et al. 2011). Hodnoty citlivosti, specifčnosti, PPV, NPV a přesnosti se pohybují v rozmezí od 0 do 1, přičemž vyšší hodnoty znamenají lepší výkonnost.

Pozitivní prediktivní hodnota (PPV) představuje podmíněnou pravděpodobnost, že škodlivé varianty ve srovnávacích datech jsou výpočetní metodou správně klasifikovány jako škodlivé. (ii)

Negativní prediktivní hodnota (NPV) představuje podmíněnou pravděpodobnost, že benigní varianty ve srovnávacích datech jsou výpočetní metodou správně klasifikovány jako benigní. (iii) Falešně negativní míra (FNR) představuje podíl škodlivých variant, které jsou nesprávně předpovězeny jako benigní varianty. (iv) Citlivost (označovaná také jako TPR) představuje tzv. podíl škodlivých variant ve srovnávacích datech, které jsou výpočetní metodou správně předpovězeny jako škodlivé varianty. FNR a citlivost jsou párové míry, jejichž součet = 100 %. (v) Falešně pozitivní míra (FPR) představuje podíl benigních variant ve srovnávacích datech, které jsou výpočetní metodou nesprávně předpovězeny jako varianty škodlivé. (vi) Specifičnost (označovaná také jako míra pravé negativy) představuje podíl benigních variant ve srovnávacích datech, které jsou výpočetní metodou správně předpovězeny jako benigní varianty. FPR a specifičnost jsou párové míry, jejichž součet = 100 %. (vii) Přesnost představuje podíl benigních variant a škodlivých variant ve srovnávacích datech, které jsou správně předpovězeny jako benigní varianty, resp. škodlivé varianty. (viii) MCC představuje korelační koeficient mezi pozorovanou a předpovězenou klasifikací v rozmezí (-1 až 1). Koeficient 1 znamená tzv. dokonalou předpověď, 0 znamená náhodnou předpověď a (-1) znamená naprostý nesoulad mezi předpovědí a skutečnou klasifikací. (ix) Dále se často používá vizualizační měření, analýza operačních charakteristik přijímače (ROC), která intuitivně porovnává plochu pod křivkou ROC (AUC) pro více prediktorů (Vihinen 2012). U neprotínajících se křivek naznačuje hodnota AUC blížící se 1 lepší celkový výkon, zatímco hodnota 0,5 znamená náhodnou a nepoužitelnou klasifikaci. Pro každou ROC je AUC jedinou hodnotou, což snižuje složitost křivky ROC. Obecně platí, že čím vyšší je AUC, tím lepší je výkon výpočetní metody. Některé z výše uvedených měř jsou odvozeny z parametrů pravdivé positivity (TP), pravdivé negativy (TN), falešné positivity (FP) a falešné negativy (FN), jak je uvedeno níže (Li et al. 2018; Qorri et al. 2022; Hassan et al. 2019b; Mueller et al. 2015; Vihinen 2012; Thusberg et al. 2011; Whalen et al. 2022) viz Rovnice 2—1 a Graf 2—3.

$$PPV = \frac{TP}{TP + FP}$$

$$NPV = \frac{TN}{TN + FN}$$

$$FNR = \frac{FN}{TP + FN}$$

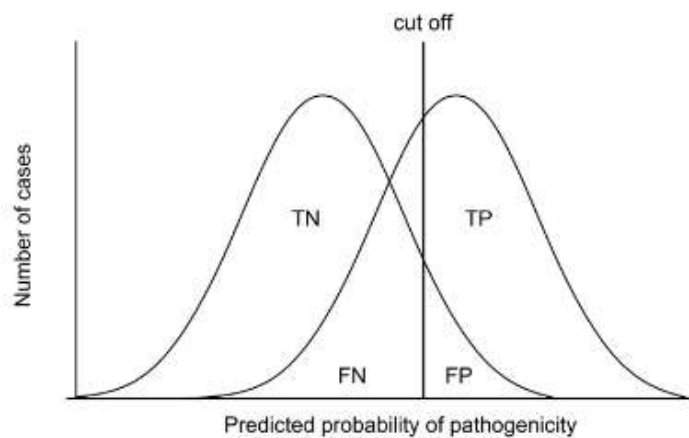
$$\text{Sensitivity} = \frac{TP}{TP + FN}$$

$$FPR = \frac{FP}{TN + FP}$$

$$\text{Specificity} = \frac{TN}{TN + FP}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}$$

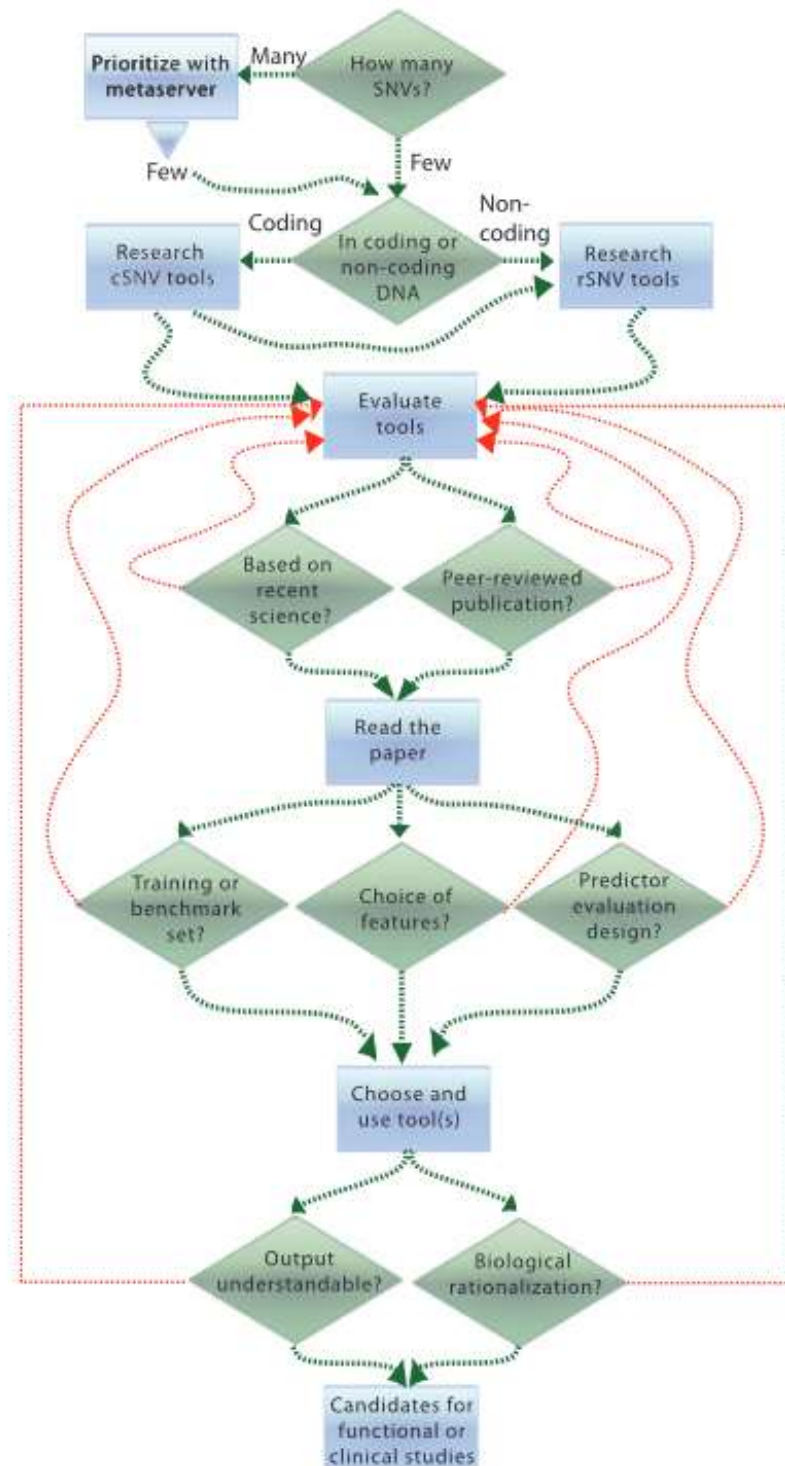
$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TN + FN) \times (TN + FP) \times (TP + FN) \times (TP + FP)}}$$



Rovnice 2—1 pozitivní prediktivní hodnota (PPV), negativní prediktivní hodnota (NPV), falešně negativní míra (FNR), citlivost, falešně pozitivní míra (FPR), specifičnost, přesnost a Matthewsův korelační koeficient (MCC) převzato z (Vihinen et al. 2012)

Graf 2—3 Mezní hodnota nástroje a poměr pravděpodobnosti škodlivosti převzato z (Vihinen et al. 2012)

Několik studií porovnávalo přesnost predikce sekvenčních (Balasubramanian et al. 2005; Mathe et al. 2006) a strukturních algoritmů (Bao a Cui 2005; Chan et al. 2007; Chao et al. 2008). Navíc existují desítky srovnávacích studií o hodnocení výkonnosti různých vybraných nástrojů pomocí různých referenčních datových sad (Chun a Fay 2009; Thusberg et al. 2011; Schwarz et al. 2010; Wei et al. 2011). Základní výběr prediktoru pro studii by měl obecně následovat schéma níže Obrázek 2—6.



Obrázek 2—6 Schéma postupu adekvátního výběru *in silico* prediktoru (Cline a Karchin 2011)

Příkladem je srovnání pro tři nástroje (SIFT, Align-GVGD a PolyPhen-2) využité i v naší studii (Hicks et al. 2011), zaměřené na přesnost a vliv použitého zarovnání při generování předpovědi funkčnosti již dobře charakterizovaných nesynonymních mutací ($n = 267$) genů BRCA1, MSH2,

MLH1 a TP53. Výsledek křivky operačních charakteristik přijímače založených na datech a využívajících nativní zarovnání pro každý algoritmus byl v rozmezí 78-79 % (AUC) pro všechny čtyři algoritmy. Předpovědi algoritmu PolyPhen-2 byly nejméně závislé na použitém zarovnání. Naproti tomu algoritmus Align-GVGD předpovídá všechny varianty neutrálně, pokud je mu poskytnuto zarovnání s velkým počtem sekvencí. Za zmínku stojí, že algoritmy provádějí různé předpovědi variant i při poskytnutí stejného zarovnání a nemusí nutně dosahovat nejlepších výsledků při použití vlastního zarovnání.

Teoreticky by strukturální prediktory popsány výše měly potenciálně poskytnout lepší vhled do mutačního efektu než prediktory založené na homologních sekvencích, protože jsou postaveny na přímém dopadu mutace na strukturu a funkci proteinu. Pravdou však je, že prediktory založené na proteinech jsou stále omezené kvůli nevyváženosti a vnitřní variabilitě termodynamických dat a jejich predikční výkonnosti. Přestože Protein Data Bank (PDB) obsahuje více než 50 000 záznamů o lidských proteinech, mnohé z nich jsou redundantní a pokrývají pouze 70 % referenčního lidského proteomu na úrovni sekvenční identity vyšší než 30 %. Celkově tak techniky založené na sekvenci za určitých okolností překonávají nástroje pro predikci stability založené na struktuře (Liu et al. 2022).

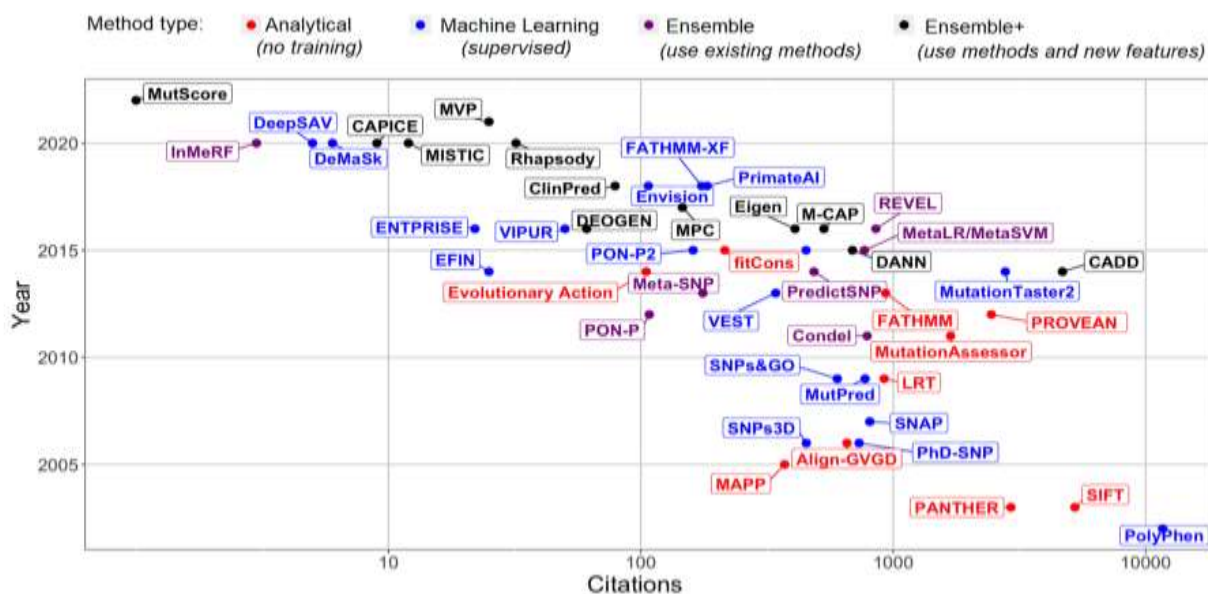
Klíčovým omezením naproti tomu u metod strojového učení s dozorem je spoléhání se na tvrzení o patogenitě z existujících databází, které mohou být nepřesné a neúplné. Nesprávná klasifikace trénovacích a testovacích variant způsobujících onemocnění anebo neutrálních variant omezuje přesnost metody predikce i výsledné odhady výkonnosti. Překvapivě vysoká heterogenita a nesoulad na základě implementované metody predikce byly demonstrovány na výsledcích panelového sekvenování pacientů s dilatační kardiomyopatií ze studie INHERITANCE (842 nsSNVs objevených u 639 pacientů), u kterých byla prověřena úplná sekvence 76 genů souvisejících s kardiomyopatiemi (Mueller et al. 2015). Predikce pomocí 13 nástrojů pro funkční odhad patogenity nsSNV odhalila, že známé asociace s onemocněním nebyly těmito nástroji identifikovány, což omezuje přímou využitelnost prediktorů v klinické praxi.

Již v roce 2015 Grimm a kolegové (Grimm et al. 2015) podrobně popsali dva potenciální zdroje chyb při vývoji nástrojů pro predikci účinků variant, a to kruhovost neboli tzv. cirkularitu typu 1 a typu 2. Obě chyby postihují jednotlivé predikční nástroje a mohou se kumulovat v integračních přístupech, jako jsou výše popsané metaprediktory. Cirkularita typu 1 (overfitting) se týká skutečného překrývání trénovacích a testovacích dat, které by mohlo vést k příliš optimistickému úsudku o prediktivní výkonnosti daného nástroje, uvedeno na příkladu nástroje ReVe (spojení

dvou nástrojů REVEL a VEST). Aby autory při vývoji nástroje obešli cirkularitu typu 1, zahrnuli pouze údaje HGMD, které byly do databáze přidány po 1. lednu 2013, protože žádný z integrovaných nástrojů nepoužil pro vývoj varianty z těchto dat. Příprava trénovací sady „benchmark data“ se obdobně jako u nástroje REVEL neobešla bez specifického vyřazení redundantních variant k finálnímu a rovnoměrnému rozdělení benigních (2782) patogenních variant (2098). Výkon nástroje byl rovněž testován na somatických variantách nádorů použitím dat u IARC TP53 databáze s obsahem 477 nefunkčních nesynonymních variant a 537 funkčních nesynonymních variant. Výsledky studie 23 metod poukázali, že některé výpočetní metody musí používat přísnější prahové hodnoty pro předpovídání škodlivých somatických nesynonymních mutací. Obecně lze říci, že metody PROVEAN, SIFT a VEST3 vykazovaly nejvyšší přesnost a skóre MCC (Li et al. 2018).

Cirkularita typu 2 se týká vnitřní tendence některých predikčních nástrojů předpovídat účinek dané varianty především podle účinku variant na stejný gen. Takové zkreslení může vzniknout, pokud jsou varianty v jednom a tomtéž genu s větší pravděpodobností zaznamenány pod stejnou funkční značkou v mutačních databázích používaných pro vývoj nástrojů. Proto v souboru dat variant, kde jsou proteiny pouze s patogenními nebo pouze s benigními variantami (nevyvážený soubor dat), mají některé algoritmy tendenci dosahovat lepších výsledků než v souboru dat, který má stejný počet patogenních a benigních variant na gen (dokonale vyvážený), i když tomu tak z biologického hlediska není. Příkladem takového nástroje je FATHMM (Ghosh et al. 2017; Grimm et al. 2015) a nebyl proto samostatně využit v našich studiích nýbrž jako relativně důležitá integrovaná složka nástroje REVEL (Ioannidis et al. 2016) v predikci vzácných variant druhé studie proveditelnosti. Jendou z možností je strategie výběru variant, kdy všechny varianty genu jsou přiřazeny ke stejnému souboru dat, čímž zajistí téměř rovnoměrné rozložení počtu variant na gen (Carter et al. 2013; Capriotti et al. 2006). Vzhledem rostoucí k převaze používání metaprediktorů oproti aplikacím založeným na konkordanci v posledních dvaceti letech bylo cílem další recentní studie z roku 2022 (Qorri et al. 2022) vyhodnotit a porovnat výkon deseti široce používaných predikčních algoritmů, včetně osmi individuálních algoritmů (Polyphen-2-HumDiv, Polyphen-2-HumVar, SIFT, PMut, PROVEAN, PhD-SNP, SNPs&GO, a PANTHER-PSEP) a dva metaprediktory (META-SNP a PredictSNP) pro 21 klinicky relevantních genů, včetně BRCA1 a BRCA2. Na hlavním souboru dat, který se skládá ze 404 nesynonymních variant, výsledky ukázaly, že výkonnost predikčních algoritmů se značně liší. Na základě MCC a AUC vykazovaly SNPs&GO a PMut konzistentně celkově nadprůměrnou výkonnost napříč různými datovými sadami. Většina nástrojů vykazovala vyšší citlivost a negativní prediktivní hodnoty na úkor nižší specificity a

pozitivních prediktivních hodnot. Autoři prokázali, že typ 1 cirkularity významně ovlivňuje výkonnost těchto nástrojů, a pokud není zohledněna, může zkreslovat výběr nejvýkonnějších algoritmů. Jako nejvýkonnější predikční algoritmy v souborech dat BRCA1 byly určeny PANTHER-PSEP, PMut a SNPs&GO, zatímco PMut byl určen jako nejvýkonnější predikční algoritmus pro soubor dat BRCA2. Autoři doporučili vyhnout se použití programu PROVEAN pro prioritizaci variant v genech BRCA1 a BRCA2, protože by to bránilo shodě mezi použitými predikčními algoritmy, čímž by se výpočetní důkaz stal zbytečným. Kromě toho, pomocí specifického souboru dat studie prokázala, že sedm z hodnocených nástrojů bylo nápadně ovlivněno cirkularitou typu 1, která, pokud není řešena před srovnávacím testem predikčních algoritmů, může vést k umělému zvýšení jejich výkonnosti a následně k výběru nesprávných algoritmů, jak již bylo zmíněno výše. Problematika správné aplikace strojového učení v genomice nadále přetrvává a pro koncového uživatele, který má přístup pouze k omezeným publikovaným informacím o vývoji specifického modelu představuje značné komplikace v rozhodování, který nástroj použít a jak výsledky predikce správně interpretovat. V přehledu (Whalen et al. 2022), autoři detailně popsali nástrahy použití strojového učení v genomice, především přímé dopady úskalí ML na výkonnost modelu. ML by se mělo obecně používat spíše k získání biologických poznatků než k predikci jako takové. V těchto případech se zpravidla interpretuje natrénovaný model k extrakci vztahů, které se naučil. Modely, které vykazují dobrou výkonnost se, ale mohly naučit nesmyslné vztahy, pokud analýzu ovlivnila jedna nebo více „nástrah“. Dále model může nadhodnocovat sílu asociace mezi znakem a výsledkem v důsledku zmatení nebo dependence. Neustálý vývoj nových bioinformatických nástrojů pro predikci genetické predispozice funkčních mutací popsanych výše a zobrazený počtem citací Obrázek 2—7, demonstrují sílu již zaběhlé počítačové doby.



Obrázek 2—7 Počet citací k primárnímu článku prediktoru v závislosti na roce zveřejnění a typu metody (Katsonis et al. 2022)

Počet citací byl získán pomocí vyhledávače Google Scholar dne 7. března 2022 převzato z (Katsonis et al. 2022).

Závěrem 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.

2.6. Hypotéza: Cíle řešení

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 antracyclinech, 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). Klinická data pacientek pilotní studie (n = 105) a validační studie (n = 805) jsou uvedena v tabulkách (Tabulka 3—1, Tabulka 3—2) níže. Klinická data pacientek studie (n = 808) genu *KIF14* jsou uvedena v tabulce č. 1 publikace (Krus et al. 2022). Klinická data pacientek studie proveditelnosti jsou uvedena v suplementu (Supplementary Table S2) publikace (Kováčová et al. 2024).

Farmakoterapie následovala běžné klinické postupy pro neoadjuvantní cytotoxickou léčbu a/nebo adjuvantní léčbu po chirurgické léčbě v době studie. Cytotoxická terapie byla založena na stejných kombinacích léčiv jak v testovacím, tak ve validačním souboru (viz 2.3). Odpověď na neoadjuvantní cytotoxickou léčbu (NACT) byla hodnocena na základě běžných zobrazovacích metod a kritérií RECIST (Response Evaluation Criteria in Solid Tumors) v1.1 (Schwartz et al. 2016).

Tabulka 3—1 Klinická data pacientek pilotní studie

Sledovaný parametr	Pacientky, N (%) ¹
Věk v době diagnózy, průměr ± S.D. ² (roky)	51.7 ± 9.4
Menopauzální stav	
Premenopauzální	46 (46)
Postmenopauzální	55 (55)
Chybějící data	4
Velikost tumoru (pT)	
pTis	8 (8)
pT1	50 (48)
pT2	40 (39)
pT3	5 (5)
pTX	2
Metastázy lymfatických uzlin (pN)	
Bez postižení (pN0)	68 (65)
S postižením (pN1-3)	37 (35)
Patologické stadium (S)	
SI	46 (44)
SII	47 (45)
SIII	12 (11)

Sledovaný parametr	Pacientky, N (%) ¹
Histologický typ	
Invazivní duktální karcinom	88 (84)
Jiný typ	17 (16) ⁴
Patologický grade (G)	
G1	11 (11)
G2	35 (35)
G3	54 (54)
GX	5
Stav exprese receptoru pro estrogen	
Pozitivní	38 (38)
Negativní	61 (62)
Chybějící data	6
Stav exprese receptoru pro progesteron	
Pozitivní	39 (39)
Negativní	60 (61)
Chybějící data	6
Stav exprese receptoru HER2/ERBB2	
Pozitivní	2 (2)
Negativní	97 (97)
Chybějící data	6
Exprese Ki-67, průměr ± S.D. ² (%)	32.9 ± 20.3
Molekulární podtyp	
Luminální A	15 (16)
Luminální B	23 (24)
Trojité negativní	58 (60)
Chybějící data	9
Odpověď na neoadjuvantní cytotoxickou léčbu	
kompletní nebo částečná	47 (69)
stabilní onemocnění nebo progresse	21 (31)
Ostatní ³	37

Poznámky pod čarou: ¹Počet pacientek s % v závorce; ²S.D. = směrodatná odchylka; ³pacientky léčené adjuvantní léčbou bez neoadjuvantní cytotoxické léčby; ⁴šest lobulárních, šest medulárních, dva metaplastické, jeden mucinózní, jeden papilární a jeden neuroendokrinní invazivní karcinom.

Tabulka 3—2 Klinická data pacientek validační studie

Sledovaný parametr	Pacientky, N (%) ¹
Věk v době diagnózy, průměr ± S.D. ² (roky)	58.9 ± 12.5
Menopauzální stav	
Premenopauzální	197 (25)
Postmenopauzální	590 (75)
Chybějící data	18
Velikost tumoru (pT)	
pTis	65 (8)
pT1	489 (62)
pT2	208 (27)
pT3	18 (2)
pTX	10 (1)
Metastázy lymfatických uzlin (pN)	
Bez postižení (pN0)	509 (67)

Sledovaný parametr	Pacientky, N (%) ¹
S postižením (pN1-3)	253 (33)
pNX	43
Patologické stadium (S)	
S0	61 (8)
SI	358 (47)
SII	282 (37)
SIII	67 (9)
Neurčeno	37
Histologický typ	
Invazivní duktální karcinom	598 (75)
Jiný typ	197 (25)
Chybějící data	10
Patologický grade (G)	
G1	177 (23)
G2	385 (50)
G3	209 (27)
GX	34
Stav exprese receptoru pro estrogen	
Pozitivní	618 (77)
Negativní	181 (23)
Chybějící data	6
Stav exprese receptoru pro progesteron	
Pozitivní	579 (73)
Negativní	220 (27)
Chybějící data	6
Stav exprese receptoru HER2/ ERBB2	
Pozitivní	194 (24)
Negativní	602 (76)
Chybějící data	9
Exprese Ki-67, průměr ± S.D. ² (%)	23.3 ± 22.6
Molekulární podtyp	
Luminální A	330 (41)
Luminální B	313 (39)
Trojitě negativní	93 (12)
HER2	63 (8)
Chybějící data	6
Odpověď na neoadjuvantní cytotoxickou léčbu	
kompletní nebo částečná	127 (75)
stabilní onemocnění nebo progresse	43 (25)
Ostatní ³	635

Poznámky pod čarou: ¹Počet pacientek s % v závorce; ²S.D. = směrodatná odchylka; ³pacientky léčené adjuvantní léčbou % v závorce; ²S.D. = směrodatná odchylka; ³pacientky léčené adjuvantní léčbou bez neoadjuvantní cytotoxické léčby.

3.1.2. Populační zdrojová data

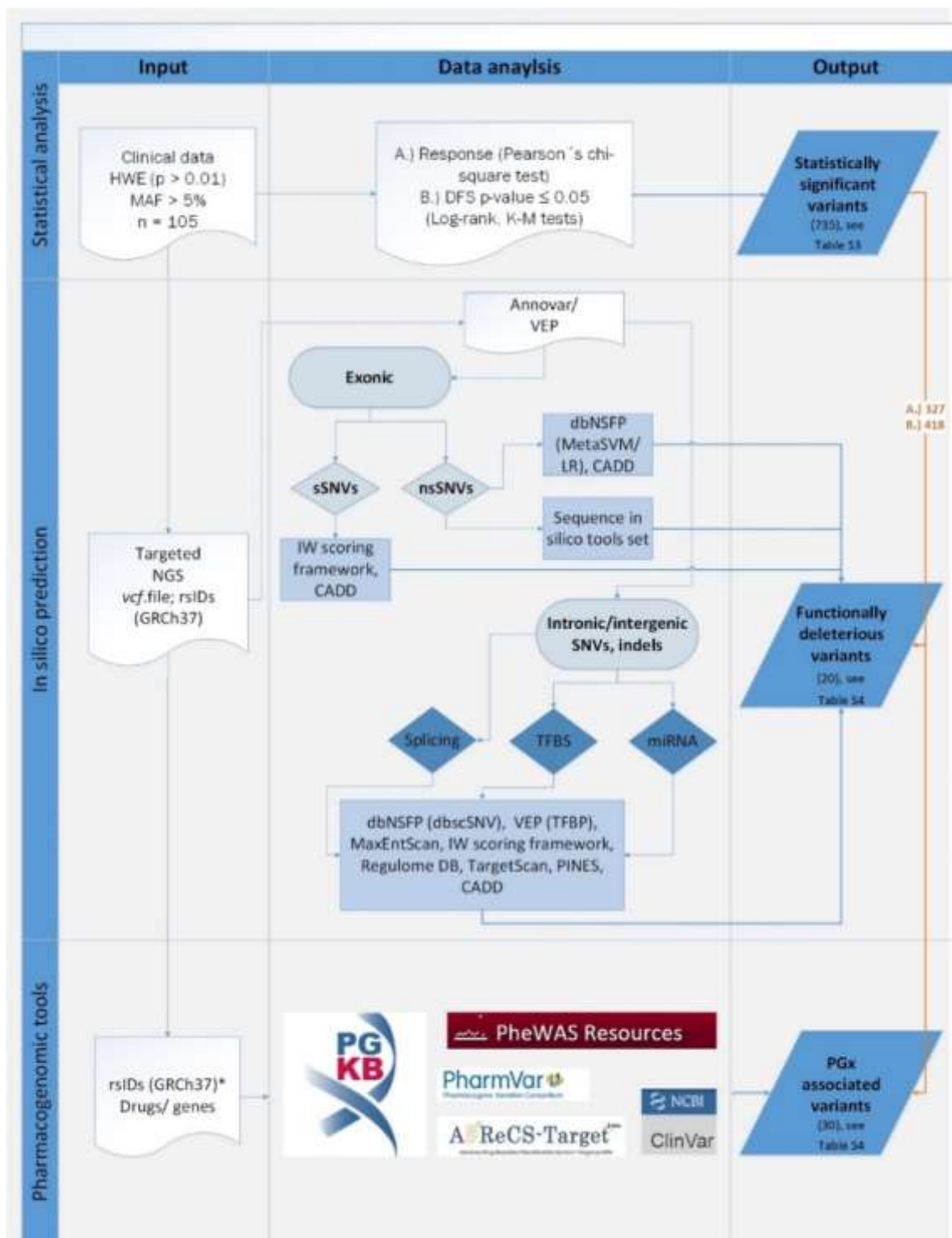
Zdrojová data pocházela z globálních veřejných uložišť genetických a farmakogenomických variant. Pro určení frekvence alel byla použita data ze světové databáze dbSNP s užším omezením na data etnicky odpovídající evropské populaci ExAC (bez Finské populace), nejaktuálnější

evropské frekvence ALFA (dbGaP) a databáze Národního centra lékařské genomiky (NCMG) pro českou subpopulaci.

