



Záverečná karta projektu

Názov projektu

Evidenčné číslo projektu

APVV-18-0319

Vývoj a testovanie molekulárnych a informatických metód na efektívnu charakterizáciu a interpretáciu klinicky relevantných mikrosatelitových repetitívnych motívov z genomických dát

Zodpovedný riešiteľ **Mgr. Jaroslav Budiš, PhD.**

Príjemca **GENETON s.r.o.**

Názov pracoviska, na ktorom bol projekt riešený

Geneton s.r.o.

Vedecký park Univerzity Komenského v Bratislave

Biomedicínske centrum Slovenskej Akadémie Vied

Názov a štát zahraničného pracoviska, ktoré spolupracovalo pri riešení

-

Udelené patenty/podané patentové prihlášky, vynálezy alebo úžitkové vzory, ktoré sú výsledkami projektu

-

Najvýznamnejšie publikácie (knihy, články, prednášky, správy a pod.) zhrňujúce výsledky projektu – uveďte aj publikácie prijaté do tlače

Radvanszky J, Hyblova M, Radvanska E, Spalek P, Valachova A, Magyarova G, Bogнар C, Polak E, Szemes T, Kadasi L. Characterisation of Non-Pathogenic Premutation-Range Myotonic Dystrophy Type 2 Alleles. Journal of clinical medicine. 2021 Jan;10(17):3934.

Styk J, Pös Z, Pös O, Radvanszky J, Turnova EH, Buglyó G, Klimova D, Budis J, Repiska V, Nagy B, Szemes T. Microsatellite instability assessment is instrumental for Predictive, Preventive and Personalised Medicine: status quo and outlook. EPMA Journal. 2023 Mar;14(1):143-65.

Sitarčík J, Vinař T, Brejová B, Krámpf W, Budiš J, Radvánszky J, Lucká M. WarpSTR: Determining tandem repeat lengths using raw nanopore signals. Bioinformatics. 2023 Jun 1;39(6):btad388.

Pös Z, Pös O, Styk J, Mocova A, Strieskova L, Budis J, Kadasi L, Radvanszky J, Szemes T. Technical and Methodological Aspects of Cell-Free Nucleic Acids Analyses. International Journal of Molecular Sciences. 2020; 21(22):8634.

Pös O, Radvanszky J, Buglyó G, Pös Z, Rusnakova D, Nagy B, Szemes T. DNA copy number variation: Main characteristics, evolutionary significance, and pathological aspects. Biomedical Journal. 2021 Oct;44(5):548.

Gažiová M, Sládeček T, Pös O, Števkó M, Krámpf W, Pös Z, Hekel R, Hlavačka M, Kucharik M, Radvanszky J, Budiš J. Automated prediction of the clinical impact of structural copy number variations. Scientific Reports. 2022 Jan 11;12(1):1-4

Sládeček T, Gažiová M, Kucharík M, Zatlková A, Pös Z, Pös O, Krampfl W, Tomková E, Hýblová M, Minárik G, Radvánszky J. Combination of expert guidelines-based and machine learning-based approaches leads to superior accuracy of automated prediction of clinical effect of copy number variations. *Scientific Reports*. 2023 Jun

Styk J. Detekcia mikrosatelitnej instability v nádoroch asociovaných s Lynchovým syndrómom. *NewsLab*. 2021, 01 : 27-31

Radvanska E, Pos Z, Zatkova A, Hyblova M, Bauer F, Szemes T, Kadasi L, Radvanszky J. Molecularly confirmed pontocerebellar hypoplasia in a large family from Slovakia with four severely affected children. *Bratislavske Lekarske Listy*. 2022 Jan 1;123(8):568-72.

Pos Z, Khedr M, Radvanszky J, Penesova A, Hekel R, Szemes T, Ranganath LR, Zatkova A. APOC3 and ABCA1 variants in unusual combined hypolipidaemia showing premature peripheral vascular disease. *Bratislava Medical Journal/Bratislavske Lekarske Listy*. 2023 May 1;124(5).

Lojova I., Budis J., Szemes T., Buchalova M., Radvanszky J. Tandem repeat motif characterization for precision medicine: A brief overview of conventional methods and massive parallel sequencing approaches. *NewsLab*, 2023; roč. 14 (1): 40-44

Holešová, Z., Gazdarica, J., Budiš, J., Pös, O., Szemes, T.: Neinvazívny skrining onkologických ochorení pomocou biomarkerov vo voľne cirkulujúcej DNA. *NewsLab*, 2023; roč. 14 (1): 36 – 39

29;13(1):10531.

Uplatnenie výsledkov projektu

Počas realizácie projektu sme navrhli a overili viaceré metodiky orientované na praktické využitie a priamu aplikáciu v klinickej diagnostike. Hlavnou vetvou výskumu bol vývoj nástroja Dante, ktorého aktuálna verzia umožňuje podrobnú analýzu STR motívov v dátach z moderného masívne paralelného sekvenovania celých genómov a exómov. Jeho veľkou výhodou je, že ako bioinformatická metóda nevyžaduje dodatočné laboratórne spracovanie vzorky, a tým podstatne rozširuje informačnú hodnotu sekvenačných údajov bez dodatočných laboratórnych nákladov. Dôsledným literárnym rešeršom sme zhromaždili viac ako 60 klinicky relevantných STR lokusov, ktorých isté formy (delécie/expansiony STR motívu) vedú k vážnym vrodeným ochoreniam. Pre sprístupnenie a uľahčenie interpretácie analýz nástroja Dante sme významne vylepšili jeho užívateľské výstupy pre účely laboratórnej diagnostiky a hodnotení STR genotypov.

