

# The Disease Ontology: fostering interoperability between biological and clinical human disease-related data

Lynn M. Schriml<sup>1</sup> · Elvira Mitraka<sup>1</sup>

Received: 19 February 2015 / Accepted: 8 June 2015 / Published online: 21 June 2015  
© The Author(s) 2015. This article is published with open access at Springerlink.com

**Abstract** The Disease Ontology (DO) enables cross-domain data integration through a common standard of human disease terms and their etiological descriptions. Standardized disease descriptors that are integrated across mammalian genomic resources provide a human-readable, machine-interpretable, community-driven disease corpus that unifies the representation of human common and rare diseases. The DO is populated by consensus-driven disease data descriptors that incorporate disease terms utilized by genomic and genetic projects and resources engaged in studies to understand the genetics of human disease through the study of model organisms. The DO project serves multiple roles for the model organism community by providing: (1) a structured “backbone” of disease concepts represented among the model organism databases; (2) authoritative disease curation services to researchers and resource providers; and (3) development of subsets of the DO representative of human diseases annotated to animal models curated within the model organism databases.

## Introduction

The nomenclature of human disease has a long and complex history. Traditionally, diseases have been named for the discoverer of the disease (Robinow syndrome), where the disease was first discovered (Rocky Mountain spotted

fever), or the animal in which the disease was first identified (avian influenza). A disease may initially be called a syndrome to describe a collection of symptoms. As genes, genetic variants and the etiology of a disease are discerned; the disease name may or may not be changed to reflect the new level of knowledge. Consequently, there is a significant need for a standardized representation of human diseases to map disease concepts across resources, to connect related information, and to support the development of computational tools that will enable robust data analysis and integration. Utilizing a standardized ontological representation of human diseases creates a rigorous knowledge backbone for the annotation of biomedical data through defined concepts connected by specified relations (Hoehndorf et al. 2012; Musen et al. 2012). The Disease Ontology (DO) (Schriml et al. 2012; Kibbe et al. 2015) was developed to connect clinical vocabulary concepts through the creation of a single unifying representation of human disease. The DO simplifies and facilitates the exchange and comparison of disease-related information between biomedical and clinical resources. The scope of this review is to highlight the roles of the DO project contributing to human disease and mammalian genetic research, specifically describing the DO’s activities to support mouse model organism research at the mouse genome informatics (MGI) database. The DO project provides on-going expert curation of human diseases annotated within the model organism databases (MOD) (Kibbe et al. 2015) and has begun to create MOD-specific subsets of DO to facilitate the use of DO within the MODs. The DO team has created a DO\_MGI\_slim as a subset of the DO for human disease terms associated with mouse animal models within MGI. A slim of the DO is a smaller version of the DO’s Human-DO.obo file that contains a subset of terms from the entire DO file.

---

✉ Lynn M. Schriml  
lschriml@som.umaryland.edu

<sup>1</sup> Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, MD 21201, USA

## DO: a centralized disease term resource

The DO provides a homogeneous, centralized ontological representation of human disease that fosters precise identification and interpretation of disease data elements and incorporates the changing nomenclature of human disease terms over time along with community-specific naming conventions and synonyms.

The DO project was established in 2003 at Northwestern University to address the need for a purpose-built ontology that covers the full spectrum of disease concepts annotated within biomedical repositories within an ontological framework that is extensible to meet community needs. Disease concepts are defined for a variety of reporting and indexing purposes that are represented in numerous medical vocabularies. The DO was initially built from the union of disease terms extracted from the ICD mortality and morbidity classifications (Ayme et al. 2010), SNOMED-CT clinical and event healthcare reports (Donnelly 2006), and MeSH Medline index terms (Nelson et al. 2004; Kibbe et al. 2015). Expansion of disease sub-types has been driven by the disease data needs of the user community. The DO team works with several MODs to integrate their human disease terms into DO. The DO project additionally integrates human disease terms represented in biomedical resources including the Online Mendelian Inheritance in Man (OMIM) (Amberger et al. 2015), the experimental factor ontology (EFO) (Parkinson et al. 2011), the protein ontology (PRO) (Natale et al. 2014), and orphanet (Ayme et al. 2010) to provide authoritative and rigorously reviewed disease terms. DO's inclusion of unique identifiers from the medical vocabularies facilitates cross-mapping between resources and enables the DO team to quickly identify and avoid inclusion of overlapping (not novel) content.

