Commentary on Shimoyama et al. (2012): three ontologies to define phenotype measurement data

John M. Hancock *

Department of Physiology, Development and Neuroscience, University of Cambridge, Cambridge, UK
*Correspondence: jmhancock@gmail.com

A commentary on

Three ontologies to define phenotype measurement data
by Shimoyama, M., Nigam, R., Mcintosh, L. S., Nagarajan, R., Rice, T., Rao, D. C., et al. (2012). Front. Genet. 3:87. doi: 10.3389/fgene.2012.00087

As human genomics moves into a mass-scale era, whereby millions of genome sequences will soon become available, new opportunities are opening up to use these very large samples better to understand the relationship between genotype and phenotype.

A fundamental problem for such studies in the past has been lack of standardization in the description of the phenotypes used. Not only have disease concepts often been confused with phenotypes (most diseases manifest numerous, distinct phenotypes which not only makes up the disease description but often can be observed in a number of diseases), but the concepts used have at times been vague (Hancock et al., 2009; Schofield et al., 2010).

In addition, model organisms are frequently used to study disease-, and more broadly phenotype-related phenomena in systems with applicability to humans but which are not subject to equivalent ethical problems or issues of data protection. A key requirement for future computational analysis of the relationship between genotype and phenotype in human will therefore be to include knowledge from model organisms (Hancock et al., 2009).

Significant progress has been made in using ontologies to describe human and model organism phenotypes in recent years. Many computational biology communities, serving particular model organism experimental communities, have developed approaches to the ontological description of phenotypes, often associated with community databases. As an example, the Mammalian Phenotype ontology (MP) (Smith and Eppig, 2012) was developed in association with the Mouse Genome Database (Blake et al., 2014) to facilitate consistent annotation of phenotypes associated with genomic differences. The MP, although originally developed for mouse, was subsequently applied to rat in the context of the Rat Genome Database (Nigam et al., 2013). For humans the Human Phenotype Ontology (HPO) (Kohler et al., 2014) has been developed to reflect the atomic features of diseases, initially making use of the OMIM (Online Mendelian Inheritance in Man) resource (Amberger et al., 2011).

A drawback of these approaches to ontological description, which make use of so-called pre-composed ontologies which are prepared in advance to the annotation process, is that they are unable to represent the full range of phenotypic observations including "normal" states, or of representing subtle differences or numerical values. To address this, an alternative approach, often known as the PATO (Phenotype and Trait Ontology) approach, has been developed (Bard and Rhee, 2004; Gkoutos et al., 2004, 2009). This aims to make use of a group of compatible ontologies to produce combinatorial expressions of the type:

*Entity **E** has attribute **A** of value **V** when measured in organism **O** using test **T** under conditions **C**.*

where elements in bold are terms from an appropriate ontology. The full implementation of such an approach is yet to be realized, although starts have been made in databases such as Zfin (Howe et al., 2013) and Europhenome (Morgan et al., 2010).

A key missing element in such a compositional approach is standard descriptions of experimental methods and conditions. Over the last decade or so a number of ontological approaches to defining experimental conditions have been developed. The MGED Ontology (Whetzel et al., 2006b) was developed to underpin the fulfillment of the MIAME (Minimum Information About a Microarray Experiment) metadata criteria (Brazma et al., 2001). The HUPO (Human Proteome Organization) PSI (Proteomics Standards Initiative) Mass Spectroscopy Vocabularies (Mayer et al., 2013) facilitate the description of experiments in proteomics and mass spectroscopy. The Metabolomics Standards Initiative has established COSMOS (COordination of Standards in MetabOlomicS) (Steinbeck et al., 2012) to describe metabolomics experiments, making use of the ISA (Investigation/Study/Assay) framework (Sansone et al., 2012). FuGO (the Functional Genomics Investigation Ontology) was developed to provide a broader structure for functional genomics experiments (Whetzel et al., 2006a). EXPO (Soldatova and King, 2006) attempts to provide a higher level ontology to which such domain-specific ontologies can be integrated.

An important addition to the armory of ontological frameworks that can be used to describe phenotypes is provided by Shimoyama et al. (2012) who describe a set of three ontologies that can be used to describe clinical measurements, measurement methods and experimental conditions for traits common to rat and man (and, by extension, to other mammalian

model systems such as mouse and, potentially, more distantly related species). Their approach extends the availability of experimental description ontologies in a whole new direction and, crucially the types of measurement they can describe using these ontologies are similar to those used in large-scale phenotyping experiments on mouse models of human disease (Hancock and Dobbie, 2014). Their ontology system can be used to describe both human and model mammal phenotyping measurements (**T** and **C** in the above expression). It therefore provides an underpinning component for the computational study of genotype-phenotype relations in humans and model mammals. At the same time it provides a valuable set of terms and relations to facilitate more systematic annotation and searching of phenotype terms across human and model organism databases. This opens up exciting new opportunities for the unified analysis of human and mouse disease and phenotype data.

