INFORMATICS

Human Mutation

OFFICIAL JOURNAL
HGVS
HUMAN GENOME VARIATION SOCIETY
www.hgvs.org

Ontological Phenotype Standards for Neurogenetics

Sebastian Köhler,[1,2] Sandra C. Doelken,[1] Ana Rath,[3] Ségolène Aymé,[3] and Peter N. Robinson[1,2,4]*

[1] *Institute for Medical and Human Genetics, Charité-Universitätsmedizin Berlin, Berlin, Germany;* [2] *Berlin Center for Regenerative Therapies, Charité-Universitätsmedizin Berlin, Berlin, Germany;* [3] *INSERM, US14, Paris, France;* [4] *Max Planck Institute for Molecular Genetics, Berlin, Germany*

For the Databases in Neurogenetics Special Issue

Received 16 January 2012; accepted revised manuscript 13 April 2012.
Published online 9 May 2012 in Wiley Online Library (www.wiley.com/humanmutation).DOI: 10.1002/humu.22112

ABSTRACT: Neurological disorders comprise one of the largest groups of human diseases. Due to the myriad symptoms and the extreme degree of clinical variability characteristic of many neurological diseases, the differential diagnosis process is extremely challenging. Even though most neurogenetic diseases are individually rare, collectively, the subgroup of neurogenetic disorders is large, comprising more than 2,400 different disorders. Recently, increasing efforts have been undertaken to unravel the molecular basis of neurogenetic diseases and to correlate pathogenetic mechanisms with clinical signs and symptoms. In order to enable computer-based analyses, the systematic representation of the neurological phenotype is of major importance. We demonstrate how the Human Phenotype Ontology (HPO) can be incorporated into these efforts by providing a systematic semantic representation of phenotypic abnormalities encountered in human genetic diseases. The combination of the HPO together with the Orphanet disease classification represents a promising resource for automated disease classification, performing computational clustering and analysis of the neurogenetic phenome. Furthermore, standardized representations of neurologic phenotypic abnormalities employing the HPO link neurological phenotypic abnormalities to anatomical and functional entities represented in other biomedical ontologies through the semantic references provided by the HPO.

Hum Mutat 33:1333–1339, 2012.

KEY WORDS: phenotype; ontology; neurogenetics; data integration; orphan diseases

Representation of Phenotypic Data in Clinical Neurology

The differential diagnostic process for neurological disorders is one of the most challenging in medical practice, which is especially true for rare neurological diseases, defined as those that affect less than 1 in 2,000 persons. Unlike many other clinical disciplines, the day-to-day practice of neurology involves a pantheon of symptoms and long lists of differential diagnoses; the challenge is often to identify rare neurological diseases, especially those that require an individualized approach to clinical management [Chinnery, 2010]. Many rare neurological diseases are genetic in origin, and are associated with one or more functional and structural abnormalities of the nervous system. They originate from a wide spectrum of molecular and cellular defects and include skeletal muscle channelopathies, neuropathies, epilepsies, neurovascular disorders, neurocutaneous syndromes, mitochondrial diseases, nondegenerative movement disorders, inherited ataxias, neurodegenerative disorders, dementias, myopathies, and muscular atrophies. Currently, the classification of rare neurological diseases in the Orphanet database lists more than 2,400 different diseases. The differential diagnosis can be a daunting challenge for physicians, and affected patients may wait years before receiving a correct and precise diagnosis. For many patients, the underlying cause of the clinical manifestations may remain unknown.

Several major international initiatives in neuroscience have emerged with the goal of developing computational models of the brain including the Neuroscience Information Framework and the Human Brain Project [Gardner et al., 2008]. Ontologies are playing a major role in these efforts to ensure interoperability across neuroscience databases, to integrate neuroscience information from disparate sources to improve our understanding of the brain, and to enable computer reasoning. In computer science, the word "ontology" is used to describe a structured, automated representation of the knowledge within a certain domain in fields such as science, government, industry, and healthcare. Ontology provides a classification of the entities within a domain, whereby each entity is said to make up a term of the ontology. Furthermore, ontology must specify the semantic relationships between the entities, whereby any term may have multiple parents reflecting multiple semantic relationships. Thus, ontology can be used to define a standard, controlled vocabulary for a scientific field [Robinson and Bauer, 2011].

A number of different ontologies are relevant for the neurosciences and clinical neurology. The Foundational Model of Anatomy (FMA) [Rosse and Mejino, 2003] provides a unifying framework for human anatomy that describes the various entities that make up the body together with the relations between them. The FMA is designated as foundational because many of its classes generalize to vertebrates other than humans. The FMA arranges its classes in an inheritance hierarchy or taxonomy in a strictly structural context. This means that the definitions of the FMA are to be understood within the context of the ontology structure. For instance, the heart is defined as an "Organ with cavitated organ parts, which is continuous with the systemic and pulmonary arterial and venous trees." The function of the heart to pump blood plays no role in the definition. Similarly, the definition for brain "Segment of neuraxis that has as its parts gray matter and white matter that surround the cerebral ventricular system" does not include

*Correspondence to: Peter N. Robinson, Charité Universitätsmedizin Berlin, Institute for Medical Genetics and Human Genetics, Augustenburger Platz 1, Berlin 13353, Germany. E-mail: peter.robinson@charite.de

Contract grant sponsor: Berlin-Brandenburg Center for Regenerative Therapies (BCRT, Bundesministerium für Bildung und Forschung; 0313911); Deutsche Forschungsgemeinschaft (DFG RO 2005/4-1).

