# RDmap: A Map for Exploring Rare Diseases

Jian Yang, Bs[#1,2]; Cong Dong, Ms[#1,2]; Huilong Duan, PhD[2]; Qiang Shu, MD[1]; Haomin Li, PhD[1*]

1. The Children's Hospital, Zhejiang University School of Medicine, National Clinical Research Center for Child Health, Zhejiang, China
2. The College of Biomedical Engineering and Instrument Science, Zhejiang University, Zhejiang, China

***Address correspondence to:** Haomin Li, The Children's Hospital, Zhejiang University School of Medicine, Binsheng Road 3333#, Hangzhou, China 310052, [hmli@zju.edu.cn], 086-13867445504;

# Jiang Yang and Cong Dong contributed equally to this study.

## Abstract

**Background:** The complexity of the phenotypic characteristics and molecular bases of many rare human genetic diseases make the diagnosis of such diseases a challenge for clinicians. A map for visualizing, locating and navigating rare diseases based on similarity will help clinicians and researchers understand and easily explore these diseases.

**Methods:** By defining the quantitative distance among phenotypes and pathogenic genes based on corresponding ontology systems, the distance matrix of rare diseases included in

Orphanet was calculated and mapped into Euclidean space. Enhanced by clustering classes and disease information, a rare disease map was developed based on ECharts.

**Results:** The rare disease map called RDmap was published at http://rdmap.nbscn.org. The phenotype-based map comprises 3,287 rare diseases and the gene-based map comprises 3,789 rare genetic diseases and they were bridged by 1,718 overlapping diseases. RDmap works similar to the widely used Google map and supports zooming and panning. The phenotype similarity base disease location function performed better than traditional keyword search in an in-silico evaluation and 20 published cases of rare diseases also demonstrated that RDmap can be used by clinicians to improve diagnosis.

**Conclusion:** RDmap is the first user-interactive map-style rare disease knowledgebase. It will help clinicians and researchers explore the increasing complicated rare genetic diseases.

**Keywords:** rare disease; phenotype; pathogenetic gene; disease map; clinical decision support

## Background

Rare diseases, most of which are caused by underlying genetic factors, often occur in infants or young children and affect the patients' whole life. Although rare, studies involving them

have revealed important insights about normal physiology that, in turn, have provided a better understanding of common disorders, universal mechanisms, critical pathways, and therapies that are useful to treat more than one disease. However, making the correct diagnosis for rare genetic diseases is extremely complicated and remains a challenge in both developed and developing countries. According to a survey from EURORDIS, it takes 5 to 30 years for a quarter of patients with rare genetic diseases from onset to diagnosis. During this period, the rate of first misdiagnosis is as high as 40%. These misdiagnoses would lead to a large number of invalid medical treatment or even unnecessary surgery, seriously endangering the health of the patients and wasting medical resources at the same time.

More than 7,000 known rare diseases have been identified, and more than 100 novel disease-gene associations have been identified per year since the introduction of next-generation sequencing technologies[1]. An important challenge of rare disease practice is to establish relationships among so many rare and diverse diseases from different levels. Accumulating studies have found that genetic diseases that are caused by similar molecules[2–4] can be diagnosed by similar phenotypic characteristics[5,6], and finally can be treated using similar drugs through corresponding targets[7–10]. Network-based medicine has emerged as a

complementary approach to identify disease-causing genes, genetic mediators, disruptions in the underlying cellular functions and for drug repositioning. Therefore, exploring the relationships of rare diseases can help to reveal the common attributes of similar rare genetic diseases. For example, the classification of rare diseases, phenotypic characteristics of diseases, and pathogenic genes of genetic diseases can improve the probability of discovering potential pathogenic mechanisms or drugs and, most importantly, can help with the clinical diagnosis of rare genetic disease and improve treatment plans.

In this study, we aimed to propose a method to construct two rare human disease maps based on the semantic similarities of both phenotypic characteristics and pathogenetic genes of rare diseases. Using advanced visualization technologies, the disease map can be used to reveal the complex relationships among different rare human genetic diseases and support the clinical diagnosis process.

