Large-scale integration of model organism and clinical research data can provide a breadth of knowledge not available from individual sources and can provide contextualization of data back to these sources.

The Monarch Initiative (www.monarchinitiative.org) is a collaborative, open science effort that aims to semantically integrate genotype-phenotype data from many species and sources in order to support precision medicine, disease modeling, and mechanistic exploration. Our integrated knowledge graph, analytic tools, and web services enable diverse users to explore relationships between phenotypes and genotypes across species.

The Monarch Initiative:
An integrative data and analytic platform connecting phenotypes to genotypes across species

Christopher J Mungall[1], Julie A McMurry[2], Sebastian Köhler[3], James P. Balhoff[4], Charles Borromeo[5], Matthew Brush[2], Seth Carbon[1], Tom Conlin[2], Nathan Dunn[1], Mark Engelstad[2], Erin Foster[2], JP Gourdine[2], Julius O.B. Jacobsen[6], Dan Keith[2], Bryan Laraway[2], Suzanna E. Lewis[1], Jeremy Nguyen Xuan[1], Kent Shefchek[2], Nicole Vasilevsky[2], Zhou Yuan[5], Nicole Washington[1], Harry Hochheiser[5], Tudor Groza[7], Damian Smedley[8], Peter N. Robinson[3], Melissa A Haendel[2*]

1. Environmental Genomics and Systems Biology, Lawrence Berkeley National Laboratory, Berkeley, CA, USA
2. Department of Medical Informatics and Clinical Epidemiology and OHSU Library, Oregon Health & Science University, Portland, USA.
3. Institute for Medical Genetics and Human Genetics, Charité-Universitätsmedizin Berlin, Augustenburger Platz 1, 13353 Berlin, Germany
4. RTI International, Research Triangle Park, NC, USA
5. Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, PA
6. Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK
7. Kinghorn Centre for Clinical Genomics, Garvan Institute of Medical Research, Darlinghurst, NSW 2010, Australia
8. William Harvey Research Institute, Barts & The London School of Medicine & Dentistry, Queen Mary University of London, Charterhouse Square, London, EC1M 6BQ, UK

*To whom correspondence should be addressed. Tel: 1-503-407-5970; Email: Haendel@ohsu.edu
Present Address: Melissa Haendel, Department of Medical Informatics and Clinical Epidemiology and OHSU Library, Oregon Health & Science University, Portland, USA.

Abstract

The correlation of phenotypic outcomes with genetic variation and environmental factors is a core pursuit in biology and biomedicine. Numerous challenges impede our progress: patient phenotypes may not match known diseases, candidate variants may be in genes that have not been characterized, model organisms may not recapitulate human or veterinary diseases, filling evolutionary gaps is difficult, and many resources must be queried to find potentially significant genotype-phenotype associations. Since non-human organisms have proven instrumental in revealing biological mechanisms, it is necessary to provide advanced informatics tools to identify phenotypically relevant disease models. Large-scale integration of model organism and clinical research data can provide a breadth of knowledge not available from individual sources and can provide contextualization of data back to these sources. The Monarch Initiative (www.monarchinitiative.org) is a collaborative, open science effort that aims to semantically integrate genotype-phenotype data from many species and sources in order to support precision medicine, disease modeling, and mechanistic exploration. Our integrated knowledge graph, analytic tools, and web services enable diverse users to explore relationships between phenotypes and genotypes across species.

Introduction

A fundamental axiom of biology is that phenotypic manifestations of an organism are due to interaction between genotype and environmental factors over time. In the rapidly advancing era of genomic medicine, a critical challenge is to identify the genetic etiologies of Mendelian disease, cancer, and common and complex diseases, and translate basic science to better treatments. Currently, available human data associates less than approximately 51% of known human coding genes with phenotype data (based on OMIM (1), ClinVar (2), Orphanet (3), CTD (4), and the GWAS catalog (5)). See Table 1 for a list of database abbreviations. This coverage can be extended to approximately 89% if phenotypic information from orthologous genes from five of the most well-studied model organisms is included (**Figure 1**).
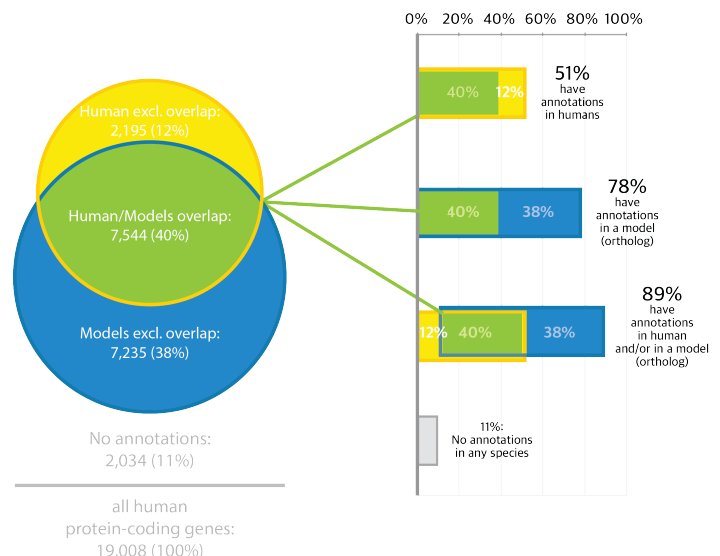

Figure 1. The phenotype annotation coverage of human coding genes. Yellow bars show that 51% of those genes have at least one phenotype association reported in humans (HPO annotations of OMIM, ClinVar, Orphanet, CTD, and GWAS). The blue bars show that 58% of human coding genes have orthologs with causal phenotypic associations reported in at least one non-human model (MGI, Wormbase, Flybase, and ZFIN). The green bars show that 40% of human coding genes have annotations both in human and in non-human orthologs. There are phenotypic associations from humans and/or non-human orthologs that cover 89% of human coding genes.

Similarly, of the 72% of the 3,230 genes in ExAC with 'near-complete depletion of predicted protein-truncating variants have no currently established human disease phenotype' (6), where 88% of these genes without a human phenotype have a phenotype in a non-human organism. However, leveraging these model data for computational use is non-trivial primarily because the relationships between gene and disease (7) and between model system and disease phenotypes (8) are not straightforward.

In recent years, there has been a growth in the number of genotype-phenotype databases available covering a diversity of domain areas for human, model organisms, and veterinary species. While providing quality inventories of the relevant species and phenotypic data types, most resources are limited to a single species or limit cross-species comparison to direct assertions (e.g. Organism X is a model of Disease Y) or based upon orthology relations (e.g. organism Y' is a model of Disease Y due to Y and Y' being orthologs). While great strides have been made in text-based search engines, phenotype data remains difficult to search and use computationally due to its complexity and in the use of different phenotype standards and terminologies. Such barriers have made linking and integration with the precision and richness needed for mechanistic discovery across species a significant challenge (9). A newer method to aid identifying models of disease and to discovery underlying mechanisms is to utilize ontologies to describe the set of phenotypes that present for a given genotype or disease, what we call a 'phenotypic profile'. A phenotypic profile is the subject of non-exact matching within and across species using ontology integration and semantic similarity algorithms (10)(11) in software applications such as Exomiser (12) and Genomiser (13), and this approach has been shown assist disease diagnosis (14–16). The Monarch Initiative uses an ontology-based strategy to deeply integrate genotype-phenotype data from many species and sources, thereby enabling computational interrogation of disease models and complex relationships between genotype and phenotype to be revealed. The Monarch Initiative name was chosen because it is a community effort to create paths for diverse data to be put to use for disease discovery, not unlike the navigation routes that a monarch butterfly would take.