Table 1. HPO Phenotype Classes

"Class" of phenotype	HPO examples
Morphological abnormality	Broad nose (HP:0000445), Arachnodactyly (HP:0001166)
Abnormal process (organ)	Epistaxis (HP:0000421), Ileus (HP:0002595)
Abnormal process (cellular)	Abnormality of amino acid metabolism (HP:0004337), Abnormality of Krebs cycle metabolism (HP:0000816)
Abnormal laboratory finding	Hyperlipidemia (HP:0003077), Glycosuria (HP:0003076)
Electrophysiological abnormality	Decreased nerve conduction velocity (HP:0000762), Hypsarrhythmia (HP:0002521)
Abnormality by medical imaging	Butterfly vertebrae (HP:0003316), Choroid plexus cyst (HP:0002190)
Behavioral abnormality	Head nodding (HP:0001361), Self-mutilation (HP:0000742)

Different types of phenotypic abnormalities covered by the HPO.

mention of the control and coordination of mental and physical actions [Cook et al., 2004; Smith and Rosse, 2004]. The FMA incorporates numerous neuroanatomical concepts that provide the foundation to which other information may be linked [Martin et al., 2001, 2003].

Various other ontologies, some of which extend the FMA's representation of structural neuroanatomy, cover several neuroscience applications, including functional neuroimaging [Turner et al., 2010], functional neuroanatomy [Talos et al., 2008], neural connectivity in health and disease states [Rubin et al., 2009], subcellular neuroanatomy [Larson et al., 2007], electroencephalographic and other brain electromagnetic data [Dou et al., 2007], and many others. The Neuroscience Information Framework (NIF) intends to provide a consistent and flexible terminology for describing and searching neuroscience concepts and resources. The NIF has developed the NIF standardized ontology, a comprehensive collection of Web Ontology Language modules covering distinct domains of biomedical reality [Bug et al., 2008]. Additionally, the International Neuroinformatics Coordinating Facility (INCF) has been founded to coordinate and foster international activities in neuroinformatics. One of the goals of the INCF is to ensure interoperability of the various nomenclatures and ontologies in neuroinformatics [Bjaalie and Grillner, 2007].

None of the above ontologies were specifically designed to represent the phenotypic abnormalities seen in neurogenetic disorders. Late onset, variable expressivity, nonpenetrance, and the need for a long-term clinical observation including brain imaging, biopsies, and laboratory tests in order to reach the correct diagnosis are both time consuming and strenuous for patients and physicians alike. In clinical practice, the role of molecular genetic analysis using traditional methods such as Sanger sequencing has been limited because of genetic heterogeneity and clinical variability. In the future era of new genomic screening techniques including whole-exome and whole-genome sequencing, the challenge will not be to find the correct molecular genetic diagnosis, but rather the correct interpretation of the findings [Robinson et al., 2011]. Neurogenetic diseases have been shown to present with remarkable genotype–phenotype relationships. Better understanding of gene function and individual genotype–phenotype correlations provide novel insights into the molecular pathogenesis of these disorders, thereby opening up new approaches toward therapy and prevention in a personalized medicine framework for neurogenetic diseases.

Accurate standardized descriptions of phenotypic features, clinical course, laboratory findings as well as molecular genetic findings are needed in order to enable the community to harness the already existing data as well as information yet to come. Up to date, accessing existing knowledge has been hampered by the shear amount of available, unstructured data. Universal standardized vocabularies for describing phenotypes, clinical course, and laboratory findings are needed. Ontologies such as the Human Phenotype Ontology (HPO) developed by our group [Robinson et al., 2008; Robinson and Mundlos, 2010] can provide a basis for making the best use of available resources and information to help with patient care, diagnosis, and treatment. Furthermore, basic research can benefit from standardized knowledge representation, especially in discovering and understanding gene-to-phenotype relationships [Webb et al., 2011; Köhler et al., 2008].

In this article, we explain how the HPO has been used to annotate neurogenetic diseases and demonstrate how the annotations can be used together with the HPO and the rare neurological disease classification provided by Orphanet [Aymé, 2003; Aymé and Schmidtke, 2007] to perform computational clustering and analysis. Both terminologies can be used as a basis for describing disease entities and phenotypic abnormalities in databases and patient registries for neurogenetic diseases. Work is undergoing to connect the HPO to other ontologies of interest such as the FMA, which will allow computational bridges between disparate datasets in medically relevant neurological research.

The Human Phenotype Ontology

In the field of human genetics, the most important sources of information about hereditary diseases are the Online Mendelian Inheritance in Man (OMIM) [Amberger et al., 2009] database and Orphanet [Aymé, 2003; Aymé and Schmidtke, 2007]. Each of these knowledge bases, as well as the commercial diagnosis support programs of the London Dysmorphology Database [Fryns and de Ravel, 2002], and the Pictures of Standard Syndromes and Undiagnosed Malformations database (http://www.possum.net.au) have used their own vocabularies to describe phenotypic abnormalities. The multiplicity of mutually incompatible vocabularies in databases and publications on human genetics has hindered integrative research in clinical genetics [Lindblom and Robinson, 2011].