## DO development and content

As a community-driven standards-based ontology project, the DO is focused on representing common and rare disease concepts captured across biomedical resources with the mission of providing a broadly useful disease interface between data resources through on-going support (term review and integration) of disease terminology needs. The DO includes 6582 terms (38 % defined, revision 2821) representing inherited, developmental, and acquired human diseases. The HumanDO.obo file additionally includes the set of 2373 obsoleted DO terms to enable ID tracking. As terms are added or reviewed within the DO, textual definitions are added. The DO team is actively working to define all terms within DO. We have increased the percent

of defined DO terms by 16 % since 2012. This includes the addition of 163 DO terms and an increase of defined DO terms by 6 % overall between October 2014 and May 2015. The DO team will continue to address the need to add definitions for our backlog of undefined terms.

## DO: representation of disease terms

DO disease descriptors represent individual disease terms and are provided in a computable format where each DO term is represented by unique uniform resource identifiers (URI's) (e.g., [http://www.purl.obolibrary.org/obo/DOID\\_4](http://www.purl.obolibrary.org/obo/DOID_4)). A DOID is a disease ontology ID. The DO's HumanDO.obo file can be downloaded from our Sourceforge repository at: <http://www.sourceforge.net/p/diseaseontology/code/HEAD/tree/trunk/HumanDO.obo>. The OWL version of the HumanDO file, produced by the OBO Foundry, can be found at: <http://www.purl.obolibrary.org/obo/doid.owl>. The DO project publishes our primary data files and software in public repositories to ensure that they will remain publicly available. While maintaining DO's Sourceforge archive (<http://www.sourceforge.net/p/diseaseontology>) for current DO users, the DO project has begun to transition our production activities to a GitHub repository (<http://www.github.com/obophenotype/human-disease-ontology>) in 2015. The DO project's ontology and database files are freely available for copying, redistribution, and adaption for any purpose under the Creative Commons Attribution 3.0 Unported License (<http://www.creativecommons.org/licenses/by/3.0/>).

The DO website (<http://www.disease-ontology.org>) contains a web-based application (Fig. 1) that enables full-text searching of DO disease terms, definitions, and related metadata. The DO includes a subset of terms utilized within MGI. The set of MGI human disease terms can be retrieved through the DO web interface by searching on the slim name DO\_MGI\_slim in the search box or as an Advanced Search by selecting subset. This query will retrieve the set of disease terms utilized for the annotation of human diseases for mouse animal models in MGI. Additionally, the set of DO diseases with an associated OMIM cross-reference can be retrieved by an Advanced Search of Xrefs and typing in OMIM to the query box.

The DO web browser's RESTful API service enables direct querying of a single DOID (e.g., sick sinus syndrome <http://www.disease-ontology.org/term/DOID:13884/>) and programmatic access of the DO database metadata (<http://www.disease-ontology.org/api/metadata/DOID:13884>) to produce:

```
{“xrefs”: [“ICD10CM:I49.5”, “MSH:D012804”, “NCI:C62244”, “OMIM:163800”, “OMIM:608567”, “OMIM:
```



The screenshot shows the Disease Ontology website interface. At the top, there is a search bar with the text 'DO\_MGI\_slim' and a 'Go »' button. To the right of the search bar is a link for 'Advanced Search »'. Below the search bar, there are two tabs: 'Welcome' and 'Sick Sinus Syndr'. The 'Sick Sinus Syndr' tab is active, displaying a 'Metadata' section. On the left side, there is a 'Navigation' panel with a tree view of disease categories. The 'Sick Sinus Syndr' term is selected in the tree. The 'Metadata' section contains the following information:

