

Introduction to Biological Databases

Bioinformatics Resources

Alessandro Romanel
2019-2020

Biological data

- Amino acid sequences (DNA/RNA)
- Protein sequences
- Structure of nucleic acids and proteins
- Functional data
 - Proteins functions
 - PPI (protein-protein interactions)
 - Post-transcriptional modifications
 - Expression
 - Polymorphisms
 - ...

Motivation

- Huge amount of biological data
- Store efficiently these data
- Accessibility: researches, doctors, biologists, students, ...
- Recover and analyze the data
- Generate new knowledge

Biological databases

- Huge number of biological databases available today
- We can distinguish between
 - Primary databases (sequences of nucleotides and aminoacids)
 - Derived and specialized databases (protein domains and motifs, protein structures, genes, transcripts, expression profiles, variations, pathways,...)

Biological databases

- Information are derived from
 - Literature
 - Lab analyses (in-vitro and in-vivo)
 - Bioinformatics analyses (in-silico)
- Each database is characterized by a central biological element which constitutes the object around which the principal entry of the database is constructed

Nomenclature

- One of the main problems related with biological database is nomenclature
- Same gene, different names (eg. TRF2 also known as TLP or TLF)
- Same name, different genes (eg. TRF stands for TBP Related Factor but also Transferrin or Telomeric Repeat Binding Factor)
- It is therefore needed a way to uniquely identify genes and proteins, and to manage the large amount of information related to them
- In primary data banks (but also derived ones) each element (gene, sequence, etc) is uniquely identified by an *accession number*

Biological sequences

- Sequences of nucleotides are represented by strings on an alphabet of 4 characters
- Sequences of amino acids are represented by strings on an alphabet of 20 characters

Alphabets

| DNA | |
|-----|----------|
| A | Adenine |
| C | Cytosine |
| G | Guanine |
| T | Thymine |

| RNA | |
|-----|----------|
| A | Adenine |
| C | Cytosine |
| G | Guanine |
| U | Uracil |

Alphabets

| | | | | | |
|---|-----------|---|-------------|---|------------|
| A | Alanine | C | Cysteine | D | Aspartate |
| E | Glutamate | F | Phenylanine | G | Glicine |
| H | Istdidine | I | Isoleucine | K | Lysine |
| L | Leucine | M | Methionine | N | Asparagine |
| P | Proline | Q | Glutamine | R | Arginine |
| S | Serine | T | Threonine | Y | Tyrosine |
| V | Valine | W | Tryptophane | | |

Standard IUB/IUPAC

Nucleic acids

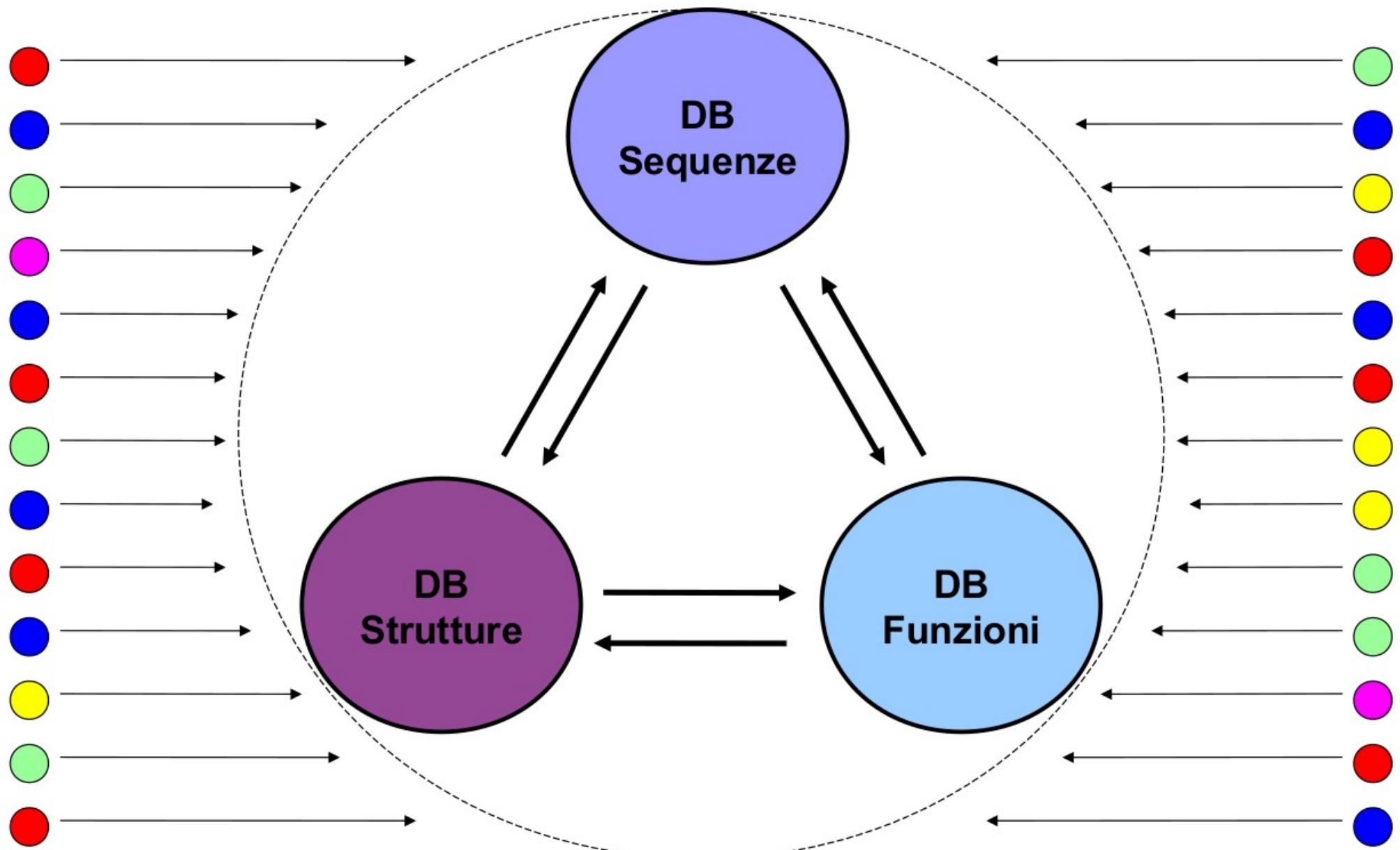
| | | | |
|---|--------------------|---|---------------|
| R | G o A (Purine) | B | G T C |
| Y | T o C (Pirimidine) | D | G A T |
| K | G o T | H | A C T |
| M | A o C | V | G C A |
| W | A o T | N | A C G T (Any) |
| | | - | Gap |

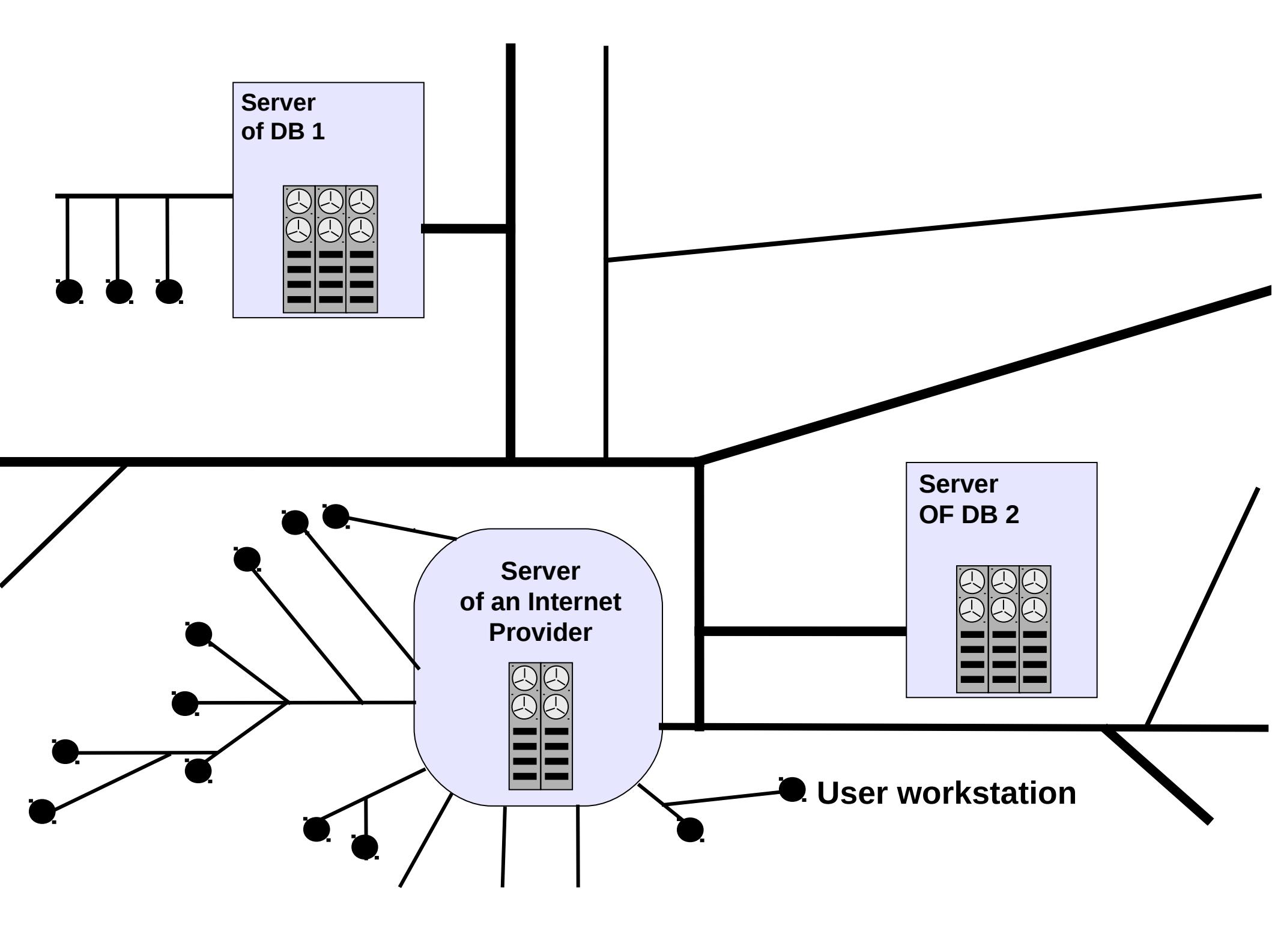
Amino acids

| | | | | |
|---|--------------------------|---|-------------------------|-----------------|
| U | Selenocysteine | Z | Glutamate and Glutamine | Qualsiasi (Any) |
| B | Aspartate and Asparagine | * | Stop traduzione | Gap |

