

Genome Browser

Jakub Zárybnický (xzaryb00@stud.fit.vutbr.cz)

March 19, 2020

1 Úkol 1

Zjistěte, zda obsahuje gen BRCA1 u myši SNP elementy. Pokud ano, znázorněte v Genome Browseru synonymní SNP modrou barvou a nesynonymní červenou barvou.

Postup:

1. Otevřete si prohlížeč se stránkou UCSC Genome Browser (<http://genome.ucsc.edu/>) a přejděte do části pro prohlížení genomů (záložka Genomes).
2. Zvolte group: **Mammal**, genome: **Mouse**, assembly: **July 2007**.
3. Do textového pole vložte název **Brcal** a stiskněte tlačítko **Submit**.
4. V seznamu zobrazených výsledků (sekce UCSC Genes) si vyberte ten, který je uveden jako první. Kliknutím na tuto položku v hlavním okně Genome Browseru se dozvíte podrobné informace o tomto genu včetně struktury jeho proteinu. Vraťte se zpět.
5. Přejděte níže do sekce obsahující seznam stop a jejich konfiguraci. Klikněte na odkaz **SNPs (128)** ze skupiny **Variation and Repeats**. V této části si můžete prohlédnout aktuální nastavení pro zobrazení této stopy včetně dalších možností její vizualizace.
6. V sekci **Coloring Options** vyberte modrou barvu pro **Coding - Synonymous** a červenou barvu pro **Coding - Non-Synonymous**. Potvrďte tlačítkem **Submit**.
7. V grafickém prohlížeči prozkoumejte upravenou stopu **SNPs (128)**. Měli by jste nyní přehledně vidět výskyty synonymních i nesynonymních mutací genu **Brcal**.

8. Kliknutím na libovolnou položku stopy **SNPs (128)** si můžete prohlédnout podrobnější informace o příslušné mutaci.

=> **Nesynonymních SNP jsem napočítal 13, synonymních 15**

2 Úkol 2

S využitím Table Browser-u získáte seznam SNP elementů pro gen Clock u člověka.

Postup:

1. Otevřete si prohlížeč se stránkou UCSC Genome Browser (<http://genome.ucsc.edu/>) a přejděte do části pro práci s tabulkami (záložka Tables).
2. Zvolte clade: **Mammal**, genome: **Human**, assembly: **Feb. 2009**.
3. Vyberte skupinu **Variation**, stopu: **Common SNPs(138)** a tabulku **snp138Common**.
4. Do textového pole pro pozici vložte název **Clock** a stiskněte tlačítko **lookup**.
5. Zobrazí se vám seznam výsledků se záznamem **Clock**. Vyberte odkaz **Homo sapiens clock circadian regulator (CLOCK), transcript variant 1, mRNA** na pozici **chr4:56294068-56412099**.
6. Ponechte nastavení filtru a průsečíku nevyplněné.
7. Jako **Output Format** vyberte **selected fields from primary and related tables**. Stiskněte tlačítko **get output**.
8. Ve výsledné nabídce zaškrtněte políčka: **chrom, chromStart, chromEnd, name, strand, observed** a **func**. Stiskněte tlačítko **get output**.
9. Zobrazí se vám výsledný seznam položek, které si můžete zkopírovat nebo uložit pro další studium. Kolik jste našli položek?

=> **550 řádků výstupu (1 řádek hlavičky, 549 SNP)**

3 Úkol 3

S využitím Table Browser-u získajte seznam známých genů člověka vyskytujících se na konci chromozomu 22, ve kterých se nacházejí CpG ostrůvky (CpG islands). Získejte sekvence těchto genů ve formátu FASTA.

