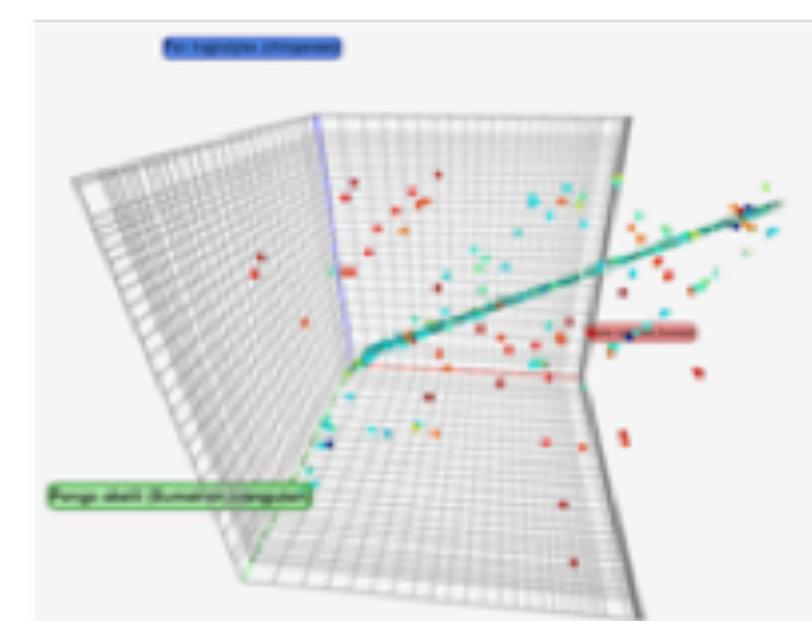
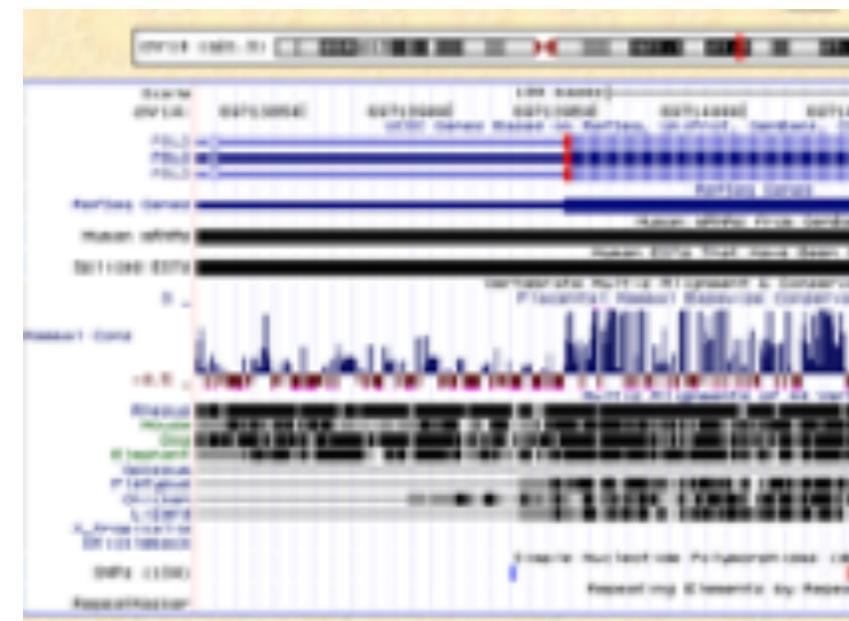


Computational Genomics

Introduction to Genome Browsers

ENSEMBL



Training materials



- Ensembl training materials are protected by a CC BY license:
creativecommons.org/licenses/by/4.0/
- If you wish to re-use these materials, please credit Ensembl for their creation
- If you use Ensembl for your work, please cite our papers:
ensembl.org/info/about/publications.html

Exploring the Ensembl genome browser

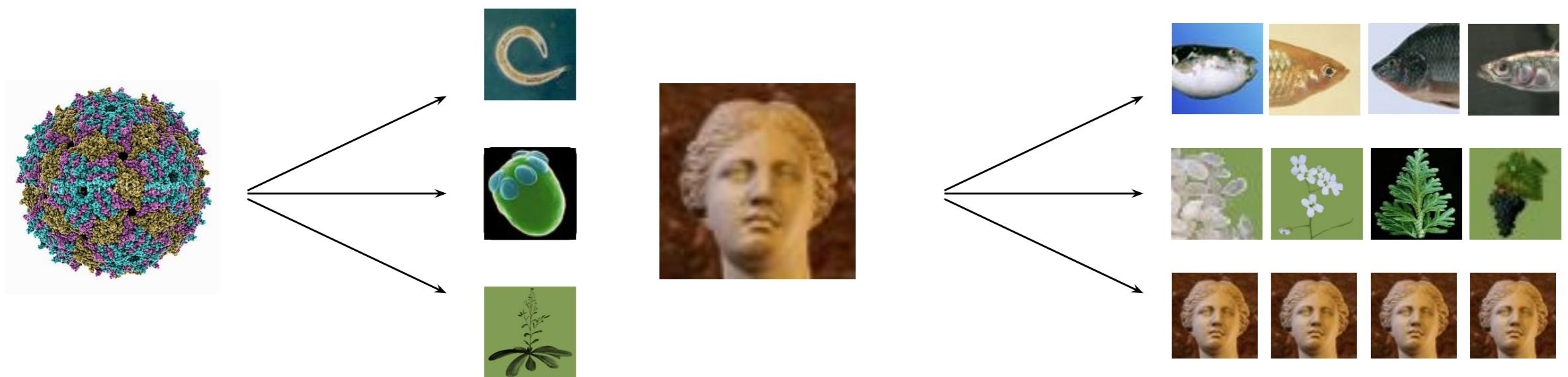
The screenshot shows the Ensembl homepage. At the top, there's a search bar with placeholder text "Search: All species for" and a "Go" button. Below the search bar, there's a section titled "Browse a Genome" with links to "Human (GRCh37)", "Mouse (GRCm38)", and "Zebrafish (Zv9)". There's also a dropdown menu for "All genomes". To the right of this, there are several boxes: "ENCODE data in Ensembl" (with a DNA helix icon), "Variant Effect Predictor (VeP)" (with a blue and red "VeP" logo), "Gene expression in different tissues" (with a tissue sample image), "Find SNPs and other variants for my gene" (with a DNA sequence diagram), "Retrieve gene sequence" (with a DNA sequence snippet), "Compare genes across species" (with a brain and eye icon), "Use my own data in Ensembl" (with a bar chart icon), and "Learn about a disease or phenotype" (with an eye icon). On the far right, there's a sidebar for "What's New in Release 73 (September 2013)" listing updates like "Updated patches for the human assembly (GRCh37.p11)", "New variation citation page and individual genotype search box", and "Upload VEP output to the ensembl website". It also includes links to "Full details of this release", "More release news on our blog →", "Latest blog posts" (listing three recent posts), and a "Did you know...?" section with a "FAQs" link.

Introduction

Why do we need genome browsers?

1977: 1st genome to be sequenced (5 kb)

2004: finished human sequence (3 Gb)



CGGCCTTGGGCTCCGCCTCAGCTCAAGACTTAACCTCCCTCCCAGCTGTCCCAGATGACGCCATCTGAAATTCTGGAA
ACACGATCACTTAAACGGAATATTGCTTTGGGGAAAGTGTTCAGCTGCTGGCACGCTGTATTGCCTACTTAAGC
CCCTGGTAATTGCTGTATTCCGAAGACATGCTGATGGGAATTACCAGGCAGCGTTGGTCTCTAACTGGAGCCCTCTGTCCCC
ACTAGCCACCGTCAGGGTAGCGTGATTGAAACTAAATCGTATGAAAATCCTCTCTAGTCGCACTAGCCACGTTCG
AGTGCTTAATGTGGCTAGTGGCACCGGTTGGACAGCACAGCTGTAAAATGTTCCCATCCTCACAGTAAGCTGTTACCGTT
CAGGAGATGGGACTGAATTAGAACAAACATTCCAGCGCTCTGAGTTTACCTCAGTCACATAATAAGGAATGCAT
CCCTGTGTAAGTGCATTGGTCTTGCAGACTTACCAAGCATTGGAGGAATATCGTAGGTAAAAATGCCTA
TTGGATCCAAGAGAGGCCAACATTGGAAATTAAAGACACGCTGCAACAAAGCAGGTATTGACAATTATATAAC
TTTATAAATTACACCGAGAAAGTGTCTAAAAAATGCTGCTAAAAACCCAGTACGTACAGTGTGCTTAGAACCAA
ACTGTCCTATGTGTATAAACATCCAGTTAACACATAATCATCGTTGCAGGTTAACCATGATAATAGAACGTCT
AGTGGATAAAGAGGAAACTGGCCCTTGACTAGCAGTAGGAACAATTACTAACAAATCAGAACGATTAATGTTACTTATGG
CAGAAGTTGTCACCTGGTTTCAGTACTCCTTACTCTTAAAGATCTAGGACCCCCGGAGTGCTTGTGTTATG
TAGCTTACCATATTAGAAATTAAAGAATTAAAGGCTGGCGTGGCTCACGCCTGTAATCCCAGCACTTGGGA
GGCGAGGTGGCGGATCACTTGAGGCCAGAAGTTGAGACCAGCCTGCCAACATGGTAAACACCCTACTCTACTAAAAAAT
ACAAAAAAATGTGCTGCGTGGTGCCTGTAATCCCAGCTACACGGGAGGTGGAGGCAGGAGAACGCTTGAACCC
TGGAGGCAGAGGTTGCAGTGAGCCAAGATCATGCCACTGCACTCTAGCCTGGCCACATAGCATGACTCTGTCTAAAACAA
ACAAACAAACAAAAACTAAGAATTAAAGTTAAACTTAAAGGTTAAAGGCTGGCGTGGCTCACGCCTGTAATCCCAGCACTTGGGA
TCTTAGGAAAAATAACTTTGAAAACAAGTGAGTGGAAATAGTTTACATTTCAGTTCTCTTTAATGTCTGGCTAAAT
AGAGATAGCTGGATTCACTTATCTGTCTAATCTGTTATTGGTAGAAGTATGTGAAAAAAATTAACCTCACGTTGAAA
AAAGGAATATTAAATAGTTTCAGTTACTTTGGTATTTCCTGTTACTTCAGTTCTCTTTAATGAAAGCTAACCCATTGCATATTATCACAA
ATACCATAGGTCTTCCATGTCGCAACATCATGCAGTGATTATTGGAAGATAGGGTGTCTGAATTACAAAGATTCC
AAATATTGATAATTGCATTAAACTATTAAAAATCTCATTCAATTACCAACATGGATGTCAGAAAGTCTTTAAGAT
TGGGTAGAAATGCCACTGGAAATTCTAATTTCAATTGAAAGTTCACATTGTCATTGACAACAAACTGTTCTTGC
AGCAACAAAGATCACTCATTGATTGTGAGAAAATGTCTACCAAATTATTAAAGTGAATAACTTGTCAAGCTGTTCTTC
AAGTAAAAATGACTTTCAATTGAAAAATTGCTTGTTCAGATCACAGCTAACATGAGTGCTTTCTAGGCAGTATTGACT
TCAGTATGCAGAAGTGCTTATGTATGCTTCTATTGTCAGAGATTATTAAAAGAAGTGCTAACAGCATTGAGCTCGAAA
TTAATTAACTGCTTCATTAGGACATTACATTAAACTGGCATTATTACTATTATTAAACAGGACACTCAGTG
GTAAGGAATATAATGGCTACTAGTATTAGTTGGGCCACTGCCATAACTCATGCAAATGTGCCAGCAGTTTACCCAGCAT
CATCTTGCAGTGTGATAACAAATGTCAACATCATGAAAAAGGGTTGAAAAAAAGGAATATTAAATAGTTCACTT