| | | | | | | | | | |
|-------|---|---|---|---|---|---|---|---|---|
| G | C | C | G | G | A | A | G | T | G |
| A | C | C | G | G | A | A | G | C | A |
| G | C | C | G | G | A | T | G | T | A |
| A | C | C | G | G | A | A | G | C | T |
| A | C | C | G | G | A | T | A | T | A |
| C | C | C | G | G | A | A | G | T | G |
| A | C | A | G | G | A | A | G | T | C |
| G | C | C | G | G | A | T | G | C | A |
| T | C | C | G | G | A | A | G | T | A |
| A | C | A | G | G | A | A | G | C | G |
| A | C | A | G | G | A | T | A | T | G |
| T | C | C | G | G | A | A | A | C | C |
| A | C | A | G | G | A | T | A | T | C |
| C | A | A | G | G | A | C | G | A | C |
| T | C | T | G | G | A | C | C | C | T |
| <hr/> | | | | | | | | | |
| N | C | M | G | G | A | W | G | Y | N |

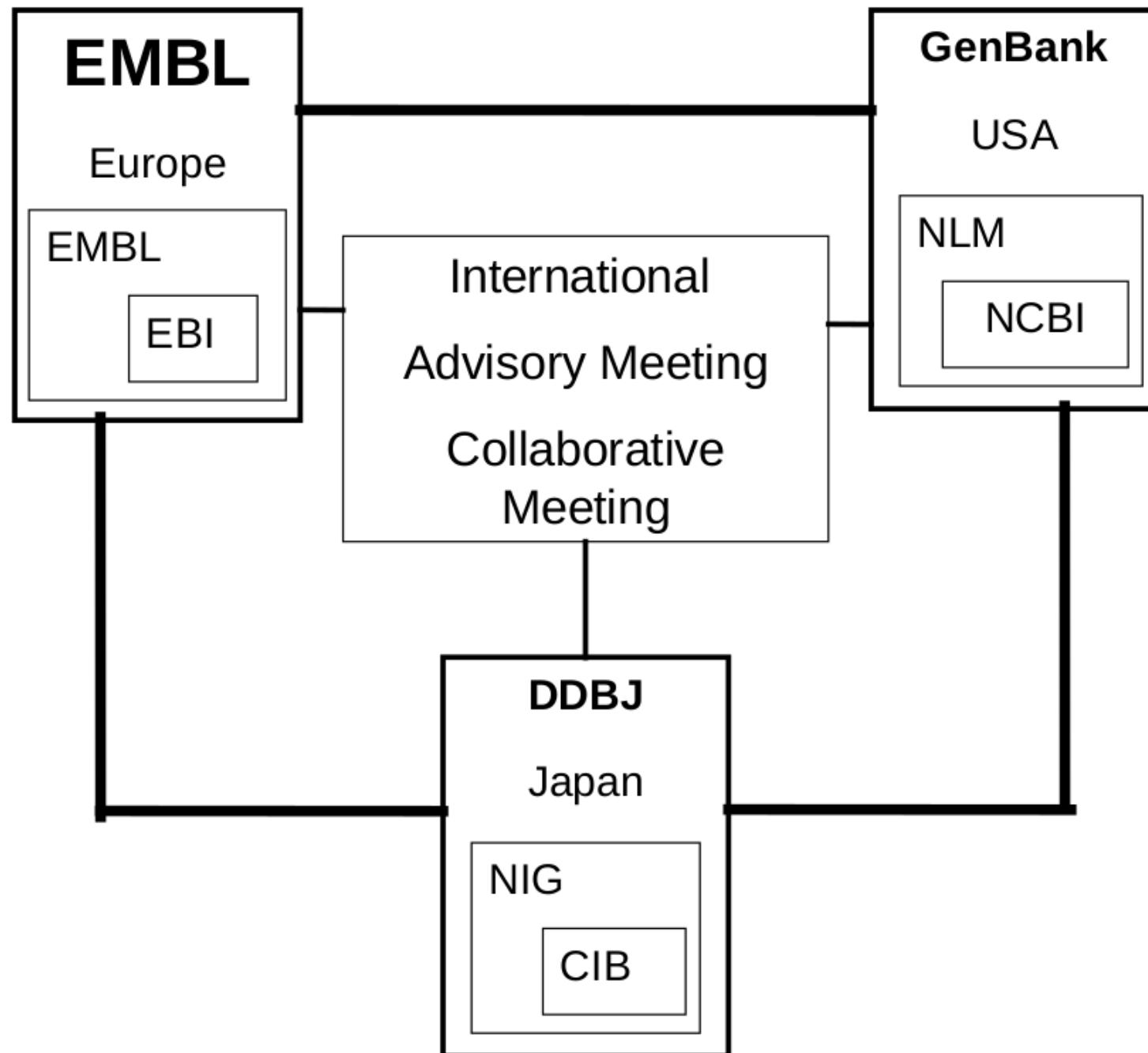
| | |
|---|--------------------|
| A | adenine |
| C | cytosine |
| G | guanine |
| T | thymine |
| U | uracile |
| R | G A (purine) |
| Y | T C (pyrimidine) |
| K | G T (groupe keto) |
| M | A C (groupe amino) |
| S | G C (strong) |
| W | A T (weak) |
| B | G T C (pas A) |
| D | G A T (pas C) |
| H | A C T (pas G) |
| V | G C A (pas T) |
| N | A G C T |





Primary databases

- There are three main primary databases:
 - GenBank (NCBI – National Center for Biotechnology Information).
 - EMBL datalibrary (EMBL – European Molecular Biology Laboratory).
 - DDBJ (DNA Database of Japan)
- Cooperation and share biological data
 - Data format differs



GenBank – primary databank

- Only nucleotide sequences
- Just store and archive
 - Storically important
 - Submitter point of view
 - Redundant
- Data
 - Direct submission
 - Batch submission
 - FTP accounts (genomic data)

GenBank record

| | | | | | | |
|------------|---|---------|------|--------|-----|-------------|
| LOCUS | HSHMLHI | 2503 bp | mRNA | linear | PRI | 31-MAR-1994 |
| DEFINITION | Human DNA mismatch repair (hmlh1) mRNA, complete cds. | | | | | |
| ACCESSION | U07418 | | | | | |
| VERSION | U07418.1 GI:466461 | | | | | |
| KEYWORDS | . | | | | | |
| SOURCE | Homo sapiens (human) | | | | | |
| ORGANISM | <u>Homo sapiens</u> Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. | | | | | |
| REFERENCE | 1 (bases 1 to 2503) | | | | | |
| AUTHORS | Papadopoulos,N., Nicolaides,N.C., Wei,Y.F., Ruben,S.M., Carter,K.C., Rosen,C.A., Haseltine,W.H., Fleischmann,R.D., Fraser,C.M., Adams,M.D., Venter,J.C., Hamilton,S.R., Petersen,G.M., Watson,P., Lynch,H.T., Peltomaki,P., Mecklin,J.P., Chapelle,A.D., Kinzler,K.W. and Vogelstein,B. | | | | | |
| TITLE | Mutation of a mutL homolog in hereditary colon cancer | | | | | |
| JOURNAL | Science 263 (5153), 1625-1629 (1994) | | | | | |
| MEDLINE | <u>94174309</u> | | | | | |
| PUBMED | <u>8128251</u> | | | | | |
| REFERENCE | 2 (bases 1 to 2503) | | | | | |
| AUTHORS | Wei,Y.F. | | | | | |
| TITLE | Direct Submission | | | | | |
| JOURNAL | Submitted (04-MAR-1994) Ying-Fei Wei, Molecular Biology, Human Genome Sciences, Inc., 9620 Medical Center Drive, Rockville, MD 20850, USA | | | | | |

GenBank record

| | | | | | | | |
|------------|---|-----------|------|--------|-----|-------------|--|
| LOCUS | HSHMLHI | 2503 bp | mRNA | linear | PRI | 31-MAR-1994 | |
| DEFINITION | Human DNA mismatch repair (hmlh1) mRNA, complete genome | | | | | | |
| ACCESSION | U07418 | | | | | | |
| VERSION | U07418.1 | GI:466461 | | | | | |

ACCESSION **U07418**

Accession

- Stable
- Universal
- Simple

VERSION **U07418.1 GI:466461**

REFERENCE 1 (bases 1 to 2503)

AUTHORS Papadimitriou, N., Nicolaides, N.C., Wei, Y.F., Ruben, S.M.,

Posen, C.A., Haseltine, W., Fleischmann, R.D.,
Adams, M.D., Venter, J.C., Peltomaki, P., Petersen, G.M.,
Lynch, H.T., Peltomaki, P., Vogelstein, B., Capelle, A.D.,

Version

Modification to the sequence

GI number

NCBI internal use

TITLE Mutation of a mutL homolog in hereditary colon cancer

JOURNAL Science 263 (5153), 1625-1629 (1994)

MEDLINE [94174309](#)

PUBMED [8128251](#)

REFERENCE 2 (bases 1 to 2503)

AUTHORS Wei, Y.F.