Metadata	
DOID	DOID:13884
Name	sick sinus syndrome
Synonyms	sinus node infection [EXACT]
Xrefs	ICD10CM:I49.5 <a href="#">MSH:DO12804</a> <a href="#">NCI:C62244</a> <a href="#">OMIM:163800</a> <a href="#">OMIM:608567</a> <a href="#">OMIM:614090</a> ORDO:166282 SNOMEDCT_2010_1_31:155373001 SNOMEDCT_2010_1_31:266307005 SNOMEDCT_2010_1_31:36083008 UMLS_CUI:C0037052
Subsets	DO_MGI_slim
Relationships	is_a <a href="#">sinoatrial node disease</a>

Below the 'Relationships' section, there is a link: 'Add an item to the term tracker'.

**Fig. 1** The disease ontology website. A query for all DO terms included in the DO\_MGI\_slim is displayed

614090”, “ORDO:166282”, “SNOMEDCT\_2010\_1\_31:155373001”, “SNOMEDCT\_2010\_1\_31:266307005”, “SNOMEDCT\_2010\_1\_31:36083008”, “UMLS\_CUI:C0037052”], “name”: “sick sinus syndrome”, “subsets”: [“DO\_MGI\_slim”], “synonyms”: [“sinus node infection EXACT “], “parents”: [“is\_a”, “sinoatrial node disease”, “DOID:0050824”]], “id”: “DOID:13884” }.

## DO: disease curation service and community collaborations

The DO project established a user support workflow in 2007 to invite DO users to submit DO term requests and updates. The DO team reviews, classifies, and integrates requested terms into DO. The DO team has established curation ‘best practices’ to field and respond to user requests, to coordinate disease representation within biomedical resources, and to provide subject matter and ontology expert curation

guidance to the user community. DO curator’s Style Guide is available on the DO website under the Resources tab. The DO Style Guide provides documentation on defining new terms, cleaning up term format, creating definitions, dbxrefs (IDs as references to external resources), and provenience of term definition sources as URLs.

The DO project has established a network of collaborative disease annotation efforts where the community members contribute their expertise and area-specific disease content to DO and the DO team provides expert disease classifications to the community members (Kibbe et al. 2015). These on-going efforts ensure the accuracy of DO annotations through on-going community review and feedback provided to DO. Additionally, the DO team receives novel disease terms, suggestions, and advices from subject matter experts in an on-going weekly, sometimes daily, process. For example, the neuroscience information framework (NIF) (Imam et al. 2012) and NeuroDevNet (Portales-Casamar et al. 2013) have been major

contributors to DO's representation of neurological diseases. The DO project collaborates with a number of NIH-funded genomic resources including: MGI, Protein Ontology (PRO), the Immune Epitope Database (IEDB), Reactome, WormBase (Harris et al. 2014), FlyBase (Dos Santos et al. 2015), and the NHLBI LINCS project (Vempati et al. 2014). For example, the Reactome project (Croft et al. 2014) has a total of 1226 pathways and reactions (out of 120,965 total human reactions and pathways) that are annotated with a DO disease term, and 2167 proteins and complexes (out of 40,349 human proteins and complexes) annotated with one.

