



# Лекция № 6

## Геномные базы данных Геномные браузеры

Разработал профессор П. М. Васильев  
Кафедра фармакологии и биоинформатики

Для студентов, обучающихся по направлению 06.03.01 «Биология»  
профили Биохимия, Генетика  
при изучении дисциплины «Биоинформатика»

# П л а н л е к ц и и

- **Геномные базы данных**
- **Геномные браузеры**
- **Аннотация генома**
- **Распознавание структурно-функциональных мотивов в генетических текстах**

# Геномные базы данных

## GenBank

The screenshot shows the NCBI GenBank website. At the top, there is a navigation bar with "NCBI Resources" and "How To" menus, and a "Sign in to NCBI" link. Below this is a search bar with "GenBank" selected and a "Nucleotide" dropdown menu. A "Search" button is to the right of the search bar. Below the search bar is a horizontal menu with various database categories: GenBank, Submit, Genomes, WGS, Metagenomes, TPA, TSA, INSDC, and Other. The main content area is divided into two columns. The left column contains a "GenBank Overview" section with a sub-heading "What is GenBank?". This section includes a paragraph explaining that GenBank is the NIH genetic sequence database, an annotated collection of all publicly available DNA sequences. It also mentions that GenBank is part of the International Nucleotide Sequence Database Collaboration, which includes the DNA DataBank of Japan (DDBJ), the European Nucleotide Archive (ENA), and GenBank at NCBI. A second paragraph states that a GenBank release occurs every two months and provides detailed information about the release and notifications of upcoming changes. A third paragraph mentions that GenBank growth statistics are available from each release. A fourth paragraph provides an example of an annotated sample GenBank record for a *Saccharomyces cerevisiae* gene. The right column contains a "GenBank Resources" section with links to "GenBank Home", "Submission Types", "Submission Tools", "Search GenBank", and "Update GenBank Records". Below the "What is GenBank?" section is an "Access to GenBank" section with a paragraph stating that there are several ways to search and retrieve data from GenBank. This is followed by a bulleted list of search methods: searching for sequence identifiers and annotations with Entrez Nucleotide, searching and aligning GenBank sequences to a query sequence using BLAST, and downloading sequences programatically using NCBI e-utilities. The list also mentions that ASN.1 and flatfile formats are available at NCBI's anonymous FTP server.

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

copying, or distribution of the information contained in GenBank.

### Confidentiality

Some authors are concerned that the appearance of their data in GenBank prior to publication will compromise their work. GenBank will, upon request, withhold release of new submissions for a specified period of time. A date must be specified; we can not hold a sequence indefinitely pending publication. However, if a paper citing the sequence or accession number is published prior to the specified date, the sequence will be released upon publication. In order to prevent the delay in the appearance of published sequence data, we urge authors to inform us of the appearance of the published data. As soon as it is available, please send the full publication data--all authors, title, journal, volume, pages and date--to the following address: [update@ncbi.nlm.nih.gov](mailto:update@ncbi.nlm.nih.gov)

# Геномные базы данных

## 1000 Genomes

IGSR: The International Genome Sample Resource

Providing ongoing support for the 1000 Genomes Project data



[Home](#) [About](#) [Data](#) [Portal](#) [Analysis](#) [Contact](#) [Browser](#) [FAQ](#)

Search 1000genomes



### IGSR and the 1000 Genomes Project



### Links

- [Announcements](#)
- [IGSR Sample Collection Principles](#)
- [1000 Genomes Project Publications](#)
- [File formats](#)
- [Software tools](#)
- [Download data](#)
- [Twitter](#)

<http://www.internationalgenome.org/>

Populations: ● - African; ● - American; ● - East Asian; ● - European; ● - South Asian;

The International Genome Sample Resource (IGSR) was established to ensure the ongoing usability of data generated by the 1000 Genomes Project and to extend the data set. More information is available [about the IGSR](#).

# Геномные базы данных

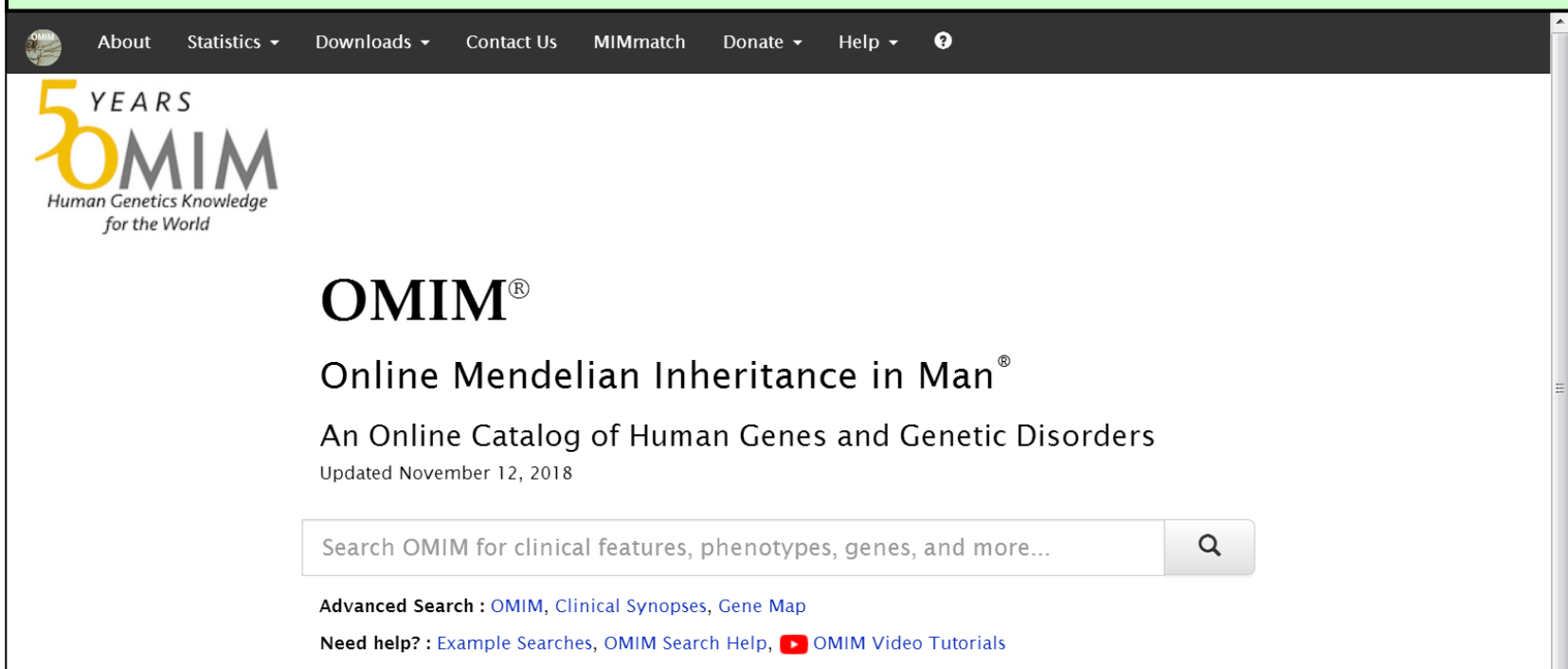
## The GDB Human Genome Database

The screenshot shows the homepage of the GDB Human Genome Database. At the top center, there are three blue circles containing the letters 'G', 'D', and 'B'. Below this is the title 'The GDB Human Genome Database' and a subtitle 'The Official World-Wide Database for the Annotation of the Human Genome'. It mentions it is hosted by RTI International, North Carolina USA, and available at mirror sites worldwide. There are links for 'New Database Information' and 'Log in'. A search section is visible with 'Simple Search' and 'Advanced Search' options. The search criteria include 'Genomic Segments', 'All Biological Data', 'People', 'Citations', 'Name/GDB ID', and 'DNA Sequence ID'. There are 'Submit' and 'Reset' buttons. A note states: 'Note: For Name/ID searches, adding \* to the end of the search text may improve results.' Below the search section, there are links for 'Preferences', 'Other Search Options', 'Submit and Edit Data', 'How to submit data, get D seg #s NEW!', and 'Insert Object'. On the left side, there is a navigation menu with links for 'Help', 'Site Map', 'Reports', 'Resources', 'GDB Operations', 'CF Mutations', 'HUGO Chromosomes', and 'GDB Redevelopment Survey'.

<http://morissardjerome.free.fr/infobiogen/>  
[www.gdb.org/gdb/](http://www.gdb.org/gdb/)

# Геномные базы данных

## Online Mendelian Inheritance in Man




The screenshot shows the homepage of the Online Mendelian Inheritance in Man (OMIM) website. At the top, there is a navigation bar with links for 'About', 'Statistics', 'Downloads', 'Contact Us', 'MIMmatch', 'Donate', and 'Help'. Below the navigation bar is a logo celebrating '5 YEARS OMIM' with the tagline 'Human Genetics Knowledge for the World'. The main heading is 'OMIM® Online Mendelian Inheritance in Man®', followed by the subtitle 'An Online Catalog of Human Genes and Genetic Disorders' and the update date 'Updated November 12, 2018'. A search bar is present with the placeholder text 'Search OMIM for clinical features, phenotypes, genes, and more...'. Below the search bar, there are links for 'Advanced Search : OMIM, Clinical Synopses, Gene Map' and 'Need help? : Example Searches, OMIM Search Help, OMIM Video Tutorials'.

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

Описание генетических заболеваний и генов, их вызывающих  
Фенотипические проявления генетических синдромов  
Генетические карты

# Геномные базы данных

## Human Gene Mutation Database



**The Human Gene Mutation Database**  
at the Institute of Medical Genetics in Cardiff

Home Search help Statistics New genes What is new Background Publications Contact Register Login LSDBs Other links

Gene symbol  Go! Symbol:  Missense/nonsense  Go!

The Human Gene Mutation Database (HGMD®) represents an attempt to collate all known (published) gene lesions responsible for human inherited disease and is maintained in Cardiff by D.N. Cooper, E.V. Ball, P.D. Stenson, A.D. Phillips, K. Evans, S. Heywood, M.J. Hayden, M.M. Chapman, M.E. Mort, L. Azevedo and M. Mort

**Get HGMD Professional** \*Please note that this less up-to-date public version of our database is freely available only to [registered](#) users from academic institutions/non-profit organisations. All commercial users are required to purchase a license from QIAGEN®, our commercial partner. A license to [HGMD Professional](#) is available to both commercial and academic non-profit users wishing to access the most up-to-date version of the database (visit QIAGEN® to request a [free trial](#) of HGMD Professional). Read more about how HGMD is [funded](#). You may not copy, store or re-distribute HGMD data without express written permission (i) from the curators or (ii) via your license agreement. Copyright © Cardiff University 2017. All rights reserved. [Register for Public Version](#)

Table:	Description:	Public entries: <small>This site. Academic non-profit users only</small>	Total entries: <small>HGMD Professional 2018.3</small>
<b>Mutation totals (as of 2018-11-13)</b>		<b>157114</b>	<b>240269</b>
Gene symbol	The gene description, gene symbol (as recommended by the HUGO Nomenclature Committee) and chromosomal location is recorded for each gene. In cases where a gene symbol has not yet been made official, a provisional symbol has been adopted which is denoted by lower-case letters.	6531	9976
cDNA sequence	cDNA reference sequences are provided, numbered by codon.	6531	10339
Genomic coordinates	Genomic (chromosomal) coordinates have been calculated for missense/nonsense, splicing, regulatory, small deletions, small insertions and small indels.	0	214308
HGVS nomenclature	Standard HGVS nomenclature has been obtained for missense/nonsense, splicing, regulatory, small deletions, small insertions and small indels.	0	214691
Missense/nonsense	Single base-pair substitutions in coding regions are presented in terms of a triplet change with an additional flanking base included if the mutated base lies in either the first or third position in the triplet.	87397	137354
Splicing	Mutations with consequences for mRNA splicing are presented in brief with information specifying the relative position of the lesion with respect to a numbered intron donor or acceptor splice site. Positions given as positive integers refer to a 3' (downstream) location, negative integers refer to a 5' (upstream) location.	14317	21222
Small insertions	Missense insertions (20 bp or less) are presented in terms of the inserted bases in lower case plus, in upper case, 10 bp DNA sequence flanking both sides of the lesion. The numbered codon is preceded in the given sequence by the caret character (^).	9844	14746
Complex rearrangements	Information regarding the nature and location of each lesion is logged in narrative form because of the extremely variable quality of the original data reported.	1575	2031

<http://www.hgmd.cf.ac.uk/ac/index.php>

Описание мутаций ядерных генов,  
приводящих к наследственным заболеваниям

# GenBank

Файл Правка Вид Журнал Закладки Инструменты Справка

genbank - Google Search x FDA Whole Genome Sequencin... x GenBank Overview x +

https://www.ncbi.nlm.nih.gov/genbank/ 110% Поиск

Часто посещаемые

NCBI Resources How To Sign in to NCBI

GenBank Nucleotide Search

GenBank Submit Genomes WGS Metagenomes TPA TSA INSDC Other

### GenBank Overview

#### What is GenBank?

GenBank® is the NIH genetic collection of all publicly available DNA sequences (*Nucleic Acids Research*, 2013 Jan;41(D1):D3-D7). It is a part of the International Nucleotide Sequence Database Collaboration, which comprises the DNA DataBank of Japan (DDB), the European Nucleotide Archive (ENA), and GenBank at NCBI. These three organizations exchange data on a daily basis.

A GenBank release occurs every two months and is available from the [ftp site](#). The [release notes](#) for the current version of GenBank provide detailed information about the release and notifications of upcoming changes to GenBank. Release notes for [previous GenBank releases](#) are also available. GenBank growth statistics for both the traditional GenBank divisions and the WGS division are available from each release. GenBank growth [statistics](#) for both the traditional GenBank divisions and the WGS division are available from each release.

An [annotated sample GenBank record](#) for a *Saccharomyces cerevisiae* gene demonstrates many of the features of the GenBank flat file format.

#### Access to GenBank

There are several ways to search and retrieve data from GenBank.

- Search GenBank for sequence identifiers and annotations with [Entrez Nucleotide](#), which is divided into three divisions: [CoreNucleotide](#) (the main collection), [dbEST](#) (Expressed Sequence Tags), and [dbGSS](#) (Genome Survey Sequences).
- Search and align GenBank sequences to a query sequence using [BLAST](#) (Basic Local Alignment Search Tool). BLAST searches CoreNucleotide, dbEST, and dbGSS independently; see [BLAST info](#) for more information about the numerous BLAST databases.
- Search, link, and download sequences programatically using [NCBI e-utilities](#). The ASN.1 and flatfile formats are available at NCBI's anonymous FTP server (<ftp://ftp.ncbi.nlm.nih.gov/ncbiasn1/>) and (<ftp://ftp.ncbi.nlm.nih.gov/genbank/>).

#### GenBank Data Usage

The GenBank database is designed to provide and encourage access within the scientific community to the most up-to-date and comprehensive DNA sequence information. Therefore, NCBI places no restrictions on the use or distribution of the GenBank data. However, some submitters may claim patent, copyright, or other intellectual property rights in all or a portion of the data they have submitted. NCBI is not in a position to assess the validity of such claims, and therefore cannot provide comment or unrestricted

## Genomes → Complete Genome



# GenBank

Файл Правка Вид Журнал Закладки Инструменты Справка

genebank - Google Search x FDA Whole Genome Sequencin... x Prokaryotic and Eukaryotic ... x +

https://www.ncbi.nlm.nih.gov/genbank/genomesubmit/ 110% Поиск

Часто посещаемые

NCBI Resources How To Sign in to NCBI

GenBank Nucleotide Search

GenBank Submit Genomes WGS Metagenomes TPA TSA INSDC Other

## Prokaryotic and Eukaryotic Genomes Submission Guide

Both WGS and non-WGS genomes, including gapless complete bacterial chromosomes, can be submitted via the Submission Portal. You will be asked to choose whether the genome being submitted is considered WGS or not. The differences for GenBank purposes are:

**non-WGS genome**

**WGS genome**

Chr1 p1 p2 Chr2

Chr1 p1 Chr2

= Gap (Ns)

**non-WGS**

- Each chromosome is in a single sequence and there are no extra sequences
- Each sequence in the genome must be assigned to a chromosome, plasmid or organelle
- Plasmids and organelles can still be in multiple pieces.

**WGS**

- One or more chromosomes are in multiple pieces and/or some sequences are not assembled into chromosomes

**In both cases**

- There can still be gaps within the sequences; you will supply that information in the submission
- Plasmids and organelles can still be in multiple pieces.
- Internal sequences must be arranged in the correct order and orientation.
- Sequences concatenated in unknown order are not allowed.

## Genome Resources

- [About WGS](#)
- [WGS Browser](#)
- [Genome Submission Guide](#)
- [Genome Submission Portal](#)
- [Update Genome Records](#)
- [FAQ](#)
- [tbl2asn](#)
- [Create Submission Template](#)
- [Eukaryotic Annotation Guide](#)
- [Prokaryotic Annotation Guide](#)
- [Annotation Example Files](#)
- [Discrepancy Report](#)
- [NCBI Prokaryotic Genome Annotation Pipeline](#)
- [AGP Format](#)
- [Complex Assembly Submission Guide](#)
- [Metagenome Submission Guide](#)
- [BioProject](#)
- [Structured Comment](#)

# WGS Browser

# (Whole Genome Shotgun)

# GenBank — WGS Browser

The screenshot shows the NCBI Sequence Set Browser interface. On the left, the 'Facet Panel' has 'Organism' selected. In the 'Description' section, the 'Project type' filter is set to 'All'. The search results table shows 1,344 projects, with the first row highlighted. The first row is circled in red, and the text 'Organism — Homo sapiens' is overlaid in purple at the bottom.

Available Facets

- Type
- Source database
- Targeted Locus Name
- Project type
- Organism
- Bioproject
- Biosample
- Strain
- Breed
- Cultivar
- Isolate
- Tissue Type
- Host
- Isolation Source
- Dev Stage

Description

This site is for browsing WGS (Whole Genome Shotgun) genomes, TSA (Transcriptome Shotgun Assemblies) and TLS (Targeted Locus Study) sets. WGS sequences are incomplete genomes that have been sequenced by a whole genome shotgun strategy. TSA sequences are transcript sequences that have been computationally assembled from primary RNA sequence data. TLS sequences are large-scale marker gene sequencing studies. Please consult [WGS Submission](#) or [TSA Submission](#) pages for more details.

Project type:  All  WGS  TLS  TSA

Search – search in all fields. Use wildcard '\*' to search in the middle of a field's text.

Term

Found 1,344 projects   Page  2 3 4 5 6 7 8 ... 27 ( per page)

#	Prefix	Type	DIV	Organism	Bioproject	Biosample	Intraspecific Name	Other Source	Contigs					
									Total Length (Mbases)	#	# Prot	Has Annot	#	
1	AACC02	WGS	PRI	Homo sapiens	PRJNA10793	SAMN02981216			137.7	171	109	Y	20	
2	AADB02	WGS	PRI	Homo sapiens	PRJNA1431	SAMN02981219			3,087.3	401,931			14,095	64,...
3	AADD01	WGS	PRI	Homo sapiens	PRJNA1431	SAMN02981219			2,756.6	61,715			2,639	
4	AADD01	WGS	PRI	Homo sapiens	PRJNA1431	SAMN02981219			2,695.6	211,493			1,372	

Organism — Homo sapiens

# GenBank — Project: AACCC02

Файл Правка Вид Журнал Закладки Инструменты Справка

GenBank - Поиск в Google x AACCC00000000.2 Homo sa... x +

https://www.ncbi.nlm.nih.gov/Traces/wgs/AACCC02?display=contigs&page=1 97% Поиск

NCBI Sequence Set Browser ? Sign In to NCBI

Project: AACCC02 Search List of all Projects

### AACCC0000000.2 Homo sapiens

Master Contigs Proteins Download

# of Contigs:	171	On Jul 19, 2004 this sequence version replaced gi:29727032.
# of Proteins:	818	The Homo sapiens whole genome shotgun (WGS) project has the project accession AACCC02000000. This version of the project (02) has the accession number AACCC02000000, and consists of sequences AACCC02000001-AACCC02000171.
# of Scaffolds/Chrs:	20	
Total length:	137,712,494 bp	
BioProject:	<a href="#">PRJNA10793</a>	
BioSample:	<a href="#">SAMN02981216</a>	
Keywords:	WGS	
Annotation:	Contigs, Scaffolds	
Organism:	<a href="#">Homo sapiens</a> <a href="#">show lineage</a>	
Biosource:	<a href="#">/chromosome = 7</a>	
	<a href="#">/mol_type = genomic genomic</a>	
WGS:	<a href="#">AACCC02000001:AACCC02000171</a>	
Scaffolds:	<a href="#">CH236947:CH236966</a>	
	20 scaffolds, 709 proteins, total length is 123,763,301 bases	
Reference:	<a href="#">Human chromosome 7: DNA sequence and biology</a> : Science 300 (5620), 767-772 (2003) – <a href="#">show 90 authors</a>	
Submission:	Submitted (27-MAR-2003) Department of Genetics and Genomic Biology, The Hospital for Sick Children, 555 University Avenue, Toronto, Ontario M5G 1X8, Canada – <a href="#">show 88 authors</a>	
Submission:	Submitted (29-JUN-2004) Department of Genetics and Genomic Biology, The Hospital for Sick Children, 555 University Avenue, Toronto, Ontario M5G 1X8, Canada – <a href="#">show 88 authors</a>	

National Center for Biotechnology Information, U.S. National Library of Medicine  
8600 Rockville Pike, Bethesda MD, 20894 USA  
[Policies and Guidelines](#) | [Contact](#)

# Chromosome 7

NATIONAL LIBRARY OF MEDICINE NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION USA.gov