3.1.3. 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 (viz Seznam příkladových kódů). 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, ggsvplot).

Celková datová analýza a strategie výběru kandidátních variant v pilotní fázi studie byla rozdělena do tří základních kroků. Strategie postupů uvedená v diagramu níže (viz Obrázek 3—1) striktně nenásleduje časovou posloupnost bioinformatické analýzy. V prvním kroku proběhla *in silico* predikce, data mining ve farmakogenomických databázích a následně statistické zpracování pro analýzu korelace variant s klinickými parametry pacientek pilotní fáze. *In silico* predikce byla provedena pro synonymní (sSNV), nesynonymní (nsSNV) jednonukleotidové varianty, sestřihové (scSNV) varianty, miRNA a krátké inserce/delece (tzv. indel mutace) z dat sekvenování nové generace (NGS) ve formátu VCF pomocí několika webových softwarových nástrojů a nástrojů příkazového řádku za předem definovaných mezních hodnot. Pro varianty s predikcí v nástrojích bez možnosti použití kódovacího jazyka, případně hromadného výsledku v tzv. šaržích („batch analysis“), proběhla *in silico* predikce pouze pro statisticky významné varianty. Získané výsledky tvořily základ funkční evidence seřazené dle váhy námi vytvořených pravidel (viz Tabulka 3—3) pro prioritizaci variant následně určených k ověření ve validační fázi studie.



Obrázek 3—1 Schéma prioritizace variant

Počty unikátních variant sdílených se statisticky významnými výsledky jsou znázorněny v závorkách. Statistická analýza vztahů zárodečných variant z cíleného sekvenování, splňovala následující podmínky: Hardy-Weinbergova rovnováha (HWE, $p > 0,01$) a frekvence minoritních alel (MAF $> 0,05$), asociace s klinickými údaji pacientek provedena za účelem hledání souvislostí s odpovědí na léčbu (A) a (B) přežitím bez onemocnění (DFS). Ve farmakogenomických (PGx) databázích byly pro ruční kurátorství zvažovány pouze statisticky významné varianty (*). GRCh37 = Genome Reference Consortium Human Build 37 (hg19); TFBS = vazebné místo transkripčního faktoru; TFBP = profil vazby transkripčního faktoru.

Tabulka 3—3 Pravidla výběru variant do validačního setu podle váhy evidence získaných dat

Input	Priorita			
	Nejvyšší	Vysoká	Střední	Nízká
Funkce varianty	Nejvyšší	Vysoká	Střední	Nízká
Odpověď na léčbu nebo přežití bez onemocnění	signifikantní	signifikantní	signifikantní	signifikantní
PharmGKB ClinVar	související odpověď na léčivou látku/ asociace s karcinomem	související žádné údaje	žádné údaje odpověď na léčivou látku/ asociace s karcinomem	žádné údaje žádné údaje
<i>In silico</i> predikce Funkce související s karcinomem	škodlivá ano	škodlivá/neutrální ano	škodlivá žádné	škodlivá ano

Signifikantní = statisticky signifikantní ($p = 0,05$)

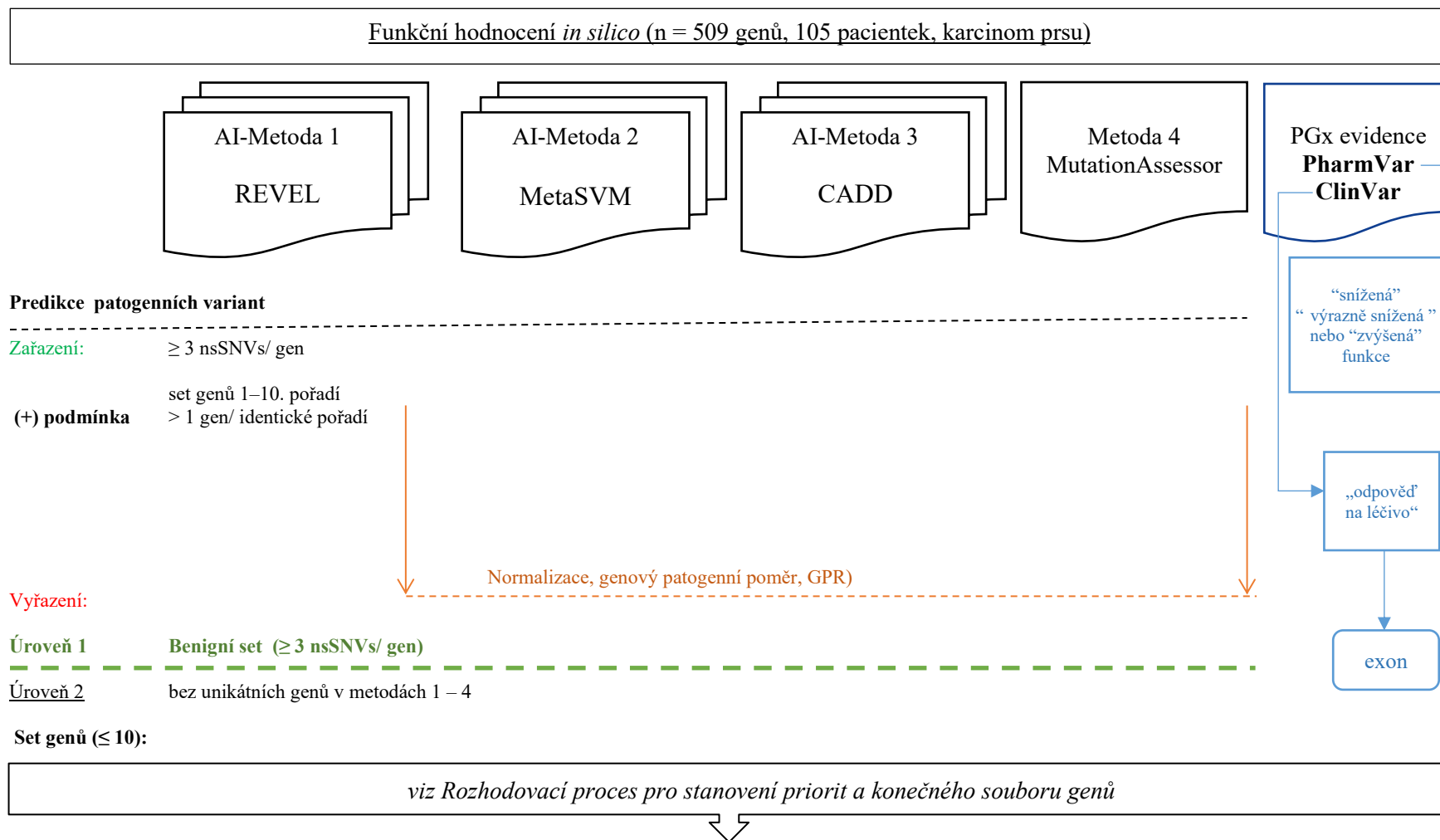
Výše uvedená pravidla pro validaci ve větším souboru ($n = 805$) pacientek s karcinomem prsu splňovalo 58 variant (56 SNV a dvě inserce/delece). Varianty jsou uvedeny sekci 4.2.

3.1.4. 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 specifičnost a citlivost, jak je definovali vývojáři nástroje.

Algoritmy a prahové hodnoty jsou popsány ve schématu Obrázek 3—2 Strategie použití *in silico* nástrojů a proces získání funkční evidence



Nástroj: Mezní hodnoty pro patogenní predikci byly následující: MetaSVM (Dong et al. 2015) (D: škodlivá), REVEL (Ioannidis et al. 2016) (score > 0.5), CADDphred (Kircher et al. 2014) (skóre ≥ 19). Mezní hodnoty pro benigní predikci byly následující: REVEL (skóre ≤ 0.5) VEST3 (skóre ≤ 0.8), MetaSVM, and MetaLR (Dong et al. 2015) (T: Tolerantní). MetaSVM and MetaLR se vzájemně doplňují a v metodě 3 této studie se zohledňují jako jeden nástroj. (Ghosh et al. 2017). Pro MutationAssessor (Reva et al. 2011) byla použita optimalizovaná mezní hodnota (> 2.0566) za účelem zvýšení přesnosti predikce specifické pro farmakogeny, jak bylo publikováno dříve (Zhou et al. 2018).

Obrázek 3—2 Strategie použití *in silico* nástrojů a proces získání funkční evidence

Metoda REVEL využívá strojové učení pod dohledem s algoritmem Random Forest a integruje 13 samostatných nástrojů. Některé z těchto nástrojů jsou založeny na naivním Bayesově klasifikátoru, skrytých Markovových modelech (HMM) a algoritmech Random Forest. Zatímco jiné, např. nástroje SIFT, PROVEAN a MutationAssessor, nejsou primárně založeny na metodách strojového učení, ale využívají přístup založený na zachování sekvence nebo metodu založenou na zachování pouze např. GERP++. Ačkoli je MutationAssessor integrován i do programu REVEL, samostatná predikce byla provedena s přísnějším a pro PGx specifickým prahem. Další vybrané nástroje řízené umělou inteligencí MetaSVM a CADD poskytují skóre souboru. MetaSVM integruje devět nezávislých skóre, frekvenci alel pro skutečně negativní běžné a vzácné varianty a používá algoritmus podpůrného vektoru (SVM). CADD integruje různé genomické anotace a funkce pomocí přístupu strojového učení logistické regrese.

Nástroje umělé inteligence využívají poměrně široké spektrum mutací ve vztahu k typu genomických dat. Funkční predikce je obecně řízená výsledkem pro škodlivé varianty. Sílu výpočetní predikce je ale možné navýšit i použitím adekvátně zvolených algoritmů k predikci variant u kterých se neočekává funkční dopad a využít tohoto kroku k filtraci. V roce 2017 Ghosh a kolegové (Ghosh et al. 2017) publikovali studii, která prokázala, že nejvýkonnější kombinace algoritmů se pro benigní a patogenní varianty navzájem liší. V naší studii proveditelnosti jsme proto se záměrem snížit počet falešných předpovědí začlenili specifickou sadu nástrojů řízených umělou inteligencí (REVEL, VEST3 3.0 a MetaSVM) se shodnou předpovědí benigních následků. Absence předpovědi některým z nástrojů v kombinaci vyřadila varianty z tohoto souhlasného hodnocení. Do procesu jsme navíc zařadili normalizaci výsledků predikce. Účelem bylo upravit výsledky algoritmů REVEL, MetaSVM, CADD a MutationAssessor a získat soubor genů s nejvyšším počtem domněle patogenních variant na gen tzv., „gene pathogenic ratio“ (GPR) pro daný algoritmus. Vzhledem k tomu, že studie je omezena na oblasti kódující proteiny, byl v tomto normalizačním kroku počítán počet nsSNVs na délku genu kbp (z anglického „kilo base pair“). Byl vypočten GPR specifický pro každý nástroj (nsSNVs/kbp) a geny byly seřazeny podle hodnot GPR.

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 (viz Rozhodovací proces pro stanovení priority konečného souboru genů). Se záměrem snížit počet falešně pozitivních

předpovědi 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*).

Proces filtrace

Patogenní varianty předpovězené pomocí REVEL, MetaSVM, CADD a MutationAssessor byly rozděleny podle genů a dále zpracovány. Geny byly seřazeny podle počtu patogenních variant s nastaveným prahem pro zařazení minimálně 3 patogenních nsSNV na gen. Dále jsme tento proces omezili na maximálně deset genů. V případech, kdy geny měly na poslední pozici souboru (n=10) stejný počet patogenních variant na gen, byla uplatněna další podmínka. To umožnilo zahrnout více než 10 genů se stejným pořadím. Proces vyřazování byl rozdělen na úroveň 1 a úroveň 2. V úrovni 1 byla vyvolána konsenzuální predikce pro benigní varianty (tj. soubor „Benign“) a zredukována na úroveň genu (≥ 3 nsSNVs/gen, až 10 (+) genů stejného pořadí). Výsledné sady patogenních genů pro každý nástroj a sadu „Benign“ byly porovnány. Tímto krokem byla predikce patogenity překrývajících se genů u všech čtyř prediktorů vyřazena. V úrovni 2 byly z konečného výběru genů vyřazeny všechny jedinečné neboli unikátní geny napříč čtyřmi prediktory z důvodu pravděpodobně nízké prediktivní síly.

Normalizované hodnoty, genový patogenní poměr ("gene pathogenic ratio", GPR)

Délka genu byla získána pomocí platformu Galaxy (Galaxy verze 0.1.2) (Goecks et al. 2010) a bioDBnet (Mudunuri et al. 2009) a integrována do celkového souboru dat pomocí SQL. Galaxy počítá počet bází ve všech exonech genu po sloučení všech překrývajících se exonů z různých transkriptů.

Rozhodovací proces pro stanovení priority konečného souboru genů

Srovnání výsledků u čtyř prediktorů s přidáním údajů o PGx (viz 3.1.6) sloužilo jako výchozí bod pro stanovení souboru genů určených pro statistickou analýzu. Skóre = 1 bylo přiřazeno pro shodně označené geny v každém prediktoru včetně prediktorů s normalizovanými výsledky (n = 8).

Klasifikace důkazů byla následující: Velmi vysoká (skóre $\geq 8-7$ a PGx evidence), Vysoká (skóre $\geq 6-5$ metod), Střední (skóre $\geq 4-3$ a silná PGx evidence) nebo Nízká (skóre ≤ 2 a/nebo silná PGx evidence). "PGx" evidence byla přiřazena genům zařazeným do databáze PharmVar

a ClinVar (viz údaje o PGx) a "silná PGx evidence" v případě dodatečného zařazení genu do skupiny VIP genů dle PharmGKB. Rozhodnutí řízené umělou inteligencí (tj., nástroje strojového učení) představovalo nejvyšší poměr ve skórovacím systému (tj. 6/8 skóre). Predikce vyvolaná dvěma nástroji (n = 2) byla považována za nedostatečnou pro provedení statistické analýzy a tyto geny byly vyloučeny.

3.1.5. 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 níže.

Obsah informací v jednotlivých souborech zahrnoval a.) „Variants“ varianty registrované v dbSNP s přiřazenými identifikátory „rsID“ a registrovány a anotovány pomocí PharmGKB „PA(ID)“, b.) „Relationships“ informace na podkladu farmakokinetické/farmakodynamické souvislosti. Vztahy s označením PK nebo PD reprezentovali dvojice entit ve farmakokinetické, respektive ve farmakodynamické dráze dle PharmGKB, nebo pokud byla anotace například varianty, haplotypu, VIP genu, či léčiva anotována ve vztahu k PK/PD. Nepřítomnost PK/PD v těchto datech na druhou stranu neznamenala, že neexistuje žádný důkaz farmakokinetického nebo farmakodynamického vztahu, a proto všechny relevantní varianty byly námi dále manuálně anotovány z dodatečných databází.

Dále soubor, c.) „Clinical Variant Data“, umožnil nalezení informace ohledně dvojice varianty – lék a úroveň důkazů pro všechny klinické anotace, kdy v souboru „Clinical Metadata“ bylo možné dohledat míru evidence důkazů („Evidence count“) včetně odkazu na publikovanou literaturu. Soubor „Var_drug_ann“ pro informaci ohledně dávky léku, odpovědi na léčbu anebo metabolismu léčiva propojil kategorii sledovaného fenotypu (tj. účinnost", "toxicita", "dávkování", "metabolismus/PK", "PD") s binární signifikancí pozorovaného fenotypu (tj. „ano“ vs „ne“) a parametrů sledovaných v daných publikovaných studiích. Schéma v sekci výsledku (Obrázek 4—1) popisuje počet nalezených variant v souboru našich pacientek se statistickou korelací k odpovědi na léčbu a/nebo k přežití bez onemocnění (DFS). Získané informace z PharmGKB, poskytly míru funkční významnosti klinické a veřejně dostupné evidence pro prioritizaci variant začleněných do validačního souboru viz Tabulka 4—12 Seznam variant ověřených ve fázi validace (n = 7) v kapitole 4.1 (sloupec „Funkce“).

Bioinformatické zpracování dat z jednotlivých souborů a vytvoření interní databáze zprostředkovalo komplexní přehled dostupných farmakogenomických a genomických informací a splnilo tak jeden z cílů studia zefektivněním výběru kandidátních variant.

3.1.6. Farmakogenomická data – vzácné varianty

Pro získání farmakogenomických dat bylo provedeno vyhledávání variant souvisejících s PGx v databázích PharmVar a ClinVar. Mnohé z těchto variant byly ověřeny *in vitro* nebo *in vivo*, a jsou tak validnější než předpovědi z *in silico* nástrojů založených na variantách souvisejících s onemocněním, případně od těchto variant pouze odvozených.

Geny a exonické varianty (GRCh37) dostupné v databázi PharmVar (verze 5.2.14) byly porovnány s naším panelem genů. Následně byly k variantám přiřazeny webové informace, které definují funkční varianty podle Clinical Pharmacogenetics Implementation Consortium (CPIC) tj., podle důkazů o ovlivnění funkce genu jako "snížená funkce", "silně snížená" nebo "zvýšená funkce". Databáze PharmGKB (soubor clinicalAnnotations) poskytla informace o VIP genech. Z účelem automatické integrace dostupných dat do relační databáze MySQL byly vynechány geny s omezeným a absentujícím výskytem funkčního dopadu nebo údaje odkazující pouze na literaturu. Data ve formátu VCF z databáze ClinVar (verze 20180603) byla integrovaná do interní databáze a následně omezena na termín klinické významnosti PGx ve vztahu k odpovědi na léčbu "drug response" (obecný termín pro variantu, která ovlivňuje odpověď na léčivo, nikoliv onemocnění). Pro prioritní geny byla interní databáze navíc doplněna aktualizovaným vyhledáváním v Národním centru pro biotechnologické informace (NCBI).

3.1.7. 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 vybraní algoritmy nastaveny v souladu se studii výše tj., častých a vzácných variant.

3.1.8. 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 FILTREM = PASS (VCF).

3.1.8.1. Statistické metody pilotní a validační studie častých variant

Porovnání odpovědi na léčbu s ohledem na skupiny pacientek (častý homozygot, heterozygot a vzácný homozygot) a jednotlivé varianty bylo založeno na Pearsonova chí-kvadrát testu. Pro každou variantu a každý z těchto testů byla vypočtena upravená p-hodnota. Výpočet upravené p-hodnoty byl následující: (1) byla vypočtena p-hodnota založená na původních datech; (2) bylo vytvořeno 1000 permutací původních dat; (3) pro každou permutaci byla vypočtena hodnota testové statistiky; (4) byl vypočten podíl p-hodnot založených na permutovaných datech (1000 p-hodnot pro každý test), které byly vyšší nebo stejné než p-hodnota založená na původních datech; a (5) byla získána upravená p-hodnota pro danou variantu. Pro vícerozměrné analýzy byla použita binární logistická regrese.

Porovnání DFS s ohledem na skupiny pacientek (častý homozygot, heterozygot a vzácný homozygot) bylo provedeno pomocí log-rank testu pro jednotlivé varianty včetně vizuálního srovnání vytvořených grafů dle Kaplan-Meierovy metody. Vzhledem k tomu, že doba sledování studie byla stanovena na 120 měsíců (10 let), v případech, kdy byla hodnota DFS nebo OS u některého ze subjektů vyšší než 120 měsíců, byla hodnota pro tento subjekt stanovena na 120 měsíců a cenzorována. Upravená p-hodnota pro log-rank test byla založena na 100 permutacích původních dat. Metodika výpočtu upravené p-hodnoty pro každou variantu byla provedena podobným způsobem, jak bylo uvedeno výše. Za statisticky významnou byla považována p-hodnota menší než 0,05 po úpravě pro vícenásobné testování pomocí 100 permutací pro přežití a 1000 permutací pro odpověď na léčbu. Analýzy byly provedeny pomocí statistického programu SPSS v16.0 (SPSS, Chicago, IL, USA) a s použitím skriptu R.

Coxova regrese byla provedena také při analýze DFS zvlášť pro varianty, které byly statisticky významné na základě upravených p-hodnot. Proměnnou odpovědi byl DFS a proměnnou prediktoru byla varianta. Na základě testu poměru pravděpodobnosti pro testování statistické významnosti varianty byla zaznamenána p-hodnota. Poté byla provedena úprava na multiplicitu pouze pro tyto p-hodnoty. V úvahu přicházely následující úpravy: Bonferroniho metoda, Hochbergova metoda, Hommelova metoda, Holmova metoda a metoda míry falešného objevu (FDR). Analýza pro srovnání DFS s ohledem na skupiny pacientek byla provedena také pro podskupiny adjuvantně léčených pacientek a neoadjuvantně léčených pacientek. Metodika pro odvození upravených p-hodnot, Coxova regresní analýza a metody úpravy na multiplicitu byly

stejně jako uvedeno výše. Permutace původních dat (100 permutací) byly vytvořeny zvlášť pro adjuvantně léčené pacientky a zvlášť pro neoadjuvantně léčené pacientky.

3.1.8.2. Statistické metody studie proveditelnosti vzácných variant

Pro studii třetího cíle této práce byly asociace mezi prioritními geny a klinickými daty (nominální proměnné), včetně terapeutické odpovědi provedeny pomocí Pearsonova chí-kvadrát testu. V případě, že následující podmínka nebyla splněna ($\geq 20,0$ % hodnotových polí mělo očekávaný počet nižší než 5), byl aplikován oboustranný Fisherův přesný (exaktní) test. Pro geny *CYP2D6*, $n = 26$ pacientek, *DPYD* $n = 10$ pacientek byly do statistické analýzy zahrnuty všechny vzácné kódující varianty včetně synonymních, inzercí/deleci bez posunu čtecího rámce na rozdíl od ostatních genů.

Testované proměnné byly následující: velikost karcinomu (pTis/pT1 vs. pT2-4), nepřítomnost/přítomnost metastáz lymfatických uzlin (pN0 vs pN1-2), patologické stadium (SI vs. SII-III), stav exprese receptoru pro estrogen (ER, pozitivní vs. negativní) a progesteron (PR, pozitivní vs. negativní), stav exprese receptoru ERBB2 (pozitivní vs. negativní), exprese Ki67 (pozitivní vs. negativní na základě podílu buněk exprimujících Ki67 s mezní hodnotou 13,25 %) (Cheang et al. 2009), molekulární podtyp (luminální A vs. luminální B a trojitě negativní) a odpověď na neoadjuvantní chemoterapii (kompletní, resp. částečná patologická odpověď vs. stabilní onemocnění nebo progresse). Přežití bez onemocnění (DFS) bylo definováno jako doba, která uplynula mezi operací a relapsem. Celkové přežití (OS) bylo považováno za čas, který uplynul od operace do úmrtí pacienta. jednotlivé varianty včetně vizuálního srovnání vytvořených grafů dle Kaplan-Meierovy metody. Vzhledem k tomu, že doba sledování studie byla stanovena na 120 měsíců, v případech, kdy byla hodnota DFS nebo OS u některého ze subjektů vyšší než 120 měsíců, byla hodnota pro tento subjekt stanovena na 120 měsíců a cenzorována. DFS a OS pacientek byly hodnoceny log-rank testem. Coxova regresní analýza pro DFS a OS byla upravená na patologický grade (1 nebo 2 vs 3) nebo patologické stadium onemocnění (karcinom in situ nebo stadium I vs stadium II-IV). Hodnoty signifikance p jsou odchylkami od oboustranných testů. 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.

3.1.8.3. Statistické metody pro funkční predikci genu KIF14

Pro statistické hodnocení byly použity genetické modely. Aditivní model byl hodnocen pomocí logistické regrese a dominantní a recesivní model pomocí Pearsonova chí-kvadrát testu. Asociace mezi kategorizovanými hodnotami jako genotyp a klinickými údaji byly analyzovány pomocí Pearsonova chí-kvadrát testu a oboustranného Fisherova přesného testu. Testované klinické proměnné byly následující: menopauzální stav (pre- vs. postmenopauzální) stadium, velikost karcinomu (pT), postižení regionálních lymfatických uzlin (pN), histologický typ (invazivní duktální vs. jiné invazivní karcinomy), stupeň exprese estrogenu (ER), progesteronu (PR) a HER2/ ERBB2 a Ki67 (negativní vs. pozitivní). Odpověď na neoadjuvantní cytotoxickou léčbu byla hodnocena jako částečná nebo úplná odpověď (odpověď na terapii) versus stabilní nebo progresivní onemocnění (bez odpovědi). Bez progresu (PFS) bylo definováno jako doba, která uplynula mezi operací a relapsem. Celkové přežití (OS) bylo považováno za dobu, která uplynula od operace do úmrtí pacienta. Studie byl stanoven na 120 měsíců (10 let), a proto byly údaje o PFS a OS cenzurovány po 120 měsících. vyhodnoceny Kaplanovou-Meierovou metodou s Breslowovou metodou. testem. Dále byla provedena multivariační Coxova regresní analýza, která zahrnovala např. varianty signifikantní v jednorozměrných analýzách a faktory ovlivňující PFS a OS. Pacienti ztracení ze sledování byli vyloučeni z analýz PFS. Hodnoty p jsou odchylky z dvoustranných testů. Benjaminiho-Hochbergova míra falešného objevení (FDR) test byl použit pro korekci vícenásobného testování. všech nových asociací v hodnoceném souboru a $q < 0,05$ bylo považováno za signifikantní. Analýzy byly prováděny v statistickém programu SPSS v16.0 (SPSS, Chicago, IL, USA).

4. Výsledky

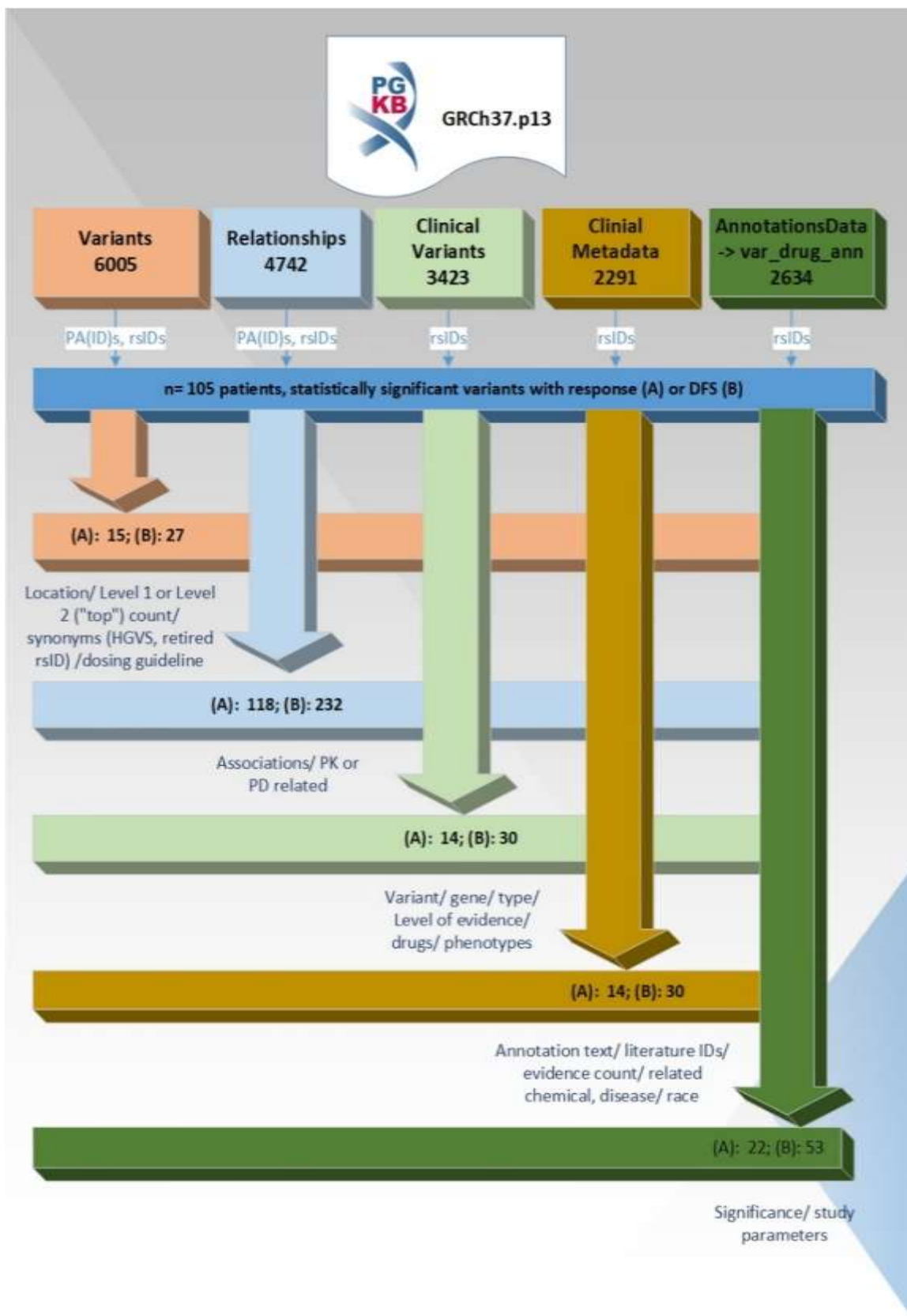
4.1. Výsledky pilotní studie častých variant – bioinformatická analýza

4.1.1. Datamining - farmakogenomické databáze

PharmGKB a PharmVar

Nedílnou součástí *in silico* predikce tvořila data získaná z farmakogenomických a klinických databází (Obrázek 4—1). Schéma popisuje počet nalezených variant uvedených v závorce v souboru našich pacientek se statistickou korelací k odpovědi na léčbu a/nebo k přežití bez onemocnění (DFS). 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 níže (viz Tabulka 4—3 až Tabulka 4—5).



