

## Zajęcia II – Sekwencjonowanie

### Pobieranie danych z bazy NCBI Short Read Archive SRA

#### Ćw. 1

##### Pobieranie sekwencji

1. Otwórz stronę główną NCBI Short Read Archive (SRA) <http://www.ncbi.nlm.nih.gov/sra>
2. Wyszukaj rekord eksperymentu o numerze SRX060173
3. Obejrzyj szczegóły eksperymentu. Zanotuj w protokole z jakiego typu komórek pochodziło RNA użyte w eksperymencie, czy pochodziło ono od osobników palących, czy osobniki były chore na raka płuc
4. Kliknij w link Study w linii Study summary
5. Przeczytaj opis badań w ramach którego zsekwencjonowana została wyżej wymieniona biblioteka cDNA (zielony link „show abstract”)
6. Wejdź w opis eksperymentu SRX060176. Zanotuj w protokole informacje analogiczne jak dla poprzedniego eksperymentu

### Analiza jakości oraz wstępna obróbka odczytów z sekwencjonowania

#### Ćw. 2

##### Analiza jakościowa sekwencji

1. Pobierz plik: <http://combio.pl/files/healthly.zip>
2. Wejdź na stronę serwisu Galaxy ( <https://usegalaxy.org/> )
3. W sekcji Tools wybierz Get Data → Upload File
4. Załaduj pobrany pliki
5. W sekcji Tools przejdź do FASTQ Quality Control
6. Przejdź do narzędzia FastQC i przeprowadź analizę dla sekwencji w pliku healthly.fastq
7. Zapisz do protokołu interpretację uzyskanych wyników. (rozkład jakości sekwencji, obecność adaptorów/primerów na końcach odczytów, obecność duplikatów)

8. Przejdź do narzędzia: FASTQ Quality Control → Trimmomatic
9. Ustaw następujące parametry:
  - a. Number of bases to average across : 10
  - b. Average quality required: 20
10. Zapisz statystyki do protokołu. 11. Powtórz analizę jakości sekwencji po filtrowaniu za pomocą FastQ
11. Zapisz do protokołu czy zmieniła się jakość sekwencji

### Ćw. 3

Samodzielna analiza jakościowa sekwencji

1. Powtórz analizę wykonaną w ćw.2 dla danych pochodzących z pacjentów z nowotworem:

<http://combio.pl/files/cancer.zip>

2. Zapisz statystyki do protokołu.

### Ćw4.

Charakterystyka odczytów uzyskiwanych z różnych technologii sekwencjonowania.

1. Załaduj przykładowe dane z różnych technologii sekwencjonowania do Galaxy
  - a. **Illumina**
    - i. Przejdź do: Shared Data -> Data Libraries -> Demonstration Dataset -> Human RNA-seq
    - ii. Wybierz plik Cd20\_rep1
    - iii. Załaduj do historii
  - b. **MinION Oxford Nanopore**
    - i. Przejdź do Get data → Upload file
    - ii. Wybierz opcję Paste/fetch data
    - iii. Wklej link  
[http://www.plabipd.de/projects/SpennLost/minion\\_albacore\\_2\\_1\\_1/20161027\\_Spenn\\_001\\_001\\_all.fastq.gz](http://www.plabipd.de/projects/SpennLost/minion_albacore_2_1_1/20161027_Spenn_001_001_all.fastq.gz)
    - iv. Pobierz dane (Build)
  - c. **454**
    - i. Shared Data -> Data Libraries -> Sample NGS Data Sets
    - ii. Wybierz pliki example\_human\_454 (.fa i .qv)
    - iii. W sekcji 'FASTA/FASTQ' wybierz 'Combine FASTA and QUAL into FASTQ'
    - iv. Utwórz plik FASTQ z example\_human\_454.fa i

example\_human\_454.qv

- v. Zmień nazwę utworzonego pliku
2. Dla każdego pliku fastq wykonaj analizę jakości za pomocą programu FastQC  
FASTQ Quality Control -> FastQC
3. Dla każdej z technologii, przygotuj zestawienie następujących informacji w formie tabeli:
  - a. Rozkład długości odczytów
  - b. Maksymalna długość odczytu
  - c. Rozkład jakości odczytów (początek, koniec odczytu)
  - d. Przedział średnich wartości jakości dla odczytu
  - e. Wystąpienie „N” w sekwencji
4. Porównaj wyniki analizy jakości uzyskane dla wyżej wymienionych danych. Zapisz do protokołu która wykazuje największe różnice pomiędzy technologiami sekwencjonowania.

#### Ćw5.

Porównanie jakości sekwencji uzyskanych z użyciem illuminy dla różnych typów eksperymentów

1. Porównaj wyniki analizy jakości uzyskane w ćwiczeniu 4 dla Illuminy z wynikami twoich analiz z ćwiczenia 3.
2. Zapisz różnice pomiędzy analizowanymi danymi do protokołu.