

I. OSOBNÍ A STUDIJNÍ ÚDAJE

Příjmení: **Snášelová** Jméno: **Irma** Osobní číslo: **492260**
 Fakulta: **Fakulta biomedicínského inženýrství**
 Studijní program: **Biomedicínská a klinická technika**
 Studijní obor: **Biomedicínská informatika**
 Název práce: **Bioinformatická analýza RNA-seq dat**

II. HODNOCENÍ BAKALÁŘSKÉ PRÁCE

Kritéria hodnocení práce		Počet bodů
1.	<p>Splnění cíle a vhodnost struktury obsahu bakalářské 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>	25
2.	<p>Teoretická úroveň a využití dostupné literatury v bakalářské 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>	20
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 bakalářské 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 bakalářské 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>	5
5.	Celkový počet bodů	70

* Slovní hodnocení uveďte v komentáři.

III. NÁVRH OTÁZEK K OBHAJOBĚ

1. Můžete prosím u obhajoby vysvětlit odstavec uvedený níže a dodat nějaký zdroj pokud existuje? Co v tomto odstavci rozumíte pod pojmy nízká kvalita genomu a proč je frekvence výskytu variant nacházejících se ve funkčně významných částech genomu v databázích populací velmi nízká? Správně nastaveným filtrem se eliminují varianty s nízkou kvalitou genotypu, ready pokrývající oblast dané varianty, případně i předpokládané modely dědičnosti na základě rodinné anamnézy. Zajímavé jsou také varianty nacházející se ve funkčně významných částech genomu, protože jejich frekvence výskytu je v databázích populací velmi nízká.

2. V práci je uvedena následující věta "K provedení analýzy byl vybrán dataset z volně dostupné online databáze obsahující čtyři vzorky, kdy každý z nich obsahoval několik již dříve klinicky ověřených fúzních genů, které byly přítomny u pacientek s karcinomem prsu." Ve které databázi se tyto vzorky nachází? Kde můžeme tyto vzorky přímo najít či stáhnout?

3.

IV. CELKOVÉ HODNOCENÍ ÚROVNĚ VYPRACOVÁNÍ BAKALÁŘSKÉ 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"/>	X	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

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

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

V. KOMENTÁŘ

Bioinformatická analýza RNA-seq dat včetně úlohy detekce fúzních genů je důležitým tématem, jak ve výzkumu, tak i pro klinickou praxi. Rovněž je také pravdou, že aktuální či nedávno publikované rozličné metody detekce fúzních genů vedou k odlišným výsledkům a i z tohoto důvodu považuji téma práce za aktuální.

Práce staví spíše na obecných literárních zdrojích. Považuji to za adekvátní přístup řešení bakalářské práce. Na základě rešerše a komunikace s klinickým pracovištěm byly pro praktickou část a analýzu zvoleny dva nástroje. Myslím, že vzhledem k povaze úlohy, by bylo u výběrů nástrojů v rámci rešerše patřičné uvést více zdrojů, které by porovnávaly jednotlivé nástroje podle různých kritérií. V práci je uvedena citace na článek [23], který se takovou analýzou zabývá. Byl jsem rychle schopný nalézt řadu dalších prací z posledních let, které se zabývají tímto tématem a důkladně porovnávají referované nástroje, což může přinést lepší náhled na výběr nástrojů. Pro referenci zde přikládám pár příkladů identifikátorů takových článků:

10.1101/2020.09.17.302307
10.1186/s13059-020-02043-x
10.1007/s00401-020-02167-1
10.1186/s13059-019-1842-9

Technická realizace, implementace a provedení RNA-seq analýzy dat v této práci je stručné a nejde do hloubky a jelikož využívá veřejně dostupná data (která jsem nedokázal lokalizovat a zjistit zda se jimi již některá práce zabývala, tudíž výsledky pro daný dataset mohou být komunitou známy), nevede podle mého názoru k unikátní aplikaci, ale slouží spíše jako realizace praktické úlohy, osvojení si metod autorkou práce a aplikace na data. Může to rovněž sloužit jako příprava pro nasazení na klinická data, což oceňuji. Pro možnosti spuštění a reprodukce výsledků by bylo dobré dodat i dockerfile s popisem prostředí a dalo by se pro účely demonstrace docker run příkaz připravit lépe. Např. při definování cest

```
/home/kbi/data2/irma/data/BT474.Right.fq.gz:/input/r2.fastq.gz
```

byla zdokumentována cesta pravděpodobně do adresáře autorky a navíc jsem v práci nenašel přímý způsob, jak se dostat k použitým vzorkům (tj. k souborům ve formátu FASTQ). Na druhou stranu, podobný typ úlohy by se mohl stát vhodnou inovační úlohou pro zařazení do předmětu Bioinformatika bakalářského programu.

Co se týká formální stránky, chápu, že určité termíny v daném oboru nemají v češtině odpovídající termín, případně se tak přímo jmenuje daná metoda (viz termín RNA-seq v tomto posudku či v názvu práce). Avšak vzhledem k tomu, že práce byla vypracována v češtině, některé termíny by podle mého názoru měly být přeloženy a jejich adekvátní termín existuje:

best practices, RT-PCR (real time polymer chain reaction), zarovnat ready kolem insercí a delecí (indely) zkoumaného vzorku

Podobně, obrázky by měly být modifikovány do jazyka práce a všechny zkratky by měly být (pokud vůbec) vysvětleny hned při prvním výskytu, např. HPO, SGS/TGS.

Našel jsem dále několik překlepů, které však nevedou k zásadnímu zhoršení čitelnosti, např.:

na fúzní gen před vybrat, interpretacevysledků, dětští pacienta jsou, jedno vláknová

Jméno a příjmení: Ing. Ondřej Klempíř, Ph.D.
Organizace: DNAnexus Czech Republic s.r.o.
Kontaktní adresa:

Podpis:

Datum: