Analiza danych uzyskanych z sekwencjonowania wysokoprzepustowego Illumina – ciąg dalszy.

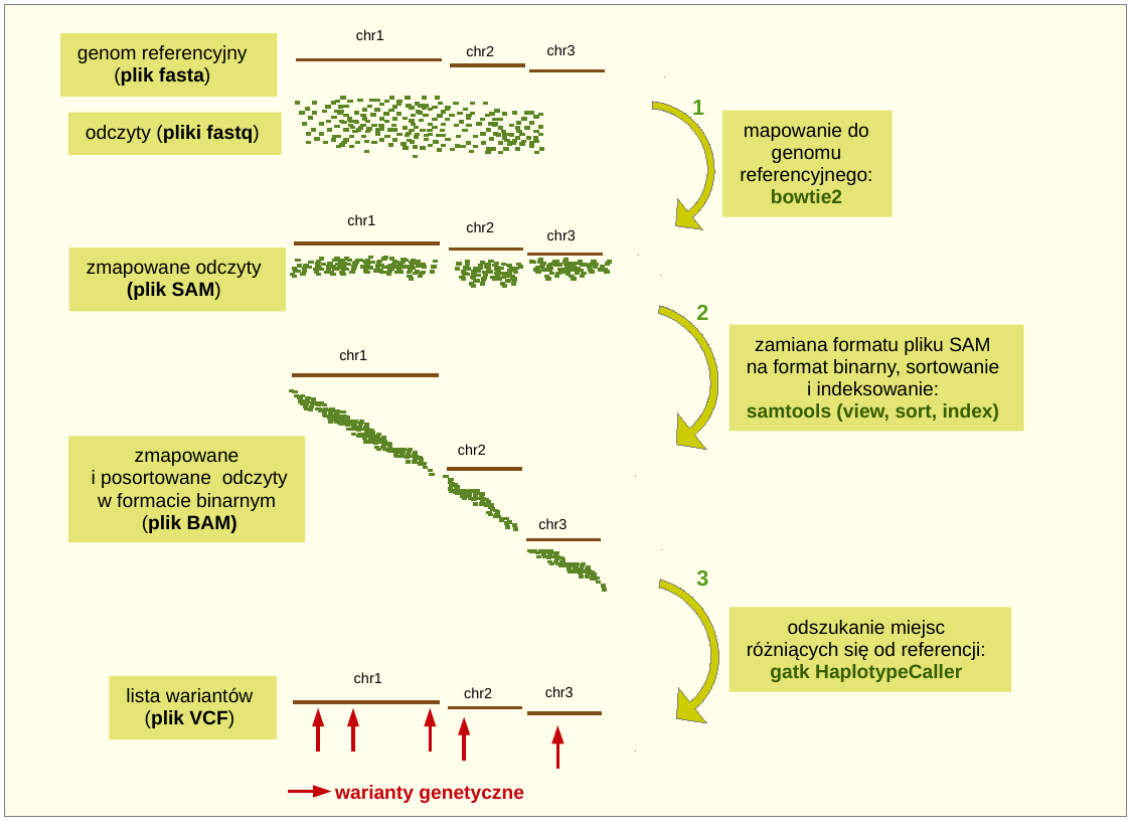
Na poprzednich zajęciach dowiedzieli się państwo, jak wyglądają pliki wynikowe z sekwencjonowania nowej generacji oraz jak zanalizować jakość otrzymanych sekwencji. Dzisiaj przeprowadzą państwo analizę sekwencji genomowego DNA uzyskanych dla trzech osób: rodziców i ich syna. Syn ten urodził się z nieznacznymi wadami (deformacje kości kciuków), nie budzącymi większego zaniepokojenia lekarzy i rodziców. Niestety, szybko okazało się, że jego rozwój nie przebiega normalnie. Z jakiegoś powodu w jego organizmie odkłada się żelazo, jednocześnie poziom erytrocytów we krwi jest bardzo niski. Wiąże się to z poważną niedokrwistością, która nie może być leczona podawaniem żelaza.

