

# 1-INF-911 Bakalársky seminár (1)

## Domáca úloha 1

Jozef Mrkvička

Školský rok 2015/16

### 1 Základné údaje o bakalárskej práci

**Školiteľ:** doc. RNDr. Mária Hrášková, PhD.

**Pracovisko školiteľa:** Katedra informatiky, FMFI UK

**Názov práce:** Použitie údajov o expresii pri hľadaní génov

**Cieľ práce:** Cieľom hľadania génov je v DNA sekvencii určiť, ktoré jej oblasti kódujú proteíny produkované daným organizmom. Hľadanie génov patrí medzi základné bioinformatické problémy. V tejto práci je úlohou do existujúceho systému na hľadanie génov ExonHunter implementovať modul, ktorý pri hľadaní génov zohľadní experimentálne merania expresie jednotlivých génov, t.j. určujúce množstvo RNA prípadne proteínu, ktoré sa na základe jednotlivých oblastí DNA produkuje.

### 2 Kľúčové zdroje

V mojej bakalárskej práci sa plánujem opierať najmä o nasledujúce informačné zdroje.

- Marketa J. Zvelebil and Jeremy O. Baum. *Understanding bioinformatics*. Garland Science, 2008

V tejto vysokolskej učebnici plánujem nájsť informácie o základných biologických a bioinformatických pojmoch, ktoré potrebujem vysvetliť vo svojej práci.

- Broňa Brejová, Daniel G. Brown, Ming Li, and Tomáš Vinař. ExonHunter: A comprehensive approach to gene finding. *Bioinformatics*, 21(Suppl 1):i57–i65, 2005

V tomto vedeckom článku autori popisujú systém ExonHunter, do ktorého idem pridávať nový modul.

- Peter Kováč. Implementácia externých zdrojov dát v hľadaní génov. Bakalárska práca, Univerzita Komenského v Bratislave, 2010

Autor tejto bakalárskej práce tiež implementoval moduly do systému ExonHunter, na jeho prácu teda priamo nadväzujem.

- Larry Wall, Tom Christiansen, and Jon Orwant. *Programming Perl*. O'Reilly, 2004

Modul budem implementovať v jazyku Perl, plánujem teda využívať túto referenčnú príručku.

- Daehwan Kim, Geo Pertea, Cole Trapnell, Harold Pimentel, Ryan Kelley, and Steven L Salzberg. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology*, 14(4):R36, 2013

Vedecký článok popisujúci softvér TopHat2. Tento softvér plánujem použiť na spracovanie vstupných dát pre môj modul.

- ENCODE consortium. RNA-seq of thymus (Mus musculus, postnatal 0 day), 2014. [Citované 2014-10-14] Dostupné z <https://www.encodeproject.org/files/ENCFF002EYX/@@download/ENCFF002EYX.fastq.gz>

Webstránka, z ktorej získam experimentálne dáta potrebné na testovanie môjho modulu.

## Literatúra

- [1] Broňa Brejová, Daniel G. Brown, Ming Li, and Tomáš Vinař. ExonHunter: A comprehensive approach to gene finding. *Bioinformatics*, 21(Suppl 1):i57–i65, 2005.
- [2] ENCODE consortium. RNA-seq of thymus (Mus musculus, postnatal 0 day), 2014. [Citované 2014-10-14] Dostupné z <https://www.encodeproject.org/files/ENCFF002EYX/@@download/ENCFF002EYX.fastq.gz>.
- [3] Daehwan Kim, Geo Pertea, Cole Trapnell, Harold Pimentel, Ryan Kelley, and Steven L Salzberg. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology*, 14(4):R36, 2013.
- [4] Peter Kováč. Implementácia externých zdrojov dát v hľadaní génov. Bakalárska práca, Univerzita Komenského v Bratislave, 2010.
- [5] Larry Wall, Tom Christiansen, and Jon Orwant. *Programming Perl*. O'Reilly, 2004.
- [6] Marketa J. Zvelebil and Jeremy O. Baum. *Understanding bioinformatics*. Garland Science, 2008.