## Methods

Methods to measure the distance between phenotypes

Human Phenotype Ontology (HPO)[11], proposed by Professor Robinson in 2008, provides a standardized vocabulary that covers all the common abnormal phenotypes in humans and has been recognized as a useful annotation of the phenotypic abnormalities of rare genetic diseases. As most of the modern ontology, HPO is structured as a directed acyclic graph (DAG), whereby the nodes of the DAG, also called HPO terms, represent abnormal phenotypic terms in humans. Additionally, these phenotypic terms are linked to their parents through subclass ("is a") relationships. Therefore, subclass phenotypic terms have more accurate definitions than parent phenotypic terms, and each phenotypic term may have multiple parents, reflecting various semantic types.

In this study, we measured the distance between different phenotype terms based on the hierarchical structure of HPO. For any two HPO terms, the distance can be quantified by the shortest distance between the corresponding two nodes of the HPO DAG:

$$Dist_p(p_1, p_2) = \frac{\min(d_1+d_2)}{d_{max}} \quad \text{(Formula 1)}$$

where $d_1$ and $d_2$ represent the distances between two child nodes and their common parent nodes of the HPO DAG, respectively. Additionally, $d_{max}$ represents the maximum distance between nodes in the HPO DAG.

Method to measure the distance between genes

The Gene Ontology (GO) knowledgebase is the world's largest source of information on the functions of genes[12]. Similar to the above process, GO can be used to compute the distance between genes. The GO describes genes from three different aspects: *molecular function*, *biological process* and *cell component*. Thus, the distance between any two genes from GO can be defined as the mean value of the shortest distance between gene nodes of the GO DAG from these three aspects:

$$Dist_g(g_1, g_2) = \frac{Dist_{cc}+Dist_{mf}+Dist_{bp}}{3} \quad \text{(Formula 2)}$$

where $Dist_{cc}$, $Dist_{mf}$ and $Dist_{bp}$ represent the distance between two genes calculated by Formula 1 from three different aspects.

Constructing the rare disease map based on Orphanet

Orphanet[13] was established in France in 1997 at the advent of the internet to gather scarce knowledge on rare diseases to improve the diagnosis, care and treatment of patients with rare diseases. Currently, Orphanet has become the reference source of information on rare diseases. In this study, 3,287 diseases with a known clinical phenotype and 3,789 diseases with known pathogenic genes, including 1,718 overlapping diseases, were used to construct the rare disease map.

Because many rare diseases in Orphanet were annotated using HPO terms and frequency, each of these diseases can be represented by a set of phenotypes with weight. The phenotypic distance between disease $d_1$ and disease $d_2$ can be measured by Formula 3:

$$Dist_{dp}(d_1, d_2) = \frac{1}{2}\left(\frac{\sum_{i=1}^{m} \min_{1 \le j \le n} (Dist_p(p_i, p_j)) * (w_i * w_j)}{m} + \frac{\sum_{i=1}^{n} \min_{1 \le j \le m} (Dist_p(p_i, p_j)) * (w_i * w_j)}{n}\right) \quad \text{(Formula 3)}$$

where m and n represent the number of phenotypes contained in disease $d_1$ and $d_2$, respectively, and $Dist(p_i, p_j)$ represents the distance between two phenotypes $p_i$ and $p_j$ as shown in Formula 1, and $w_i$ and $w_j$ are the frequencies of two phenotypes $p_i$ and $p_j$ in $d_1$ and $d_2$, respectively.

Similarly, we extracted disease gene relationships from the Orphanet knowledgebase. The genetic distance between diseases can then be transformed into the distance between genes:

$$Dist_{dg}(d_1, d_2) = \frac{1}{2}\left(\frac{\sum_{i=1}^{m} \min_{1 \le j \le n}(Dist_g(g_i, g_j))}{m} + \frac{\sum_{i=1}^{n} \min_{1 \le j \le m}(Dist_g(g_i, g_j))}{n}\right) \quad \text{(Formula 4)}$$

where m and n represent the number of genes identified as pathogenic genes in disease $d_1$ and $d_2$, respectively, and $Dist_g(g_i, g_j)$ represents the distance between two genes $g_i$ and $g_j$ as shown in Formula 2.