Lekarze podejrzewają, że dziecko cierpi na jakąś rzadką chorobę genetyczną. Wykrycie jej podłoża mogło by pomóc w opracowaniu sposobu leczenia objawowego. Co ważne, chłopiec ma liczne i całkowicie zdrowe rodzeństwo – dwie siostry i trzech braci. Podobnych objawów chorobowych nie obserwowano też u innych spokrewnionych z nim osób. Sugeruje to, że ewentualna wada genetyczna jest nową mutacją, powstałą przy tworzeniu się komórek rozrodczych u jednego z rodziców. Mutacja ta musiałaby mieć przynajmniej częściowo dominujący charakter.

Dzisiejsze zajęcia poświęcone będą wykonaniu analiz sekwencji uzyskanych dla omawianej rodziny, tak aby wynikiem końcowym była lista miejsc zmutowanych (innych niż w sekwencji referencyjnej) obecnych u każdej z badanych osób. Lista ta posłuży następnie do wykrycia mutacji obecnych tylko u chorego dziecka.

Uwaga! Wielkość plików fastq zawierających odczyty uzyskane dla wszystkich eksonów jest na tyle duża, że ich analiza zajęłaby bardzo dużo czasu. Ponadto ograniczenie stanowi ilość miejsca na serwerze. Dlatego też będą państwo pracować na okrojonych plikach. Odczyty będą państwo mapować do fragmentu chromosomu 1.

Kroki analizy



1) Krok pierwszy ze schematu. Mapowanie odczytów, czyli wskazanie dla każdego z nich miejsca w genomie referencyjnym o takiej samej, lub bardzo podobnej sekwencji.

Do mapowania można używać różnych programów. Podczas zajęć wykorzystają państwo program `bowtie2`.

i) Tworzenie indeksu genomu (lub innej sekwencji referencyjnej):

```
bowtie2-build ref.fasta ref_index
```

`ref.fasta` to plik z sekwencją referencyjną w formacie fasta (do niej chcemy zmapować odczyty)
`ref_index` to nazwa, jaką będą miały stworzone przez program bowtie2 pliki z indeksem

ii) Mapowanie odczytów zawartych w plikach fastq do sekwencji referencyjnej. Gdy analizowane są sparowane odczyty, polecenie wygląda następująco:

```
bowtie2 --rg-id nazwa --rg SM:nazwa -x ref_index -S plik.sam -1 odczyty_1.fastq \
        -2 odczyty_2.fastq
```

`--rg-id` po spacji podany jest nazwa sekwencjonowanej próbki (`nazwa`)
`--rg` po spacji podana jest nazwa sekwencjonowanej próbki (`SM:nazwa`)
Dodanie obu powyższych opcji powoduje, że w liniach wynikowego pliku SAM znajdzie się informacja, z jakiej próbki pochodzi dany odczyt (ostatnia kolumna, po znaczniku RG:Z:)

- `-x` za tą opcją, po spacji, podana jest nazwa indeksu (`ref_index`)
- `-S` po spacji podana jest nazwa wynikowego pliku w formacie SAM (`plik.sam`)
- `-1` po spacji podana jest nazwa pierwszego z pary plików fastq do analizy (`odczyty_1.fastq`)
- `-2` po spacji podana jest nazwa drugiego z pary plików fastq (`odczyty_2.fastq`)

2) Krok drugi ze schematu. Zmiana formatu pliku ze zmapowanymi odczytami (plik SAM) na format binarny (BAM). Wynikowy plik binarny jest nieczytelny dla człowieka, ale pozwala innym programom na wydajniejsze manipulacje zawartymi w nim danymi. Plik będzie następnie sortowany (według pozycji w sekwencji referencyjnej) i indeksowany (aby umożliwić szybszy dostęp do danych).

i) Do konwersji formatu SAM ↔ BAM można wykorzystać program `view` z pakietu `samtools`:

```
samtools view -b -o plik_ns.bam plik.sam
```

