Faculty Research 2023   Faculty & Staff Research

12-1-2023

The alliance of genome resources: transforming comparative genomics.

Carol J Bult

Paul W Sternberg

The alliance of genome resources: transforming comparative genomics

Carol J. Bult[1] · Paul W. Sternberg[2]

Abstract

Comparing genomic and biological characteristics across multiple species is essential to using model systems to investigate the molecular and cellular mechanisms underlying human biology and disease and to translate mechanistic insights from studies in model organisms for clinical applications. Building a scalable knowledge commons platform that supports cross-species comparison of rich, expertly curated knowledge regarding gene function, phenotype, and disease associations available for model organisms and humans is the primary mission of the Alliance of Genome Resources (the Alliance). The Alliance is a consortium of seven model organism knowledgebases (mouse, rat, yeast, nematode, zebrafish, frog, fruit fly) and the Gene Ontology resource. The Alliance uses a common set of gene ortholog assertions as the basis for comparing biological annotations across the organisms represented in the Alliance. The major types of knowledge associated with genes that are represented in the Alliance database currently include gene function, phenotypic alleles and variants, human disease associations, pathways, gene expression, and both protein–protein and genetic interactions. The Alliance has enhanced the ability of researchers to easily compare biological annotations for common data types across model organisms and human through the implementation of shared programmatic access mechanisms, data-specific web pages with a unified "look and feel", and interactive user interfaces specifically designed to support comparative biology. The modular infrastructure developed by the Alliance allows the resource to serve as an extensible "knowledge commons" capable of expanding to accommodate additional model organisms.

Introduction

The Alliance of Genome Resources (the Alliance) is a consortium of seven model organism knowledgebases and the Gene Ontology resource. The mission of the Alliance is to support comparative genomics as a means to investigate the genetic and genomic basis of human biology, health, and disease. The Alliance seeks to serve a diverse community of biomedical researchers including basic scientists, clinicians, and data scientists. To promote sustainability of core community data resources, the Alliance has implemented and maintains an extensible "knowledge commons platform" for comparative genomics using modular infrastructure components that can be used by a wide range of multiple model organism genome knowledgebases (Alliance of Genome Resources 2019, 2020; Howe et al. 2018). The history of how Model Organism Databases and the Gene Ontology Consortium united to form the Alliance of Genome Resources has been published previously (Alliance of Genome Resources 2019, 2022). In 2023, the Alliance was recognized as a Core Global Biodata Resource by the Global Biodata Coalition (Anderson et al. 2017) (Fig. 1).

The Alliance of Genome Resources is organized as two interdependent units: Alliance Central and the Alliance Knowledge Centers (Alliance of Genome Resources 2019) (Fig. 1). *Alliance Central* is responsible for developing and maintaining the software for data access and for the coordination of concept modeling and data harmonization activities across the Knowledge Centers. The ultimate goal of Alliance Central is to reduce redundancy in systems administration and software development for model organism data resources and to deploy a unified 'look and feel' for access to and display of data types and annotations in common across diverse model organisms. Model organism-specific knowledgebases serve as *Alliance Knowledge Centers*. Knowledge Centers are responsible for expert curation of data and for

✉ Carol J. Bult
  carol.bult@jax.org

1 The Jackson Laboratory, Bar Harbor, ME, USA

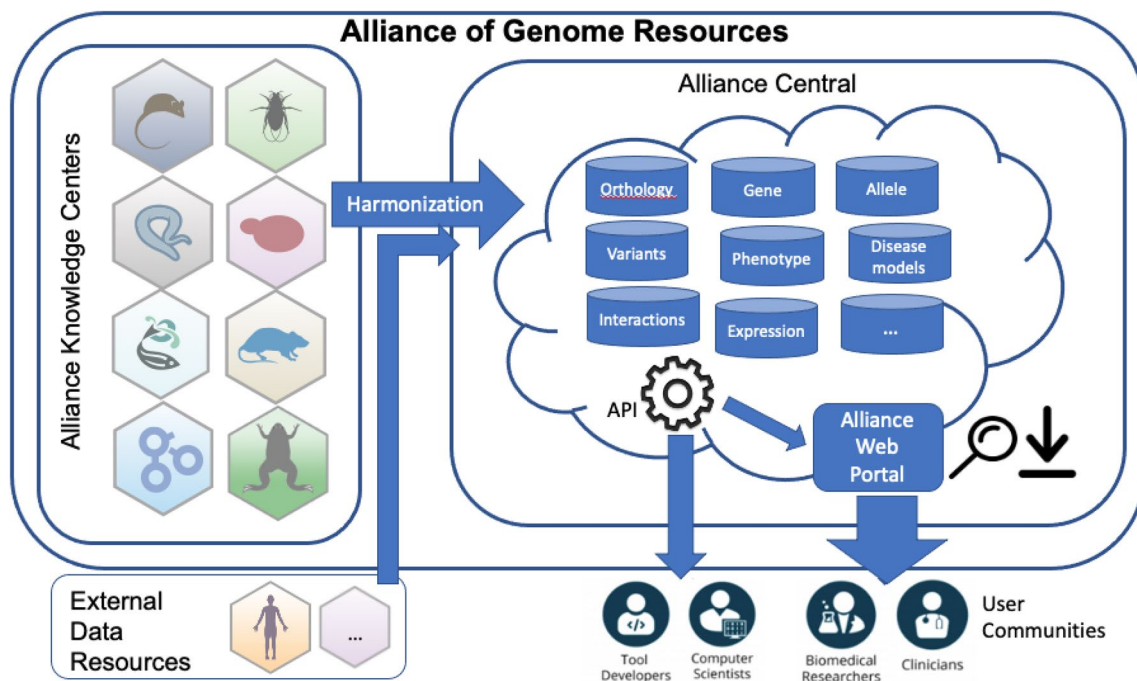2 California Institute of Technology, Pasadena, CA, USA

Fig. 1 A graphical representation of the organizational units that comprise the Alliance of Genome Resources consortium. Knowledge Centers (aka, model organism databases or MODs) are organism-specific resources for expert curation of knowledge about a model organism's genome. Data from external resources are integrated with genome annotations provided by the Knowledge Centers. Alliance Central is the platform for delivery of harmonized biological annotations to a diverse user community via the Alliance web portal and Application Programming Interfaces (APIs)

submission of data to Alliance Central using standardized data formats and annotation standards. Knowledge Centers also are responsible for organism-specific user support activities and for providing access to data types not yet supported by Alliance Central. The founding Alliance Knowledge Centers are Saccharomyces Genome Database (Engel et al. 2022), WormBase (Davis et al. 2022), FlyBase (Gramates et al. 2022), Mouse Genome Informatics (Ringwald et al. 2022), the Zebrafish Information Network (Bradford et al. 2023), Rat Genome Database (Vedi et al. 2023), and the Gene Ontology Consortium (Gene Ontology et al. 2023). The newest member, Xenbase (Fisher et al. 2023), joined the Alliance consortium in 2022. Annotations for human genes are acquired from numerous resources including the Alliance Knowledge Centers, NCBI's dbSNP (Smigielski et al. 2000), the Human Gene Nomenclature Committee (Yates et al. 2021), Disease Ontology (Schriml et al. 2022), Human Phenotype Ontology (Kohler et al. 2021), Orphanet (Rath et al. 2012), OMIM (Hamosh et al. 2021), BioGrid (Oughtred et al. 2021), and Reactome (Gillespie et al. 2022).

Although the model organism-centric knowledgebases that comprise the Alliance all contain similar data types (e.g., gene function, gene expression, genetic variation, phenotype, and human disease associations), the resources differ in how these data types are modeled and displayed to end users. These differences present significant challenges to the development of common schemas and uniform user interfaces for data types across different organisms. To address these challenges, a major activity within the Alliance is the harmonization of biological concepts which can be represented in a common schema. For example, all of the model organisms currently in the Alliance consortium have a concept of a transgene. For some model organisms, a transgene is represented as a random insertion of a construct but does not include gene trap alleles. For other organisms, gene traps are included in the representation of transgenes. For yet other model organisms, transgenes are represented as the random insertion of any foreign DNA into the genome, including the construct. To implement a common data model and unified display for transgenes, a harmonized data model was developed in which transgenes are represented by two separate concepts—the transgene construct and the transgene allele—and the specific relationships between the concepts. In the harmonized Alliance model, a transgene construct is defined as the DNA used to create a transgenic allele. The transgene construct has explicit relationships to genes, gene segments, and to the transgenic alleles created using the construct. The transgenic allele represents a construct in the context of a genome. Transgenic alleles have relationships to constructs and genomes. Because the transgene data type

a.

Species (carrying the transgene)	Allele Symbol	Transgenic Construct	Expressed Components	Knock-down Targets	Regulatory Regions	Has Disease Annotations	Has Phenotype Annotations
Drosophila melanogaster	mTor[+t9.4]	P{dTOR-9.4}	mTor (Dme)		mTor (Dme)		Yes
Drosophila melanogaster	mTor[FRB.UAS]	P{UAS-mTor.FRB}	mTor (Dme)		UASt		
Drosophila melanogaster	mTor[GD10721]	P{GD10721}		mTor (Dme)	UASt		
Drosophila melanogaster	mTor[GL00156]	P{TRiP.GL00156}		mTor (Dme)	UAS		

b.

Species (carrying the transgene)	Allele Symbol	Transgenic Construct	Expressed Components	Knock-down Targets	Regulatory Regions	Has Disease Annotations	Has Phenotype Annotations
Mus musculus	Cdh5[tm4(Cdh5/Fkbp,Ptprb/Mtor*)Dvst]		Cdh5 (Mmu), Fkbp1a (Mmu), Mtor (Mmu), Ptprb (Mmu)				Yes
Mus musculus	Gt(ROSA)26Sor[tm1(tTA,tetO-Mtor*D2338A,-IRES-EGFP)Ebm]		Mtor (Mmu)				

Fig. 2 Example of the harmonized representation of transgenic alleles at the Alliance of Genome Resources. **a** Transgenic alleles of the mTor gene in *Drosophila*. **b** Transgenic alleles of the *Mtor* gene in the laboratory mouse. In the Alliance data model, transgenic alleles are represented by various components that can be populated with information in a species-specific manner which allows for representational consistency and completeness across all model organisms. Allele symbols conform to species-specific nomenclature standards. Transgenic constructs represent the transgene construct independent of the host organism following species-specific guidelines. Expressed components represent the genetic elements expressed by the construct which may be full or partial genes and may include both protein-coding and non protein-coding genes. Knock-down targets are elements in the transgene designed to interfere with the expression of another gene. The regulatory regions component includes genetic elements driving the expression of entities in the construct (e.g., upstream activation sequence, human cytomegalovirus, etc.)

is harmonized, data from all of the model organism-specific Knowledge Centers can be represented in a uniform manner on the Alliance web portal (Fig. 2).

Biological annotations obtained from data-specific resources that are not members of the Alliance consortium are also incorporated into Alliance Central. For example, the Biological General Repository for Interaction Datasets (BioGrid) (Oughtred et al. 2021) and the International Molecular Exchange consortium (IMex) (Porras et al. 2022) are primary sources of molecular and genetic interaction data. Reactome (Gillespie et al. 2022) is leveraged as one source of pathway and reaction data. The Alliance Central practice of leveraging existing community resources also extends to software for data analysis and visualization. Externally developed tools such as Intermine (Smith et al. 2012), JBrowse (Buels et al. 2016), Apollo (Dunn et al. 2019), SequenceServer (Priyam et al. 2019), and the Reactome pathway viewer (Gillespie et al. 2022) are key components of the knowledge commons platform providing useful functionality for the Alliance user community and allowing software development efforts within Alliance Central to be focused on tools and interfaces for comparative biology and genomics that provide added value to the biomedical research community. A number of the software components used by the Alliance (e.g., Apollo, JBrowse, Intermine) were developed under the auspices of the Generic Model Organism Database project (http://gmod.org/wiki/Main_Page).

The Alliance resource has a unique and complementary role relative to other informatics resources that support comparative biology such as NCBI's new Comparative Genomics Resource (CGR; https://www.ncbi.nlm.nih.gov/comparative-genomics-resource/). Whereas the CGR is focused on developing analysis tools and resources for sequence-based genome comparisons across a large number of species, the Alliance focuses on standardized annotations, harmonized biological concepts, and comparison of biological knowledge. The CGR supports comparative sequence analysis for all eukaryotes whereas the Alliance is primarily focused on model organisms used widely in biomedical research. The CGR resource integrates the standardized gene summaries from the Alliance and follows nomenclature and ontology standards developed and maintained by Alliance members. For sequence analysis, the Alliance leverages sequence-based analysis tools developed and maintained by the CGR such as BLAST.

The approach to data management for the Alliance and its members aligns with modern FAIR principles (Findability, Accessibility, Interoperability, and Reusability) (Wilkinson

et al. 2016) which are designed to ensure that data are structured to be machine accessible with minimal or no human intervention. Examples of how the Alliance conforms to FAIR principles includes the use of unique, persistent identifiers for data entities and meta-data, the use of well-recognized and accepted community standard bio-ontologies and vocabularies for knowledge representation, clear data use licensing guidelines, and the availability of open and freely available application programming interfaces (APIs) for data retrieval.