### MGI: Disease Ontology collaboration

The DO project teamed up with MGI in the fall of 2014 to provide review and integration of MGI's genetic disease vocabulary terms (aligned to OMIM terms and IDs) into DO. The DO is being enriched by the novel content contributed by MGI representing genetic disease sub-types associated with new mouse models and recently reported human diseases. The MGI team generated and provided DO with a mapping file of 2606 disease terms with corresponding OMIM IDs. Manual review of each disease term and OMIM mapping identified a set of 1198 OMIM IDs that mapped to 260 existing DO terms and a set of 1408 OMIM IDs that had no corresponding DOIDs. These 1408 OMIM IDs were mapped to new DO terms. In total, 250 new DO terms were created, primarily representing syndromes and rare diseases. The DO term to OMIM ID mapping identified several one to many mappings between DOIDs and OMIM IDs. This work represents a substantial multi-month manual curation effort and highlights the importance of the manual reviews required to connect biological concepts. Automated mapping of disease terms between resources and vocabularies holds the promise of streamlining manual reviews (Rance et al. 2013). The DO project recognizes the potential of automated mapping of disease terms. However, the DO project has been cautious to integrate automated analysis into our workflow, as the precision (the proportion of true positives that are correctly identified as true positives) of these analyses is less than 100%. This manual review process included: the reassessment of the hierarchy of each term; the re-evaluation of its definition, if one existed; the definition of the term, if one didn't exist; addition of synonyms; and mapping of the term to other sources (ICD10 codes, Orphanet IDs) via cross-references. A small number of OMIM terms could not be included in DO as they represent a phenotype but not a disease such as the OMIM entries: "Plasmodium falciparum blood infection level" (OMIM: 248310); "Plasmodium falciparum quantitative trait locus 1"

(OMIM:611384), and "susceptibility to migraine, with or without aura" (OMIM:610208).

### DO: an ontology and resource disease hub

The DO has been widely adopted and utilized as a standard representation of human disease in biomedical ontologies including the human phenotype ontology (HPO; Köhler et al. 2014), the infectious disease ontology (IDO; <http://www.infectiousdiseaseontology.org>), the influenza ontology, the NIFSTD ontology, the cell ontology (CL), the experimental factor ontology (EFO), the protein ontology (PRO), and the cardiovascular disease ontology (<https://www.code.google.com/p/cvdo/>). The DO has become a highly utilized disease data resource for evaluating and connecting diverse sets of data. DO terms are being utilized for identifying representative phenotype sets (Köhler et al. 2014; Schofield et al. 2010; Monarch Initiative <http://www.monarchinitiative.org> and HPO Phenomizer <http://www.compbio.charite.de/phenomizer/>) and functionally similar genes (Fang and Gough 2013; Singleton et al. 2014). The DO is also being used for analyzing ontologies and protein domain annotations (dcGOR (Fang 2014)), human gene and genome annotations (Peng et al. 2013; Osborne et al. 2009), pathways (Croft et al. 2014; Hoehndorf et al. 2012), cancer variants (Wu et al. 2014; Milacic et al. 2012), and immune epitopes (Vita et al. 2014).

The extensive use of DO in data analysis workflows and the annotation of DO terms across a wide body of genomic resources enables users to readily identify data of interest within or between these resources and to identify novel connections. For example, the Jackson Laboratory for Genomic Medicine is utilizing DO within their Clinical Genome Informatics System workflow to collect clinical disease data and to translate pathology reports to Next Generation Sequencing Clinical Reports (Ananda et al. 2015). Likewise, the BioXpress, a gene expression and cancer association database, is utilizing DO to map cancer terms across cancer resources to facilitate uniform pan-cancer analysis (Wan et al. 2015).

### DO subsets: resource-specific disease representations (DO slims)

By definition biomedical ontologies capture the breadth of domain specific knowledge. In order to take further advantage of the detailed ontological information found in the leaf ontology terms subsets of biomedical ontologies, called slims, are being created to provide customized high-level views (Davis et al. 2010; Geifman et al. 2010). For example, subsets of the gene ontology (GO) (GO



Consortium 2013) have been created (<http://www.geneontology.org/page/go-slim-and-subset-guide>) to represent model organism or resource-specific subsets of GO. The DO project initiated the development of DO slims in 2014 to provide resource- (e.g., MGI) or disease-specific (e.g., cancer) subsets of DO. DO slim files can be extracted from the DO HumanDO.obo file and are also available as from DO's GitHub repository (<https://www.github.com/obophenotype/human-disease-ontology>).