TITLE Direct Submission

JOURNAL Submitted (04-MAR-1994) Ying-Fei Wei, Molecular Biology, Human
Genome Sciences, Inc., 9620 Medical Center Drive, Rockville, MD
20850, USA

GenBank record

Well annotated

| | |
|-----------------------|-----|
| LOCUS | HSH |
| DEFINITION | Hum |
| ACCESSION | U07 |
| VERSION | U07 |
| ACCESSION | UG |
| VERSION | UG |
| REFERENCE | 1 |
| AUTHORS | Pap |
| Version | |
| Modification to the s | |
| TITLE | Mut |
| JOURNAL | Sci |
| MEDLINE | 941 |
| PUBMED | 812 |
| REFERENCE | 2 |
| AUTHORS | Wei |
| TITLE | Dir |

FEATURES Location/Qualifiers
source 1..2503
/organism="Homo sapiens"
/db_xref="taxon: 9606"
/chromosome="3"
/map="p21"
/tissue_type="gall bladder"
/dev_stage="adult"
gene 1..2503
/gene="hmlh1"
CDS 42..2312
/gene="hmlh1"
/function="DNA mismatch repair"
/note="human homolog of E. coli mutL gene product,
Swiss-Prot Accession Number P23367"
/codon_start=1
/protein_id="[AAA17374.1](#)"
/db_xref="GI: 466462"
/translation="MSFVAGVIRRLDETVVNRIAAGEVIQRPANAICKEMIENCLDAKS
TSIQIVKEGGLKLIQIQDNGTGIRKEDLDIVCERFTTSKLQSFEGLASISTYGRGE
ALASISHVAHTITTKTADGKCAYRASYSDGKLAPPKPCAGNQGTQITVEDLFYNIA
TRRKALKNPSEYGYKILEVVGRYSVHNAGISFSVKKQGETVADVRTLPNASTVDNIRS
VFGNAVSRELIEIGCEDKTLAFKMNGYISNANYSVKKCIFLLFINHRLVESTSLRKA
ETVYAAYLKPKNTHPFPLYLSLEISPQNVDVNVHPTKHEVHFLHEESILERVQQHIESKL
LGSNSSRMYFTQTLLPGLAGPSGEMVKSTTSSTSGSSDKVYAHQMVRTDSREQK
LDAFLQPLSKPLSSQPQAIVTEDKTDISSGRARQDEEMLELPAPAEVAAKNQSLEGD
TTKGTSSEMSEKRGPTSSNPRKRHREDSDVEMVEDDSRKEMTAACTPRRIINLTSVLS
LQEELINEQGHEVLREMLHNHSFVGCVNPQWALAQHQTLYLLNTTKLSEELFYQILY
DFANFGVRLSEPAPLFDLAMLALDSPESGWTEEDGPKEGLAEYIVEFLKKKAEMLAD
YFSLEIDEEGNLIGPLLIDNYVPPL EGLPIFLR LATEVNWDEEKECFESLSKECAM
FYSIRKQYISEESTLSGQQSEVPGSI PNSWKWTVEHIVYKALRSHILPPKHFTEDGNI
LOLANLPDLYKVFERC"

home Sciences, Inc., 9620 Medical Center Drive, Rockville, MD
20850, USA

GenBank record

| ACCESSION U07 | | FEATURES | Location/Qualifiers |
|-------------------------------------|--|-------------|---|
| VERSION U07 | | source | 1..2503 /organism="Homo sapiens" /db_xref="taxon:9606" |
| | | /chromosome | /map="p11.2" |
| | | /tissue | BASE COUNT 723 a 539 c 599 g 642 t |
| | | /dev_st | ORIGIN |
| ACCESSION U07 | | gene | 1 gttgaacatc tagacgttcc ttggctt ctggcgccaa aatgtcggtc gtggcagggg 61 ttattcgccg gctggacag acatgttgc accgcattgc ggcggggggaa gttatccaga |
| VERSION U07 | | CDS | 121 ggccagctaa tgcttatcaa gagatgttgc agaactgttt agatgcaaaa tccacaagta 181 ttcaagtgtat tgtaaaagag ggaggcctgtat gatgttgc gatccaagac aatggcacgc 241 ggatcagaa agaagatctg gatattgtat gtgaaagggtt cactactgt aaactgcagt 301 ctttggggat tttagccgtt atttcttccat atggcttc aggtggggct ttggccggca 361 taaggccatgt ggctcatgtt acttatacaa cggaaaacaggc tgatggaaag tttgtcata 421 gagcaaggta ctcagatggaa aaaaatgttgc cccctcttcc accatgtgtt ggcaatcaag 481 ggaccccgat cacgggtggag gacctttttt acaacatgc caccggaga aaagctttaa 541 aaaatccaaat tgaatgtat gggaaaattt tggaaatgtt tgccggatgt tcagttacaca 601 atgcaggcat tagtttctca gttaaaaaaaaac aaggagac agtagctgtat gttaggacac 661 taccatcatgtt ctcacccatgtt gacaatattt gtcgttcc ttggaaatgtt gtttagtcgac 721 aactgtataga aattggatgtt gaggataaaaa cccttagccctt caaatgtat ggttacatata 781 ccaatgcataa ctactatgtt aagaatgtca ttctttactt cttcatcaac catcgcttgg 841 tagaatccaaat ttcccttgat gggaaaatccat aacatgttgc tgccggatcat ttggccaaaa 901 acacacaccc attctgtatc ctcagtttag aatattgttgc ccaatgtgtt gatgttaat 961 tgcacccccc aaaaatgttgc gttcaatcc ttggccggat gggatccat gggccggat 1021 agcagcacat cgagagcaag ctcctgggtt ccaatttcctt caggatgtac ttccaccat 1081 ctttgcatttcc aggactgtt gggccctctt gggagatgtt taaatccaca acaatgttgc 1141 ctcctgttcc tactttgttgc agtagtgc tttttttttt caccatgttgc gttcgat 1201 atccccggaa acagaatgttgc tttttttttt cttttttttt cttttttttt cttttttttt 1261 agccccccatgttgc tttttttttt cttttttttt cttttttttt cttttttttt cttttttttt 1321 aagatggat gatgttgc gttttttttt cttttttttt cttttttttt cttttttttt cttttttttt 1381 tggggggggat tttttttttt cttttttttt cttttttttt cttttttttt cttttttttt cttttttttt 1441 gcaacccccc aaaaatgttgc tttttttttt cttttttttt cttttttttt cttttttttt cttttttttt 1501 gaaaatgttgc gttttttttt cttttttttt cttttttttt cttttttttt cttttttttt cttttttttt 1561 tttttttttt cttttttttt cttttttttt cttttttttt cttttttttt cttttttttt cttttttttt 1621 accacttccatgttgc tttttttttt cttttttttt cttttttttt cttttttttt cttttttttt 1681 tttttttttt cttttttttt cttttttttt cttttttttt cttttttttt cttttttttt cttttttttt 1741 tttttttttt cttttttttt cttttttttt cttttttttt cttttttttt cttttttttt cttttttttt 1801 tttttttttt cttttttttt cttttttttt cttttttttt cttttttttt cttttttttt cttttttttt 1861 tttttttttt cttttttttt cttttttttt cttttttttt cttttttttt cttttttttt cttttttttt 1921 tttttttttt cttttttttt cttttttttt cttttttttt cttttttttt cttttttttt cttttttttt 1981 tttttttttt cttttttttt cttttttttt cttttttttt cttttttttt cttttttttt cttttttttt 2041 tttttttttt cttttttttt cttttttttt cttttttttt cttttttttt cttttttttt cttttttttt 2101 tttttttttt cttttttttt cttttttttt cttttttttt cttttttttt cttttttttt cttttttttt 2161 tttttttttt cttttttttt cttttttttt cttttttttt cttttttttt cttttttttt cttttttttt 2221 tttttttttt cttttttttt cttttttttt cttttttttt cttttttttt cttttttttt cttttttttt 2281 tttttttttt cttttttttt cttttttttt cttttttttt cttttttttt cttttttttt cttttttttt 2341 tttttttttt cttttttttt cttttttttt cttttttttt cttttttttt cttttttttt cttttttttt 2401 tttttttttt cttttttttt cttttttttt cttttttttt cttttttttt cttttttttt cttttttttt 2461 tttttttttt cttttttttt cttttttttt cttttttttt cttttttttt cttttttttt cttttttttt |
| Version | | REFERENCE | 1 |
| | | AUTHORS | Par |
| Modification to the sequence | | TITLE | Mut |
| | | JOURNAL | Sci |
| | | MEDLINE | 941 |
| | | PUBMED | 812 |
| | | REFERENCE | 2 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | REFERENCE | 1 |
| | | AUTHORS | Wei |
| | | TITLE | Dir |
| | | | |

RefSeq – derived data bank

- Curated and not redundant collection of DNA, RNA and protein sequences.
- Collection of transcripts and proteins
- Each RefSeq represent a single molecule in a particular organism
- RefSeq basis is compiled with a process of:
 - Collaboration
 - Extraction
 - Computationfrom the GenBank.
- Each molecule is accurately annotated reporting the name of the organism, the correct gene symbol for that organism and informative names of proteins when possible

Accession numbers

| Accession prefix | Molecule type |
|------------------|------------------------------------|
| NC_ | Complete genomic molecule |
| NG_ | Genomic region |
| NM_ | mRNA |
| NP_ | Protein |
| NR_ | RNA |
| NT_ ^a | Genomic contig |
| NW_ ^a | Genomic contig (WGS ^b) |
| XM_ ^a | mRNA |
| XP_ ^a | Protein |
| XR_ ^a | RNA |
| NZ_ ^c | Genomic (WGS) |
| ZP_ ^a | Protein, on NZ_ |

^a Computed.

^b Assembly of Whole Genome Shotgun (WGS) sequence data.