# GenBank — Project: AACCC02

GenBank - Поиск в Google x AACCC00000000.2 Homo sa... x +

https://www.ncbi.nlm.nih.gov/Traces/wgs/AACCC02?display=proteins&page=1

NCBI Sequence Set Browser Sign In to NCBI

Project: AACCC02 Search List of all Projects

**AACCC00000000.2 Homo sapiens**

Master **Contigs** **Proteins** Download

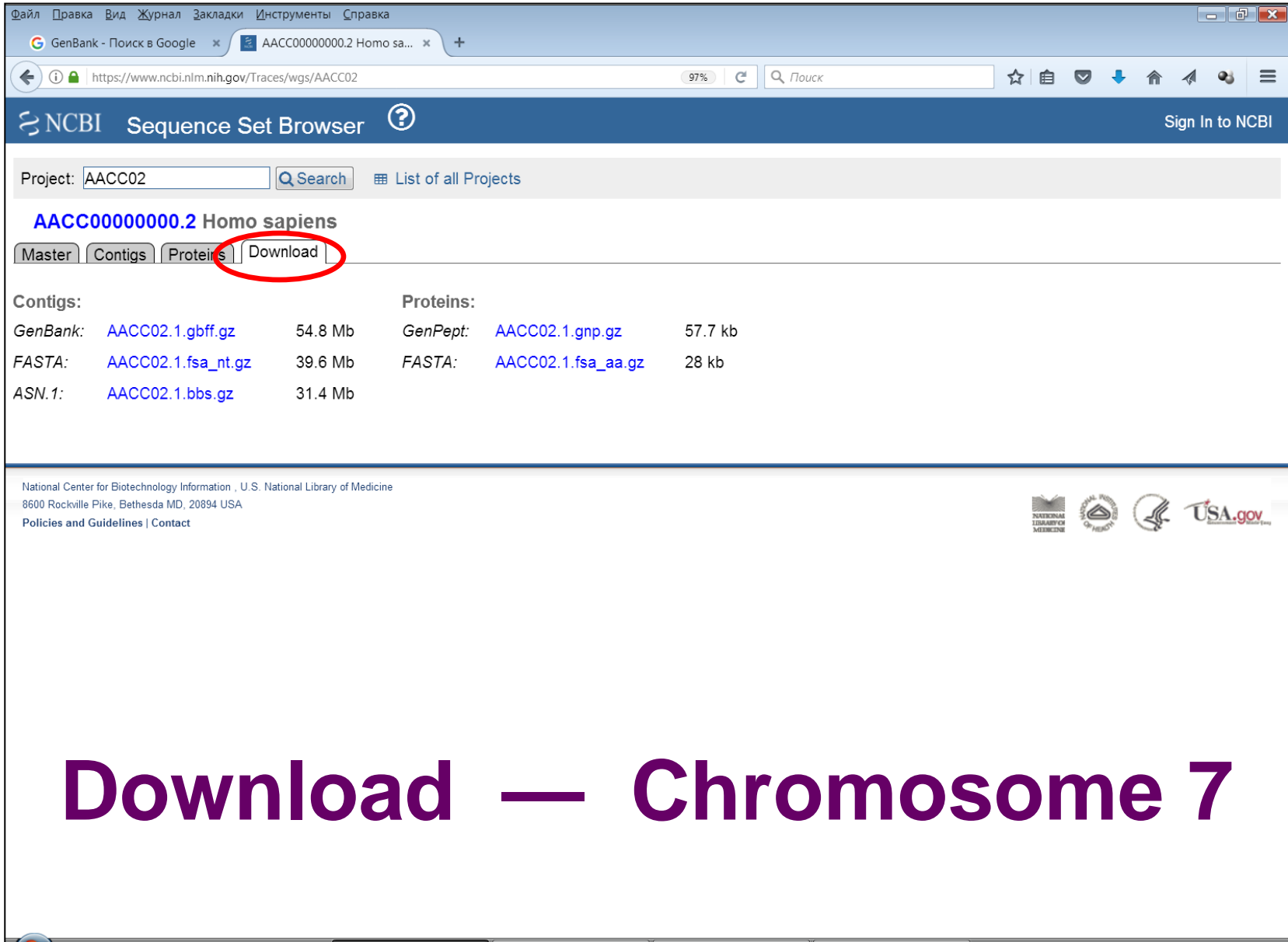
Accession Product name Search Clear

Found 109 proteins Download Page 1 2 3 (50 per page)

#	Accession	Length	Product name	View
1	EAL24413.1	519	Rho guanine nucleotide exchange factor (GEF) 5	FASTA GenPept ASN.1
2	EAL24414.1	490	similar to OG-2 homeodomain protein-like; similar to U65067 (PID:g1575526)	FASTA GenPept ASN.1
3	EAL24415.1	243	thiamin pyrophosphokinase 1	FASTA GenPept ASN.1
4	EAL24416.1	366	similar to Importin alpha-2 subunit (Karyopherin alpha-2 subunit) (SRP1-alpha) (RAG cohort protein 1)	FASTA GenPept ASN.1
5	EAL24417.1	1,331	contactin associated protein-like 2	FASTA GenPept ASN.1
6	EAL24418.1	174	similar to Mtr3 (mRNA transport regulator 3)-homolog; Mtr3 (mRNA transport regulator 3)-homolog (yeast)	FASTA GenPept ASN.1
7	EAL24419.1	132	similar to 60S ribosomal protein L32	FASTA GenPept ASN.1
8	EAL24420.1	177	chromosome 7 open reading frame 33	FASTA GenPept ASN.1
9	EAL24421.1	101	similar to Huntingtin interacting protein K	FASTA GenPept ASN.1
10	EAL24422.1	776	cullin 1	FASTA GenPept ASN.1
11	EAL24423.1	707	enhancer of zeste homolog 2 (Drosophila)	FASTA GenPept ASN.1
12	EAL24424.1	751	enhancer of zeste homolog 2 (Drosophila)	FASTA GenPept ASN.1
13	EAL24425.1	645	protein disulfide isomerase related protein (calcium-binding protein, intestinal-related)	FASTA GenPept ASN.1
14	EAL24426.1	782	hypothetical protein DKFZp7621137	FASTA GenPept ASN.1
15	EAL24427.1	752	hypothetical protein LOC155054	FASTA GenPept ASN.1

**Proteins**

# GenBank — Project: AACCC02







The screenshot shows the NCBI Sequence Set Browser interface for project AACCC02. The browser window title is "GenBank - Поиск в Google" and the address bar shows "https://www.ncbi.nlm.nih.gov/Traces/wgs/AACCC02". The page header includes the NCBI logo and "Sequence Set Browser" with a help icon and a "Sign In to NCBI" link. Below the header, there is a search bar with "Project: AACCC02" and a "Search" button, along with a "List of all Projects" link. The main content area displays the project name "AACCC0000000.2 Homo sapiens" and a navigation menu with tabs for "Master", "Contigs", "Proteins", and "Download". The "Download" tab is circled in red. Below the navigation menu, there are two columns of data: "Contigs:" and "Proteins:". The "Contigs:" column lists three files: "AACC02.1.gbff.gz" (54.8 Mb), "AACC02.1.fsa\_nt.gz" (39.6 Mb), and "AACC02.1.bbs.gz" (31.4 Mb). The "Proteins:" column lists two files: "AACC02.1.gnp.gz" (57.7 kb) and "AACC02.1.fsa\_aa.gz" (28 kb). At the bottom of the page, there is a footer with the NCBI address and contact information, and logos for the National Library of Medicine and USA.gov.

Project: AACCC02

**AACCC0000000.2 Homo sapiens**

Contigs:		Proteins:	
GenBank:	<a href="#">AACC02.1.gbff.gz</a> 54.8 Mb	GenPept:	<a href="#">AACC02.1.gnp.gz</a> 57.7 kb
FASTA:	<a href="#">AACC02.1.fsa_nt.gz</a> 39.6 Mb	FASTA:	<a href="#">AACC02.1.fsa_aa.gz</a> 28 kb
ASN.1:	<a href="#">AACC02.1.bbs.gz</a> 31.4 Mb		

National Center for Biotechnology Information, U.S. National Library of Medicine  
8600 Rockville Pike, Bethesda MD, 20894 USA  
[Policies and Guidelines](#) | [Contact](#)

## Download — Chromosome 7

# GenBank — Reference Genom

The screenshot shows the NCBI Assembly website interface. The browser address bar displays <https://www.ncbi.nlm.nih.gov/assembly>. The search bar contains the text "Homo sapiens genome", which is circled in red. The word "Assembly" in the navigation menu is circled in blue. A pink banner at the top provides information about COVID-19, including links to CDC and NIH resources. Below the banner, a section titled "Assembly" describes the database's purpose. The page is organized into three columns: "Using Assembly" with links like "Assembly Help" and "Browse by Organism"; "Submitting an Assembly" with links like "Submission Information" and "Genome Size Check"; and "Related Resources" with links like "NCBI Datasets" and "Genome Reference Consortium".

Assembly

**Homo sapiens genome** Search

COVID-19 is an emerging, rapidly evolving situation.  
Get the latest public health information from CDC: <https://www.coronavirus.gov>.  
Get the latest research from NIH: <https://www.nih.gov/coronavirus>.  
Find NCBI SARS-CoV-2 literature, sequence, and clinical content: <https://www.ncbi.nlm.nih.gov/sars-cov-2/>.

Filters removed.

### Assembly

A database providing information on the structure of assembled genomes, assembly names and other meta-data, statistical reports, and links to genomic sequence data.

#### Using Assembly

- [Assembly Help](#)
- [Browse by Organism](#)
- [NCBI Assembly Data Model](#)
- [Assembly Basics](#)
- [Nucleic Acids Research Database Issue article](#)
- [Reasons an Assembly may be Excluded from RefSeq](#)

#### Submitting an Assembly

- [Submission Information](#)
- [Submission FAQ](#)
- [Genome Size Check](#)
- [AGP Specifications](#)
- [AGP Validation](#)

#### Related Resources

- [NCBI Datasets](#) **NEW**
- [Genome](#)
- [Genome Reference Consortium](#)
- [Genome Remapping Service \(Remap\)](#)

[Prokaryotic ANI Report \(FTP file\)](#)

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

# GenBank — Reference Genom

GenBank - Поиск в Google x Homo sapiens genome - A... x +

https://www.ncbi.nlm.nih.gov/assembly/?term=Homo+sapiens+genome

NCBI Resources How To Sign in to NCBI

Assembly Assembly Homo sapiens genome Search

Create alert Advanced Browse by organism Help

COVID-19 is an emerging, rapidly evolving situation.  
Get the latest public health information from CDC: <https://www.coronavirus.gov>.  
Get the latest research from NIH: <https://www.nih.gov/coronavirus>.  
Find NCBI SARS-CoV-2 literature, sequence, and clinical content: <https://www.ncbi.nlm.nih.gov/sars-cov-2/>.

Organism group Summary 20 per page Sort by Significance Send to: Filters: [Manage Filters](#)

Animals (903)  
Customize ...

Status clear [Download Assemblies](#)

✓ Latest (903)  
✓ Latest GenBank (903)  
Latest RefSeq (1)

Assembly level

Complete genome (0)  
Chromosome (18)  
Scaffold (806)  
Contig (79)

RefSeq category

Reference (1)  
Representative (0)

Exclude clear

Exclude partial (1)  
✓ Exclude anomalous (0)  
Customize ...

Annotation status

Has annotation (3)  
GenBank has annotation (1)  
RefSeq has annotation (3)

Relation to type material

Assembly from any type (0)  
Assembly from type (0)  
Assembly from synonym type (0)  
Assembly from pathotype (0)  
Assembly designated as neotype (0)

GENOME ASSEMBLY Was this helpful?

[GRCh38.p13](#)

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

Genome Reference Consortium (March 2019)  
RefSeq GCF\_000001405.39  
[PubMed \(27\)](#) [Genome Resources](#)

[Genome Browser](#) [BLAST](#) [Get data](#)

Assembly statistics +

Search results << First < Prev Page 1 of 46 Next > Last >>

Items: 1 to 20 of 903

Filters activated: Latest, Latest GenBank, Exclude anomalous. [Clear all](#) to show 1068 items.