Data Architecture

The overall data architecture for Monarch is shown in **Figure 2**. The bulk of the data integration is carried out using our Data Ingest Pipeline (Dipper) tool (https://github.com/monarch-initiative/dipper), which

maps a variety of external data sources and databases to RDF (Resource Description Framework) graphs. RDF provides a flexible way of modeling a variety of complex datatypes, and allows entities from different databases to be connected via common instance or class URIs (Uniform Resource Indicators). We use relationship types from the Relation Ontology (RO; https://github.com/oborel/obo-relations) (17) and other vocabularies to connect entities together, along with a number of Open Biological Ontologies (18) (OBOs) to classify these entities. For example, a mouse genotype can be related to a phenotype using the *has_phenotype* relation (RO:0002200), with the genotype classified using a term from the Genotype Ontology (GENO) (19), and the phenotype classified using the Mammalian Phenotype Ontology (MP) (20). We use the Open Biomedical Annotations (OBAN; https://github.com/EBISPOT/OBAN) vocabulary to associate evidence and provenance metadata with each edge, using the Evidence and Conclusions Ontology (ECO) for types of evidence (21).The graphs produced by Dipper are available as a standalone resource in RDF/turtle format at http://data.monarchinitiative.org/ttl.
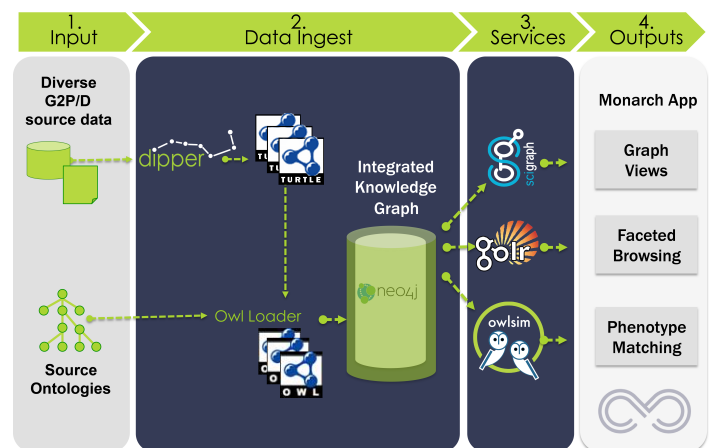

Figure 2. *Monarch Data Architecture. Structured and unstructured data sources are loaded into SciGraph via Dipper. Ontologies are also loaded into SciGraph, resulting in a combined knowledge and data graph. Data is disseminated via SciGraph Services, an ontology-enhanced Solr instance called GOlr, and to the OwlSim semantic similarity software. Monarch applications and end users access the services for graph querying, application population, and phenotype matching.*

We also import a number of external and in-house ontologies, for data description and data integration. As these ontologies are all available from the OBO Library in Web Ontology Language (OWL), no additional transformation is necessary. The combined corpus of graphs ingested using Dipper and from ontologies is referred to as the **Monarch Knowledge Graph**.
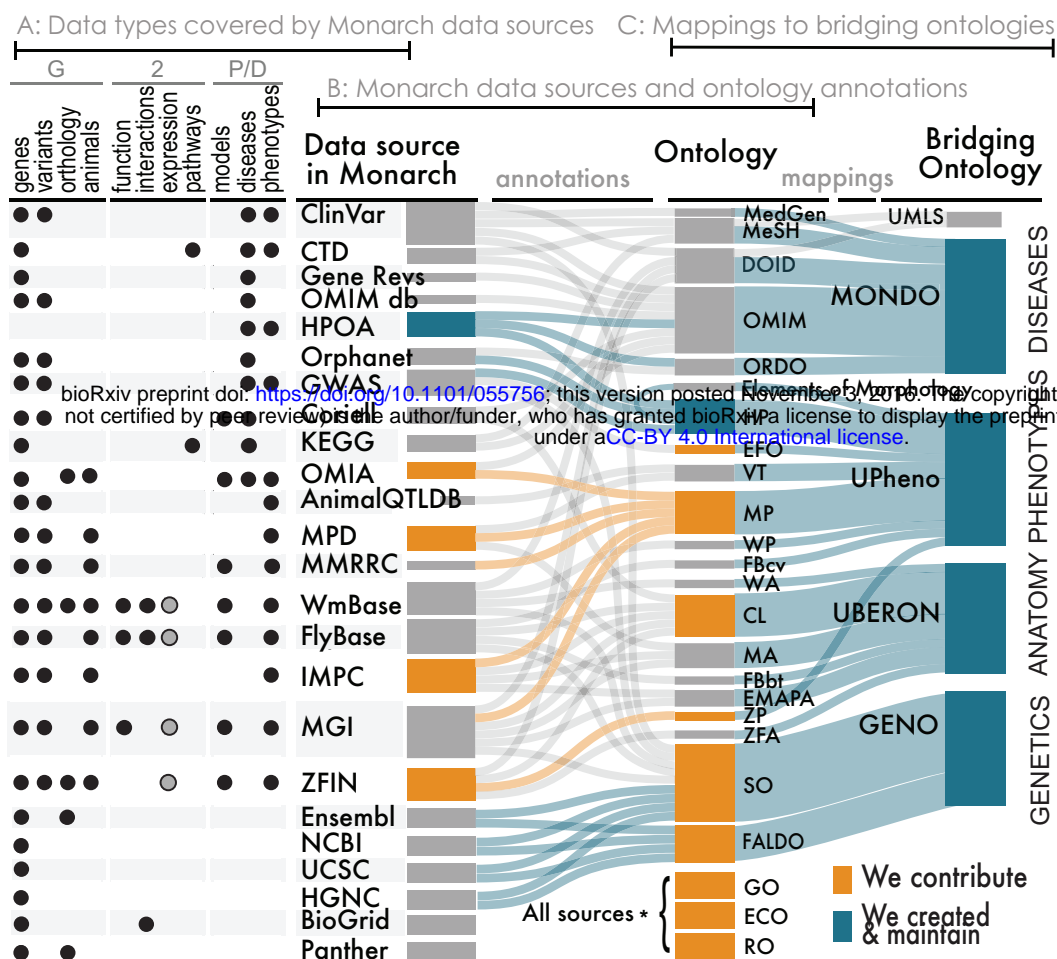
Figure 3. Data types, sources, and the ontologies used for their integration into the Monarch knowledge graph. Each data source uses or is mapped to a suite of different ontologies or vocabularies. These are in turn integrated into bridging ontologies for Genetics/Anatomy (Uberon/CL), Phenotypes (UPheno), and Diseases (MonDO).
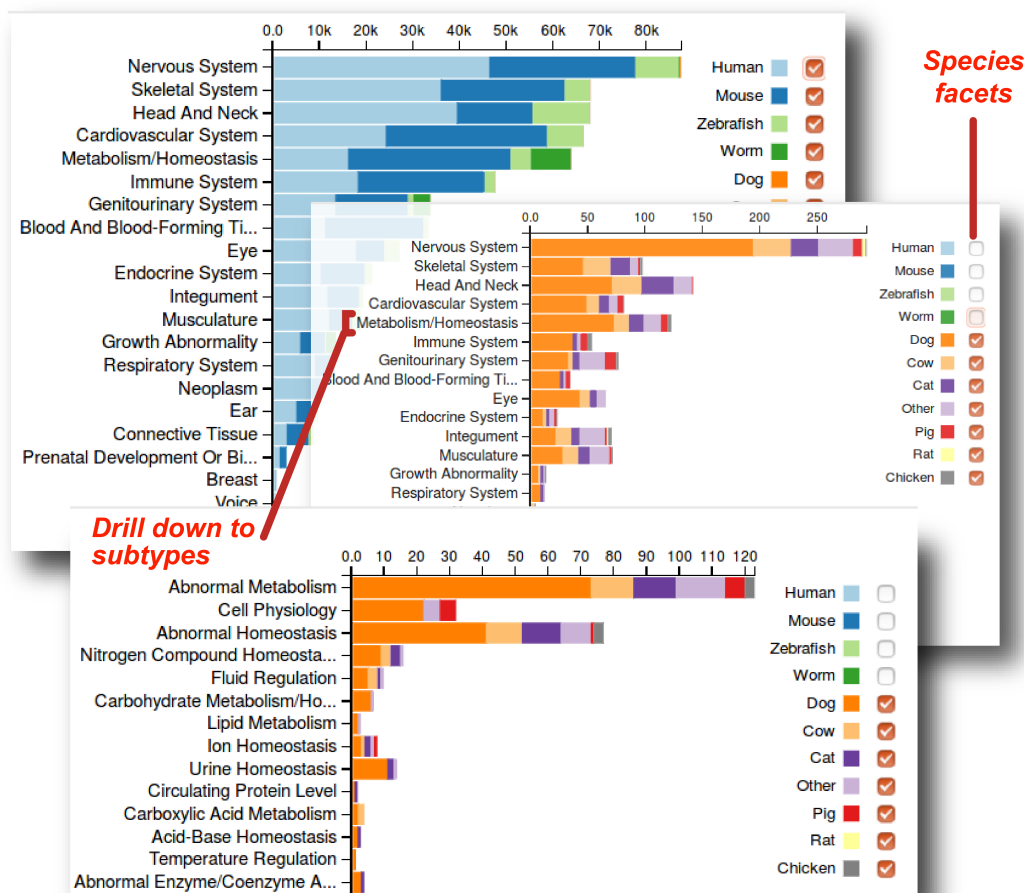

