

# Zmiany przeciwciał limfocytów B w mukowiscydozie



Hanna Pęciak  
113752

Piotr Hajduk  
121687

# Odpowiedź immunologiczna w mukowiscydozie:

## Skupienie na zakażeniach *Pseudomonas aeruginosa*

- **Mukowiscydoza (CF)** często wiąże się z przewlekłymi zakażeniami *Pseudomonas aeruginosa* (PA), które przyczyniają się do postępującego uszkodzenia płuc.
- Pomimo sprawnego działania układu odpornościowego u osób z CF obecne terapie przeciwciałami skierowanymi przeciwko czynnikom zjadliwości PA, takim jak PcrV, nie przyniosły skutecznych rezultatów klinicznych.
- **Cel:** Zbadanie, czy istnieją istotne różnice w sekwencjach receptorów komórek B (BCR) i mutacjach somatycznych pomiędzy osobami z CF przewlekle zakażonymi PA a tymi bez takiego zakażenia – w celu lepszego zrozumienia odpowiedzi adaptacyjnej i identyfikacji potencjalnych celów terapeutycznych.

# Komórki B i mechanizmy tworzenia przeciwciał swoistych dla *Pseudomonas aeruginosa*

- **Komórki B** odgrywają kluczową rolę w odpowiedzi adaptacyjnej, wytwarzając **przeciwciała**, które wiążą antygeny bakterii, takie jak białko PcrV – czynnik zjadliwości *Pseudomonas aeruginosa*.
- Przeciwciała są produkowane przez komórki B po aktywacji i dojrzewaniu, a ich swoistość zależy od unikalnych sekwencji w regionach zmiennych receptorów komórkowych – tzw. **B cell receptors (BCR)**.
- Regiony te powstają poprzez **rekombinację segmentów genów V (Variable), D (Diversity) i J (Joining)** w trakcie rozwoju komórki B w szpiku kostnym.
- Kombinacja różnych segmentów V, D i J, wraz z mutacjami somatycznymi, prowadzi do ogromnej różnorodności przeciwciał – umożliwiając organizmowi rozpoznanie szerokiego wachlarza patogenów.
- W kontekście CF i zakażeń *Pseudomonas*, analiza regionów V(D)J pozwala ocenić, **jak skutecznie organizm tworzy przeciwciała specyficzne dla PA**, oraz czy przewlekła infekcja wpływa na sposób doboru i mutacji tych regionów.

# Regiony V(D)J

Locus ciężkiego łańcucha immunoglobulin (IGH@) na **chromosomie 14** zawiera segmenty genowe dla ciężkiego łańcucha immunoglobulin.

Locus łańcucha lekkiego typu kappa ( $\kappa$ ) immunoglobulin (IGK@) na **chromosomie 2** zawiera segmenty genowe dla jednego typu ( $\kappa$ ) lekkiego łańcucha immunoglobulin.

Locus łańcucha lekkiego typu lambda ( $\lambda$ ) immunoglobulin (IGL@) na **chromosomie 22** zawiera segmenty genowe dla innego typu ( $\lambda$ ) lekkiego łańcucha immunoglobulin.

# Dane wejściowe

Dawca	ID	Wiek	Na ETI?*	Nasilenie choroby płuc	Historia zakażenia PA
1	—	33	Tak	Umiarkowane	Przewlekłe
2	SRR32572151	19	Tak	Minimalne	Ujemna hodowla w czasie pobrania krwi. Dodatnia 6 mies. wcześniej. Wcześniej 5 lat ujemnych hodowli.
3	SRR32572150	18	Tak	Minimalne	Przewlekłe

\*ETI to skrót od triple terapii składającej się z elexacaftor, tezacaftor i ivacaftor, stosowanej w leczeniu mukowiscydozy

# Trimming: parametry

**PE -phred33** – tryb parowany, jakość kodowana w formacie Phred+33

**ILLUMINACLIP:....:2:30:10** – usuwanie adapterów (max 2 niedopasowania, próg 30/10)

**LEADING:10** – usuwa niskiej jakości bazy (<10) z początku odczytu

**TRAILING:10** – usuwa bazy <10 z końca odczytu

**SLIDINGWINDOW:5:30** – tnie, gdy średnia jakość w 5-bazowym oknie spada poniżej 30

**MINLEN:70** – odrzuca odczyty krótsze niż 70 bp

# Trimming: wyniki

Próbka	Input Read Pairs	Both Surviving (%)	Forward Only Surviving (%)	Reverse Only Surviving (%)	Dropped (%)
SRR32572150	58906	12996 (22.06%)	38818 (65.90%)	846 (1.44%)	6246 (10.60%)
SRR32572151	116294	4783 (4.11%)	102274 (87.94%)	98 (0.08%)	9139 (7.86%)

## **Input Read Pairs:**

Całkowita liczba par odczytów (forward i reverse) przekazanych do Trimmomatic.

## **Both Surviving (%):**

Procent par, w których oba odczyty przeszły przycinanie i zostały zachowane jako pary.

## **Forward Only Surviving (%):**

Tylko odczyt forward przeszedł przycinanie; odczyt reverse został odrzucony. Wynik jako odczyt pojedynczy.

## **Reverse Only Surviving (%):**

Tylko odczyt reverse przeszedł przycinanie; odczyt forward został odrzucony. Wynik jako odczyt pojedynczy.

## **Dropped (%):**

Oba odczyty zostały odrzucone – nie spełniły kryteriów jakości i długości.

# Podsumowanie QC (góra przed trimmingiem, dół po trimmingu)

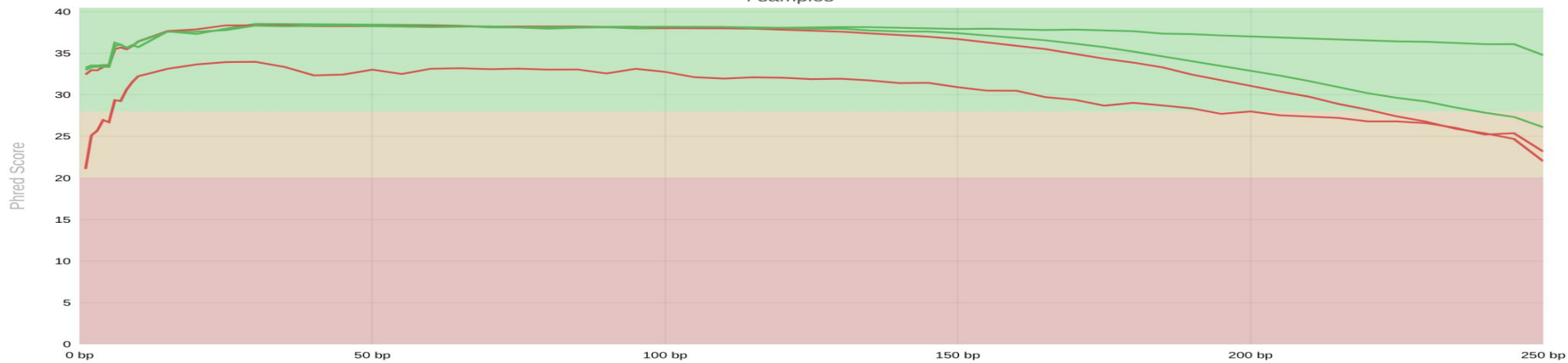
Sample Name	Dups	GC	Seqs
SRR32572150_1	88.8 %	52.0 %	0.1 M
SRR32572150_2	69.4 %	54.0 %	0.1 M
SRR32572151_1	92.4 %	57.0 %	0.1 M
SRR32572151_2	67.8 %	55.0 %	0.1 M

Sample Name	Dups	GC	Median len	Seqs
SRR32572150_forward_paired	85.0 %	58.0 %	251 bp	0.0 M
SRR32572150_forward_unpaired	90.0 %	57.0 %	157 bp	0.0 M
SRR32572150_reverse_paired	65.3 %	58.0 %	251 bp	0.0 M
SRR32572150_reverse_unpaired	46.6 %	57.0 %	251 bp	0.0 M
SRR32572151_forward_paired	75.2 %	57.0 %	251 bp	0.0 M
SRR32572151_forward_unpaired	93.1 %	57.0 %	251 bp	0.1 M
SRR32572151_reverse_paired	68.1 %	57.0 %	172 bp	0.0 M
SRR32572151_reverse_unpaired	83.7 %	59.0 %	112 bp	0.0 M



