

I. IDENTIFIKAČNÍ ÚDAJE

Název práce:	Secondary structure search in primary nucleic acid structures
Jméno autora:	Anh Vu Le
Typ práce:	diplomová
Fakulta/ústav:	Fakulta elektrotechnická (FEL)
Katedra/ústav:	Katedra počítačů
Oponent práce:	RNDr. Martin Pospíšek, Ph.D.
Pracoviště oponenta práce:	Přírodovědecká fakulta Univerzity Karlovy, Katedra genetiky a mikrobiologie

II. HODNOCENÍ JEDNOTLIVÝCH KRITÉRIÍ

Zadání	mimořádně náročné
---------------	--------------------------

Hodnocení náročnosti zadání závěrečné práce.

Hlavním cílem práce bylo otestovat hypotézu, zda lidský genom může kódovat ribonukleové kyseliny (RNA), které se skládají do sekundárních struktur podobných internímu vazebnému místu pro ribozom viru žloutenky typu C (HCV IRES) nebo jeho částí. Pro vlastní provedení bylo nutné provést důkladnou rešerši a testování dostupných softwarových nástrojů, navrhnout novou proceduru ("pipeline"), otestovat a navrhnout jednotlivé části této procedury a následně se pokusit o vlastní hledání strukturního homologu HCV IRES uvnitř lidského transkriptomu. Kromě technických dovedností byl student pro úspěšné zvládnutí projektu nucen se do značné míry seznámit též se zákonitostmi stavby nukleových kyselin, zejména RNA, získat detailní přehled o bioinformatických nástrojích v současnosti používaných pro analýzu a predikci sekundárních struktur RNA a seznámit se s postupy a nástroji používanými pro hledání strukturních motivů v souborech/databázích obsahujících primární a sekundární struktury RNA.

Splnění zadání	splněno
-----------------------	----------------

Posuďte, zda předložená závěrečná práce splňuje zadání. V komentáři případně uveďte body zadání, které nebyly zcela splněny, nebo zda je práce oproti zadání rozšířena. Nebylo-li zadání zcela splněno, pokuste se posoudit závažnost, dopady a případně i příčiny jednotlivých nedostatků.

Zadání student splnil a jako bonus získal výsledky, které mohou být rovnou předány do biologické laboratoře k otestování. Jak je zřejmé z popisu řešení projektu, cesta ke splnění zadání nebyla jednoduchá a student musel v průběhu řešení projektu několikrát změnit své plány na to, jak bude "pipeline" pro hledání částí HCV IRES v lidském genomu nakonec sestavena.

Zvolený postup řešení	vynikající
------------------------------	-------------------

Posuďte, zda student zvolil správný postup nebo metody řešení.

Zvolený postup řešení je originální a ve svém celku nekopíruje žádný ze známých a zavedených postupů. Autor se rozhodl zcela zaměřit na sekundární strukturu hledané části HCV IRES a potlačit informace o sekvenci nukleotidů. Podle dosažených výsledků se zdá, že tento postup byl správný a vedl k lepším výsledkům než dosud publikované pokusy cílící též na hledání struktur virových IRES v hostitelských genomech. Precizní přístup studenta k řešení problému je vidět i z toho, že v postupu jedné z těchto prací publikované v roce 2013 v prestižním časopise RNA Biology našel chybu, která možná vysvětluje slabé výsledky tehdy publikované procedury.

Odborná úroveň	A - výborně
-----------------------	--------------------

Posuďte úroveň odbornosti závěrečné práce, využití znalostí získaných studiem a z odborné literatury, využití podkladů a dat získaných z praxe.

Práce je na vysoké odborné úrovni. Zejména oceňuji propojení znalostí algoritmů a programování se znalostmi v oblasti struktury nukleových kyselin a jejich analýzy, což nebývá obvyklé.

Formální a jazyková úroveň, rozsah práce	A - výborně
---	--------------------

Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku.

Diplomová práce je napsána v angličtině a obsahuje jen minimum chyb a překlepů. Grafická úprava práce je výborná. Práci musím pochválit nejen za odborné řešení problému, ale též za způsob, jakým je napsána. Často se stane, že práce s výbornými výsledky jsou napsány nezajímavým a těžkým stylem. Zde to tak není. Text je velmi čtivý a možná i z důvodu, že problematika řešená v rámci projektu mne velmi zajímá, četl jsem práci s opravdovým potěšením a jedním dechem, téměř jako detektivku.

