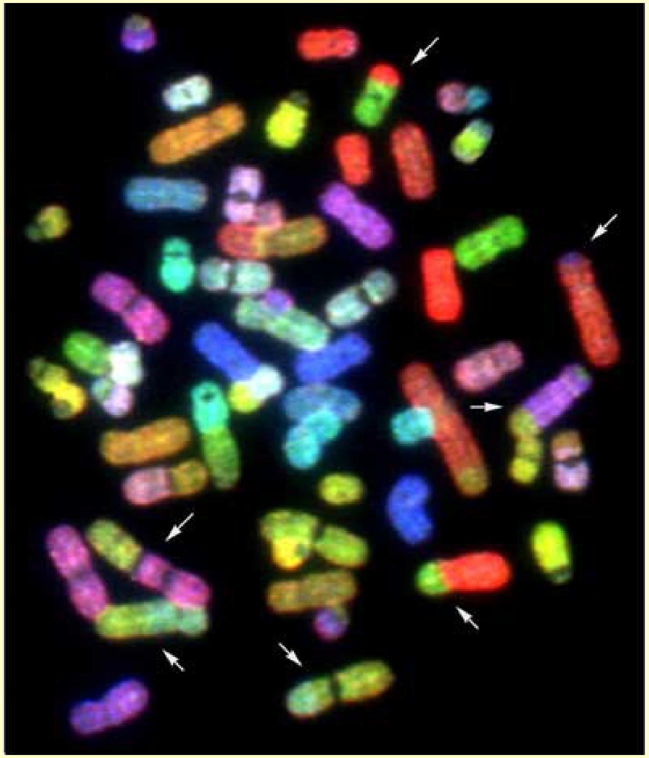
Your Genes and Your Health

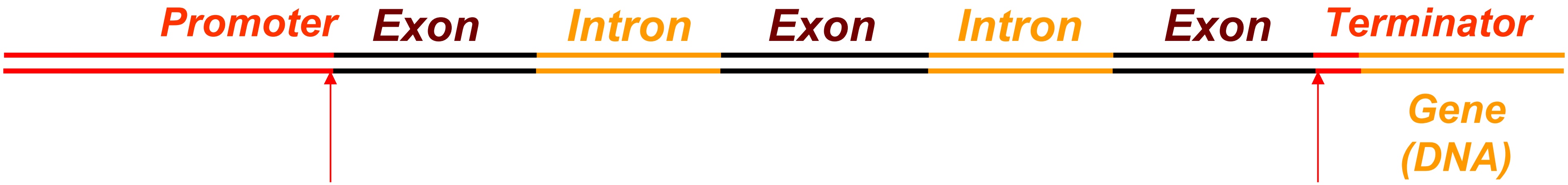
http://bio84.stanford.edu/

Genome Databases

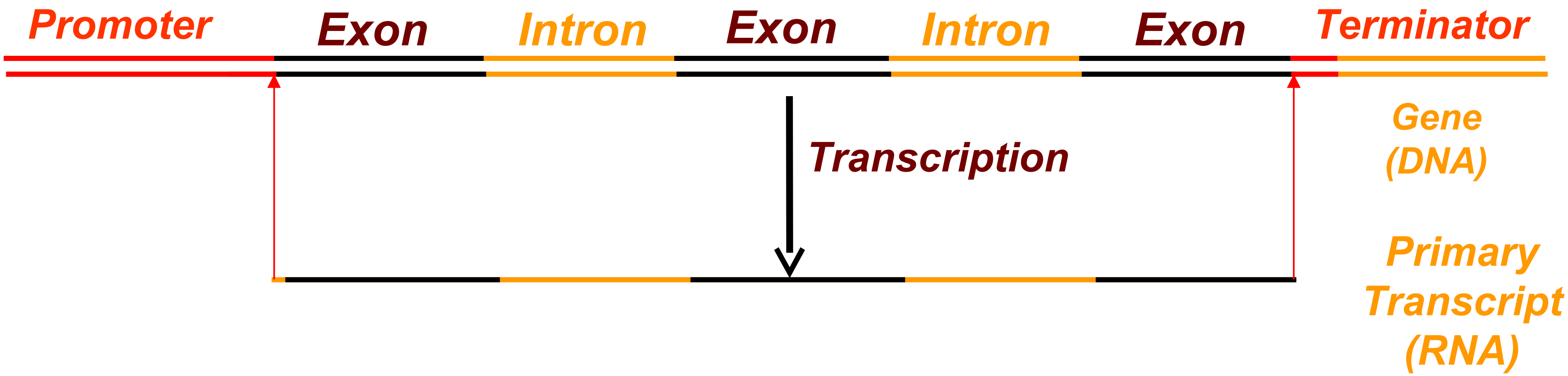


Doug Brutlag, Professor Emeritus
Biochemistry and Medicine (by courtesy)
Stanford University School of Medicine

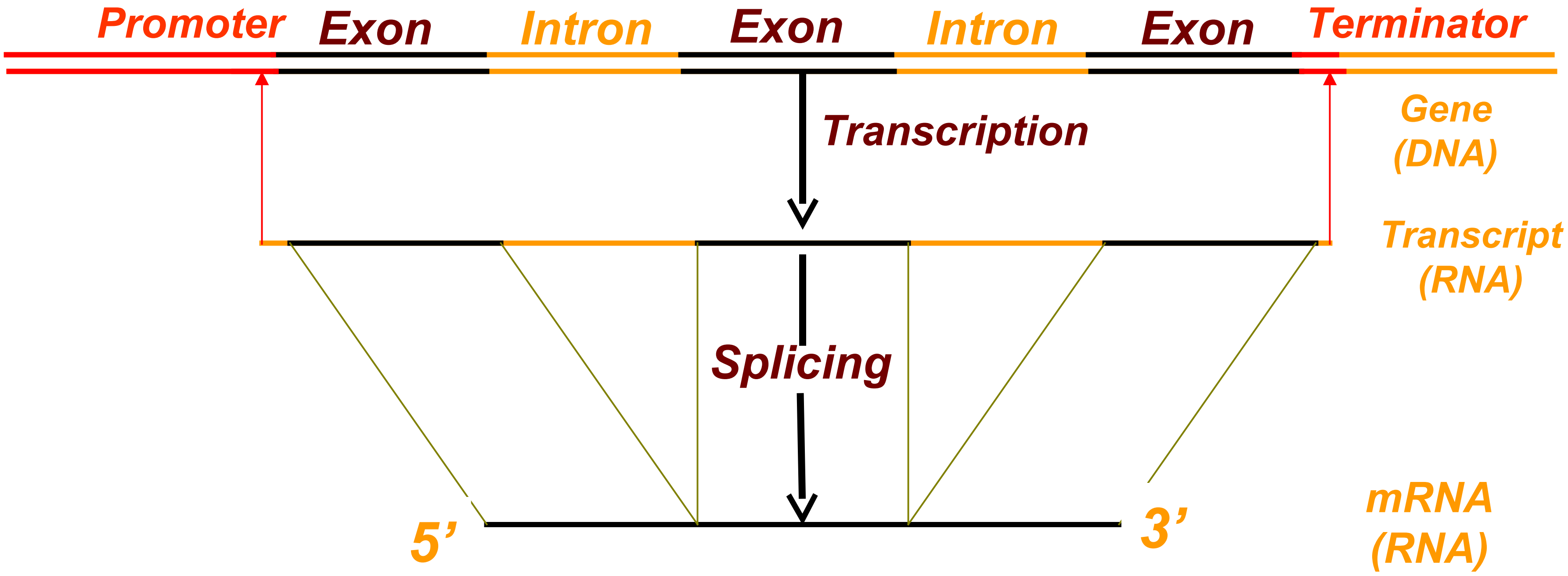
Components of a Typical Human Gene



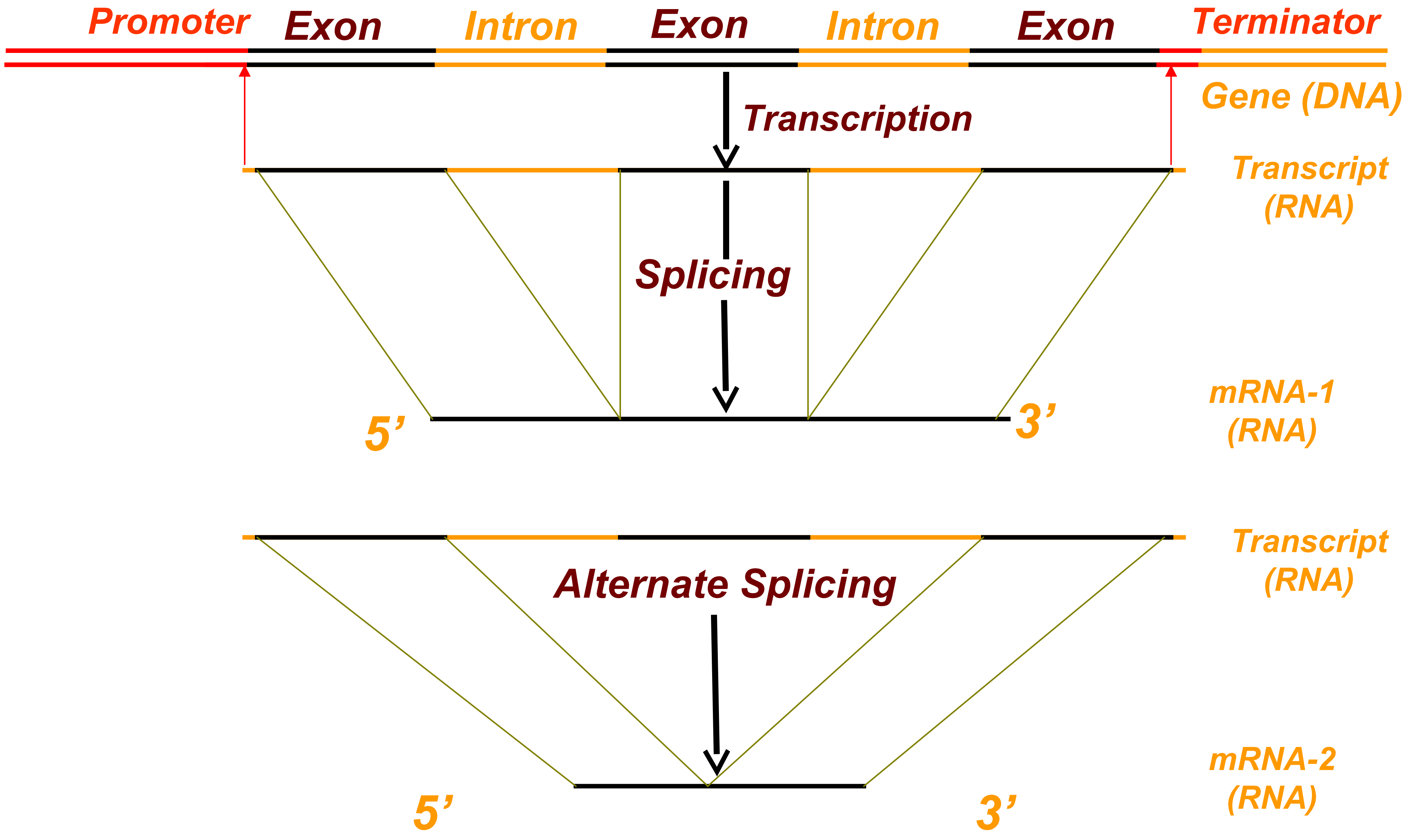
Active Genes are Transcribed into RNA



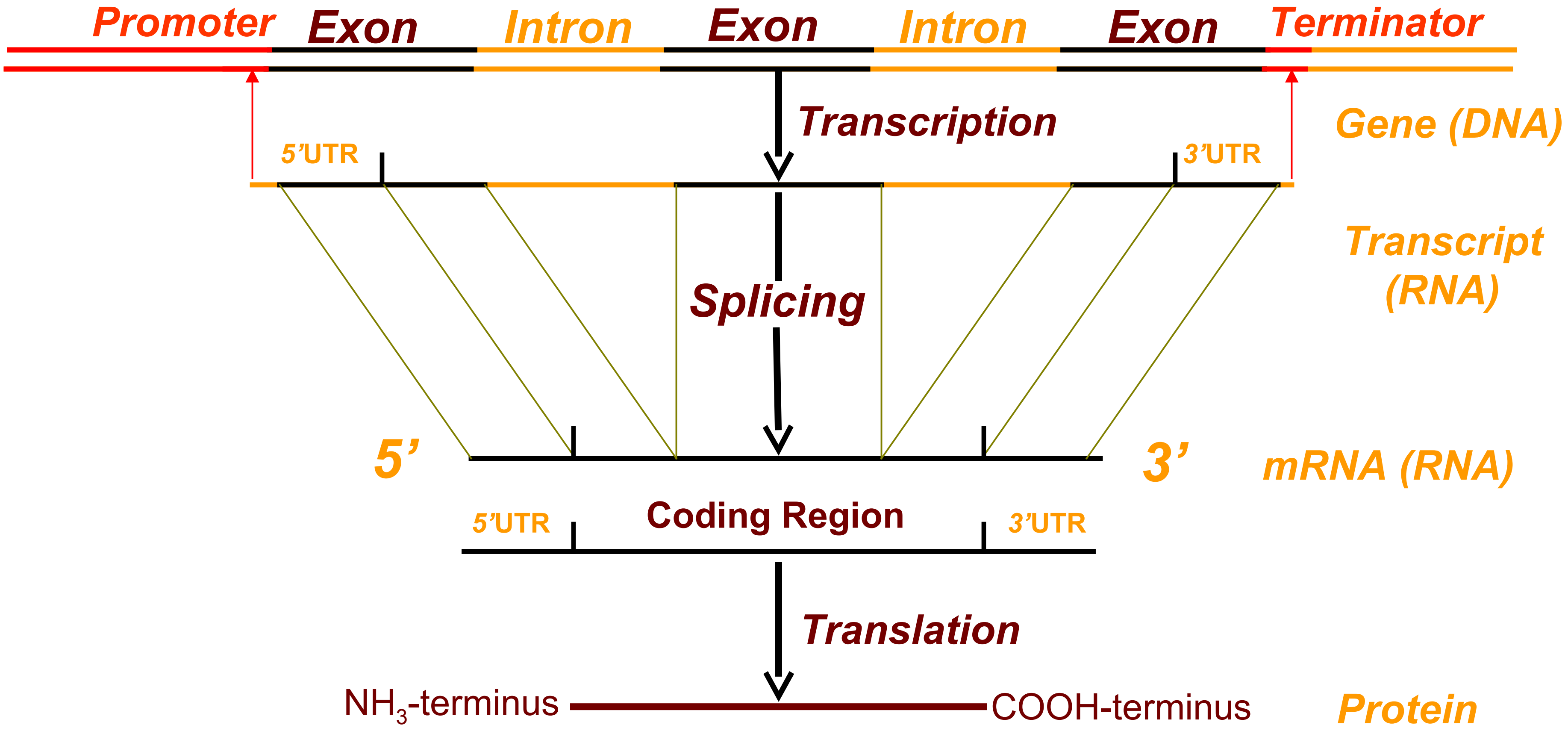
Splicing Transcript Yields Mature mRNA



Alternative Splicing Generates Distinct Proteins in Different Tissues



Mature mRNA contains Coding Region and 5′ and 3′ Untranslated Regions



Promoter **Exon** **Intron** **Exon** **Intron** **Exon** **Terminator**

Gene (DNA)

Transcription

5'UTR **3'UTR**

Transcript (RNA)

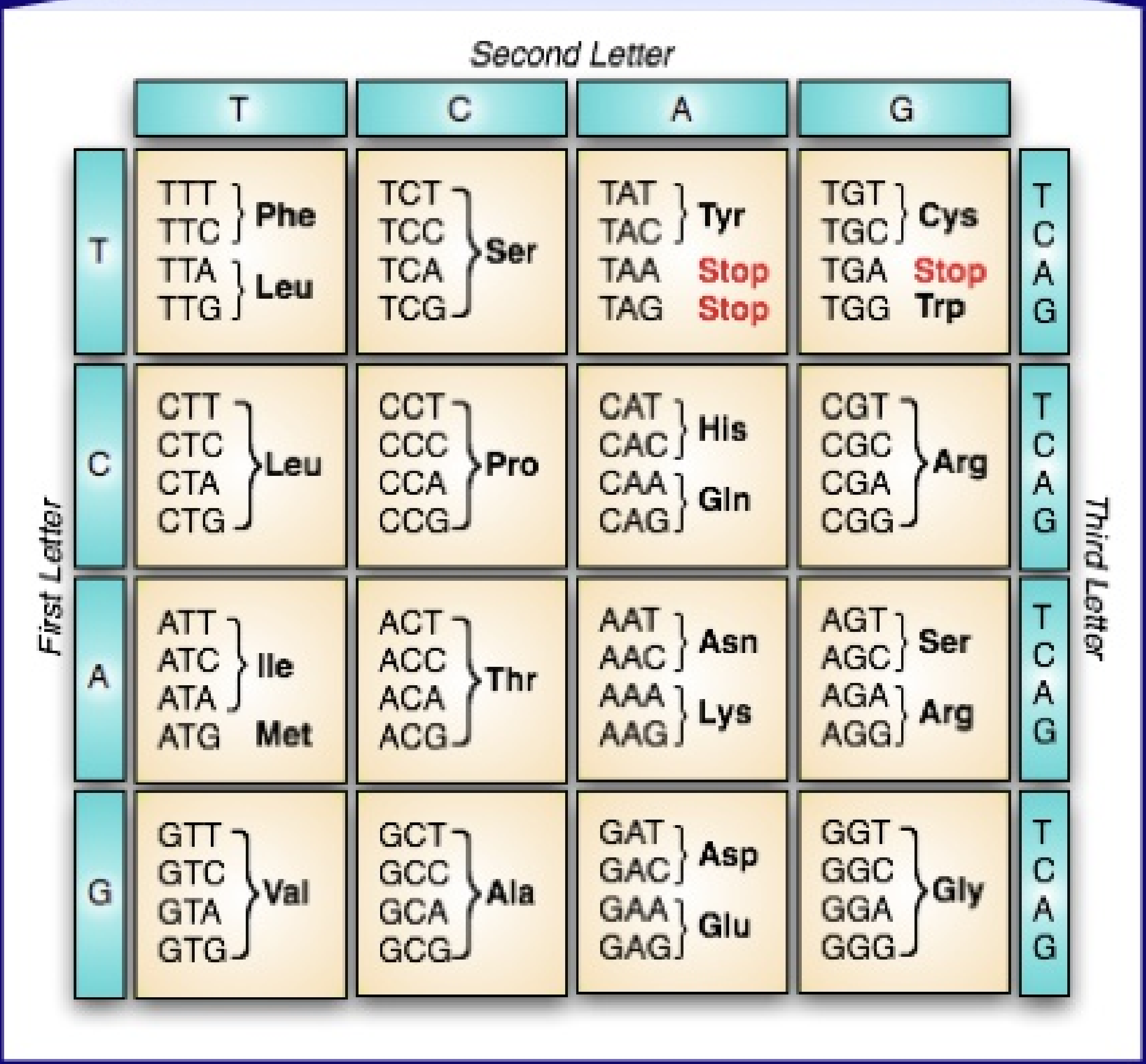
Splicing

5' **3'** **mRNA (RNA)**

5'UTR **Coding Region** **3'UTR**

Translation

NH_3-terminus COOH-terminus **Protein**

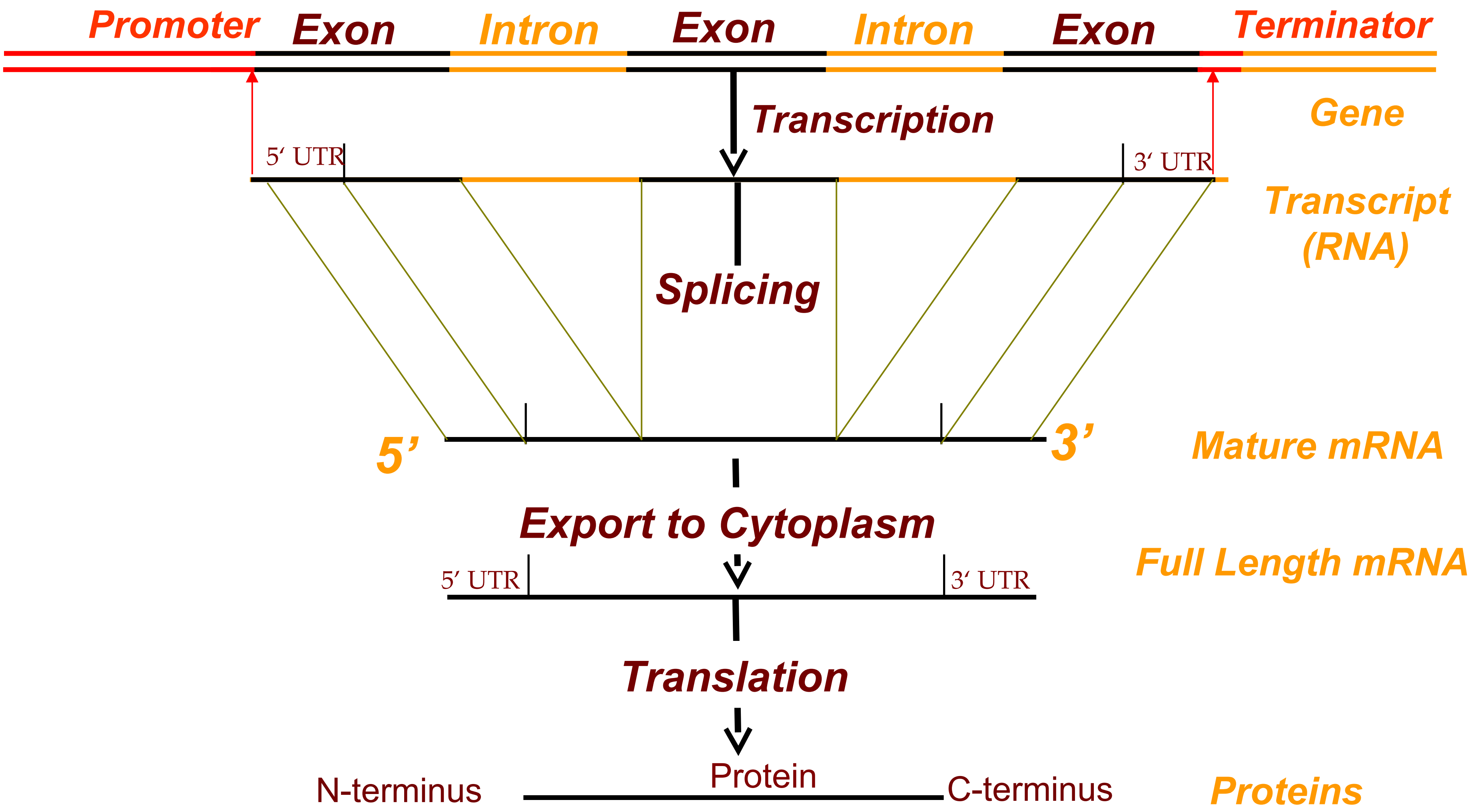


Second Letter

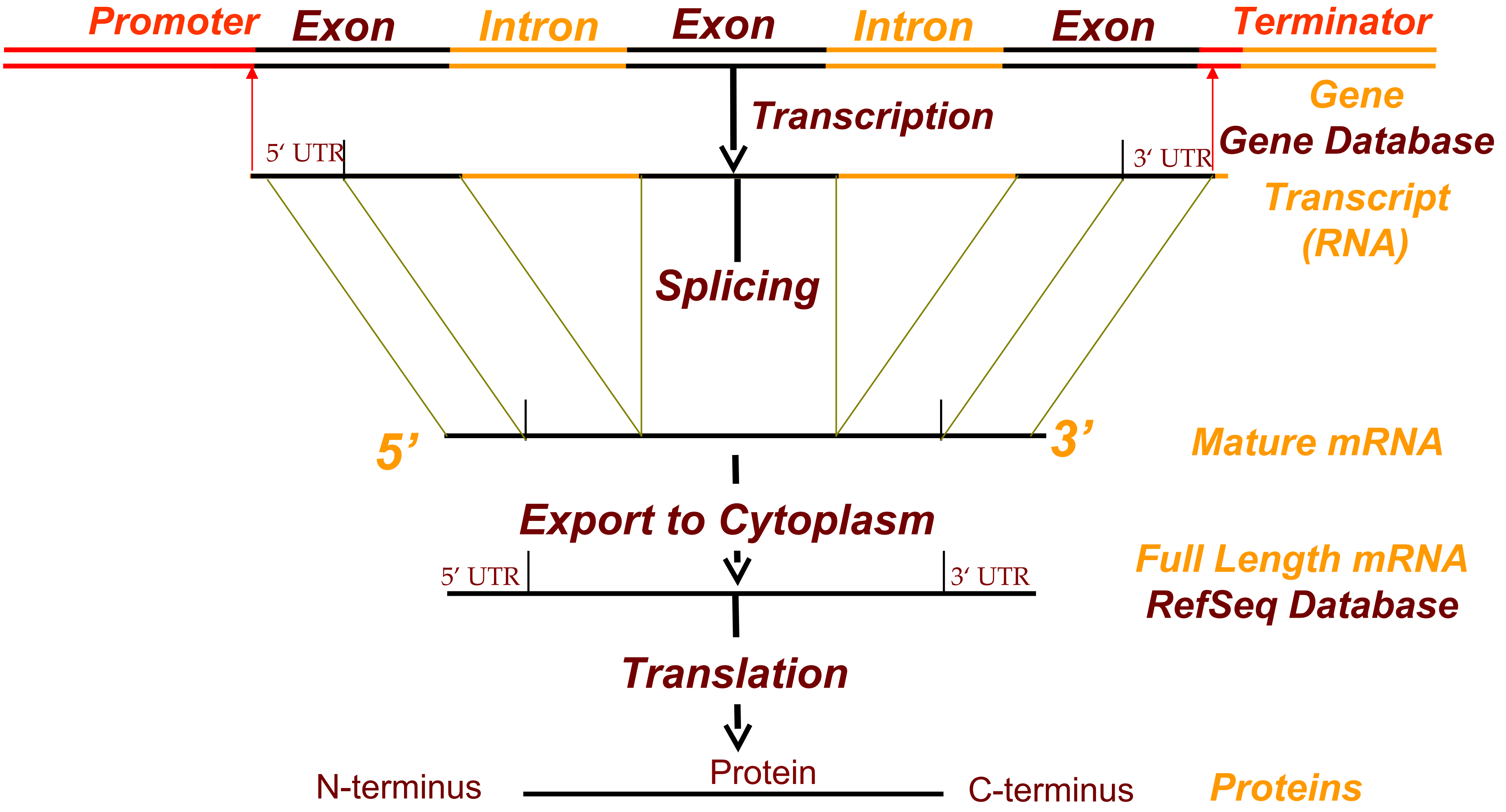
First Letter	T	C	A	G	Third Letter
T	TTT TTC } Phe TTA TTG } Leu	TCT TCC TCA TCG } Ser	TAT TAC } Tyr TAA Stop TAG Stop	TGT TGC } Cys TGA Stop TGG Trp	T C A G
C	CTT CTC CTA CTG } Leu	CCT CCC CCA CCG } Pro	CAT CAC } His CAA CAG } Gln	CGT CGC CGA CGG } Arg	T C A G
A	ATT ATC ATA } Ile ATG Met	ACT ACC ACA ACG } Thr	AAT AAC } Asn AAA AAG } Lys	AGT AGC } Ser AGA AGG } Arg	T C A G
G	GTT GTC GTA GTG } Val	GCT GCC GCA GCG } Ala	GAT GAC } Asp GAA GAG } Glu	GGT GGC GGA GGG } Gly	T C A G

The Genetic Code

Transcription, Splicing and Translation



Transcription, Splicing and Translation



Gene
Gene Database

Transcript (RNA)

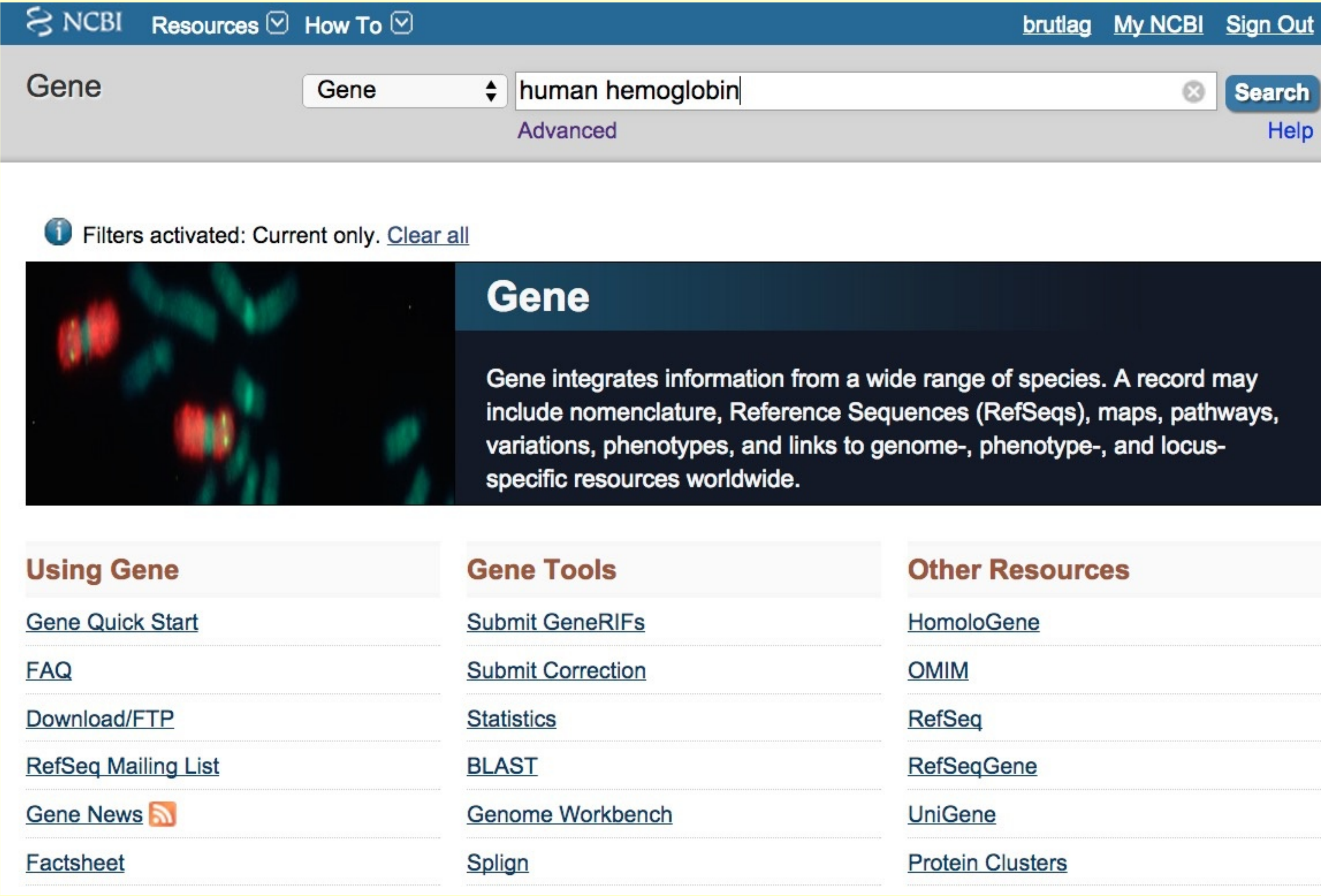
Mature mRNA

Full Length mRNA
RefSeq Database

Proteins
Protein Database

NCBI Gene Database

http://www.ncbi.nlm.nih.gov/gene



Search field: human hemoglobin

Filters activated: Current only. Clear all

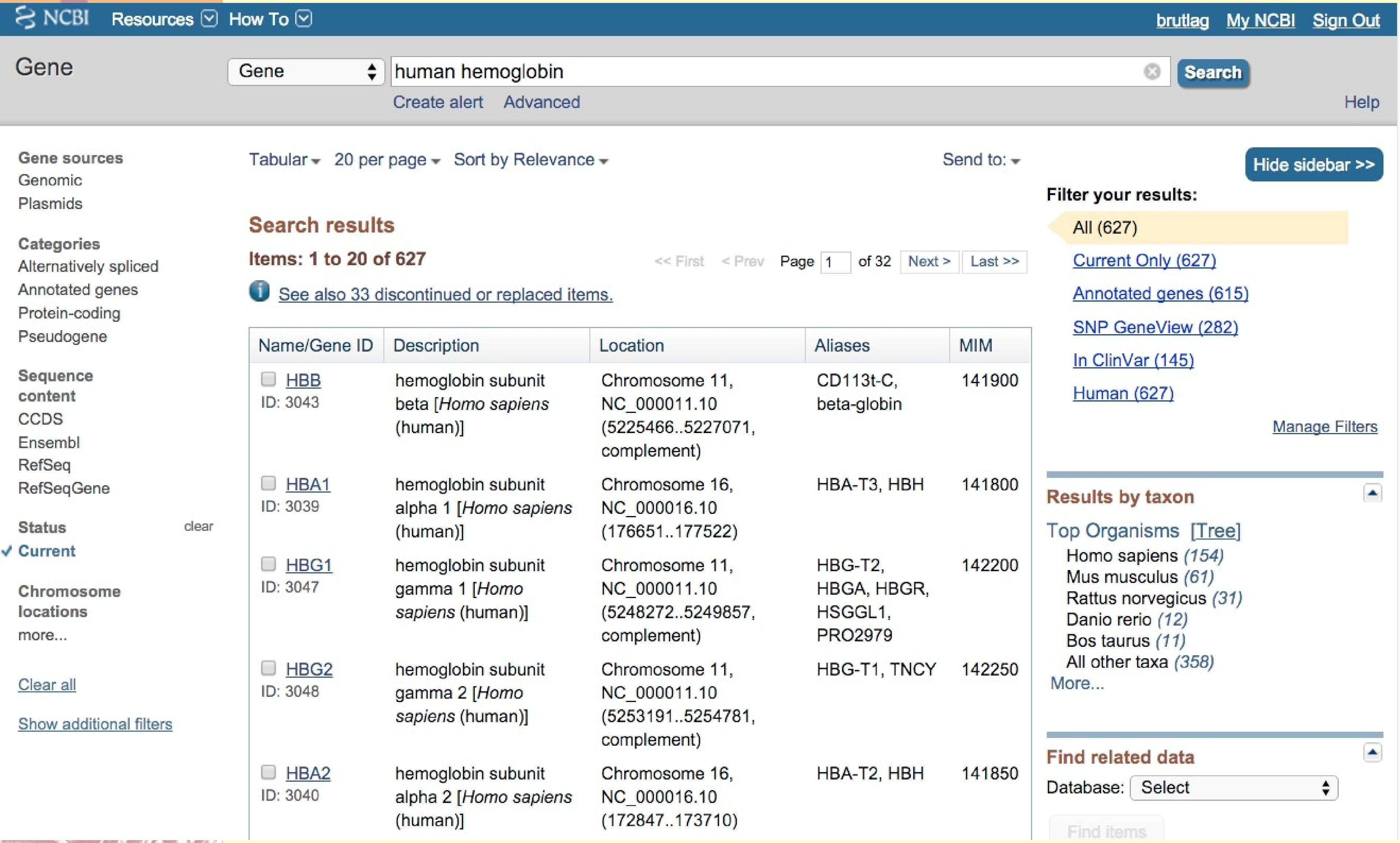
Gene

Gene integrates information from a wide range of species. A record may include nomenclature, Reference Sequences (RefSeqs), maps, pathways, variations, phenotypes, and links to genome-, phenotype-, and locus-specific resources worldwide.

Using Gene	Gene Tools	Other Resources
Gene Quick Start	Submit GeneRIFs	HomoloGene
FAQ	Submit Correction	OMIM
Download/FTP	Statistics	RefSeq
RefSeq Mailing List	BLAST	RefSeqGene
Gene News	Genome Workbench	UniGene
Factsheet	Splign	Protein Clusters

Human Beta Hemoglobin

http://www.ncbi.nlm.nih.gov/gene



Gene ▾ human hemoglobin

Gene sources
Genomic
Plasmids

Categories
Alternatively spliced
Annotated genes
Protein-coding
Pseudogene

Sequence content
CCDS
Ensembl
RefSeq
RefSeqGene

Status clear
✓ **Current**

Chromosome locations
more...

Clear all

Show additional filters

Search results

Items: 1 to 20 of 627

Page 1 of 32

See also 33 discontinued or replaced items.

Name/Gene ID	Description	Location	Aliases	MIM
☐ HBB ID: 3043	hemoglobin subunit beta [*Homo sapiens* (human)]	Chromosome 11, NC_000011.10 (5225466..5227071, complement)	CD113t-C, beta-globin	141900
☐ HBA1 ID: 3039	hemoglobin subunit alpha 1 [*Homo sapiens* (human)]	Chromosome 16, NC_000016.10 (176651..177522)	HBA-T3, HBH	141800
☐ HBG1 ID: 3047	hemoglobin subunit gamma 1 [*Homo sapiens* (human)]	Chromosome 11, NC_000011.10 (5248272..5249857, complement)	HBG-T2, HBGA, HBGR, HSGGL1, PRO2979	142200
☐ HBG2 ID: 3048	hemoglobin subunit gamma 2 [*Homo sapiens* (human)]	Chromosome 11, NC_000011.10 (5253191..5254781, complement)	HBG-T1, TNCY	142250
☐ HBA2 ID: 3040	hemoglobin subunit alpha 2 [*Homo sapiens* (human)]	Chromosome 16, NC_000016.10 (172847..173710)	HBA-T2, HBH	141850

Filter your results:
All (627)
Current Only (627)
Annotated genes (615)
SNP GeneView (282)
In ClinVar (145)
Human (627)

Results by taxon

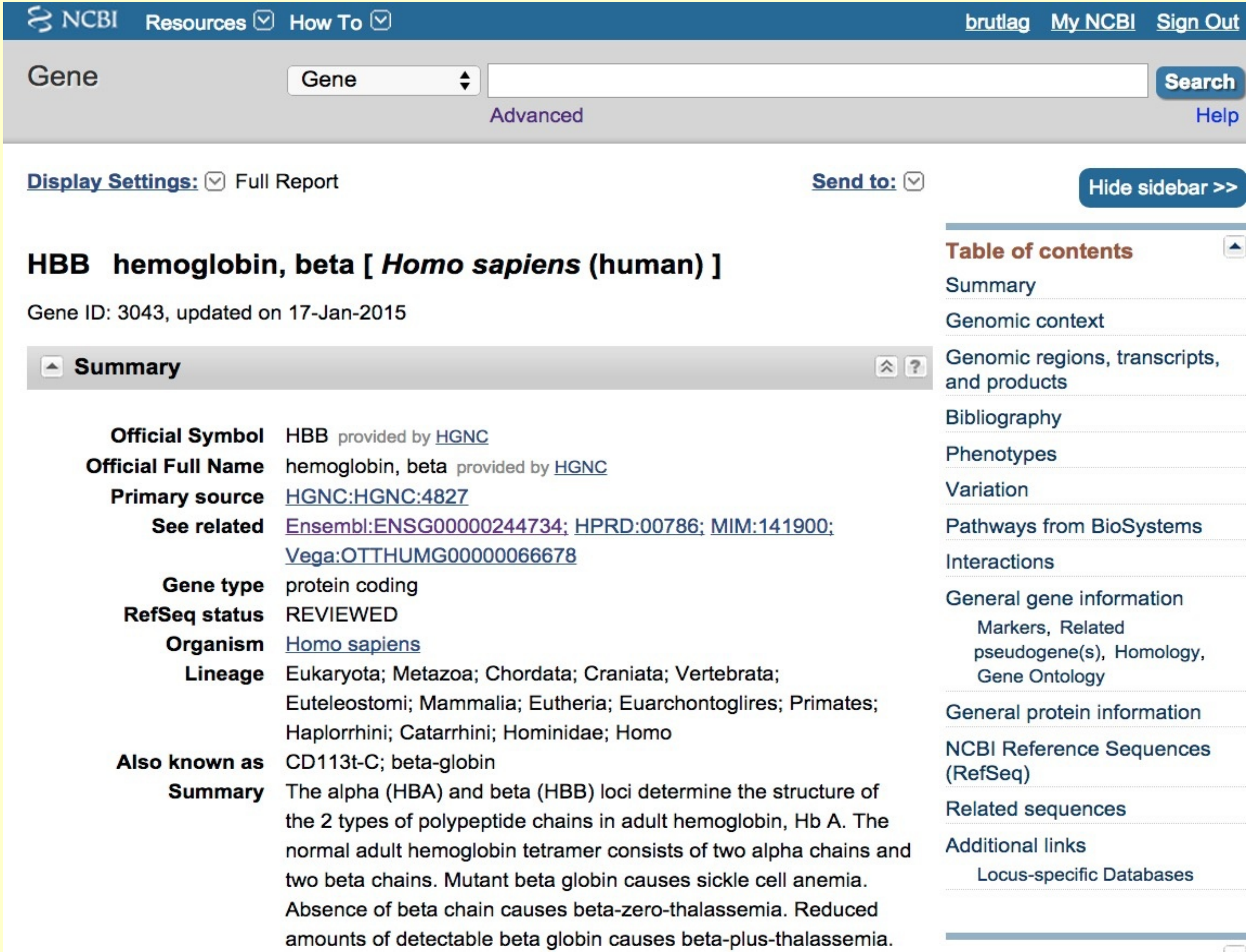
Top Organisms [Tree]
Homo sapiens *(154)*
Mus musculus *(61)*
Rattus norvegicus *(31)*
Danio rerio *(12)*
Bos taurus *(11)*
All other taxa *(358)*
More...

Find related data

Database: Select

Human Beta-Hemoglobin Gene Entry

http://www.ncbi.nlm.nih.gov/gene/3043



HBB hemoglobin, beta [*Homo sapiens* (human)]

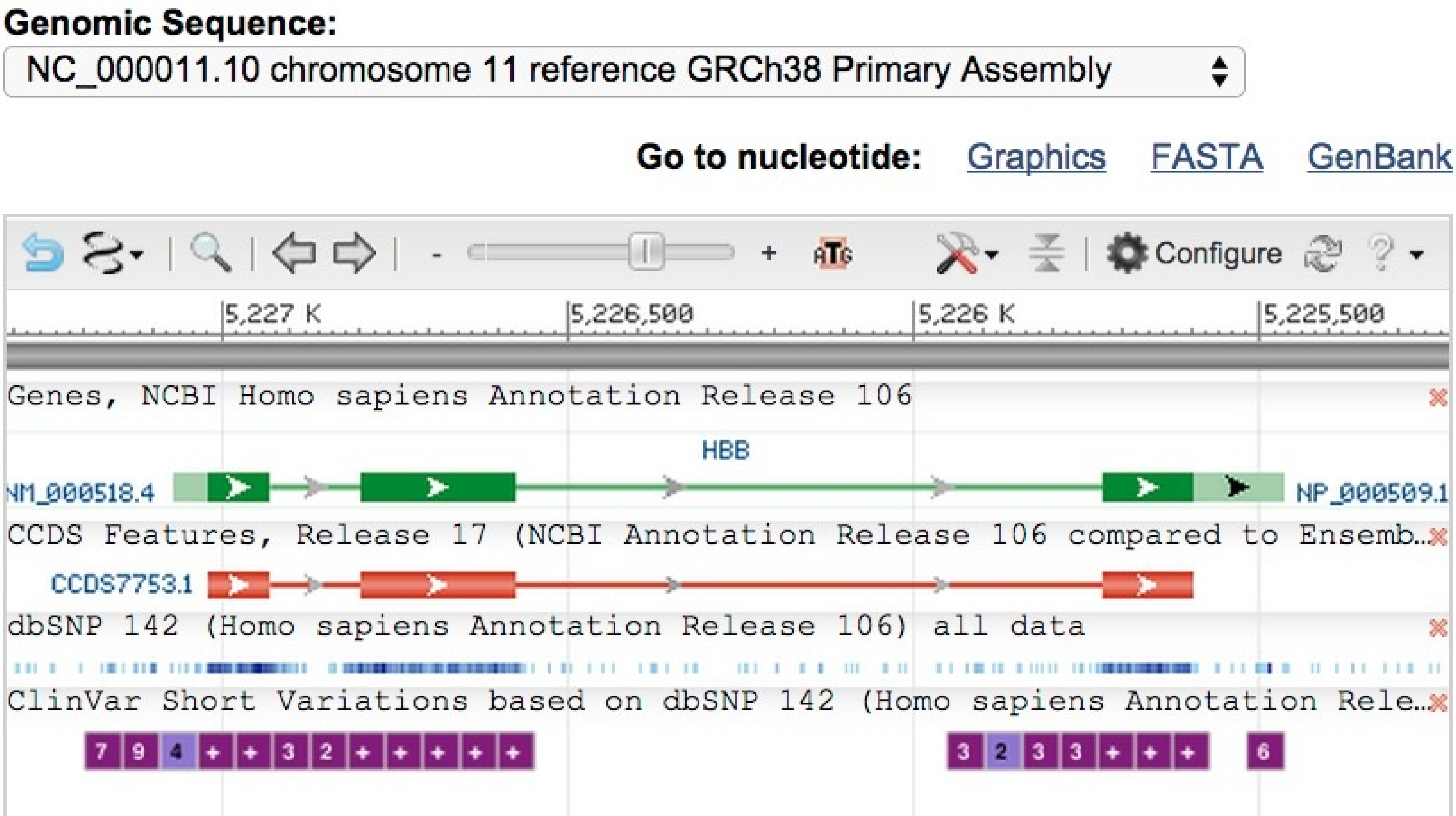
Gene ID: 3043, updated on 17-Jan-2015

Summary

Official Symbol	HBB provided by HGNC
Official Full Name	hemoglobin, beta provided by HGNC
Primary source	HGNC:HGNC:4827
See related	Ensembl:ENSG00000244734; HPRD:00786; MIM:141900; Vega:OTTHUMG00000066678
Gene type	protein coding
RefSeq status	REVIEWED
Organism	Homo sapiens
Lineage	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo
Also known as	CD113t-C; beta-globin
Summary	The alpha (HBA) and beta (HBB) loci determine the structure of the 2 types of polypeptide chains in adult hemoglobin, Hb A. The normal adult hemoglobin tetramer consists of two alpha chains and two beta chains. Mutant beta globin causes sickle cell anemia. Absence of beta chain causes beta-zero-thalassemia. Reduced amounts of detectable beta globin causes beta-plus-thalassemia.

Human Beta-Hemoglobin Gene Entry

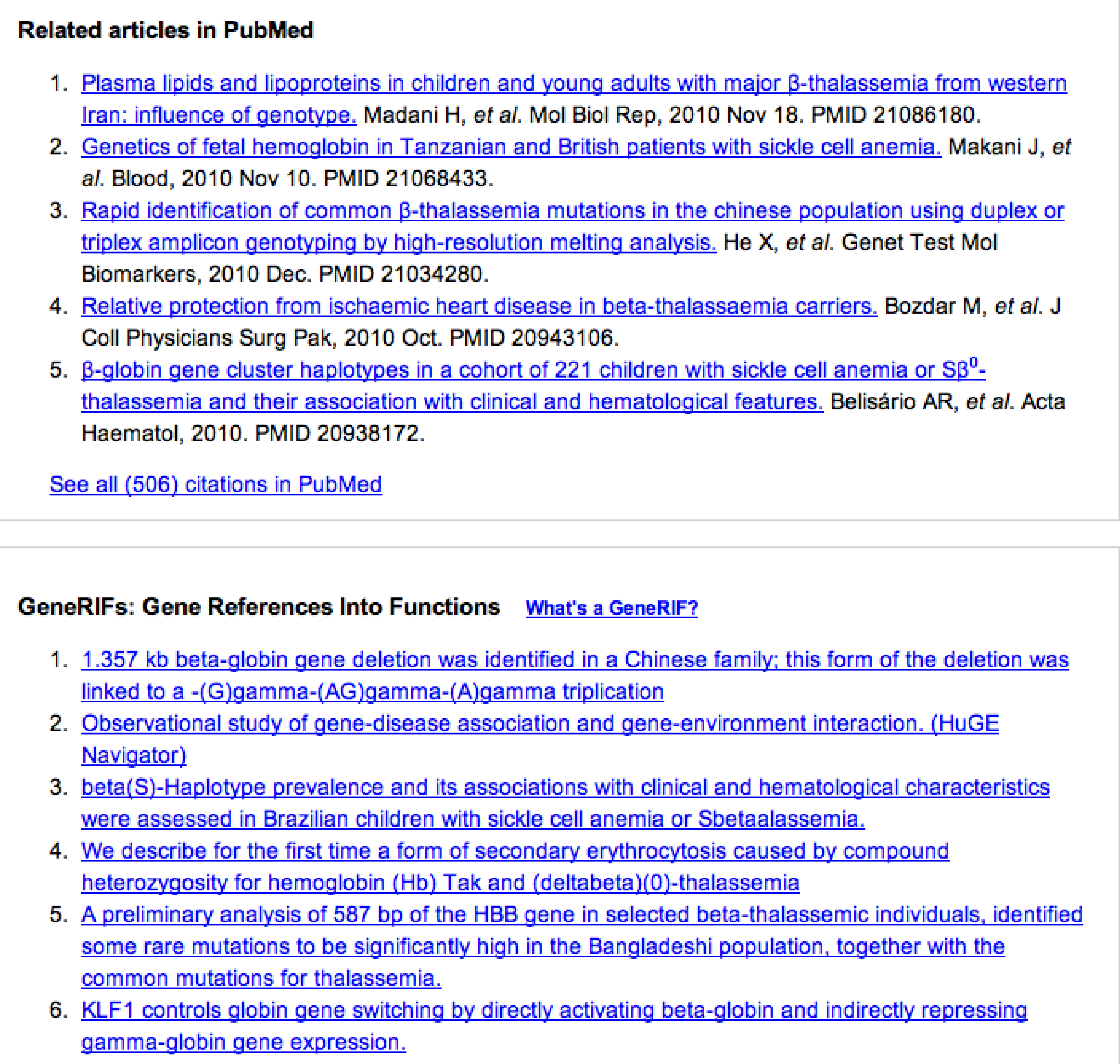
http://www.ncbi.nlm.nih.gov/gene/3043



Genomic Sequence:

NC_000011.10 chromosome 11 reference GRCh38 Primary Assembly

Go to nucleotide: Graphics FASTA GenBank



Related articles in PubMed

1. Plasma lipids and lipoproteins in children and young adults with major β-thalassemia from western Iran: influence of genotype. Madani H, *et al.* Mol Biol Rep, 2010 Nov 18. PMID 21086180.
2. Genetics of fetal hemoglobin in Tanzanian and British patients with sickle cell anemia. Makani J, *et al.* Blood, 2010 Nov 10. PMID 21068433.
3. Rapid identification of common β-thalassemia mutations in the chinese population using duplex or triplex amplicon genotyping by high-resolution melting analysis. He X, *et al.* Genet Test Mol Biomarkers, 2010 Dec. PMID 21034280.
4. Relative protection from ischaemic heart disease in beta-thalassaemia carriers. Bozdar M, *et al.* J Coll Physicians Surg Pak, 2010 Oct. PMID 20943106.
5. β-globin gene cluster haplotypes in a cohort of 221 children with sickle cell anemia or $S\beta^0$-thalassemia and their association with clinical and hematological features. Belisário AR, *et al.* Acta Haematol, 2010. PMID 20938172.

