Resolution of Disease Phenotypes Resulting from Multilocus Genomic Variation

Jennifer E. Posey, M.D., Ph.D., **Tamar Harel, M.D., Ph.D.**, **Pengfei Liu, Ph.D.**, **Jill A. Rosenfeld, M.S.**, **Regis A. James, Ph.D.**, **Zeynep H. Coban Akdemir, Ph.D.**, **Magdalena Walkiewicz, Ph.D.**, **Weimin Bi, Ph.D.**, **Rui Xiao, Ph.D.**, **Yan Ding, M.D.**, **Fan Xia, Ph.D.**, **Arthur L. Beaudet, M.D.**, **Donna M. Muzny, M.S.**, **Richard A. Gibbs, Ph.D.**, **Eric Boerwinkle, Ph.D.**, **Christine M. Eng, M.D.**, **V. Reid Sutton, M.D.**, **Chad A. Shaw, Ph.D.**, **Sharon E. Plon, M.D., Ph.D.**, **Yaping Yang, Ph.D.**, and **James R. Lupski, M.D., Ph.D., D.Sc.**

Departments of Molecular and Human Genetics (J.E.P., T.H., P.L., J.A.R., Z.H.C.A., M.W., W.B., R.X., F.X., A.L.B., D.M.M., R.A.G., C.M.E., V.R.S., C.A.S., S.E.P., Y.Y., J.R.L.) and Pediatrics (S.E.P., J.R.L.), Baylor Genetics (P.L., M.W., W.B., R.X., Y.D., F.X., R.A.G., C.M.E., Y.Y.), Program in Structural and Computational Biology and Molecular Biophysics (R.A.J.), and Human Genome Sequencing Center (D.M.M., R.A.G., E.B., S.E.P., J.R.L.), Baylor College of Medicine, the Human Genetics Center, University of Texas Health Science Center (E.B.), and the Department of Pediatrics (S.E.P., J.R.L.) and Texas Children's Cancer Center (S.E.P.), Texas Children's Hospital — all in Houston

Abstract

BACKGROUND—Whole-exome sequencing can provide insight into the relationship between observed clinical phenotypes and underlying genotypes.

METHODS—We conducted a retrospective analysis of data from a series of 7374 consecutive unrelated patients who had been referred to a clinical diagnostic laboratory for whole-exome sequencing; our goal was to determine the frequency and clinical characteristics of patients for whom more than one molecular diagnosis was reported. The phenotypic similarity between molecularly diagnosed pairs of diseases was calculated with the use of terms from the Human Phenotype Ontology.

RESULTS—A molecular diagnosis was rendered for 2076 of 7374 patients (28.2%); among these patients, 101 (4.9%) had diagnoses that involved two or more disease loci. We also analyzed parental samples, when available, and found that de novo variants accounted for 67.8% (61 of 90) of pathogenic variants in autosomal dominant disease genes and 51.7% (15 of 29) of pathogenic variants in X-linked disease genes; both variants were de novo in 44.7% (17 of 38) of patients with two monoallelic variants. Causal copy-number variants were found in 12 patients (11.9%) with multiple diagnoses. Phenotypic similarity scores were significantly lower among patients in whom the phenotype resulted from two distinct mendelian disorders that affected different organ systems

(50 patients) than among patients with disorders that had overlapping phenotypic features (30 patients) (median score, 0.21 vs. 0.36; P = 1.77×10^{-7}).

CONCLUSIONS—In our study, we found multiple molecular diagnoses in 4.9% of cases in which whole-exome sequencing was informative. Our results show that structured clinical ontologies can be used to determine the degree of overlap between two mendelian diseases in the same patient; the diseases can be distinct or overlapping. Distinct disease phenotypes affect different organ systems, whereas overlapping disease phenotypes are more likely to be caused by two genes encoding proteins that interact within the same pathway. (Funded by the National Institutes of Health and the Ting Tsung and Wei Fong Chao Foundation.)

Medical genetics focuses on the relationship between observed phenotypes and their underlying genotypes, modes of transmission, and risks of recurrence. Expected patterns of mendelian inheritance are often used to confirm the identification of disease genes, and deviations from mendelian expectations have led to the discovery of more complicated genetic underpinnings of disease (Fig. S1 in the Supplementary Appendix, available with the full text of this article at NEJM.org).[1–8] Multiple (or dual) molecular diagnoses involve more than one clinical diagnosis and more than one genetic locus (Fig. 1), each segregating independently.

Diagnostic whole-exome sequencing affords opportunities for providing insights into relationships between multilocus genomic variation and disease. In several studies of patient series, the occurrence of multiple molecular diagnoses in a single genome has been reported in 3.2 to 7.2% of cases in which a molecular diagnosis is made, but data on this phenomenon from large case series and on the associated clinical consequences are lacking (Table S1 in the Supplementary Appendix).[9–13] Here we describe a large-scale clinical analysis involving patients with multiple molecular diagnoses and an analysis of their phenotypes with the use of a structured phenotype ontology.

Methods

Patient Population

We performed a retrospective analysis involving 7374 unrelated patients who were referred to our diagnostic laboratory for proband-only or triobased whole-exome sequence analysis between October 2011 and April 2016. Whole-exome sequencing for cancer exome analysis was not included. Our laboratory is certified by the College of American Pathologists and is in compliance with the Clinical Laboratory Improvement Amendments. The reporting of deidentified demographic and molecular data was approved by the institutional review board at the Baylor College of Medicine. Of the 101 patients who received two or more genetic diagnoses, 26 have been reported previously.[9,10,12]

Whole-Exome Sequencing

Library construction, exome capture, next-generation sequencing, and data processing were performed as described previously.[9,14,15] Whole-exome sequencing included a coding single-nucleotide polymorphism (cSNP) array for quality control. Mitochondrial genome sequencing was performed for a subset of consecutive cases (4263 of 7374 [57.8%], October

2012 through December 2014). Variants described in Table S5 in the Supplementary Appendix that have not previously been reported have been submitted to the National Center for Biotechnology Information ClinVar database under accession numbers SCV000328705 to SCV000328861.

Statistical Analyses

A Poisson model and an alternative model of independently occurring multiple diagnoses were applied to analyze the proportions of patients with one to four molecular diagnoses in this cohort. This modeling was performed with the use of empirical data from the observed rate of molecular diagnoses in the study cohort, which represents a referral population, rather than a primarily healthy general population. The observed proportion of patients with at least one molecular diagnosis was 28.2% (2076 of 7374), which resulted in a total of 2182 independent molecular diagnoses in 7374 cases; this yielded a mean of 0.2959 diagnoses per patient, and this value was used as the Poisson rate parameter (Section S1 in the Supplementary Appendix). For the alternative independence model, the rate of singleton diagnoses was used, and powers of this rate were used to determine expected proportions of two, three, and four molecular diagnoses. To test the observed rate of multiple molecular diagnoses, we used the frequencies from the Poisson model and the alternative independence model to determine a null hypothesis for the expected number of patients with more than one molecular diagnosis. We tested the number of multiple diagnoses observed under both models against the null hypothesis, using a binomial test to determine the overall P values (Fig. S2 in the Supplementary Appendix). Details of the variant analyses, phenotype analyses, and statistical modeling of multiple diagnoses are provided in Sections S2, S3, and S4 in the Supplementary Appendix.

Results

Molecular Diagnoses

Among the 7374 sequential DNA samples submitted to our clinical laboratory for probandonly (7029 patients) or trio-based (345 patients) diagnostic whole-exome sequencing between October 2011 and April 2016, a molecular diagnosis involving a mendelian disease gene related to the clinical phenotype at the time of referral was reported for 2076 patients (28.2%). Two or more molecular diagnoses were reported for 101 patients (4.9%); in total, 2182 independent molecular diagnoses were reported for 7374 referred patients.