1. [GRCh38.p13](#)  
1. Genome Reference Consortium Human Build 38 patch release 13 (GRCh38.p13)

NCBI Datasets [Download Datasets](#)

Download a genome dataset including genome, transcript and protein sequence, annotation and a data report. [Learn more](#)

Find related data Database: Select [Find items](#)

Search details

```
"Homo sapiens"[Organism] AND ((latest[filter] OR "latest genbank"[filter]) AND all[filter] NOT anomalous[filter])
```

[Search](#) [See more...](#)

Recent activity [Turn Off](#) [Clear](#)

Homo sapiens genome AND ((latest[filter] OR "latest genbank"[filt... (903) Assembly

GRCh38.p13 - Genome - Assembly - NCBI Assembly

# Референсный геном человека

Файл Правка Вид Журнал Закладки Инструменты Справка

GenBank - Поиск в Google x GRCh38.p13 - Genome - A... x +

https://www.ncbi.nlm.nih.gov/assembly/GCF\_000001405.39

NCBI Resources How To Sign in to NCBI

Assembly Assembly Search Help

Advanced Browse by organism

Full Report v Send to: v

**GRCh38.p13**

**Description:** Genome Reference Consortium Human Build 38 patch release 13 (GRCh38.p13)

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

**BioProject:** [PRJNA31257](#)

**Submitter:** Genome Reference Consortium

**Date:** 2019/02/28

**Assembly type:** haploid-with-alt-loci

**Release type:** patch

**Assembly level:** Chromosome

**Genome representation:** full

**RefSeq category:** reference genome

**GenBank assembly accession:** GCA\_000001405.28 (latest)

**RefSeq assembly accession:** GCF\_000001405.39 (latest)

**RefSeq assembly and GenBank assembly identical:** no ([hide details](#))

- Only in GenBank: 1 unplaced scaffold (in primary assembly-unit)
- Data displayed for RefSeq version

IDs: 2334371 [UID] 8687898 [GenBank] 8765528 [RefSeq]

**History** ([Show revision history](#))

**Comment**

See [Genome](#) Information for **Homo sapiens**

There are 917 assemblies for this organism

[See more](#)

**Download Assembly**

**Access the data**

- Genome Data Viewer
- RefSeq Annotation Report
- BLAST the assembly
- Full sequence report
- Statistics report
- Regions report
- FTP directory for RefSeq assembly

**Assembly information**

- Assembly Help
- Assembly Basics
- NCBI Assembly Data Model

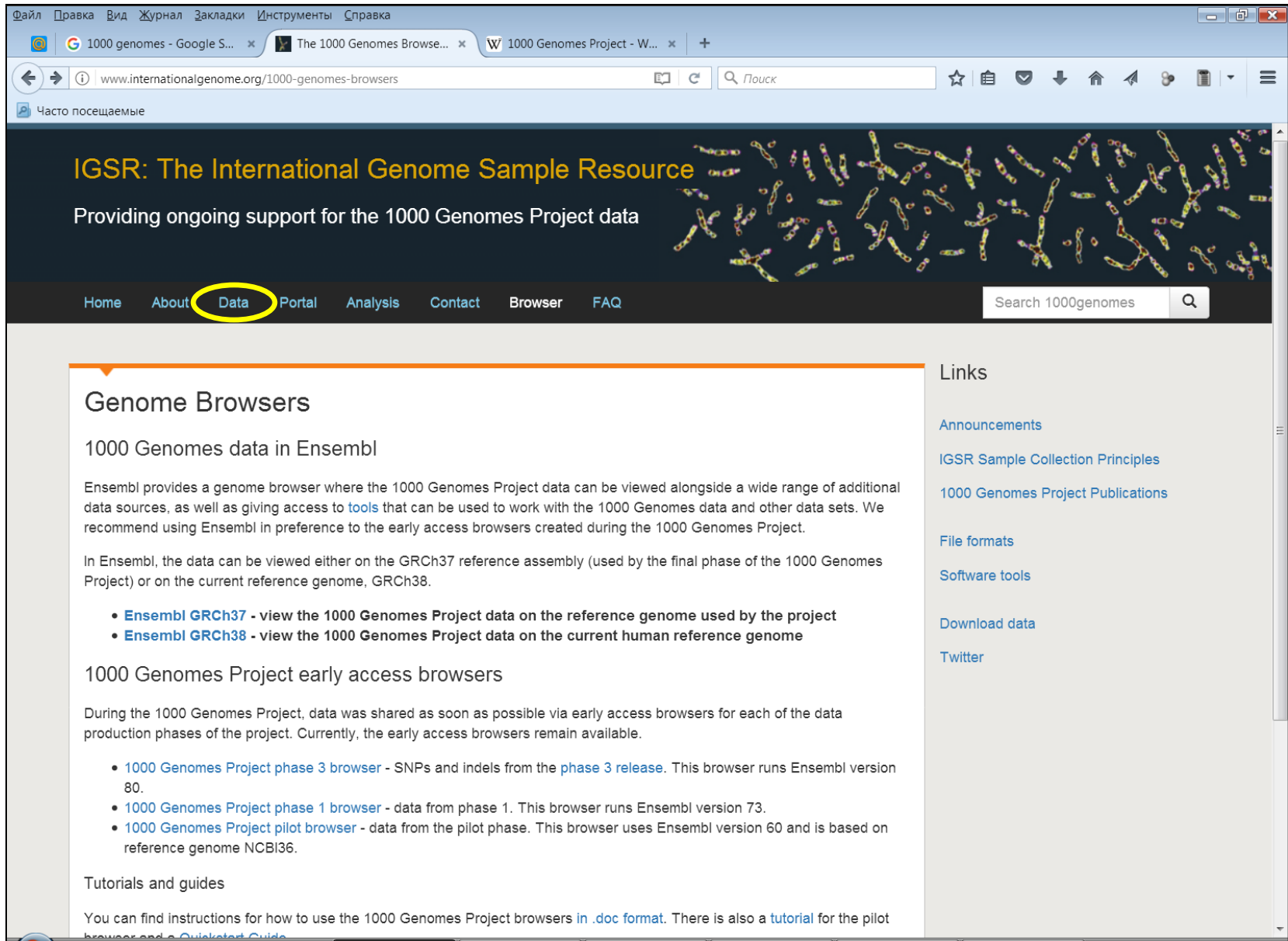
[https://www.ncbi.nlm.nih.gov/assembly/GCF\\_000001405.39](https://www.ncbi.nlm.nih.gov/assembly/GCF_000001405.39)

Всего доступно  
917 различных сборок

Сборка GRCh38.p13 — февраль 2019  
Genome Reference Consortium Human Build 38 patch release 13



# 1000 Genomes



Файл Правка Вид Журнал Закладки Инструменты Справка

1000 genomes - Google S... The 1000 Genomes Browse... 1000 Genomes Project - W... +

www.internationalgenome.org/1000-genomes-browsers Поиск

Часто посещаемые

## IGSR: The International Genome Sample Resource

Providing ongoing support for the 1000 Genomes Project data

Home About **Data** Portal Analysis Contact Browser FAQ

Search 1000genomes

### Genome Browsers

#### 1000 Genomes data in Ensembl

Ensembl provides a genome browser where the 1000 Genomes Project data can be viewed alongside a wide range of additional data sources, as well as giving access to [tools](#) that can be used to work with the 1000 Genomes data and other data sets. We recommend using Ensembl in preference to the early access browsers created during the 1000 Genomes Project.

In Ensembl, the data can be viewed either on the GRCh37 reference assembly (used by the final phase of the 1000 Genomes Project) or on the current reference genome, GRCh38.

- [Ensembl GRCh37](#) - view the 1000 Genomes Project data on the reference genome used by the project
- [Ensembl GRCh38](#) - view the 1000 Genomes Project data on the current human reference genome

#### 1000 Genomes Project early access browsers

During the 1000 Genomes Project, data was shared as soon as possible via early access browsers for each of the data production phases of the project. Currently, the early access browsers remain available.

- [1000 Genomes Project phase 3 browser](#) - SNPs and indels from the [phase 3 release](#). This browser runs Ensembl version 80.
- [1000 Genomes Project phase 1 browser](#) - data from phase 1. This browser runs Ensembl version 73.
- [1000 Genomes Project pilot browser](#) - data from the pilot phase. This browser uses Ensembl version 60 and is based on reference genome NCBI36.

#### Tutorials and guides

You can find instructions for how to use the 1000 Genomes Project browsers [in .doc format](#). There is also a [tutorial](#) for the pilot browser and a [Quickstart Guide](#).

### Links

- [Announcements](#)
- [IGSR Sample Collection Principles](#)
- [1000 Genomes Project Publications](#)
- [File formats](#)
- [Software tools](#)
- [Download data](#)
- [Twitter](#)

# 1000 Genomes — Data

The screenshot shows the IGSR website interface. At the top, there's a navigation bar with links: Home, About, Data, Portal, Analysis, Contact, Browser, and FAQ. A search bar is also present. The main content area features a section titled "Using data from IGSR" with a sub-section "Data portal beta". Below this is a table of genomic data. The "Population" column in the table is circled in red. To the right of the table is a "Links" sidebar with various resource links.

IGSR: The International Genome Sample Resource  
Providing ongoing support for the 1000 Genomes Project data

Home About Data Portal Analysis Contact Browser FAQ

Search 1000genomes

## Using data from IGSR

IGSR provides open data to support the community's research efforts. You can see our terms of use in our [data disclaimer](#).

### Data portal *beta*

We are developing a [new data portal](#) to make it easier to find and browse data in IGSR. You can use the development version to [explore the data set](#). Let us know what you think at [info@1000genomes.org](mailto:info@1000genomes.org).

Sample	Sex	Population	Exome	Low cov WLL	High cov WLL	HD genotyp	Complete
HG00513	Female	CHS	•	•	•	•	
HG01112	Male	CLM	•	•	•	•	
HG00759	Female	CDX	•	•	•	•	
HG01500	Male	IBS	•	•	•	•	
HG03006	Male	BEB	•	•	•		
NA18525	Female	CHB	•	•	•	•	
NA19648	Female	MXL	•	•	•	•	
HG00096	Male	GBR	•	•	•	•	
HG00419	Female	CHS	•	•	•	•	

### Links

- [Announcements](#)
- [IGSR Sample Collection Principles](#)
- [1000 Genomes Project Publications](#)
- [File formats](#)
- [Software tools](#)
- [Download data](#)
- [Twitter](#)

# 1000 Genomes — Population

IGSR: The International Genome Sample Resource  
Providing ongoing support for the 1000 Genomes Project data

Home About Data Portal Analysis Contact Browser FAQ

Search 1000genomes

Samples Populations Data collections Search

Filter by population Filter by analysis group Filter by data collection Toggle table view Download the list

Filter by population

- ACB
- ASW
- BEB
- CDX
- CEU
- CHB
- CHD
- CHS
- CLM
- ESN
- FIN
- GBR
- GIH
- GWD
- GWF
- GWJ
- GWW
- IBS
- ITU
- JPT
- KHV
- LWK
- MSL
- MXL
- PEL
- PJL
- PUR
- STU
- TSI
- YRI

Filters: GBR

Showing 1 to 50 of 107 samples

« Previous Next »

Sample	Sex	Population	GRCCh1	Phase 2	Phase 1	Platinum pedigree	Structural variation	Gambian variation	90 Han Chinese	Geuvadis
HG00099	Female	GBR	.	.	.				.	
HG00102	Female	GBR	.	.	.				.	

**British**

# 1000 Genomes — Sample HG00099

The screenshot shows a web browser window with the URL [www.internationalgenome.org/data-portal/sample/HG00099](http://www.internationalgenome.org/data-portal/sample/HG00099). The page header includes the IGSR logo and navigation links: Home, About, Data, Portal, Analysis, Contact, Browser, and FAQ. A search bar is located in the top right corner.

The main content area features a navigation menu with 'Samples', 'Populations', 'Data collections', and 'Search'. The 'Sample HG00099' title is circled in red. A 'Data portal beta' badge is visible in the top right of the content area.

Two summary boxes are displayed:

- HG00099 details:**
  - Sex: Female
  - Biosample ID: [SAME123271](#)
  - Search Coriell: [HG00099](#)
- GBR population:**
  - Population: [British](#)
  - Code: GBR
  - Description: British in England and Scotland
  - Superpopulation: European
  - Superpopulation code: EUR

Below these boxes, the section 'Data collections for HG00099' is shown. It includes tabs for '1000 Genomes on GRCh38', '1000 Genomes phase 3 release', '1000 Genomes phase 1 release', and 'Geuvadis'. The '1000 Genomes on GRCh38' tab is active, showing a link to the 'Data reuse policy for 1000 Genomes on GRCh38', a count of '12 matching data files', and a 'Download the list' button.

A table of data files is partially visible at the bottom, with columns for 'Data types', 'File', and 'Analysis group'. The first row shows a file path: `ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR741/SRR741411/SRR741411_2.fastq.gz` with an analysis group of 'Low coverage'.

