Nucleic Acids Research, 2025, gkaf1239
https://doi.org/10.1093/nar/gkaf1239
Database issue

Ensembl 2026

Andrew D. Yates[1,*], Olanrewaju Austine-Orimoloye[1], Andrey G. Azov[1], Matthieu Barba[1], If Barnes[1], Vianey Paola Barrera-Enriquez[1], Arne Becker[1], Ruth Bennett[1], Andrew Berry[1], Jyothish Bhai[1], Simarpreet Kaur Bhurji[1], Paulo R. Branco Lins[1], Lucy Brooks[1], Shashank Budhanuru Ramaraju[1], Lahcen I. Campbell[1], Manuel Carbajo Martinez[1], Jack Carpenter[1,2], Mehrnaz Charkhchi[1], Lucas A. Cortes[1,3], Claire Davidson[1], Suzanna Dickson[1,4], Kamalkumar Dodiya[1], Sarah Donaldson[1], Bilal El Houdaigui[1], Tamara El Naboulsi[1], Aine Fairbrother-Browne[1], Oluwadamilare Falola[1], Reham Fatima[1], Jose Gonzalez Martinez[1], Tatiana Gurbich[1], Holly Hall[1], Matthew Hardy[1], Zoe Hollis[1], Toby Hunt[1], Mike Kay[1], Vinay Kaikala[1], Anna Lazar[1], Diana Lemos[1], Disha Lodha[1], Nourhen Mathlouthi[1], Gabriela A. Merino[1], Ryan Merritt[1], Louisse Paola Mirabueno[1], Aleena Mushtaq[1], Syed Nakib Hossain[1], José G. Pérez-Silva[1], Ivana Piližota[1], Daniel Poppleton[1], Irina Prosovetskaia[1], Shriya Raj[1], Ahamed Imran Abdul Salam[1], Shradha Saraf[1], Swati Sinha[1], Botond Sipos[1], Vasily Sitnik[1], Marie-Marthe Suner[1], Likhitha Surapaneni[1], Jack A.S. Tierney[1], David Urbina-Gómez[1], Andres Veidenberg[1], Thomas A. Walsh[1], Jamie M. Allen[1], Jorge Alvarez-Jarreta[1], Jitender Cheema[1], Jorge Batista da Rocha[1], Nishadi H. De Silva[1], Francesca Floriana Tricomi[1], Stefano Giorgetti[1], Garth R. Ilsley[1], Jon Keatley[1], Jane E. Loveland[1], Jonathan M. Mudge[1], Guy Naamati[1], John Tate[1], Natalie L. Willhoft[1], Andrea Winterbottom[1], Bethany R. Flint[1], Adam Frankish[1], Leanne Haggerty[1], Sarah E. Hunt[1], Emily L. Clark[1], Sarah C. Dyer[1], Mallory A. Freeberg[1], Fergal J. Martin[1], Robert D. Finn[1]

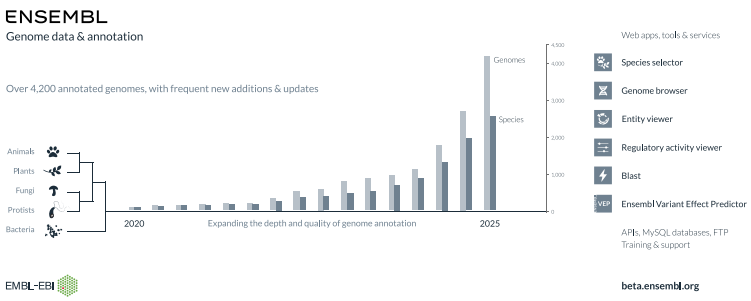
[1]European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SD, United Kingdom
[2]School of Biosciences, Cardiff University, Cardiff CF10 3AX, United Kingdom
[3]Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne NE1 7RU, United Kingdom
[4]School of Life Sciences, University of Warwick, Coventry CV4 7AL, United Kingdom

*To whom correspondence should be addressed. Email:ayates@ebi.ac.uk

Abstract

The Ensembl project (https://www.ensembl.org) is a public and open resource providing access to genomes, annotations, high-quality tools, and methods applicable to species from across the tree of life. This year has witnessed nearly a doubling in our rate of annotation and genome release, with 1927 new genomes released, with the total number of genomes now standing at 37 546. This includes expanded support for the human and barley pangenomes. We also present two new interfaces providing improved mechanisms to explore and interrogate genome regulation annotations. As our focus remains on sustainable scaling, we have archived Ensembl Rapid Release and accelerated the move to the new Ensembl platform. Ensembl release 116 (Q1-2026) will be the last release on the current platform.

Graphical abstract



Received: September 17, 2025. **Revised:** October 16, 2025. **Accepted:** October 16, 2025

Introduction

For 26 years, the Ensembl project (www.ensembl.org) has been providing high-quality reference genome annotation across the taxonomic space. Originating from the human genome project, the resource has grown to provide support for model organisms, vertebrates, microbes, plants, fungi, and other metazoans producing one of the world's most comprehensive genomic annotation resources. We annotate genes and transcripts, small- and large-scale genetic variation, genome regulatory elements, and comparative genomics (orthology and whole-genome alignments). These data are produced by our open, in-house analysis methods and supplemented with other annotation providers such as RefSeq and VEuPathDB [1, 2]. All genome sequences must be submitted to the International Nucleotide Sequence Database Collaboration and be publicly available as required by a joint browser agreement with the National Center for Biotechnology Information (NCBI) and the University of California, Santa Cruz (UCSC) Genome Browser [3–5]. Ensembl is an ELIXIR Core Data Resource and a Global Core Biodata Resource, highlighting the centrality of the resource to global research.

