

Wprowadzenie do analizy danych RNA-Seq

Michał Szczęśniak

Wprowadzenie, źródła i formaty danych

Genomy, transkryptomy, przeglądarki
genomowe,
źródła danych NGS

Technologie i platformy NGS

illumina®



SOLiD®



ion torrent
⬇ * ⬆ ○ × □ + ≈

ThermoFisher
SCIENTIFIC



Sekwencjonowanie
genomów

Sekwencjonowanie
transkryptomów
(RNA-Seq)

Interakcje
białko – RNA
(CLIP-seq)

Interakcje
białko – DNA
(ChIP-seq)

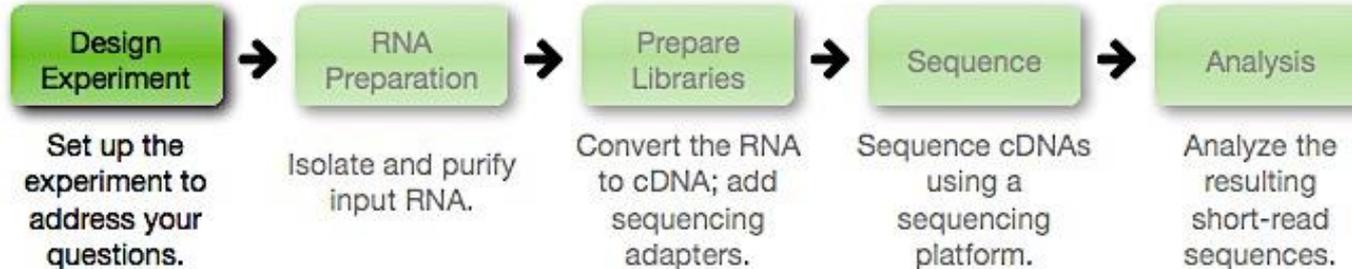


Badanie metylacji DNA

Metagenomika

Sekwencjonowanie transkryptomów RNA-Seq

- ✓ Kodujący mRNA
- ✓ Niekodujący RNA → miRNA, małe RNA, długie niekodujące RNA
- ✓ Cały transkryptom



<http://allseq.com/>

<http://rnaseq.uoregon.edu>

Jakich danych potrzebujemy?

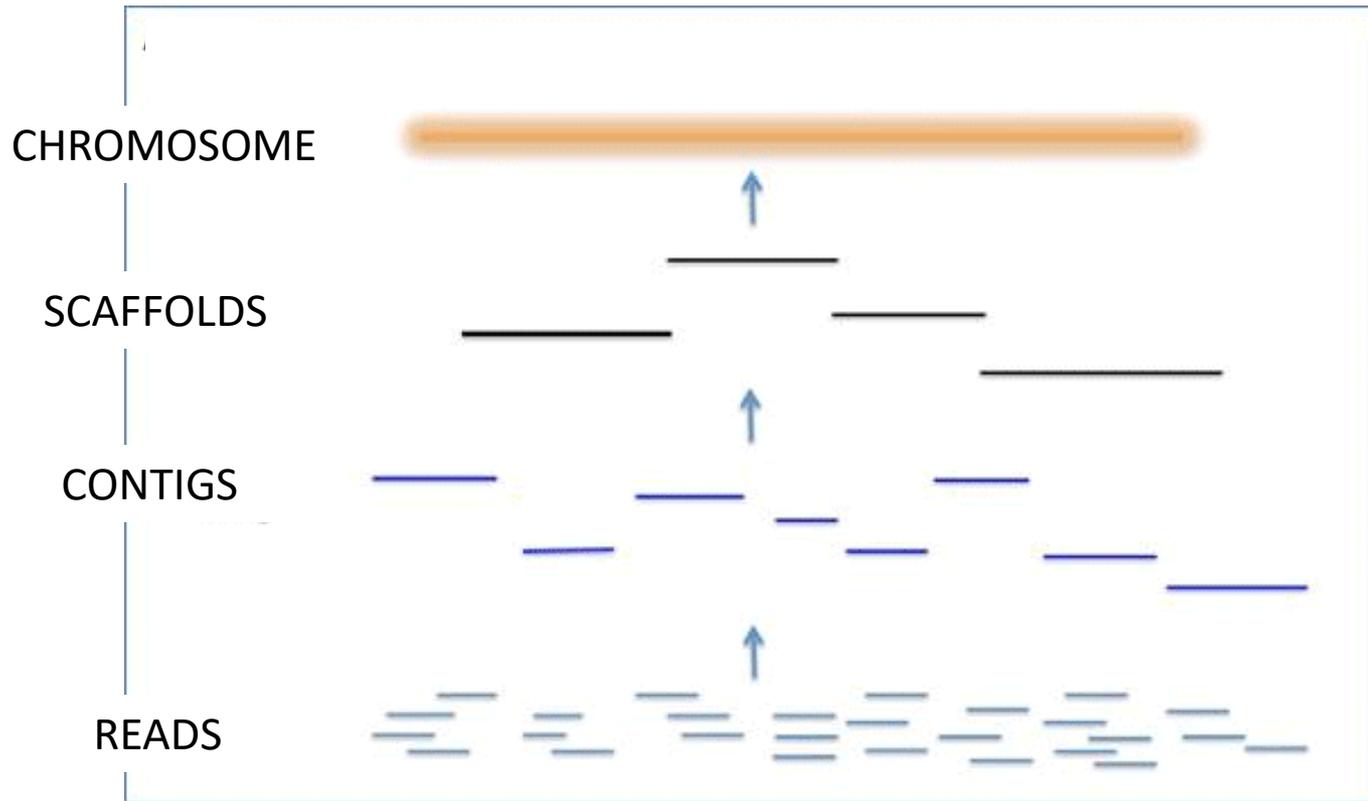
- ✓ Sekwencje genomowe
- ✓ Sekwencje transkryptomów
- ✓ Adnotacje
- ✓ Odczyty z sekwencjonowania RNA-seq

Adnotacje genomów

- W przypadku genomów, które nie są w pełni zsekwencjonowane lub złożone (ang. *assembled*), dysponujemy sekwencjami **kontigów** (ang. *contigs*) i **skafoldów** (ang. *scaffolds*).
- Dodatkowo, w przypadku dobrze zaadnotowanych genomów (np. człowiek, mysz) do **chromosomów** dodawane są **łatki** (ang. *patches*), zawierające sekwencje, których nie włączono do chromosomów.
- Niekiedy dysponujemy nie tylko genomem **referencyjnym** ale również **alternatywnymi** złożeniami genomów (inne grupy badawcze, konsorcja).

Składanie genomu - genome assembly

ASSEMBLING = SKŁADANIE



<http://www.ddbj.nig.ac.jp/sub/assembly-e.html>

PRZYDATNE DEFINICJE: <https://www.ncbi.nlm.nih.gov/grc/help/definitions>

Wybrane formaty danych

1. FASTA → sekwencje
2. FASTQ → odczyty NGS
3. BAM, SAM → zmapowania NGS
4. BED → koordynaty
5. GTF, GFF3 → adnotacje
6. Plik tekstowy tabularny (TSV) → np. wyniki działania programu BLAST, wyniki analiz ekspresji
7. PSL → wynik działania programu BLAT

Wybrane formaty danych

FASTA

Plik FASTA składa się z **nagłówek** (zaczyna się od znaku '>') po którym następuje **sekwencja**. Może zawierać wiele sekwencji (MULTIFASTA). Rozszerzenie **.fasta** albo **.fa**

```
>gi|453055269|gb|KC207570.1| Homo sapiens transient receptor potential cation channel  
subfamily C member 3 variant c (TRPC3) mRNA  
GGATATAGAAATGGGAATGGGTAAGTCAAAGTCCAGGCAGATAATGAAAAGACTTATAAAGCGGTATGTTTTGAAAGCACAAGTAGACAA  
AGAAAATGATGAAGTTAATGAAGGTGAATAAAAGAAATCAAGCAAGATATCTCCAGCCTTCGTTATGAACTTTTGAAGACAAGAGCCAA  
GCAACTGAGGAATTAGCCATTCTAATTCATAAACTTAGTGAGAAACTGAATCCCAG
```

```
>XP_007907608.1 PREDICTED: keratin, type I cytoskeletal 19-like [Callorhinchus milii]  
MSRSVYSANIGGSVIVSSNQRRSFASTSSSLFGSGAPSRRAISVYNIGANRGKRISAGGSWNASYASLG  
GDAGILCNDEKQTMQNLNARLSSYMEKVRSLKSNRQLEFQIHEFYEKKAPVSTKDLTVYEGNISDCHLQ  
IYAI SLHNAKLMLQIDNARLAADDFRIKYESELAIRKGVEADIQGLRKVMDLSLTKRGLESQVTALKED  
LVYLHRSHKEELSLRTGMGGSVTVDL DSTPATDLNKILSNLRIEYETIAEKNRDKDVEAWYLEKCHTLNQ
```

Wybrane formaty danych

FASTA

Kod IUPAC

IUPAC nucleotide code	Base
A	Adenine
C	Cytosine
G	Guanine
T (or U)	Thymine (or Uracil)
R	A or G
Y	C or T
S	G or C
W	A or T
K	G or T
M	A or C
B	C or G or T
D	A or G or T
H	A or C or T
V	A or C or G
N	any base

Wybrane formaty danych

BED

Plik z rozszerzeniem **.bed** zawiera koordynaty czyli informację o lokalizacji jakiegoś rejonu (genu, egzonu, intronu) w genomie.

Wymagane pola to chromosom, start i koniec rejonu.

Może zawierać więcej kolumn.

```
chr7 127474697 127475864
chr7 127475864 127477031
chr7 127477031 127478198
chr7 127478198 127479365
chr7 127479365 127480532
chr7 127480532 127481699
```

↓
Chromosom START END

Wybrane formaty danych

GTF, GFF3

Pliki z rozszerzeniem **.gtf .gff3** zawierają informacje o elementach w genomie i ich adnotacjach. Danemu elementowi przypisana jest jedna linia w pliku składająca się z dziewięciu kolumn.

GTF - General Transfer Format

```
#!genome-build GRCh38.p7
#!genome-version GRCh38
#!genome-date 2013-12
#!genome-build-accession NCBI:GCA_000001405.22
#!genebuild-last-updated 2016-06
```

1 2 3 4 5 6 7 8 9

```
1 havana gene 11869 14409 . + . gene_id "ENSG00000223972"; gene_version "5"; gene_name "DDX11L1"; gene_source "havana"; gene_biotype "transcribed_u
1 havana transcript 11869 14409 . + . gene_id "ENSG00000223972"; gene_version "5"; transcript_id "ENST00000456328"; transcript_version "2"; gene_name
1 havana exon 11869 12227 . + . gene_id "ENSG00000223972"; gene_version "5"; transcript_id "ENST00000456328"; transcript_version "2"; exon_number "
1 havana exon 12613 12721 . + . gene_id "ENSG00000223972"; gene_version "5"; transcript_id "ENST00000456328"; transcript_version "2"; exon_number "
1 havana exon 13221 14409 . + . gene_id "ENSG00000223972"; gene_version "5"; transcript_id "ENST00000456328"; transcript_version "2"; exon_number "
1 havana transcript 12010 13670 . + . gene_id "ENSG00000223972"; gene_version "5"; transcript_id "ENST00000450305"; transcript_version "2"; gene_name
1 havana exon 12010 12057 . + . gene_id "ENSG00000223972"; gene_version "5"; transcript_id "ENST00000450305"; transcript_version "2"; exon_number "
1 havana exon 12179 12227 . + . gene_id "ENSG00000223972"; gene_version "5"; transcript_id "ENST00000450305"; transcript_version "2"; exon_number "
1 havana exon 12613 12697 . + . gene_id "ENSG00000223972"; gene_version "5"; transcript_id "ENST00000450305"; transcript_version "2"; exon_number "
1 havana exon 12975 13052 . + . gene_id "ENSG00000223972"; gene_version "5"; transcript_id "ENST00000450305"; transcript_version "2"; exon_number "
1 havana exon 13221 13374 . + . gene_id "ENSG00000223972"; gene_version "5"; transcript_id "ENST00000450305"; transcript_version "2"; exon_number "
1 havana exon 13453 13670 . + . gene_id "ENSG00000223972"; gene_version "5"; transcript_id "ENST00000450305"; transcript_version "2"; exon_number "
```

GTF - General Transfer Format

1 2 3 4 5 6 7 8 9

```
1 havana gene 11869 14409 . + . gene_id "ENSG00000223972"; gene_version "5"; gene_name "DDX11L1"; gene_source "havana"; gene_biotype "transcribed_u
1 havana transcript 11869 14409 . + . gene_id "ENSG00000223972"; gene_version "5"; transcript_id "ENST00000456328"; transcript_version "2"; gene_name
1 havana exon 11869 12227 . + . gene_id "ENSG00000223972"; gene_version "5"; transcript_id "ENST00000456328"; transcript_version "2"; exon_number "
1 havana exon 12613 12721 . + . gene_id "ENSG00000223972"; gene_version "5"; transcript_id "ENST00000456328"; transcript_version "2"; exon_number "
1 havana exon 13221 14409 . + . gene_id "ENSG00000223972"; gene_version "5"; transcript_id "ENST00000456328"; transcript_version "2"; exon_number "
1 havana transcript 12010 13670 . + . gene_id "ENSG00000223972"; gene_version "5"; transcript_id "ENST00000450305"; transcript_version "2"; gene_name
1 havana exon 12010 12057 . + . gene_id "ENSG00000223972"; gene_version "5"; transcript_id "ENST00000450305"; transcript_version "2"; exon_number "
1 havana exon 12179 12227 . + . gene_id "ENSG00000223972"; gene_version "5"; transcript_id "ENST00000450305"; transcript_version "2"; exon_number "
1 havana exon 12613 12697 . + . gene_id "ENSG00000223972"; gene_version "5"; transcript_id "ENST00000450305"; transcript_version "2"; exon_number "
1 havana exon 12975 13052 . + . gene_id "ENSG00000223972"; gene_version "5"; transcript_id "ENST00000450305"; transcript_version "2"; exon_number "
1 havana exon 13221 13374 . + . gene_id "ENSG00000223972"; gene_version "5"; transcript_id "ENST00000450305"; transcript_version "2"; exon_number "
1 havana exon 13453 13670 . + . gene_id "ENSG00000223972"; gene_version "5"; transcript_id "ENST00000450305"; transcript_version "2"; exon_number "
```

- 1. seqname** - name of the chromosome or scaffold; chromosome names can be given with or without the 'chr' prefix (the convention in Ensembl is to omit the 'chr' prefix).
- 2. source** - name of the program that generated this feature, the data source (database or project name) or gene status
- 3. feature** - feature type name, e.g. Gene, Variation, Similarity
- 4. start** - start position of the feature, with sequence numbering starting at 1.
- 5. end** - end position of the feature, with sequence numbering starting at 1.
- 6. score** - a floating point value.
- 7. strand** - defined as + (forward) or - (reverse).
- 8. frame** - one of '0', '1' or '2'. '0' indicates that the first base of the feature is the first base of a codon, '1' that the second base is the first base of a codon, and so on..
- 9. attribute** - a semicolon-separated list of tag-value pairs, providing additional information about each feature.

Przegląd podstawowych źródeł danych

dotyczących genomów i ich adnotacji

Trzy metabazy i przeglądarki



National Center for Biotechnology Information



Ensembl Genome Browser

European Bioinformatics Institute, Wellcome Trust Sanger Institute



University of California, Santa Cruz Genome Browser

RefSeq - sekwencje – identyfikatory

NM_..., **X**M_... → mRNA

NP_..., **X**P_... → białka

NR_..., **X**R_... → niekodujące RNA

NC_..., **N**G_..., **N**T_..., **N**W_..., **A**C_... → kontigi, sekwencje genomowe

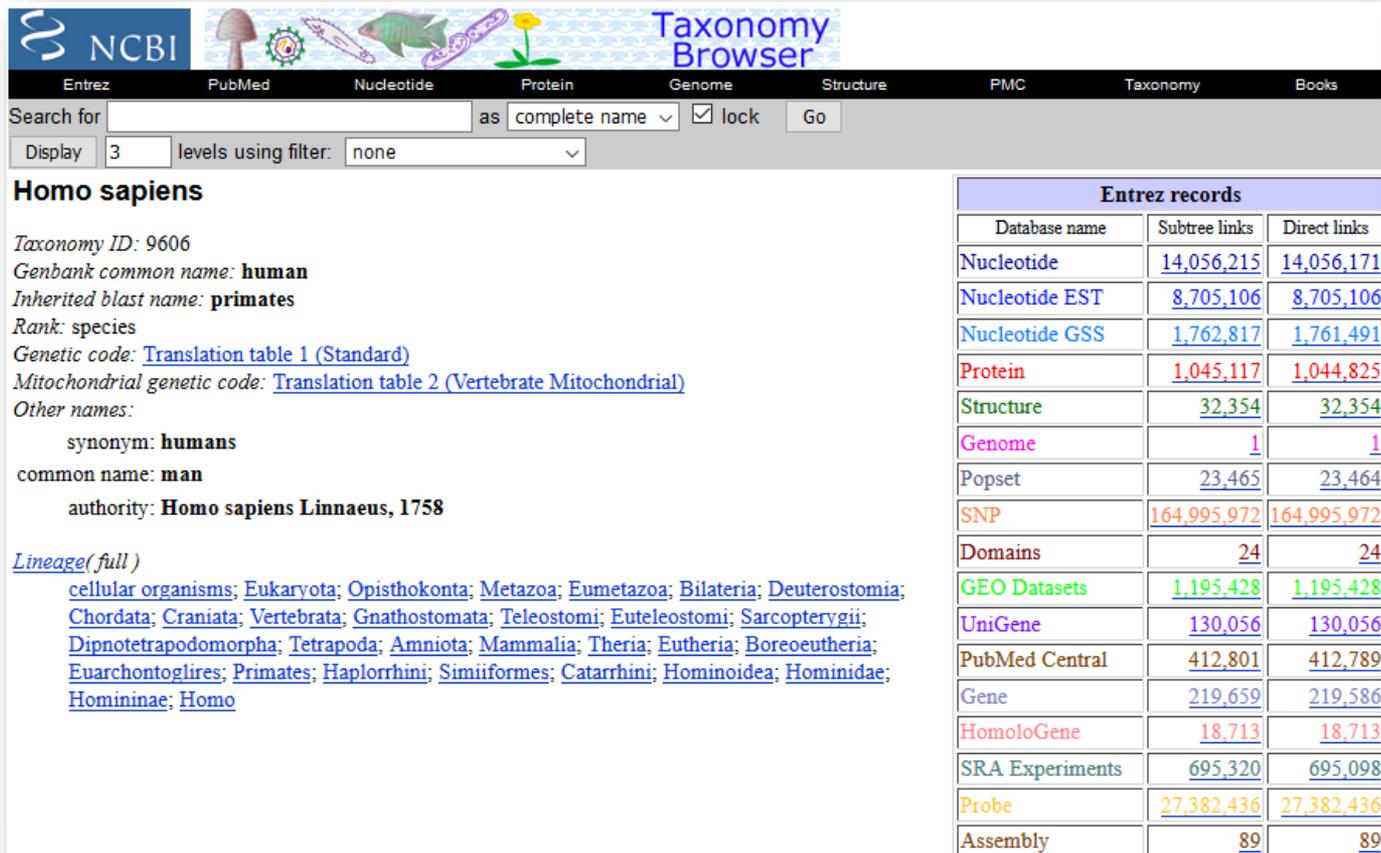
N oznacza sekwencje uzyskane z wyników eksperymentów

X oznacza, że cząsteczka została przewidziana bioinformatycznie np. przez zmapowanie białka do genomu spokrewnionego organizmu.

RefSeq Database – zbiór sprawdzonych, niepowtarzalnych sekwencji, często weryfikowanych manualnie.

NCBI (Taxonomy)

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



The screenshot displays the NCBI Taxonomy Browser interface. At the top, there is a navigation bar with icons for Entrez, PubMed, Nucleotide, Protein, Genome, Structure, PMC, Taxonomy, and Books. Below this is a search bar with the text "Search for" and a dropdown menu set to "as complete name". There is also a "lock" checkbox and a "Go" button. Below the search bar, there is a "Display" field set to "3" and a "levels using filter" dropdown set to "none".

The main content area is titled "Homo sapiens". It provides the following information:

- Taxonomy ID:** 9606
- Genbank common name:** human
- Inherited blast name:** primates
- Rank:** species
- Genetic code:** [Translation table 1 \(Standard\)](#)
- Mitochondrial genetic code:** [Translation table 2 \(Vertebrate Mitochondrial\)](#)
- Other names:**
 - synonym: humans
 - common name: man
 - authority: **Homo sapiens Linnaeus, 1758**

Below this information is a section titled "Lineage(full)" with a list of taxonomic ranks: [cellular organisms](#); [Eukaryota](#); [Opisthokonta](#); [Metazoa](#); [Eumetazoa](#); [Bilateria](#); [Deuterostomia](#); [Chordata](#); [Craniata](#); [Vertebrata](#); [Gnathostomata](#); [Teleostomi](#); [Euteleostomi](#); [Sarcopterygii](#); [Dipnotetrapodomorpha](#); [Tetrapoda](#); [Amniota](#); [Mammalia](#); [Theria](#); [Eutheria](#); [Boreoeutheria](#); [Euarchontoglires](#); [Primates](#); [Haplorrhini](#); [Simiiformes](#); [Catarrhini](#); [Hominoidea](#); [Hominidae](#); [Homininae](#); [Homo](#)

On the right side of the page, there is a table titled "Entrez records" with three columns: "Database name", "Subtree links", and "Direct links".

Database name	Subtree links	Direct links
Nucleotide	14,056,215	14,056,171
Nucleotide EST	8,705,106	8,705,106
Nucleotide GSS	1,762,817	1,761,491
Protein	1,045,117	1,044,825
Structure	32,354	32,354
Genome	1	1
Popset	23,465	23,464
SNP	164,995,972	164,995,972
Domains	24	24
GEO Datasets	1,195,428	1,195,428
UniGene	130,056	130,056
PubMed Central	412,801	412,789
Gene	219,659	219,586
HomoloGene	18,713	18,713
SRA Experiments	695,320	695,098
Probe	27,382,436	27,382,436
Assembly	89	89

NCBI (Downloads FTP)

NCBI Home	<h2>All Resources</h2> <p>All Databases Downloads Submissions Tools How To</p> <h3>Downloads</h3> <p>BLAST (Stand-alone) BLAST executables for local use are provided for Solaris, LINUX, Windows, and MacOS information. Pre-formatted databases for BLAST nucleotide, protein, and translated sea subdirectory.</p> <p>FTP: BLAST Databases Sequence databases for use with the stand-alone BLAST programs. The files in this dir BLAST.</p> <p>FTP: CDD This site provides full data records for CDD, along with individual Position Specific Score data for each conserved domain. See the README file for full details.</p> <p>FTP: ClinVar Data This site provides full data extractions in XML and summary data in VCF format. It cont in ClinVar, MedGen, and GTR.</p> <p>FTP: FASTA BLAST Databases Sequence databases in FASTA format for use with the stand-alone BLAST programs. T can be used with BLAST.</p> <p>FTP: GenBank This site contains files for all sequence records in GenBank in the default flat file format contents are described in the README.genbank file.</p> <p>FTP: GenPept The protein sequences corresponding to the translations of coding sequences (CDS) in the README file in the directory for more information.</p>
Resource List (A-Z)	
All Resources	
Chemicals & Bioassays	
Data & Software	
DNA & RNA	
Domains & Structures	
Genes & Expression	
Genetics & Medicine	
Genomes & Maps	
Homology	
Literature	
Proteins	
Sequence Analysis	
Taxonomy	
Training & Tutorials	
Variation	

NCBI (Downloads FTP)

Proszę poszukać sekwencję genomową chromosomu 22 u człowieka i pobrać ją spakowaną w formacie FASTA (.fa) dla najnowszej wersji genomu referencyjnego.

Trzy metabazy i przeglądarki



National Center for Biotechnology Information



Ensembl Genome Browser

European Bioinformatics Institute, Wellcome Trust Sanger Institute



University of California, Santa Cruz Genome Browser

Ensembl

<http://www.ensembl.org/>
<http://ensemblgenomes.org/>

*e!*EnsemblGenomes

*e!*Ensembl

*e!*EnsemblFungi

*e!*EnsemblMetazoa

*e!*EnsemblProtists

*e!*EnsemblPlants

*e!*EnsemblBacteria

*Pre!*Ensembl

Ensembl - identyfikatory

Show/hide columns (1 hidden)		Filter						
Name	Transcript ID	bp	Protein	Biotype	CCDS	UniProt	RefSeq	Flags
CFTR-001	ENST00000003084.10	6132	1480aa	Protein coding	CCDS5773	A0A024R730 P13569	NM_000492 NP_000483	TSL:1 GENCODE basic APPRIS P1
CFTR-005	ENST00000426809.5	4316	1438aa	Protein coding	-	E7EPB6	-	CDS 3' incomplete TSL:5
CFTR-002	ENST00000468795.1	682	190aa	Protein coding	-	H0Y8A9	-	CDS 5' incomplete TSL:5
CFTR-004	ENST00000446805.1	575	36aa	Protein coding	-	C9J6L5	-	CDS 3' incomplete TSL:4
CFTR-008	ENST00000600166.1	559	156aa	Protein coding	-	M0QYZ3	-	CDS 5' incomplete TSL:5
CFTR-009	ENST00000608965.5	896	No protein	Processed transcript	-	-	-	TSL:5
CFTR-010	ENST00000610149.1	519	No protein	Processed transcript	-	-	-	TSL:5
CFTR-007	ENST00000429014.1	423	No protein	Processed transcript	-	-	-	TSL:5
CFTR-003	ENST00000546407.1	222	No protein	Processed transcript	-	-	-	TSL:1
CFTR-006	ENST00000472848.1	148	No protein	Processed transcript	-	-	-	TSL:5
CFTR-011	ENST00000621535.1	657	No protein	Retained intron	-	-	-	TSL:5

ENSG → Gene

ENST → Transcript (.ver)

ENSP → Protein (.ver)

ENSE → Exon

} *Homo sapiens*

Protein_coding

Processed_transcript

Processed_pseudogene

miRNA

rRNA, scRNA

snoRNA, snRNA ...

ENSXETG00000003158

Xenopus tropicalis

ENSMUST00000044620.1

Mus musculus

ENSDARG00000079015

Danio rerio

ENSECAT00000016282.5

Equus caballus

<http://www.ensembl.org/Help/Faq?id=468>

http://vega.sanger.ac.uk/info/about/gene_and_transcript_types.html

Ensembl – pobieranie danych

1. Pojedyncze rekordy w bazie danych
(Uwaga! Na różnym poziomie – różne dane!)
2. Download → FTP
3. BioMart

Ensembl: ćwiczenie

W przeglądarce Ensembl proszę znaleźć **gen *BRCA1* u człowieka**.

1. Ile ma form splicingowych?
2. Na którym chromosomie i na jakiej nici się znajduje?

Proszę wybrać jeden transkrypt tego genu, a następnie:

1. Znaleźć kodowane przez nie białko
2. Obejrzeć strukturę intronowo-egzonową
3. Znaleźć sekwencję pierwszego egzonu
4. Pobrać jego sekwencję mRNA
5. Pobrać sekwencję genomową wraz z 100 nt rejonami otaczającymi

Ensembl – pobieranie danych (FTP)

Downloads



Download a sequence or region

Export data

```
CAGAATGAT
AAATGTTCT
AAAGAAGCA
CTGTCATGC
ATAAAAGAA
AGTGATACT
```



Click on the 'Export data' button in the lefthand menu of most pages to export:

- FASTA sequence
- GTF or GFF features

...and more!

Customise your download



Custom datasets can be retrieved using the BioMart data-mining tool.

You may find exploring this web-based query tool easier than extracting information direct from our databases.

Fetch data programmatically



Write your own Perl scripts to retrieve small-to-medium datasets. All our data, as well as added functionality, is available through the Ensembl Perl API.

Use the API to retrieve gene and transcript sets, fetch alignments between sequences, compare allele frequencies and much more!

You can also use our [REST API](#) (currently in Beta) to retrieve data to process in the programming language of your choice.

Download databases & software



All of our data and software, including pipelines and web code, is available free.

- [Download data via FTP](#)
- [Ensembl pipeline in CVS](#)
- [Set up your own Ensembl website](#)

Multi-species data

Database	MySQL	EMF	MAF	BED	XML	Ancestral Alleles
Comparative genomics	MySQL	EMF	MAF	BED	XML	Ancestral Alleles
BioMart	MySQL	-	-	-	-	-
Stable ids	MySQL	-	-	-	-	-

Single species data

Popular species are listed first. You can customise this list via our [home page](#).

Show 10 entries

Show/hide columns

Filter

★	Species	DNA (FASTA)	cDNA (FASTA)	CDS (FASTA)	ncRNA (FASTA)	Protein sequence (FASTA)	Annotated sequence (EMBL)	Annotated sequence (GenBank)	Gene sets	Whole databases	Variation (GVF)	Variation (VCF)	Variation (VEP)	Regulation (GFF)	Data files	BAM/BigV
Y	Human <i>Homo sapiens</i>	FASTA	EMBL	GenBank	GTF GFF3	MySQL	GVF	VCF	VEP	Regulation (GFF)	Regulation on data files	BAM/BigV				
Y	Mouse <i>Mus musculus</i>	FASTA	EMBL	GenBank	GTF GFF3	MySQL	GVF	VCF	VEP	Regulation (GFF)	Regulation on data files	BAM/BigV				

Ensembl – pobieranie danych

1. Proszę poszukać pliki zawierające sekwencje:

- **Genomową (DNA) chromosomu 22**
- **białkowe** w formacie FASTA
- **Adnotacje** w formacie GTF **człowieka**

- *Jak nazywają się te pliki?*

- *Jak nazywa się plik z całą sekwencją genomową w formacie FASTA?*

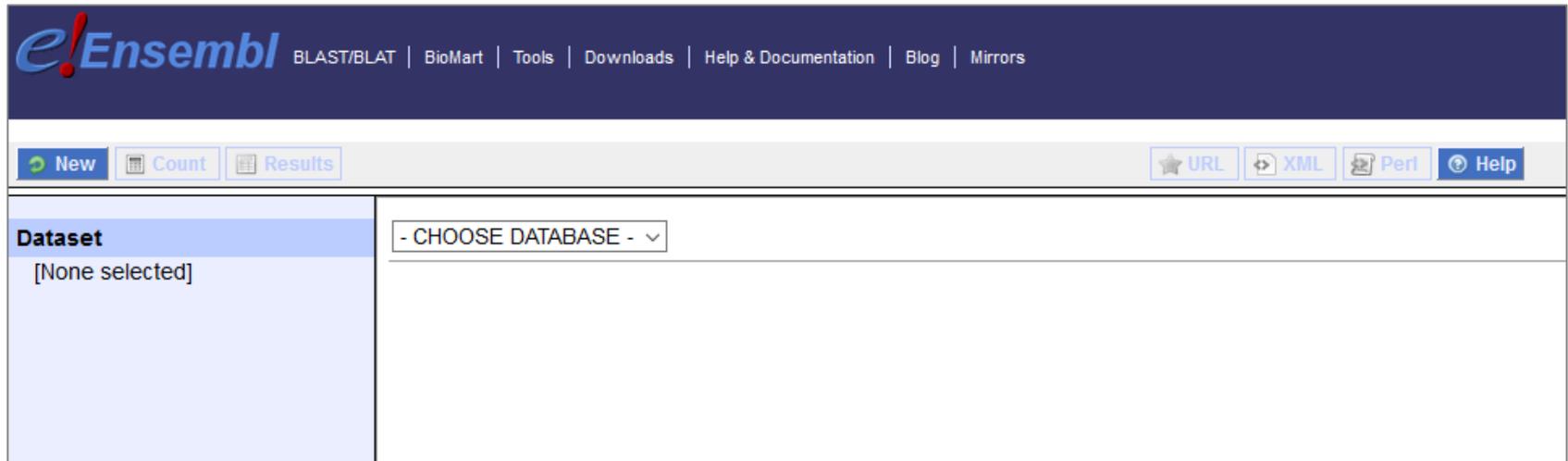
2. Proszę poszukać sekwencje **cDNA** w formacie FASTA dla bakterii ***Bacillus subtilis*** [*Ensembl Bacteria*]

3. Proszę poszukać sekwencje **ncRNA** dla myszy ***Mus musculus*** w bazie Ensembl [*wersja archiwalna 85*]

Ensembl – pobieranie danych

UWAGA! Operujemy nie tylko na identyfikatorach charakterystycznych dla Ensembl

Ensembl BioMart



The screenshot displays the Ensembl BioMart web interface. At the top, the Ensembl logo is followed by navigation links: BLAST/BLAT, BioMart, Tools, Downloads, Help & Documentation, Blog, and Mirrors. Below the navigation bar is a toolbar with buttons for 'New', 'Count', and 'Results'. On the right side of the toolbar are buttons for 'URL', 'XML', 'Perl', and 'Help'. The main content area is divided into two sections. The left section, titled 'Dataset', shows '[None selected]'. The right section contains a dropdown menu with the text '- CHOOSE DATABASE -' and a downward arrow.

Ensembl – pobieranie danych

The screenshot shows the Ensembl BioMart interface. At the top, there is a navigation bar with links for BLAST/BLAT, BioMart, Tools, Downloads, Help & Documentation, Blog, and Mirrors. Below this is a toolbar with buttons for New, Count, Results, URL, XML, Perl, and Help. The main configuration area is divided into several sections: 'Dataset' (set to 'Ensembl Genes 87'), 'Human genes (GRCh38.p7)', 'Filters' (set to '[None selected]'), 'Attributes' (with 'Gene ID' and 'Transcript ID' selected), and another 'Dataset' section (set to '[None Selected]'). A red circle highlights the 'Dataset', 'Filters', and 'Attributes' sections.

- **Dataset** → wybieramy wersję bazy Ensembl oraz organizm
- **Filters** → możliwość zawężania przeszukiwania do określonych cech, identyfikatorów, zbiorów danych itp.
- **Attributes** → określamy jak mają wyglądać dane wyjściowe, jakie cechy, identyfikatory, informacje są dla nas ważne.
- **Count** → informacja o tym ile rekordów o określonych parametrach znaleziono w porównaniu z całym zestawem genów dla danego organizmu.
- **Results** → informacja o tym jak wygląda fragment danych wynikowych, opcje zapisu i eksportu danych

ĆWICZENIE

Proszę sprawdzić **ile** jest genów kodujących białka u myszy (*Mus musculus*) i **pobrać** listę identyfikatorów Gene ID i Transcript ID, nazwę genu i opis (*Filters* → *Gene* → *Gene type* → ‘protein coding’ → *Count*)

New
Count
Results

★ URL
XML
Perl
Help

Dataset 22212 / 52386 Genes
Mouse genes (GRCm38.p5)

Filters

Gene type: protein_coding

Attributes

Gene stable ID
Transcript stable ID
Gene description
Gene name

Dataset
[None Selected]

Export all results to TSV Unique results only Go

Email notification to

View rows as HTML Unique results only

Gene stable ID	Transcript stable ID	Gene description	Gene name
ENSMUSG00000064370	ENSMUST00000082421	mitochondrially encoded cytochrome b [Source:MGI Symbol;Acc:MGI:102501]	mt-Cytb
ENSMUSG00000064368	ENSMUST00000082419	mitochondrially encoded NADH dehydrogenase 6 [Source:MGI Symbol;Acc:MGI:102495]	mt-Nd6
ENSMUSG00000064367	ENSMUST00000082418	mitochondrially encoded NADH dehydrogenase 5 [Source:MGI Symbol;Acc:MGI:102496]	mt-Nd5
ENSMUSG00000064363	ENSMUST00000082414	mitochondrially encoded NADH dehydrogenase 4 [Source:MGI Symbol;Acc:MGI:102498]	mt-Nd4
ENSMUSG00000065947	ENSMUST00000084013	mitochondrially encoded NADH dehydrogenase 4L [Source:MGI Symbol;Acc:MGI:102497]	mt-Nd4l
ENSMUSG00000064360	ENSMUST00000082411	mitochondrially encoded NADH dehydrogenase 3 [Source:MGI Symbol;Acc:MGI:102499]	mt-Nd3
ENSMUSG00000064358	ENSMUST00000082409	mitochondrially encoded cytochrome c oxidase III [Source:MGI Symbol;Acc:MGI:102502]	mt-Co3
ENSMUSG00000064357	ENSMUST00000082408	mitochondrially encoded ATP synthase 6 [Source:MGI Symbol;Acc:MGI:99927]	mt-Atp6
ENSMUSG00000064356	ENSMUST00000082407	mitochondrially encoded ATP synthase 8 [Source:MGI Symbol;Acc:MGI:99926]	mt-Atp8
ENSMUSG00000064354	ENSMUST00000082405	mitochondrially encoded cytochrome c oxidase II [Source:MGI Symbol;Acc:MGI:102503]	mt-Co2

ĆWICZENIE

- Proszę pobrać sekwencje nukleotydowe (Unspliced transcript) miRNA u człowieka (Transcript type) pochodzące z chromosomu 22 wraz z informacjami w linii definicji na temat: typu transkryptu, długości transkryptu i nici (strand).

The screenshot displays the Ensembl BioMart interface. On the left, the 'Filters' section is set to 'Chromosome/scaffold: 22' and 'Transcript type: miRNA'. The 'Attributes' section includes 'Gene stable ID', 'Transcript stable ID', 'Unspliced (Transcript)', 'Transcript type', 'Strand', and 'Transcript length (including UTRs and CDS)'. The 'Dataset' section shows 'None Selected'. On the right, the 'Export all results to' dropdown is set to 'File', and the 'FASTA' checkbox is selected. The 'Email notification to' field is empty. Below, the 'View' section shows '10 rows as FASTA' with the 'Unique results only' checkbox unselected. The main content area displays a list of FASTA entries for miRNA transcripts, including their Ensembl IDs and coordinates.

```
>ENSG00000280968|ENST00000625572|miRNA|-1|110
TCCTTGGGGACCCCTGGCAGCCCCCTCTGATGATTCCTTCTTCTGAGCAGCCTCATGATG
AGCAAACCTGAGCCTCTAAGAAGTTGACTGAAGGGGCTGCTTCCCAAGGA
>ENSG00000275818|ENST00000621576|miRNA|1|65
CTATTTTGTGTGAGTACAGAGAGCATCTGAATGGGTACAGTTGTTGTCTCTTGTTCCTCA
CACAG
>ENSG00000264757|ENST00000583102|miRNA|-1|80
GACTGTGCTCTCACTGTTCAACCAGCACTAGCAGTACCAGACGGTCTGTGGAGTCTTGG
GGAATGGAGAGAGCACAGTC
>ENSG00000278420|ENST00000622122|miRNA|-1|61
GAGGGTTGGGGTGGAGGGCAAGGAGCTGGGTGGGGTGCCAAGCCTCTGTCCCCACCCCA
G
>ENSG00000284520|ENST00000385140|miRNA|1|83
CGGGGTGAGGTAGTAGGTTGTGTGGTTTCAGGGCAGTGATGTTGCCCTCGGAAGATAAC
TATAACAACCTACTGCCTTCCCTG
>ENSG00000276162|ENST00000619803|miRNA|1|80
GGTTGGCTATAACTATCATTTCCAAGTTGTGCTTTTAGGAAATGTTGGCTGTCTTCCGG
AGAGAGAATGGGGAGCCAGG
>ENSG00000266320|ENST00000579518|miRNA|1|119
GGTATGCTGTGTGCGCTGTCTTCTCTGTTGGGAGCAGGCTCCGGGGGACAGGGAAAAGCAC
ACAAGGAACTTGTCCTCTAGGGCTGCAGTCTCATGGGAGAGTGACATGCACCAGGACC
>ENSG00000264073|ENST00000582434|miRNA|-1|88
GGTGACTCCAGGGACTGCCTTAGGAGAAAAGTTTCTGGAAGTTCTGACATCCAGAACTT
TCTCCTAAGGCAGTCCCCTGGGAGTCACT
>ENSG00000283871|ENST00000385018|miRNA|1|82
GGCCTGCCGACACTCTTCCCTGTTCAGTACTATAGGCCGCTGGGAAGCAGTGCAATG
ATGAAAGGGCATCGGTCAGGTC
>ENSG00000266533|ENST00000579667|miRNA|1|83
ACGGCATCTTTGCACTCAGCAGGCAGGCTGGTGCAGCCGTTGGTGGGGACCATCCTGCC
```

ĆWICZENIE

Proszę pobrać sekwencje nukleotydowe (cDNA) wszystkich genów kodujących białka u *Arabidopsis thaliana*, które zlokalizowane są na chromosomie 3 i mają swoje ortologiczne geny u *Vitis vinifera*.

The screenshot shows the Ensembl BioMart interface. The left sidebar displays the dataset 'Arabidopsis thaliana genes (TAIR10)' with filters for 'Chromosome/scaffold: 3', 'Orthologous Vitis vinifera Genes: Only', and 'Gene type: protein_coding'. The main area shows the export options (File, FASTA, Unique results only) and the view settings (10 rows as FASTA, Unique results only). The results are displayed in FASTA format, showing the cDNA sequences for two genes: AT3G01070 and AT3G01050.

```
>AT3G01070 | AT3G01070.1
AAAAAAGGTAAGAAATATGACCAAAATGAAAAGGAAAATATTACTTTTGGTTAATTACAT
AACGCTTATCTTGTACTTTTCGTCAACCTCGAGGTTAATTAATAATTCGAAGGTTAGTC
GAAGGAATTACGGATCTAGAGAGATCTTTCCCTGAAAAAGAGACATATTTATCTCTCAG
ATTGAACCTTCGATAAACCCATCCCCCCCCACAAACGACGACTCATCCTTTTTTTTTTAGA
GTACTGATCGATGGCGAGAGTTGCGGTTTTGGTTGCAGGGGCGGTGCTGGCTTTTCTGCT
GGCGGCAGCAATGTGACGGCGAAGAGGTGGACGGTGGGTGATAACAAGTTCGGAATCC
TAATATCAACTATACCATTTGGGCTCAGGACAAACATTTCTACCTCGACGATTGGCTCTA
TTTTGTATATGAGAGAAACCAATACAACGTTATAGAAGTGAACGAGACGAACACATAAG
CTGCAATCCCAATAACCCGATCGCAAATGGAGTCGTGGAGCTGGAAGGGACTTGGTTCA
TCTCAATGTGACAAGGCATTACTATCTGATTAGTGGTAATGGTGGTGGATGTTACGGAGG
TATGAAGCTCGCTGTTTTAGTTGAGAAACCGCCTCCTCCACCAGCTGCAGCACATAACAA
GAACAGTGCAAGAAGAACCCTTCTCGGTCTCCGGTTTTGCCTATCAGTTTCTTATTCCAGT
AGCTGTTTTCGCAGCAGTTGGGACACGGTACTGAGGTTGTGGTAACAACTTTAAAGAGA
AAACGATTGTTATCGGTGAAATCCGATTCGATTGATATCTTCGCAAAGTGAAGCTTTTT
GTAGTTTTAGTGTTACTCTTTTATTATGTTGGGTTCTTGGATCTTTCTAGTCATTTCA
CACTCGGGTGTGTGTGTGTCTTTGGAGATTGTCATGCTTTTGCTTTTGTTCAGGGGG
GGAATCTTTGTGAAATAACAATTGTAATATAATTTGTTGAACAATCTTTAAAACCT
TTTTTGGCCATATGATTGGGCTTCGATTTTTTTTGTAGTTGTAATTTGTTTGTTCATGC
TTTTGCTTTTGTAGGGGGAAAAATCATGTGTTGAA
>AT3G01050 | AT3G01050.3
CATTGAACACTCAGAGATCTTCTTCTTCTTGTATTTCTTCTCTCTCTCTCTGCTGATG
CGTGTCTTGATTCATCGGCTCAGATCAGTCACCAAATAGAATAAATAGCAAACAAGATC
CTCTGTGGAATTTGATACATAGAGAACTGTCAGTGCTGTGGTGTCTTTTGGATTTTC
TTGTGGTTCGTGGATTGTTTGTGGTTGGTTGCTTCTTGATAAATGCCAGAAGTCCAT
```

Uwaga!
Identyfikatory specyficzne dla
rzodkiewnika

Trzy metabazy i przeglądarki



National Center for Biotechnology Information



Ensembl Genome Browser

European Bioinformatics Institute, Wellcome Trust Sanger Institute

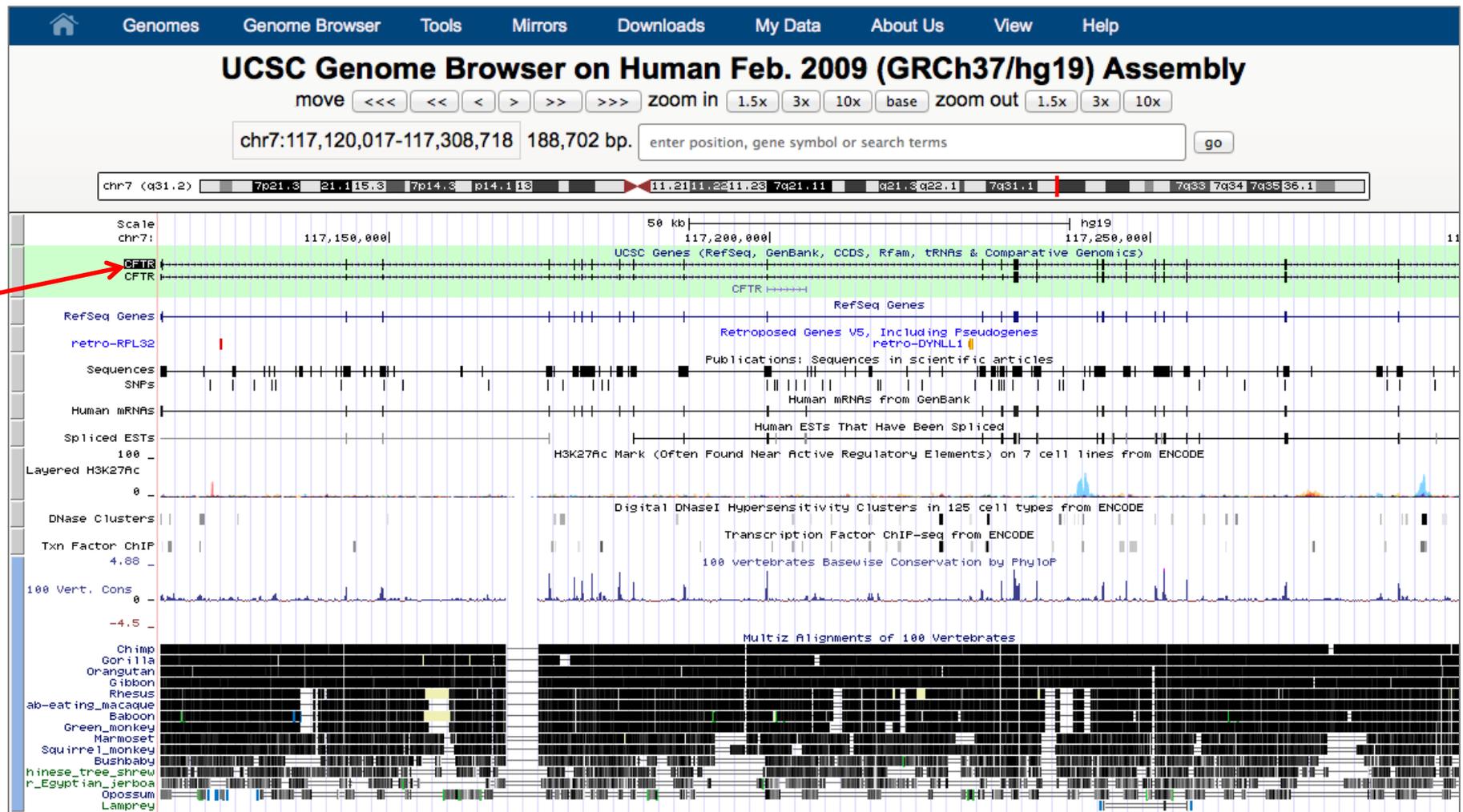


University of California, Santa Cruz Genome Browser

UCSC Genes

UCSC Transcript ID ale bardzo często ID zaczerpnięte jest z innej bazy

CFTR	(uc011knq.2)	at chr7:117120017-117308718	- Homo sapiens cystic fibrosis transmembrane conductance regulator (ATP-binding cassette sub-family C (CFTR/MRP), member 1 (ABCC1), mRNA.
CFTR	(uc003vje.1)	at chr7:117199518-117204842	- Homo sapiens cystic fibrosis transmembrane conductance regulator (ATP-binding cassette sub-family C (CFTR/MRP), member 1 (ABCC1), mRNA.
CFTR	(uc003vjd.3)	at chr7:117120017-117308718	- Homo sapiens cystic fibrosis transmembrane conductance regulator (ATP-binding cassette sub-family C (CFTR/MRP), member 1 (ABCC1), mRNA.
ABCC1	(uc021tdt.1)	at chr16:16208623-16225792	- Homo sapiens ATP-binding cassette, sub-family C (CFTR/MRP), member 1 (ABCC1), mRNA.
ABCC1	(uc021tds.1)	at chr16:16208623-16225792	- Homo sapiens ATP-binding cassette, sub-family C (CFTR/MRP), member 1 (ABCC1), mRNA.
ABCC1	(uc021tdr.1)	at chr16:16149949-16170258	- Homo sapiens ATP-binding cassette, sub-family C (CFTR/MRP), member 1 (ABCC1), mRNA.
ABCC1	(uc021tdg.1)	at chr16:16146581-16170258	- Homo sapiens ATP-binding cassette, sub-family C (CFTR/MRP), member 1 (ABCC1), mRNA.
ABCC8	(uc021qej.1)	at chr11:17491648-17498449	- Homo sapiens ATP-binding cassette, sub-family C (CFTR/MRP), member 8 (ABCC8), mRNA.
ABCC5	(uc011bqt.2)	at chr3:183637724-183732235	- Homo sapiens ATP-binding cassette, sub-family C (CFTR/MRP), member 5 (ABCC5), transcript variant 2 (ABCC5), mRNA.
ABCC11	(uc010vxl.1)	at chr16:48250025-48281478	- Homo sapiens ATP-binding cassette, sub-family C (CFTR/MRP), member 11 (ABCC11), transcript variant 1 (ABCC11), mRNA.



move start

Click on a feature for details. Click or drag in the base position track to zoom in. Click side bars for track options. Drag side bars or labels up or down to reorder tracks. Drag tracks left or right to new position.

track search default tracks default order hide all add custom tracks track hubs configure reverse resize refresh

collapse all Use drop-down controls below and press refresh to alter tracks displayed. expand all Tracks with lots of items will automatically be displayed in more compact modes.

Mapping and Sequencing Tracks refresh

Base Position dense	Chromosome Band hide	STS Markers hide	FISH Clones hide	Recomb Rate hide	deCODE Recomb hide
ENCODE Pilot hide	Map Contigs hide	Assembly hide	GRC Map Contigs hide	INSDC hide	Gap hide
BAC End Pairs hide	Fosmid End Pairs hide	GC Percent hide	GRC Patch Release hide	Hg18 Diff hide	GRC Incident hide
Hi Seq Depth hide	Wiki Track hide	BU ORChID hide	Mapability hide	Short Match hide	Restr Enzymes hide

Phenotype and Disease Associations refresh

GAD View hide	DECIPHER hide	OMIM AV SNPs hide	OMIM Genes hide	OMIM Pheno Loci hide	COSMIC hide
LOVD Variants hide	HGMD Variants hide	UniProt Variants hide	ClinVar Variants hide	GWAS Catalog hide	ISCA hide
Coriell CNVs hide	RGD Human QTL hide	RGD Rat QTL hide	MGI Mouse QTL hide	GeneReviews hide	

Genes and Gene Prediction Tracks refresh

UCSC Genes pack	GENCODE... hide	Old UCSC Genes hide	UCSC Alt Events hide	CCDS hide	RefSeq Genes dense
Other RefSeq hide	MGC Genes hide	ORFeome Clones hide	TransMap... hide	Vega Genes hide	Pfam in UCSC Gene hide

Zmiany ustawień w widoku poszczególnych „ścieżek” w przeglądarce (ang. tracks)

UCSC - przeszukiwanie bazy

ĆWICZENIE

Proszę poszukać gen kodujący insulinę (*INS*) u człowieka w najnowszej wersji genomu i odpowiedzieć na pytania:

1. Na którym chromosomie jest zlokalizowany?
2. Jakiej długości jest sekwencja genomowa?
3. Ile ma wariantów splicingowych w bazie RefSeq?
4. Ile ma egzonów?
5. Proszę poszukać identyfikator z RefSeq dla sekwencji nt
6. Gdzie ulega najwyższej ekspresji?
7. Czy są jakieś zidentyfikowane elementy powtarzalne w obrębie tego genu?

UCSC – pobieranie danych

1. Pojedyncze rekordy w bazie danych,
z poziomu przeglądarki genomowej
2. Download → FTP
3. TableBrowser

UCSC – Pobieranie danych (FTP)

Home → Downloads → Genome Data

UCSC Genome Bioinformatics

[Home](#) - [Genomes](#) - [Blat](#) - [Tables](#) - [Gene Sorter](#) - [PCR](#) - [FAQ](#) - [Help](#)

Sequence and Annotation Downloads

This page contains links to sequence and annotation data downloads for the genome assemblies featured in the UCSC Genome Browser. For quick access to the most recent assembly of each genome, see the [current genomes](#) directory. This directory may not always contain the most recent assembly.

To view the current descriptions and formats of the tables in the annotation database, use the "describe table schema" button (no longer maintained) also provides descriptions of selected tables in the database.

All tables in the Genome Browser are freely usable for any purpose except as indicated in the README.txt files in the download directory. Please use the corresponding download link and review the README text. These data were contributed by many researchers, as listed on the [Genome Browser](#) data you use.

VERTEBRATES - Complete annotation sets

Human	Green Monkey	Platypus
Alpaca	Guinea pig	Rabbit
American alligator	Hedgehog	Rat
Armadillo	Horse	Rhesus
Atlantic cod	Kangaroo rat	Rock hyrax
Baboon	Lamprey	Sheep
Bonobo	Lizard	Shrew
Brown kiwi	Malayan flying lemur	Sloth
Budgerigar	Manatee	Squirrel
Bushbaby	Marmoset	Squirrel monkey
Cat	Medaka	Stickleback
Chicken	Medium ground finch	Tarsier
Chimpanzee	Megabat	Tasmanian devil
Chinese hamster	Microbat	Tenrec
Coelacanth	Minke whale	Tetraodon
Cow	Mouse	Tree shrew

UCSC – Pobieranie danych (FTP)

ĆWICZENIE

Proszę **poszukać** zbiór sekwencji mRNA (1) oraz plik z całą sekwencją genomową (2) dla makaka (*Rhesus macaca*) – **UWAGA! wersja genomu z października 2010. [Full dataset]**

- Jak oznaczona jest ta wersja genomu makaka i gdzie go sekwencjonowano?
- Jak nazywają się poszczególne pliki?
- Proszę się zapoznać z opisem danych, nazw plików.

UCSC – Pobieranie danych (FTP)

ĆWICZENIE

Proszę **poszukać** sekwencję genomową chromosomu 22 człowieka (*Homo sapiens*) w formacie FASTA.

UWAGA! Wersja genomu hg38

[Data set by chromosome]

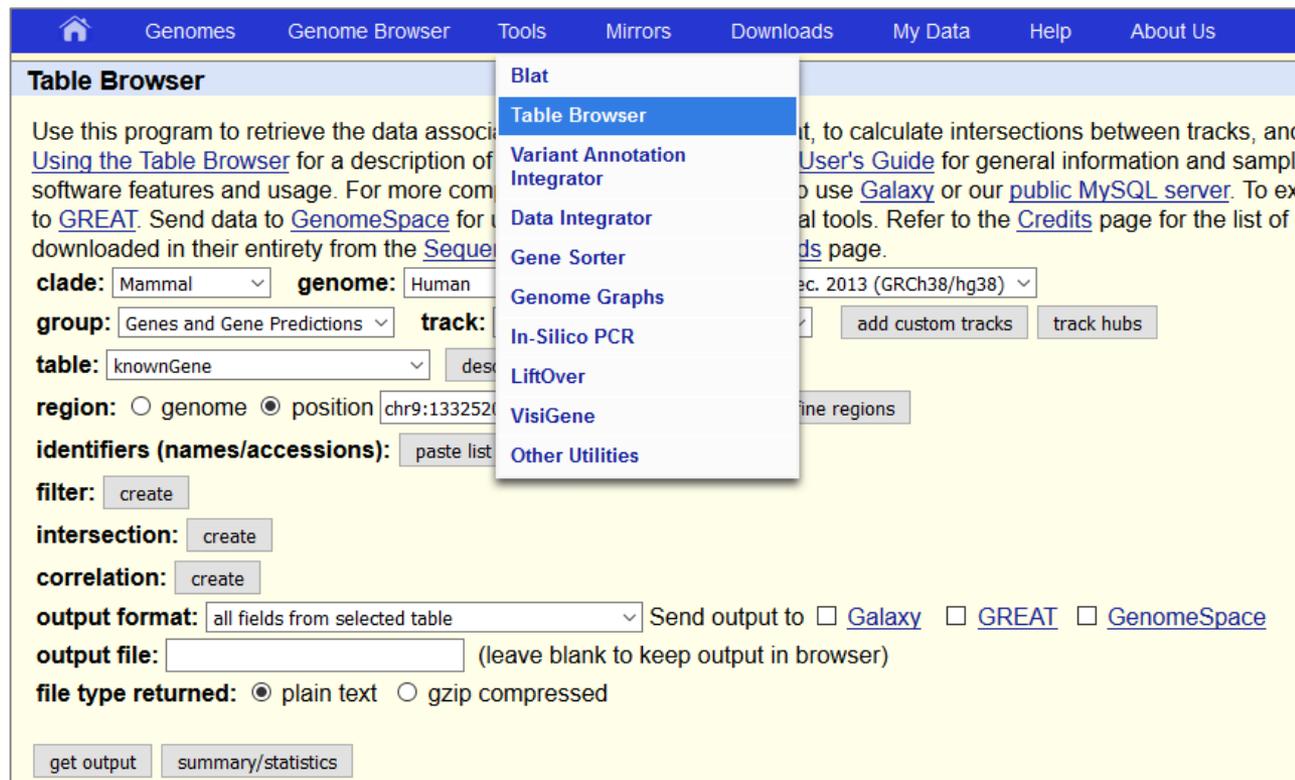
- Jak nazywa się pobrany plik?
- Skąd pochodzą inne nazwy plików?
- Proszę się zapoznać z opisem danych, nazw plików.

UCSC – pobieranie danych

UWAGA! Operujemy nie tylko na identyfikatorach charakterystycznych dla UCSC

UCSC Table Browser

<http://genome.ucsc.edu/cgi-bin/hgTables>



The screenshot displays the UCSC Table Browser interface. At the top, a navigation bar includes links for Genomes, Genome Browser, Tools, Mirrors, Downloads, My Data, Help, and About Us. The main content area is titled 'Table Browser' and contains a detailed description of the tool's purpose: to retrieve data associated with a specific genomic region. Below the description, there are several input fields and buttons for configuring the query. A dropdown menu is open, showing a list of tools including Blat, Table Browser (highlighted), Variant Annotation Integrator, Data Integrator, Gene Sorter, Genome Graphs, In-Silico PCR, LiftOver, VisiGene, and Other Utilities. The main form includes fields for 'clade' (Mammal), 'genome' (Human), 'group' (Genes and Gene Predictions), 'track' (knownGene), 'region' (position chr9:133252), 'identifiers (names/accessions)', 'filter', 'intersection', 'correlation', 'output format' (all fields from selected table), 'output file', and 'file type returned' (plain text). Buttons for 'get output' and 'summary/statistics' are located at the bottom.

Table Browser

Use this program to retrieve the data associated with a specific genomic region. For a description of software features and usage. For more complete information, refer to the [User's Guide](#) for general information and sample queries, or the [FAQ](#) for more information. To export data, you can use [Galaxy](#) or our [public MySQL server](#). To export data to other formats, refer to the [Credits](#) page.

clade: genome:

group: track:

table:

region: genome position

identifiers (names/accessions):

filter:

intersection:

correlation:

output format: Send output to [Galaxy](#) [GREAT](#) [GenomeSpace](#)

output file: (leave blank to keep output in browser)

file type returned: plain text gzip compressed

UCSC – Pobieranie danych Table Browser

Home Genomes Genome Browser Tools Mirrors Downloads My Data About Us Help

Table Browser

Use this program to retrieve the data associated with a track in text format, to calculate intersections between tracks, and to retrieve DNA sequence covered by a track. For help in using this application see [Using the Table Browser](#) for a description of the controls in this form, the [User's Guide](#) for general information and sample queries, and the OpenHelix Table Browser [tutorial](#) for a narrated presentation of the software features and usage. For more complex queries, you may want to use [Galaxy](#) or our [public MySQL server](#). To examine the biological function of your set through annotation enrichments, send the data to [GREAT](#). Refer to the [Credits](#) page for the list of contributors and usage restrictions associated with these data. All tables can be downloaded in their entirety from the [Sequence and Annotation Downloads](#) page.

clade: Mammal **genome:** Human **assembly:** Feb. 2009 (GRCh37/hg19)

group: Custom Tracks **track:** User Track [manage custom tracks](#) [track hubs](#)

table: ct_UserTrack_3545 [describe table schema](#)

region: genome ENCODE Pilot regions **position:** chr7:127469864-127481543 [lookup](#) [define regions](#)

identifiers (names/accessions): [paste list](#) [upload list](#)

filter: [create](#)

intersection: [create](#)

correlation: [create](#)

output format: BED - browser extensible data Send output to [Galaxy](#) [GREAT](#)

output file: (leave blank to keep output in browser)

file type returned: plain text gzip compressed

[get output](#) [summary/statistics](#)

To reset all user cart settings (including custom tracks), [click here](#).

Using the Table Browser

This section provides brief line-by-line descriptions of the Table Browser controls. For more information on using this program, see the [Table Browser User's Guide](#).

UCSC – Pobieranie danych Table Browser

ĆWICZENIE

Proszę pobrać sekwencje mRNA szympansa w formacie **FASTA**, które pochodzą z chromosomu X.

Ile jest tych sekwencji?

clade: Mammal genome: Chimp assembly: May 2016 (Pan_tro 3.0/panTro5)
group: mRNA and EST track: Chimp mRNAs
table: all_mrna
region: genome position chrX
identifiers (names/accession): filter: create
intersection: create
correlation: create
output format: sequence
output file:
file type returned: plain text
get output summary/statistics

Sequence Retrieval Region Options:

- Promoter/Upstream by 100 bases
- Blocks
- Regions between blocks
- Downstream by 100 bases
- One FASTA record per item.
- One FASTA record per region (block/between blocks) with 0 extra bases upstream (5') and 0 extra downstream (3')

Note: if a feature is close to chromosome.

Sequence Formatting

- Blocks in upper case, even if they are lower case.
- All upper case.
- All lower case.
- Mask repeats: to low complexity

get sequence cancel

Chimp mRNAs (all_mrna) Summary

item count	150
item bases	5,474,437 (3.65%)
item total	5,552,068 (3.70%)
smallest item	20
average item	37,014
biggest item	1,157,324
block count	746
block bases	138,497 (0.09%)
block total	153,985 (0.10%)
smallest block	4
average block	206
biggest block	3,148

and upstream/downstream bases are added, they may be truncated.

```
>panTro5_all_mrna_AK305659_range=chrX:1147656-1177541_5'pad=0_3'pad=0_strand=-  
GGGTCTCAGGTTTGAGGTGGTCCCCTCCAAGTTTAAAGAGAAGCTGGACA  
AAGCCTCCTTCGCTACTCCGTATGGGTACGCCATGGAGACCGCAAGCAG  
AAGGCCCTGGAGGTGGCAACCGGATGCACCAGgcaagggggccctgctc  
tttggcccttatgaaaattccttggctgtgatgaggggatttgcaggcag  
taaatatacatgatgatacaataaacatcattcagaagatgcagttgggtgc  
acagcttagctattgctatgttataatgtaatgcaactattgoggtaacct  
aataataaccatgatacagaataatgtagcttactctgttactactgct  
gttataatattctcttgcaattgatacagatcattgatgatagatt  
aatgtagcttacttactatgtgttataatattctcttgcaattgatt  
tcatagatcattcatgatagataaattgtagcttactctgttactatg  
ttataatattctcacaattgatacagatcattgatgatacagaattaat  
gtagttatcttactatgctgtgtaatatctcttgcaattgatacagat  
atcattgatgatacagaataatgtagcttactctgttactatgtttata  
atattctcacaattgatcatagatcattgatgatagataaattgatt  
agttatcttactatgtgttataatattctcttgcaattgatagat  
cattgatgatacagaataatgtagcttactctgttactatgtttataat  
attctcacaattgatcacagatcattgatgatagataaattgtag
```

UCSC – Pobieranie danych Table Browser

ĆWICZENIE

Proszę pobrać informacje w formacie **GTF** o genach z chromosomu X z bazy RefSeq u kota.

clade: Mammal genome: Cat assembly: Nov. 2014 (ICGSC Felis_catus_8.0/felCat8)
group: Genes and Gene Predictions track: RefSeq Genes add custom tracks track hubs
table: refGene describe table schema
region: genome position chrA2:53484451-53597660 lookup define regions
identifiers (names/accessions): paste list upload list
filter: create
intersection: create
correlation: create
output format: GTF - gene transfer format Send output to Galaxy GREAT GenomeSpace
(leave blank to keep output in browser)
 gzip compressed

RefSeq Genes (refGene) Summary

item count	11
item bases	183,230 (0.15%)
item total	183,230 (0.15%)

smallest item	chrX	felCat8_refGene	exon	21883079	21883169	0.000000	+	.	gene_id	"NM_001009262"; transcript_id "NM_001009262";
	chrX	felCat8_refGene	start_codon	21890310	21890312	0.000000	+	.	gene_id	"NM_001009262"; transcript_id "NM_001009262";
	chrX	felCat8_refGene	CDS	21890310	21891434	0.000000	+	0	gene_id	"NM_001009262"; transcript_id "NM_001009262";
average item	chrX	felCat8_refGene	stop_codon	21891435	21891437	0.000000	+	.	gene_id	"NM_001009262"; transcript_id "NM_001009262";
	chrX	felCat8_refGene	exon	21890249	21891487	0.000000	+	.	gene_id	"NM_001009262"; transcript_id "NM_001009262";
	chrX	felCat8_refGene	exon	21891501	21891669	0.000000	+	.	gene_id	"NM_001009262"; transcript_id "NM_001009262";
biggest item	chrX	felCat8_refGene	stop_codon	42849134	42849136	0.000000	-	.	gene_id	"NM_001083952"; transcript_id "NM_001083952";
	chrX	felCat8_refGene	CDS	42849137	42849283	0.000000	-	0	gene_id	"NM_001083952"; transcript_id "NM_001083952";
	chrX	felCat8_refGene	exon	42849134	42849283	0.000000	-	.	gene_id	"NM_001083952"; transcript_id "NM_001083952";
	chrX	felCat8_refGene	CDS	42849493	42849594	0.000000	-	0	gene_id	"NM_001083952"; transcript_id "NM_001083952";
block count	chrX	felCat8_refGene	exon	42849493	42849594	0.000000	-	.	gene_id	"NM_001083952"; transcript_id "NM_001083952";
	chrX	felCat8_refGene	CDS	42850334	42850410	0.000000	-	2	gene_id	"NM_001083952"; transcript_id "NM_001083952";
block bases	chrX	felCat8_refGene	exon	42850334	42850410	0.000000	-	.	gene_id	"NM_001083952"; transcript_id "NM_001083952";
	chrX	felCat8_refGene	CDS	42851084	42851234	0.000000	-	0	gene_id	"NM_001083952"; transcript_id "NM_001083952";
block total	chrX	felCat8_refGene	exon	42851084	42851234	0.000000	-	.	gene_id	"NM_001083952"; transcript_id "NM_001083952";
	chrX	felCat8_refGene	exon	42851084	42851084	0.000000	-	.	gene_id	"NM_001083952"; transcript_id "NM_001083952";
smallest block	chrX	felCat8_refGene	CDS	42852193	42852273	0.000000	-	0	gene_id	"NM_001083952"; transcript_id "NM_001083952";
	chrX	felCat8_refGene	exon	42852193	42852273	0.000000	-	.	gene_id	"NM_001083952"; transcript_id "NM_001083952";
	chrX	felCat8_refGene	CDS	42852462	42852549	0.000000	-	1	gene_id	"NM_001083952"; transcript_id "NM_001083952";
average block	chrX	felCat8_refGene	exon	42852462	42852549	0.000000	-	.	gene_id	"NM_001083952"; transcript_id "NM_001083952";
	chrX	felCat8_refGene	exon	42853442	42853546	0.000000	-	1	gene_id	"NM_001083952"; transcript_id "NM_001083952";
biggest block	chrX	felCat8_refGene	CDS	42853442	42853546	0.000000	-	.	gene_id	"NM_001083952"; transcript_id "NM_001083952";

UCSC – Pobieranie danych Table Browser

ĆWICZENIE

Proszę pobrać wszystkie elementy powtarzalne (repeats) w formacie **BED** dla człowieka pochodzące z chromosomu 22 (UWAGA! najnowsza wersja genomu)

through annotation enrichments, send the data to [GREAT](#). Send data to [GenomeSpace](#) for use with the restrictions associated with these data. All tables can be downloaded in their entirety from the [SnpEff](#)

clade: Mammal **genome:** Human **assembly:** Dec. 2013 (GRCh38/hg38)

group: Repeats **track:** RepeatMasker

table: rmsk

region: genome position chr22

identifiers (names/accessions):

filter:

intersection:

output format: BED - browser extensible data

output file: (leave blank for default)

file type returned: plain text gzip compressed

RepeatMasker (rmsk) Summary Statistics

item count	79,521
item bases	20,908,038 (53.39%)
item total	20,908,038
smallest item	chr22 10511018 10511332
average item	chr22 10511479 10511791
biggest item	chr22 10511878 10512212
smallest score	chr22 10512454 10512692
average score	chr22 10512706 10514778
biggest score	chr22 10514778 10515050

chr22	10510227	10510528	AluSx1	2021	+
chr22	10511018	10511332	L1MC5a	781	-
chr22	10511479	10511791	L1MB1	524	+
chr22	10511878	10512212	L1MB1	313	+
chr22	10512454	10512692	L1MB1	656	+
chr22	10512706	10514778	L1MB1	11092	+
chr22	10514778	10515050	AluSx1	1933	+
chr22	10515050	10515074	L1MB1	11092	+
chr22	10515074	10515121	(GAAG)n	52	+
chr22	10515121	10516103	L1MB1	11092	+
chr22	10516114	10516222	(TA)n	47	+
chr22	10516223	10516285	LTR66	237	-
chr22	10516287	10516630	L1MB1	1504	+

Inne bazy danych sekwencji i adnotacji



Genomes Online Database

<http://genomesonline.org/cgi-bin/GOLD/index.cgi>



Wellcome Trust Sanger Institute

<https://www.sanger.ac.uk/resources/downloads/>



<http://www.cbs.dtu.dk/services/GenomeAtlas/>

Inne bazy danych sekwencji i adnotacji



Genomes – EMBL - EBI

<http://www.ebi.ac.uk/genomes/index.html>

Genomes Pages - At the EBI

- Complete genomes
- Archaea
- Archaeal virus
- Bacteria
- Eukaryota
- Organelle
- Phage
- Plasmid
- Viroid
- Virus
- Links
 - WGS info
 - EnsemblGenomes
 - Genome Reviews
 - Integr8 (proteomes)
 - Fasta33 Server
 - Ensembl

Databases > Nucleotide > The European Nucleotide Archive > Complete Genomes

Access to Completed Genomes

The first completed genomes from [viruses](#), [phages](#) and [organelles](#) were deposited into the EMBL Database in the early 1980's. Since then, molecular biology's shift to obtain the complete sequences of as many genomes as possible combined with major developments in sequencing technology resulted in hundreds of complete genome sequences being added to the database, including [Archaea](#), [Bacteria](#) and [Eukaryota](#). These web pages give access to a large number of complete genomes, [help](#) is available to describe the layout.

Whole Genome Shotgun Sequences (WGS)

Methods using whole genome shotgun data are used to gain a large amount of genome coverage for an organism. WGS data for a growing number of organisms are being submitted to DDBJ/EMBL/GenBank.

[More information about WGS projects...](#)

Last 40 Genome Entries

Date	Accession	Description
27-JUN-2014	AP014565.1	Salmonella enterica subsp. enterica serovar Typhimurium str. L-3553
27-JUN-2014	AP014566.1	Salmonella enterica subsp. enterica serovar Typhimurium str. L-3553 plasmid pST3553
27-JUN-2014	KJ634409.1	Yerba mate endornavirus strain INTA
27-JUN-2014	KJ739609.1	Pomacea canaliculata mitochondrion
26-JUN-2014	KJ562277.1	Amphilophus citrinellus mitochondrion
26-JUN-2014	KJ650081.1	Paratiroza sinica mitochondrion
26-JUN-2014	KJ668270.1	Ecotropic murine leukemia virus
26-JUN-2014	KJ668271.1	Polytropic murine leukemia virus
25-JUN-2014	HF920637.1	Armadillidium vulgare iridescent virus
25-JUN-2014	HG938353.1	Rhizobium galegae str. HAMB1 540
25-JUN-2014	HG938355.1	Rhizobium galegae str. HAMB1 1141
25-JUN-2014	HG938356.1	Rhizobium galegae str. HAMB1 1141 chromid pHAMB11141a
25-JUN-2014	HG938357.1	Rhizobium galegae str. HAMB1 1141 plasmid pHAMB11141b
25-JUN-2014	HG975439.1	Solanum pennellii chromosome ch00
25-JUN-2014	HG975440.1	Solanum pennellii chromosome ch01
25-JUN-2014	HG975441.1	Solanum pennellii chromosome ch02



Przegląd podstawowych źródeł danych

dotyczących projektów
sekwencjonowania nowej generacji

SRA NCBI

ENA EBI

GEO NCBI



SRA

Sequence Read Archive (SRA) makes biological sequence data available to the research community to enhance reproducibility and allow for new discoveries by comparing data sets. The SRA stores raw sequencing data and alignment information from high-throughput sequencing platforms, including Roche 454 GS System®, Illumina Genome Analyzer®, Applied Biosystems SOLiD System®, Helicos Heliscope®, Complete Genomics®, and Pacific Biosciences SMRT®.

Sequence Read Archive (SRA)

- **Główne źródło danych** z sekwencjonowania nowej generacji
- Wygodny w użyciu sposób pobierania oraz przeszukiwania danych
- Możliwość wysyłania własnych danych
- Dane przechowywane w formacie SRA

Sequence Read Archive przechowuje dane z wielu platform: 454, IonTorrent, Illumina, SOLiD, Helicos, Complete Genomics, Pacific Biosciences oraz Oxford Nanopore Technologies.

SRA przechowuje również przyrównania w postaci pozycji odczytów na sekwencji referencyjnej a także dane z bazy ENA.

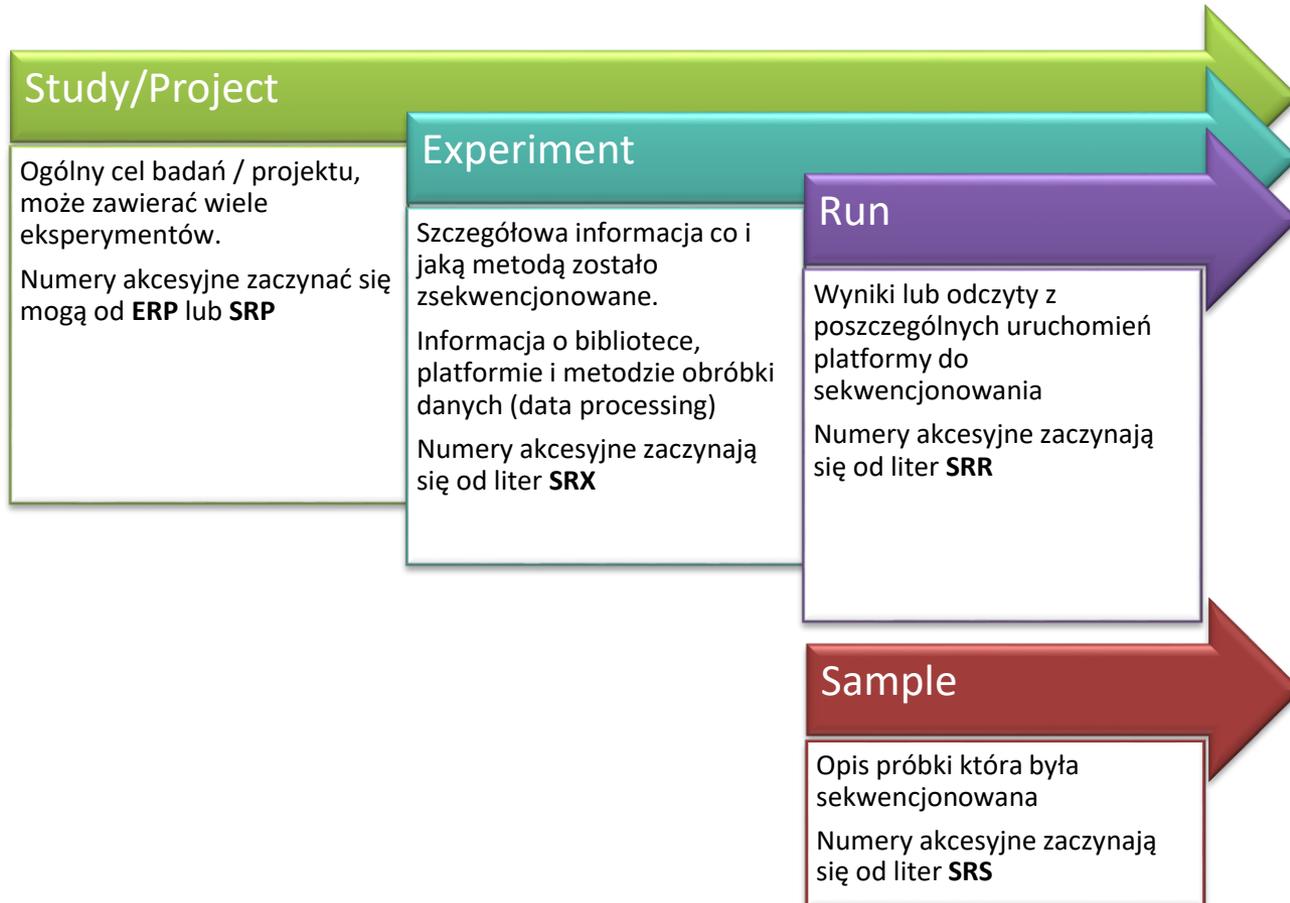
<http://www.ncbi.nlm.nih.gov/sra/>



SRA

Sequence Read Archive (SRA) makes biological sequence data available to the research community to enhance reproducibility and allow for new discoveries by comparing data sets. The SRA stores raw sequencing data and alignment information from high-throughput sequencing platforms, including Roche 454 GS System®, Illumina Genome Analyzer®, Applied Biosystems SOLiD System®, Helicos Heliscope®, Complete Genomics®, and Pacific Biosciences SMRT®.

Sequence Read Archive (SRA)



SRA: search builder



<http://www.ncbi.nlm.nih.gov/sra/>

Advanced Search > Organism: **Arabidopsis thaliana (1)**

Platform: „illumina” (2), Properties: „instrument illumina hiseq 2000” (3)

SRA
Advanced

SRA Advanced Search Builder

((Arabidopsis thaliana[Organism]) AND "illumina"[Platform]) AND "instrument illumina hiseq 2000"[Properties]

[Edit](#)

[Clear](#)

Builder

Organism

AND [Platform]

AND [Properties]

- instrument illumina hiscansq (2393)
- instrument illumina hiseq 1000 (11425)
- instrument illumina hiseq 1500 (4716)
- instrument illumina hiseq 2000 (1145604)**
- instrument illumina hiseq 2500 (297832)
- instrument illumina hiseq 3000 (1534)
- instrument illumina hiseq 4000 (2450)
- instrument illumina miseq (343892)
- instrument ion torrent pgm (21355)
- instrument ion torrent proton (3165)

AND

or [Add to history](#)

SRA: search builder

<http://www.ncbi.nlm.nih.gov/sra/>

SRA [Create alert](#) [Advanced](#) [Help](#)

Access: Public (7,748) | Source: DNA (4,142), RNA (3,564) | Type: genome (2,100) | Other: aligned data (166) | [Clear all](#) | [Show additional filters](#)

Summary ▾ 20 per page ▾ | Send to: ▾ | Filters: [Manage Filters](#)

View results as an expanded interactive table using the RunSelector. [Send results to Run selector](#)

Search results

Items: 1 to 20 of 7748 | << First < Prev Page 1 of 388 Next > Last >>

- [Illumina HiSeq 2000 sequencing: Ribosome footprinting upon oxidative stress treatment for wild-type and catalase2 knock-out mutants](#)
1 ILLUMINA (Illumina HiSeq 2000) run: 160.4M spots, 16.2G bases, 9Gb downloads
Accession: ERX1495854
- [Illumina HiSeq 2000 sequencing: Ribosome footprinting upon oxidative stress treatment for wild-type and catalase2 knock-out mutants](#)
1 ILLUMINA (Illumina HiSeq 2000) run: 180.9M spots, 18.3G bases, 10Gb downloads
Accession: ERX1495853

Find related data | Database: |

Search details
{"Arabidopsis thaliana"[Organism] AND "illumina"[Platform]) AND "instrument illumina hiseq 2000"[Properties]} | | [See more...](#)

Recent activity

SRA: Run

Run Browser > [SRR4333274](#)

([SRR4333274](#))

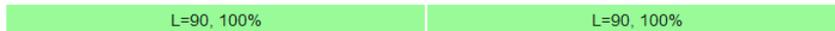
[Metadata](#) [Analysis](#) [Reads](#) [Data access](#) [FASTA/FASTQ download](#)

Run

Run	Spots	Bases	Size	GC Content	Published	Access Type
SRR4333274	32.4M	5.8G	3.6G	36.8%	2016-10-07	public

Quality graph ([bigger](#))

This run has 2 reads per spot:



[Legend](#)

Experiment

Experiment	Library Name	Platform	Strategy	Source	Selection	Layout	Action
SRX2206057	Cvi-0	Illumina	WGS	GENOMIC	RANDOM	PAIRED	BLAST

Biosample

Biosample	Sample Description	Organism
-----------	--------------------	----------

SRA: Reads

(SRR4333274)

 Metadata

 Analysis

 Reads

 Data access

 FASTA/FASTQ download

Filter

Search by sub-sequence, spot ID, etc 

Find

[What can the filter be applied to?](#)

Reads: 32,394,270 reads

< Page 1 / 3,239,427 >

quality scores advanced options

1 SRR4333274.1
name: FCD0T67ACXX:8:1101:1243:1935
member: GCCAATATx: 1243, y: 1935
< >

2 SRR4333274.2
name: FCD0T67ACXX:8:1101:1250:1955
member: GCCAATATx: 1250, y: 1955
< >

3 SRR4333274.3
name: FCD0T67ACXX:8:1101:1249:1982
member: GCCAATATx: 1249, y: 1982
< >

4 SRR4333274.4
name: FCD0T67ACXX:8:1101:1169:1988
member: GCCAATATx: 1169, y: 1988
< >

Reads (separated)

>gnl|SRA|SRR4333274.1.1 FCD0T67ACXX:8:1101:1243:1935 forward (Biological)
NCAATAACCT CAGACCAATC AATCGAATAT TGATTAATAC ATAATTGATC GAACACTACT
TGAAAACGGC TCTTCCGCTC AGAAACGAAA

>gnl|SRA|SRR4333274.1.2 FCD0T67ACXX:8:1101:1243:1935 reverse (Biological)
ACCGAGAGAT CCATAAATCG GGATCCTAAT GCATATAGAT ACAATGGTC CAATGGGAGC
AAGAATTTCC AGGAGCATTT GGAACATTTT

SRA – NCBI

ĆWICZENIE

- Proszę poszukać w bazie SRA danych dla człowieka z sekwencjonowania transkryptomu (Properties → strategy rna-seq), gdzie wynikiem są odczyty sparowane (Layout → paired)
- Dla jednego wybranego rekordu proszę poszukać identyfikatory Study, Experiment, Run, Sample oraz krótki opis eksperymentu.

Search

Examples: [BN000065](#), [histone](#)

[Advanced](#)

[Sequence](#)

European Nucleotide Archive

The European Nucleotide Archive (ENA) provides a comprehensive record of the world's nucleotide sequencing information, covering raw sequencing data, sequence assembly information and functional annotation. [More about ENA](#)

Access to ENA data is provided through the browser, through search tools, large scale file download and through the API.

Text Search

Examples: [BN000065](#), [histone](#)

Search

[Advanced search](#)

Sequence Search

Enter or paste a nucleotide sequence or accession number

Search

[Advanced search](#)

Popular

- [Submit and update](#)
- [Sequence submissions](#)
- [Genome assembly submissions](#)
- [Submitting environmental sequences](#)
- [Citing ENA data](#)
- [Rest URLs for data retrieval](#)
- [Rest URLs to search ENA](#)

Latest ENA news

09 Dec 2014: [ENA release 122](#)

Release 122 of ENA's assembled/annotated sequences is now available.

12 Nov 2014: [Simplification of data release procedures](#)

The European Nucleotide Archive will couple the public release of sequence records and the release of study records that contain these sequence records, with immediate effect.

11 Nov 2014: [ENA/EMG Sample Record Annotation Workshop](#)

European Nucleotide Archive (ENA) and EBI Metagenomics Portal (EMG), are organising the ENA/EMG Sample Record Annotation Workshop on the 1-5 December 2014 to enrich the environmental sample records.

Identyfikatory – numery akcesyjne

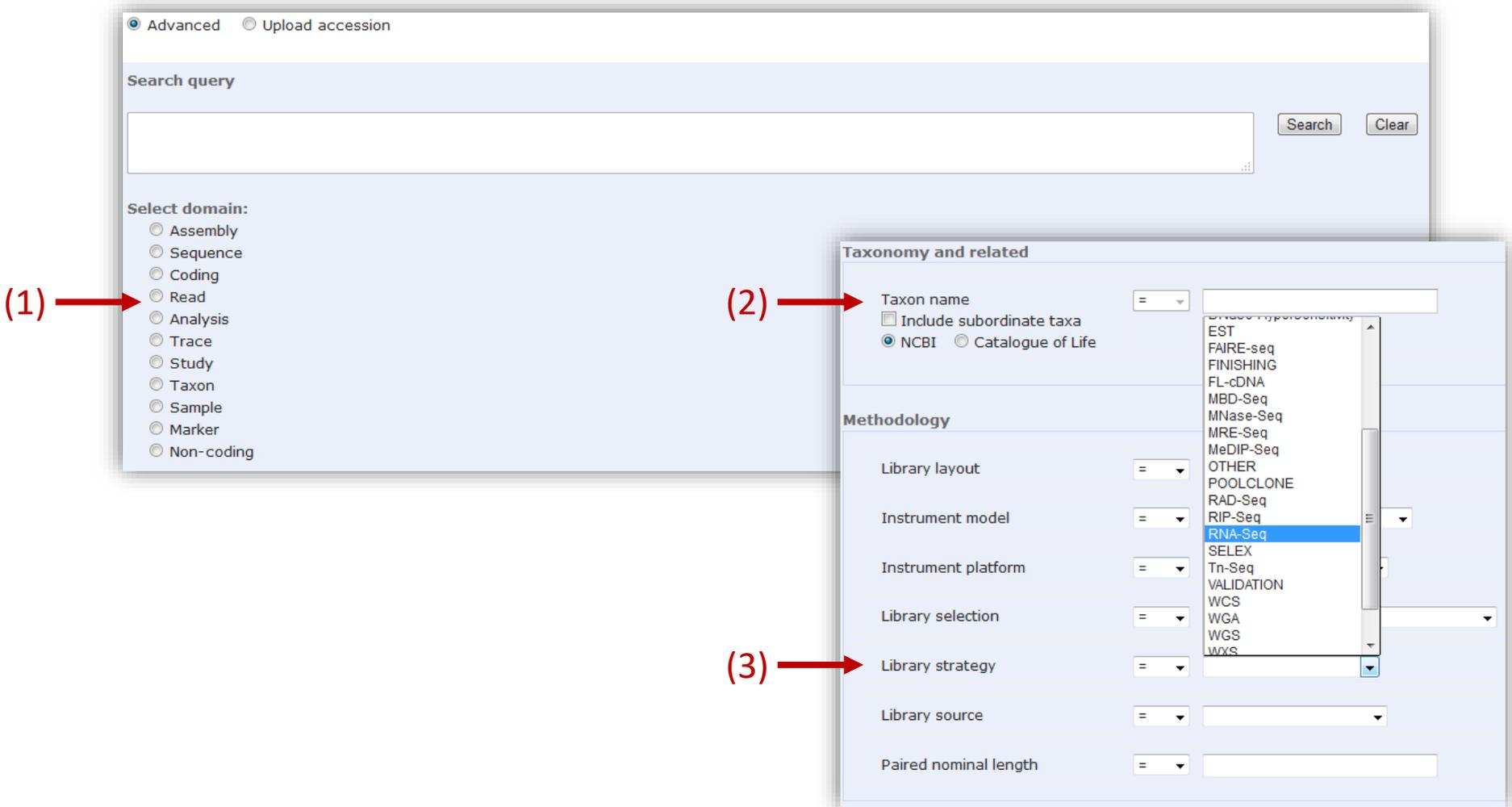
Metadata object	Accession prefix	Number of digits	Example
Submission	ERA, SRA, DRA	6	ERA000092
Sample	ERS, SRS, DRS	6	ERS000081
Study	ERP, SRP, DRP	6	ERP000016
Experiment	ERX, SRX, DRX	6	ERX000398
Run	ERR, SRR, DRR	6	ERR003990
Analysis	ERZ, SRZ, DRZ	6	ERZ000001
Project/Study	PRJ		

ER - EBI

SR – SRA NCBI

DR - DDBJ

Select domain: **read** (1), Taxon name: **Bos taurus** (2), Library strategy = **RNA-Seq** (3)



The screenshot shows the ENA search interface with three red arrows pointing to specific settings:

- (1)** Points to the "Read" option under "Select domain:".
- (2)** Points to the "Taxon name" field, which contains "Bos taurus".
- (3)** Points to the "Library strategy" dropdown menu, which is open and shows "RNA-Seq" selected.

The interface includes a search query field, a "Search" button, and a "Clear" button. The "Select domain:" section lists various options: Assembly, Sequence, Coding, Read, Analysis, Trace, Study, Taxon, Sample, Marker, and Non-coding. The "Taxonomy and related" section includes "Taxon name" and "Include subordinate taxa" (with "NCBI" selected). The "Methodology" section includes "Library layout", "Instrument model", "Instrument platform", "Library selection", "Library strategy", "Library source", and "Paired nominal length".

Advanced Upload accession

Search query

[Help](#)

tax_eq(9913) AND library_strategy="RNA-Seq"

Search

Clear

[Query builder](#)

[Edit query](#)

Search results for *tax_eq(9913) AND library_strategy="RNA-Seq"*

Read

Study (46)

Experiment (797)

Run (824)

Study (46 results found)

PRJDA72405 RNA-sequencing of bovine GV and M2 oocytes and 8-cell stage embryos.

[View all 46 results](#)

Experiment (797 results found)

DRX000961 Illumina Genome Analyzer II sequencing; RNA-sequencing of bovine granulosa cells from young cows (28.3?0.7 months)

[View all 797 results](#)

Run (824 results found)

DRR001364 Illumina Genome Analyzer II sequencing; RNA-sequencing of bovine granulosa cells from young cows (28.3?0.7 months)

[View all 824 results](#)

Read Files

Portal

Attributes

Publications

[Download files](#)

View: [TEXT](#)

Download: [TEXT](#)

[Select columns](#)

Showing results 1 - 10 of 16 results

[Next](#)

Study accession	Secondary study accession	Sample accession	Secondary sample accession	Experiment accession	Run accession	Tax ID	Scientific name	Instrument model	Library layout	Fastq files (ftp)	Fastq files (galaxy)	Submitted files (ftp)	Submitted files (galaxy)
PRJDA72405	DRP000449	SAMD00000728	DRS000859	DRX000961	DRR001364	9913	Bos taurus	Illumina Genome Analyzer II	SINGLE	File 1	File 1		
PRJDA72405	DRP000449	SAMD00000728	DRS000859	DRX000962	DRR001365	9913	Bos taurus	Illumina Genome Analyzer II	SINGLE	File 1	File 1		
PRJDA72405	DRP000449	SAMD00000728	DRS000859	DRX000962	DRR001366	9913	Bos taurus	Illumina Genome Analyzer II	SINGLE	File 1	File 1		
PRJDA72405	DRP000556	SAMD00011319	DRS001301	DRX001352	DRR001889	9913	Bos taurus	Illumina Genome Analyzer II	SINGLE	File 1	File 1		
PRJDA72405	DRP000556	SAMD00011318	DRS001302	DRX001354	DRR001890	9913	Bos taurus	Illumina Genome Analyzer II	SINGLE	File 1	File 1		
PRJDA72405	DRP000556	SAMD00011318	DRS001302	DRX001354	DRR001891	9913	Bos taurus	Illumina Genome Analyzer II	SINGLE	File 1	File 1		

ENA – EBI

ĆWICZENIE

- Proszę poszukać w bazie ENA danych dla człowieka, z sekwencjonowania transkryptomu (Library strategy → RNA-seq), gdzie wynikiem są odczyty sparowane (Library layout → paired)
- Dla jednego wybranego rekordu (eksperymentu) proszę poszukać identyfikatory Study, Experiment, Run, Sample oraz krótki opis eksperymentu.

GEO – Gene Expression Omnibus (NCBI)

<http://www.ncbi.nlm.nih.gov/gds>



GEO DataSets

This database stores curated gene expression DataSets, as well as original Series and Platform records in the Gene Expression Omnibus (GEO) repository. Enter search terms to locate experiments of interest. DataSet records contain additional resources including cluster tools and differential expression queries.

Advanced Search > Organism: **Mus musculus (1)**, DataSet Type: „non coding rna profiling by high throughput sequencing” (2)

(mus musculus[Organism]) AND "non coding rna profiling by high throughput sequencing"[DataSet Type]

[Edit](#) [Clear](#)

Builder

Organism ← (1) [Show index list](#)

AND [DataSet Type] [Hide index list](#)

methylation profiling by high throughput sequencing (632)

methylation profiling by snp array (9)

non coding rna profiling by array (1975)

non coding rna profiling by genome tiling array (105)

non coding rna profiling by high throughput sequencing (1341) ← (2.2)

other (999)

protein profiling by mass spec (4)

protein profiling by protein array (151)

snp genotyping by snp array (483)

third party reanalysis (118)

[Previous 200](#)
[Next 200](#) ↑ (2.1)

[Refresh index](#)

AND [Show index list](#)

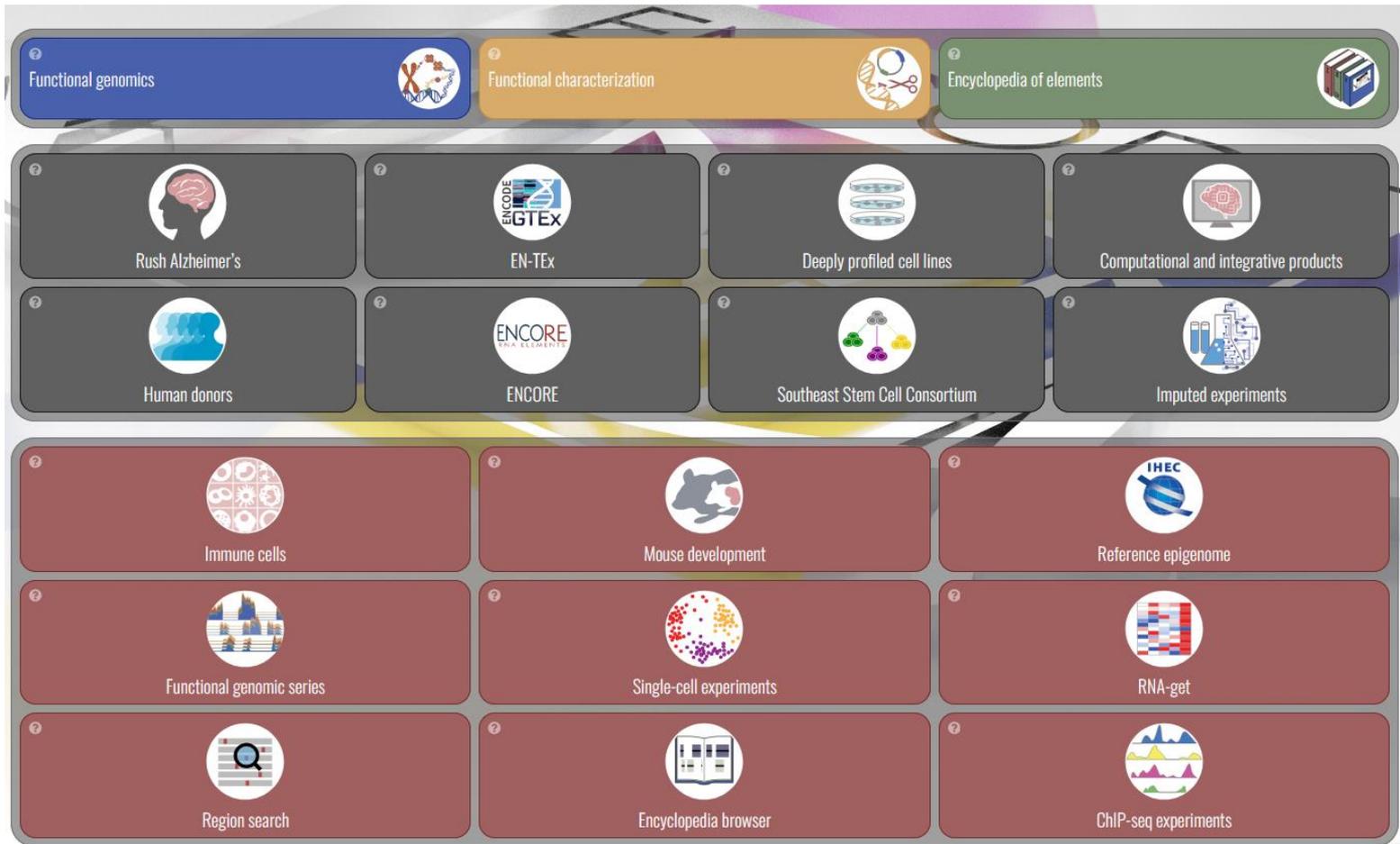
or [Add to history](#)

Inne bazy danych

- ENCODE
<https://www.encodeproject.org/>
- The Cancer Genome Atlas:
<https://tcga-data.nci.nih.gov/tcga/>
- Expression Atlas
<https://www.ebi.ac.uk/gxa/home>
- GENT2:
<http://gent2.appex.kr/gent2/>
- SCDevDB
<https://scdevdb.deepomics.org/>

ENCODE

<https://www.encodeproject.org/>



RNA-Seq Blog

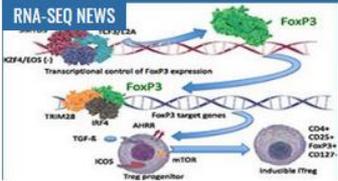
<http://www.rna-seqblog.com/>

DONT MISS Online Workshop – Bioinformatics Pipeline Development with Nextflow (November 22-25, 2021) 📶 f t in Search...



HOME NEWS ▾ EVENTS ▾ JOBS TECHNOLOGY ▾ DATA ANALYSIS ▾ BLOG READER POSTS CONTACT ▾

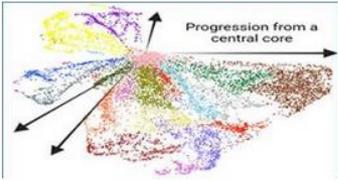
RNA-SEQ NEWS



A new blood-based RNA sequencing method to detect coronary artery disease
3 hours ago • Leave a comment • 81 Views
Diagnosis of CAD is typically based on coronary angiography, an imaging process used to detect restricted blood flow in the heart's major arteries.

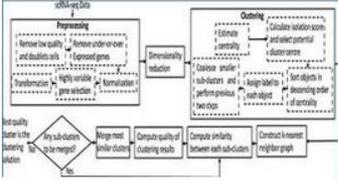
Up to 40% of the 1 million ...

[Read More »](#)



Crafting a blueprint for single-cell RNA sequencing
4 hours ago • Leave a comment • 104 Views
Droplet-based single-cell RNA sequencing (scRNA-Seq) has moved rapidly from a technology with great potential to a method applied to ever-broader questions. The detailed information that scRNA-Seq offers has proven incredibly ...

[Read More »](#)



UICPC – Centrality-based clustering for scRNA-seq data analysis without user input
1 day ago • Leave a comment • 258 Views
scRNA-seq data analysis enables new possibilities for identification of novel cells, specific characterization of known cells and study of cell heterogeneity...

STAY CONNECTED

[RSS](#) [Facebook](#) [Twitter](#) [LinkedIn](#)

RNA-SEQ & COVID-19

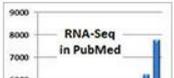


[SUBMIT NEWS](#)

RNA-SEQ FORUM



PUBLICATIONS TREND



Year	Publications
2000	~5000
2005	~6000
2010	~7000
2015	~8000
2020	~9000

SUBSCRIBE TO THE RNA-SEQ BLOG

email address

[Subscribe](#)

RNA-SEQ PRODUCTS & SERVICES



Zymo-Seq RiboFree™ Total RNA Library Kit
Compatible With Any Organism

[LEARN MORE](#)





Remove >95% rRNA from plant RNA-seq samples

[GET A TRIAL KIT](#)

RNA-seqlopedia



RNA-seq produces millions of sequences from complex RNA samples. With this powerful approach, you can:

1. Measure gene expression.
2. Discover and annotate complete transcripts.
3. Characterize alternative splicing and polyadenylation.

The RNA-seqlopedia provides an overview of RNA-seq and of the choices necessary to carry out a successful RNA-seq experiment.

RNA-seqlopedia is written by the [Cresko Lab](#) of the [University of Oregon](#) and was funded by grant R24 RR032670 (NIH, National Center for Research Resources).

[Credits.](#)

SEQanswers

<http://seqanswers.com/>



SEQanswers
the next generation sequencing community

SEQanswers Home

User Name Remember Me?
Password

[Register](#)

[FAQ](#)

[Community](#) ▾

[Calendar](#)

[Today's Posts](#)

[Search](#)

You are currently viewing the SEQanswers forums as a guest, which limits your access. [Click here to register now](#), and join the discussion

» Site Navigation

- » [About SEQanswers](#)
- » [Next Gen Summaries](#)
- » **Forums**
- » [Wiki](#)
- » [Instrument Map](#)

» Log in

User Name:

Password:

 Remember Me?

Not a member yet?
[Register Now!](#)

» Online Users: 222

11 members and 211 guests
[AndrewO](#), [boucceerb](#),
[Chipper](#), [dsenalik](#), [Michael](#)
[L. Altshuler](#), [netpumber](#),
[nrd](#), [vasha](#), [wen vuan](#),
[vzhang](#), [zirouc](#)

Most users ever online was
1,120 - 06-26-2014 at 02:21

» New Posts

Title, Username, & Date	Last Post	Replies	Views	Forum
Short RNAseq results for metabolic pathways netpumber	Today 02:22 PM by netpumber	0	1	RNA Sequencing
BLASTp parameter -dbsize problems (blastall -z) fireLog2	Today 01:24 PM by dsenalik	5	1,125	Bioinformatics
Wood Frog Genome ver 0.1 cement_head	Today 01:10 PM by cement_head	8	405	General
NCBI taxid download Pol8	Today 12:56 PM by GenoMax	1	45	Bioinformatics
Lotsa new toys from Illumina: HiSeq X Five, 3000, 4000, NextSeq 550 GW_OK	Today 12:56 PM by kcchan	42	3,233	Illumina/Solexa
DESeq2 Error nw328	Today 12:52 PM by nw328	0	35	Bioinformatics

Lotsa new toys from Illumina: HiSeq X Five, 3000, 4000, NextSeq 550

Jan 12, 2015 - 11:36 AM - by [GW_OK](#)

HiSeq 4000
HiSeq 3000
NextSeq 550
HiSeqX 5

<http://www.illumina.com/company/news...newsid=2006979>

42 Replies | 3,233 Views

» Our Sponsors

» Recent Job Postings

[Quality Engineer - Manufacturing](#)
01-21-2015 04:50 PM
by [Pacific Biosciences](#)

[Quality Engineer - R&D](#)
01-21-2015 04:37 PM
by [Pacific Biosciences](#)

[Scientist, Next-Generation Applications and...](#)
01-15-2015 10:00 AM
by [Pacific Biosciences](#)

[Senior Scientist, Applications](#)
01-13-2015 05:18 PM
by [Pacific Biosciences](#)

[Senior Bioinformatics Test Engineer](#)
01-09-2015 04:45 PM
by [Pacific Biosciences](#)

Technical Guides & Resources

[Beginner's Handbook of Next Generation Sequencing](#)
[Designing Your Sequencing Runs](#)
[Sequencing Coverage Guide](#)
[NGS Library Preparation Kit Guide](#)
[NGS Instrument Guide](#)
[Best Practices for Shipping RNA/DNA/Tissue for NGS](#)
[Cluster Density Optimization](#)
[External Next-Generation Sequencing Resources](#)

Application Guides

[Whole Genome Sequencing \(WGS\) and Re-sequencing](#)
[Whole Exome Sequencing](#)
[Whole Transcriptome and mRNA Sequencing](#)
[Shotgun Metagenomics Sequencing](#)
[Small RNA \(miRNA\) Sequencing](#)
[Whole Genome Sequencing on HiSeq X Ten](#)
[Sequencing Applications on HiSeq 3000, HiSeq 4000](#)

1. Find and select matching services (optional):

Sequencing & library prep Sequencing only Library prep only

Project type: DNA (whole genome) ▾

Number of samples: 1

Match by coverage Match by number of reads Help me

Targeted genome size: 3,200,000,000

Minimum coverage per sample: e.g. 30

[Find Services](#)

2. Submit your request:

We typically respond within 2 business hours (US Central Time)

First name

Last name

Your email

Tell us about your project

[Submit Request](#)

Dziękuję za uwagę