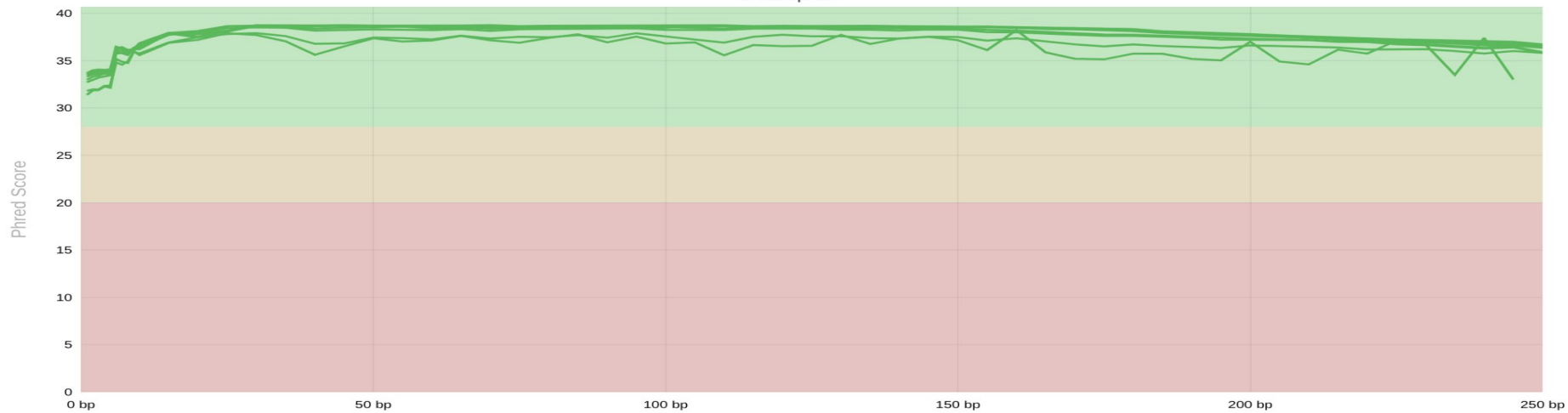
## FastQC: Mean Quality Scores

4 samples

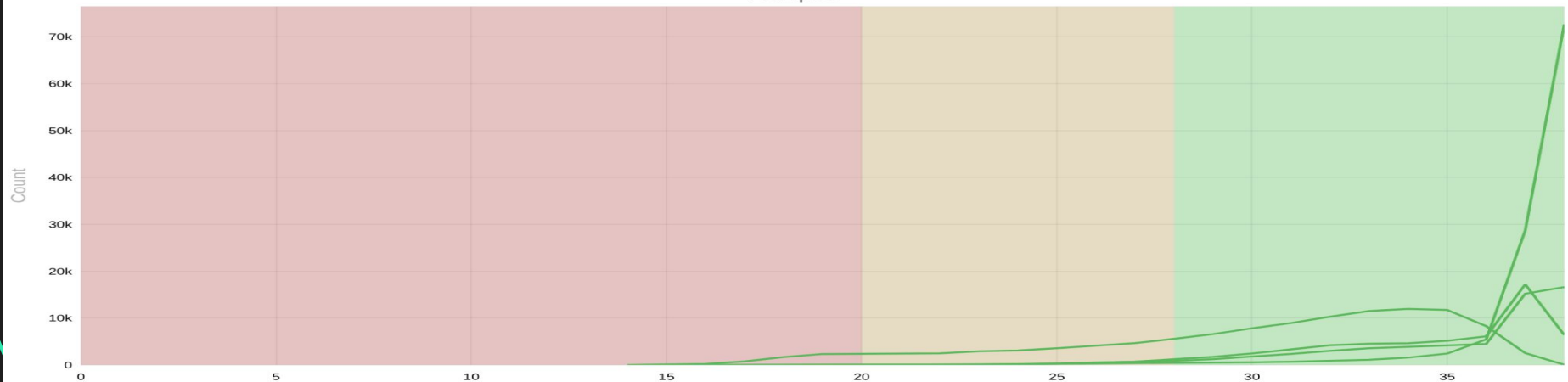


## FastQC: Mean Quality Scores

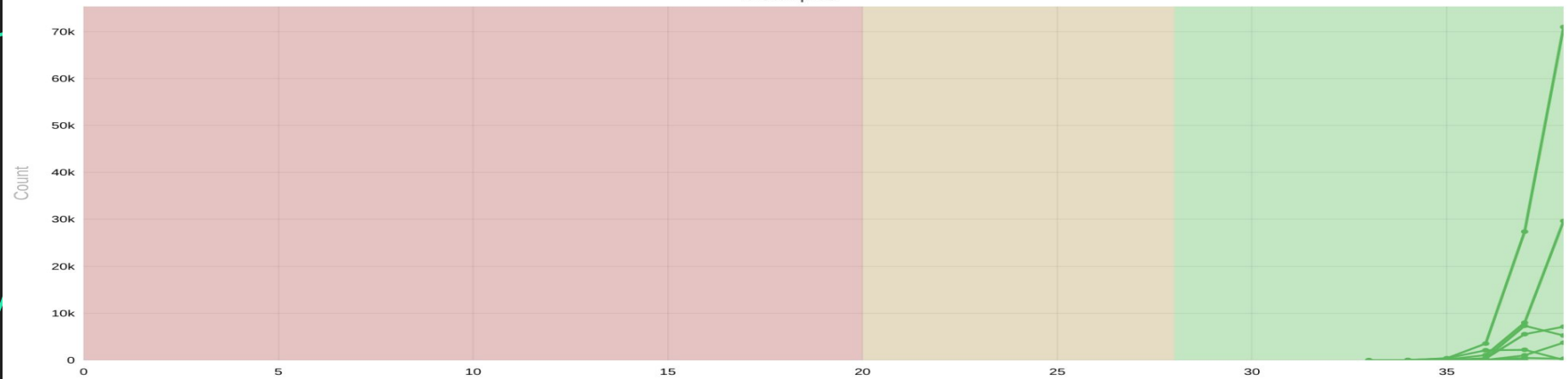
8 samples



FastQC: Per Sequence Quality Scores  
4 samples

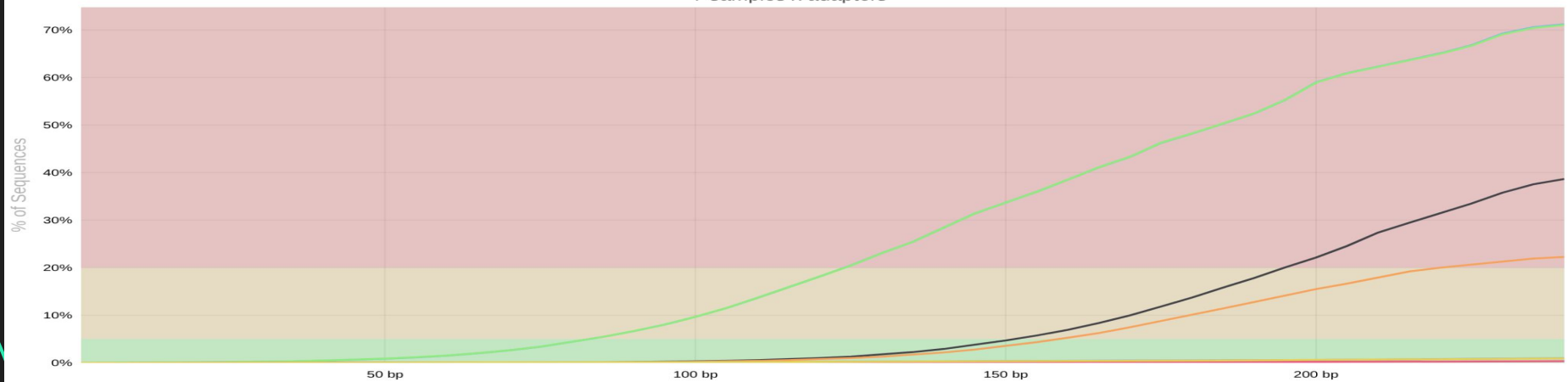


FastQC: Per Sequence Quality Scores  
8 samples



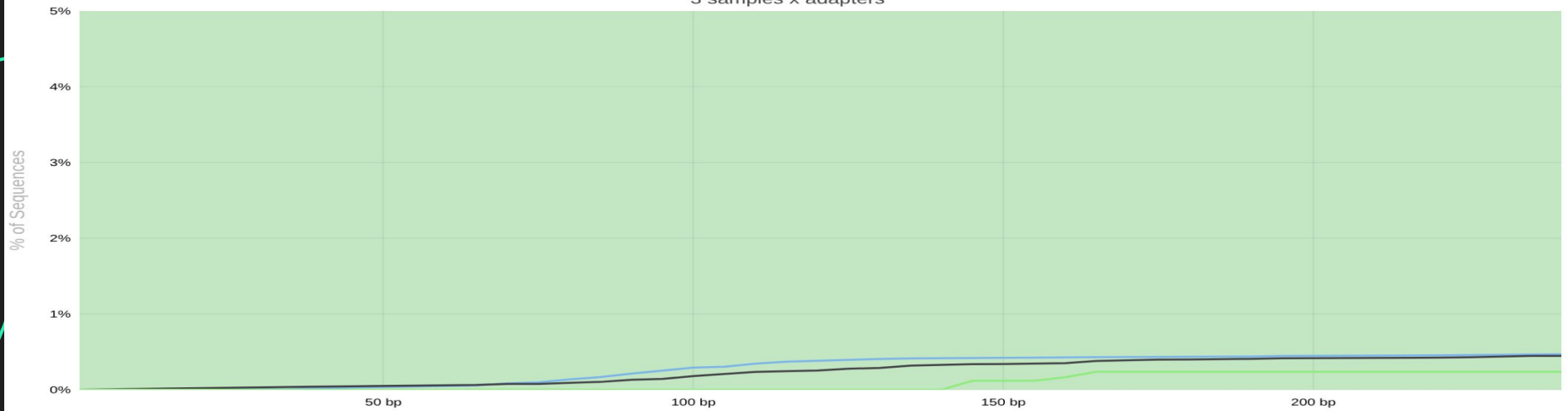
## FastQC: Adapter Content

7 samples x adapters



## FastQC: Adapter Content

3 samples x adapters



# Alignment | Post-alignment | BQSR

## Alignement *bwa mem*

Przypisanie odczytów do sekwencji referencyjnej i dodanie RG (read groups)