Ensembl data are available in standard file formats, without restrictions, through our websites, BioMart, REST application programming interfaces (APIs) (https://rest.ensembl.org), MySQL databases, and FTP site [6, 7]. Ensembl software is available from GitHub (https://github.com/Ensembl and https://github.com/EnsemblGenomes) including the Ensembl Variant Effect Predictor (Ensembl VEP) [8]. All Ensembl data are made available under the EMBL-EBI Terms of Use and software under an Apache 2.0 license

Ensembl Beta, our unified platform for accessing genomes across the tree of life, houses both newly generated datasets and the majority of data hosted in our dedicated portals: Vertebrates, Plants (plants.ensembl.org), Metazoa (metazoa.ensembl.org), Fungi (fungi.ensembl.org), Protists (protists.ensembl.org), and GRCh37 (grch37.ensembl.org). Data hosted in our Bacteria (bacteria.ensembl.org) and SARS-CoV-2 (covid-19.ensembl.org) portals will migrate during 2026. Our new platform receives updates approximately every two weeks and provides access to over 4200 genomes. Ensembl releases will continue on the current platform until release 116 (Q1-2026), after which all future data releases will be made through our new platform.

Broadening genome annotation resources

Ensembl continues to support an expanding community through increasing coverage of biodiversity projects including Darwin Tree of Life, European Reference Genome Atlas, Canada Biogenome Project, Aquatic Symbiosis Genomics Project, and Earth BioGenome Project (EBP) [9–12]. Over the past year, we have released 1927 new genomes representing 946 unique species across 20 phyla. New barley, oat, grape, pea, and lablab bean assemblies are available alongside large-scale variation for wheat and rice [13–18]. Livestock, companion animals, and rodents have received updates for sheep (new breed and reference) and cattle (new breeds and regulation annotation), cat, and rat, respectively (Table 1) [19–21]. We have imported new assemblies for key disease vectors and pests alongside alignment of genomes with VEuPathDB. All genomes made available from Ensembl Beta have homology predictions based on reciprocal best BLAST hits to align genomes with one of 11 pre-defined taxonomic collections, e.g. *Liliopsida* or *Mammalia*, or with a default set of 36 representative genomes if a specific group is not applicable [22].

To meet challenges in structurally annotating genes from genomes lacking transcriptomic data, we have adopted two new machine learning and hidden Markov model-based tools: Helixer and Tiberius [23, 24]. These methods model gene structure directly from a genomic sequence and are applicable across a wide taxonomic spread with minimal parameter tuning, in contrast to BRAKER2, which relies on species-specific training data and external homology evidence. *Umbilicaria deusta* (GCA_964340765.1) is our first released annotation based on Helixer. To help differentiate between methods, we include the annotation methodology as part of dataset metadata available from the Ensembl Beta website. These two new methods supersede BRAKER2 as our default annotation system for such cases [25].

Human annotation

The integration of long-read transcriptomic data has led to a significant increase in annotation of full-length protein-coding transcripts in GRCh38. Ensembl 115 contained ~121 000 new protein-coding transcripts (a 2.4-fold increase) added to the GRCh38 human reference gene set via the GENCODE TAGENE pipeline [26, 27]. This increase necessitated the use of our previously described GENCODE Primary subset methodology to identify and rank transcripts with high functional potential through expression and evolutionary constraint data [28]. Since Ensembl 114, GENCODE Primary is the default annotation for both human GRCh38 and mouse GRCm39 in our analysis and visualization tools, including Ensembl VEP, and is flagged in our GFF3 files [28]. GENCODE Primary enables faster downstream analysis by excluding exons and splice sites that lack evidence of evolutionary conservation or constraint or with low expression/inclusion. Additionally, the Matched Annotation from NCBI and EMBL-EBI (MANE) collaborative dataset has been updated to version 1.4 [29] and contains 50 disease-associated non-coding genes, including RNU4ATAC, which is implicated in RNU4ATAC Spectrum Disorder [30]. The Ensembl Transcript Archive (Tark) has been updated to include these collective annotation updates. Continuing our support of human pangenomics, we have annotated the second release of the Human Pangenome Reference Consortium representing the largest public set of assembled human genomes from 232 individuals of diverse ancestries and 464 haploid genomes annotated using projection from GENCODE 47 [31].

We continue to update variation and phenotype resources, providing a comprehensive description of human variation integrating data from dbSNP, gnomAD, NHGRI-EBI GWAS Catalog, COSMIC, and UniProt [32–35]. Phenotype association pages now display clinical impact classifications for somatic variants from ClinVar, enhancing support for cancer variant interpretation. GRCh37 and CHM13-T2T have been updated with population allele frequencies from gnomAD v4.1.

GRCh38 regulatory feature annotation has been refined by incorporating additional open chromatin and histone ChIP-seq data from the Encyclopedia of DNA Elements (ENCODE), with promoter annotation updated to match gene annotation from GENCODE 48. This led to a small increase in total features from 370 183 to 380 818, but notably a reduction in unclassified open chromatin regions from 19 029

Table 1. Summary of new assemblies, annotations, and datasets across plants, animals, insects, and microbes

