ORPHANET JOURNAL OF RARE DISEASES

ORAL PRESENTATION

Open Access

Use of animal models for exome prioritization of rare disease genes

Damian Smedley[1*], Sebastian Kohler[2,3], William Bone[4], Anika Oellrich[1], Jules Jacobsen[1],
Sanger Mouse Genetics Project[1], Kai Wang[5], Chris Mungall[6], Nicole Washington[6], Sebastian Bauer[2,3],
Dominic Seelow[7], Peter Krawitz[2,3,8], Cornelius Boerkel[4], Christian Gilissen[9], Melissa Haendel[10], Suzanna E Lewis[5],
Peter N Robinson[2,3,7]

From 7th European Conference on Rare Diseases and Orphan Products (ECRD 2014)

Over 100 disease-gene associations have been identified by whole-exome sequencing since the first reports in 2010, leading to a revolution in rare disease-gene discovery [1,2]. However, many cases remain unsolved due to the fact that ~100-1000 loss of function, candidate variants remain after removing those deemed as common, low quality or non-pathogenic. In some cases it may be possible to use multiple affected individuals, linkage data, identity-by-descent inference, identification of *de novo* heterozygous mutations from trio analysis, or prior knowledge of affected pathways to narrow down to the causative variant [3]. Where this is not possible or successful, one approach is to use phenotype data to evaluate whether a particular candidate variant is likely to result in the patient's clinical manifestations.

Model organism phenotype data represents a highly pertinent but under-utilised resource for such disease gene discovery. Whilst some 1800 human genes were associated with human phenotype ontology annotations (HPO) at the time of publication, a further 5700 genes have been shown to have phenotype data available from mouse and zebrafish model organism databases [4]. We have previously developed algorithmic approaches to semantically compare disease phenotypes with mouse and zebrafish phenotypes for disease candidate gene identification [5-8].

We have previously reported that comparisons to mouse phenotype data can dramatically increase the performance of exome analysis prioritization [9]. In the work presented here we combine the comparison of patient phenotypes to known disease as well as mouse and zebrafish phenotypes for each candidate variant in the exome. Where phenotype data is not available for a candidate we use proximity in protein-protein networks to genes with phenotype data to inform on candidacy based on guilt-by-association. The output is combined with measures of variant candidacy such as pathogenicity and allele frequency and synergistically improves performance: the causative variant is identified as the top hit in up to 96% of exomes for known associations and 49% of exomes for previously undescribed associations.

Our software, Exomiser, is openly available to use at our website [http://www.sanger.ac.uk/resources/databases/exomiser/query] and for download to perform local analysis. We are currently collaborating with the NIH Undiagnosed Disease Program to achieve diagnosis of problematic cases through exome analysis. In conclusion, our results clearly show the value of collecting comprehensive clinical phenotype data for translational bioinformatics and future work will focus on producing a robust solution for clinical diagnostics.

Authors' details
[1]Mouse Genetics Project, Wellcome Trust Sanger Institute, Hinxton, UK. [2]Institute for Medical and Human Genetics, Charite-Universitatsmedizin Berlin, Augustenburger Platz 1, 13353 Berlin, Germany. [3]Max Planck Institute for Molecular Genetics, Ihnestr. 63–73, 14195 Berlin, Germany. [4]National Institutes of Health, NHGRI, UDP Translational Laboratory, Rockville, USA. [5]Center for Applied Genomics, Childrens Hospital of Philadelphia, University of Pennsylvania, Philadelphia, PA 19104, USA. [6]Genomics Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. [7]Department of Neuropaediatrics Charite-Universitatsmedizin Berlin, Augustenburger Platz 1, 13353 Berlin, Germany. [8]Berlin Brandenburg Center for Regenerative

[1]Mouse Genetics Project, Wellcome Trust Sanger Institute, Hinxton, UK
Full list of author information is available at the end of the article

Therapies, Charite-Universitatsmedizin Berlin, Augustenburger Platz 1, 13353 Berlin, Germany. [9]Department of Human Genetics, Nijmegen Centre for Molecular Life Sciences and Institute for Genetic and Metabolic Disorders, Radboud University Nijmegen Medical Centre, Nijmegen, the Netherlands. [10]University Library and Department of Medical Informatics and Epidemiology, Oregon Health & Sciences University, Portland, OR, USA.

References

1. Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, Huff CD, Shannon PT, Jabs EW, Nickerson DA, *et al*: Exome sequencing identifies the cause of a Mendelian disorder. *Nat Genet* 2010, **42**:30-35.
2. Rabbani B, Mahdieh N, Hosomichi K, Nakaoka H, Inoue I: Next-generation sequencing: Impact of exome sequencing in characterizing Mendelian disorders. *J Hum Genet* 2012, **57**:621-632.
3. Robinson PN, Krawitz P, Mundlos S: Strategies for exome and genome sequence data analysis in disease-gene discovery projects. *Clin Genet* 2011, **80**:127-132.
4. Doelken SC, Köhler S, Mungall CJ, Gkoutos GV, Ruef BJ, Smith C, Smedley D, Bauer S, Klopocki E, Schofield PN, *et al*: Phenotypic overlap in the contribution of individual genes to CNV pathogenicity revealed by cross-species computational analysis of single-gene mutations in humans, mice and zebrafish. *Dis Model Mech* 2013, **6**:358-372.
5. Smedley D, Oellrich A, Köhler S, Ruef B, Sanger Mouse Genetics Project, Westerfield M, Robinson P, Lewis S, Mungall C: PhenoDigm: Analyzing curated annotations to associate animal models with human diseases. *Database (Oxford)* 2013, bat025.
6. Chen CK, Mungall CJ, Gkoutos GV, Doelken SC, Köhler S, Ruef BJ, Smith C, Westerfield M, Robinson PN, Lewis SE, Schofield PN, Smedley D: MouseFinder: Candidate disease genes from mouse phenotype data. *Hum Mutat* 2012, **33(5)**:858-66.
7. Mungall CJ, Gkoutos GV, Smith CL, Haendel MA, Lewis SE, Ashburner M: Integrating phenotype ontologies across multiple species. *Genome Biol* 2010, **11(1)**:R2.
8. Washington NL, Haendel MA, Mungall CJ, Ashburner M, Westerfield M, Lewis SE: Linking human diseases to animal models using ontology-based phenotype annotation. *PLoS Biol* 2009, **7(11)**:e1000247.
9. Robinson PN, Köhler S, Oellrich A, Sanger Mouse Genetics Project, Wang K, Mungall CJ, Lewis SE, Washington N, Bauer S, Seelow D, Krawitz P, Gilissen C, Haendel M, Smedley D: Improved exome prioritization of disease genes through cross-species phenotype comparison. *Genome Res* 2014, **24(2)**:340-8.