Training materials

If you use Ensembl for your work, please cite our papers

http://www.ensembl.org/info/about/publications.html

Browsing Genes and Genomes with Ensembl

Academia Sinica
Taiwan

Dr Denise Carvalho-Silva

European Molecular Biology Laboratory

European Bioinformatics Institute

Today 09:00-12:00

- Introduction to Ensembl
- Browser walkthrough

10:30-10:50 coffee/tea

- Browser exercises
- Ensembl tools (Talk + Exercises)
- Wrap up, photo opportunity & feedback survey

Materials

http://www.ebi.ac.uk/~denise/workshops/2016/taiwan/sinica

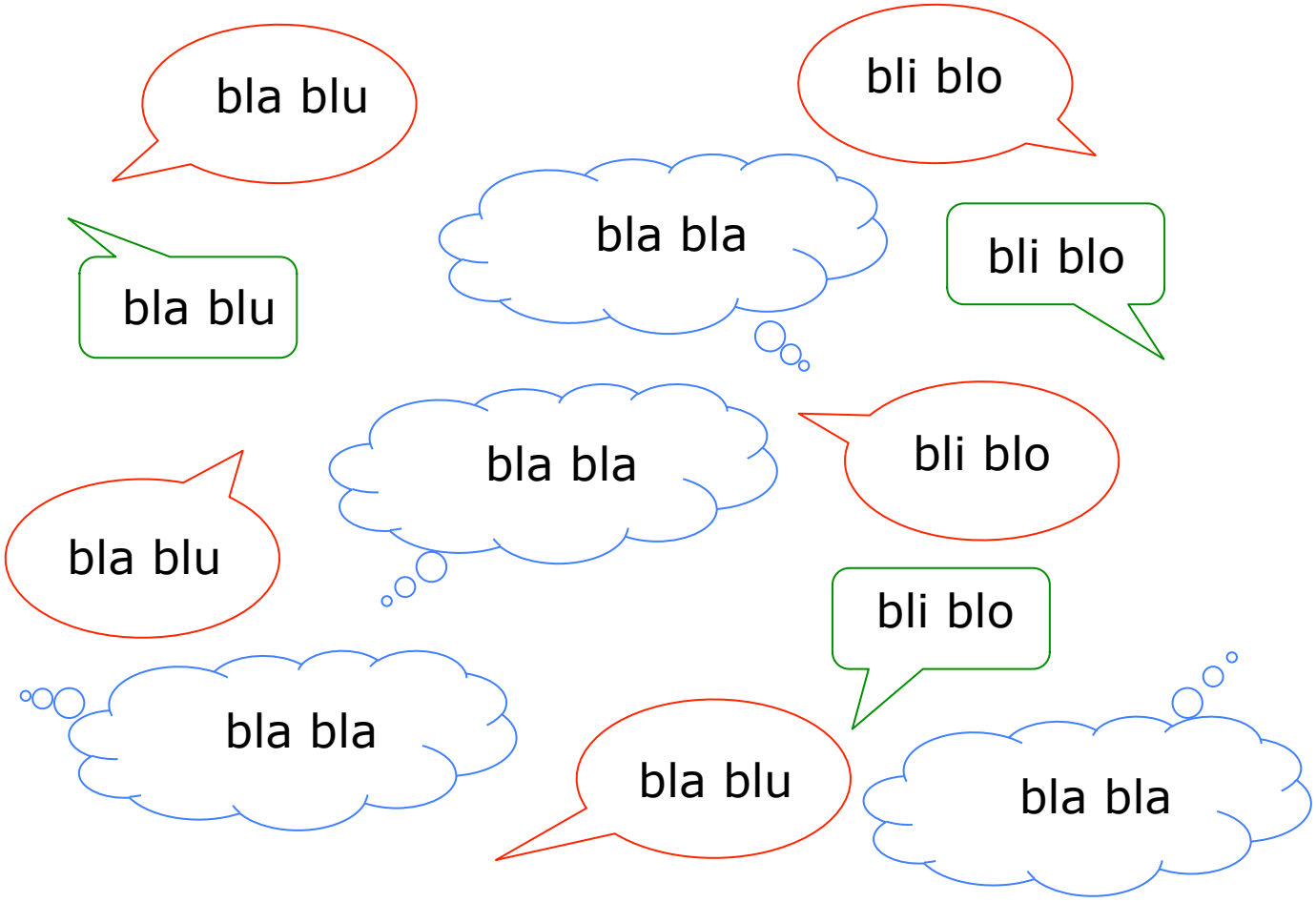
Course Objectives

What is Ensembl?

What type of data can you get in Ensembl?

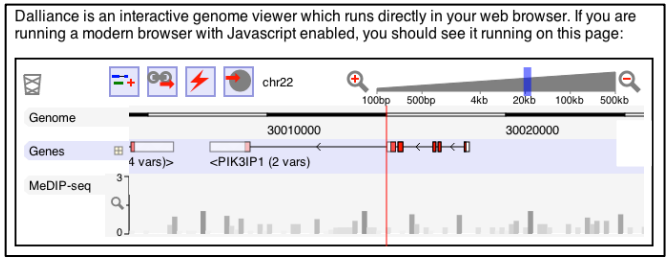
How to navigate the Ensembl browser website?

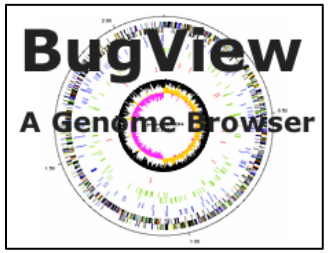
How to connect with Ensembl

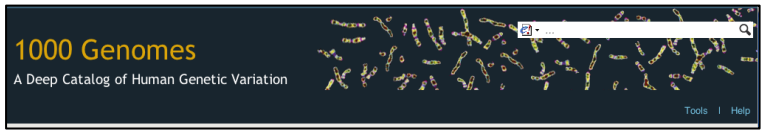

bla blu
bli blo
bla bla
bla blu
bli blo
bla bla
bli blo
bla blu
bli blo
bla bla
bla blu
bla bla

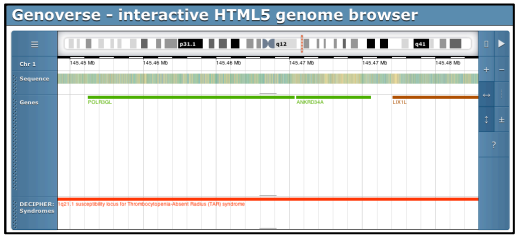
Introduction

Why do we need/have genome browsers?









Genome sequencing

1977: 1st genome to be sequenced (5 kb)
2000: draft human sequence (3 gb)

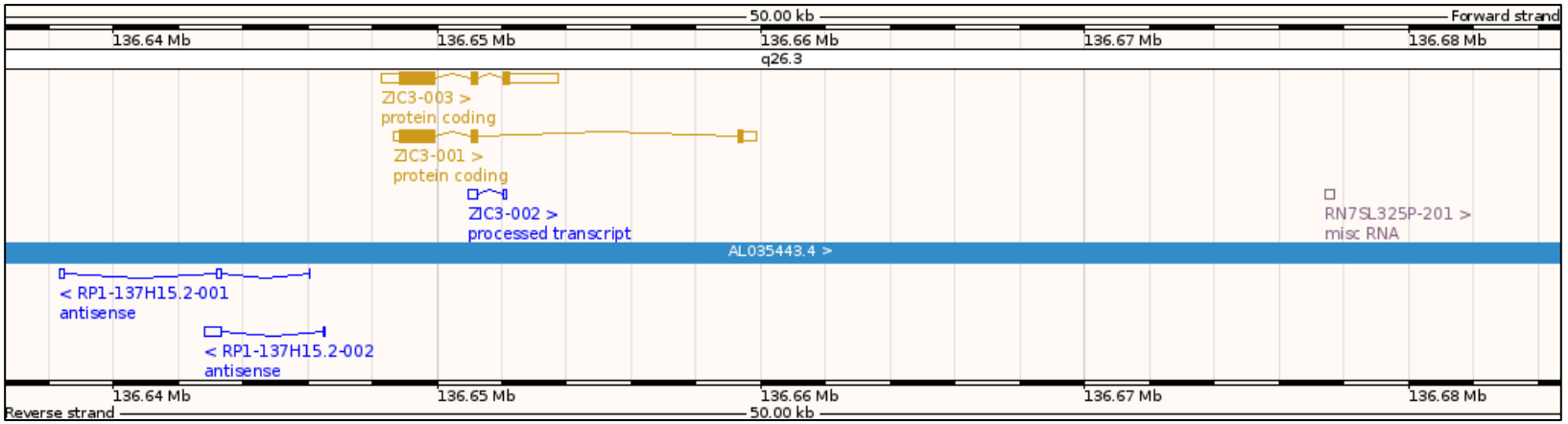
Large amounts of raw DNA sequence data

Raw DNA sequence data

```
>chromosome:GRCh37:18:45357922:45457515:-1
GCCGGGAGGCGGGGCGGGCCGTAGGCAAAGGGAGGTGGGGAGGCGGTGGCCGGCGACTCCCCGCGCCCCGCTCGC
CCCCCGGCCCTTCCCGCGGTGCTCGGCCTCGTTCCTTTCCTCCTCCGCTCCCTCCGTCTTCCATACCCGCCCCGC
GCGGCTTTCGGCCGGCGTGCCTCGCGCCCTAACGGGCGGCTGGAGGCGCCAATCAGCGGGCGGCAGGGTGCCAGC
CCCGGGGCTGCGCCGGCGAATCGGCGGGGCCCGCGGCCCAGGGTGGCAGGCGGGTCTACCCGCGCGGCCGCGGCG
GCGGAGAAGCAGCTCGCCAGCCAGCAGCCCGCCAGCCGCCGGGAGGTGGGTGCGTGGCGCCGCGGCGGCCGGCGG
CCGAGGGCGGAGGGCGGAAGCGGAGGTGGGCTGGCGGGGGAGGGCGCGGCCGTGCGGGCGGCCGGTAGGGCTGCG
GGCGCGCGCCTGAGGGGAGGAGGGGCAGCGCGGGCGCGCGCGTCCTCACCCCCTCCTTCCCCGCGGGCGGCGGCC
AGGCTCCCTCCCCTCCCCTTCCCTCTCCTCCCCTCCCCTCCCCTCTCTTCCCCTACCCTCCCGCGCGCCCGGGCC
GCCGGCCGGGCCCGGGCCTGGGGGCGGGGCGGGAAGACGGCGGCCGGGAGTGTTTTCAGTTCCGCCTCCAATCGC
CCATTCCCCTCTTCCCCTCCCAGCCCCCTCCATCCCATCGGAAGAGGAAGGAACAAAAGGTCCCGGACCCCCCGG
ATCTGACGGGGCGGGACCTGGCGCCACCTTGCAGGTAAAGCCTGGGCGCCCGCGGGCCTCCAGCTAGGGAAGTGT
TTGCGTGCGTCCGCGGCCGGGGCGATGGGCCGTGTCACATGGCCGCTGCGGGTGGGGGCTGGGGTGTGGTGAGTT
CGGGGGCTGTGGGTGCGCCGGCCCGGGCGTGCGGGTTCGGGGCCGGAGAGCCGGGAAGGGACGGGGCTCGGTTGC
ACTGCGCTGCCGCCCAGGCTGACGGGGGCGGGGGCTGCCTGCGTCCCTTCCCCTCGCTGCTCTCACACTCCATAG
TTTGTGGTTCTGATTTTTAAAGCCGAAGGGCCAGAGCTCTTGTTTCAGGAGTTGTGGGGAAGCCTTGTTAGGGAC
GCGTAATACTTTGCCTCCACTTTTTTTTTTTGGTCTCGTAACCTTTGTTAACAGTGGTAGTCTCGGGTTTTCCAT
GGCTTGTGTGTACTCACAAACACGCACACAGCTGATAACTCTGTGATCATCCTTCACTCACTTGTGAAAGTTGTT
CTGCGGGTGGAGCCTATTAGTTTTTATCACACGCCTTTGGAAAGCCCTCGAAGTGATTTCTCGTATTTCAAACTT
GGTTTTTAAAATTGCAACTTACTTTGTTCTCTTGGAGGGGGATATTGTCTTTGGGTCAGGTACTTTTGTCTTTTA
CTTATCCTTAAAATGTCCTTTTGATCCTAAACAGAAAAAGAATGCCTGGGTTGGTTTTGTTATTCCTTTCTCTTG
ACCCTTTAAAAGCAAAATACCACCAGTGTGTTGTCAACCATACCTTTAAAAAAGAAATTTCCAGGTAAAACGAAT
TTGCAAGCAGCATTTTCCATCAAAGTCATGTCTTGTTTGTTGCAATTTGGACTATCTTAAATTTGGGTTTGTGAA
ACTTTTAAAACGACATGTGTAAAATAATTGTTAATAAATAAATAACTTGGAATAATTACATCTTTTTAGAAATGG
TCTTTGAAATGTGATTTAATTTCCAAAATTCTTTTTCTCTCTTCTGACTTCATATGAATTGAACCTGATAGTTTC
GTATGAATGAATCGCTGATAGGATGTTTTCCTTGAGCCTAGTAATAAATTTGCTCTTATTATGCTGAAAATTTGC
TATTCTACTTTAACACCCTTTAAAAAGTTTCCATCTTACAGAGAAGTTAGTCAAATAGTTAAATGGAGTTTCAAC
AGTTTTAATACATGAATTAGTGAAGCAGATGAAGCAGAGATGGGAGAAGTAATTTAAAATGGGCTTCATAAATGT
```

