

Niniejsze opracowanie zostało stworzone przez dr Magdę Mielczarek, pracownika Uniwersytetu Przyrodniczego we Wrocławiu w ramach wykonywania obowiązków związanych z kształceniem studentów i jest przeznaczone dla studentów Bioinformatyki (Wydział Biologii i Hodowli Zwierząt) na potrzeby dydaktyczne bez prawa do dalszego rozpowszechniania.

WYBRANE BAZY DANYCH

Analiza danych NGS
Wykład 2

SUROWE DANE NGS

SRA
ENA

SEQUENCE READ ARCHIVE (SRA)



National Library of Medicine
National Center for Biotechnology Information

Log in

SRA

SRA ▼

Advanced

Search

Help



SRA - Now available on the cloud

Sequence Read Archive (SRA) data, available through multiple cloud providers and NCBI servers, is the largest publicly available repository of high throughput sequencing data. The archive accepts data from all branches of life as well as metagenomic and environmental surveys. SRA stores raw sequencing data and alignment information to enhance reproducibility and facilitate new discoveries through data analysis.

Getting Started

[Documentation](#)

[How to submit](#)

[How to search and download](#)

[How to use SRA in the cloud](#)

[Submit to SRA](#)

Tools and Software

[Download SRA Toolkit](#)

[SRA Toolkit Documentation](#)

[SRA-BLAST](#)

[SRA Run Browser](#)

[SRA Run Selector](#)

Related Resources

[Submission Portal](#)

[dbGaP Home](#)

[BioProject](#)

[BioSample](#)

SEQUENCE READ ARCHIVE (SRA)



National Library of Medicine
National Center for Biotechnology Information

Log in

SRA

SRA

((human[Organism]) AND 150[ReadLength]) AND WGS[Strategy]

Search

Create alert Advanced

Help

Access

Controlled (20,838)

Public (19,366)

Source

DNA (40,084)

RNA (31)

Type

genome (40,204)

Library Layout

paired (39,358)

single (846)

Platform

ABI SOLiD (12)

BGISeq (359)

Complete Genomics (81)

Illumina (39,469)

Ion Torrent (30)

Summary 20 per page

Send to:

Filters: [Manage Filters](#)

Search results

Items: 1 to 20 of 40204

<< First < Prev Page 1 of 2011 Next > Last >>

☐ [Whole exome sequencing of adult males: peripheral blood](#)

1. 1 DNBSEQ (DNBSEQ-T7) run: 50.9M spots, 15.3G bases, 8.1Gb downloads

Accession: SRX23909514

☐ [WGS of Homo sapiens: adult male prostate](#)

2. 1 ILLUMINA (Illumina NovaSeq 6000) run: 511.4M spots, 153.4G bases, 46.6Gb downloads

Accession: SRX23851466

☐ [WGS of Homo sapiens: adult male prostate](#)

3. 1 ILLUMINA (Illumina NovaSeq 6000) run: 501.6M spots, 150.5G bases, 46.4Gb downloads

Accession: SRX23851465

Search in related databases

Database	Access		all
	public	controlled	
BioSample	195	1	196
BioProject			
dbGaP		3	3
GEO Datasets	3		3

Find related data

Database:

Find items

SEQUENCE READ ARCHIVE (SRA)



National Library of Medicine
National Center for Biotechnology Information

Log in

SRA

SRA

((human[Organism]) AND 150[ReadLength]) AND WGS[Strategy]

Search

Sequence Read Archive

Search

Run Browser

Analyses

Study

Help

Access

Controlled (20,838)

Public (19,366)

Source

DNA (40,084)

RNA (31)

Type

genome (40,204)

Library Layout

paired (39,358)

single (846)

Platform

ABI SOLiD (12)

BGISeq (359)

Complete Genomics

Illumina (39,469)

Ion Torrent (30)

Study > SRP493619

SRP493619

Search

[? What can be entered in this field?](#)

Prostate sarcoma genomic sequencing

Identifiers

SRA: SRP493619

BioProject: [PRJNA1084752](#)

Study Type

Other

Abstract

A 45 year-old male of samples from prostate.

Filters: [Manage Filters](#)

Search in related databases

Database	Access		all
	public	controlled	
BioSample	195	1	196
BioProject			
dbGaP		3	3
GEO Datasets	3		3

Find related data

Database:

Find items

Malignant solitary fibrous tumor of the prostate genomic sequencing

A 45 year-old male of samples from prostate.

Accession	PRJNA1084752
Data Type	Raw sequence reads
Scope	Multispecies
Submission	Registration date: 6-Mar-2024 The Second Hospital of Dalian Medical University

Project Data:

Resource Name	Number of Links
SEQUENCE DATA	
SRA Experiments	2
OTHER DATASETS	
BioSample	2

SRA Data Details

Parameter	Value
Data volume, Gbases	304
Data volume, Mbytes	99804

Links from BioProject

Items: 2

- ☐ [WGS of Homo sapiens: adult male prostate](#)
1. 1 ILLUMINA (Illumina NovaSeq 6000) run: 511.4M spots, 153.4G bases, 46.6Gb downloads
Accession: SRX23851466
- ☐ [WGS of Homo sapiens: adult male prostate](#)
2. 1 ILLUMINA (Illumina NovaSeq 6000) run: 501.6M spots, 150.5G bases, 46.4Gb downloads
Accession: SRX23851465



SRA

SRA

▼

Advanced

Full ▼

Send to: ▼

SRX23851465: WGS of Homo sapiens: adult male prostate

1 ILLUMINA (Illumina NovaSeq 6000) run: 501.6M spots, 150.5G bases, 46.4Gb downloads

Design: After the DNA samples pass quality assessment, they undergo random fragmentation using a Covaris ultrasonic disruptor. Subsequent steps involve end repair, A-tailing, ligation of sequencing adapters, fragment size selection, PCR amplification, and purification, culminating in the completion of the sequencing library preparation process. This results in the generation of the final DNA library.

Submitted by: The Second Hospital of Dalian Medical University

Study: Prostate sarcoma genomic sequencing
[PRJNA1084752](#) • [SRP493619](#) • [All experiments](#) • [All runs](#)
[show Abstract](#)

Sample:
[SAMN40279114](#) • SRS20671550 • [All experiments](#) • [All runs](#)
[Organism:](#) [Homo sapiens](#)

Library:
[Name:](#) Tumor
[Instrument:](#) Illumina NovaSeq 6000
[Strategy:](#) WGS
[Source:](#) GENOMIC
[Selection:](#) RANDOM
[Layout:](#) PAIRED

Runs: 1 run, 501.6M spots, 150.5G bases, [46.4Gb](#)

Run	# of Spots	# of Bases	Size	Published
SRR28240330	501,557,640	150.5G	46.4Gb	2024-03-06



SRA

SRA

▼

Advanced

Full ▼

Send to: ▼

SRX23851466: WGS of Homo sapiens: adult male prostate

1 ILLUMINA (Illumina NovaSeq 6000) run: 511.4M spots, 153.4G bases, 46.6Gb downloads

Design: After the DNA samples pass quality assessment, they undergo random fragmentation using a Covaris ultrasonic disruptor. Subsequent steps involve end repair, A-tailing, ligation of sequencing adapters, fragment size selection, PCR amplification, and purification, culminating in the completion of the sequencing library preparation process. This results in the generation of the final DNA library.

Submitted by: The Second Hospital of Dalian Medical University

Study: Prostate sarcoma genomic sequencing
[PRJNA1084752](#) • [SRP493619](#) • [All experiments](#) • [All runs](#)
[show Abstract](#)

Sample:
[SAMN40279115](#) • SRS20671551 • [All experiments](#) • [All runs](#)
[Organism:](#) [Homo sapiens](#)

Library:
[Name:](#) Normal
[Instrument:](#) Illumina NovaSeq 6000
[Strategy:](#) WGS
[Source:](#) GENOMIC
[Selection:](#) RANDOM
[Layout:](#) PAIRED

Runs: 1 run, 511.4M spots, 153.4G bases, [46.6Gb](#)

Run	# of Spots	# of Bases	Size	Published
SRR28240329	511,434,907	153.4G	46.6Gb	2024-03-06



Results found in 2 databases

BIOPROJECT

Malignant solitary fibrous tumor of the prostate genomic sequencing

A 45 year-old male of samples from prostate.

PRJNA1084752

BioSample

Literature

Bookshelf	0
MeSH	0
NLM Catalog	0
PubMed	0
PubMed Central	0

Genes

Gene	0
GEO DataSets	0
GEO Profiles	0
Protein	error
ProteinSet	0

Genomes

Assembly	0
BioCollections	0
BioProject	1
BioSample	0
Genome	0
Nucleotide	0
SRA	2
Taxonomy	0

PRJNA1084752

Examples: histone, BN000065

Enter accession

View 

Examples: Taxon:9606, BN000065, PRJEB402

Text Search

Uses [EBI Search](#) to perform a free text search across ENA data. For more detailed usage please refer to the [help & documentation section](#).

Search term:

PRJNA1084752

Search 

Search results for PRJNA1084752

- **Read**
 - [Experiment](#) (2)
 - [Run](#) (2)
- **Study**
 - [Study](#) (1)

Experiment

Download ENA records: [XML](#) [TSV](#)

Accession	Description/Title
SRX23851465	Illumina NovaSeq 6000 paired end sequencing; Illumina NovaSeq 6000 sequencing: WGS of Homo sapiens: ...
SRX23851466	Illumina NovaSeq 6000 paired end sequencing; Illumina NovaSeq 6000 sequencing: WGS of Homo sapiens: ...