What is Ensembl?

```
AGTGCTTAATGTGGCTAGTGGCACCGGTTGGACAGCACAGCTGTAA  
AATGTTCCCATCCTCACAGTAAGCTGTTACCGTTCCAGGAGATGGGA  
CTGAATTAGAACAAACAAATTTCAGCGCTTCTGAGTTTACCT  
CAGTCACATAATAAGGAATGCATCCCTGTGTAAGTGCATTTGGTCT  
TCTGTTTGCAGACTTATTACCAAGCATTGGAGGAATATCGTAGGT  
AAAAATGCCTATTGGATCCAAGAGAGGCCAACATTTTGAAATT  
TTAAGACACGCTGCAACAAAGCAGGTATTGACAAATTTATATAACT  
TTATAAAATTACACCGAGAAAGTGTCTAAAAATGCTGCTAAAA  
ACCCAGTACGTACAGTGTGCTTAGAACCATAAACTGTTCTTATG  
TGTGTATAAAATCCAGTTAACACATAATCATCGTTGCAGGTTAAC  
ACATGATAAAATATAGAACGTCTACTGGATAAGAGGAAACTGGCCCC  
TTGACTAGCAGTAGGAACAATTACAAACAAATCAGAACATTAATGT
```



Ensembl annotates and maps genomic features from genome sequences

What is Ensembl?

```
AGTGCTTAATGTGGCTAGTGGCACCGGTTGGACAGCACAGCTGTAA  
AATGTTCCCATCCTCACAGTAAGCTGTTACCGTTCCAGGAGATGGGA  
CTGAATTAGAACAAACAAATTTCAGCGCTTCTGAGTTTACCT  
CAGTCACATAATAAGGAATGCATCCCTGTGTAAGTGCATTTGGTCT  
TCTGTTTGCAGACTTATTACCAAGCATTGGAGGAATATCGTAGGT  
AAAAATGCCTATTGGATCCAAGAGAGGCCAACATTTTGAAATT  
TTAAGACACGCTGCAACAAAGCAGGTATTGACAAATTATATAACT  
TTATAAAATTACACCGAGAAAGTGTCTAAAAATGCTGCTAAAA  
ACCCAGTACGTACAGTGTGCTTAGAACCATAAACTGTTCTTATG  
TGTGTATAAAATCCAGTTAACACATAATCATCGTTGCAGGTTAAC  
ACATGATAAAATATAGAACGTCTACTGGATAAGAGGAAACTGGCCCC  
TTGACTAGCAGTAGGAACAATTACAAACAAATCAGAACATTATGT
```



Ensembl is an ‘added value resource’ bringing together information from a wide range of other databases in a single site

What is Ensembl?

```
AGTGCTTAATGTGGCTAGTGGCACCGGTTGGACAGCACAGCTGTAA  
AATGTTCCCATCCTCACAGTAAGCTGTTACCGTTCCAGGAGATGGGA  
CTGAATTAGAACAAACAAATTTCAGCGCTTCTGAGTTTACCT  
CAGTCACATAATAAGGAATGCATCCCTGTGTAAGTGCATTTGGTCT  
TCTGTTTGCAGACTTATTACCAAGCATTGGAGGAATATCGTAGGT  
AAAAATGCCTATTGGATCCAAGAGAGGCCAACATTTTGAAATT  
TTAAGACACGCTGCAACAAAGCAGGTATTGACAAATTTATATAACT  
TTATAAAATTACACCGAGAAAGTGTCTAAAAATGCTGCTAAAA  
ACCCAGTACGTACAGTGTGCTTAGAACCATAAACTGTTCTTATG  
TGTGTATAAAATCCAGTTAACACATAATCATCGTTGCAGGTTAAC  
ACATGATAAAATATAGAACGTCTACTGGATAAGAGGAAACTGGCCCC  
TTGACTAGCAGTAGGAACAATTACAAACAAATCAGAACATTAATGT
```



NIH U.S. National Library of Medicine NCBI

Genome Data Viewer

[www.ncbi.nlm.nih.gov/
genome/gdv/](http://www.ncbi.nlm.nih.gov/genome/gdv/)