The Alliance of Genome Resources web portal

The Alliance web portal (www.alliancegenome.org) provides a single point of access to the expertly curated and harmonized annotations from diverse model organisms and humans. The portal supports keyword searching within six categories: Gene, Allele/Variant, Disease Models, Gene Ontology annotation, Disease, and High Throughput Data (HTP) (Fig. 3). Results of keyword searches are displayed as faceted counts for the six categories. The counts are updated as search parameters are refined by the user. The Alliance

Table 1 Content summary for the alliance of genome resources web portal (v. 5.4.0)

Data type	Count
Genes	352,073
Alleles/variants	401,287,981
Disease models	142,147
Functional annotations	43,095
Disease ontology (DO) terms	11,237
Annotations using DO terms	351,137
High throughput datasets	10,753

Data at the portal is available currently for human, mouse, rat, zebrafish, frog, nematode, yeast, and fruit fly

database content summary as of the most recent release of the portal (v. 5.4.0) is provided in Table 1.

Search results are displayed with a consistent look and feel and layout of data for all organisms represented in the Alliance. To facilitate the comparison of biological knowledge across multiple species, annotations for orthology, function, phenotype, and disease are displayed in the portal use an interactive comparative annotation ribbon (Fig. 4). The ribbon display allows users to quickly assess the degree to which annotations are similar across multiple species. The cells in the ribbon are linked to tabular summaries with

Fig. 3 Partial screenshot of the Alliance web portal results page showing the search results for the *BMP4* gene. The six major categories of data with counts of database records are provided on the left navigation panel. Users can refine their searches by selecting a category of interest which leads to the display of new facet categories and record counts. Alternatively, users can scroll through the search results and select a gene of interest to navigate to the relevant detail page in the web portal

Fig. 4 Screenshot showing the interactive annotation summary ribbon for functional annotations of orthologs of the human *BMP4* gene. Each row in the ribbon is a different species. Each column is a grouping of terms from the molecular function, biological process, and cellular component categories of the Gene Ontology. The shading in the cells reflects the number of annotations to terms represented by the column. The darker the shading, the more annotations. Selecting a cell in the ribbon results in a table that shows the specific ontology terms and evidence used to make the annotation. Users can use the orthology settings to control which organisms are shown in the graphic

details about the relevant ontology terms and sources of evidence for the annotations.

In addition to support for keyword searches, the Alliance web portal provides users with downloadable files of gene description summaries and annotations in commonly used data formats (e.g., JSON, tab-separated, GFF, etc.). Downloads are currently available for disease annotations, gene expression, molecular and genetic interactions, orthology, alleles, and short gene descriptions. Sequence variants that are associated with documented phenotypic consequences are available as files in Variant Call Format (VCF) format. The downloadable data files are updated regularly. The file headers display the Alliance database version and the date the file was generated. Programmatic access to annotations in the Alliance is provided through an OpenAPI Specification (OAS). The schemas for the API-accessible data classes are available in a browsable format on the website.

Users looking to search for more than one gene at a time or in managing lists of genes can use AllianceMine. AllianceMine uses the InterMine data warehouse system (Smith et al. 2012). Although AllianceMine can be used without creating an account, having an account allows users to save gene lists and the outputs of gene list operations (i.e., intersection, combine, difference, subtraction).

Schedules for public data releases at the Alliance Knowledge Centers range from daily to monthly. Data submitted from Alliance Knowledge Centers and other external data sources are refreshed monthly at the Alliance web portal. These monthly data releases are occasionally suspended when major software changes to the Alliance infrastructure are being implemented. Release notes are accessed from the News menu in the header that document changes to data release frequencies, user interfaces, portal functionality, and any known issues with Alliance resources (https://www.alliancegenome.org/release-notes).

Orthology

Gene orthology is fundamental to comparative genomics. The Alliance uses a common set of orthologs as the foundation for comparing functional, phenotype, and disease annotations across model organisms and humans. Alliance orthology assertions are based on outputs from algorithms/methods that have been benchmarked by the Quest for Orthologs Consortium (Nevers et al. 2022) and integrated using the DRSC Integrative Ortholog Prediction Tool (DIOPT) (Hu et al. 2011). These ortholog assertions are subsequently supplemented with manually curated ortholog inferences from the Human Gene Nomenclature Committee (for human and mouse genes) (Yates et al. 2021), Xenbase (for frog genes), and ZFIN (for zebrafish genes). Manual curation is particularly useful for ensuring accuracy and completeness of orthology representation for species such as *Xenopus* and *Danio* where there has been extensive genome duplication.

Gene detail pages

Organism-specific gene detail pages in the Alliance web portal are the primary 'hub' of functional and biological annotations. All gene pages include a summary section with a short description of what the gene does and its association with phenotypes and/or human disease. The gene function descriptions are generated automatically by an algorithm that leverages expertly curated structured ontology term annotations associated with genes (Kishore et al. 2020). In addition to a gene summary, the standard sections of gene pages include Orthology, Function, Pathways, Phenotypes, Disease Associations, Models, Alleles and Variants, Transgenic Alleles, Sequence Feature Viewer, Gene Expression, Molecular Interactions, and Genetic Interactions. The two primary means for displaying annotations in these categories are a table view and an annotation ribbon display. Brief descriptions for each section of the gene detail pages and examples of the display paradigms are provided below. A list of the specific ontologies used at the Alliance along with licensing information is provided in Table 2 and on the Privacy, Warranty, and Licensing page (https://www.alliancegenome.org/privacy-warranty-licensing) at the Alliance web portal.

Orthology

The default display of orthologs reflects the output of the most stringent criteria based on DIOPT score; however, options are provided for researchers to select less stringent criteria and/or orthologs predicted by a user-selected subset of the available inference methods (Fig. 5). The Alliance orthologs are available as a downloadable file and can also be accessed programmatically via the Alliance Central API service.

Functional annotations (Gene Ontology)

Annotations to terms for Biological Process, Molecular Function, and Cellular Component from the Gene Ontology are summarized for high level GO categories using the Alliance annotation ribbon display paradigm (Fig. 4). Each cell in the ribbon is shaded if there is an annotation for a term in the category. The deeper the color of the shading, the more annotations are associated with the terms in the category. Selecting a cell generates a table that lists all of the annotation terms with evidence codes and sources for the annotation(s). By expanding the display to include additional organisms, the functional annotations for orthologs are displayed as additional rows.

Pathways

Representation of pathways is supported using visualization widgets from Reactome (Gillespie et al. 2022) and from GO Causal Annotation Model (GO-CAM) curation (Thomas et al. 2019) which have been integrated into relevant gene pages on the Alliance web portal (Fig. 6). GO-CAMs are models of biological processes constructed by linking together individual GO annotations. The simplified models shown on Alliance gene pages are linked to model details at the Gene Ontology resource. The interactive Reactome pathways and reaction graphics on Alliance gene pages are linked to the Reactome database for additional information about the reactions and pathways.

