

Oponentský posudek na diplomovou práci

Jméno oponenta: Mgr. **Marian Novotný**, PhD.

Datum: 21.5.2010

Autor: Pavel Pitule

Název práce: **Komparativní genomická předpověď nových transmembránových adaptorových proteinů a jejich expresní analýza**

Cílem předkládané diplomové práce byla identifikace a následná charakterizace nových transmembránových adaptorových proteinů důležitých pro imunitní odpověď. Hlavní půvab této práce spočívá v dosud bohužel stále poměrně vzácné kombinaci přístupů bioinformatických a experimentálních.

V první části se Pavel Pitule pokusil za pomoci programů (které sám napsal) a relační databáze MySQL zautomatizovat postup na vyhledávání transmembránových adaptorových proteinů v lidském proteomu. Kandidátní proteiny, objevené za pomoci nástrojů popsaných v první části, pak dále charakterizoval – zajímal se především o jejich lokalizaci a tkáňovou specifitu exprese. Takto získané experimentální výsledky pak opět použil pro další optimalizaci bioinformatických nástrojů a zpřesnění svých modelů. A to několikrát a tato iterace může pokračovat i nadále. Jde tedy o krásný příklad symbiozy vytváření modelu (bioinformatická část) a jeho ověřování (experimentální část). Tato kombinace však samozřejmě klade velké nároky na studenta, který se pokouší o obé. Pavel Pitule se této úlohy zhostil, dle mého názoru, rozhodně se ctí.

Práce samotná je přehledně členěna na předepsané části. Po velmi podrobném a úplném seznamu zkratk následuje teoretický úvod. V teoretickém úvodu autor představuje koncept adaptorových proteinů a lineárních motivů. Dále nás seznamuje s jejich členěním a jejich rolí v regulaci buňkových procesů.

Teoretický úvod je také velmi dobrým a současným přehledem o dostupných databázích a nástrojích na analýzu krátkých lineárních motivů a nástrojích na predikci transmembránových helixů.

Poté následuje velmi pečlivě zpracovaná kapitola metody. Podrobněji jsem se však věnoval jen bioinformatickým metodám, neboť experimentální část metod si netroufám sám posuzovat. Programy, které napsal netrhají rekordy v délce, naopak překvapují tím, jak mnoho užitečného lze udělat s několika řádky skriptu.

V kapitole Výsledky se Pavel pokusil podrobně popsat postup své práce. Ukazuje v ní velmi pěkně i některé slepé uličky a dobře demonstruje už zmiňovaný iterativní charakter jeho práce, který velmi oceňuji. Osobně jsem

však měl problém ve výsledcích oddělit ve své podstatě metodické výsledky (nové programy, postupy) od výsledků biologických (kandidátní adaptorové proteiny). Líbí se mi naopak pokus o integraci co největšího množství dat.

Diskuze a závěr adekvátně shrnují předložené výsledky.

Jazyková úroveň v celém textu je velmi vysoká, autor se velmi pěkně vyhýbá zbytečnému užívání anglicismů a také jeho obrazová dokumentace je na velmi vysoké úrovni.

Ve své práci autor obsáhl skutečně široké spektrum metod – od psaní programů jazykem Perl, přes komunikaci s databází až po qPCR na populacích imunitních buněk z myši sleziny. Jeho diplomová práce je dle mého názoru vyjímečná především nelehkou kombinací bioinformatiky a experimentů. **Práce podle mě jednoznačně splňuje požadavky na diplomovou práci a navrhuji její přijetí.**

K práci mám následující připomínky:

V práci se vyskytují občas překlepy, ale vzhledem k délce textu jsou v míře více než únosné. Autor není vždy konzistentní v některých formálních aspektech práce. Např. mezery před pomlčkami se objevují respektive neobjevují ve všech možných kombinacích.

Práce s literárními zdroji je v této práci velmi dobrá, ale ani tady se autor nevyhnul drobným lapsům a navíc měl velkou smůlu, že mě zaujal zrovna článek, který má v seznamu literatury neúplnou citaci (citace 33). Citace 31 a 32 odkazují, dle mého skromného úsudku, na úplně stejnou publikaci, akorát citace 32 opět postrádá rok publikace.

Postrádal jsem citaci na str. 57, kde autor zmiňuje, že na základě literárních údajů vybral další motivy.

Str. 22 – třetí podmínka nepopisuje vlastnost motivu (jak deklarujete), ale jde o podmínku hledání motivu.

Dále mám velmi mnoho otázek:

- 1) V úvodu píšete, že C1 doména váže estery forbolu. Můžete nám říci k čemu estery forbolu slouží a zda jsou vůbec přirozenou součástí buněk?
- 2) Jak vypadá ITAM motiv?
- 3) Jsou si navzájem TRAP 2. třídy sekvenčně podobné?
- 4) Jak často vznikají znaky na molekulární úrovni konvergentní evolucí, ak píšete? Můžete uvést příklad?
- 5) Související otázka – proč jsou lineární motivy náchylné k zániku bodovou mutací, když píšete, že nejsou příliš konzervovane?
- 6) Chtěl bych se zeptat, co přesně je těch více než 70000 „proteinů“ z Ensembl, které jste použili pro vyhledávání. Poslední verze Ensembl mi totiž ukazuje jen něco přes 22000 proteiny-kódujících genů a obdobně proteinová databáze Uniprot, která deklaruje, že popsala kompletní lidský

proteom, obsahuje dokonce o 2000 proteinů méně. Jde i o sestřihové varianty proteinů?

- 7) Nemáte představu, jak často chybí myši ortolog lidského genu nebo naopak jako v případě Vašeho kandidátního genu C1orf186?
- 8) Máte vysvětlení pro nízké skóre některých známých TRAP ve Vašem hodnocení? Zkoušeli jste optimalizovat parametry tak, aby vyskočily výš? Jak podobné jsou si lineární motivy známých TRAP II. třídy?
- 9) Zaujala mě i Vaše snaha udělat systém robustnější přidáním více ortologů. Jak rozvinutý imunitní systém lze očekávat u *Ceanorhabditis* či *Drosophily*? Měli by tyto organismy mít TRAP II. třídy.
- 10) Jak rámcově funguje md5sum, které používáte na srovnávání proteinových sekvencí?
- 11) Neuvažujete o ověřování predikcí u kandidátních genů např. několika algoritmy pro predikci transmembránových helixů, abyste šetřili experimentální síly?

Mám i další otázky, ale těmi už nebudu zdržovat na obhajobě.

Podpis oponenta: 