`-b` opcja informująca, że plik wynikowy ma być w formacie BAM.
Domyślnie program `view` zmienia format BAM w SAM.
`-o` po spacji podana jest nazwa wynikowego pliku w formacie BAM (`plik_ns.bam`)
Na końcu polecenia podana jest nazwa pliku, którego format ma być zmieniony (`plik.sam`).

ii) Sortowanie pliku BAM, z użyciem programu `sort` z pakietu `samtools`:

```
samtools sort plik_ns.bam -o plik.bam
```

`plik_ns.bam`, to nieposortowany jeszcze plik BAM
Po opcji `-o` podana jest nazwa wynikowego, już posortowanego pliku (`plik.bam`).

iii) Tworzenie indeksu dla pliku BAM, z wykorzystaniem programu `index` z pakietu `samtools`:

```
samtools index plik.bam
```

`plik.bam`, to posortowany plik BAM; program utworzy dodatkowy plik z indeksem, o nazwie `plik.bam.bai`

3) Krok trzeci ze schematu. Odnalezienie miejsc, w których odczyty różnią się od sekwencji referencyjnej. Tę operację po angielsku określa się jako *variant calling* albo *SNP calling*.

Do szukania miejsc zmienionych wykorzystają państwo program `HaplotypeCaller` (pakiet `gatk` w wersji `4.2.2.0`):

i) Przygotowanie pomocniczych plików referencyjnych dla programu `HaplotypeCaller` (program `faidx` z pakietu `Samtools` i program `CreateSequenceDictionary` z pakietu `gatk`):

```
samtools faidx ref.fasta
gatk CreateSequenceDictionary -R ref.fasta -O ref.dict
```

Pierwsze polecenie tworzy indeks dla pliku referencyjnego (zapisany w dodatkowym pliku `ref.fasta.fai`). Drugie polecenie tworzy plik `ref.dict`, w którym znajdują się dodatkowe informacje o sekwencji referencyjnej wymagane przez program `HaplotypeCaller`, przy czym:
po opcji `-R` podajemy nazwę pliku z sekwencją referencyjną w formacie fasta (do niej chcemy zmapować odczyty, tutaj `ref.fasta`); natomiast po opcji `-O` podana jest nazwa drugiego pliku pomocniczego, stworzonego przez program `CreateSequenceDictionary`.

ii) Wyszukiwanie miejsc różniących się pomiędzy odczytami i sekwencją referencyjną:

```
gatk HaplotypeCaller -R ref.fasta -I plik.bam -O plik.vcf
```

Program `HaplotypeCaller` odnajduje te pozycje w sekwencji referencyjnej, które są zmienione przynajmniej w części zmapowanych odczytów.
Po opcji `-R` podana jest nazwa sekwencji referencyjnej w formacie fasta (`ref.fasta`).
Plik z danymi do analizy to `plik.bam` (podany po opcji `-I`)
Plik wynikowy to `plik.vcf` (podany po opcji `-O`)

Zadanie1

Proszę przejść do katalogu `ngs/ref` i wyświetlić znajdujące się w nim pliki. Jeśli wykonali państwo zadanie domowe z pierwszych zajęć :), to powinni państwo odnaleźć plik `chr1_fragment.fasta`. Plik ten zawiera fragment referencyjnej sekwencji chromosomu 1 człowieka (pozycje od 91 500 000 do 94 000 000).
i) Proszę otworzyć plik w edytorze `nano` i zmienić wiersz opisu na `>chr1`. Ułatwi to odrobinę dalszą pracę. Proszę zapisać wprowadzone zmiany.
ii) Proszę zbudować wymagany przez program bowtie2 indeks dla sekwencji z pliku `chr1_fragment.fasta`.
Indeks proszę zapisać w katalogu `ngs/ref`, proszę nadać mu nazwę: `chr1_index`
Polecenie:

Czy po wydaniu polecenia w katalogu pojawiły się jakieś nowe pliki?

