Л е к ц и я № 6

Геномные базы данных
Геномные браузеры

Разработал профессор П. М. Васильев
Кафедра фармакологии и биоинформатики

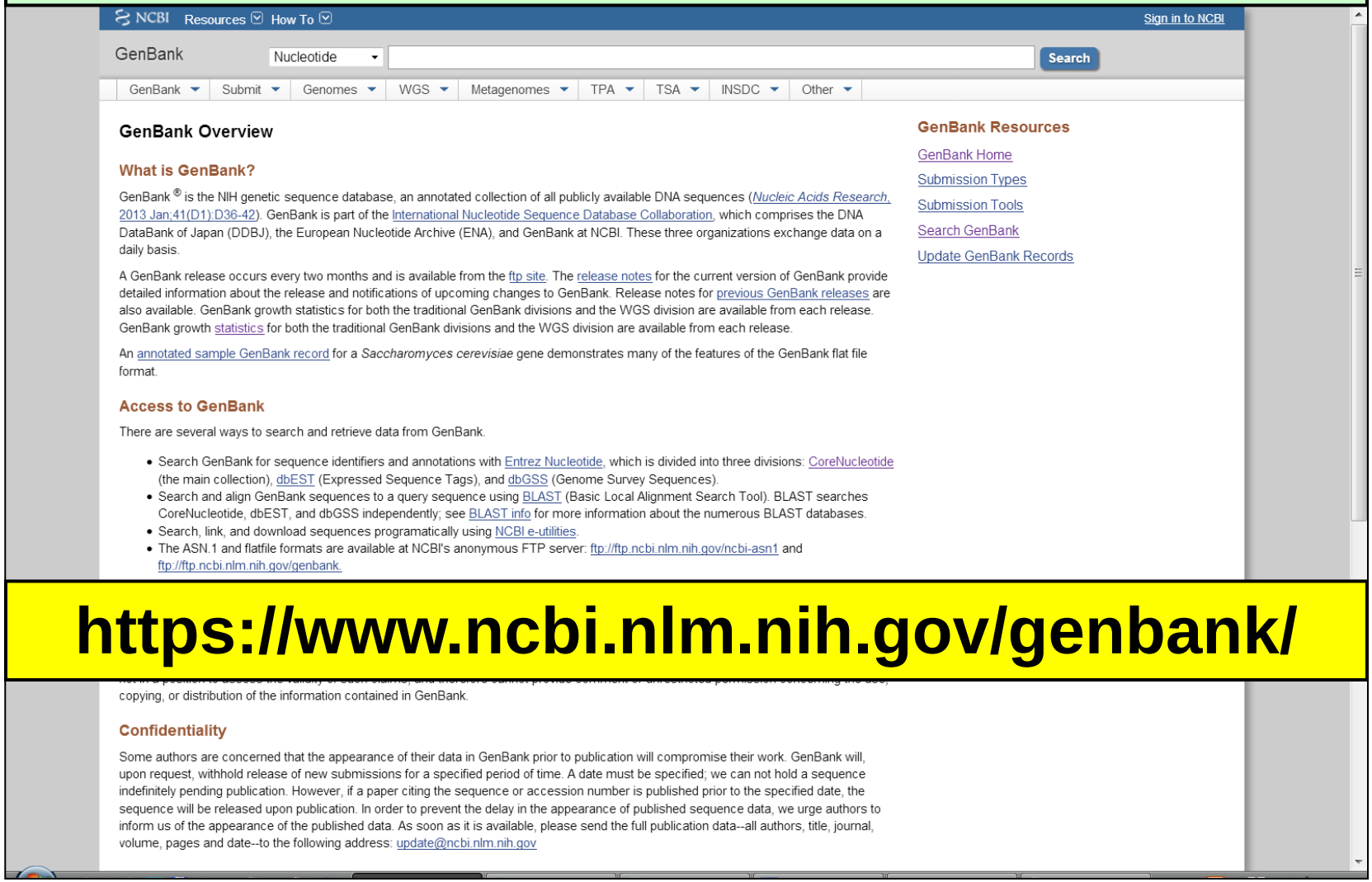
Для студентов, обучающихся по направлению 06.03.01 «Биология»
профили Биохимия, Генетика
при изучении дисциплины «Биоинформатика»

План лекции

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

Геномные базы данных

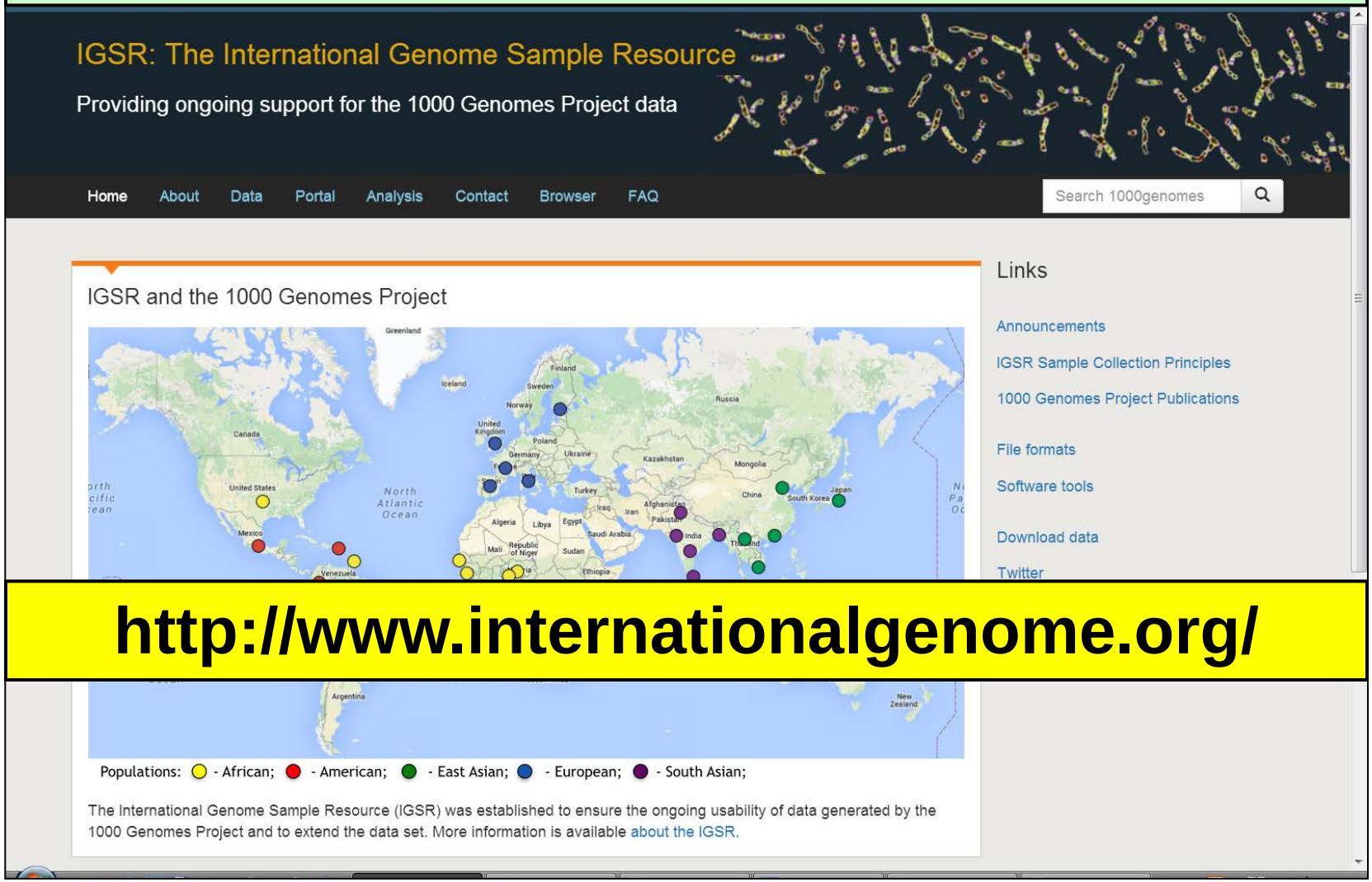
GenBank



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

Геномные базы данных

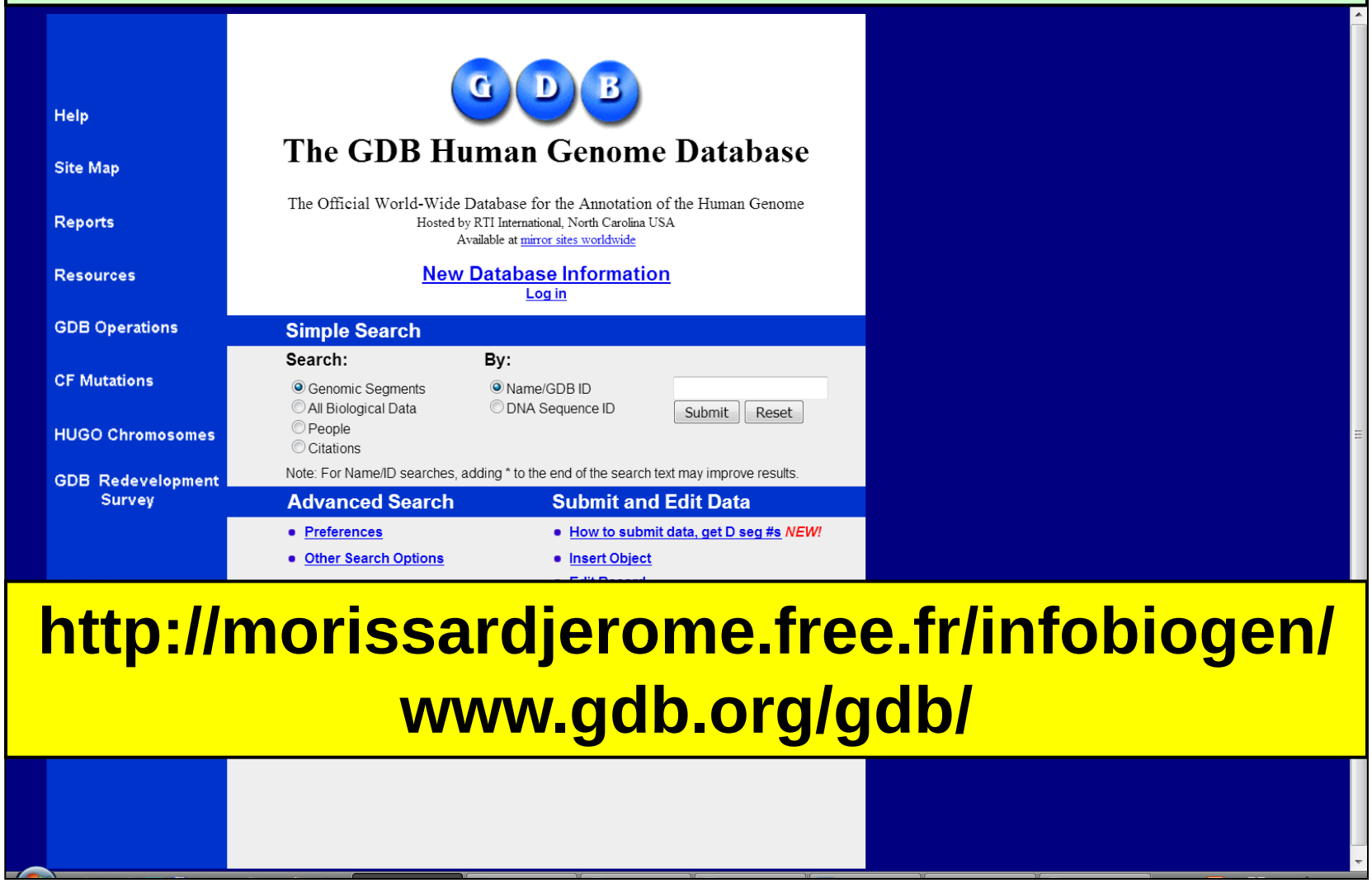
1000 Genomes



http://www.internationalgenome.org/

Геномные базы данных

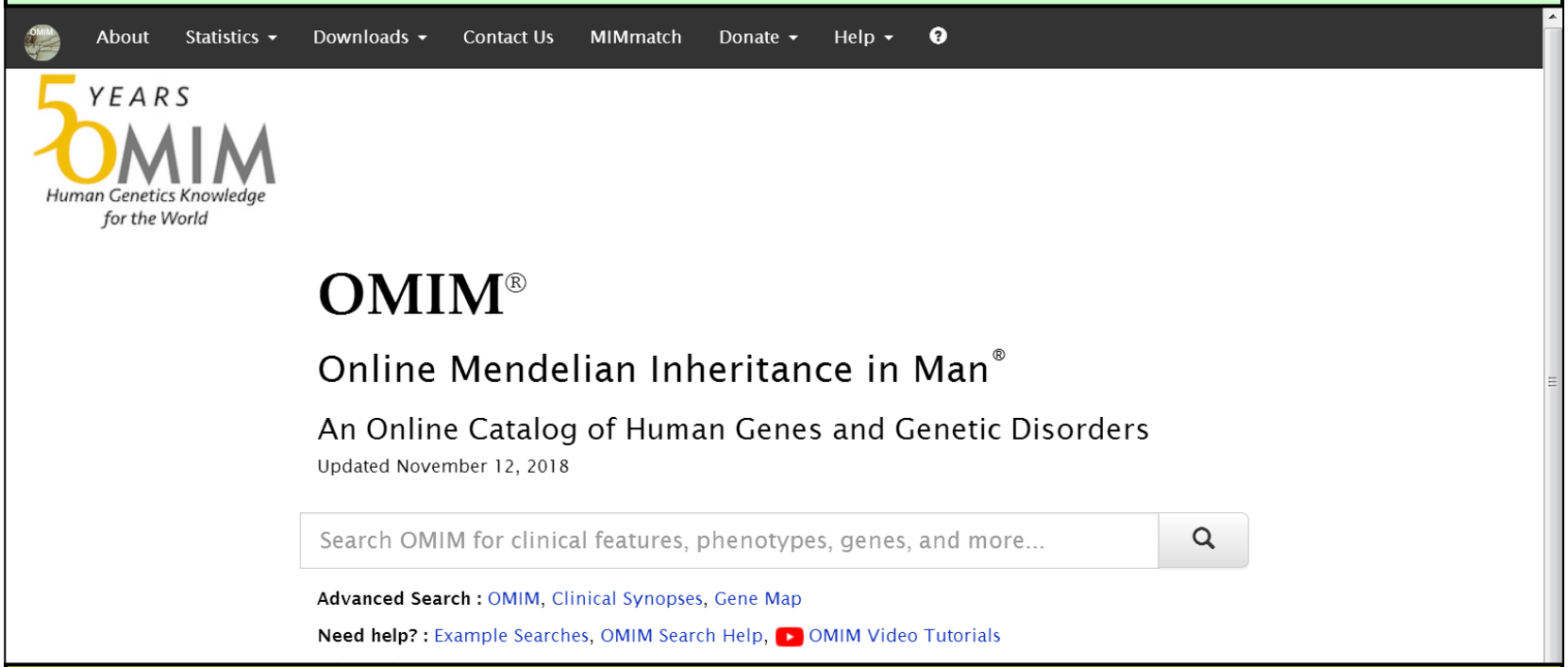
The GDB Human Genome Database



http://morissardjerome.free.fr/infobiogen/
www.gdb.org/gdb/

Геномные базы данных

Online Mendelian Inheritance in Man

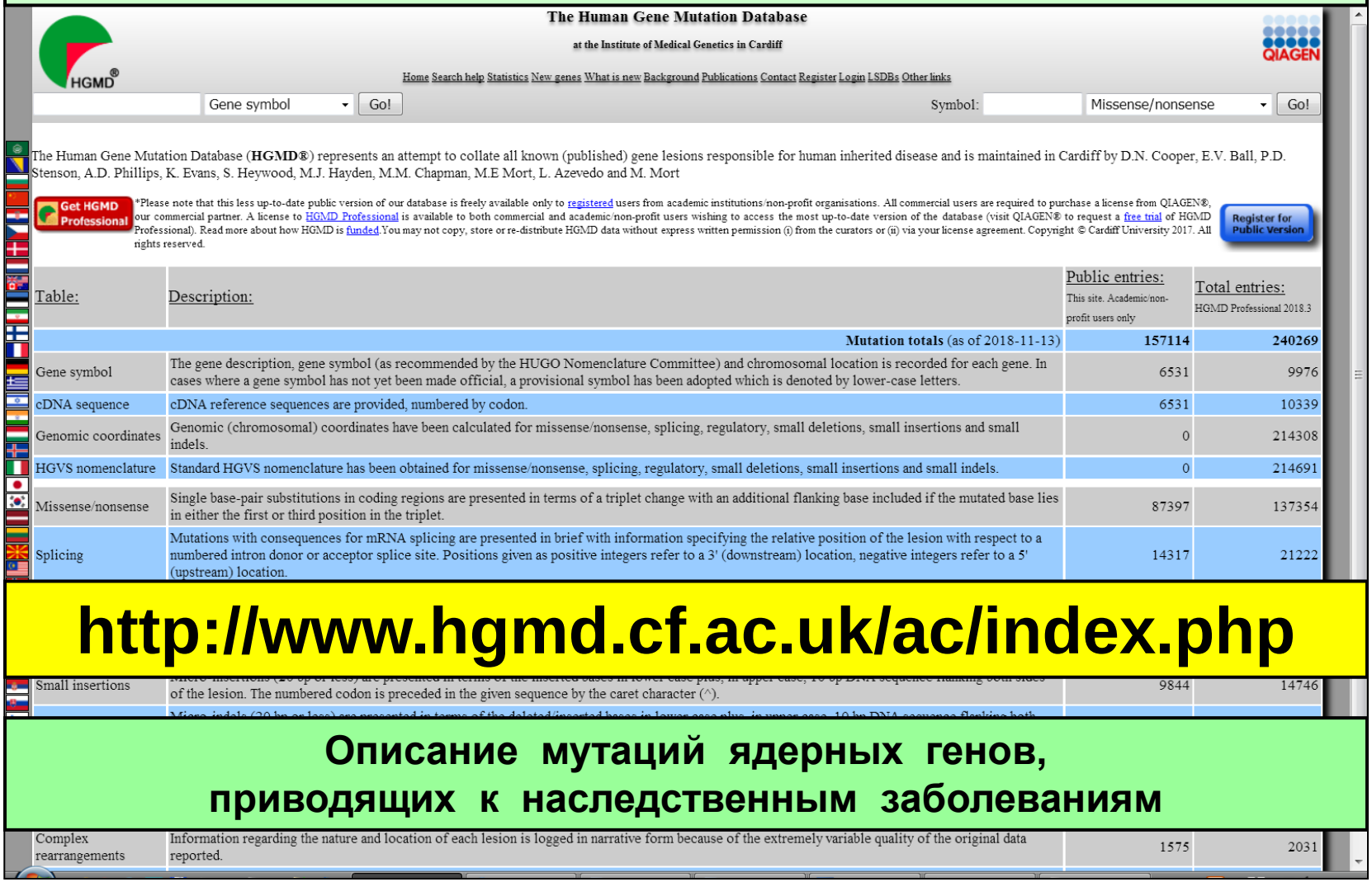


https://www.omim.org/

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.

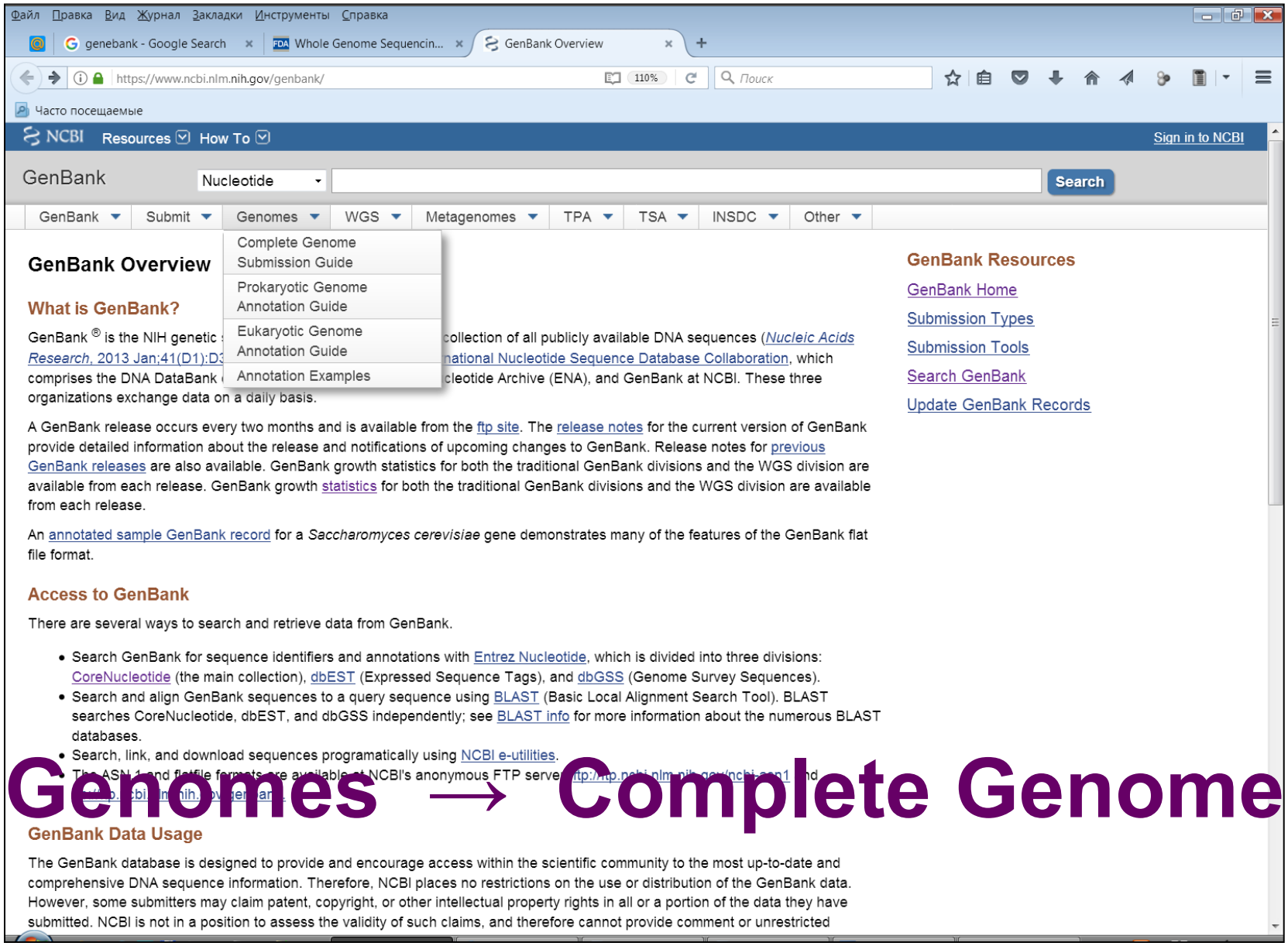
Register for Public Version

Table:	Description:	Public entries: This site. Academic/non-profit users only	Total entries: HGMD Professional 2018.3
	Mutation totals (as of 2018-11-13)	**157114**	**240269**
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	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



Genomes → Complete Genome

GenBank

Файл Правка Вид Журнал Закладки Инструменты Справка

genebank - Google Search | Whole Genome Sequencin... | Prokaryotic and Eukaryotic ...

https://www.ncbi.nlm.nih.gov/genbank/genomesubmit/

Часто посещаемые

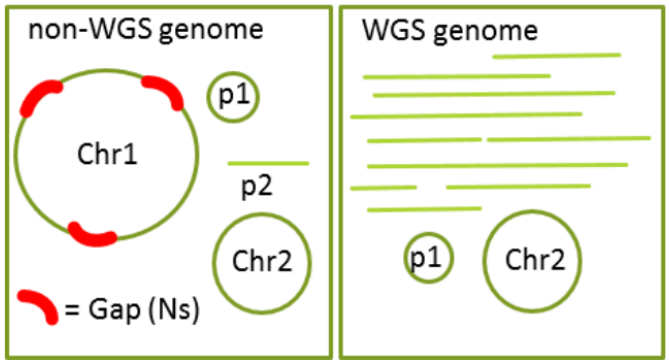
NCBI Resources How To Sign in to NCBI

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

- Each chromosome is in a single sequence and there are no extra sequences
- Each sequence in the genome must be assigned to a chromosome or 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

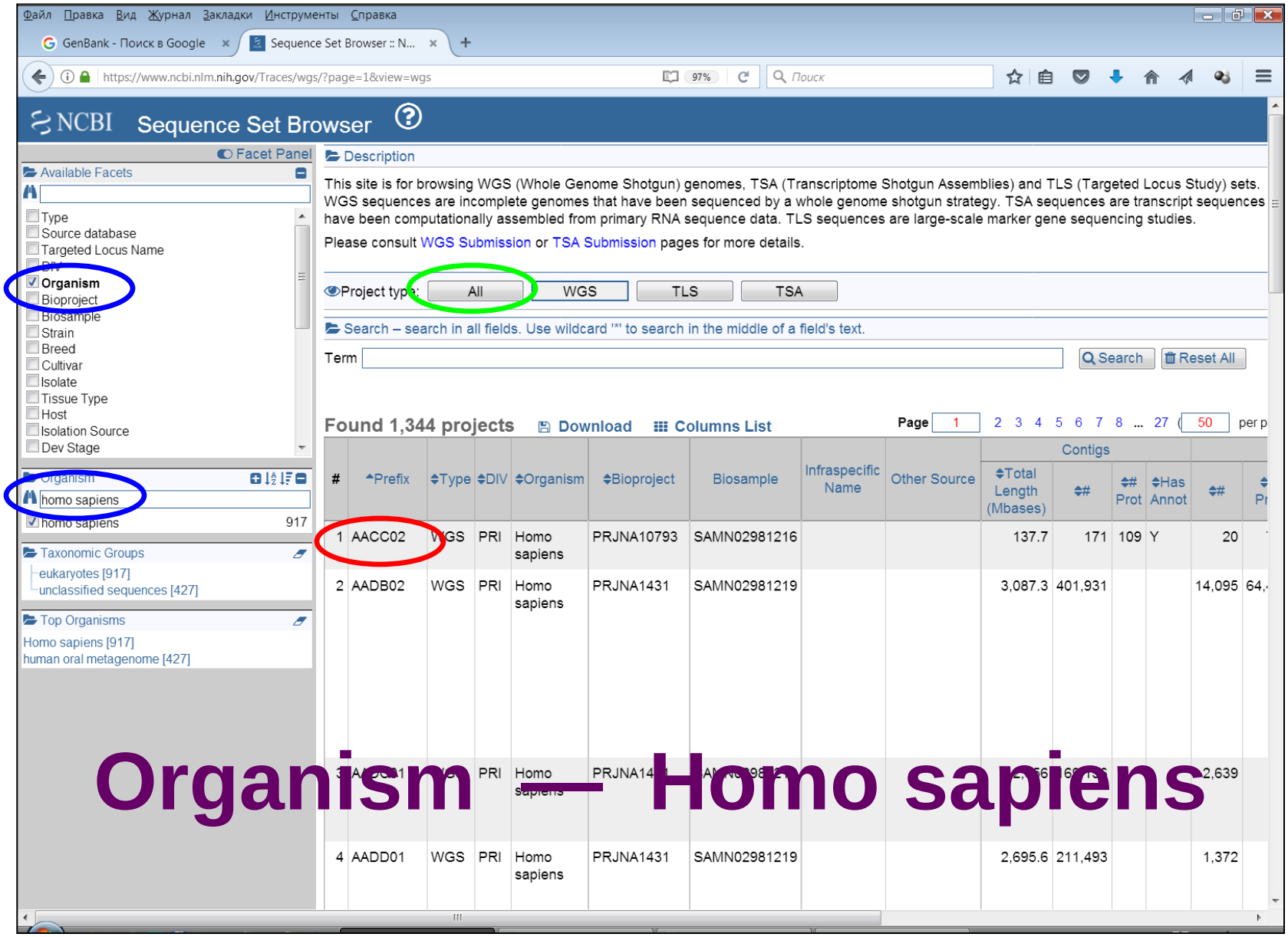
- There can still be gaps within the sequences; you will supply that information in the submission
- Plasmids and organelles can still be in multiple pieces.
- Internal sequences must be arranged in the correct order and orientation.
- Sequences concatenated in unknown order are not allowed.

Genome Resources

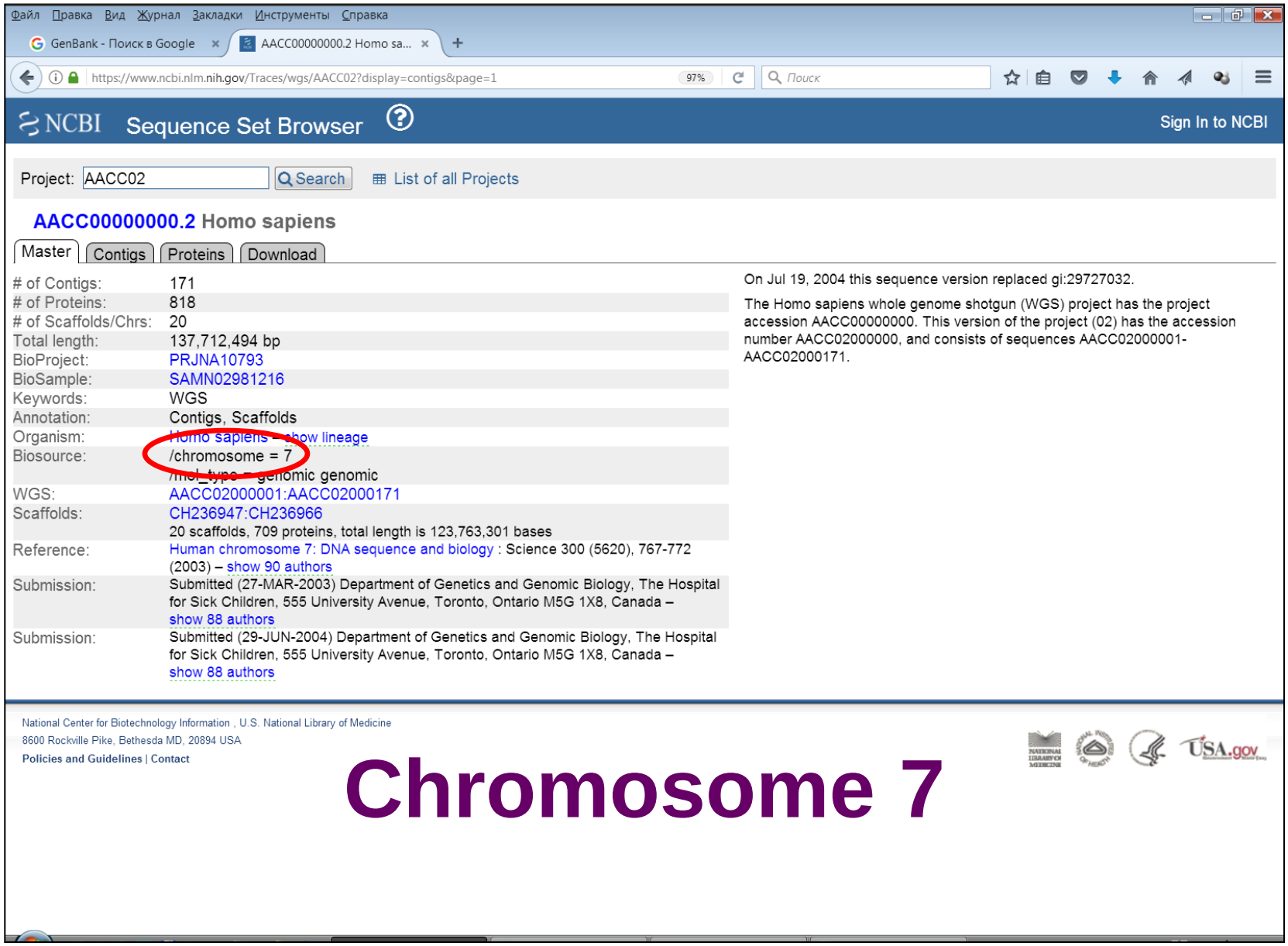
About WGS
WGS Browser
Genome Submission Guide
Genome Submission Portal
Update Genome Records
FAQ
tbl2asn
Create Submission Template
Eukaryotic Annotation Guide
Prokaryotic Annotation Guide
Annotation Example Files
Discrepancy Report
NCBI Prokaryotic Genome Annotation Pipeline
AGP Format
Complex Assembly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

WGS Browser (Whole Genome Shotgun)

GenBank — WGS Browser

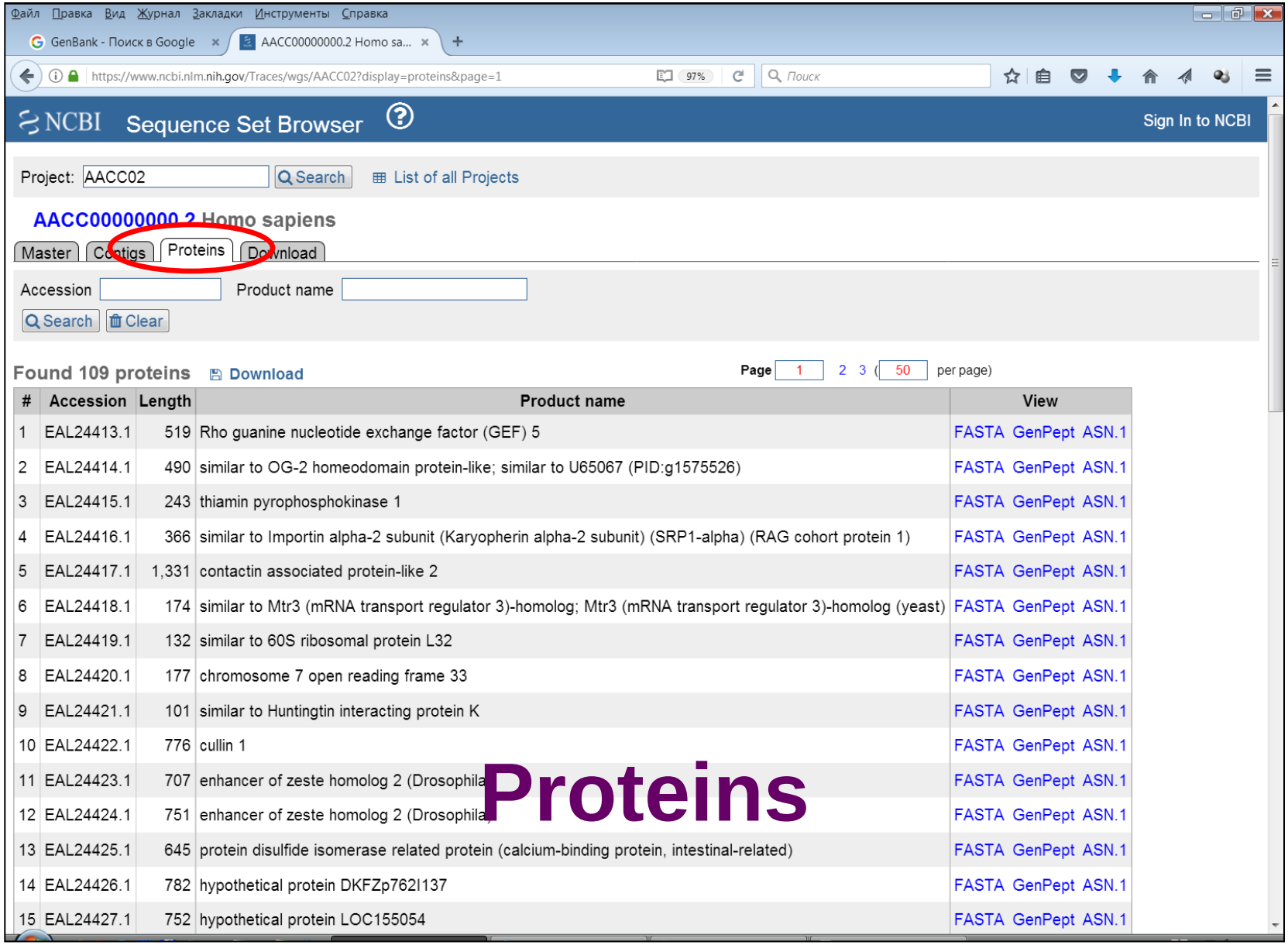


GenBank — Project: AACC02



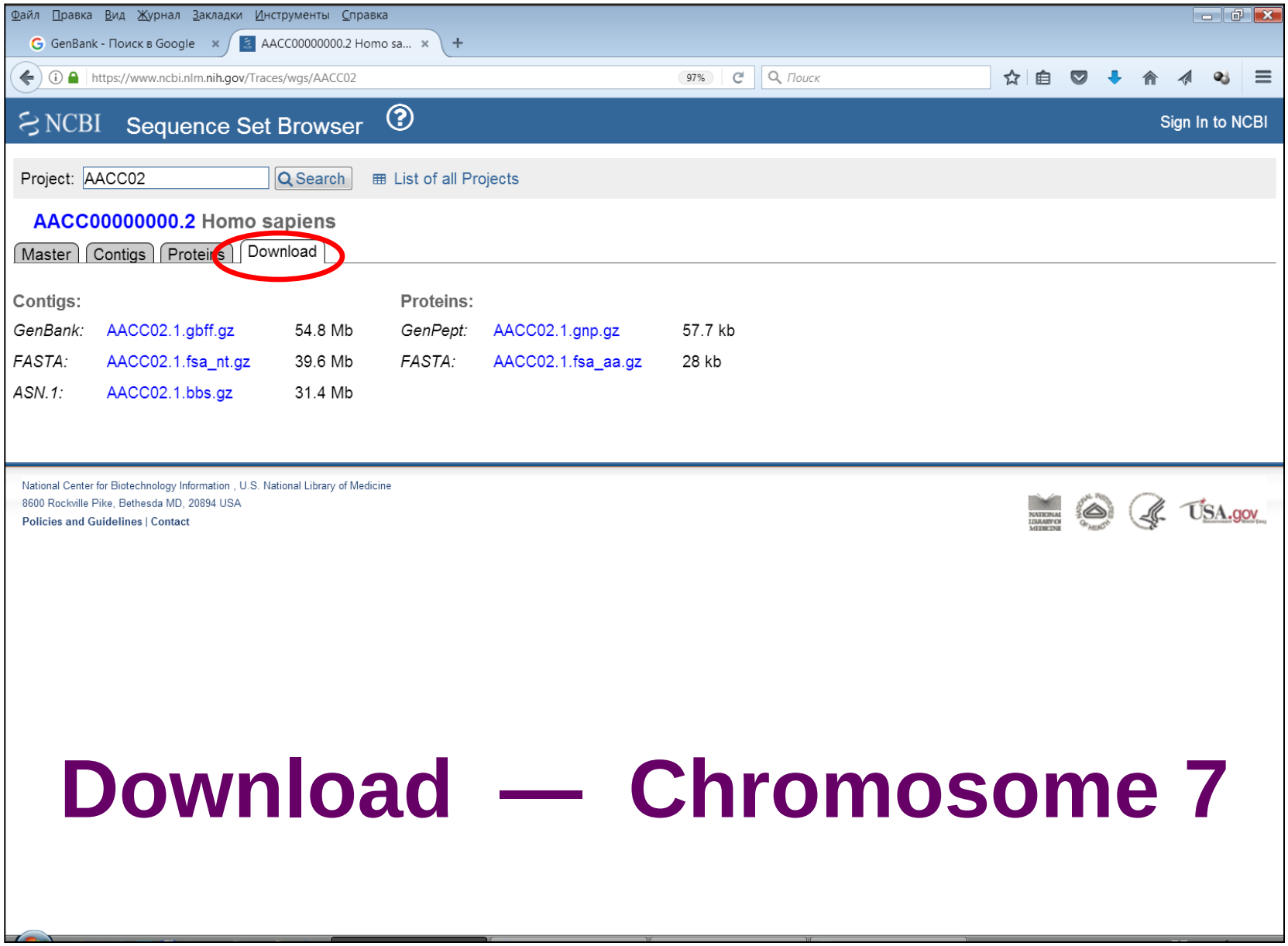
Chromosome 7

GenBank — Project: AACC02



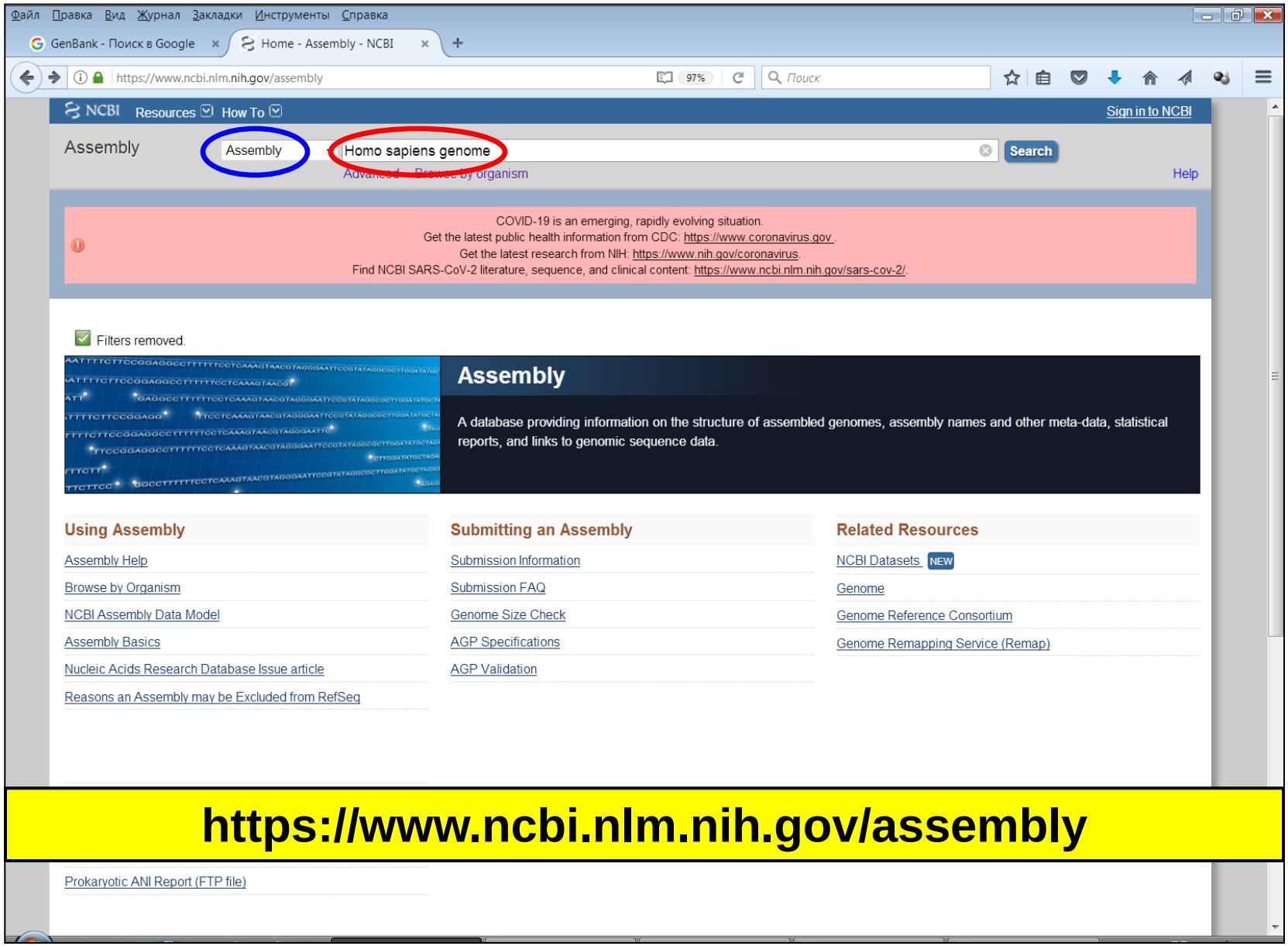
#	Accession	Length	Product name	View
1	EAL24413.1	519	Rho guanine nucleotide exchange factor (GEF) 5	FASTA GenPept ASN.1
2	EAL24414.1	490	similar to OG-2 homeodomain protein-like; similar to U65067 (PID:g1575526)	FASTA GenPept ASN.1
3	EAL24415.1	243	thiamin pyrophosphokinase 1	FASTA GenPept ASN.1
4	EAL24416.1	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 DKFZp762I137	FASTA GenPept ASN.1
15	EAL24427.1	752	hypothetical protein LOC155054	FASTA GenPept ASN.1