See all (506) citations in PubMed

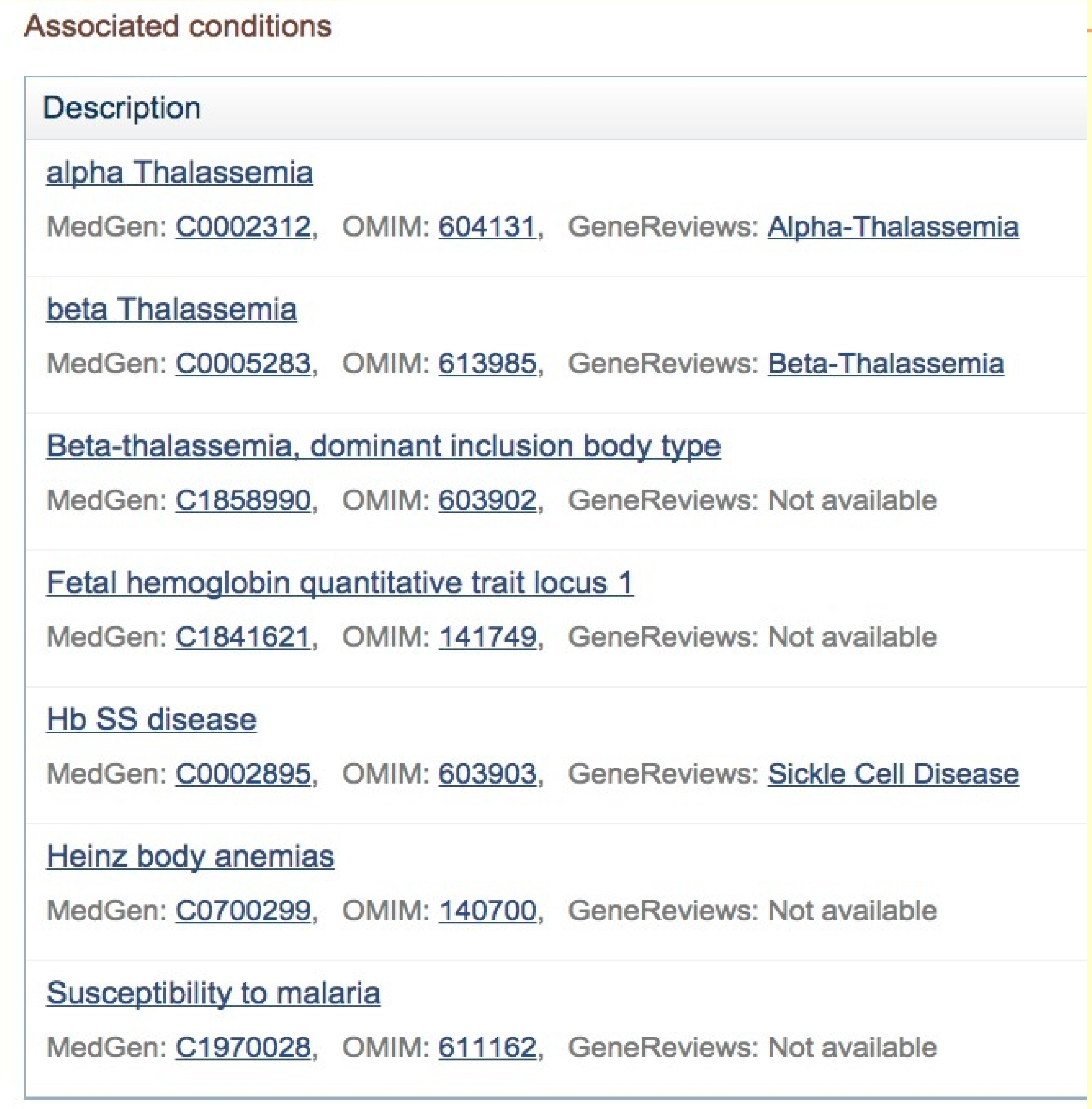
GeneRIFs: Gene References Into Functions

What's a GeneRIF?

1. 1.357 kb beta-globin gene deletion was identified in a Chinese family; this form of the deletion was linked to a -(G)gamma-(AG)gamma-(A)gamma triplication
2. Observational study of gene-disease association and gene-environment interaction. (HuGE Navigator)
3. beta(S)-Haplotype prevalence and its associations with clinical and hematological characteristics were assessed in Brazilian children with sickle cell anemia or Sbetaalassemia.
4. We describe for the first time a form of secondary erythrocytosis caused by compound heterozygosity for hemoglobin (Hb) Tak and (deltabeta)(0)-thalassemia
5. A preliminary analysis of 587 bp of the HBB gene in selected beta-thalassemic individuals, identified some rare mutations to be significantly high in the Bangladeshi population, together with the common mutations for thalassemia.
6. KLF1 controls globin gene switching by directly activating beta-globin and indirectly repressing gamma-globin gene expression.

Human Beta-Hemoglobin Phenotypes

http://www.ncbi.nlm.nih.gov/gene/3043

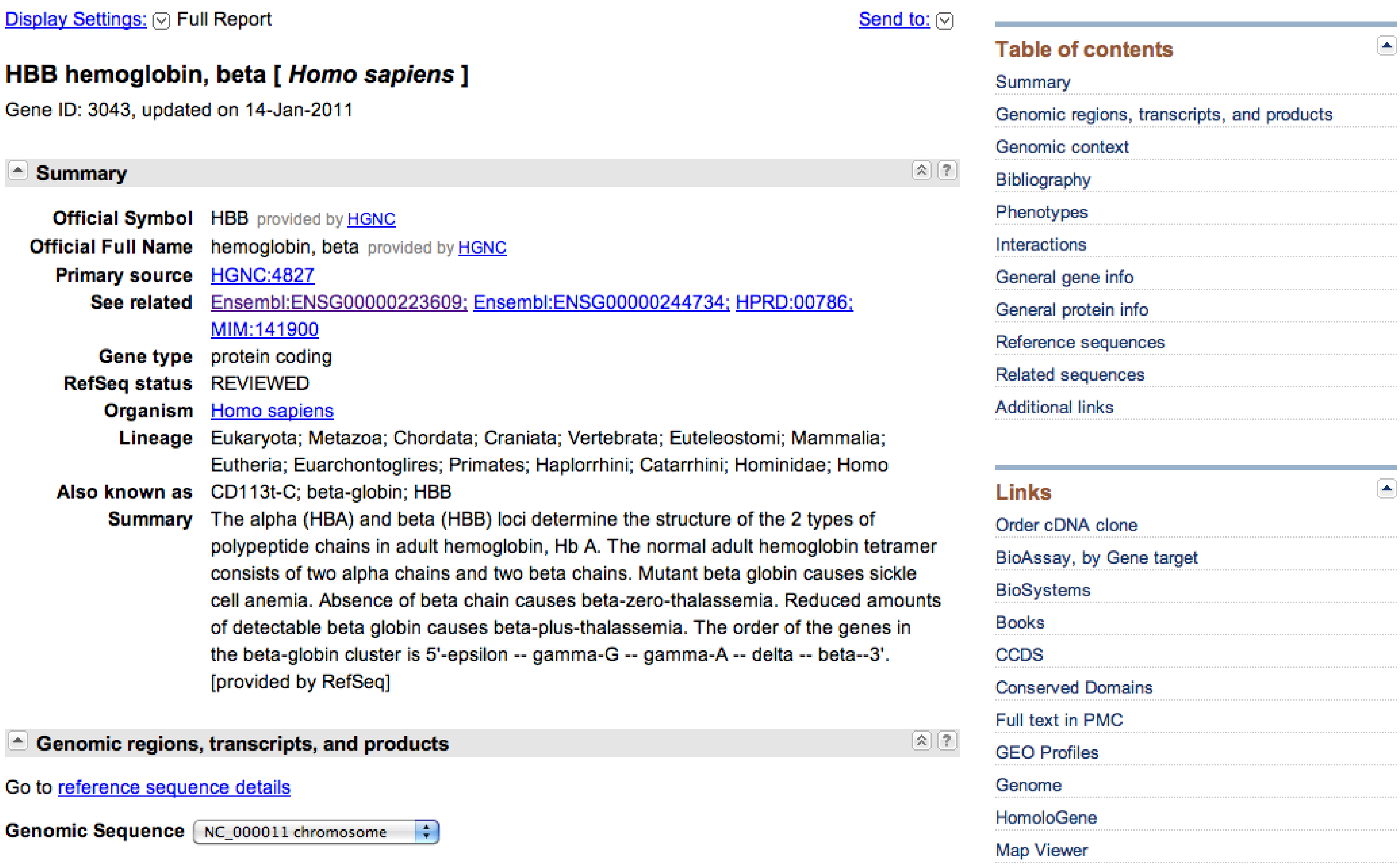


Associated conditions

Description
alpha Thalassemia MedGen: C0002312, OMIM: 604131, GeneReviews: Alpha-Thalassemia
beta Thalassemia MedGen: C0005283, OMIM: 613985, GeneReviews: Beta-Thalassemia
Beta-thalassemia, dominant inclusion body type MedGen: C1858990, OMIM: 603902, GeneReviews: Not available
Fetal hemoglobin quantitative trait locus 1 MedGen: C1841621, OMIM: 141749, GeneReviews: Not available
Hb SS disease MedGen: C0002895, OMIM: 603903, GeneReviews: Sickle Cell Disease
Heinz body anemias MedGen: C0700299, OMIM: 140700, GeneReviews: Not available
Susceptibility to malaria MedGen: C1970028, OMIM: 611162, GeneReviews: Not available

Human Beta-Hemoglobin Gene Entry

http://www.ncbi.nlm.nih.gov/gene/3043



HBB hemoglobin, beta [*Homo sapiens*]

Gene ID: 3043, updated on 14-Jan-2011

Summary

Official Symbol	HBB provided by HGNC
Official Full Name	hemoglobin, beta provided by HGNC
Primary source	HGNC:4827
See related	Ensembl:ENSG00000223609; Ensembl:ENSG00000244734; HPRD:00786; MIM:141900
Gene type	protein coding
RefSeq status	REVIEWED
Organism	Homo sapiens
Lineage	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo
Also known as	CD113t-C; beta-globin; HBB
Summary	The alpha (HBA) and beta (HBB) loci determine the structure of the 2 types of polypeptide chains in adult hemoglobin, Hb A. The normal adult hemoglobin tetramer consists of two alpha chains and two beta chains. Mutant beta globin causes sickle cell anemia. Absence of beta chain causes beta-zero-thalassemia. Reduced amounts of detectable beta globin causes beta-plus-thalassemia. The order of the genes in the beta-globin cluster is 5'-epsilon -- gamma-G -- gamma-A -- delta -- beta--3'. [provided by RefSeq]

Genomic regions, transcripts, and products

Go to reference sequence details

Genomic Sequence NC_000011 chromosome

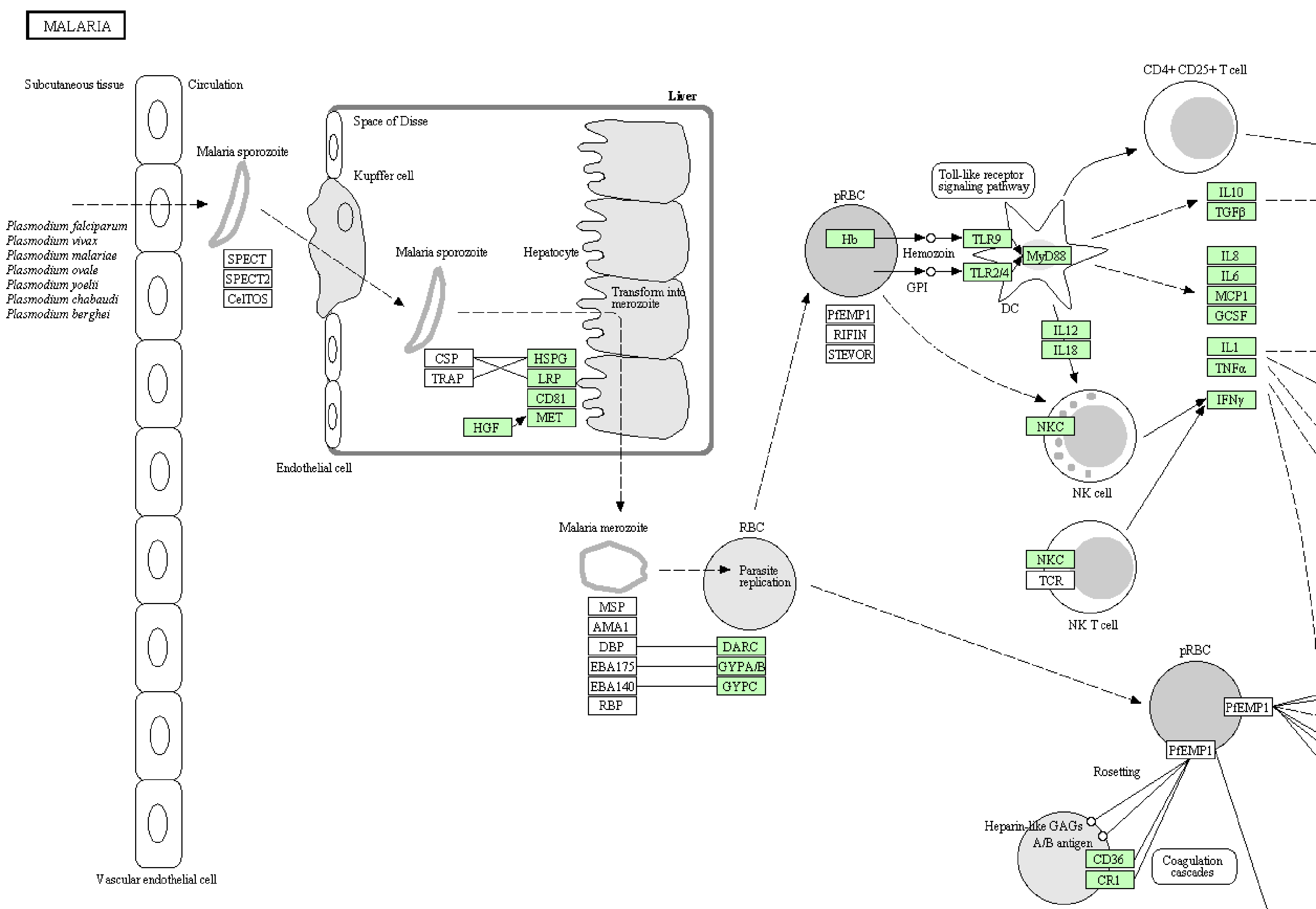
Table of contents

Links

- Order cDNA clone
- BioAssay, by Gene target
- BioSystems
- Books
- CCDS
- Conserved Domains
- Full text in PMC
- GEO Profiles
- Genome
- HomoloGene
- Map Viewer

