

# Bazy danych stosowane w analizie wyników badań molekularnych

## Część praktyczna

Mateusz Dawidziuk  
Zakład Genetyki Medycznej  
Instytut Matki i Dziecka w Warszawie

# Identyfikatory sekwencji w NCBI

- Zadanie - znajdź identyfikatory sekwencji dla ludzkiego genu **AGPS**
  - Podaj identyfikator Accession version oraz GI dla tego genu
  - Podaj identyfikator Accession version oraz GI dla najnowszej wersji jego transkryptu
  - Podaj identyfikator Accession version oraz GI dla najstarszej wersji jego transkryptu

| Accession version | GI       |
|-------------------|----------|
| NC_000001.8       | 51511461 |
| NC_000001.7       | 42406218 |
| NC_000001.6       | 42405892 |
| NC_000001.5       | 37623929 |
| NC_000001.4       | 29824572 |
| NC_000001.3       | 29824110 |
| NC_000001.2       | 27777714 |
| NC_000001.1       | 22539468 |

<https://www.ncbi.nlm.nih.gov/gene/>

# Identyfikatory sekwencji w NCBI

- Odpowiedzi - identyfikatory sekwencji dla ludzkiego genu **AGPS**
  - Podaj identyfikator Accession version oraz GI dla tego genu  
**NG\_008968.1 - 8540**
  - Podaj identyfikator Accession version oraz GI dla najnowszej wersji jego transkryptu  
**NM\_003659.4 - 1519315989**
  - Podaj identyfikator Accession version oraz GI dla najstarszej wersji jego transkryptu  
**NM\_003659.1 - 4501992**

| Accession version | GI       |
|-------------------|----------|
| NC_000001.8       | 51511461 |
| NC_000001.7       | 42406218 |
| NC_000001.6       | 42405892 |
| NC_000001.5       | 37623929 |
| NC_000001.4       | 29824572 |
| NC_000001.3       | 29824110 |
| NC_000001.2       | 27777714 |
| NC_000001.1       | 22539468 |

# Informacje z bazy NCBI

- Zadanie

- Podaj numery Accession version dla wszystkich transkryptów ludzkiego genu **ZEB2**
- Podaj numery Accession version odpowiadających tym transkryptom produktom białkowym
- Podaj numery Accession version dla niekodujących produktów tego genu
- Na jakim chromosomie zlokalizowany jest ten gen? Podaj numer Accession version odpowiadający kontigowi tego chromosomu
- W jakim prążku cytogenetycznym zlokalizowany jest ten gen?
- Podaj liczbę eksonów tego genu
- W jakiej tkance ulega on największej ekspresji?

# Informacje z bazy NCBI

- Odpowiedzi
  - Podaj numery Accession version dla wszystkich transkryptów ludzkiego genu **ZEB2** oraz odpowiadających tym transkryptom produktom białkowym  
**NM\_001171653.2 → NP\_001165124.1**  
**NM\_014795.4 → NP\_055610.1**
  - Podaj numery Accession version dla niekodujących produktów tego genu  
**NR\_033258.2**
  - Na jakim chromosomie zlokalizowany jest ten gen? Podaj numer Accession version kontigowi tego chromosomu  
**Chromosom 2 - NC\_000002.12 (GRCh38)**
  - W jakim prążku cytogenetycznym zlokalizowany jest ten gen?  
**2q22.3**
  - Podaj liczbę eksonów tego genu - **11**
  - W jakiej tkance ulega on największej ekspresji?  
**Mózg**

# Wyszukiwanie w bazach NCBI - PubMed

- Stwórz możliwie najdokładniejsze zapytanie w bazie PubMed szukając artykułów odnośnie badań klinicznych szczepionek mRNA opublikowanych od roku 2000
  - Podaj liczbę takich artykułów w bazie PubMed
  - Ile było opublikowanych artykułów przed 2020 rokiem? Jakich chorób dotyczyły?
- Wyszukaj publikacje autora o nazwisku Shearer w których tytule albo w abstrakcie znajdziemy słowo głuchota (deafness)
  - Ile było takich artykułów?
  - W którym roku ukazało się najwięcej publikacji
  - Czy wszystkie należą do jednego autora

# OMIM

## Online Mendelian Inheritance in Man

- Zadania:
  - Jaki gen znajduje się pod identyfikatorem OMIM #113705? Co możemy powiedzieć o tym genie i chorobach przez niego wywoływanych?
  - Podobnie jak powyżej, ale dla genu *ALB*
  - Podaj liczbę zespołów związanych z głuchotą, która powodowana jest przez patogenne warianty w genach na chromosomie X (tylko geny ze znanym molekularnym podłożem choroby)

[www.omim.org](http://www.omim.org)

# dbSNP

- Sprawdź jaki wariant kryje się pod numerem identyfikacyjnym rs773571137
  - Jakie substytucje znajdują się pod tym identyfikatorem?
  - Jakie są ich konsekwencje na poziomie DNA a jakie na poziomie białkowym?
  - W jakiej populacji występuje ten wariant?
- Znajdź ile jest wariantów powodujących delecje w ludzkim genie *LPL* o częstości (MAF) pomiędzy 0.001-0.01

<https://www.ncbi.nlm.nih.gov/snp>

# ClinVar

- Co możemy powiedzieć o statusie patogenności wariantu o numerze identyfikacyjnym VCV000006940.45?
- Ile jest wariantów patogennych w ludzkim genie *FOXP1*, które powodują przesunięcie ramki odczytu i zostały zaraportowane przez kilka ośrodków

<https://www.ncbi.nlm.nih.gov/clinvar>

# dbNSFP

**database of Non-Synonymous SNPs and their Functional Predictions**

- Sprawdź wyniki algorytmów predykcyjnych dla wariantu o współrzędnych 14:28767838 A>G (zaznacz opcje od 8 do 31)

<http://database.liulab.science/dbNSFP>

# Identyfikatory sekwencji w Ensembl

- Zadanie - znajdź identyfikatory sekwencji dla ludzkiego genu **AGPS**
  - Podaj identyfikator ensembl dla tego genu
  - Podaj identyfikator ensembl dla trzech najdłuższych transkryptów tego genu
  - Który z tych transkryptów ma swój odpowiednik w bazie RefSeq?
  - Podaj identyfikator ensembl najkrótszego transkryptu, który koduje białko. Jaki poziom wsparcia TSL ma ten transkrypt?