www.ensembl.org



www.genome.ucsc.edu

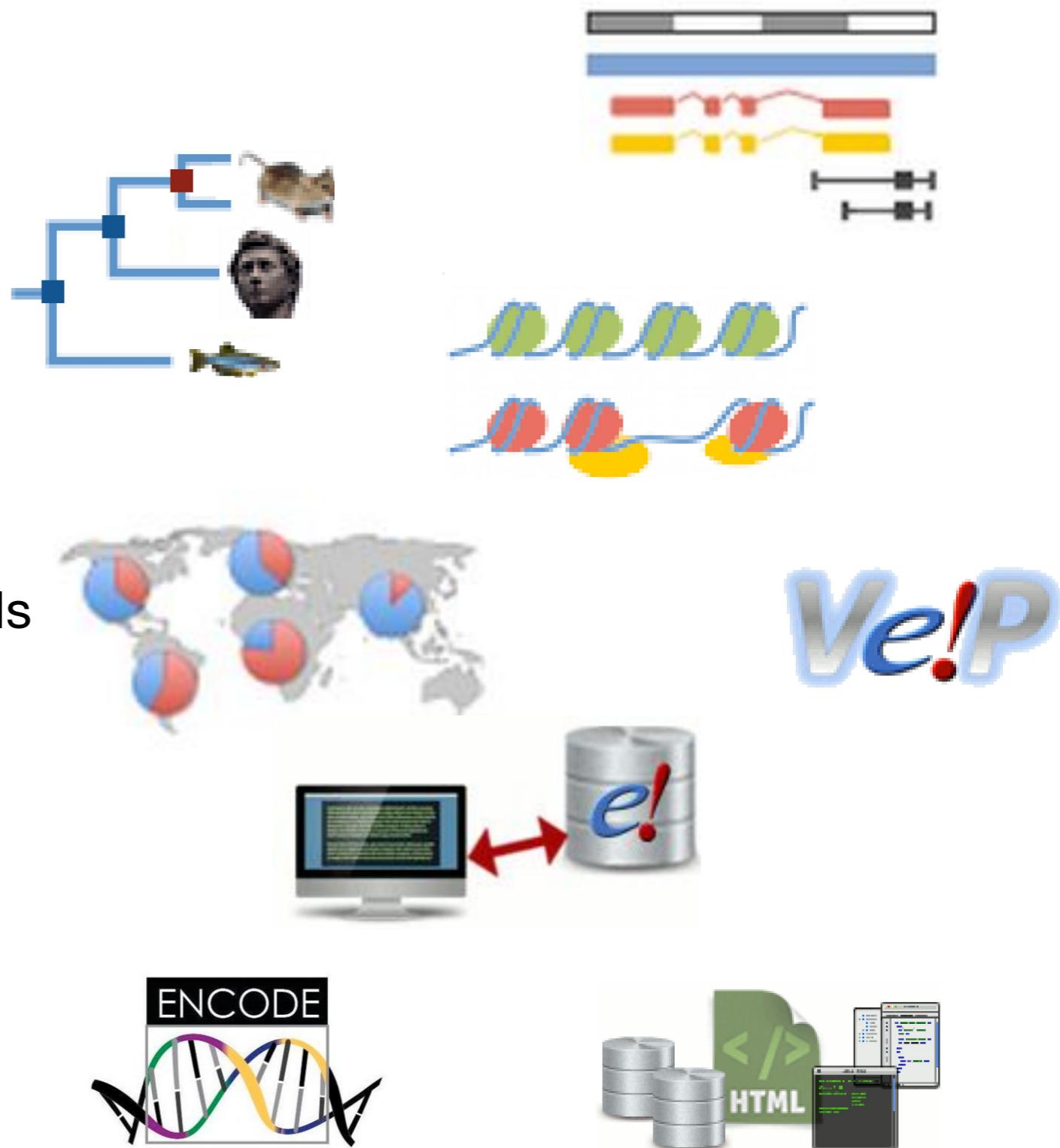


www.ensemblgenomes.org

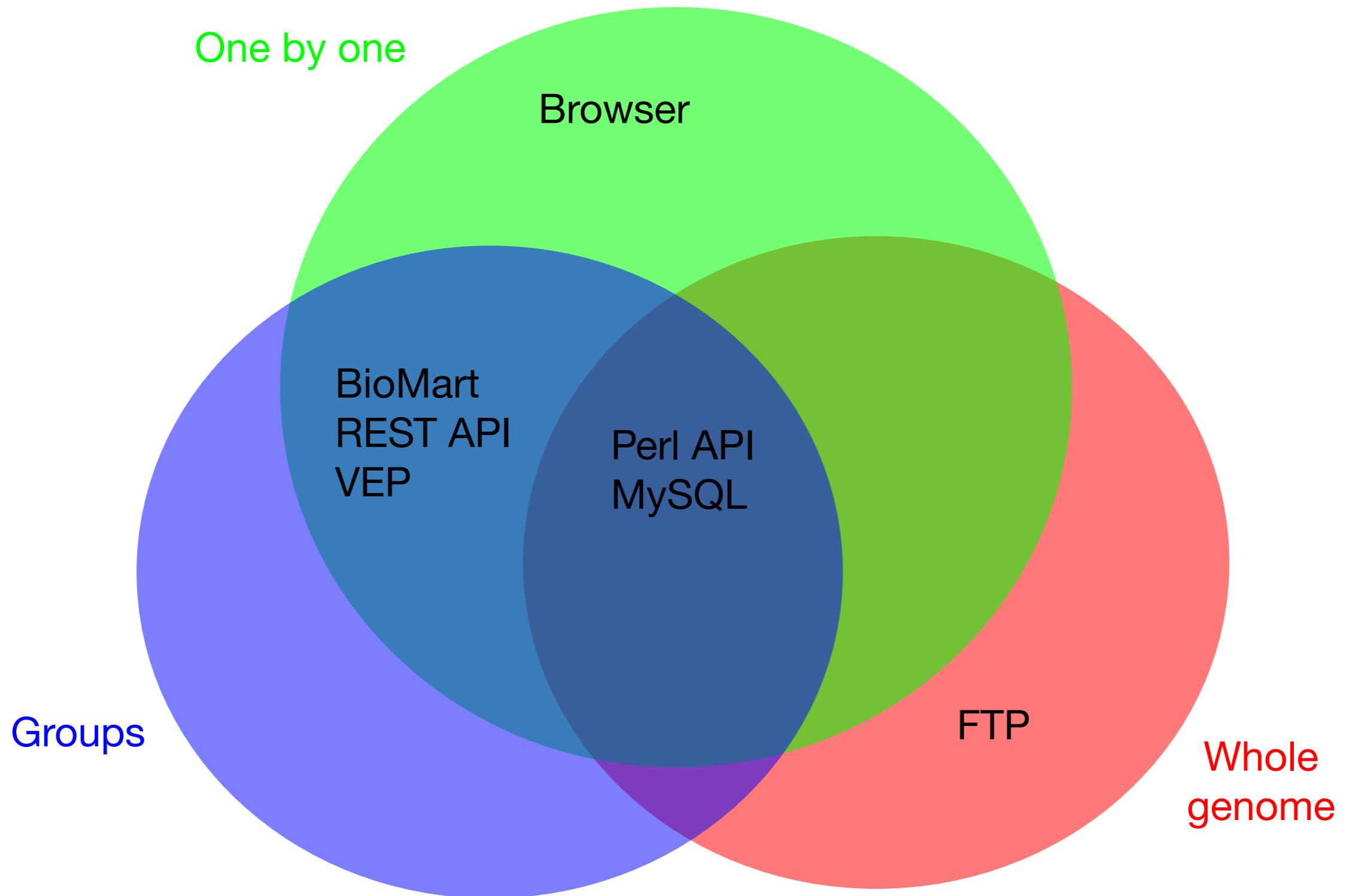
q

Ensembl features

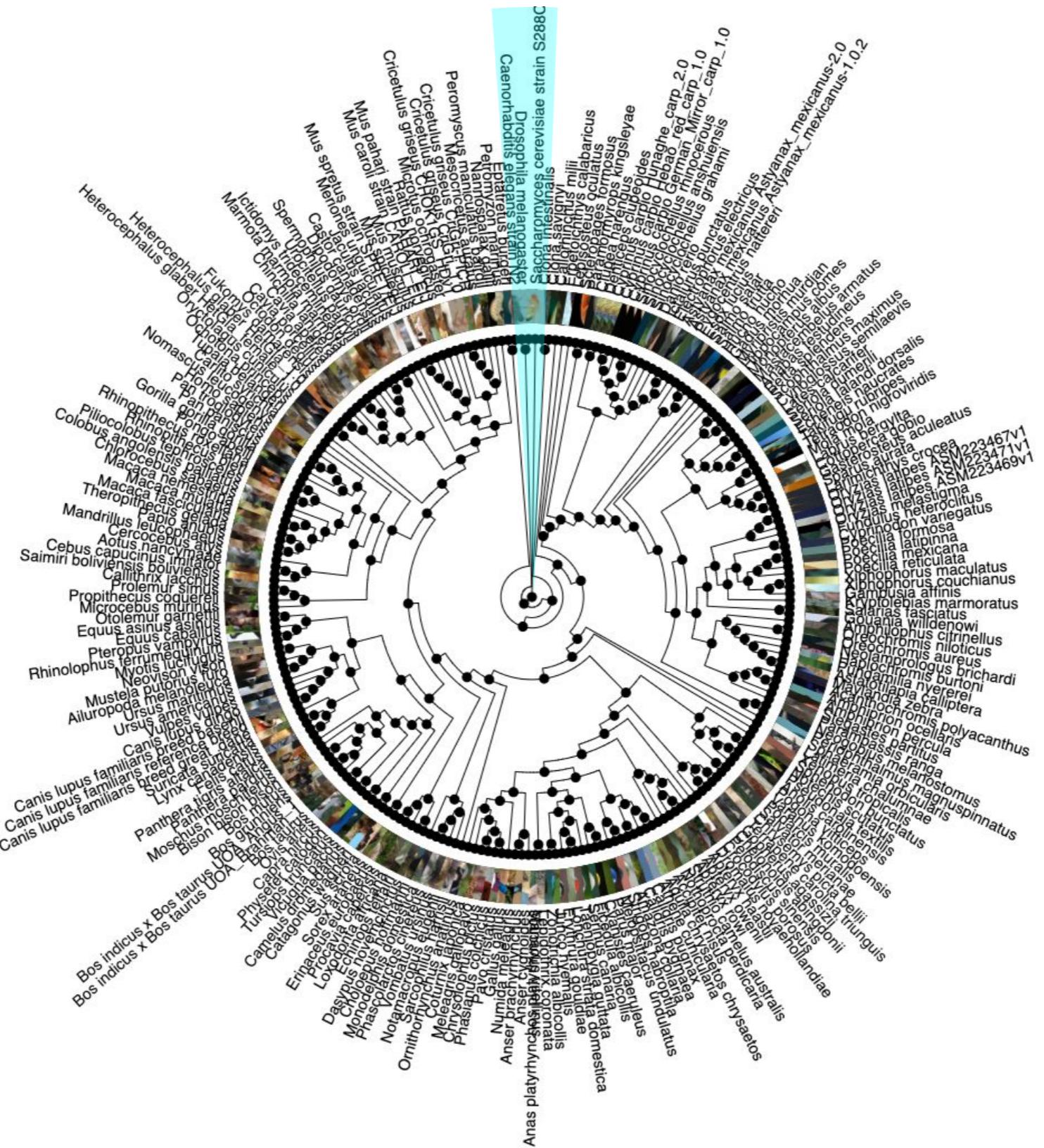
- Gene builds for >300 species
- Gene trees
- Regulatory build (ENCODE)
- Variation display and VEP
- Display of user data
- BioMart (data export)
- Programmatic access via the APIs
- Completely Open Source



Access scales



Vertebrate species in Ensembl



Non-vertebrates on Ensembl genomes

[e! EnsemblBacteria](#) ▾ BLAST | More ▾ [Search Ensembl Bacteria species...](#)

Help & Documentation Species List

Find a Species

Ensembl Bacteria Species

Bacillus collection 78 genomes

Bacillus amyloliquefaciens European Nucleotide Archive	Bacillus anthracis A0248 European Nucleotide Archive	Bacillus anthracis Ames European Nucleotide Archive
Bacillus anthracis Ames ancestor European Nucleotide Archive	Bacillus anthracis CDC 684 European Nucleotide Archive	Bacillus anthracis Sterne European Nucleotide Archive
Bacillus cereus 03BB102 European Nucleotide Archive	Bacillus cereus 172560W European Nucleotide Archive	Bacillus cereus 95/8201 European Nucleotide Archive
Bacillus cereus AH1271 European Nucleotide Archive	Bacillus cereus AH1272 European Nucleotide Archive	Bacillus cereus AH1273 European Nucleotide Archive
Bacillus cereus AH187 European Nucleotide Archive	Bacillus cereus AH603 European Nucleotide Archive	Bacillus cereus AH621 European Nucleotide Archive
Bacillus cereus AH676 European Nucleotide Archive	Bacillus cereus AH820 European Nucleotide Archive	Bacillus cereus ATCC 10876 European Nucleotide Archive
Bacillus cereus ATCC 10987 European Nucleotide Archive	Bacillus cereus ATCC 14579 European Nucleotide Archive	Bacillus cereus ATCC 4342 European Nucleotide Archive
Bacillus cereus B4264 European Nucleotide Archive	Bacillus cereus BDRD-Bcer4 European Nucleotide Archive	Bacillus cereus BDRD-ST196 European Nucleotide Archive
Bacillus cereus BDRD-ST24 European Nucleotide Archive	Bacillus cereus BDRD-ST26 European Nucleotide Archive	Bacillus cereus BG6E1 European Nucleotide Archive
Bacillus cereus F65185	Bacillus cereus G9842	Bacillus cereus MM3

Bacteria

[e! EnsemblProtists](#) ▾ BLAST | More ▾ [Search Ensembl Protists species...](#)

Help & Documentation Species List

Find a Species

Ensembl Protists Species

Alveolata

Plasmodium berghei GeneDB Plasmodium berghei ANKA	Plasmodium knowlesi Wellcome Trust Sanger Institute Plasmodium knowlesi	Toxoplasma gondii ToxoDB Toxoplasma gondii
Plasmodium chabaudi GeneDB Plasmodium chabaudi	Plasmodium vivax The Institute for Genomic Research Plasmodium vivax	
Plasmodium falciparum GeneDB Plasmodium falciparum 3D7	Tetrahymena thermophila The Institute for Genomic Research Tetrahymena thermophila SB210	

Amoebozoa

Dictyostelium discoideum DictyBase Dictyostelium discoideum	Entamoeba histolytica AmoebaDB Entamoeba histolytica HM-1:IMSS	
---	--	--

Stramenopiles

Albugo laibachii The Sainsbury Laboratory Albugo laibachii Nc14	Phytophthora infestans BROAD Phytophthora infestans	Pythium ultimum Pythium Genome Database Pythium ultimum
---	---	---

Protists

[e! EnsemblFungi](#) ▾ BLAST | More ▾ [Search Ensembl Fungi species...](#)

Help & Documentation Species List

Find a Species

Ensembl Fungi Species

Capnodiidae

Mycosphaerella graminicola JGI Mycosphaerella graminicola IPO323
--

Eurotiidae

Aspergillus clavatus CADRE Aspergillus clavatus	Aspergillus fumigatus1163 CADRE Aspergillus fumigatus1163 A1163	Aspergillus oryzae CADRE Aspergillus oryzae
Aspergillus flavus CADRE Aspergillus flavus	Aspergillus nidulans CADRE Aspergillus nidulans FGSC A4	Aspergillus terreus CADRE Aspergillus terreus
Aspergillus fumigatus CADRE Aspergillus fumigatus A123	Aspergillus niger CADRE Aspergillus niger CBS 513.88	Neosartorya fischeri CADRE Neosartorya fischeri

Hypocreales

Fusarium oxysporum Broad Institute Fusarium oxysporum 4287	Gibberella zeae Broad Institute Gibberella zeae PH-1	Trichoderma virens JGI Trichoderma virens C8
--	--	--