@SRR32572147.1 **M02172**:273:000000000-DJPRD:1:1101:15616:1365 length=251

## Post-alignment *samtools*

- sort - sortowanie po nazwie odczytu
- fixmate - dodanie informacji o parach odczytów
- sort - sortowanie po pozycji w genomie
- markdup - usunięcie duplikatów
- index - indeksowanie
- flagstat - ocena jakości dopasowania
- depth - głębokość dopasowania

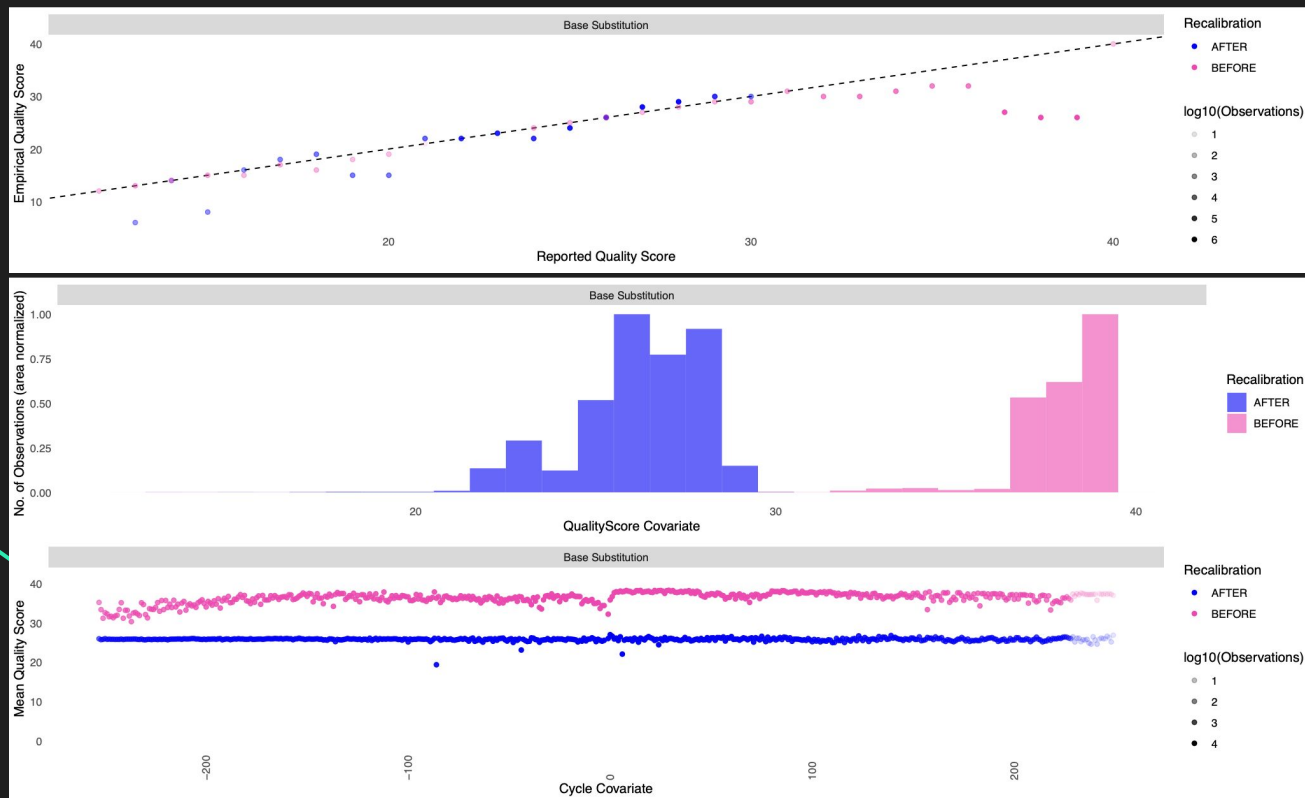
## BGQR *applyBGQSR*

Korekta błędów w ocenie jakości wynikających z konkretnego sekwenatora i z wykorzystaniem danych o znanych wariantach (VCF)