^c An ordered collection of WGS for a genome.

Other derived databanks

- DbGap
- Structure
- Gene
- BioSystems
- E altro ancora...

All NCBI databases are accessible through a unique search engine called **Entrez**:

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



Search NCBI

Search NCBI

Search

NCBI Databases

Literature

The World's largest repository of medical and scientific abstracts, full-text articles, books and reports

[Bookshelf](#)

Books and reports

[MeSH](#)

Ontology used for PubMed indexing

[NLM Catalog](#)

Books, journals and more in the NLM Collections

[PubMed](#)

Scientific and medical abstracts/citations

[PubMed Central](#)

Full-text journal articles

Genes

Gene sequences and annotations used as references for the study of orthologs structure, expression, and evolution

[Gene](#)

Collected information about gene loci

[GEO DataSets](#)

Functional genomics studies

[GEO Profiles](#)

Gene expression and molecular abundance profiles

[HomoloGene](#)

Homologous genes sets for selected organisms

[PopSet](#)

Sequence sets from phylogenetic and population studies

[UniGene](#)

Clusters of expressed transcripts

Genetics

Heritable DNA variations, associations with human pathologies, and clinical diagnostics and treatments

[ClinVar](#)

Human variations of clinical significance

[dbGaP](#)

Genotype/phenotype interaction studies

[dbSNP](#)

Short genetic variations

[dbVar](#)

Genome structural variation studies

[GTR](#)

Genetic testing registry

[MedGen](#)

Medical genetics literature and links

[OMIM](#)

Online mendelian inheritance in man

Proteins

Protein sequences, 3-D structures, and tools for the study of functional protein domains and active sites

[Conserved Domains](#)

Conserved protein domains

[Identical Protein Groups](#)

Protein sequences grouped by identity

[Protein](#)

Protein sequences

[Protein Clusters](#)

Sequence similarity-based protein clusters

[Sparcle](#)

Functional categorization of proteins by domain architecture

[Structure](#)

Experimentally-determined biomolecular structures

Genomes

Genome sequence assemblies, large-scale functional genomics data, and source biological samples

[Assembly](#)

Genome assembly information

[BioCollections](#)

Museum, herbaria, and other biorepository collections

[BioProject](#)

Biological projects providing data to NCBI

[BioSample](#)

Descriptions of biological source materials

[Genome](#)

Genome sequencing projects by organism

[Nucleotide](#)

DNA and RNA sequences

[Probe](#)

Sequence-based probes and primers

[SRA](#)

High-throughput sequence reads

[Taxonomy](#)

Taxonomic classification and nomenclature

Chemicals

Repository of chemical information, molecular pathways, and tools for bioactivity screening

[BioSystems](#)

Molecular pathways with links to genes, proteins and chemicals

[PubChem BioAssay](#)

Bioactivity screening studies

[PubChem Compound](#)

Chemical information with structures, information and links

[PubChem Substance](#)

Deposited substance and chemical information



Search NCBI

Tp53



Search

NCBI Databases

Results found in 32 databases for: Tp53

Literature

| | |
|----------------|--------|
| Bookshelf | 989 |
| MeSH | 18 |
| NLM Catalog | 18 |
| PubMed | 16,869 |
| PubMed Central | 39,142 |

Genes

| | |
|--------------|---------|
| Gene | 4,640 |
| GEO DataSets | 6,174 |
| GEO Profiles | 193,438 |
| HomoloGene | 13 |
| PopSet | 38 |
| UniGene | 141 |

Genetics

| | |
|---------|-------|
| ClinVar | 1,738 |
| dbGaP | 0 |
| dbSNP | 6,148 |
| dbVar | 3,663 |
| GTR | 556 |
| MedGen | 86 |
| OMIM | 364 |

Proteins

| | |
|--------------------------|-------|
| Conserved Domains | 4 |
| Identical Protein Groups | 261 |
| Protein | 4,187 |
| Protein Clusters | 3 |
| Sparcle | 22 |
| Structure | 174 |

Genomes

| | |
|----------------|--------|
| Assembly | 0 |
| BioCollections | 0 |
| BioProject | 466 |
| BioSample | 1,154 |
| Genome | 33 |
| Nucleotide | 10,242 |
| Probe | 1,450 |
| SRA | 2,900 |
| Taxonomy | 0 |

Chemicals

| | |
|-------------------|-------|
| BioSystems | 3,874 |
| PubChem BioAssay | 476 |
| PubChem Compound | 0 |
| PubChem Substance | 621 |

| Name/Gene ID | Description | Location | Aliases | MIM |
|--|---|--|---|--------|
| <input type="checkbox"/> TP53 ID: 7157 | tumor protein p53 [<i>Homo sapiens</i> (human)] | Chromosome 17, NC_000017.11 (7668402..7687550, complement) | BCC7, LFS1, P53, TRP53 | 191170 |
| <input type="checkbox"/> Tp53 ID: 24842 | tumor protein p53 [<i>Rattus norvegicus</i> (Norway rat)] | Chromosome 10, NC_005109.4 (56187013..56198449) | Trp53, p53 | |
| <input type="checkbox"/> tp53 ID: 30590 | tumor protein p53 [<i>Danio rerio</i> (zebrafish)] | Chromosome 5, NC_007116.6 (23582427..23594007) | brp53, drp53, etlD22686.5, fb40d06, p53, wu:fb40d06, zgc:111919 | |
| <input type="checkbox"/> TP53 ID: 403869 | tumor protein p53 [<i>Canis lupus familiaris</i> (dog)] | Chromosome 5, NC_006587.3 (32561406..32565149, complement) | P53 | |
| <input type="checkbox"/> TP53 ID: 397276 | tumor protein p53 [<i>Sus scrofa</i> (pig)] | Chromosome 12, NC_010454.3 (55220545..55234668) | P53 | |
| <input type="checkbox"/> TP53 ID: 281542 | tumor protein p53 [<i>Bos taurus</i> (cattle)] | Chromosome 19, AC_000176.1 (27985492..27997883, complement) | | |
| <input type="checkbox"/> TP53 ID: 396200 | tumor protein p53 [<i>Gallus gallus</i> (chicken)] | | | |
| <input type="checkbox"/> tp53 ID: 397926 | tumor protein p53 [<i>Xenopus laevis</i> (African clawed frog)] | | Trp248, Xp53, Xrel3, p53 | |
| <input type="checkbox"/> TP53 ID: 493847 | tumor protein p53 [<i>Felis catus</i> (domestic cat)] | Chromosome E1, NC_018736.2 (2488175..2500104, complement) | | |
| <input type="checkbox"/> TP53 ID: 100062044 | tumor protein p53 [<i>Equus caballus</i> (horse)] | Chromosome 11, NC_009154.2 (50611931..50616087, complement) | p53 | |
| <input type="checkbox"/> TP53 ID: 100009292 | tumor protein p53 [<i>Oryctolagus cuniculus</i> (rabbit)] | Chromosome 19, NC_013687.1 (11519989..11531435) | | |
| <input type="checkbox"/> Tp53 ID: 100682525 | tumor protein p53 [<i>Cricetulus griseus</i> (Chinese hamster)] | NW_003613643.1 (1638391..1643801) | I79_002739, P53 | |
| <input type="checkbox"/> tp53 ID: 100049321 | tumor protein p53 [<i>Oryzias latipes</i> (Japanese medaka)] | Chromosome 18, NC_019876.1 (15005140..15014025) | p53 | |
| <input type="checkbox"/> TP53 ID: 716170 | tumor protein p53 [<i>Macaca mulatta</i> (Rhesus monkey)] | Chromosome 16, NC_027908.1 (7652077..7670934, complement) | EGK_08142 | |
| <input type="checkbox"/> Tp53 ID: 100379269 | tumor protein p53 [<i>Cavia porcellus</i> (domestic guinea pig)] | | p53 | |

TP53 tumor protein p53 [*Homo sapiens* (human)]

Gene ID: 7157, updated on 17-Feb-2019

Summary

| | |
|--------------------|--|
| Official Symbol | TP53 provided by HGNC |
| Official Full Name | tumor protein p53 provided by HGNC |
| Primary source | HGNC:HGNC:11998 |
| See related | Ensembl:ENSG00000141510 MIM:191170 |
| 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 | P53; BCC7; LFS1; BMFS5; TRP53 |
| Summary | This gene encodes a tumor suppressor protein containing transcriptional activation, DNA binding, and oligomerization domains. The encoded protein responds to diverse cellular stresses to regulate expression of target genes, thereby inducing cell cycle arrest, apoptosis, senescence, DNA repair, or changes in metabolism. Mutations in this gene are associated with a variety of human cancers, including hereditary cancers such as Li-Fraumeni syndrome. Alternative splicing of this gene and the use of alternate promoters result in multiple transcript variants and isoforms. Additional isoforms have also been shown to result from the use of alternate translation initiation codons from identical transcript variants (PMIDs: 12032546, 20937277). [provided by RefSeq, Dec 2016] |
| Expression | Ubiquitous expression in spleen (RPKM 13.2), lymph node (RPKM 13.1) and 25 other tissues See more |
| Orthologs | mouse all |

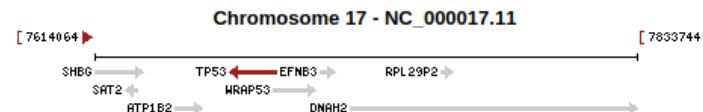
Genomic context

Location: 17p13.1

See TP53 in [Genome Data Viewer](#)