GenBank — Project: AACC02



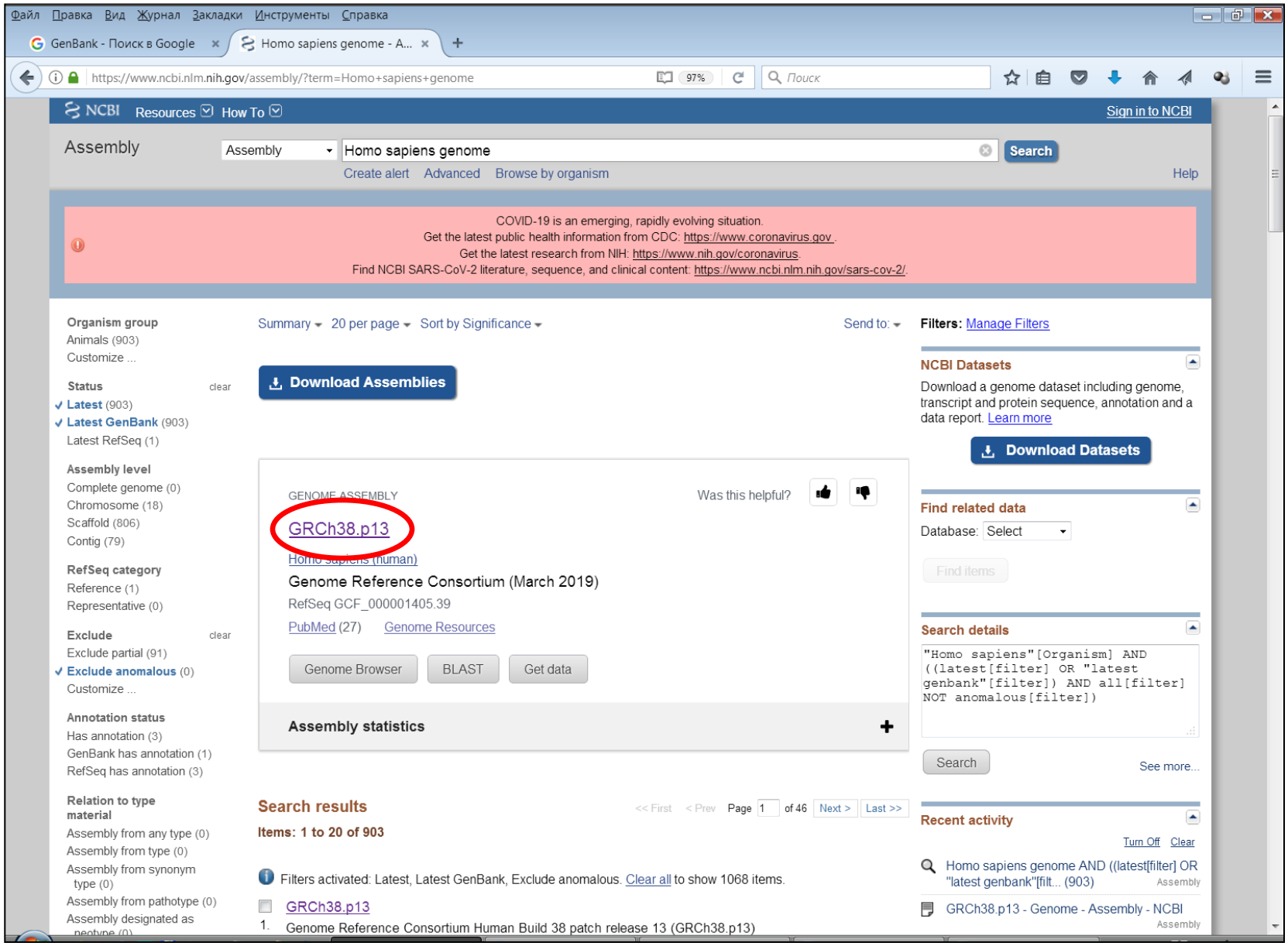
Download — Chromosome 7

GenBank — Reference Genom

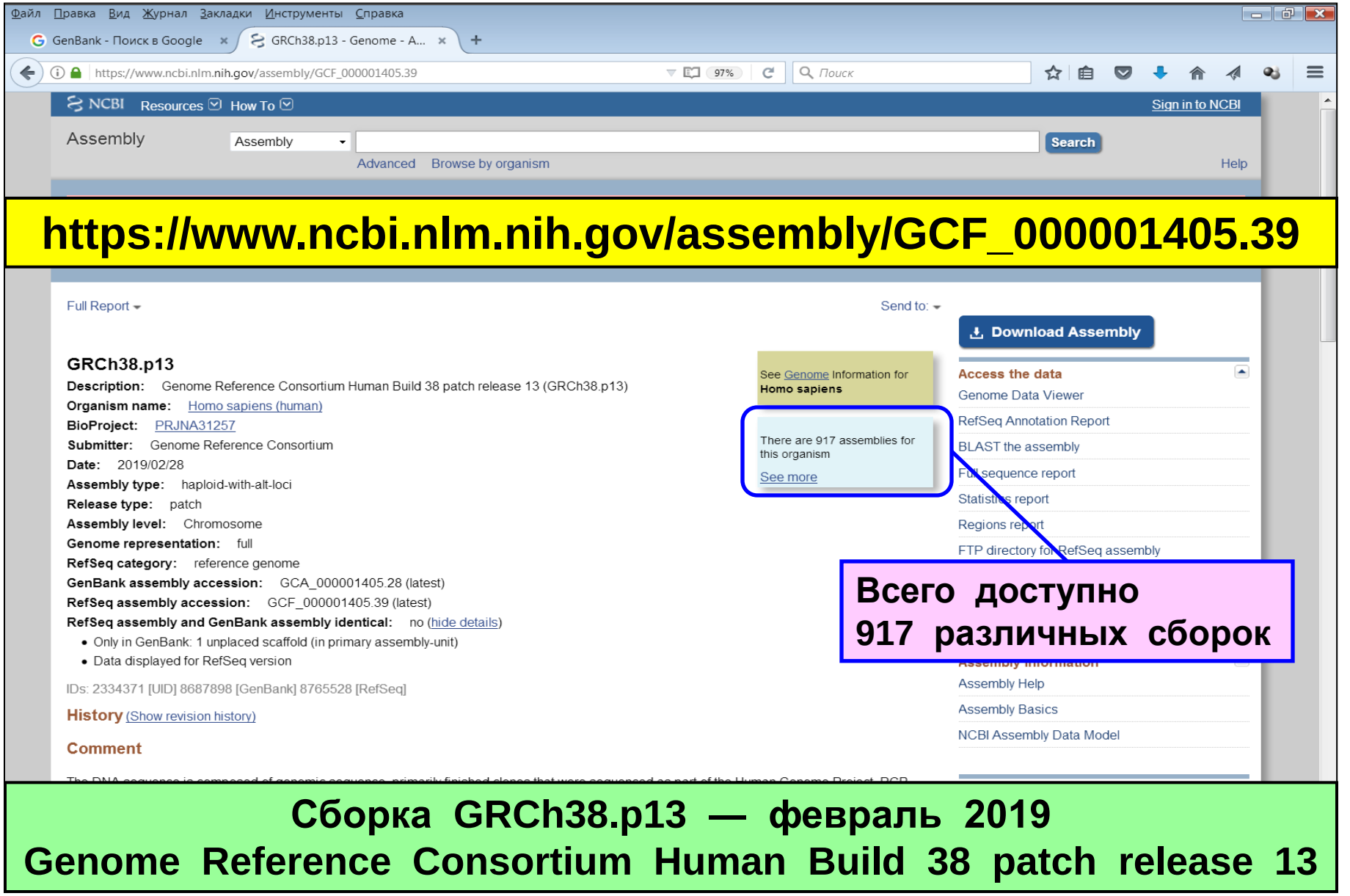


https://www.ncbi.nlm.nih.gov/assembly

GenBank — Reference Genom

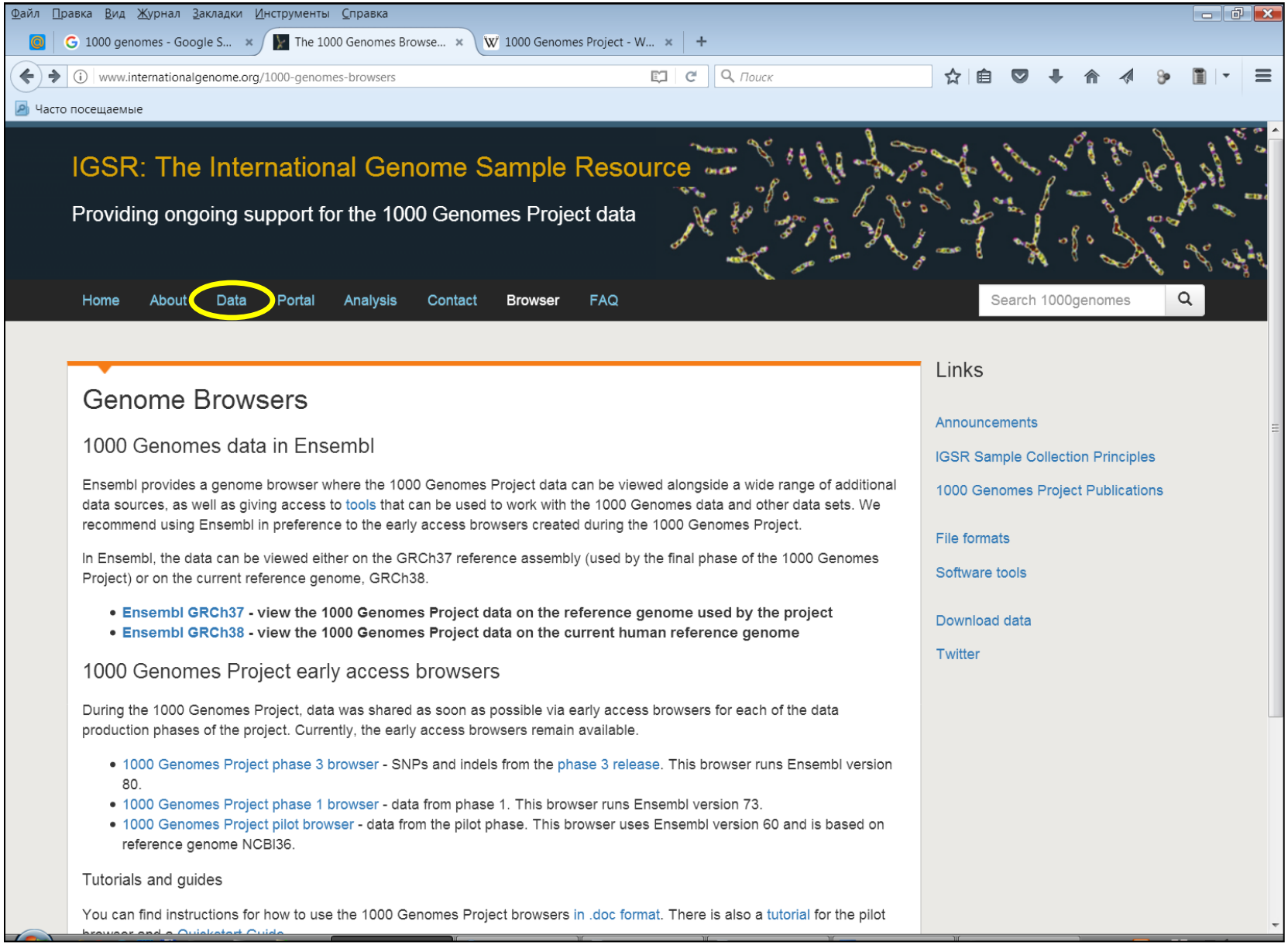


Референсный геном человека

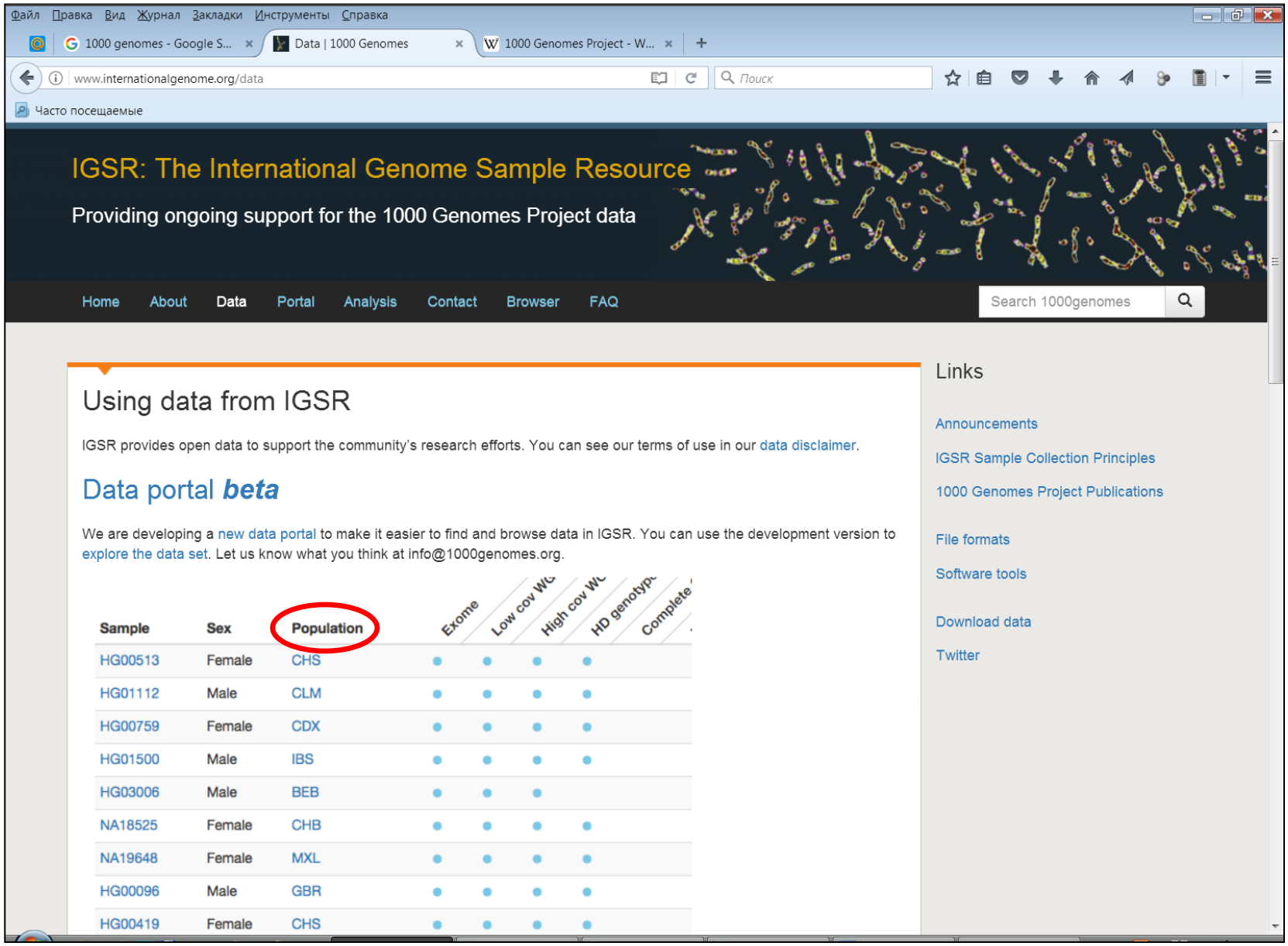


Сборка GRCh38.p13 — февраль 2019
Genome Reference Consortium Human Build 38 patch release 13