Fungi

[e! EnsemblMetazoa](#) ▾ BLAST | More ▾ [Search Ensembl Metazoa species...](#)

Help & Documentation Species List

Find a Species

Ensembl Metazoa Species

Diptera

Aedes aegypti VectorBase Aedes aegypti	Drosophila grimshawi FlyBase Drosophila grimshawi	Drosophila simulans FlyBase Drosophila simulans
Anopheles darlingi European Nucleotide Archive Anopheles darlingi	Drosophila melanogaster FlyBase Drosophila melanogaster	Drosophila virilis FlyBase Drosophila virilis
Anopheles gambiae VectorBase Anopheles gambiae	Drosophila mojavensis FlyBase Drosophila mojavensis	Drosophila willistoni FlyBase Drosophila willistoni
Culex quinquefasciatus VectorBase Culex quinquefasciatus	Drosophila persimilis FlyBase Drosophila persimilis	Drosophila yakuba FlyBase Drosophila yakuba
Drosophila ananassae FlyBase Drosophila ananassae	Drosophila pseudoobscura FlyBase Drosophila pseudoobscura	
Drosophila erecta FlyBase Drosophila erecta	Drosophila sechellia FlyBase Drosophila sechellia	

Metazoa

[e! EnsemblPlants](#) ▾ BLAST | More ▾ [Search Ensembl Plants species...](#)

Help & Documentation Species List

Find a Species

Ensembl Plants Species

Liliopsida

Brachypodium distachyon Brachypodium.org Brachypodium distachyon (L.) Beauvois	Oryza glaberrima AGI Oryza glaberrima	Sorghum bicolor JGI Sorghum bicolor BT
Hordeum vulgare IBSC Hordeum vulgare	Oryza sativa MSU Oryza sativa Nipponbare (Japonica rice)	Zea mays MaizeSequence Zea mays
Musa acuminata CIRAD Musa acuminata Doubled-haplloid Pahang (DH-Pahang)	Oryza sativa Indica Group RIS Oryza indica 93-11 (Indica rice)	
Oryza brachyantha OGE Oryza brachyantha	Setaria italica JGI Setaria italica	

eudicotyledons

Arabidopsis lyrata JGI Arabidopsis lyrata	Glycine max JGI Glycine max	Solanum tuberosum PGSC Solanum tuberosum
Arabidopsis thaliana TAIR Arabidopsis thaliana	Populus trichocarpa JGI Populus trichocarpa	Vitis vinifera Genoscope Vitis vinifera

Plants

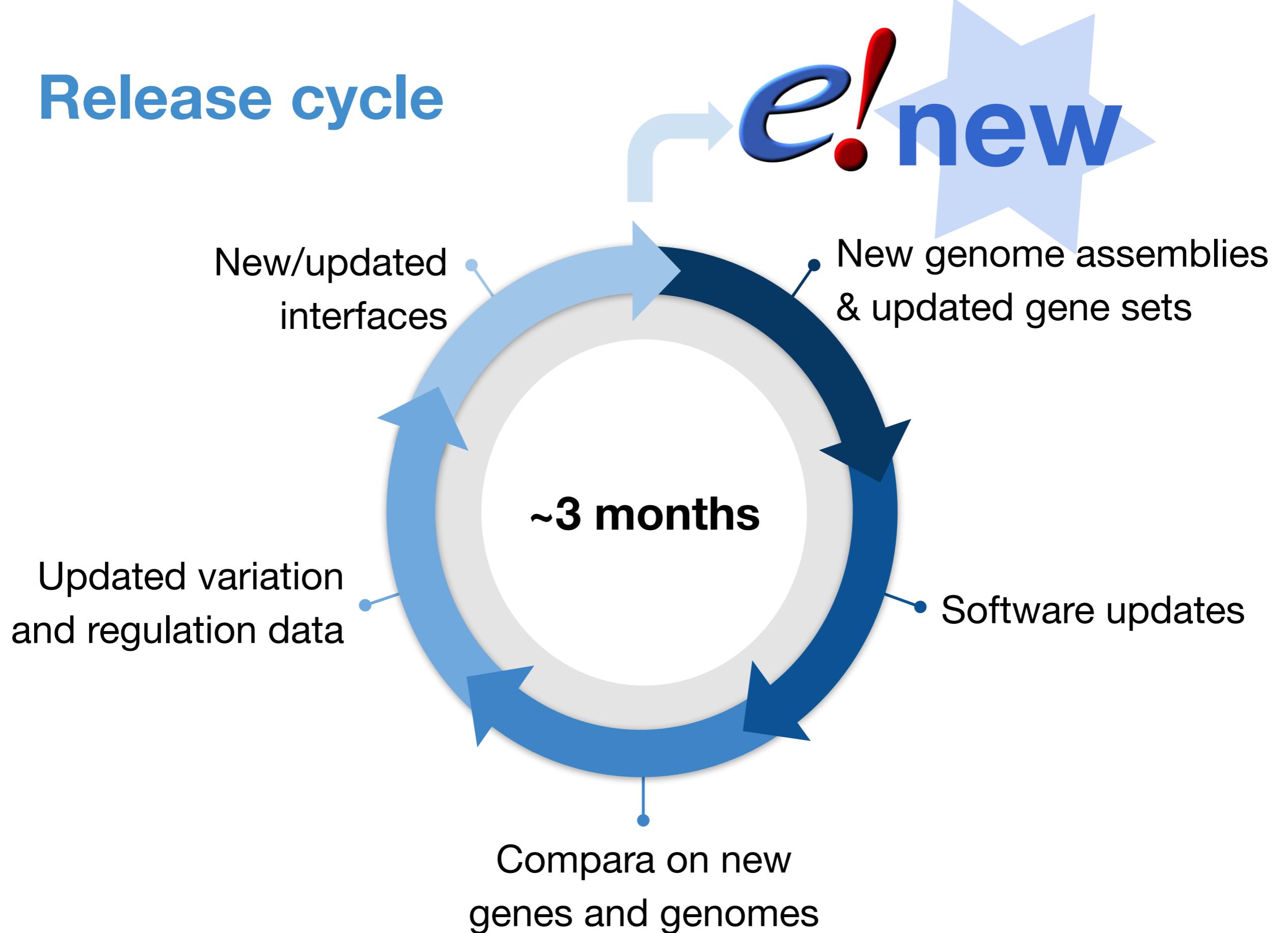
www.ensemblgenomes.org



Ensembl and Ensembl Genomes

	Ensembl	Ensembl Genomes
Released	2000	2009
Species	Vertebrates (fly, worm and yeast as outgroups)	Non-vertebrates (protists, plants, fungi, metazoa, bacteria)
Annotation	by Ensembl	in collaboration with the scientific communities
URL	www.ensembl.org	www.ensemblgenomes.org

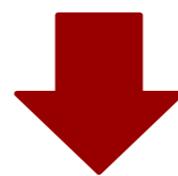
Release cycle



Ensembl Rapid Release

- Released every two weeks 
- Genome with gene annotation only 
- BLAST 
- No BioMart 
- No gene trees/homologues 
- No variation 

Cloning into BACs



Cut into chunks and
clone into BACs

BL001

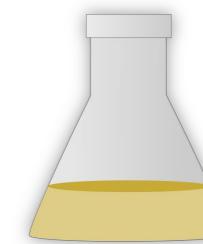
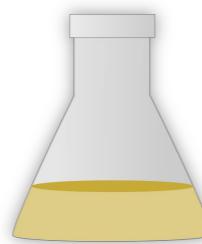
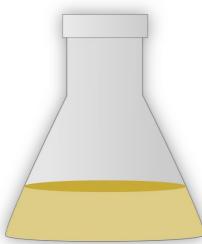
BL002

BL003

BL004



Grow up in bacterial
cultures



Making a contig

Sequence reads

CGGCCTTGGGCTCCGCCTCAGCTCAAGA

CAGCTGTCCCAGATGAC ACTTAACCTCCCTCCCAGCTGTCC

GGGCTCCGCCTCAGCTC AACTCCCTCCCAGCT TCCCAGCTGTCCCAGATGACGCCATC

CGGCCTTGGGCTCC CAGATGACGCC TCCGCCTTCAGCTCAAGACTTAACCTTC

Match up overlaps

CGGCCTTGGGCTCCGCCTCAGCTCAAGA AACTCCCTCCCAGCT CAGATGACGCC
TCCGCCTTCAGCTCAAGACTTAACCTTC TCCCAGCTGTCCCAGATGACGCCATC
GGGCTCCGCCTCAGCTC ACTTAACCTCCCTCCCAGCTGTCC
CGGCCTTGGGCTCC CAGCTGTCCCAGATGAC