Exon count: 12

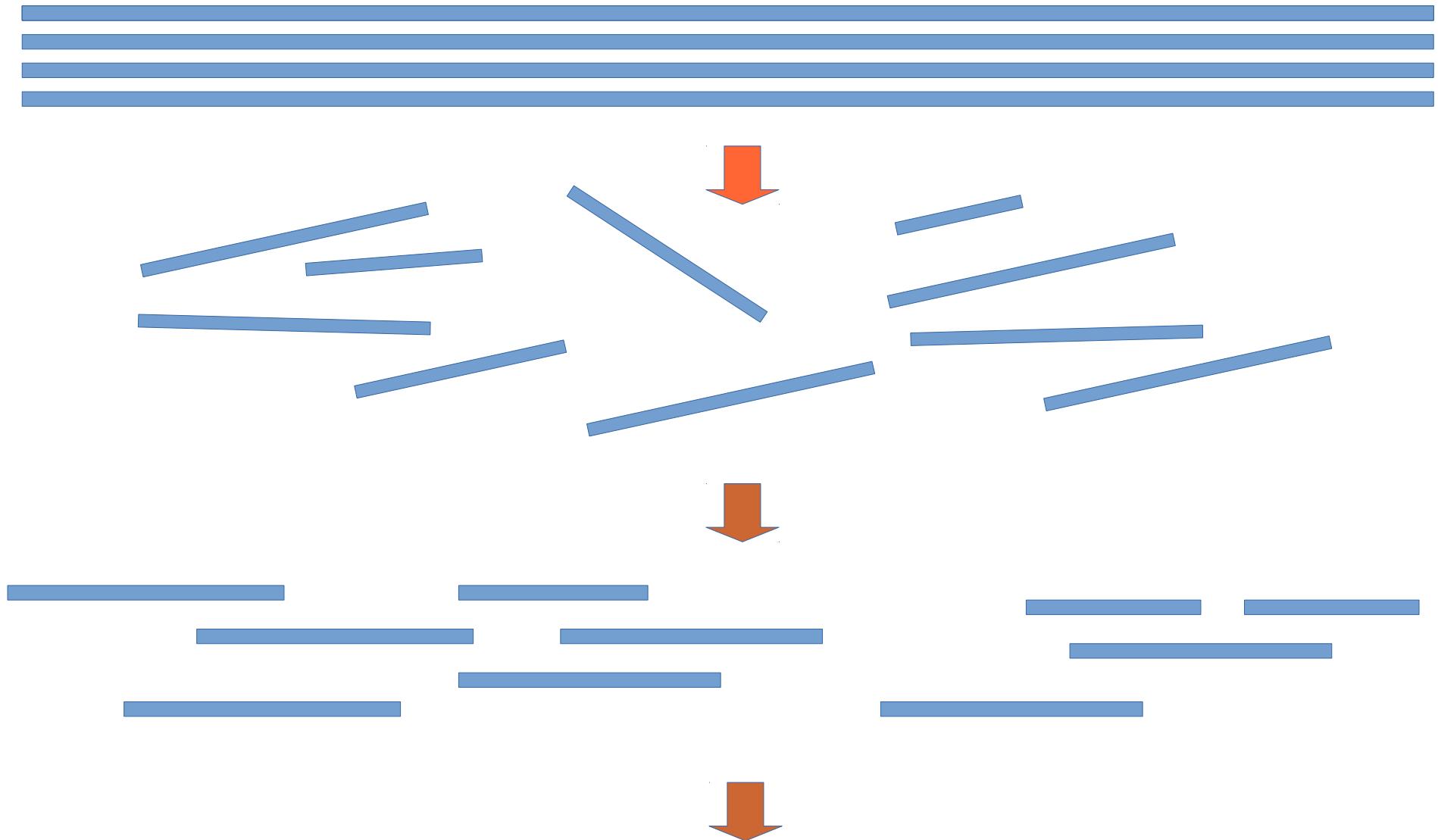
| Annotation release | Status | Assembly | Chr | Location |
|---------------------|-------------------|---|-----|---|
| 109 | current | GRCh38.p12 (GCF_000001405.38) | 17 | NC_000017.11 (7668402..7687550, complement) |
| 105 | previous assembly | GRCh37.p13 (GCF_000001405.25) | 17 | NC_000017.10 (7571720..7590868, complement) |



Reference genome

- A “digital” sequence of nucleic acids, assembled to be a representative sequence for a given species.
- Assembled from DNA sequencing of a set of donors
- GRCh37 → hg19 → 13 donors
- Now at GRCh38 → hg38 → many donors

Genome assembly

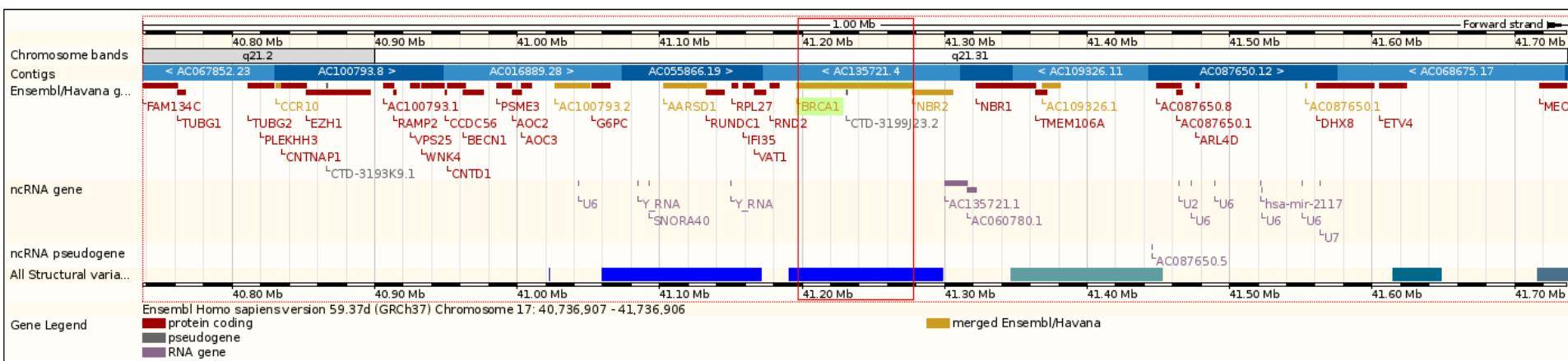


ACTCCGGTTAAATGCGCGCGTCCGCGCTTATTNNATTGCCCGGAAGGTTGGGT

List of biological databases

https://en.wikipedia.org/wiki/List_of_biological_databases

What is a genome browser ?



- A database containing reference sequence assemblies for one or more genomes
- Allows to browse data at various detail levels, from chromosome to gene and down to a single exon/intron
- Let you compare different species and extract data

A quick list

Multi-species genome browsers:

- UCSC (<http://genome.ucsc.edu>)
- ENSEMBL (<http://www.ensembl.org>)
- Entrez (<http://www.ncbi.nlm.nih.gov>)

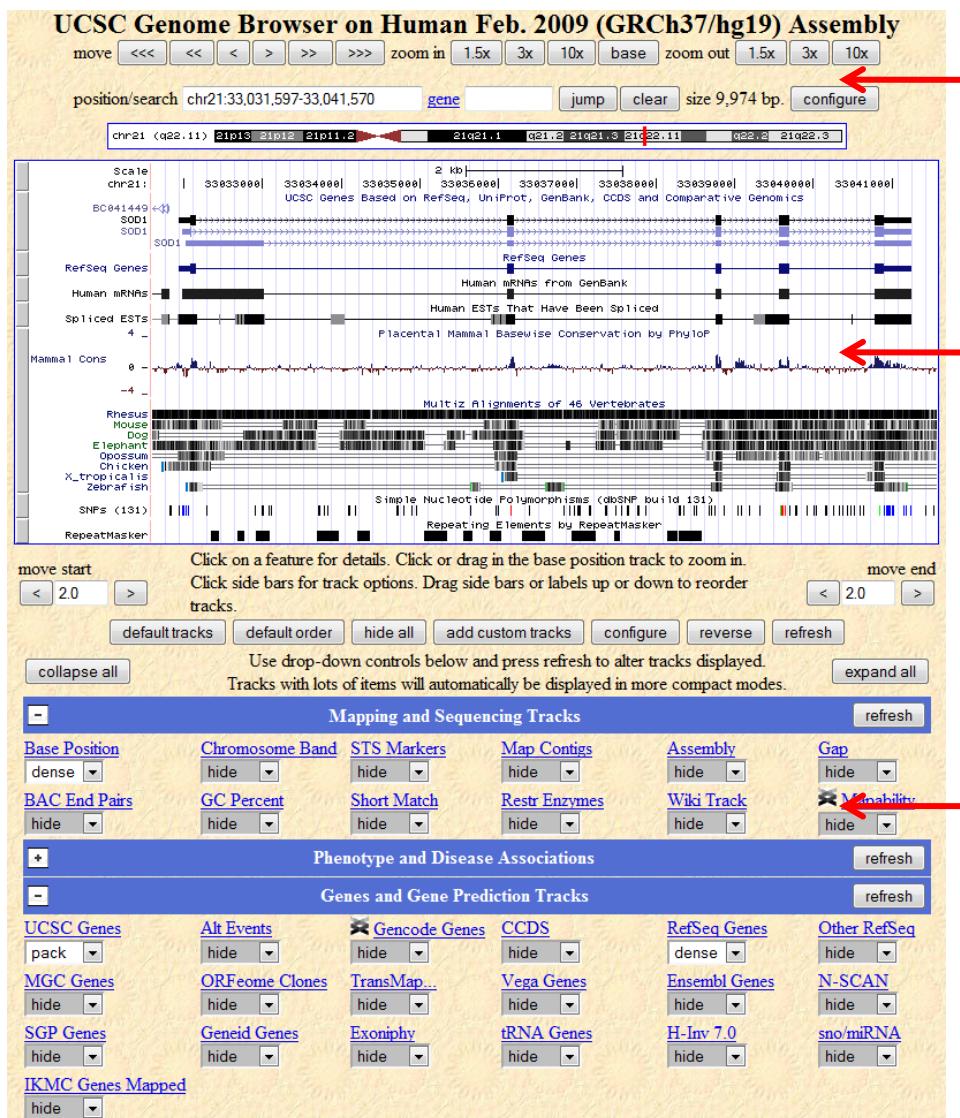
Some organism-specific genome browsers:

- ZFIN – zebrafish (<http://www.zfin.org>)
- SGD – yeast (<http://www.yeastgenome.org>)
- Flybase – fruitfly (<http://www.flybase.org>)
- Wormbase – worm (<http://www.wormbase.org>)

The UCSC Genome browser

- Contains genome of ~100 species, yet does not provide a browser for all of them (*data is accessible, though*)
- Integrates informations such as SNPs, sequence conservation, regulatory elements (ENCODE) and more
- It has a *doppelgänger*: the **Table Browser** allows you to extract data directly from the database tables rather than having to graphically browse it
- Additional tools also allow to align sequences, annotate SNPs, convert data between genome versions, etc.