Figure 4. Distribution of phenotypic annotations across species in Monarch, broken down by the top levels of the phenotype ontology. The graph can be interactively explored at https://monarchinitiative.org/phenotype/. Note that annotations are currently dominated by human, mouse, zebrafish and C elegans (top panel); the chart is faceted allowing individual species to be switched on and off to see contributions for less data-rich species such as veterinary animals and monkeys (middle panel). Clicking on a given phenotype text allows drilling down to its subtypes (lower panel).

The data integrated in Monarch wide range of sources, including human clinical knowledge sources as well as genetic and genomic resources covering organismal biology. The list of data sources and ontologies integrated is shown in **Figure 3**, with a species distribution illustrated in **Figure 4**. The knowledge graph is loaded into an instance of a SciGraph database (https://github.com/SciGraph/SciGraph/), which embeds and extends a Neo4J database, allowing for complex queries and ontology-aware data processing

and **Named Entity Recognition**. We provide two public endpoints for client software to query these services:

https://scigraph-ontology.monarchinitiative.org/scigraph/docs (for ontology access) and

https://scigraph-data.monarchinitiative.org/scigraph/docs (for ontology plus data access).

These SciGraph instances provide powerful graph querying capabilities over the complete knowledge graph. Many of the common query patterns are executed in advance and stored in an Apache Solr index, making use of the Gene Ontology 'GOlr' indexing strategy, allowing for fast queries of ontology-indexed associations.

Finally, we also load a subset of the graph into an OwlSim instance, which provides phenotype matching services as well as the ability to perform fuzzy phenotype searches based on a phenotype profile. We also provide phenotype matching services via the Global Alliance for Genomes and Health (GA4GH) Matchmaker Exchange (MME) API MME (22), available at https://mme.monarchinitiative.org.

Many of the data sources we integrate make use of their own terminologies and ontologies. We aggregate these into a unified ontology (https://github.com/monarch-initiative/monarch-ontology/) and make use of bridging ontologies and our curated integrative ontologies to connect these together. In particular:
- The Uber-anatomy ontology (Uberon) bridges species-specific and clinical anatomical and tissue ontologies (23)
- The unified phenotype ontology bridges model organism and human phenotype ontologies and terminologies, using techniques described in (24)(25)
- The Monarch Merged Disease Ontology (MonDO) uses a Bayes ontology merging algorithm (26) to integrate multiple human disease resources into a single ontology, and additionally includes animal diseases from OMIA.

- The Genotype Ontology (GENO) (19) defines genotypic elements and bridges the Sequence Ontology (SO) (27) and FALDO (28). GENO allows the propagation of phenotypes that are annotated differently in different sources to genotypic elements.

Entity resolution and unification

One of the many challenges faced when integrating bioinformatics resources is the presence of the same entity in multiple databases, designated with different identifiers (29). This problem is compounded by the different ways the same identifier can be written, using different prefixes or no prefix at all. Taking a Monarch page for a single gene, for example 'fibrinogen gamma chain', *FGG*, (https://monarchinitiative.org/gene/NCBIGene:2266). Monarch has integrated data from a variety of human, model organism, and other biomedical sources such as OMIM (1), Orphanet (3), ClinVar (2), HPO (30), KEGG (31), CTD (4), MyGene (32), BioGrid (33), and via orthology in PantherDB (34) we also incorporate *Fgg* gene data from ZFIN (35) and from MGI (36). No two of these sources represents the identifier for *FGG* in precisely the same way. As part of our data ingest process, we normalize all identifiers using a curated set of database prefixes. These have a defined mapping to an http URI. These curated prefixes have been deposited in the Prefix Commons (https://github.com/prefixcommons), which similarly contains identifier prefixes used within the Gene Ontology (37) and Bio2RDF (38).

In post-processing equivalent identifiers, we perform clique-merging (https://github.com/SciGraph/SciGraph/wiki/Post-processors). We take all edges labeled with either the owl:sameAs or owl:equivalentClasses property and calculate equivalence cliques, based on the symmetric and transitive nature of these properties. We then merge these cliques together, taking a designated 'clique leader' (for instance, NCBI for genes) and mapping all edges in the monarch graph such that they point to a clique leader.

In-house curation

In addition to ingest of external sources and ontologies, we perform in-house data and ontology curation. For curation of ontology-based genotype-phenotype associations (including disease-phenotypic profiles), we are transitioning to the WebPhenote platform (http://create.monarchinitiative.org), which allows a variety of disease entities to be connected to phenotypic descriptors. We also make use of text

mining to create seed disease-phenotype associations using the Bio-Lark toolkit (39), which are then manually curated. Most recently we have performed a large-scale annotation of PubMed to extract common disease-phenotype associations (40). Most of the in-house curation work involves making smaller resources with free text description of phenotypic information computable, for example, the Online Mendelian Inheritance in Animals (OMIA) resource, with whom we have been collaborating to support this curation (41).

Quality control

External resources and datasets that are incorporated into Monarch are evaluated before incorporation into the Dipper pipeline – we primarily integrate high-quality curated resources. For all ontologies we bring in, we apply automated reasoning to detect inconsistencies between different ontologies. For each release, we perform high-level checks on each integrated resource to ensure no errors in the extraction process occurred, but we do not perform in-depth curation checks of integrated resources. Each release happens once every one to two months.

In order to measure annotation richness, we have also created an annotation sufficiency meter web service (42) available at https://monarchinitiative.org/page/services; this service determines whether a given phenotype profile for any organism is sufficiently broad and deep to be of diagnostic utility. The sufficiency score can be displayed as a five star scale as in PhenoTips (43) and in the Monarch web portal (see below) to aid curation or data entry, and can also be used to suggest additional phenotypic assays to be performed – whether in a patient or in a model organism.

Monarch Web Portal

The Monarch portal is designed with a number of different use cases in mind, including:
- A researcher interested in a human gene, its phenotypes, and the phenotypes of orthologs in model organisms and other species
- Patients or researchers interested in a particular disease or phenotype (or groups of these), together with information on all implicated genes
- A clinical scenario in which a patient has an undiagnosed disease showing a spectrum of phenotypes, with no definitive candidate gene demonstrated by sequencing; in this scenario the

clinician wishes to search for either known diseases that have a similar presentation, or model organism genes that demonstrate homologous phenotypes when the gene is perturbed
- Researcher looking for diseases that have similar phenotypic feature to a newly identified model organism mutant identified in a screen
- Researchers or clinicians who need to identify potentially informative phenotyping assays for differential diagnosis or to identify candidate genes

Features

Integrated information on entities of interest
We provide overview pages for entities such as genes, diseases, phenotypes, genotypes, variants, and publications. Each page highlights the provenance of the data from the diverse clinical, model organism, and non-model organism sources. These pages can be found either via search (see below) or through an entity resolver. For example, the URL https://monarchinitiative.org/OMIM:266510 will redirect to a page about the disease 'Peroxisome biogenesis disorder type 3B' from the OMIM resource, showing its relationships to other content within the Monarch knowledge graph, such as phenotypes and genes associated with the disease. We make use of MonDO (the Monarch merged disease ontology (26)) to group similar diseases together. **Figure 5** shows an example page for Marfan syndrome with related phenotype, gene, model and variant data.

Basic Search
The portal provides different means of searching over integrated content. In cases where a user is interested in a specific disease, gene, phenotype etc., these can usually be found via autocomplete. Site-wide synonym-aware text search can also be used to find pages of interest. Because the knowledgebase combines information from multiple species, entities such as genes often have ambiguous symbols. We provide species information to help disambiguate in a search.

Search by phenotype profile
One of the most innovative features of Monarch is the ability to query within and across species to look for diseases or organisms that share a set of similar but non-exact set of phenotypes (phenotypic profile). This feature uses a semantic similarity algorithm available from the OWLsim package (http://owlsim.org). Users can launch searches against specific targets: organisms,

sets of named gene models, or against all models and diseases available in the Monarch repository. The Monarch Analyze Phenotypes interface (https://monarchinitiative.org/analyze/phenotypes) allows the user to build up a 'cart' of phenotypes, and then perform a comparison against phenotypes related to genes and diseases. Results are ranked according to closeness of match, partitioned by species, and are displayed as both a list and in the Phenogrid widget (**Figure 6**).
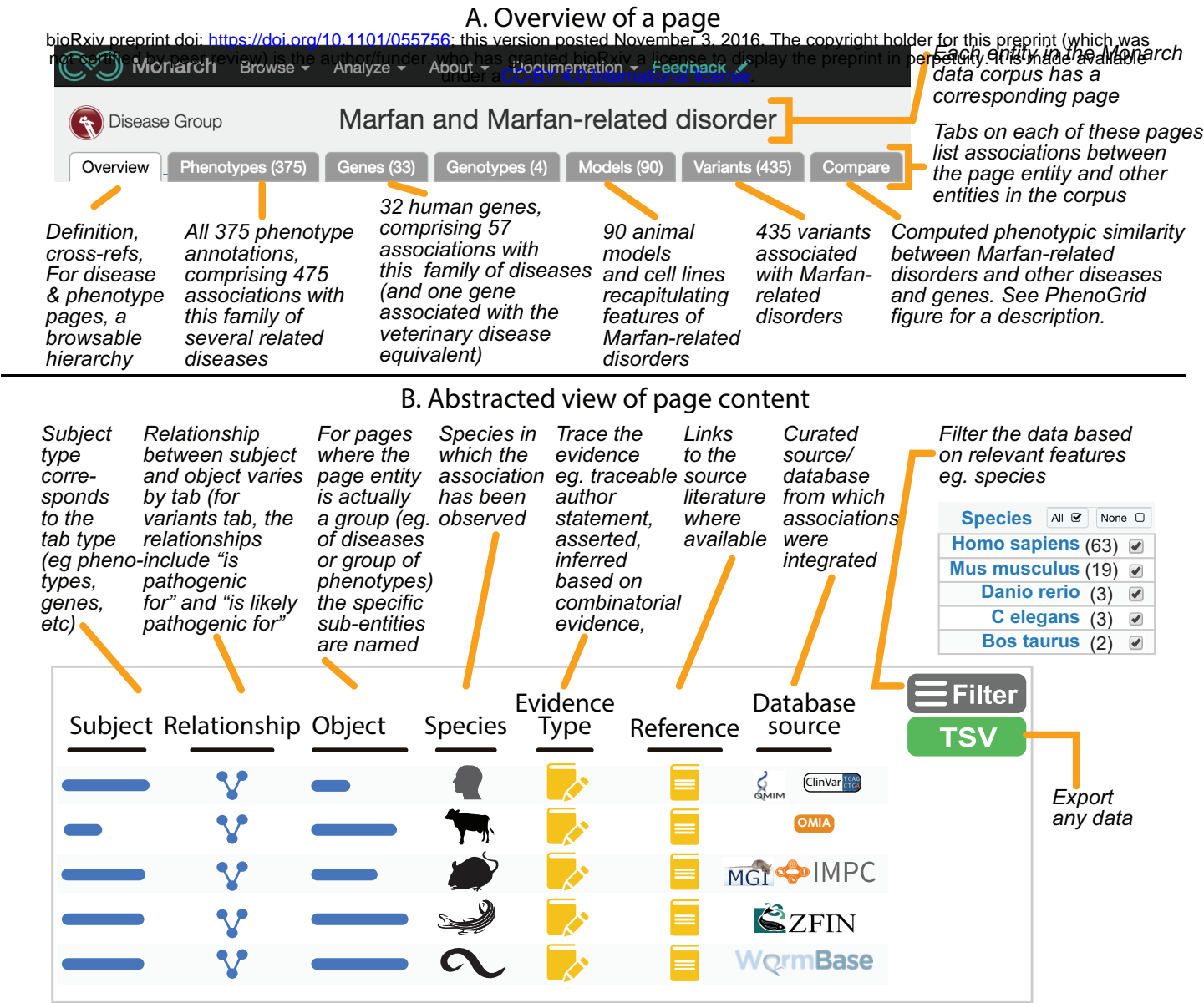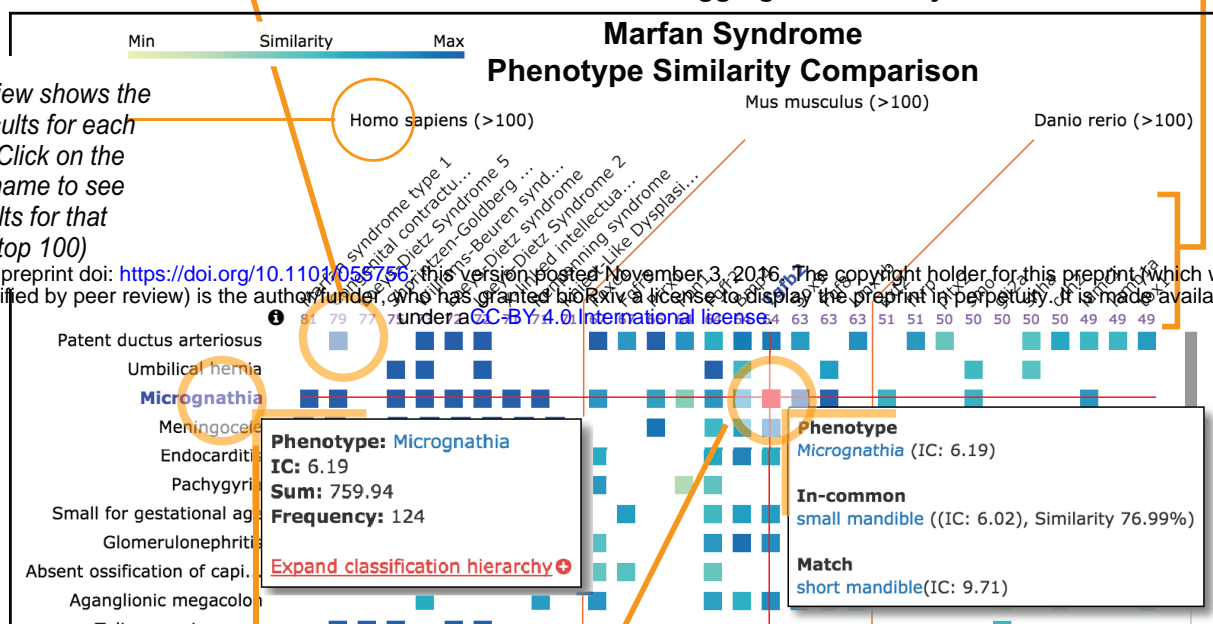
Figure 5. *Annotated Monarch webpage for Marfan and Marfan Related syndrome. This group of syndromic diseases has a number of different associations spanning multiple entity types – disease phenotypes, implicated human genes, variants and animal models and other model systems. An abstraction of the contents and features of the tabs is shown in the lower panel. Actual contents of the tabs are best viewed in the context of the web app at https://monarchinitiative.org/DOID:14323.*

