

Zajęcia I – PRZEGLĄDARKI GENOMOWE: UCSC oraz ENSEMBL

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** (www.genome.ucsc.edu)
2. Przejdź do zakładki **Genomes** i wybierz genom myszy, **assembly July 2007**
3. Wyszukaj gen **Brca1** (użyj podpowiedzi wyświetlanych podczas wpisywania)
4. Zresetuj ustawienia (przyciski „**default tracks**” oraz „**default order**”)
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. 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 w której przeglądarce genomowej łatwiej było znaleźć poszukiwaną zmianę SNP?

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). Zanotuj 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**. 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

Znajdź gen ulegający alternatywnemu składaniu mRNA na podstawie danych z wysokoprzepustowego sekwencjonowania

1. Otwórz genome browser (human, assembly hg19) wyszukaj ludzki gen **TSPAN2** i przywróć domyślne ustawienia wyświetlania.
2. Z dostępnych ścieżek znajdź **Burge RNA-seq** (w grupie **Expression**)
3. Wejdź w ustawienia ścieżki i wybierz następujące opcje:
 - Ustaw **Raw Signal** na **full**
 - Ustaw **Alignments** na **hide**

- Wybierz podścieżki (subtracks): **Adipose, Brain, Breast, Colon, Heart, Liver, LymphNode, SkelMuscle, Testes**
 - Zatwierdź zmiany przyciskiem **submit**
4. Zidentyfikuj różnice w ekspresji genu **TSPAN2** pomiędzy tkankami. Zapisz do protokołu ile egzonów ulega alternatywnemu składaniu?

ćw. 6

Wizualizacja własnych ścieżek z danymi

1. W nowym oknie przeglądarki pobierz pliki do ćwiczeń z adresu: www.combio.pl, zakładka Teaching
2. Otwórz **genome browser** i zresetuj ustawienia.
3. Wybierz genom człowieka **Homo sapiens (assembly Feb. 2009)**
4. W **genome browser** wybierz przycisk **add custom tracks**
5. Wybierz przycisk **wybierz plik**
6. Wybierz plik z rozszerzeniem **.wig.tar.gz**
7. Kliknij **submit**
8. Kliknij nazwę ścieżki i edytuj konfigurację dodając:


```
track type=wiggle_0 name="normal_MCF" description="normal" visibility=full autoScale=on
alwaysZero=on maxHeightPixels=50:50:5 color=0,120,0
```
9. Powtórz czynność z plikiem dla linii komórkowej HCC1937 zmieniając nazwę na „tumor_HCC1937”
10. Pojawi się lista dodanych ścieżek, kliknij **go to genome browser**
11. Zlokalizuj dodaną ścieżkę na wykresie i w liście ścieżek
12. Wyszukaj gen **AFAP-AS1**
13. Znajdź różnicę pomiędzy wprowadzonymi ścieżkami i zinterpretuj wyniki

Szczegółowe opcje konfiguracyjne opisane są pod adresem:

<http://genome.ucsc.edu/goldenPath/help/hgTracksHelp.html#CustomTracks>

ćw. 7

Tworzenie własnych ścieżek w formacie BED

UWAGA: sposób tworzenia własnych ścieżek wraz z formatami danych opisany jest na stronie

<http://genome.ucsc.edu/goldenPath/help/hgTracksHelp.html#CustomTracks>

1. Otwórz notatnik

2. Wprowadź trzy linie zawierające zakresy danych w taki sposób, żeby wyświetlić na chromosomie 1 trzy linie o długości 100 pb z odstępem 100 pb. Minimalny format linii danych BED to :
chromosom start koniec nazwa punktacja nić np. chr22 1000 1100
linia1 0 +
3. W przypadku wyświetlania plików BED w UCSC Genome Browser można ustawić sposób graficznego wyświetlania ścieżki. Jest to możliwe dzięki dodaniu pierwszej konfiguracyjnej linii uzupełnionej swoimi danymi. Zwróć uwagę na fakt, że wokół znaku „=” nie powinno być spacji. Jeśli w polach **name** lub **description** chcesz umieścić kilka słów, musisz umieścić je w cudzysłowie:
track name= "description="
4. Możesz również skonfigurować ścieżkę tak, aby **genome browser** automatycznie przeniósł cię do właściwego miejsca na chromosomie po załadowaniu ścieżki. W tym celu jako pierwszą linię wprowadź linię **browser** w następujący sposób (dla powyższego przykładu linii BED, w ćwiczeniu wprowadź własną nazwę chromosomu i pozycję) : np. browser position chr22:1000-1100
5. Zapisz plik z rozszerzeniem .bed
6. Załaduj go do genome browser
7. Sprawdź czy wyświetlony obraz odpowiada zamierzeniom
8. Spróbuj wprowadzić modyfikacje do twojego pliku BED – zmiany ustawień, kolorów oraz zmiany ilości danych. Wynik swoich eksperymentów wklej do protokołu w formacie pliku BED (zawartość).

Ćw 8

Wyświetlanie własnych ścieżek w Ensembl

1. Załaduj stworzoną przez siebie ścieżkę BED do Ensembl. W tym celu wybierz genom dla którego tworzona była ścieżka. Wybierz opcję **Add your data** z menu. Jako **data format** wybierz BED. Załaduj plik.
2. Zapisz do protokołu czy udało się wyświetlić dane zawarte w pliku BED? Jakie różnice w wyświetlaniu ścieżek udało ci się zauważyć? Który ze sposobów jest dla ciebie bardziej czytelny?