The HPO is being developed with the goal of providing a standard for describing the manifestations of human disease that can be used in databases, registries, publications, as well as for sophisticated bioinformatics analysis [Köhler et al., 2009]. The word phenotype is used with many different meanings. In biology, the most widely accepted definition of phenotype is "the observable traits of an organism," or perhaps the collection of observable traits of an organism, which result from the interaction of the genetic constitution of the organism and the environment [Mahner and Kary, 1997]. In medical contexts, the "phenotype" more often refers to some deviation from "normal," and this is the definition taken by the HPO. Many different types of phenotypic abnormality are represented in the HPO, including not only morphological signs, but also cellular, physiological, behavioral abnormalities (see Table 1). In general, HPO terms describe the abnormality and not the diagnostic procedure used to identify it. Thus, the term *Agenesis of corpus callosum* (HP:0001274) does not state whether the abnormality was found by computer tomography or magnetic resonance imaging. On the other

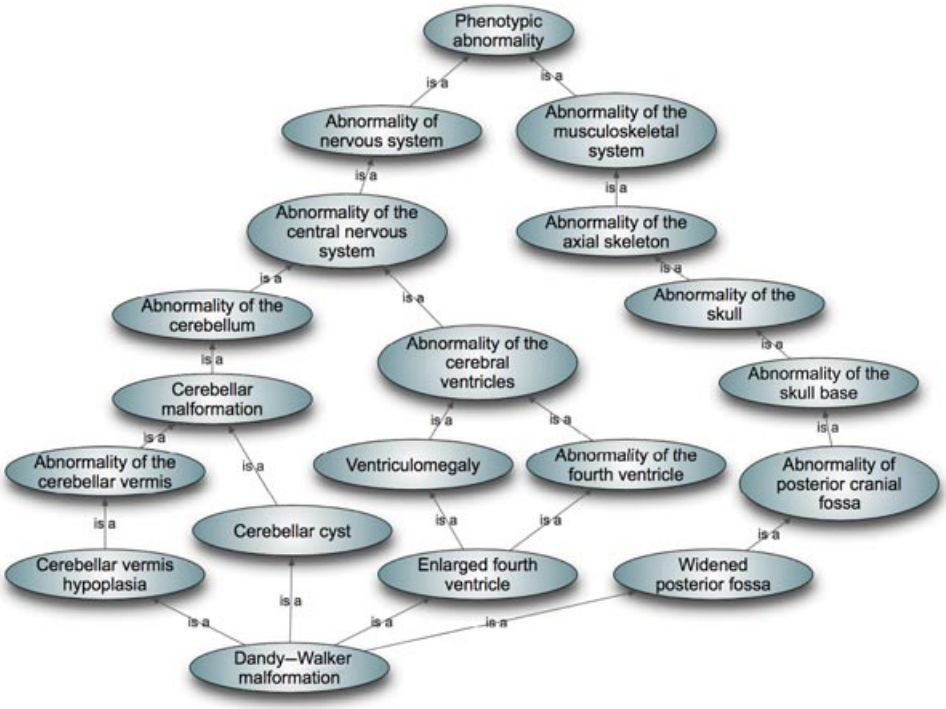

Figure 1. An excerpt of the HPO showing the relations between the terms. Except for the root term, which has no parent, each term in the HPO has one or more parent terms, to which it is related by a subclass relation ("is_a"). For instance, Dandy–Walker malformation is defined as a congenital cystic malformation of the cerebellum characterized by incomplete formation of the cerebellar vermis, dilation of the fourth ventricle, and enlargement of the posterior fossa. Therefore, Dandy–Walker malformation is a subclass of cerebellar vermis hypoplasia, cerebellar cyst, enlarged fourth ventricle, and widened posterior fossa in the sense that any patient with Dandy–Walker malformation can also be said to have these four abnormalities.

hand, there are abnormalities that are only observable with one kind of diagnostic modality (e.g., EEG with hyperventilation-induced focal epileptiform discharges, HP:0011183), and, in these cases, the diagnostic modality is generally included in the name of the term.

The HPO is available under an open-source license and it currently contains over 10,000 terms, each describing a phenotypic abnormality. The terms are related to one another by subclass ("is_a") relations, such that the ontology can be represented as a so-called directed acyclic graph in which any term except the root can have one or more parent terms and cycles are forbidden (Fig. 1).