Contig

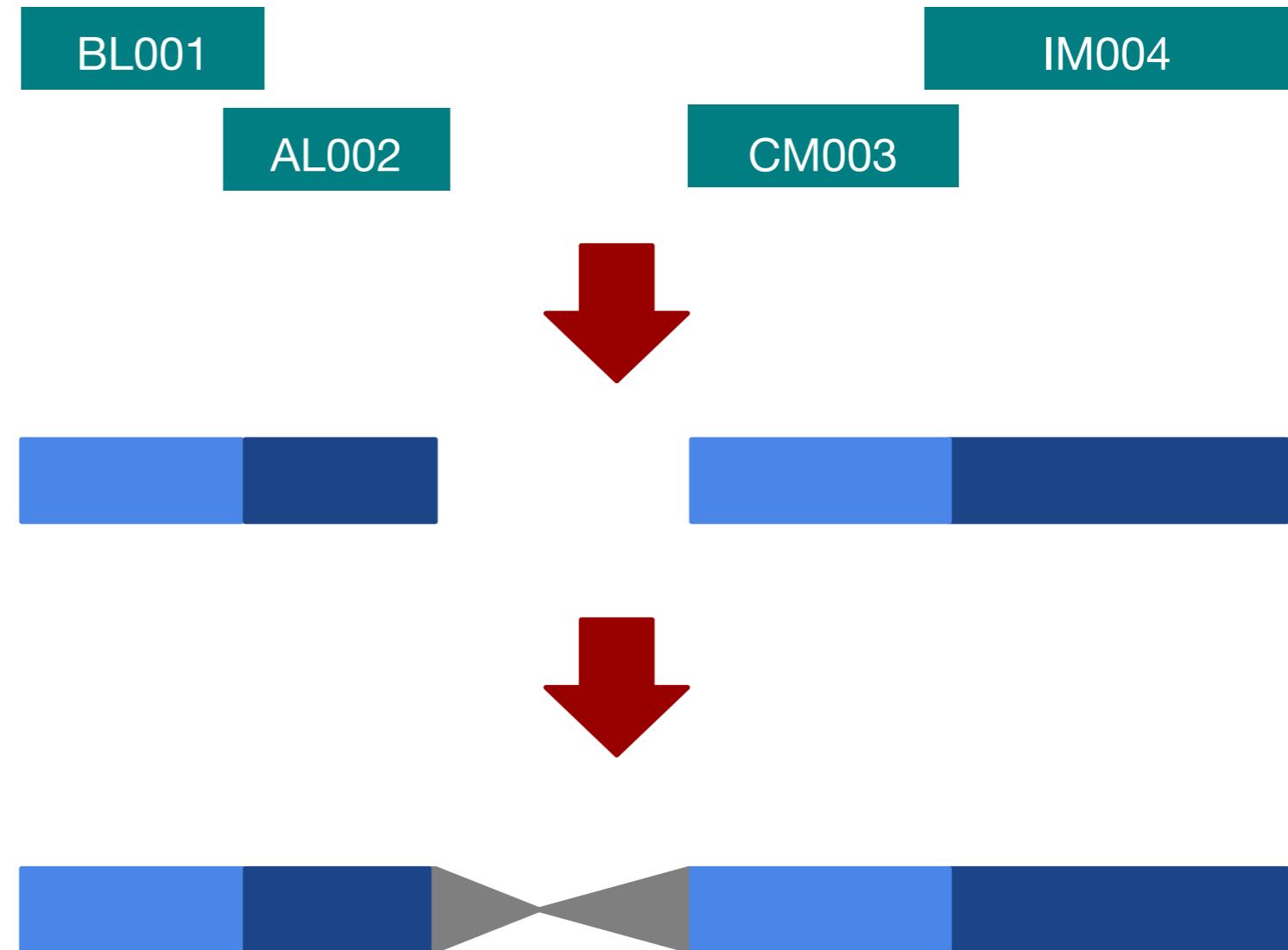
CGGCCTTGGGCTCCGCCTCAGCTCAAGACTTAACCTCCCTCCCAGCTGTCCCAGATGACGCCATC

Contigs to scaffolds

BACs from different individuals assembled together with overlaps
Tilepath

Overlaps trimmed to give **contigs**. A run of contigs with no gaps is a **scaffold**.

Genetic maps are used to assemble scaffolds into a chromosome

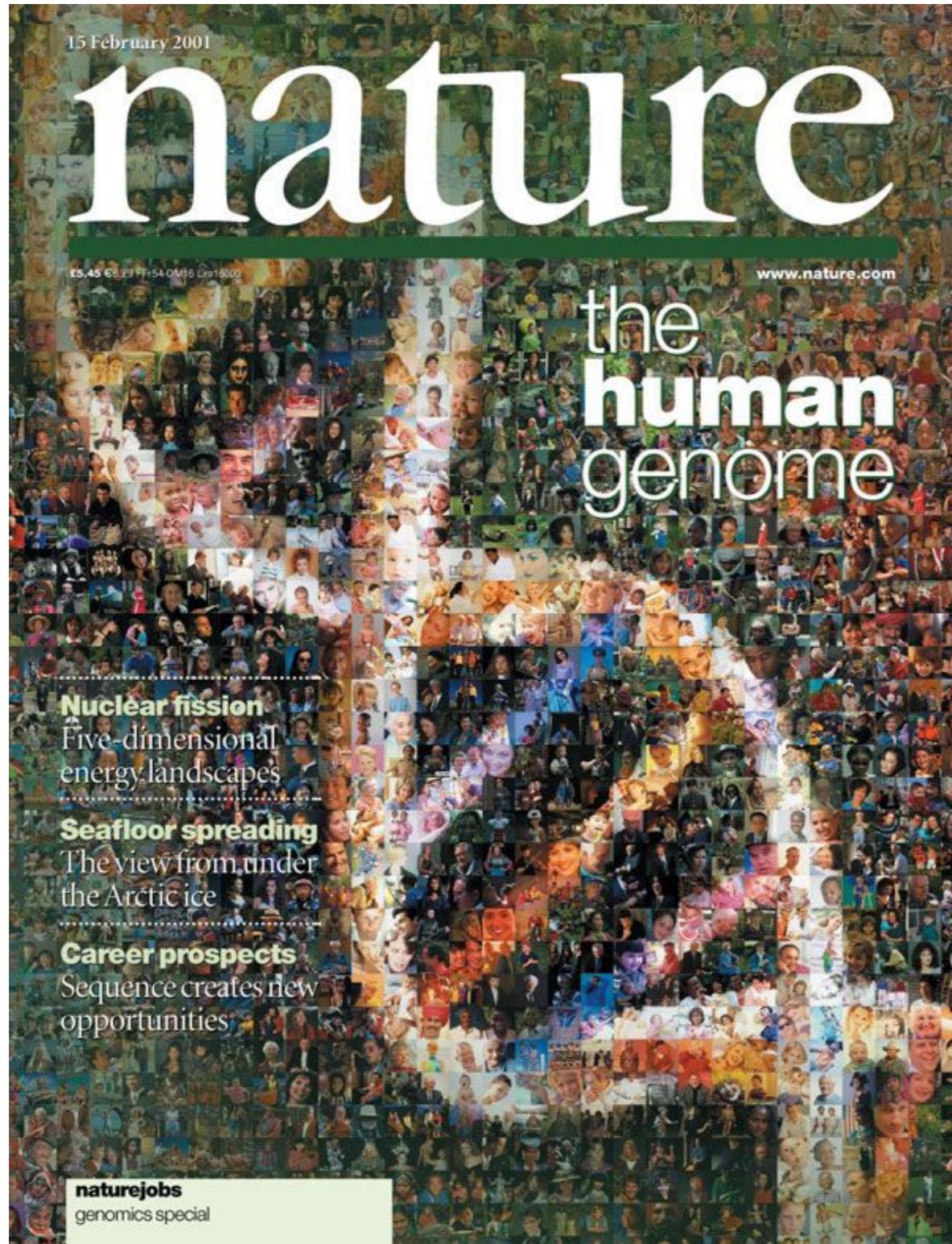


Tilepath

Contig/
scaffold

Chromosome

Genome assemblies



BL



AL



CM



IM

Genome contigs



BL001

AL002

CM003

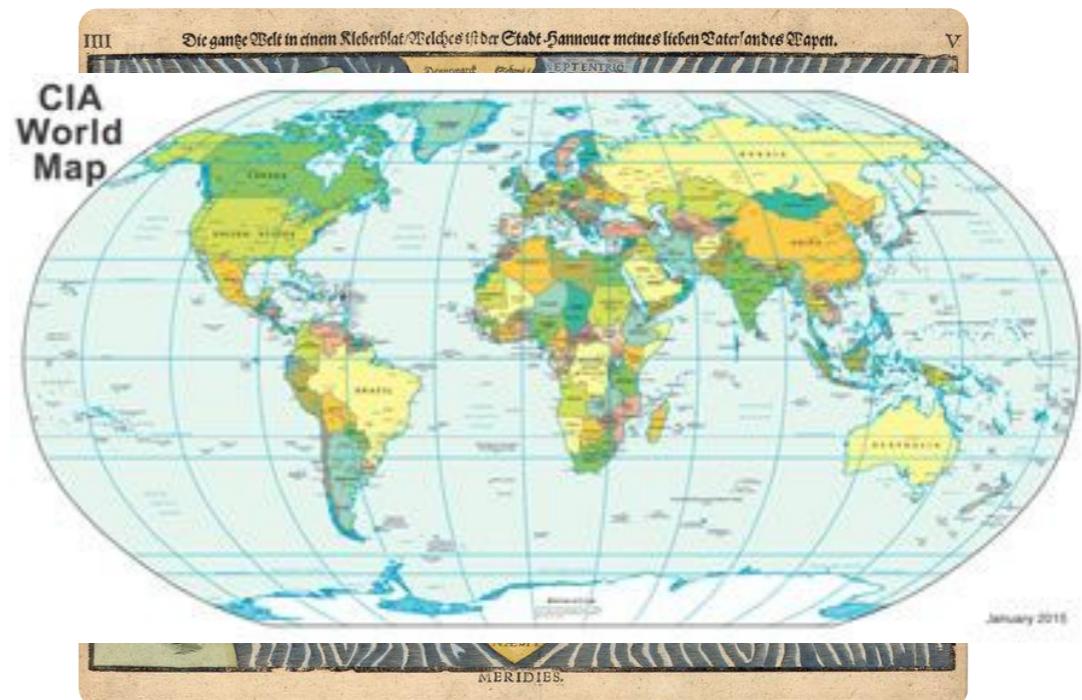
IM004

Genome assemblies

Genome
“DNA within a cell”



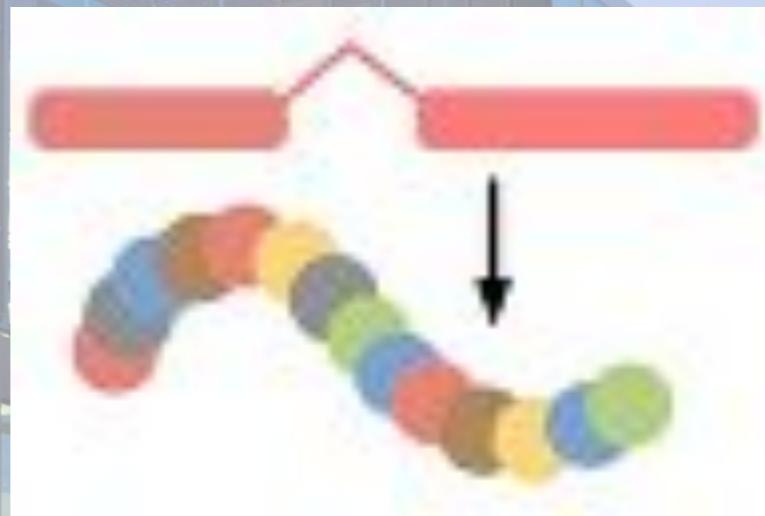
Genome assembly
Representation of a genome
Contains errors and gaps
Coordinate system