# 1000 Genomes — Exome

Sample HG00099 Data portal beta

**HG00099 details**

Sex: Female  
Biosample ID: SAME123271  
Search Coriell: HG00099 [Q](#)

**GBR population**

Population: [British](#)  
Code: GBR  
Description: British in England and Scotland  
Superpopulation: European  
Superpopulation code: EUR

Data collections for HG00099

1000 Genomes on GRCh38    1000 Genomes phase 3 release    1000 Genomes phase 1 release    [Geuvadis](#)

[Data reuse policy for 1000 Genomes on GRCh38](#)    2 matching data files    [Download the list](#)

**Data types**

Alignment  
 **Sequence**

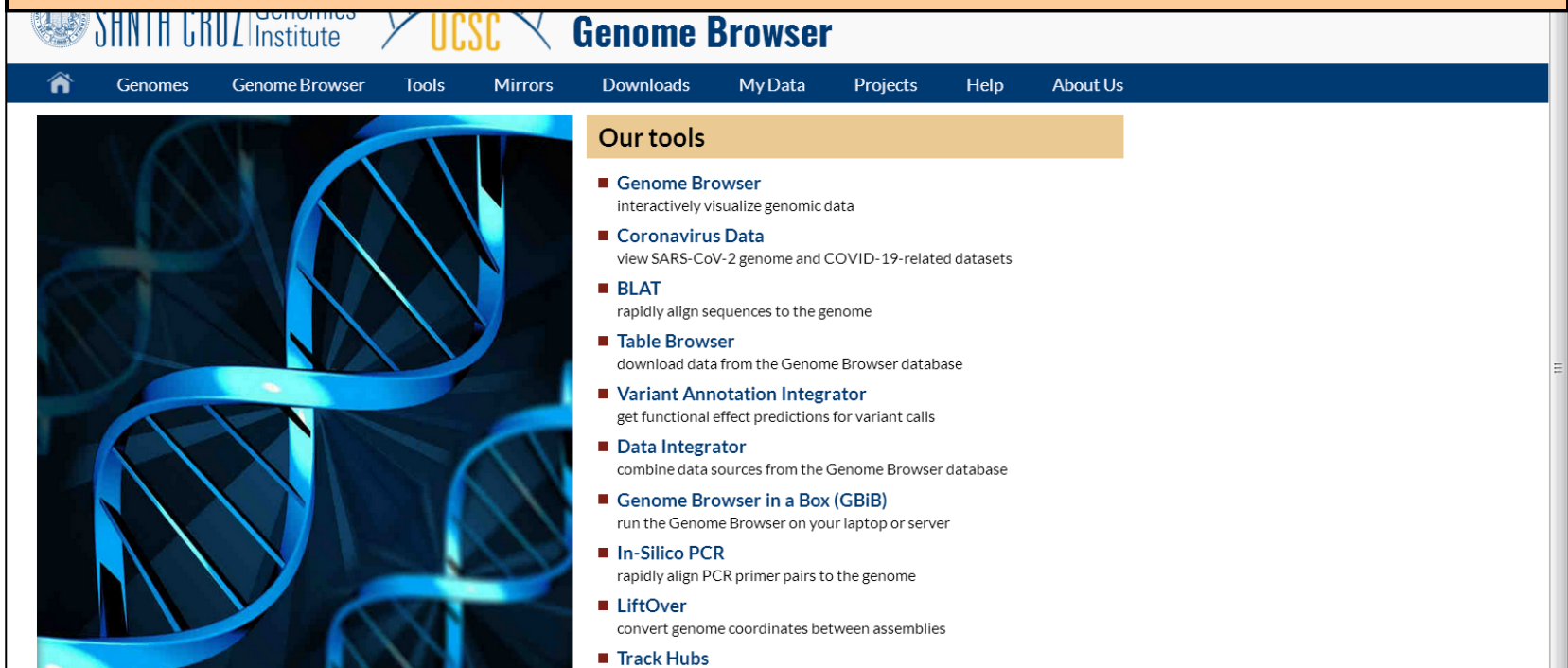
**Analysis groups**

**Exome**  
 Low coverage WGS

File	Analysis group
<a href="ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR765/SRR765993/SRR765993_2.fastq.gz">ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR765/SRR765993/SRR765993_2.fastq.gz</a>	Exome
<a href="ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR765/SRR765993/SRR765993_1.fastq.gz">ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR765/SRR765993/SRR765993_1.fastq.gz</a>	Exome

# Геномные браузеры

## UCSC Genome Browser



**Our tools**

- **Genome Browser**  
interactively visualize genomic data
- **Coronavirus Data**  
view SARS-CoV-2 genome and COVID-19-related datasets
- **BLAT**  
rapidly align sequences to the genome
- **Table Browser**  
download data from the Genome Browser database
- **Variant Annotation Integrator**  
get functional effect predictions for variant calls
- **Data Integrator**  
combine data sources from the Genome Browser database
- **Genome Browser in a Box (GBiB)**  
run the Genome Browser on your laptop or server
- **In-Silico PCR**  
rapidly align PCR primer pairs to the genome
- **LiftOver**  
convert genome coordinates between assemblies
- **Track Hubs**

<http://genome.ucsc.edu/>

### Our story

On June 22, 2000, UCSC and the other members of the International Human Genome Project consortium completed the first working draft of the human genome assembly, forever ensuring free public access to the genome and the information it contains. A few weeks later, on July 7, 2000, the newly assembled genome was released on the web at <http://genome.ucsc.edu>, along with the initial prototype of a

### What's new

Nov 3, 2020 - [New gnomAD pext track for GRCh37/hg19](#)  
Oct 23, 2020 - [New GRCm39 Mouse Genome Browser is here](#)  
Oct 21, 2020 - [Oct 21st SARS-CoV-2 release](#)  
[More news...](#)

# Геномные браузеры

## Ensembl Genome Browser



The screenshot shows the Ensembl Genome Browser homepage. At the top, there is a navigation bar with the Ensembl logo and links for BLAST/BLAT, VEP, Tools, BioMart, Downloads, Help & Docs, and Blog. A search bar is located on the right side of the navigation bar. Below the navigation bar, there is a main content area with several sections:

- Tools:** A section with four columns:
  - 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:** A search box with a dropdown menu set to "All species" and a "Go" button. Below the search box, there is a text input field and a "Go" button. Below the text input field, there is a text input field with the text "e.g. BRCA2 or rat 5:62797383-63627669 or rs699 or coronary heart disease".
- All genomes:** A dropdown menu with the text "-- Select a species --".
- Favourite genomes:** A section with a small image of a human head and the text "Human GRCh38.p12".
- Ensembl Release 94 (October 2018):** A section with a list of bullet points:
  - New fish: 38 new and updated fish genomes
  - GENCODE update 29 for human and M19 for mouse
  - Additional pathogenicity predictors for missense variants
  - New transcription factor binding motifs from SELEX
  - Gene trees using HMMs
- Other news from our blog:** A section with a list of bullet points:
  - 13 Nov 2018: [What's coming in Ensembl 95 / Ensembl Genomes 42](#)
  - 06 Nov 2018: [Job: Regulation Project Leader](#)
  - 06 Nov 2018: [Job: Bioinformatician – comparative genomics](#)

At the bottom of the page, there is a yellow banner with the URL <https://www.ensembl.org/index.html>. Below the banner, there is a section with a small image of a zebrafish and the text "Zebrafish GRCz11". At the very bottom of the page, there is a footer with a cookie notice: "This website requires cookies, and the limited processing of your personal data in order to function. By using the site you are agreeing to this as outlined in our [Privacy Policy](#) and [Terms of Use](#)." and a button that says "I Agree".

# Геномные браузеры

## Vista

The screenshot shows the VISTA web interface. At the top, there is a navigation menu with links for VISTA Home, Custom Alignment, Browser (selected), Enhancer DB, Downloads, Publications, and Help. Below the menu is a search form with the following fields: Clade (Vertebrate), Genome (Human), Release (Feb. 2009), and Position (chr9:107,543,285-107,690,436). There is a Submit button and two radio buttons for VISTA-Point and VISTA Browser (Requires Java2). A Help link is also present.

**Whole Genome Comparative Analysis of the *Human Feb. 2009* Genome**

From this page you can access the results of:

- the **multiple alignments**
  - [Human Feb. 2009](#), [Chimp Mar. 2006](#), [Callithrix jacchus v.2.0.2](#), [Rhesus Jan. 2006](#) and [Pongo pygmaeus abelii v.2.0.2](#) genomes.
- the **pairwise alignments of the**
  - Human Feb. 2009 genome produced by the [Genome Reference Consortium](#) with the following genomes:
    - The [Chimp Feb. 2011](#) genome produced by the [Chimpanzee Sequencing and Analysis Consortium](#)
    - The [Cow Oct. 2011](#) genome produced by the [Baylor College of Medicine -- Cow](#)
    - The [Gorilla Dec. 2009](#) genome produced by the [The Wellcome Trust Sanger Institute](#)

Annotation for this genome comes from the [UCSC](#).

[Methods and references](#)

**<http://pipeline.lbl.gov/cgi-bin/gateway2>**



# Геномные браузеры

## GBrowse



[Log in / create account](#)

Page [Discussion](#)Read [View form](#) [View source](#) [View history](#)

### GBrowse



**Contents** [\[hide\]](#)

- 1 About Generic Genome Browser (GBrowse)
  - 1.1 Screenshots
- 2 Downloads
- 3 Using GBrowse
  - 3.1 System Requirements
  - 3.2 Installation
  - 3.3 Configuration
- 4 Documentation
  - 4.1 POD documentation
  - 4.2 Installation
  - 4.3 Configuration
  - 4.4 Documentation
  - 4.5 FAQ

**Status**

Mature release

Development: active

Support: active

---

**Licence**

 [GPL 2.0](#) [Artistic License 2.0](#)

---

**Resources**

[Home page](#)

<http://gmod.org/wiki/GBrowse>

What links here

Related changes

Special pages

Printable version

Permanent link

Page information

Browse properties

Print as PDF

- 7 GBrowse in the wild
- 8 See also
- 9 More on GBrowse
  - 9.1 GBrowse Logo

### About Generic Genome Browser (GBrowse)

GBrowse is a combination of database and interactive web pages for manipulating and displaying annotations on genomes. Features include:

# UCSC Genome Browser

Уайл Правка Вид Журнал Закладки Инструменты Справка

GenBank - Поиск в Google UCSC Genome Browser Ho... +

genome.ucsc.edu Поиск

UNIVERSITY OF CALIFORNIA SANTA CRUZ Genomics Institute UCSC Genome Browser

Genomes Genome Browser Tools Mirrors Downloads MyData Projects Help About Us

- Human GRCh38/hg38
- Human GRCh37/hg19
- Mouse GRM39/mm39
- Mouse GRM38/mm10
- Mouse: 16 strains
- SARS-CoV-2 (COVID-19)
- Other

### Our tools

- Genome Browser**  
interactively visualize genomic data
- Coronavirus Data**  
view SARS-CoV-2 genome and COVID-19-related datasets
- BLAT**  
rapidly align sequences to the genome
- Table Browser**  
download data from the Genome Browser database
- Variant Annotation Integrator**  
get functional effect predictions for variant calls
- Data Integrator**  
combine data sources from the Genome Browser database
- Genome Browser in a Box (GBIB)**  
run the Genome Browser on your laptop or server
- In-Silico PCR**  
rapidly align PCR primer pairs to the genome
- LiftOver**  
convert genome coordinates between assemblies
- Track Hubs**

**Загружаем референсный геном**

### Our story

On June 22, 2000, UCSC and the other members of the International Human Genome Project consortium completed the first working draft of the human genome assembly, forever ensuring free public access to the genome and the information it contains. A few weeks later, on July 7, 2000, the first assembled genome was released on the web at [genome.ucsc.edu/cgi-bin/hgGateway](http://genome.ucsc.edu/cgi-bin/hgGateway) long with the initial prototype of a

### What's new

- Nov 3, 2020 - [New gnomAD pext track for GRCh37/hg19](#)
- Oct 23, 2020 - [New GRCm39 Mouse Genome Browser is here](#)
- Oct 21, 2020 - [Oct 21st SARS-CoV-2 release](#)
- [More news...](#)