Phenotype annotations

If an organism has curated phenotype annotations, the annotations are displayed in a tabular format in the Phenotypes section of the gene detail page. The table includes columns for the phenotype term from the relevant phenotype ontology, annotation details, and the reference(s) for the annotation. Experimental conditions such as chemical, dietary, or physical interventions that contribute to or modify an observed phenotype are included in such annotation details are available. Although the display format is uniform for all organisms represented in the Alliance, the details displayed for phenotype annotations differ by organism. In mouse, for example, phenotype annotations are associated with genotypes and genetic backgrounds. In zebrafish, the phenotype annotations are associated with a fish. In *Drosophila*, phenotype annotations are associated with alleles.

Disease associations and models

Similar to the display for functional annotations, the summary of disease associations for a gene across available data across model organisms is displayed as an interactive ribbon (Fig. 7). Selecting a column in the ribbon generates a tabular summary of the annotations that includes annotation type, evidence, and source. Of the more than 350,000 disease model annotations available in version 5.4.0 of the Alliance website (Table 1), over 60,000 are from experimentally derived models. More than 28,800 annotations represent either biomarkers of disease or disease associations based on orthology to a human gene.

Disease Models are specific strains, genotypes, animals, etc. that support investigation into the genetic and genomic basis of phenotypes and disease. Models defined as genotypes that are associated with specific observable

Table 2 Ontologies used within the alliance of genome resources to annotate function, phenotype, and disease properties associated with genome features, gene products, genetically defined populations, and other biological entities

Ontology	Abbreviation	Alliance knowledge center						
		ZFIN	MGD	SGD	WormBase	FlyBase	RGD	Xenbase
Ascomycete phenotype ontology	APO			✓				
Biological spatial ontology	BSPO	✓						
C. elegans (nematode) life stage	WBls				✓			
C. elegans anatomy	WBbt				✓			
C. elegans phenotype	WBPhenotype				✓			
Cell ontology	CL		✓				✓	
Chemical entities of biological interest	ChEBI	✓	✓	✓	✓		✓	
Clinical measurement ontology	CMO						✓	
Drosophila development	FBdv					✓		
Drosophila gross anatomy	FBdt					✓		
Drosophila phenotype ontology	DPO					✓		
Embrace data and methods	EDAM			✓				
Evidence and conclusion ontology	ECO	✓	✓	✓			✓	
Experimental condition ontology	XCO						✓	
FlyBase controlled vocabulary	FBcv					✓		
Gene ontology	GO	✓	✓	✓	✓	✓	✓	✓
Human disease ontology	DOID	✓	✓	✓		✓	✓	✓
Human phenotype ontology	HP		✓				✓	
Mammalian phenotype ontology	MP		✓				✓	
Measurement method ontology	MMO						✓	
Molecular interactions	MI			✓		✓	✓	
Mouse adult gross anatomy	MA		✓				✓	
Mouse developmental stage ontology	Mmusdv		✓					
Mouse gross anatomy and development, timed	EMAPA		✓					
Mouse pathology	MPATH	✓	✓					
Ontology for biomedical investigations	OBI			✓				
Pathway ontology	PW						✓	
Phenotype and trait ontology	PATO	✓	✓		✓			
Protein modification	MOD			✓				
Protein ontology	PRO		✓					
Rat Strain ontology	RS						✓	
Relations ontology	RO	✓	✓	✓	✓			
Sequence types and features	SO	✓	✓	✓		✓	✓	
Uberon	Uberon		✓				✓	
Vertebrate trait ontology	VT		✓					
Xenopus anatomy ontology	XAO							✓
Xenbase experimental data ontology	XBED							✓
Xenopus phenotype ontology	XPO							✓
Xenopus small molecule ontology	XSMO							✓
Zebrafish anatomy	ZFA	✓						
Zebrafish developmental stages	ZFS	✓						
Zebrafish experimental conditions ontology	ZECO	✓						

phenotypes and/or that have characteristics that reflect biological properties of specific human diseases or syndromes. As the harmonization for the concept of a model across different model systems is still being discussed, the details for model genotypes displayed on Alliance gene detail page are available as links back to the relevant Knowledge Center.

Species	Gene Symbol	Count	Best	Best Reverse	Ensembl Compara	HGNC	Hieranoid	InParanoid	OMA	OrthoFinder	OrthoInspector	PANTHER	PhylomeDB	SonicParanoid	Xenbase	ZFIN
Mus musculus	Polr2g	10 of 10	Yes	Yes	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑		
Rattus norvegicus	Polr2g	10 of 10	Yes	Yes	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑		
Xenopus laevis	polr2g.L	1 of 1	Yes	Yes											☑	
Xenopus tropicalis	polr2g	8 of 9	Yes	Yes	☑		☑	☑	☑	☑	☑	☑	☑	□		
Danio rerio	polr2g	10 of 10	Yes	Yes	☑		☑	☑	☑	☑	☑	☑	☑	☑		☑
Drosophila melanogaster	Polr2G	8 of 9	Yes	Yes	☑		☑	☑	☑	☑	☑	☑	☑	□		
Caenorhabditis elegans	rpb-7	9 of 9	Yes	Yes	☑		☑	☑	☑	☑	☑	☑	☑	☑		
Saccharomyces cerevisiae	RPB7	9 of 9	Yes	Yes	☑		☑	☑	☑	☑	☑	☑	☑	☑		

Fig. 5 Screenshot of the orthology summary for the human *POLR2G* gene. By default, only orthologs with high stringency (as determined by DIOPT score) are displayed. Users can adjust the settings for stringency, species displayed, and orthology call methods used. Alliance orthologs are also available as a downloadable file and via the Alliance API

Alleles, variants, and transgenes

Alleles and sequence variants for a gene are provided in a table format. Variants are defined as sequence differences at a single position or in contiguous nucleotides relative to a reference sequence. Variants are expressed in standard Human Genome Variation Society (den Dunnen et al. 2016) format and annotated with the variant type (e.g., SNP), variant identifier, and molecular consequence annotations generated by the Variant Effect Predictor (VEP) tool (McLaren et al. 2016). Alleles are defined as alternative forms of a gene and may be associated one or more sequence variants. Alleles are displayed with official nomenclature and synonyms and are linked to detail pages at the Alliance that summarize phenotype and disease associations when relevant.

The details provided for transgenic alleles varies by model organism but may include the symbol, the transgene construct, expressed components, knock-down targets, and regulatory regions. Transgenic alleles are linked to Alliance detail pages that provide transgene construct details and summaries of any annotated phenotype and/or disease associations.