Výběr zdrojů, korektnost citací

A - výborně

Vyjádřete se k aktivitě studenta při získávání a využívání studijních materiálů k řešení závěrečné práce. Charakterizujte výběr pramenů. Posuďte, zda student využil všechny relevantní zdroje. Ověřte, zda jsou všechny převzaté prvky řádně odlišeny od vlastních výsledků a úvah, zda nedošlo k porušení citační etiky a zda jsou bibliografické citace úplné a v souladu s citačními zvyklostmi a normami.

Student použil relevantní zdroje, které jsou i řádně citovány. Pouze bych doporučil číslovat citace postupně podle jejich prvního uvedení v textu. Celkově se práce odkazuje na 86 pramenů.

Další komentáře a hodnocení

Vyjádřete se k úrovni dosažených hlavních výsledků závěrečné práce, např. k úrovni teoretických výsledků, nebo k úrovni a funkčnosti technického nebo programového vytvořeného řešení, publikačním výstupům, experimentální zručnosti apod.

Řešení projektu i jeho presentaci v diplomové práci považuji za velmi zdařilé. Další komentáře uvádím v celkovém hodnocení níže.

III. CELKOVÉ HODNOCENÍ, OTÁZKY K OBHAJOBĚ, NÁVRH KLASIFIKACE

Shrňte aspekty závěrečné práce, které nejvíce ovlivnily Vaše celkové hodnocení. Uveďte případné otázky, které by měl student zodpovědět při obhajobě závěrečné práce před komisí.

Práci považuji za velmi zdařilou. Rozhodně oceňuji studentův pohyb na rozhraní dvou velmi nepříbuzných oborů a výsledky, které v průběhu řešení projektu získal. S velkým zájmem jsem si pro sebe narychlo analyzoval nejperspektivnější RNA strukturu nalezenou v průběhu řešení diplomové práce. Jedná se o strukturu v oblasti 5'UTR mRNA přepisované z genu *DRC3*. 5' UTR genu *DRC3* je dlouhá a obsahuje několik dalších krátkých otevřených čtecích rámců, z nichž některé se vzájemně překrývají. Takováto 5'UTR brání normální translaci v eukaryontní buňce a splňuje všechny učebnicové předpoklady pro přítomnost nějakého translačního regulátoru, například IRES. Potvrzení případné biologické aktivity struktury nalezené v 5' UTR *DRC3*, případně dalších nalezených struktur, musí být provedeno v laboratoři, ale již nyní je zřejmé, že postup a výsledky předložené v diplomové práci mají velký potenciál k publikaci v kvalitním vědeckém časopise. Z těchto důvodů si dovoluji doporučit pozdržení zveřejnění práce. V souladu s názorem studenta si myslím, že budoucím výsledkem by měla být veřejná webová aplikace / server, které by umožňovaly uživatelům vyhledávat v genomech a transkriptomech hlavních modelových organismů oblasti kódující RNA o požadované sekundární struktuře.

K práci mám následující dotazy:

- Při prohledávání databáze lidského genomu nebyly nalezeny stejné struktury jako při prohledávání databáze mRNA RefSeq. Předpokládám, že je to tím, že bylo testováno pouze 6,5 % inverzních sekvencí. Máte nějaký odhad, jak dlouho by trvalo prohledat celý lidský genom s použitím stejného počtu inverzních sekvencí, který jste použil pro lidský transkriptom (mRNA RefSeq). Bylo by možné tuto práci nějakým způsobem urychlit, například použitím větší výpočetní kapacity, paralelizací výpočtu atd?

- Libil se mi Váš způsob otestování programů pro inverzní skládání struktur RNA ("inverse folding"), ve kterém jste ukázal, že program RNAiFOLD poskytuje artefakty. Neplánujete podobným způsobem otestovat více těchto programů? Myslím, že by stálo za to podobný test dostupných softwarových řešení provést a publikovat.

Předloženou závěrečnou práci hodnotím klasifikačním stupněm **A - výborně**.

Datum: 13.6.2019

Podpis: