

ANALIZA DANYCH Z UŻYCIEM PRZEGLĄDAREK GENOMOWYCH

Ćw. 1

Sprawdź jaki efekt na sekwencję białka mają wariacje SNP w mysim genie BRCA1 (breast cancer type 1 susceptibility protein)

1. Otwórz **Genome Browser** (genome.ucsc.edu)
2. Zresetuj ustawienia („[click here to reset browser](#)”)
3. Wybierz genom myszy, **assembly July 2007**
4. Wyszukaj gen **Brca1** (użyj podpowiedzi wyświetlanych podczas wpisywania)
5. Obejrzyj strukturę genu, zwróć uwagę na ścieżkę **SNPs (128)**
6. Obejrzyj szczegóły ścieżki **SNPs (128)** (na dole ekranu)
7. Rozwiń i obejrzyj dostępne opcje formatowania ścieżki. Z opcji kolorowania wybierz **Color Specification: Function**. Zmień **wszystkie typy SNP** na **czarny**, oprócz **Coding-NonSynonymous** dla których wybierz **blue**
8. Wybierz tryb wyświetlania ścieżki na **pack** (na górze strony)
9. Sprawdź różnice w prezentowaniu ścieżki
10. Wybierz pierwszy od lewej **niebieski SNP (rs28273098)**
11. Obejrzyj informacje które dla niego zostały zgromadzone w **genome browser**
12. Zapisz pozycję genomową oraz w którym eksonie znajduje się rs28273098.
13. Z której wersji bazy dbSNP pochodzą informacje widoczne w przeglądarce UCSC Genome Browser?
14. Przejdź do bazy danych **dbSNP**. Sprawdź jaka zmiana aminokwasu jest powiązana z tym SNP i zapisz do protokołu.

Ćw. 2

Wizualizacja SNP w Ensembl

1. Przejdź do strony ensembl: www.ensembl.org
2. Wybierz genom myszy i znajdź gen **Brca1**. Z karty genu przejdź do jego lokalizacji (Location)
3. Włącz wyświetlanie ścieżki **Sequence variants**. Aby to zrobić, wybierz z menu po lewej „Configure this page”. Po lewej stronie znajduje się lista dostępnych ścieżek. Wybierz grupę **Variation** i **sequence variants**. Zaznacz ścieżkę **Sequence variants (all sources)** żeby wyświetlona została w stylu **expanded with names**.
4. Postaraj się znaleźć tą samą zmianę SNP jak w ćwiczeniu 3. Zauważ, że nazwy SNP wyświetlane będą jeżeli okno widoku obejmować będzie mniej niż 10 kb sekwencji.
5. Zapisz do protokołu czy udało ci się znaleźć tą zmianę SNP. W której przeglądarce genomowej było to łatwiejsze?

Ćw. 3

Znajdź dane wskazujące na lokalizację promotora ludzkiego genu **NRAS (neuroblastoma RAS viral (v-ras) oncogene homolog)**

1. Analogicznie jak powyżej zresetuj **genome browser** i znajdź w **genomie ludzkim (assembly Feb. 2009)** gen **NRAS** (użyj podpowiedzi wyświetlanych podczas wpisywania)
2. Zwiększ widziany obszar o 1000 pz z obu stron genu, modyfikując zakres wyświetlanej pozycji
3. Pośród listy dostępnych ścieżek znajdź **TFBS Conserved** (w grupie **Regulation**) i wybierz opcję wyświetlania **full**. Pamiętaj że każda zmiana sposobu wyświetlania ścieżek musi zostać zatwierdzona przyciskiem **refresh**
4. Przyjrzyj się danym które pojawiły się na wykresie. Zwróć uwagę na kierunek transkrypcji genu.
5. Obejrzyj szczegóły dla pierwszego od góry czynnika transkrypcyjnego (**V\$CDC5_01**). Użyj jego identyfikatora **SwissProt** aby dowiedzieć się o jego funkcji w bazie **Uniprot** (www.uniprot.org). Zanonuj znalezione funkcje w protokole.