By calculating these distances among all rare diseases from Orphanet, we generated two distance matrices with the sizes of $3287 \times 3287$and $3789 \times 3789$ for phenotype and gene, respectively. We used multidimensional scaling[14] (*cmdscale* from the package *stats* in R) to convert the distance matrix into 2D points which can be visualized as a map.

To further explore the internal relationship between phenotypes and genes of rare genetic diseases, we divided the rare disease map into several disease clusters using the K-means clustering method. To determine the optimal k for disease clustering, a bootstrap approach implemented in the *clusterboot* function from the *fpc* package in R (version 3.4.0) was used.

Thus, we developed a web-based interactive rare disease map called RDMap (http://RDMap.nbscn.org) based on ECharts[15]. All the data precessions were under R, and the online web service was developed using Node.js and Python.

To evaluate the RDMap in clinical diagnosis, an in-silico test and a case report test were used. The targeted disease ranked in the recommend disease list was used to evaluate the performance of RDMap in clinical practice using only clinical information.

## Results

In this study, 3,287 diseases in Orphanet with a clinical phenotype and 3,789 diseases with known pathogenic genes in Orphanet were plotted into the Euclidean space, as shown in Fig 1. In total, 17 phenotype-based disease clusters and 18 gene-based disease clusters were generated and highlighted by different colors. The detailed information of disease clustering is explained in the supplemental material.

We published RDMap online (http://RDMap.nbscn.org) to help the user to explore rare disease relationships interactively. The map supports zooming and panning to find and locate special diseases like the widely used google map (Fig 2). It also supports a feature-based exploration such as one or more phenotypes will locate the most likely rare diseases on the

map and filter by the similarity score (Fig 2A). The detailed information about the disease will be shown when the disease is selected on the RDmap (Fig 2B). When a disease was selected on RDmap, user can jump between phenotype map and gene map through a toolbar button. This will help user to explore the interested diseases from different levels. The user guide for RDmap is provided in the supplemental material.

In the traditional knowledge base, the entries were indexed by keywords, and users are required to use the exact term used in the knowledge base to query the knowledge. However, obtaining the exact phenotype features in a particular patient clinically and matching it with the standard phenotype terms used to annotated diseases in knowledgebases remain challenges[16]. Because thousands of genetic diseases are known, their clinical presentations often overlap in patients and are typically abridged with respect to classical descriptions. The incompleteness, heterogeneity, imprecision, and noise in phenotype description sometimes miss the right diagnosis and even to wrong diagnoses.

To compare the performance of RDMap and direct simple matching of phenotype vectors (Jaccard matching), we designed an in silico evaluation test in which 1,000 rare genetic diseases from the Orphanet database are taken as the target diseases, and each disease is

represented as a set of four characteristic phenotypes with the highest frequency of the disease. Then, the adjacent node or parent node of the phenotype in the HPO DAG is defined as the imprecise phenotype of the target phenotype. The rank of the target diseases in the query results was used to evaluate the performance. The performance of the D-P method decreases significantly as the number of imprecise phenotypes increases (Fig 3). This finding also explains why it is very difficult to diagnose a rare genetic disease accurately in clinical practice using only clinical phenotypes. The RDmap-proposed methods Similarity (one-way distance calculation) and Similarity-Avg (average of two-way distance calculation) both have an obvious advantage over the Jaccard matching method, particularly regarding imprecision phenotypes. We also noticed that the one-way distance algorithm (Similarity) is more stable in the disease recommendation than the Similarity-Avg in this scenario. This one-way distance algorithm had been implemented in published RDmap.