The HPO terms themselves do not describe any specific disease. To describe a disease, annotations to HPO terms can be used. For instance, to assert that patients with sialidosis type I have cherry red spot of the macula, we annotate the disease sialidosis type I with the corresponding HPO term, that is, we associate these two entities with each other in the annotation file (see Table 2). Because such an annotation is not a simple binary relation, HPO annotations also include a number of metadata items that allow further specifications to be added. For example, the evidence code of an annotation indicates how the annotation to a particular term is supported. When the HPO was initially constructed and until 2008, most annotations were extracted by parsing the Clinical Synopsis sections of OMIM [Amberger et al., 2009]. These annotations are assigned the evidence code "IEA" (inferred from electronic annotation). Since 2008, many records have been revised and extended by expert biocuration. These annotations are given the code "PCS" (published clinical study), and the source of the study is indicated (usually, this is a PubMed ID). The evidence code "ICE" (individual clinical experience) can be used for annotations based on individual clinical experience. This may be appropriate for disorders with a limited amount of published data for annotations by an experienced clinician. This must be accompanied by an entry in the DB:Reference field denoting the individual or center performing the annotation together with an identifier. Additionally, the evidence code "ITM" (inferred by text mining) can be used to mark annotations retrieved by new text-mining efforts.

Another important characteristic of an annotation is the frequency with which individuals with a given disease have a certain phenotypic feature. For instance, according to a study by Caciotti et al. (2009), 9 of 43 persons with sialidosis type II have cherry red spot of the macula. This can be important for differential diagnostic purposes. For instance, if 100% of patients with some other disease have cherry red spot of the macula, then all else being equal, a person with cherry red spot of the macula would be more likely to have that disease than sialidosis type II. In many cases, exact numerical information on a frequency is not available, and more or less vague terms such as "occasional" are used in medical textbooks and articles. Descriptions such as these have been used in extremely variable ways in the medical literature. The HPO has defined a set of eight such categories to describe the frequency of features (Table 3).

Each annotation indicates the source of the annotation (e.g., the name of the biocurator). Recently, Orphanet has made a large number of annotations for rare diseases available. These data are particularly valuable because most of the phenotypic features have been assigned to one of three frequency categories by expert annotation. The HPO and the Orphanet teams have developed a mapping from the Orphanet thesaurus of Clinical Signs to corresponding HPO terms and used this to explore the annotations in the setting of neurogenetic disease in this work. These data have been used to

Table 2. Annotation Example

HPO term name	HPO term ID	DB:Reference	Evidence code	Frequency	Assigned by	Date
Intellectual disability	HP:0001249	OMIM:256550	IEA	NA	HPO	Feb. 17, 2009
		PMID:19568825	PCS	19/44	Sdoelken	June 21, 2011
Dysostosis multiplex	HP:0000943	OMIM:256550	IEA	NA	HPO	Feb. 17, 2009
		PMID:19568825	PCS	21/40	Sdoelken	June 21, 2011
Inguinal hernia	HP:0000023	OMIM:256550	IEA	NA	HPO	Feb. 17, 2009
		PMID:19568825	PCS	4/42	Sdoelken	June 21, 2011
Coarse facial features	HP:0000280	OMIM:256550	IEA	NA	HPO	Feb. 17, 2009
		ORPHANET:812	TAS	Hallmark	Orphanet	June 17, 2011
		PMID:19568825	PCS	26/41	Sdoelken	June 21, 2011
Sensorineural hearing impairment	HP:0000407	OMIM:256550	IEA	NA	HPO	Feb. 17, 2009
		ORPHANET:812	TAS	Hallmark	Orphanet	June 17, 2011
		PMID:19568825	PCS	10/17	Sdoelken	June 21, 2011
Hydrops fetalis	HP:0001789	OMIM:256550	IEA	NA	HPO	Feb. 17, 2009
		ORPHANET:87876	TAS	Hallmark	Orphanet	June 17, 2011
		PMID:19568825	PCS	22/27	Sdoelken	June 21, 2011
Cherry red spot of the macula	HP:0010729	OMIM:256550	IEA	NA	Skoehler	Feb. 17, 2009
		PMID:19568825 [Caciotti et al., 2009]	PCS	9/43	Sdoelken	June 21, 2011
Inheritance: autosomal recessive (HP:0000007)						

Excerpt from the annotation file for sialidosis [MIM: 256550]. Note that for every given phenotypic feature of the disease apart from the HPO terms and HPO term IDs, the reference, the evidence code, the frequency of the feature, the curator, as well as the date of annotation are stated. For more detailed information on annotation guidelines and standards such as the correct nomenclature for frequency data and the evidence codes also see http://www.human-phenotype-ontology.org/index.php/annotation-guide.html.

Table 3. Textual Frequency Information for Annotations with HPO Terms

Description	Percentage of patients
Very rare	1
Rare	5
Occasional	7.5
Frequent	33
Typical	50
Common	75
Hallmark	90
Obligate	100

If more detailed frequency information is not available (e.g., 23 of 65 patients have feature X), the HPO annotation guide suggests eight frequency categories for disease annotations. For instance, if we say that feature α is hallmark in disease X, then we mean that 90% of individuals with disease X have feature α. Numerical values are given for the categories to provide a rough guide.

supplement data from OMIM and biocuration by the HPO team (Table 2 provides an example).