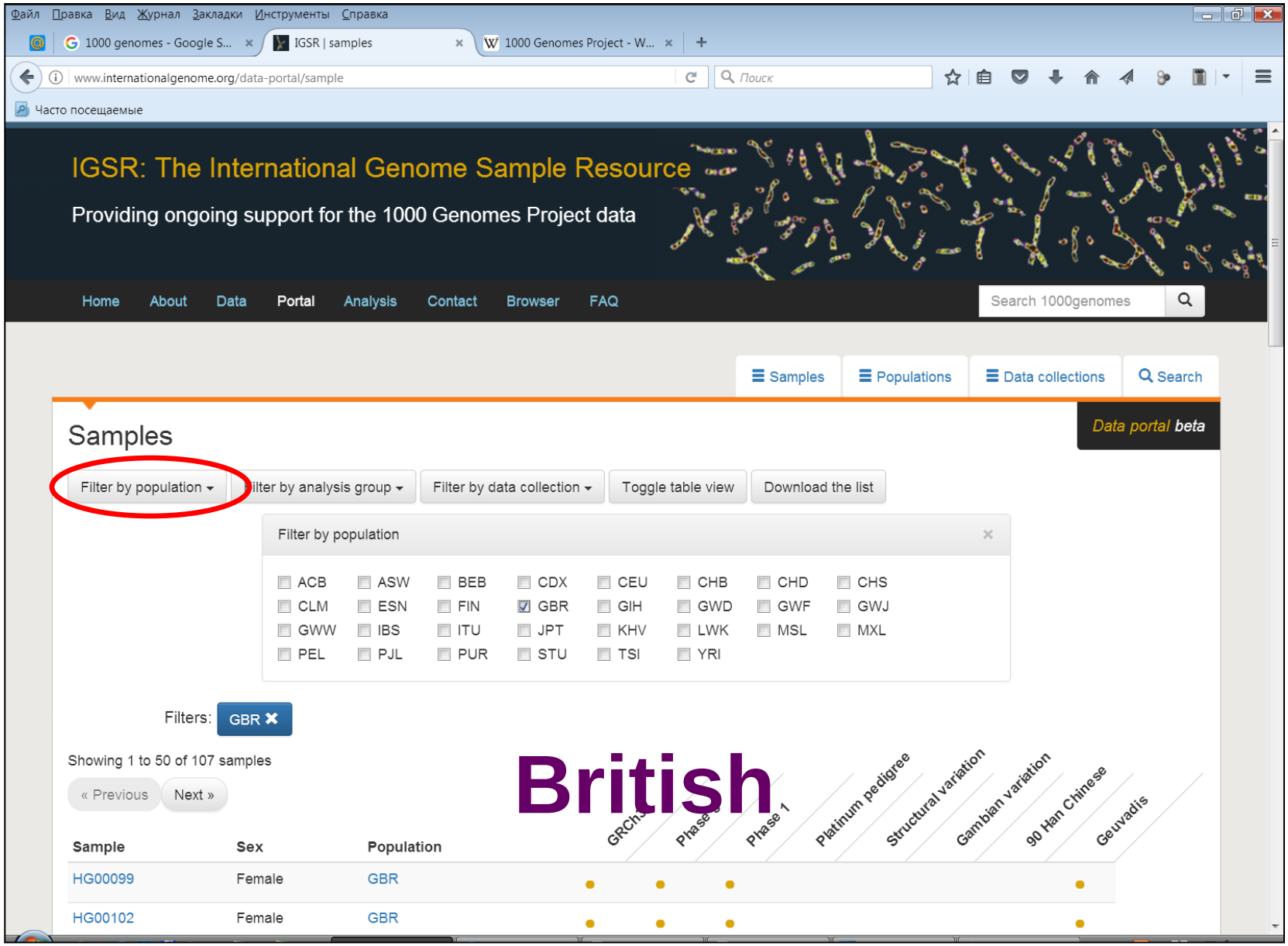
1000 Genomes



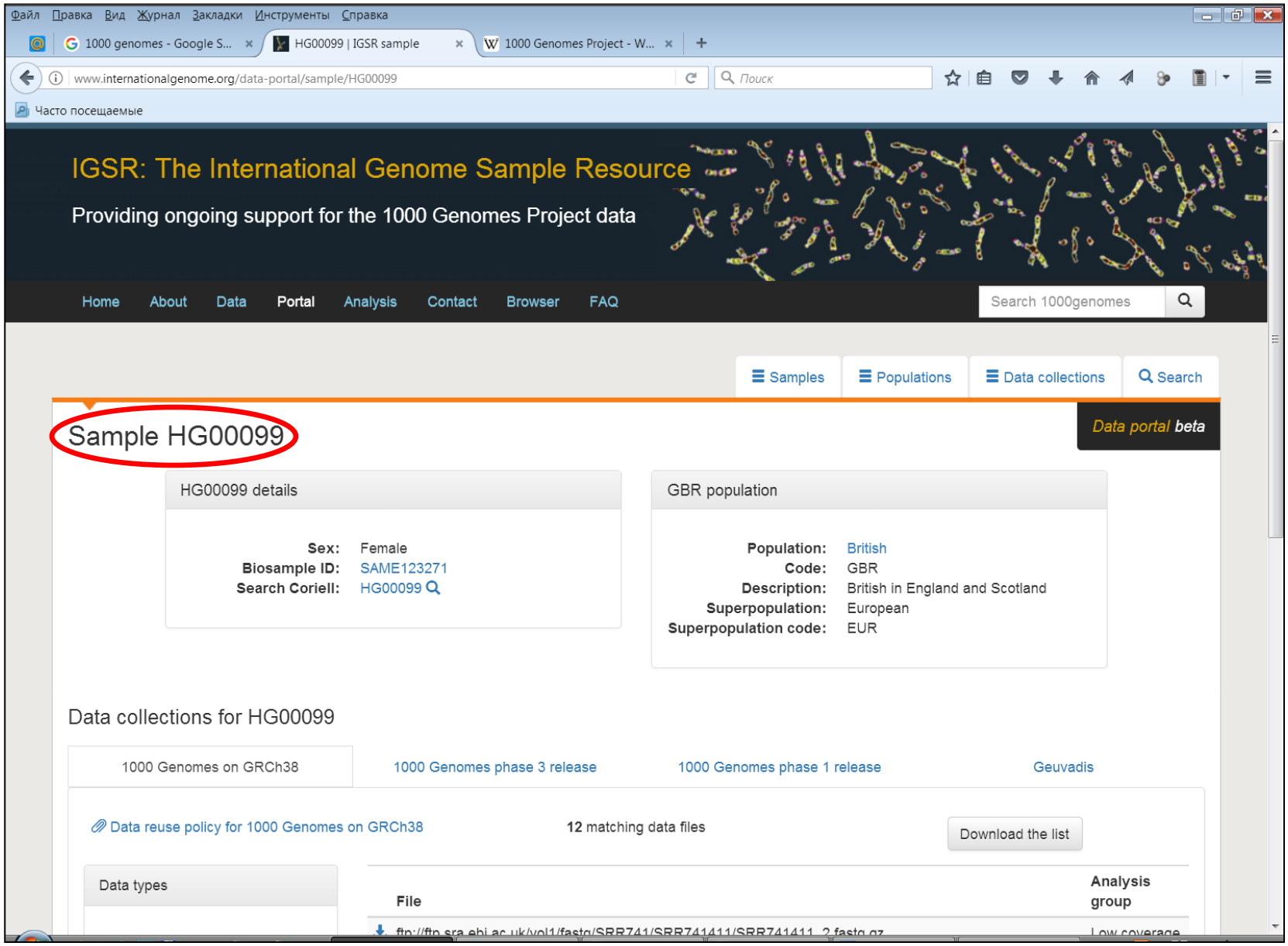
1000 Genomes — Data



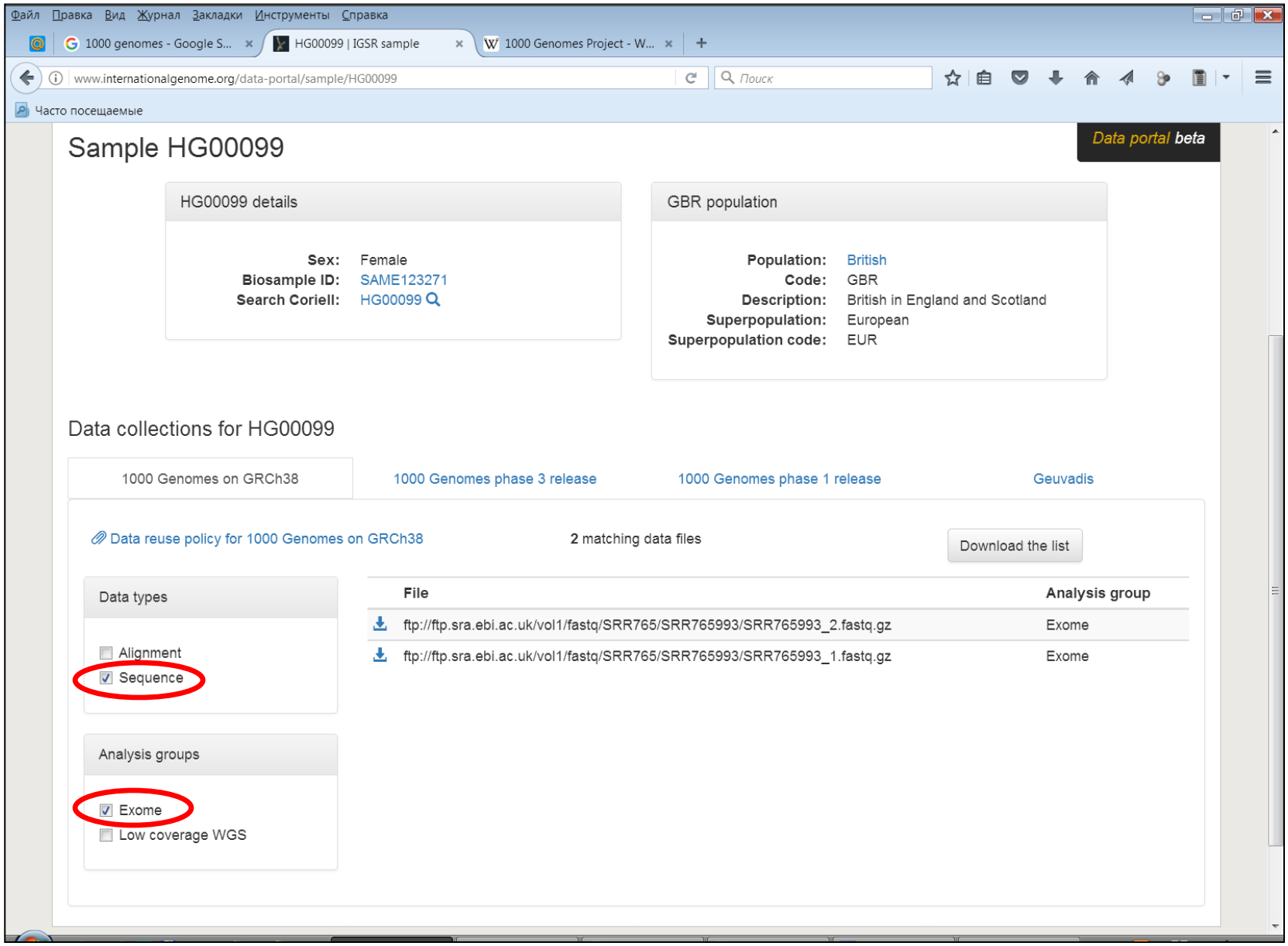
1000 Genomes — Population



1000 Genomes — Sample HG00099

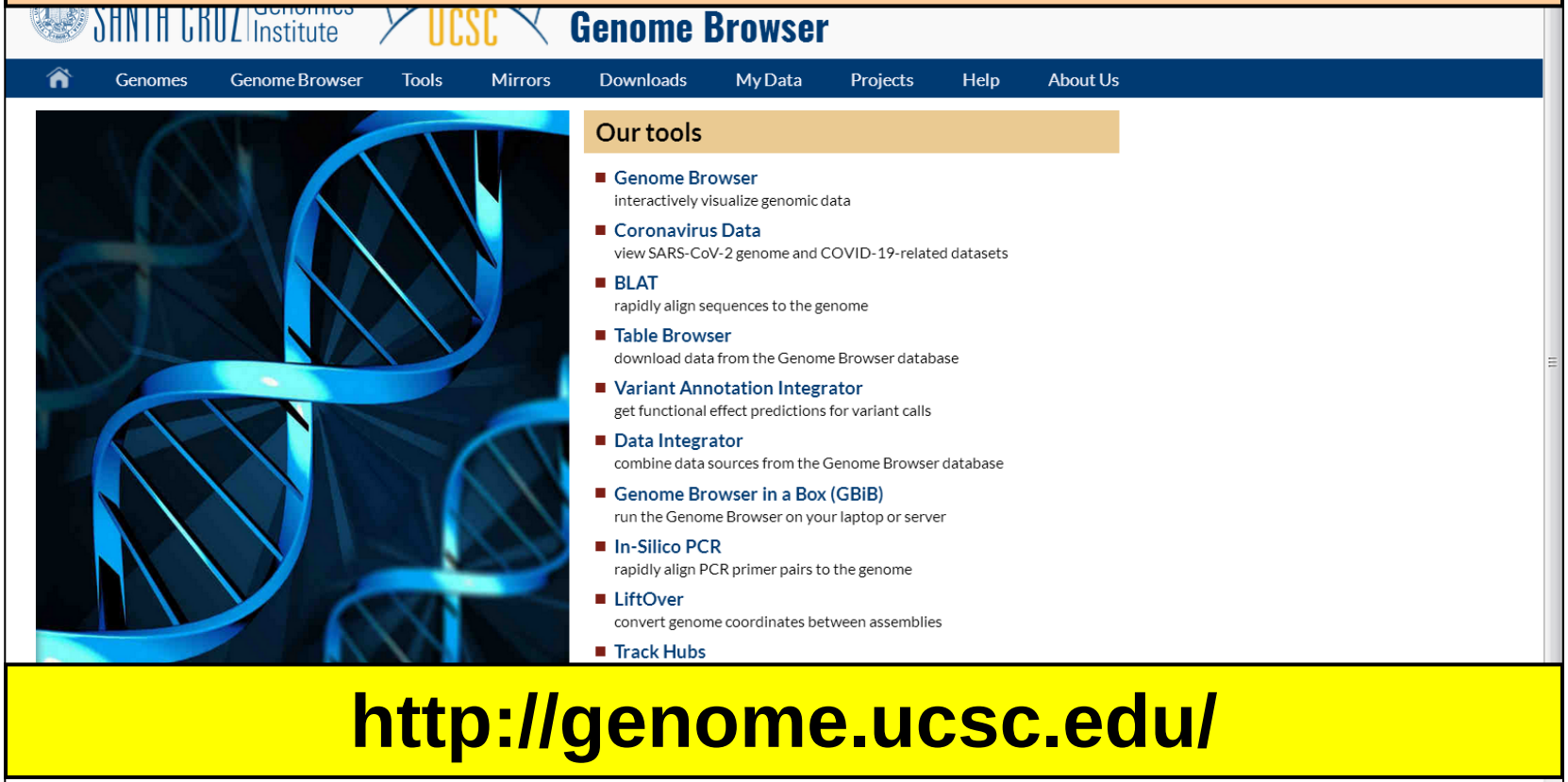


1000 Genomes — Exome



Геномные браузеры

UCSC Genome Browser



http://genome.ucsc.edu/

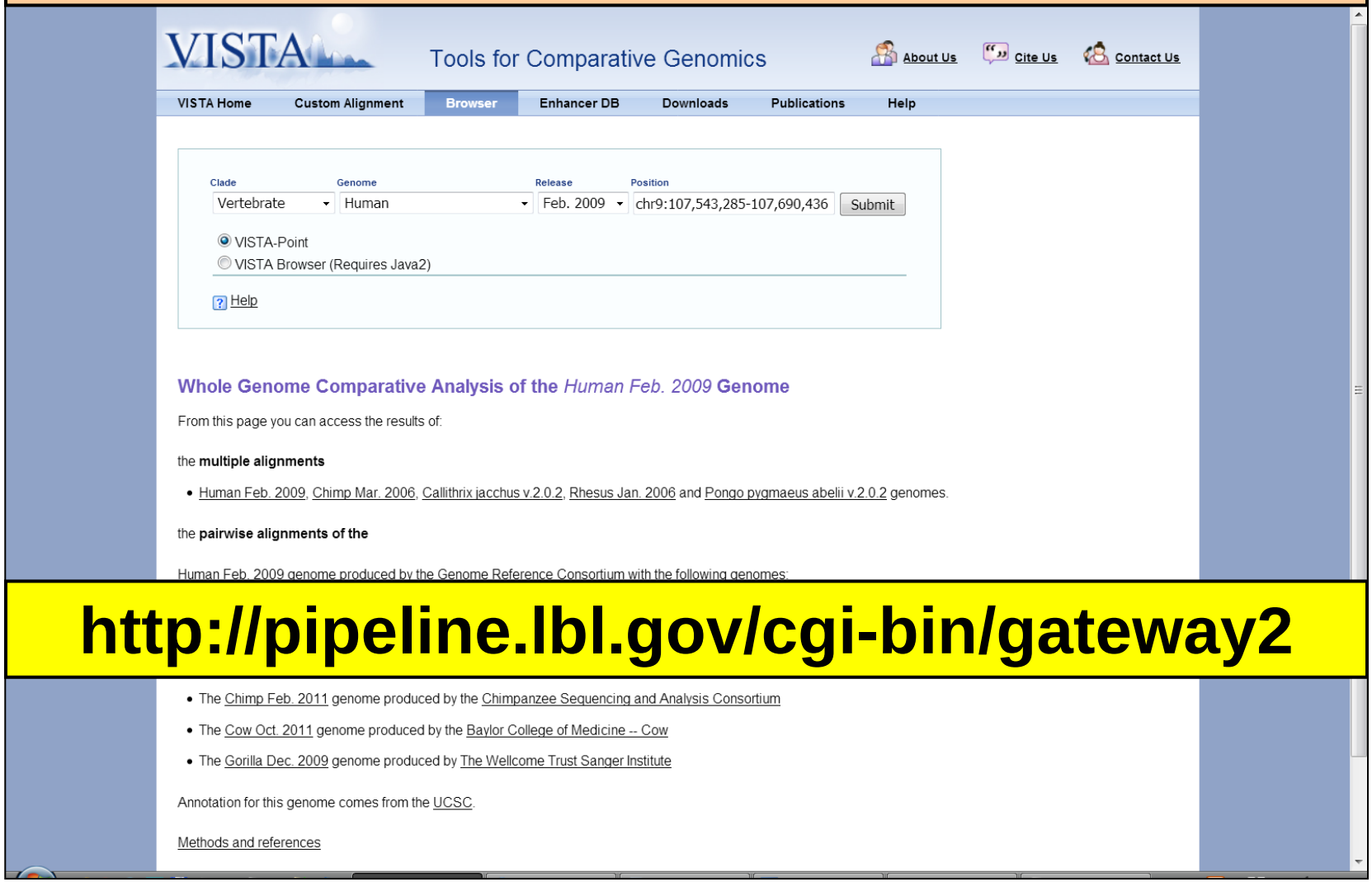
Геномные браузеры

Ensembl Genome Browser

https://www.ensembl.org/index.html

Геномные браузеры

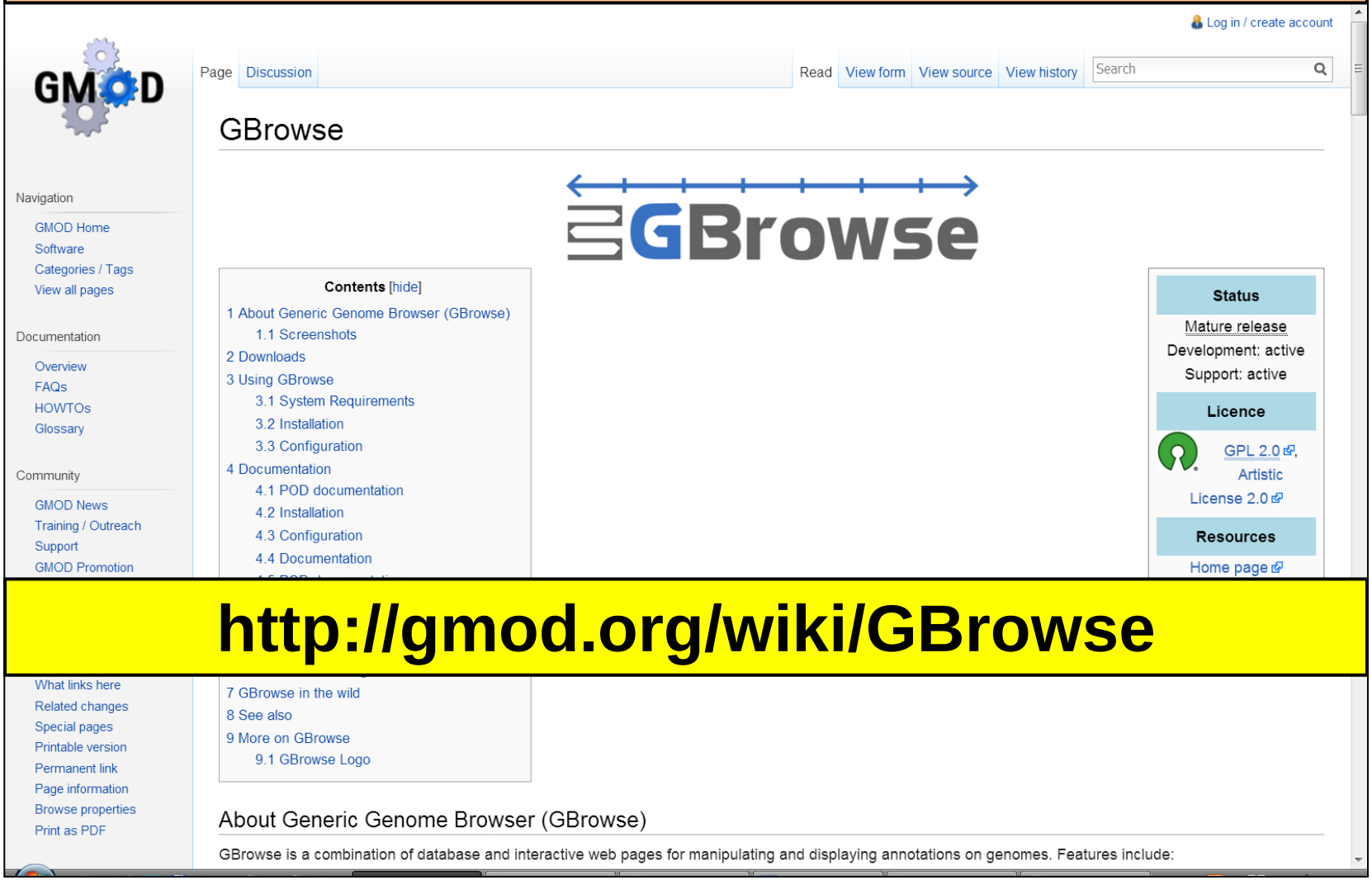
Vista



http://pipeline.lbl.gov/cgi-bin/gateway2

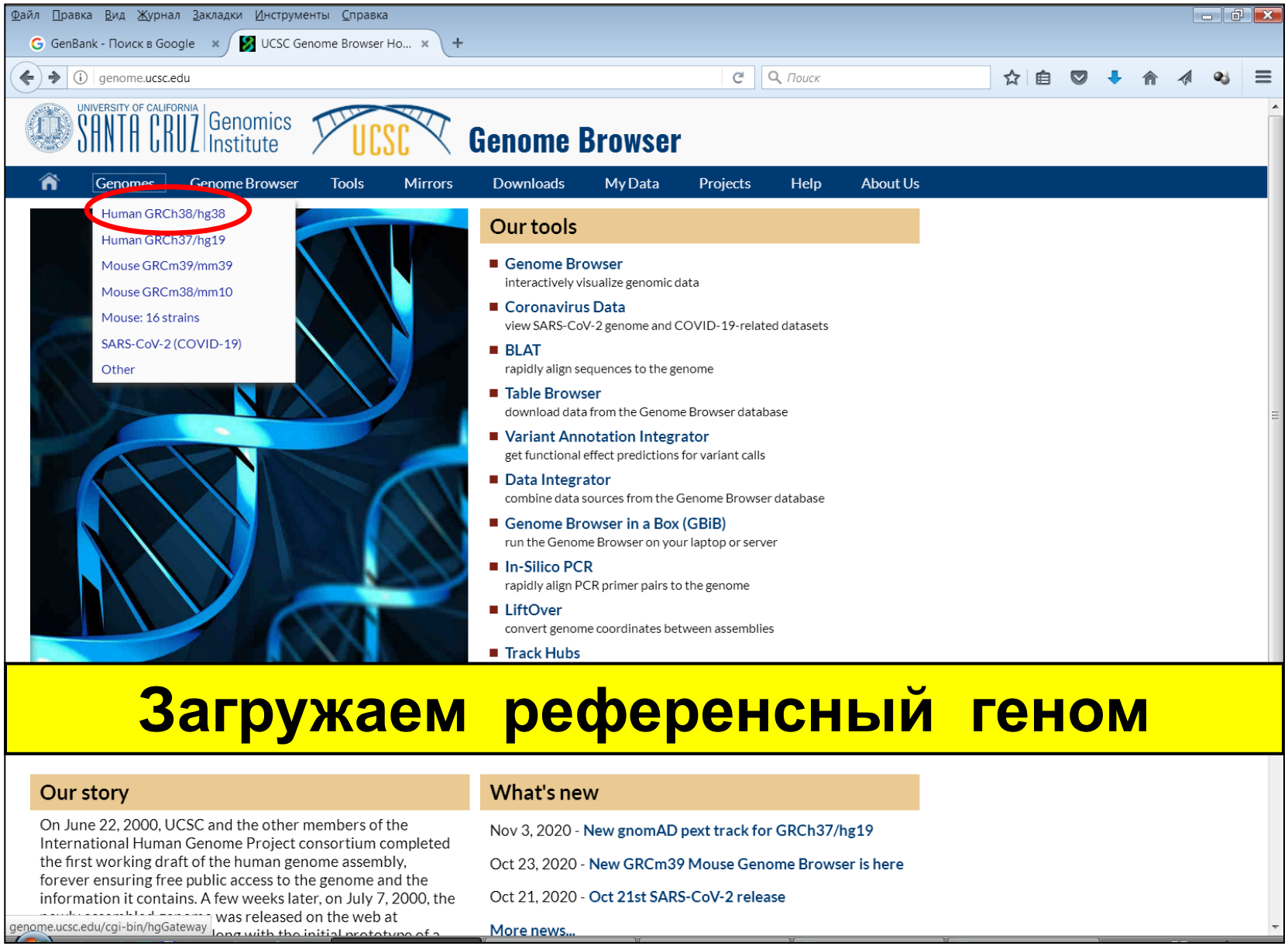
Геномные браузеры

GBrowse

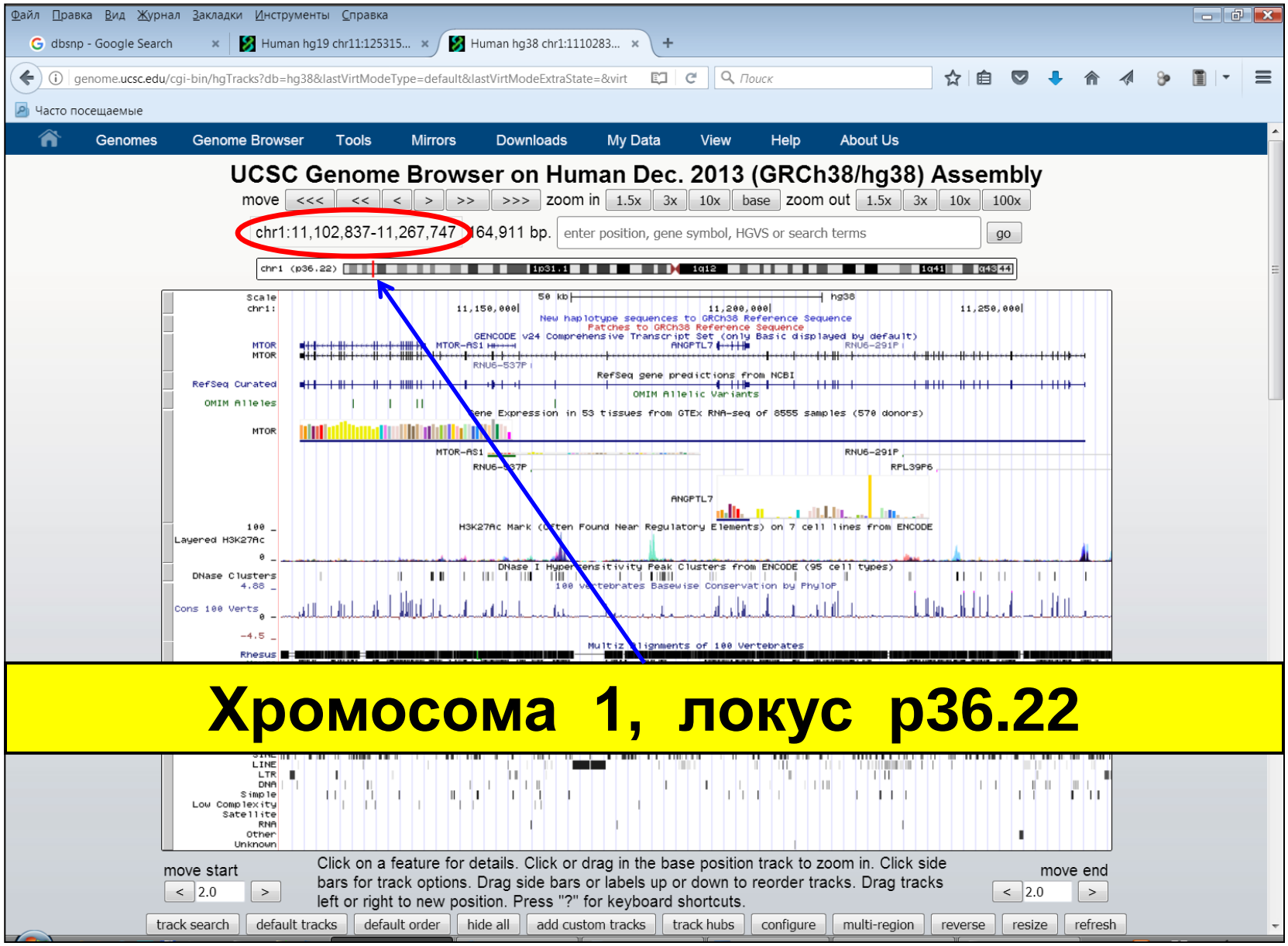


http://gmod.org/wiki/GBrowse

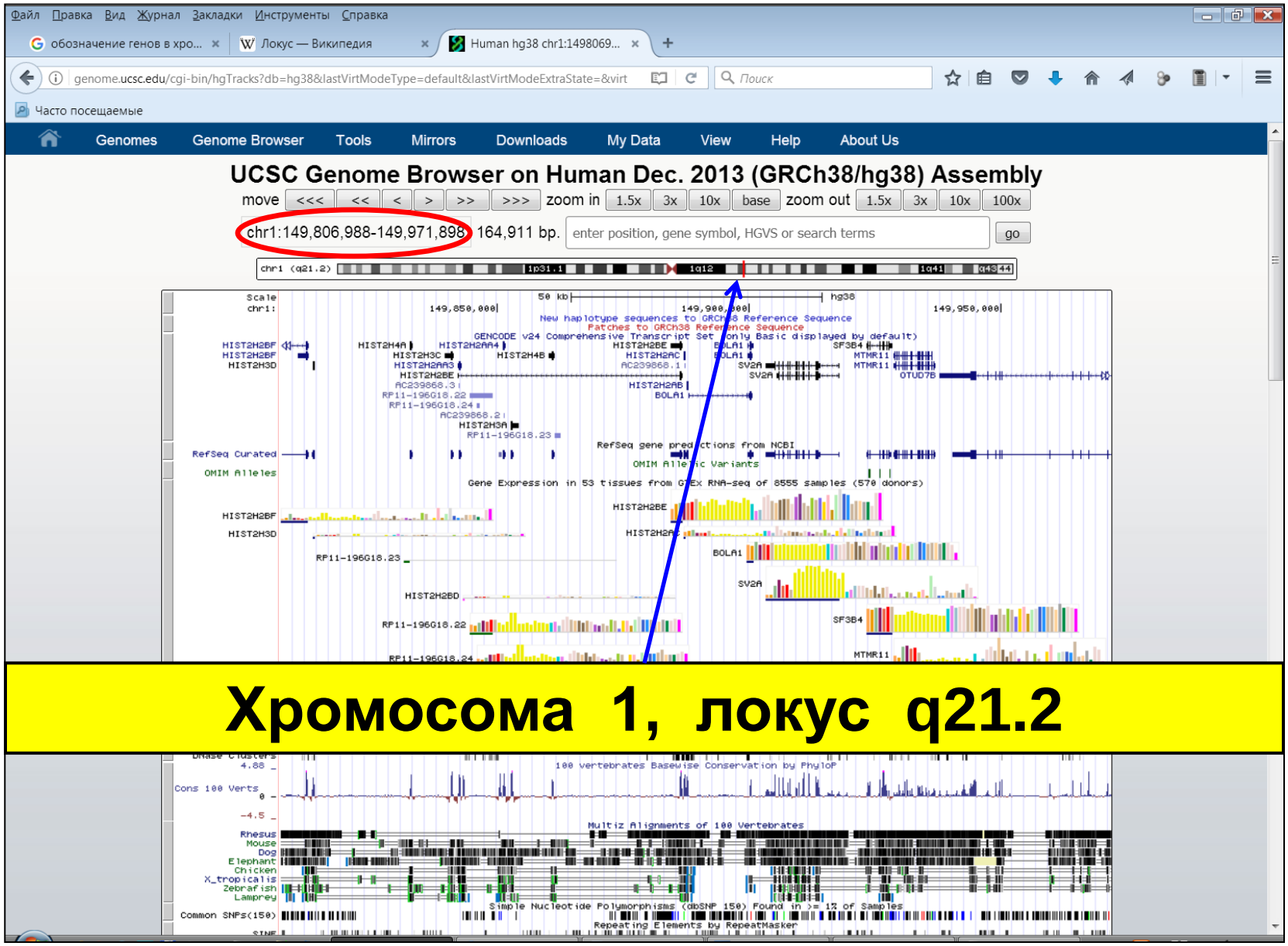
UCSC Genome Browser



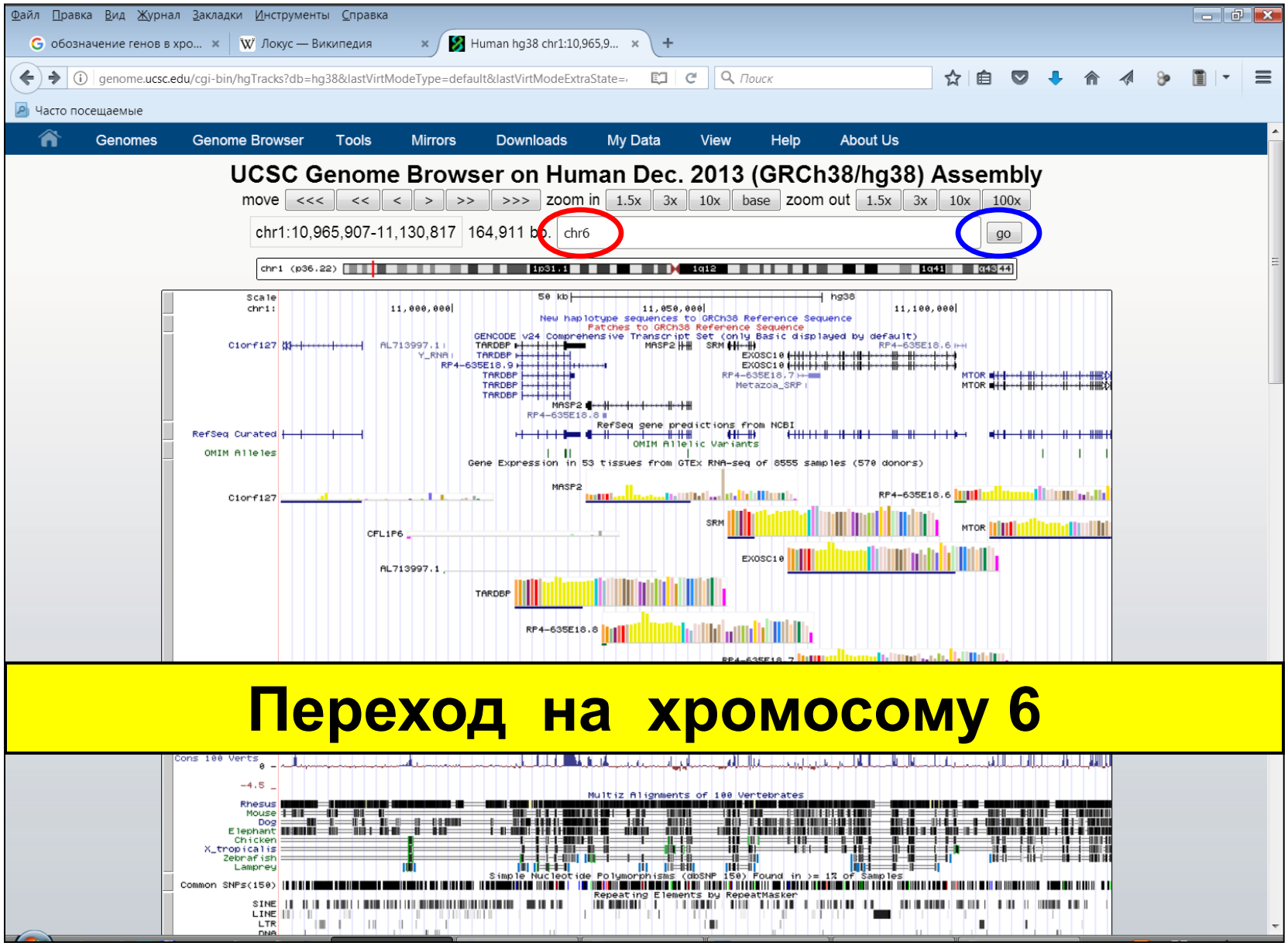
UCSC Genome Browser



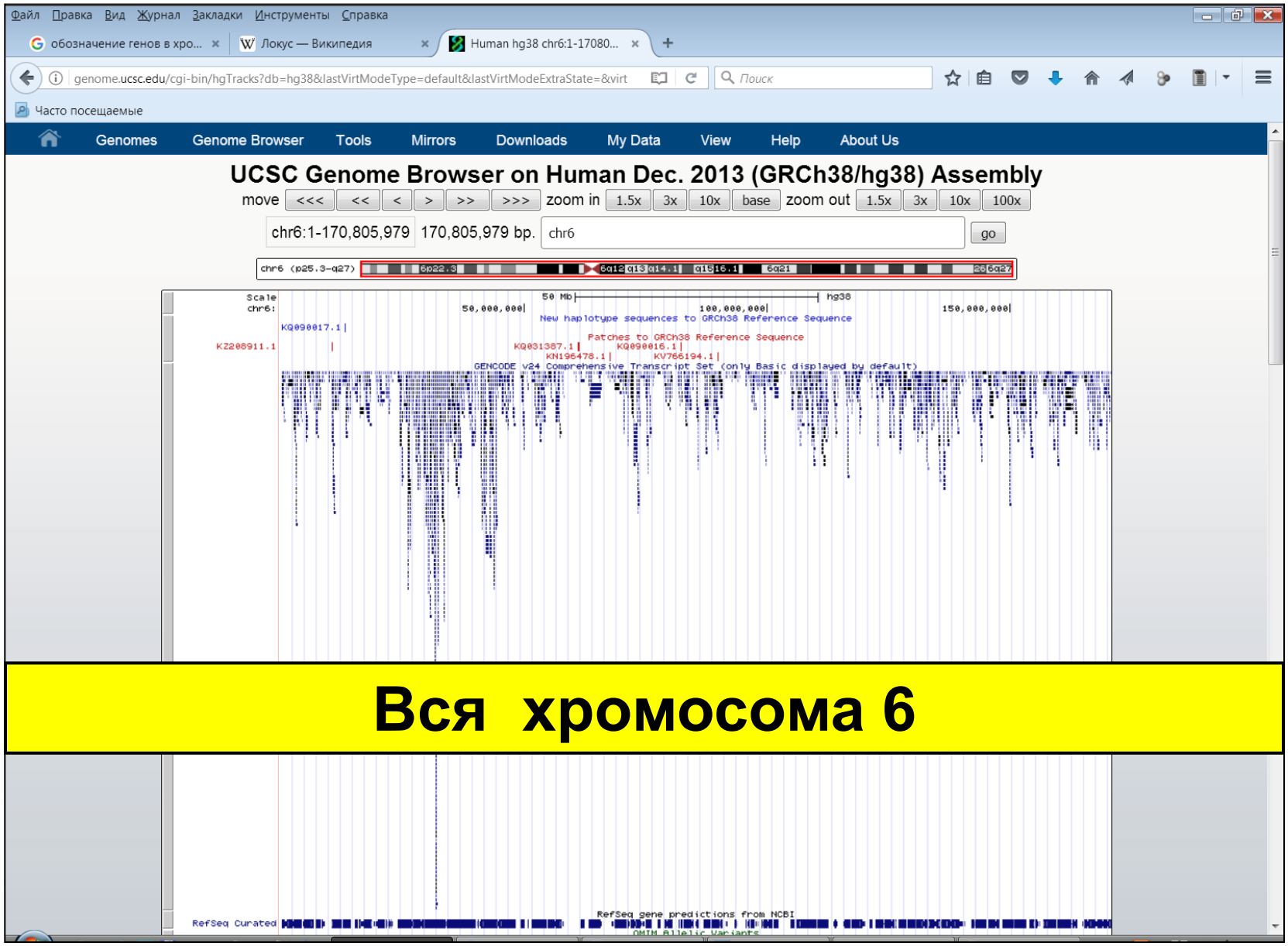
UCSC Genome Browser



UCSC Genome Browser

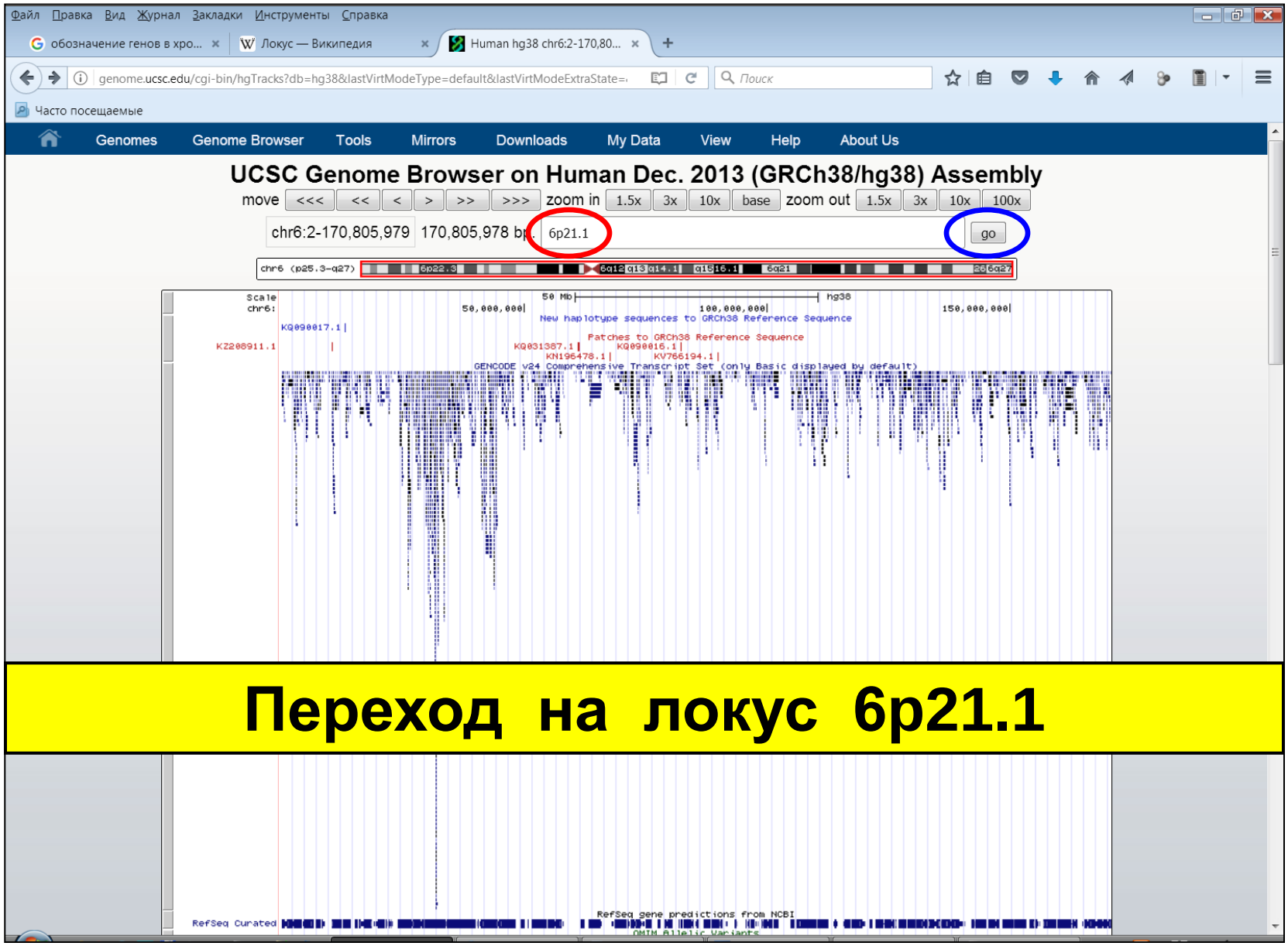


UCSC Genome Browser



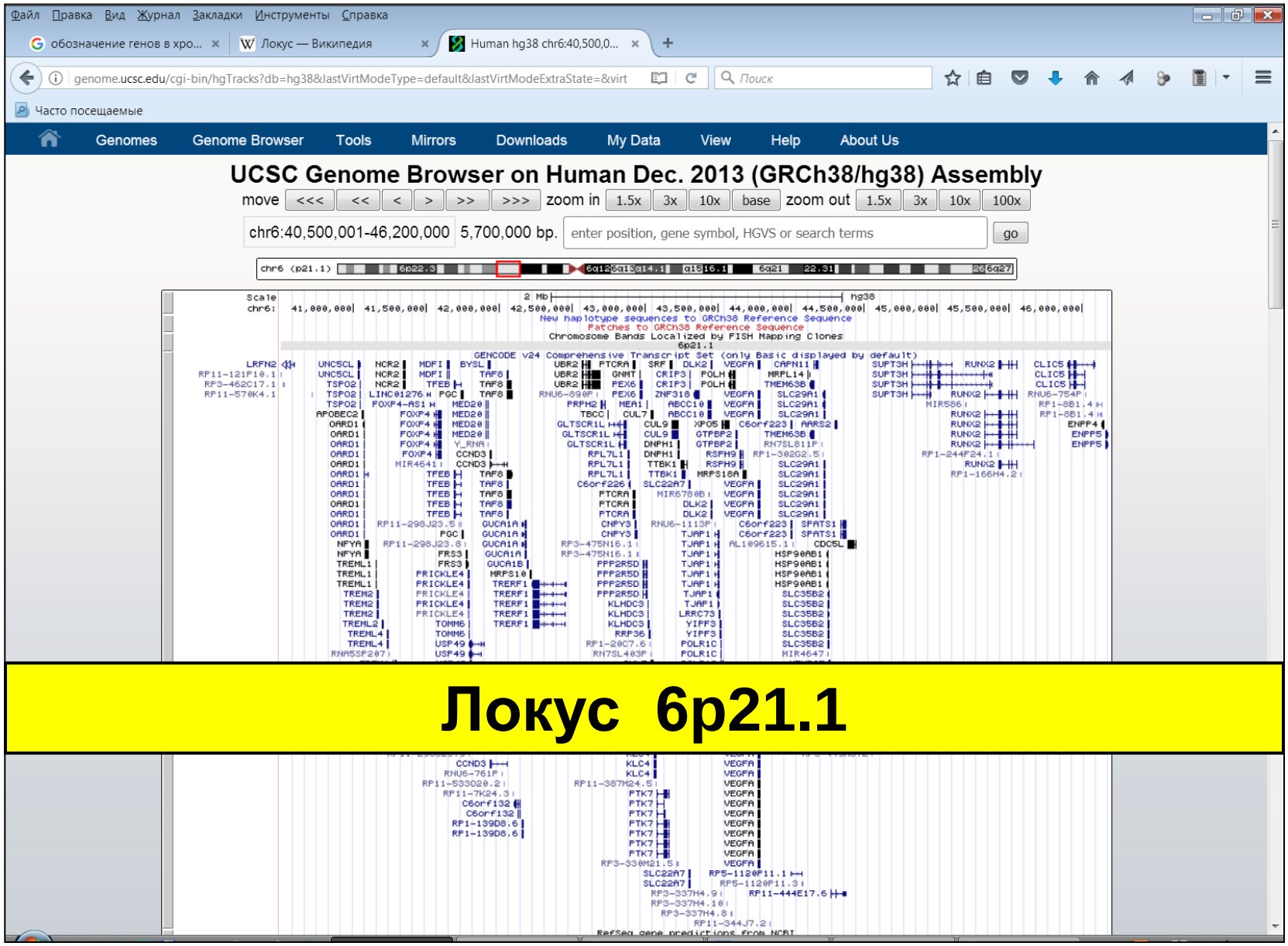