# UCSC Genome Browser

UCSC Genome Browser on Human Dec. 2013 (GRCh38/hg38) Assembly

move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x 100x

chr1:11,102,837-11,267,747 164,911 bp. enter position, gene symbol, HGVS or search terms go

chr1 (p36.22) 1p31.1 1q12 1q41 1q42.44

Scale chr1: 11,150,000 50 kb hg38 11,200,000 11,250,000

MTOR MTOR New haplotype sequences to GRCh38 Reference Sequence Patches to GRCh38 Reference Sequence

GENCODE v24 Comprehensive Transcript Set (only Basic displayed by default) RNU6-291P

RefSeq Curated RefSeq gene predictions from NCBI RNU6-537P

OMIM Alleles OMIM Allelic Variants

gene Expression in 53 tissues from GTEx RNA-seq of 8555 samples (570 donors)

MTOR MTOR-RS1 RNU6-537P RNU6-291P RPL39P6

Layered H3K27Ac H3K27Ac Mark (Often Found Near Regulatory Elements) on 7 cell lines from ENCODE

DNase Clusters DNase I Hypersensitivity Peak Clusters from ENCODE (95 cell types)

Cons 100 Verts 100 vertebrates Basewise Conservation by PhyloP

Rhesus Multiz Alignments of 100 Vertebrates

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. Press "?" for keyboard shortcuts.

move start < 2.0 > move end < 2.0 >

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

**Хромосома 1, локус p36.22**

# UCSC Genome Browser

File Правка Вид Журнал Закладки Инструменты Справка

обозначение генов в хро... x Локус — Википедия x Human hg38 chr1:1498069... x +

genome.ucsc.edu/cgi-bin/hgTracks?db=hg38&lastVirtModeType=default&lastVirtModeExtraState=&virt Поиск

Часто посещаемые

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

## UCSC Genome Browser on Human Dec. 2013 (GRCh38/hg38) Assembly

move <<< << < > >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x 100x

**chr1:149,806,988-149,971,898** 164,911 bp. enter position, gene symbol, HGVS or search terms go

chr1 (q21.2) 1p31.1 1q12 1q41 1q43.44

Scale chr1: 149,850,000 | 50 kb | 149,900,000 | hg38

RefSeq Curated: HIST2H2BF, HIST2H2BF, HIST2H3D, HIST2H4A, HIST2H2AA4, HIST2H3C, HIST2H2A3, HIST2H2BE, AC239868.3, RP11-196G18.22, RP11-196G18.24 s, AC239868.2, HIST2H3A, RP11-196G18.23, BOLA1, SV2A, SF3B4, MTMR11, OTUD7B

OMIM Alleles: HIST2H2BF, HIST2H3D, RP11-196G18.23, HIST2H2BD, RP11-196G18.22, RP11-196G18.24

Gene Expression in 53 tissues from G-Ex RNA-seq of 8555 samples (570 donors): HIST2H2BE, HIST2H2A3, BOLA1, SV2A, SF3B4, MTMR11

## Хромосома 1, локус q21.2

Database Clusters: 4.65

Cons 100 Verts: 0

Multiz Alignments of 100 Vertebrates: Rhesus, Mouse, Dog, Elephant, Chicken, X\_tropicalia, Zonotrichia, Lamprey

Common SNPs (150): chr1

# UCSC Genome Browser

Файл Правка Вид Журнал Закладки Инструменты Справка

обозначение генов в хро... Локус — Википедия Human hg38 chr1:10,965,9...

genome.ucsc.edu/cgi-bin/hgTracks?db=hg38&lastVirtModeType=default&lastVirtModeExtraState=

Часто посещаемые

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

## UCSC Genome Browser on Human Dec. 2013 (GRCh38/hg38) Assembly

move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x 100x

chr1:10,965,907-11,130,817 164,911 bp. chr6 go

chr1 (p36.22) 1p31.1 1q12 1q41 1q43.44

Scale chr1: 11,000,000 | 11,050,000 | 11,100,000 hg38

New haplotype sequences to GRCh38 Reference Sequence  
Patches to GRCh38 Reference Sequence  
Comprehensive Transcript Set (only Basic displayed by default)

RefSeq gene predictions from NCBI  
OMIM Allelic Variants

Gene Expression in 53 tissues from GTEx RNA-seq of 8555 samples (570 donors)

C1orf127 AL713997.1 Y\_RNA TARDBP MASP2 SRM EXOSC10 RP4-635E18.6 MTOR  
RP4-635E18.9 TARDBP RP4-635E18.7 Metazoa\_SRP  
MASP2 RP4-635E18.8  
CFL1P6 SRM MTOR  
AL713997.1 TARDBP EXOSC10  
RP4-635E18.6  
RP4-635E18.7

## Переход на хромосому 6

Cons 100 Verts -4.5

Multiz Alignments of 100 Vertebrates

Rhesus Mouse Dog Elephant Chicken X\_tropicalis Zebrafish Lamprey

Simple Nucleotide Polymorphisms (dbSNP 150) Found in >= 1% of Samples

Repeating Elements by RepeatMasker

Common SNPs (150)

SINE LINE LTR DNA

# UCSC Genome Browser

The screenshot displays the UCSC Genome Browser interface for Human hg38. The browser window shows the URL `genome.ucsc.edu/cgi-bin/hgTracks?db=hg38&lastVirtModeType=default&lastVirtModeExtraState=&virt`. The page title is "UCSC Genome Browser on Human Dec. 2013 (GRCh38/hg38) Assembly". The navigation bar includes "Genomes", "Genome Browser", "Tools", "Mirrors", "Downloads", "My Data", "View", "Help", and "About Us".

The main content area shows a genomic track for chromosome 6. The track is titled "chr6: 1-170,805,979 170,805,979 bp. chr6". Below the title, there are navigation controls for moving and zooming. The track itself displays various genomic features, including "New haplotype sequences to GRCh38 Reference Sequence" and "Patches to GRCh38 Reference Sequence". The track is labeled "Scale chr6:" and "50 Mb". The track shows a dense collection of blue vertical lines representing genomic data, with some red vertical lines indicating specific features. The track is labeled "GENCODE v24 Comprehensive Transcript Set (only Basic displayed by default)".

A prominent yellow banner at the bottom of the screenshot contains the text "Вся хромосома 6" (The whole chromosome 6).

# UCSC Genome Browser

The screenshot displays the UCSC Genome Browser interface. At the top, the browser's address bar shows the URL: `genome.ucsc.edu/cgi-bin/hgTracks?db=hg38&lastVirtModeType=default&lastVirtModeExtraState=`. The page title is "UCSC Genome Browser on Human Dec. 2013 (GRCh38/hg38) Assembly".

Navigation controls include a "move" section with arrows and a "zoom in" section with buttons for 1.5x, 3x, 10x, and 100x. The current coordinates are `chr6:2-170,805,979 170,805,978 bp`, and the selected locus is `6p21.1`. A "go" button is located to the right of the coordinates.

Below the navigation controls is a track labeled "chr6 (p25.3-q27)" showing a genomic map with bands for `6p22.3`, `6q12`, `6q13`, `6q14.1`, `6q15.1`, `6q21`, and `6q25`. The main track displays a 50 Mb scale for chromosome 6, with coordinates from 50,000,000 to 150,000,000. The track shows various annotations, including "New haplotype sequences to GRCh38 Reference Sequence" and "Patches to GRCh38 Reference Sequence".

A yellow banner at the bottom of the image contains the text: **Переход на локус 6p21.1**

# UCSC Genome Browser

Файл Правка Вид Журнал Закладки Инструменты Справка

обозначение генов в хро... Локус — Википедия Human hg38 chr6:40,500,000...

genome.ucsc.edu/cgi-bin/hgTracks?db=hg38&lastVirtModeType=default&lastVirtModeExtraState=&virt Поиск

Часто посещаемые

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

## UCSC Genome Browser on Human Dec. 2013 (GRCh38/hg38) Assembly

move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x 100x

chr6:40,500,001-46,200,000 5,700,000 bp. enter position, gene symbol, HGVS or search terms go

chr6 (p21.1) 6p22.3 6q12q13q14.1 q15.1 6q21 22.31 22q27

Scale chr6: 41,000,000 41,500,000 42,000,000 42,500,000 43,000,000 43,500,000 44,000,000 44,500,000 45,000,000 45,500,000 46,000,000

GENCODE v24 Comprehensive Transcript Set (only basic displayed by default)

UNC5CL NCR2 MDF1 BYSL UBR2 PTCRA SRF DLK2 VEGFA CAPN11 SUPT3H CLIC5  
UNC5CL NCR2 MDF1 TAF8 UBR2 GNMT CRIP3 POLH MRPL14 SUPT3H CLIC5  
TSFO2 NCR2 TFEB TAF8 UBR2 PEVE CRIP3 POLH TMEM38 SUPT3H RUNX2  
LINC01276 W POC TAF8 RNUS-990F PEVE ZNF318 VEGFA SLC29A1 SUPT3H RUNX2  
TSFO2 FOXF4-RS1 H MED20 PRFH2 MER1 ABCC18 VEGFA SLC29A1 MIR5861 RP1-861.4 H  
APOBEC2 FOXF4 H MED20 TBC1 CULT7 ABCC18 VEGFA SLC29A1 RUNX2 RP1-861.4 H  
OARD1 FOXF4 H MED20 GLTSCR1L H CUL9 XPO5 C6orf223 AARS2 RUNX2 ENPF4  
OARD1 FOXF4 H MED20 GLTSCR1L H CUL9 GTFEF2 TMEM38 RUNX2 ENPF5  
OARD1 Y\_RNA DNPH1 GTFEF2 RN7SL11P1 RUNX2 ENPF5  
OARD1 FOXF4 CCND3 RPL7L1 DNPH1 RSPH9 RP1-38202.5 RP1-244F24.1 RUNX2  
OARD1 MIR4641 CCND3 H RPL7L1 TTBK1 MRPS18A SLC29A1 RP1-166H4.2  
OARD1 TFEB TAF8 C6orf226 SLC22A7 PTCRA MIR675B VEGFA SLC29A1  
OARD1 TFEB H TAF8 PTCRA DLK2 VEGFA SLC29A1  
OARD1 TFEB H TAF8 PTCRA DLK2 VEGFA SLC29A1  
OARD1 RP11-299J23.5 GUCR1A CNPY3 RNUS-1113P1 C6orf223 SPATS1  
OARD1 PGC GUCR1A H NPYA RP3-475N16.1 TJAP1 H C6orf223 SPATS1  
NPYA RP11-299J23.2 GUCR1A H RP3-475N16.1 TJAP1 H AL109515.1 CDOSL  
NPYA FRSS3 GUCR1A H RP3-475N16.1 TJAP1 H HSF90B1  
TREML1 FRSS3 GUCR1B PFP2R5D TJAP1 H HSF90B1  
TREML1 PRICKLE4 MRPS10 PFP2R5D TJAP1 H HSF90B1  
TREML1 PRICKLE4 TRERF1 PFP2R5D TJAP1 H SLC35B2  
TREML2 PRICKLE4 TRERF1 PFP2R5D TJAP1 H SLC35B2  
TREML2 PRICKLE4 TRERF1 KLHDC3 TJAP1 SLC35B2  
TREML2 TONM6 TRERF1 KLHDC3 LRRC73 SLC35B2  
TREML4 TONM6 RRP36 YIPF3 SLC35B2  
TREML4 USF49 H RFP1-29C7.5 POLR1C YIPF3 SLC35B2  
RNP55P2871 USF49 H RNTSL43P1 POLR1C NTR4647

# Локус 6p21.1

CCND3 RNU6-761P RP11-533020.21 RP11-7K24.31 C6orf132 C6orf132 RP1-199D8.6 RP1-199D8.6 KLC4 KLC4 PTK7 PTK7 PTK7 PTK7 PTK7 PTK7 SLC22A7 SLC22A7 RPS-337H4.9 RP1-444E17.6 RPS-337H4.10 RP1-344J7.2 RPS-337H4.6 RP11-338M2.5 RPS-1120P11.3 RP1-444E17.6 RPS-337H4.9 RP1-444E17.6 RPS-337H4.10 RP1-344J7.2 RPS-337H4.6 RP11-338M2.5 RPS-1120P11.3 RP1-444E17.6

RefSeq gene predictions from NCBI



# UCSC Genome Browser

The screenshot displays the UCSC Genome Browser interface. At the top, the browser's navigation bar includes menu items like "Genomes", "Genome Browser", "Tools", "Mirrors", "Downloads", "My Data", "View", "Help", and "About Us". The main heading reads "UCSC Genome Browser on Human Dec. 2013 (GRCh38/hg38) Assembly". Below this, navigation controls include "move" buttons and "zoom in" options (1.5x, 3x, 10x, base, zoom out, 1.5x, 3x, 10x, 100x). A search bar contains the coordinates "chr2:161,992,241-162,074,542 82,302 bp" and the gene name "DPP4 (Homo sapiens dipeptidyl peptidase 4 (DPP4), mRNA. (from Re" with a "go" button. A red box highlights the search bar and the "go" button. Below the search bar, a chromosome map shows the location of the gene on chromosome 2 (p21.1). The main display area shows a genomic track with various annotations, including gene models, transcripts, and other genomic features. A yellow banner at the bottom of the image contains the text "Поиск гена DPP4".

