Drug Discovery Today: Disease Models
Vol. xxx, No. xx 2017

Tools for exploring mouse models of human disease

Melissa Haendel[1], Irene Papatheodorou[2], Anika Oellrich[3], Christopher J. Mungall[4], Nicole Washington[4], Suzanna E. Lewis[4], Peter N. Robinson[5], Damian Smedley[6,*]

[1]University Library and Department of Medical Informatics and Epidemiology, Oregon Health & Science University, Portland, OR, USA
[2]European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK
[3]Institute of Psychiatry, Psychology & Neuroscience, NIHR Biomedical Research Centre, Kings College London, London, UK
[4]Genomics Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA
[5]Jackson Laboratory for Genomic Medicine, Farmington, CT, USA
[6]Queen Mary University of London — William Harvey Research Institute, Barts & The London School of Medicine & Dentistry, London, UK

Despite significant computational challenges, a number of tools have been developed recently to leverage the mouse to model human disease. Here we review these tools and show how they can be applied in the identification of candidate genes and therapeutic targets as well as mouse models for mechanistic studies and drug validation.

Section editor: Steve Brown – MRC Harwell Institute, Mammalian Genetics Unit, Oxfordshire, OX11 0RD, UK.

Introduction

The first step towards understanding the aetiology of human heritable diseases and developing potential new treatments is to unravel the relationship between genotype and phenotype. Despite new sequencing technologies leading to an explosion in the number of diseases for which we now know the causative variant [1] and the efforts of the Gene Ontology consortium [2][GO; 2], for many human genes we do not have functional or phenotype data. The International Mouse Phenotype Consortium (IMPC) aims to complete the functional catalogue of all protein-coding genes in the genome by 2020 [3], making the mouse an invaluable resource for understanding human disease and developing new therapies. However, until recently this data has been under-utilised due to the challenges of comparing human and mouse phenotypes and the lack of tools allowing researchers to perform these comparisons [4]. Here we discuss these challenges and some of the solutions that have been developed.

Sources of phenotype data

The Online Inheritance in Man (OMIM) knowledgebase catalogues almost 8000 known genetic human diseases along with any known associated genes [URL: http://www.omim.org]. Orphanet focuses on the rare disease component and orphan drugs that could potentially be used to treat them [5] [5,URL: http://www.orpha.net]. Both resources contain substantial amounts of descriptive signs and symptoms data for diseases. The development of the Human Phenotype Ontology (HP) to describe phenotypes in OMIM and Orphanet in a standardised manner has made such descriptions amenable to computational analysis [6]. HP now contains almost

*Corresponding author: D. Smedley (d.smedley@qmul.ac.uk)

 http://dx.doi.org/10.1016/j.ddmod.2017.07.004

Please cite this article in press as: Haendel M, et al. Tools for exploring mouse models of human disease, Drug Discov Today: Dis Model (2017), http://dx.doi.org/10.1016/j.ddmod.2017.07.004

12,000 phenotype terms and HP annotations exist for 7927 OMIM and Orphanet disorders.

The Mouse Genome Database (MGD) curates phenotype associations for mouse mutants described in literature and makes these annotations publically available [7]. MGD contains some 304,000 phenotype annotations for over 59,000 different mouse lines involving disruptions in some 12,000 genes. Although this phenotype data is fairly comprehensive across organ systems, phenotypes that were not the focus of any given published study can obviously be missing. The IMPC is attempting to counteract this by performing a standardised pipeline of phenotyping assays on all mouse lines and currently contains phenotype data for 3868 genes. The Mouse Phenome Database (MPD) is a complementary database that assembles the results of quantitative assays on standard (non-mutagenized) laboratory mouse strains, measuring both baseline phenotypes and effects of drug, diet, disease, and aging. MPD contains more than 1,206,730 quantitative measurements on nearly 1549 strains. All three resources, MGD, IMPC, and MPD, annotate phenotypes using the Mammalian Phenotype ontology (MP), which contains almost 12,000 terms [8].