Obrázek 4—1 Výsledek data mining procesu použitím interní databáze PharmGKBdownload pro statisticky významné varianty v odpovědi na léčbu (A) anebo DFS (B) ilustrován ve schématu

Tabulka 4—1 Vybrané varianty (n = 12) v klinickém PK/PD vztahu (protinádorové léčivo, nádor) dle PharmGKBdownload („Clinical Metadata“)

PGKB ID/rsID/Gen/Populace	Klinická anotace	Text anotace
PA166155049/ rs3743527/ <i>ABCC1</i> / Neznámá	Toxicita/ADR ("adverse drug reaction" = nežádoucí účinek)	CC:U pacientů s genotypem CC a kolorektálním nádorem může být závažnost neutropenie při užívání irinotekanu vyšší než u pacientů s genotypy CT a TT. Závažnost neutropenie při užívání irinotekanu mohou ovlivnit i další klinické a genetické faktory.", "CT:Patienti s genotypem CT a kolorektálním nádorem mohou mít při užívání irinotekanu zhoršenou závažnost neutropenie ve srovnání s pacienty s genotypem CC. Závažnost neutropenie při užívání irinotekanu mohou ovlivnit i další klinické a genetické faktory.", "TT:Patienti s genotypem TT a kolorektálním nádorem mohou mít při užívání irinotekanu pozdější závažnost neutropenie ve srovnání s pacienty s genotypem CC. Závažnost neutropenie při užívání irinotekanu mohou ovlivnit i další klinické a genetické faktory."
PA166157866/ rs2227291/ <i>ATP7A</i>/ Neznámá	Toxicita/ADR	CC:U pacientů s genotypem CC může být riziko toxicity docetaxelu a thalidomidu snižené, ale nikoli chybějící, ve srovnání s pacienty s genotypy CG nebo GG. Odpověď na léčbu mohou ovlivnit i další genetické a klinické faktory.", "CG:Patienti s genotypem CG mohou mít zvýšené riziko toxicity docetaxelu a thalidomidu ve srovnání s pacienty s genotypem CC. Odpověď na léčbu mohou ovlivnit i další genetické a klinické faktory.", "GG:Patienti s genotypem GG mohou mít zvýšené riziko toxicity docetaxelu a thalidomidu ve srovnání s pacienty s genotypem CC. Odpověď na léčbu mohou ovlivnit i další genetické a klinické faktory."
PA166153667/ rs2072671/ <i>CDA</i> / Evropská	Toxicita/ADR	AA:U pacientů s genotypem AA, kteří jsou léčeni cytarabinem, může být riziko toxicity snižené, nikoli chybějící, ve srovnání s pacienty s genotypem AC nebo CC. Riziko toxicity vyvolané léčivem mohou u pacienta ovlivnit i další genetické a klinické faktory.", "AC:Patienti s genotypem AC, kteří jsou léčeni cytarabinem, mohou mít zvýšené riziko toxicity ve srovnání s pacienty s genotypem AA. Další genetické a klinické faktory mohou rovněž ovlivnit riziko toxicity vyvolané léčivem.", "CC:Patienti s genotypem CC, kteří jsou léčeni cytarabinem, mohou mít zvýšené riziko toxicity ve srovnání s pacienty s genotypem AA. Riziko toxicity vyvolané léčivem mohou u pacienta ovlivnit i další genetické a klinické faktory."
PA166153793/ rs7543016/ <i>CMPK1</i> / Asijská	Účinnost	CC:Patienti s genotypem CC a karcinomem pankreatu mohou mít při léčbě gemcitabinem delší dobu progresu než pacienti s genotypem CG nebo GG. Dobu progresu u pacientů s karcinomem pankreatu mohou ovlivňovat i další genetické a klinické faktory, CG:Patienti s genotypem CG a karcinomem pankreatu mohou mít kratší dobu progresu při léčbě gemcitabinem ve srovnání s pacienty s genotypem CC. Na dobu progresu u pacientů s karcinomem pankreatu mohou mít vliv i další genetické a klinické faktory. ", "GG:Patienti s genotypem GG a karcinomem pankreatu mohou mít kratší dobu progresu při léčbě gemcitabinem ve srovnání s pacienty s genotypem CC. NA dobu progresu u pacientů s karcinomem pankreatu mohou mít vliv i další genetické a klinické faktory. "
PA166154017/ rs2515641/ <i>CYP2E1</i> / Neznámá	Toxicita/ADR	CC:U pacientů s genotypem CC může být při léčbě cytarabinem, fludarabinem, gemtuzumab ozogamicinem a idarubicinem zvýšená pravděpodobnost toxického onemocnění jater ve srovnání s pacienty s genotypem TT. Další genetické a klinické faktory mohou rovněž ovlivnit odpověď pacienta na cytarabin, fludarabin, gemtuzumab ozogamicin a idarubicin., "CT:Patienti s genotypem CT mohou mít při léčbě cytarabinem, fludarabinem, gemtuzumab ozogamicinem a idarubicinem sníženou pravděpodobnost toxického onemocnění jater ve srovnání s pacienty s genotypem CC a zvýšenou pravděpodobnost toxického onemocnění jater ve srovnání s pacienty s genotypem TT. Na odpověď pacienta na cytarabin, fludarabin, gemtuzumab ozogamicin a idarubicin mohou mít vliv i další genetické a klinické faktory.", "TT:U pacientů s genotypem TT může být při léčbě cytarabinem, fludarabinem, gemtuzumab ozogamicinem a idarubicinem ve srovnání s pacienty s genotypem CC snížená pravděpodobnost toxického onemocnění jater. Odpověď pacienta na cytarabin, fludarabin, gemtuzumab ozogamicin a idarubicin mohou ovlivnit i další genetické a klinické faktory."

PGKB ID/rsID/Gen/Populace	Klinická anotace	Text anotace
PA166153647/ rs1801160*/ DPYD/ Neznámá	Toxicita/ADR," Metabolismus/PK"	CC:U pacientů s genotypem CC (<i>DPYD</i> *1/*1) a nádorovým onemocněním, kteří jsou léčeni chemoterapií založenou na fluoropyrimidínech, může být 1) zvýšený metabolismus fluorouracilu a 2) snížené riziko toxicity léku ve srovnání s pacienty s genotypem CT nebo TT (<i>DPYD</i> *1/*6 nebo *6/*6). Fluoropyrimidiny se často používají v kombinované chemoterapii, jako je FOLFOX (fluorouracil, leukovorin a oxaliplatin), nebo s dalšími léčivými, jako je bevacizumab, cetuximab, raltitrexed. Odpověď na chemoterapii založenou na fluoropyrimidínech mohou ovlivnit i další genetické a klinické faktory.", "CT:Pacienti s genotypem CT (<i>DPYD</i> *1/*6) a karcinomem, kteří jsou léčeni chemoterapií založenou na fluoropyrimidínech, mohou mít 1) snížený metabolismus fluorouracilu a 2) zvýšené riziko toxicity léku ve srovnání s pacienty s genotypem CC (<i>DPYD</i> *1/*1). Fluoropyrimidiny se často používají v kombinované chemoterapii, jako je FOLFOX (fluorouracil, leukovorin a oxaliplatin), nebo s dalšími léky, jako je bevacizumab, cetuximab, raltitrexed. Odpověď na chemoterapii založenou na fluoropyrimidínech mohou ovlivnit i další genetické a klinické faktory.", "TT:Pacienti s genotypem TT (<i>DPYD</i> *6/*6) a nádorovým onemocněním, kteří jsou léčeni chemoterapií založenou na fluoropyrimidínech, mohou mít 1) snížený metabolismus fluorouracilu a 2) zvýšené riziko toxicity léku při léčbě ve srovnání s pacienty s genotypem CC (<i>DPYD</i> *1/*1). Fluoropyrimidiny se často používají v kombinované chemoterapii, jako je FOLFOX (fluorouracil, leukovorin a oxaliplatin), nebo s dalšími léky, jako je bevacizumab, cetuximab, raltitrexed. Odpověď na chemoterapii založenou na fluoropyrimidínech mohou ovlivnit i další genetické a klinické faktory."
PA166153874/ rs17376848/ DPYD/ Neznámá	Toxicita/ADR	AA:Pacienti s genotypem AA a nádorovým onemocněním mohou mít při léčbě chemoterapií na bázi fluoropyrimidinů nižší riziko lékové toxicity ve srovnání s pacienty s genotypy AG a GG. Jiné studie nepopisují asociaci s lékovou toxicitou a chemoterapií na bázi fluoropyrimidinu. Fluoropyrimidiny se často používají v kombinované chemoterapii, jako je FOLFOX (fluorouracil, leukovorin a oxaliplatin). Odpověď na chemoterapii založenou na fluoropyrimidínech mohou ovlivnit i další genetické a klinické faktory.", "AG:Pacienti s genotypem AG a karcinomem mohou mít při léčbě chemoterapií založenou na fluoropyrimidínech zvýšené riziko lékové toxicity ve srovnání s pacienty s genotypy AA nebo GG. Jiné studie nepopisují žádnou asociaci s lékovou toxicitou a chemoterapií na bázi fluoropyrimidinu. Fluoropyrimidiny se často používají v kombinované chemoterapii, jako je FOLFOX (fluorouracil, leukovorin a oxaliplatin). Odpověď na chemoterapii založenou na fluoropyrimidínech mohou ovlivnit i další genetické a klinické faktory.", "GG:Pacienti s genotypem GG a nádorovým onemocněním mohou mít při léčbě chemoterapií založenou na fluoropyrimidínech snížené riziko lékové toxicity ve srovnání s pacienty s genotypy AA a AG. Jiné studie nepopisují žádnou asociaci s lékovou toxicitou a chemoterapií na bázi fluoropyrimidinu. Fluoropyrimidiny se často používají v kombinované chemoterapii, jako je FOLFOX (fluorouracil, leukovorin a oxaliplatin). Odpověď na chemoterapii založenou na fluoropyrimidínech mohou ovlivnit i další genetické a klinické faktory."
PA166153696/ rs2297595/ DPYD/ Neznámá	Toxicita/ADR, "Metabolismus/PK"	CC:Pacienti s genotypem CC mohou mít 1) zvýšené riziko závažné toxicity při léčbě fluoropyrimidiny a 2) snížený metabolismus fluorouracilu ve srovnání s pacienty s genotypem TT. Pro tyto asociace však existují rozporuplné nebo negativní výsledky. Fluoropyrimidiny se často používají v kombinované chemoterapii, jako je FOLFOX (fluorouracil, leukovorin a oxaliplatin), nebo s dalšími léky, jako je bevacizumab, cetuximab, raltitrexed. Odpověď na chemoterapii založenou na fluoropyrimidínech mohou ovlivnit i další genetické a klinické faktory. „CT:Pacienti s genotypem CT mohou mít 1) zvýšené riziko závažné toxicity při léčbě fluoropyrimidiny a 2) snížený metabolismus fluorouracilu ve srovnání s pacienty s genotypem TT. Pro tyto asociace však existují rozporuplné nebo negativní výsledky. Fluoropyrimidiny se často používají v kombinované chemoterapii, jako je FOLFOX (fluorouracil, leukovorin a oxaliplatin), nebo s dalšími léky, jako je bevacizumab, cetuximab, raltitrexed. Odpověď na chemoterapii založenou na fluoropyrimidínech mohou ovlivnit i další genetické a klinické faktory. ", "TT:Pacienti s genotypem TT mohou mít 1) snížené, ale nikoliv chybějící riziko závažné toxicity při léčbě fluoropyrimidiny a 2) zvýšený metabolismus fluorouracilu ve srovnání s pacienty s genotypem CC nebo CT. Pro tyto

PGKB ID/rsID/Gen/Populace	Klinická anotace	Text anotace
		asociace však existují rozporuplné nebo negativní výsledky. Fluoropyrimidiny se často používají v kombinované chemoterapii, jako je FOLFOX (fluorouracil, leukovorin a oxaliplatin), nebo s dalšími léčivými, jako je bevacizumab, cetuximab, raltitrexed. Odpověď na chemoterapii založenou na fluoropyrimidinech mohou ovlivnit i další genetické a klinické faktory. "
PA166154805/ rs4986938/ ESR2/ Neznámá	Toxicita/ADR	CC:U pacientek s genotypem CC může dojít k vyššímu zvýšení triglyceridů vyvolanému tamoxifenem u žen po menopauze ve srovnání s pacientkami s genotypem TT nebo CT. Odpověď na tamoxifen mohou ovlivňovat i další genetické a klinické faktory.", "CT: Pacientky s genotypem CT mohou mít nižší tamoxifenem indukované zvýšení triglyceridů u postmenopauzální ženy ve srovnání s pacientkami s genotypem CC. Odpověď na tamoxifen mohou ovlivnit i další genetické a klinické faktory.", "TT:Pacientky s genotypem TT mohou mít nižší tamoxifenem indukované zvýšení triglyceridů u postmenopauzálních žen ve srovnání s pacientkami s genotypem CC. Odpověď na tamoxifen mohou ovlivňovat i další genetické a klinické faktory."
PA166157094/ rs3957357/ GSTA1, GSTA6P/ Neznámá	Toxicita/ADR	AA:Pacienti s genotypem AA mohou mít při léčbě cisplatinou a cyklofosfamidem zvýšené riziko anémie ve srovnání s pacienty s genotypem GG. Odpověď na režimy cisplatinu mohou ovlivnit i další genetické a klinické faktory.", "AG:Pacienti s genotypem AG mohou mít zvýšené riziko anémie při léčbě cisplatinou a cyklofosfamidem ve srovnání s pacienty s genotypem GG. Odpověď na režimy cisplatinu mohou ovlivnit i další genetické a klinické faktory.", "GG:Pacienti s genotypem GG mohou mít snížené riziko anémie při léčbě cisplatinou a cyklofosfamidem ve srovnání s pacienty s genotypem AA nebo AG. Odpověď na režimy s cisplatinou mohou ovlivnit i další genetické a klinické faktory."
PA166154249/ rs1695/ GSTP1/ Evropská	Toxicita/ADR	AA:Pacienti s genotypem AA a nádorovým onemocněním, kteří jsou léčeni fluorouracilem, mohou mít vyšší riziko hematologické toxicity ve srovnání s pacienty s genotypem GG. Riziko hematologické toxicity u pacienta vystaveného fluorouracilu mohou ovlivnit i další genetické a klinické faktory,AG:Pacienti s genotypem AG a karcinomem, kteří jsou léčeni fluorouracilem, mohou mít nižší, ale nikoliv chybějící riziko hematologické toxicity ve srovnání s pacienty s genotypem AA, nebo mohou mít vyšší riziko hematologické toxicity ve srovnání s pacienty s genotypem GG. Riziko hematologické toxicity u pacienta vystaveného působení fluorouracilu mohou ovlivnit i další genetické a klinické faktory.", "GG:Pacienti s genotypem GG a karcinomem, kteří jsou léčeni fluorouracilem, mohou mít nižší, ale ne žádné, riziko hematologické toxicity ve srovnání s pacienty s genotypem AA. Riziko hematologické toxicity u pacienta vystaveného působení fluorouracilu mohou ovlivnit i další genetické a klinické faktory."
PA166155965/ rs10485828/ TUBB1/ Neznámá	Jiné	CC:Pacientky s genotypem CC a ER a/nebo PR pozitivním karcinomem prsu mohou mít snížené riziko zlomenin kostí při užívání inhibitorů aromatázy anastrozolu a exemestanu ve srovnání s pacientkami s genotypem CG nebo GG. Riziko zlomenin kostí u osob s ER a/nebo PR pozitivním karcinomem prsu při expozici exemestanu nebo anastrozolu mohou ovlivňovat i další klinické a genetické faktory,CG:Pacientky s genotypem CG a ER a/nebo PR pozitivním karcinomem prsu mohou mít zvýšené riziko zlomenin kostí při expozici inhibitorům aromatázy anastrozolu a exemestanu ve srovnání s pacientkami s genotypem CC a snížené riziko zlomenin kostí ve srovnání s pacientkami s genotypem GG. Riziko zlomenin kostí u osob s ER a/nebo PR pozitivním karcinomem prsu při expozici exemestanu nebo anastrozolu mohou ovlivnit i další klinické a genetické faktory.", "GG:Pacientky s genotypem GG a ER a/nebo PR pozitivním karcinomem prsu mohou mít zvýšené riziko zlomenin kostí při expozici inhibitorům aromatázy anastrozolu a exemestanu ve srovnání s pacientkami s genotypem CG nebo GG. Riziko zlomenin kostí u osob s ER a/nebo PR pozitivním karcinomem prsu při expozici exemestanu nebo anastrozolu mohou ovlivňovat i další klinické a genetické faktory."

Poznámky pod čarou: Klinicky validované varianty na vzorku populace 805 pacientek s nádorem prsu je označena tučným písmem. *Varianta rs1801160 (c.2194G>A) byla detekována s jinou záměnou alel nežli u pacientů z dat PharmGKB (CC/CT/TT). Varianta rs1801160 je považovaná za experimentálně i klinicky validovanou což dokládá studie, kde byla zařazena mezi 24 funkčních variant farmakogenů (Kozyra et al. 2017) k testování výkonosti prediktorů.

Tabulka 4—2 Výsledky pro data mining z integrované databáze „PharmVarDownload“ pro gen CYP2D6, haplotypy pro 2 varianty (rs1065852, rs28371725), #version=pharmvar-1.1.9.2

haplotyp	Gen	rsID	refsequence	start	stop	REF	ALT	konsekvence
<i>CYP2D6*4A</i>	<i>CYP2D6</i>	rs1065852	NC_000022.10	42526694	42526694	G	A	substituce
<i>CYP2D6*4B</i>	<i>CYP2D6</i>	rs1065852	NC_000022.10	42526694	42526694	G	A	substituce
<i>CYP2D6*4C</i>	<i>CYP2D6</i>	rs1065852	NC_000022.10	42526694	42526694	G	A	substituce
<i>CYP2D6*4D</i>	<i>CYP2D6</i>	rs1065852	NC_000022.10	42526694	42526694	G	A	substituce
<i>CYP2D6*4E</i>	<i>CYP2D6</i>	rs1065852	NC_000022.10	42526694	42526694	G	A	substituce
<i>CYP2D6*4F</i>	<i>CYP2D6</i>	rs1065852	NC_000022.10	42526694	42526694	G	A	substituce
<i>CYP2D6*4G</i>	<i>CYP2D6</i>	rs1065852	NC_000022.10	42526694	42526694	G	A	substituce
<i>CYP2D6*4H</i>	<i>CYP2D6</i>	rs1065852	NC_000022.10	42526694	42526694	G	A	substituce
<i>CYP2D6*4J</i>	<i>CYP2D6</i>	rs1065852	NC_000022.10	42526694	42526694	G	A	substituce
<i>CYP2D6*4K</i>	<i>CYP2D6</i>	rs1065852	NC_000022.10	42526694	42526694	G	A	substituce
<i>CYP2D6*4L</i>	<i>CYP2D6</i>	rs1065852	NC_000022.10	42526694	42526694	G	A	substituce
<i>CYP2D6*4N</i>	<i>CYP2D6</i>	rs1065852	NC_000022.10	42526694	42526694	G	A	substituce
<i>CYP2D6*4P</i>	<i>CYP2D6</i>	rs1065852	NC_000022.10	42526694	42526694	G	A	substituce
<i>CYP2D6*10A</i>	<i>CYP2D6</i>	rs1065852	NC_000022.10	42526694	42526694	G	A	substituce
<i>CYP2D6*10B</i>	<i>CYP2D6</i>	rs1065852	NC_000022.10	42526694	42526694	G	A	substituce
<i>CYP2D6*10D</i>	<i>CYP2D6</i>	rs1065852	NC_000022.10	42526694	42526694	G	A	substituce
<i>CYP2D6*14A</i>	<i>CYP2D6</i>	rs1065852	NC_000022.10	42526694	42526694	G	A	substituce
<i>CYP2D6*36</i>	<i>CYP2D6</i>	rs1065852	NC_000022.10	42526694	42526694	G	A	substituce
<i>CYP2D6*37</i>	<i>CYP2D6</i>	rs1065852	NC_000022.10	42526694	42526694	G	A	substituce
<i>CYP2D6*41</i>	<i>CYP2D6</i>	rs28371725	NC_000022.10	42523805	42523805	C	T	substituce
<i>CYP2D6*47</i>	<i>CYP2D6</i>	rs1065852	NC_000022.10	42526694	42526694	G	A	substituce
<i>CYP2D6*49</i>	<i>CYP2D6</i>	rs1065852	NC_000022.10	42526694	42526694	G	A	substituce
<i>CYP2D6*52</i>	<i>CYP2D6</i>	rs1065852	NC_000022.10	42526694	42526694	G	A	substituce
<i>CYP2D6*54</i>	<i>CYP2D6</i>	rs1065852	NC_000022.10	42526694	42526694	G	A	substituce
<i>CYP2D6*56B</i>	<i>CYP2D6</i>	rs1065852	NC_000022.10	42526694	42526694	G	A	substituce
<i>CYP2D6*57</i>	<i>CYP2D6</i>	rs1065852	NC_000022.10	42526694	42526694	G	A	substituce
<i>CYP2D6*64</i>	<i>CYP2D6</i>	rs1065852	NC_000022.10	42526694	42526694	G	A	substituce
<i>CYP2D6*65</i>	<i>CYP2D6</i>	rs1065852	NC_000022.10	42526694	42526694	G	A	substituce
<i>CYP2D6*69</i>	<i>CYP2D6</i>	rs28371725	NC_000022.10	42523805	42523805	C	T	substituce
<i>CYP2D6*69</i>	<i>CYP2D6</i>	rs1065852	NC_000022.10	42526694	42526694	G	A	substituce
<i>CYP2D6*72</i>	<i>CYP2D6</i>	rs1065852	NC_000022.10	42526694	42526694	G	A	substituce
<i>CYP2D6*87</i>	<i>CYP2D6</i>	rs1065852	NC_000022.10	42526694	42526694	G	A	substituce
<i>CYP2D6*91</i>	<i>CYP2D6</i>	rs28371725	NC_000022.10	42523805	42523805	C	T	substituce
<i>CYP2D6*94A</i>	<i>CYP2D6</i>	rs1065852	NC_000022.10	42526694	42526694	G	A	substituce

haplotyp	Gen	rsID	refsequence	start	stop	REF	ALT	konsekvence
<i>CYP2D6</i> *94B	<i>CYP2D6</i>	rs1065852	NC_000022.10	42526694	42526694	G	A	substituce
<i>CYP2D6</i> *95	<i>CYP2D6</i>	rs1065852	NC_000022.10	42526694	42526694	G	A	substituce
<i>CYP2D6</i> *99	<i>CYP2D6</i>	rs1065852	NC_000022.10	42526694	42526694	G	A	substituce
<i>CYP2D6</i> *100	<i>CYP2D6</i>	rs1065852	NC_000022.10	42526694	42526694	G	A	substituce
<i>CYP2D6</i> *101	<i>CYP2D6</i>	rs1065852	NC_000022.10	42526694	42526694	G	A	substituce

Kombinaci získaných dat z PharmGKB a PharmVar databází pro statisticky významné varianty a v konkordanci s metaprediktorem CADD bylo vybráno celkově 28 variant do fáze validace (viz Tabulka 4—3).

Tabulka 4—3 Vybrané PGx (PharmGKB, PharmVar) varianty pro fázi validace (n = 28) a zároveň predikované jako škodlivé dle nástroje CADD (n = 8)

Gen	HGVS	HGVS protein	Klasifikace Annovar ¹	rsID	ClinVar	DFS ²	Odpověď na léčbu ²	Funkce ³	AF ⁴	ExAC ⁵	NCMG ⁶
<i>ABCB1</i>	.	.	intron	rs2032583	.	0,03		PA166157317	0,106	0,123	0,093
<i>ABCC1</i>	c.*1512T>C	.	UTR3	rs212091	.	0,05		PA166154987	0,18	.	0,114
<i>ABCC1</i>	c.*543C>T	.	UTR3	rs3743527	.	0,05		PA166155049	0,205	.	.
<i>AKR7A2</i>	c.424G>A	p.A142T	NS	rs1043657	.	0,01		PA166161794; CADD	0,095	0,093	0,081
<i>AKT1</i>	.	.	intron	rs3803304	.	0,05		PA166154802; 1f	0,292	.	0,289
<i>ATP7A</i>	c.2299G>C	p.V767L	NS	rs2227291	benigní		0,003	PA166157866; CADD	0,26	0,217	0,225
<i>CDA</i>	c.79A>C	p.K27Q	NS	rs2072671	.	0,01		PA166153667	0,362	0,343	0,296
<i>CES1</i>	c.-75T>G	.	UTR5	rs3815583	.		0,038	PA166155058	0,202	.	0,162
<i>CES1</i>	c.224G>A	p.S75N	NS	rs2307240	.	0,01		PA166155039	0,067	0,054	0,063
<i>CMPK1</i>	c.22G>C	p.G8R	NS	rs7543016	.	0,05		PA166153793	0,451	0,538	0,32
<i>CYP2C9</i>	.	.	intron	rs1934969	.	0,03		PA166153986	0,613	.	0,658
<i>CYP2D6</i>	c.100C>T	p.P34S	NS	rs1065852	pravděp. benigní		0,021	PA166156062; PharmVar; CADD	0,214	0,249	0,206
<i>CYP2D6</i>	c.985+39G>A	.	ncRNA_intron	rs28371725	.		0,011	PA166156155; PharmVar	0,059	0,095	0,084
<i>CYP2E1</i>	c.1263T>C	p.F421F	synonymní	rs2515641	benigní	0,03		PA166154017; CADD	0,856	0,887	0,88

Gen	HGVS	HGVS protein	Klasifikace Annovar ¹	rsID	ClinVar	DFS ²	Odpověď na léčbu ²	Funkce ³	AF ⁴	ExAC ⁵	NCMG ⁶
DPYD	c.2194G>A	p.V732I	NS	rs1801160	pravděp. benigní	0,02		PA166153647; CADD	0,052	0,047	0,045
DPYD	c.1896T>C	p.F632F	synonymní	rs17376848	pravděp. benigní	0,03		PA166153874; CADD	0,057	0,037	0,047
DPYD	c.496A>G	p.M166V	NS	rs2297595	odpověď na léčivo			PA166153696; D, CADD	0,148	0,103	0,125
DPYS	.	.	intron	rs2669429	.		0,01	PA166157579	0,54	.	0,557
ESR2	c.*39G>A	.	UTR3	rs4986938	.		0,05	PA166154805	0,371	0,379	0,347
GSTAI	c.-9630T>C	.	UTR5	rs3957357	.		0,047	PA166157094	0,591	.	.
GSTA2	c.335G>C	p.S112T	NS	rs2180314	.		0,017	PA166157020	0,592	0,59	0,581
GSTP1	c.313A>G	p.I105V	NS	rs1695	odpověď na léčivo	0,05		PA166154249	0,329	0,319	0,32
PPARG	c.1347C>T	p.H449H	synonymní	rs3856806	pravděp. benigní		0,046	PA166156388	0,12	0,125	0,133
SLC22A1	c.1222A>G	p.M408V	NS	rs628031	.		0,028	PA166156933	0,6	0,592	0,605
SLC28A3	c.338A>G	p.Y113C	NS	rs10868138	.		0,022	PA166157820; CADD	0,067	0,085	0,091
SLC2A1	c.*462G>C	.	UTR3	rs4658	benigní	0,01		PA166153544	0,21	.	0,178
SLCO1A2	c.-189_-188insA	.	UTR5	rs3834939	.	0,05		PA166163600	0,295	.	.
TUBB1	c.*817G>C	.	UTR3	rs10485828	.	0,03		PA166155965	0,212	.	.

Poznámky pod čarou: ¹NS = nesynonymní; ²p-hodnota uvedená pro klinické asociace; ³Předpověď založená na kombinaci farmakogenomických databází PharmGKB (označení "PA" znamená konkrétní onemocnění, geny a léčiva v databázi), PharmVar a *in silico* nástrojů: CADD v případech škodlivé predikce; MetaLR/SVM („D“ = deleterious = škodlivá), ⁴MAF = hlavní frekvence alel v testovacím souboru; ⁵Exome Aggregation Consortium, alelické frekvence v Evropské (bez Finské) populaci; ⁶Národní centrum lékařské genomiky, alelické frekvence v běžné České populaci. Varianty označené *kurzívou* nebylo možné validovat z důvodu selhávání detekční techniky i navzdory několika pokusům optimalizace metody. Klinicky validované varianty na vzorku populace 805 pacientek s nádorem prsu jsou označena **tučným** písmem.

