


COMMENT OPEN



Early illustrations of the importance of systematic phenotyping

Reuben J. Pengelly ¹✉

© The Author(s) 2022

European Journal of Human Genetics (2022) 30:1102; <https://doi.org/10.1038/s41431-022-01165-z>

The Online Mendelian Inheritance in Man (OMIM) database (www.omim.org) [1] (and its offline predecessor) has long been a ubiquitous reference for medical genetics professionals and researchers. At its migration from a textbook to an online resource in the 1980s, it was ahead of its time in many ways, though of course the fast-evolving field of medical genetics and genomics now has considerations that would not have been imagined at the time. Notably, the human-readable information in OMIM at the time was not well suited for computational analysis.

To address the challenges of computational analysis of OMIM data at the time, van Driel et al. [2] undertook a text-mining analysis, published in 2006. In this, they extracted MeSH terms present in the text of 5080 OMIM records for disease phenotypes and clustered these records based on similarity in MeSH terms within the records. On exploring the similar OMIM disease phenotypes, they identified that phenotypically linked diseases were more likely to have a shared molecular basis, and furthermore, pathogenic variants in genes in the same family (or containing same functional domains) were more likely to underlie phenotypically similar diseases. Exploring their dataset may support the identification of candidate genes for diseases with unknown molecular bases. The results of their analysis were released in their online tool MimMiner (<http://www3.cmbi.umcn.nl/MimMiner>), allowing the user to explore an OMIM record in relation to phenotypically similar records.

Since 2006, there has been substantial progress in systematic phenotyping, notably including the development of the Human Phenotype Ontology (HPO) [3]. This ontology has been developed with reference to databases including OMIM. HPO is increasingly being incorporated into routine genomics, and is underpinning the implementation of national genomics programmes such as the 100,000 Genomes Project [4] and the subsequent business as usual NHS Genomic Medicine Service.

Software utilising HPO for diagnostics, such as Exomiser and PhenIX, utilise HPO terms for a patient to identify variants in a patient in genes that would be expected to cause the phenotype from the literature. These software allow for the systematic interpretation of genomic data, streamlining interpretation and allowing efficient use of time for busy clinical staff [5]. Systematic collection of detailed phenotypes when patients are referred for diagnostic testing further allows the leveraging of patient cohorts for research; this allows long-term diagnostic uplift through gene discovery [6].

Phenotyping now with systematic taxonomies in the first instance renders text-mining approaches, such as that applied by van Driel et al. [2] unnecessary, but they are still a useful tool to extract information *post hoc*, allowing valuable information to be pulled into databases going forwards. In the world of ‘big data’, consistency and robustness of data is key. With any data analysis, the results are

always limited by the quality of the information going in, so any work we do to improve data quality will pay dividends.

REFERENCES

1. Amberger JS, Bocchini CA, Scott AF, Hamosh A. OMIM.org: leveraging knowledge across phenotype–gene relationships. *Nucleic Acids Res.* 2018;47:D1038–D1043.
2. van Driel MA, Bruggeman J, Vriend G, Brunner HG, Leunissen JAM. A text-mining analysis of the human phenome. *Eur J Hum Genet.* 2006;14:535–42.
3. Köhler S, Gargano M, Matentzoglou N, Carmody LC, Lewis-Smith D, Vasilevsky NA, et al. The human phenotype ontology in 2021. *Nucleic Acids Res.* 2020;49:D1207–D1217.
4. 100,000 Genomes Project Pilot Investigators. 100,000 Genomes pilot on rare-disease diagnosis in health care — preliminary report. *N Engl J Med.* 2021;385:1868–80.
5. Pengelly RJ, Alom T, Zhang Z, Hunt D, Ennis S, Collins A. Evaluating phenotype-driven approaches for genetic diagnoses from exomes in a clinical setting. *Sci Rep.* 2017;7:13509.
6. Son JH, Xie G, Yuan C, Ena L, Li Z, Goldstein A, et al. Deep phenotyping on electronic health records facilitates genetic diagnosis by clinical exomes. *Am J Hum Genet.* 2018;103:58–73.

FUNDING

No specific funding.

COMPETING INTERESTS

The author declares no competing interests.

ADDITIONAL INFORMATION

Correspondence and requests for materials should be addressed to Reuben J. Pengelly.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2022

¹Faculty of Medicine, University of Southampton, Southampton, UK. ✉email: R.J.Pengelly@soton.ac.uk

Received: 19 July 2022 Accepted: 21 July 2022