# Identyfikatory sekwencji w Ensembl

- Odpowiedzi - znajdź identyfikatory sekwencji dla ludzkiego genu **AGPS**
  - Podaj identyfikator ensembl dla tego genu  
**ENSG00000018510**
  - Podaj identyfikator ensembl dla trzech najdłuższych transkryptów tego genu  
**ENST00000642466.2**  
**ENST00000264167.11**  
**ENST00000679459.1**
  - Który z tych transkryptów ma swój odpowiednik w bazie RefSeq?  
**ENST00000264167.11**
  - Podaj identyfikator ensembl najkrótszego transkryptu, który koduje białko. Jaki poziom wsparcia TSL ma ten transkrypt?  
**ENST00000409888.1**

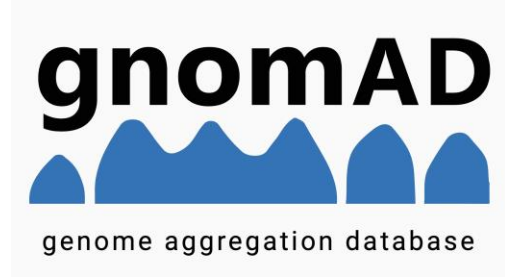
# Ensembl i UCSC

- W bazie Ensembl wyszukaj gen *SLC25A1* (wybierz transkrypt kanoniczny)
  - Dostosuj widok w taki sposób, aby widać było 250 nukleotydów w głąb sekwencji intronowej, warianty były widoczne zarówno w eksonach jak i intronach, oraz ukryj najrzadsze warianty z MAF poniżej 0.01%
- Wyszukaj koordynaty 4 eksonu tego genu i użyj ich do podejrzenia sekwencji w przeglądarce UCSC
  - Włącz informacje o wariantach z bazy ClinVar i HGMD – ile jest zaraportowanych wariantów w tym eksonie w poszczególnych bazach?

<https://genome-euro.ucsc.edu>

# gnomAD

## Genome Aggregation Database



- Sprawdź pokrycie genu *FOXP1* w bazie gnomAD v2.1.1
  - Jakie są różnice w pokryciu pomiędzy sekwencjami pochodzącymi z WES (kolor niebieski) a WGS (kolor zielony)?
  - Przyglądając się wykresowi wariantów raportowanych w ClinVar, w którym miejscu w tym genie znajduje się mutacyjny „hot-spot”
  - Który wariant występuje w największej liczbie w postaci homozygotycznej? Podaj jego identyfikator w bazie dbSNP i ClinVar
- Co możemy powiedzieć o jakości wariantu **14-29236656-C-A**

<https://gnomad.broadinstitute.org>

# COSMIC i Mastermind

- Wyszukaj zmianę c.4598G>T w genie *BRCA1* w bazie COSMIC
  - W których tkankach wykrywano ten wariant i jakie może powodować nowotwory?
  - Ile publikacji dotyczących tego wariantu znajduje się na stronie COSMIC?
  - Znajdź publikacje opisujące ten wariant w bazie Mastermind, podaj liczbę znalezionych artykułów
  - W jakim roku opublikowano pierwszy artykuł opisujący ten wariant?

<https://cancer.sanger.ac.uk/cosmic>

<https://mastermind.genomenon.com>



MASTERMIND  
By: GENOMENON

# VarSome

- Wyszukaj wariant NM\_007327:c.2417C>A dla genomu referencyjnego hg38
  - W jakim genie zlokalizowany jest ten wariant? Jaki ma identyfikator w bazie dbSNP?
  - Jak zmieni się pozycja nukleotydowa tego wariantu gdy wybierzemy wariant kanoniczny RefSeq?
  - Jakie ma predykcje wg. ACMG? Jak zmieni się ta predykcja, jeśli ustaliliśmy że wariant nie został odziedziczony od rodziców i powstał spontanicznie (*de novo*)?
  - Czy był raportowany w bazie ClinVar? Jaką ma klasyfikację?
  - Czy był raportowany w bazie populacyjnej gnomAD?
  - Jak wyglądają wyniki algorytmów predykcyjnych dla tego wariantu?

<https://varsome.com/>



The human genetics search engine  
Supported by the global community of geneticists

# BioMart

- Zadanie – sprawdź w jakich genach lokalizują się podane poniżej warianty (chromosom : start : koniec : nić):

16:2084316:2084316:1

16:2496492:2496492:1

16:29813694:29813694:1

2:166251876:166251876:-1

8:60856167:60856167:1

<https://www.ensembl.org/biomart/martview>

# **Dziękuję za uwagę**

dr n. med. Mateusz Dawidzук  
Zakład Genetyki Medycznej  
kontakt: [mateusz.dawidziuk@imid.med.pl](mailto:mateusz.dawidziuk@imid.med.pl)