Classification of rare diseases: The case of the ultra- and hyper-rare

27 June 2023

Dr C. I. Edvard Smith, Dr Daniel W. Hagey and Dr Rula Zain, from the Karolinska Institutet in Sweden, note that there is an almost infinite number of unique diseases, contributing to the challenges with the classification of rare diseases

It was recently estimated that there are many more unique diseases than people living on this planet (Smith et al., 2022). Therefore, how do we conduct the identification and classification of ultra and hyper-rare diseases?

Diseases are either common or rare. Even if most people are affected by the common ones, it is among the rare ones that the most significant number of disease entities are found. This is because there are many mechanisms underlying disease. To provide two examples:

- Human beings have more than 20,000 genes; alterations in a single gene can cause several diseases. On top of that, defects in different genes can be combined in many ways that generate phenotypes and unique disease symptoms.
- There is an estimated up to one thousand billion (1,000,000,000,000) microorganism species on this planet (Locey & Lennon, 2016). Therefore, if only one microorganism in a million can cause disease in any potential context, this would still equal one million different infections.

The Online Mendelian Inheritance of Man

A subclass of rare diseases belongs to the hyper-rare group, which affects fewer than one in a hundred million people (Smith et al., 2022). This demonstrates the considerable challenge of the classification of rare diseases.

Many different tools have been developed for creating disease nomenclatures. Such initiatives are essential for cataloguing the growing number of diagnoses discovered by novel diagnostic techniques. For example, Victor A. McKusick pioneered what is today known as Online Mendelian Inheritance of Man (OMIM), a freely available, daily updated, comprehensive collection of human genes and genetic phenotypes.

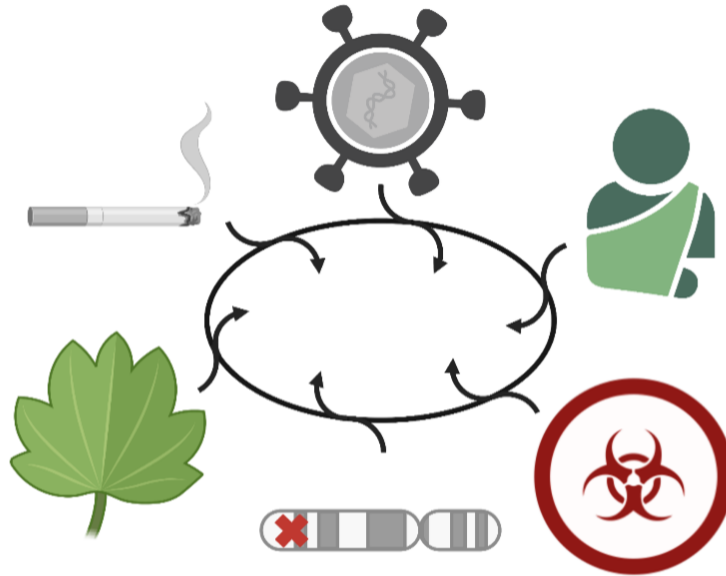It contains information on all known Mendelian disorders and over 16,000 genes.

Figure 1: The interaction between external elements such as the environment, lifestyle risk factors, infections, injuries, hazardous agents, and genetic factors causes disease

A special initiative for rare diseases (affecting fewer than 1/2000 individuals) is the ORPHAcode system, based on the Orphanet classification of rare diseases. This is a unique and multilingual standardised system aimed at providing specific terminology for rare diseases. As defined by Orphanet, each clinical entity is assigned a unique and time-stable ORPHAcode, around which the rest of the data present in the Orphanet database is structured. This coding system provides a common language across healthcare systems, registries, and research.

Biesecker et al., 2021 have formalised a versatile approach to assert that only by a combinatorial, or dyadic, approach taking both genotype and phenotype into account can a unitary, distinct genetic disorder be designated.

Furthermore, they argue that this approach to delineating and naming disorders reconciles the complexity of gene-to-phenotype relationships simply and clearly. They also make a case for GeneReviews®, which is defined as an international point-of-care resource for busy clinicians and provides clinically relevant and medically actionable information for inherited conditions in a standardised journal-style format. We have previously contributed such a review for a monogenic disease, the primary immunodeficiency, named X-linked agammaglobulinemia (Smith & Berglöf, 2016).

The International Classification of Diseases (ICD) has been the basis for comparable statistics on causes of mortality and morbidity between places for more than a century. The latest version of the ICD, ICD-11, was adopted in 2019 and came into effect on 1

January 2022. While useful, this taxonomy is much more limited regarding the classification of rare diseases and certainly not applicable when discussing ultra- and hyper-rare diseases.

Patients affected by concurrent diseases causing a distinct phenotype

While the existing nomenclatures are useful, it is still essential to develop classifications to include a simple system for the inclusion of patients affected by concurrent diseases causing a distinct phenotype, what we have named 'phenotype conversion' (Smith et al., 2022).

Here, it is essential to clearly distinguish between concurrent diseases causing a distinct illness, i.e., when their pathological mechanisms interact (phenotype conversion) or not (phenotype maintenance). For example, phenotype maintenance is when two diseases do not interact, such as when someone breaks a leg and has a mouth ulcer.

In contrast, patients inheriting two disease genes, each of which will cause a heart abnormality, may acquire a novel disease entity distinct from each of the individual illnesses. As such, when these layers of potential disease diversity are considered on top of one another, the number of definable, unique diseases becomes functionally infinite.

References

1. Biesecker LG, Adam MP, Alkuraya FS, et al. A dyadic approach to the delineation of diagnostic entities in clinical genomics. Am J Hum Genet. 2021 Jan 7;108(1):8- 15.
2. Locey KJ, Lennon JT. Scaling laws predict global microbial diversity. Proc Natl Acad Sci U S A. 2016 May 24;113(21):5970-5.
3. Smith CIE, Berglöf A. X-Linked Agammaglobulinemia. 2001 Apr 5 [updated 2016 Aug 4]. In: Adam MP, Mirzaa GM, Pagon RA, Wallace SE, Bean LJH, Gripp KW, Amemiya A, editors. GeneReviews® [Internet]. Seattle (WA): University of Washington, Seattle; 1993–2023.
4. Smith CIE, Bergman P, Hagey DW. Estimating the number of diseases – the concept of rare, ultra-rare, and hyper- rare. iScience. 2022 Jul 1;25(8):104698.