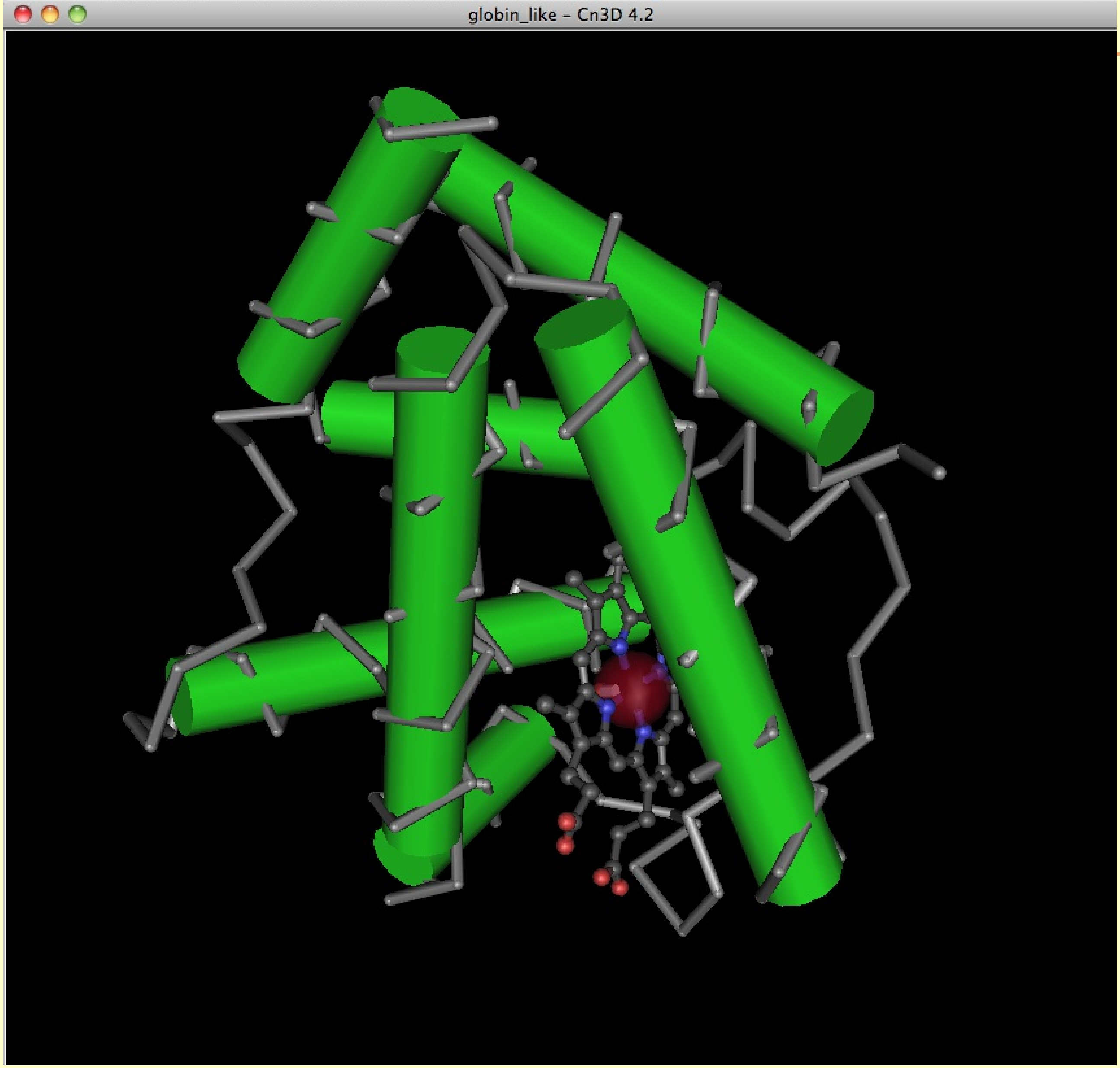
Human Beta-Hemoglobin Biosystems

http://www.ncbi.nlm.nih.gov/gene/3043



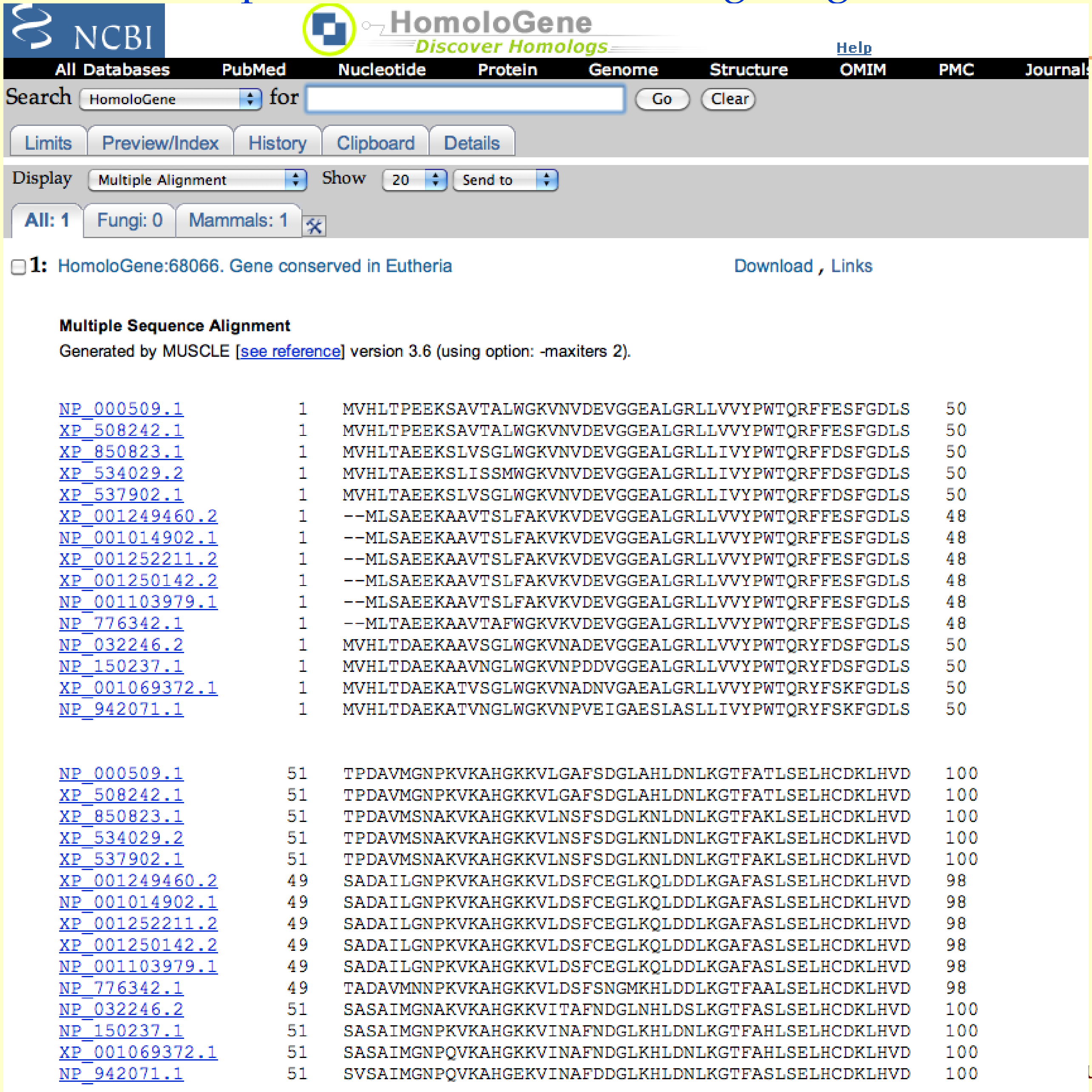
Human Beta-Hemoglobin Conserved Domain

http://www.ncbi.nlm.nih.gov/gene/3043



Human Beta-Hemoglobin Homologene

http://www.ncbi.nlm.nih.gov/gene/3043



☐ **1:** HomoloGene:68066. Gene conserved in Eutheria

Download , Links

Multiple Sequence Alignment

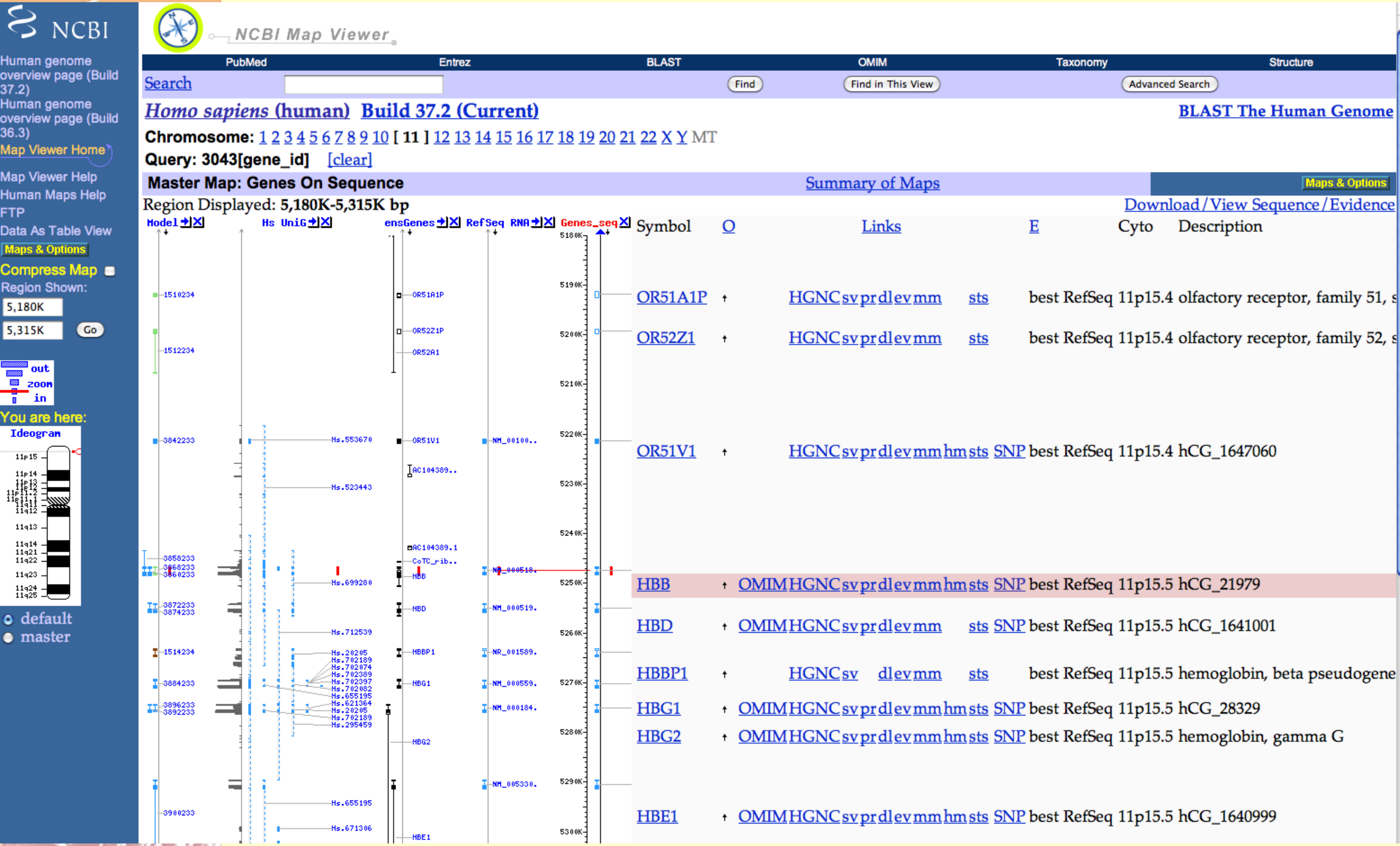
Generated by MUSCLE [see reference] version 3.6 (using option: -maxiters 2).

```
NP_000509.1          1    MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLS    50
XP_508242.1          1    MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLS    50
XP_850823.1          1    MVHLTAEEKSLVSGLWGKVNVDEVGGEALGRLLIVYPWTQRFFDSFGDLS    50
XP_534029.2          1    MVHLTAEEKSLISSMWGKVNVDEVGGEALGRLLIVYPWTQRFFDSFGDLS    50
XP_537902.1          1    MVHLTAEEKSLVSGLWGKVNVDEVGGEALGRLLIVYPWTQRFFDSFGDLS    50
XP_001249460.2       1    --MLSAEEKAAVTSLFAKVKVDEVGGEALGRLLVVYPWTQRFFESFGDLS    48
NP_001014902.1       1    --MLSAEEKAAVTSLFAKVKVDEVGGEALGRLLVVYPWTQRFFESFGDLS    48
XP_001252211.2       1    --MLSAEEKAAVTSLFAKVKVDEVGGEALGRLLVVYPWTQRFFESFGDLS    48
XP_001250142.2       1    --MLSAEEKAAVTSLFAKVKVDEVGGEALGRLLVVYPWTQRFFESFGDLS    48
NP_001103979.1       1    --MLSAEEKAAVTSLFAKVKVDEVGGEALGRLLVVYPWTQRFFESFGDLS    48
NP_776342.1          1    --MLTAEEKAAVTAFWGKVKVDEVGGEALGRLLVVYPWTQRFFESFGDLS    48
NP_032246.2          1    MVHLTDAEKAAVSGLWGKVNADEVGGEALGRLLVVYPWTQRYFDSFGDLS    50
NP_150237.1          1    MVHLTDAEKAAVNGLWGKVNPDDVGGEALGRLLVVYPWTQRYFDSFGDLS    50
XP_001069372.1       1    MVHLTDAEKATVSGLWGKVNADNVGAEALGRLLVVYPWTQRYFSKFGDLS    50
NP_942071.1          1    MVHLTDAEKATVNGLWGKVNPVEIGAESLASLLIVYPWTQRYFSKFGDLS    50


NP_000509.1         51    TPDAVMGNPKVKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVD    100
XP_508242.1         51    TPDAVMGNPKVKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVD    100
XP_850823.1         51    TPDAVMSNAKVKAHGKKVLNSFSDGLKNLDNLKGTFAKLSELHCDKLHVD    100
XP_534029.2         51    TPDAVMSNAKVKAHGKKVLNSFSDGLKNLDNLKGTFAKLSELHCDKLHVD    100
XP_537902.1         51    TPDAVMSNAKVKAHGKKVLNSFSDGLKNLDNLKGTFAKLSELHCDKLHVD    100
XP_001249460.2      49    SADAILGNPKVKAHGKKVLDSFCEGLKQLDDLKGAFASLSELHCDKLHVD    98
NP_001014902.1      49    SADAILGNPKVKAHGKKVLDSFCEGLKQLDDLKGAFASLSELHCDKLHVD    98
XP_001252211.2      49    SADAILGNPKVKAHGKKVLDSFCEGLKQLDDLKGAFASLSELHCDKLHVD    98
XP_001250142.2      49    SADAILGNPKVKAHGKKVLDSFCEGLKQLDDLKGAFASLSELHCDKLHVD    98
NP_001103979.1      49    SADAILGNPKVKAHGKKVLDSFCEGLKQLDDLKGAFASLSELHCDKLHVD    98
NP_776342.1         49    TADAVMNNPKVKAHGKKVLDSFSNGMKHLDDLKGTFAALSELHCDKLHVD    98
NP_032246.2         51    SASAIMGNAKVKAHGKKVITAFNDGLNHLDSLKGTFASLSELHCDKLHVD    100
NP_150237.1         51    SASAIMGNPKVKAHGKKVINAFNDGLKHLDNLKGTFAHLSELHCDKLHVD    100
XP_001069372.1      51    SASAIMGNPQVKAHGKKVINAFNDGLKHLDNLKGTFAHLSELHCDKLHVD    100
NP_942071.1         51    SVSAIMGNPQVKAHGEKVINAFDDGLKHLDNLKGTFASLSELHCDKLHVD    100
```