Domain	Species/group	Update type
Companion animals	*Felis catus*	Assembly update (Fca126_mat1.0)
Crops	*Avena* spp. (oats)	4 new assemblies
Crops	*Hordeum vulgare* (barley)	75 cultivars, pangenome expansion, multiple whole-genome alignment (WGA) and gene trees
Crops	*Oryza sativa* (rice)	Variation from 3024 accessions (3K project)
Crops	*Pisum sativum* (pea), *S. stenocarpa* (African yam bean), *L. purpureus* (lablab bean)	New assemblies
Crops	*Triticum aestivum* (wheat)	New cultivar (Alchemy); variation from TaNG SNP and Watkins Core
Crops	*Vitis vinifera* (grape)	PN40024 telomere-to-telomere assembly
Disease vectors	*Amblyomma americanum* (lone star tick), *Ornithodoros turicata* (softbacked tick)	New assemblies
Fungi/Protists	*Plasmodium falciparum*, *Fusarium graminearum*	24 genomes imported from VEuPathDB
Insects	*Drosophila* spp.	6 new assemblies and pangenome WGA
Livestock	*Bos taurus* (cattle)	2 new breeds; ARS-UCD2.0 regulatory data (functional annotation of animal genomes)
Livestock	*Ovis aries* (sheep)	New breed annotations; Rambouillet update incl. Y chr.
Livestock	*Sus scrofa* (pigs)	Pangenome WGA update
Pests	*Vespa mandarinia* (Asian giant hornet), *Bactrocera oleae* (olive fruit fly), *Citripestis eutraphera* (mango seed moth)	New assemblies
Rodents	*Mus musculus*	Mouse strains WGA update
Rodents	*Rattus norvegicus*	Assembly update (GRCr8)

to 7541, with most now classified as enhancers. Both motif features and regulatory annotation from GRCh38 have been projected to GRCh37 using UCSC liftover with <1% of features failing to project.

Enhancing support for genomic interpretation

A new interface for genome regulation investigation and interpretation

Our new visualization (regulation.ensembl.org/115/regulatory_activity) is a reimagining of how to explore regulatory annotation (Fig. 1). We prioritized three enhancements: visualization of signal and peaks across all available epigenomes in the context of gene and regulatory annotation; filtering and sorting epigenomes by key metadata, e.g. life stage, organ; allowing epigenomes to be combined or split based on a dimension of interest, e.g. splitting liver activity by sex or combining into a single liver epigenome. Researchers can choose the attributes to filter displayed epigenomes, combine epigenomes based upon a shared set of attributes, and configure the order of tracks displayed based on the relative importance of selected attributes. The interface is capable of displaying annotations across hundreds of epigenomes within a few seconds. While currently only available for GRCh38, we plan to expand this visualization and roll it out to the remaining nine reference epigenomes (mouse, pig, cattle, chicken, Atlantic salmon, turbot, rainbow trout, common carp, and European seabass) supported by Ensembl in 2026.

Improving the exploration of epigenome catalogues

The new Ensembl regulation subsite (regulation.ensembl.org) provides a streamlined interface for exploring primary experimental data and sources that underpin our regulatory annotation. It allows researchers to browse annotation, high-level statistics, and key sample metadata sources (including BioSamples, European Nucleotide Archive, and ENCODE) across 3 Ensembl releases and our 10 supported species [36, 37]. Similar to our previously described regulation activity viewer, our regulation subsite allows filtering by multiple attributes, e.g. organ and life stage. The interface gives researchers access to experimental details for each regulatory feature, specifying in which epigenomes it is active or inactive, and the specific experiments and biological replicates used to make that determination.

Ensembl VEP

The Ensembl VEP, available online via a RESTful API, web tool, and command-line application, has been significantly enhanced with new features for interpreting genetic variants. Ensembl VEP includes NIH All Of Us allele frequency data and new somatic classifications from ClinVar [38, 39]. For human structural variants, we report clinical significance assertions from ClinVar and frequency information from gnomAD on overlapping structural variant sites. Ensembl VEP can indicate when a variant falls within a GENCODE promoter, a key feature for identifying potential non-coding disease associations.

Multiplexed assays of variant effects (MAVEs) from the MaveDB resource measure the effect of all possible variants in a region on a cellular phenotype, which provides insights into possible disease impact when the assay method reflects the gene-disease mechanism [40]. We have imported the latest MaveDB version, which represents a 6.4-fold increase in variants covered to ~7.7 million. Results from MaveDB are mapped to genomic coordinates to enable annotation of variants via Ensembl VEP with MAVE results [41]. Our online tool has been enhanced by embedding publication links, enabling easier interrogation of results. Finally, we have developed a new Ensembl VEP extension to integrate predictions of likely gene disease mechanism derived from a support vector classification model [42].

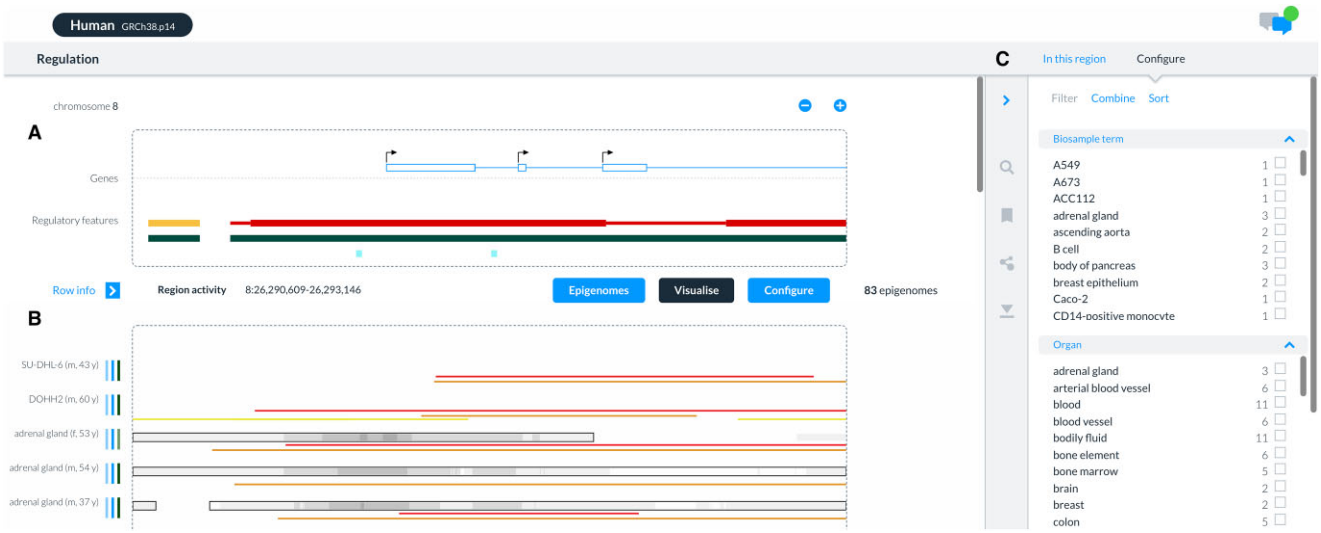


Figure 1. The regulatory activity viewer. (**A**) The gene of interest (PPP2R2A), exon and transcription start sites, and the mapping of regulatory features. Promoters are indicated in red, enhancers yellow, epigenetically modified accessible regions dark green, open chromatin light grey, and CTCF light blue. Each feature can be clicked on to display additional information. The panel can be panned and zoomed by clicking on the + and − buttons. A new focus gene can be selected by clicking on a gene and selecting 'Make focus'. (**B**) Epigenomic activity across cell/tissue (*y*-axis) with open chromatin and histone marks (*x*-axis). Open chromatin signal is indicated in greyscale, with peaks as black rectangles. Histone mark peaks are shown as lines: H3K4me3 (red), H3K27ac (orange), and H3K4me (lime). Cell/tissue metadata are summarized on the left, with further details available via the 'Epigenomes' button. (**C**) Configuration panel displayable by clicking the 'Configure' button, offering options to combine or sort epigenomes based on attributes.

Ensembl beta

In September 2025, we archived Ensembl Rapid Release (rapid-archive.ensembl.org). All data in Rapid Release are now available from Ensembl Beta, and all Rapid URLs redirect to the most appropriate Beta page using the Resolver API (resolver.ensembl.org). The rate of data release has now increased to ∼150 genomes every two weeks. Functionality has been extended to display population frequency data from EVA release 7 [43]. We have also made the first release of our core data model (CDM, https://github.com/Ensembl/ensembl-cdm-docs) and variant data model (VDM, https://github.com/Ensembl/ensembl-vdm-docs), a complete rebuild of the Ensembl data model for gene models, assemblies, metadata, and variant events. These new models drive our GraphQL API (https://beta.ensembl.org/data/graphql), providing a robust and modern foundation for data access and integration.

Supporting integrated and partial data release

The Ensembl project has been historically characterized by stable, versioned, and fully integrated data releases providing consistent datasets that can be cited in long-term analyses and support reproducible research. The time between these release cycles was significant in length, slowing access to new data. By contrast, data made available through our rapid release platform offered timely access to new genomes but with limitations: data were not systematically integrated across the resource and lacked long-term persistence. Ensembl Beta now supports two types of release cycles: an integrated release and a partial release. The integrated release corresponds to the traditional Ensembl release cycle, a fully synchronized update of Ensembl where data are coordinated approximately every 3 to 4 months. The partial releases provide genomes or subsets of annotations without a full integration across Ensembl but provide access to new data every ∼2 to 3 weeks (Fig. 2). In February 2025, the first Ensembl Beta integrated release made 2919 genomes available for long-term use; >1200 genomes were subsequently published in partial releases. Our strategy provides researchers with a choice between stable, consistent datasets or more immediate updates, depending on their requirements.

Training and support

We continue to offer a comprehensive training programme, delivering 71 in-person and virtual events to 4326 participants over the past year. This year witnessed the launch of our first virtual Train the Trainer (TtT) course to build capacity in bioinformatics education using Ensembl training materials. The initial course was delivered in collaboration with the Kano Independent Research Center Trust in Nigeria with 15 participants. Our second course expanded to a pan-African group combining recorded lectures and assignments to increase accessibility to our materials across 10 countries, reaching 15 participants. In June, our first workshop on using Ensembl Beta was delivered to another pan-African cohort of over 1700 participants across 50 classrooms in collaboration with H3ABioNet and the African Genomics Data Hub. We also piloted our first public engagement activity in Latin America, introducing students to bioinformatics by demonstrating DNA extraction from local fruits and exploring the genomics of regional dish ingredients. Materials from this course are available from our training website in both English and, for the first time, Spanish. Our outreach team is available to deliver courses on all Ensembl platforms, tools, and TtT to both in-person and virtual audiences. Support for Ensembl tools and data is available via our helpdesk and developer mailing list. Our training materials (available under a CC-BY 4.0 license) and calendar of upcoming events are available from https://training.ensembl.org, with additional events available from the EMBL-EBI training portal (https://www.ebi.ac.uk/training/events). Additional information concerning Ensembl

