

I. OSOBNÍ A STUDIJNÍ ÚDAJE

Příjmení: **Koudelka** Jméno: **Matěj** Osobní číslo: **474368**
 Fakulta: **Fakulta biomedicínského inženýrství**
 Studijní program: **Biomedicínská a klinická informatika**
 Název práce: **Webová aplikace pro detekci somatických bodových variant u pacientů s leukémií**

II. HODNOCENÍ DIPLOMOVÉ PRÁCE

Kritéria hodnocení práce		Počet bodů
1.	<p>Splnění cíle a vhodnost struktury obsahu diplomové práce z hlediska zadaného tématu (splnění zadání). (0 - 30)*</p> <p>Každá část či věta ze zadání musí mít jasný odraz ve zpracované práci. Excelentně splněné zadání může být ohodnoceno maximálním počtem bodů. V poměru rozsahu části v zadání, která není zcela vhodně či úplně zpracována, se hodnocení odpovídajícím způsobem snižuje.</p>	19
2.	<p>Teoretická úroveň a využití dostupné literatury v diplomové práci. (0 - 30)*</p> <p>Oponent posuzuje relevantnost teoretické části k zadání, rozsah rešerší a systematické uspořádání zjištěných poznatků. Pokud převažuje doslovné převzetí textů, snižuje oponent hodnocení až o 15 bodů (přirozeně za předpokladu dodržení autorských práv). Důvodem pro snížení celkového hodnocení je dále nedostatečný výběr teoretických poznatků, literatury a zdrojů.</p>	17
3.	<p>Rozsah realizačních prací (SW, HW), aplikovaných vědomostí a znalostí, úroveň metodologického zpracování a závěrů práce. (0 - 30)*</p> <p>Maximální počet bodů lze udělit práci, která je vhodná k publikování. Tento aspekt se posuzuje zejména z hlediska významu pro obohacení teoretických poznatků a má praktický význam. Obzvláště pozitivně je hodnoceno vytvoření modelu, SW produktu a též technická realizace. Za drobné metodologické nedostatky se hodnocení snižuje až o 5 bodů. Nekonzistentnost zpracování s teoretickými východiskami a nejasný či ne zcela odborný metodologický přístup vede ke snížení minimálně o 15 bodů. Další snížení hodnocení lze udělit za nedostatečnou diskusi k závěrům. Celkem 30 bodů za velmi komplexní a bezchybnou práci včetně dalších aktivit jako je účast na vědecko-výzkumném projektu či grantu, aktivní účast na tvorbě publikací, patentů či užitečných vzorů.</p>	20
4.	<p>Formální náležitosti a úprava diplomové práce (úroveň psaní, označení struktury textu, grafy, tabulky, citace v textu, seznam použité literatury apod.). (0 - 10)*</p> <p>Oponent hodnotí formální náležitosti z pohledu dodržení pravidel o psaní, atributů závěrečných prací, tj. formátování textu, struktury práce, seznamu použité literatury, vybavenosti diplomové práce grafy a tabulkami, způsobu citování. Za nedodržení jednotlivých pravidel snižuje maximální hodnocení o 2 body za každý nerespektovaný atribut. Rovněž za výskyt gramatických chyb, překlepů a nevhodné stylistiky a terminologie se snižuje hodnocení o 2-4 body. V práci by se měla objevovat pouze standardní odborná terminologie a to zejména v českém jazyce (je třeba hodnotit schopnost vyjadřovat se technickým jazykem - 2 body), grafy jsou tvořeny podle zásad (viz tolerance a vliv statistického zpracování - 2 body), u grafů a tabulek jsou patřičné legendy a vše je čitelné (2 body), jsou dodržena citační pravidla podle ISO690 a ISO690-2 (2 body).</p>	3
5.	Celkový počet bodů	59

* Slovní hodnocení uveďte v komentáři.

III. NÁVRH OTÁZEK K OBHAJOBĚ

1. Jakým způsobem jsou vzorky a data zpracovávána předtím, než se dostanou do formátu NextDom? Popište celý proces od přípravy vzorku po způsob sekvenování přes následnou bioinformatickou analýzu.

2. Dílčím úkolem bylo seznámení se se základními formáty dat využívaných v sekvenování nové generace. Tuto kapitolu jsem nikde v práci nenašla. Můžete prosím popsat základní formáty NGS dat?

3. V práci uvádíte, že při "hlubokém amplikonu sekvencí provedené pomocí NGS vznikají záměny a mutace na určitých pozicích v nukleotidech DNA sekvence". Uveďte prosím, jaké jsou mechanismy vzniku těchto mutací/záměn a ve kterých všech fázích zpracování vzorku, sekvenování a analýzy získaných dat vznikají. Vznikají mutace na některých pozicích častěji než na jiných?

IV. CELKOVÉ HODNOCENÍ ÚROVNĚ VYPRACOVÁNÍ DIPLOMOVÉ PRÁCE

Hodnocení**:	A (výborně)	B (velmi dobře)	C (dobře)	D (uspokojivě)	E (dostatečně)	F (nedostatečně)
Počet bodů:	100 - 90	89 - 80	79 - 70	69 - 60	59 - 50	< 50
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>

** v případě hodnocení F (nedostatečně) uveďte podrobný komentář

Diplomovou práci hodnotím výše uvedeným klasifikačním stupněm a doporučuji/nedoporučuji k obhajobě.