File Правка Вид Журнал Закладки Инструменты Справка

Uniprot - Google Search x dpp4 in UniProtKB x Локус — Википедия x Human hg38 chr6:40,500,0... x +

genome.ucsc.edu/cgi-bin/hgTracks?db=hg38&lastVirtModeType=default&lastVirtModeExtraState=... Поиск

Часто посещаемые

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

UCSC Genome Browser on Human Dec. 2013 (GRCh38/hg38) Assembly

move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x 100x

chr2:161,992,241-162,074,542 82,302 bp DPP4 (Homo sapiens dipeptidyl peptidase 4 (DPP4), mRNA. (from Re go

chr6 (p21.1) 6022.3 64126q13q14.1 q1516.1 6q21 22.31 256q27

Scale 2 Mb hg38  
chr6: 41,000,000 | 41,500,000 | 42,000,000 | 42,500,000 | 43,000,000 | 43,500,000 | 44,000,000 | 44,500,000 | 45,000,000 | 45,500,000 | 46,000,000

GENCODE v24 Comprehensive Transcript Set (only Basic displayed by default)

UNC5CL NCR2 MDF1 BYSL UBR2 PTCRA SRF DLK2 VEGFA C6orf223 SPATS1 SUPT3H RUNK2 CLIC5  
UNC5CL NCR2 MDF1 TAF8 UBR2 GNMT CRIP3 POLH MRPL14 SUPT3H RUNK2 CLIC5  
TSFO2 NCR2 TFEB TAF8 RNU6-761P FEVE ZNF318 VEGFA SLC29A1 SUPT3H RUNK2 CLIC5  
TSFO2 LINC01276 W FOC TAF8 RNU6-761P FEVE ZNF318 VEGFA SLC29A1 SUPT3H RUNK2 CLIC5  
TFOP2 FOXF4-AS1 MED20 PRPH2 MER1 ABC110 VEGFA SLC29A1 SUPT3H RUNK2 CLIC5  
APOBEC2 FOXF4 MED20 TBC1 CULT ABC110 VEGFA SLC29A1 SUPT3H RUNK2 CLIC5  
OARD1 FOXF4 MED20 GLTSCR1L CUL9 XPO5 C6orf223 AARS2 SUPT3H RUNK2 CLIC5  
OARD1 FOXF4 MED20 GLTSCR1L CUL9 GTFP2 TME63B SUPT3H RUNK2 CLIC5  
OARD1 FOXF4 Y\_RNA GLTSCR1L CUL9 GTFP2 TME63B SUPT3H RUNK2 CLIC5  
OARD1 FOXF4 CCND3 RPL7L1 DNPH1 RSPH9 RPL1-38202.5 SUPT3H RUNK2 CLIC5  
OARD1 MIR4641 CCND3 RPL7L1 TTBK1 RSPH9 SLC29A1 SUPT3H RUNK2 CLIC5  
OARD1 TFEB TAF8 RPL7L1 TTBK1 MRPS18A SLC29A1 SUPT3H RUNK2 CLIC5  
OARD1 TFEB TAF8 C6orf226 SLC22A7 VEGFA SLC29A1 SUPT3H RUNK2 CLIC5  
OARD1 TFEB TAF8 PTCRA HIR675B VEGFA SLC29A1 SUPT3H RUNK2 CLIC5  
OARD1 TFEB TAF8 PTCRA DLK2 VEGFA SLC29A1 SUPT3H RUNK2 CLIC5  
OARD1 TFEB TAF8 PTCRA DLK2 VEGFA SLC29A1 SUPT3H RUNK2 CLIC5  
OARD1 RFP11-298J23.5 GUCR1A CNPY3 RNU6-1113P C6orf223 SPATS1 SUPT3H RUNK2 CLIC5  
OARD1 RFP11-298J23.3 GUCR1A RFP3-475N16.1 TJAP1 H AL189515.1 CDC2L SUPT3H RUNK2 CLIC5  
NFYA RFP11-298J23.3 GUCR1A RFP3-475N16.1 TJAP1 H HSP90AB1 SUPT3H RUNK2 CLIC5  
NFYA FRSS3 GUCR1B PFP2R5D TJAP1 H HSP90AB1 SUPT3H RUNK2 CLIC5  
TREML1 PRICKLE4 MRPS10 PFP2R5D TJAP1 H HSP90AB1 SUPT3H RUNK2 CLIC5  
TREML1 PRICKLE4 TRERF1 PFP2R5D TJAP1 H HSP90AB1 SUPT3H RUNK2 CLIC5  
TREML2 PRICKLE4 TRERF1 PFP2R5D TJAP1 H HSP90AB1 SUPT3H RUNK2 CLIC5  
TREML2 PRICKLE4 TRERF1 KLHDC3 TJAP1 H HSP90AB1 SUPT3H RUNK2 CLIC5  
TREML2 PRICKLE4 TRERF1 KLHDC3 LRRC73 SLC35B2 SUPT3H RUNK2 CLIC5  
TREML2 TOMM6 TRERF1 KLHDC3 YIPF3 SLC35B2 SUPT3H RUNK2 CLIC5  
TREML4 TOMM6 RRP36 POLR1C SLC35B2 SUPT3H RUNK2 CLIC5  
RNF53P297 USP49 RFP11-2907.6 POLR1C SLC35B2 SUPT3H RUNK2 CLIC5  
RNF53P297 USP49 RNF53P297 MIR4647 NFKBIE SUPT3H RUNK2 CLIC5  
RNF53P297 USP49 RNF53P297 MIR4647 NFKBIE SUPT3H RUNK2 CLIC5  
RNF53P297 USP49 RNF53P297 MIR4647 NFKBIE SUPT3H RUNK2 CLIC5

RefSeq gene predictions from NCBI

Поиск гена DPP4

# UCSC Genome Browser

Файл Правка Вид Журнал Закладки Инструменты Справка

Uniprot - Google Search x dpp4 in UniProtKB x Локус — Википедия x Human hg38 chr2:1619922... x +

genome.ucsc.edu/cgi-bin/hgTracks?db=hg38&lastVirtModeType=default&lastVirtModeExtraState=&virt Поиск

Часто посещаемые

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

## UCSC Genome Browser on Human Dec. 2013 (GRCh38/hg38) Assembly

move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x 100x

chr2:161,992,241-162,074,542 82,302 bp. enter position, gene symbol, HGVS or search terms go

chr2 (q24.2) 21 14 p12 11.2 q34 q35

Scale chr2: 162,000,000 162,010,000 162,020,000 162,030,000 162,040,000 162,050,000 162,060,000 162,070,000 hg38

New haplotype sequences to GRCh38 Reference Sequence  
Patches to GRCh38 Reference Sequence  
Chromosome Bands Localized by FISH Mapping Clones  
2q24.2

GENCODE v24 Comprehensive Transcript Set (only Basic displayed by default)  
RefSeq Curated RefSeq gene predictions from NCBI  
OMIM Alleles OMIM Allelic Variants  
Gene Expression in 53 tissues from GTEx RNA-seq of 8555 samples (570 donors)  
DPP4 AC008063.2 (4)

Layered H3K27Ac H3K27Ac Mark (Often Found Near Regulatory Elements) on 7 cell lines from ENCODE  
AC008063.2

DNase I Hypersensitivity Peak Clusters from ENCODE (95 cell types)  
100 vertebrates Basewise Conservation by PhyloP

Cons 100 Verts  
-4.5  
Multiz Alignments of 100 Vertebrates  
Rhesus  
Mouse  
Dog  
Elephant  
Chicken  
X\_tropicalis  
Zebrafish

## Ген DPP4 — локус 2q24.2

Satellite  
RNA  
Other  
Unknown

move start < 2.0 > move end < 2.0 >

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

collapse all

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

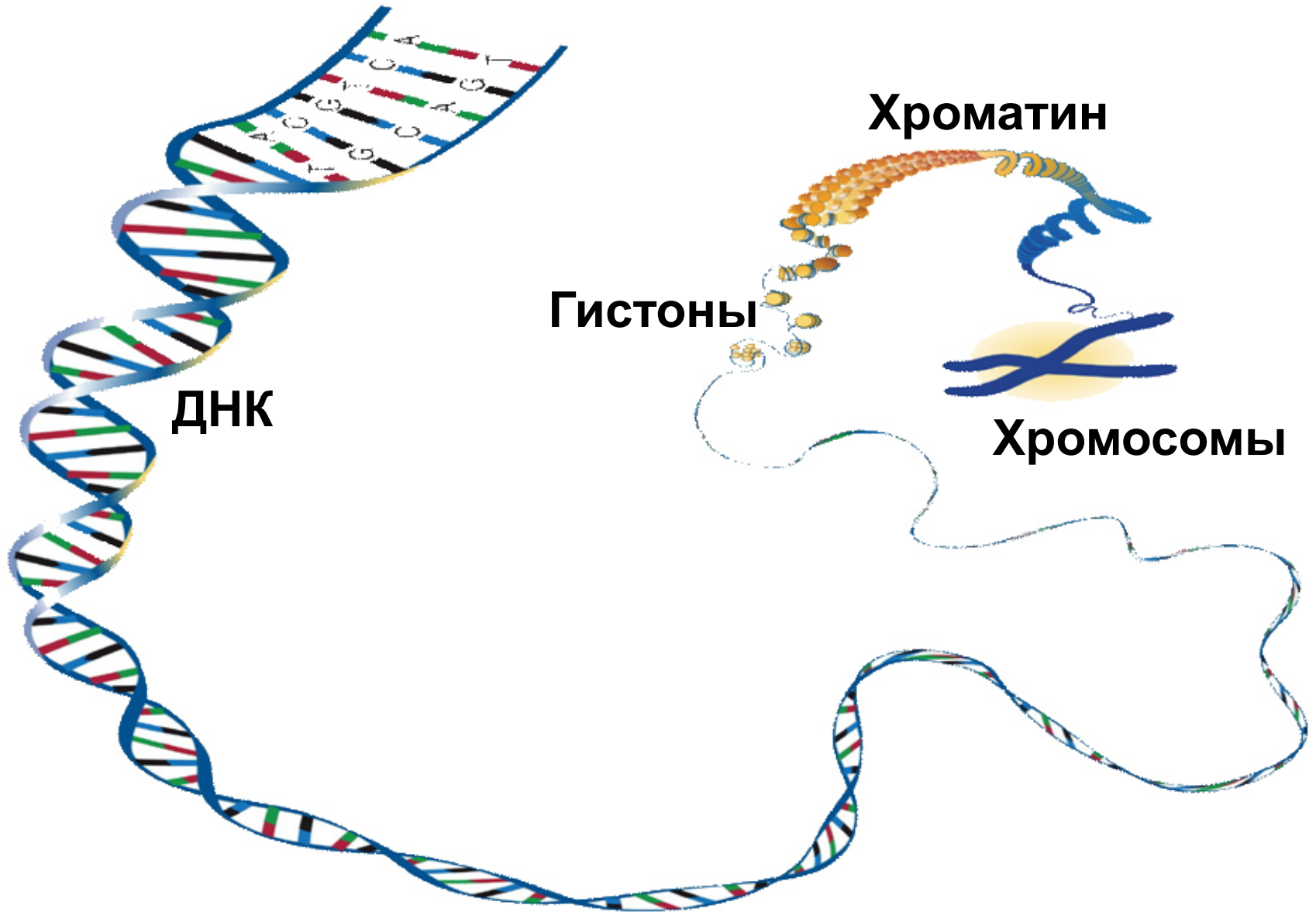
# Определения

- Лocus — местоположение определённого гена на генетической или цитологической карте хромосомы
- Аллель — вариант последовательности ДНК в данном локусе
- Генетическая карта — упорядоченный перечень локусов генома
- Генное картирование — определение локуса для специфического биологического признака