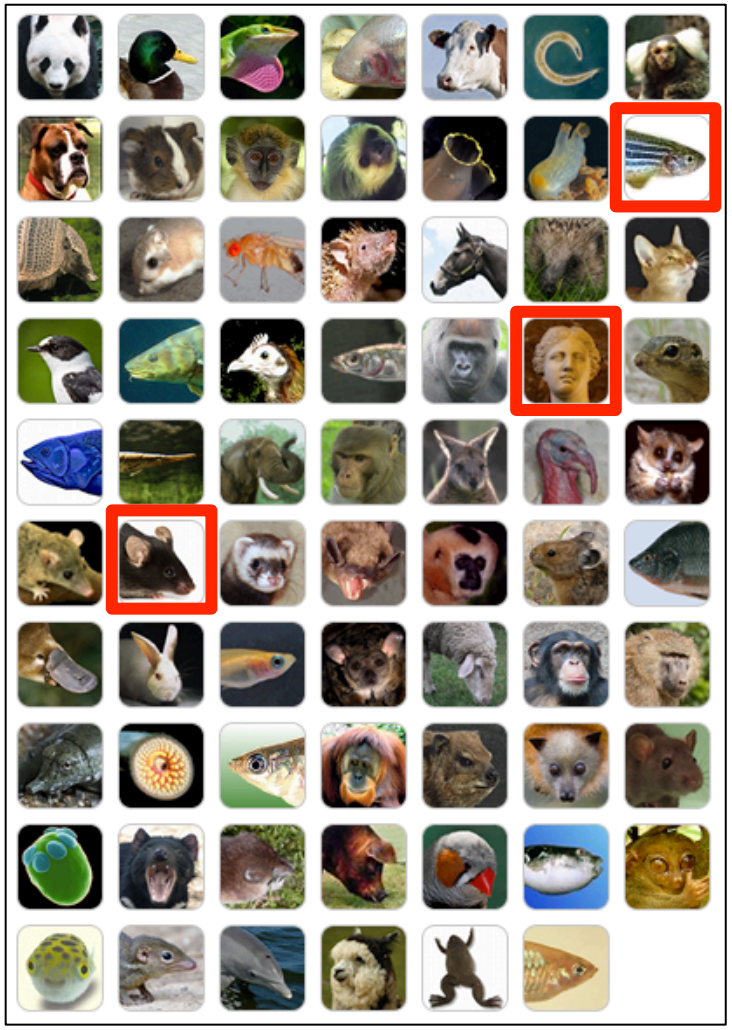
Annotation: making sense

```
>chromosome:GRCh37:18:45357922:45457515:-1
GCCGGGAGGCGGGGCGGGCCGTAGGCAAAGGGAGGTGGGGAGGCGGTGGCCGGCGACTCCCCGCGCCCCGCTCGC
CCCCCGGCCCTTCCCGCGGTGCTCGGCCTCGTTCCTTTCCTCCTCCGCTCCCTCCGTCTTCCATACCCGCCCCGC
GCGGCTTTCGGCCGGCGTGCCTCGCGCCCTAACGGGCGGCTGGAGGCGCCAATCAGCGGGCGGCAGGGTGCCAGC
CCCGGGGCTGCGCCGGCGAATCGGCGGGGCCCGCGGCCCAGGGTGGCAGGCGGGTCTACCCGCGCGGCCGCGGCG
GCGGAGAAGCAGCTCGCCAGCCAGCAGCCCGCCAGCCGCCGGGAGGTGGGTGCGTGGCGCCGCGGCGGCCGGCGG
CCGAGGGCGGAGGGCGGAAGCGGAGGTGGGCTGGCGGGGGAGGGCGCGGCCGTGCGGGCGGCCGGTAGGGCTGCG
GGCGCGCGCCTGAGGGGAGGAGGGGCAGCGCGGGCGCGCGCGTCCTCACCCCCTCCTTCCCCGCGGGCGGCGGCC
AGGCTCCCTCCCCTCCCCTTCCCTCTCCTCCCCTCCCCTCCCCTCTCTTCCCCTACCCTCCCGCGCGCCCGGGCC
GCCGGCCGGGCCCGGGCCTGGGGGCGGGGCGGGAAGACGGCGGCCGGGAGTGTTTTCAGTTCCGCCTCCAATCGC
CCATTCCYCTCTTCCCCTCCCAGCCCCCTCCATCCCRTCGGAAGAGGAAGGAACAAAAGGTCCCGGACCCCCCGG
ATCTGACGGGGCGGGACCTGGYGCCACCTTGCAGGTAAAGCCTGGGCGCCMGCGGGCCKCCAGCTAGGGAAGTGT
TTGYGTGCGTCCGCGGCCGGGGCGATGGGCCGTGTCACATGGCCGCTGCGGGTGGGGGCTGGGGTGTGGTGAGTT
CGGGGGCTGTGGGTGCGCCGGCCCGGGCGTGCGGGTTCGGGGCCGGAGAGCCGGGAAGGGACGGGGCTCGGTTGC
```



Annotation of vertebrate genomes

www.ensembl.org



e!Ensembl

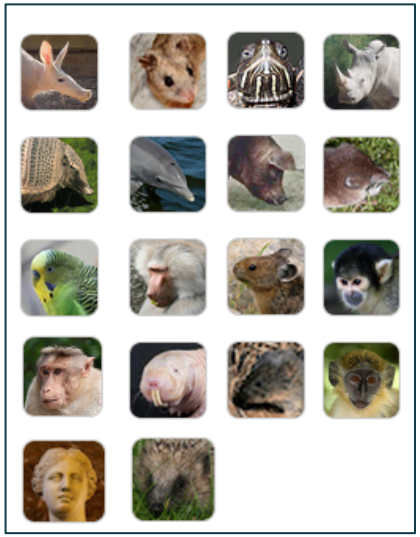
>80 genomes*

D. melanogaster

C. elegans

S. cerevisae

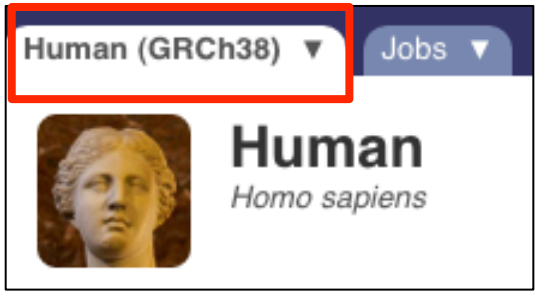
pre.ensembl.org



Pre!Ensembl

**Release 83*
Dec 2015

1 human genome → 3 assemblies



www.ensembl.org

grch37.ensembl.org

e54.ensembl.org

Non-vertebrate genomes

Extends the use of Ensembl to other species

Wider taxonomic range (> 30K genomes*)

*Release 30 December 2015

EnsemblMetazoa	x 55
EnsemblPlants	x 39
EnsemblFungi	x 564
EnsemblProtists	x 152
EnsemblBacteria	x 29,777

www.ensemblgenomes.org

The Ensembl projects

www.ensembl.org

- launched in 1999
- vertebrates
- Ensembl gene annotation
- EBI and WTSI

www.ensemblgenomes.org

- launched in 2009
- non-vertebrates
- community gene annotation
- EBI

Bacteria | Protists | Fungi | Plants | Metazoa | Vertebrates

Ensembl features

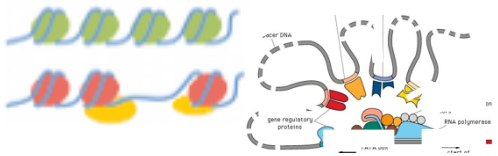
Gene models

Comparative Genomics

Variation

Tools

Regulation

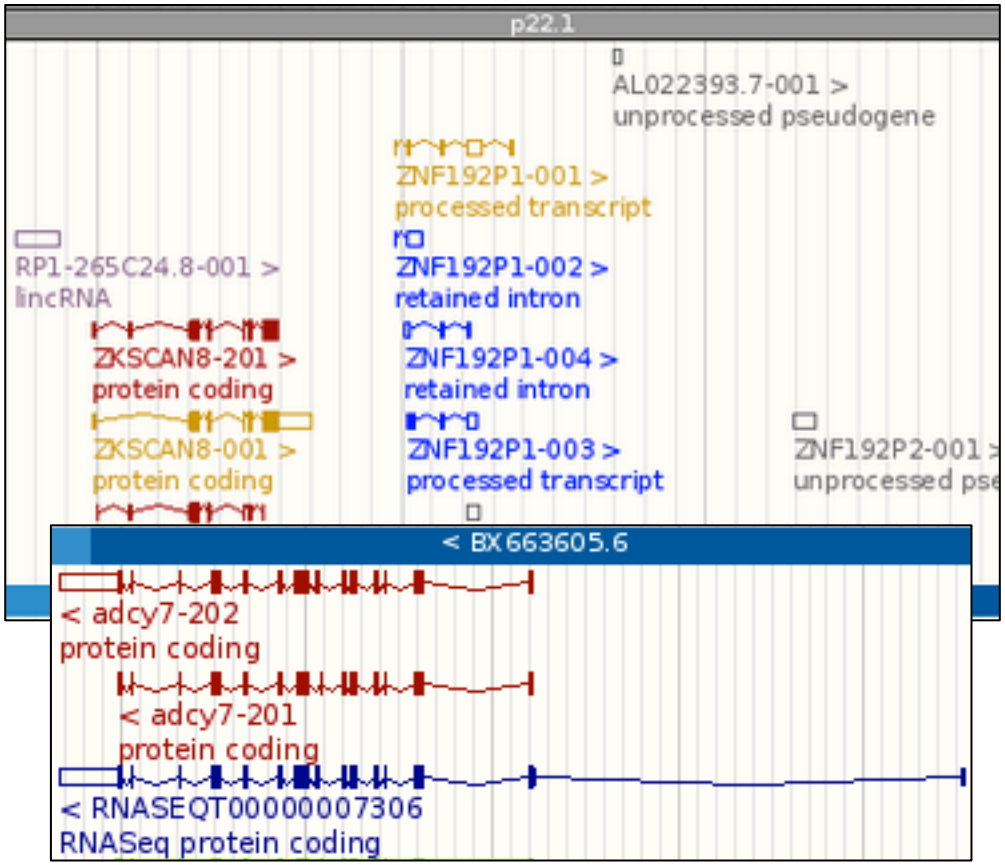


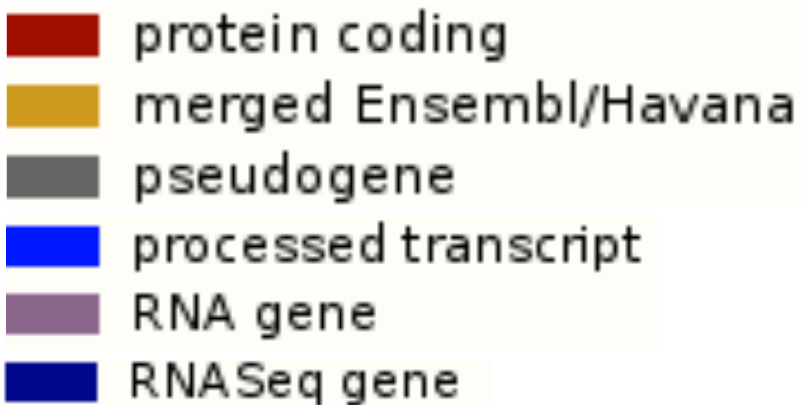
Programmatic access

Custom data display

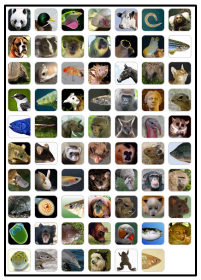
Free code & data

Gene models in Ensembl





Automatic

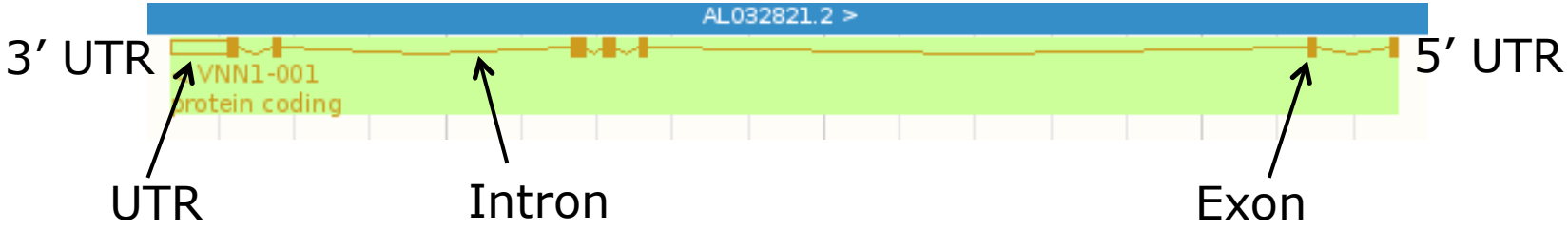


Manual

Goal: Generate set of well-supported genes

Ensembl genes & transcripts*

- merged annotation
- higher confidence and quality
- comprehensive: alternatively spliced transcripts



Gold (identical annotation) = Automatic + Manual

* based on experimental, biological evidence (INSDC, UniProtKB…)

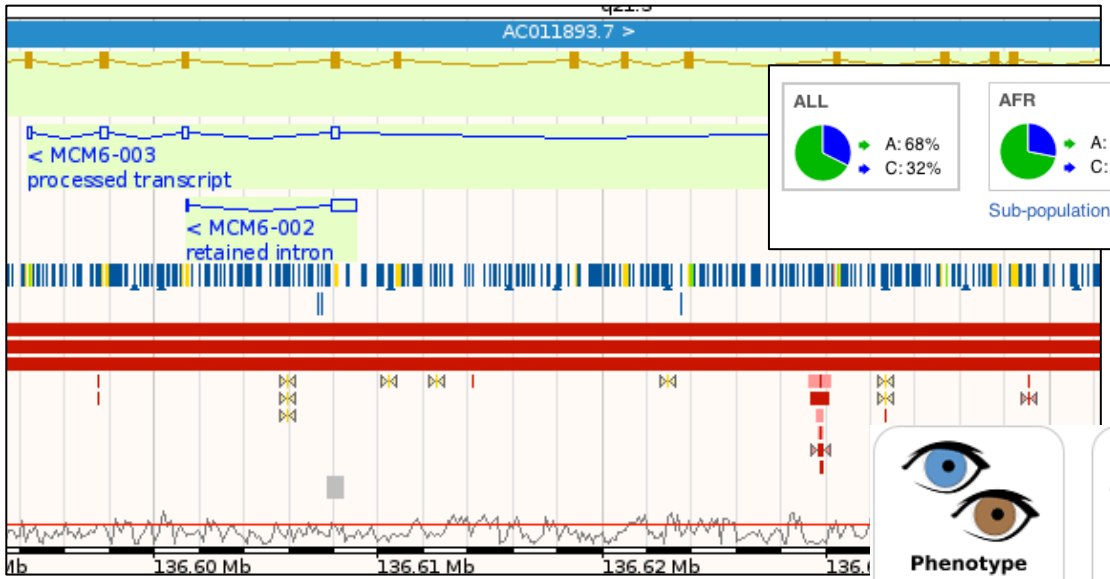
Ensembl stable identifiers

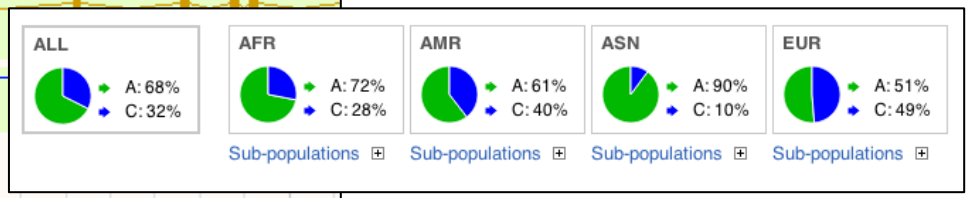
- ENSG########## Ensembl Gene ID
- ENST########## Ensembl Transcript ID
- ENSP########## Ensembl Peptide ID
- ENSE########## Ensembl Exon ID