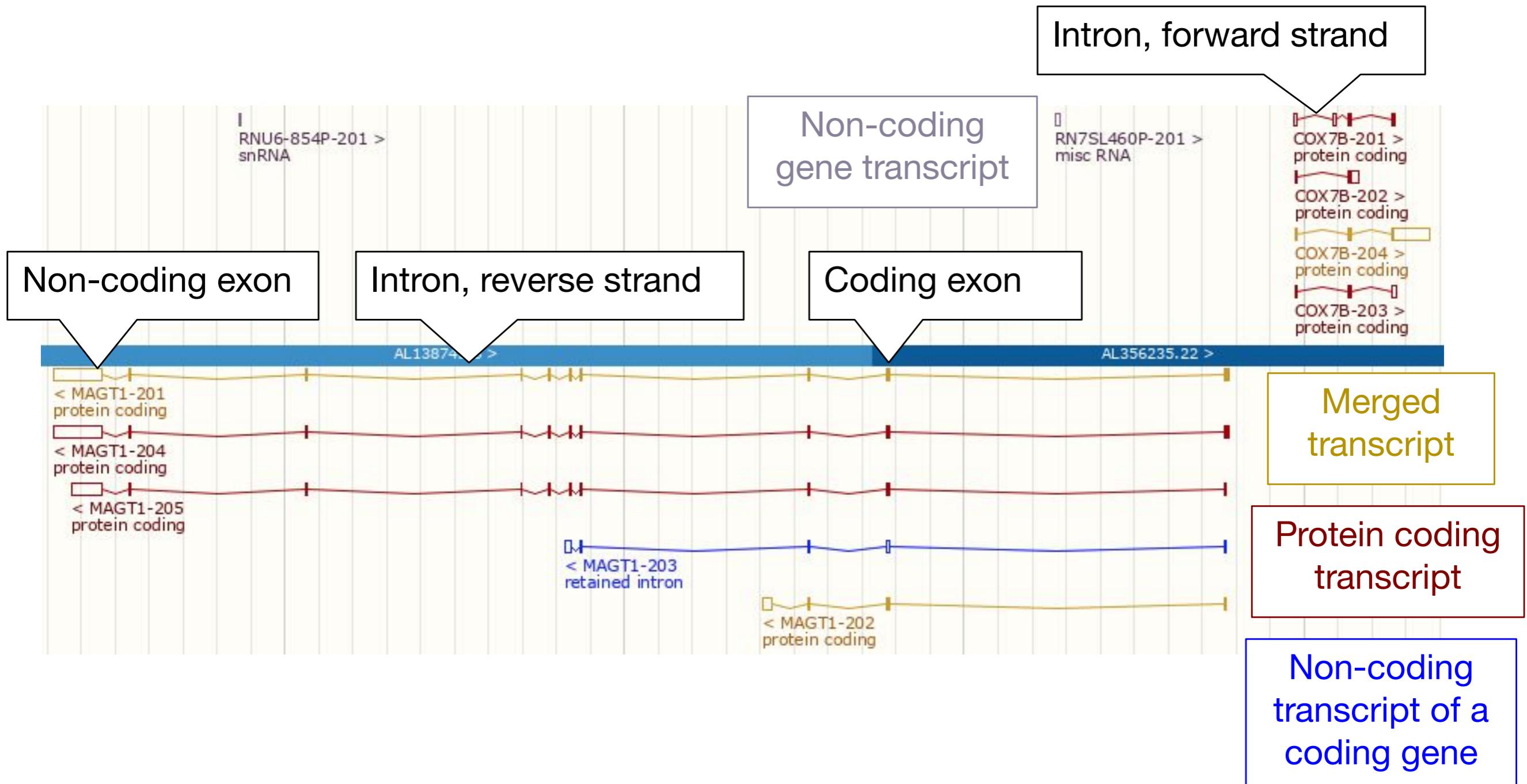
Human genome assemblies

- GRCh38 (aka hg38) 
 - Many rare/private alleles replaced.
 - www.ensembl.org
 - Most up-to-date and supported
- GRCh37 (aka hg19) 
 - Some large gaps
 - grch37.ensembl.org
 - Limited data and software updates
 - Still the preferred genome of the clinical community
- NCBI36 (aka hg18) 
 - Many gaps
 - ncbi36.ensembl.org
 - No longer updated

Genes and Transcripts

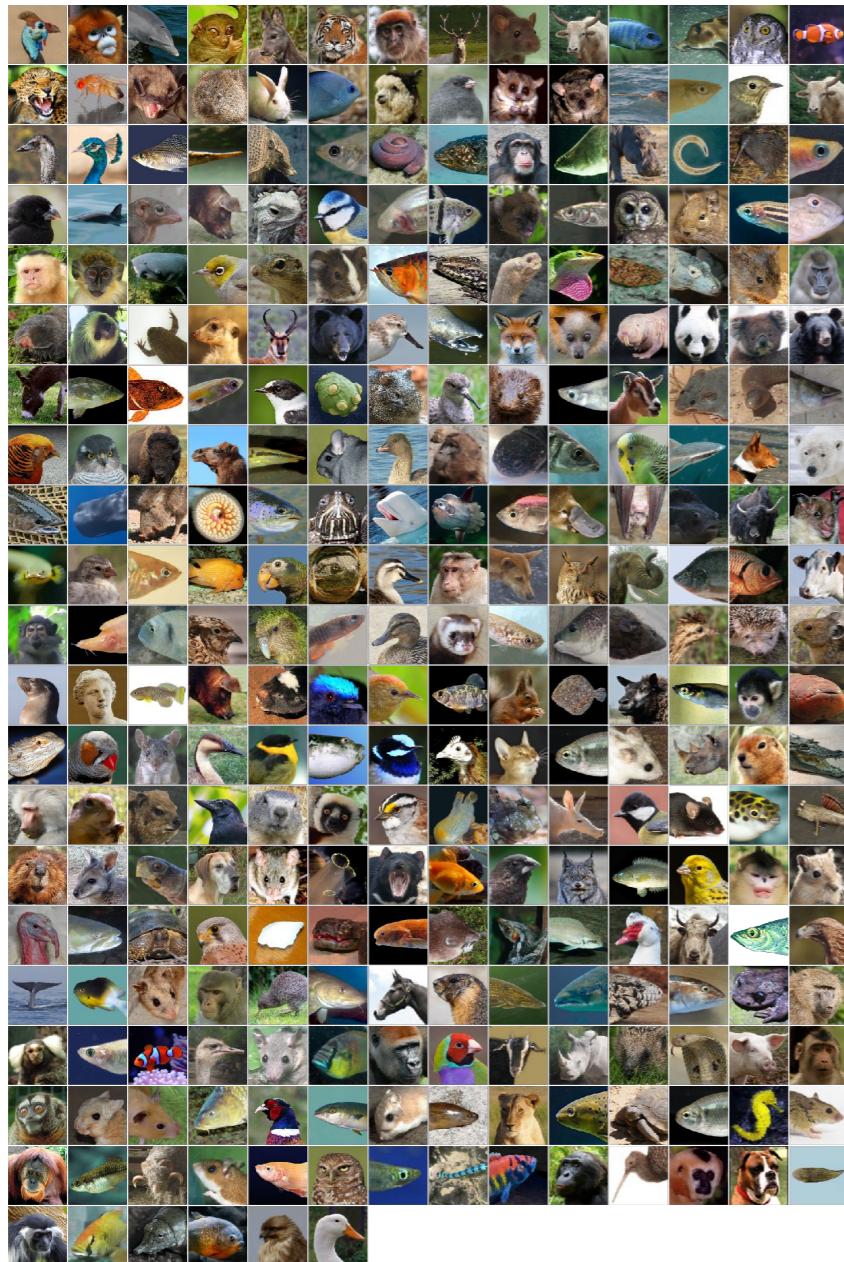


Gene views

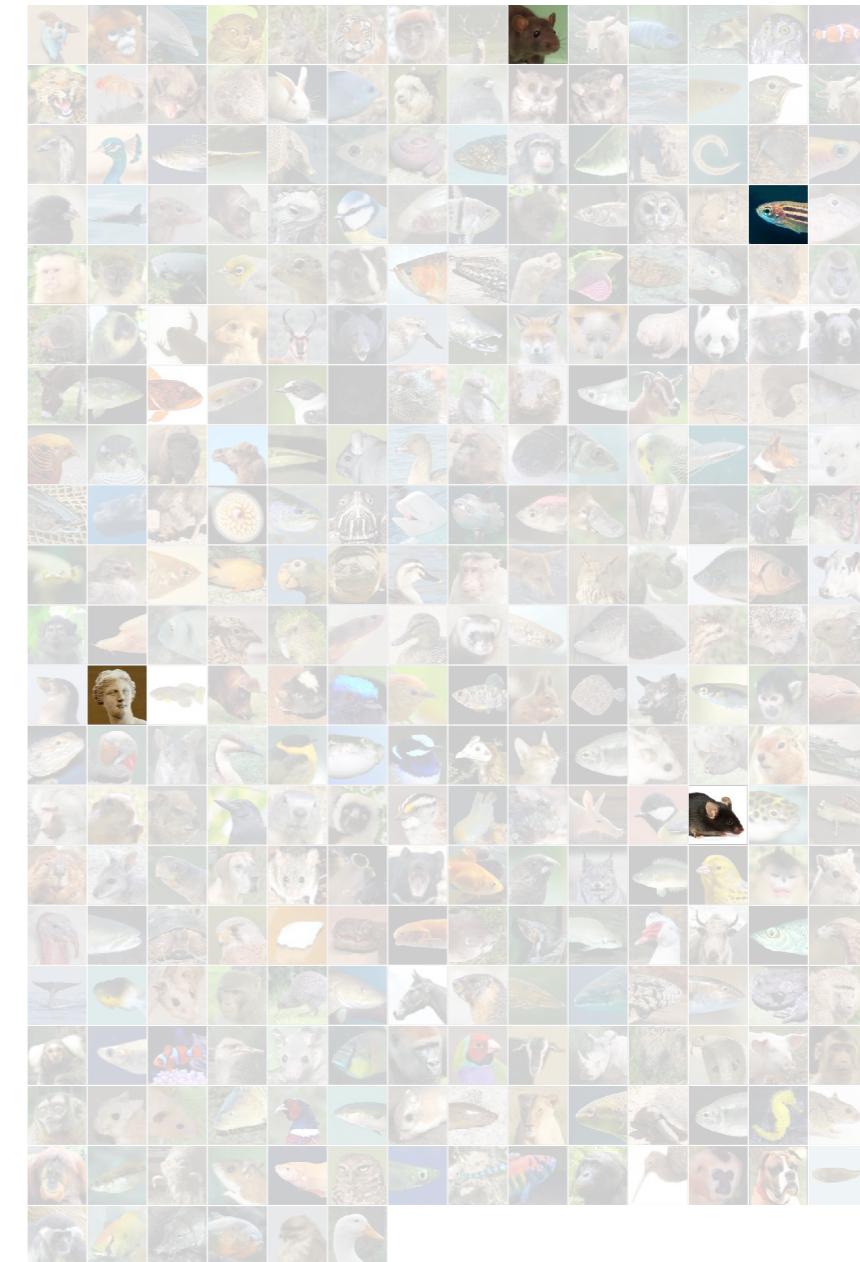


Automatic and manual annotation

Automatic annotation



Manual annotation



Automatic and manual annotation

Automatic annotation

- computational pipeline
 - one genome = two weeks

Manual annotation



- human annotator
 - one genome = several years

Annotation sources

Automatic annotation

- INSDC
 - cDNAs
 - ESTs
 - RNA-seq
- Protein sequence databases
 - Swiss-Prot: manually curated
 - TrEMBL: unreviewed translations