Postup:

1. Otevřete si prohlížeč se stránkou UCSC Genome Browser (<http://genome.ucsc.edu/>) a přejděte do části pro práci s tabulkami (záložka Tables).
2. Zvolte clade: **Mammal**, genome: **Human**, assembly: **Feb. 2009**.
3. Vyberte skupinu **Genes and Gene Predictions Tracks**, stopu: **UCSC Genes** a tabulku **knownGene**.
4. Do textového pole pro pozici vložte název **chr22** a stiskněte tlačítko **lookup**. Počáteční pozici upravte na **40.000.000**.
5. Stiskněte tlačítko **intersection**.
6. Vyberte skupinu **Regulation** a stopu **CpG Islands**. Ostatní volby ponechte v původním nastavení a stiskněte tlačítko **submit**.
7. Jako **Output Format** vyberte **sequence**. Stiskněte tlačítko **get output**.
8. Ve výsledné nabídce vyberte **genomic**. Stiskněte tlačítko **submit**.
9. Ujistěte se, že volby **5' UTR Exons**, **CDS Exons**, **3' UTR Exons** jsou zaškrtnuty a volba **Introns** odškrtnuta. Zvolte **One FASTA record per region** a ponechte ostatní volby v původním nastavení. Stiskněte tlačítko **Get Sequence**.
10. Zobrazí se vám výsledný seznam položek, které si můžete zkopírovat nebo uložit pro další studium. Kolik jste našli položek?

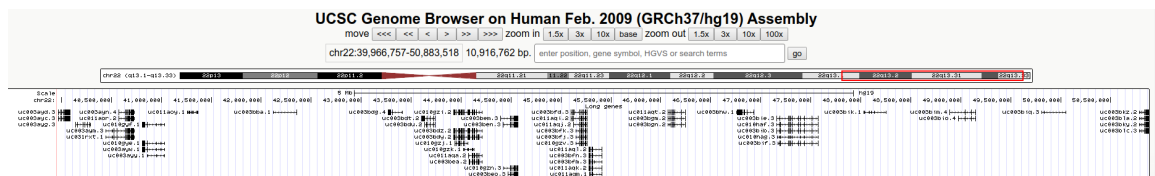
=> **4021 fasta záznamů**

4 Úkol 4

Prostudujte si návod na vytváření vlastních stop v Genome Browser-u a vytvořte vlastní stopu, která bude znázorňovat geny člověka na chromozomu 22, jejichž délka je větší než 100k bází.

Postup:

1. Otevřete si prohlížeč se stránkou UCSC Genome Browser (<http://genome.ucsc.edu/>) a prostudujte si návod na vytváření vlastních stop (záložka **Help**).
2. Přejděte do části pro práci s tabulkami (záložka **Tables**).
3. Zvolte clade: **Mammal**, genome: **Human**, assembly: **Feb. 2009**.
4. Vyberte skupinu **Genes and Gene Predictions Tracks**, stopu: **UCSC Genes** a tabulku **knownGene**.
5. Do textového pole pro pozici vložte název **chr22** a stiskněte tlačítko **lookup**. Počáteční pozici upravte na **40.000.000**.
6. Volby **filter** a **intersection** ponechte prázdné.
7. Jako **Output Format** vyberte **BED - browser extensible data**. Do položky **output file** vložte název výsledného souboru. Stiskněte tlačítko **get output**.
8. Ve výsledné nabídce zvolte volbu **Whole Gene**. Stiskněte tlačítko **get BED**.
9. Vytvořte si jednoduchý skript (Python, Perl, AWK, ...), který přečte soubor ve formátu **BED** a vypíše pouze ty geny, jejichž délka je větší než 100k bází. Výstup skriptu opět uložte do souboru ve formátu **BED**.
10. Vytvořený soubor vložte jako novou stopu do Geonome Browseru.



5 Vysvětlivky

- **SNP (Single-nucleotide polymorphism)**: označuje takový nukleotid v genomu, který se liší mezi jednotlivými členy určitého druhu nebo mezi párovými chromozomy příslušného jedince.

- **Synonymní mutace:** mutace kodonu, která nezpůsobí změnu aminokyseliny výsledného proteinu.
- **Nesynonymní mutace:** mutace kodonu, která způsobí změnu aminokyseliny výsledného proteinu.
- **BRCA1:** jedná se o gen, který pomáhá opravovat poškozenou DNA v buňkách prsu. Pokud je poškozen zvyšuje se riziko rakoviny prsu. Zdroje: wiki
- **CLOCK:** jedná se o gen, který kóduje protein související s režimem spánku. Jeho mutace mohou vést k poruchám spánku a k bipolárním poruchám. Zdroje: wiki, youtube video.
- **CpG ostrůvek:** jedná se o oblasti genomu, ve kterých je vysoká koncentrace CpG bází (kde p označuje fosfátovou vazbu).