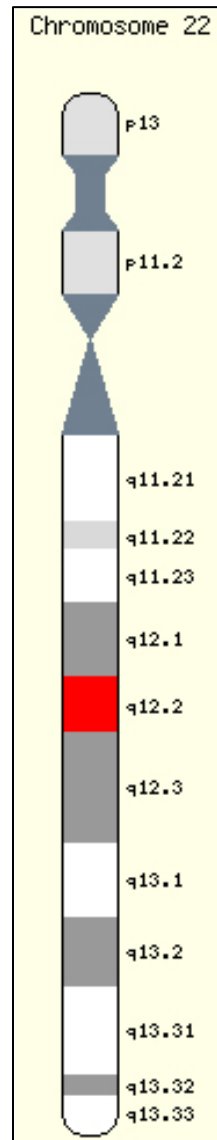
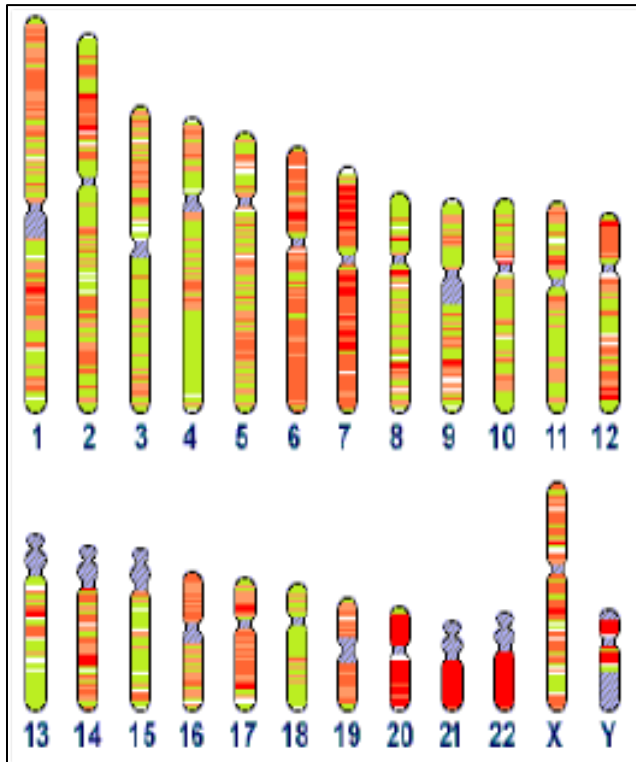
# Аннотация генома

- Описание функциональных и структурных характеристик генома
- Местонахождение кодирующих участков генов в геноме, регуляторных элементов, регулирующих транскрипцию, другие функции генома и особенности функционирования генома
- Тканеспецифичность экспрессии, профили экспрессии и коэкспрессию генов

# Структура генома



# Обозначения локусов



## 6p21.3

6 – номер хромосомы

p – плечо хромосомы:

p – короткое

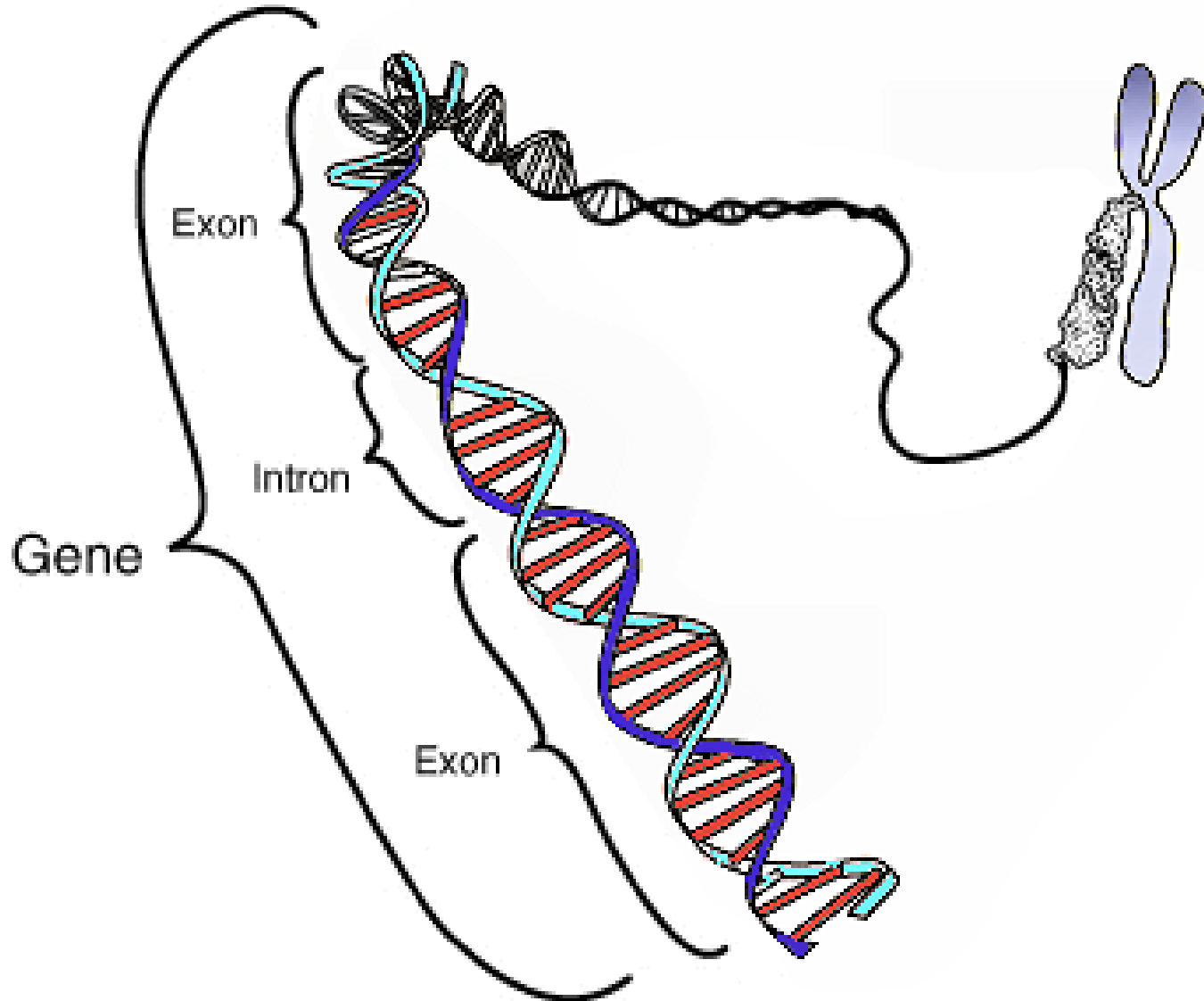
q – длинное

2 – область

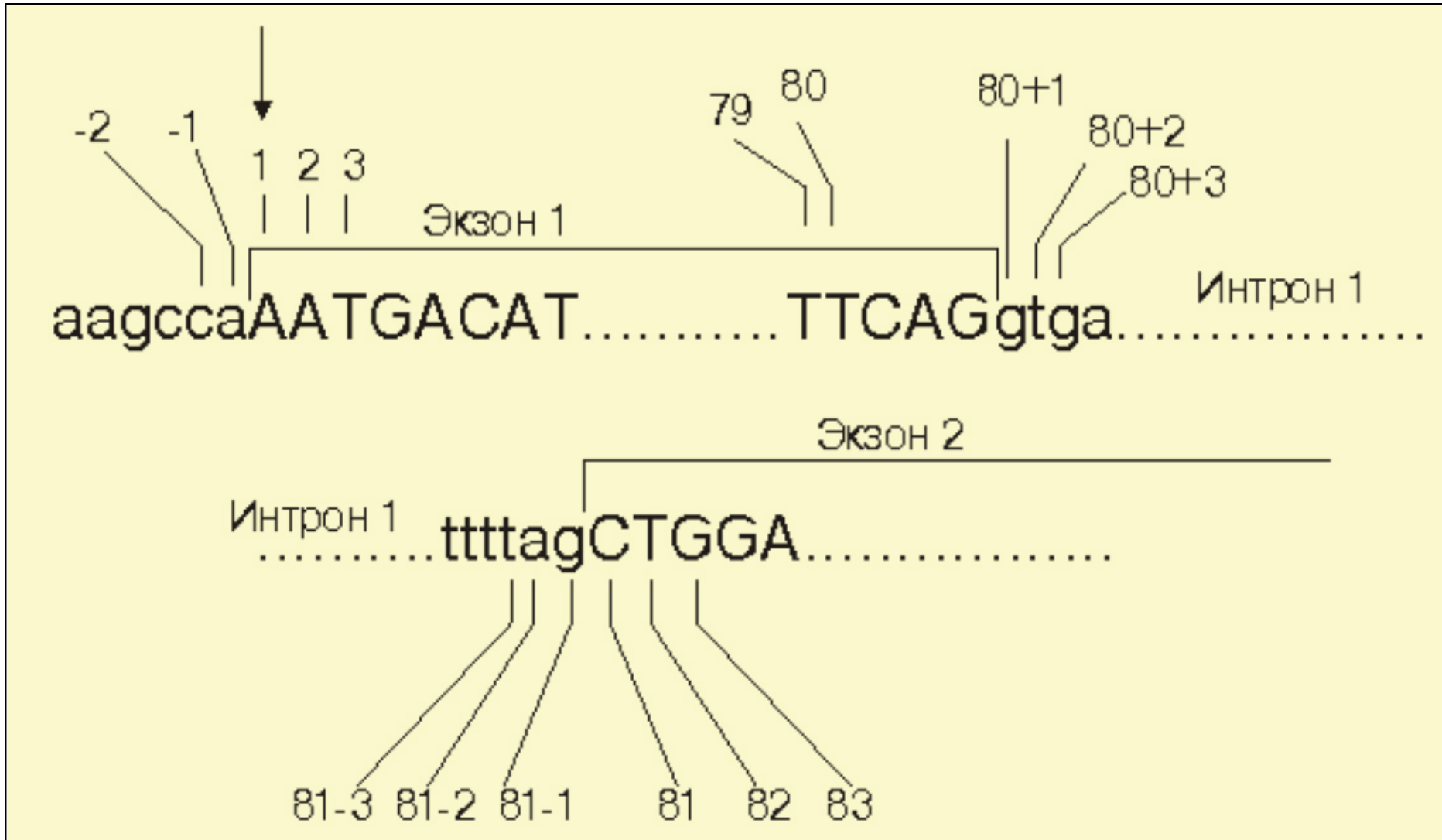
1 – бэнд

.3 – суб-бэнд

# Структура генома



# Нумерация нуклеотидов в гене





# Характеристики генома

- 24 хромосомы
- ~3.2 млрд. пар нуклеотидов
- ~21 тыс. генов
- Один ген — 1000-3000 оснований
- Два генома идентичны на ~99.9 %
- Фенотипические признаки и все заболевания — ~0.1 % генома, т.е. его 1/1000 мутированная часть

**To be continued ...**