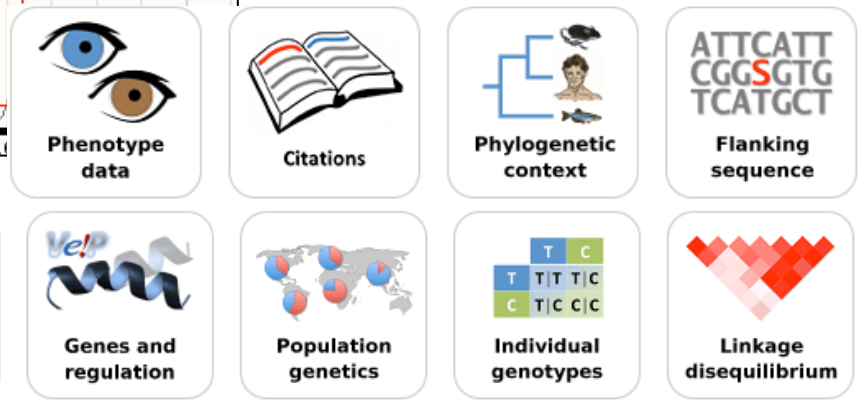
- For non-human species a suffix is added:

ENSMUSG	MUS (*Mus musculus*)	for mouse
ENSRNOG	RNO (*Rattus norvegicus*)	for rat

Genetic variation

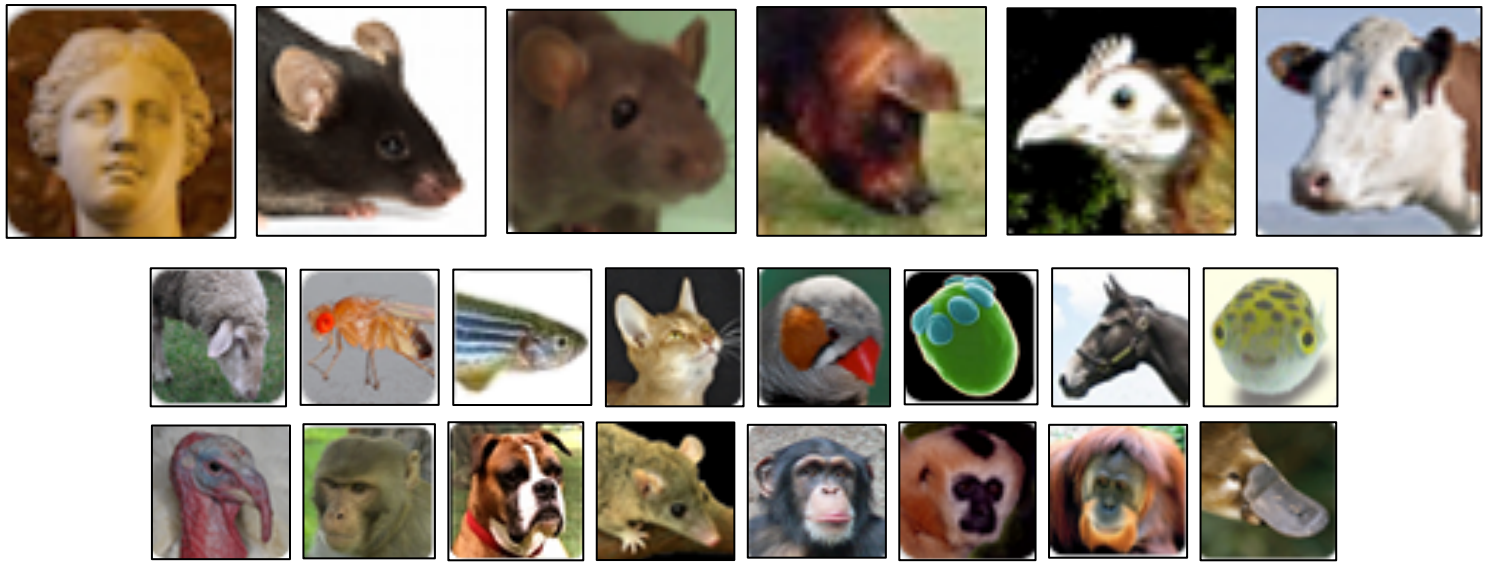






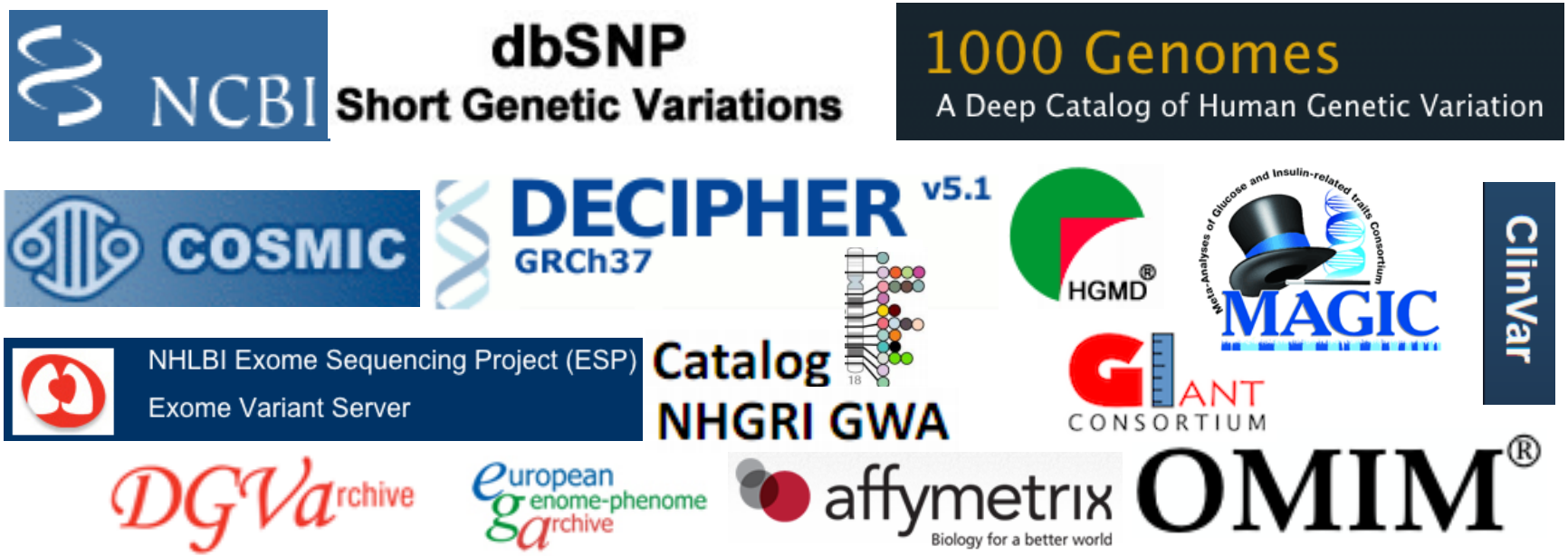
rs56404215			0.06	0.891	ENST00000380152
rs56400215		?	0.05	0.477	ENST00000380152
rs202155613		?	0.01	0.994	ENST00000380152

Species with variation data



Understand the types of genetic variation data and how to view them in the context of our genomes

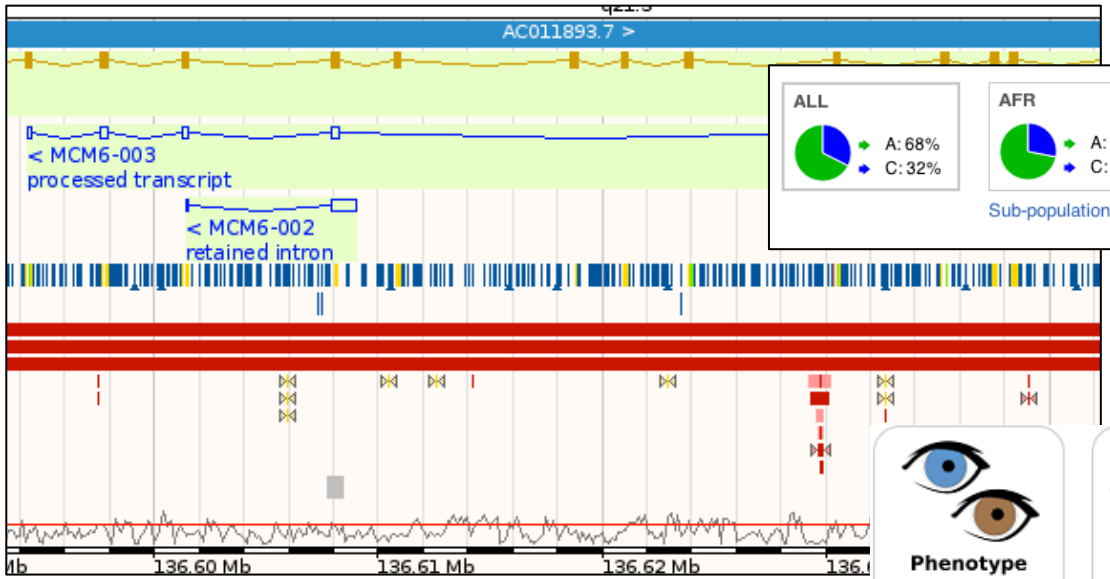
Sources of variation data

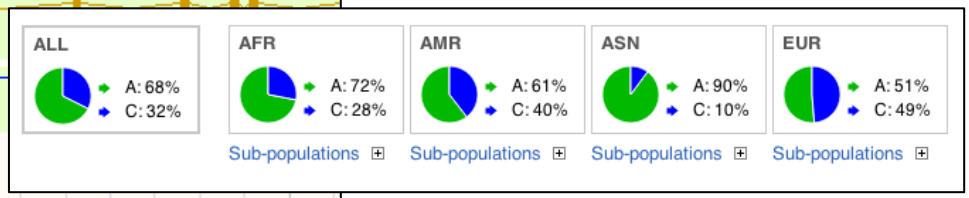


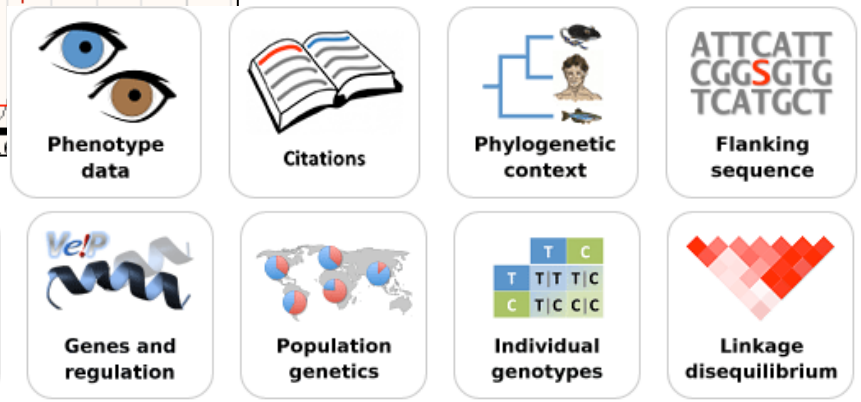
http://www.ensembl.org/info/docs/variation/sources_documentation.html

- Import alleles and frequencies
- Annotate variants VeP

Genetic variation



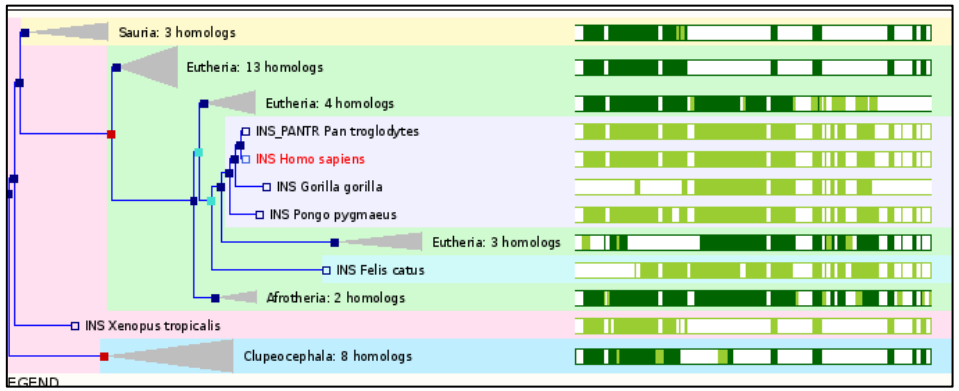


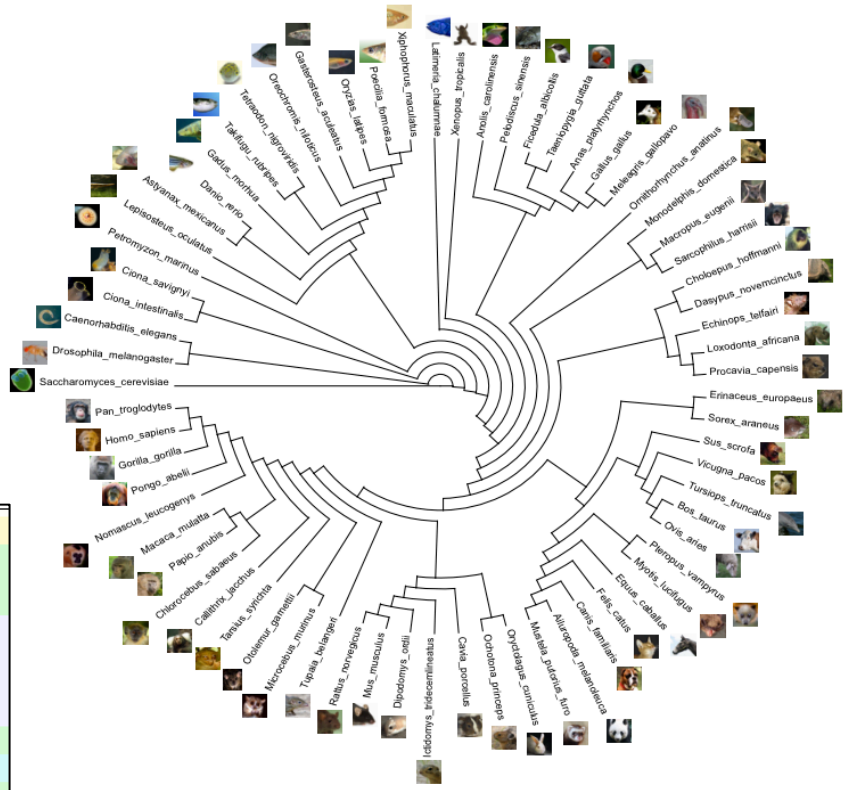


rs56404215			0.06	0.891	ENST00000380152
rs56400215		?	0.05	0.477	ENST00000380152
rs202155613		?	0.01	0.994	ENST00000380152