To further evaluate the performance of RDmap in clinical practice, we collected 20 rare disease cases reported by the Orphanet Journal of Rare Diseases as test cases. The case presentations in publication were converted into HPO terms manually by one of the authors. The targeted diseases ranked in the similarity search results on RDmap are shown in Table 1

(the detailed information of each test case is shown in the supplemental material). RDmap worked very well in most cases with clear clinical phenotype descriptions and the targeted disease rank average 1.8 and median top 1 in all diseases. If you check the detail test case in supplemental material, some other similar diseases for some test cases remain under consideration for the clinician. These results were all highlighted on RDmap, and a quick check of a typical phenotype and its frequency in these potential diagnoses on RDmap will support clinicians to make a decision.

## Discussion

In this study, we constructed two maps of rare human genetic diseases based on phenotypic characteristics and genes and divided these genetic diseases into several disease clusters. Because diseases from the same cluster are related in phenotypic characteristics or gene functions, correlating clusters between two maps will be helpful to understand the physiological and pathological bases of related genetic diseases. Consistent with the results of Goh et al.[17], most of the diseases in the same phenotype-based cluster tend to have similar phenotypic characteristics. In total, 1,718 diseases overlapped in the two maps, and the relationship between 17 phenotype-based clusters and 18 gene-based clusters are shown in an

alluvial diagram in Fig 4 and supplemental material. The complicated branches among these clusters further confirmed the complicated relationships among the pathogenic genes and phenotypes of rare genetic diseases. Diseases with similar phenotypes may be divided into different gene-based disease clusters. However, diseases from the same gene-based clusters also present diverse phenotypes. However, we also noticed mainstreams among different clusters. RDmap also provide a button to jump from disease selected in phenotype-based map to same disease in gene-based map and vice versa. So, there are 1,718 bridges between two maps. These will inspire researchers to evaluate the inner relationships among pathogenic genes and phenotypes.

In recent years, to reveal the similar relationships between different human genetic diseases, many studies have used various ways to construct a human genetic disease network. For example, Goh et al. extracted known disease-gene associations from the OMIM database and constructed the human disease network[17]. The core idea of their method is that two diseases are related if they share at least one common gene. Lee et al. constructed a human disease network based on cell metabolism, and the core idea of this method is that two diseases are related if the related mutant enzyme catalyzes the adjacent metabolism

reaction[18]. Zhang et al. constructed a disease phenotype network using the similarity between phenotypes to obtain the gene function module[19]. However, RDmap shows a complicated disease relationship in a user-interactive map that we believe will be conducive to the discovery of potential relationships among pathogenic genes and phenotypic characteristics among many genetic diseases. The map-style visualization that reflects the distance of disease more intuitively will inspire investigators to understand the inner relationships among these diseases and their potential treatments and identify new pathogenic genes. Moreover, this tool can help the clinician or genetic counselor accurately diagnose rare genetic diseases effectively, especially when the clinical phenotypes are incomplete, imprecise or noise.

This study has some limitations. First, the two disease maps still not covering all rare genetic diseases. It based on a history version of Ophanet when this project started. Second, when a novel disease is enrolled in the map, all the disease maps and disease clustering need to be recalculated and updated. But we will annually update it based on the feedbacks from the community.

## Acknowledgements

Not applicable

## Authors' contributions

JY and CD developed the website and the initial draft of the manuscript. HD and QS provided data and developed the analysis protocol. HL concepted this project, analysis the data, design the website and revised the manuscript.

## Funding

This study was supported by the National Natural Science Foundation of China (81871456) and National Key R&D Program of China (2016YFC0901905).

## Availability of data and materials

All data generated or analyzed during this study are published online or included in this published article and its supplementary files.

## Ethics approval and consent to participate

Not applicable

## Consent for publication

Not applicable

## Competing interests

The authors declare no conflict of interest.

## References

1. Boycott KM, Rath A, Chong JX, Hartley T, Alkuraya FS, Baynam G, et al. International Cooperation to Enable the Diagnosis of All Rare Genetic Diseases. Am J Hum Genet. 2017;100:695–705.