Experiment: SRX23851465

WGS of Homo sapiens: adult male prostate



Organism: [Homo sapiens \(human\)](#)

Instrument Platform: ILLUMINA

Instrument Model: Illumina NovaSeq 6000

Center Name: SUB14263517

Library Layout: PAIRED

Library Strategy: WGS

Library Source: GENOMIC

Library Name: Tumor

Library Selection: RANDOM

Experiment Accession: SRX23851465

Run Accession: SRR28240330

Download report: [JSON](#) [TSV](#)

Get d

Study Accession	Sample Accession	Experiment Accession	Run Accession	Tax Id	Sci Name
PRJNA1084752	SAMN40279114	SRX23851465	SRR28240330	9606	Homo sapiens

Experiment: SRX23851466

WGS of Homo sapiens: adult male prostate



Organism: [Homo sapiens \(human\)](#)

Instrument Platform: ILLUMINA

Instrument Model: Illumina NovaSeq 6000

Center Name: SUB14263517

Library Layout: PAIRED

Library Strategy: WGS

Library Source: GENOMIC

Library Name: Normal

Library Selection: RANDOM

Experiment Accession: SRX23851466

Run Accession: SRR28240329

Download All						
Study Accession	Sample Accession	Experiment Accession	Run Accession	Tax Id	Scientific Name	Generated FASTQ files: FTP
PRJNA1084752	SAMN40279115	SRX23851466	SRR28240329	9606	Homo sapiens	☐ SRR28240329_1.fastq.gz
						☐ SRR28240329_2.fastq.gz



Results found in 3 databases (1 error)

BIOPROJECT

RNA-seq of longuissimus dorsi muscle in an F2 Duroc-Pietrian pig resource population

RNA-seq of longuissimus dorsi muscle for 168 F2 pigs from the Michigan State University Duroc x Pietrain Pig Resource Population

PRJNA403969

[BioSample](#)

Literature	
Bookshelf	0
MeSH	0
NLM Catalog	0
PubMed	0
PubMed Central	3

Genes	
Gene	0
GEO DataSets	0
GEO Profiles	0
HomoloGene	error
PopSet	0

Genomes	
Assembly	0
BioCollections	0
BioProject	1
BioSample	0
Genome	0
Nucleotide	0
SRA	168
Taxonomy	0



BioProject

BioProject

PRJNA403969

[Create alert](#)[Advanced](#)[Browse by Project attributes](#)

Display Settings: ▾

Send to: ▾

RNA-seq of longuissimus dorsi muscle in an F2 Duroc-Pietrian pig resource population

Accession: PRJNA403969 ID: 403969

RNA-seq of longuissimus dorsi muscle for 168 F2 pigs from the Michigan State University Duroc x Pietrain Pig Resource Population

Accession	PRJNA403969
Data Type	Raw sequence reads
Scope	Multispecies
Grants	"Integrated analysis of genetic marker, mRNA and miRNA data to unravel mechanisms controlling growth and meat quality traits in pigs" (Grant ID 2014-67015-21619, National Institute of Food and Agriculture)
Submission	Registration date: 9-Sep-2017 Michigan State University
Relevance	Agricultural

Project Data:

Resource Name	Number of Links
SEQUENCE DATA	
SRA Experiments	168
OTHER DATASETS	
BioSample	168

▾ SRA Data Details

Parameter	Value
Data volume, Gbases	1,126
Data volume, Tbytes	0.62

Results found in 3 databases (1 error)

BIOPROJECT

RNA-seq of longuissimus dorsi muscle in an F2 Duroc-Pietrian pig resource population

RNA-seq of longuissimus dorsi muscle for 168 F2 pigs from the Michigan State University Duroc x Pietrain Pig Resource Population

PRJNA403969

[BioSample](#)

Literature	
Bookshelf	0
MeSH	0
NLM Catalog	0
PubMed	0
PubMed Central	3

Genes	
Gene	0
GEO DataSets	0
GEO Profiles	0
HomoloGene	error
PopSet	0

Genomes	
Assembly	0
BioCollections	0
BioProject	1
BioSample	0
Genome	0
Nucleotide	0
SRA	168
Taxonomy	0



SRA

PRJNA403969

Create alertAdvanced

Summary20 per pageSend to:

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

Search results

Items: 1 to 20 of 168

<< First < Prev Page 1 of 9 Next > Last >>

☐

[RNA-seq of Sus Scrofa: longuissimus dorsi muscle](#)

1. 1 ILLUMINA (Illumina HiSeq 2500) run: 21.9M spots, 5.5G bases, 2.4Gb downloads
Accession: SRX3174577

☐

[RNA-seq of Sus Scrofa: longuissimus dorsi muscle](#)

2. 1 ILLUMINA (Illumina HiSeq 2500) run: 18.5M spots, 4.6G bases, 2.1Gb downloads
Accession: SRX3174576

☐

[RNA-seq of Sus Scrofa: longuissimus dorsi muscle](#)

3. 1 ILLUMINA (Illumina HiSeq 2000) run: 61.9M spots, 12.4G bases, 7.9Gb downloads
Accession: SRX3174575

☐

[RNA-seq of Sus Scrofa: longuissimus dorsi muscle](#)

4. 1 ILLUMINA (Illumina HiSeq 2500) run: 17.7M spots, 4.4G bases, 2Gb downloads
Accession: SRX3174574

☐

[RNA-seq of Sus Scrofa: longuissimus dorsi muscle](#)

5. 1 ILLUMINA (Illumina HiSeq 2500) run: 22.3M spots, 5.6G bases, 2.5Gb downloads
Accession: SRX3174573

☐

[RNA-seq of Sus Scrofa: longuissimus dorsi muscle](#)

6. 1 ILLUMINA (Illumina HiSeq 2000) run: 99.7M spots, 19.9G bases, 12.7Gb downloads
Accession: SRX3174572

SRA

SRA

Advanced

FullSend to:

SRX3174577: RNA-seq of Sus Scrofa: longuissimus dorsi muscle
1 ILLUMINA (Illumina HiSeq 2500) run: 21.9M spots, 5.5G bases, 2.4Gb downloads

Design: miRNeasy Mini Kit RNA extraction, RIN greater than 7, Illumina TrueSeq Stranded mRNA Kit libraries sequenced on Illumina HiSeq 2500 platform, 2 x 125bp reads

Submitted by: Michigan State University

Study: RNA-seq of longuissimus dorsi muscle in an F2 Duroc-Pietrian pig resource population
[PRJNA403969](#) • [SRP117198](#) • [All experiments](#) • [All runs](#)
[show Abstract](#)

Sample:
[SAMN07626885](#) • [SRS2504484](#) • [All experiments](#) • [All runs](#)
[Organism: Sus scrofa](#)

Library:
[Name: 1082](#)
[Instrument: Illumina HiSeq 2500](#)
[Strategy: RNA-Seq](#)
[Source: TRANSCRIPTOMIC](#)
[Selection: RANDOM](#)
[Layout: PAIRED](#)

Runs: 1 run, 21.9M spots, 5.5G bases, [2.4Gb](#)

Run	# of Spots	# of Bases	Size	Published
SRR6023669	21,936,651	5.5G	2.4Gb	2017-09-09

Results found in 3 databases (1 error)

BIOPROJECT

RNA-seq of longuissimus dorsi muscle in an F2 Duroc-Pietrian pig resource population

RNA-seq of longuissimus dorsi muscle for 168 F2 pigs from the Michigan State University Duroc x Pietrain Pig Resource Population

PRJNA403969

BioSample

Literature	
Bookshelf	0
MeSH	0
NLM Catalog	0
PubMed	0
PubMed Central	3



Genes	
Gene	0
GEO DataSets	0
GEO Profiles	0
HomoloGene	error
PopSet	0

Genomes	
Assembly	0
BioCollections	0
BioProject	1
BioSample	0
Genome	0
Nucleotide	0
SRA	168
Taxonomy	0

You searched 8+ million full text articles

Try this search in 34+ million citations and abstracts

PMC Full-Text Search Results

Items: 3

- ☐ [An effect of large-scale deletions and duplications on tran](#)
- 1. [Magda Mielczarek, Magdalena Frąszczak, Anna E. Zielak-Jagoda, Pierścińska, Wojciech Kruszyński, Joanna Szyda](#)
Funct Integr Genomics. 2023; 23(1): 19. Published online 2022 Dec
PMCID: PMC9789009
[Article](#) [PDF-1003K](#)
- ☐ [Genetic control of longissimus dorsi muscle gene expres](#)
- 2. [phenotypic quantitative trait loci in pigs](#)
Deborah Velez-Irizarry, Sebastian Casiro, Kaitlyn R. Daza, Ronald O. Bates, Nancy E. Raney, Juan P. Steibel, Catherine W. Ernst
BMC Genomics. 2019; 20: 3. Published online 2019 Jan 3. doi: 10.1186/s12864-018-5386-2
PMCID: PMC6319002
[Article](#) [PDF-1.5M](#)
- ☐ [Mapping splice QTLs reveals distinct transcriptional and post-transcriptional regulatory variation of](#)
- 3. [gene expression and identifies putative alternative splicing variation mediating complex trait variation in pigs](#)
Fei Zhang, Deborah Velez-Irizarry, Catherine W. Ernst, Wen Huang
BMC Genomics. 2023; 24: 240. Published online 2023 May 5. doi: 10.1186/s12864-023-09314-4
PMCID: PMC10161646
[Article](#) [PDF-1.3M](#)

RNA extraction and RNA sequencing

Tissue samples were collected immediately post mortem from the *longissimus dorsi* muscle, flash frozen in liquid nitrogen and stored at -80°C until processed [17]. RNA extraction was performed with the miRNeasy Mini Kit (Qiagen, Germantown, MD) following the manufacturer's protocol. Quality and quantity of extracted total RNA were determined using the Agilent 2100 Bioanalyzer ($\text{RIN} \geq 7$). Sequencing was performed at the Michigan State University Research Technology Support Facility. Libraries for 24 samples were prepared using the Illumina TruSeq RNA Library Prep Kit v2, and sequenced on the Illumina HiSeq 2000 platform (2×100 bp paired-end reads). The remaining 152 libraries were prepared using the Illumina TrueSeq Stranded mRNA Kit, and sequenced on the Illumina HiSeq 2500 platform (2×125 bp, paired-end reads). Base calling was performed with the Illumina Real Time Analysis v1.18.61 software, and the Illumina Bc12fastq v1.8.4 was used for conversion to FastQ format. A total of 96 sequence files (741Gb) consisting of ~ 63 million short-reads per library were obtained from the HiSeq 2000 platform, and 1218 sequence files (~ 2 Tb) of ~ 23 million short-reads per library were obtained from the HiSeq 2500 platform. Eight samples were removed from further analysis due to low sequence quality, leaving a total of 168 samples for subsequent analyses. Sequence data has been deposited in the NCBI Sequence Read Archive accession number PRJNA403969.

Search

See more...

Recent activity

Turn Off Clear

- PRJNA403969 (3) PMC
- PRJNA1084752 (2) SRA
- SRA Links for BioProject (Select 1084752) (2) SRA
- BioSample for BioProject (Select 1084752) (2) BioSample

POBIERANIE DANYCH

Dane z bazy SRA
należy pobrać za
pomocą narzędzia
SRA-toolkit.



SRA

Full

SRX23851466: WGS of Homo sapiens: adult male prostate
1 ILLUMINA (Illumina NovaSeq 6000) run: 511.4M spots, 153.4G bases, 46.6Gb downloads

Design: After the DNA samples pass quality assessment, they undergo random fragmentation using a Covaris ultrasonic disruptor. Subsequent steps involve end repair, A-tailing, ligation of sequencing adapters, fragment size selection, PCR amplification, and purification, culminating in the completion of the sequencing library preparation process. This results in the generation of the final DNA library.

Submitted by: The Second Hospital of Dalian Medical University

Study: Prostate sarcoma genomic sequencing
[PRJNA1084752](#) • [SRP493619](#) • [All experiments](#) • [All runs](#)
[show Abstract](#)

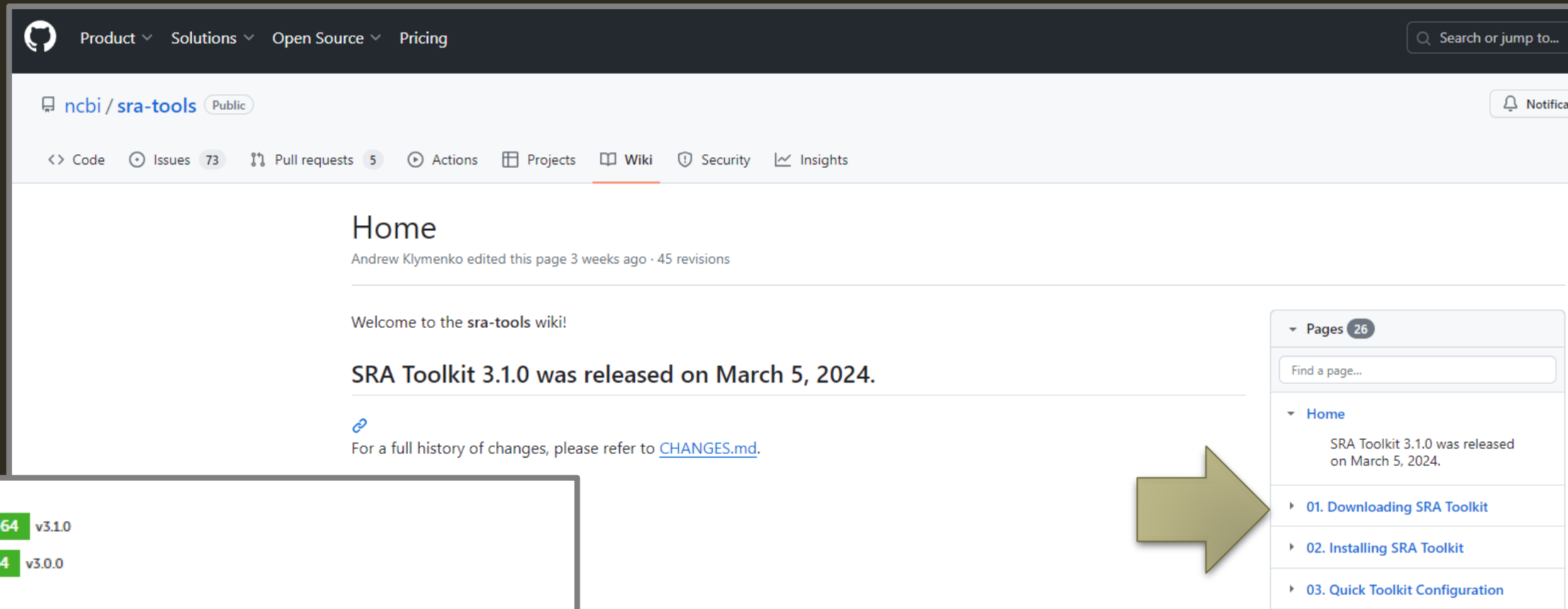
Sample:
[SAMN40279115](#) • [SRS20671551](#) • [All experiments](#) • [All runs](#)
Organism: [Homo sapiens](#)

Library:
Name: Normal
Instrument: Illumina NovaSeq 6000
Strategy: WGS
Source: GENOMIC
Selection: RANDOM
Layout: PAIRED

Runs: 1 run, 511.4M spots, 153.4G bases, [46.6Gb](#)

Run	# of Spots	# of Bases	Size	Published
SRR28240329	511,434,907	153.4G	46.6Gb	2024-03-06

POBIERANIE DANYCH: SRA-TOOLKIT



ncbi / [sra-tools](#) Public

<> Code Issues 73 Pull requests 5 Actions Projects Wiki Security Insights

Home

Andrew Klymenko edited this page 3 weeks ago · 45 revisions

Welcome to the **sra-tools** wiki!

SRA Toolkit 3.1.0 was released on March 5, 2024.

[For a full history of changes, please refer to CHANGES.md.](#)

linux-64 v3.1.0

osx-64 v3.0.0

Pages 26

Find a page...

- Home
 - SRA Toolkit 3.1.0 was released on March 5, 2024.
- 01. Downloading SRA Toolkit
- 02. Installing SRA Toolkit
- 03. Quick Toolkit Configuration

conda install ?

To install this package run one of the following:

```
conda install bioconda::sra-tools
```

```
conda install bioconda/label/cf201901::sra-tools
```

POBIERANIE DANYCH: SRA-TOOLKIT

```
(base) mielczarek@eagle:/mnt/storage_3/home/mielczarek/pl0199-01/scratch/data_sra$ fastq-dump --stdout -X 2 SRR28240329
Read 2 spots for SRR28240329
Written 2 spots for SRR28240329
@SRR28240329.1 A00821:1505:H3LNWDSXC:3:1101:1570:1016 length=300
AAAGAGCGTTTCAAACTTCTCTATGAAAAGAAAGTTCTACTCCTTTAGTTGAGGACACACATCACGAGTAAGTTTCTGAGAATGCTTCTGTCTAGTTTTTATGGGAAGATATTTCTTTTTTCATCATAGGCCTCAAAGCGCT
TGAGACACTCTTTCTGCACTATCTGGAAGTGGACATTTGGAGCGCTTTGAGGCCTATGATGAAAAAGGAAATATCTTCCCATAAAAACTAGACAGAAGCATTCTCAGAACTAACTCGTGATGTGTGTCCTCA
+SRR28240329.1 A00821:1505:H3LNWDSXC:3:1101:1570:1016 length=300
FFFFFFFFFFFF:FFFFFF,FFFFF:F,FFFF:F:F:FFFFFFFFFFFF::,F::FFF:FFF,FF:FF,FF:F:,,:FFF,FFFFFFFF:FFFFFFFF:FFFFF:FFF:F:FFF,:::F:FFFFFFFFFFFF,FFF,FFFFFFFF,F,FF,,
F:FFFFFF:FFFFFFFFFFFF:FFFFFFFFFFFF:FFFFFFFFFFFF:FFFFFFFF:FFFFFFFFFFFF,FFFFFFFFFFFFFFFFFFFF:FFF,FFFFFF,FFFFFFFFFFFF,FFFFF::FFFF,F,,,:FFFF:FFFFFFFFFFFF
@SRR28240329.2 A00821:1505:H3LNWDSXC:3:1101:1588:1016 length=300
AATTATTTCAAGATGGATTAAAACTTAAATGTTAGGCCTAAAACCATAAAAAACCCTAGAAGAAATCCTATGCAATACCATTACAGGGCATAGGCGTGAGTAAGGGCTTCATGTCTAAAACACCAAAAAGCAATGGCAACAAAAGC
TTTGTCAGATGAGTAGATTGCAAAAATTTCTCCATTCTGTAGGTTGCCTGTTCACTCTGATGGTAGTTTCTTTTGCTGTGCAGAAGCTCTTTAGTTTAATTAGATCCCACTTGTCAATTTTGGCTTTTGT
+SRR28240329.2 A00821:1505:H3LNWDSXC:3:1101:1588:1016 length=300
FFFFFFFFFFFF:FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF::FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF:FFFFFFFFFFFFFFFFFFFF:FFFFFFFFFFFF:FFFFFFFFFFFF:FFFFFFFFFFFF:
FF:FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF
(base) mielczarek@eagle:/mnt/storage_3/home/mielczarek/pl0199-01/scratch/data_sra$ fastq-dump --split-files -X 1000000 --gzip SRR28240329
Read 1000000 spots for SRR28240329
Written 1000000 spots for SRR28240329
(base) mielczarek@eagle:/mnt/storage_3/home/mielczarek/pl0199-01/scratch/data_sra$ ls -lpth
total 132M
-rw-r--r-- 1 mielczarek pl0199-01 71M Mar 29 14:20 SRR28240329_2.fastq.gz
-rw-r--r-- 1 mielczarek pl0199-01 70M Mar 29 14:20 SRR28240329_1.fastq.gz
(base) mielczarek@eagle:/mnt/storage_3/home/mielczarek/pl0199-01/scratch/data_sra$ zcat SRR28240329_1.fastq.gz | wc -l
4000000
(base) mielczarek@eagle:/mnt/storage_3/home/mielczarek/pl0199-01/scratch/data_sra$ zcat SRR28240329_2.fastq.gz | wc -l
4000000
```

→ Fastq-dump & fasterq-dump

EUROPEAN NUCLEOTIDE ARCHIVE (ENA)



European Nucleotide Archive

Home Submit Search Rulespace About

We recommend that you subscribe to the [ENA-announce mailing list](#).

Effective September 1st, 2023, our data retrieval APIs will implement optimized and efficient access to our APIs.

Submit Search



European Nucleotide Archive

Home Submit Search Rulespace About Support

Searching ENA

ENA data can be searched and retrieved interactively and programmatically and visualized using the ENA Browser. Please refer to the following sections for more information about the ENA data access functionality with links to more detailed documentation.

Search term:

Uses [this](#) to perform a free text search across ENA data.

Free Text Search

Advanced Search

Cross References

Sequence Similarity Search

Sequence Version Archive

Magda Mielczarek

Advanced Search

[How to use this service](#) [API Documentation](#)

Customise your own search query and retrieve a set of ENA records tailored to your search criteria.

All searches are performed against a subset of the archive specified by the *Data type* you choose to search against. You can then build your search query to specify what data you are looking for and select what fields you want to retrieve from your search. There are additional options to include/exclude specific datasets as well as filter the number of results you wish to return.

If you intend to repeat the same search at a later date, you can save this as a Rule using [Rulespace](#). If you want to access the same data programmatically, you can copy the produced curl request and run this yourself against the [ENA Portal API](#).

1

Data Type

2

Query

3

Inclusion/Exclusion

4

Fields

5

Data Filters

Select datatype to build a new search

Data type:

Nucleotide sequence analyses from reads
Studies used for nucleotide sequence analyses from reads
Genome assemblies
Coding sequences
Non-coding sequences
Experiments used for raw reads
Raw reads
Studies used for raw reads
Samples
Nucleotide sequences
Studies
Taxonomic classification
Targeted locus study contig sets (TLS)
Transcriptome assembly contig sets (TSA)
Genome assembly contig set (WGS)

Use a previously defined Rulespace search

Rule ID/Name:

Reset

read_run

Next



The European Nucleotide Archive (ENA)
The ENA is an ELIXIR Core Data Resource. [Learn more](#)

- Taxonomy and related
- Geographical location
- Geography
- Collection event information
- Sampling information
- Sample state and conditions
- Host information
- Methodology
- Sequencing information



AND OR + Add rule + Add group

Instrument platform = ABI_SOLID, 61879 results ✕ Delete

instrument platform used in sequencing experiment

- ABI_SOLID, 61879 results
- BGISEQ, 110805 results
- CAPILLARY, 355205 results
- COMPLETE_GENOMICS, 7709 results
- DNBSEQ, 88482 results
- element, 36898 results
- genapsys, 9 results
- genemind, 54 results
- HELICOS, 4977 results
- ILLUMINA, 27448398 results**
- ION_TORRENT, 454563 results
- LS454, 414994 results
- OXFORD_NANOPORE, 777734 results
- PACBIO_SMRT, 771428 results
- tapestri, 38 results
- vela_diagnostics, 10 results

AND OR + Add rule + Add group

Library layout = PAIRED, 22897198 results ✕ Delete

sequencing library layout

AND OR + Add rule + Add group

Library source = GENOMIC, 11575760 results ✕ Delete

source material being sequenced

- GENOMIC, 11575760 results
- GENOMIC SINGLE CELL, 116850 results
- METAGENOMIC, 5455858 results
- METATRANSCRIPTOMIC, 140094 results
- OTHER, 259031 results
- SYNTHETIC, 111558 results
- TRANSCRIPTOMIC, 5545423 results
- TRANSCRIPTOMIC SINGLE CELL, 562839 results
- VIRAL RNA, 6765770 results

Text Search

Uses [EBI Search](#) to perform a free text search across ENA data. For more detailed usage please refer to the [help & documentation section](#).

Search term:

[Search](#)

Search results for PRJNA1084752

- **Read**
 - [Experiment](#) (2)
 - [Run](#) (2)
- **Study**
 - [Study](#) (1)

Experiment

Download ENA records: [XML](#) [TSV](#)

Accession	Description/Title
SRX23851465	Illumina NovaSeq 6000 paired end sequencing; Illumina NovaSeq 6000 sequencing: WGS of Homo sapiens: ...
SRX23851466	Illumina NovaSeq 6000 paired end sequencing; Illumina NovaSeq 6000 sequencing: WGS of Homo sapiens: ...

Experiment: SRX23851465

WGS of Homo sapiens: adult male prostate



Organism: [Homo sapiens \(human\)](#)

Instrument Platform: ILLUMINA

Instrument Model: Illumina NovaSeq 6000

Center Name: SUB14263517

Library Layout: PAIRED

Library Strategy: WGS

Library Source: GENOMIC

Library Name: Tumor

Library Selection: RANDOM

Experiment Accession: SRX23851465

Run Accession: SRR28240330

Download report: [JSON](#) [TSV](#)

Get d

Study Accession	Sample Accession	Experiment Accession	Run Accession	Tax Id	Sci Name
PRJNA1084752	SAMN40279114	SRX23851465	SRR28240330	9606	Homo sapiens

Experiment: SRX23851466

WGS of Homo sapiens: adult male prostate



Organism: [Homo sapiens \(human\)](#)

Instrument Platform: ILLUMINA

Instrument Model: Illumina NovaSeq 6000

Center Name: SUB14263517

Library Layout: PAIRED

Library Strategy: WGS

Library Source: GENOMIC

Library Name: Normal

Library Selection: RANDOM

Experiment Accession: SRX23851466

Run Accession: SRR28240329

Study Accession	Sample Accession	Experiment Accession	Run Accession	Tax Id	Scientific Name	Generated FASTQ files: FTP
PRJNA1084752	SAMN40279115	SRX23851466	SRR28240329	9606	Homo sapiens	☐ SRR28240329_1.fastq.gz
						☐ SRR28240329_2.fastq.gz



Download All

[nature](#) > [news](#) > article

NEWS | 30 November 2023

World's biggest set of human genome sequences opens to scientists

The whole genomes of 500,000 people in the UK Biobank will help researchers to probe our genetic code for links to disease.

UK BIOBANK

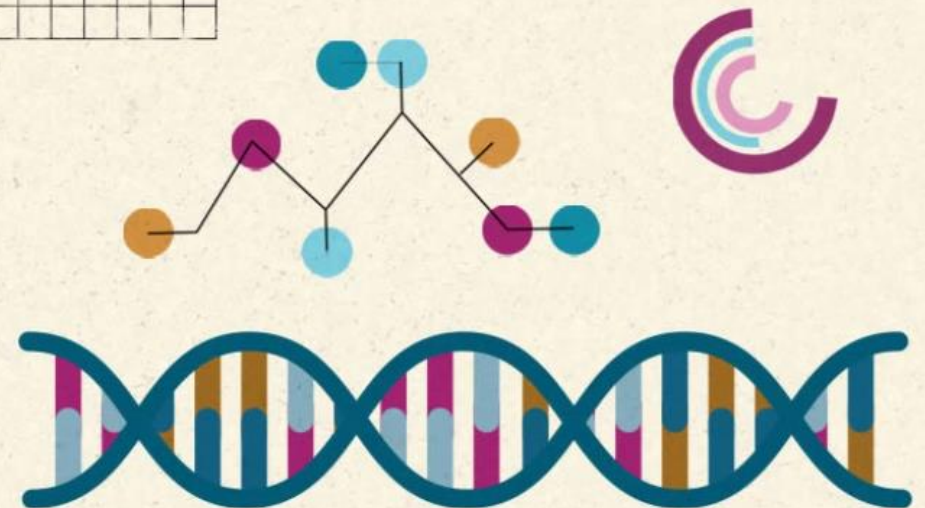
[Researcher log in](#)[Participant log in](#)[Contact us](#)[Enable your research](#)[Explore your participation](#)[Learn more about UK Biobank](#)

500,000 genomes sequenced

We've sequenced the full genetic code of our 500,000 volunteers. Combining this genetic information with the other data in UK Biobank – from health records, to lifestyle information and imaging data – provides researchers with unprecedented insight into how genetics plays a role in our health.

[The story](#)[The data](#)

023
938
34234
5
6777568568
345000234
2344009681



UK BIOBANK



Researcher lo

Enable your res

500,000 genomes sequenced

We've sequenced the full genetic code of our 500,000 volunteers. Combining this genetic information with the other data in UK Biobank – from health records, to lifestyle information and imaging data – provides researchers with unprecedented insight into how genetics plays a role in our health.

The story

The data

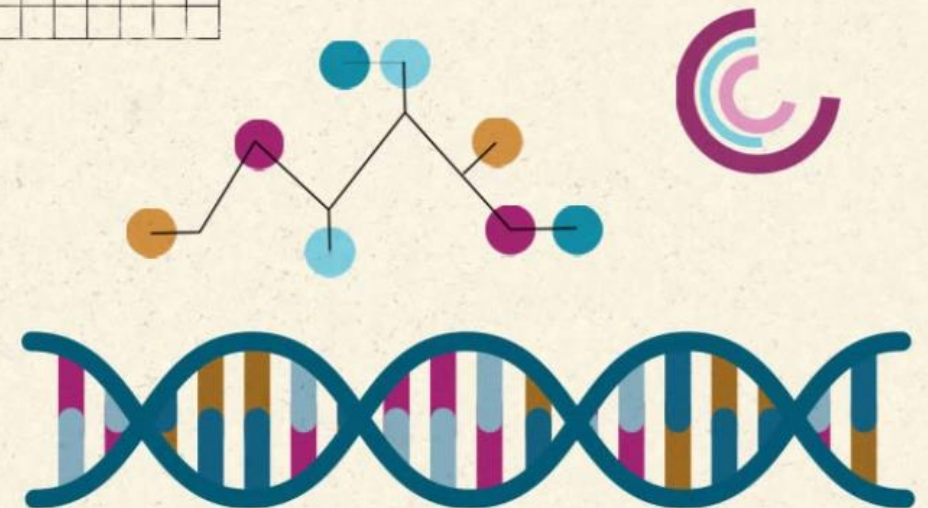
What is whole genome sequencing?

DNA sequencing means reading DNA to determine the order of the four 'letters', called base pairs, which make up DNA (A,T,G and C).

Whole genome sequencing is reading the genetic sequence for a person's *entire* genetic code: the unique genetic code of three billion building blocks that makes each individual who they are.

Both the DNA which forms instructions for producing proteins (the exome, around 2% of the genome) and the 'non-coding' region (the remaining 98%) have been sequenced for each participant.

023
938
34234
5
6777568568
345000234
2344009681



UK BIOBANK

- Obecnie największy na świecie zestaw danych sekwencji ludzkiego genomu → 500 000 osób z brytyjskiego Biobanku
- Badanie kodu genetycznego pod kątem występowania chorób → identyfikacja, które choroby mają podłoże genetyczne, a które są wynikiem stylu życia czy środowiska, w którym żyje dana osoba.
- Zawiera **całe genomy** i **dane fenotypowe**. Niemożliwe jest zidentyfikowanie tożsamości osób badanych
- Koszt: 200 milionów funtów (250 milionów dolarów). Sponsorzy: fundator badań biomedycznych Wellcome, rząd Wielkiej Brytanii i firmy farmaceutyczne, które w zamian uzyskały dostęp do danych na dziewięć miesięcy przed ich szerszym udostępnieniem.
- Obecnie, naukowcy z całego świata mogą ubiegać się o dostęp do danych (dostęp do danych jest płatny).

UK BIOBANK: RARE VARIANTS

- Wcześniej informacje genetyczne brytyjskiego Biobanku obejmowały całe „egzomy” (2% genomu) → brak niekodujących regionów genomu!
- WGS:
 - Pokrywa miejsca niekodujące
 - Pozwala na detekcję bardzo rzadkich mutacji, które zwykle mają silniejszy wpływ na cechę niż typowe zmiany obecne w populacji.
- Przebadano już 200 000 pierwszych kompletnych genomów z Biobanku i znaleziono 29 rzadkich wariantów DNA, które są powiązane z dziedziczeniem wzrostu; wariantów tych nie wykryto w poprzednich badaniach genetycznych.

ALL OF US

All of Us Research Hub

NIH National Institutes of Health
All of Us Research Program

[ABOUT](#)

[DATA & TOOLS](#)

[DISCOVER](#)

[SUPPORT](#)



[REGISTER](#)



[View Data Availability in the Roadmap >>](#)



[Explore Genomic Data in the Data Browser >>](#)

Welcome to the *All of Us* Research Hub

The National Institutes of Health's *All of Us* Research Program is building one of the largest biomedical data resources of its kind. The *All of Us* Research Hub stores health data from participants from across the United States.

Register for the Researcher Workbench to access data and tools to conduct health research and improve understanding of health and disease.

[REGISTER FOR ACCESS](#)



PRZEGLĄDARKA GENOMOWA

Ensembl

PRZEGLĄDARKA GENOMOWA

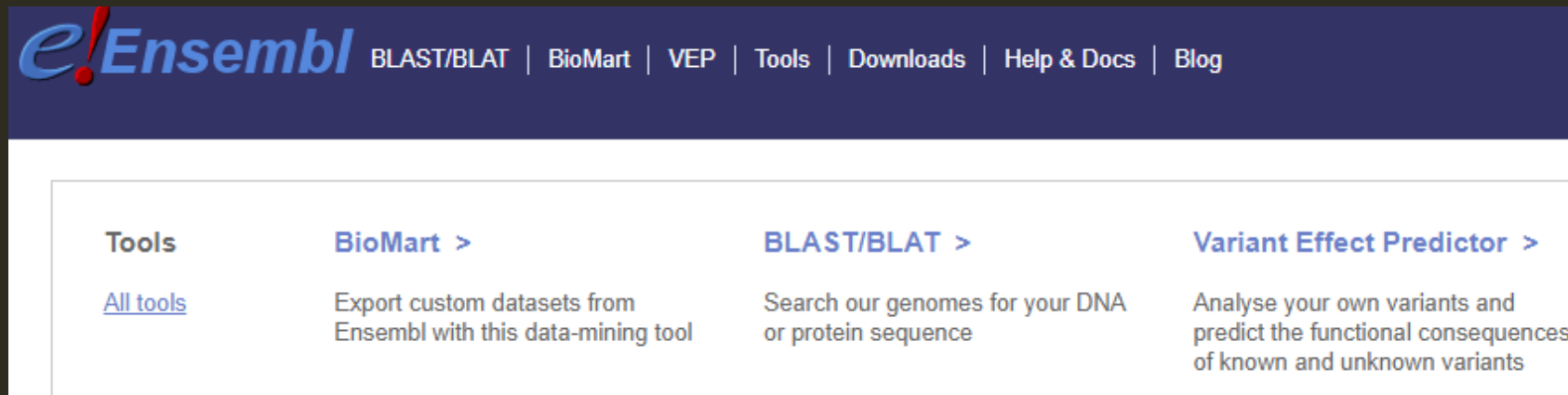
- Czym jest genom?
- Czym jest genom referencyjny?



<https://pl.freepik.com/>

ENSEMBL

- „Przeglądarka” genomów **kręgowców**
- Wspiera badania z zakresu genomiki porównawczej, ewolucji i regulacji ekspresji genów
- Dostępne narzędzia: **BioMart**, Variant Effect Predictor (**VEP**)



ENSEMBL

- Ensembl
- EnsemblGenomes
- EnsemblFungi
- EnsemblMetazoa
- EnsemblProtists
- EnsemblBacteria
- EnsemblPlants



[Caenorhabditis elegans](#)
WBcel235



[Drosophila melanogaster](#)
BDGP6.28



[Bombyx mori](#)
ASM15162v1

[Go to Ensembl Metazoa](#)



[Plasmodium falciparum 3D7](#)
ASM276v2



[Dictyostelium discoideum](#)
dicty_2.7



[Phytophthora infestans](#)
ASM14294v1

[Go to Ensembl Protists](#)

EnsemblFungi



[Magnaporthe oryzae](#)
MG8



[Saccharomyces cerevisiae](#)
R64-1-1



[Aspergillus nidulans](#)
ASM1142v1

[Go to Ensembl Fungi](#)

EnsemblBacteria

[Streptococcus pneumoniae](#)
ASM688v1

[Escherichia coli](#)
ASM584v2

[Bacillus subtilis](#)
ASM73511v1

[Go to Ensembl Bacteria](#)

ENSEMBL

- Ensembl
- EnsemblGenomes
- EnsemblFungi
- EnsemblMetazoa
- EnsemblProtists
- EnsemblBacteria
- EnsemblPlants



ENSEMBL

Identyfikatory (*Homo sapiens*):

- ENS**G** → Gene
- EN**S****T** → Transcript
- EN**S****P** → Protein
- EN**S****E** → Exon

Identyfikatory (*inne gatunki*):

- ENS**F****C****A****T**00000032635 *Felis catus*
- ENS**R****N****O****G**00000050313 *Rattus norvegicus*
- ENS**C****A****F****G**00000022708 *Canis lupus familiaris*
- ENS**B****T****A****T**00000064726 *Bos taurus*

Zagadka:

ENST00000471181.7

? ?

ENSRNOT00000075759.1

? ?

ENSSSCG00000018060

? ?