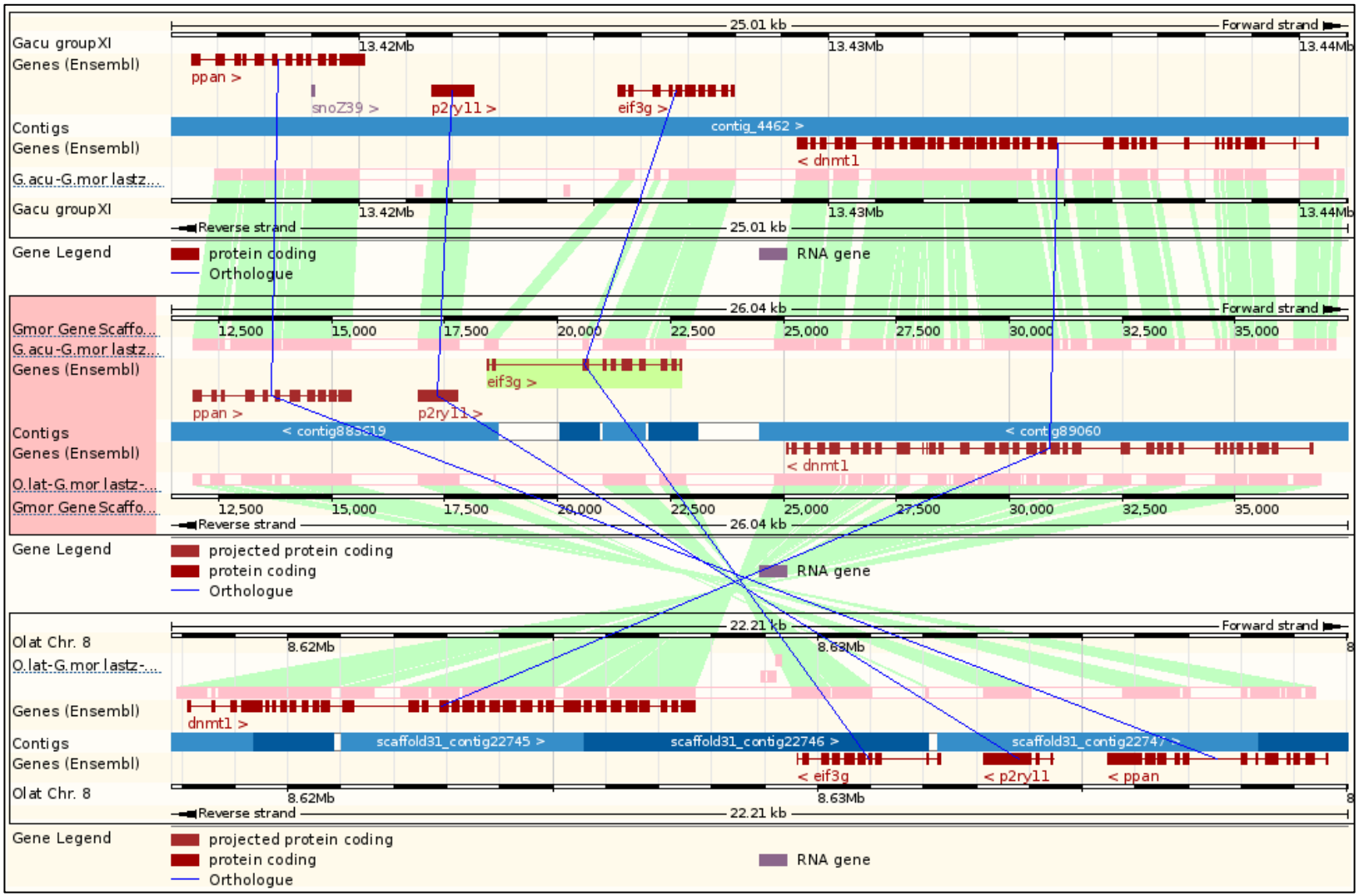
Comparative Genomics

- Gene Trees
- Orthologues/Paralogues
- Protein Families
- Whole Genome Alignments





Region Comparison



Gene Regulation

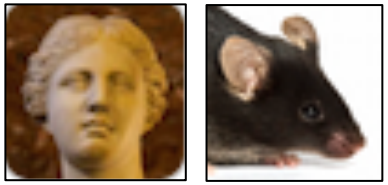
Goal: Annotate the genome with features that may play a role in the transcriptional regulation of genes

Multiple data sources: collection and summary

http://www.ensembl.org/info/docs/funcgen/regulation_sources.html
http://www.ensembl.org/Homo_sapiens/Experiment/

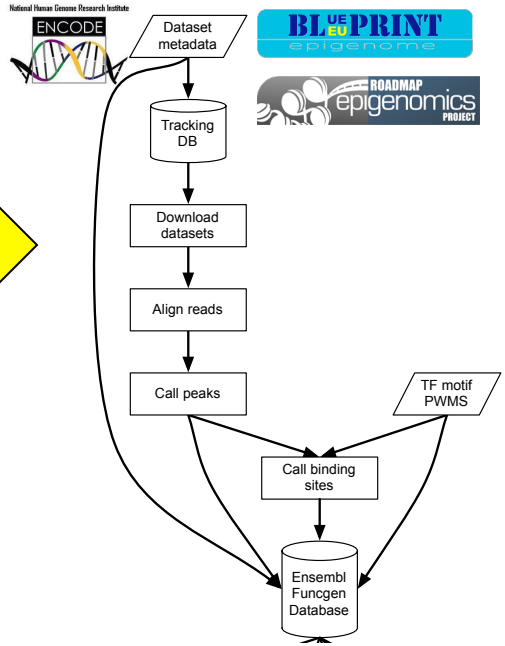
Ensembl Regulatory build

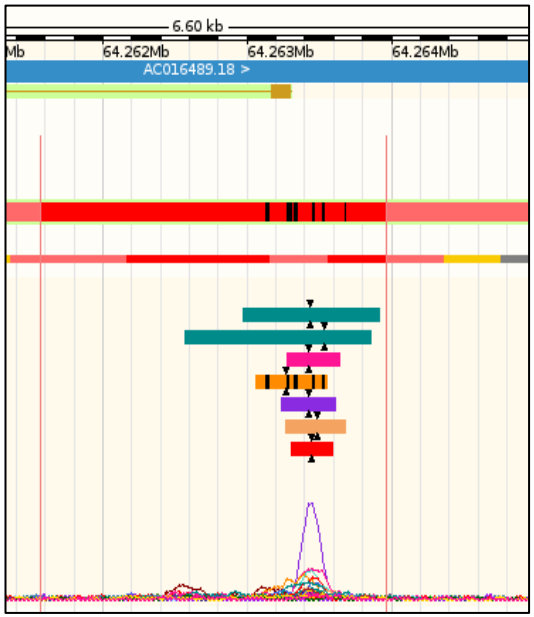
raw data → Ensembl pipeline → Ensembl annotation



CGCTT
GAACA
CGCTT GAACA
ACGTC ACGTC

ChIP-Seq
DNase1-Seq



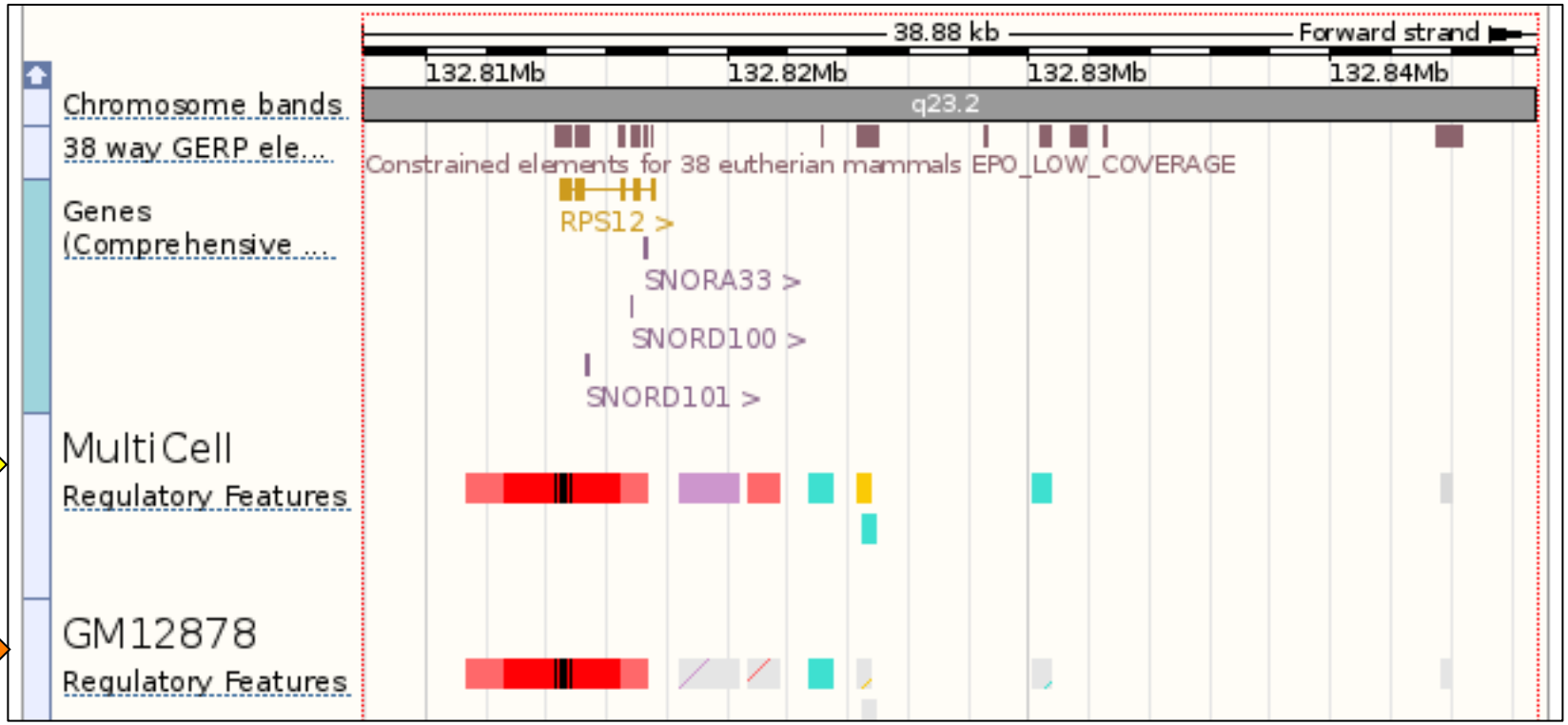


Biomart | R | Ensembl.org | Ensembl API | MySQL | REST

Regulatory features: view

For MultiCell and individual cell lines, e.g. GM12878

Configure this page → Regulation → Regulatory features



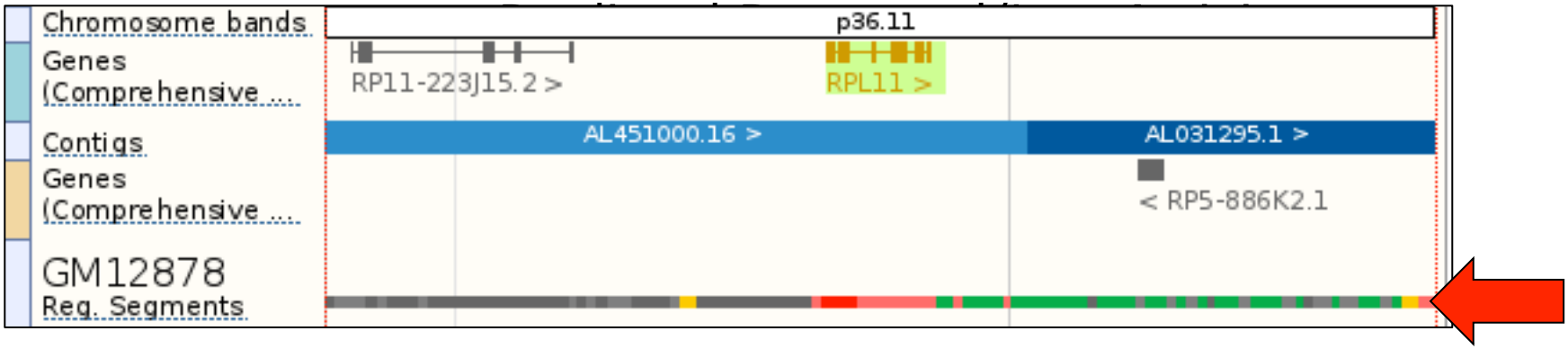
Segmentation data in Ensembl

ENCODE

17 cell types

categories of combined segments

- CTCF enriched
- Predicted Weak Enhancer/Cis-reg element
- Predicted Transcribed Region
- Predicted Enhancer
- Predicted Promoter Flank



Configure this page → Regulation → Regulatory features

Ensembl Browser

Live demo:

Walking through the website

pages 14-37

Before we start: background

The *ESPN* gene products are active in the inner ear, where it appears to play an essential role in normal hearing and balance.