Zadanie2

Państwa zadaniem będzie napisanie skryptu, który przeprowadza wstępne kroki analizy sekwencji uzyskanych w technologii ngs: mapowanie odczytów, sortowanie i indeksowanie plików z dopasowaniem (BAM).
W katalogu `/dane/ngs` znajdują się pliki z sekwencjami uzyskanymi dla matki (`matka_1.fastq`, `matka_2.fastq`), ojca (`ojciec_1.fastq`, `ojciec_2.fastq`) i ich syna (`syn_1.fastq`, `syn_2.fastq`). Dodatkowo znajdują się tam dwa niewielkie pliki, które posłużą państwu do

przetestowania napisanego skryptu (`test_1.fastq`, `test_2.fastq`). Proszę o skopiowanie tych plików do katalogu `ngs` na swoim koncie. Proszę skopiować też plik `komendy2.txt`. Znajdą w nim państwo wszystkie polecenia, które przeprowadzają kolejne kroki analizy.

Proszę otworzyć plik `komendy2.txt` w edytorze `nano` i zmienić go w skrypt do analizy danych ngs.
i) Skrypt ten powinien jako argument wiersza poleceń przyjmować nazwę osoby, której sekwencje analizujemy (np. `matka`).
ii) Nazwy wszystkich plików wynikowych również powinny zawierać tę nazwę (np. `matka.sam`, `matka_ns.bam`, `matka.bam` itd.), Proszę także, aby w ten sam sposób nazwać analizowane próbki (np. `--rg-id matka`, `--rg SM:matka`)
iii) Proszę pamiętać o zmianie ścieżek dostępu i nazw plików z sekwencją referencyjną i indeksem programu bowtie2:
plik referencyjny to `chr1_fragment.fasta`
utworzony dzisiaj przez państwa indeks ma nazwę `chr1_index`
pliki znajdują się w katalogu `ngs/ref`)
iv) Proszę dopisać komentarze krótko wyjaśniające, co robią poszczególne polecenia skryptu oraz, co przyjmuje on jako argument wiersza poleceń.

Proszę zmienić nazwę skryptu na `mapowanie.sh` i zapisać go w katalogu `ngs`. Proszę nadać wszystkim użytkownikom skryptu prawo do jego wykonywania i sprawdzić, czy skrypt działa z argumentem `test`. (Czy otrzymali państwo odpowiednie `pliki test.bam`, `test.bam.bai`?) Proszę zademonstrować skrypt osobie prowadzącej ćwiczenia.

Jeśli skrypt działa poprawnie, można go zastosować do analizy sekwencji uzyskanych dla rodziców i ich syna.

Gdy spodziewają się państwo, że jakieś zadanie, które ma być wykonane w powłoce Shell, zabierze sporo czasu, warto skorzystać z programu `screen`:

```
screen -S nazwa_terminala
```

Komenda ta spowoduje otwarcie sesji „wirtualnego” terminala. Można w nim teraz pracować tak samo, jak w zwykłym terminalu. Co ważne, można w nim wydać polecenie (lub uruchomić skrypt), którego wykonanie wymaga dużo czasu, po czym opuścić „wirtualny” terminal, nie przerywając wykonywania polecenia przez komputer (wszystkie procesy będą kontynuowane). Aby to zrobić, należy nacisnąć jednocześnie klawisze ***Ctrl*** i ***a***, po czym ***Ctrl*** i ***d***. Po opuszczeniu „wirtualnego” terminala można zamknąć także zwykły terminal (`Exit`) i iść do domu lub też otworzyć następny wirtualny terminal, w którym wykonywane będą inne zadania.

Aby ponownie powrócić do któregoś z wirtualnych terminali programu screen, należy wpisać komendę:

```
screen -r nazwa_terminala
```

Aby zobaczyć, nazwy wszystkich wytworzonych „wirtualnych” terminali należy wpisać:

```
screen -ls
```

Aby pozbyć się któregoś z „wirtualnych” terminali (gdy wydane w nim polecenia zostały wykonane) należy wejść do niego, po czym wpisać komendę `Exit`.

Zadanie3

Proszę uruchomić napisany przez siebie skrypt dla odczytów uzyskanych dla matki w „wirtualnym” terminalu programu `screen`. Proszę opuścić ten „wirtualny” terminal, bez przerywania zachodzących w nim procesów. Podobnie proszę w dwóch nowych „wirtualnych” terminalach ponownie uruchomić skrypt dla ojca i dla syna.

Zadanie4

Pliki w formacie BAM zawierają informacje o odczytach (sekwencja, jakość dla każdego nukleotydu) oraz informacje dotyczące ich mapowania (położenie odczytu względem sekwencji referencyjnej, jakość zmapowania, itp). Pliki te są binarne i aby je obejrzeć, należy użyć odpowiednich programów:

i) jeśli chcemy obejrzeć zawarte w pliku BAM informacje w formie tabeli, zmieniamy format pliku z BAM na SAM (domyślne działanie programu `samtools view`) :

```
samtools view plik.bam > plik.sam
samtools view plik.bam | less
samtools view plik.bam chr1:20000-80000 | less
```

Pierwsze polecenie zamieni binarny format pliku `plik.bam`, na czytelny dla człowieka format SAM i wynik zapisze do pliku `plik.sam`.
Drugie polecenie także przeprowadzi konwersję formatu, ale wynik zostanie wyświetlony w programie `less`.
Trzecie z poleceń pozwoli obejrzeć w programie `less` odczyty zmapowane do wybranego odcinka sekwencji referencyjnej (proszę zwrócić uwagę, jak zapisać interesujące nas położenie).

Dane w pliku zapisane są w formie tabeli – format SAM:
kol1: identyfikator odczytu (taki, jak w pliku fastq)
kol2: flaga binarna (*ang. SAM flag*) (tu zakodowane są informacje o zmapowaniu, np. do której nici odczyt został zmapowany, czy jest on unikalny itp.)
kol3: nazwa sekwencji referencyjnej
kol4: pozycja w sekwencji referencyjnej, od której zaczyna się zmapowanie
kol5: jakość zmapowania ($-10log_{10}p$, gdzie p to prawdopodobieństwo, że odczyt jest zmapowany niepoprawnie – to znaczy nie do tego miejsca, skąd naprawdę pochodzi)
kol10: sekwencja odczytu
kol11: informacja o jakości poszczególnych nukleotydów

Proszę obejrzeć jeden z otrzymanych przez państwa plików BAM (po zamianie formatu na tekstowy) w programie `less`.
Polecenie:

Proszę zanalizować jeden z wierszy:
Nazwa odczytu:

Miejsce zmapowania:

Flaga binarna:

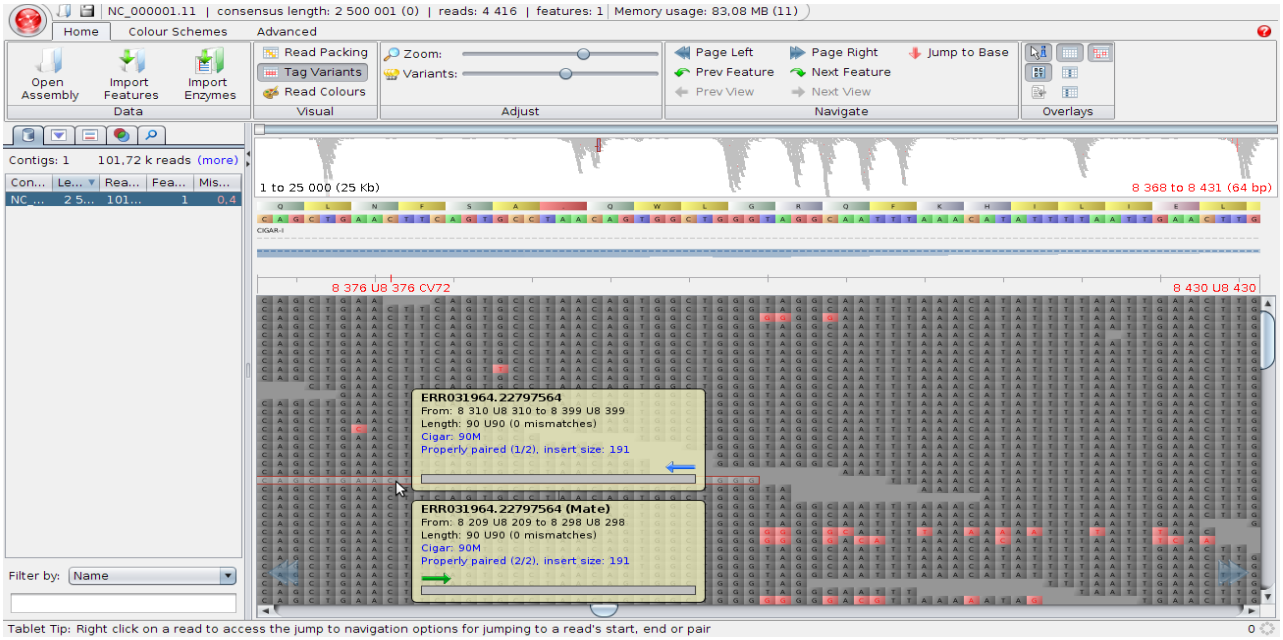
Czy odczyt wchodził w skład prawidłowo zmapowanej pary?

Do której nici został zmapowany odczyt?

Aby odpowiedzieć na dwa ostatnie pytania trzeba rozszyfrować znaczenie flagi binarnej. Proszę wykorzystać stronę: https://broadinstitute.github.io/picard/explain-flags.html
(odczyt po angielsku to *read*, drugi odczyt z pary to *mate*)

ii) Jeśli chcemy obejrzeć uzyskane zmapowanie w formie graficznej, można użyć programu ***Tablet***.

Proszę przenieść na komputer lokalny jeden z otrzymanych plików BAM, plik z jego indeksem (np. `matka.bam`; `matka.bam.bai`) oraz plik fasta z sekwencją referencyjną (`chr1_fragment.fasta`) i umieścić je w jednym katalogu.
Proszę uruchomić program ***Tablet***. Następnie proszę nacisnąć ikonkę ***Open Assembly***. Pojawi się teraz okno dialogowe. Proszę wskazać plik BAM, np. `matka.bam` (w oknie ***Primary assembly file or URL***), plik z referencją (w oknie ***Reference/consensus file or URL***), po wybrać przycisk ***open***.
W oknie po lewej stronie pojawi się teraz lista sekwencji referencyjnych (u państwa jednoelementowa). Proszę kliknąć w sekwencję chr1 i obejrzeć zmapowanie:



Proszę odpowiedzieć na pytania:
Czy odczyty pokrywają całą sekwencję referencyjną jednolicie?

Czy jest to spodziewany wynik?

Proszę ustawić kursor na jednym z odczytów i podać następujące informacje:
Jak nazywa się ten odczyt?

Gdzie został zmapowany?

Jaką ma długość?

Do której nici został zmapowany?

Jaka jest przerwa pomiędzy nim i drugim odczytem z pary?

Czy zawiera zmienione nukleotydy?

Zadanie5

Kolejnym etapem analizy, mającym na celu odnalezienie mutacji wywołującej chorobę syna, jest wyszukanie miejsc w genomach wszystkich trzech osób, które różnią się od sekwencji referencyjnej (trzeci krok ze schematu). Do przeprowadzenia tego zadania wykorzystają państwo program `HaplotypeCaller` (program z pakietu `gatk`). Analizę tą wykonają państwo dla wszystkich osób naraz – w tym celu trzeba będzie przygotować wspólny (zindeksowany) plik BAM. Posłuży do tego program `MergeSamFiles` (także z pakietu `gatk`).

Proszę o skopiowanie do katalogu `ngs` pliku `komendy5.txt` (z `/dane/ngs`), gdzie znajdą państwo wszystkie potrzebne polecenia (bez zmienionych ścieżek dostępu i nazw plików) oraz pomocnicze pytania, które powinny ułatwić wykonanie tego ćwiczenia. Proszę otworzyć i przeanalizować plik z poleceniami.