Novo publikovaný nástroj WarpSTR má navyše veľký potenciál pre zvýšenie kvality genotypizácie STR lokusov z údajov čoraz častejšie používanej treťogeneračnej platformy Nanopore. Tento nástroj je verejne dostupný a je teda otvorený pre účely laboratórnej diagnostiky a vedeckú komunitu.

Celogenómová analýza STR lokusov má veľký potenciál v oblasti modernizácie diagnostických metód s cieľom charakterizácie onkologických ochorení, pričom posúva individualizovanú liečbu prebiehajúcich rakovinových ochorení za hranice konvenčných prístupov. Detekcia a agregácia instabilit v miliónoch relevantných pozícií STR lokusov v celom genóme má potenciál určiť špecifické subtypy rakoviny spojené s mikrosatelitovou instabilitou (MSI) a umožniť tak lepšie a efektívnejšie nastavenie manažmentu protirakovinovej terapie u pacientov s prebiehajúcimi nádorovými ochoreniami. Predpokladáme, že takýto rozsiahly celogenómový test by mohol personalizovať rozhodovanie o následnej liečbe a dlhodobom sledovaní pacienta, pretože prítomnosť alebo neprítomnosť MSI v nádore bude smerovať pacienta k adekvátnej protirakovinovej terapii (napríklad adjuvančná chemoterapia, ak v nádore nie je MSI, alebo zahájenie liečby inhibítormi imunitných kontrolných bodov, ak je MSI prítomná). Tým sa zvýšia šance na prevenciu recidívy ochorenia a zároveň znížia finančné náklady v dôsledku nastolenia neadekvátnej a pre pacienta neúčinnnej liečby.

Súhrn výsledkov riešenia projektu a naplnenia cieľov projektu v slovenskom jazyku (max. 20 riadkov)

Riešený projekt bol členený na dve hlavné vetvy, diagnostiku ochorení spojených s expanziou repetitívnych motívov a určovania mikrosatelitovej instability (MSI) u onkologických pacientov. Pre obe vetvy bolo vyvinutých a validovaných viacero laboratórnych a výpočtových metód.

V prvej vetve boli vyvinuté softvérové nástroje na spracovanie sekvenačných údajov z

technológií druhogeneračných (Dante 2) a treťogeneračných sekvenátorov (WarpSTR). Na základe podrobnej rešerše literatúry bola vytvorená interná databáza klinicky relevantných mikrosatelitových motívov, ktorá obsahuje informácie o lokalizáciách, sekvenciách, nomenklatúre, charakterizácii a ďalších parametroch. Vyvinuté nástroje boli validované na viac ako 60 vzorkách genetického materiálu s expanziami motívov a kontrolnými vzorkami a porovnávané s konvenčnými laboratórnymi metódami.

Pre detekciu MSI zo sekvenačných dát s nízkym pokrytím sme vytvorili metódu založenú na analýze miliónov lokusov na úrovni celých genómov. Navrhli sme sekvenačný panel a príslušné vyhodnocovacie nástroje pre vybrané homopolymérne markery s následnou sekvenačnou analýzou na platforme Illumina MiSeq. Vyvinuli sme navyše výpočtové postupy pre analýzu celogenómových a celoexómových sekvenačných dát.

Výstupom projektu je viacero nových laboratórnych a naväzujúcich výpočtových postupov, bioinformatických metodík a algoritmov, čo viedlo k množstvu publikácií a prezentácií na odborných konferenciách a genetických kongresoch.

Súhrn výsledkov riešenia projektu a naplnenia cieľov projektu v anglickom jazyku (max. 20 riadkov)

The project was divided into two main branches, the diagnosis of diseases associated with the expansion of repetitive motifs and the determination of microsatellite instability (MSI) in cancer patients. Several laboratory and computational methods were developed and validated for both branches.

In the first branch, software tools were developed to process sequencing data from second-generation (Dante 2) and third-generation sequencer technologies (WarpSTR). Based on a detailed literature search, an in-house database of clinically relevant microsatellite motifs was created, containing information on localizations, sequences, nomenclature, characterization, and other parameters. The developed tools were validated on more than 60 samples of genetic material with motif expansions and control samples and compared with conventional laboratory methods.

For MSI detection from low-coverage sequencing data, we developed a method based on the analysis of millions of loci at the whole-genome level. We designed a sequencing panel and corresponding evaluation tools for selected homopolymeric markers followed by sequencing analysis on the Illumina MiSeq platform. In addition, we developed computational procedures for the analysis of whole-genome and whole-exome sequencing data.

The project has resulted in a number of novel laboratory and downstream computational approaches, bioinformatics methodologies and algorithms, leading to a number of publications and presentations at professional conferences and genetics congresses.