Manual annotation

- RNA-seq transcriptome data
 - Illumina short read
 - Oxford Nanopore long read
 - PacBio long read
- Transcript structure data
 - Introns
 - CAGE transcription start sites
 - PolyA-Seq transcription ends
- Mass Spec protein data
- publications

The Three Layers of Genome Annotation

What Are Features?

protein_coding	23249
pseudogene	17311
lncRNA	19399
miRNA	2137
transcribed_pseudogene	1584
tRNA	691
snoRNA	1301
V_segment	365
V_segment_pseudogene	299
J_segment	117
snRNA	175
D_segment	61
C_region	33
ncRNA	54
other	30
rRNA	81
antisense_RNA	20
misc_RNA	65
J_segment_pseudogene	11
C_region_pseudogene	7
vault_RNA	4
scRNA	4
Y_RNA	4
telomerase_RNA	1
ncRNA_pseudogene	1
RNase_P_RNA	2
RNase_MRP_RNA	1

Features Present in
Homo sapiens
GRCh38.p14
GFF3

Manual vs automatic

- Manual annotation is more comprehensive
 - More transcripts per gene, especially non-coding transcripts
 - More genes overall, especially non-coding genes
- More biotypes
 - e.g. polymorphic pseudogene, NMD, non-stop decay, stop codon read-through
- Manual annotation can be more accurate for difficult to annotate features such as:
 - UTRs
 - Splice sites
 - Single exon transcripts
 - Exceptions, such as immunoglobulins, stop codon readthroughs



GENCODE == Ensembl human and mouse genes

GENCODE is the default gene set used by major projects such as:

- gnomAD/ ExAC: Exome Aggregation Consortium, Genome Aggregation
- GTEx: Genotype-Tissue Expression
- Decipher
- 100,000 Genomes Project, Genomics England
- ENCODE: Encyclopedia of DNA Elements
- TCGA: The Cancer Genome Atlas
- ICGC: International Cancer Genome Consortium
- Roadmap: NIH Roadmap Epigenomics Mapping Consortium
- Blueprint: Blueprint Epigenome
- 1000 Genomes Project
- HCA: Human Cell Atlas

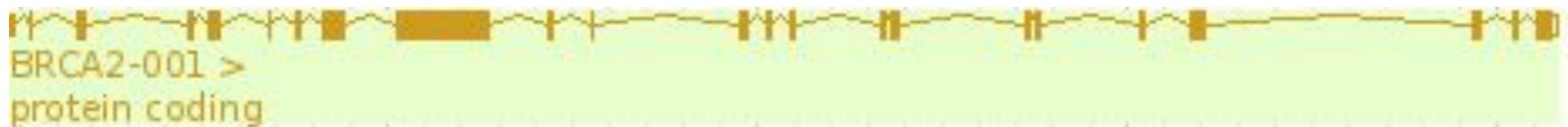


Golden transcripts

- Identical annotation



- Higher confidence and quality



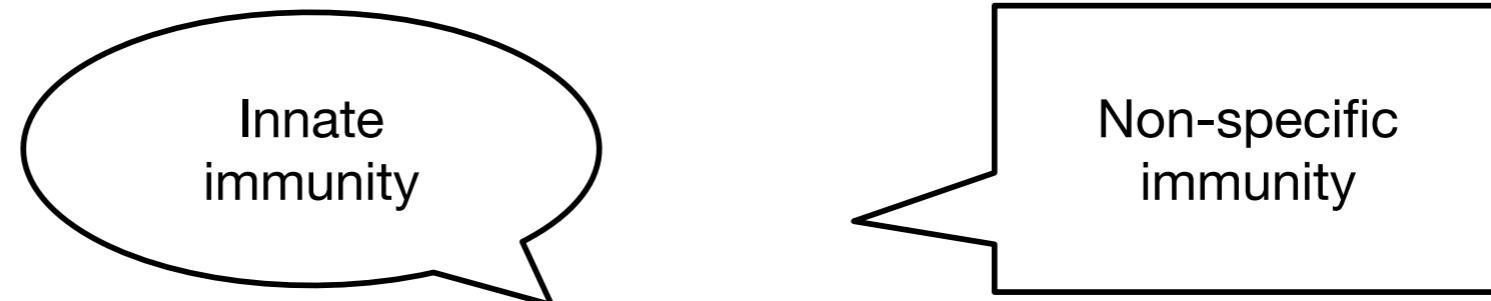
Ensembl stable IDs

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

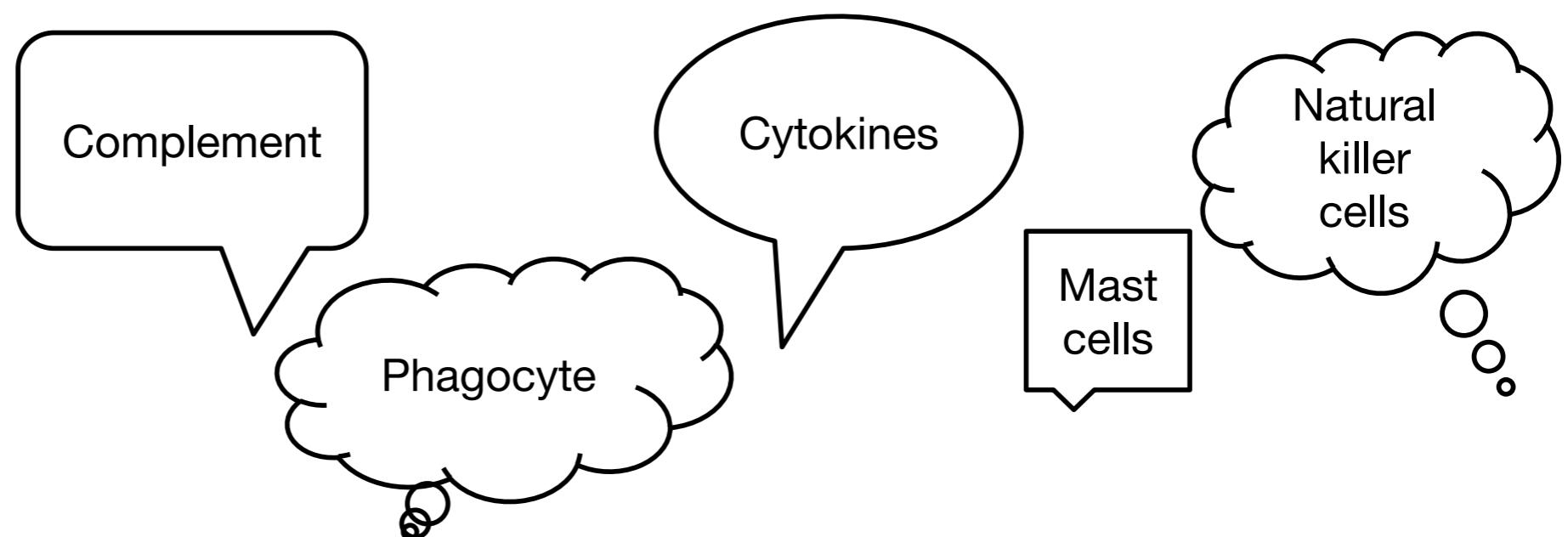
http://www.ensembl.org/info/genome/stable_ids/index.html

Why Gene Ontology (GO)?

Multiple terms for
the same thing



Gene descriptions
too specific



THE GENE ONTOLOGY RESOURCE

The mission of the GO Consortium is to develop a comprehensive, **computational model of biological systems**, ranging from the molecular to the organism level, across the multiplicity of species in the tree of life.

The Gene Ontology (GO) knowledgebase is the world's largest source of information on the functions of genes. This knowledge is both human-readable and machine-readable, and is a foundation for computational analysis of large-scale molecular biology and genetics experiments in biomedical research.

Search GO term or Gene Product in AmiGO ...


 Any Ontology Gene Product

GO Enrichment Analysis

Powered by PANTHER

Your gene IDs here...

biological process

Homo sapiens

Examples

Launch

Hint: can use UniProt ID/AC, Gene Name, Gene Symbols, MOD IDs


The network of biological classes describing the current best representation of the “universe” of biology: the molecular functions, cellular locations, and processes gene products may carry out.

- [GO Ontology Overview](#)
- [Browse in AmiGO](#)
- [Download](#)



Statements, based on specific, traceable scientific evidence, asserting that a specific gene product is a real exemplar of a particular GO class.

- [GO Annotations Overview](#)
- [Browse in AmiGO](#)
- [Download](#)



GO Causal Activity Model (GO-CAM) provides a structured framework to link standard GO annotations into a more complete model of a biological system.

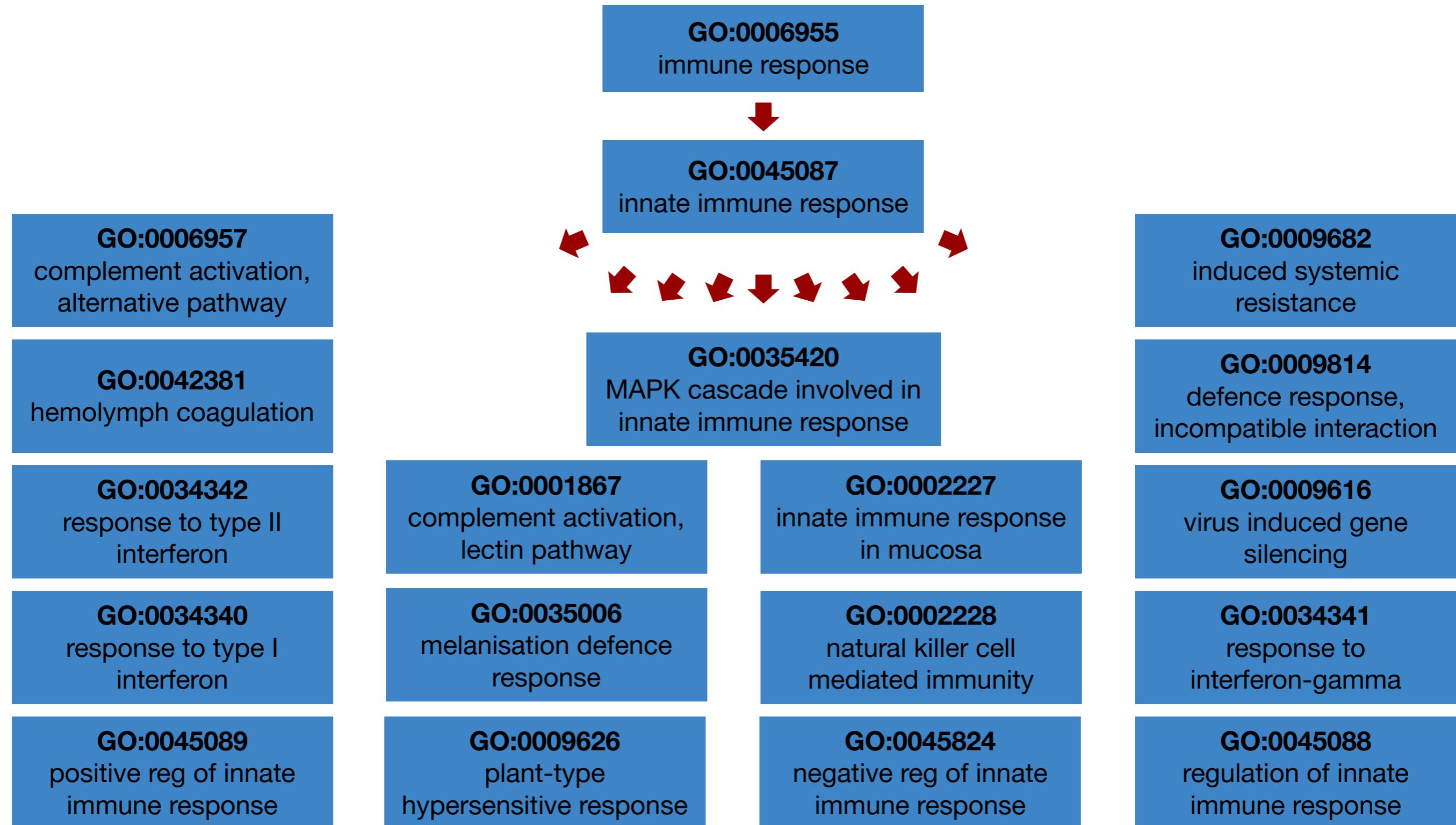
- [GO-CAM Overview](#)
- [Browse GO-CAMs](#)
- [Download](#)