Sequence feature viewer

Every gene detail page includes a graphical summary of transcripts annotated to the gene. When relevant, the location of sequence variants associated with alleles of the gene are also displayed. Population level variants (polymorphisms) determined by high-throughput sequencing and/or large-scale genotyping technologies are not displayed in the feature viewer because of the volume and density of these data. High throughput variants and additional genome features can be viewed using the Alliance JBrowse instance (Buels et al. 2016). A link to JBrowse is provided under the sequence feature viewer widget.

Gene expression

Expression data displayed using the standard Alliance interactive ribbon display. The data reflect developmental and cellular/anatomical expression of genes in wild-type backgrounds. As with the other comparative annotation ribbons on the gene detail pages, the shading of the cells is indicative of the number of annotations supporting the expression information, not levels of transcription or translation. Cells with red slashes indicate that a particular anatomical structure is not biologically relevant for the organism.
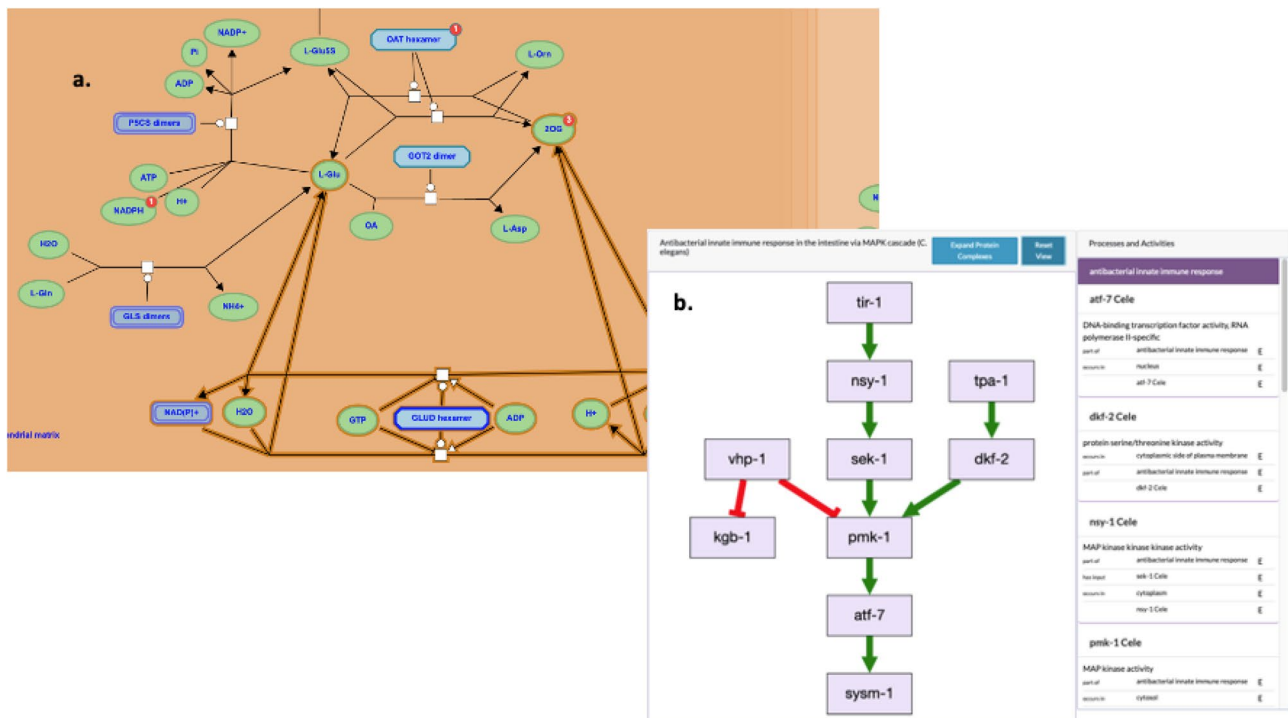
Fig. 6 Screenshots showing pathway representation in the Alliance web portal. **a** Pathway summary display for glutamate and glutamine metabolism for mouse *Glud1* from Reactome. The display is integrated into the gene detail pages and links out to the Reactome resource. **b** Gene Ontology causal activity model (GO-CAM) representation of the antibacterial innate immune response in the intestine via MAPK cascade biological process associated with the nsy-1 gene from *Caenorhabditis elegans*. The GO-CAM models link out to the Gene Ontology website

Molecular and genetic interactions

Information about molecular and genetic interactions are available for genes from humans and all seven model organisms in the Alliance. The interaction data include curated information provided by two Alliance Knowledge Centers (WormBase and FlyBase) and two external interaction data resources: BioGrid (Oughtred et al. 2021) and IMex (Porras et al. 2022). Currently the interaction data are presented as a table but future plans for Alliance Central include the implementation of a graphical display for these data.

Relating model organisms to human disease

Human disease annotations in the Alliance use standardized terms from the Disease Ontology (Schriml et al. 2022). As of version 5.4.0 of the Alliance database, over 11,200 DO terms are associated with model organism and/or human annotations (Table 1). Using the Alliance web portal, researchers can access disease annotations in one of two ways: keyword searching or via the interactive annotation ribbon graphic provided on gene detail pages.

To facilitate keyword searches for a disease of interest, an autocomplete function dynamically generates a list of potential matching terms as the user types a term into the search form. Disease detail pages include a summary header containing the disease term definition from DO with cross references to other terminologies and ontologies. The multi-organism annotations available on disease detail pages include associated genes, alleles, and disease models. For associated genes, the nature of the association, the type of evidence for the association and the source(s) used to support the gene association are provided in an interactive table (Fig. 8). The types of disease gene associations listed include genes *implicated* as causal for a disease and those that are associated as *biomarkers* for a disease. Users can filter the rows displayed in the associated genes table by any of the column headers and can customize the sorting order by disease, gene, or species. Similar information and filter and sort functionality are provided for the associated alleles and associated models tables. Each of the tables shown on the disease detail pages can be downloaded as a tab-separated file.

Fig. 7 Partial screenshot of the comparative disease annotation summary on the *BMP4* gene detail page. The comparative annotation ribbon display is linked to a summary of disease associations for the *BMP4* gene. Disease annotations include the type of association, the disease ontology term, the evidence for the association, and the source(s) of evidence

Fig. 8 Partial screenshot of the results for a search of the disease term "congenital diaphragmatic hernia". The results include genes associated with the disease, type of association and evidence, and the source(s) for the evidence

Extending the platform

One of the overarching goals of Alliance Central is to establish a knowledgebase platform capable of supporting model

organism communities beyond the founding members of the Alliance consortium. A number of the software components developed by the Alliance have been adopted by external database resources, including the short gene descriptions, the Sequence Feature Viewer, and the annotation ribbon

display. To demonstrate the extensibility of the platform to other model systems, the Alliance family of model organism databases recently was extended to include Xenbase, the model organism database for *Xenopus* sp.