To investigate whether the proportion of multiple molecular diagnoses in this series is similar to that expected by chance alone in a referral population, we considered two models: a Poisson model, in which it was assumed that pathogenic variants arose independently at different loci within each patient's genome, and an alternative, independence model that used the observed rate of singleton diagnosis. Under both models, the observed proportion of patients with a diagnosis who had multiple molecular diagnoses (4.9%) was significantly lower than that expected among patients with a diagnosis in a referral population (Poisson model, 14.0%; independence model, 26.4%; P<0.0001 by one-sided binomial test for both models) (Fig. S2 in the Supplementary Appendix), which suggests that pathogenic variants

at multiple molecular loci are underascertained in a population of patients who are referred for diagnostic whole-exome sequencing or that they do not truly occur independently. Because these analyses are based on empirical observations from a referral population, they are limited by ascertainment bias.

Among the 101 patients with multiple diagnoses, 97 had two (dual) molecular diagnoses (Table 1), 3 had three molecular diagnoses, and 1 had four molecular diagnoses (Fig. 2A). Medically actionable secondary findings[16] contributed an additional two molecular diagnoses; however, because these were considered to be incidental, they were not analyzed further. Patient age and sex were not associated with the occurrence of multiple diagnoses. The specialty of the referring physician was medical genetics in 81.2% of cases (82 of 101), neurology in 12.9% (13 of 101), neurogenetics in 2.0% (2 of 101), allergy or immunology in 1.0% (1 of 101), and unknown for the remainder (Table S2 in the Supplementary Appendix). For 6 patients, a previous molecular diagnosis was known at the time of referral but was not believed to account for the disease phenotype in its entirety; for another 11 patients, a presumptive genetic diagnosis was available (Table S3 in the Supplementary Appendix). For 34 patients, a family history suggestive of an inherited genetic condition was reported.

Modes of Inheritance in Patients with Multiple Diagnoses

Variants in autosomal dominant disease genes were the most common pathogenic variants among patients with multiple diagnoses (112 of 207 diagnoses, 54.1%) (Fig. 2B, and Table S4 in the Supplementary Appendix). Pathogenic variants in X-linked disease genes were reported in 31 patients, including 3 patients who had two such variants. Among patients for whom parental samples were available, de novo variants accounted for 67.8% (61 of 90) of pathogenic variants in autosomal dominant disease genes and 51.7% (15 of 29) of pathogenic variants in X-linked disease genes (Fig. 2B and 2C, and Table S4 in the Supplementary Appendix); 10 female patients had a de novo pathogenic variant in X-linked disease genes (Table S5 in the Supplementary Appendix). Among patients with two molecular diagnoses, two pathogenic variants in autosomal dominant disease genes was the most common observed pattern (Table S6 in the Supplementary Appendix), and de novo pathogenic variants contributed to all combinations involving autosomal dominant or X-linked genes (Fig. 2C). Among patients in whom monoallelic variants at two loci (in combinations of autosomal dominant and X-linked genes) were found and for whom parental samples were available, 44.7% (17 of 38) had de novo variants at both disease loci (Fig. 2D). Maternal and paternal ages were provided for 95 and 91 patients, respectively. The mean paternal age did not differ significantly between patients with two de novo pathogenic variants and patients with no de novo pathogenic variants (35.5 and 32.8 years, respectively; P = 0.14 by two-tailed t-test) (Table S7 in the Supplementary Appendix); the parental origin for de novo variants was not determined.