Další využitou farmakogenomickou databází byla, an Adverse Drug Reaction Classification System-Target, ADRCS-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. Filtrace dat následovala výběr léčiv v souladu s farmakoterapií administrovanou pacientkám naší studie. Varianty uvedené v tabulce níže (viz Tabulka 4—4) jsou omezeny pouze na statisticky korelované varianty ve vztahu k léčbě anebo délce přežití.

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 (viz Tabulka 4—5). Pro vybrané varianty databáze neuváděla fenotypovou asociaci ve statistické významnosti na hladině alfa 5%.

ADReCS-Target

Tabulka 4—4 Výsledky pro data mining z integrované databáze „ADReCS-Target“ pro variantu (rs1695) pro gen *GSTP1*

ID varianty	ADR (“adverse drug reaction”) Termín	Toxicita dle ADReCS-Target	léčivo	Organismus
rs1695	Porucha krve	Hematologická toxicita, neutropenie	fluorouracil	člověk
rs1695	Tinnitus	Tinnitus, porucha sluchu, Raynaudův syndrom	cisplatina	člověk
rs1695	Gastrointestinální toxicita	Hematologická toxicita, gastrointestinální toxicita	fluorouracil	člověk
rs1695	Gastrointestinální toxicita	Významné zvýšení produkce tumor nekrotizujícího faktoru alfa a snížení hladin glutathionu a aktivity glutathionreduktázy po léčbě racemátem a (R)-(-)-ketoprofenem	fluorouracil	potkan

Varianta rs1695 je považovaná za experimentálně i klinicky validovanou což potvrzuje studie, kde byla zařazena mezi 24 funkčních variant farmakogenů (Kozyra et al. 2017) k testování prediktivních metod.

PheWAS

Tabulka 4—5 Databáze PheWAS obsahovala jednu z prioritizovaných variant bez statistické korelace k antineoplastické léčbě způsobující nežádoucí účinky v přepočtu na 205 případů.

basepair	gene	rsID	phewas_string	cases	odds_ratio	p
NULL	<i>BAKI</i>	rs210134	Antineoplastické a imunosupresivní léčiva způsobující nežádoucí účinky	205	0,9654	0,7531

4.1.2. *In silico* predikce

Sekvenční sada nástrojů

Pro funkční predikci nesynonymních variant byla sestavena samostatná sada nástrojů využívající algoritmy sekvenčních srovnání. Algoritmus sady není založen na metodách strojového učení, a překonává tak obavy z kruhovosti typu 2 v důsledku nedostatečného rozlišení či nepoměru škodlivých od neutrálních variant daného proteinu a genu v trénovacím souboru dat (Grimm et al. 2015). Pro interpretaci predikce byl použit přísný konsenzus ve výsledcích všech nástrojů za vyloučení všech variant s chybějící predikcí minimálně v jednom z nástrojů. Tato sada zahrnuje následující softwarové nástroje a podmínku pro predikci, případně mezní hodnotu uvedenou v závorce: MutationAssessor (H/M = funkční), SIFT (D: škodlivá $\leq 0,05$), LRT (D: „škodlivá“) a Provean (D: „škodlivá“).

Tabulka 4—6 Varianty predikované *in silico* sekvenční sadou pro fázi validace (n = 3)

a.) frekvence ve 3 populačních datech (pacientky studie (MAF), obecná Evropská (ExAC) a Česká populace (NCMG))

Gen	HGVS	HGVS protein	Klasifikace Annovar ¹	rsID	DFS ²	Odpověď na léčbu ²	MAF ³	ExAC ⁴	NCMG ⁵
<i>EPHX2</i>	c.662G>A	p.R221Q	NS	rs751141	0,03		0,114	0,095	0,115
<i>PIK3C2G</i>	c.2732C>T	p.P911L	NS	rs12312266	0,04		0,205	0,298	0,242
<i>UGT2A1</i>	c.949G>A	p.G317R	NS	rs4148301		0,033	0,11	0,096	0,104

Poznámky pod čarou: ¹NS = nesynonymní; ²p-hodnota uvedená pro klinické asociace; ³MAF = hlavní frekvence alel v testovaném souboru; ⁴Exome Aggregation Consortium, alelické frekvence v Evropské (bez Finské) populaci; ⁵Národní centrum lékařské genomiky, alelické frekvence v běžné České populaci

Tabulka 4—7 Varianty predikované *in silico* sekvenční sadou pro fázi validace (n = 3)

b.) výsledky jednotlivých nástrojů

rsID	Pozice varianty	Alela	SIFT (skóre)	LRT	MutationAssessor	PROVEAN	SIFT
rs751141	8:27373865-27373865	A	škodlivá (0,03)	D	M	D	D
rs12312266	12:18649057-18649057	T	škodlivá (0,03)	D	M	D	D
rs4148301	4:70462042-70462042	T	škodlivá (0)	D	H	D	D

Poznámky pod čarou: D=škodlivá, M=střední a H=vysoká hodnota pravděpodobnosti že je varianta škodlivá

TargetScan, Meta SVM/LR, CADD, Nexus IW score, RegulomeDB a dbSNV

Výběr *in silico* nástrojů byl založen na rozsahu predikce pro daný software se záměrem zajistit predikci pro všechny typy variant v našem souboru, tj. pro kódující i nekódující. Biologické cíle miRNA a konzervovaná místa dané varianty (3'-nepřepisovaný konec - UTR3) byly porovnány pomocí programu TargetScan (verze 7.2, Cambridge, MA, USA). Pro nesynonymní varianty a varianty sestřihového místa tzv. „splice site“ poskytl Annovar binární predikci pomocí ansámblového skóre MetaSVM/LR a dbSNV skóre. Nástroj CADD v1.3 (mezní hodnota ≥ 19 , VEP) byl zvolen pro všechny typy variant (tj. kódující, nekódující SNV a krátké inserce/delece). CADD poskytl společně s metaprediktorem MetaSVM/LR doplňkové skóre k souboru nsSNVs vzhledem k omezení výběru predikce pouze sekvenční sadou, která opomíjí na rozdíl od metaprediktorů predikci strukturální změny v proteinu. Známé regulační prvky v mezigenových („intergenic“) oblastech např. „DNázová“, tj., endonukleázová hypersenzitivita, vazebná místa transkripčních faktorů a promotorové oblasti, které byly biochemicky charakterizovány jako důležité pro regulaci transkripce, byly předpovězeny za škodlivé pomocí skóre nástroje RegulomeDB (Boyle et al. 2012) za podmínky splňující klasifikační kategorii 1. Tyto varianty byly považovány za pravděpodobně ovlivňující vazbu na transkripční faktory a expresi cílového genu. Dále byl použit komplexní webový rámec IW-Scoring (<https://snp-nexus.org/IW-Scoring/>) (Wang et al. 2018b). IW skóre (p-hodnota $\leq 0,05$) určuje škodlivost pro známé (IW skóre K11) a nové (IW skóre N8) nekódující a kódující synonymní varianty.

Tabulka 4—8 Vybrané varianty pro fázi validace (n = 22) predikované za škodlivé dle *in silico* nástrojů (bez sekvenčního setu) případně dle klinických dat v ClinVar (n = 1)

Gen	HGVS	HGVS protein	Klasifikace Annovar ¹	rsID	ClinVar	DFS ²	Odpověď na léčbu ²	Funkce ³	AF ⁴	ExAC ⁵	NCMG ⁶
<i>ABCA13</i>	c.5603A>T	p.N1868I	NS	rs1801466	pravděp. benigní	0,03		D; CADD	0,105	0,066	0,065
<i>ABCB8</i>	.	.	intron	rs3214587	.	0,01		miR-670-3p	0,115	.	.
<i>ABCC3</i>	.	.	intron	rs4148413	.		0,002	1f, PINES	0,168	.	.
<i>ABCC6</i>	c.2835C>T	p.P945P	synonymní	rs2856585	škodlivá	0,03		ClinVar	0,064	0,099	0,044
<i>AHRR</i>	.	.	intron	rs2013782	.	0,02		1f	0,587	0,623	0,597
<i>BAK1</i>	dist=114	.	downstream	rs210134	.		0,033	1f, PINES, PheWAS	0,75	.	.
<i>BIRC7</i>	c.528C>T	p.S176S	synonymní	rs2273487	.	<0,001		1b	0,486	0,467	0,464
<i>BLK</i>	c.-53667C>T	.	UTR5	rs922483	benigní		0,023	1f	0,229	.	.
<i>CYP2E1</i>	.	.	intron	rs2070677	.	0,05		1f	0,856	.	0,867
<i>ENOSF1</i>	.	.	intron	rs2612083	.	0,01		1f	0,381	.	0,322
<i>EPHX2</i>	dist=55	.	downstream	rs4149259	.	0,03		1f	0,167	.	.
<i>GSTP1</i>	.	.	intron	rs762803	.	<0,001		1f; PINES	0,386	0,40	0,323
<i>IRS1</i>	c.*4476A>G	.	UTR3	rs2288587	.	0,05		1f	0,057	.	.
<i>MADD</i>	.	.	intron	rs10501320	.	0,05		IW; PINES	0,281	.	.
<i>PIP4K2B</i>	.	.	intron	rs2075061	.	0,02		1f	0,605	.	0,592
<i>PPARA</i>	c.*5977G>A	.	UTR3	rs9626814	.		0,019	1d	0,101	.	.
<i>RALBP1</i>	c.*756G>A	.	UTR3	rs3322	.	0,03		1f	0,095	0,09	0,072
<i>RARB</i>	c.*1287T>G	.	UTR3	rs1058378	.		0,013	IW, miR-665	0,091	.	0,083

Gen	HGVS	HGVS protein	Klasifikace Annovar ¹	rsID	ClinVar	DFS ²	Odpověď na léčbu ²	Funkce ³	AF ⁴	ExAC ⁵	NCMG ⁶
<i>RPTOR</i>	c.90T>C	p.F30F	synonymní	rs61750765	.	0,03		IW	0,238	0,12	0,142
<i>RRAGD</i>	c.*1105 T>A	.	UTR3	rs1555403	.		0,019	1f	0,238	.	.
<i>DFFB</i>	.	.	intron	rs4376673	.	0,04	0,047	PINES DFS + odpověď	0,909	0,947	0,901
NR5A2	dist=45	.	upstream	rs2816948	.	0,05	0,036	PINES DFS + odpověď	0,13	.	.
<i>KCNABI</i>	.	.	intron	rs2293194	.	0,04	0,013	PINES, DFS + odpověď	0,476	.	.

Poznámky pod čarou: ¹NS = nesynonymní; ²p-hodnota uvedená pro klinické asociace; ³Předpověď založená na kombinaci farmakogenomických databází, např. PheWAS (p-hodnota = 0,75), ClinVar (uveden v případech škodlivé predikce a nástrojů *in silico* TargetScan (miRNA např. miR-665), MetaLR/SVM (D = škodlivá), CADD (uveden v případech škodlivé predikce), Nexus IW (IW, p-hodnota ≤ 0,05); nebo RegulomeDB (značení kategorie „číslo a písmeno“), PINES (p-hodnota ≤ 0,05); ⁴MAF = hlavní frekvence alel v testovacím souboru; ⁵Exome Aggregation Consortium, alelické v Evropské (bez Finské) populaci; ⁶Národní centrum lékařské genomiky, alelické frekvence v běžné České populaci, *nesignifikantní. Variantu označenou *kurzívou* nebylo možné validovat z důvodu selhávání detekční techniky i navzdory několika pokusům optimalizace metody. Klinicky validované varianty na vzorku populace 805 pacientek s nádorem prsu jsou označena **tučným** písmem.

Predikované microRNA, konkrétně miR-665 se nacházela na pozici místa 7mer-m8 a miR-670-3p na místě 7mer-A1.

Nástroj dbSNV (dbNSFP 3.5a), poskytující skóre pro nekódující a synonymní kódující varianty, predikoval varianty, které nebyly statisticky významné ve vztahu k odpovědi na léčbu nebo DFS. Nesplnil tak podmínku prioritizace variant pro validační sadu.

MaxEntScan a TFBP (VEP)

Software VEP umožnil predikci sestřihového defektu dle označení varianty ve vysoké informační pozici profilu vazby transkripčních faktorů (TFBP). Varianty donorů/akceptorů sestřihu byly anotovány nástrojem MaxEntScan založeným na principu maximální entropie a neuronových sítí se skóre pro referenční a alternativní varianty. Vyšší skóre v MaxEntScan znamenalo vyšší pravděpodobnost, že sekvence je skutečným místem sestřihu.

Ze statisticky významných nekódujících variant ve vztahu k odpovědi na léčbu, molekulárnímu subtypu či DFS bylo predikováno 6 variant nástroji MaxEntScan anebo TFBP (VEP) Tabulka 4—9. Pro tyto varianty však heterozygotní genotyp prezentoval statisticky výraznější vliv ve srovnání s oběma homozygoty, pročež byly vyloučeny z konečného souboru statisticky významných variant určených k validaci.

Tabulka 4—9 Varianty predikované nástroji MaxEntScan a z TFBP (VEP)

parametr	rsID
DFS	
MaxEntScan	rs609290
Odpověď na léčbu	
TFBP	rs9658058
Molekulární subtyp	
TFBP	rs201711208
TFBP	rs34372448
MaxEntScan	rs3892097
MaxEntScan	rs776746

PINES

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. Významnost predikce hodnocena na hladině alfa = 5% (Bodea et al. 2018).

Tabulka 4—10 Varianty (n = 7) predikované *in silico* nástrojem PINES

SNP_ID	Nevážená hodnota	Vážená hodnota
rs10501320	0,000846667	0,000846667
rs210134	0,002466667	0,002466667
rs2293194	0,03448667	0,03448667
rs2816948	0,02390667	0,02390667
rs4148413	0,03230667	0,03230667
rs4376673	0,00304	0,00304
rs762803	8,00E-05	8,00E-05

Klinicky validované varianty na vzorku populace 805 pacientek s nádorem prsu jsou označeny **tučným** písmem

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 dopovědi na léčbu, tak zkrácením časového období bez nemoci. Tyto varianty nezískaly žádnou *in silico* evidenci v predikci škodlivosti dle vybraných nástrojů.

Tabulka 4—11 Statisticky a klinicky významné varianty pro fázi validace (n = 4) bez podpory evidence z farmakogenomických, případně klinických databází a *in silico* predikce

Gen	HGVS	HGVS protein	Klasifikace Annovar	rsID	DFS ¹	Odpověď na léčbu ¹	Funkce	AF4 ²	ExAC ³	NCMG ⁴
<i>ABC5</i>	.	.	intron	s11983326	0,04	0,038	DFS + odpověď	0,279	0,297	0,246
<i>CES1</i>	.	.	intron	s76336259	0,001	0,046	DFS + odpověď	0,063	.	0,06
<i>CYP4F12</i>	.	.	intron	s12460651	0,02	0,029	DFS + odpověď	0,882	.	.
<i>SLCO1C1</i>	.	.	intron	s34288910	0,03	0,028	DFS + odpověď	0,144	0,128	0,152

Poznámky pod čarou: ¹p-hodnota uvedená pro klinické asociace; ²MAF = hlavní frekvence alel v testovacím souboru; ³Exome Aggregation Consortium, alelické frekvence v Evropské (bez Finské) populaci; ⁴Národní centrum lékařské genomiky, alelické frekvence v běžné České populaci.

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 (viz Tabulka 4—12), důležitá z hlediska individualizované terapie, byla potvrzena ve validační studii. Všechny validované varianty byly rovněž predikovány za škodlivé aspoň jedním ze zvolených nástrojů.

Tabulka 4—12 Seznam variant ověřených ve fázi validace (n = 7)

Gen	HGVS protein	Klasifikace Annovar ¹	rsID	ClinVar	DFS ²	Odpověď na léčbu ²	Funkce ³	AF ⁴	ExAC ⁵	NCMG ⁶
<i>ATP7A</i>	p.V767L	NS	rs2227291	benigní		0,003	PA166157866; CADD	0,26	0,217	0,225
<i>DPYD</i>	p.V732I	NS	rs1801160	pravděp. benigní	0,02		PA166153647; CADD	0,052	0,047	0,045
<i>SLC28A3</i>	p.Y113C	NS	rs10868138	.		0,022	PA166157820; CADD	0,067	0,085	0,091
<i>DFFB</i>	.	intron	rs4376673	.	0,04	0,047	PINES DFS + odpověď	0,909	0,947	0,901
<i>DPYD</i>	p.F632F	synonymní	rs17376848	pravděp. benigní	0,03		PA166153874; CADD	0,057	0,037	0,047
<i>IRSI</i>	.	UTR3	rs2288587	.	0,05		1f	0,057	.	.
<i>KCNAB1</i>	.	intron	rs2293194	.	0,04	0,013	PINES, DFS + odpověď	0,476	.	.

Poznámky pod čarou: ¹NS = nesynonymní; ²p-hodnota uvedená pro klinické asociace; ³Předpověď založená na kombinaci farmakogenomických databází, např. PharmGKB (označení "PA" znamená konkrétní onemocnění, geny léčiva v databázi) a *in silico* nástrojů: CADD v případech škodlivé predikce, RegulomeDB (značení skóre „číslo a písmeno“ zde 1f) a PINES ($\leq 0,05$); ⁴MAF = hlavní frekvence alel v testovacím souboru; ⁵Exome Aggregation Consortium, alelické frekvence v Evropské (bez Finské) populaci; ⁶Národní centrum lékařské genomiky, alelické frekvence v běžné České populaci.

Čtyři varianty v genech *SLC28A3*, *ATP7A*, *KCNAB1* a *DFFB* souvisely s odpovědí pacientek na neoadjuvantní cytotoxickou léčbu (Tabulka 4—13).

Tabulka 4—13 Validované varianty významně asociující s odpovědí pacientek na neoadjuvantní cytotoxickou léčbu ve validační fázi

SNV/Gen	Genotyp	Odpověď na léčbu ¹	Bez odpovědi na léčbu ¹	p-hodnota	p-hodnota adj. ⁴
rs10868138/ <i>SLC28A3</i>	Častý homozygot	102	41	0,013	0,266
	Vzácná alela ³	23	1		
rs2227291/ <i>ATP7A</i>	Častý homozygot	88	16	0,001	0,004
	Heterozygot	29	24		
	Vzácný homozygot	9	2		
rs2293194/ <i>KCNAB1</i>	Častý homozygot	42	9	0,003	0,030
	Heterozygot	66	17		
	Vzácný homozygot	19	17		
rs4376673/ <i>DFFB</i>	Častý homozygot	115	32	0,007	0,017
	Vzácná alela ³	12	11		

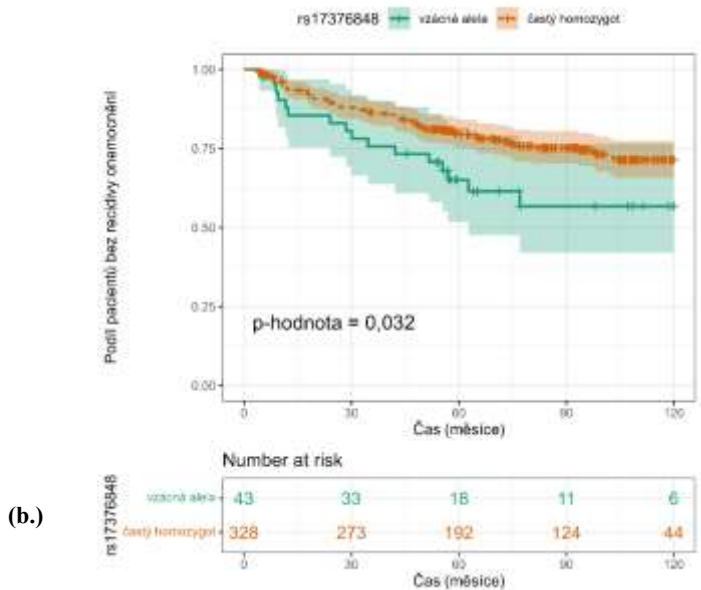
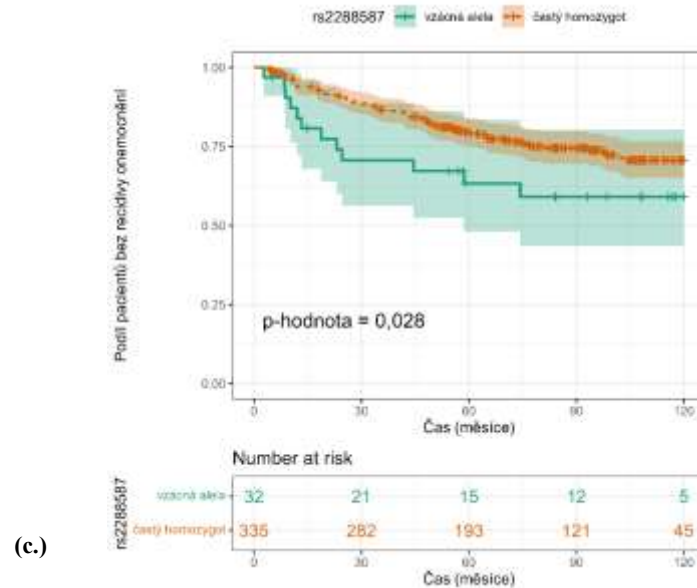
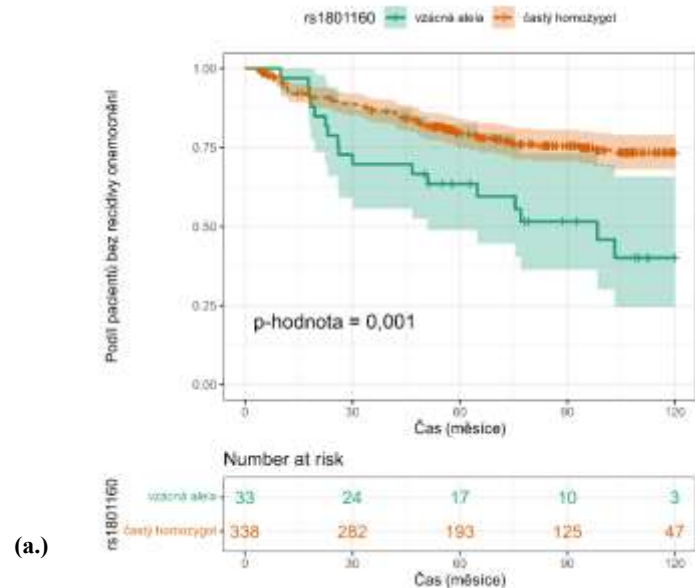
Poznámky pod čarou: ¹ počty pacientek s odpovědí na léčbu (kompletní nebo parciální remise) anebo bez odpovědi na léčbu (stabilní nebo progresivní onemocnění); ² p-hodnota Pearsonova testu; ³ při absenci vzácných homozygotů v žádné ze srovnávaných skupin byl hodnocen vliv vzácné alely; ⁴ upravená p-hodnota multivariační logistickou regresí upravenou podle stádia onemocnění.

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).

Tabulka 4—14 Validované varianty významně asociující s odpovědí pacientek vzhledem k molekulárnímu podtypu karcinomu

SNV/Gen	Genotyp	Luminální A ¹	Luminální B ¹	HER2 ²	TN ^{2,3}
rs1801160/ <i>DPYD</i>		NS	<0,001	NS	0,018
	Častý homozygot	90	150	33	63
	Vzácná alela ³	11	16	3	3
rs17376848/ <i>DPYD</i>		NS	NS	0,012	NS
	Častý homozygot	88	146	33	59
	Vzácná alela ³	13	20	3	7
rs2288587/ <i>IRS1</i>		0,002	NS	NS	NS
	Častý homozygot	93	150	33	57
	Vzácná alela ³	7	14	3	8

Poznámky pod čarou: ¹ V případě nepřítomnosti vzácných homozygotů v žádné ze srovnávaných skupin byl hodnocen vliv vzácné alely; ² p-hodnota podle log-rank testu a počty pacientek; NS = nevýznamný



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

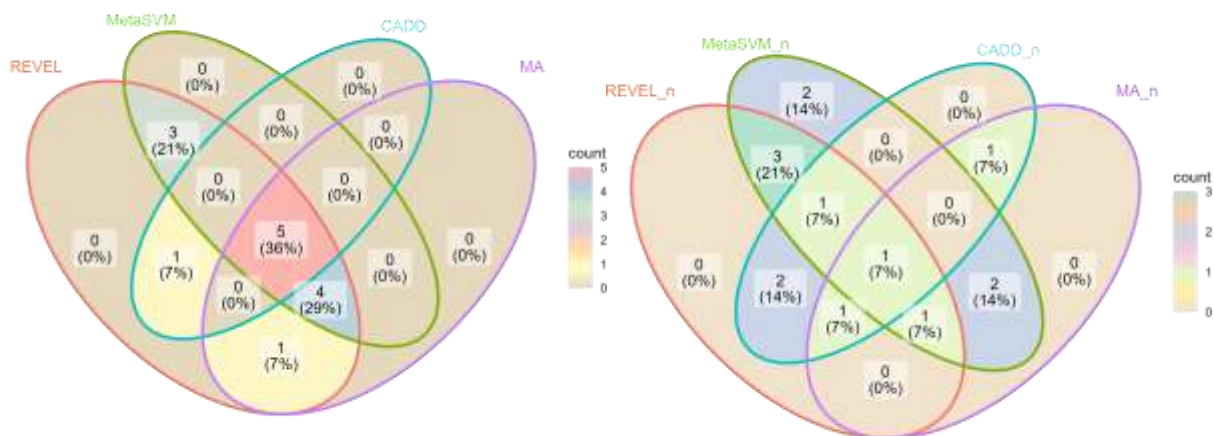
Bioinformatická funkční predikce odhalila 1080 vzácných nsSNV u 365 z 509 sekvenovaných genů. Z těchto genů bylo 34,5 % (n = 126) postiženo pouze jednou vzácnou nsSNV, tj. singletonem. Celkem 7,8 % (n = 84) vzácných nsSNV bylo nových, tj. postrádajících referenční číslo v dbSNP151. Frekvence alel pro 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. Na úrovni genů byl nejvyšší počet nsSNVs u genů *ABCA13* (n = 34), *ABCA4* (n = 21) a *ABCC1* (n = 14).

Patogenní sada

Po filtraci úrovně 1 (tj., benigní soubor) a následně úrovně 2 (tj., vyřazení unikátních genů) bylo nejvyšší patogenní skóre získáno algoritmy řízenými umělou inteligencí (REVEL, MetaSVM a CADD) a MutationAssessor pro pět (36 %) genů (*CFTR*, *ABCC1*, *SLC22A1*, *TUBB1* a *ABCA1*; viz Graf 4—2 (A)). Po úpravě výsledků na délku genu zůstal pouze jeden gen *SLC22A1* (5 %) klasifikovaný všemi nástroji jako nejvíce patogenní viz Graf 4—2 (B). Seznam genů patogenní sady je uveden v doplňkovém materiálu k článku (Kováčová et al. 2024) (Supplementary Table S5, A). Konečný soubor prioritních genů a celkové skóre pro každý gen jsou uvedeny v tabulkách (Tabulka 4—15, Tabulka 4—16).

Benigní sada

Ve filtraci predikce 1. úrovně bylo 300 genů předpovězeno jako benigní (délka genu 3785 ± 2553 bp, průměr \pm SD). Celkem 737 nsSNV (nově detekovaných, n=50, 6,8 %) bylo v deseti nejčastěji mutovaných genech rozloženo v rozmezí 7-23 nsSNVs na gen. Zde bylo identifikováno 95 genů (4533 bp \pm 3143, průměr \pm SD) s alespoň 3 nsSNVs na gen. Šest genů (*ABCA13*, *ABCA12*, *NCOR2*, *PIK3C2B*, *CIT* a *ABCA4*) bylo společně předpovězeno s patogenními geny a dva geny (*ABCA4* a *ABCC2*) byly společně se soubory normalizovaných genů, a proto byly vyloučeny, viz doplňkový materiál k článku (Kováčová et al. 2024) (Supplementary Table S5, A a B).



Graf 4—2 Vennův diagram shodně predikovaných genů po filtraci úrovně 1 a 2, (A) prediktory řízené strojovým učení, 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 zabarvení diagramu dle počtu sdílených genů (R script, knihovna: ggVennDiagram, ggplot2).

Tabulka 4—15 Získaná váha důkazů patogenity genů

Velmi vysoká (++++) skóre: 8-7	Vysoká (+++) skóre: 6-5	Střední (++) skóre: 4-3	Nízká (+) skóre: ≤2 a/ nebo PGx silné důkazy
<i>SLC22A1</i> skóre: 8	<i>TUBB1</i> skóre: 6 PGx	<i>CYP4F3</i> skóre: 4	<i>DPYD</i> PGx silné důkazy VIP gen skóre: 2
<i>ABCC1</i> skóre: 7 PGx ClinVar	<i>CFTR</i> skóre: 6 PGx VIP gen	<i>ABCC4</i> skóre: 4	<i>CYP2D6</i> PGx silné důkazy VIP gen skóre: 0
	<i>ABCB6</i> skóre: 5 <i>SLC22A18</i> skóre: 5	<i>ABCA1</i> skóre: 4 <i>ATP7B</i> skóre: 3 <i>ABCA9</i> skóre: 3	

Tabulka 4—16 Konečný soubor nejlépe hodnocených genů a počet variant (nsSNV) ve srovnání s poměrem genových variant (GVR) a poměrem patogenních genů (GPR) pro každý nástroj.