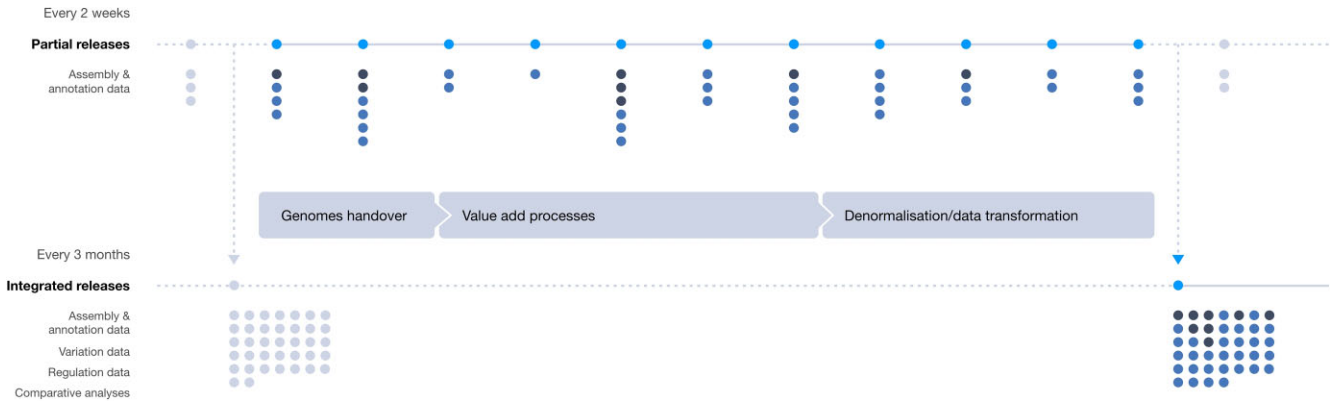


Figure 2. Two release processes are shown executing in parallel. Release points are indicated as light blue circles. Below each release point, new assemblies (dark blue dots) and annotation updates (black dots) are indicated. Genomes that were published prior to the start of an integrated release cycle are collated, and an integrated release is minted. In parallel, partial releases of data are made. After an integrated release is made, the next integrated cycle starts and brings in all partially released data.

can be found on our blog (https://www.ensembl.info), including release announcements and tool guides.

Future directions

Recently, the EBP has announced it will increase its efforts 10-fold in pace in order to reach its goal of sequencing all 1.67 million known species by 2035 [44]. As such, we recognize a need to accelerate our transition to our new platform. We plan to complete the migration of all genome assemblies and annotations by mid-2026 and have targeted Ensembl 116 (Q2-2026) as the last release on our current platform. All existing Ensembl sites and tools will become archives and will receive only high-priority data patches and infrastructure maintenance for one year. Our probe mapping resources will receive their last update in Ensembl 116 and remain accessible only via our archives. Researchers who work with biodiversity data and human pangenomes and require access to our latest data will be best served through our new infrastructure. Those reliant on our existing whole-genome comparative visualization, existing regulation interfaces, AlphaFold Variant visualization, and functional annotation including phenotype associations, and Gene Ontology annotations should remain on our current infrastructure while these visualizations and data are migrated [45]. Our new infrastructure's FTP site will provide access to data not yet displayable on our resource. We will develop a new set of interfaces and enhancements, including visualizations of large structural variants and sequence alignments; extend species search to include more filtering facets, including taxonomy, project, and thematic group; and bring our new data warehouse query service online. We will also identify currently available datasets to migrate to the new platform, including support for transcriptomic data and alternative gene models. Progress of the migration to the new infrastructure will be disseminated by our blog and social media channels, which provide advanced announcements concerning major changes when they are released and address frequently asked questions.

Acknowledgements

We wish to thank our user community and data providers for making their data available for reuse within Ensembl. We also wish to thank the following members of EMBL-EBI's IT & Technical Services for their continued support: Jonathan Barker, Sarah Butcher, Andy Cafferkey, Tim Dyce, Santiago Ramon Insua Fernandez, Somayeh Hajiahmadi, Atefeh Hasibi Taheri, Hasaranga Madhushan Liyana Pathirannehelage, Manuela Menchi, Ania Niewielska, David Ocaña, Tim Porter, Vidya Sreedevi Sankaran Potti, Marc Riera Duocastella, Karthick Subramanian, and C.D. Tiwari. We thank Terence Murphy and RefSeq curators for their collaboration on the MANE Project, Doreen Ware and the Gramene team at CSHL, and Wafaa M. Rashed and the Pan-African PGS Education and Research Initiative (PAPERI) community for their engagement in our virtual TtT workshop. 'Ensembl' and 'Ensembl VEP' are registered trademarks of EMBL. For the purpose of open access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Author contributions: Andrew D.Yates (Conceptualization [equal], Funding acquisition [equal], Methodology [equal], Project administration [equal], Supervision [equal], lead] Writing—original draft [lead] Writing—review & editing), Olanrewaju Austine-Orimoloye (Software [equal]), Andrey G. Azov (Data curation [equal], Software [equal]), If Barnes (Data curation [equal]), Vianey Paola Barrera-Enriquez (Software [equal]), Arne Becker (Data curation [equal], Software [equal]), Ruth Bennett (Data curation [equal]), Andrew Berry (Data curation [equal]), Jyothish Bhai (Software [equal]), Simarpreet Kaur Bhurji (Software [equal]), Paulo R. Branco Lins (Data curation [equal], Methodology [equal], Methodology [equal], Software [equal]), Lucy Brooks (Data curation [equal], Formal analysis [equal], Visualization [equal]), Shashank Budhanuru Ramaraju (Software [equal]), Lahcen I. Campbell (Software [equal]), Manuel Carbajo Martinez (Software [equal]), Jack Carpenter (Formal analysis [equal], Software [equal]), Mehrnaz Charkhchi (Software [equal]), Lucas A. Cortes (Software [equal]), Claire Davidson (Data curation [equal]), Suzanna Dickson (Formal analysis [equal], Formal analysis [equal], Validation [equal]), Kamalkumar Dodiya (Software [equal]), Sarah Donaldson (Data curation [equal]), Bilal El Houdaigui (Data curation [equal], Software [equal]), Tamara El Naboulsi (Data curation [equal], Software [equal]), Aine Fairbrother-Browne (Software [equal]), Oluwadamilare Falola (Software [equal]), Re-