Semantic Similarity and Disease Classification

In order to demonstrate the usefulness of the HPO in neurogenetics, we downloaded the Orphanet classification of rare neurological diseases from the Web site www.orphadata.org. Each disease in this graph was assigned to the most general disease classification(s) (i.e., level one diseases). For example, the disease X-linked distal spinal muscular atrophy is classified as both neuromuscular disease and rare peripheral neuropathy. The 10 categories with the highest numbers of diseases were used for the analysis (see legend of Fig. 2). We then included all diseases that belonged to at least one but not more than four categories. For each of the selected diseases, we transferred all annotations from Orphanet to HPO terms and added these to the annotations that we already have from the corresponding OMIM entries. We used the Orphanet cross-reference file for this purpose, which maps Orphanet diseases to one or multiple OMIM entries. Diseases with less than three annotations were excluded at this step. Afterward, we calculated the semantic similarity between all pairs of diseases, using the symmetric similarity measure as described for the Phenomizer [Köhler et al., 2009].

At first, the similarity between two terms of the HPO has to be defined. For this, we determine the specificity of each term t of the HPO by calculating the information content (IC[t]). The IC of t is calculated as $-\log(p(t))$, where $p(t)$ is the probability of an annotation to t among all diseases. For each term, the set of all ancestors can be determined, which enables us to determine the set of common ancestors of two terms $CA(t_1,t_2)$. The similarity between two terms is calculated as the maximum information content of a common ancestor of the two terms: $\mathrm{sim}(t_1,t_2) = \max\{\mathrm{IC}(a) | a \in \mathrm{CA}(t_1,t_2)\}$.

Let be the set of terms annotated to disease D and let $|D| = n$ be the size of annot(D). The asymmetric semantic similarity (best-match average, $\mathrm{BMA_a}$) between two diseases $D1$ and $D2$ can be calculated as

$$\mathrm{BMA_a}(D1, D2) \frac{1}{|D1|} \sum_{t_1 \in \mathrm{annot}(D1)} \max_{t_2 \in \mathrm{annot}(D2)} \mathrm{sim}(t_1, t_2).$$

For the analysis presented here, the symmetric similarity score is used and calculated as:

$$\mathrm{BMA_s}(D1, D2) = \frac{1}{2}\mathrm{BMA_a}(D1, D2) + \frac{1}{2}\mathrm{BMA_a}(D2, D1).$$

If a pair of diseases ($D1$, $D2$) had a semantic similarity ($\mathrm{BMA_s}[D1,D2]$) of 3.0 or higher, we added an edge between the two diseases to the graph shown in Figure 2. The resulting network consists of 354 nodes and 1,316 edges. The nodes are colored according to their classification membership (using the *MultiColoredNodes Plugin* [Warsow et al., 2010] for Cytoscape [Smoot et al., 2011]). If an edge is drawn between two diseases that share the same category, the edge is colored using the color for that category. For edges connecting diseases that share multiple categories, we choose yellow (see legend). The edge thickness correlates to the semantic similarity between the connected diseases.

