

## I. OSOBNÍ A STUDIJNÍ ÚDAJE

Příjmení: **Součková** Jméno: **Michaela** Osobní číslo: **483028**  
 Fakulta: **Fakulta biomedicínského inženýrství**  
 Studijní program: **Biomedicínská a klinická informatika**  
 Název práce: **Vytvoření nástroje pro zpracování dat z RNAseq analýz pacientů s leukemickým onemocněním**

## 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>	29
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>	26
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>	25
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>	8
5.	<b>Celkový počet bodů</b>	88

\* Slovní hodnocení uveďte v komentáři.

### III. NÁVRH OTÁZEK K OBHAJOBĚ

1. Při zkoušení pipeline v lokálním režimu na počítači se systémem Linux jsem narazil na několik problémů, které mi bránily v získání výsledků. Nejprve, pipeline se zavěsí hned po spuštění, ještě před startem prvního nástroje fastp. Překvapivě, odeslání signálu SIGKILL do procesu Bash příkazu pipeline ve skutečnosti zahájilo její běžení. Mám podezření, že tento problém nějak souvisí s konfigurací Nextflow. Poté, po úspěšném spuštění modulu fastp, se pipeline ukončila se stavem 127 a zprávou ".command.sh: line 3: STAR: command not found". Docker kontejner "souckmi2/star:1.0" však nemá problém spustit STAR, když je spuštěn nezávisle. Předpokládám, že tyto problémy se při spuštění v prostředí MetaCentra nevyskytují. Vzhledem k tomu, že uvádíte, že pipeline může být spuštěna v lokálním režimu, napadlo by vás řešení těchto problémů na lokálním počítači?

2. Když spouštím pipeline lokálně přes Bash, jak upravím počet jader CPU a přidělenou paměť? Je úprava konfiguračních souborů jediný způsob, jak toho dosáhnout?

3. Proč se počet hlášených fúzí mezi Arriba/STAR-Fusion a dalšími připravovanými nástroji tak liší? Jinými slovy, pokud Arriba a STAR-Fusion detekují 10 fúzí a jiné nástroje detekují stovky, měli bychom brát v úvahu pouze těchto 10?

### 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"/>	X	<input type="checkbox"/>	<input type="checkbox"/>	<input 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

Cílem práce bylo vyvinout pipeline pro zpracování dat sekvenování RNA od vstupního FASTQ až po vizualizaci a tabulku fúzních genů připravenou pro odbornou klinickou analýzu. Autorka tento cíl úspěšně splňuje vytvořením pipeline pro detekci fúzních genů a jejím otestováním na klinických datech. Pipeline je k dispozici pro instalaci na jakémkoliv linuxovém počítači s nainstalovaným frameworkem Nextflow. Autorka uvádí výsledky spuštění pipeline na vzorcích od pacientů s leukémií, kde úspěšně detekuje klinicky relevantní genové fúze. Získání know-how a úspěšná implementace robustního klastrového řešení problému detekce fúzních genů je cennou snahou, která vyžaduje racionální přístup a dobré schopnosti řešení problémů. Cíle práce jsou jasně definované a jsou autorkou splněné.

Autorka v práci rozebírá všechny důležité kroky experimentu sekvenování RNA od přípravy knihovny a sekvenování až po zpracování dat a detekci genových fúzí. Pozoruhodné je, že autorka také implementuje nejlepší postupy vývoje bioinformatického softwaru s využitím existujících kontejnerizačních řešení.

Na stranách 22-23 však autorka popisuje "nejjednodušší" přístup k normalizaci počtu readů a kritizuje jej za přílišnou prostotu, ale žádnou alternativu neuvádí. Až mnohem později, na straně 28, je představen koncept fúzního fragmentu na milion (FFPM), který však není propojen s odstavcem o normalizaci.

Po technické stránce autorka zvolila racionální a škálovatelný přístup k návrhu pipeline pro detekci fúzních genů. Pokud je mi známo, softwarové nástroje použité v této studii jsou v oblasti analýzy hromadného transkriptomu špičkové. Avšak, dle mého názoru, práce postrádá rozsáhlejší diskuse variací mezi výsledky různých nástrojů integrovaných do pipeline.

Jazyk autorky je jasný, stručný a informativní. V úvodní části jsou velmi dobře vysvětleny teoretické pojmy. Praktická část je dobře strukturovaná a formátovaná. Celkově je formální a jazyková úroveň práce velmi dobrá.

Text má však několik drobných problémů. Zaprvé, obsahuje několik nekritických pravopisných chyb. Zadruhé, některé zkratky (např. COSMIC na s. 28, BAM a SAM na s. 37 nebo jména organizátorů soutěže DREAM na s. 39) nejsou podle mého chápání nikde v textu rozšířeny. Avšak, dá se říct, že tyto zkratky jsou buď v této oblasti dobře známé, jak to je v případě prvních dvou, nebo, upřímně řečeno, nejsou pro cíle práce tak podstatné.

Horším přestupkem je způsob, jakým je prezentován obrázek 3.2. Nejenže zkratky TP a FP nejsou nikde uvedené, ale celá část b tohoto obrázku není dostatečně popsána, což zanechává dojem, že obrázek 3.2b by mohl být z práce odstraněn aniž by postrádala jakýkoli smyslový obsah, nebo lépe, mohl by být opatřen odstavcem textu, který by poskytl kontext k jednotlivým grafům.

Celkově hodnotím tuto práci jako velmi dobrou a doporučuji ji k obhajobě.

Jméno a příjmení: Mgr. Dmitry Kazantsev, Ph.D.  
Organizace: Ústav patologické fyziologie 1. LF UK  
Kontaktní adresa:

Podpis: .....

Datum: .....