# Ensembl Genome Browser Workshop

## Liverpool, United Kingdom

**Aleena Mushtaq**

**Ensembl, EMBL-EBI**

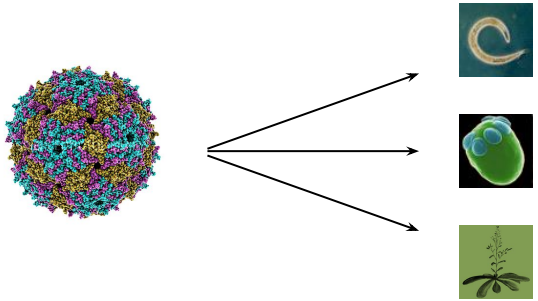
July 2025

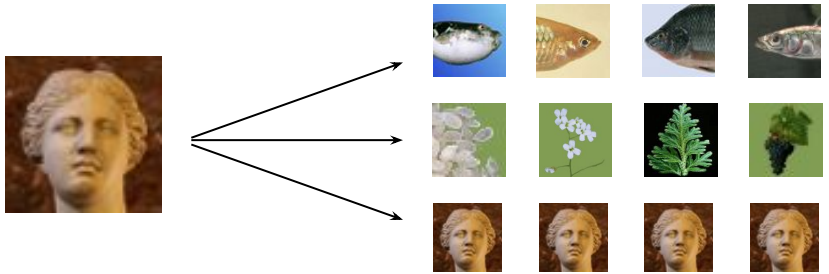
# Learning outcomes

- Differentiate the types of genomic data available in Ensembl
- Perform an analysis of a variation dataset using the Ensembl VEP
- View the output from the Ensembl VEP and filter to prioritise variants of interest
- Identify how to find the help and documentation pages for the available data types

# Sequencing genomes

- 1977: first genome sequenced (5 kb)
- 2004: first human genome sequenced (3 Gb)