Balanced levels of Espin are critical for stereociliary growth and length maintenance†

Agnieszka Rzadzinska[1,‡], Mark Schneider[1], Konrad Noben-Trauth[2], James R. Bartles[3], Bechara Kachar[1,*]

Article first published online: 3 OCT 2005
DOI: 10.1002/cm.20094

CYTOSKELETON

Journal of MEDICAL GENETICS

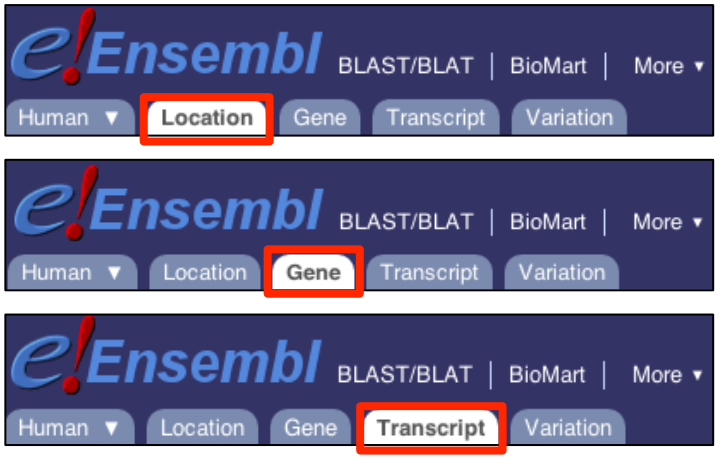
Current TOC | Instructions for authors

J Med Genet. 2006 February; 43(2): 157–161.

Espin gene (*ESPN*) mutations associated with autosomal dominant hearing loss cause defects in microvillar elongation or organisation

F Donaudy, L Zheng, R Ficarella, E Ballana, M Carella, S Melchionda, X Estivill, J R Bartles, and P Gasparini

Let's explore *ESPN*



Human ESPN: location

A) What is the location and strand of the human *ESPN* gene?

B) How can I view protein alignments and variants mapped to this location?

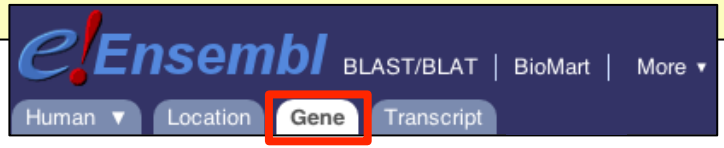
C) Can I move data tracks up and down, share and delete tracks?

Human ESPN: gene

A) How can I find the genomic sequence of this gene? What is the ID of its first exon?

B) Can I display the genomic coordinates and variants on this sequence?

C) Can I find the orthologues of this gene in other vertebrates?



Human ESPN: transcript

A) How many exons does the longest *ESPN* transcript have? Are there any completely untranslated exons?

B) Can I find its cDNA sequence?

C) What are the UniProt and RefSeq entries cross referenced to this transcript?

Ensembl Browser

Exercises
pages 38-40

Answers
www.ebi.ac.uk/~denise/workshops/2016/taiwan/sinica/answers

Feel free to explore your favourite gene/genome too!

Ensembl Tools: BioMart

EMBL-EBI

wellcome trust sanger institute

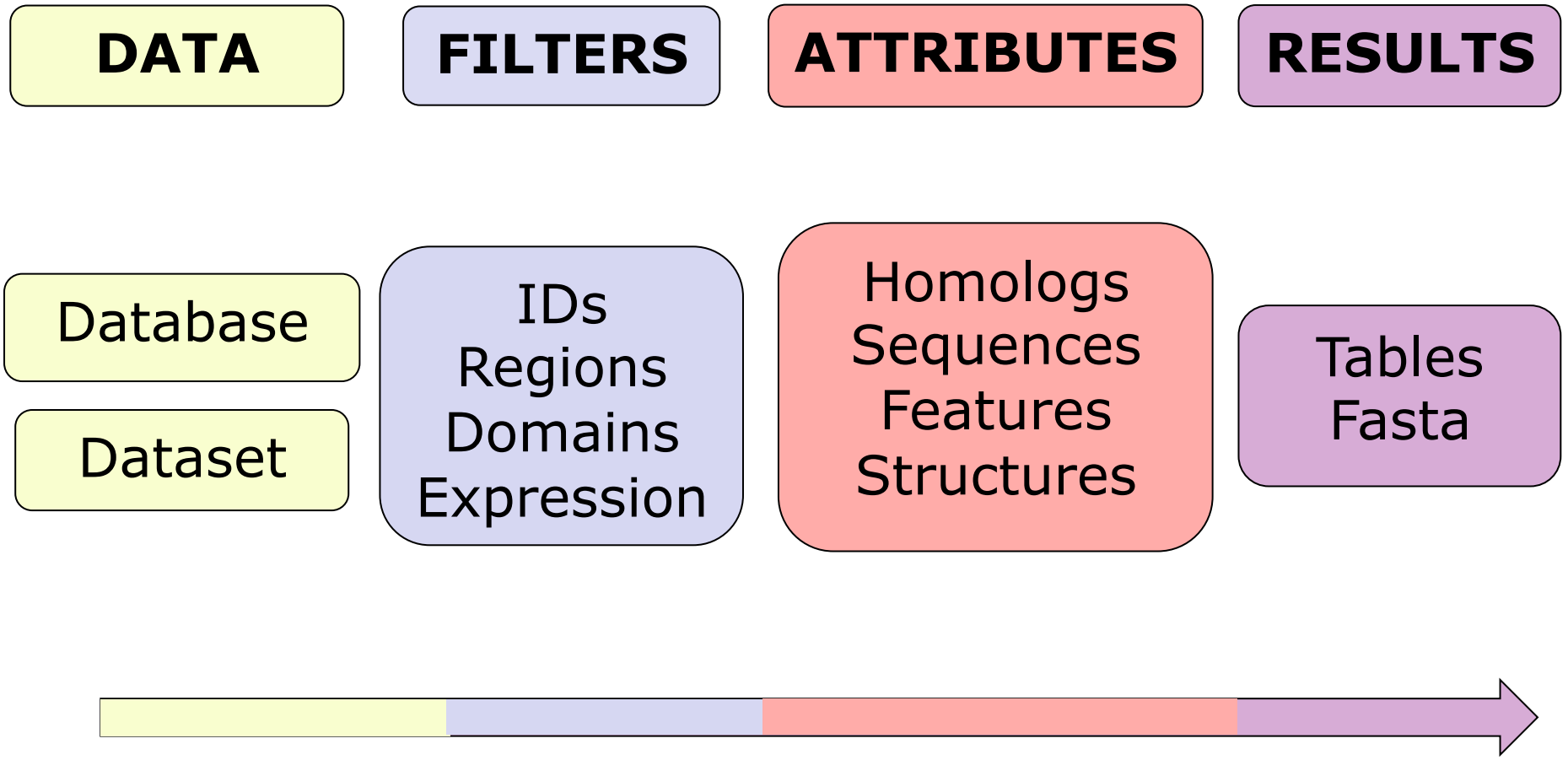
Outline

- Definitions
- The principle: 4 steps
- Tutorial: simple query in human
- Find Ensembl BioMart and BioMart elsewhere
- Sophisticated platforms: mart services, APIs, etc…
- Exercises

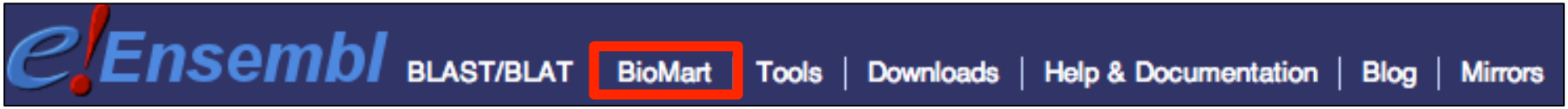
What is BioMart?

- Free service for easy retrieval of Ensembl data
- Data export tool with little/no programming required
- Complex queries with a few mouse clicks
- Output formats (.xls, .csv, fasta, tsv, html)

The four-step principle



Find BioMart



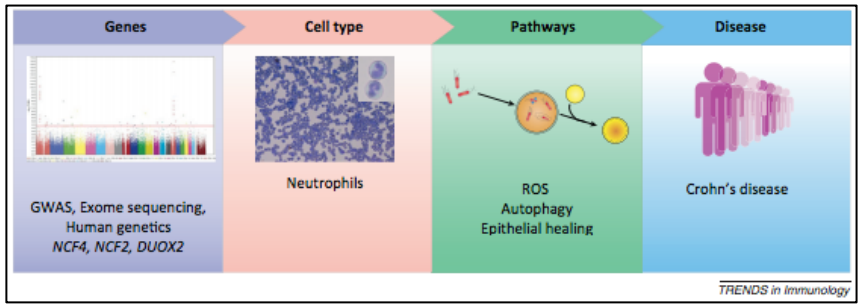
www.ensembl.org/biomart/martview

Review

From genetics of inflammatory bowel disease towards mechanistic insights

Daniel B. Graham[1,2] and Ramnik J. Xavier[1,3]

Advancements in human genetics now poise the field to illuminate the pathophysiology of complex genetic disease. In particular, genome-wide association studies (GWAS) have generated insights into the mechanisms driving inflammatory bowel disease (IBD) and implicated genes shared by multiple autoimmune and autoinflammatory diseases. Thus, emerging evidence suggests a central role for the mucosal immune system in mediating immune homeostasis and highlights the complexity of genetic and environmental interactions that collectively modulate the risk of disease. Nevertheless, the challenge remains to determine how genetic variation can precipitate and sustain the inappropriate inflammatory response to commensals that is observed in IBD. Here, we highlight recent advancements in immunogenetics and provide a forward-looking view of the innovations that will deliver mechanistic insights from human genetics.



Selected IBD genes

IL23R, *PTPN22*,
CUL2, *C1orf106*, *IL18RAP*

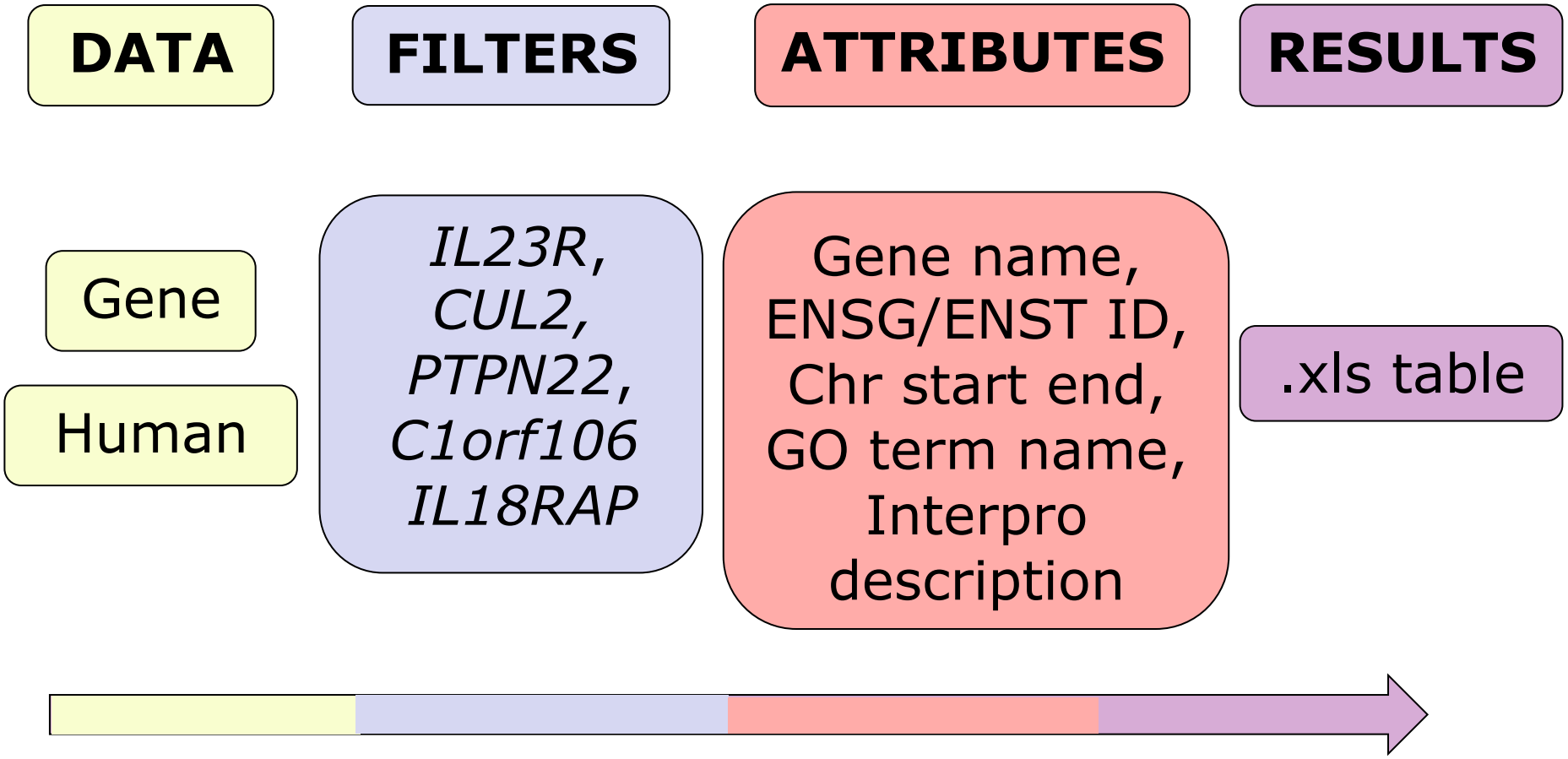
Trends in Immunology August 2013, Vol. 34, No. 8

Tutorial: BioMart

For the *IL23R*, *PTPN22*, *CUL2, C1orf106* and *IL18RAP genes,* use BioMart to retrieve a table (.xls) containing:

- Associated gene name, ENSG and ENST IDs
- Chromosome name, gene start and end
- GO term name and Interpro description

The four-step principle



Ensembl BioMart

Live demo

Ensembl BioMarts

DATABASE
The Journal of Biological Databases and Curation

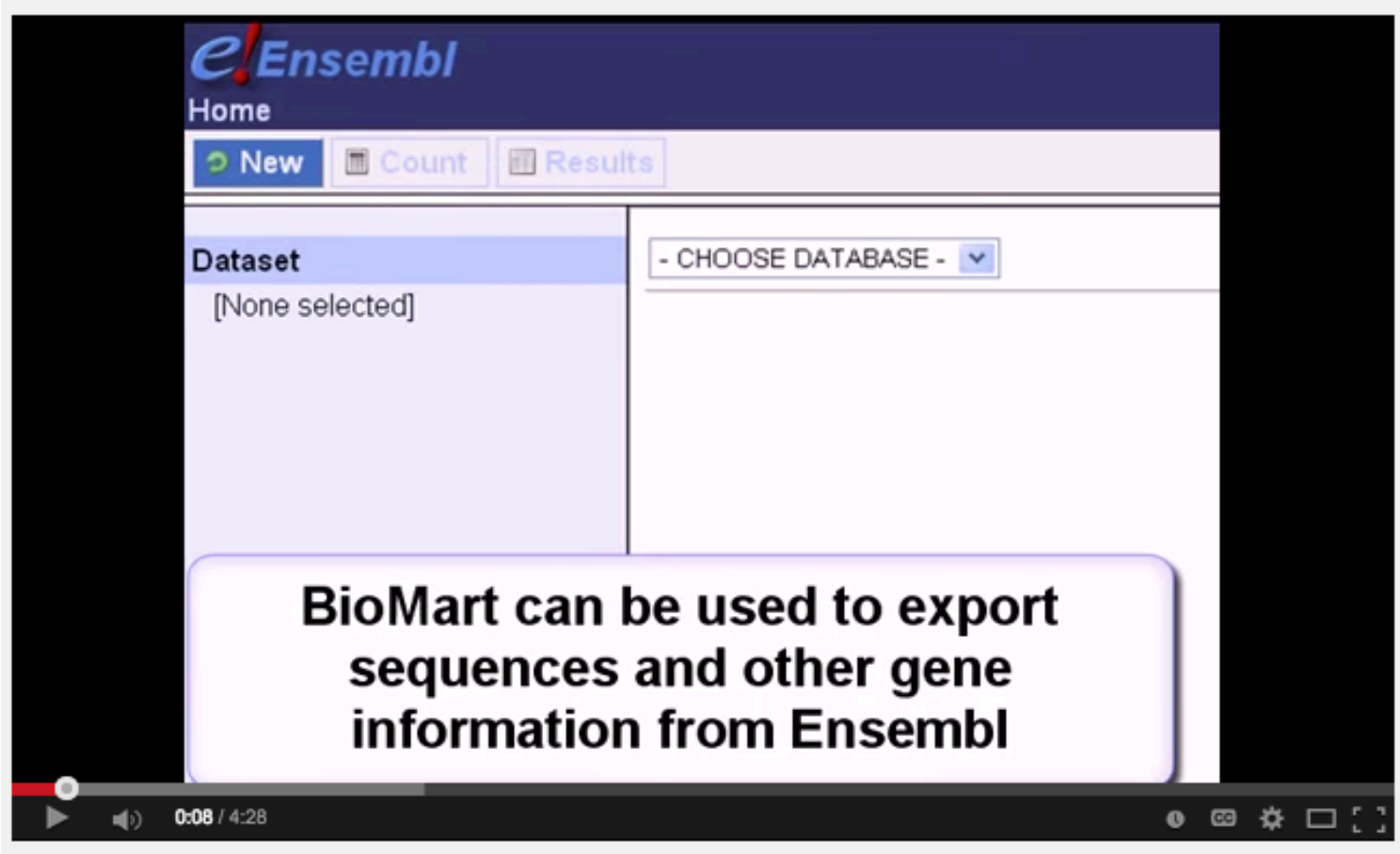
Database, Vol. 2011, Article ID bar030, doi:10.1093/database/bar030

Original article

Ensembl BioMarts: a hub for data retrieval across taxonomic space

Rhoda J. Kinsella[1,*], Andreas Kähäri[1], Syed Haider[2], Jorge Zamora[1], Glenn Proctor[1], Giulietta Spudich[1], Jeff Almeida-King[1], Daniel Staines[1], Paul Derwent[1], Arnaud Kerhornou[1], Paul Kersey[1] and Paul Flicek[1,*]

BioMart YouTube video



http://tinyurl.com/video-biomart

More sophisticated platforms

- BioMart 

queries: MartService

  www.biomart.org/martservice.html

- APIs: PERL, Java, Web Services
- Third party softwares

taverna.org.uk

biomaRt

bioconductor.org

galaxyproject.org

Ensembl Tools: BioMart

Step-by-step example
pages 45-49

Exercises
pages 50-52

Answers
www.ebi.ac.uk/~denise/workshops/2016/taiwan/sinica/answers

Feel free to explore BioMart in other contexts too!

Ensembl Tools: The VEP

EMBL-EBI

wellcome trust sanger institute

Annotating your own variants

- Variant Effect Predictor
- Different input formats
- SIFT/PolyPhen for missense variants

BIOINFORMATICS APPLICATIONS NOTE Vol. 26 no. 16 2010, pages 2069–2070
doi:10.1093/bioinformatics/btq330

Databases and ontologies Advance Access publication June 18, 2010

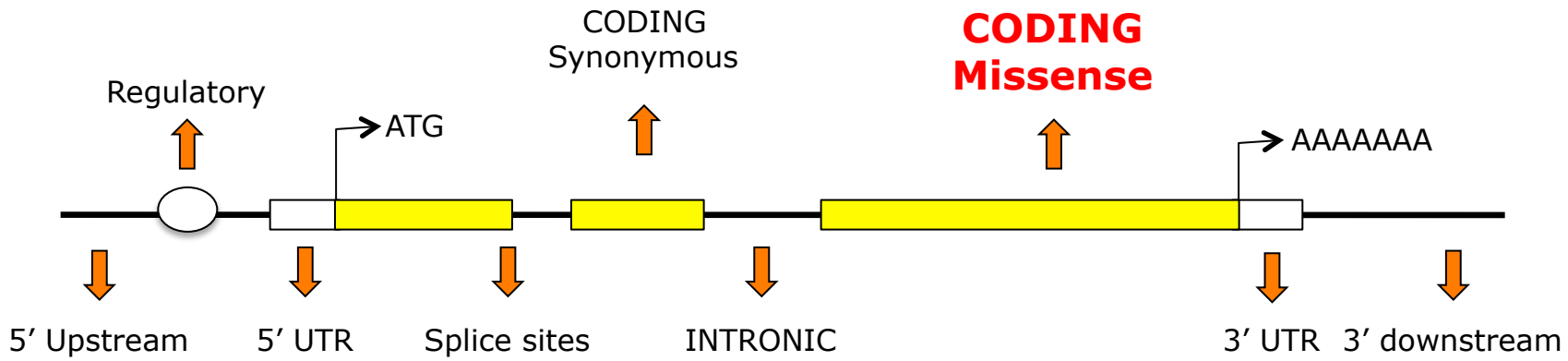
Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor

William McLaren[1,*], Bethan Pritchard[2], Daniel Rios[1], Yuan Chen[1], Paul Flicek[1] and Fiona Cunningham[1,*]

[1]European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SD and [2]Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SA, UK

PMID: 20562413

Mapping variants on transcripts



Identify transcripts that overlap variants and predict the consequence of these on Ensembl (or RefSeq) transcripts using VeP

Consequence terms for variants

* SO term	SO description
transcript_ablation	A feature ablation whereby the deleted region includes a transcript feature
splice_donor_variant	A splice variant that changes the 2 base region at the 5' end of an intron
splice_acceptor_variant	A splice variant that changes the 2 base region at the 3' end of an intron
stop_gained	A sequence variant whereby at least one base of a codon is changed, resulting in a premature stop codon, leading to a shortened transcript
frameshift_variant	A sequence variant which causes a disruption of the translational reading frame, because the number of nucleotides inserted or deleted is not a multiple of three
stop_lost	A sequence variant where at least one base of the terminator codon (stop) is changed, resulting in an elongated transcript
initiator_codon_variant	A codon variant that changes at least one base of the first codon of a transcript
inframe_insertion	An inframe non synonymous variant that inserts bases into in the coding sequence
inframe_deletion	An inframe non synonymous variant that deletes bases from the coding sequence
missense_variant	A sequence variant, where the change may be longer than 3 bases, and at least one base of a codon is changed resulting in a codon that encodes for a different amino acid
transcript_amplification	A feature amplification of a region containing a transcript
splice_region_variant	A sequence variant in which a change has occurred within the region of the splice site, either within 1-3 bases of the exon or 3-8 bases of the intron
incomplete_terminal_codon_variant	A sequence variant where at least one base of the final codon of an incompletely annotated transcript is changed
synonymous_variant	A sequence variant where there is no resulting change to the encoded amino acid
stop_retained_variant	A sequence variant where at least one base in the terminator codon is changed, but the terminator remains
coding_sequence_variant	A sequence variant that changes the coding sequence
mature_miRNA_variant	A transcript variant located with the sequence of the mature miRNA
5_prime_UTR_variant	A UTR variant of the 5' UTR
3_prime_UTR_variant	A UTR variant of the 3' UTR
intron_variant	A transcript variant occurring within an intron
NMD_transcript_variant	A variant in a transcript that is the target of NMD
non_coding_exon_variant	A sequence variant that changes non-coding exon sequence
nc_transcript_variant	A transcript variant of a non coding RNA
upstream_gene_variant	A sequence variant located 5' of a gene
downstream_gene_variant	A sequence variant located 3' of a gene

http://www.ensembl.org/info/genome/variation/predicted_data.html#consequence_type_table

* defined by the Sequence Ontology (SO) project (http://www.sequenceontology.org/)

Consequence: missense

GAG >GGG
Glu > Gly

Show 10 | Show/hide columns | Filter

ID			Global MAF	Class	Source	Evidence	Type	AA	AA coord	SIFT	Poly-Phen	Transcript
rs80359214			-	SNP	dbSNP	-	Missense variant	W/C	31	0	0.103	ENST00000380152
rs80359214			-	SNP	dbSNP	-	Missense variant	W/C	31	0	0.103	ENST00000544455
rs397508057	13:32893240	T/C	-	SNP	dbSNP_ClinVar	-	Missense variant	F/L	32	0	0.775	ENST00000380152
rs397508057	13:32893240	T/C	-	SNP	dbSNP_ClinVar	-	Missense variant	F/L	32	0	0.775	ENST00000544455
rs80358415	13:32893267	C/T	-	SNP	dbSNP	-	Missense variant	P/S	41	0.01	0.39	ENST00000380152
rs80358415	13:32893267	C/T	-	SNP	dbSNP	-	Missense variant	P/S	41	0.01	0.39	ENST00000544455
rs4987046	13:32893271	A/G	0.057 (G)	SNP	dbSNP		Missense variant	Y/C	42	0.11	0.032	ENST00000380152
rs4987046	13:32893271	A/G	0.057 (G)	SNP	dbSNP		Missense variant	Y/C	42	0.11	0.032	ENST00000544455
rs80358425	13:32893282	C/T	-	SNP	dbSNP	-	Missense variant	P/S	46	0.71	0.801	ENST00000380152
rs80358425	13:32893282	C/T	-	SNP	dbSNP	-	Missense variant	P/S	46	0.71	0.801	ENST00000544455

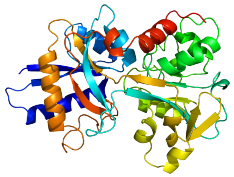
Showing 31 to 40 of 3,698 entries | << < 2 3 4 5 6 > >>

- SIFT

sift.jcvi.org/

- PolyPhen-2

genetics.bwh.harvard.edu/pph2/



SIFT predictions: Prediction only

PolyPhen predictions:

Frequency filtering of existing variants (human only)

- ✓ No
- Prediction only
- Score only
- Prediction and score

VeP

Ensembl tools

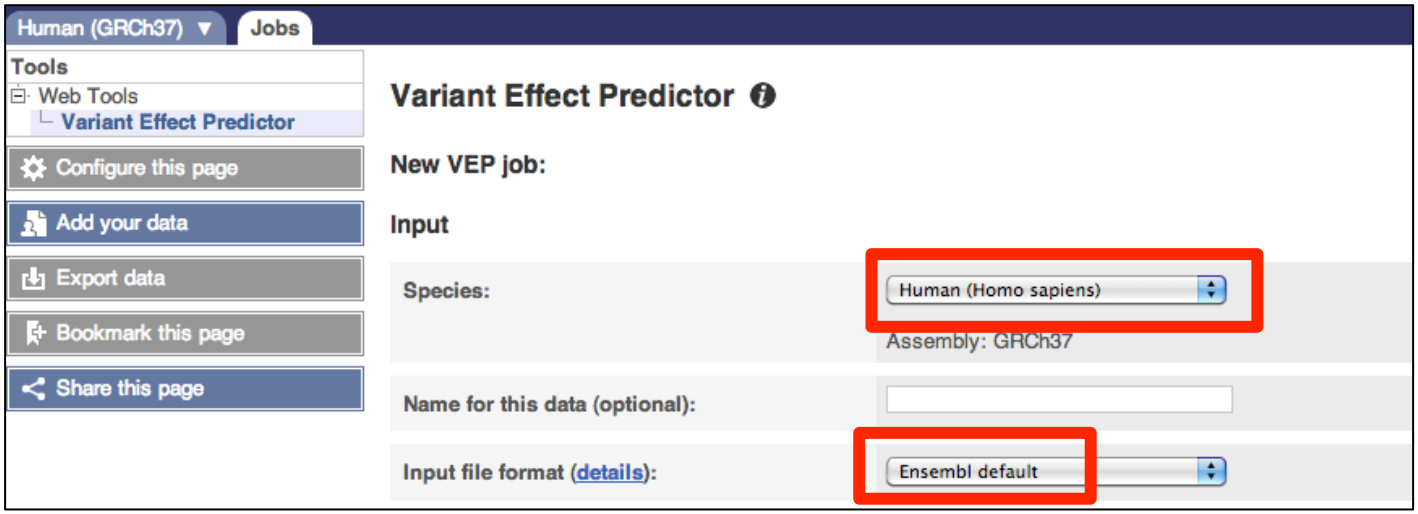
http://www.ensembl.org/tools.html

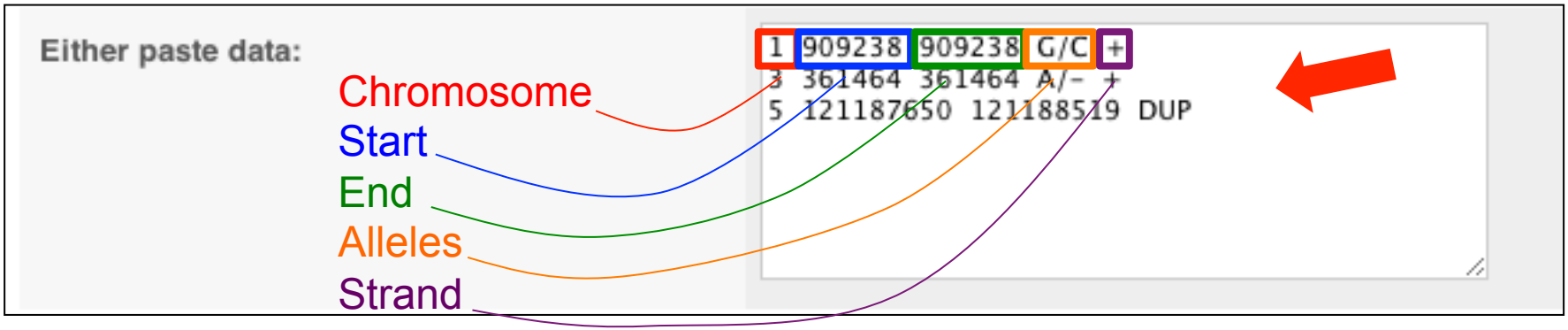
Name	Description	Online tool	Download code	Documentation
Variant Effect Predictor VeP	Analyse your own variants and predict the functional consequences of known and unknown variants via our Variant Effect Predictor (VEP) tool.			
BLAST/BLAT	Search our genomes for your DNA or protein sequence.			
BioMart	Use this data-mining tool to export custom datasets from Ensembl.			
Assembly converter	Map (liftover) your data's coordinates to the current assembly.			
ID History converter	Convert a set of Ensembl IDs from a previous release into their current equivalents.			
Ensembl Virtual Machine	VirtualBox virtual Machine with Ubuntu desktop and pre-configured with the latest Ensembl API plus Variant Effect Predictor (VEP). **NB: download is >1 GB**	-		

http://www.ensembl.org/vep

Launch VeP

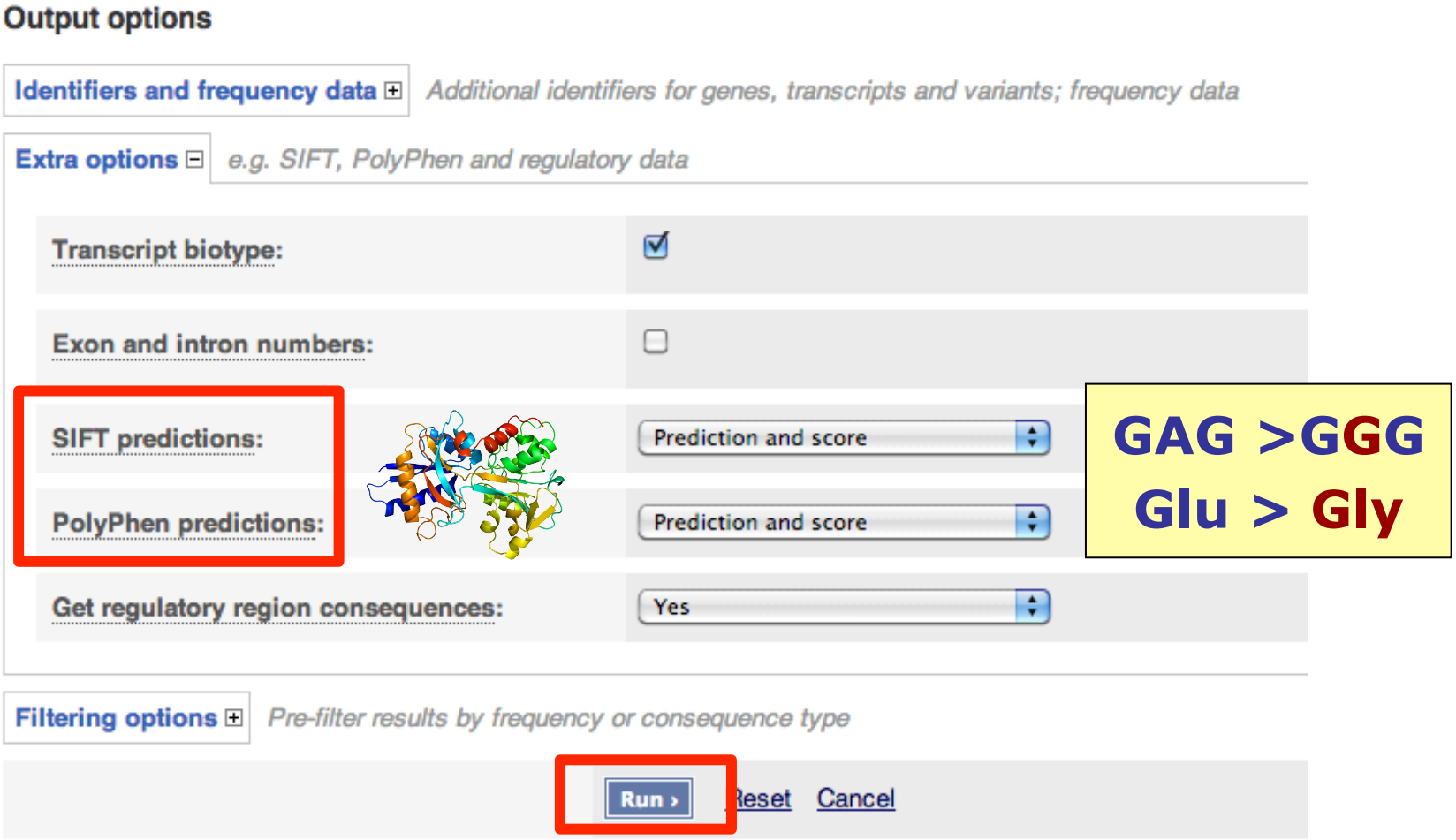
Inputting data into VeP



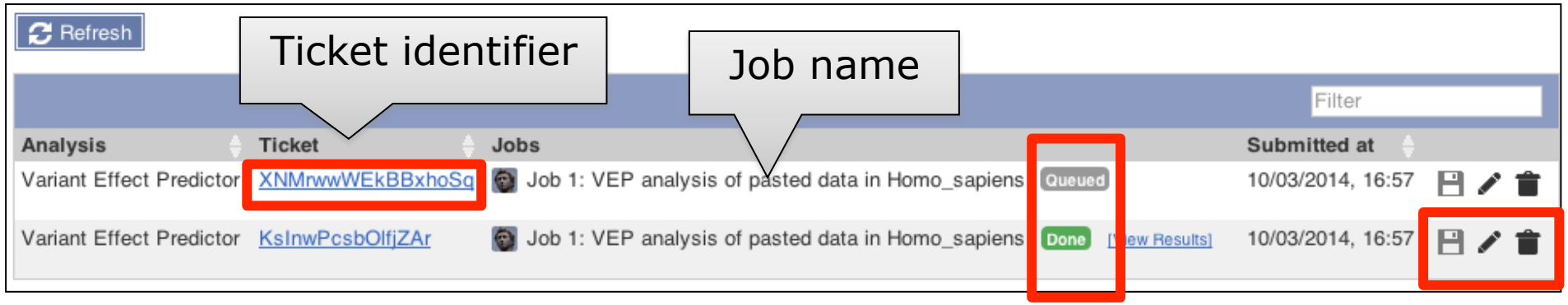


Output options in VeP

http://www.ensembl.org/info/docs/variation/vep/index.html



Ticket system in VeP



Save to your account (log in)

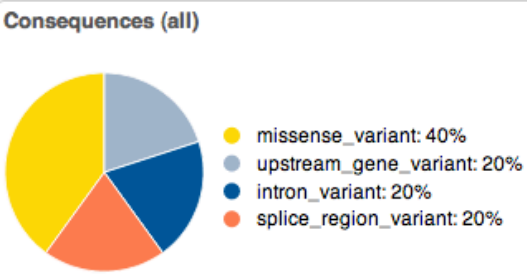
Edit and resubmit your job

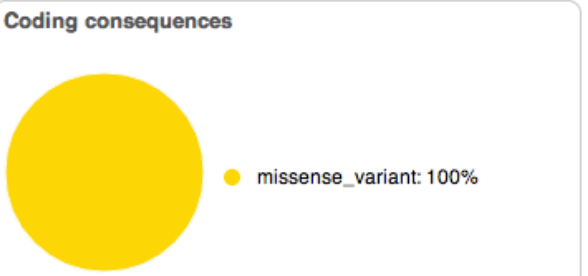
Delete job

Queued
Running
Done
Failed

Viewing VeP results

Category	Count
Variants processed	4
Variants remaining after filtering	4
Novel / existing variants	-
Overlapped genes	2
Overlapped transcripts	2
Overlapped regulatory features	-





SO consequence terms*

Edit & resubmit

Results preview

Navigation — Showing 4 results for variants 1-4 of 4 | **Show** 1 All |

Filters — Uploaded variation | is | defined | Add

Download — All VCF VEP TXT

Show/hide columns

Uploaded variation	Location	Allele	Gene	Feature	Consequence	Protein position	Amino acids	Codons	Symbol	Symbol source	SIFT
2_83820006_T/C	2:83820006	C	ENSGALG00000013135	ENSGALT00000021443	intron_variant	-	-	-	GALNT1	Uniprot_gn	-
2_83896185_T/-	2:83896184-83896185	-	ENSGALG00000013135	ENSGALT00000021443	upstream_gene_variant	-	-	-	GALNT1	Uniprot_gn	-
7_19676386_G/A	7:19676386	A	ENSGALG00000020703	ENSGALT00000033140	missense_variant	34	A/T	GCT/ACT	GRB14	HGNC	1
7_19684533_G/A	7:19684533	A	ENSGALG00000020703	ENSGALT00000033140	missense_variant,	115	G/S	GGC/AGC	GRB14	HGNC	0.87

*http://www.sequenceontology.org/index.html

Viewing VeP results

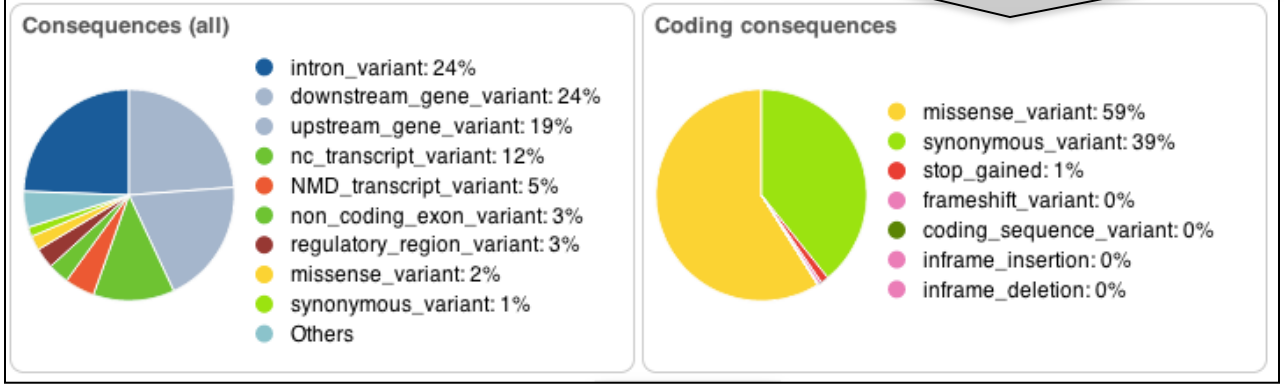
Variant Effect Predictor results

Summary statistics for ticket uTPZdaCscZ7TeOiX:

Category	Count
Variants processed	19976
Variants remaining after filtering	19976
Novel / existing variants	914 (4.6%) / 19062 (95.4%)
Overlapped genes	116
Overlapped transcripts	428
Overlapped regulatory features	373

Table
- Before / after filtering
- novel / existing variants

Pie charts (consequence terms)
- total observed (more than one per variant)
- Separate chart: coding consequences



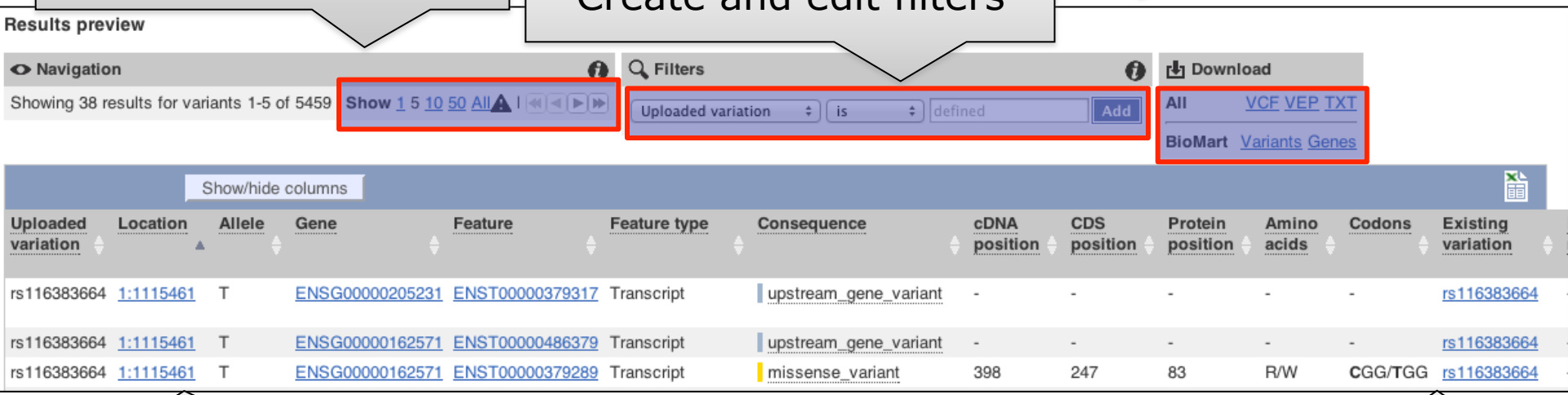
ensembl.org/info/docs/tools/vep/online/results.html#summary

VeP results table

Navigate results (one row per variant/ transcript overlap)