Each square is colored according to how closely matched the pair of phenotypes is

Names of the entities (diseases, genes ...) that are phenotypically similar to Marfan; below each label is its corresponding aggregate similarity score

Default view shows the top 10 results for each species. Click on the species name to see only results for that species (top 100)

**Marfan Syndrome
Phenotype Similarity Comparison**

Min   Similarity   Max

Homo sapiens (>100)   Mus musculus (>100)   Danio rerio (>100)

Patent ductus arteriosus
Umbilical hernia
Micrognathia
Meningocele
Endocarditis
Pachygyria
Small for gestational age
Glomerulonephritis
Absent ossification of capi.
Aganglionic megacolon
Talipes equinovarus

Phenotype: Micrognathia
IC: 6.19
Sum: 759.94
Frequency: 124

Expand classification hierarchy ⊕

Phenotype
Micrognathia (IC: 6.19)

In-common
small mandible ((IC: 6.02), Similarity 76.99%)

Match
short mandible (IC: 9.71)

All of the phenotypes associated with Marfan Syndrome in humans (filtered to show only those shared in phenotypically similar diseases and models)

Hover for phenotype term info

Hover over a match to see the pair of phenotypes that were matched between the disease and the matched model as well as the intermediate phenotype that both have in common, if applicable.

The mouse gene Tgfb2 has a phenotype profile that is 64% similar to that of Marfan. Click on the gene symbol to explore the specific Tgfb2 variants and Tgfb2 genotypes in which the matching phenotypes were described.

Figure 6. *Partial screenshot of PhenoGrid showing Marfan syndrome. PhenoGrid shows input phenotypes in rows, models in columns, and cell contents color-coded with greater saturation indicating greater similarity. Disease phenotypes are shown as rows, and phenotypically matching human diseases and model organism genes are shown as columns – the saturation of a cell correlates with strength if phenotypic match. Mouse-over tooltips highlight diseases associated with a selected phenotype (or vice-versa), or details (including similarity scores) of any match between a phenotype and a model. User controls support the selection of alternative sort orders, similarity metrics, and displayed organism(s) (mouse, human, zebrafish, or the 10 most similar models for each). Here we see all diseases or genes that exhibit 'Hypoplasia of the mandible' with the matching mouse gene Tfgb2. Actual PhenoGrid data is best viewed in the context of the web app at https://monarchinitiative.org/Orphanet:284993#compare.*

Note matches do not need to be exact – here the mouse phenotype of 'small mandible' (Mouse Phenotype Ontology) has a high scoring match to "micrognathia" (Human Phenotype Ontology) based on the fact that both phenotypes are related to "small mandible" (Mouse Phenotype Ontology). Advanced PhenoGrid features (not displayed) include the ability to alter the scoring and sorting methods, as well as zoomed-out map-style navigation.

Phenogrid

Given a set of input phenotypes, as associated with a patient or a disease, Monarch phenotypic profile similarity calculations can generate results involving hundreds of diseases and models. The PhenoGrid visualization widget (**Figure 6**) provides an overview of these similarity results, implemented using the D3 javascript library (44). Phenotypes and models are frequently too numerous to fit on the initial display; thus scrolling, dragging, and filtering have been implemented. PhenoGrid is available as an open-source widget suitable for integration in third-party web sites, such as for model organism databases as done in the International Mouse Phenotyping Consortium (IMPC) or clinical comparison tools. Download and installation instructions are available on the Monarch Initiative web site.

Text Annotation

The Monarch annotation service allows a user to enter free text (for example, a paper abstract or a

clinical narrative) and perform an automated annotation on this text, with entities in the text marked up with terms from the Monarch knowledge graph, such as genes, diseases, and phenotypes. Once the text is marked up, the user has the option of turning the recognized phenotype terms into a phenotype profile, and performing a profile search, or to link to any of the entity pages identified in the annotation. This tool is also available via services.

Inferring causative variants

The Exomiser (45) and more recently, Genomiser tools (46) make use of the Monarch platform phenotype matching algorithms to rank putative causative variants using a combined variant and phenotype score. These tools have been used to diagnose patients as part of the NIH Undiagnosed Diseases Project (14) and are the first examples of using model organism phenotype data to aid rare disease diagnostics.

Discussion

The Monarch Initiative provides a system to organize and harmonize the heterogeneous genotype-phenotype data found across clinical and model and non-model organism resources (such as veterinary species), creating a unified overview of this rich landscape of data sources. Some of the challenges we have had to address are that each resource shares data via different mechanisms and uses a different data model. It is particularly important to note that each organism annotates phenotypic data to different aspects of the genotype – one resource might be to a gene, another an allele, another to a set of alleles, a full genotype or a SNP. This not only makes data integration difficult, but it also means that computation over the genotype-phenotype associations must be done with care. Similar issues at MGI have been described (47). In addition, since most anatomy, phenotype, and disease ontologies describe the biology of one species, it has traditionally been quite difficult to 'map' across species. Some examples are the Human Phenotype Ontology (HPO) (30) and the Mouse Anatomy Ontology (48). Monarch uses four species-neutral ontologies that unify their species-specific counterparts (as shown in Figure 3): GENO for genotypes (19), UPheno for phenotypes (49), UBERON for anatomy (23), and MonDO for diseases (26). Prior efforts to map or integrate species-specific anatomical ontologies (24,50), for example, have been utilized in the construction of these species-neutral ontologies. The end result is a translational platform that allows a unified view of human, model and non-model organism biology.

A comparison between Monarch and existing resources is warranted. InterMine is an open-source data warehouse system used for disseminating data from large, complex biological heterogeneous data sources (51). InterMine provides sophisticated web services to support denormalized query and has been used to improve query and data access to model organism databases (52) and non-model organisms (53). InterMine is a federated approach where individual databases each can adopt and populate their own object-oriented data model, but can also align on certain aspects such as having genomic data models aligned using the SO. However, as yet genotype and phenotype modeling is not aligned, and Intermine does not provide disease matching or phenotypic search. We are currently working with InterMine to achieve harmonization in this area. Other resources, such as KaBOB (54) and Bio2RDF (38) semantically integrate various resources into large triplestores. Bio2RDF typically retains the source vocabulary of the integrated resources, whereas KaBOB is more similar to Monarch in that it maps OBO ontologies (18). Other data integration approaches include the BioThings API, exemplified by the MyVariant system (32) which aggregates variant data from multiple sources. We are currently working with the BioThings API developers to integrate these different approaches within the Dipper framework. Monarch is unique in that it aims to align both genotypic and phenotypic modeling across species and sources.

Future directions

Future directions include bringing in phenotypic data from specialized sources and databases, incorporating a wider range of datatypes, and to extend and improve analytic methods for making cross-species inferences. Currently the core of Monarch includes primarily qualitative phenotypes described using terms from existing phenotypic vocabularies – we are starting to bring in more quantitative data, from sources such as the MPD (55) and GeneNetwork (56), in addition to expression data annotated to Uberon in BgeeDb (57). We are also extending our phenotypic search methods to incorporate Phenologs, phenotypic groupings inferred on the basis of orthologous genes (58,59). Early comparisons suggest that addition of phenologs to our suite of tools to enable genotype-phenotype inquiry across species will extend our reach in a synergistic manner (60). We therefore plan to implement this type of approach into the Monarch tool suite and website. One of the most important realizations we came across in constructing the Monarch platform was the need to better represent scientific evidence of genotype-phenotype

associations. We are currently developing a Scientific Evidence and Provenance Information Ontology (SEPIO) (61) in collaboration with the Evidence and Conclusion Ontology consortium (21) and ClinGen (62) in order to classify associations as complementary, confirmatory, or contradictory. SEPIO will also integrate biological assays from the Ontology of Biomedical Investigations (63). Monarch has also been collaborating with the US National Cancer Institute's Thesaurus (NCIT) team to integrate cancer phenotypes. Finally, Monarch has been working in the context of the Global Alliance for Genomics and Health (GA4GH) to develop a formal phenotype exchange format (www.phenopackets.org) that can aid phenotypic data sharing in numerous contexts such as clinical, model organism research, biodiversity, veterinary, and evolutionary biology.