Xenopus is a tetrapod model organism that occupies a key evolutionary position between the mammalian models and zebrafish already represented in the Alliance. Two species of Xenopus are now represented in the Alliance: The African clawed frog (*X. laevis*) and the Western clawed frog (*X. tropicalis*). Both *Xenopus* species are widely studied as models for developmental and cell biology. The African clawed frog, *X. laevis* (abbreviated Xla in the Alliance) is an allotetraploid ($2n = 36$) of hybrid origin. The resulting *X. laevis* genome has a set of 'long' and 'short' chromosomes and gene symbols are therefore appended with '.L' or '.S' denoting on which chromosome pair they reside. The second *Xenopus* species, the Western clawed frog *X. tropicalis* (Xtr), is a conventional diploid ($2n = 20$), and is increasingly used in modeling of human disease.

A key step in the integration of Xenopus into the Alliance was the modification of the representation of orthologs in the Alliance. Orthology assertions for *X. tropicalis* were generated from DIOPT. Orthology assertions for *X. laevis* were provided by Xenbase curators and are displayed as coming from the source, "Xenbase" on the orthology summary table. Data for both of the *Xenopus* species are available on gene detail pages, including feature gene descriptions, relationships to orthologs in other model organisms, disease associations for frog genes, gene expression, and a Sequence Feature Viewer.

User support and community engagement

User support and engagement for the Alliance features a Help Desk, tutorials, an active social media presence, and an on-line discussion forum. Through the on-line forum researchers can share announcements about upcoming meetings and job postings and initiate dialog about organism-specific reagents and methods on the forum. From the Help menu on the Alliance home page researchers will find an extensive FAQ, glossary of terms, video tutorials, and on-line documentation describing how to access and use Alliance resources.

The Alliance offers workshops comprised lectures, demos, and interactive tutorials on a regular basis. Workshops are customized to the research interests and needs of the audience. To inquire about hosting an Alliance workshop (virtually or in person), email help@alliancegenome.org.

Access to the primary community engagement sites for the Alliance of Genome Resources consortium are as follows:

- Email access: help@alliancegenome.org
- Discussion forum: https://community.alliancegenome.org/
- Facebook: https://www.facebook.com/AllianceOfGenomeResources
- Twitter: https://twitter.com/alliancegenome
- YouTube: https://www.youtube.com/@allianceofgenomeresources9696/featured
- GitHub: https://github.com/alliance-genome

Citing the alliance

For a general citation of the Alliance, researchers can cite this manuscript. For citing specific data or annotations, the recommended citation format is as follows:

[Type of] data for this paper were retrieved from the Alliance of Genome Resources, URL: https://www.alliancegenome.org; [date the data were retrieved and the release version of the resource].

The release version of the resource is found in the header of every web page (currently, 5.4.0).

Summary and future directions

Prior to the formation of the Alliance of Genome Resources consortium and the Alliance Central knowledge commons platform, researchers seeking to compare biological and functional annotations across different model organisms were typically faced with the daunting task of navigating multiple web sites, each with its own unique style for user interfaces and APIs for programmatic data access. The Alliance of Genome Resources is transforming comparative genomics through the implementation of uniform display of and access to harmonized genetic and genomic data across diverse model organisms and human. Alliance resources allow researchers to easily find, access, compare, and analyze data across multiple species. The modular nature of the Alliance Central platform is designed specifically to allow extension of the resource to other model organisms which will benefit model organism research communities that lack centralized informatics resources by providing cost-effective infrastructure and data management practices that conform to FAIR principles (Wilkinson et al. 2016).

Future directions for the Alliance include the incorporation of additional intensively-studied model organisms into the platform, the continued harmonization of biological

concepts, and refinement and expansion of novel interfaces and analysis tools in support of comparative biology and genomics. A major initiative currently underway within Alliance Central is the implementation of a centralized literature curation system that uses machine learning and artificial intelligence methods to (1) identify published manuscripts with data relevant to the mission of the Alliance and (2) map concepts and entities described in scientific publications to standard nomenclatures and ontology terms. This initiative builds on a large body of prior work among Alliance members to improve the efficiency and scalability of expert curation of knowledge published in the scientific literature (Hirschman et al. 2010; Karamanis et al. 2008; Liu et al. 2015; Muller et al. 2018; Ringwald et al. 2022).

Data availability The annotations available from the Alliance of Genome Resources web portal (https://alliancegenome.org) are distributed under a CC BY 4.0 license.

Declarations

Conflict of interest The authors declare no competing interests.

References

Alliance of Genome Resources C (2019) The alliance of genome resources: building a modern data ecosystem for model organism databases. Genetics 213:1189–1196

Alliance of Genome Resources C (2020) Alliance of genome resources portal: unified model organism research platform. Nucleic Acids Res 48:D650–D658

Alliance of Genome Resources C (2022) Harmonizing model organism data in the alliance of genome resources. Genetics. https://doi.org/10.1093/genetics/iyad022

Anderson WP, Apweiler R, Bateman A, Bauer GA, Herman B, Blake JA, Blomberg N, Burley SK, Cochrane G, Di Francesco V, Donohue T, Durinx C, Game A, Green E, Gojobori T, Goodhand P, Hamosh A, Hermjakob H, Kanehisa M, Kiley R, McEntyre J, McKibbin R, Miyano S, Pauly B, Perrimon N, Ragan MA, Richards G, Teo Y-Y, Westerfield M, Westhof E, Lasko PF (2017) Data management: a global coalition to sustain core data. Nature 543:179

Bradford YM, Van Slyke CE, Howe DG, Fashena D, Frazer K, Martin R, Paddock H, Pich C, Ramachandran S, Ruzicka L, Singer A, Taylor R, Tseng WC, Westerfield M (2023) From multiallele fish to nonstandard environments, how ZFIN assigns phenotypes, human disease models, and gene expression annotations to genes. Genetics. https://doi.org/10.1093/genetics/iyad032

Buels R, Yao E, Diesh CM, Hayes RD, Munoz-Torres M, Helt G, Goodstein DM, Elsik CG, Lewis SE, Stein L, Holmes IH (2016) JBrowse: a dynamic web platform for genome visualization and analysis. Genome Biol 17:66