Browsing a gene

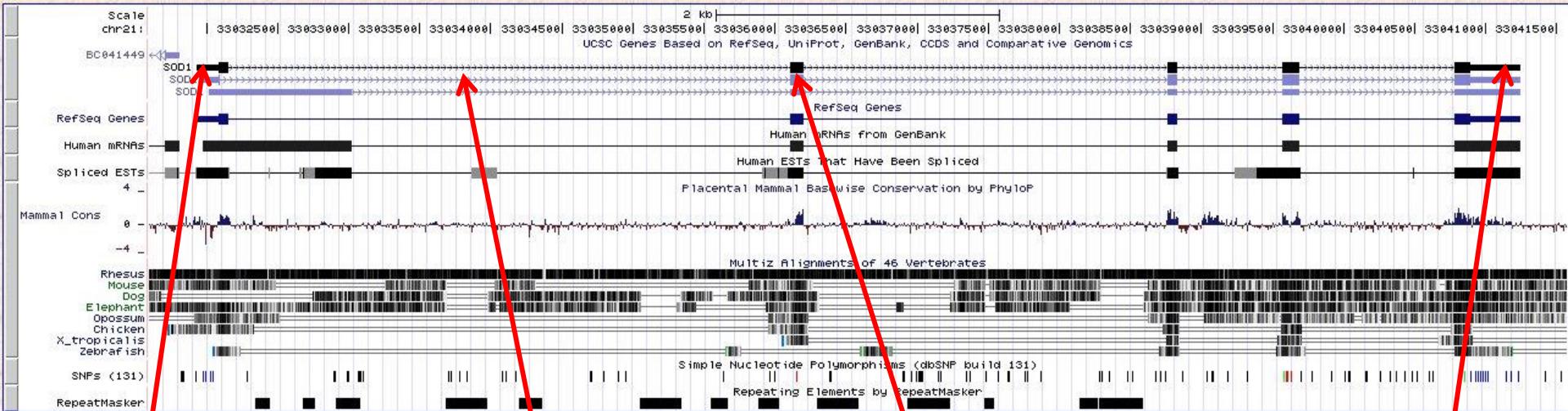


Zoom in/out area

Tracks display box

Selection of which
data
tracks are shown
(and how)

Reading the gene structure

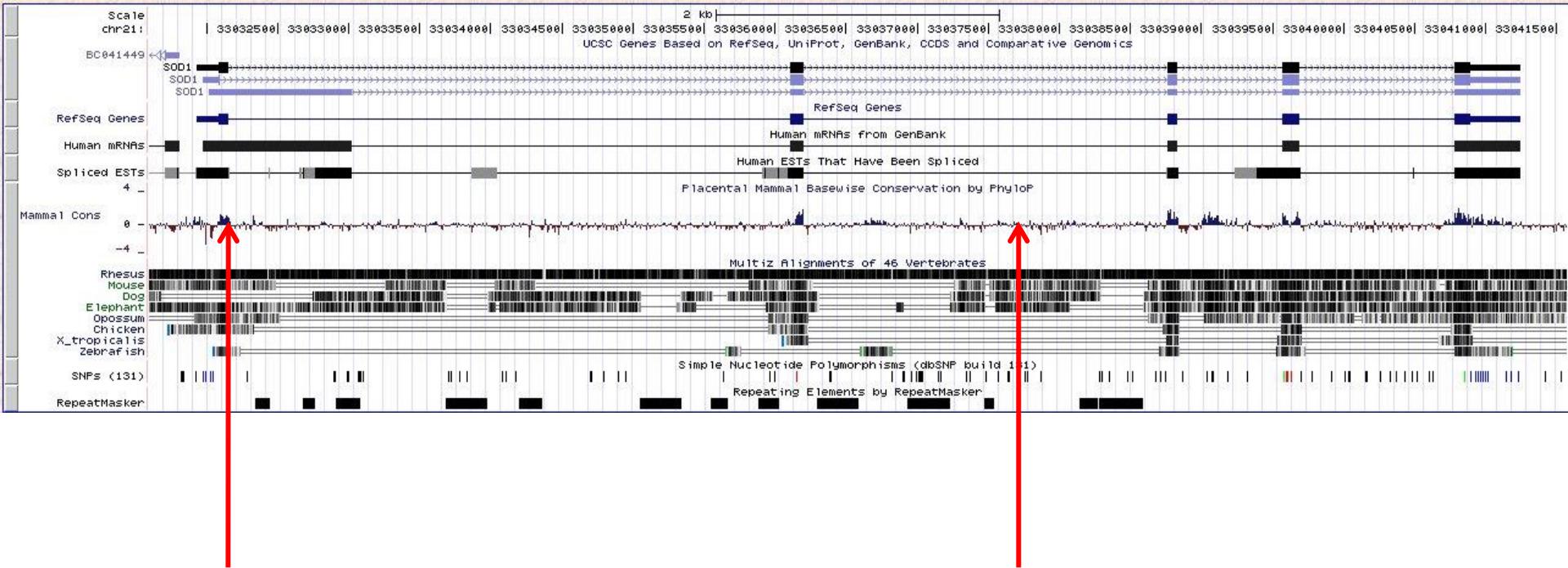


Horizontal line with arrows indicating direction of transcription: **INTRON**

Dark block: **EXON**

Thick and arrowless lines: **UTR**

Reading phylogenetic conservation



High conservation region
(in mammal alignment)

Low conservation region
(in mammal alignment)

Track display modes

- The UCSC browser offers multiple visualization modes for each track:

Hide: *the track is not displayed.*

Dense: *the track is displayed into a single line.*

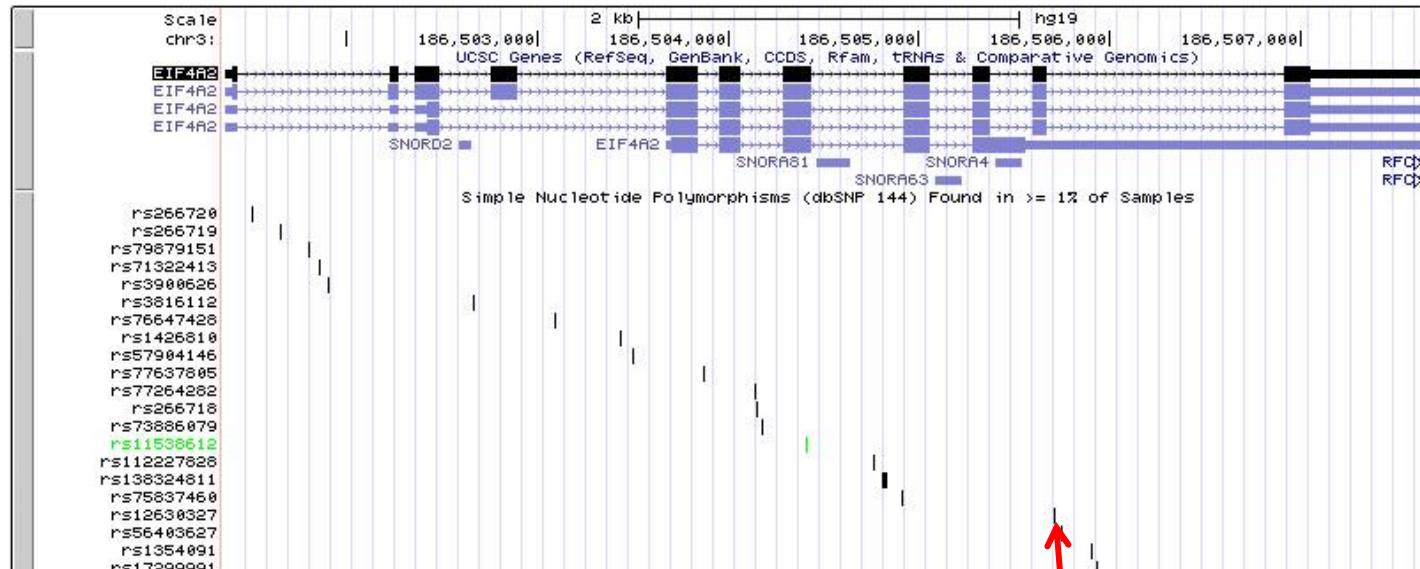
Pack: *each annotation feature is shown separately and labeled, but not always on a separate line.*

Full: *each annotation feature on a separate line.*

- **Hide -> Full:** *more detailed*
- **Full -> Hide:** *more compact*

Reading SNPs

- Change *Common SNP* track display mode to *full*



Identifier of the SNP

Position of the SNP
on the gene

Practical example

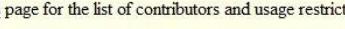
- Display the RARA gene with the browser. How many alternative transcripts does it have ? Considering the first transcript, how many introns and exons?
- Now add to the displayed tracks the GC-percent track (it shows the percentage of GC bases along the sequence). Drag it and move it just under the sequence.
- Now hide the alignment track which is shown by default
- Zoom out on the 5'UTR region of the transcript and check if there are any known SNPs in that region