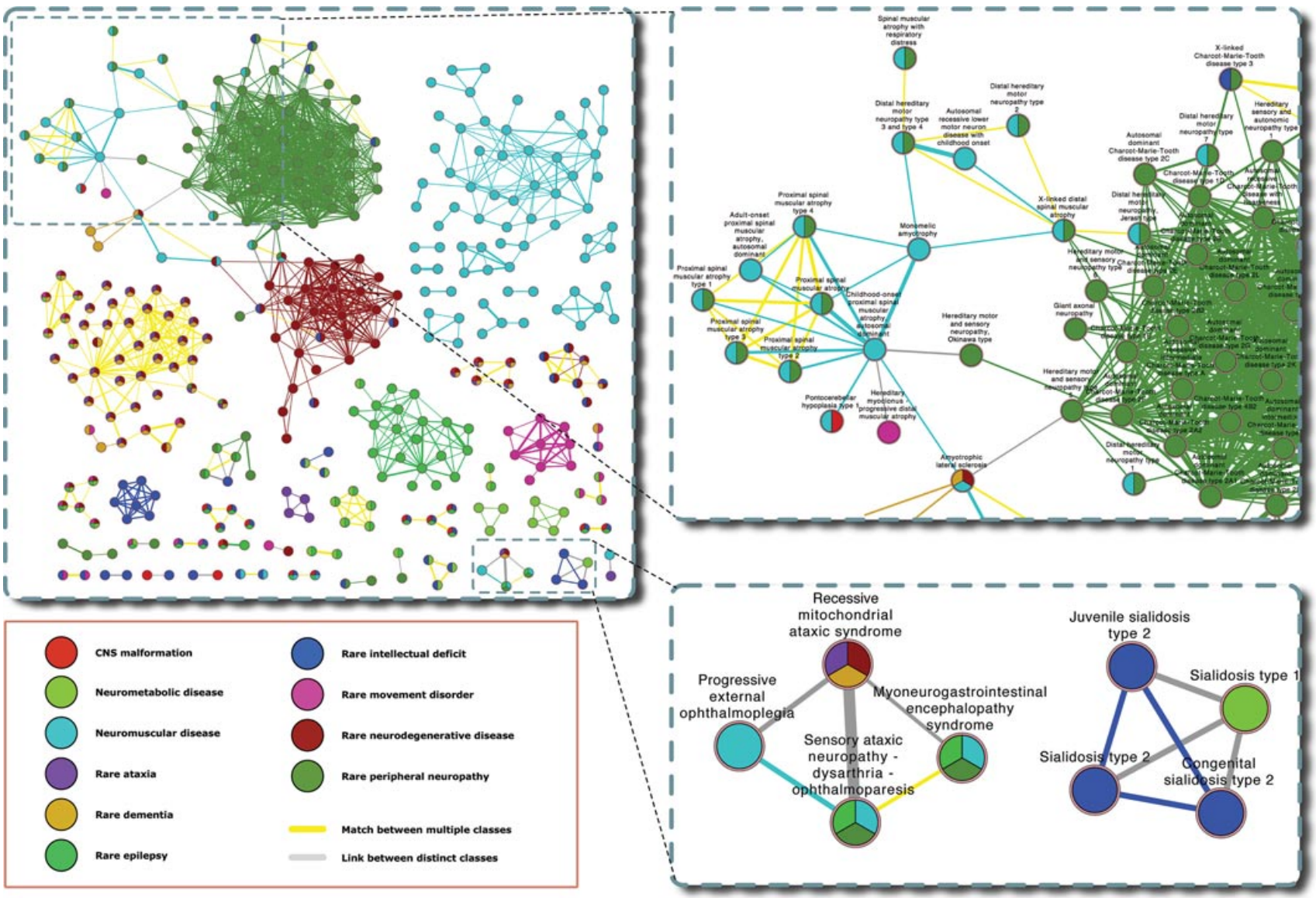

Figure 2. Network of rare neurological diseases based on semantic similarity in the HPO. The "organic layout" feature of Cytoscape was used to create the graph, and some nodes were moved manually to improve readability.

Discussion

Many different disease classifications have evolved over time, either based on clinical manifestations or on pathogenetic mechanisms. These classifications are an important tool for gaining insights into the relationship between pathomechanisms and phenotypic disease manifestations [Brunner and van Driel, 2004].

The neurogenetic disease network presented in this study was created on the basis of the phenotypic similarities between the combined annotations of Orphanet and OMIM neurogenetic diseases. A phenotypic abnormality in the HPO comprises more than simple morphological abnormalities (see Table 1) so that the resulting network also reflects similarities based on abnormalities found by medical imaging, abnormal laboratory findings, and so on.

The clustering shown in Figure 2 is a visualization of the phenotypic network, the "phenome," that exists between 10 classes of rare neurological diseases. Although our clustering procedure knew nothing of the assignment of the diseases to the 10 classes, the resulting network clearly shows that the diseases cluster largely according to the assignment by the Orphanet disease classification. This is an indirect confirmation of the correctness of the disease classification, in the sense that the classification of the diseases reflects the spectrum of shared and distinct phenotypic abnormalities that characterize the diseases. In addition to this, there appear to be some interesting interconnections between some of the clusters. For instance, a cluster of rare peripheral neuropathy diseases (dark green) shows a number of links to the diseases that are classified as neuromuscular diseases (turquoise). Note that in the Orphanet classification of rare neurological diseases, rare peripheral neuropathy diseases and neuromuscular diseases are separate groups. The links between these two classes in our network reflect the well known similarity between phenotypic abnormalities observed in muscular and neurologic diseases in this special field of neurogenetics. A phenotype such as muscle weakness may be either caused by a primarily muscular problem or a deficit of the nerves that innervate the muscles. The links between those two clusters reflect the situation in clinical practice and show that there is much more phenotypic similarity between peripheral neuropathies and neuromuscular diseases than between these two and, for example, ataxias (shown in purple).

In some cases, phenotypic similarity is present even between diseases that belong to different Orphanet classifications. In Figure 2, this is shown by a gray edge. For instance, type 1 sialidosis is classified as a neurometabolic disease, whereas sialidosis of type 2 is classified as rare intellectual deficit. Another example is Niemann–Pick disease type A, which is classified as a neurometabolic disease, rare epilepsy, and rare movement disorder. According to our analysis, this disease shows a high phenotypic similarity to Niemann–Pick disease type B, which is classified as rare peripheral neuropathy and thus shares no classification with the Niemann–Pick disease type A. Thus, phenotypic analysis can be used to point out areas that require curator attention, and in the future might also help to place novel neurogenetic diseases correctly—according to their phenotypic similarity to already existing structures and nosologies.

The HPO can thus be used to provide a useful measure of phenotypic similarity in the field of neurogenetics. It can provide a basis for computational biomedical research involving human phenotype analysis by linking phenotypic aspects to the genetic networks involving the disease genes. For instance, HPO terms were coupled with gene-set enrichment analysis in a recent study that identified phenotypes most relevant for proteins expressed in the human postsynaptic density (hPSD) from the human neocortex, with hPSD proteins being significantly enriched for a number of neural phenotypes that primarily involved cognition and motor functions, which was interpreted as suggesting that a subset of hPSD proteins are important for these functions (Bayés et al., 2011).