Among the 42 patients for whom parental samples were available and who had pathogenic variants in one or more autosomal recessive disease genes (53 variants in total), 28 variants (52.8%), which were found in 20 patients, were homozygous variants that were confirmed to have been inherited from two heterozygous parents. A review of cSNP data for all 27 patients with at least one diagnosis that was associated with a homozygous variant revealed

22 patients with one or more regions of absence of heterozygosity larger than 10 Mb, totaling between 54 and 610 Mb per personal genome (Table S5 in the Supplementary Appendix). Consanguinity was reported in 15 of these 22 patients, and homozygous variants in two or more autosomal recessive disease genes were reported in 36.4% (8 of 22). The absence of heterozygosity involved multiple chromosomes in all patients, which argued against uniparental disomy. Only Patient 76 harbored 2 pathogenic variants within a single region of absence of heterozygosity. One patient (Patient 100) had homozygous variants in three autosomal recessive disease genes, each located in a different region of absence of heterozygosity. Although homozygous variants in autosomal recessive genes accounted for the majority (31 of 47, 66.0%) of diagnoses in these 22 patients who had genomic intervals of absence of heterozygosity that were larger than 10 Mb, we also found 6 de novo variants: 5 in autosomal dominant genes (in patients with a variant in an autosomal dominant gene plus a variant in an autosomal recessive gene) and 1 in an X-linked gene (in a patient with a variant in an X-linked gene plus a variant in an autosomal recessive gene), which showed that de novo variation can cause disease in patients with multiple regions of absence of heterozygosity.

A combination of copy-number variants (CNVs) and single-nucleotide variants (SNVs) contributed to multiple molecular diagnoses in 12 of 101 patients (11.9%). Three patients (Patients 47, 51, and 89) had homozygous intragenic deletions involving one to three exons (Table S5 in the Supplementary Appendix).

Phenotype Description

On the basis of clinical evaluations, blended phenotypes among patients with dual diagnoses could be divided into two major categories: distinct phenotypes, wherein individual phenotypic features were clearly attributable to only one of the two diagnoses, and overlapping phenotypes, wherein phenotypic features could be attributable to either one of the diagnoses (Fig. 3). We hypothesized that an objective, computational analysis of phenotypic similarity could be used to quantitatively differentiate distinct and overlapping phenotypes. We devised a phenotypic similarity score to objectively quantify the degree of overlap between two sets of disease phenotypes. The similarity score was calculated by the symmetric Resnik method (see Section S4 in the Supplementary Appendix) for each of 80 pairs of disease diagnoses for which both diseases in the Online Mendelian Inheritance in Man (OMIM) database (www.omim.org) had been mapped to the Human Phenotype Ontology (http://human-phenotype-ontology.github.io) (Fig. 4A and Table 1).[17–19] CNVs attributable to a single disease gene were included in this analysis. Disease pairs with the lowest scores (0.01) included transposition of the great arteries (OMIM 608808) and X-linked, syndromic, Turner-type mental retardation (OMIM 300706) (Patient 72), and the Coffin–Siris syndrome (OMIM 135900) and nonspherocytic hemolytic anemia (OMIM 300908) (Patient 61). In contrast, the disease pairs with the highest phenotypic similarity scores (>0.60) were the Coffin–Siris syndrome (OMIM 135900) and the Wiedemann–Steiner syndrome (OMIM 605130) (Patient 4) and two types of epileptic encephalopathy (OMIM 613720 and OMIM 614558) (Patient 27).

To assess whether an objective computational assessment of phenotypic similarity can closely model a subjective, human clinical assessment of phenotype, two physician scientists (the first two authors) independently assigned all patients with two molecular diagnoses to distinct or overlapping categories after a review of the phenotypic features that were provided in the OMIM database for each disease (Table 1). Patients were considered to have overlapping phenotypes if one or more phenotypic features were reported in OMIM as being associated with both molecular diagnoses; patients for whom no phenotypic features were shared between molecular diagnoses were categorized as having distinct phenotypes. The physician scientists were unaware of the results of the computational phenotypic similarity analyses described above. For 77 of 92 patients (83.7%) with dual diagnoses who were included in the analysis, the category assignments were concordant between the two physician scientists. Discrepancies were resolved through a joint review of the molecular diagnoses and their respective OMIM entries for each case. Next, phenotypic similarity scores were graphed according to clinical categorization made by the physician scientists; the mean (±SD) phenotype similarity score was 0.39±0.13 (SE, 0.024) for overlapping diagnoses, as compared with 0.20±0.12 (SE, 0.018) for distinct diagnoses (Fig. 4B). Phenotypic similarity scores for clinically categorized overlapping diagnoses were significantly greater than the scores for clinically categorized distinct diagnoses (median, 0.36 vs. 0.21; P = 1.77×10^{-7} by the Wilcoxon signed rank test) (Fig. 4C).