Gen (celkový počet nsSNVs; počet nových nsSNVs)	Počet pacientek	GVR	GPR REVEL	GPR MetaSVM	GPR CADD	GPR MutationAssessor
<i>ABCC1 (14;0)</i>	26	2.15	1.54	1.54	1.85	1.54
<i>ABCA1 (11;1)</i>	19	1.05	0.29	0.48	0.67	0.38
<i>TUBB1 (7;1)</i>	13	2	0.86	1.43	2	2
<i>CFTR (12;1)</i>	16	1.96	1.79	1.79	1.63	1.14
<i>CYP4F3 (6;1)</i>	6	2.01	1	1.34	1	1.34
<i>SLC22A1 (8;0)</i>	26	4.2	1.57	2.62	3.67	3.15
<i>ABCB6 (7;1)</i>	13	2.41	2.06	2.06	2.06	1.03
<i>ABCC4 (11;0)</i>	13	3.86	1.4	1.05	NA	1.05
<i>ATP7B (7;1)</i>	10	1.16	1	0.83	0.83	0.66
<i>SLC22A18 (5;1)</i>	9	3.24	1.95	NA	2.59	2.59
<i>ABCA9 (7;1)</i>	7	1.1	0.63	0.63	0.63	0.63

Poznámky pod čarou: Gene (GVR) byl vypočítán na úrovni genů pro celkový počet nsSNV. zjištěných u pacientek s karcinomem prsu ve studii na gen a délku genu. Gen (GPR) byl vypočten jako celkový počet patogenních nsSNVs na délku genu. předpovězených uvedeným nástrojem a délkou genu. NA - není k dispozici

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 dat viz Tabulka 4—17). Varianty výrazně snižující nebo zvyšující funkci genu nebyly identifikovány.

Data mining v databázi ClinVar s filtrem pro „odpověď na léčivo“, poskytlo potenciálně důležité varianty u 12 pacientů. (opětovně v genu) *CYP2D6* a *ABCC1* (viz Tabulka 4—17). Získaná váha důkazů patogenity genů a konečný soubor prioritních genů včetně celkového skóre pro každý gen jsou uvedeny v tabulkách (viz **Tabulka 4—15** a **Tabulka 4—16**).

Frekvence alel pro nsSNV pacientek s karcinomem prsu v prioritizovaných genech byla srovnatelná z hlediska stanovené prahové hodnoty, tj. pod $MAF \leq 0,05$ s údaji z ExAC-NFE a NCMG.

Tabulka 4—17 PharmVar and ClinVar "odpověď na léčivo" data mining

Gen	rsID/typ varianty	HGVS (GRCh37)	ALFA ¹	CPIC Klinická Funkce	Ref	Alt	Počet pacientek
PharmVar							
DPYD	rs67376798 ² / nesynonymní SNV	NC_000001.10:g.97547947T>A	T=0,995 A=0,005	snížená funkce	T	A	1
	rs56038477/ synonymní SNV	NC_000001.10:g.98039419C>T	C=0,980 T=0,019	snížená funkce	C	T	4 [#]
	rs1081003/*10 synonymní SNV	NC_000022.10:g.42525756G>A	G=0,972 A=0,028	snížená funkce	G	A	5 [#]
CYP2D6	rs5030656/*109 non-frameshift delece	NC_000022.10:g.42524178_42524180del	CTT=0,02 2	snížená funkce	CTT	-	2
	rs79292917/*59 synonymní SNV	NC_000022.10:g.42523854C>T	C=0,995 T=0,005	snížená funkce	C	T	5
ClinVar							
CYP2D6	rs5030655/*6 frameshift delece	NC_000022.11:g.42129084del	A= 0,999 =0,0017	NA	A	-	2
	rs35742686/*3 frameshift delece	NC_000022.11:g.42128242del	T= 0,986 =0,014	NA	T	-	4 [#]
ABCC1	rs45511401/ nesynonymní SNV	NC_000016.10:g.16079375G>T	G=0,954 T=0,0484	NA	G	T	7 [#]

Poznámky pod čarou:

¹ALPHA verze: 20201027095038, Evropská populace

² Klinický význam v "odpovědi na léčivo" podle ClinVar alely s hvězdičkou (*) *3/6/59 CYP2D6 (referenční materiál vytvořený PharmGKB a CPIC)

2 pacientky s kombinací variant rs56038477-rs35742686 a rs1081003-rs45511401

NA - není k dispozici

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* poskytují 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], $p = 0,008$).

Pro gen ABCC1 bylo zjištěno čtrnáct vzácných nsSNV u dvaceti šesti pacientů (z toho 8 tzv., unikátních variant, tj. variant pozorovaných pouze u jednoho pacienta v souboru dat). U třinácti pacientek (12,4 %) bylo identifikováno jedenáct nsSNVs (n = 7 unikátních variant) pro gen ABCC4. Maximálně u jedné pacientky byly detekovány dvě nsSNVs na gen pro ABCC1 a ABCC4. 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] (viz Tabulka 4—18).

Naopak varianty v ABCC4 vykazovaly asociaci naznačující jejich protektivní roli, pokud jde o nízký stupeň nádoru (G1 nebo G2) a jiný než invazivní duktální 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, viz Tabulka 4—18).

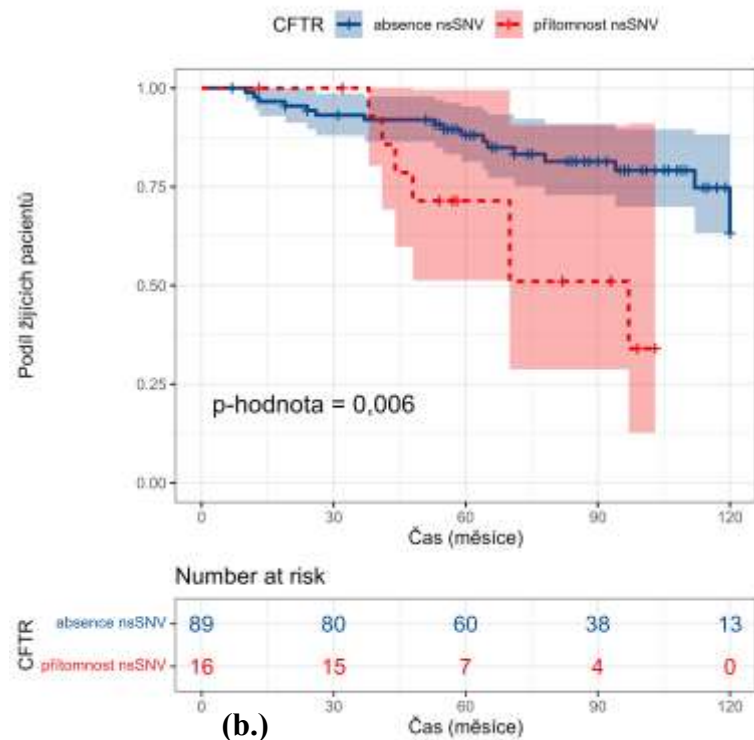
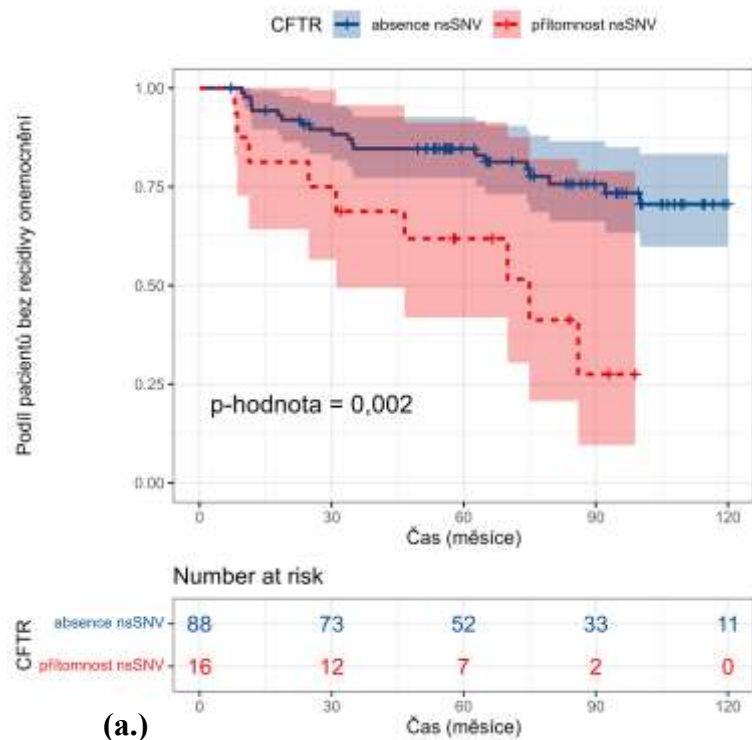
Pro gen ATP7B bylo zjištěno sedm nsSNVs (5 unikátních variant) u deseti pacientů (9,5 %) a u CYP2D6 byly zjištěny tři nsSNVs (2 unikátní varianty) u pěti pacientů (4,8 %). U žádného z těchto genů nebyla zjištěna více než jedna varianta na pacienta. Přítomnost nsSNVs se vzácnou frekvencí ve farmakogenu 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). Celkově byl u šesti pacientek (5,7 %) zjištěn výskyt pouze unikátních variant v genu CYP4F3 a sedm nsSNV (4 unikátních varianty) u dvanácti (11,43 %) pacientek v genu ABCB6 (viz Tabulka 4—18).

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.

Statistická významnost mezi souborem prioritních genů a klinickými a klinicko-patologickými parametry pacientek, včetně odpovědi na léčbu, byla nalezena pro několik srovnání, viz Tabulka 4—18.

Tabulka 4—18 Vztah vzácných nsSNVs klinicko-patologickými údaji pacientek s karcinomem prsu

Gen	nsSNV přítomna/ absence	<u>klinicko-patologický stav</u> počet pacientek		OR	95% (CI)	P- hodnota
<i>ABCC1</i>		<u>pGrade 1-2</u>	<u>pGrade 3</u>			
	absence	41	34			
	přítomna	7	18	3,1	(1,16 – 8,30)	0,021
<i>ABCC4</i>		<u>pGrade 1-2</u>	<u>pGrade 3</u>			
	absence	39	49			
	přítomna	9	3	0,27	(0,07 – 1,05)	0,046
		<u>jiný histologický typ karcinomu</u>	<u>IDC</u>			
	absence	13	79			
	přítomna	5	8	0,26	(0,07 – 0,93)	0,045
		<u>stabilní/ progresivní onemocnění</u>	<u>kompletní/ částečná odpověď na léčbu</u>			
	absence	14	45			
	přítomna	6	2	9,64	(1,75 – 3,26)	0,007
<i>CYP2D6</i>		<u>pGrade 1-2</u>	<u>pGrade 3</u>			
	absence	43	52			
	přítomna	5	0	0,45	(0,36 – 0,56)	0,023
<i>ATP7B</i>		<u>pGrade 1-2</u>	<u>pGrade 3</u>			
	absence	40	51			
	přítomna	8	1	0,1	(0,01 – 0,82)	0,013
		<u>LA or LB</u>	<u>TNBC</u>			
	absence	39	57			
	přítomna	7	2	0,2	(0,04 – 0,99)	0,04
<i>CYP4F3</i>		<u>LB</u>	<u>TNBC</u>			
	absence	26	58			
	přítomna	4	1	0,11	(0,01 – 1,05)	0,042
<i>ABCB6</i>		<u>stabilní/ progresivní onemocnění</u>	<u>kompletní/ částečná odpověď na léčbu</u>			
	absence	15	44			
	přítomna	5	3	4,89	(1,04 – 22,96)	0,045
		<u>premenopauzální</u>	<u>postmenopauzální</u>			
	absence	39	54			
	přítomna	9	3	0,24	(0,06 – 0,95)	0,03

Poznámky pod čarou: nsSNV = nesynonymní jednonukleotidová varianta, OR = poměr šancí, CI = konfidenční interval, IDC = intraduktální histologický typ, LA/LB = luminální A nebo B molekulární podtyp, TNBC = triple negativní karcinom prsu molekulární podtyp, pGrade (patologický grade) karcinomu (1 a 2 vs. 3), odpověď na léčbu NACT (kompletní nebo částečná patologická odpověď vs. stabilní nebo progresivní onemocnění).

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). Nicméně, "zlatý standard" všeobecně uznávaného nástroje stále chybí a zmíněné metody pro interpretaci výsledků jsou neharmonizovaně aplikovány. V předkládané práci jsme využili databázového relačního programovacího jazyka SQL a vyvinuli interní databázi, která disponuje daty dostupnými na stránkách PharmGKB a poskytuje automatické a rychlé vyhledávání potřebné k prioritizaci variant. V našem výzkumu jsme navíc aplikovali značné množství dostupných a obecně využívaných bioinformatických metod k získání komplexních informací pro dané varianty. Predikce nebo znalosti získané *in silico* strategií nebyly dále ručně dohledávány v ostatních zdrojích jako je LOVD atd. Důvodem bylo plné využití automatické, a tak časově efektivní prioritizace k odhadu přidané hodnoty nástrojů *in silico* pro další studie laboratoře školitele.

Za posledních 20 let bylo vyvinuto mnoho bioinformatických nástrojů, které pomáhají s predikcí především pro kódující varianty. Evoluční charakteristiky variant ve farmakogenech zapojených do biotransformace a transportu léčiv jsou však odlišné. Tento fakt komplikuje přesnost odhadů predikce *in silico* metodami většinou postavenými na principech dědičnosti nemocí (Ingelman-Sundberg et al. 2018). V důsledku toho nebyly do naší sady sekvenčních nástrojů zahrnuty hlavně algoritmy zaměřené na profilování evolučního aspektu genomu anebo algoritmy hodnotící evoluční omezení, stejně jako nástroje vyškolené na patogenních/neutrálních variantách pro onemocnění. Nicméně, brali jsme v potaz několikanásobné ověření obecných publikovaných výsledků výkonnosti, kdy CADD anebo MetaSVM/LR nástroje značně

převyšovaly ostatní dostupné metody a predikci pro nsSNVs jsme doplnili o tyto prediktory, minoritně využívající evoluční data a data fenotypově podléhajícím Mendelovským zákonům dědičnosti.

Dalším aspektem ve výběru prediktoru byla nastavená hranice frekvence alel (MAF nad 5 %). Polymorfismus je definován jako změna v sekvenci DNA, která se v obecné populaci vyskytuje s MAF větším než 1 %. Podle široce zaužívaných pokynů American College of Medical Genetics (ACMG) and Association for Molecular Pathology (AMP) pro interpretaci klinických variant, se varianta vyšší frekvence > 5 % MAF považuje za samostatnou podporu pro benigní interpretaci (Richards 2015). Benigní charakter varianty je podpořen "neutrální teorií", která definuje neutrální varianty jako varianty objevující se v populaci náhodným genetickým driftem a nezpůsobující ani škodlivý, ani prospěšný vliv na přežití jednotlivých organismů (Kimura 1979). Zmíněná teorie ale není platná pro farmakogenomické varianty s frekvencí nad 5 % kdy se zaměřujeme na fenotyp odpovědi na léčbu. Dalším faktem je skutečnost, že MAF je do značné míry závislá na velikosti populace, která je rozdílná mezi subpopulacemi, což vede k důležité stratifikaci populace opomíjené v prediktorech (Eilbeck et al. 2017). Například rs79444516 genu USH2A, varianta běžná v Africké populaci (12 %) vykazuje v Evropské a Asijské populaci extrémní vzácnost. Při odhadu ve smíšené populaci je MAF 1,2 %, což způsobí nesprávnou, tj. benigní klasifikaci při volání dat, v našem zájmu pro Evropskou populaci (Niroula a Vihinen 2019). Problematika filtrace variant se přenesla i do *in silico* nástrojů. Při trénování prediktorů jsou varianty s frekvencí alel z populačních databází vyšší než stanovenou (často 5 % nebo 1 %) obvykle označovány jako benigní nebo neutrální. Pro stanovení našeho prvního cíle zaměřeného na potenciální škodlivé varianty s MAF nad 5 % to znamenalo komplikaci ve výběru nástroje v predikci, hlavně u nesynonymních variant. Sekvenční sada nástrojů měla tento problém omezit vzhledem k převaze nástrojů bez filtrace frekvence.

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 *KCNAB1*. 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 luminálním B a triple negativním molekulárním podtypem.

Zajímavého výsledku jsme dosáhli pro následující dvě intronové varianty. Pro gen *DFFB* (rs4376673) a *KCNAB1* (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 ($p < 0,001$). Gen *KCNABI* kóduje draslíkový kanál zapojený do důležité dopaminové dráhy, chemického přenosu signálu přes synapse a různých drah CYP450 (<https://www.genecards.org/cgi-bin/carddisp.pl?gene=KCNABI>). 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). Bylo zjištěno, že *DFFB* spouští fragmentaci DNA i kondenzaci chromatinu během apoptózy ([https://www.genecards.org/cgi-bin/carddisp.pl?gene=DFFB &keywords=DFFB](https://www.genecards.org/cgi-bin/carddisp.pl?gene=DFFB&keywords=DFFB)). 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 analogů je dihydropyrimidin dehydrogenáza (*DPYD*). Protein kódovaný tímto genem je pyrimidinový katabolický enzym, rychlost limitující faktor v dráze katabolismu uracilu a thymidinu (<https://www.genecards.org/cgi-bin/carddisp.pl?gene=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 režimem sestávajícím 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 epitelialním karcinomem vaječníků 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 specificitě 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 škodlivě *in silico*, byly klinicky ověřeny.

5.1.2. Predikce vzácných variant

Hodnocení fenotypových důsledků nízkofrekvenčních a vzácných variant je stále pracné a nemá jednotný přístup. Celosvětově platná směrnice American College of Medical Genetics and Genomics (ACMG) a Association of Molecular Pathologists (AMP), ačkoli není přímo určena pro genetickou interpretaci odpovědi na léčbu, doporučuje podmínky pro predikci prostřednictvím počítačové analýzy (Richards 2015). V posledních dvou desetiletích byla vyvinuta široká škála nástrojů pro předpovídání funkčních důsledků variant, které využívají různé přístupy, jako je analýza sekvence a struktury proteinu, evoluční zachování a strojové učení. Ve studii dílčího cíle pro vzácné varianty jsme zkoumali přístup, který integruje interpretaci funkčních důsledků řízenou umělou inteligencí. pro následnou prioritizaci genů a klinických korelací u pacientek s nádorem prsu.

Výsledky *in silico* predikce ukázali že geny *ABCA13*, *NCOR2* a *ABCA4* lze považovat za nevíce patogenní ale i benigní co se týče počtu detekovaných nsSNVs na gen. Tyto geny byly v procesu úrovně 1 vyloučeny z celkového rozhodnutí pro statistickou analýzu kvůli pravděpodobně falešně pozitivním předpovědím, a to u všech čtyř algoritmů (REVEL, MetaSVM, CADD a MutationAssessor). Vysoký počet pacientů s variantami v *ABCA13* a *ABCA4* (35 resp. 45 ze 105 pacientek) dále naznačoval nepatogenní povahu těchto genů. V naší dříve publikované pilotní studii (Hlavac et al. 2018) byly *NCOR2*, *ABCA13*, *RPTOR*, *ABCA4* a *CIT* identifikovány jako pět nejčastěji polymorfních genů u pacientek s karcinomem prsu.

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 (n=28, 26,6 %), 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. Predikce byly v souladu s databázemi a a

varianty byly vyhodnocené jako patogenní. 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.

Celkově náš přístup poskytl rychlý a efektivní způsob vyhodnocení funkčnosti vzácných variant. Pro sedm z dvanácti prioritizovaných genů (*ABCC1*, *ABCC4*, *ABCB6*, *ATP7B*, *CFTR*, *CYP2D6* a *CYP4F3*) jsme odhalili statistickou významnost s klinicko-patologickými parametry nebo s prognózu onkologické léčby (*CFTR*).

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 naší 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, které 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*

Specifické zaměření na všechny varianty v genu *KIF14*, který je považován za jeden ze 70 prognostických biomarkerů u karcinomu prsu, byly nástrojem RegulomeDB objeveny funkční nekódující varianty (rs17448931 a rs3806362). Varianty byly následně úspěšně reprodukovány na rozsáhlých klinických datech (n = 808) pacientek studie. Celkově tak 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 zařazená 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

Cílem předkládané práce bylo vybrat vhodné algoritmy a navrhnout efektivní postup pro zpracování a zhodnocení kandidátních variant z testovací studie do validační studie. Pomocí *in silico* bioinformatické analýzy jsme dospěli k rychlejší a komplexnější predikci funkční konsekvence vrozených variant všech typů, tj. v kódující i nekódující části genomu. Z epidemiologického hlediska jsou všechny stávající hodnocení pouze retrospektivní a nevyhnutelně trpí řadu výběrových zkreslení. Vrozené varianty, avšak nevykazují, na rozdíl od somatických, závislost v čase a částečně tak redukuje zkreslení výstupů v procesu hodnocení dopadu na progresi onemocnění.

Výběr nástrojů obsahoval specifický set algoritmů a nové metody strojového učení doplněné o farmakogenomická a klinická data největších světových projektů. V jednotlivých částech práce byly objeveny významné vztahy variant k odpovědi na léčbu a přežití bez relapsu onemocnění u pacientek s karcinmem prsu léčených cytotoxickými léky.

5.2.1. Validační studie

Prostřednictvím masivního paralelního sekvenování byla u českých pacientek s karcinmem 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 *KCNABI* (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 zvažování

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. 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 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. RegulomeDB je postaven na heuristickém bodovacím systému

založeném na funkční důvěryhodnosti varianty. Kvantitativní skórovací systém, který určuje, zda se varianta nachází ve funkčním místě a pravděpodobně vede k funkčnímu důsledku, tj. ke změně vazby transkripčního faktoru a regulačnímu účinku genu.

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 publikací

Publikace, které jsou podkladem práce s IF

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, 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%.

Seznam příkladových kódů

Příkladový kód 1 Tvorba interní databáze s názvem „PharmGKBdownload” v MySQL, farmakogenomická data ze souboru „Relationships“ (PharmGKB)

```
create database PharmGKBdownload;
show tables;
create table relationships (Entity1_id varchar(100) NOT NULL, Entity1_name varchar(100),
Entity1_type varchar(100), Entity2_id varchar(100), Entity2_name varchar(100),
Entity2_type varchar(100), Evidence varchar(100), Association varchar(100), PK
varchar(100), PD varchar(100), PMIDs int(100), PRIMARY KEY (Entity1_id));
load data local infile 'C:/ProgramData/MySQL/MySQL Server 5.7/Uploads/relationships.tsv'
INTO TABLE relationships IGNORE 1 LINES.
alter table relationships add column ident MEDIUMINT NOT NULL AUTO_INCREMENT
PRIMARY KEY FIRST;
```

Příkladový kód 2 Úprava dat pro varianty z anotačních souborů dbNSFP a VEP v systému Unix pomocí jazyka AWK k detekci shodných variant s varianty pacientů (konkrétně pacientek) studie a varianty zdrojových dat „PharmGKBdownload”

```
bcftools view SeqCap_GATK_All.hg19_multianno.vcf.gz | bcftools query -H -f
"%CHROM\t%POS\t%REF\t%ALT\t%ID\t%Gene.refGene\t%avsnp150\t%AC\t%AF\t%AN
\t%ExAC_NFE\t%ExAC_nontcga_NFE\t%NCMG2017\t%NCMG2018\t%1000g2015aug_e
ur\t%gnomAD_exome_NFE[\t%GT\t]\n"> SeqCap_GATK_GT_AF # filtrace alelické
frekvence našich pacientek a ExAC_NFE databáze
```

```
/cat outputdbNSFP | awk ' {if (i == 303 || ($7=="missense_variant")) {print $0 }; i++ } ' >
missense_variant_dbNSFP_VEPcmd. #výběr nesynonymních jednonukleotidových variant
a průnik variant v souboru anotovaných dat dle ANNOVAR, obdobně voláno ze souboru
VEP
```

```
cat SeqCapNEW.hg19_multianno.txt | awk '{if (i == 0 || ($6=="CYP2C9")) {print $0};
i=1}' > annovar_cyp2c9.txt #výběr variant ze souboru na příkladu genu CYP2C9
```

```
cat missense_variant_dbNSFP_VEPcmd | awk '{if (i == 0 || ($20=="rs1799853")) {print
$0}; i=1}' > rs1799853_dbNSFP_VEPcmd #výběr pacientů se specifickou variantou
identifikátoru rs1799853 ze souboru anotovaných dat dle ANNOVAR, obdobně voláno ze
souboru VEP
```

```
# grep "associated" PacAll_CYPs_annovarWT_MySQL_avsnp150vsrelationships.txt >
sharedCYPs_PharmGKB_associated #výběr pacientů s varianty v genech CYP rodiny a
informací uvedenou v PharmGKB (relationships) pro tyto geny
```

```
cat annovarWT_PharmGKB_shared_CYPsOnly_clinicalVariations | awk -F $'\t' 'BEGIN
{OFS = FS} {if (i == 1 || ($4=="1A")) {print $0}; i=1}' #výběr pacientů s variantami
v genech CYP rodiny a klinickou evidencí 1A uvedenou v PharmGKB (ClinicalVariations)
pro tyto geny
```

Příkladový kód 3 Volání predikce sekvenční sadou

```
cat outputdbNSFP_final_all | awk -F '\t' 'BEGIN {OFS = FS} {if (i == 303 || ($184 ~ /D/ && $184 !~ /T/ && $124 ~ /D/ && $124 !~ /T/ && $147 ~ /D/ && $147 !~ /T/ && $174 ~ /D/ && $174 !~ /B/ && $177 ~ /D/ && $177 !~ /B/ && $145 ~ /D/ && $145 !~ /N/ && $210 ~ /D/ && $210 !~ /N/ && $172 ~ /D/ && $172 !~ /N/ && $162 ~ /H/ && $162 !~ /N/)) {print $184, $185, $124, $125, $153, $155, $150, $152, $147, $149, $174, $176, $177, $179, $145, $146, $210, $212, $172, $173, $162, $163, $0}; i++}' > sekvencni_sada_consensus.txt
```

Příkladový kód 4 Přidání délky genu využitím interaktivního nástroje Galaxy

Platforma Galaxy verze 0.1.2, nástroj pro délku gen (R verze 3.3.2, CentOS Linux 7) poskytl tabulkový soubor s údaji názvu genu a délky genu na základě zadaného GTF a souboru FASTA. FASTA soubor byl zvolen pro lidský genom hg19.