ham Fatima (Software [equal]), Jose Gonzalez Martinez (Software [equal]), Tatiana Gurbich (Software [equal]), Holly Hall (Software [equal]), Matthew Hardy (Data curation [equal]), Zoe Hollis (Data curation [equal]), Toby Hunt (Data curation [equal]), Mike Kay (Data curation [equal]), Vinay Kaikala (Software [equal]), Anna Lazar (Formal analysis [equal], Software [equal]), Diana Lemos (Software [equal]), Disha Lodha (Software [equal]), Nourhen Mathlouthi (Project administration [equal]), Gabriela A. Merino (Data Curation [equal], Formal analysis [equal], Methodology [equal], Software [equal]), Ryan Merritt (Data curation [equal]), Louisse Paola Mirabueno (Resources [equal]), Aleena Mushtaq (Resources [equal]), Syed Nakib Hossain (Software [equal]), José G. Pérez-Silva (Software [equal]), Ivana Piližota (Software [equal]), Daniel Poppleton (Software [equal]), Irina Prosovetskaia (Software [equal]), Shriya Raj (Project administration [equal]), Ahamed Imran Abdul Salam (Software [equal]), Shradha Saraf (Software [equal]), Swati Sinha (Software [equal]), Botond Sipos (Software [equal]), Vasily Sitnik (Software [equal]), Marie-Marthe Suner (Data curation [equal]), Likhitha Surapaneni (Software [equal]), Jack A.S. Tierney (Software [equal]), David Urbina-Gómez (Data curation [equal], Formal analysis [equal], Methodology [equal], Software [equal]), Andres Veidenberg (Software [equal]), Thomas A. Walsh (Software [equal]), Jamie M. Allen (Project administration [equal], Software [equal], Supervision [equal]), Jorge Alvarez-Jarreta (Project administration [equal], Software [equal], Supervision [equal]), Jitender Cheema (Project administration [equal], Software [equal], Supervision [equal]), Jorge Batista da Rocha (Project administration [equal], Resources [equal], Supervision [equal]), Francesca Floriana Tricomi (Project administration [equal], Software [equal], Supervision [equal], Project administration [equal]), Stefano Giorgetti (Data curation [equal], Methodology [equal], Project administration [equal], Software [equal], Supervision [equal], Methodology [equal], Data curation [equal]), Garth R. Ilsley (Formal analysis [equal], Methodology [equal], Project administration [equal], Supervision [equal]), Jon Keatley (Data curation [equal], Project administration [equal], Software [equal], Data curation [equal], Supervision [equal]), Jane E. Loveland (Data curation [equal], Methodology [equal], Project administration [equal], Supervision [equal]), Jonathan M. Mudge (Data Curation [equal], Methodology [equal], Project administration [equal], Supervision [equal]), Guy Naamati (Project administration [equal], Resources [equal], Software [equal], Supervision [equal], Software [equal], Supervision [equal]), John Tate (Methodology [equal], Project administration [equal], Software [equal], Supervision [equal], Methodology [equal]), Natalie L. Willhoft (Methodology [equal], Project administration [equal], Software [equal], Supervision [equal]), Andrea Winterbottom (Methodology [equal], Methodology [equal], Visualization [equal]), Bethany R. Flint (Conceptualization [equal], Methodology [equal], Project Administration [equal], Supervision [equal], Visualization [equal], Writing—review & editing [equal]), Adam Frankish (Data curation [equal], lead] Funding acquisition, Methodology [equal], Project administration [equal], Supervision [equal], Writing—review & editing [equal], lead] Funding acquisition, Methodology [equal], Project administration [equal], Supervision [equal], Writing—review & editing [equal]), Leanne Haggerty (Methodology [equal], Project administration [equal], Software [equal], Supervision [equal], Writing—review & editing [equal], Methodology [equal], Project administration [equal], Software [equal], Supervision [equal], Writing—review & editing [equal]), Sarah E. Hunt (Funding acquisition [equal], Methodology [equal], Project administration [equal], Software [equal], Supervision [equal], Writing—review & editing [equal]), Emily L. Clark (Funding acquisition [equal], Methodology [equal], Project administration [equal], Methodology [equal], Project administration [equal], Supervision [equal]), Sarah C. Dyer (Funding acquisition [equal], Methodology [equal], Project administration [equal], Supervision [equal]), Mallory A. Freeberg (Funding acquisition [equal], Methodology [equal], Project administration [equal], Supervision [equal], Writing—review & editing [equal]), Fergal J. Martin (Funding acquisition [equal], Methodology [equal], Project administration [equal], Supervision [equal], Writing—review & editing [equal], Methodology [equal], Project administration [equal], Supervision [equal]), and Robert D. Finn (Funding acquisition [equal], Methodology [equal], Project administration [equal], Supervision [equal])

Conflict of interest

None declared.

Funding

Ensembl receives majority funding from Wellcome Trust [222155/Z/20/Z] with additional funding for specific project components. Research reported in this publication was supported by National Human Genome Research Institute of the National Institutes of Health under award number 2U24HG007234-09, U41HG010972, R01 HG010485, U24HG011451, and 2U24HG007497-05. Ensembl receives further funding from The Biotechnology and Biological Sciences Research Council [BB/W019108/1, BB/P016855/1, BB/T015608/1, BB/T01461X/1]; Open Targets; Wellcome Trust [212925/Z/18/Z, 226458/Z/22/Z, 226083/Z/22/Z]; ELIXIR: the research infrastructure for life-science data, and the European Molecular Biology Laboratory. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 733161 (MultipleMS), No. 825575 (EJP RD), No. 817923 (AQUA-FAANG), No. 817998 (GENE-SWitCH), and No. 815668 (BovReg). This project is funded in part by the Gordon and Betty Moore Foundation through grant CBMF8897. Funding to pay the Open Access publication charges for this article was provided by Wellcome Trust [222155/Z/20/Z].