Currently, logical definitions are being developed for the HPO that will link each term to concepts from other ontologies including the FMA, the Gene Ontology [Blake et al., 2012], ChEBI [Degtyarenko et al., 2008], PATO, and others [Gkoutos et al., 2009]. This will enable integrative computational research in neuroscience, including interspecies phenotypic matching [Mungall et al., 2010; Washington et al., 2009], and can be used for improving the quality and agreement of ontologies [Köhler et al., 2011]. One of the major challenges in computational neuroscience is the interoperability of nomenclature. Major initiatives such as the NIF, the INCF, and the human brain project [Schutter et al., 2006] are developing methodologies for making the brain computable by providing compatible ontologies for describing the brain and neurosciences. The HPO will benefit from such initiatives, and future work will concentrate on aligning and extending the HPO's representation of neurological phenotypes according to other major ontologies in the computational neurosciences.

Mutation databases in neurogenetics, known as locus-specific databases (LSDBs), collect data on pathogenic mutations in individual genes. In some cases, clinical phenotypes and information about neutral polymorphisms are also offered [Cotton et al., 2008]. However, in the past, different LSDBs have used mutually incompatible standards for phenotype reporting, which has hampered large-scale computational research. Following the completion of the Human Genome Project, the need for a Human Phenome Project has been widely recognized [Butte and Kohane, 2006; Freimer and Sabatti, 2003], but there are still enormous obstacles. The Orphanet rare disease ontology and the HPO represent standards for reporting disease entities and phenotypic abnormalities that represent a first step toward creating a mutual resource for computational phenotype research and diagnostics in neurogenetics.

Acknowledgments

We would like to thank Melissa Haendel and Chris Mungall for useful comments on a draft version of this manuscript. We would also like to thank Gregor Warsow for his immediate support with the Cytoscape Plugin.

References

Amberger J, Bocchini CA, Scott AF, Hamosh A. 2009. McKusick's Online Mendelian Inheritance in Man (OMIM). Nucleic Acids Res 37(Database issue):D793–D796.

Aymé S. 2003. Orphanet, an information site on rare diseases. Soins; la revue de référence infirmière (672):46–47.

Aymé S, Schmidtke J. 2007. Networking for rare diseases: a necessity for Europe. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 50(12):1477–1483.

Bayés A, van de Lagemaat LN, Collins MO, Croning MDR, Whittle IR, Choudhary JS, Grant SGN. 2011. Characterization of the proteome, diseases and evolution of the human postsynaptic density. Nat Neurosci 14(1):19–21.

Bjaalie JG, Grillner S. 2007. Global neuroinformatics: the international neuroinformatics coordinating facility. J Neurosci 27(14):3613–3615.

Blake JA, Dolan M, Drabkin H, Hill DP, Ni L, Sitnikov D, Burgess S, Buza T, Gresham C, McCarthy F, Pillai L, Wang H, et al. 2012. The Gene Ontology: enhancements for 2011. Nucleic Acids Res 40(Database issue):D559–D564

Brunner HG, van Driel MA. 2004. From syndrome families to functional genomics. Nat Rev Genet 5(7):545–551.

Bug WJ, Ascoli GA, Grethe JS, Gupta A, Fennema-Notestine C, Laird AR, Larson SD, Rubin D, Shepherd GM, Turner JA, Martone ME. 2008. The NIFSTD and BIRNLex vocabularies: building comprehensive ontologies for neuroscience. Neuroinformatics 6(3):175–194.

Butte AJ, Kohane IS. 2006. Creation and implications of a phenome–genome network. Nat Biotechnol 24(1):55–62.

Caciotti A, Rocco MD, Filocamo M, Grossi S, Traverso F, d'Azzo A, Cavicchi C, Messeri A, Guerrini R, Zammarchi E, Donati MA, Morrone A. 2009. Type II sialidosis: review of the clinical spectrum and identification of a new splicing defect with chitotriosidase assessment in two patients. J Neurol 256(11):1911–1915.

Chinnery PF. 2010. Defining neurogenetic phenotypes (or how to compare needles in haystacks). Brain 133(Pt 3):649–651.

Cook DL, Mejino JLV, Rosse C. 2004. Evolution of a foundational model of physiology: symbolic representation for functional bioinformatics. Stud Health Technol Inform 107(Pt 1):336–340.

Cotton RGH, Auerbach AD, Beckmann JS, Blumenfeld OO, Brookes AJ, Brown AF, Carrera P, Cox DW, Gottlieb B, Greenblatt MS, Hilbert P, Lehvaslaiho H, et al. 2008. Recommendations for locus-specific databases and their curation. Hum Mutat 29(1):2–5.

Degtyarenko K, de Matos P, Ennis M, Hastings J, Zbinden M, McNaught A, Alcàntara R, Darsow M, Guedj M, Ashburner M. 2008. ChEBI: a database and ontology for chemical entities of biological interest. Nucleic Acids Res 36(Database issue):D344–D350.

Dou D, Frishkoff G, Rong J, Frank R, Malony A, Tucker D. 2007. Development of neuroelectromagnetic ontologies (nemo): a framework for mining brainwave ontologies. Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.

Freimer N, Sabatti C. 2003. The human phenome project. Nat Genet 34(1):15–21.

Fryns JP, de Ravel TJL. 2002. London dysmorphology database, London neurogenetics database and dysmorphology photo library on cd-rom [version 3] 2001. Hum Genet 111(1):113.