+ indeksowanie

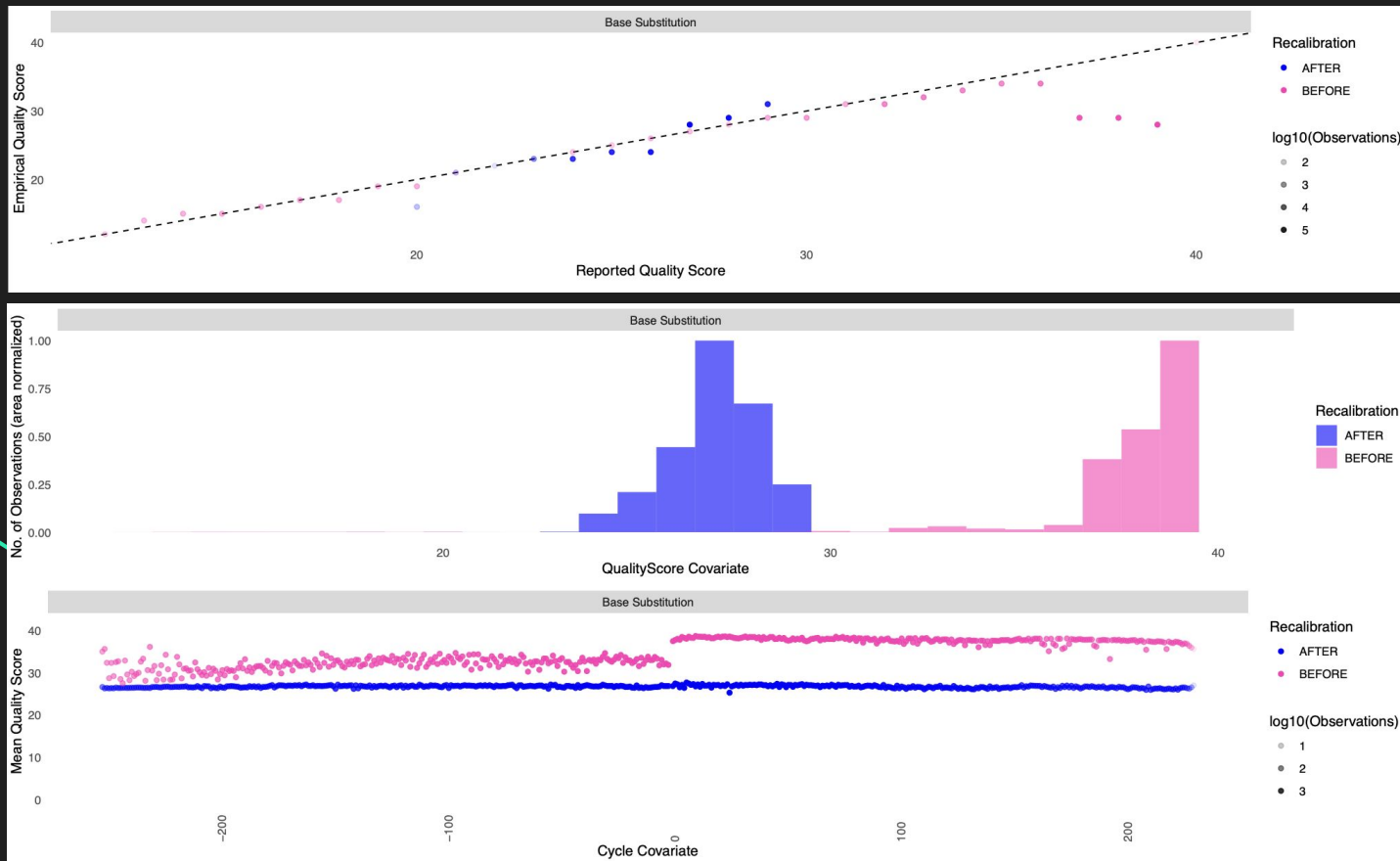
# Base Quality Score Recalibration

SRR32572150



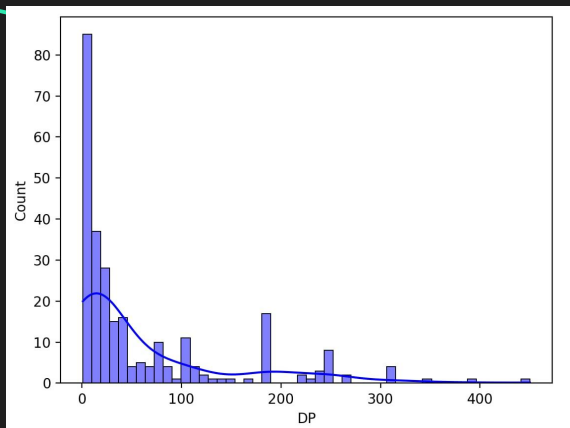
# Base Quality Score Recalibration

SRR32572151



# Zidentyfikowane warianty

Sample	Tool	Variants identified	Varsians filered	Variants after filtration
SRR32572150	BCF	182	- 57% (78)	104
	GATK	208	- 65% (71)	137
SRR32572151	BCF	96	- 42% (55)	41
	GATK	115	- 46% (61)	54



Narzędzia użyte do identyfikacji wariantów:

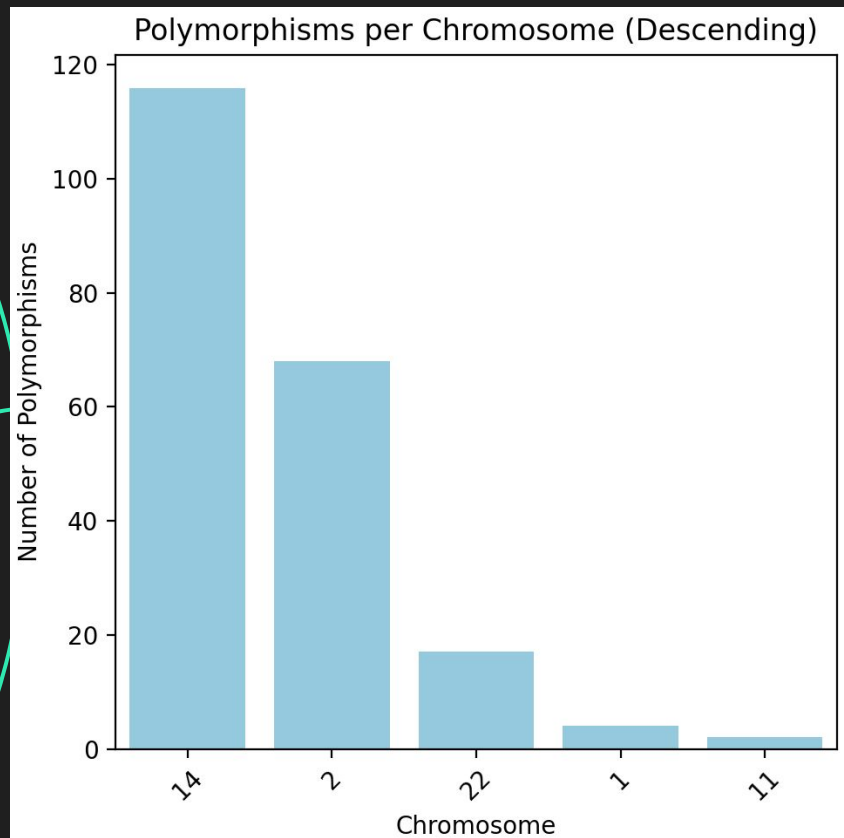
- **GATK HaplotypeCaller**
- **BCF mpileup**