Data availability

All Ensembl data is made available without restriction from our main website (https://www.ensembl.org), Ensembl Beta (https://beta.ensembl.org) and portals (https://plants.ensembl.org, https://metazoa.ensembl.org, https://fungi.ensembl.org, https://protists.ensembl.org, https://bacteria.ensembl.org, https://grch37.ensembl.org, https://covid-19.ensembl.org). Data is also available for bulk access via our FTP site (https://ftp.ensembl.org) and programmatically (https://rest.ensembl.org,https://beta.ensembl.org/data/graphql). Ensembl code is available from GitHub (https://github.com/Ensembl and https://github.com/EnsemblGenomes) under an open source Apache 2.0 licence. News about our releases and services can be found

on our blog (https://www.ensembl.info), our announce mailing list (https://lists.ensembl.org/mailman/listinfo/announce), X (@ensembl; https://x.com/ensembl), LinkedIn (https://www.linkedin.com/company/ensemblgenomebrowser) and Facebook (https://facebook.com/Ensembl.org).

References

1. Goldfarb T, Kodali VK, Pujar S *et al.* NCBI RefSeq: reference sequence standards through 25 years of curation and annotation. *Nucleic Acids Res* 2025;53:D243–57.
2. Alvarez-Jarreta J, Amos B, Aurrecoechea C *et al.* VEuPathDB: the eukaryotic pathogen, vector and host bioinformatics resource center in 2023. *Nucleic Acids Res* 2024;52:D808–16. https://doi.org/10.1093/nar/gkad1003
3. Karsch-Mizrachi I, Arita M, Burdett T *et al.* The international nucleotide sequence database collaboration (INSDC): enhancing global participation. *Nucleic Acids Res* 2024; 53:D62–6. https://doi.org/10.1093/nar/gkae1058
4. Sayers EW, Beck J, Bolton EE *et al.* Database resources of the National Center for Biotechnology Information in 2025. *Nucleic Acids Res* 2025;53:D20–9. https://doi.org/10.1093/nar/gkae979
5. Perez G, Barber GP, Benet-Pages A *et al.* The UCSC Genome Browser database: 2025 update. *Nucleic Acids Res* 2025;53:D1243–9. https://doi.org/10.1093/nar/gkae974
6. Kinsella RJ, Kähäri A, Haider S *et al.* Ensembl BioMarts: a hub for data retrieval across taxonomic space. *Database* 2011;2011:bar030.
7. Yates A, Beal K, Keenan S *et al.* The Ensembl REST API: ensembl data for any language. *Bioinformatics* 2015;31:143–5. https://doi.org/10.1093/bioinformatics/btu613
8. McLaren W, Gil L, Hunt SE *et al.* The Ensembl Variant Effect Predictor. *Genome Biol* 2016;17:122. https://doi.org/10.1186/s13059-016-0974-4
9. The Darwin Tree of Life Project Consortium, Blaxter M, Mieszkowska N *et al.* Sequence locally, think globally: the Darwin Tree of Life Project. *Proc Natl Acad Sci USA* 2022;119:e2115642118.
10. Mazzoni CJ, Ciofi C, Waterhouse RM Biodiversity: an atlas of European reference genomes. *Nature* 2023;619:252. https://doi.org/10.1038/d41586-023-02229-w.
11. McKenna V, Archibald JM, Beinart R *et al.* The Aquatic Symbiosis Genomics Project: probing the evolution of symbiosis across the Tree of Life. *Wellcome Open Res* 2024;6:254. https://doi.org/10.12688/wellcomeopenres.17222.2
12. Lewin HA, Robinson GE, Kress WJ *et al.* Earth BioGenome Project: sequencing life for the future of life. *Proc Natl Acad Sci USA* 2018;115:4325–33. https://doi.org/10.1073/pnas.1720115115
13. Feng J-W, Pidon H, Cuacos M *et al.* A haplotype-resolved pangenome of the barley wild relative *Hordeum bulbosum*. *Nature* 2025;645:429–38. https://doi.org/10.1038/s41586-025-09270-x
14. Wang W, Mauleon R, Hu Z *et al.* Genomic variation in 3,010 diverse accessions of Asian cultivated rice. *Nature* 2018;557:43–9. https://doi.org/10.1038/s41586-018-0063-9
15. 3,000 rice genomes project. The 3,000 rice genomes project. *Gigascience* 2014;3:7.
16. Cheng S, Feng C, Wingen LU *et al.* Harnessing landrace diversity empowers wheat breeding. *Nature* 2024;632:823–31. https://doi.org/10.1038/s41586-024-07682-9
17. Burridge AJ, Winfield M, Przewieslik-Allen A *et al.* Development of a next generation SNP genotyping array for wheat. *Plant Biotechnol J* 2024;22:2235–47. https://doi.org/10.1111/pbi.14341
18. Shi X, Cao S, Wang X *et al.* The complete reference genome for grapevine (*Vitis vinifera* L.) genetics and breeding. *Hortic Res* 2023;10:uhad061. https://doi.org/10.1093/hr/uhad061
19. Clark EL, Archibald AL, Daetwyler HD *et al.* From FAANG to fork: application of highly annotated genomes to improve farmed animal production. *Genome Biol* 2020;21:285. https://doi.org/10.1186/s13059-020-02197-8
20. Halstead MM, Kern C, Saelao P *et al.* A comparative analysis of chromatin accessibility in cattle, pig, and mouse tissues. *BMC Genomics* 2020;21:698. https://doi.org/10.1186/s12864-020-07078-9
21. Kern C, Wang Y, Xu X *et al.* Functional annotations of three domestic animal genomes provide vital resources for comparative and agricultural research. *Nat Commun* 2021;12:1821. https://doi.org/10.1038/s41467-021-22100-8
22. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL, BLAST+: architecture and applications.. *BMC Bioinformatics* 2009;10:421.
23. Gabriel L, Becker F, Hoff KJ *et al.* Tiberius: end-to-end deep learning with an HMM for gene prediction. *Bioinformatics* 2024;40:btae685. https://doi.org/10.1093/bioinformatics/btae685
24 Holst F, Bolger A, Günther C *et al.* Helixer—*de novo* prediction of primary eukaryotic gene models combining deep learning and a hidden Markov model. bioRxiv, https://doi.org/10.1101/2023.02.06.527280, 06 February 2023, preprint: not peer reviewed.
25. Brůna T, Hoff KJ, Lomsadze A *et al.* BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genomics Bioinforma* 2021;3:lqaa108. https://doi.org/10.1093/nargab/lqaa108
26. Mudge JM, Carbonell-Sala S, Diekhans M *et al.* GENCODE 2025: reference gene annotation for human and mouse. *Nucleic Acids Res* 2025;53:D966–75. https://doi.org/10.1093/nar/gkae1078
27. Kaur G, Perteghella T, Carbonell-Sala S *et al.* GENCODE: massively expanding the lncRNA catalog through capture long-read RNA sequencing. bioRxiv,. https://doi.org/10.1101/2024.10.29.620654, 31 October 2024, preprint: not peer reviewed
28. Dyer SC, Austine-Orimoloye O, Azov AG *et al.* Ensembl 2025. *Nucleic Acids Res* 2025;53:D948–57. https://doi.org/10.1093/nar/gkae1071
29. Morales J, Pujar S, Loveland JE *et al.* A joint NCBI and EMBL-EBI transcript set for clinical genomics and research. *Nature* 2022;604:310–5. https://doi.org/10.1038/s41586-022-04558-8
30. Duker A, Velasco D, Robertson N *et al.* RNU4atac-opathy. InAdam MP, Feldman J, Mirzaa GM et al. (eds), *GeneReviews*®. Seattle (WA): University of Washington, 1993.
31. Liao W-W, Asri M, Ebler J *et al.* A draft human pangenome reference. *Nature* 2023;617:312–24. https://doi.org/10.1038/s41586-023-05896-x
32. Resource Coordinators NCBI, Agarwala R, Barrett T *et al.* Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 2018;46:D8–D13. https://doi.org/10.1093/nar/gkx1095
33. Chen S, Francioli LC, Goodrich JK *et al.* A genomic mutational constraint map using variation in 76,156 human genomes. *Nature* 2024;625:92–100. https://doi.org/10.1038/s41586-023-06045-0
34. Sollis E, Mosaku A, Abid A *et al.* The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. *Nucleic Acids Res* 2023;51:D977–85. https://doi.org/10.1093/nar/gkac1010
35. Sondka Z, Dhir NB, Carvalho-Silva D *et al.* COSMIC: a curated database of somatic variants and clinical data for cancer. *Nucleic Acids Res* 2024;52:D1210–7. https://doi.org/10.1093/nar/gkad986
36. Courtot M, Gupta D, Liyanage I *et al.* BioSamples database: fAIRer samples metadata to accelerate research data management. *Nucleic Acids Res* 2022;50:D1500–7. https://doi.org/10.1093/nar/g1046
37. Luo Y, Hitz BC, Gabdank I *et al.* New developments on the Encyclopedia of DNA Elements (ENCODE) data portal. *Nucleic Acids Res* 2020;48:D882–9. https://doi.org/10.1093/nar/gkz1062
38. The All of Us Research Program Investigators. The "All of Us" Research Program. *N Engl J Med* 2019;381:668–76. https://doi.org/10.1056/NEJMsr1809937.