Currently around 3400 human genes have HP annotations via their association to one or more diseases. In contrast, mouse mutants involving only a single gene disruption and with MP annotation exist for 11,155 genes, with only 2597 overlapping with the human set. Therefore even in these early stages of the IMPC project, there is an abundance of genotype-phenotype information that is available only in the mouse and potentially translatable to human disease studies.

Cross-species phenotype mapping

The major challenge in using the mouse genotype-phenotype associations in the context of human disease research comes from the fact that the human and mouse communities use different vocabularies, as evidenced by the content of the MP and HP ontologies. For example, a computer or even a non-specialist researcher would typically not know that the HP term *craniosynostois* (HP:0001363) is equivalent to the MP term *premature suture closure* (MP:0000081). An approach termed decomposition has been developed to logically define these species-specific phenotype terms as a combination of an entity (E), representing the anatomical structure or biological process, and a quality (Q), representing what is abnormal about the entity [9,10]. These EQ statements use species-agnostic terms from standardised ontologies such as GO, CHEBI or the UBERON multi-species anatomy ontology [11] for the entities and the Phenotype and Trait Ontology (PATO) for the qualities. In the above example, both the HP and MP terms are represented by the *premature closure* (PATO:0002166) of the *suture* (UBERON:0000969).

Taking the annotated human and mouse phenotype datasets and these EQ statements, software can automatically detect the similarities between a human disease and mouse mutant [9,12]. The hierarchical structure of the ontologies can be taken advantage of to detect non-exact matches; e.g. a clinical phenotype of *speech articulation problems* and a mouse mutant exhibiting abnormal *larynx morphology* would share a common phenotype of *abnormality of the larynx*. Finally, measures of semantic similarity [13] such as the Jaccard index can be combined with the Information Content of the match to score the similarity between a pair of human and mouse phenotype terms. Further, these can be aggregated to compare how similar a human disease and mouse mutant are, based upon their sets of phenotypes.

Tools for finding mouse models of human disease

Publically available websites utilise the above approach to generate a ranked list of mouse models for a chosen human disease e.g. MouseFinder [14], PhenoDigm [15], PhenomeNET [16], and the Monarch Initiative [URL: http://monarchinitiative.org; 17]. In the context of diseases with no known molecular association, these tools can suggest candidate disease genes based upon comparing a set of patient phenotypes against mouse models. Where the gene is known these sites can be queried to see if a mouse model exists that closely phenocopies the clinical phenotypes or contains specific phenotypes of interest (Fig. 1). This computational approach to compare patient phenotypes against mouse models enables visualization tools illustrating the ways in which a model recapitulates disease (Fig. 2). Furthermore, one can similarly compare mutants involving different alleles, zygosity, and genetic backgrounds to optimize model selection. Many of these mouse models can then be ordered from public repositories for mechanistic studies and therapeutic target validation experiments e.g. European Mouse Mutant Archive [URL: https://www.infrafrontier.eu/resources-and-services/access-emma-mouse-resources] or JAX mice [URL: http://jaxmice.jax.org].

Another study used the cross-species phenotype comparison approach to assess the contribution of multiple genes within copy number variation (CNV) regions to the disease phenotype [18]. Evidence was presented for multi-gene contributions to certain CNV syndromes as well as examples where a single-gene was responsible as has traditionally been described for CNV disorders. A Cytoscape plug-in has now been developed for researchers to explore their own CNV patients using this approach [URL: http://compbio.charite.de/contao/index.php/phenoviz.html]

Disease models from the International Mouse Phenotyping Consortium

As described elsewhere in this issue, the IMPC is performing a standardised pipeline of phenotyping experiments on mouse knockouts for every protein-coding gene. These assays are tightly controlled with numerous controls and a robust

Please cite this article in press as: Haendel M, et al. Tools for exploring mouse models of human disease, Drug Discov Today: Dis Model (2017), http://dx.doi.org/10.1016/j.ddmod.2017.07.004