The Table Browser

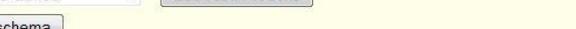
- Very flexible tool that can automatically retrieve data for one or more genes in a variety of formats
- If submitting heavy tasks (e.g. many genes), will redirect you to an online workflow system: Galaxy

Table Browser

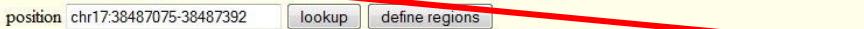
Use this program to retrieve the data associated with a track in text format, to calculate intersections between tracks, and to retrieve DNA sequence covered by a track. For help in using this application see [Using the Table Browser](#) for a description of the controls in this form, the [User's Guide](#)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

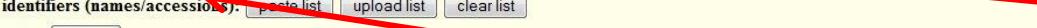
clade: Mammal genome: Human assembly: Feb. 2009 (GRCh37/hg19)  **Select the organism**

group: Genes and Gene Prediction Tracks track: UCSC Genes add custom tracks

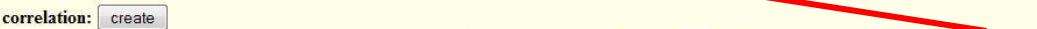
table: knownGene  **Select GROUP, TRACK and TABLE**

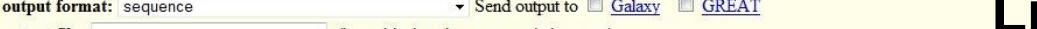
region: genome position chr17:38487075-38487392 lookup define regions

identifiers (names/accessions):  **Limit your search by chromosomal positions or by gene identifiers**

filter:  **create**

intersection:  **create**

correlation:  **create**

output format: sequence  Send output to Galaxy GREAT

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

file type returned: plain text gzip compressed

get output **summary/statistics**

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

Choose the type and the format of desired output data

Practical example

- We want to retrieve the 3'UTR sequence for our gene, RARA:

Paste In Identifiers for UCSC Genes

Please paste in the identifiers you want to include. The items must be values of the **name** field of the **Gene** object.

Some example values:

uc001nez.2
uc009xng.1
uc010lxr.2
NM_032996
CP4B1_HUMAN
uc001lims.1

RARA



Select sequence type for UCSC Genes

UCSC Genes Genomic Sequence

Sequence Retrieval Region Options:

Promoter/Upstream by 1000 bases

5' UTR Exons

CDS Exons

3' UTR Exons

Introns

Downstream by 1000 bas

- One FASTA record per gene.

- One FASTA record per region

■ Split UTR and CDS parts of an exon into separate FASTA records

Note: if a feature is close to the beginning or end of a chromosome and upstream/downstream bases are added, they may be truncated in order to avoid extending

Sequence Formatting Options:

- Exons in upper case, everything else in lower case.

- CDS in upper case, UTR in lower case.

- All upper case.
- All 1

All lower case.

Mask repeats: to lower case to N

get sequence **cancel**



The ENSEMBL genome browser



- European genome browser, focuses on vertebrate genomes: contains a lot of them!
- Includes genomic variants (somatic and structural) and regulatory elements data
- Like the UCSC genome browser, *it offers an interface to access data directly, BioMart*

Ensembl Species

| | |
|---|---|
|  Alpaca <i>Vicugna pacos</i> |  Guinea Pig <i>Cavia porcellus</i> |
|  Anole Lizard <i>Anolis carolinensis</i> |  Hedgehog <i>Erinaceus europaeus</i> |
|  Armadillo <i>Dasypus novemcinctus</i> |  Horse <i>Equus caballus</i> |
|  Bushbaby <i>Otolemur garnettii</i> |  Human <i>Homo sapiens</i> |
|  Caenorhabditis elegans |  Hyrax <i>Procavia capensis</i> |
|  Ciona intestinalis |  Kangaroo rat <i>Dipodomys ordii</i> |
|  Ciona savignyi |  Lamprey (preview - assembly only) <i>Petromyzon marinus</i> |
|  Cat <i>Felis catus</i> |  Lesser hedgehog tenrec <i>Echinops telfairi</i> |
|  Chicken <i>Gallus gallus</i> |  Macaque <i>Macaca mulatta</i> |
|  Chimpanzee <i>Pan troglodytes</i> |  Marmoset <i>Callithrix jacchus</i> |
|  Cow <i>Bos taurus</i> |  Medaka <i>Oryzias latipes</i> |
|  Dog <i>Canis familiaris</i> |  Megabat <i>Pteropus vampyrus</i> |
|  Dolphin <i>Tursiops truncatus</i> |  Microbat <i>Myotis lucifugus</i> |
|  Drosophila melanogaster |  Mouse <i>Mus musculus</i> |
|  Elephant <i>Loxodonta africana</i> |  Mouse Lemur <i>Microcebus murinus</i> |
| |  Pig <i>Sus scrofa</i> |
| |  Pika <i>Ochotona princeps</i> |
| |  Platypus <i>Ornithorhynchus anatinus</i> |
| |  Rabbit <i>Oryctolagus cuniculus</i> |
| |  Rat <i>Rattus norvegicus</i> |
| |  Saccharomyces cerevisiae |
| |  Shrew <i>Sorex araneus</i> |
| |  Sloth <i>Choloepus hoffmanni</i> |
| |  Squirrel <i>Spermophilus tridecemlineatus</i> |
| |  Stickleback <i>Gasterosteus aculeatus</i> |
| |  Tarsier <i>Tarsius syrichta</i> |
| |  Tetraodon <i>Tetraodon nigroviridis</i> |
| |  Tree Shrew <i>Tupaia belangeri</i> |
| |  Wallaby <i>Macropus eugenii</i> |
| |  Xenopus tropicalis |

e!



What Distinguishes Ensembl from the UCSC and NCBI Browsers?

- The gene set. Automatic annotation based on mRNA and protein information.
- Programmatic access via the Perl API (open source)
- BioMart
- Integration with other databases (DAS)

e!



Ensembl Genes – biological basis

All Ensembl transcripts are based on proteins and mRNAs in:

- UniProt/Swiss-Prot (*manually curated*)
- UniProt/TrEMBL

www.uniprot.org

- NCBI RefSeq (*manually curated*)

www.ncbi.nlm.nih.gov/RefSeq

e!

How is this information organised?

- Ensembl Views (Website)



- Ensembl Database (open source)
- BioMart ‘DataMining tool’

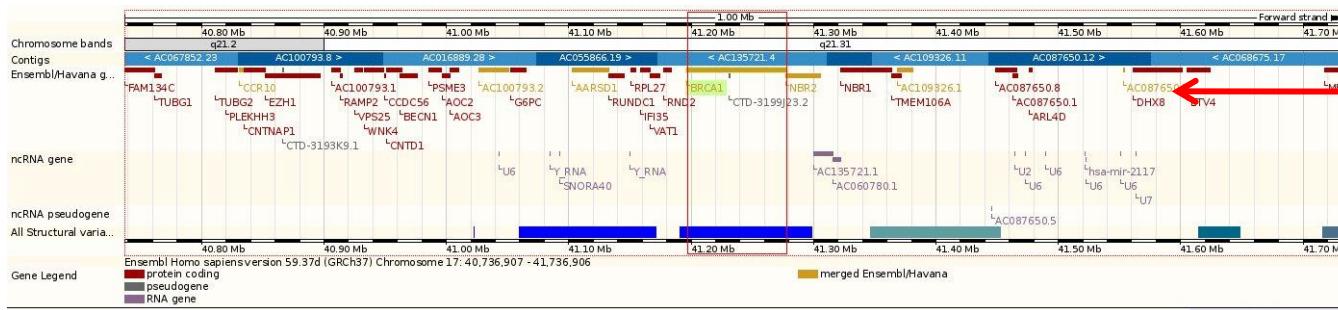
e!



Names in Ensembl

- **ENSG###** Ensembl Gene ID
 - **ENST###** Ensembl Transcript ID
 - **ENSP###** Ensembl Peptide ID
 - **ENSE###** Ensembl Exon ID
-
- For other species than human a suffix is added:
MUS (*Mus musculus*) for mouse: **ENSMUSG###**
DAR (*Danio rerio*) for zebrafish: **ENSDARG###**, etc.