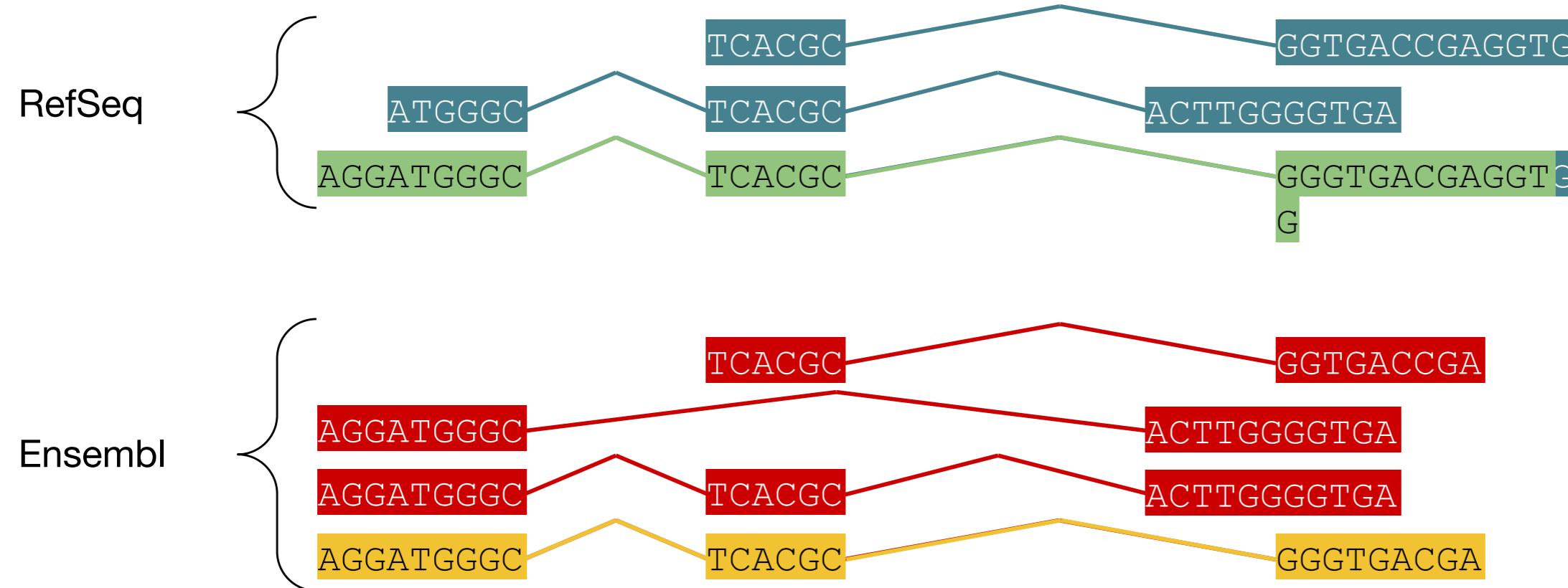
Tools to curate, browse, search, visualize and download both the ontology and annotations. Includes bioinformatic guides (Notebooks) and simple API access to integrate the GO into your research.

- [GO Tools Overview](#)
- [GO APIs Guide](#)
- [GO GitHub](#)

GO terms are hierarchical



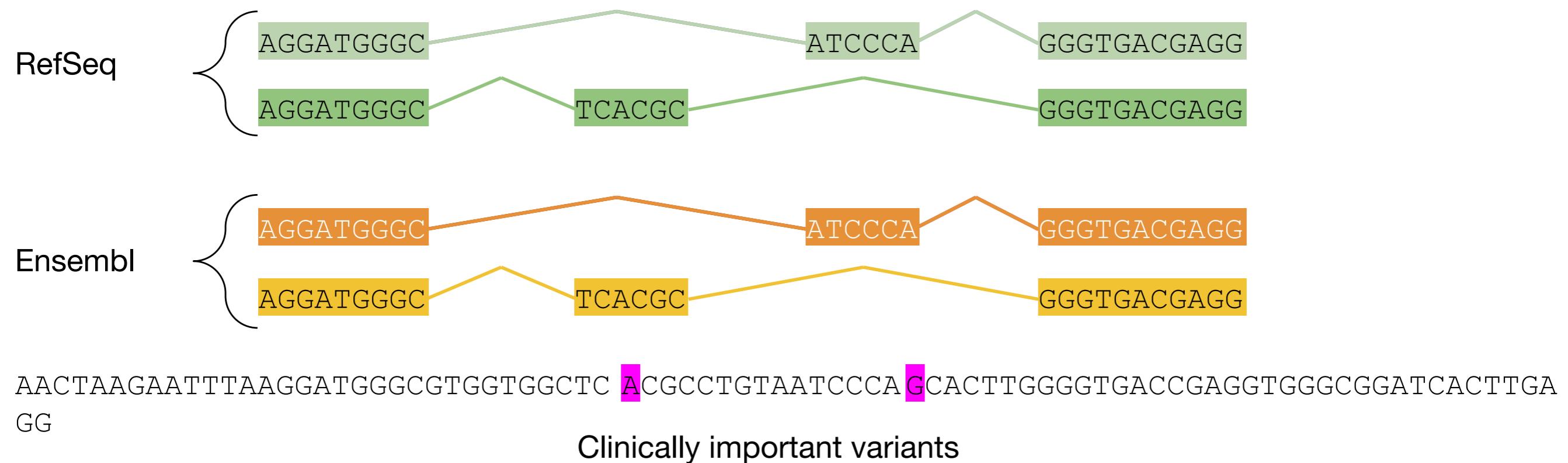
MANE Select



AACTAAGAATTAAAGGATGGCGTGGTGGCTCACGCCTGTAATCCCAGCACTTGGGGTACCGAGGTGGCGGATCACTTGA
GG

Select = most clinically relevant, based on expression
levels, clinically important genetic variation and
conservation between species

MANE Plus Clinical



Canonical transcript

APPRIS P1 AGGATGGGC TCACGC GGGTGACGA

Uniprot canonical -M--G-- S--R-- G--*--

Clinical variants T/- C/A

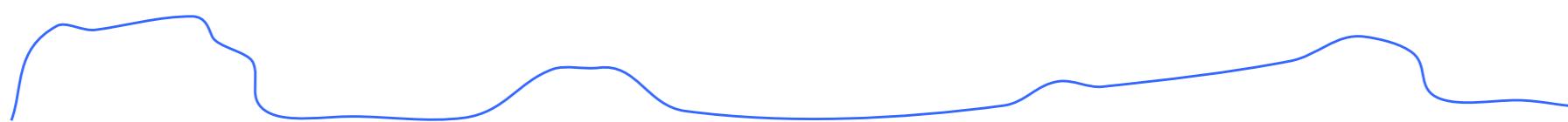
PhyloCSF conservation

A G G A T G G G C
A G G A T G G G C
A G G A T G G G C
A G G A T G G G C

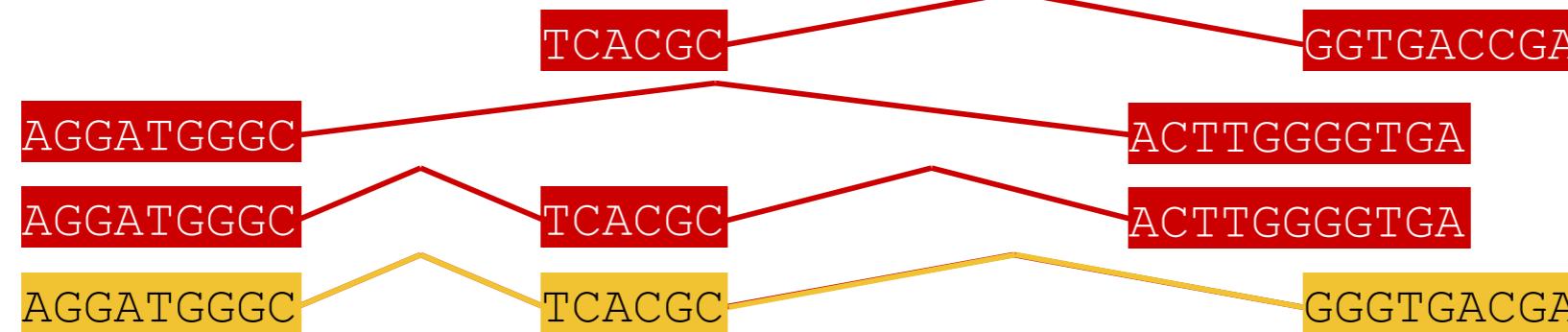
T C A C G G C
T C A C G G C
T C A C G G C
T C A C G G C

A C T T G G G G T G A C G A
A G T T G G G G T G A C G A
A G T T G G G G T G A C G A
A C T T G G G G T G A C G A

RNA-seq



Intron-spanning reads



AACTAAGAATTAAAGGATGGCGTGGCTCACGCCTGTAATCCCAGCACTTGGGGTACCGAGGTGGCGGATCACTTGA
GG