Davis P, Zarowiecki M, Arnaboldi V, Becerra A, Cain S, Chan J, Chen WJ, Cho J, da Veiga BE, Diamantakis S, Gao S, Grigoriadis D, Grove CA, Harris TW, Kishore R, Le T, Lee RYN, Luypaert M, Muller HM, Nakamura C, Nuin P, Paulini M, Quinton-Tulloch M, Raciti D, Rodgers FH, Russell M, Schindelman G, Singh A, Stickland T, Van Auken K, Wang Q, Williams G, Wright AJ, Yook K, Berriman M, Howe KL, Schedl T, Stein L, Sternberg PW (2022) WormBase in 2022-data, processes, and tools for analyzing *Caenorhabditis elegans*. Genetics. https://doi.org/10.1093/genetics/iyad003

den Dunnen JT, Dalgleish R, Maglott DR, Hart RK, Greenblatt MS, McGowan-Jordan J, Roux AF, Smith T, Antonarakis SE, Taschner PE (2016) HGVS recommendations for the description of sequence variants: 2016 update. Hum Mutat 37:564–569

Dunn NA, Unni DR, Diesh C, Munoz-Torres M, Harris NL, Yao E, Rasche H, Holmes IH, Elsik CG, Lewis SE (2019) Apollo: democratizing genome annotation. PLoS Comput Biol 15:e1006790

Engel SR, Wong ED, Nash RS, Aleksander S, Alexander M, Douglass E, Karra K, Miyasato SR, Simison M, Skrzypek MS, Weng S, Cherry JM (2022) New data and collaborations at the *Saccharomyces* genome database: updated reference genome, alleles, and the alliance of genome resources. Genetics. https://doi.org/10.1093/genetics/iyad224

Fisher M, James-Zorn C, Ponferrada V, Bell AJ, Sundararaj N, Segerdell E, Chaturvedi P, Bayyari N, Chu S, Pells T, Lotay V, Agalakov S, Wang DZ, Arshinoff BI, Foley S, Karimi K, Vize PD, Zorn AM (2023) Xenbase: key features and resources of the Xenopus model organism knowledgebase. Genetics. https://doi.org/10.1093/genetics/iyad018

Gene Ontology C, Aleksander SA, Balhoff J, Carbon S, Cherry JM, Drabkin HJ, Ebert D, Feuermann M, Gaudet P, Harris NL, Hill DP, Lee R, Mi H, Moxon S, Mungall CJ, Muruganugan A, Mushayahama T, Sternberg PW, Thomas PD, Van Auken K, Ramsey J, Siegele DA, Chisholm RL, Fey P, Aspromonte MC, Nugnes MV, Quaglia F, Tosatto S, Giglio M, Nadendla S, Antonazzo G, Attrill H, Dos Santos G, Marygold S, Strelets V, Tabone CJ, Thurmond J, Zhou P, Ahmed SH, Asanitthong P, Luna Buitrago D, Erdol MN, Gage MC, Ali Kadhum M, Li KYC, Long M, Michalak A, Pesala A, Pritazahra A, Saverimuttu SCC, Su R, Thurlow KE, Lovering RC, Logie C, Oliferenko S, Blake J, Christie K, Corbani L, Dolan ME, Drabkin HJ, Hill DP, Ni L, Sitnikov D, Smith C, Cuzick A, Seager J, Cooper L, Elser J, Jaiswal P, Gupta P, Jaiswal P,

Naithani S, Lera-Ramirez M, Rutherford K, Wood V, De Pons JL, Dwinell MR, Hayman GT, Kaldunski ML, Kwitek AE, Laulederkind SJF, Tutaj MA, Vedi M, Wang SJ, D'Eustachio P, Aimo L, Axelsen K, Bridge A, Hyka-Nouspikel N, Morgat A, Aleksander SA, Cherry JM, Engel SR, Karra K, Miyasato SR, Nash RS, Skrzypek MS, Weng S, Wong ED, Bakker E, Berardini TZ, Reiser L, Auchincloss A, Axelsen K, Argoud-Puy G, Blatter MC, Boutet E, Breuza L, Bridge A, Casals-Casas C, Coudert E, Estreicher A, Livia Famiglietti M, Feuermann M, Gos A, Gruaz-Gumowski N, Hulo C, Hyka-Nouspikel N, Jungo F, Le Mercier P, Lieberherr D, Masson P, Morgat A, Pedruzzi I, Pourcel L, Poux S, Rivoire C, Sundaram S, Bateman A, Bowler-Barnett E, Bye AJH, Denny P, Ignatchenko A, Ishtiaq R, Lock A, Lussi Y, Magrane M, Martin MJ, Orchard S, Raposo P, Speretta E, Tyagi N, Warner K, Zaru R, Diehl AD, Lee R, Chan J, Diamantakis S, Raciti D, Zarowiecki M, Fisher M, James-Zorn C, Ponferrada V, Zorn A, Ramachandran S, Ruzicka L, Westerfield M (2023) The gene ontology knowledge-base in 2023. Genetics. https://doi.org/10.1093/genetics/iyad031

Gillespie M, Jassal B, Stephan R, Milacic M, Rothfels K, Senff-Ribeiro A, Griss J, Sevilla C, Matthews L, Gong C, Deng C, Varusai T, Ragueneau E, Haider Y, May B, Shamovsky V, Weiser J, Brunson T, Sanati N, Beckman L, Shao X, Fabregat A, Sidiropoulos K, Murillo J, Viteri G, Cook J, Shorser S, Bader G, Demir E, Sander C, Haw R, Wu G, Stein L, Hermjakob H, D'Eustachio P (2022) The reactome pathway knowledgebase 2022. Nucleic Acids Res 50:D687–D692

Gramates LS, Agapite J, Attrill H, Calvi BR, Crosby MA, Dos Santos G, Goodman JL, Goutte-Gattat D, Jenkins VK, Kaufman T, Larkin A, Matthews BB, Millburn G, Strelets VB, The FlyBase C (2022) FlyBase: a guided tour of highlighted features. Genetics. https://doi.org/10.1093/genetics/iyac035

Hamosh A, Amberger JS, Bocchini C, Scott AF, Rasmussen SA (2021) Online mendelian inheritance in man (OMIM(R)): victor McKusick's magnum opus. Am J Med Genet A 185:3259–3265

Hirschman J, Berardini TZ, Drabkin HJ, Howe D (2010) A MOD(ern) perspective on literature curation. Mol Genet Genomics 283:415–425

Howe DG, Blake JA, Bradford YM, Bult CJ, Calvi BR, Engel SR, Kadin JA, Kaufman TC, Kishore R, Laulederkind SJF, Lewis SE, Moxon SAT, Richardson JE, Smith C (2018) Model organism data evolving in support of translational medicine. Lab Anim (NY) 47:277–289

Hu Y, Flockhart I, Vinayagam A, Bergwitz C, Berger B, Perrimon N, Mohr SE (2011) An integrative approach to ortholog prediction for disease-focused and other functional studies. BMC Bioinform 12:357

Karamanis N, Seal R, Lewin I, McQuilton P, Vlachos A, Gasperin C, Drysdale R, Briscoe T (2008) Natural language processing in aid of FlyBase curators. BMC Bioinform 9:193