ENSEMBL

 [BLAST/BLAT](#) | [BioMart](#) | [VEP](#) | [Tools](#) | [Downloads](#) | [Help & Docs](#) | [Blog](#)

Tools
[All tools](#)

BioMart >
Export custom datasets from Ensembl with this data-mining tool

BLAST/BLAT >
Search our genomes for your DNA or protein sequence

Variant Effect Predictor >
Analyse your own variants and predict the functional consequences of known and unknown variants

Search

All species ▼

for

Go

e.g. BRCA2 or rat 5:62797383-63627669 or rs699 or coronary heart disease

All genomes

-- Select a species -- ▼

- [View full list of all Ensembl species](#)
- [Edit your favourites](#)

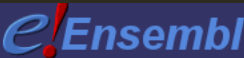
Favourite genomes

**Human**
GRCh38.p12
[Still using GRCh37?](#)

**Mouse**
GRCm38.p6

**Zebrafish**
GRCz11

ENSEMBL

 [BLAST/BLAT](#) | [BioMart](#) | [VEP](#) | [Tools](#) | [Downloads](#) | [Help & Docs](#) | [Blog](#)

Tools
[All tools](#)

BioMart >
Export custom datasets from Ensembl with this data-mining tool

BLAST/BLAT >
Search our genomes for your DNA or protein sequence

Variant Effect Predictor >
Analyse your own variants and predict the functional consequences of known and unknown variants



Search

Cow

▼ for

19:12,571,690-12,571,788

Go

e.g. [BRCA2](#) or [rat 5:62797383-63627669](#) or [rs699](#) or [coronary heart disease](#)

All genomes

-- Select a species --

▼

- [View full list of all Ensembl species](#)
- [Edit your favourites](#)

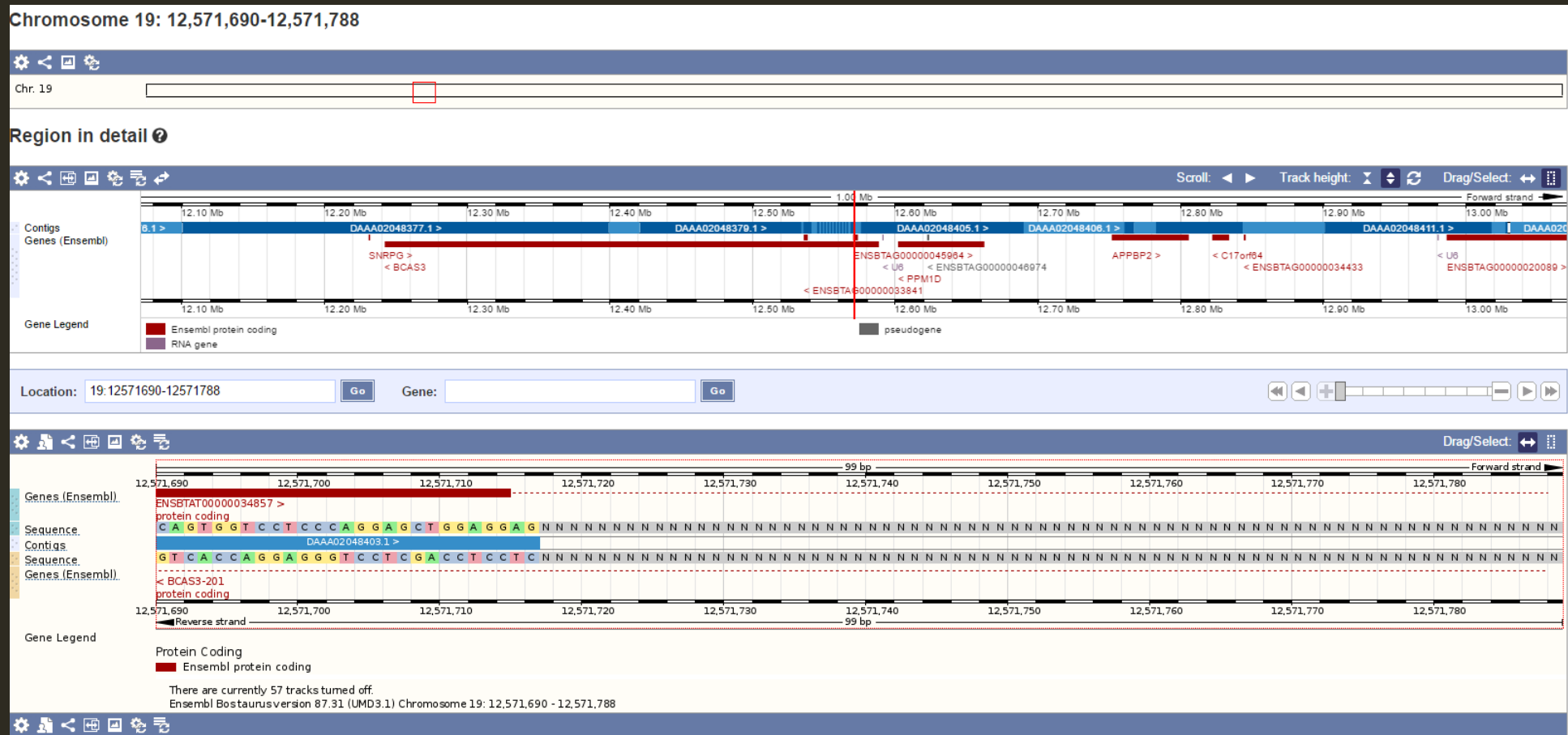
Favourite genomes

 **Human**
GRCh38.p12
[Still using GRCh37?](#)

 **Mouse**
GRCm38.p6

 **Zebrafish**
GRCz11

ENSEMBL



ENSEMBL

 [BLAST/BLAT](#) | [BioMart](#) | [VEP](#) | [Tools](#) | [Downloads](#) | [Help & Docs](#) | [Blog](#)

[Tools](#)
[All tools](#)

[BioMart >](#)
Export custom datasets from Ensembl with this data mining tool

[BLAST/BLAT >](#)
Search our genomes for your DNA or protein sequence

[Variant Effect Predictor >](#)
Analyse your own variants and predict the functional consequences of known and unknown variants

Search

Human ▼

for
BRCA1 [Go](#)

e.g. [BRCA2](#) or [rat 5:62797383-63627669](#) or [rs699](#) or [coronary heart disease](#)

All genomes

-- Select a species -- ▼

- [View full list of all Ensembl species](#)
- [Edit your favourites](#)

Favourite genomes

 **Human**
GRCh38.p12
[Still using GRCh37?](#)

 **Mouse**
GRCm38.p6

 **Zebrafish**
GRCz11

ENSEMBL

Na którym chromosomie leży gen BRCA1?
Ile ma form splicingowych?
Który transkrypt jest najdłuższy?

Gene: BRCA1 ENSG00000012048

Description

BRCA1 DNA repair associated [Source:HGNC Symbol;Acc:[HGNC:1100](#)]

Gene Synonyms

BRCC1, FANCS, PPP1R53, RNF53

Location

[Chromosome 17: 43,044,295-43,170,245](#) reverse strand.
GRCh38:CM000679.2

About this gene

This gene has 40 transcripts ([splice variants](#)), [196 orthologues](#) and is associated with [111 phenotypes](#).

Transcripts

Hide transcript table

Show/hide columns (1 hidden)

Transcript ID	Name	bp	Protein	Biotype	CCDS	UniProt Match	RefSeq Match
ENST00000357654.9	BRCA1-203	7088	1863aa	Protein coding	CCDS11453	P38398-1	NM_007294.4
ENST00000618469.2	BRCA1-226	7613	1863aa	Protein coding	CCDS11453	G1UI37	-
ENST00000652672.2	BRCA1-232	7320	1816aa	Protein coding	CCDS11459	A0A494C182	-
ENST00000471181.7	BRCA1-209	7270	1884aa	Protein coding	CCDS11456	P38398-7	-
ENST00000494123.6	BRCA1-221	7168	1863aa	Protein coding	CCDS11453	Q3B891	-
ENST00000470026.6	BRCA1-208	7147	1863aa	Protein coding	CCDS11453	E7EWN5	-

ENSEMBL (DOWNLOADS → FTP DOWNLOAD)

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/BigWig
Y	Human <i>Homo sapiens</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	GVF	VCF	VEP	Regulation (GFF)	Regulation data files	BAM/BigWig
Y	Mouse <i>Mus musculus</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	GVF	VCF	VEP	Regulation (GFF)	Regulation data files	BAM/BigWig
Y	Zebrafish <i>Danio rerio</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	GVF	VCF	VEP	-	-	BAM/BigWig
	Algerian mouse <i>Mus spretus</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	-	-	VEP	-	-	-
	Alpaca <i>Vicugna pacos</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	-	-	VEP	-	-	-
	Amazon molly <i>Poecilia formosa</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	-	-	VEP	-	-	BAM/BigWig
	Angola colobus <i>Colobus angolensis palliatus</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	-	-	VEP	-	-	-
	Anole lizard <i>Anolis carolinensis</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	-	-	VEP	-	-	BAM/BigWig

ENSEMBL

Show 10 entries		Show/hide columns												mus		
★	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/BigWig
	Mouse <i>Mus musculus</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	GVF	VCF	VEP	Regulation (GFF)	Regulation data files	BAM/BigWig
	Algerian mouse <i>Mus spretus</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	-	-	VEP	-	-	-
	Ferret <i>Mustela putorius furo</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	-	-	VEP	-	-	BAM/BigWig
	Mouse 129S1/SvImJ <i>Mus musculus</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	-	-	VEP	-	-	-
	Mouse A/J <i>Mus musculus</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	-	-	VEP	-	-	-
	Mouse AKR/J <i>Mus musculus</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	-	-	VEP	-	-	-
	Mouse BALB/cJ <i>Mus musculus</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	-	-	VEP	-	-	-
	Mouse C3H/HeJ	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	-	-	VEP	-	-	-