Human Beta-Hemoglobin Genome Map Viewer

http://www.ncbi.nlm.nih.gov/gene/3043



Homo sapiens (human) Build 37.2 (Current)

Chromosome: 1 2 3 4 5 6 7 8 9 10 [11] 12 13 14 15 16 17 18 19 20 21 22 X Y MT

Query: 3043[gene_id] [clear]

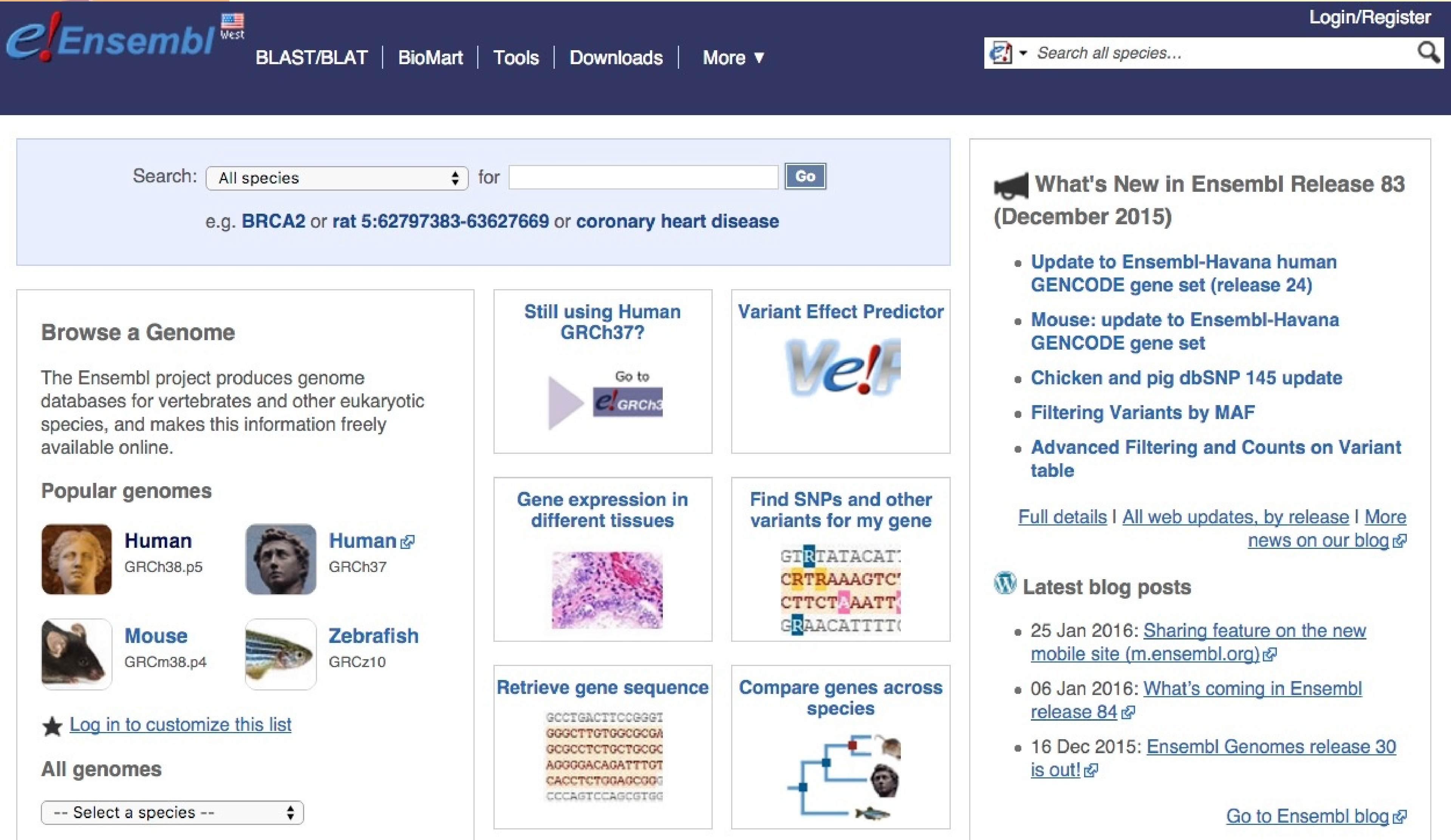
Master Map: Genes On Sequence

Region Displayed: 5,180K-5,315K bp

Symbol	O	Links	E	Cyto	Description
OR51A1P	↑	HGNC sv pr dl ev mm sts	best RefSeq	11p15.4	olfactory receptor, family 51, s
OR52Z1	↑	HGNC sv pr dl ev mm sts	best RefSeq	11p15.4	olfactory receptor, family 52, s
OR51V1	↑	HGNC sv pr dl ev mm hm sts SNP	best RefSeq	11p15.4	hCG_1647060
HBB	↑	OMIM HGNC sv pr dl ev mm hm sts SNP	best RefSeq	11p15.5	hCG_21979
HBD	↑	OMIM HGNC sv pr dl ev mm sts SNP	best RefSeq	11p15.5	hCG_1641001
HBBP1	↑	HGNC sv dl ev mm sts	best RefSeq	11p15.5	hemoglobin, beta pseudogene
HBG1	↑	OMIM HGNC sv pr dl ev mm hm sts SNP	best RefSeq	11p15.5	hCG_28329
HBG2	↑	OMIM HGNC sv pr dl ev mm hm sts SNP	best RefSeq	11p15.5	hemoglobin, gamma G
HBE1	↑	OMIM HGNC sv pr dl ev mm hm sts SNP	best RefSeq	11p15.5	hCG_1640999

EBI Genomes Home Page

http://www.ensembl.org/



Search: All species for Go

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

Browse a Genome

The Ensembl project produces genome databases for vertebrates and other eukaryotic species, and makes this information freely available online.

Popular genomes

Human GRCh38.p5	**Human** GRCh37
Mouse GRCm38.p4	**Zebrafish** GRCz10

★ Log in to customize this list

All genomes

-- Select a species --

Still using Human GRCh37?

Go to GRCh37

Variant Effect Predictor

Gene expression in different tissues

Find SNPs and other variants for my gene

Retrieve gene sequence

Compare genes across species

What's New in Ensembl Release 83 (December 2015)

- Update to Ensembl-Havana human GENCODE gene set (release 24)
- Mouse: update to Ensembl-Havana GENCODE gene set
- Chicken and pig dbSNP 145 update
- Filtering Variants by MAF
- Advanced Filtering and Counts on Variant table

Full details | All web updates, by release | More news on our blog

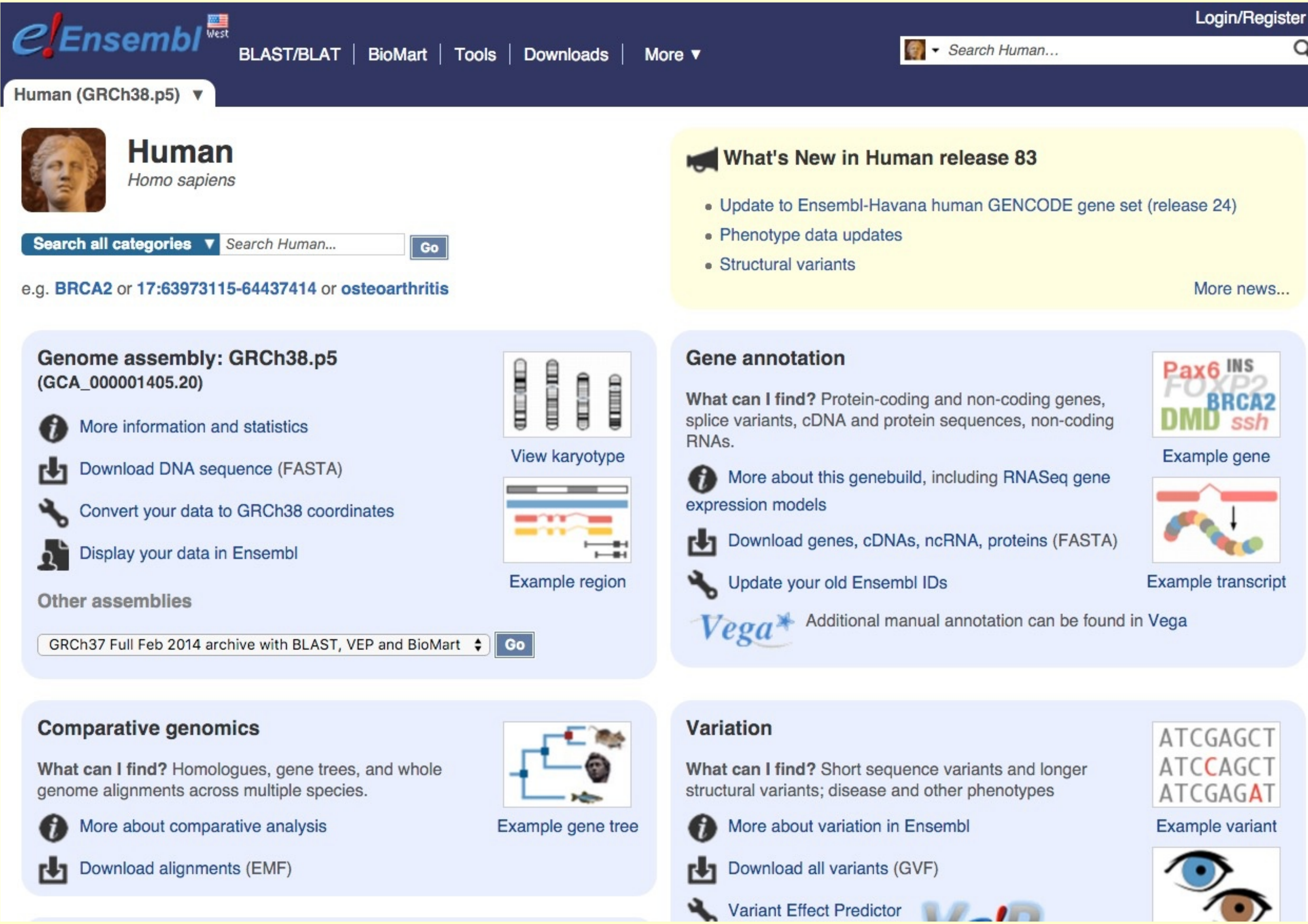
Latest blog posts

- 25 Jan 2016: Sharing feature on the new mobile site (m.ensembl.org)
- 06 Jan 2016: What's coming in Ensembl release 84
- 16 Dec 2015: Ensembl Genomes release 30 is out!

Go to Ensembl blog

Ensembl Human Genome

http://uswest.ensembl.org/Homo_sapiens/Info/Index



Human

Homo sapiens

e.g. **BRCA2** or **17:63973115-64437414** or **osteoarthritis**

What's New in Human release 83

- Update to Ensembl-Havana human GENCODE gene set (release 24)
- Phenotype data updates
- Structural variants

More news...

Genome assembly: GRCh38.p5
(GCA_000001405.20)

More information and statistics

Download DNA sequence (FASTA)

Convert your data to GRCh38 coordinates

Display your data in Ensembl

Other assemblies

GRCh37 Full Feb 2014 archive with BLAST, VEP and BioMart ⬍ Go

Gene annotation

What can I find? Protein-coding and non-coding genes, splice variants, cDNA and protein sequences, non-coding RNAs.