Zastosowane parametry filtrowania:

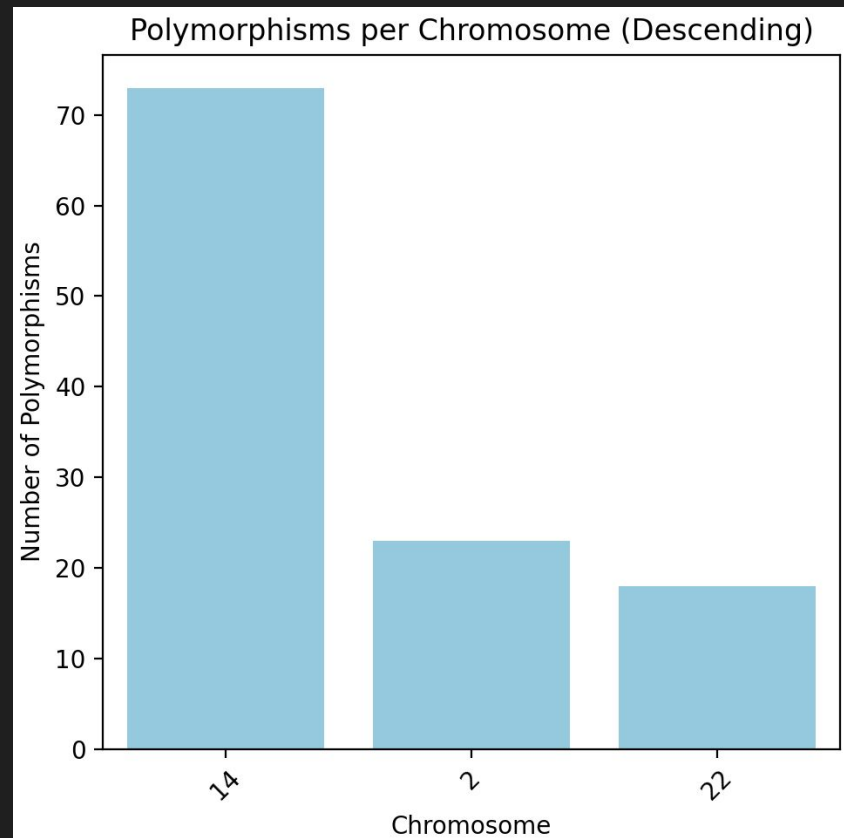
- Jakość wariantu  $QUAL > 30$
- Pokrycie  $10 < DP < 500$
- Allele alternatywne  $< 2$

# Liczba wariantów na chromosom

SRR32572150



SRR32572151





# Variant Effect Predictor – BCF

SRR32572150

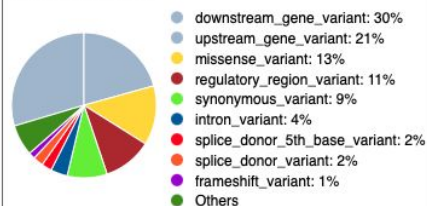
## Variant Effect Predictor results ?

[Job details](#) ⓘ

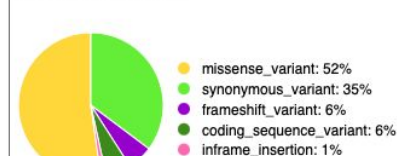
[Summary statistics](#) ⓘ

Category	Count
Variants processed	104
Variants filtered out	0
Novel / existing variants	22 (21.2) / 82 (78.8)
Overlapped genes	61
Overlapped transcripts	75
Overlapped regulatory features	10

### Consequences (all)



### Coding consequences



SRR32572151

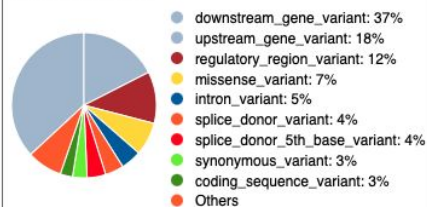
## Variant Effect Predictor results ?

[Job details](#) ⓘ

[Summary statistics](#) ⓘ

Category	Count
Variants processed	41
Variants filtered out	0
Novel / existing variants	10 (24.4) / 31 (75.6)
Overlapped genes	61
Overlapped transcripts	64
Overlapped regulatory features	13

### Consequences (all)



### Coding consequences



# Variant Effect Predictor – GATK

SRR32572150

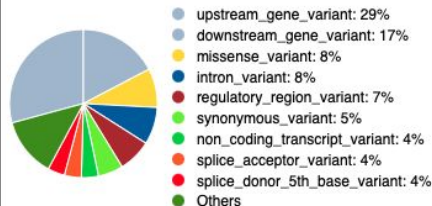
## Variant Effect Predictor results ?

[Job details](#) +

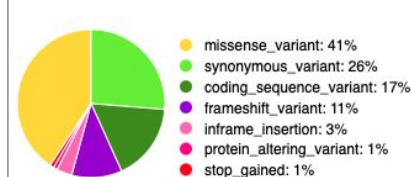
[Summary statistics](#) -

Category	Count
Variants processed	137
Variants filtered out	0
Novel / existing variants	56 (40.9) / 81 (59.1)
Overlapped genes	84
Overlapped transcripts	137
Overlapped regulatory features	15

### Consequences (all)



### Coding consequences



SRR32572151

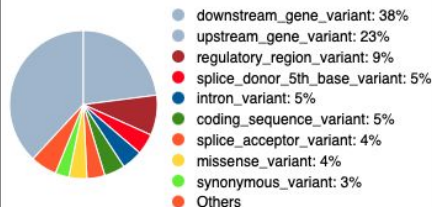
## Variant Effect Predictor results ?