- Download results
- Send results to BioMart

Create and edit filters



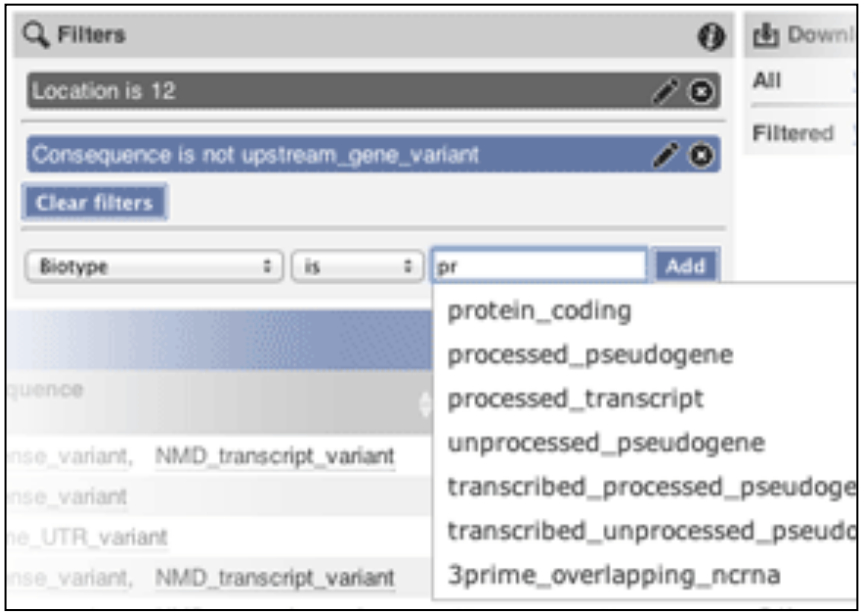
Uploaded variation	Location	Allele	Gene	Feature	Feature type	Consequence	cDNA position	CDS position	Protein position	Amino acids	Codons	Existing variation
rs116383664	1:1115461	T	ENSG00000205231	ENST00000379317	Transcript	upstream_gene_variant	-	-	-	-	-	rs116383664
rs116383664	1:1115461	T	ENSG00000162571	ENST00000486379	Transcript	upstream_gene_variant	-	-	-	-	-	rs116383664
rs116383664	1:1115461	T	ENSG00000162571	ENST00000379289	Transcript	missense_variant	398	247	83	R/W	CGG/TGG	rs116383664

Show/hide columns in results table

more columns: scroll right

ensembl.org/info/docs/tools/vep/online/results.html#table

Filtering VeP results



Filters consist of three components

Field

- e.g. Consequence, biotype

Operator

- e.g. is, matches (partial string matches)

Value

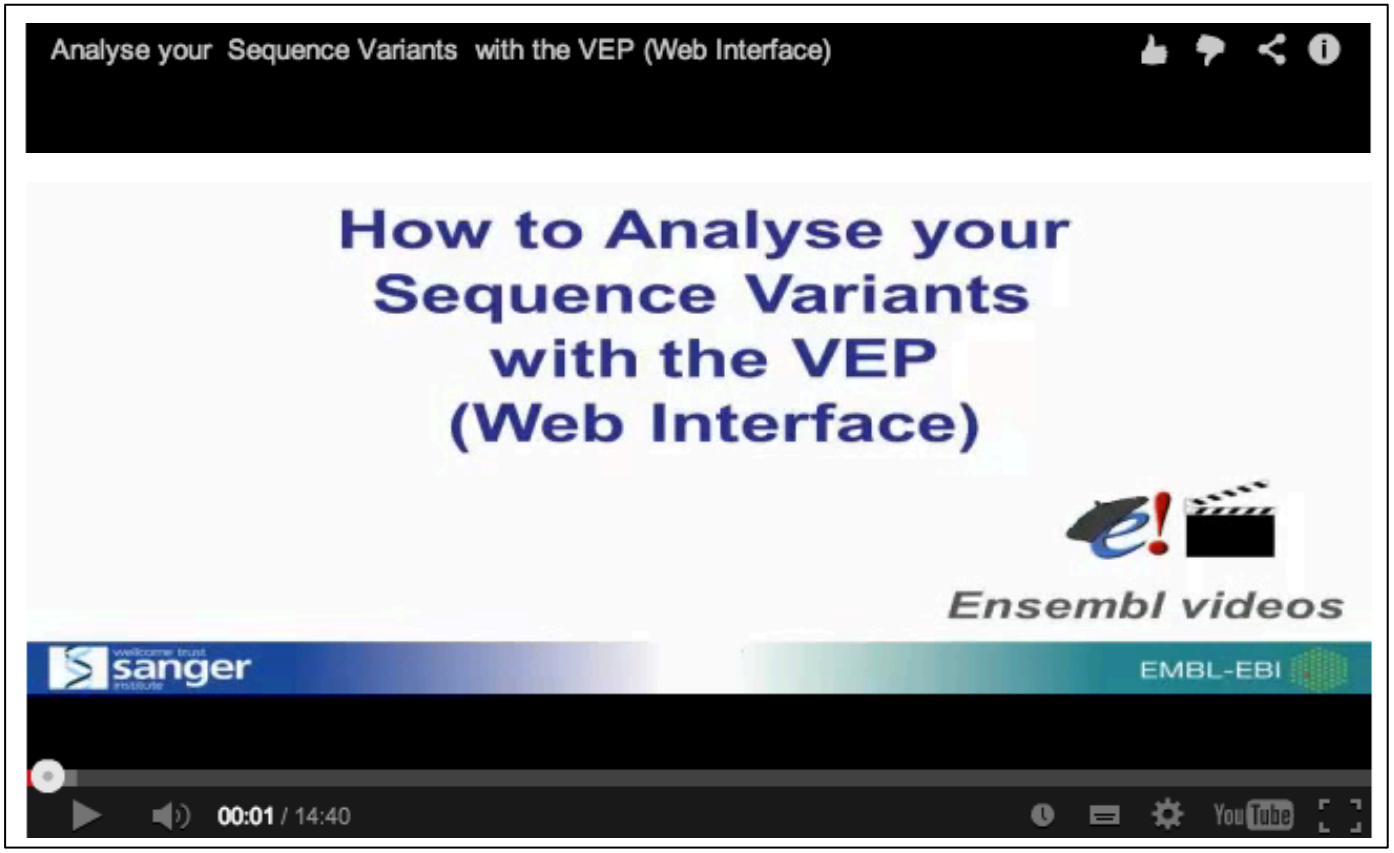
- the value to compare against
- some fields have autocomplete values

Multiple filters allowed with logical relationship (AND, OR)

Active filters can be edited too!

ensembl.org/info/docs/tools/vep/online/results.html#filter

VEP YouTube video



http://tinyurl.com/vep-video

Ensembl Tools: VEP

Exercises
pages 53-54

Answers
www.ebi.ac.uk/~denise/workshops/2016/taiwan/sinica/answers

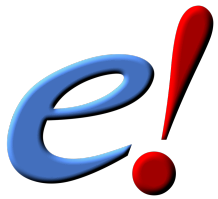
Feel free to explore your favourite variant/phenotype too!

Wrap up

Ensembl is the place!

Genes, genomes, variants, tools and more

Biomart	R	Ensembl.org	Ensembl API	MySQL	REST



Oh Yes!

And all is 100% free

Ensembl Retreat June 2015

Latest publication

Nucleic Acids Research

Ensembl 2016

Volume 44, Issue D1 › Pp. D710-D716.

Andrew Yates[1], Wasiu Akanni[1], M. Ridwan Amode[1], Daniel Barrell[1,2], Konstantinos Billis[1], Denise Carvalho-Silva[1], Carla Cummins[1], Peter Clapham[2], Stephen Fitzgerald[1], Laurent Gil[1], Carlos García Girón[1], Leo Gordon[1], Thibaut Hourlier[1], Sarah E. Hunt[1], Sophie H. Janacek[1], Nathan Johnson[1], Thomas Juettemann[1], Stephen Keenan[1], Ilias Lavidas[1], Fergal J. Martin[1], Thomas Maurel[1], William McLaren[1], Daniel N. Murphy[1], Rishi Nag[1], Michael Nuhn[1], Anne Parker[1], Mateus Patricio[1], Miguel Pignatelli[1], Matthew Rahtz[2], Harpreet Singh Riat[1], Daniel Sheppard[1], Kieron Taylor[1], Anja Thormann[1], Alessandro Vullo[1], Steven P. Wilder[1], Amonida Zadissa[1], Ewan Birney[1], Jennifer Harrow[2], Matthieu Muffato[1], Emily Perry[1], Magali Ruffier[1], Giulietta Spudich[1], Stephen J. Trevanion[1], Fiona Cunningham[1], Bronwen L. Aken[1], Daniel R. Zerbino[1] and Paul Flicek[1,2,*]

*To whom correspondence should be addressed. Tel: +44 1223 492581; Fax: +44 1223 494494 Email: flicek@ebi.ac.uk

Acknowledgements

Ensembl 2016

Andrew Yates[1], Wasiu Akanni[1], M. Ridwan Amode[1], Daniel Barrell[1,2], Konstantinos Billis[1], Denise Carvalho-Silva[1], Carla Cummins[1], Peter Clapham[2], Stephen Fitzgerald[1], Laurent Gil[1], Carlos García Girón[1], Leo Gordon[1], Thibaut Hourlier[1], Sarah E. Hunt[1], Sophie H. Janacek[1], Nathan Johnson[1], Thomas Juettemann[1], Stephen Keenan[1], Ilias Lavidas[1], Fergal J. Martin[1], Thomas Maurel[1], William McLaren[1], Daniel N. Murphy[1], Rishi Nag[1], Michael Nuhn[1], Anne Parker[1], Mateus Patricio[1], Miguel Pignatelli[1], Matthew Rahtz[2], Harpreet Singh Riat[1], Daniel Sheppard[1], Kieron Taylor[1], Anja Thormann[1], Alessandro Vullo[1], Steven P. Wilder[1], Amonida Zadissa[1], Ewan Birney[1], Jennifer Harrow[2], Matthieu Muffato[1], Emily Perry[1], Magali Ruffier[1], Giulietta Spudich[1], Stephen J. Trevanion[1], Fiona Cunningham[1], Bronwen L. Aken[1], Daniel R. Zerbino[1] and Paul Flicek[1,2,*]

[1]European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SD, UK and [2]Wellcome Trust Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SA, UK

Funding

wellcome trust

National Human Genome Research Institute

CTTV Centre for Therapeutic Target Validation

BLUEPRINT epigenome

EMBL

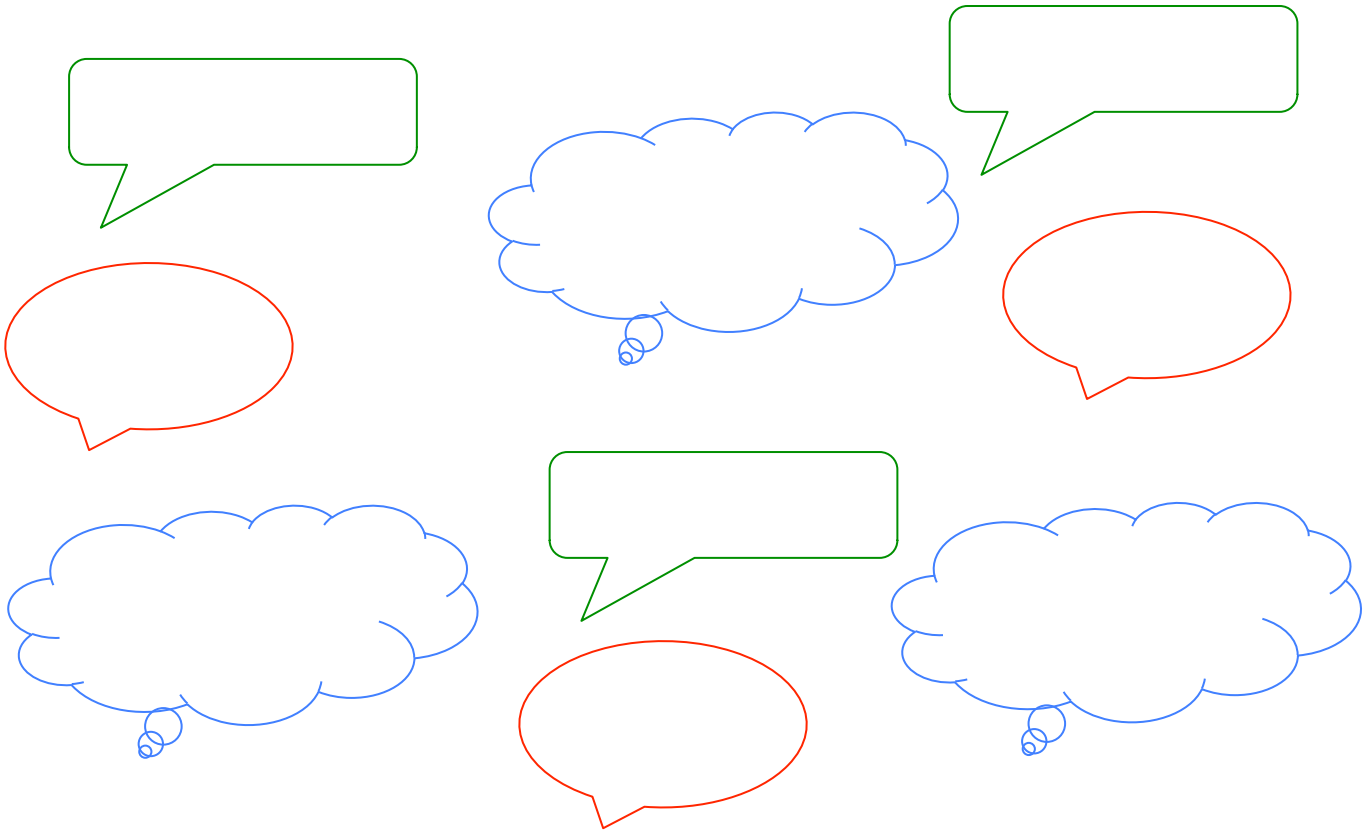
BBSRC bioscience for the future

BioMedBridges

MedBioinformatics

Co-funded by the European Union

Your take home message



Feedback survey

http://tinyurl.com/taiwan220216

Connect with Ensembl

helpdesk@ensembl.org

https://www.youtube.com/user/EnsemblHelpdesk

Training materials

If you use Ensembl for your work, please cite our papers

http://www.ensembl.org/info/about/publications.html