ENSEMBL

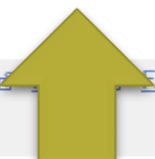
[illegible]

Index of /pub/release-110/fasta/mus_musculus/dna

	<u>Name</u>	<u>Last modified</u>	<u>Size</u>	<u>Description</u>
	Parent Directory		-	
	CHECKSUMS	2023-05-02 09:13	4.2K	
	Mus_musculus.GRCm39.dna.chromosome.1.fa.gz	2023-04-21 17:21	56M	
	Mus_musculus.GRCm39.dna.chromosome.2.fa.gz	2023-04-21 17:21	52M	
	156.17.191.216 - PuTTY			
	N>1 dna:chromosome chromosome:GRCm38:1:1:195471971:1 REF			
	N			
	N			
	N			
	Mus_musculus.GRCm39.dna.chromosome.7.fa.gz	2023-04-21 17:21	41M	
	Mus_musculus.GRCm39.dna.chromosome.8.fa.gz	2023-04-21 17:21	36M	
	Mus_musculus.GRCm39.dna.chromosome.9.fa.gz	2023-04-21 17:21	35M	
	Mus_musculus.GRCm39.dna.chromosome.10.fa.gz	2023-04-21 17:21	37M	

ENSEMBL

Show	10	entries	Show/hide columns	mus												
★	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/BigWig
Y	Mouse <i>Mus musculus</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	GVF	VCF	VEP	Regulation (GFF)	Regulation data files	BAM/BigWig
	Algerian mouse <i>Mus spretus</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	-	-	VEP	-	-	-
	Ferret <i>Mustela putorius furo</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	-	-	VEP	-	-	BAM/BigWig
	Mouse	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF	MySQL	-	-	VEP	-	-	-



```

156.17.191.216 - PuTTY
>ENSMUST00000196221.1 cdna chromosome:GRCm38:14:54113468:54113476:1 gene:ENSMUSG00000096749.2 gene_biotype:TR_D_gene trans
cript_biotype:TR_D_gene gene_symbol:Trdd1 description:T cell receptor delta diversity 1 [Source:MGI Symbol;Acc:MGI:4439547
]
ATGGCATAT
>ENSMUST00000179664.1 cdna chromosome:GRCm38:14:54113468:54113478:1 gene:ENSMUSG00000096749.2 gene_biotype:TR_D_gene trans
cript_biotype:processed_transcript gene_symbol:Trdd1 description:T cell receptor delta diversity 1 [Source:MGI Symbol;Acc:
MGI:4439547]
ATGGCATATCA
>ENSMUST00000177564.1 cdna chromosome:GRCm38:14:54122226:54122241:1 gene:ENSMUSG00000096176.1 gene_biotype:TR_D_gene trans
cript_biotype:TR_D_gene gene_symbol:Trdd2 description:T cell receptor delta diversity 2 [Source:MGI Symbol;Acc:MGI:4439546
]
ATCGGAGGGATACGAG

```


ENSEMBL

Show	10	entries	Show/hide columns													mus	
★	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/BigWig	
Y	Mouse <i>Mus musculus</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	GVF	VCF	VEP	Regulation (GFF)	Regulation data files	BAM/BigWig	
	Algerian mouse <i>Mus spretus</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	-	-	VEP	-	-	-	
	Ferret <i>Mustela putorius furo</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	-	-	VEP	-	-	BAM/BigWig	
	Mouse	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF	MySQL	-	-	VEP	-	-	-	



```

156.17.191.216 - PuTTY
>ENSMUST00000196221.1 cds chromosome:GRCm38:14:54113468:54113476:1 gene:ENSMUSG00000096749.2 gene_biotype:TR_D_gene transcr
ipt_biotype:TR_D_gene gene_symbol:Trdd1 description:T cell receptor delta diversity 1 [Source:MGI Symbol;Acc:MGI:4439547]
ATGGCATAT
>ENSMUST00000177564.1 cds chromosome:GRCm38:14:54122226:54122241:1 gene:ENSMUSG00000096176.1 gene_biotype:TR_D_gene transcr
ipt_biotype:TR_D_gene gene_symbol:Trdd2 description:T cell receptor delta diversity 2 [Source:MGI Symbol;Acc:MGI:4439546]
ATCGGAGGGGATACGAG
>ENSMUST00000178537.1 cds chromosome:GRCm38:6:41533201:41533212:1 gene:ENSMUSG00000095668.1 gene_biotype:TR_D_gene transcr
ipt_biotype:TR_D_gene gene_symbol:Trbd1 description:T cell receptor beta, D region 1 [Source:MGI Symbol;Acc:MGI:4439571]
GGGACAGGGGGC
>ENSMUST00000178862.1 cds chromosome:GRCm38:6:41542163:41542176:1 gene:ENSMUSG00000094569.1 gene_biotype:TR_D_gene transcr
ipt_biotype:TR_D_gene gene_symbol:Trbd2 description:T cell receptor beta, D region 2 [Source:MGI Symbol;Acc:MGI:4439727]
GGGACTGGGGGGGC
  
```

ENSEMBL

Show	10	entries	Show/hide columns														mus	
★	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/BigWig		
Y	Mouse <i>Mus musculus</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	GVF	VCF	VEP	Regulation (GFF)	Regulation data files	BAM/BigWig		
	Algerian mouse <i>Mus spretus</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	-	-	VEP	-	-	-		
	Ferret <i>Mustela putorius furo</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	-	-	VEP	-	-	BAM/BigWig		
	Mouse <i>129/SvEv</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	-	-	VEP	-	-	-		



```

156.17.191.216 - PuTTY
>ENSMUST00000083969.1 ncRNA chromosome:GRCm38:X:64800511:64800616:1 gene:ENSMUSG00000065903.1 gene_biotype:snRNA transcript_biotype:snRNA gene_symbol:Gm26111 description:predicted gene, 26111 [Source:MGI Symbol;Acc:MGI:5455888]
GTGCTCACTTCCGCAGCACATACTAAAATTGGAACAAAACAGAGAAGGTTAGCATGGC
CCTTGTACAAGTATGACATGCACGTTTCATGAAGCTTCCATATTTTT
>ENSMUST000000157441.1 ncRNA chromosome:GRCm38:2:71495271:71495374:1 gene:ENSMUSG00000088066.1 gene_biotype:snRNA transcript_biotype:snRNA gene_symbol:Gm23253 description:predicted gene, 23253 [Source:MGI Symbol;Acc:MGI:5453030]
GCTGTGCTCACTTCAGCAGCAAACAGTAAAATCGGAATGAGCAGAGGTTAGTATGACCTT
TGTACAAGGGTGACACAGAAAATTGTGAGGTGCTCCATTAAAAA
>ENSMUST000000178877.1 ncRNA chromosome:GRCm38:16:14807163:14807269:-1 gene:ENSMUSG00000095165.1 gene_biotype:snRNA transcript_biotype:snRNA gene_symbol:Gm25547 description:predicted gene, 25547 [Source:MGI Symbol;Acc:MGI:5455324]
GTGCTCGCTTCGGCAGCACATATACTAAAATTGGAACGATACAGAGAAGATTAGCATGGC
CCCTGCGCAAGGATGACACGCAAATTCGTGAAGCGTTCATATTTTT
  
```

ENSEMBL

Show	10	entries	Show/hide columns												mus		
★	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/BigWig	
Y	Mouse <i>Mus musculus</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	GVF	VCF	VEP	Regulation (GFF)	Regulation data files	BAM/BigWig	
	Algerian mouse <i>Mus spretus</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	-	-	VEP	-	-	-	
	Ferret <i>Mustela putorius furo</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	-	-	VEP	-	-	BAM/BigWig	
	Mouse 129S1/SvJmJ	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	-	-	VEP	-	-	-	



```

156.17.191.216 - PuTTY
>ENSMUSP00000142546.1 pep chromosome:GRCm38:14:54113468:54113476:1 gene:ENSMUSG00000096749.2 transcript:ENSMUST00000196221
.1 gene_biotype:TR_D_gene transcript_biotype:TR_D_gene gene_symbol:Trddl description:T cell receptor delta diversity 1 [Source:MGI Symbol;Acc:MGI:4439547]
MAY
>ENSMUSP00000142955.1 pep chromosome:GRCm38:14:54122226:54122241:1 gene:ENSMUSG00000096176.1 transcript:ENSMUST00000177564
.1 gene_biotype:TR_D_gene transcript_biotype:TR_D_gene gene_symbol:Trdd2 description:T cell receptor delta diversity 2 [Source:MGI Symbol;Acc:MGI:4439546]
IGGIR
>ENSMUSP00000141764.1 pep chromosome:GRCm38:6:41533201:41533212:1 gene:ENSMUSG00000095668.1 transcript:ENSMUST00000178537.
1 gene_biotype:TR_D_gene transcript_biotype:TR_D_gene gene_symbol:Trbd1 description:T cell receptor beta, D region 1 [Source:MGI Symbol;Acc:MGI:4439571]
GTGG
  
```

ENSEMBL

More extensive sequence annotation by means of feature tables and contain thus the genome sequence as annotated by the automated Ensembl genome annotation pipeline. Each nucleotide sequence record in a flat file represents a 1 Mb slice of the genome sequence.