2. Aerts S, Lambrechts D, Maity S, Van Loo P, Coessens B, De Smet F, et al. Gene prioritization through genomic data fusion. Nat Biotechnol. 2006;24:537–44.

3. Chavali S, Barrenas F, Kanduri K, Benson M. Network properties of human disease genes with pleiotropic effects. BMC Syst Biol. 2010;4:78.

4. Franke L, Van Bakel H, Fokkens L, De Jong ED, Egmont-Petersen M, Wijmenga C. Reconstruction of a functional human gene network, with an application for prioritizing positional candidate genes. Am J Hum Genet. 2006;78:1011–25.

5. Robinson P, Mundlos S. The Human Phenotype Ontology. Clin Genet [Internet]. 2010;77:525–34. Available from: http://doi.wiley.com/10.1111/j.1399-0004.2010.01436.x

6. Robinson PN, Köhler S, Bauer S, Seelow D, Horn D, Mundlos S. The Human Phenotype Ontology: A Tool for Annotating and Analyzing Human Hereditary Disease. Am J Hum Genet. 2008;83:610–5.

7. Yu L, Ma X, Zhang L, Zhang J, Gao L. Prediction of new drug indications based on clinical data and network modularity. Sci Rep. 2016;6:32530.

8. Gottlieb A, Stein GY, Ruppin E, Sharan R. PREDICT: a method for inferring novel drug indications with application to personalized medicine. Mol Syst Biol. 2011;7:496.

9. Luo H, Wang J, Li M, Luo J, Peng X, Wu F-X, et al. Drug repositioning based on comprehensive similarity measures and Bi-Random walk algorithm. Bioinformatics. 2016;32:2664–71.

10. Yu L, Wang B, Ma X, Gao L. The extraction of drug-disease correlations based on module distance in incomplete human interactome. BMC Syst Biol. 2016;10:111.

11. Köhler S, Doelken SC, Mungall CJ, Bauer S, Firth H V., Bailleul-Forestier I, et al. The Human Phenotype Ontology project: Linking molecular biology and disease through phenotype data. Nucleic Acids Res. 2014;

12. Carbon S, Douglass E, Dunn N, Good B, Harris NL, Lewis SE, et al. The Gene Ontology Resource: 20 years and still GOing strong. Nucleic Acids Res. 2019;47:D330–8.

13. Rath A, Olry A, Dhombres F, Brandt MM, Urbero B, Ayme S. Representation of rare diseases in health information systems: The orphanet approach to serve a wide range of end users. Hum Mutat. 2012;33:803–8.

14. Mead A. Review of the Development of Multidimensional Scaling Methods. Stat. 1992;41:27.

15. Li D, Mei H, Shen Y, Su S, Zhang W, Wang J, et al. ECharts: A declarative framework for rapid construction of web-based visualization. Vis Informatics. 2018;2:136–46.

16. Eldomery MK, Coban-Akdemir Z, Harel T, Rosenfeld JA, Gambin T, Stray-Pedersen A, et al. Lessons learned from additional research analyses of unsolved clinical exome cases. Genome Med. 2017;9:26.

17. Goh K-I, Cusick ME, Valle D, Childs B, Vidal M, Barabasi A-L. The human disease network. Proc Natl Acad Sci. 2007;104:8685–90.

18. Lee DS, Park J, Kay KA, Christakis NA, Oltvai ZN, Barabasi A-L. The implications of human metabolic network topology for disease comorbidity. Proc Natl Acad Sci. 2008;105:9880–5.

19. Zhang S-H, Wu C, Li X, Chen X, Jiang W, Gong B-S, et al. From phenotype to gene: Detecting disease-specific gene functional modules via a text-based human disease phenotype network construction. FEBS Lett. 2010;584:3635–43.

20. Al-Owain M, Mohamed S, Kaya N, Zagal A, Matthijs G, Jaeken J. A novel mutation and first report of dilated cardiomyopathy in ALG6-CDG (CDG-Ic): a case report. Orphanet J Rare Dis. 2010;5:7.