Glossary of Acronyms

Acronym	Name	URL	Ref	PMID	NAR DB
Bgee	BgeeDb	http://bgee.org/	57		
BioGrid	Biological General Repository for Interaction Datasets.	https://thebiogrid.org/	33	PMID:25428363	D470-D478
CL	Cell Ontology	http://obofoundry.org/ontology/cl.html	64	PMID:27377652	42/D1/D950
ClinVar	ClinVar	https://www.ncbi.nlm.nih.gov/clinvar/	2	PMID:26582918	D862-8
CTD	Clinical Toxicology Database	http://ctdbase.org/	4	PMID:27651457	D914-20
ECO	Evidence and Conclusions Ontology	http://obofoundry.org/ontology/eco.html	21	PMID:25052702	
ExAC	Exome Aggregation Consortium	http://exac.broadinstitute.org/	6	PMID:27535533	
FlyBase	FlyBase	http://flybase.org	65	PMID:26467478	D786-92
GeneNetwork	Gene Network	http://genenetwork.org	56		
GENO	Genotype Ontology	https://github.com/monarch-initiative/GENO-ontology/	19		
GO	Gene Ontology	http://geneontology.org	37	PMID:25428369	D1049-56
GWAS	GWAS Catalog	https://www.ebi.ac.uk/gwas/	5	PMID:24316577	D1001-6
HP	Human Phenotype Ontology	http://human-phenotype-ontology.org/	30	PMID:24217912	D966-D974
KEGG	Kyoto Encyclopedia of Genes and Genomes	http://www.kegg.jp/	31	PMID:24214961	D199-205
MGI	Mouse Genome Informatics	http://www.informatics.jax.org/	36	PMID:26238262	
MonDO	Monarch Merged Disease Ontology	https://github.com/monarch-initiative/monarch-disease-ontology/	26		
MP	Mammalian Phenotype Ontology	http://obofoundry.org/ontology/mp.html	20	PMID:26238262	
MPD	Mouse phenome database	http://phenome.jax.org/	55	PMID:24243846	D825-34
MyGene	MyGene	http://mygene.info	32	PMID:27154141	
OMIA	Online Mendelian Inheritance in Animals	http://omia.angis.org.au/home/	41		
OMIM	Online Mendelian Inheritance in Man	http://omim.org	1	PMID:25428349	D789-98
OrphaNet	Portal for rare diseases and orphan drugs	http://www.orpha.net	3	PMID:22422702	
Panther	PantherDB	http://pantherdb.org	34	PMID:26578592	D336-42
RO	Relation Ontology	http://obofoundry.org/ontology/ro.html	17	PMID: 15892874	
SEPIO	Scientific Evidence and Provenance Information Ontology	https://github.com/monarch-initiative/SEPIO-ontology/	61		
SO	Sequence Ontology	http://www.sequenceontology.org/	27	PMID:26229585	
Uberon	Uber-anatomy ontology	http://uberon.org	23	PMID:25009735	
Upheno	Unified Phenotype Ontology	https://github.com/obophenotype/upheno/	49	PMID:24358873	
WormBase	WormBase	http://wormbase.org	66	PMID:26578572	D774-80
ZFIN	Zebrafish Information Resource	http://zfin.org	35	PMID:26097180	

Funding

References

1. Amberger JS, Bocchini CA, Schiettecatte F, Scott AF, Hamosh A. OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. Nucleic Acids Res [Internet]. 2015 Jan 28 [cited 2015 Apr 1];43(Database issue):D789-98. Available from: http://nar.oxfordjournals.org/content/43/D1/D789.short

2. Landrum MJ, Lee JM, Benson M, Brown G, Chao C, Chitipiralla S, et al. ClinVar: public archive of interpretations of clinically relevant variants. Nucleic Acids Res [Internet]. 2016 Jan 4;44(D1):D862-8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26582918

3. Rath A, Olry A, Dhombres F, Brandt MM, Urbero B, Ayme S. Representation of rare diseases in health information systems: the Orphanet approach to serve a wide range of end users. Hum Mutat [Internet]. 2012 May [cited 2015 Feb 23];33(5):803–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22422702

4. Davis AP, Grondin CJ, Johnson RJ, Sciaky D, King BL, McMorran R, et al. The Comparative Toxicogenomics Database: update 2017. Nucleic Acids Res [Internet]. 2016 Sep 19 [cited 2016 Oct 24]; Available from: http://www.ncbi.nlm.nih.gov/pubmed/27651457

5. Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Res [Internet]. 2014 Jan;42(Database issue):D1001-6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24316577

6. Lek M, Karczewski KJ, Minikel E V., Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature [Internet]. 2016 Aug 17 [cited 2016 Oct 21];536(7616):285–91. Available from: http://www.nature.com/doifinder/10.1038/nature19057

7. Strohman R. Maneuvering in the complex path from genotype to phenotype. Science (80-) [Internet]. 2002/04/27. 2002;296(5568):701–3. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11976445

8. Houle D, Govindaraju DR, Omholt S. Phenomics: the next challenge. Nat Rev Genet [Internet]. 2010/11/19. 2010;11(12):855–66. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21085204

9. McMurry JA, Köhler S, Washington NL, Balhoff JP, Borromeo C, Brush M, et al. Navigating the Phenotype Frontier: The Monarch Initiative. Genetics [Internet]. 2016 Aug [cited 2016 Aug 13];203(4):1491–5. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27516611

10. Smedley D, Oellrich A, Köhler S, Ruef B, Westerfield M, Robinson P, et al. PhenoDigm: analyzing curated annotations to associate animal models with human diseases. Database (Oxford) [Internet]. 2013 Jan [cited 2013 Nov 4];2013:bat025. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3649640&tool=pmcentrez&rendertype=abstract

11. Washington NL, Haendel M a, Mungall CJ, Ashburner M, Westerfield M, Lewis SE. Linking human diseases to animal models using ontology-based phenotype annotation. PLoS Biol [Internet]. 2009 Nov [cited 2012 Jul 30];7(11):e1000247. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2774506&tool=pmcentrez&rendertype=abstract

12. Robinson PN, Kohler S, Oellrich A, Genetics SM, Wang K, Mungall CJ, et al. Improved exome prioritization of disease genes through cross-species phenotype comparison. Genome Res [Internet]. 2014 Feb [cited 2015 Feb 18];24(2):340–8. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3912424&tool=pmcentrez&rendertype=abstract

13. Smedley D, Schubach M, Jacobsen JOBOB, Köhler S, Zemojtel T, Spielmann M, et al. A

Whole-Genome Analysis Framework for Effective Identification of Pathogenic Regulatory Variants in Mendelian Disease. Am J Hum Genet [Internet]. 2016 Sep 3 [cited 2016 Aug 27];99(3):595–606. Available from: http://dx.doi.org/10.1016/j.ajhg.2016.07.005

14. Bone WP, Washington NL, Buske OJ, Adams DR, Davis J, Draper D, et al. Computational evaluation of exome sequence data using human and model organism phenotypes improves diagnostic efficiency. Genet Med [Internet]. 2016 Nov 12 [cited 2015 Nov 10]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26562225

15. Zemojtel T, Kohler S, Mackenroth L, Jager M, Hecht J, Krawitz P, et al. Effective diagnosis of genetic disease by computational phenotype analysis of the disease-associated genome. Sci Transl Med [Internet]. 2014 Sep 3 [cited 2014 Sep 4];6(252):252ra123-252ra123. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25186178

16. Wright CF, Fitzgerald TW, Jones WD, Clayton S, McRae JF, van Kogelenberg M, et al. Genetic diagnosis of developmental disorders in the DDD study: a scalable analysis of genome-wide research data. Lancet (London, England) [Internet]. 2015 Apr 4 [cited 2016 Jul 4];385(9975):1305–14. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25529582

17. Smith B, Ceusters W, Klagges B, Köhler J, Kumar A, Lomax J, et al. Relations in biomedical ontologies. Genome Biol [Internet]. 2005 [cited 2016 Oct 27];6(5):R46. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15892874

18. Smith B, Ashburner M, Rosse C, Bard J, Bug W, Ceusters W, et al. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. Nat Biotechnol. 2007 Nov;25(11):1251–5.