39. Landrum MJ, Lee JM, Benson M *et al.* ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res* 2018;46:D1062–7. https://doi.org/10.1093/narx1153
40. Rubin AF, Stone J, Bianchi AH *et al.* MaveDB 2024: a curated community database with over seven million variant effects from multiplexed functional assays. *Genome Biol* 2025;26:13. https://doi.org/10.1186/s13059-025-03476-y
41. Arbesfeld JA, Da EY, Stevenson JS *et al.* Mapping MAVE data for use in human genomics applications. *Genome Biol* 2025;26:179. https://doi.org/10.1186/s13059-025-03647-x
42. Badonyi M, Marsh JA Proteome-scale prediction of molecular mechanisms underlying dominant genetic diseases. *PLoS One* 2024;19:e0307312. https://doi.org/10.1371/journal.pone.0307312
43. Cezard T, Cunningham F, Hunt SE *et al.* The European Variation Archive: a FAIR resource of genomic variation for all species. *Nucleic Acids Res* 2022;50:D1216–20. https://doi.org/10.1093/nar/gkab960
44. Blaxter M, Lewin HA, DiPalma F *et al.* The Earth BioGenome Project Phase II: illuminating the eukaryotic tree of life. *Front Sci* 2025;3:1514835. https://doi.org/10.3389/fsci.2025.1514835
45. Camon E. The Gene Ontology Annotation (GOA) Database: sharing knowledge in Uniprot with Gene Ontology. *Nucleic Acids Res* 2004;32:262D–266. https://doi.org/10.1093/nar/gkh021

Received: September 17, 2025. **Revised:** October 16, 2025. **Accepted:** October 16, 2025