V. KOMENTÁŘ

Ad 1) Cíle práce a dílčí úkoly jsou odlišně definované v zadání diplomové práce a v samotné diplomové práci (viz. kapitola 3). Cíle práce by měly být definovány jednotně a neměnně. Pokud vycházím z dílčích úkolů definovaných v zadání práce, nebyl splněn úkol 1 - "Seznámení se se základními formáty dat využívaných v sekvenování nové generace (NGS)". Práce popisuje pouze formát dat vygenerovaných aplikací NextGENe (Part1 a Part2) - nejedná se tedy o základní formáty NGS dat. Úkol 2 - "Studium problematiky genetické podstaty leukémii prostřednictvím metod NGS" byl splněn částečně. Očekávala bych, že se dozvím o této problematice více. Za velmi dobře splněné cíle pokládám úkol 3 - "Implementace statistické metody pro stanovení individuálních prahů pro detekci bodových variant" a úkol 4 - "Návrh a implementace webové aplikace pro stanovení individuálních prahů včetně jednoduchého grafického rozhraní". Úkol 5 - "Testování na simulovaných a reálných datech" byl dle informací v diplomové práci splněn pouze částečně - testování proběhlo pouze na reálných datech. Primární cíl, tedy "návrh a vývoj webové aplikace pro stanovení individuálních prahů při detekci bodových variant" byl splněn dobře.

Ad 2) Teoretická část práce představuje základy genetiky, základy bioinformatiky, sekvenování nové generace, statistické zpracování genomických dat, stručně přibližuje problematiku leukémií a současný stav aplikace NextDOM2. Témata teoretické části jsou vhodně zvolené, nicméně jejich obsah není vždy vhodně zvolen. Kapitola "2.1.1 Základní pojmy genetiky" je uvedena jako slovníček jednotlivých pojmů. Bylo by výrazně lepší jednotlivé koncepty vysvětlit v celých větách a vysvětlit jak spolu jednotlivé pojmy souvisejí. Kapitola "2.2.1 Sekvenování" vysvětluje detailně Sangerovu metodu v její původní podobě, nicméně bych pokládala za vhodnější vysvětlit detailně současný princip založený na PCR a kapilární elektroforéze. Dále mi v práci chybí kapitola o bioinformatickém zpracování NGS dat a příslušných datových formátech. Kapitola "4 Metody", detailně a vhodně srovnává programovací jazyky MATLAB a Python. Kapitola "5 Návrh aplikace" rozebírá funkční a technické požadavky na aplikaci a předkládá rešerši technologických řešení obdobných projektů, které slouží pro následný výběr vhodných technologií pro vlastní aplikaci.

Ad 3) Výsledkem práce je webová aplikace pro stanovení individuálních prahů při detekci bodových variant u pacientů s leukémií. Předkládané řešení využívá následující technologie: Python, Python Flask, HTML, CSS, Bootstrap, JavaScript, Git a Docker. Aplikace je veřejně dostupná přes GitHub a DockerHub. Na základě instrukcí dostupných v diplomové práci jsem byla schopná aplikaci úspěšně spustit a otestovat. V kapitole "7.1.1 Prerekvizity nutné k správné funkci" odkazy pro instalaci WSL2 a Docker Desktopu vedou na neexistující stránky. Při instalaci jsem dále

musela ručně doinstalovat modul 'psutil'. Při testování aplikace jsem narazila na chybu "FileNotFoundException" při ukládání souboru vypočtených prahových hodnot (Download thresholds) a při nastavování vypočtených prahových hodnot jako výchozích (Set these thresholds as default). V obou případech se jednalo o tréninková data.

Ad 4) Struktura práce odpovídá požadavkům pro diplomovou práci, nicméně obsahuje velké množství gramatických chyb, překlepů a nevhodné stylistiky. Terminologie je nejednoznačně používána a místy nepřesná (např. cytosin ale i cytozin, vložení ale i inserce, "SNP array" přeloženo jako "SNP pole" místo "SNP čip" či "DNA čip", "read mapping" přeloženo jako "čtecí mapy" místo "mapování/zarovnávání čtení", "base quality" přeloženo jako "základní mapování" místo "kvalita báze". Některým částem textu je tak velmi obtížně rozumět. V seznamu zkratk některé zkratky chybí (Fj, Rj, NPN, CC, TIGR) a naopak jsou uvedeny jiné, které se v textu vůbec nevyskytují (OOP). Práce obsahuje chyby a nepřesnosti v citacích. Např. citaci [2] se mi nepodařilo vůbec dohledat, odkaz vede na jiný článek, u citace [6] chybí název článku, citace [10] - autoři jsou F. Sanger a A.R. Coulson, nikoliv SANDER, F a A R GOULSON. U citace [12] chybí povinný údaj a to datum citace a požadovanou stránku se mi nepodařilo dohledat - bylo by vhodné uvést plnou URL adresu webové stránky. Popisky obrázků a tabulek by bylo vhodné rozšířit a umístit u obou formátů stejně - tedy buď nahoře, nebo dole.

I přes výše uvedené se jedná o velmi zajímavé téma a výsledné řešení výrazně usnadní rozšíření aplikace NextDOM mezi další uživatele.

Diplomovou práci Bc. Koudelky doporučuji k obhajobě.

Jméno a příjmení: Mgr. et Mgr. Anna Přistoupilová, Ph.D.
Organizace: DNAnexus Czech Republic s.r.o.
Kontaktní adresa:

Podpis:

Datum: