

MBI - Lab 4 - Raport

Artur Godlewski

Adam Grabski

19.05.21r.

3. Mapowanie

Jaka jest typowa długość odczytów?

Odczyty mają długość około 70 nukleotydów.

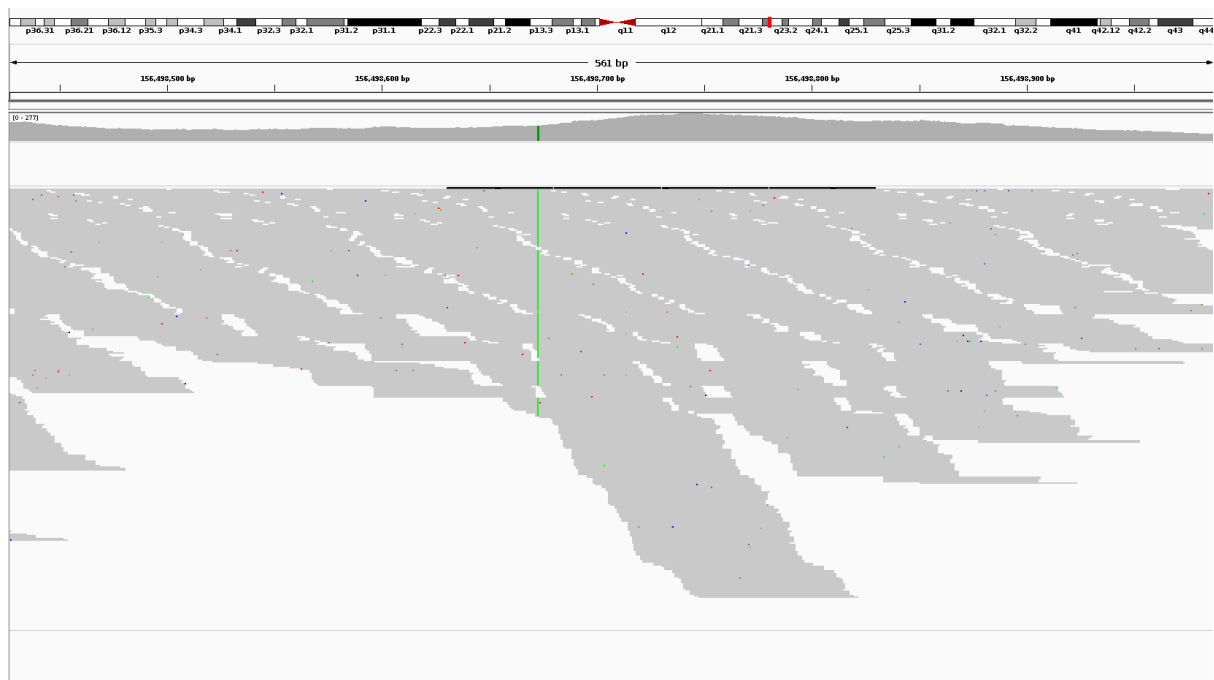
Jak jest różnica w wielkości plików FASTQ, BAM, SAM?

Plik FASTQ ma 57Mb, BAM 14Mb, SAM 75,9Mb.

4. Wizualizacja zawartości pliku BAM w programie IGV

Jaka jest pozycja tego wariantu? Ile odczytów wskazuje na wariant a ile na referencje? Czy jest to wariant homo czy heterozygotyczny? Załącz zrzut ekranu z programu IGV pokazujący wybrany wariant.

Pozycja wariantu o pokryciu powyżej 10x to 156,498,673. Z 148 odczytów 24 odczyty wskazują na wariant, a 124 na referencje. Jest to wariant homozygotyczny.



5. Wykrywanie wariantów

Ile wariantów zawiera plik?

Plik zawiera 6050 wariantów.

Ile wariantów zostało po filtracji? Jakich innych parametrów możemy użyć do dalszej filtracji liczby wariantów?

Po filtracji zostało 241 wariantów. Do dalszej filtracji można użyć:

POS - do ograniczenia wariantów w danym przedziale pozycji

REF - typ nukleotydu w referencji

ALT - typ nukleotydu w wariancie

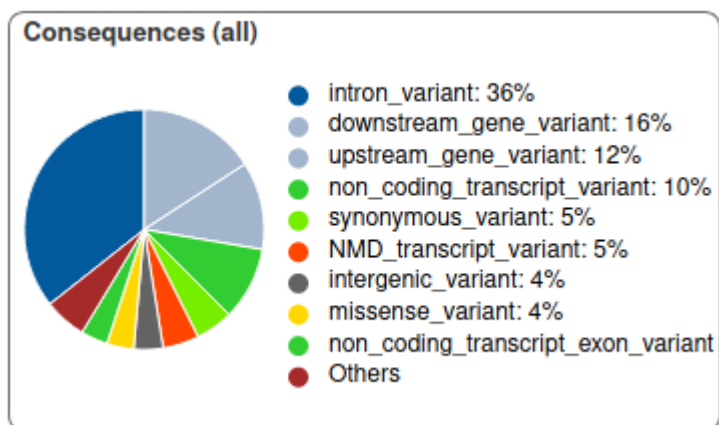
CHROM - ograniczenie do danego chromosomu (nie w naszym przypadku ponieważ mamy tylko jeden chromosom)

QUAL - ograniczenie do wariantów z danego przedziału jakości

VDB - (Variant distance bias) ograniczenie pod względem prawdopodobieństwa przypadkowości zmian w wariantach

6. Adnotacje wariantów

Jaki typ wariantów przeważa?



Przeważają warianty intronów.

Załącz do sprawozdania wiersze odpowiadające temu wariantowi. Czy jest to wariant w części kodującej?

156498673-156498673 A	intron_variant	IQGAP3	ENSG00000183856	Transcript	ENST00000361170.2	protein_coding	-	-	-
156498673-156498673 A	upstream_gene_variant	snoU13	ENSG00000238843	Transcript	ENST00000458777.1	snoRNA	-	-	-
156498673-156498673 A	intron_variant, NMD_transcript_variant	IQGAP3	ENSG00000183856	Transcript	ENST00000491900.1	nonsense_mediated_decay	-	-	-

Wariant ten znajduje się w intronie więc jest w niekodującej części DNA.

Wnioski

Resekwencjonowanie pozwala na znalezienie zmian w genomie co następnie może być użyte do szybkiego stwierdzenia jakie są konsekwencje takiej zmiany za pomocą np. <http://grch37.ensembl.org/>. Unifikacja formatów oraz narzędzia dostępne online pozwalają przeprowadzać takie analizy z minimalnym doświadczeniem nie dysponującymi (do pewnego stopnia) dużą mocą obliczeniową.