Proszę następnie:
i) przygotować wspólny (połączony) plik BAM dla rodziców i syna (`razem.bam)` W tym celu proszę wykorzystać polecenie:

```
gatk MergeSamFiles -I plik1.bam -I plik2.bam -I plik3.bam -O wyjściowy.bam
```

Nazwy plików BAM, które chcemy połączyć należy podać po opcji `-I`
Nazwę pliku wyjściowego należy po opcji `-O`

ii) **utworzyć indeks dla pliku `razem.bam`** (wykorzystane już przez państwa polecenie `samtools index`).

iii) **przygotować pomocnicze pliki referencyjne dla programu `HaplotypeCaller`** (opis poleceń potrzebnych do wykonania tego kroku znajduje się przed treścią pierwszego zadania). Uzyskane pliki referencyjne powinny być zapisane w katalogu `ngs/ref` (lub tam, gdzie znajduje się u państwa plik z sekwencja referencyjną!). Plik wyjściowy uzyskany w wyniku działania programu `CreateSequenceDictionary` musi mieć nazwę taką, jak plik fasta z sekwencją referencyjną, dlatego też po opcji `-O` programu `CreateSequenceDictionary` proszę wpisać nazwę `chr1_fragment.dict`

iv) **wyszukać miejsca różniące w genomach analizowanych osób od sekwencji referencyjnej człowieka** (program `HaplotypeCaller`). Uzyskana lista mutacji powinna być zawarta w pliku o nazwie `razem.vcf`. Ten punkt proszę wykonać w programie screen.

Zadanie6

Pliki w formacie VCF (*ang. variant calling format*) są wykorzystywane do opisu miejsc zmiennych u jednej lub wielu osób. Proszę obejrzeć plik `razem.vcf` w programie `less` (jest to plik tekstowy).
Plik VCF zawiera wiersze z opisem (zaczynające się od znaków `##`), wiersze z danymi o miejscach zmiennych (tabela) oraz linię opisującą poszczególne kolumny tabeli z danymi (linia ta zaczyna się pojedynczym znakiem `#`).
Proszę odpowiedzieć na pytania dotyczące formatu VCF:
W których kolumnach podane jest położenie zmienionego nukleotydu?

nazwa chromosomu:
pozycja:
Które kolumny informują, do jakiej zmiany doszło?
sekwencja/nukleotyd w referencji:
sekwencja/nukleotyd w odczytach:

Zadanie7

Znalezione przez program `HaplotypeCaller` miejsca zmienne nie zawsze mają znaczenie biologiczne (to znaczy, nie zawsze mutacja była obecna u sekwencjonowanej osoby). Część odczytów może zawierać zmienione nukleotydy z powodu:
- błędów powstałych podczas sekwencjonowania (przy namnażaniu matrycy, przy odczytywaniu sygnałów fluorescencyjnych)
- nieprawidłowego zmapowania odczytów.

Do odrzucenia „fałszywych" mutacji najczęściej wykorzystuje się informacje zawarte w następujących kolumnach pliku VCF:
i) QUAL (kol. 6); QUAL to $-10log_{10}p$, tym razem p to prawdopodobieństwo, że w danej pozycji występuje NIEZMIENIONY nukleotyd. Im wyższa wartość QUAL, tym mutacja bardziej prawdopodobna.
ii) INFO (kol. 7) ; zawarta jest tu między innymi informacja o pokryciu miejsca zmiennego przez odczyty
iii) kolumnach dotyczących poszczególnych osób; opis oznaczeń w kol9 FORMAT; dane w kolumnach 10 i dalszych – nazwy tych kolumn odpowiadają nazwom, które podali państwo jako nazwę próbki podczas mapowania plików fastq).

Proszę w liniach opisu pliku VCF (linie poprzedzone dwoma znakami ##) odszukać informacji o znaczeniu pól DP i GT z kolumn INFO/FORMAT.
DP:
AD:
GT:
Jak oznaczone są możliwe genotypy (homozygota REF/REF, heterozygota, homozygota ALT/ALT)?
0/0
1/1
0/1
Odpowiedź mogą państwo znaleźć po porównaniu oznaczeń z wartościami pola AD.

Zadanie8

Państwa zadaniem będzie znalezienie mutacji, która mogła spowodować chorobę u chłopca. Proszę sprawdzić, ile mutacji opisuje plik VCF (na początku takich linii nie ma znaku `#`)
Polecenie:

Liczba linii:

Dlaczego miejsc zmiennych jest tak dużo? Czy każda z opisanych zmian opisuje jakaś szkodliwą mutację? Proszę jeszcze raz przeczytać opis choroby chłopca i opisać, w jaki sposób można spośród wszystkich linii pliku `razem.vcf` wybrać takie, które mogą opisywać szkodliwą zmianę (słowny opis algorytmu).
Czy mutacja powinna występować u rodziców? Dlaczego?

Czy u syna mutacja ta powinna występować w stanie homo, czy heterozygotycznym? Dlaczego?
Jakie inne informacje mogliby państwo wykorzystać (niekoniecznie zawarte w pliku vcf)?

Zadanie9. Automatyczne filtrowanie listy mutacji

Do wybrania tylko interesujących mutacji (to znaczy takich, które mogą być przyczyną choroby chłopca) wykorzystają państwo programu `SnpSift filter` (z pakietu `snpEff`).
Komenda:

```
cat razem.vcf | SnpSift filter 'wzorzec filtra'
```

wyświetli tylko te linie pliku VCF, które pasują do zadanego wzorca.

Na stronie programu (https://pcingola.github.io/SnpEff/ss_filter/) znajdą państwo przykłady, jak wpisać wzorzec. Proszę zaproponować wzorzec, który pozwoli wybrać linie, które mogą opisywać szukaną mutację (filtrowanie po genotypach). Dodatkowo proszę zostawić tylko te zmiany, dla których wartość QUAL przekracza 30.
wzorzec:

Proszę wynik zapisać do pliku (`syn.vcf`).

Zadanie domowe

Proszę dowiedzieć się, jakie efekty ma znaleziona mutacja. Czy występuje wewnątrz genu, czy zmienia aminokwas, czy jej chorobotwórczość była już opisana?
i) Proszę zmienić pozycję zmutowanego nukleotydu w pliku `syn.vcf`, tak aby opisywała rzeczywiste położenie mutacji w chromosomie1. Przy przygotowywaniu pliku z sekwencją referencyjną wybrali państwo fragment chromosomu1 od 91500000 do 94000000. Aby wrócić do prawidłowych koordynat należy do położenia mutacji dodać 91499999. Można to zrobić „ręcznie" - w edytorze nano. Albo wykorzystać skrypt `change-position.sh` (`/dane/change-position.sh`). Skrypt jako pierwszy argument pozycyjny przyjmuje nazwę pliku vcf (w którym ma być poprawiona pozycja), jako drugi argument pozycyjny należy podać liczbę nukleotydów, którą trzeba dodać. Wynikowy plik będzie nazywał się `syn-corect-pos.vcf` (jeśli plik wejściowy to syn.vcf).

Efekty mutacji można zanalizować z wykorzystaniem narzędzia ***Variant Effect Predictor***, dostępnego na stronie głównej ***Ensembl***.

ii) Proszę wybrać program ***Variant Effect Predictor***, po czym wkleić linie z opisem mutacji otrzymanej dla syna w odpowiednim polu (lub podpiąć plik vcf z poprawionymi pozycjami mutacji)
Czy mutacja ta leży wewnątrz genu? Jeśli tak, to proszę podać:

nazwę zmutowanego genu:

jak mutacja wpływa na produkt genu:

czy już wcześniej opisano jej związek z jakąś chorobą:

Do zapamiętania:

1) Kroki analizy danych ngs.
2) Formaty SAM/BAM i VCF.
3) Kiedy warto użyć programu screen
4) Do czego służy program Tablet
5) Jak można przeanalizować efekty mutacji.
6) Dlaczego pliki VCF wymagają filtrowania, jak można je odfiltrować automatycznie.