290 21. Böhm J, Yiş U, Ortaç R, Çakmakçı H, Kurul S, Dirik E, et al. Case report of intrafamilial
291 variability in autosomal recessive centronuclear myopathy associated to a novel BIN1 stop
292 mutation. Orphanet J Rare Dis. 2010;5:35.
293 22. Acién P, Galán F, Manchón I, Ruiz E, Acién M, Alcaraz LA. Hereditary renal adysplasia,
294 pulmonary hypoplasia and Mayer-Rokitansky-Küster-Hauser (MRKH) syndrome: a case
295 report. Orphanet J Rare Dis. 2010;5:6.
296 23. Mejia-Gaviria N, Gil-Pẽa H, Coto E, Pérez-Menéndez TM, Santos F. Genetic and clinical
297 peculiarities in a new family with hereditary hypophosphatemic rickets with hypercalciuria:
298 A case report. Orphanet J Rare Dis. 2010;
299 24. Joy T, Cao H, Black G, Malik R, Charlton-Menys V, Hegele RA, et al. Alstrom
300 syndrome (OMIM 203800): a case report and literature review. Orphanet J Rare Dis.
301 2007;2:49.
302 25. Zhu Y, Zou Y, Yu Q, Sun H, Mou S, Xu S, et al. Combined surgical-orthodontic
303 treatment of patients with cleidocranial dysplasia: case report and review of the literature.
304 Orphanet J Rare Dis. 2018;13:217.

26. Zamel R, Khan R, Pollex RL, Hegele RA. Abetalipoproteinemia: two case reports and literature review. Orphanet J Rare Dis. 2008;3:19.

27. Vroegindeweij LHP, Boon AJW, Wilson JHP, Langendonk JG. Effects of iron chelation therapy on the clinical course of aceruloplasminemia: an analysis of aggregated case reports. Orphanet J Rare Dis. 2020;15:105.

28. Zhou L, Ouyang R, Luo H, Ren S, Chen P, Peng Y, et al. Efficacy of sirolimus for the prevention of recurrent pneumothorax in patients with lymphangioleiomyomatosis: a case series. Orphanet J Rare Dis. 2018;13:168.

29. Dias RP, Buchanan CR, Thomas N, Lim S, Solanki G, Connor SEJ, et al. Os odontoideum in wolcott-rallison syndrome: A case series of 4 patients. Orphanet J Rare Dis. 2016;

30. Valayannopoulos V, Nicely H, Harmatz P, Turbeville S. Mucopolysaccharidosis VI. Orphanet J Rare Dis. 2010;5:5.

31. Biesecker LG. The Greig cephalopolysyndactyly syndrome. Orphanet J Rare Dis. 2008;

32. Germain DP. Fabry disease. Orphanet J Rare Dis. 2010;5:30.

320 33. Drera B, Ritelli M, Zoppi N, Wischmeijer A, Gnoli M, Fattori R, et al. Loeys-Dietz
321 syndrome type i and type II: Clinical findings and novel mutations in two Italian patients.
322 Orphanet J Rare Dis. 2009;
323 34. Reibel A, Manière M-C, Clauss F, Droz D, Alembik Y, Mornet E, et al. Orodental
324 phenotype and genotype findings in all subtypes of hypophosphatasia. Orphanet J Rare Dis.
325 2009;4:6.
326 35. Sarfati J, Bouvattier C, Bry-Gauillard H, Cartes A, Bouligand J, Young J. Kallmann
327 syndrome with FGFR1 and KAL1 mutations detected during fetal life. Orphanet J Rare Dis.
328 2015;10:71.
329 36. Weisfeld-Adams JD, Mehta L, Rucker JC, Dembitzer FR, Szporn A, Lublin FD, et al.
330 Atypical Chédiak-Higashi syndrome with attenuated phenotype: three adult siblings
331 homozygous for a novel LYST deletion and with neurodegenerative disease. Orphanet J Rare
332 Dis. 2013;8:46.
333 37. Mowat DR, Wilson MJ, Goossens M. Mowat-Wilson syndrome. J. Med. Genet. 2003.
334 38. Chrzanowska KH, Gregorek H, Dembowska-Bagińska B, Kalina MA, Digweed M.
335 Nijmegen breakage syndrome (NBS). Orphanet J Rare Dis. 2012;7:13.