GTF soubor byl vygenerován z anotační databáze (UCSC Genome Browser database).

```
awk -F '\t' '$3== "CDS" {print $9, $5-$4}' Gene_length > output.txt # filtrace získaného souboru pro exomové varianty
```

```
cat Gene_length | awk -F '\t' 'BEGIN {OFS = FS} {if (i == 0 || ($3== "exon")) {print $31, $0}; i++}' > SeqCap_GATK_All.hg19 # filtrace genů v souladu se souborem genů v panelu naší studie
```

```
cat table1_geneLength.txt | awk '{if (i == 0 || ($1~/^NM_000667$/||$1~/^NM_000668$/||$1~/^NM_000669$/||$1~/^NM_000670 (...)||$1~/^NM_030773$/||$1~/^NM_001143991$/)) {print $0}; i=1}' > 460_genes_galaxy # výslední soubor dostupných genů, n = 460 z nástroje Galaxy
```

#50 genů v našem panelu a nedostupných pro délku genu z nástroje Galaxy bylo získáno z databáze NCBI Reference Sequence Database, např., NM_000015.2, 1317 bp a integrováno do MySQL

```
create table gene_length (Gene_refGene varchar(100), RefSeq_mRNA_ac varchar(100), RefSeq_mRNA varchar(100), GeneLength varchar(100)); #vytvoření tabulky v MxSQL
```

```
load data local infile 'C:/ProgramData/MySQL/MySQL Server 5.7/Uploads/gene_length_MySQL.csv' INTO TABLE gene_length FIELDS TERMINATED BY ',' ENCLOSED BY '"' LINES TERMINATED BY '\n' IGNORE 1 ROWS; # úprava tabulky
```

```
SELECT * FROM gene_length WHERE Gene_refGene rlike "ABCBI"; # kontrola volání údajů na příkladu genu ABCBI
```

```
SELECT g.GeneLength, a.Gene_refGene, COUNT(a.ident) AS Variants FROM
GATK_Annovar_wt AS a JOIN gene_length AS g ON g.Gene_refGene = a.Gene_refGene
GROUP BY Gene_refGene ORDER BY Variants DESC; #spojení údajů pro délku geny
společně s anotovanými daty pacientek podle Annovaru
```

```
SELECT g.GeneLength, a.Gene_refGene, COUNT(a.ident) AS Variants FROM
GATK_Annovar_wt AS a JOIN gene_length AS g ON g.Gene_refGene = a.Gene_refGene
where (REVEL_score > 0.5) AND AF <= 0.05 AND FILTR = 'PASS' GROUP BY
Gene_refGene ORDER BY Variants DESC; #seřazení délky genu s nejvyšším počtem
patogenních vzácných nsSNVs dle predikce nástrojem REVEL a mezní hodnoty 0.5 predikce
```

```
SELECT g.GeneLength, a.Gene_refGene, COUNT(a.ident) AS Variants FROM
GATK_Annovar_wt AS a RIGHT JOIN gene_length AS g ON g.Gene_refGene =
a.Gene_refGene where (MutationAssessor_score > 2.0566) AND ExonicFunc_refGene like
'nonsynonymous SNV' AND FILTR = 'PASS' AND AF <= 0.05 GROUP BY Gene_refGene
ORDER BY Variants DESC; seřazení délky genu s nejvyšším počtem patogenních vzácných
nsSNVs dle predikce nástrojem MutationAssessor a mezní optimalizované hodnoty (2.0566)
predikce
```

```
SELECT g.GeneLength, a.Gene_refGene, COUNT(a.ident) AS Variants FROM
GATK_Annovar_wt AS a RIGHT JOIN gene_length AS g ON g.Gene_refGene =
a.Gene_refGene where (REVEL_score <= 0.5) AND VEST3_score <= 0.8 AND
MetaSVM_pred rlike 'T' AND ExonicFunc_refGene like 'nonsynonymous SNV' AND AF <=
0.05 AND FILTR = 'PASS' GROUP BY Gene_refGene ORDER BY Variants DESC;
#seřazení délky genu s nejvyšším počtem benigních vzácných nsSNVs dle predikce koncensu
třech nástrojů REVEL, VEST3 a MetaSVM
```

```
SELECT Gene_refGene, (IF(ID_v151 = ".", CONCAT(Var_id), ID_v151)), sample_ID,
ExonicFunc_refGene FROM GATK_Annovar_wt WHERE (REVEL_score <= 0.5) AND
VEST3_score <= 0.8 AND MetaSVM_pred rlike 'T' AND (AF <= 0.05) AND
ExonicFunc_refGene like 'nonsynonymous SNV' AND FILTR = 'PASS'; #seřazení genů
s nejvyšším počtem benigních vzácných nsSNVs dle predikce koncensu třech nástrojů
REVEL, VEST3 a MetaSVM a zároveň s uvedením identifikace varianty dle rsID anebo
umístění na chromosomu a identifikačního čísla pacienta
```

```
SELECT LENGTH(sample_ID) - LENGTH(REPLACE(sample_ID, ';', '')), Gene_refGene,
sample_ID, count(ident) AS Variants FROM GATK_Annovar_wt WHERE
ExonicFunc_refGene like 'nonsynonymous SNV' AND (AF < 0.05) AND FILTR = 'PASS'
GROUP BY ident ORDER BY Variants DESC; #seskupení počtu všech nsSNVs na gen
s uvedením identifikačního čísla pacienta
```

Seznam webových zdrojů

Český statistický úřad, Zemřelí podle seznamu příčin smrti v ČR, Statistická ročenka České republiky – 2019, Zveřejněno dne: 25.11.2019. Dostupný z www: <http://www.czso.cz>

<https://www.sukl.cz/tamoxifen-a-cytochrom-p450-cyp2d6?highlightWords=tamoxifen>

https://www.ema.europa.eu/en/documents/report/genomics-genetics-transcriptomics-epigenetics-subgroup-report_en.pdf

<https://www.ema.europa.eu/en/news/reflection-paper-use-artificial-intelligence-lifecycle-medicines>

<https://omim.org/statistics/geneMap>

<https://www.orpha.net/consor/cgi-bin/index.php>

<http://CFTR2.org>

<https://www.pharmgkb.org>

<https://cpicpgx.org/genes-drugs>

<https://www.pharmvar.org/>

<https://www.oncokb.org>

<https://www.mycancergenome.org/>

<https://www.cancergenomeinterpreter.org/home>

<https://candl.osu.edu/>

<https://cancervar.wglab.org/about.php>

<https://cancer.sanger.ac.uk/cosmic>

<https://www.cbioportal.org/>

<https://dcc.icgc.org/>

<https://www.cancer.gov/ccg/research/genome-sequencing/tcga>

<http://agvgd.iarc.fr>

<http://genetics.bwh.harvard.edu/pph2/index.shtml>

<http://genetics.bwh.harvard.edu/pph2/index.shtml>

<http://sift.jcvi.org/>

<http://genes.mit.edu/burgelab/>

https://www.targetscan.org/vert_80/

<http://bioinf.xmu.edu.cn/ADReCS-Target>

<https://snp-nexus.org/IW-Scoring/>

<https://www.pharmgkb.org/vip>

https://www.ensembl.org/Homo_sapiens

<https://www.genecards.org/cgi-bin/carddisp.pl?gene=KCNAB1>

<https://www.genecards.org/cgi-bin/carddisp.pl?gene=DFFB &keywords=DFFB>

<https://www.genecards.org/cgi-bin/carddisp.pl?gene=DPYD>

<https://www.ncbi.nlm.nih.gov/refseq/>

<https://cpicpgx.org/guidelines/guideline-for-fluoropyrimidines-and-dpyd/>

Seznam použité literatury

- ABUGESSAISA, Imad, Jordan A RAMIŁOWSKI, Marina LIZIO, Jesicca SEVERIN, Akira HASEGAWA, Jayson HARSHBARGER, Atsushi KONDO, Shuhei NOGUCHI, Chi Wai YIP, Jasmine Li Ching OOI, Michihira TAGAMI, Fumi HORI, Saumya AGRAWAL, Chung Chau HON, Melissa CARDON, Shuya IKEDA, Hiromasa ONO, Hidemasa BONO, Masaki KATO, Kosuke HASHIMOTO, Alessandro BONETTI, Masaki KATO, Norio KOBAYASHI, Jay SHIN, Michiel DE HOON, Yoshihide HAYASHIZAKI, Piero CARNINCI, Hideya KAWAJI a Takeya KASUKAWA, 2020. FANTOM enters 20th year: expansion of transcriptomic atlases and functional annotation of non-coding RNAs. *Nucleic Acids Research* [online]. **49**(D1), D892–D898 [vid. 2023-08-13]. ISSN 0305-1048. Dostupné z: doi:10.1093/nar/gkaa1054
- 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/>
- ADZHUBEI, Ivan A, Steffen SCHMIDT, Leonid PESHKIN, Vasily E RAMENSKY, Anna GERASIMOVA, Peer BORK, Alexey S KONDRASHOV a Shamil R SUNYAEV, 2010. A method and server for predicting damaging missense mutations. *Nature Methods* [online]. **7**(4), 248–249 [vid. 2018-07-04]. ISSN 1548-7091, 1548-7105. Dostupné z: doi:10.1038/nmeth0410-248
- 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
- AGARWAL, Vikram, George W. BELL, Jin-Wu NAM a David P. BARTEL, 2015. Predicting effective microRNA target sites in mammalian mRNAs. *eLife* [online]. **4**, e05005. ISSN 2050-084X. Dostupné z: doi:10.7554/eLife.05005
- 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>
- ALTSCHUL, Stephen F., Warren GISH, Webb MILLER, Eugene W. MYERS a David J. LIPMAN, 1990. Basic local alignment search tool. *Journal of Molecular Biology* [online]. **215**(3), 403–410 [vid. 2023-07-07]. ISSN 0022-2836. Dostupné z: doi:10.1016/S0022-2836(05)80360-2
- 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
- ARAUJO, Fabricio Almeida, Debmalya BARH, Artur SILVA, Luis GUIMARÃES a Rommel Thiago Juca RAMOS, 2018. GO FEAT: a rapid web-based functional annotation tool for genomic and transcriptomic data. *Scientific Reports* [online]. **8**(1), 1794 [vid. 2023-07-31]. ISSN 2045-2322. Dostupné z: doi:10.1038/s41598-018-20211-9
- ASHBURNER, Michael, Catherine A. BALL, Judith A. BLAKE, David BOTSTEIN, Heather BUTLER, J. Michael CHERRY, Allan P. DAVIS, Kara DOLINSKI, Selina S. DWIGHT, Janan T. EPPIG, Midori A. HARRIS, David P. HILL, Laurie ISSEL-TARVER, Andrew KASARSKIS, Suzanna LEWIS, John C. MATESE, Joel E. RICHARDSON, Martin RINGWALD, Gerald M. RUBIN a Gavin SHERLOCK, 2000. Gene Ontology: tool for the unification of biology. *Nature Genetics* [online]. **25**(1), 25–29 [vid. 2023-07-07]. ISSN 1546-1718. Dostupné z: doi:10.1038/75556
- 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 STOESSERT 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
- BALASUBRAMANIAN, Suganthi, Yu XIA, Elizaveta FREINKMAN a Mark GERSTEIN, 2005. Sequence variation in G-protein-coupled receptors: analysis of single nucleotide polymorphisms. *Nucleic Acids Research* [online]. **33**(5), 1710–1721 [vid. 2023-07-08]. ISSN 0305-1048. Dostupné z: doi:10.1093/nar/gki311
- BAO, Lei a Yan CUI, 2005. Prediction of the phenotypic effects of non-synonymous single nucleotide polymorphisms using structural and evolutionary information. *Bioinformatics (Oxford, England)* [online]. **21**(10), 2185–2190. ISSN 1367-4803. Dostupné z: doi:10.1093/bioinformatics/bti365
- BEER, Tjaart A. P. de, Roman A. LASKOWSKI, Sarah L. PARKS, Botond SIPOS, Nick GOLDMAN a Janet M. THORNTON, 2013. Amino Acid Changes in Disease-Associated Variants Differ Radically from Variants Observed in the 1000 Genomes Project Dataset.

- PLOS Computational Biology* [online]. **9**(12), e1003382 [vid. 2023-07-07]. ISSN 1553-7358. Dostupné z: doi:10.1371/journal.pcbi.1003382
- 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
- BENDL, Jaroslav, Miloš MUSIL, Jan ŠTOURAC, Jaroslav ZENDULKA, Jiří DAMBORSKÝ a Jan BREZOVSKÝ, 2016. PredictSNP2: A Unified Platform for Accurately Evaluating SNP Effects by Exploiting the Different Characteristics of Variants in Distinct Genomic Regions. *PLoS Computational Biology* [online]. **12**(5) [vid. 2017-11-18]. ISSN 1553-734X. Dostupné z: doi:10.1371/journal.pcbi.1004962
- 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
- BLANCHETTE, Mathieu, W. James KENT, Cathy RIEMER, Laura ELNITSKI, Arian F. A. SMIT, Krishna M. ROSKIN, Robert BAERTSCH, Kate ROSENBLOOM, Hiram CLAWSON, Eric D. GREEN, David HAUSSLER a Webb MILLER, 2004. Aligning Multiple Genomic Sequences With the Threaded Blockset Aligner. *Genome Research* [online]. **14**(4), 708–715 [vid. 2023-07-07]. ISSN 1088-9051, 1549-5469. Dostupné z: doi:10.1101/gr.1933104
- 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
- BOYLE, Alan P., Eurie L. HONG, Manoj HARIHARAN, Yong CHENG, Marc A. SCHAUB, Maya KASOWSKI, Konrad J. KARCZEWSKI, Julie PARK, Benjamin C. HITZ, Shuai WENG, J. Michael CHERRY a Michael SNYDER, 2012. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Research* [online]. **22**(9), 1790–1797 [vid. 2018-01-02]. ISSN 1088-9051. Dostupné z: doi:10.1101/gr.137323.112
- 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

- BREITLING, Rainer, Yang LI, Bruno M. TESSON, Jingyuan FU, Chunlei WU, Tim WILTSHIRE, Alice GERRITS, Leonid V. BYSTRYKH, Gerald de HAAN, Andrew I. SU a Ritsert C. JANSEN, 2008. Genetical Genomics: Spotlight on QTL Hotspots. *PLoS Genetics* [online]. **4**(10), e1000232 [vid. 2023-08-23]. ISSN 1553-7404. Dostupné z: doi:10.1371/journal.pgen.1000232
- BRENNECKE, Julius, Alexander STARK, Robert B. RUSSELL a Stephen M. COHEN, 2005. Principles of MicroRNA–Target Recognition. *PLoS Biology* [online]. **3**(3), e85 [vid. 2023-08-05]. ISSN 1545-7885. Dostupné z: doi:10.1371/journal.pbio.0030085
- 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
- CALABRESE, Remo, Emidio CAPRIOTTI, Piero FARISELLI, Pier Luigi MARTELLI a Rita CASADIO, 2009. Functional annotations improve the predictive score of human disease-related mutations in proteins. *Human Mutation* [online]. **30**(8), 1237–1244 [vid. 2023-07-07]. ISSN 1098-1004. Dostupné z: doi:10.1002/humu.21047
- CAPRIOTTI, E., R. CALABRESE a R. CASADIO, 2006. Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. *Bioinformatics* [online]. **22**(22), 2729–2734 [vid. 2023-07-07]. ISSN 1367-4803. Dostupné z: doi:10.1093/bioinformatics/btl423
- CAPRIOTTI, Emidio, Russ B. ALTMAN a Yana BROMBERG, 2013. Collective judgment predicts disease-associated single nucleotide variants. *BMC Genomics* [online]. **14**(3), S2 [vid. 2023-07-07]. ISSN 1471-2164. Dostupné z: doi:10.1186/1471-2164-14-S3-S2
- 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
- CARTER, Hannah, Christopher DOUVILLE, Peter D. STENSON, David N. COOPER a Rachel KARCHIN, 2013. Identifying Mendelian disease genes with the Variant Effect Scoring Tool. *BMC Genomics* [online]. **14**(3), S3 [vid. 2018-06-25]. ISSN 1471-2164. Dostupné z: doi:10.1186/1471-2164-14-S3-S3
- CINGOLANI, Pablo, Adrian PLATTS, Le Lily WANG, Melissa COON, Tung NGUYEN, Luan WANG, Susan J. LAND, Xiangyi LU a Douglas M. RUDEN, 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w¹¹¹⁸; iso-2; iso-3. *Fly* [online]. **6**(2), 80–92 [vid. 2017-11-13]. ISSN 1933-6934, 1933-6942. Dostupné z: doi:10.4161/fly.19695
- 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

- CRONA, Daniel a Federico INNOCENTI, 2012. Can knowledge of germline markers of toxicity optimize dosing and efficacy of cancer therapy? *Biomarkers in Medicine* [online]. **6**(3), 349–362. ISSN 1752-0371. Dostupné z: doi:10.2217/bmm.12.19
- DAVID, Alessia, Rozami RAZALI, Mark N. WASS a Michael J.E. STERNBERG, 2012. Protein–protein interaction sites are hot spots for disease-associated nonsynonymous SNPs. *Human Mutation* [online]. **33**(2), 359–363 [vid. 2023-07-07]. ISSN 1098-1004. Dostupné z: doi:10.1002/humu.21656
- DAYEM ULLAH, Abu Z., Jorge OSCANO, 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
- DEHOUCK, Yves, Aline GROSFILS, Benjamin FOLCH, Dimitri GILIS, Philippe BOGAERTS a Marianne ROOMAN, 2009. Fast and accurate predictions of protein stability changes upon mutations using statistical potentials and neural networks: PoPMuSiC-2.0. *Bioinformatics* [online]. **25**(19), 2537–2543 [vid. 2023-07-07]. ISSN 1367-4803. Dostupné z: doi:10.1093/bioinformatics/btp445
- DELPPIERRE, Cyrille a Thomas LEFÈVRE, 2023. Precision and personalized medicine: What their current definition says and silences about the model of health they promote. Implication for the development of personalized health. *Frontiers in Sociology* [online]. **8** [vid. 2023-07-31]. ISSN 2297-7775. Dostupné z: <https://www.frontiersin.org/articles/10.3389/fsoc.2023.1112159>
- 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 SCHAİK, R. H. J. MATHIJSSSEN, 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
- DICKSON, Samuel P., Kai WANG, Ian KRANTZ, Hakon HAKONARSON a David B. GOLDSTEIN, 2010. Rare Variants Create Synthetic Genome-Wide Associations. *PLOS Biology* [online]. **8**(1), e1000294 [vid. 2023-07-26]. ISSN 1545-7885. Dostupné z: doi:10.1371/journal.pbio.1000294

- DONG, Chengliang, Peng WEI, Xueqiu JIAN, Richard GIBBS, Eric BOERWINKLE, Kai WANG a Xiaoming LIU, 2015. Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. *Human Molecular Genetics* [online]. **24**(8), 2125–2137 [vid. 2018-01-12]. ISSN 0964-6906. Dostupné z: doi:10.1093/hmg/ddu733
- 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
- DOUVILLE, Christopher, David L. MASICA, Peter D. STENSON, David N. COOPER, Derek M. GYGAX, Rick KIM, Michael RYAN a Rachel KARCHIN, 2016. Assessing the Pathogenicity of Insertion and Deletion Variants with the Variant Effect Scoring Tool (VEST-Indel). *Human Mutation* [online]. **37**(1), 28–35. ISSN 1098-1004. Dostupné z: doi:10.1002/humu.22911
- 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
- EDGAR, Robert C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* [online]. **32**(5), 1792–1797 [vid. 2023-07-07]. ISSN 0305-1048. Dostupné z: doi:10.1093/nar/gkh340
- EILBECK, Karen, Suzanna E. LEWIS, Christopher J. MUNGALL, Mark YANDELL, Lincoln STEIN, Richard DURBIN a Michael ASHBURNER, 2005. The Sequence Ontology: a tool for the unification of genome annotations. *Genome Biology* [online]. **6**(5), R44 [vid. 2023-07-07]. ISSN 1474-760X. Dostupné z: doi:10.1186/gb-2005-6-5-r44
- EILBECK, Karen, Aaron QUINLAN a Mark YANDELL, 2017. Settling the score: variant prioritization and Mendelian disease. *Nature Reviews Genetics* [online]. **18**(10), 599–612 [vid. 2023-07-07]. ISSN 1471-0064. Dostupné z: doi:10.1038/nrg.2017.52
- 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
- EVANS, William E. a Howard L. MCLEOD, 2003. Pharmacogenomics — Drug Disposition, Drug Targets, and Side Effects. *New England Journal of Medicine* [online]. **348**(6), 538–549 [vid. 2018-01-15]. ISSN 0028-4793. Dostupné z: doi:10.1056/NEJMra020526

- FILIPSKI, Kelly K., Leah E. MECHANIC, Rochelle LONG a Andrew N. FREEDMAN, 2014. Pharmacogenomics in oncology care. *Frontiers in Genetics* [online]. **5** [vid. 2018-06-30]. ISSN 1664-8021. Dostupné z: doi:10.3389/fgene.2014.00073
- FINDLAY, Gregory M., Riza M. DAZA, Beth MARTIN, Melissa D. ZHANG, Anh P. LEITH, Molly GASPERINI, Joseph D. JANIZEK, Xingfan HUANG, Lea M. STARITA a Jay SHENDURE, 2018. Accurate classification of BRCA1 variants with saturation genome editing. *Nature* [online]. **562**(7726), 217–222 [vid. 2023-07-07]. ISSN 1476-4687. Dostupné z: doi:10.1038/s41586-018-0461-z
- 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
- FLOCKHART, David A., Dennis O’KANE, Marc S. WILLIAMS, Michael S. WATSON, David A. FLOCKHART, Brian GAGE, Roy GANDOLFI, Richard KING, Elaine LYON, Robert NUSSBAUM, Dennis O’KANE, Kevin SCHULMAN, David VEENSTRA, Marc S. WILLIAMS a Michael S. WATSON, 2008. Pharmacogenetic testing of CYP2C9 and VKORC1 alleles for warfarin. *Genetics in Medicine* [online]. **10**(2), 139–150 [vid. 2023-10-30]. ISSN 1530-0366. Dostupné z: doi:10.1097/GIM.0b013e318163c35f
- 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 crowdsourcing efforts. *Nucleic Acids Research* [online]. **49**(D1), D1144–D1151. ISSN 1362-4962. Dostupné z: doi:10.1093/nar/gkaa1084
- FU, Wenqing, Timothy D. O’CONNOR, Goo JUN, Hyun Min KANG, Goncalo ABECASIS, Suzanne M. LEAL, Stacey GABRIEL, Mark J. RIEDER, David ALTSHULER, Jay SHENDURE, Deborah A. NICKERSON, Michael J. BAMSHAD, NHLBI EXOME SEQUENCING PROJECT a Joshua M. AKEY, 2013. Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. *Nature* [online]. **493**(7431), 216–220 [vid. 2023-08-23]. ISSN 1476-4687. Dostupné z: doi:10.1038/nature11690
- 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
- GALPERIN, Michael Y., Xosé M. FERNÁNDEZ-SUÁREZ a Daniel J. RIGDEN, 2017. The 24th annual Nucleic Acids Research database issue: a look back and upcoming changes.

Nucleic Acids Research [online]. **45**(Database issue), D1–D11 [vid. 2017-11-14]. ISSN 0305-1048. Dostupné z: doi:10.1093/nar/gkw1188

GARCIA, Felipe Antonio de Oliveira, Edilene Santos DE ANDRADE a Edenir Inez PALMERO, 2022. Insights on variant analysis in silico tools for pathogenicity prediction. *Frontiers in Genetics* [online]. **13**, 1010327 [vid. 2023-04-03]. ISSN 1664-8021. Dostupné z: doi:10.3389/fgene.2022.1010327

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

GIBBS, Richard A., John W. BELMONT, Paul HARDENBOL, Thomas D. WILLIS, Fuli YU, Huanming YANG, Lan-Yang CH'ANG, Wei HUANG, Bin LIU, Yan SHEN, Paul Kwong-Hang TAM, Lap-Chee TSUI, Mary Miu Yee WAYE, Jeffrey Tze-Fei WONG, Changqing ZENG, Qingrun ZHANG, Mark S. CHEE, Luana M. GALVER, Semyon KRUGLYAK, Sarah S. MURRAY, Arnold R. OLIPHANT, Alexandre MONTPETIT, Thomas J. HUDSON, Fanny CHAGNON, Vincent FERRETTI, Martin LEBOEUF, Michael S. PHILLIPS, Andrei VERNER, Pui-Yan KWOK, Shenghui DUAN, Denise L. LIND, Raymond D. MILLER, John P. RICE, Nancy L. SACCONI, Patricia TAILLON-MILLER, Ming XIAO, Yusuke NAKAMURA, Akihiro SEKINE, Koki SORIMACHI, Toshihiro TANAKA, Yoichi TANAKA, Tatsuhiko TSUNODA, Eiji YOSHINO, David R. BENTLEY, Panos DELOUKAS, Sarah HUNT, Don POWELL, David ALTSHULER, Stacey B. GABRIEL, Houcan ZHANG, Changqing ZENG, Ichiro MATSUDA, Yoshimitsu FUKUSHIMA, Darryl R. MACER, Eiko SUDA, Charles N. ROTIMI, Clement A. ADEBAMOWO, Toyin ANIAGWU, Patricia A. MARSHALL, Olayemi MATTHEW, Chibuzor NKWODIMMAH, Charmaine D. M. ROYAL, Mark F. LEPPERT, Missy DIXON, Lincoln D. STEIN, Fiona CUNNINGHAM, Ardavan KANANI, Gudmundur A. THORISSON, Aravinda CHAKRAVARTI, Peter E. CHEN, David J. CUTLER, Carl S. KASHUK, Peter DONNELLY, Jonathan MARCHINI, Gilean A. T. MCVEAN, Simon R. MYERS, Lon R. CARDON, Gonçalo R. ABECASIS, Andrew MORRIS, Bruce S. WEIR, James C. MULLIKIN, Stephen T. SHERRY, Michael FEOLO, David ALTSHULER, Mark J. DALY, Stephen F. SCHAFFNER, Renzong QIU, Alastair KENT, Georgia M. DUNSTON, Kazuto KATO, Norio NIIKAWA, Bartha M. KNOPPERS, Morris W. FOSTER, Ellen Wright CLAYTON, Vivian Ota WANG, Jessica WATKIN, Richard A. GIBBS, John W. BELMONT, Erica SODERGREN, George M. WEINSTOCK, Richard K. WILSON, Lucinda L. FULTON, Jane ROGERS, Bruce W. BIRREN, Hua HAN, Hongguang WANG, Martin GODBOUT, John C. WALLENBURG, Paul

L'ARCHEVÊQUE, Guy BELLEMARE, Kazuo TODANI, Takashi FUJITA, Satoshi TANAKA, Arthur L. HOLDEN, Eric H. LAI, Francis S. COLLINS, Lisa D. BROOKS, Jean E. MCEWEN, Mark S. GUYER, Elke JORDAN, Jane L. PETERSON, Jack SPIEGEL, Lawrence M. SUNG, Lynn F. ZACHARIA, Karen KENNEDY, Michael G. DUNN, Richard SEABROOK, Mark SHILLITO, Barbara SKENE, John G. STEWART, David L. VALLE (CHAIR), Ellen Wright CLAYTON (CO-CHAIR), Lynn B. JORDE (CO-CHAIR), John W. BELMONT, Aravinda CHAKRAVARTI, Mildred K. CHO, Troy DUSTER, Morris W. FOSTER, Marla JASPERSE, Bartha M. KNOPPERS, Pui-Yan KWOK, Julio LICINIO, Jeffrey C. LONG, Patricia A. MARSHALL, Pilar N. OSSORIO, Vivian Ota WANG, Charles N. ROTIMI, Charmaine D. M. ROYAL, Patricia SPALLONE, Sharon F. TERRY, Eric S. LANDER (CHAIR), Eric H. LAI (CO-CHAIR), Deborah A. NICKERSON (CO-CHAIR), Gonçalo R. ABECASIS, David ALTSHULER, David R. BENTLEY, Michael BOEHNKE, Lon R. CARDON, Mark J. DALY, Panos DELOUKAS, Julie A. DOUGLAS, Stacey B. GABRIEL, Richard R. HUDSON, Thomas J. HUDSON, Leonid KRUGLYAK, Pui-Yan KWOK, Yusuke NAKAMURA, Robert L. NUSSBAUM, Charmaine D. M. ROYAL, Stephen F. SCHAFFNER, Stephen T. SHERRY, Lincoln D. STEIN, Toshihiro TANAKA, †THE INTERNATIONAL HAPMAP CONSORTIUM, GENOTYPING CENTRES: BAYLOR COLLEGE OF MEDICINE AND PARALLELE BIOSCIENCE, CHINESE HAPMAP CONSORTIUM, ILLUMINA, MCGILL UNIVERSITY AND GÉNOME QUÉBEC INNOVATION CENTRE, UNIVERSITY OF CALIFORNIA AT SAN FRANCISCO AND WASHINGTON UNIVERSITY, UNIVERSITY OF TOKYO AND RIKEN, WELLCOME TRUST SANGER INSTITUTE, WHITEHEAD INSTITUTE/MIT CENTER FOR GENOME RESEARCH, COMMUNITY ENGAGEMENT/PUBLIC CONSULTATION AND SAMPLE-COLLECTION GROUPS: BEIJING NORMAL UNIVERSITY AND BEIJING GENOMICS INSTITUTE, Eubios Ethics Institute and Shinshu University HEALTH SCIENCES UNIVERSITY OF HOKKAIDO, HOWARD UNIVERSITY AND UNIVERSITY OF IBADAN, UNIVERSITY OF UTAH, ANALYSIS GROUPS: COLD SPRING HARBOR LABORATORY, JOHNS HOPKINS UNIVERSITY SCHOOL OF MEDICINE, UNIVERSITY OF OXFORD, Wellcome Trust Centre for Human Genetics UNIVERSITY OF OXFORD, US NATIONAL INSTITUTES OF HEALTH, Legal and Social Issues: Chinese Academy of Social Sciences ETHICAL, GENETIC INTEREST GROUP, HOWARD UNIVERSITY, KYOTO UNIVERSITY, NAGASAKI UNIVERSITY, UNIVERSITY OF MONTRÉAL, UNIVERSITY OF OKLAHOMA, VANDERBILT UNIVERSITY, WELLCOME TRUST, SNP DISCOVERY: BAYLOR COLLEGE OF MEDICINE, WASHINGTON UNIVERSITY, SCIENTIFIC MANAGEMENT: CHINESE ACADEMY OF SCIENCES, CHINESE MINISTRY OF SCIENCE AND TECHNOLOGY, GENOME CANADA, GÉNOME QUÉBEC, Culture JAPANESE MINISTRY OF EDUCATION Sports, Science and Technology, THE SNP CONSORTIUM, Legal and Social Issues Group INITIAL PLANNING GROUPS: POPULATIONS AND ETHICAL, a METHODS GROUP, 2003. The International HapMap Project. *Nature* [online]. **426**(6968), 789–796 [vid. 2023-08-23]. ISSN 1476-4687. Dostupné z: doi:10.1038/nature02168

GOECKS, Jeremy, Anton NEKRUTENKO, James TAYLOR, a GALAXY TEAM, 2010. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biology* [online]. **11**(8), R86. ISSN 1474-760X. Dostupné z: doi:10.1186/gb-2010-11-8-r86

GONZÁLEZ-DEL POZO, María, Elena FERNÁNDEZ-SUÁREZ, Nereida BRAVO-GIL, Cristina MÉNDEZ-VIDAL, Marta MARTÍN-SÁNCHEZ, Enrique RODRÍGUEZ-DE LA