More about this genebuild, including RNASeq gene expression models

Download genes, cDNAs, ncRNA, proteins (FASTA)

Update your old Ensembl IDs

Additional manual annotation can be found in Vega

Comparative genomics

What can I find? Homologues, gene trees, and whole genome alignments across multiple species.

More about comparative analysis

Download alignments (EMF)

Variation

What can I find? Short sequence variants and longer structural variants; disease and other phenotypes

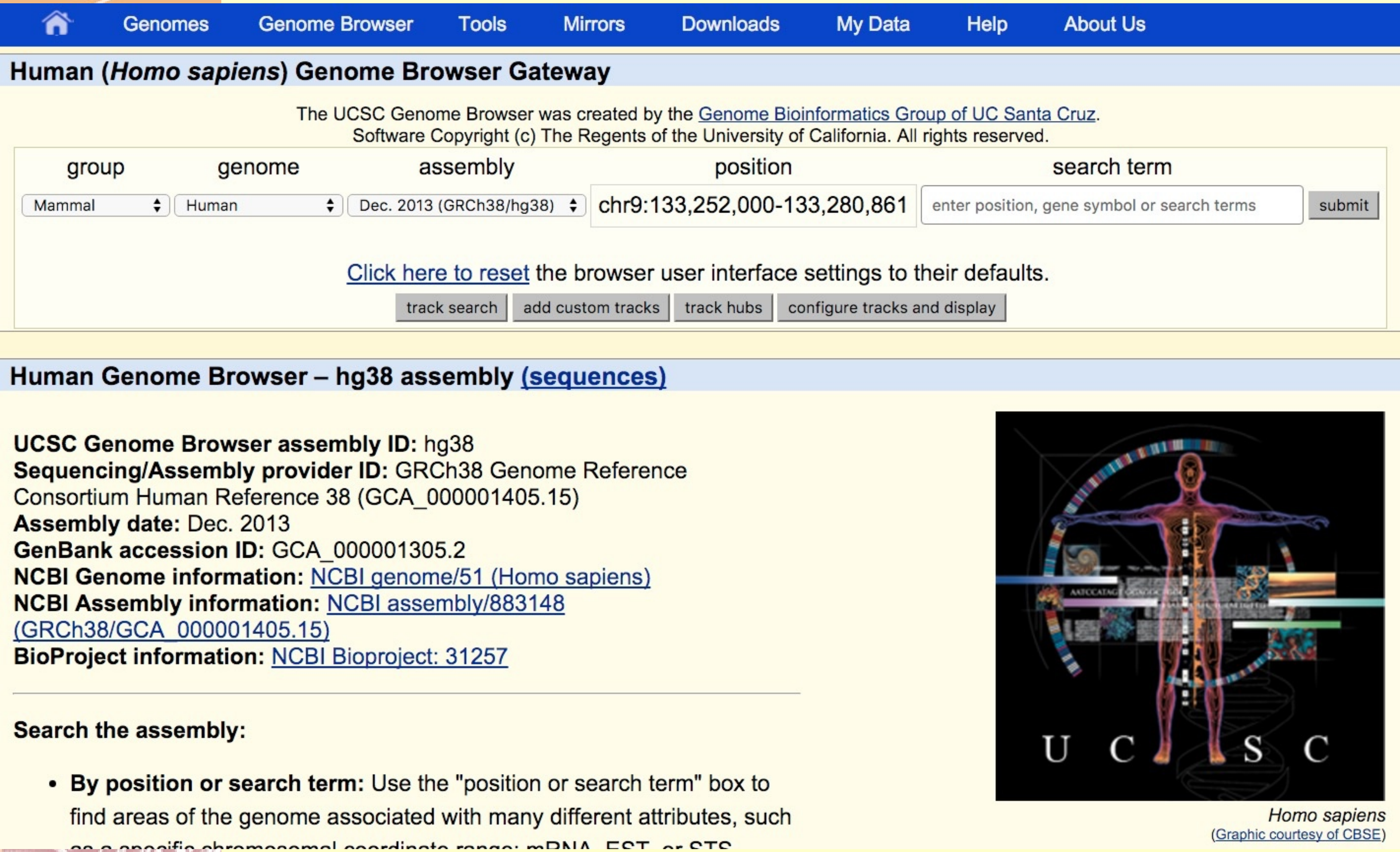
More about variation in Ensembl

Download all variants (GVF)

Variant Effect Predictor

UCSC Genome Browser

http://genome.ucsc.edu/



Human (*Homo sapiens*) Genome Browser Gateway

The UCSC Genome Browser was created by the Genome Bioinformatics Group of UC Santa Cruz.
Software Copyright (c) The Regents of the University of California. All rights reserved.

group	genome	assembly	position	search term
Mammal	Human	Dec. 2013 (GRCh38/hg38)	chr9:133,252,000-133,280,861	enter position, gene symbol or search terms

Click here to reset the browser user interface settings to their defaults.

Human Genome Browser – hg38 assembly (sequences)

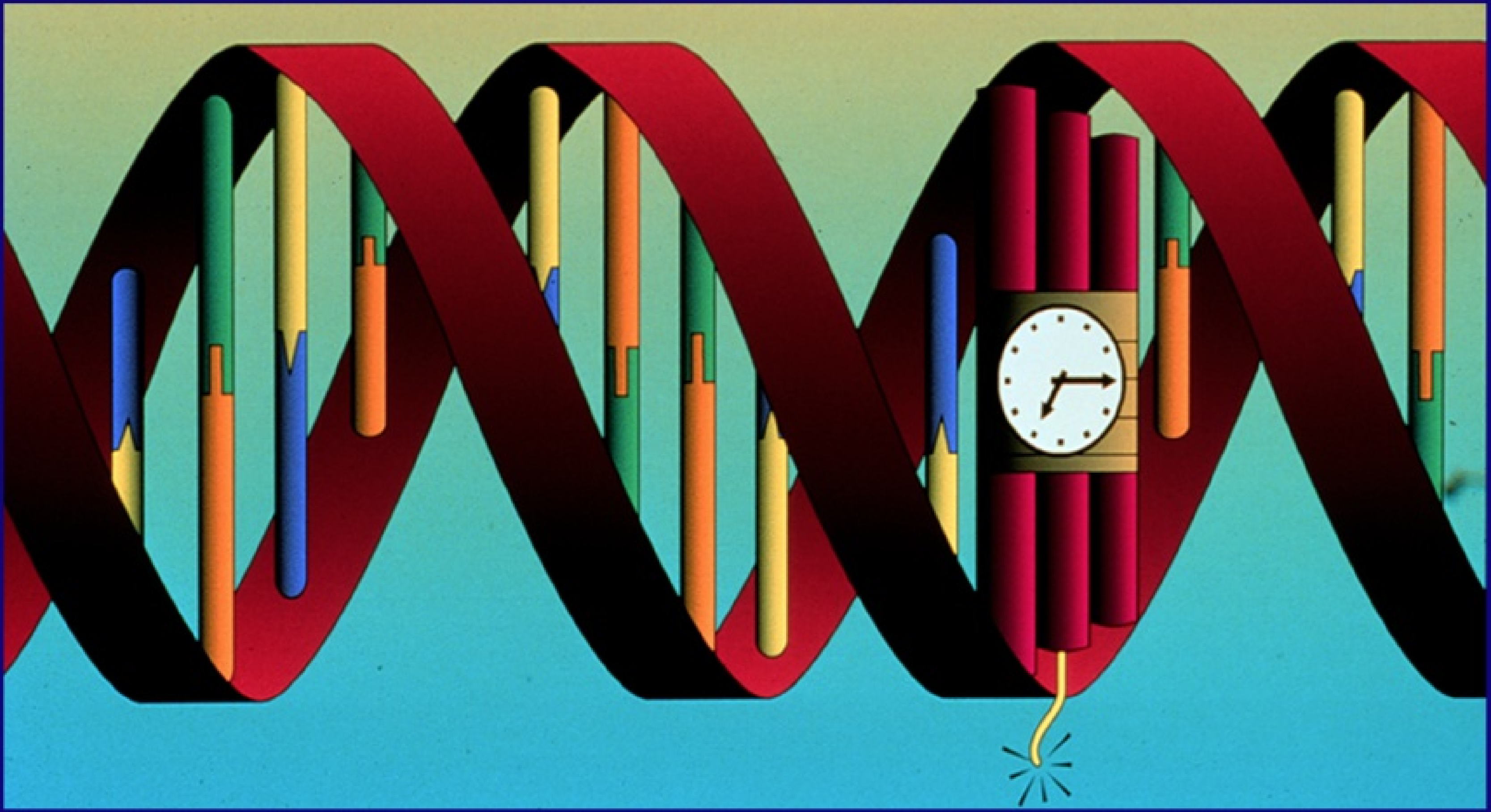
UCSC Genome Browser assembly ID: hg38
Sequencing/Assembly provider ID: GRCh38 Genome Reference Consortium Human Reference 38 (GCA_000001405.15)
Assembly date: Dec. 2013
GenBank accession ID: GCA_000001305.2
NCBI Genome information: NCBI genome/51 (Homo sapiens)
NCBI Assembly information: NCBI assembly/883148 (GRCh38/GCA_000001405.15)
BioProject information: NCBI Bioproject: 31257

Search the assembly:

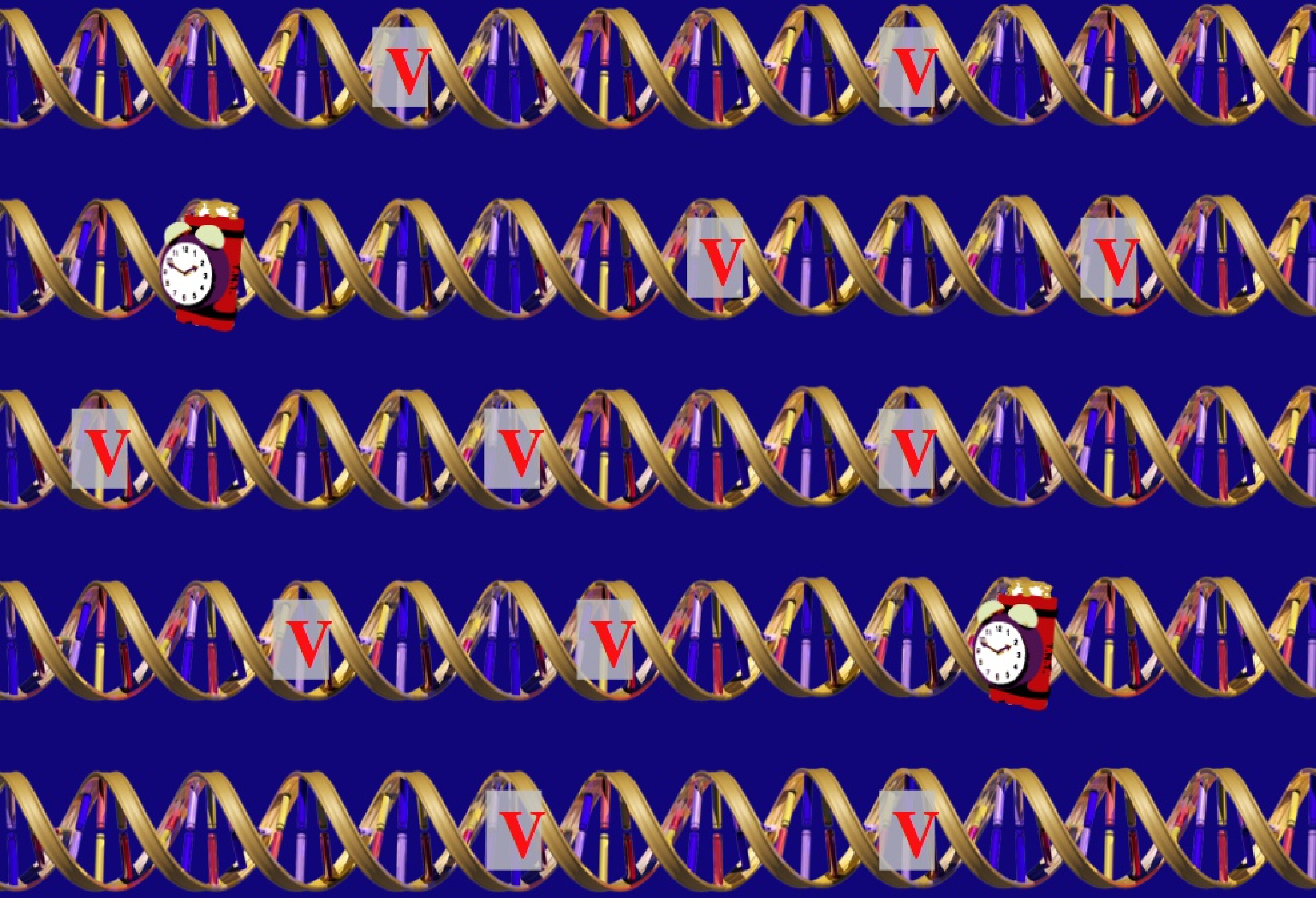
- **By position or search term:** Use the "position or search term" box to find areas of the genome associated with many different attributes, such as a specific chromosomal coordinate range; mRNA, EST, or STS

Homo sapiens
(Graphic courtesy of CBSE)

All humans are ~99.7% identical at the DNA sequence level, and yet…



all of us carry a significant number of ‘glitches’ in our genomes.


V
V
V
V
V
V
V
V
V
V
V