The DO\_MGI\_slim represents a focused view of the subset of genetic diseases in DO that are represented in MGI and enables researchers to identify and explore mouse model organism associated diseases. The DO\_MGI\_slim subset currently contains 354 terms (DO revision 2821). Additional terms are being reviewed for inclusion in DO from a set of 510 MGI human disease terms. DO slim files are available from the DO SVN repository.

The DO\_cancer\_slim provides a broad view of the set of DO cancer terms represented within the catalog of somatic mutations in cancer (COSMIC), the international cancer genome consortium (ICGC), the cancer genome atlas (TCGA), and the integrative oncogenomics (IntOGen) database to support the coordinated effort of pan-genome analysis of cancer genetic variants (Wu et al. 2015). The DO\_cancer\_slim represents the classification of a subset of 393 cancer terms into a cohesive set of 192 DO terms. This collaborative classification of cancer terms resulted in 43 new cancer terms being added to DO and the addition of definitions and references for 63 DO parent terms and 187 DO child node (leaf) terms. The DO\_cancer\_slim terms were further categorized by the type of cancer they represent into a selection of 64 upper-level DO terms (TopNodes\_DOcancerslim). This project has produced a unifying ontology of cancer terms that could be used to map data across various resources, representing a multitude of cancer types, and be utilized for pan-cancer analysis.

## Conclusions

The DO project is highly committed in our efforts to support genetic and genomic research as disease data and curation resource. As a centralized disease resource DO provides stable, long-term support of disease terms to be used by the MOD research community. The DO team provides reliable and responsive curation services to researchers and data providers. Requests for DO curation may be submitted through the DO website (<http://www.disease-ontology.org>), DO's SVN term tracker (<http://www.sourceforge.net/p/diseaseontology/feature-requests/>), DO's Contact Us web form, or by emailing the DO team members (email: DO PI - Lynn Schriml: [lschriml@som.umaryland.edu](mailto:lschriml@som.umaryland.edu)).

The DO project will be creating additional DO slims for each of the MODs (e.g., Reactome, FlyBase, WormBase,

RGD, SGD, and ZFIN) in the near future. These subsets of the DO will enable DO users to review the representation and classification of associated diseases, to compare the diseases represented between MODs and to compare the different animal models associated with a particular disease or types of diseases across species.