336 39. Marshall BA, Paciorkowski AR, Hoekel J, Karzon R, Wasson J, Viehover A, et al.

337 Phenotypic characteristics of early Wolfram syndrome. Orphanet J Rare Dis. 2013;

338

339

**Figure Legends:**

**Fig. 1. Rare disease maps and clusters. The locations reflect the distance among diseases, and the size of the points reflect the prevalence of rare diseases** A. Rare disease map and clusters based on phenotype. B. Rare disease map and clusters based on gene.

**Fig. 2. Rare disease map zooming, panning, location, filtering and disease detail information.** A. The RDmap locates similar diseases based on phenotype search. The slider in the left bottom corner can control the similarity filtering threshold by the user. The prevalence options switch at the bottom right can filter the results based on prevalence. B. When a disease was selected on RDmap, its detail information will be shown like this.

**Fig. 3. In-slico test of RDMap.** Performance of RDMap under conditions with different numbers of imprecision phenotypes for the search

**Fig. 4. Alluvial diagram between 17 phenotype-based rare disease clusters and 18 gene-based rare disease clusters.**

**Table 1 Evaluation of RDmap based on cases from publications**

| Ref. | Disease (OMIM) | phenotypes | rank | Sim. score |
|---|---|---|---|---|
| Case1[20] | Congenital disorder of glycosylation (OMIM 603147) | HP:0025356 Psychomotor retardation/Psychomotor<br>HP:0001252 Muscular hypotonia<br>HP:0001644 Dilated cardiomyopathy<br>HP:0001250 Seizures<br>HP:0000486 Strabismus<br>HP:0006610 Wide intermammillary distance | 6 | 0.0625 |
| Case2[21] | Centronuclear myopathies (OMIM 255200) | HP:0009073 Progressive proximal muscle weakness<br>HP:0000297 Facial hypotonia<br>HP:0000508 Ptosis<br>HP:0000602 Ophthalmoplegia<br>HP:0001315 Reduced tendon reflexes<br>HP:0001256 Intellectual disability, mild | 4 | 0.0486 |
| Case3[22] | Mayer-Rokitansky-Küster-Hauser syndrome (OMIM 277000) | HP:0002089 Pulmonary hypoplasia<br>HP:0000122 Unilateral renal agenesis<br>HP:0000151 Aplasia of the uterus<br>HP:0008726 Hypoplasia of the vagina | 4 | 0.0937 |
| Case4[23] | Hereditary hypophosphatemic rickets with hypercalciuria (OMIM 241530) | HP:0002148 Hypophosphatemia<br>HP:0002150 Hypercalciuria | 1 | 0 |
| Case5[24] | Alström syndrome (OMIM 203800) | HP:0000662 Night blindness<br>HP:0000618 Blindness<br>HP:0012330 Pyelonephritis<br>HP:0000822 Hypertension<br>HP:0000819 Diabetes mellitus<br>HP:0000510 Retinitis pigmentosa<br>HP:0000518 Cataract | 1 | 0.535 |
| Case6[25] | Cleidocranial dysplasia (OMIM 119600) | HP:0000684 Delayed eruption of teeth<br>HP:0000164 Abnormality of the teeth<br>HP:0000316 Hypertelorism<br>HP:0011069 Increased number of teeth | 1 | 0 |
| Case7[26] | Abetalipoproteinemia (OMIM 200100) | HP:0002630 Fat malabsorption<br>HP:0001251 Ataxia<br>HP:0001324 Muscle weakness<br>HP:0001315 Reduced tendon reflexes | 4 | 0.0416 |
| Case8[27] | Aceruloplasminemia (OMIM 604290) | HP:0001935 Microcytic anemia<br>HP:0001260 Dysarthria<br>HP:0001288 Gait disturbance<br>HP:0000819 Diabetes mellitus<br>HP:0001903 Anemia<br>HP:0001300 Parkinsonism | 1 | 0.0416 |
| Case9[28] | Lymphangioleiomyomatosis (OMIM 606690) | HP:0100749 Chest pain<br>HP:0002094 Dyspnea<br>HP:0002107 Pneumothorax | 1 | 0 |
| Case10[29] | Wolcott–Rallison syndrome (OMIM 226980) | HP:0006554 Acute hepatic failure<br>HP:0001298 Encephalopathy<br>HP:0000083 Renal insufficiency<br>HP:0002654 Multiple epiphyseal dysplasia | 1 | 0.0208 |
| Case11[30] | Mucopolysaccharidosis type 6 (OMIM 253200) | HP:0000280 Coarse facial features<br>HP:0000470 Short neck<br>HP:0000158 Macroglossia<br>HP:0002808 Kyphosis<br>HP:0012471 Thick vermilion border | 1 | 0.0083 |