[Job details](#) +

[Summary statistics](#) -

Category	Count
Variants processed	54
Variants filtered out	0
Novel / existing variants	20 (37.0) / 34 (63.0)
Overlapped genes	72
Overlapped transcripts	77
Overlapped regulatory features	11

### Consequences (all)



### Coding consequences



- **SRR32572150:**

- Większa liczba wariantów ogółem.
- Wysoki udział **wariantów missense** oraz **frameshift** – mogą silnie wpływać na funkcję białek.
- Obecność **wariantów stop-gained** i **protein-altering** – potencjalnie szkodliwe mutacje.

- **SRR32572151:**

- Mniej wariantów, ale większy udział w **regionach regulatorowych**.
- W kodujących regionach dominują **warianty typu coding\_sequence** i **missense**.
- Obecne **warianty start lost** – mogą zakłócać inicjację translacji białka

**Wniosek:**

SRR32572150 wykazuje silniejszy wpływ na regiony kodujące, natomiast SRR32572151 może oddziaływać na regulację ekspresji genów.

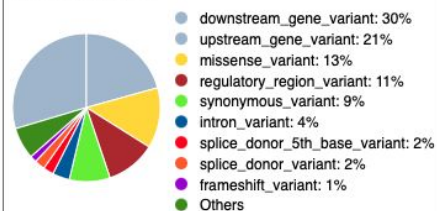
BCF

## Variant Effect Predictor results ⓘ

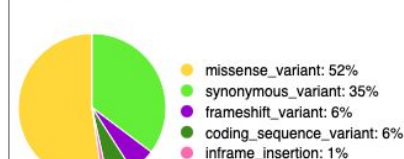
[Job details](#) ⓘ[Summary statistics](#) ⓘ

Category	Count
Variants processed	104
Variants filtered out	0
Novel / existing variants	22 (21.2) / 82 (78.8)
Overlapped genes	61
Overlapped transcripts	75
Overlapped regulatory features	10

## Consequences (all)



## Coding consequences



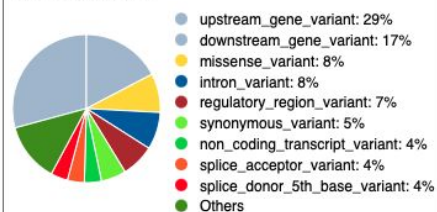
GATK

## Variant Effect Predictor results ⓘ

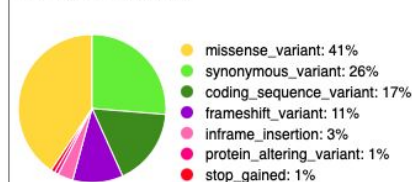
[Job details](#) ⓘ[Summary statistics](#) ⓘ

Category	Count
Variants processed	137
Variants filtered out	0
Novel / existing variants	56 (40.9) / 81 (59.1)
Overlapped genes	84
Overlapped transcripts	137
Overlapped regulatory features	15

## Consequences (all)



## Coding consequences



# Porównanie wyników BCF vs GATK (VEP analiza)

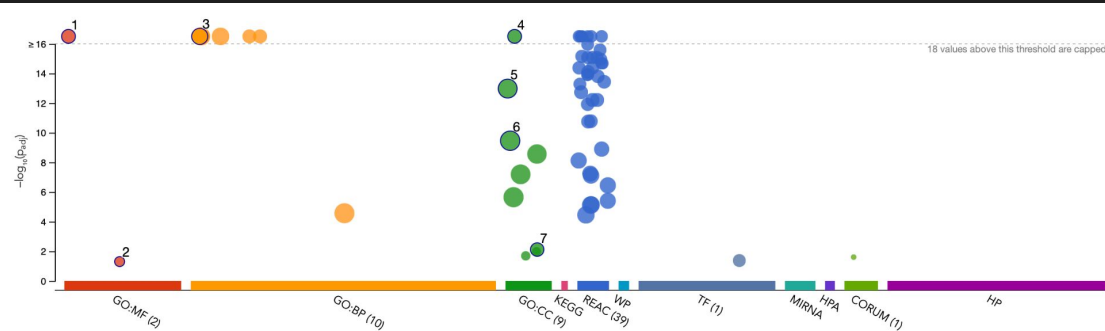
- **Liczba wykrytych wariantów:**  
GATK zidentyfikował więcej wariantów (137 vs. 104) i wyższy udział nowych wariantów (41% vs. 21%).
- **Pokrycie genomowe:**  
GATK lepiej pokrywa geny, transkrypty i elementy regulatorowe.
- **Konsekwencje funkcjonalne:**  
BCF – więcej wariantów *missense* i *synonymous* (łagodniejsze).  
GATK – większa różnorodność, więcej wariantów o silniejszym wpływie (*frameshift*, *stop gained*).
- **Wniosek:**  
**GATK oferuje większą czułość i głębszą anotację**, podczas gdy **BCF** daje bardziej zachowawcze wyniki.