**Acknowledgments** This work is partially funded under NIH/NIGMS R01 GM089820 by the University of Maryland, Institute for Genome Sciences. The DO team is deeply indebted and acknowledges the long-term support of our collaborators and contributors for their tireless efforts to improve this project.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Amberger J, Bocchini C, Schiettecatte F, Scott AF, Hamosh A (2015) OMIM.org: online Mendelian inheritance in man (OMIM<sup>®</sup>), an online catalog of human genes and genetic disorders. *Nucl Acids Res* 43:D789–798
- Ananda G, Mockus S, Lundquist M, Spotlow V, Simons A, Mitchell T, Stafford G, Philip V, Stearns T, Srivastava A, Barter M, Rowe L, Malcolm J, Bult C, Karuturi RK, Rasmussen K, Hinerfeld D (2015) Development and validation of the JAX Cancer Treatment Profile<sup>TM</sup> for detection of clinically actionable mutations in solid tumors. *Exp Mol Pathol* 98:106–112
- Ayme S, Rath A, Bellet B (2010) WHO international classification of diseases (ICD) revision process: incorporating rare diseases into the classification scheme: state of art. *Orphanet J Rare Dis* 5:P1
- Croft D, Mundo AF, Haw R, Milacic M, Weiser J, Wu G, Caudy M, Garapati P, Gillespie M, Kamdar MR, Jassal B, Jupe S, Matthews L, May B, Palatnik S, Rothfels K, Shamovsky V, Song H, Williams M, Birney E, Hermjakob H, Stein L, D'Eustachio P (2014) The reactome pathway knowledgebase. *Nucl Acids Res* 42:D472–D477
- Davis MJ, Sehgal MS, Ragan MA (2010) Automatic, context-specific generation of gene ontology slims. *BMC Bioinform* 11:498
- Donnelly K (2006) SNOMED-CT: the advanced terminology and coding system for eHealth. *Stud Health Technol Inform* 121:279–290
- Fang H (2014) dcGOR: an R package for analysing ontologies and protein domain annotations. *PLoS Comput Biol* 10:e1003929
- Fang H, Gough J (2013) DcGO: database of domain-centric ontologies on functions, phenotypes, diseases and more. *Nucl Acids Res* 41:D536–D544
- FlyBase Consortium, Dos Santos G, Schroeder AJ, Goodman JL, Strelets VB, Crosby MA, Thurmond J, Emmert DB, Gelbart WM (2015) FlyBase: introduction of the *Drosophila melanogaster* Release 6 reference genome assembly and large-scale migration of genome annotations. *Nucl Acids Res* 43:D690–D697
- Geifman N, Monsonego A, Rubin E (2010) The neural/immune gene ontology: clipping the gene ontology for neurological and immunological systems. *BMC Bioinform* 11:458
- Gene Ontology Consortium, Blake JA, Dolan M, Drabkin H, Hill DP, Li N, Sitnikov D, Bridges S, Burgess S, Buza T, McCarthy F, Peddinti D, Pillai L, Carbon S, Dietze H, Ireland A, Lewis SE, Mungall CJ, Gaudet P, Chrisholm RL, Fey P, Kibbe WA, Basu S, Siegle DA, McIntosh BK, Renfro DP, Zweifel AE, Hu JC,