REFERENCES

Amberger, J., Bocchini, C., and Hamosh, A. (2011). A new face and new challenges for Online Mendelian Inheritance in Man (OMIM(R)). *Hum. Mutat.* 32, 564–567. doi: 10.1002/humu.21466

Bard, J. B., and Rhee, S. Y. (2004). Ontologies in biology: design, applications and future challenges. *Nat. Rev. Genet.* 5, 213–222. doi: 10.1038/nrg1295

Blake, J. A., Bult, C. J., Eppig, J. T., Kadin, J. A., and Richardson, J. E. (2014). The Mouse Genome Database: integration of and access to knowledge about the laboratory mouse. *Nucleic Acids Res.* 42, D810–D817. doi: 10.1093/nar/gkt1225

Brazma, A., Hingamp, P., Quackenbush, J., Sherlock, G., Spellman, P., Stoeckert, C., et al. (2001). Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat. Genet.* 29, 365–371. doi: 10.1038/ng1201-365

Gkoutos, G. V., Green, E. C. J., Mallon, A.-M., Hancock, J. M., and Davidson, D. (2004). Using ontologies to describe mouse phenotypes. *Genome Biol.* 6, R8. doi: 10.1186/gb-2004-6-1-r8

Gkoutos, G. V., Mungall, C., Dolken, S., Ashburner, M., Lewis, S., Hancock, J., et al. (2009). Entity/quality-based logical definitions for the human skeletal phenome using PATO. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 2009, 7069–7072. doi: 10.1109/IEMBS.2009.533336

Hancock, J. M., and Dobbie, M. (2014). "Phenomics of the laboratory mouse," in *Phenomics*, ed J. M. Hancock (Enfield, NH: Science Publishers/CRC Press), 24–64.

Hancock, J. M., Mallon, A. M., Beck, T., Gkoutos, G. V., Mungall, C., and Schofield, P. N. (2009). Mouse, man, and meaning: bridging the semantics of mouse phenotype and human disease. *Mamm. Genome* 20, 457–461. doi: 10.1007/s00335-009-9208-3

Howe, D. G., Bradford, Y. M., Conlin, T., Eagle, A. E., Fashena, D., Frazer, K., et al. (2013). ZFIN, the Zebrafish Model Organism Database: increased support for mutants and transgenics. *Nucleic Acids Res.* 41, D854–D860. doi: 10.1093/nar/gks938

Kohler, S., Doelken, S. C., Mungall, C. J., Bauer, S., Firth, H. V., Bailleul-Forestier, I., et al. (2014). The Human Phenotype Ontology project: linking molecular biology and disease through phenotype data. *Nucleic Acids Res.* 42, D966–D974. doi: 10.1093/nar/gkt1026

Mayer, G., Montecchi-Palazzi, L., Ovelleiro, D., Jones, A. R., Binz, P. A., Deutsch, E. W., et al. (2013). The HUPO proteomics standards initiative- mass spectrometry controlled vocabulary. *Database (Oxford)* 2013, bat009. doi: 10.1093/database/bat009

Morgan, H., Beck, T., Blake, A., Gates, H., Adams, N., Debouzy, G., et al. (2010). EuroPhenome: a repository for high-throughput mouse phenotyping data. *Nucleic Acids Res.* 38, D577–D585. doi: 10.1093/nar/gkp1007

Nigam, R., Lauiederkind, S. J., Hayman, G. T., Smith, J. R., Wang, S. J., Lowry, T. F., et al. (2013). Rat Genome Database: a unique resource for rat, human, and mouse quantitative trait locus data. *Physiol. Genomics* 45, 809–816. doi: 10.1152/physiolgenomics.00065.2013

Sansone, S. A., Rocca-Serra, P., Field, D., Maguire, E., Taylor, C., Hofmann, O., et al. (2012). Toward interoperable bioscience data. *Nat. Genet.* 44, 121–126. doi: 10.1038/ng.1054

Schofield, P. N., Gkoutos, G. V., Gruenberger, M., Sundberg, J. P., and Hancock, J. M. (2010). Phenotype ontologies for mouse and man; bridging the semantic gap. *Dis. Model. Mech.* 3, 281–289. doi: 10.1242/dmm.002790

Shimoyama, M., Nigam, R., Mcintosh, L. S., Nagarajan, R., Rice, T., Rao, D. C., et al. (2012). Three ontologies to define phenotype measurement data. *Front. Genet.* 3:87. doi: 10.3389/fgene.2012.00087

Smith, C. L., and Eppig, J. T. (2012). The Mammalian Phenotype Ontology as a unifying standard for experimental and high-throughput phenotyping data. *Mamm. Genome* 23, 653–668. doi: 10.1007/s00335-012-9421-3

Soldatova, L. N., and King, R. D. (2006). An ontology of scientific experiments. *J. R. Soc. Interface* 3, 795–803. doi: 10.1098/rsif.2006.0134

Steinbeck, C., Conesa, P., Haug, K., Mahendraker, T., Williams, M., Maguire, E., et al. (2012). MetaboLights: towards a new COSMOS of metabolomics data management. *Metabolomics* 8, 757–760. doi: 10.1007/s11306-012-0462-0

Whetzel, P. L., Brinkman, R. R., Causton, H. C., Fan, L., Field, D., Fostel, J., et al. (2006a). Development of FuGO: an ontology for functional genomics investigations. *OMICS* 10, 199–204. doi: 10.1089/omi.2006.10.199

Whetzel, P. L., Parkinson, H., Causton, H. C., Fan, L., Fostel, J., Fragoso, G., et al. (2006b). The MGED Ontology: a resource for semantics-based description of microarray experiments. *Bioinformatics* 22, 866–873. doi: 10.1093/bioinformatics/btl005