Show	10	entries	Show/hide columns	mus												
★	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/BigWig
Y	Mouse <i>Mus musculus</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	GVF	VCF	VEP	Regulation (GFF)	Regulation data files	BAM/BigWig
	Algerian mouse <i>Mus spretus</i>	FASTA	FASTA	FASTA	FASTA	FASTA		GenBank	GTF GFF3	MySQL	-	-	VEP	-	-	-
	Ferret <i>Mustela putorius furo</i>	FASTA	FASTA	FASTA	FASTA	FASTA										
	Mouse 129S1/SvImJ <i>Mus musculus</i>	FASTA	FASTA	FASTA	FASTA	FASTA										
	Mouse A/J <i>Mus musculus</i>	FASTA	FASTA	FASTA	FASTA	FASTA										
	Mouse AKR/J <i>Mus musculus</i>	FASTA	FASTA	FASTA	FASTA	FASTA										
	Mouse BALB/cJ <i>Mus musculus</i>	FASTA	FASTA	FASTA	FASTA	FASTA										
	Mouse C3H/HeJ <i>Mus musculus</i>	FASTA	FASTA	FASTA	FASTA	FASTA										

156.17.191.216 - PuTTY
LOCUS 1 195471971 bp DNA HTG 9-MAR-2018
DEFINITION Mus musculus chromosome 1 GRCm38 full sequence 1..195471971 reannotated via Ensembl
ACCESSION chromosome:GRCm38:1:1:195471971:1
VERSION 1GRCm38
KEYWORDS .
SOURCE house mouse
ORGANISM Mus musculus
Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria;
Deuterostomia; Chordata; Craniata; Vertebrata; Gnathostomata;
Teleostomi; Euteleostomi; Sarcopterygii; Dipnotetrapodomorpha;
Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Boreoeutheria;
Euarchontoglires; Glires; Rodentia; Myomorpha; Muroidea; Muridae.
COMMENT This sequence was annotated by Ensembl(www.ensembl.org). Please visit the Ensembl or EnsemblGenomes web site, http://www.ensembl.org/ or http://www.ensemblgenomes.org/ for more information.

ENSEMBL

GTF → Gene sets for each species. These files include annotations of both coding and non-coding genes.

GFF3 → provides access to all annotated transcripts which make up an Ensembl gene set.

Show	10	entries	Show/hide columns													mus	
★	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/BigWig	
Y	Mouse <i>Mus musculus</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	GVF	VCF	VEP	Regulation (GFF)	Regulation data files	BAM/BigWig	
	Algerian mouse <i>Mus spretus</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	-	-	VEP	-	-	-	
	Ferret <i>Mustela putorius furo</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	-	-	VEP	-	-	BAM/BigWig	
	Mouse 129S1/SvImJ <i>Mus musculus</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	-	-	VEP	-	-	-	
	Mouse A/J <i>Mus musculus</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	-	-	VEP	-	-	-	
	Mouse AKR/J <i>Mus musculus</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	-	-	VEP	-	-	-	
	Mouse BALB/cJ <i>Mus musculus</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	-	-	VEP	-	-	-	
	Mouse C3H/HeJ <i>Mus musculus</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	-	-	VEP	-	-	-	

GTF → Gene sets for each species. These files include annotations of both coding and non-coding genes

ENSEMBL

```
156.17.191.216 - PuTTY
#!genome-build GRCh38.p6
#!genome-version GRCh38
#!genome-date 2012-01
#!genome-build-accession NCBI:GCA_000001635.8
#!genebuild-last-updated 2018-01
1       havana  gene      3073253 3074322 .      +      .      gene_id "ENSMUSG00000102693"; gene_version "1"; gene_name
"4933401J01Rik"; gene_source "havana"; gene_biotype "TEC";
1       havana  transcript 3073253 3074322 .      +      .      gene_id "ENSMUSG00000102693"; gene_version "1"; tr
anscript_id "ENSMUST00000193812"; transcript_version "1"; gene_name "4933401J01Rik"; gene_source "havana"; gene_biotype "T
EC"; transcript_name "4933401J01Rik-201"; transcript_source "havana"; transcript_biotype "TEC"; tag "basic"; transcript_su
pport_level "NA";
1       havana  exon      3073253 3074322 .      +      .      gene_id "ENSMUSG00000102693"; gene_version "1"; transcript
_id "ENSMUST00000193812"; transcript_version "1"; exon_number "1"; gene_name "4933401J01Rik"; gene_source "havana"; gene_b
iotype "TEC"; transcript_name "4933401J01Rik-201"; transcript_source "havana"; transcript_biotype "TEC"; exon_id "ENSMUSE0
0001343744"; exon_version "1"; tag "basic"; transcript_support_level "NA";
1       ensembl gene      3102016 3102125 .      +      .      gene_id "ENSMUSG00000064842"; gene_version "1"; gene_name
"Gm26206"; gene_source "ensembl"; gene_biotype "snRNA";
1       ensembl transcript 3102016 3102125 .      +      .      gene_id "ENSMUSG00000064842"; gene_version "1"; tr
anscript_id "ENSMUST00000082908"; transcript_version "1"; gene_name "Gm26206"; gene_source "ensembl"; gene_biotype "snRNA"
; transcript_name "Gm26206-201"; transcript_source "ensembl"; transcript_biotype "snRNA"; tag "basic"; transcript_support_
level "NA";
1       ensembl exon      3102016 3102125 .      +      .      gene_id "ENSMUSG00000064842"; gene_version "1"; transcript
_id "ENSMUST00000082908"; transcript_version "1"; exon_number "1"; gene_name "Gm26206"; gene_source "ensembl"; gene_biotyp
e "snRNA"; transcript_name "Gm26206-201"; transcript_source "ensembl"; transcript_biotype "snRNA"; exon_id "ENSMUSE0000052
2066"; exon_version "1"; tag "basic"; transcript_support_level "NA";
```

GFF3 → provides access to all annotated transcripts which make up an Ensembl gene set

ENSEMBL

```
156.17.191.216 - PuTTY
##gff-version 3
##sequence-region 1 1 195471971
#!genome-build Ensembl GRCm38.p6
#!genome-version GRCm38
#!genome-date 2012-01
#!genome-build-accession NCBI:GCA_000001635.8
#!genebuild-last-updated 2018-01

1 ensembl_havana gene 3205901 3671498 . - . ID=gene:ENSMUSG00000051951;Name=Xkr4;biotype=protein;description=X-linked Kx blood group related 4 [Source:MGI Symbol%3BAcc:MGI:3528744];gene_id=ENSMUSG00000051951;logically_name=ensembl_havana_gene;version=5
1 havana lnc_RNA 3205901 3216344 . - . ID=transcript:ENSMUST00000162897;Parent=gene:ENSMUSG00000051951;Name=Xkr4-203;biotype=processed_transcript;transcript_id=ENSMUST00000162897;transcript_support_level=1;version=1
1 havana exon 3205901 3207317 . - . Parent=transcript:ENSMUST00000162897;Name=ENSMUSE00000866652;constitutive=0;ensembl_end_phase=-1;ensembl_phase=-1;exon_id=ENSMUSE00000866652;rank=2;version=1
1 havana exon 3213609 3216344 . - . Parent=transcript:ENSMUST00000162897;Name=ENSMUSE00000858910;constitutive=0;ensembl_end_phase=-1;ensembl_phase=-1;exon_id=ENSMUSE00000858910;rank=1;version=1
1 havana lnc_RNA 3206523 3215632 . - . ID=transcript:ENSMUST00000159265;Parent=gene:ENSMUSG00000051951;Name=Xkr4-202;biotype=processed_transcript;transcript_id=ENSMUST00000159265;transcript_support_level=1;version=1
1 havana exon 3206523 3207317 . - . Parent=transcript:ENSMUST00000159265;Name=ENSMUSE00000867897;constitutive=0;ensembl_end_phase=-1;ensembl_phase=-1;exon_id=ENSMUSE00000867897;rank=2;version=1
1 havana exon 3213439 3215632 . - . Parent=transcript:ENSMUST00000159265;Name=ENSMUSE00000863980;constitutive=0;ensembl_end_phase=-1;ensembl_phase=-1;exon_id=ENSMUSE00000863980;rank=1;version=1
1 ensembl_havana mRNA 3214482 3671498 . - . ID=transcript:ENSMUST00000070533;Parent=gene:ENSMUSG00000051951;Name=Xkr4;biotype=mRNA;transcript_id=ENSMUST00000070533;transcript_support_level=1;version=1
```


ENSEMBL

Show 10 entries		Show/hide columns														mus	
★	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/BigWig	
Y	Mouse <i>Mus musculus</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	GVF	VCF	VEP	Regulation (GFF)	Regulation data files	BAM/BigWig	
	Algerian mouse <i>Mus spretus</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	-		VEP	-	-	-	
	Ferret <i>Mustela putorius furo</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	-	-	VEP	-	-	BAM/BigWig	
	Mouse 129S1/SvImJ <i>Mus musculus</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	-	-	VEP	-	-	-	
	Mouse A/J <i>Mus musculus</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	-	-	VEP	-	-	-	
	Mouse AKR/J <i>Mus musculus</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	-	-	VEP	-	-	-	
	Mouse BALB/cJ <i>Mus musculus</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	-	-	VEP	-	-	-	
	Mouse C3H/HeJ <i>Mus musculus</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	-	-	VEP	-	-	-	