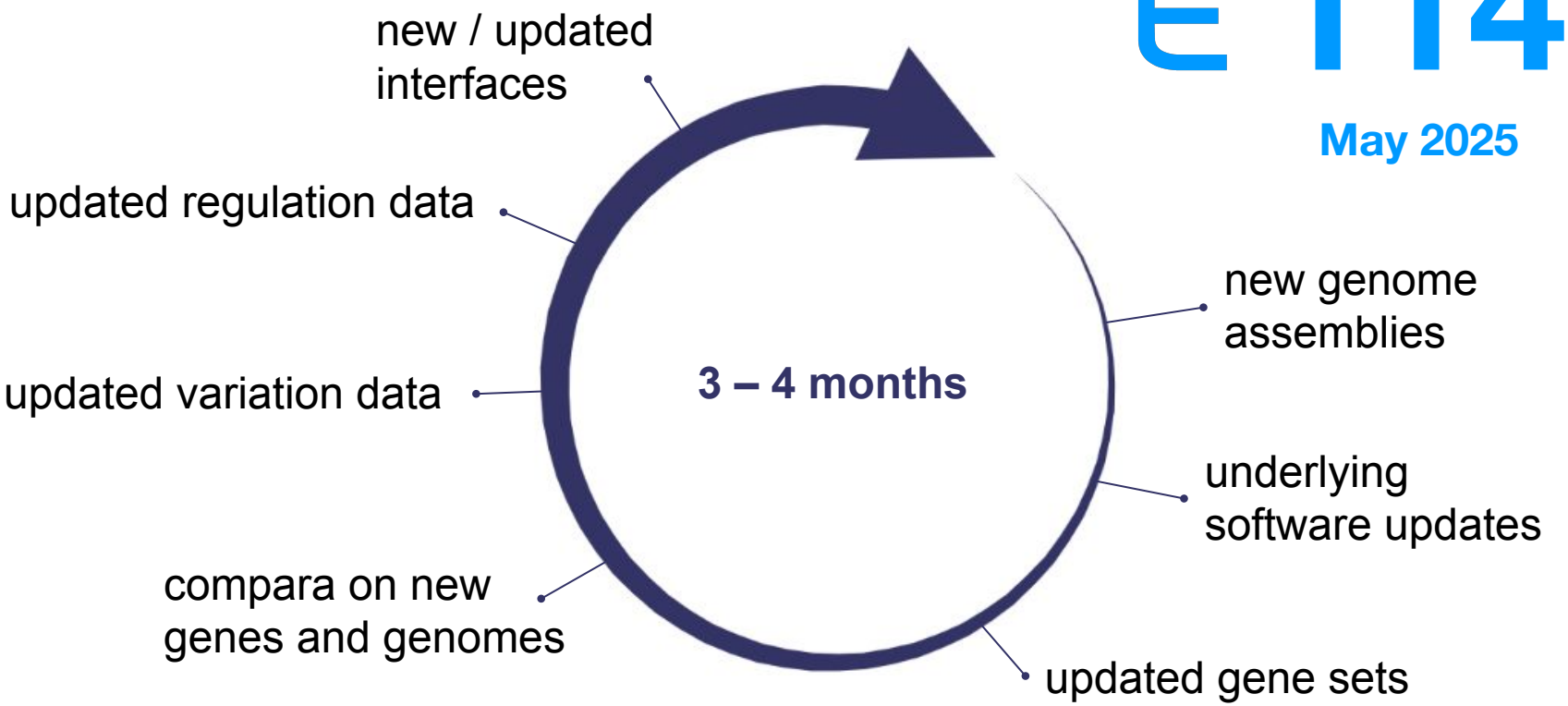




# Ensembl release cycle

E 114

May 2025

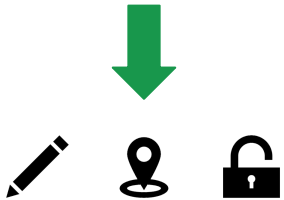


# Why do we need genome browsers?

# What is Ensembl?

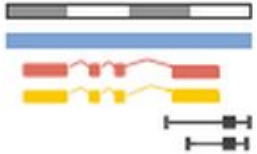
Ensembl annotates and maps genomic features from genome sequences, and brings together information from a wide range of other databases in a single site.

```
AGTGCTTAATGTGGCTAGTGGCACCGGTTTGGACAGCACAGCTGTAAAATGTTCCC
ATCCTCACAGTAAGCTGTTACCGTTCCAGGAGATGGGACTGAATTAGAATTCAAAC
AAATTTTCCAGCGCTTCTGAGTTTTACCTCAGTCACATAATAAGGAATGCATCCCT
GTGTAAGTGCATTTTGGTCTTCTGTTTTGCAGACTTATTTACCAAGCATTGGAGGA
ATATCGTAGGTAAAAATGCCTATTGGATCCAAAGAGAGGCCAACATTTTTTGAAAT
TTTTAAGACACGCTGCAACAAAGCAGGTATTGACAAATTTTATATAACTTTATAAA
TTACACCGAGAAAGTGTTTTCTAAAAAATGCTTGCTAAAAACCCAGTACGTCACAG
TGTTGCTTAGAACCATAAACTGTTCCTTATGTGTGTATAAATCCAGTTAACAACAT
AATCATCGTTTGCAGGTTAACCACATGATAAATATAGAACGTCTAGTGGATAAAGA
GGAAACTGGCCCCTTGACTAGCAGTAGGAACAATTACTAACAAATCAATTACTAAC
```



# Ensembl features

- Genomes and gene builds
- Comparative genomics (homologues, gene trees, alignments)
- Regulatory builds
- Data export with BioMart
- Tools for data processing, e.g. Ensembl VEP
- Display your own data
- Programmatic access via APIs
- Completely open-source (FTP, GitHub)



The mission of Ensembl is to provide a comprehensive and accurate annotation of genomes, integrating this information with other biological data and making it publicly accessible via the web.



# ENSEMBL

Genome data & annotation

About using Ensembl

**Species selector**

Create & manage your own species list

**Genome browser**

Look at genes & transcripts in their genomic context

**Entity viewer**

Get gene & transcript information

# Ensembl Beta

- New Ensembl genome browser platform with an app-based approach to functionality
- Infrastructure continues to provide the community with access to more intuitive interfaces and new feature developments
- Highly responsive and modern user interface
- User will have the ability to browse multiple species of interest simultaneously
- All Ensembl sites now in a single platform
- Currently under Beta testing - we want to hear the users' thoughts!

# Available human genome assemblies

99 human genome assemblies, including:

- GRCh37 (hg19)
- GRCh38 (hg38)
- T2T-CHM13
- Genomes of 48 individuals (96 haplotypes) from the Human Pangenome Reference Consortium (HPRC)

TOWARDS A COMPLETE REFERENCE OF HUMAN GENOME DIVERSITY

HUMAN PANGENOME

# New features to expect in Ensembl Beta

- Regulatory data (Available now!)
- Cross-references (XRefs) (Available now!)
- Additional genome assemblies and gene builds
- Gene expression, ontology and pathways
- Associated phenotypes
- Additional variation data

# The Ensembl Variant Effect Predictor (VEP)

## Ensembl Genome Browser Workshop

# Ensembl VEP

An analysis tool to predict the **molecular consequences** of variants and report essential information about a variant/ location.

# What can you do with Ensembl VEP?

**Variant input:**

- Alleles + location
- VCF
- HGVS
- Variant IDs
- SPDI


**Output:**

- Genes, transcripts affected
- Pathogenicity
- Frequency data
- Protein information
- PubMed citations

# Data input formats

**HGVS notation:**

```
ENST00000641515.2:c.4A>C
ENST00000272065.10:c.43+16C>T
ENST00000256509.7:c.5del
9:g.22125504G>C
```

**Variant IDs:**

```
rs1156485833
COSM327779
rs867704559
```

**SPDI:**

```
NC_000001.11:65567:A:C
NC_000002.12:265022:C:T
NC_000003.12:319780:A:
```

# Data input formats

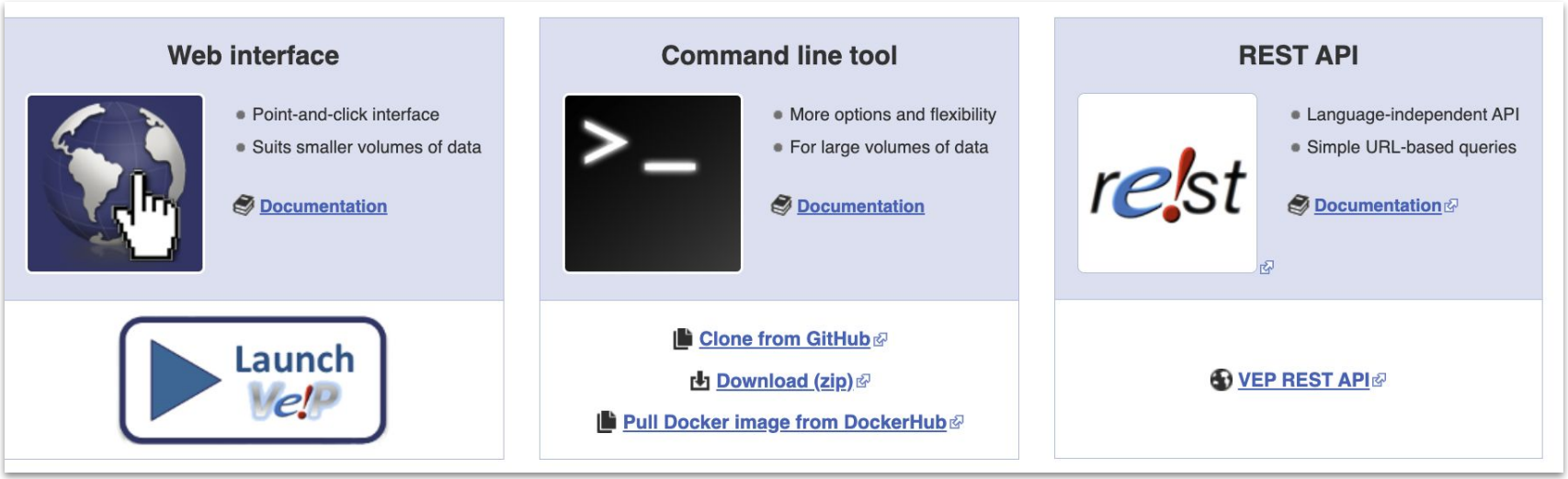
## Ensembl default input:

```
#chr    start   end allele strand id
1   65568   65568   A/C 1
2   265023  265023  C/T 1
3   319781  319781  A/- 1
```

## VCF:

```
#chr    pos id  ref alt qual    filt    info
1   65568   .   A   C   .   .   .
2   265023  .   C   T   .   .   .
3   319780  .   GA  G   .   .   .
```

# Ensembl VEP interfaces



# Ensembl VEP REST API

**`GET`** and **`POST`** requests are supported for single variants or small batches using HGVS notation, variant IDs or variant coordinates.

| Resource | Description |
|---|---|
| GET vep/:species/hgvs/:hgvs_notation | Fetch variant consequences based on a HGVS notation |
| POST vep/:species/hgvs | Fetch variant consequences for multiple HGVS notations |
| GET vep/:species/id/:id | Fetch variant consequences based on a variant identifier |
| POST vep/:species/id | Fetch variant consequences for multiple ids |
| GET vep/:species/region/:region/:allele/ | Fetch variant consequences |
| POST vep/:species/region | Fetch variant consequences for multiple regions |

# Command-line Ensembl VEP

- Simple installation
- Can handle genome-wide data
- Can work 100% offline
- Highly customisable
- Data bundles (caches) available for GRCh38 / hg38, GRCh37 / hg19, T2T-CHM13 and more

# Deleteriousness predictions

Ensembl VEP plugins are available to integrate predictive scores from tools including:

- Missense variant scores:
  - AlphaMissense
  - REVEL
  - EVE
  - ClinPred
- Genome wide change tolerance:
  - CADD
- Splicing impact:
  - SpliceAI
  - MaxEntScan
- Impact on gene expression:
  - Enformer

# Phenotype/functional evidence plugins

Phenotype and functional information is available from sources including :


results of highly parallel assays into cellular variant effect


evidence of variant impact on interactions

NHGRI-EBI GWAS Catalog

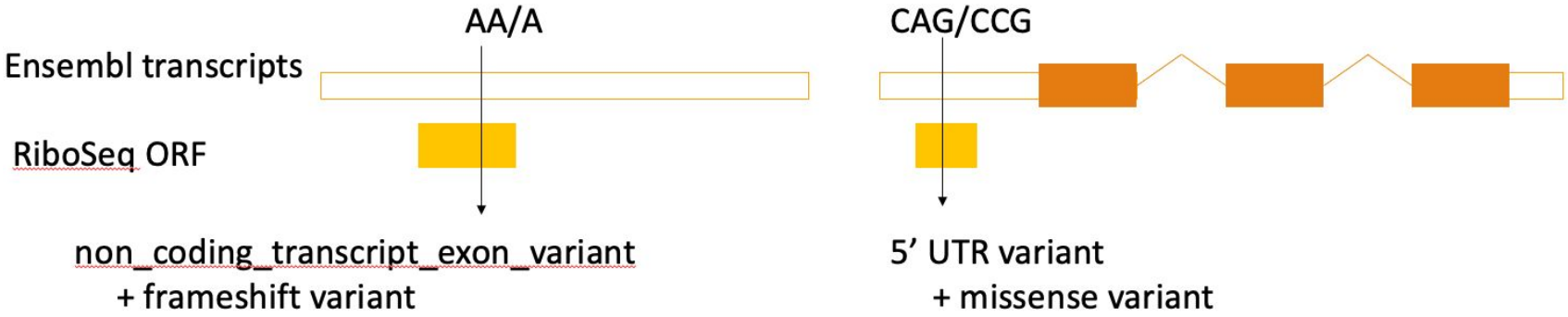
disease associations


the function, the cellular location and the associated biological processes of the product of any gene a variant lies within

# Additional analysis

Ensembl VEP plugins can run additional analysis, for example-

- Prediction of molecular consequences with respect to open reading frames discovered via RiboSeq

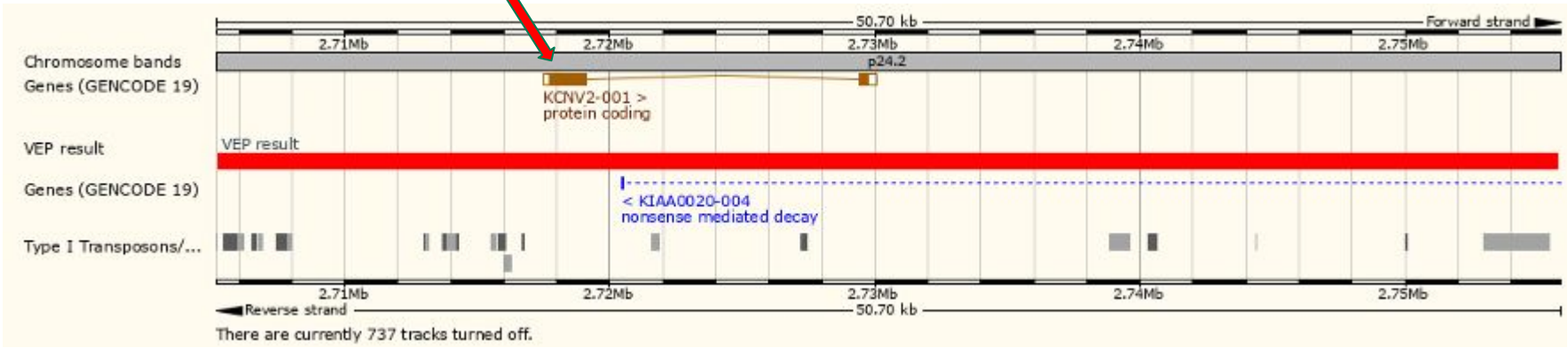



- Reporting ClinVar assertions associated with locations in paralogous genes

# Structural variant annotation

Ensembl VEP annotates SVs, reporting when transcripts and regulatory elements are deleted, truncated, amplified etc

