

Architektura bazy danych GMP

MACIEJ ANTCZAK, MARCIN BOROWSKI I JAROSŁAW SYNAK

Części składowe bazy

- Dane z Ensembla
- Anotacje (dane o mutacjach)
- Dane z plików VCF
- Dane z FenGen

Role w bazie danych

- Owner – dostęp do wszystkich relacji w bazie
- Anonymous – dostęp do wszystkich danych poza relacjami zawierającymi identyfikatory próbek (GMP_XXXXX) oraz dane osobowe

Liczba próbek (stan na 2023-05-31)

- Dane z FenGen – 1000
- Dane o anotacjach – 2000

Na obecną chwilę są to zupełnie rozłączne zbiory próbek!

Liczba anotacji (stan na 2023-05-31)

- Liczba anotacji – 67,8 mln
- Liczba plików źródłowych – 7999

Pochodzenie (stan na 2023-05-31)

- Znane – 1739 próbek

- Nieznane – 261 próbek

MAŁOPOLSKIE	153
PODLASKIE	206
PODKARPACKIE	34
POMORSKIE	302
MAZOWIECKIE	210
ŚWIĘTOKRZYSKIE	218
OPOLSKIE	33
LUBELSKIE	110
ŁÓDZKIE	61
LUBUSKIE	83
ZACHODNIOPOMORSKIE	61
WIELKOPOLSKIE	63
KUJAWSKO-POMORSKIE	155
WARMIŃSKO-MAZURSKIE	50

Kopie zapasowe

- Przyrostowa (codzienna), obecnie wyłączona
- Pełna (cotygodniowa), jej aktualny rozmiar to 38,07GiB

Komunikacja – REST API

The screenshot shows the Postman interface with a GET request to `192.168.105.103:3000/annot/transcript?limit=10`. The request is configured with a query parameter `limit=10`. The response is a JSON array of 5 objects, each containing `annot_transcript_id` and `stable_id`.

KEY	VALUE	DESCRIPTION
<input type="checkbox"/> and	(gene_id.gt.5,gene_id.eq.6)	
<input type="checkbox"/> select	value	
<input type="checkbox"/> order	gene_id.dec	
<input checked="" type="checkbox"/> limit	10	

```
1 {
2   "annot_transcript_id": 1,
3   "stable_id": "ENST00000335137"
4 },
5 {
6   "annot_transcript_id": 2,
7   "stable_id": "ENST00000466557"
8 },
9 {
10  "annot_transcript_id": 3,
11  "stable_id": "ENST00000491962"
12 },
13 {
14  "annot_transcript_id": 4,
15  "stable_id": "ENST00000623983"
16 },
17 {
18  "annot_transcript_id": 5,
19  "stable_id": "ENST00000623834"
20 },
21 },
22 {
23  "annot_transcript_id": 6,
24  "stable_id": "ENST00000624431"
25 }
```

Przykładowe zapytanie w programie Postman o dostępne transkrypty. API jest dostępne tylko za pośrednictwem VPN lub wewnątrz sieci.

Technologie