Kishore R, Arnaboldi V, Van Slyke CE, Chan J, Nash RS, Urbano JM, Dolan ME, Engel SR, Shimoyama M, Sternberg PW, Genome Resources TAO (2020) Automated generation of gene summaries at the alliance of genome resources. Database. https://doi.org/10.1093/database/baaa037

Kohler S, Gargano M, Matentzoglu N, Carmody LC, Lewis-Smith D, Vasilevsky NA, Danis D, Balagura G, Baynam G, Brower AM, Callahan TJ, Chute CG, Est JL, Galer PD, Ganesan S, Griese M, Haimel M, Pazmandi J, Hanauer M, Harris NL, Hartnett MJ, Hastreiter M, Hauck F, He Y, Jeske T, Kearney H, Kindle G, Klein C, Knoflach K, Krause R, Lagorce D, McMurry JA, Miller JA, Munoz-Torres MC, Peters RL, Rapp CK, Rath AM, Rind SA, Rosenberg AZ, Segal MM, Seidel MG, Smedley D, Talmy T, Thomas Y, Wiafe SA, Xian J, Yuksel Z, Helbig I, Mungall CJ, Haendel MA, Robinson PN (2021) The human phenotype ontology in 2021. Nucleic Acids Res 49:D1207–D1217

Liu W, Laulederkind SJ, Hayman GT, Wang SJ, Nigam R, Smith JR, De Pons J, Dwinell MR, Shimoyama M (2015) OntoMate: a text-mining tool aiding curation at the rat genome database. Database. https://doi.org/10.1093/database/bau129

McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, Flicek P, Cunningham F (2016) The ensembl variant effect predictor. Genome Biol 17:122

Muller HM, Van Auken KM, Li Y, Sternberg PW (2018) Textpresso central: a customizable platform for searching, text mining, viewing, and curating biomedical literature. BMC Bioinform 19:94

Nevers Y, Jones TEM, Jyothi D, Yates B, Ferret M, Portell-Silva L, Codo L, Cosentino S, Marcet-Houben M, Vlasova A, Poidevin L, Kress A, Hickman M, Persson E, Pilizota I, Guijarro-Clarke C, EttQfOC O, Iwasaki W, Lecompte O, Sonnhammer E, Roos DS, Gabaldon T, Thybert D, Thomas PD, Hu Y, Emms DM, Bruford E, Capella-Gutierrez S, Martin MJ, Dessimoz C, Altenhoff A (2022) The quest for orthologs orthology benchmark service in 2022. Nucleic Acids Res 50:W623–W632

Oughtred R, Rust J, Chang C, Breitkreutz BJ, Stark C, Willems A, Boucher L, Leung G, Kolas N, Zhang F, Dolma S, Coulombe-Huntington J, Chatr-Aryamontri A, Dolinski K, Tyers M (2021) The BioGRID database: a comprehensive biomedical resource of curated protein, genetic, and chemical interactions. Protein Sci 30:187–200

Porras P, Orchard S, Licata L (2022) IMEx databases: displaying molecular interactions into a single, standards-compliant dataset. Methods Mol Biol 2449:27–42

Priyam A, Woodcroft BJ, Rai V, Moghul I, Munagala A, Ter F, Chowdhary H, Pieniak I, Maynard LJ, Gibbins MA, Moon H, Davis-Richardson A, Uludag M, Watson-Haigh NS, Challis R, Nakamura H, Favreau E, Gomez EA, Pluskal T, Leonard R, Rumpf W, Wurm Y (2019) Sequenceserver: a modern graphical user interface for custom BLAST databases. Mol Biol Evol 36:2922–2924

Rath A, Olry A, Dhombres F, Brandt MM, Urbero B, Ayme S (2012) Representation of rare diseases in health information systems: the orphanet approach to serve a wide range of end users. Hum Mutat 33:803–808

Ringwald M, Richardson JE, Baldarelli RM, Blake JA, Kadin JA, Smith C, Bult CJ (2022) Mouse genome informatics (MGI): latest news from MGD and GXD. Mamm Genome 33:4–18

Schriml LM, Munro JB, Schor M, Olley D, McCracken C, Felix V, Baron JA, Jackson R, Bello SM, Bearer C, Lichenstein R, Bisordi K, Dialo NC, Giglio M, Greene C (2022) The human disease ontology 2022 update. Nucleic Acids Res 50:D1255–D1261

Smigielski EM, Sirotkin K, Ward M, Sherry ST (2000) dbSNP: a database of single nucleotide polymorphisms. Nucleic Acids Res 28:352–355

Smith RN, Aleksic J, Butano D, Carr A, Contrino S, Hu F, Lyne M, Lyne R, Kalderimis A, Rutherford K, Stepan R, Sullivan J, Wakeling M, Watkins X, Micklem G (2012) InterMine: a flexible data warehouse system for the integration and analysis of heterogeneous biological data. Bioinformatics 28:3163–3165

Thomas PD, Hill DP, Mi H, Osumi-Sutherland D, Van Auken K, Carbon S, Balhoff JP, Albou LP, Good B, Gaudet P, Lewis SE, Mungall CJ (2019) Gene ontology causal activity modeling (GO-CAM) moves beyond GO annotations to structured descriptions of biological functions and systems. Nat Genet 51:1429–1433

Vedi M, Smith JR, Thomas Hayman G, Tutaj M, Brodie KC, De Pons JL, Demos WM, Gibson AC, Kaldunski ML, Lamers L, Laulederkind SJF, Thota J, Thorat K, Tutaj MA, Wang SJ, Zacher S, Dwinell MR, Kwitek AE (2023) 2022 updates to the rat genome database: a findable, accessible, interoperable, and reusable (FAIR) resource. Genetics. https://doi.org/10.1093/genetics/iyad042

Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten JW, da Silva Santos LB, Bourne PE,

Bouwman J, Brookes AJ, Clark T, Crosas M, Dillo I, Dumon O, Edmunds S, Evelo CT, Finkers R, Gonzalez-Beltran A, Gray AJ, Groth P, Goble C, Grethe JS, Heringa J, Hoen PA, Hooft R, Kuhn T, Kok R, Kok J, Lusher SJ, Martone ME, Mons A, Packer AL, Persson B, Rocca-Serra P, Roos M, van Schaik R, Sansone SA, Schultes E, Sengstag T, Slater T, Strawn G, Swertz MA, Thompson M, van der Lei J, van Mulligen E, Velterop J, Waagmeester A, Wittenburg P, Wolstencroft K, Zhao J, Mons B (2016) The FAIR guiding principles for scientific data management and stewardship. Sci Data 3:160018

Yates B, Gray KA, Jones TEM, Bruford EA (2021) Updates to HCOP: the HGNC comparison of orthology predictions tool. Brief Bioinform. https://doi.org/10.1093/bib/bbab155

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.