Вся хромосома 6

UCSC Genome Browser

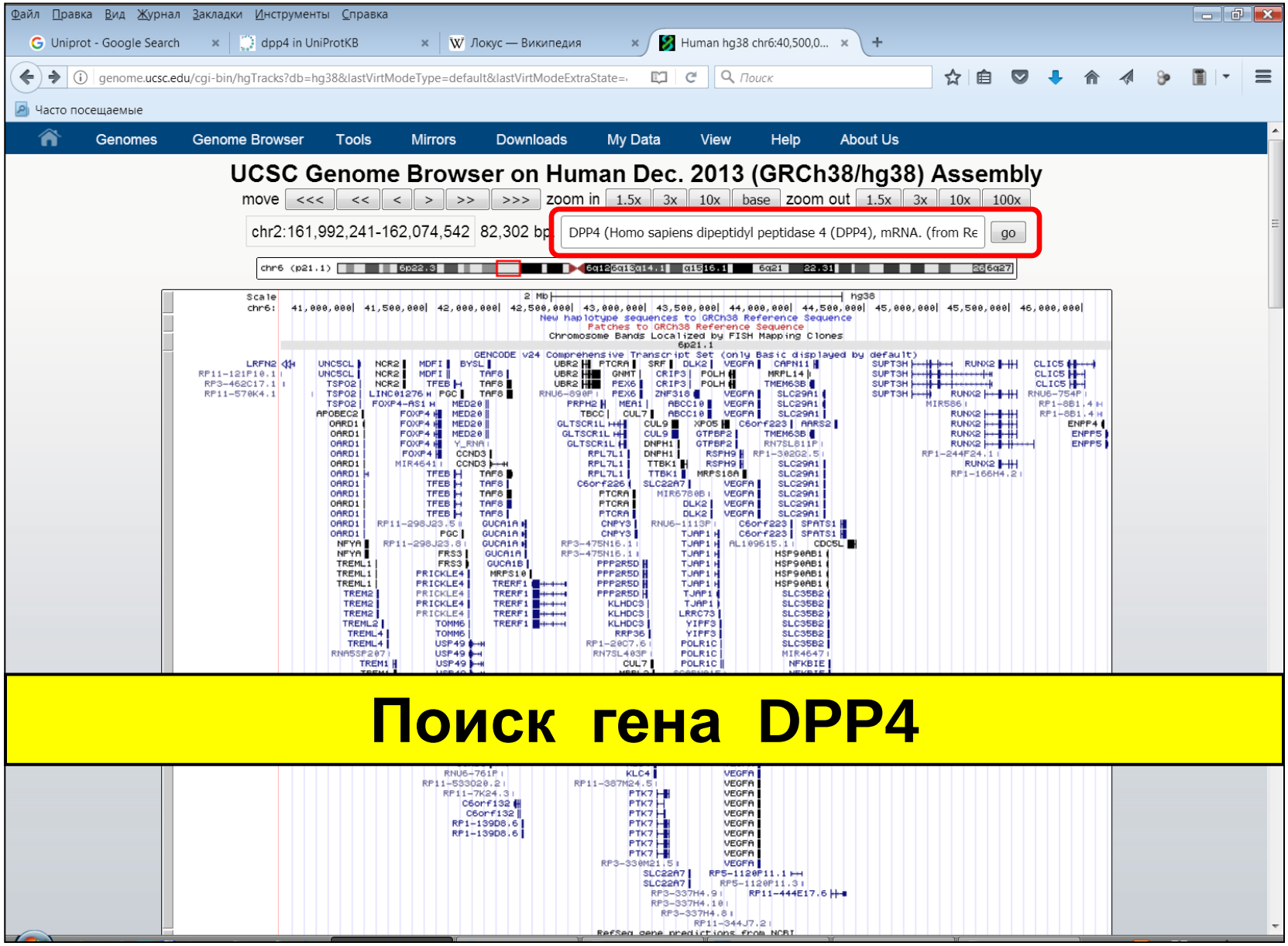


Переход на локус 6p21.1

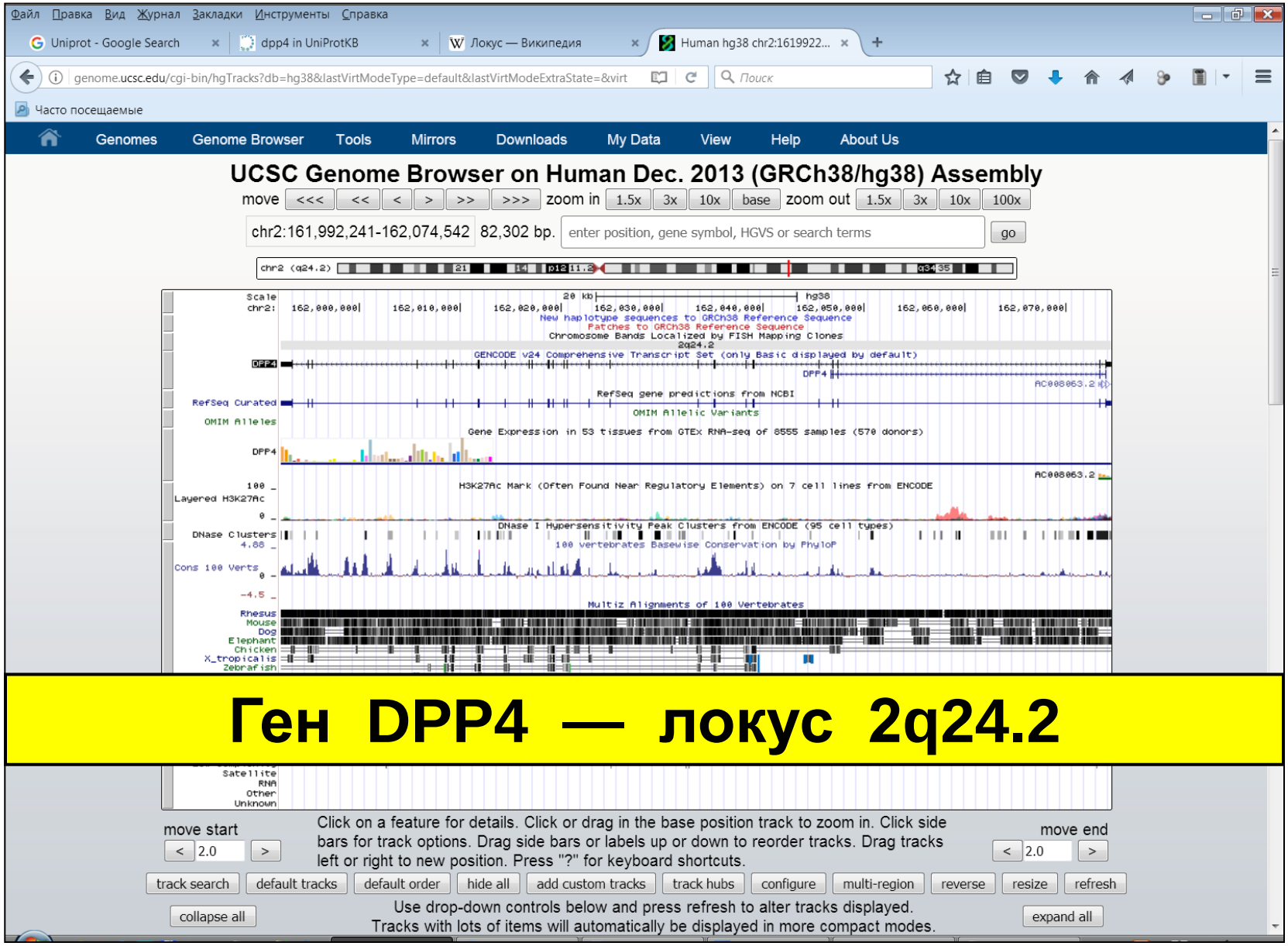
UCSC Genome Browser



UCSC Genome Browser



UCSC Genome Browser



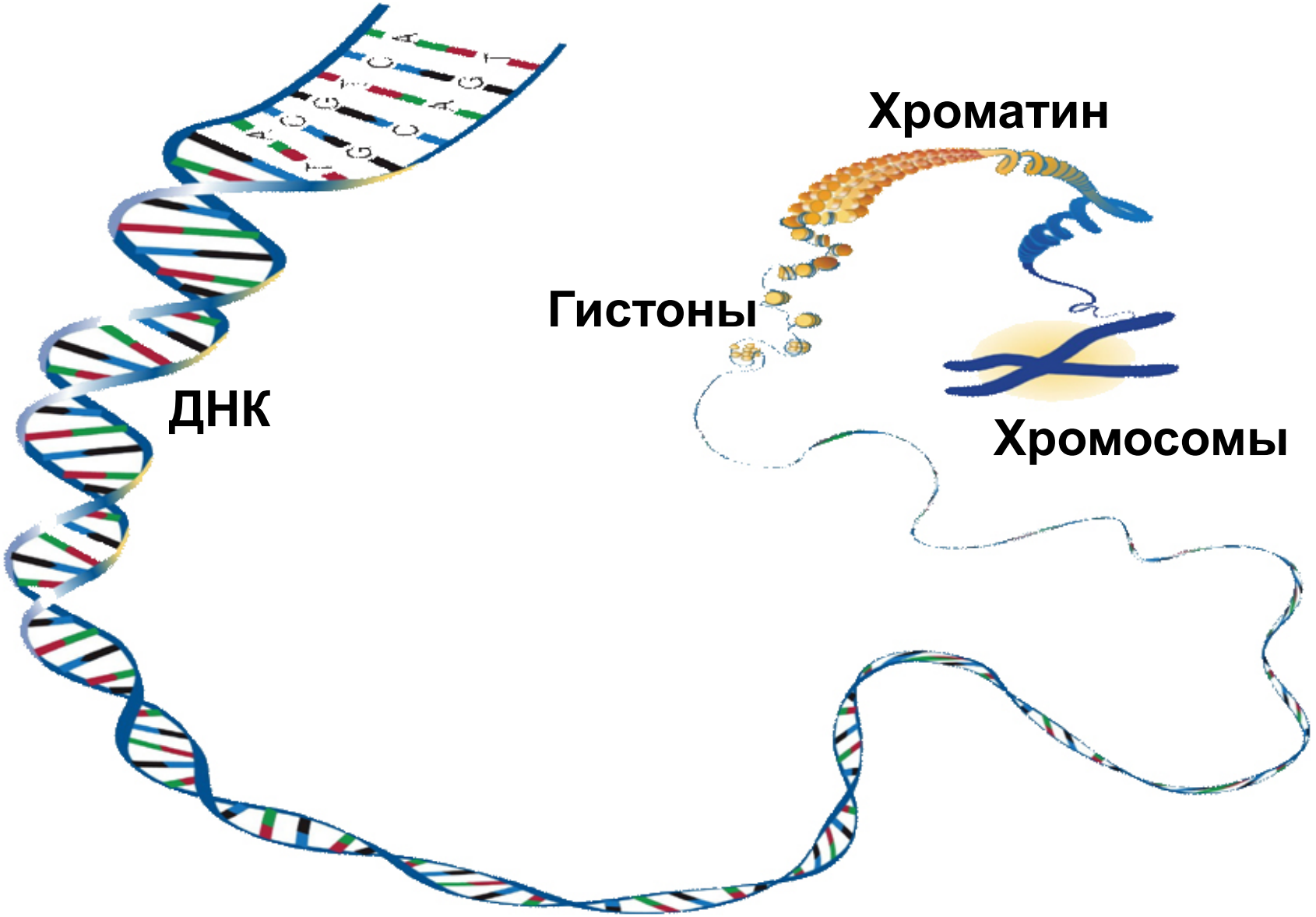
Определения

- Локус — 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

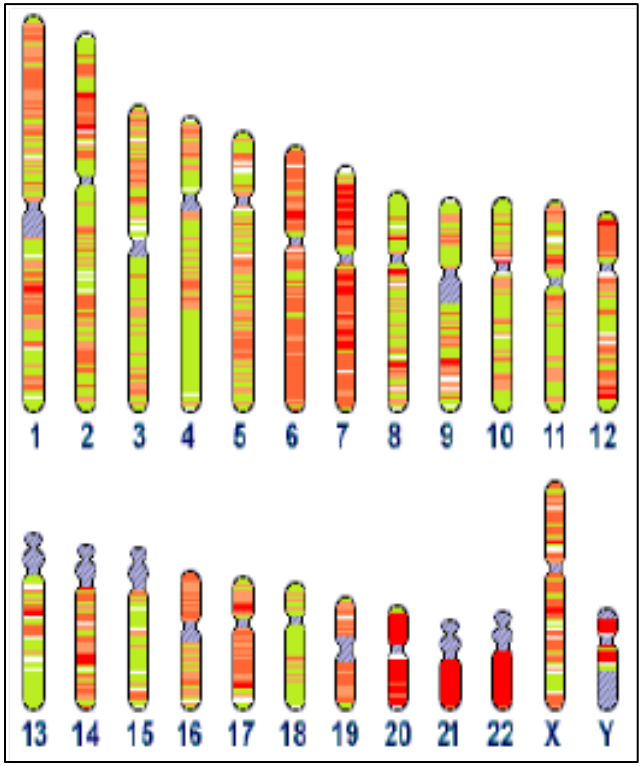
Аннотация генома

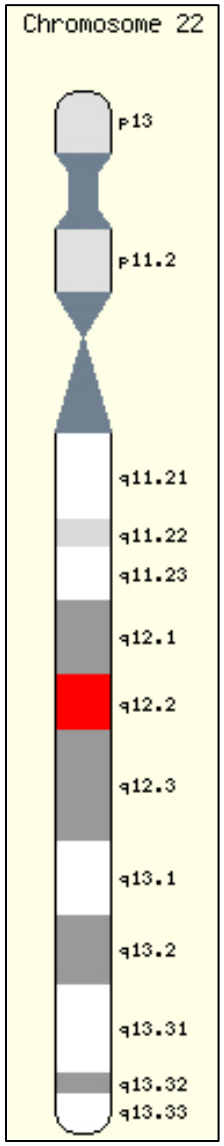
- Описание функциональных и структурных характеристик генома