- RÚA, Manuel RAMOS-JIMÉNEZ, María José MORILLO-SÁNCHEZ, Salud BORREGO a Guillermo ANTIÑOLO, 2022. A comprehensive WGS-based pipeline for the identification of new candidate genes in inherited retinal dystrophies. *npj Genomic Medicine* [online]. **7**(1), 1–15 [vid. 2023-04-03]. ISSN 2056-7944. Dostupné z: doi:10.1038/s41525-022-00286-0
- GONZÁLEZ-PÉREZ, Abel a Nuria LÓPEZ-BIGAS, 2011. Improving the Assessment of the Outcome of Nonsynonymous SNVs with a Consensus Deleteriousness Score, *Condel. American Journal of Human Genetics* [online]. **88**(4), 440–449 [vid. 2017-11-19]. ISSN 0002-9297. Dostupné z: doi:10.1016/j.ajhg.2011.03.004
- GRANTHAM, R., 1974. Amino acid difference formula to help explain protein evolution. *Science (New York, N.Y.)* [online]. **185**(4154), 862–864. ISSN 0036-8075. Dostupné z: doi:10.1126/science.185.4154.862
- GRIBSKOV, M, A D MCLACHLAN a D EISENBERG, 1987. Profile analysis: detection of distantly related proteins. *Proceedings of the National Academy of Sciences* [online]. **84**(13), 4355–4358 [vid. 2023-07-07]. Dostupné z: doi:10.1073/pnas.84.13.4355
- 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
- GUEROIS, Raphael, Jens Erik NIELSEN a Luis SERRANO, 2002. Predicting Changes in the Stability of Proteins and Protein Complexes: A Study of More Than 1000 Mutations. *Journal of Molecular Biology* [online]. **320**(2), 369–387 [vid. 2023-07-07]. ISSN 0022-2836. Dostupné z: doi:10.1016/S0022-2836(02)00442-4
- HARAKSINGH, Rajini R. a Michael P. SNYDER, 2013. Impacts of Variation in the Human Genome on Gene Regulation. *Journal of Molecular Biology* [online]. **425**(21), Understanding Molecular Effects of Naturally Occurring Genetic Differences, 3970–3977 [vid. 2023-07-07]. ISSN 0022-2836. Dostupné z: doi:10.1016/j.jmb.2013.07.015
- 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
- HECHT, Maximilian, Yana BROMBERG a Burkhard ROST, 2015. Better prediction of functional effects for sequence variants. *BMC Genomics* [online]. **16**(8), S1 [vid. 2018-07-06]. ISSN 1471-2164. Dostupné z: doi:10.1186/1471-2164-16-S8-S1

- 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 SCHAIK, 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
- HERTZ, Daniel L, 2013. Germline pharmacogenetics of paclitaxel for cancer treatment. *Pharmacogenomics* [online]. **14**(9), 1065–1084 [vid. 2018-01-13]. ISSN 1462-2416. Dostupné z: doi:10.2217/pgs.13.90
- HICKS, Stephanie, David A. WHEELER, Sharon E. PLON a Marek KIMMEL, 2011. Prediction of missense mutation functionality depends on both the algorithm and sequence alignment employed. *Human Mutation* [online]. **32**(6), 661–668 [vid. 2023-04-09]. ISSN 1098-1004. Dostupné z: doi:10.1002/humu.21490
- 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
- HON, Kha Wai, Nadiah ABU, Nurul-Syakima AB MUTALIB a Rahman JAMAL, 2018. miRNAs and lncRNAs as Predictive Biomarkers of Response to FOLFOX Therapy in Colorectal Cancer. *Frontiers in Pharmacology* [online]. **9** [vid. 2018-08-29]. ISSN 1663-9812. Dostupné z: doi:10.3389/fphar.2018.00846
- HOPF, Thomas A., John B. INGRAHAM, Frank J. POELWIJK, Charlotta P. I. SCHÄRFE, Michael SPRINGER, Chris SANDER a Debora S. MARKS, 2017. Mutation effects

predicted from sequence co-variation. *Nature Biotechnology* [online]. **35**(2), 128–135 [vid. 2023-07-07]. ISSN 1546-1696. Dostupné z: doi:10.1038/nbt.3769

- HOUDAYER, Claude, Virginie CAUX-MONCOUTIER, Sophie KRIEGER, Michel BARROIS, Françoise BONNET, Violaine BOURDON, Myriam BRONNER, Monique BUISSON, Florence COULET, Pascaline GAILDRAT, Cédrick LEFOL, Mélanie LÉONE, Sylvie MAZOYER, Danielle MULLER, Audrey REMENIERAS, Françoise RÉVILLION, Etienne ROULEAU, Joanna SOKOLOWSKA, Jean-Philippe VERT, Rosette LIDEREAU, Florent SOUBRIER, Hagay SOBOL, Nicolas SEVENET, Brigitte BRESSAC-DE PAILLERETS, Agnès HARDOUIN, Mario TOSI, Olga M. SINILNIKOVA a Dominique STOPPA-LYONNET, 2012. Guidelines for splicing analysis in molecular diagnosis derived from a set of 327 combined in silico/in vitro studies on BRCA1 and BRCA2 variants. *Human Mutation* [online]. **33**(8), 1228–1238 [vid. 2023-08-05]. ISSN 1098-1004. Dostupné z: doi:10.1002/humu.22101
- HU, Taishan, Nilesh CHITNIS, Dimitri MONOS a Anh DINH, 2021. Next-generation sequencing technologies: An overview. *Human Immunology* [online]. **82**(11), Next Generation Sequencing and its Application to Medical Laboratory Immunology, 801–811 [vid. 2023-06-27]. ISSN 0198-8859. Dostupné z: doi:10.1016/j.humimm.2021.02.012
- 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
- CHAN, Philip A., Sekhar DURAISAMY, Peter J. MILLER, Joan A. NEWELL, Carole MCBRIDE, Jeffrey P. BOND, Tiina RAEVAARA, Saara OLLILA, Minna NYSTRÖM, Andrew J. GRIMM, John CHRISTODOULOU, William S. OETTING a Marc S. GREENBLATT, 2007. Interpreting missense variants: comparing computational methods in human disease genes CDKN2A, MLH1, MSH2, MECP2, and tyrosinase (TYR). *Human Mutation* [online]. **28**(7), 683–693. ISSN 1098-1004. Dostupné z: doi:10.1002/humu.20492
- CHAO, Elizabeth C., Jonathan L. VELASQUEZ, Mavee S.L. WITHERSPOON, Laura S. ROZEK, David PEEL, Pauline NG, Stephen B. GRUBER, Patrice WATSON, Gad RENNERT, Hoda ANTON-CULVER, Henry LYNCH a Steven M. LIPKIN, 2008. Accurate classification of MLH1/MSH2 missense variants with multivariate analysis of protein polymorphisms–mismatch repair (MAPP-MMR). *Human Mutation* [online]. **29**(6), 852–860 [vid. 2023-07-08]. ISSN 1098-1004. Dostupné z: doi:10.1002/humu.20735
- CHEANG, Maggie C. U., Stephen K. CHIA, David VODUC, Dongxia GAO, Samuel LEUNG, Jacqueline SNIDER, Mark WATSON, Sherri DAVIES, Philip S. BERNARD, Joel S. PARKER, Charles M. PEROU, Matthew J. ELLIS a Torsten O. NIELSEN, 2009. Ki67 Index, HER2 Status, and Prognosis of Patients With Luminal B Breast Cancer. *JNCI Journal of the National Cancer Institute* [online]. **101**(10), 736–750 [vid. 2022-11-27]. ISSN 0027-8874. Dostupné z: doi:10.1093/jnci/djp082
- 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

- CHIN, Lauren, Beth DEVINE, Sarah BARADARAN, Katelyn KEYLOUN, William CANESTARO a Jonathan PHAM, 2017. Characterizing the Strength of Evidence in FDA Labels for Pharmacogenomic Biomarker-Guided Medication Use. *AMIA Joint Summits on Translational Science proceedings. AMIA Joint Summits on Translational Science*. **2017**, 30–39. ISSN 2153-4063.
- CHOI, Yongwook, Gregory E. SIMS, Sean MURPHY, Jason R. MILLER a Agnes P. CHAN, 2012. Predicting the Functional Effect of Amino Acid Substitutions and Indels. *PLOS ONE* [online]. **7**(10), e46688 [vid. 2019-10-30]. ISSN 1932-6203. Dostupné z: doi:10.1371/journal.pone.0046688
- CHOUDHURY, Arunabh, Taj MOHAMMAD, Farah ANJUM, Alaa SHAFIE, Indrakant K. SINGH, Bekhzod ABDULLAEV, Visweswara Rao PASUPULETI, Mohd ADNAN, Dharmendra Kumar YADAV a Md Imtaiyaz HASSAN, 2022. Comparative analysis of web-based programs for single amino acid substitutions in proteins. *PLOS ONE* [online]. **17**(5), e0267084 [vid. 2023-03-12]. ISSN 1932-6203. Dostupné z: doi:10.1371/journal.pone.0267084
- CHUN, Sung a Justin C. FAY, 2009. Identification of deleterious mutations within three human genomes. *Genome Research* [online]. **19**(9), 1553–1561 [vid. 2017-12-29]. ISSN 1088-9051. Dostupné z: doi:10.1101/gr.092619.109
- IKEDIABI, O. N., 2008. Somatic pharmacogenomics in cancer. *The pharmacogenomics journal*. **8**(5), 305–314.
- INGELMAN-SUNDBERG, Magnus, Souren MKRTCHIAN, Yitian ZHOU a Volker M. LAUSCHKE, 2018. Integrating rare genetic variants into pharmacogenetic drug response predictions. *Human Genomics* [online]. **12**(1) [vid. 2018-08-13]. ISSN 1479-7364. Dostupné z: doi:10.1186/s40246-018-0157-3
- 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
- IOANNIDIS, Nilah M., Joseph H. ROTHSTEIN, Vikas PEJAVER, Sumit MIDDHA, Shannon K. MCDONNELL, Saurabh BAHETI, Anthony MUSOLF, Qing LI, Emily HOLZINGER, Danielle KARYADI, Lisa A. CANNON-ALBRIGHT, Craig C. TEERLINK, Janet L. STANFORD, William B. ISAACS, Jianfeng XU, Kathleen A. COONEY, Ethan M. LANGE, Johanna SCHLEUTKER, John D. CARPTEN, Isaac J. POWELL, Olivier CUSSENOT, Geraldine CANCEL-TASSIN, Graham G. GILES, Robert J. MACINNIS, Christiane MAIER, Chih-Lin HSIEH, Fredrik WIKLUND, William J. CATALONA, William D. FOULKES, Diptasri MANDAL, Rosalind A. EELES, Zsofia KOTE-JARAI, Carlos D. BUSTAMANTE, Daniel J. SCHAID, Trevor HASTIE, Elaine A. OSTRANDER, Joan E. BAILEY-WILSON, Predrag RADIVOJAC, Stephen N. THIBODEAU, Alice S. WHITTEMORE a Weiva SIEH, 2016. REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. *American Journal of Human Genetics* [online]. **99**(4), 877–885. ISSN 1537-6605. Dostupné z: doi:10.1016/j.ajhg.2016.08.016
- IORIO, Francesco, Theo A. KNIJNENBURG, Daniel J. VIS, Graham R. BIGNELL, Michael P. MENDEN, Michael SCHUBERT, Nanne ABEN, Emanuel GONÇALVES, Syd BARTHORPE, Howard LIGHTFOOT, Thomas COKELAER, Patricia GRENINGER,

- Ewald VAN DYK, Han CHANG, Heshani DE SILVA, Holger HEYN, Xianming DENG, Regina K. EGAN, Qingsong LIU, Tatiana MIRONENKO, Xenia MITROPOULOS, Laura RICHARDSON, Jinhua WANG, Tinghu ZHANG, Sebastian MORAN, Sergi SAYOLS, Maryam SOLEIMANI, David TAMBORERO, Nuria LOPEZ-BIGAS, Petra ROSS-MACDONALD, Manel ESTELLER, Nathanael S. GRAY, Daniel A. HABER, Michael R. STRATTON, Cyril H. BENES, Lodewyk F.A. WESSELS, Julio SAEZ-RODRIGUEZ, Ultan MCDERMOTT a Mathew J. GARNETT, 2016. A Landscape of Pharmacogenomic Interactions in Cancer. *Cell* [online]. **166**(3), 740–754 [vid. 2017-11-13]. ISSN 0092-8674. Dostupné z: doi:10.1016/j.cell.2016.06.017
- JAGADEESH, Karthik A., Aaron M. WENGER, Mark J. BERGER, Harendra GUTURU, Peter D. STENSON, David N. COOPER, Jonathan A. BERNSTEIN a Gill BEJERANO, 2016. M-CAP eliminates a majority of variants of uncertain significance in clinical exomes at high sensitivity. *Nature Genetics* [online]. **48**(12), 1581–1586. ISSN 1546-1718. Dostupné z: doi:10.1038/ng.3703
- JALALI SEFID DASHTI, Mahjoubeh a Junaid GAMIELDIEN, 2017. A practical guide to filtering and prioritizing genetic variants. *BioTechniques* [online]. **62**(1) [vid. 2017-11-13]. ISSN 1940-9818, 0736-6205. Dostupné z: doi:10.2144/000114492
- JIAN, Xueqiu, Eric BOERWINKLE a Xiaoming LIU, 2014. In silico tools for splicing defect prediction - A survey from the viewpoint of end-users. *Genetics in medicine : official journal of the American College of Medical Genetics* [online]. **16**(7), 497–503 [vid. 2018-01-14]. ISSN 1098-3600. Dostupné z: doi:10.1038/gim.2013.176
- JONES, Clyde M., Samuel CHARACHE a Patricia J. HATHAWAY, 1979. The Effect of Hemoglobin F-Chesapeake ($\alpha 292$ Arg. \rightarrow Leu $\gamma 2$) 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
- KARCZEWSKI, Konrad J., Laurent C. FRANCIOLI, Grace TIAO, Beryl B. CUMMINGS, Jessica ALFÖLDI, Qingbo WANG, Ryan L. COLLINS, Kristen M. LARICCHIA, Andrea GANNA, Daniel P. BIRNBAUM, Laura D. GAUTHIER, Harrison BRAND, Matthew SOLOMONSON, Nicholas A. WATTS, Daniel RHODES, Moriel SINGER-BERK, Eleina M. ENGLAND, Eleanor G. SEABY, Jack A. KOSMICKI, Raymond K. WALTERS, Katherine TASHMAN, Yossi FARJOUN, Eric BANKS, Timothy POTERBA, Arcturus WANG, Cotton SEED, Nicola WHIFFIN, Jessica X. CHONG, Kaitlin E. SAMOCHA, Emma PIERCE-HOFFMAN, Zachary ZAPPALA, Anne H. O'DONNELL-LURIA, Eric Vallabh MINIKEL, Ben WEISBURD, Monkol LEK, James S. WARE, Christopher VITTAL, Irina M. ARMEAN, Louis BERGELSON, Kristian CIBULSKIS, Kristen M. CONNOLLY, Miguel COVARRUBIAS, Stacey DONNELLY, Steven FERRIERA, Stacey GABRIEL, Jeff GENTRY, Namrata GUPTA, Thibault JEANDET, Diane KAPLAN, Christopher LLANWARNE, Ruchi MUNSHI, Sam NOVOD, Nikelle PETRILLO, David

- ROAZEN, Valentin RUANO-RUBIO, Andrea SALTZMAN, Molly SCHLEICHER, Jose SOTO, Kathleen TIBBETTS, Charlotte TOLONEN, Gordon WADE, Michael E. TALKOWSKI, GENOME AGGREGATION DATABASE (GNOMAD) CONSORTIUM, Benjamin M. NEALE, Mark J. DALY a Daniel G. MACARTHUR, 2019. *The mutational constraint spectrum quantified from variation in 141,456 humans* [online]. preprint. B.m.: Genomics [vid. 2023-06-24]. Dostupné z: doi:10.1101/531210
- KATSONIS, Panagiotis, Kevin WILHELM, Amanda WILLIAMS a Olivier LICHTARGE, 2022. Genome interpretation using in silico predictors of variant impact. *Human Genetics* [online]. **141**(10), 1549–1577 [vid. 2023-05-28]. ISSN 1432-1203. Dostupné z: doi:10.1007/s00439-022-02457-6
- 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
- KIM, Yul, Chiyong KANG, Bumki MIN a Gwan-Su YI, 2015. Detection and analysis of disease-associated single nucleotide polymorphism influencing post-translational modification. *BMC Medical Genomics* [online]. **8**(2), S7 [vid. 2023-07-07]. ISSN 1755-8794. Dostupné z: doi:10.1186/1755-8794-8-S2-S7
- KIMURA, Motoo, 1979. The Neutral Theory of Molecular Evolution. *Scientific American* [online]. **241**(5), 98–129 [vid. 2023-07-08]. ISSN 0036-8733. Dostupné z: <https://www.jstor.org/stable/24965339>
- KIRCHER, Martin, Daniela M WITTEN, Preti JAIN, Brian J O'ROAK, Gregory M COOPER a Jay SHENDURE, 2014. A general framework for estimating the relative pathogenicity of human genetic variants. *Nature Genetics* [online]. **46**(3), 310–315 [vid. 2018-07-04]. ISSN 1546-1718. Dostupné z: doi:10.1038/ng.2892
- 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/>
- KNECHT, Carolin, Matthew MORT, Olaf JUNGE, David N. COOPER, Michael KRAWCZAK a Amke CALIEBE, 2017. IMHOTEP—a composite score integrating popular tools for predicting the functional consequences of non-synonymous sequence variants. *Nucleic Acids Research* [online]. **45**(3), e13–e13 [vid. 2018-02-25]. ISSN 0305-1048. Dostupné z: doi:10.1093/nar/gkw886
- 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
- KOZYRA, Mikael, Magnus INGELMAN-SUNDBERG a Volker M. LAUSCHKE, 2017. Rare genetic variants in cellular transporters, metabolic enzymes, and nuclear receptors can be important determinants of interindividual differences in drug response. *Genetics in Medicine* [online]. **19**(1), 20–29 [vid. 2019-01-06]. ISSN 1530-0366. Dostupné z: doi:10.1038/gim.2016.33
- KRANZLER, Henry R., Rachel V. SMITH, Robert SCHNOLL, Afaf MOUSTAFA a Emma GREENSTREET-AKMAN, 2017. Precision Medicine and Pharmacogenetics: What Does Oncology Have that Addiction Medicine Doesn't? *Addiction (Abingdon, England)* [online]. **112**(12), 2086–2094 [vid. 2023-08-05]. ISSN 0965-2140. Dostupné z: doi:10.1111/add.13818
- KREK, Azra, Dominic GRÜN, Matthew N. POY, Rachel WOLF, Lauren ROSENBERG, Eric J. EPSTEIN, Philip MACMENAMIN, Isabelle DA PIEDADE, Kristin C. GUNSALUS, Markus STOFFEL a Nikolaus RAJEWSKY, 2005. Combinatorial microRNA target predictions. *Nature Genetics* [online]. **37**(5), 495–500 [vid. 2023-08-05]. ISSN 1546-1718. Dostupné z: doi:10.1038/ng1536
- 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
- KUMAR, Prateek, Steven HENIKOFF a Pauline C. NG, 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature Protocols* [online]. **4**(7), 1073–1081. ISSN 1750-2799. Dostupné z: doi:10.1038/nprot.2009.86
- LAIMER, Josef, Heidi HOFER, Marko FRITZ, Stefan WEGENKITTL a Peter LACKNER, 2015. MAESTRO - multi agent stability prediction upon point mutations. *BMC Bioinformatics* [online]. **16**(1), 116 [vid. 2023-07-07]. ISSN 1471-2105. Dostupné z: doi:10.1186/s12859-015-0548-6

- LANDRUM, Melissa J, Jennifer M LEE, Mark BENSON, Garth R BROWN, Chen CHAO, Shanmuga CHITIPIRALLA, Baoshan GU, Jennifer HART, Douglas HOFFMAN, Wonhee JANG, Karen KARAPETYAN, Kenneth KATZ, Chunlei LIU, Zenith MADDIPATLA, Adriana MALHEIRO, Kurt MCDANIEL, Michael OVETSKY, George RILEY, George ZHOU, J Bradley HOLMES, Brandi L KATTMAN a Donna R MAGLOTT, 2018. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Research* [online]. **46**(Database issue), D1062–D1067 [vid. 2023-07-02]. ISSN 0305-1048. Dostupné z: doi:10.1093/nar/gkx1153
- LEMAN, Raphaël, Pascaline GAILDRAT, Gérald LE GAC, Chandran KA, Yann FICHO, Marie-Pierre AUDREZET, Virginie CAUX-MONCOUTIER, Sandrine M CAPUTO, Nadia BOUTRY-KRYZA, Mélanie LÉONE, Sylvie MAZOYER, Françoise BONNET-DORION, Nicolas SEVENET, Marine GUILLAUD-BATAILLE, Etienne ROULEAU, Brigitte BRESSAC-DE PAILLERETS, Barbara WAPPENSCHMIDT, Maria ROSSING, Danielle MULLER, Violaine BOURDON, Françoise REVILLON, Michael T PARSONS, Antoine ROUSSELIN, Grégoire DAVY, Gaia CASTELAIN, Laurent CASTÉRA, Joanna SOKOLOWSKA, Florence COULET, Capucine DELNATTE, Claude FÉREC, Amanda B SPURDLE, Alexandra MARTINS, Sophie KRIEGER a Claude HOUDAYER, 2018. Novel diagnostic tool for prediction of variant spliceogenicity derived from a set of 395 combined in silico/in vitro studies: an international collaborative effort. *Nucleic Acids Research* [online]. **46**(15), 7913–7923 [vid. 2022-08-06]. ISSN 0305-1048. Dostupné z: doi:10.1093/nar/gky372
- LEWIS, Benjamin P., Christopher B. BURGE a David P. BARTEL, 2005. Conserved Seed Pairing, Often Flanked by Adenosines, Indicates that Thousands of Human Genes are MicroRNA Targets. *Cell* [online]. **120**(1), 15–20 [vid. 2023-08-05]. ISSN 0092-8674. Dostupné z: doi:10.1016/j.cell.2004.12.035
- LI, Biao, Vidhya G. KRISHNAN, Matthew E. MORT, Fuxiao XIN, Kishore K. KAMATI, David N. COOPER, Sean D. MOONEY a Predrag RADIVOJAC, 2009. Automated inference of molecular mechanisms of disease from amino acid substitutions. *Bioinformatics* [online]. **25**(21), 2744–2750 [vid. 2023-06-23]. ISSN 1367-4803. Dostupné z: doi:10.1093/bioinformatics/btp528
- LI, Jinchun, Tingting ZHAO, Yi ZHANG, Kun ZHANG, Leisheng SHI, Yun CHEN, Xingxing WANG a Zhongsheng SUN, 2018. Performance evaluation of pathogenicity-computation methods for missense variants. *Nucleic Acids Research* [online]. **46**(15), 7793–7804 [vid. 2019-01-04]. ISSN 0305-1048. Dostupné z: doi:10.1093/nar/gky678
- 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
- LI, W. H., C. I. WU a C. C. LUO, 1984. Nonrandomness of point mutation as reflected in nucleotide substitutions in pseudogenes and its evolutionary implications. *Journal of Molecular Evolution* [online]. **21**(1), 58–71. ISSN 0022-2844. Dostupné z: doi:10.1007/BF02100628

- LIANTO, Priscilia, Samantha A. HUTCHINSON, J. Bernadette MOORE, Thomas A. HUGHES a James L. THORNE, 2021. Characterization and prognostic value of LXR splice variants in triple-negative breast cancer. *iScience* [online]. **24**(10), 103212 [vid. 2023-10-30]. ISSN 2589-0042. Dostupné z: doi:10.1016/j.isci.2021.103212
- 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
- LUO, Li, Eric BOERWINKLE a Momiao XIONG, 2011. Association studies for next-generation sequencing. *Genome Research* [online]. **21**(7), 1099–1108. ISSN 1549-5469. Dostupné z: doi:10.1101/gr.115998.110
- MACARTHUR, Daniel G., Suganthi BALASUBRAMANIAN, Adam FRANKISH, Ni HUANG, James MORRIS, Klaudia WALTER, Luke JOSTINS, Lukas HABEGGER, Joseph K. PICKRELL, Stephen B. MONTGOMERY, Cornelis A. ALBERS, Zhengdong D. ZHANG, Donald F. CONRAD, Gerton LUNTER, Hancheng ZHENG, Qasim AYUB, Mark A. DEPRISTO, Eric BANKS, Min HU, Robert E. HANDSAKER, Jeffrey A. ROSENFELD, Menachem FROMER, Mike JIN, Ximeng Jasmine MU, Ekta KHURANA, Kai YE, Mike KAY, Gary Ian SAUNDERS, Marie Marthe SUNER, Toby HUNT, If H.A. BARNES, Clara AMID, Denise R. CARVALHO-SILVA, Alexandra H. BIGNELL, Catherine SNOW, Bryndis YNGVADOTTIR, Suzannah BUMPSTEAD, David N. COOPER, Yali XUE, Irene Gallego ROMERO, Jun WANG, Yingrui LI, Richard A. GIBBS, Steven A. MCCARROLL, Emmanouil T. DERMITZAKIS, Jonathan K. PRITCHARD, Jeffrey C. BARRETT, Jennifer HARROW, Matthew E. HURLES, Mark B. GERSTEIN a Chris TYLER-SMITH, 2012. A systematic survey of loss-of-function variants in human protein-coding genes. *Science* [online]. **335**(6070), 823–828 [vid. 2023-07-07]. ISSN 0036-8075. Dostupné z: doi:10.1126/science.1215040
- 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

- MATHE, Ewy, Magali OLIVIER, Shunsuke KATO, Chikashi ISHIOKA, Pierre HAINAUT a Sean V. TAVTIGIAN, 2006. Computational approaches for predicting the biological effect of p53 missense mutations: a comparison of three sequence analysis based methods. *Nucleic Acids Research* [online]. **34**(5), 1317–1325. ISSN 1362-4962. Dostupné z: doi:10.1093/nar/gkj518
- MATHELIER, Anthony, Xiaobei ZHAO, Allen W. ZHANG, François PARCY, Rebecca WORSLEY-HUNT, David J. ARENILLAS, Sorana BUCHMAN, Chih-yu CHEN, Alice CHOU, Hans IENASESCU, Jonathan LIM, Casper SHYR, Ge TAN, Michelle ZHOU, Boris LENHARD, Albin SANDELIN a Wyeth W. WASSERMAN, 2014. JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. *Nucleic Acids Research* [online]. **42**(Database issue), D142-147. ISSN 1362-4962. Dostupné z: doi:10.1093/nar/gkt997
- 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
- MCLAREN, William, Laurent GIL, Sarah E. HUNT, Harpreet Singh RIAT, Graham R. S. RITCHIE, Anja THORMANN, Paul FLICEK a Fiona CUNNINGHAM, 2016. The Ensembl Variant Effect Predictor. *Genome Biology* [online]. **17** [vid. 2017-12-30]. ISSN 1474-7596. Dostupné z: doi:10.1186/s13059-016-0974-4
- 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
- MEULENDIJKS, Didier, Linda M. HENRICKS, Gabe S. SONKE, Maarten J. DEENEN, Tanja K. FROEHLICH, Ursula AMSTUTZ, Carlo R. LARGIADÈR, Barbara A. JENNINGS, Anthony M. MARINAKI, Jeremy D. SANDERSON, Zdenek KLEIBL, Petra KLEIBLOVA, Matthias SCHWAB, Ulrich M. ZANGER, Claire PALLES, Ian TOMLINSON, Eva GROSS, André B. P. van KUILENBURG, Cornelis J. A. PUNT, Miriam KOOPMAN, Jos H. BEIJNEN, Annemieke CATS a Jan H. M. SCHELLENS, 2015. Clinical relevance of DPYD variants c.1679T>G, c.1236G>A/HapB3, and c.1601G>A as predictors of severe fluoropyrimidine-associated toxicity: a systematic review and meta-analysis of individual patient data. *The Lancet Oncology* [online]. **16**(16), 1639–1650 [vid. 2017-11-03]. ISSN 1470-2045, 1474-5488. Dostupné z: doi:10.1016/S1470-2045(15)00286-7
- 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 [.beta.138 (H16) Ala .fwdarw. Pro]: an unstable variant near the C-terminus of the .beta.-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