19. Brush M, Mungall CJ, Washington NL, Haendel MA. What's in a Genotype?: An Ontological Characterization for Integration of Genetic Variation Data. In: International Conference on Biomedical Ontology 2013 [Internet]. 2013. Available from: http://ceur-ws.org/Vol-1060/icbo2013_submission_60.pdf

20. Smith CL, Eppig JT. Expanding the mammalian phenotype ontology to support automated exchange of high throughput mouse phenotyping data generated by large-scale mouse knockout screens. J Biomed Semantics [Internet]. 2015 Mar 25 [cited 2015 Mar 26];6(1):11. Available from: http://www.pubmedcentral.nih.gov/articlerende

r.fcgi?artid=4378007&tool=pmcentrez&rendert ype=abstract

21. Chibucos MC, Mungall CJ, Balakrishnan R, Christie KR, Huntley RP, White O, et al. Standardized description of scientific evidence using the Evidence Ontology (ECO). Database [Internet]. 2014 Jul 22 [cited 2014 Jul 22];2014(0):1–11. Available from: http://database.oxfordjournals.org/content/201 4/bau075.full

22. Mungall CJ, Washington NL, Nguyen-Xuan J, Condit C, Smedley D, Köhler S et al. Use of model organism and disease databases to support matchmaking for human disease gene discovery. Hum Mutat [Internet]. 2015 Oct [cited 2015 Dec 19];36(10):979–84. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26269093

23. Haendel MA, Balhoff JP, Bastian FB, Blackburn DC, Blake JA, Bradford Y, et al. Unification of multi-species vertebrate anatomy ontologies for comparative biology in Uberon. J Biomed Semantics [Internet]. 2014 Jan [cited 2015 Apr 30];5(1):21. Available from: http://www.jbiomedsem.com/content/5/1/21

24. Mungall C, Gkoutos G, Smith C, Haendel M, Lewis S, Ashburner M. Integrating phenotype ontologies across multiple species. Genome Biol [Internet]. 2010;11(1):R2. Available from: http://genomebiology.com/2010/11/1/R2

25. Köhler S, Doelken SC, Ruef BJ, Bauer S, Washington N, Westerfield M, et al. Construction and accessibility of a cross-species phenotype ontology along with gene annotations for biomedical research. F1000Research [Internet]. 2013 Feb 1 [cited 2013 Mar 6];2. Available from: http://f1000research.com/articles/2-30/v1

26. Mungall CJ, Koehler S, Robinson P, Holmes I, Haendel M. k-BOOM: A Bayesian approach to ontology structure inference, with applications in disease ontology construction. In: Phenotype Day, ISMB [Internet]. 2016 [cited 2016 Oct 27]. Available from: http://phenoday2016.bio-lark.org/pdf/2.pdf

27. Mungall CJ, Batchelor C, Eilbeck K. Evolution of the Sequence Ontology terms and relationships. J Biomed Inform [Internet]. 2011 Feb [cited 2015 May 3];44(1):87–93. Available from: http://www.pubmedcentral.nih.gov/articlerende r.fcgi?artid=3052763&tool=pmcentrez&rendert ype=abstract

28. Bolleman JT, Mungall CJ, Strozzi F, Baran J, Dumontier M, Bonnal RJP, et al. FALDO: a semantic standard for describing the location

of nucleotide and protein feature annotation. J Biomed Semantics [Internet]. 2016 Dec 13 [cited 2016 Aug 19];7(1):39. Available from: http://jbiomedsem.biomedcentral.com/articles/10.1186/s13326-016-0067-z

29. McMurry J, Muilu J, Dumontier M, Hermjakob H, Conte N, Gormanns P, et al. 10 Simple rules for design, provision, and reuse of identifiers for web-based life science data. 2015 Oct 2 [cited 2016 May 31]; Available from: http://zenodo.org/record/31765

30. Köhler S, Doelken SC, Mungall CJ, Bauer S, Firth H V, Bailleul-Forestier I, et al. The Human Phenotype Ontology project: linking molecular biology and disease through phenotype data. Nucleic Acids Res [Internet]. 2014 Jan [cited 2015 Mar 11];42(Database issue):D966-74. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3965098&tool=pmcentrez&rendertype=abstract

31. Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M. Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic Acids Res [Internet]. 2014 Jan [cited 2016 Oct 27];42(Database issue):D199-205. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24214961

32. Xin J, Mark A, Afrasiabi C, Tsueng G, Juchler M, Gopal N, et al. High-performance web services for querying gene and variant annotation. Genome Biol [Internet]. 2016 May 6 [cited 2016 May 8];17(1):91. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4858870&tool=pmcentrez&rendertype=abstract

33. Chatr-Aryamontri A, Breitkreutz B-J, Oughtred R, Boucher L, Heinicke S, Chen D, et al. The BioGRID interaction database: 2015 update. Nucleic Acids Res [Internet]. 2015 Jan [cited 2016 Oct 24];43(Database issue):D470-8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25428363

34. Mi H, Poudel S, Muruganujan A, Casagrande JT, Thomas PD. PANTHER version 10: expanded protein families and functions, and analysis tools. Nucleic Acids Res [Internet]. 2016 Jan 4 [cited 2016 Oct 27];44(D1):D336-42. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26578592

35. Ruzicka L, Bradford YM, Frazer K, Howe DG, Paddock H, Ramachandran S, et al. ZFIN, The zebrafish model organism database: Updates and new directions. Genesis [Internet]. 2015 Aug [cited 2016 Oct 23];53(8):498–509. Available from:

http://www.ncbi.nlm.nih.gov/pubmed/26097180

36. Eppig JT, Richardson JE, Kadin JA, Ringwald M, Blake JA, Bult CJ. Mouse Genome Informatics (MGI): reflecting on 25 years. Mamm Genome [Internet]. 2015 Aug [cited 2016 Oct 23];26(7–8):272–84. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26238262

37. The Gene Ontology Consortium. Gene Ontology Consortium: going forward. Nucleic Acids Res [Internet]. 2014 Nov 26 [cited 2015 Jan 5];43(Database issue):D1049-56. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25428369

38. Dumontier M, Callahan A, Cruz-Toledo J, Ansell P, Emonet V, Belleau F, et al. Bio2RDF release 3: a larger connected network of linked data for the life sciences. Proceedings of the 2014 International Conference on Posters & Demonstrations Track - Volume 1272. CEUR-WS.org; 2014. p. 401–4.