| | | | | |
|---|---|---|---|---|
| Case12[31] | Greig cephalopolysyndactyly syndrome (OMIM 175700) | HP:0000256 Macrocephaly<br>HP:0011304 Broad thumb<br>HP:0001159 Syndactyly<br>HP:0001162 Postaxial hand polydactyly<br>HP:0005873 Polysyndactyly of hallux | 1 | 0.016 |
| Case13[32] | Fabry disease (OMIM 301500) | Angiokeratoma (HP:0001014) | 1 | 0 |
| Case14[33] | Loeys-Dietz syndrome (OMIM 609192) | Camptodactyly of finger (HP:0100490)<br>Ulnar deviation of the hand or fingers of the hand (HP:0001193)<br>Bilateral talipes equinovarus (HP:0001776)<br>Blue sclerae (HP:0000592)<br>Microretrognathia (HP:0000308)<br>High palate (HP:0000218)<br>Bifid uvula (HP:0000193) | 1 | 0.0628 |
| Case15[34] | Hypophosphatasia (OMIM 146300) | Recurrent fractures (HP:0002757)<br>Craniosynostosis (HP:0001363)<br>Premature loss of teeth (HP:0006480) | 1 | 0.0138 |
| Case16[35] | Kallmann syndrome (OMIM 308700) | Oligomenorrhea (HP:0000876)<br>Breast hypoplasia (HP:0003187)<br>Anosmia (HP:0000458)<br>Hearing impairment (HP:0000365)<br>Reduced number of teeth (HP:0009804) | 1 | 0.0249 |
| Case17[36] | Chédiak–Higashi syndrome (OMIM 214500) | Lower limb muscle weakness (HP:0007340)<br>Dementia (HP:0000726)<br>Ataxia (HP:0001251)<br>Hypermetric saccades (HP:0007338)<br>Bradykinesia (HP:0002067)<br>Periodontitis (HP:0000704) | 4 | 0.0972 |
| Case18[37] | Mowat–Wilson syndrome (OMIM 235730) | Open mouth (HP:0000194)<br>Abnormality of the eyebrow (HP:0000534)<br>Frontal bossing (HP:0002007)<br>Deeply set eye (HP:0000490)<br>Wide nasal bridge (HP:0000431)<br>Strabismus (HP:0000486) | 1 | 0 |
| Case19[38] | Nijmegen breakage syndrome (OMIM 251260) | Microcephaly (HP:0000252)<br>Sloping forehead (HP:0000340)<br>Retrognathia (HP:0000278)<br>Macrotia (HP:0000400)<br>Bulbous nose (HP:0000414) | 1 | 0.0116 |
| Case20[39] | Wolfram syndrome (OMIM 222300) | Diabetes mellitus (HP:0000819)<br>Optic atrophy (HP:0000648)<br>Diabetes insipidus (HP:0000873)<br>Hearing impairment (HP:0000365)<br>Gastroesophageal reflux (HP:0002020) | 1 | 0.0333 |

358