Browsing a gene



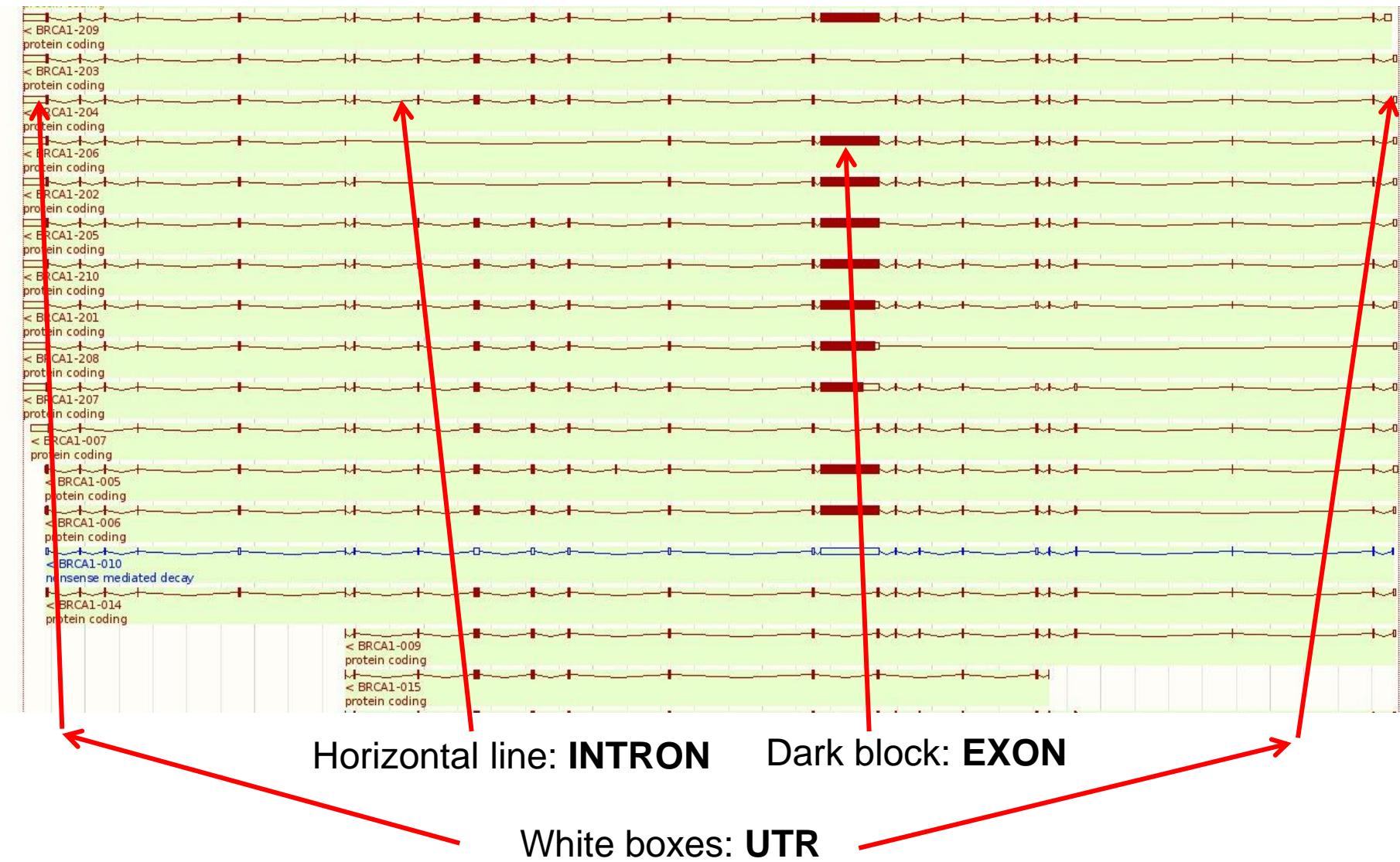
Region summary



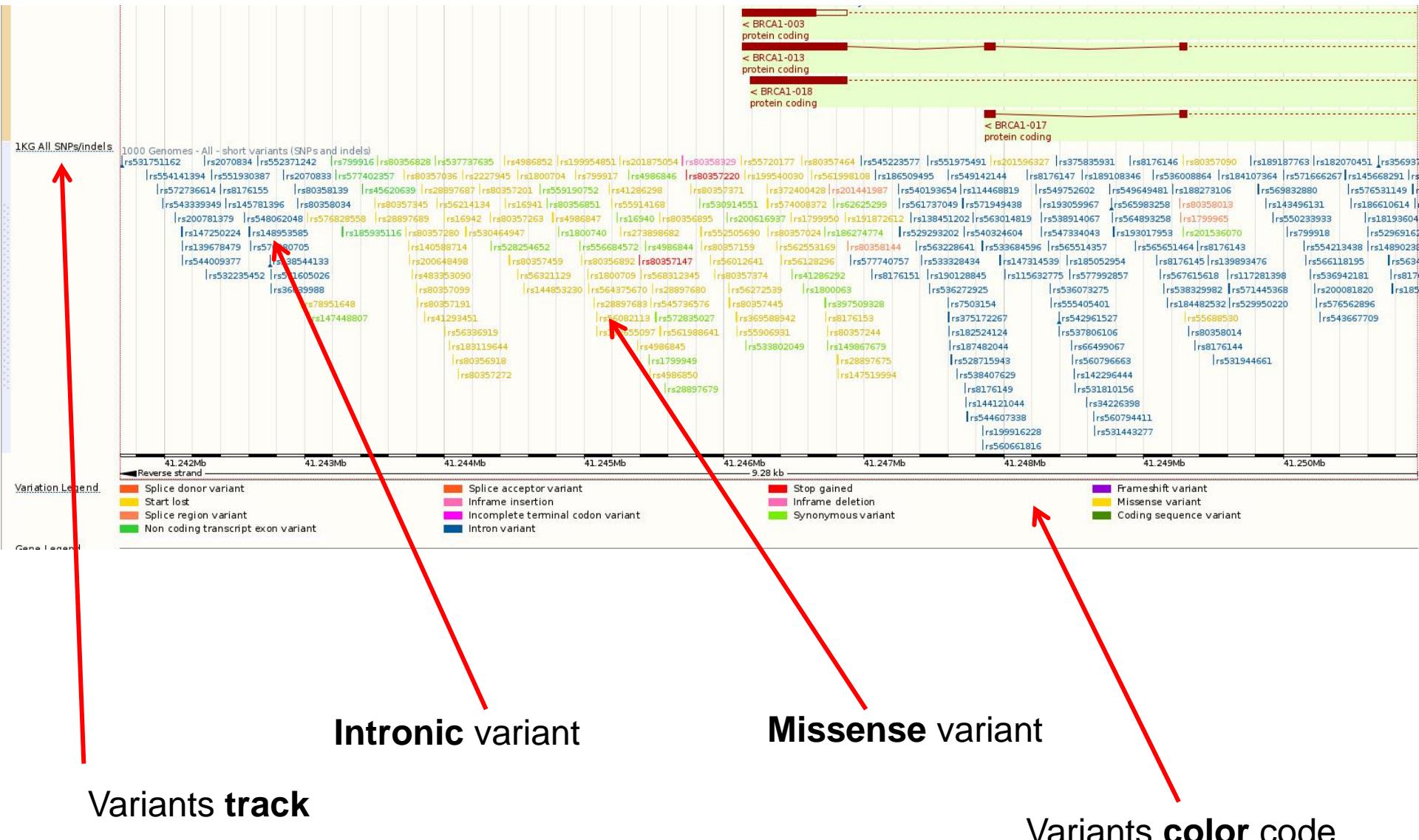
Zoom / position controls

Sequence and tracks

Reading the gene structure



Reading somatic variants



Practical example

- Display the KRAS gene with the browser. How many alternative transcripts does it have ? Considering the first transcript, how many introns and exons?
- Now add the 1000 Genomes – EUR common variants track to the display. Are there 5'UTR variants in KRAS?
- Now zoom to the second exon of the first isoform of KRAS. How many missense variants are there ?

BioMart

- A data mining platform allowing to perform complex query on the Ensembl database
- Similar to the UCSC Table Browser, but much more powerful
- Can retrieve both annotation and sequences with no programming required!
- E.g. retrieve all genes contained in a specific region of the genome
- E.g. retrieve all sort of information on a specified list of genes

BioMart

Organism/dataset selection

The screenshot shows the BioMart interface. At the top, there are navigation buttons: New, Count, and Results. On the right, there are links for URL, XML, Perl, and Help. Below these are sections for 'Dataset' and 'Attributes'. The 'Dataset' section shows 'Homo sapiens genes (GRCh38.p5)' and a 'Filters' button. The 'Attributes' section lists various Ensembl and transcript-related attributes with checkboxes. A large red arrow points from the 'Organism/dataset selection' text to the 'Dataset' section. Another red arrow points from the 'Select subsets of data to retrieve' text to the 'Attributes' section.

New Count Results

Dataset

Homo sapiens genes (GRCh38.p5)

Filters

[None selected]

Attributes

Ensembl Gene ID
Ensembl Transcript ID

Dataset

[None Selected]

Please select columns to be included in the output and hit 'Results' when ready

Features Variant (Germline)
Structures Variant (Somatic)
Homologs Sequences

GENE:

Ensembl

Ensembl Gene ID
 Ensembl Transcript ID
 Ensembl Protein ID
 Ensembl Exon ID
 Description
 Chromosome Name
 Gene Start (bp)
 Gene End (bp)
 Strand
 Band
 Transcript Start (bp)
 Transcript End (bp)
 Transcription Start Site (TSS)
 Transcript length (including UTRs and CDS)
 Transcript Support Level (TSI)

APPRI annotation
Associated Gene Name
Associated Gene Source
Associated Transcript Name
Associated Transcript Source
Transcript count
% GC content
Gene type
Transcript type
Source (gene)
Source (transcript)
Status (gene)
Status (transcript)
Version (gene)
Version (transcript)

Choose **attributes** that must be retrieved

Select **subsets** of data to retrieve

Available attributes
(multiple categories available)

BioMart

Number of genes in output

Dataset 37 / 66232 Genes

Homo sapiens genes (GRCh38.p5)

Filters

Chromosome: MT

Attributes

Ensembl Gene ID
Chromosome Name
Strand
HGNC symbol

Dataset [None Selected]

Export all results to File TSV Unique results only Go

Email notification to

View 10 rows as HTML Unique results only

| Ensembl Gene ID | Chromosome Name | Strand | HGNC symbol |
|-----------------|-----------------|--------|-------------|
| ENSG00000210049 | MT | 1 | MT-TF |
| ENSG00000211459 | MT | 1 | MT-RNR1 |
| ENSG00000210077 | MT | 1 | MT-TV |
| ENSG00000210082 | MT | 1 | MT-RNR2 |
| ENSG00000209082 | MT | 1 | MT-TL1 |
| ENSG00000198888 | MT | 1 | MT-ND1 |
| ENSG00000210100 | MT | 1 | MT-TI |
| ENSG00000210107 | MT | -1 | MT-TQ |
| ENSG00000210112 | MT | 1 | MT-TM |
| ENSG00000198763 | MT | 1 | MT-ND2 |

Recap of selected filters and attributes

Results export options

Actual results of the query

Practical example

- Get the Ensembl ID and MGI symbol for all mouse genes on chromosome 19. How many genes are there ?
- Now limit this to protein coding genes. How many genes are there ?
- Now save the results to a *comma-separated values* file