39. Groza T, Köhler S, Doelken S, Collier N, Oellrich A, Smedley D, et al. Automatic concept recognition using the Human Phenotype Ontology reference and test suite corpora. Database (Oxford) [Internet]. 2015 Jan [cited 2015 Mar 3];2015. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4343077&tool=pmcentrez&rendertype=abstract

40. Groza T, Köhler S, Moldenhauer D, Vasilevsky N, Baynam G, Zemojtel T, et al. The Human Phenotype Ontology: Semantic Unification of Common and Rare Disease. Am J Hum Genet [Internet]. 2015 Jun 24 [cited 2015 Jun 30];97(1):111–24. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26119816

41. Faculty of Veterinary Science U of S http://omia. angis. org. au. Online Mendelian Inheritance in Animals.

42. Washington N, Haendel M, Köhler S. How good is your phenotyping? Methods for quality assessment. In: Phenoday2014Bio-LarkOrg [Internet]. 2013. p. 1–4. Available from: http://phenoday2014.bio-lark.org/pdf/6.pdf

43. Girdea M, Dumitriu S, Fiume M, Bowdin S, Boycott KM, Chénier S, et al. PhenoTips: patient phenotyping software for clinical and research use. Hum Mutat [Internet]. 2013 Aug [cited 2015 Mar 30];34(8):1057–65. Available from:
http://www.ncbi.nlm.nih.gov/pubmed/23636887

44. Bostock M, Ogievetsky V, Heer J. D^3: Data-Driven Documents. IEEE Trans Vis Comput Graph [Internet]. 2011 Dec 1 [cited 2015 Feb 12];17(12):2301–9. Available from: http://ieeexplore.ieee.org/articleDetails.jsp?arn

umber=6064996

45. Robinson PN, Kohler S, Oellrich A, Genetics SM, Wang K, Mungall CJ, et al. Improved exome prioritization of disease genes through cross-species phenotype comparison. Genome Res [Internet]. 2014 Feb [cited 2014 Jul 12];24(2):340–8. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3912424&tool=pmcentrez&rendertype=abstract

46. Smedley D, Schubach M, Jacobsen JOB, Kohler S, Zemojtel T, Spielmann M, et al. A Whole-Genome Analysis Framework for Effective Identification of Pathogenic Regulatory Variants in Mendelian Disease. Am J Hum Genet [Internet]. 2016 Sep [cited 2016 Sep 22];99(3):595–606. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0002929716302786

47. Bello SM, Smith CL, Eppig JT. Allele, phenotype and disease data at Mouse Genome Informatics: improving access and analysis. Mamm Genome [Internet]. 2015 Aug [cited 2016 Aug 19];26(7–8):285–94. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26162703

48. Hayamizu TF, Baldock RA, Ringwald M. Mouse anatomy ontologies: enhancements and tools for exploring and integrating biomedical data. Mamm Genome [Internet]. 2015 Oct [cited 2016 Aug 19];26(9–10):422–30. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26208972

49. Köhler S, Doelken SC, Ruef BJ, Bauer S, Washington N, Westerfield M, et al. Construction and accessibility of a cross-species phenotype ontology along with gene annotations for biomedical research. F1000Research [Internet]. 2013 Jan [cited 2014 Aug 19];2:30. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3799545&tool=pmcentrez&rendertype=abstract

50. Hayamizu TF, de Coronado S, Fragoso G, Sioutos N, Kadin JA, Ringwald M. The mouse-human anatomy ontology mapping project. Database (Oxford) [Internet]. 2012 Jan [cited 2015 Apr 12];2012:bar066. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3308156&tool=pmcentrez&rendertype=abstract

51. Kalderimis A, Lyne R, Butano D, Contrino S, Lyne M, Heimbach J, et al. InterMine: extensive web services for modern biology. Nucleic Acids Res [Internet]. 2014 Jul [cited 2016 Oct 21];42(Web Server issue):W468-72. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24753429

52. Lyne R, Sullivan J, Butano D, Contrino S, Heimbach J, Hu F, et al. Cross-organism analysis using InterMine. Genesis [Internet]. 2015 Aug [cited 2016 Oct 21];53(8):547–60. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26097192

53. Elsik CG, Tayal A, Diesh CM, Unni DR, Emery ML, Nguyen HN, et al. Hymenoptera Genome Database: integrating genome annotations in HymenopteraMine. Nucleic Acids Res [Internet]. 2016 Jan 4 [cited 2016 Oct 21];44(D1):D793-800. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26578564

54. Livingston KM, Bada M, Baumgartner WA, Hunter LE, Galperin M, Rigden D, et al. KaBOB: ontology-based semantic integration of biomedical databases. BMC Bioinformatics [Internet]. 2015 Dec 23 [cited 2016 Oct 21];16(1):126. Available from: http://www.biomedcentral.com/1471-2105/16/126

55. Grubb SC, Bult CJ, Bogue MA. Mouse phenome database. Nucleic Acids Res [Internet]. 2014 Jan [cited 2015 May 3];42(Database issue):D825-34. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3965087&tool=pmcentrez&rendertype=abstract

56. Mulligan MK, Mozhui K, Prins P, Williams RW. GeneNetwork – A Toolbox for Systems Genetics. Syst Genet Methods Mol Biol. 2016;9.

57. Bastian F, Parmentier G, Roux J, Moretti S, Laudet V, Robinson-Rechavi M. Bgee: Integrating and Comparing Heterogeneous Transcriptome Data Among Species. In: Data Integration in the Life Sciences. 2008. p. 124–31.

58. McGary KL, Park TJ, Woods JO, Cha HJ, Wallingford JB, Marcotte EM. Systematic discovery of nonobvious human disease models through orthologous phenotypes. Proc Natl Acad Sci U S A [Internet]. 2010 Apr 6 [cited 2015 May 1];107(14):6544–9. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2851946&tool=pmcentrez&rendertype=abstract

59. Woods JO, Singh-Blom UM, Laurent JM, McGary KL, Marcotte EM. Prediction of gene-phenotype associations in humans, mice, and plants using phenologs. BMC Bioinformatics [Internet]. 2013 Jan [cited 2015 May

20];14:203. Available from: http://www.pubmedcentral.nih.gov/articlerende r.fcgi?artid=3704650&tool=pmcentrez&rendert ype=abstract

60. Laraway B. Comparative analysis of semantic similarity and gene orthology tools for identification of gene candidates for human diseases [Internet]. Oregon Health & Science University; 2015. Available from: http://digitalcommons.ohsu.edu/etd/3741

61. Brush M, Shefchek K, Haendel MA. SEPIO: A Semantic Model for the Integration and Analysis of Scientific Evidence. In: International Conference on Biomedical Ontology and BioCreative (ICBO BioCreative 2016) [Internet]. Corvallis, Oregon; 2016. Available from: http://icbo.cgrb.oregonstate.edu/

62. Rehm HL, Berg JS, Brooks LD, Bustamante CD, Evans JP, Landrum MJ, et al. ClinGen-- the Clinical Genome Resource. N Engl J Med [Internet]. 2015 Jun 4 [cited 2016 Mar 22];372(23):2235–42. Available from: http://www.pubmedcentral.nih.gov/articlerende r.fcgi?artid=4474187&tool=pmcentrez&rendert ype=abstract

63. Bandrowski A, Brinkman R, Brochhausen M, Brush MH, Bug B, Chibucos MC, et al. The Ontology for Biomedical Investigations. PLoS One [Internet]. 2016;11(4):e0154556. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27128319

64. Diehl AD, Meehan TF, Bradford YM, Brush MH, Dahdul WM, Dougall DS, et al. The Cell Ontology 2016: enhanced content, modularization, and ontology interoperability. J Biomed Semantics. 2016 Jul 4;7(1):44. Available from: https://www.ncbi.nlm.nih.gov/pubmed/2737765 2

65. Attrill H, Falls K, Goodman JL, Millburn GH, Antonazzo G, Rey AJ, et al. FlyBase: establishing a Gene Group resource for Drosophila melanogaster. Nucleic Acids Res. 2016 Jan 4;44(D1):D786-92. Available from: https://www.ncbi.nlm.nih.gov/pubmed/?term=P MID%3A26467478

66. Howe KL, Bolt BJ, Cain S, Chan J, Chen WJ, Davis P, et al. WormBase 2016: expanding to enable helminth genomic research. Nucleic Acids Res. 2016 Jan 4;44(D1):D774-80. Available from: https://www.ncbi.nlm.nih.gov/pubmed/2657857 2