- Местонахождение кодирующих участков генов в геноме, регуляторных элементов, регулирующих транскрипцию, другие функции генома и особенности функционирования генома
- Тканеснецифичность экспрессии, профили экспрессии и коэкспрессию генов

Структура генома



Обозначения локусов





6p21.3

6 – номер хромосомы

p – плечо хромосомы:

p – короткое

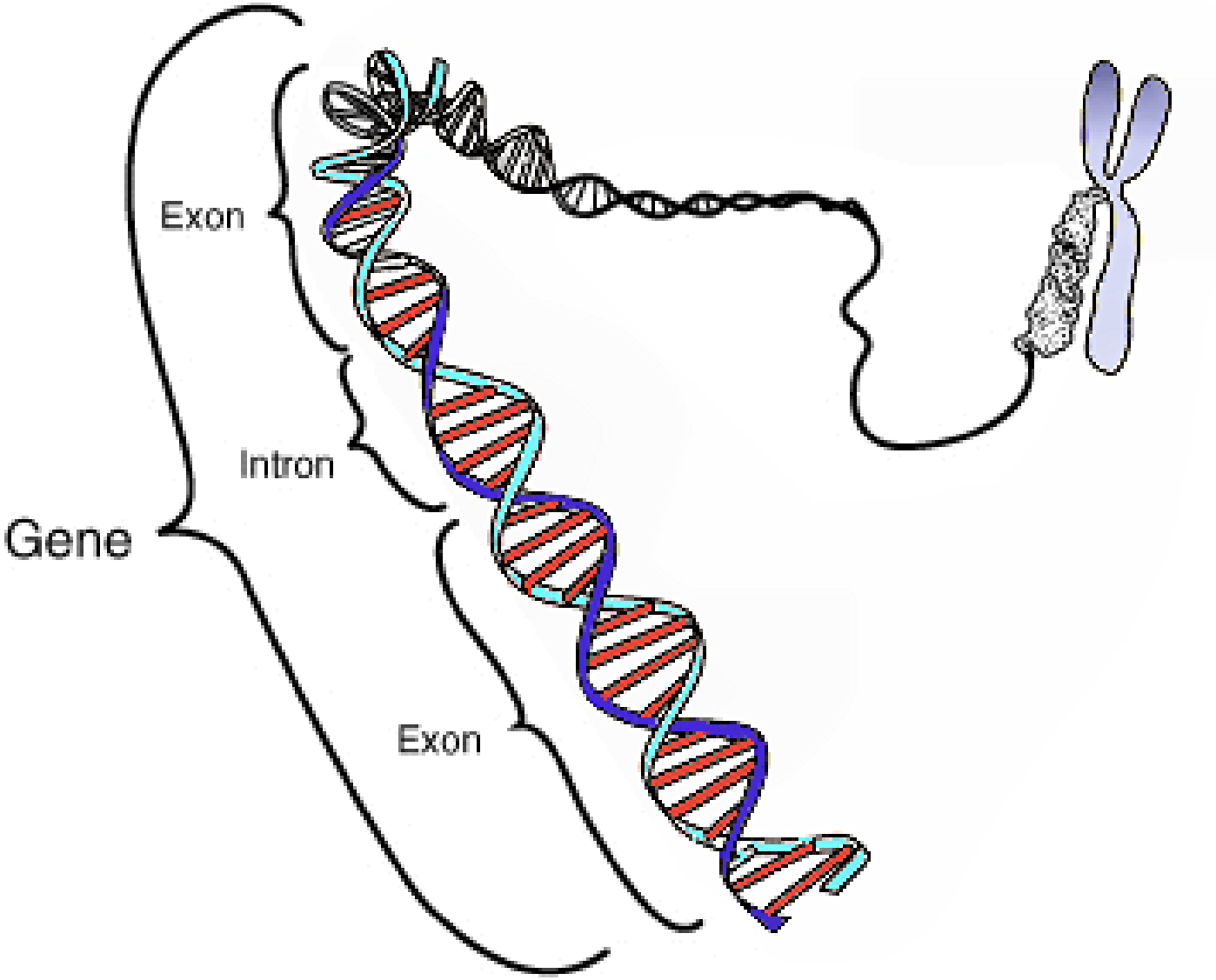
q – длинное

2 – область

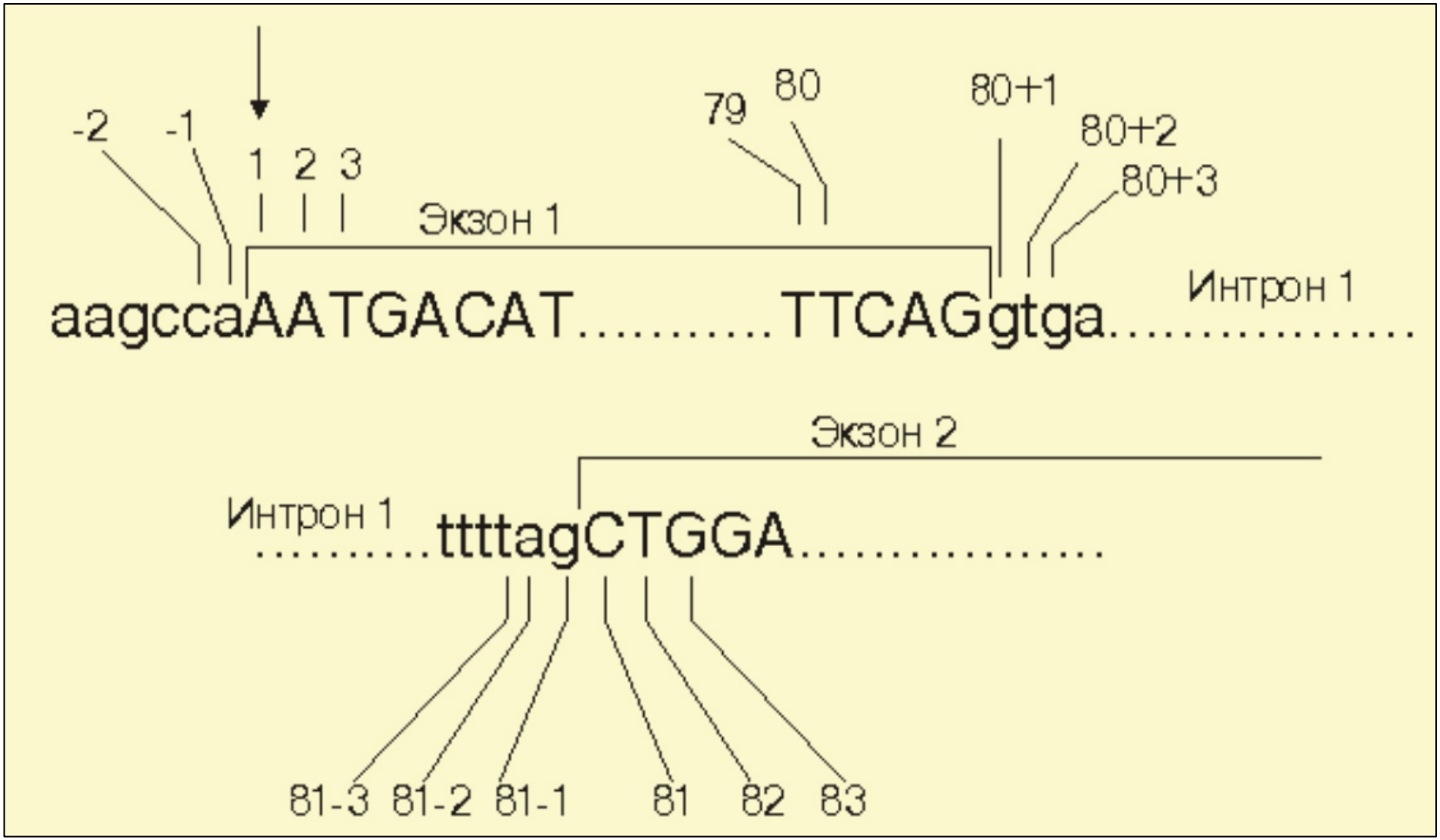
1 – бэнд

.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 ...