- Brown NH, Tweedie S, Alam-Faruque Y, Apweiler R, Auchinchloss A, Axelsen K, Bely B, Blatter M, Bonilla C, Bouguerlet L, Boutet E, Breuza L, Bridge A, Chan WM, Chavali G, Coudert E, Dimmer E, Estreicher A, Famiglietti L, Feuerhann M, Gos A, Gruaz-Gumowski N, Hieta R, Hinz C, Hulo C, Huntley R, James J, Jungo F, Keller G, Laiho K, Legge D, Lemercier P, Lieberherr D, Magrane M, Martin MJ, Masson P, Mutowo-Muellenet P, O'Donovan C, Pedruzzi I, Pichler K, Poggioni D, Porras Millán P, Poux S, Rivoire C, Roehert B, Sawford T, Schneider M, Stutz A, Sundaram S, Tognolli M, Xenarios I, Foulgar R, Lomax J, Roncaglia P, Khodiyar VK, Lovering RC, Talmud PJ, Chibucos M, Giglio MG, Chang H, Hunter S, McAnulla C, Mitchell A, Sangrador A, Stephan R, Harris MA, Oliver SG, Rutherford K, Wood V, Bahler J, Lock A, Kersey PJ, McDowall DM, Staines DM, Dwinell M, Shimoyama M, Laulederkind S, Hayman T, Wang S, Petri V, Lowry T, D'Eustachio P, Matthews L, Balakrishnan R, Binkley G, Cherry JM, Costanzo MC, Dwight SS, Engel SR, Fisk DG, Hitz BC, Hong EL, Karra K, Miyasato SR, Nash RS, Park J, Skrzypek MS, Weng S, Wong ED, Berardini TZ, Huala E, Mi H, Thomas PD, Chan J, Kishore R, Sternberg P, Van Auken K, Howe D, Westerfield M (2013) Gene Ontology annotations and resources. *eic Acids Res* 41:D530–D535
- Harris TW, Baran J, Bieri T, Cabunoc A, Chan J, Chen WJ, Davis P, Done J, Grove C, Howe K, Kishore R, Lee R, Li Y, Muller HM, Nakamura C, Ozersky P, Paulini M, Raciti D, Schindelman G, Tuli MA, Van Auken K, Wang D, Wang X, Williams G, Wong JD, Yook K, Schedl T, Hodgkin J, Berriman M, Kersey P, Spieth J, Stein L, Sternberg PW (2014) WormBase 2014: new views of curated biology. *Nucl Acids Res* 42:D789–D793
- Hoehndorf R, Dumontier M, Gkoutos GV (2012) Identifying aberrant pathways through integrated analysis of knowledge in pharmacogenomics. *Bioinformatics* 28:2169–2175
- Imam FT, Larson SD, Bandrowski A, Grethe JS, Gupta A, Martone ME (2012) Development and use of ontologies inside the neuroscience information framework: a practical approach. *Front Genet* 3:111
- Kibbe WA, Arze C, Felix V, Mitraka E, Bolton E, Fu G, Mungall CJ, Binder JX, Malone J, Vasant D, Parkinson H, Schriml LM (2015) Disease ontology 2015 update: an expanded and updated database of human diseases for linking biomedical knowledge through disease data. *Nucl Acids Res* 43:D1071–D1078
- Köhler S, Doelken SC, Mungall CJ, Bauer S, Firth HV, Bailleul-Forestier I, Black GC, Brown DL, Brudno M, Campbell J, FitzPatrick DR, Eppig JT, Jackson AP, Freson K, Girdea M, Helbig I, Hurst JA, Jähn J, Jackson LG, Kelly AM, Ledbetter DH, Mansour S, Martin CL, Moss C, Mumford A, Ouwehand WH, Park SM, Riggs ER, Scott RH, Sisodiya S, Van Vooren S, Wapner RJ, Wilkie AO, Wright CF, Vulto-van Silfhout AT, de Leeuw N, de Vries BB, Washington NL, Smith CL, Westerfield M, Schofield P, Ruef BJ, Gkoutos GV, Haendel M, Smedley D, Lewis SE, Robinson PN (2014) The human phenotype ontology project: linking molecular biology and disease through phenotype data. *Nucl Acids Res* 4:D966–D974
- Milacic M, Haw R, Rothfels K, Wu G, Croft D, Hermjakob H, D'Eustachio P, Stein L (2012) Annotating cancer variants and anti-cancer therapeutics in reactome. *Cancers* 4:1180–1211
- Natale DA, Arighi CN, Blake JA, Bult CJ, Christie KR, Cowart J, D'Eustachio P, Diehl AD, Drabkin HJ, Helfer O, Huang H, Masci AM, Ren J, Roberts NV, Ross K, Ruttenberg A, Shamovsky V, Smith B, Yerramalla MS, Zhang J, AlJanahi A, Çelen I, Gan C, Lv M, Schuster-Lezell E, Wu CH (2014) Protein ontology: a controlled structured network of protein entities. *Nucl Acids Res* 42:D415–D421