| Uploaded variant | Location | Consequence | Symbol | Gene | Feature type | Feature | Biotype | Exon | Intron |
|---|---|---|---|---|---|---|---|---|---|
| 9_2705212_deletion | 9:2705211-2755784 | transcript_ablation | KCNV2 | ENSG00000168263 | Transcript | ENST00000382082 | protein_coding | 1-2/2 | 1/1 |
| 9_2705212_deletion | 9:2705211-2755784 | 3_prime_UTR_variant, intron_variant, NMD_transcript_variant, feature_truncation | KIAA0020 | ENSG00000080608 | Transcript | ENST00000490444 | nonsense_mediated_decay | 4/4 | 3/3 |
| 9_2705212_deletion | 9:2705211-2755784 | regulatory_region_ablation, regulatory_region_variant | - | - | RegulatoryFeature | ENSR00001445356 | - | - | - |
| 9_2705212_deletion | 9:2705211-2755784 | regulatory_region_ablation, regulatory_region_variant | - | - | RegulatoryFeature | ENSR00001445357 | - | - | - |



# More information

Publications:

Hunt SE, Moore B, Amode RM, et al. Annotating and prioritizing genomic variants using the Ensembl Variant Effect Predictor - A tutorial. Human Mutation. 2022 Aug;43(8):986-997. DOI: 10.1002/humu.24298. PMID: 34816521; PMCID: PMC7613081.

McLaren W, Gil L, Hunt SE, et al. The Ensembl Variant Effect Predictor. Genome Biology. 2016 Jun;17(1):122. DOI: 10.1186/s13059-016-0974-4. PMID: 27268795; PMCID: PMC4893825.

# Demonstration

We're going to use Ensembl VEP to look at a set of variant:

14 73219188 73219188 C/T
6 161785778 161785778 A/C
7 117480097 117480097 G/T
7 117587806 117587806  G/A
10 62813413 62813413 G/A

# Recommend us

**Browser course**
One day course on the Ensembl browser, aimed at wet-lab scientists.

**REST API course**
Half day course on the Ensembl REST API, aimed at bioinformaticians.

**Train the trainer course**
One day course on delivering the Ensembl browser course.

We can teach an Ensembl course at any institute for free (plus trainers' expenses in high income countries).

helpdesk@ensembl.org

https://training.ensembl.org/hosting

# Reach out to us

/Ensembl.org

/EnsemblGenomeBrowser

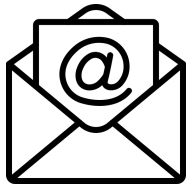
@Ensembl

www.ensembl.info

# Help and documentation

| | |
|---|---|
| Courses online | https://www.ebi.ac.uk/training/ |
| Tutorials | https://www.ensembl.org/info/website/tutorials/ |
| Flash animations | https://www.youtube.com/user/EnsemblHelpdesk |
| | https://youku.com/u/UMzM1NjkzMTI0 |

| | |
|---|---|
| Email | helpdesk@ensembl.org |
| Mailing lists | dev@ensembl.org |
| | announce@ensembl.org |



# Publications

- List of Ensembl publications: https://www.ensembl.org/info/about/publications.html
- Topic-specific articles mentioned throughout the workshop

Harrison PW, Amode MR, Harrison PW, Austine-Orimoloye O et al. Ensembl 2024. Nucleic Acids Research. 2024 Jan;52(D1):D891-D899. DOI: 10.1093/nar/gkad1049. PMID: 37953337; PMCID: PMC10767893.

Yates AD, Allen J, Amode RM, et al. Ensembl Genomes 2022: an expanding genome resource for non-vertebrates. Nucleic Acids Research. 2022 Jan;50(D1):D996-D1003. DOI: 10.1093/nar/gkab1007. PMID: 34791415; PMCID: PMC8728113.

# Acknowledgements and funding

EMBL
wellcome
UKRI Biotechnology and Biological Sciences Research Council
UKRI Medical Research Council
GORDON AND BETTY MOORE FOUNDATION
elixir
NIH National Human Genome Research Institute
NIH National Institute of Allergy and Infectious Diseases
NIH National Institute of General Medical Sciences

Funded by the European Union
Horizon 2020
Horizon Europe Programme

# Training materials

- Ensembl training materials are protected by a CC BY license: https://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: https://www.ensembl.org/info/about/publications.html