- MORNINGSTAR-KYWI, Noam, Ian S. HAWORTH a Scott A. MOSLEY, 2021. Ligand-specific pharmacogenetic effects of nonsynonymous mutations. *Pharmacogenetics and Genomics* [online]. **31**(4), 75 [vid. 2023-07-07]. ISSN 1744-6872. Dostupné z: doi:10.1097/FPC.0000000000000424
- MORT, Matthew, Timothy STERNE-WEILER, Biao LI, Edward V. BALL, David N. COOPER, Predrag RADIVOJAC, Jeremy R. SANFORD a Sean D. MOONEY, 2014. MutPred Splice: machine learning-based prediction of exonic variants that disrupt splicing. *Genome Biology* [online]. **15**, R19 [vid. 2018-07-10]. ISSN 1474-760X. Dostupné z: doi:10.1186/gb-2014-15-1-r19
- MUDUNURI, Uma, Anney CHE, Ming YI a Robert M. STEPHENS, 2009. bioDBnet: the biological database network. *Bioinformatics* [online]. **25**(4), 555–556 [vid. 2019-11-01]. ISSN 1367-4803. Dostupné z: doi:10.1093/bioinformatics/btn654
- MUELLER, Sabine C., Christina BACKES, Jan HAAS the INHERITANCE Study Group, Hugo A. KATUS, Benjamin MEDER, Eckart MEESE a Andreas KELLER, 2015. Pathogenicity prediction of non-synonymous single nucleotide variants in dilated cardiomyopathy. *Briefings in Bioinformatics* [online]. **16**(5), 769–779 [vid. 2023-03-12]. ISSN 1467-5463. Dostupné z: doi:10.1093/bib/bbu054
- 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
- NG, Pauline C. a Steven HENIKOFF, 2003. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Research* [online]. **31**(13), 3812–3814 [vid. 2023-04-01]. ISSN 0305-1048. Dostupné z: doi:10.1093/nar/gkg509
- NG, Pauline C. a Steven HENIKOFF, 2006. Predicting the effects of amino acid substitutions on protein function. *Annual Review of Genomics and Human Genetics* [online]. **7**, 61–80. ISSN 1527-8204. Dostupné z: doi:10.1146/annurev.genom.7.080505.115630
- NIROULA, Abhishek a Mauno VIHINEN, 2017. PON-P and PON-P2 predictor performance in CAGI challenges: Lessons learned. *Human mutation* [online]. **38**(9), 1085–1091 [vid. 2022-11-13]. ISSN 1059-7794. Dostupné z: doi:10.1002/humu.23199
- NIROULA, Abhishek a Mauno VIHINEN, 2019. How good are pathogenicity predictors in detecting benign variants? *PLOS Computational Biology* [online]. **15**(2), e1006481 [vid. 2022-11-06]. ISSN 1553-7358. Dostupné z: doi:10.1371/journal.pcbi.1006481
- 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
- ORGOGOZO, Virginie, Baptiste MORIZOT a Arnaud MARTIN, 2015. The differential view of genotype–phenotype relationships. *Frontiers in Genetics* [online]. **6** [vid. 2023-07-02]. ISSN 1664-8021. Dostupné z: <https://www.frontiersin.org/articles/10.3389/fgene.2015.00179>
- PARK, Jihye, Soo Youn LEE, Su Youn BAIK, Chan Hee PARK, Jun Hee YOON, Brian Y. RYU a Ju Han KIM, 2020. Gene-Wise Burden of Coding Variants Correlates to Noncoding Pharmacogenetic Risk Variants. *International Journal of Molecular Sciences* [online]. **21**(9), 3091 [vid. 2023-07-14]. ISSN 1422-0067. Dostupné z: doi:10.3390/ijms21093091
- 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
- PETUKH, Marharyta, Tugba G. KUCUKKAL a Emil ALEXOV, 2015. On Human Disease-Causing Amino Acid Variants: Statistical Study of Sequence and Structural Patterns. *Human Mutation* [online]. **36**(5), 524–534 [vid. 2023-07-07]. ISSN 1098-1004. Dostupné z: doi:10.1002/humu.22770
- PIRES, Allan S., William F. PORTO, Octavio L. FRANCO a Sérgio A. ALENCAR, 2017. In silico analyses of deleterious missense SNPs of human apolipoprotein E3. *Scientific Reports* [online]. **7** [vid. 2018-01-21]. ISSN 2045-2322. Dostupné z: doi:10.1038/s41598-017-01737-w
- QORRI, Erda, Bertalan TAKÁCS, Alexandra GRÁF, Márton Zsolt ENYEDI, Lajos PINTÉR, Ernő KISS a Lajos HARACSKA, 2022. A Comprehensive Evaluation of the Performance of Prediction Algorithms on Clinically Relevant Missense Variants. *International Journal of Molecular Sciences* [online]. **23**(14), 7946 [vid. 2023-03-12]. ISSN 1422-0067. Dostupné z: doi:10.3390/ijms23147946
- QUANG, Daniel, Yifei CHEN a Xiaohui XIE, 2015. DANN: a deep learning approach for annotating the pathogenicity of genetic variants. *Bioinformatics* [online]. **31**(5), 761–763 [vid. 2023-07-07]. ISSN 1367-4803. Dostupné z: doi:10.1093/bioinformatics/btu703
- 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
- RENTZSCH, Philipp, Daniela WITTEN, Gregory M. COOPER, Jay SHENDURE a Martin KIRCHER, 2019. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Research* [online]. **47**(D1), D886–D894 [vid. 2019-03-06]. ISSN 0305-1048. Dostupné z: doi:10.1093/nar/gky1016
- REVA, Boris, Yevgeniy ANTIPIN a Chris SANDER, 2007. Determinants of protein function revealed by combinatorial entropy optimization. *Genome Biology* [online]. **8**(11), R232 [vid. 2023-07-07]. ISSN 1474-760X. Dostupné z: doi:10.1186/gb-2007-8-11-r232
- REVA, Boris, Yevgeniy ANTIPIN a Chris SANDER, 2011. Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Research* [online]. **39**(17), e118 [vid. 2023-02-28]. ISSN 0305-1048. Dostupné z: doi:10.1093/nar/gkr407
- RIEDER, Mark J., Alexander P. REINER, Brian F. GAGE, Deborah A. NICKERSON, Charles S. EBY, Howard L. MCLEOD, David K. BLOUGH, Kenneth E. THUMMEL, David L. VEENSTRA a Allan E. RETTIE, 2005. Effect of VKORC1 haplotypes on transcriptional regulation and warfarin dose. *The New England Journal of Medicine* [online]. **352**(22), 2285–2293. ISSN 1533-4406. Dostupné z: doi:10.1056/NEJMoa044503
- RICHARDS, Sue, 2015. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology.
- ROBINSON, Peter N., Sebastian KÖHLER, Sebastian BAUER, Dominik SEELow, Denise HORN a Stefan MUNDLOS, 2008. The Human Phenotype Ontology: A Tool for Annotating and Analyzing Human Hereditary Disease. *The American Journal of Human Genetics* [online]. **83**(5), 610–615 [vid. 2023-07-07]. ISSN 0002-9297. Dostupné z: doi:10.1016/j.ajhg.2008.09.017
- 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

- SAITO, Seiji, Kazuki OHNO a Hitoshi SAKURABA, 2011. Fabry-database.org: database of the clinical phenotypes, genotypes and mutant α -galactosidase A structures in Fabry disease. *Journal of Human Genetics* [online]. **56**(6), 467–468 [vid. 2023-07-07]. ISSN 1435-232X. Dostupné z: doi:10.1038/jhg.2011.31
- 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
- SHIOVITZ, S. a L. A. KORDE, 2015. Genetics of breast cancer: a topic in evolution. *Annals of Oncology* [online]. **26**(7), 1291–1299 [vid. 2023-07-29]. ISSN 0923-7534. Dostupné z: doi:10.1093/annonc/mdv022
- SCHAID, Daniel J., Wenan CHEN a Nicholas B. LARSON, 2018. From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nature Reviews Genetics* [online]. **19**(8), 491–504 [vid. 2023-07-11]. ISSN 1471-0064. Dostupné z: doi:10.1038/s41576-018-0016-z
- 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
- SCHWARTZ, Lawrence H., Saskia LITIÈRE, Elisabeth DE VRIES, Robert FORD, Stephen GWYTHYER, Sumithra MANDREKAR, Lalitha SHANKAR, Jan BOGAERTS, Alice CHEN, Janet DANCEY, Wendy HAYES, F. Stephen HODI, Otto S. HOEKSTRA, Erich P. HUANG, Nancy LIN, Yan LIU, Patrick THERASSE, Jedd D. WOLCHOK a Lesley SEYMOUR, 2016. RECIST 1.1-Update and clarification: From the RECIST committee. *European Journal of Cancer (Oxford, England: 1990)* [online]. **62**, 132–137. ISSN 1879-0852. Dostupné z: doi:10.1016/j.ejca.2016.03.081
- SCHWARZ, Jana Marie, Christian RÖDELSPERGER, Markus SCHUELKE a Dominik SEELOW, 2010. MutationTaster evaluates disease-causing potential of sequence alterations. *Nature Methods* [online]. **7**(8), 575–576. ISSN 1548-7105. Dostupné z: doi:10.1038/nmeth0810-575
- 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
- SMIGIELSKI, E. M., K. SIROTKIN, M. WARD a S. T. SHERRY, 2000. dbSNP: a database of single nucleotide polymorphisms. *Nucleic Acids Research* [online]. **28**(1), 352–355. ISSN 0305-1048. Dostupné z: doi:10.1093/nar/28.1.352
- 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. PharmacDB: 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

- SMITH, Cynthia L., Carroll-Ann W. GOLDSMITH a Janan T. EPPIG, 2004. The Mammalian Phenotype Ontology as a tool for annotating, analyzing and comparing phenotypic information. *Genome Biology* [online]. **6**(1), R7 [vid. 2023-07-07]. ISSN 1474-760X. Dostupné z: doi:10.1186/gb-2004-6-1-r7
- STARK, Chris, Bobby-Joe BREITKREUTZ, Teresa REGULY, Lorrie BOUCHER, Ashton BREITKREUTZ a Mike TYERS, 2006. BioGRID: a general repository for interaction datasets. *Nucleic Acids Research* [online]. **34**(Database issue), D535-539. ISSN 1362-4962. Dostupné z: doi:10.1093/nar/gkj109
- STRIMBU, Kyle a Jorge A. TAVEL, 2010. What are Biomarkers? *Current opinion in HIV and AIDS* [online]. **5**(6), 463–466 [vid. 2023-07-31]. ISSN 1746-630X. Dostupné z: doi:10.1097/COH.0b013e32833ed177
- SUDMANT, Peter H., Tobias RAUSCH, Eugene J. GARDNER, Robert E. HANDSAKER, Alexej ABYZOV, John HUDDLESTON, Yan ZHANG, Kai YE, Goo JUN, Markus HSI-YANG FRITZ, Miriam K. KONKEL, Ankit MALHOTRA, Adrian M. STÜTZ, Xinghua SHI, Francesco PAOLO CASALE, Jieming CHEN, Fereydoun HORMOZDIARI, Gargi DAYAMA, Ken CHEN, Maika MALIG, Mark J. P. CHAISSON, Klaudia WALTER, Sascha MEIERS, Seva KASHIN, Erik GARRISON, Adam AUTON, Hugo Y. K. LAM, Ximeng JASMINE MU, Can ALKAN, Danny ANTAKI, Taejeong BAE, Eliza CERVEIRA, Peter CHINES, Zechen CHONG, Laura CLARKE, Elif DAL, Li DING, Sarah EMERY, Xian FAN, Madhusudan GUJRAL, Fatma KAHVECI, Jeffrey M. KIDD, Yu KONG, Eric-Wubbo LAMEIJER, Shane MCCARTHY, Paul FLICEK, Richard A. GIBBS, Gabor MARTH, Christopher E. MASON, Androniki MENELAOU, Donna M. MUZNY, Bradley J. NELSON, Amina NOOR, Nicholas F. PARRISH, Matthew PENDLETON, Andrew QUITADAMO, Benjamin RAEDER, Eric E. SCHADT, Mallory ROMANOVITCH, Andreas SCHLÄTTL, Robert SEBRA, Andrey A. SHABALIN, Andreas UNTERGASSER, Jerilyn A. WALKER, Min WANG, Fuli YU, Chengsheng ZHANG, Jing ZHANG, Xiangqun ZHENG-BRADLEY, Wanding ZHOU, Thomas ZICHNER, Jonathan SEBAT, Mark A. BATZER, Steven A. MCCARROLL, Ryan E. MILLS, Mark B. GERSTEIN, Ali BASHIR, Oliver STEGLE, Scott E. DEVINE, Charles LEE, Evan E. EICHLER a Jan O. KORBEL, 2015. An integrated map of structural variation in 2,504 human genomes. *Nature* [online]. **526**(7571), 75–81 [vid. 2023-08-23]. ISSN 1476-4687. Dostupné z: doi:10.1038/nature15394
- SUNYAEV, Shamil R., Frank EISENHABER, Igor V. RODCHENKOV, Birgit EISENHABER, Vladimir G. TUMANYAN a Eugene N. KUZNETSOV, 1999. PSIC: profile extraction from sequence alignments with position-specific counts of independent observations. *Protein Engineering, Design and Selection* [online]. **12**(5), 387–394 [vid. 2018-07-04]. ISSN 1741-0134, 1741-0126. Dostupné z: doi:10.1093/protein/12.5.387
- SUZEK, Baris E., Hongzhan HUANG, Peter MCGARVEY, Raja MAZUMDER a Cathy H. WU, 2007. UniRef: comprehensive and non-redundant UniProt reference clusters. *Bioinformatics* [online]. **23**(10), 1282–1288 [vid. 2023-08-11]. ISSN 1367-4803. Dostupné z: doi:10.1093/bioinformatics/btm098
- TAVTIGIAN, Sean V., Graham B. BYRNES, David E. GOLDGAR a Alun THOMAS, 2008. Classification of Rare Missense Substitutions, Using Risk Surfaces, With Genetic- and Molecular-Epidemiology Applications. *Human mutation* [online]. **29**(11), 1342–1354 [vid. 2023-07-07]. ISSN 1059-7794. Dostupné z: doi:10.1002/humu.20896

- THUSBERG, Janita, Ayodeji OLATUBOSUN a Mauno VIHINEN, 2011. Performance of mutation pathogenicity prediction methods on missense variants. *Human Mutation* [online]. **32**(4), 358–368 [vid. 2017-11-23]. ISSN 1098-1004. Dostupné z: doi:10.1002/humu.21445
- TIAN, Yuan, Tina PESARAN, Adam CHAMBERLIN, R. Bryn FENWICK, Shuwei LI, Chia-Ling GAU, Elizabeth C. CHAO, Hsiao-Mei LU, Mary Helen BLACK a Dajun QIAN, 2019. REVEL and BayesDel outperform other in silico meta-predictors for clinical variant classification. *Scientific Reports* [online]. **9**(1), 12752. ISSN 2045-2322. Dostupné z: doi:10.1038/s41598-019-49224-8
- UZILOV, Andrew V., Wei DING, Marc Y. FINK, Yevgeniy ANTIPIN, Andrew S. BROHL, Claire DAVIS, Chun Yee LAU, Chetanya PANDYA, Hardik SHAH, Yumi KASAI, James POWELL, Mark MICCHELLI, Rafael CASTELLANOS, Zhongyang ZHANG, Michael LINDERMAN, Yayoi KINOSHITA, Micol ZWEIG, Katie RAUSTAD, Kakit CHEUNG, Diane CASTILLO, Melissa WOOTEN, Imane BOURZGUI, Leah C. NEWMAN, Gintaras DEIKUS, Bino MATHEW, Jun ZHU, Benjamin S. GLICKSBERG, Aye S. MOE, Jun LIAO, Lisa EDELMANN, Joel T. DUDLEY, Robert G. MAKI, Andrew KASARSKIS, Randall F. HOLCOMBE, Milind MAHAJAN, Ke HAO, Boris REVA, Janina LONGTINE, Daniela STARCEVIC, Robert SEBRA, Michael J. DONOVAN, Shuyu LI, Eric E. SCHADT a Rong CHEN, 2016. Development and clinical application of an integrative genomic approach to personalized cancer therapy. *Genome Medicine* [online]. **8** [vid. 2017-11-13]. ISSN 1756-994X. Dostupné z: doi:10.1186/s13073-016-0313-0
- 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
- VIHINEN, Mauno, 2012. How to evaluate performance of prediction methods? Measures and their interpretation in variation effect analysis. *BMC Genomics* [online]. **13**(4), S2 [vid. 2023-04-09]. ISSN 1471-2164. Dostupné z: doi:10.1186/1471-2164-13-S4-S2
- 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
- WALTER, Klaudia, Josine L. MIN, Jie HUANG, Lucy CROOKS, Yasin MEMARI, Shane MCCARTHY, John R. B. PERRY, ChangJiang XU, Marta FUTEMA, Daniel LAWSON, Valentina IOTCHKOVA, Stephan SCHIFFELS, Audrey E. HENDRICKS, Petr DANECEK, Rui LI, James FLOYD, Louise V. WAIN, Inês BARROSO, Steve E. HUMPHRIES, Matthew E. HURLES, Eleftheria ZEGGINI, Jeffrey C. BARRETT, Vincent PLAGNOL, J. BRENT RICHARDS, Celia M. T. GREENWOOD, Nicholas J. TIMPSON, Richard DURBIN, Nicole SORANZO, Senduran BALA, Peter CLAPHAM, Guy COATES, Tony COX, Allan DALY, Petr DANECEK, Yuanping DU, Richard DURBIN, Sarah EDKINS, Peter ELLIS, Paul FLICEK, Xiaosen GUO, Xueqin GUO, Liren HUANG, David K.

JACKSON, Chris JOYCE, Thomas KEANE, Anja KOLB-KOKOCINSKI, Cordelia LANGFORD, Yingrui LI, Jieqin LIANG, Hong LIN, Ryan LIU, John MASLEN, Shane MCCARTHY, Dawn MUDDYMAN, Michael A. QUAIL, Jim STALKER, Jianping SUN, Jing TIAN, Guangbiao WANG, Jun WANG, Yu WANG, Kim WONG, Pingbo ZHANG, Inês BARROSO, Ewan BIRNEY, Chris BOUSTRED, Lu CHEN, Gail CLEMENT, Massimiliano COCCA, Petr DANECEK, George DAVEY SMITH, Ian N. M. DAY, Aaron DAY-WILLIAMS, Thomas DOWN, Ian DUNHAM, Richard DURBIN, David M. EVANS, Tom R. GAUNT, Matthias GEIHS, Celia M. T. GREENWOOD, Deborah HART, Audrey E. HENDRICKS, Bryan HOWIE, Jie HUANG, Tim HUBBARD, Pirro HYSI, Valentina IOTCHKOVA, Yalda JAMSHIDI, Konrad J. KARCZEWSKI, John P. KEMP, Genevieve LACHANCE, Daniel LAWSON, Monkol LEK, Margarida LOPES, Daniel G. MACARTHUR, Jonathan MARCHINI, Massimo MANGINO, Iain MATHIESON, Shane MCCARTHY, Yasin MEMARI, Sarah METRUSTRY, Josine L. MIN, Alireza MOAYYERI, Dawn MUDDYMAN, Kate NORTHSTONE, Kalliope PANOUTSOPOULOU, Lavinia PATERNOSTER, John R. B. PERRY, Lydia QUAYE, J. BRENT RICHARDS, Susan RING, Graham R. S. RITCHIE, Stephan SCHIFFELS, Hashem A. SHIHAB, So-Youn SHIN, Kerrin S. SMALL, María SOLER ARTIGAS, Nicole SORANZO, Lorraine SOUTHAM, Timothy D. SPECTOR, Beate ST POURCAIN, Gabriela SURDULESCU, Ioanna TACHMAZIDOU, Nicholas J. TIMPSON, Martin D. TOBIN, Ana M. VALDES, Peter M. VISSCHER, Louise V. WAIN, Klaudia WALTER, Kirsten WARD, Scott G. WILSON, Kim WONG, Jian YANG, Eleftheria ZEGGINI, Feng ZHANG, Hou-Feng ZHENG, Richard ANNEY, Muhammad AYUB, Jeffrey C. BARRETT, Douglas BLACKWOOD, Patrick F. BOLTON, Gerome BREEN, David A. COLLIER, Nick CRADDOCK, Lucy CROOKS, Sarah CURRAN, David CURTIS, Richard DURBIN, Louise GALLAGHER, Daniel GESCHWIND, Hugh GURLING, Peter HOLMANS, Irene LEE, Jouko LÖNNQVIST, Shane MCCARTHY, Peter MCGUFFIN, Andrew M. MCINTOSH, Andrew G. MCKECHANIE, Andrew MCQUILLIN, James MORRIS, Dawn MUDDYMAN, Michael C. O'DONOVAN, Michael J. OWEN, Aarno PALOTIE, Jeremy R. PARR, Tiina PAUNIO, Olli PIETILAINEN, Karola REHNSTRÖM, Sally I. SHARP, David SKUSE, David ST CLAIR, Jaana SUVISAARI, James T. R. WALTERS, Hywel J. WILLIAMS, Inês BARROSO, Elena BOCHUKOVA, Rebecca BOUNDS, Anna DOMINICZAK, Richard DURBIN, I. Sadaf FAROOQI, Audrey E. HENDRICKS, Julia KEOGH, Gaëlle MARENNE, Shane MCCARTHY, Andrew MORRIS, Dawn MUDDYMAN, Stephen O'RAHILLY, David J. PORTEOUS, Blair H. SMITH, Ioanna TACHMAZIDOU, Eleanor WHEELER, Eleftheria ZEGGINI, Saeed AL TURKI, Carl A. ANDERSON, Dinu ANTONY, Inês BARROSO, Phil BEALES, Jamie BENTHAM, Shoumo BHATTACHARYA, Mattia CALISSANO, Keren CARSS, Krishna CHATTERJEE, Sebahattin CIRAK, Catherine COSGROVE, Richard DURBIN, David R. FITZPATRICK, James FLOYD, A. REGHAN FOLEY, Christopher S. FRANKLIN, Marta FUTEMA, Detelina GROZEVA, Steve E. HUMPHRIES, Matthew E. HURLES, Shane MCCARTHY, Hannah M. MITCHISON, Dawn MUDDYMAN, Francesco MUNTONI, Stephen O'RAHILLY, Alexandros ONOUFRIADIS, Victoria PARKER, Felicity PAYNE, Vincent PLAGNOL, F. LUCY RAYMOND, Nicola ROBERTS, David B. SAVAGE, Peter SCAMBLER, Miriam SCHMIDTS, Nadia SCHOENMAKERS, Robert K. SEMPLE, Eva SERRA, Olivera SPASIC-BOSKOVIC, Elizabeth STEVENS, Margriet VAN KOGELBERG, Parthiban VIJAYARANGAKANNAN, Klaudia WALTER, Kathleen A. WILLIAMSON, Crispian WILSON, Tamiaka WHYTE, Antonio CIAMPI, Celia M. T. GREENWOOD, Audrey E. HENDRICKS, Rui LI, Sarah METRUSTRY, Karim OUALKACHA, Ioanna TACHMAZIDOU, ChangJiang XU, Eleftheria ZEGGINI, Martin BOBROW, Patrick F. BOLTON, Richard DURBIN, David R. FITZPATRICK, Heather

- GRIFFIN, Matthew E. HURLES, Jane KAYE, Karen KENNEDY, Alastair KENT, Dawn MUDDYMAN, Francesco MUNTONI, F. LUCY RAYMOND, Robert K. SEMPLE, Carol SMEE, Timothy D. SPECTOR, Nicholas J. TIMPSON, Ruth CHARLTON, Rosemary EKONG, Marta FUTEMA, Steve E. HUMPHRIES, Farrah KHAWAJA, Luis R. LOPES, Nicola MIGONE, Stewart J. PAYNE, Vincent PLAGNOL, Rebecca C. POLLITT, Sue POVEY, Cheryl K. RIDOUT, Rachel L. ROBINSON, Richard H. SCOTT, Adam SHAW, Petros SYRRIS, Rohan TAYLOR, Anthony M. VANDERSTEEN, Jeffrey C. BARRETT, Inês BARROSO, George DAVEY SMITH, Richard DURBIN, I. Sadaf FAROOQI, David R. FITZPATRICK, Matthew E. HURLES, Jane KAYE, THE UK10K CONSORTIUM, WRITING GROUP, PRODUCTION GROUP, COHORTS GROUP, NEURODEVELOPMENTAL DISORDERS GROUP, OBESITY GROUP, RARE DISEASE GROUP, STATISTICS GROUP, ETHICS GROUP, INCIDENTAL FINDINGS GROUP, a MANAGEMENT COMMITTEE, 2015. The UK10K project identifies rare variants in health and disease. *Nature* [online]. **526**(7571), 82–90 [vid. 2023-08-23]. ISSN 1476-4687. Dostupné z: doi:10.1038/nature14962
- WANG, Jun, Abu Z. DAYEM ULLAH a Claude CHELALA, 2018a. IW-Scoring: an Integrative Weighted Scoring framework for annotating and prioritizing genetic variations in the noncoding genome. *Nucleic Acids Research* [online]. **46**(8), e47–e47 [vid. 2018-05-09]. ISSN 0305-1048. Dostupné z: doi:10.1093/nar/gky057
- WANG, Jun, Abu Z DAYEM ULLAH a Claude CHELALA, 2018b. IW-Scoring: an Integrative Weighted Scoring framework for annotating and prioritizing genetic variations in the noncoding genome. *Nucleic Acids Research* [online]. **46**(8), e47 [vid. 2023-06-30]. ISSN 0305-1048. Dostupné z: doi:10.1093/nar/gky057
- WANG, Kai, Mingyao LI a Hakon HAKONARSON, 2010. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Research* [online]. **38**(16), e164–e164 [vid. 2019-10-31]. ISSN 0305-1048. Dostupné z: doi:10.1093/nar/gkq603
- WEI, Peng, Xiaoming LIU a Yun-Xin FU, 2011. Incorporating predicted functions of nonsynonymous variants into gene-based analysis of exome sequencing data: a comparative study. *BMC Proceedings* [online]. **5**(Suppl 9), S20 [vid. 2017-12-29]. ISSN 1753-6561. Dostupné z: doi:10.1186/1753-6561-5-S9-S20
- WENG, Liming, Li ZHANG, Yan PENG a R Stephanie HUANG, 2013. Pharmacogenetics and pharmacogenomics: a bridge to individualized cancer therapy. *Pharmacogenomics* [online]. **14**(3), 315–324 [vid. 2017-11-13]. ISSN 1462-2416. Dostupné z: doi:10.2217/pgs.12.213
- WHALEN, Sean, Jacob SCHREIBER, William S. NOBLE a Katherine S. POLLARD, 2022. Navigating the pitfalls of applying machine learning in genomics. *Nature Reviews Genetics* [online]. **23**(3), 169–181 [vid. 2023-05-28]. ISSN 1471-0064. Dostupné z: doi:10.1038/s41576-021-00434-9
- WHEELER, Heather E., Michael L. MAITLAND, M. Eileen DOLAN, Nancy J. COX a Mark J. RATAIN, 2013. Cancer pharmacogenomics: strategies and challenges. *Nature Reviews Genetics* [online]. **14**(1), 23–34 [vid. 2022-01-09]. ISSN 1471-0056, 1471-0064. Dostupné z: doi:10.1038/nrg3352

- WINGENDER, E., P. DIETZE, H. KARAS a R. KNÜPPEL, 1996. TRANSFAC: a database on transcription factors and their DNA binding sites. *Nucleic Acids Research* [online]. **24**(1), 238–241. ISSN 0305-1048. Dostupné z: doi:10.1093/nar/24.1.238
- 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
- WU, Wen-Tao, Yuan-Jie LI, Ao-Zi FENG, Li LI, Tao HUANG, An-Ding XU a Jun LYU, 2021. Data mining in clinical big data: the frequently used databases, steps, and methodological models. *Military Medical Research* [online]. **8**(1), 44. ISSN 2054-9369. Dostupné z: doi:10.1186/s40779-021-00338-z
- 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.compbiomed.2022.105434
- YUAN, Xiao, Jing WANG, Bing DAI, Yanfang SUN, Keke ZHANG, Fangfang CHEN, Qian PENG, Yixuan HUANG, Xinlei ZHANG, Junru CHEN, Xilin XU, Jun CHUAN, Wenbo MU, Huiyuan LI, Ping FANG, Qiang GONG a Peng ZHANG, 2022. Evaluation of phenotype-driven gene prioritization methods for Mendelian diseases. *Briefings in Bioinformatics* [online]. **23**(2), bbac019 [vid. 2023-04-01]. ISSN 1477-4054. Dostupné z: doi:10.1093/bib/bbac019
- YUE, Peng, Zhaolong LI a John MOULT, 2005. Loss of Protein Structure Stability as a Major Causative Factor in Monogenic Disease. *Journal of Molecular Biology* [online]. **353**(2), 459–473 [vid. 2023-07-07]. ISSN 0022-2836. Dostupné z: doi:10.1016/j.jmb.2005.08.020
- YUE, Wyatt W., D. Sean FROESE a Paul E. BRENNAN, 2014. The Role of Protein Structural Analysis in the Next Generation Sequencing Era. In: Nelson L.S. TANG a Terence POON, ed. *Chemical Diagnostics: From Bench to Bedside* [online]. Berlin, Heidelberg: Springer, Topics in Current Chemistry, s. 67–98 [vid. 2023-07-07]. ISBN 978-3-642-39942-8. Dostupné z: doi:10.1007/128_2012_326
- ZEGGINI, Eleftheria, 2011. Next-generation association studies for complex traits. *Nature Genetics* [online]. **43**(4), 287–288. ISSN 1546-1718. Dostupné z: doi:10.1038/ng0411-287
- 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, Hongyi, Mu GAO a Jeffrey SKOLNICK, 2016. ENTPRISE: An Algorithm for Predicting Human Disease-Associated Amino Acid Substitutions from Sequence Entropy and Predicted Protein Structures. *PLOS ONE* [online]. **11**(3), e0150965 [vid. 2023-07-07]. ISSN 1932-6203. Dostupné z: doi:10.1371/journal.pone.0150965

ZHOU, Yitian a Volker M. LAUSCHKE, 2021. Computational Tools to Assess the Functional Consequences of Rare and Noncoding Pharmacogenetic Variability. *Clinical Pharmacology & Therapeutics* [online]. **110**(3), 626–636 [vid. 2023-05-21]. ISSN 1532-6535. Dostupné z: doi:10.1002/cpt.2289

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