- NCBO team, Musen MA, Noy NF, Shah NH, Whetzel PL, Chute CG, Story MA, Smith B (2012) The national center for biomedical ontology. *J Am Med Inform Assoc* 19:190–195
- Nelson SJ, Schopen M, Savage AG, Schulman J, Arluk N (2004) The MeSH translation maintenance system: structure, interface design, and implementation. In: Fieschi M, et al., (eds) Proceedings of the 11th world congress on medical informatics. IOS Press, Amsterdam, San Francisco, pp 67–69
- Osborne JD, Flatow J, Holko M, Lin SM, Kibbe WA, Zhu LJ, Danila MI, Feng G, Chisholm RL (2009) Annotating the human genome with disease ontology. *BMC Genomics* 10(Suppl 1):S6
- Parkinson H, Sarkans U, Kolesnikov N, Abeygunawardena N, Burdett T, Dylag M, Emam I, Farne A, Hastings E, Holloway E, Kurbatova N, Lukk M, Malone J, Mani R, Pilicheva E, Rustici G, Sharma A, Williams E, Adamusiak T, Brandizi M, Sklyar N, Brazma A (2011) Arrayexpress update—an archive of microarray and high-throughput sequencing-based functional genomics experiments. *Nucl Acids Res* 39:D1002–D1004
- Peng K, Xu W, Zheng J, Huang K, Wang H, Tong J, Lin Z, Liu J, Cheng W, Fu D, Du P, Kibbe WA, Lin SM, Xia T (2013) The disease and gene annotations (DGA): an annotation resource for human disease. *Nucl Acids Res* 41:D553–D560
- Portales-Casamar E, Ch'ng C, Lui F, St-Georges N, Zoubarev A, Lai AY, Lee M, Kwok C, Kwok W, Tseng L, Pavlidis P (2013) Neurocarta: aggregating and sharing disease-gene relations for the neurosciences. *BMC Genomics* 14:129
- Rance B, Snyder M, Lewis J, Bodenreider O (2013) Leveraging terminological resources for mapping between rare disease information sources. *Stud Health Technol Inform* 192:529–533
- Schofield PN, Gkoutos GV, Gruenberger M, Sundberg JP, Hancock JM (2010) Phenotype ontologies for mouse and man: bridging the semantic gap. *Dis Model Mech* 3:281–289
- Schriml LM, Arze C, Nadendla S, Chang YW, Mazaitis M, Felix V, Feng G, Kibbe WA (2012) Disease ontology: a backbone for disease semantic integration. *Nucl Acids Res* 40:D940–D946
- Singleton MV, Guthery SL, Voelkerding KV, Chen K, Kennedy B, Margraf RL, Durtschi J, Eilbeck K, Reese MG, Jorde LB, Huff CD, Yandell M (2014) Phevor combines multiple biomedical ontologies for accurate identification of disease-causing alleles in single individuals and small nuclear families. *Am J Hum Genet* 94:599–610
- Vempati UD, Chung C, Mader C, Koleti A, Datar N, Vidović D, Wrobel D, Erickson S, Muhlich JL, Berriz G, Benes CH, Subramanian A, Pillai A, Shamu CE, Schürer SC (2014) Metadata standard and data exchange specifications to describe, model, and integrate complex and diverse high-throughput screening data from the library of integrated network-based cellular signatures (LINCS). *J Biomol Screen* 19:803–816
- Vita R, Overton JA, Greenbaum JA, Ponomarenko J, Clark JD, Cantrell JR, Wheeler DK, Gabbard JL, Hix D, Sette A, Peters B (2014) The immune epitope database (IEDB) 3.0. *Nucl Acids Res* 43:D405–D412
- Wan Q, Dingerdissen H, Fan Y, Gulzar N, Pan Y, Wu TJ, Yan C, Zhang H, Mazumder R (2015) BioXpress: an integrated RNA-seq-derived gene expression database for pan-cancer analysis. *Database (Oxf)* 2015:bav019
- Wu TJ, Shamsaddini A, Pan Y, Smith K, Crichton DJ, Simonyan V, Mazumder R (2014) A framework for organizing cancer-related variations from existing databases, publications and NGS data using a High-performance Integrated Virtual Environment (HIVE). *Database (Oxf)* 2014:bau022
- Wu TJ, Schriml LM, Chen QR, Colbert M, Crichton DJ, Finney R, Hu Y, Kibbe WA, Kincaid H, Meerzaman D, Mitraka E, Pan Y, Smith KM, Srivastava S, Ward S, Yan C, Mazumder R (2015) Generating a focused view of disease ontology cancer terms for pan-cancer data integration and analysis. *Database (Oxf)* 2015:32