Ćw. 4

Dane regulatorowe w Ensembl

1. Spróbuj powtórzyć ćwiczenie 5 w bazie Ensembl. W celu wyświetlenia miejsc wiązania czynników transkrypcyjnych użyj grupy ścieżek **regulation** i w jej ramach **Open chromatin and TFBS**.
2. Podczas włączania ścieżki włącz filtr dostępnych danych ograniczając je do **Transcription Factor**. Zapisz do protokołu czy możesz znaleźć ten sam czynnik transkrypcyjny który analizowaliśmy w ćwiczeniu 5? Następnie włącz wszystkie dostępne dane (**Enable/disable all**) i ustaw sposób wyświetlania na **Both (Peaks oraz Signal)**.
3. Porównaj sposób prezentacji danych pomiędzy przeglądarkami genomowymi. Zapisz do protokołu którą z przeglądarek uważasz za bardziej przydatną do analizy miejsc wiązania czynników transkrypcyjnych i dlaczego?

Ćw. 5

Identyfikacja genów regulowanych przez czynnik regulatorowy SMARCB1.

Przeczytaj o funkcji czynnika CREB1 wyszukując jego nazwę w bazie Uniprot (<http://uniprot.org>).

1. W przeglądarce **UCSC Genome Browser** wybierz z menu **Tools -> Table Browser**.
Pobierz dane dla genomu człowieka (złożenie **hg19, chromosom 22**) dotyczące pozycji wiązania czynnika transkrypcyjnego **SMARCB1**. W tym celu wybierz:
 - a. Group: **Regulation**
 - b. Track: **Txn Factor ChIP**
 - c. Table: **WgEncodeRegTfbsClusteredV3**
 - d. Region: **position: chr22**
 - e. Filter: wciśnij **create**, następnie w pozycji **name** ustaw **SMARCB1** i wciśnij **submit**
 - f. Output format: **BED**
 - g. Output file: **factors.bed.gz**
 - h. File type returned: **gzip compressed**Kliknij **get output** i na następnej stronie potwierdź pobranie danych.
2. Pobierz również pozycje znanych genów wg adnotacji ENSEMBL. W tym celu wybierz:
 - a. Group: **Gene and gene prediction tracks**
 - b. Track: **Ensembl Genes**
 - c. Table: **ensGene**
 - d. Region: **position: chr22**
 - e. Output format: **BED**
 - f. Output file: **genes.bed.gz**
 - g. File type returned: **gzip compressed**Kliknij **get output** i na następnej stronie potwierdź pobranie danych.
3. Otwórz stronę publicznego serwera Galaxy: <http://usegalaxy.org> W panelu narzędzi po prawej stronie wybierz narzędzie **Get Data -> Upload File**. Prześlij na serwer oba pliki pobrane w poprzednim kroku.
4. Posortuj pliki BED. W tym celu wybierz narzędzie **BEDTools -> Sort BED Files**.
Uruchom narzędzie dwukrotnie z domyślnymi opcjami wybierając każdy z załadowanych plików.
5. Wyszukaj geny najbliższe miejscom wiązania czynników transkrypcyjnych. W tym celu uruchom narzędzie **BEDTools -> ClosestBed**. Jako pierwszy plik wybierz plik z posortowanymi pozycjami czynnika transkrypcyjnego **SMARCB1**, natomiast w drugim oknie wybierz plik z posortowanymi pozycjami genów. Opcje pozostaw domyślne za wyjątkiem **Ignore Features in B that overlap A** którą należy ustawić na **Yes**.
6. Wytnij z powstałej tabeli kolumnę zawierającą identyfikatory genów (kolumna 9). W tym celu uruchom narzędzie **Text Manipulation -> Cut columns from a table**. Jako warunek wycinania w pozycji **Cut columns** wpisz **c9**.

7. Otwórz podgląd uzyskanego wyniku (ikona „oczka” przy nazwie pliku w historii) i przekopiuj listę genów do schowka. Następnie przejdź do strony serwisu **GProfiler**: <http://biit.cs.ut.ee/gprofiler/index.cgi> i wklej listę genów w odpowiednie okno, ustaw organizm na Homo sapiens i uruchom narzędzie. Przeanalizuj uzyskane wyniki pod kątem udziału genów z listy w znanych sieciach regulatorowych oraz sprawdź czy lista jest wzbogacona w geny należące do opisanych kategorii funkcjonalnych.