Gardner D, Akil H, Ascoli GA, Bowden DM, Bug W, Donohue DE, Goldberg DH, Grafstein B, Grethe JS, Gupta A, Halavi M, Kennedy DN, et al. 2008. The neuroscience information framework: a data and knowledge environment for neuroscience. Neuroinformatics 6(3):149–160.

Gkoutos GV, Mungall C, Dölken S, Ashburner M, Lewis S, Hancock J, Schofield P, Köhler S, Robinson PN. 2009. Entity/quality-based logical definitions for the human skeletal phenome using PATO. Conf Proc IEEE Eng Med Biol Soc 2009:7069–7072.

Köhler S, Bauer S, Horn D, Robinson PN. 2008. Walking the interactome for prioritization of candidate disease genes. Am J Hum Genet 82(4):949–958.

Köhler S, Bauer S, Mungall CJM, Carletti G, Smith CL, Schofield P, Gkoutos GV, Robinson PN. 2011. Improving ontologies by automatic reasoning and evaluation of logical definitions. BMC Bioinformatics 12(1):418.

Köhler S, Schulz MH, Krawitz P, Bauer S, Dölken S, Ott CE, Mundlos C, Horn D, Mundlos S, Robinson PN. 2009. Clinical diagnostics in human genetics with semantic similarity searches in ontologies. Am J Hum Genet 85(4):457–464.

Larson SD, Fong LL, Gupta A, Condit C, Bug WJ, Martone ME. 2007. A formal ontology of subcellular neuroanatomy. Front Neuroinform 1:3.

Lindblom A, Robinson PN. 2011. Bioinformatics for human genetics: promises and challenges. Hum Mutat 32(5):495–500.

Mahner M, Kary M. 1997. What exactly are genomes, genotypes and phenotypes? and what about phenomes? J Theor Biol 186(1):55–63.

Martin RF, Mejino JL, Bowden DM, Brinkley JF, Rosse C. 2001. Foundational model of neuroanatomy: implications for the human brain project. Proc AMIA Symp:438–442.

Martin RF, Rickard K, Mejino JLV, Agoncillo AV, Brinkley JF, Rosse C, Group SI. 2003. The evolving neuroanatomical component of the foundational model of anatomy. AMIA Annual Symposium Proc:927.

Mungall CJ, Gkoutos GV, Smith CL, Haendel MA, Lewis SE, Ashburner M. 2010. Integrating phenotype ontologies across multiple species. Genome Biol 11(1):R2.

Robinson PN, Bauer S. 2011. Introduction to bio-ontologies. CRC Press Inc. Boca Raton, FL.

Robinson PN, Köhler S, Bauer S, Seelow D, Horn D, Mundlos S. 2008. The human phenotype ontology: a tool for annotating and analyzing human hereditary disease. Am J Hum Genet 83(5):610–615.

Robinson PN, Krawitz P, Mundlos S. 2011. Strategies for exome and genome sequence data analysis in disease-gene discovery projects. Clin Genet 80(2):127–132.

Robinson, PN Mundlos S. 2010. The human phenotype ontology. Clin Genet 77(6):525–534.

Rosse C, Mejino JLV. 2003. A reference ontology for biomedical informatics: the Foundational Model of Anatomy. J Biomed Inform 36: 478–500.

Rubin DL, Talos IF, Halle M, Musen MA, Kikinis R. 2009. Computational neuroanatomy: ontology-based representation of neural components and connectivity. BMC Bioinformatics 10 (Suppl 2):S3.

Schutter ED, Ascoli GA, Kennedy DN. 2006. On the future of the human brain project. Neuroinformatics 4(2):129–130.

Smith B, Rosse C. 2004. The role of foundational relations in the alignment of biomedical ontologies. Stud Health Technol Inform 107(Pt 1):444–448.

Smoot ME, Ono K, Ruscheinski J, Wang PL, Ideker T. 2011. Cytoscape 2.8: new features for data integration and network visualization. Bioinformatics 27(3):431–1432.

Talos IF, Rubin DL, Halle M, Musen M, Kikinis R. 2008. A prototype symbolic model of canonical functional neuroanatomy of the motor system. J Biomed Inform 41(2):251–263.

Turner JA, Mejino JLV, Brinkley JF, Detwiler LT, Lee HJ, Martone ME, Rubin DL. 2010. Application of neuroanatomical ontologies for neuroimaging data annotation. Front Neuroinform 4:1–4.

Warsow G, Greber B, Falk SS, Harder C, Siatkowski M, Schordan S, Som A, Endlich N, Schöler H, Repsilber D, Endlich K, Fuellen G. 2010. ExprEssence—revealing the essence of differential experimental data in the context of an interaction/regulation net-work. BMC Syst Biol 4:164.

Washington NL, Haendel MA, Mungall CJ, Ashburner M, Westerfield M, Lewis SE. 2009. Linking human diseases to animal models using ontology-based phenotype annotation. PLoS Biol 7(11):e1000247.

Webb AJ, Thorisson GA, Brookes AJ, GEN2PHEN Consortium. 2011. An informatics project and online "knowledge centre" supporting modern genotype-to-phenotype research. Hum Mutat 32(5):543–550.