# Analiza funkcjonalna – GProfiler

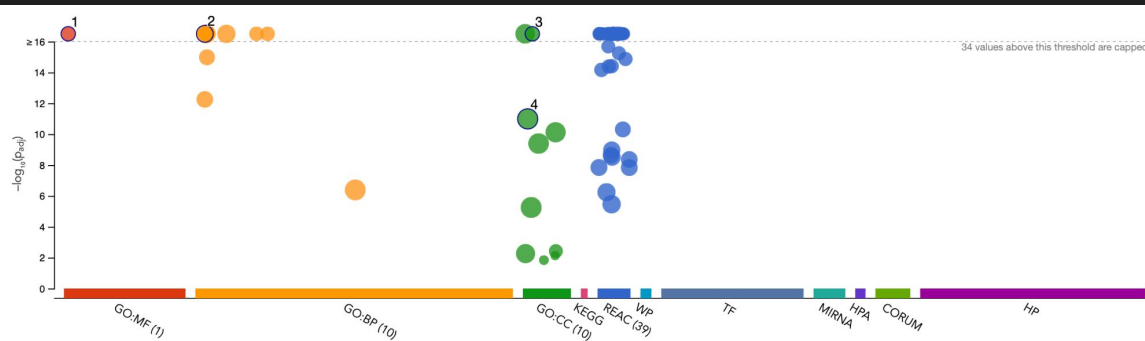
BCF

SRR32572150

SRR32572151



ID	Source	Term ID	Term Name	P <sub>adj</sub> (query_1)
1	GO:MF	GO:0003823	antigen binding	$8.020 \times 10^{-41}$
2	GO:MF	GO:0034987	immunoglobulin receptor binding	$4.998 \times 10^{-2}$
3	GO:BP	GO:0002250	adaptive immune response	$2.952 \times 10^{-34}$
4	GO:CC	GO:0019814	immunoglobulin complex	$4.454 \times 10^{-55}$
5	GO:CC	GO:0005576	extracellular region	$1.053 \times 10^{-13}$
6	GO:CC	GO:0005886	plasma membrane	$3.494 \times 10^{-10}$
7	GO:CC	GO:0072562	blood microparticle	$7.697 \times 10^{-3}$



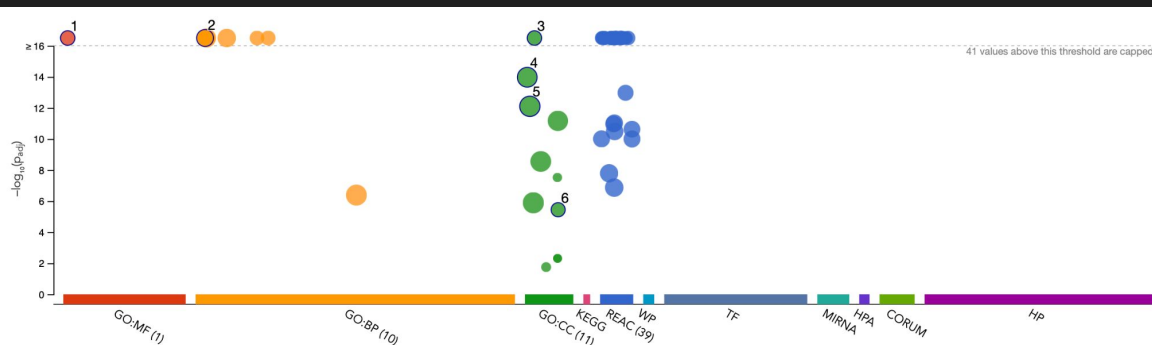
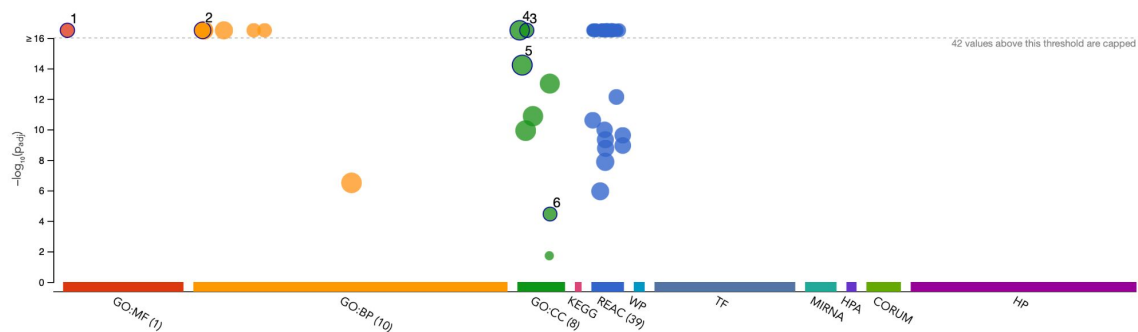
ID	Source	Term ID	Term Name	P <sub>adj</sub> (query_1)
1	GO:MF	GO:0003823	antigen binding	$8.222 \times 10^{-36}$
2	GO:BP	GO:0002250	adaptive immune response	$2.051 \times 10^{-35}$
3	GO:CC	GO:0019814	immunoglobulin complex	$2.454 \times 10^{-54}$
4	GO:CC	GO:0005886	plasma membrane	$1.025 \times 10^{-11}$

# Analiza funkcjonalna – GProfiler

GATK

SRR32572150

SRR32572151



# Interpretacja wyników

W obu próbkach (SRR32572150 i SRR32572151) najistotniejsze zmiany były w kategoriach:

- **antigen binding / adaptive immune response** (GO:MF:0003823, GO:BP:0002250)  
Powiązane z rozpoznawaniem antygenów i inicjację odpowiedzi immunologicznej, np. w genach kodujących przeciwciała. Takie polimorfizmy mogą wpływać na zdolność do rozpoznawania patogenów i efektywność odpowiedzi odpornościowej
- **immunoglobulin complex** (GO:CC:0019814)  
Zmiany składu, stabilności lub funkcji kompleksów immunoglobulinowych, mogące zaburzać zdolność wiązania patogenów oraz zmiany w poziomie aktywności układu odpornościowego.
- **extracellular region / plasma membrane / blood microparticle**  
Te zmiany mogą powodować zaburzenie komunikacji między komórkami odpornościowymi i przekazywania sygnałów zapalnych, a nawet zwiększoną przepuszczalnością nabłonka oddechowego (co sprzyja kolonizacji przez bakterie).

Metryka (g:Profiler)	SRR32572150	SRR32572151	Interpretacja
Liczba znaczących terminów GO:CC	43	17	Większa liczba struktur komórkowych związanych z odpornością w próbce <b>SRR32572150</b> .
Najsilniejsze p_adj (GO:MF)	$3.3 \times 10^{-48}$	$5.7 \times 10^{-40}$	Różnica 8 rzędów wielkości, nadal ekstremalnie istotne statystycznie.





**Dziękujemy za uwagę**

**Hanna Pęciak**  
**113752**

**Piotr Hajduk**  
**121687**