

Posudek oponenta na doktorskou dizertační práci Ing. et. Ing. Davida Staňka s názvem „Objasňování příčin neurogenetických onemocnění analýzou dat z MPS pomocí moderních algoritmů“

Předkládaná dizertační práce je tematicky zaměřena na masivní paralelní sekvenování, jako moderní nástroj pro detekci variant a jejich identifikaci a klasifikaci u neurogenetických onemocnění studovaných na pracovišti Davida Staňka.

Práce je členěna tradičně, tedy na literární úvod, cíle práce, metodickou část, výsledky, diskuzi a závěr. Práce je podpořena i dostatečnou publikační aktivitou článků s prvoautorstvím a to včetně publikací v impaktovaném časopise. Publikační výstupy jsou „pouhou“ přílohou této práce a netvoří jednotlivé kategorie, což velmi oceňuji a práci to dodalo konzistenci a celkovou logiku.

Jedná se tedy o téma vysoce aktuální v dynamicky se rozvíjejícím oboru, což práci přidává na zajímavosti ale i obtížnosti. Zároveň si práce krom vlastního bioinformatického zpracování dat sekvenování klade za cíl vytvořit i strukturu kolem tohoto vyhodnocení jak na straně databází, tak i na straně informačního systému pro správu dat. V souvislosti s tím by mě zajímalo, zda by David dokázal procentuálně vyčíslit svou participaci na řešení jednotlivých cílů práce.

Úvod práce je přehledný a s dobrou definicí, jak klasifikačních jednotek, tak i základních bioinformatických pojmů. Uvítal bych i více přehledněji řazených epidemiologických dat., která jsou uvedena pro CMT, ale pro EE chybí.

Cíle práce jsou jasně strukturovány a definovány poměrně ambiciózně, což dle mého názoru dokonce překračuje očekávání u dizertační práce.

Metodické části nelze nic výrazného vytknout. Metody byly, zajisté po diskuzi se školitelem, zvoleny vhodně a směřovány k naplnění cílů práce. Rád bych vyzdvihl metodickou šíři práce, kdy David testoval i mnoho bioinformatických nástrojů, před vlastním nastavení jednotlivých analýz např. v případě detekce CNV.

Výsledky dobře kopírují linii studia a pokroků v odkrývání jednotlivých problémů, tak jak David postupoval ve své práci stále hlouběji. Objasněnost, cca 28%, je podle mě výtečná známka laboratoře, ostatně z tohoto pohledu nadprůměrně vychází i v porovnání s literaturou, což je v práci uvedeno. Zajímavé též bylo nalezení statistické souvislosti mezi věkem pacienta při prvním záchvatu a nalezením patogenní varianty. Oceňuji též vytvoření databáze variant propojených se studovaných fenotypem dědičných neuropatií a též oceňuji snahu i realizaci podívat se trochu komplexněji na projev detekované varianty na úrovni proteinu a tedy vytvoření a použití Databáze proteinových domén Prot2HG.

V diskuzi jsou dle mého názoru všechny výsledky dány do kontextu s cíli práce a literárními údaji a opět přehledně kopírují strukturu předchozích kapitol.

Formální stránka práce je na vysoké úrovni, práce je sepsána s minimem překlepů, přehledně a též typologicky přesně a jednoduše. Snad jen drobnost bych, jako čtenář přivítal. V textu je sice každá zkratka správně použita až po plném vysvětlení termínu, ale v seznamu zkratek některé s nich chybí (např. chybí označení variant v databázi WES). Též uvedení seznamu na začátku textu by se mi zdálo

přehlednější. Z formálního pohledu se mi zdálo naopak výtečné až odvážné některé užití českých výrazů pro obecně užívané anglicismy např. mě nadchl výraz magnetické korálky (magnetic beads) nebo reakční komora (flowcell).

Co se týče věcné části textu, tak bych měl několik následujících otázek.

1. V úvodu na str. 7 je uvedena prevalence CMT 1:2500 podle dvou zahraničních pramenů. Jaká je prevalence v ČR evnt. jaký je druhá nejčastější klasifikace v rámci dědičných neuropatií a jaká je její prevalence?
2. V části Metody/3.1 Pacienti jsou zařazeny tři diagnózy CMT, Epilepsie a Hluchota. Jaká je jejich provázanost s v úvodu definovanými klasifikací EE?
3. V metodické části 3.2.7. jsou vyjmenovány různé nástroje a přístupy pro detekci CNV. Jak byla data normalizována evnt. jaká byla použita reference při jejich využití? Nemohlo být chybné určení reference příčinou selhání detekce CNV?
4. Kupodivu jsem nenašel nebo přehlédl alespoň základní parametry nastavení sekvenačních běhů. Jaké bylo průměrné pokrytí u jednotlivých sekvenačních strategií – panely, WES a WGS?
5. Hezký výsledek je závislost pravděpodobnosti objasnění EE na věku pacienta při prvním záchvatu (viz. Obrázek 4.3). Nebyly tyto „věkové skupiny“ nějak dále charakterizovány evnt. neuvažujete o další funkční charakterizaci patogenních variant v těchto skupinách?
6. Databáze WES variant je hezký souhrný výsledek, který je snadno dosažitelný, i když je potřeba znát přesnou URL, jinak je nedohledatelný, chápu to tedy jako primárně interní nástroj. Otestoval jsem výskyt variant u genu MUC3A v databázi (378) a liší se od počtu uvedeného v přehledu práce (311, str. 84) – prosím o vysvětlení tohoto rozdílu.

Závěrem tohoto posudku bych ještě jednou ocenil rozsah práce, systematickosti řešení problematiky a významný příspěvek k posunu detekce p594in neurogenetických onemocnění v ČR. Celkově proto doporučuji dizertační práci Davida Staňka k úspěšné obhajobě.

V Praze 20.1 2020

RNDr. Zbyněk Halbhuber, Ph.D.