Protein–protein interaction databases and pathway databases were interrogated for physical interactions between the encoded proteins in a pair. No direct interactions between the proteins encoded by the disease loci were revealed. However, an extension of our analysis to include in silico predicted second-degree and third-degree physical interactors revealed nine dual diagnoses for which interactions were predicted (Fig. S3 in the Supplementary Appendix). Four of these pairs ranked in the top seven cases in terms of phenotypic similarity scores, a finding consistent with their proposed interaction (Fig. 4B), and the phenotypic features in these patients could be attributed to either disease diagnosis within the overlapping disease pair and may have been more severe than that observed with either molecular diagnosis alone (intellectual disability with neurodevelopmental delay in Patients 3 and 4, seizures in Patient 27, and skin photosensitivity in Patient 78). Patients 3 and 4 shared a molecular diagnosis of the Coffin–Siris syndrome (OMIM 135900) caused by pathogenic variants in *ARID1B*. The second diagnoses, the KBG syndrome (OMIM 148050, *ANKRD11*, Patient 3) and the Wiedemann–Steiner syndrome (OMIM 605130, *KMT2A*, Patient 4) share features of short stature, delayed bone age, intellectual disability, and developmental delay, resulting in shared phenotypes between these two patients. No primary interactors were observed among our patients with dual diagnoses; we speculate that such interactors are more likely to be represented by a digenic inheritance model (Fig. 4B).

Discussion

The clinical implementation of whole-exome sequencing as a molecular diagnostic assay allows interrogation of the interplay between pathogenic variants in multiple genes and the resulting complex spectrum of observed phenotypes within one patient. Our study is therefore limited by the depth and breadth of reported phenotypes of the patients whose exomes we analyzed. However, phenotype analysis with a structured phenotype ontology

and an objective computational analysis supports a framework in which dual molecular diagnoses lead to either distinct or overlapping categories of disease expression at extreme ends of phenotype similarity scores. This was supported by manual assessment of cases by two physician scientists, whose category assignments were 83.7% in concordance with one another; this illustrates the need for more objective phenotype analysis tools.

We found multiple molecular diagnoses in 4.9% of cases in which whole-exome sequencing was informative, a frequency similar to that previously reported[9–13] and lower than expected on the basis of a Poisson model and an alternative model of independently occurring molecular diagnoses; this suggests that patients with multiple diagnoses are underrecognized, that the pathogenic variants in each disease gene do not occur in an independent fashion, or that there is a synthetic lethal effect. These analyses are limited by the ascertainment bias that is inherent in the study of a referral population, as well as by the circumstances leading to inclusion in or exclusion from this population and the simplicity of the epidemiologic models of mutational events within a population. Nonetheless, these findings support the notion that a diagnostic evaluation is not necessarily complete with the identification of an initial molecular diagnosis and that genomewide analyses may reveal more than one mendelian disease that is relevant for a patient and the patient's family.

Pathogenic de novo variants were found in both autosomal dominant disease genes and X-linked disease genes and were reported for both molecular diagnoses in 17 patients. Despite the presumption of pathogenic variants in recessive disease genes in families with consanguinity, six diagnoses were associated with de novo variants in patients with documented absence of heterozygosity. Other investigators have described the occurrence of de novo variants in consanguineous populations.[20–22] These findings support the hypothesis that recently arisen, private variants play a substantial role in human disease.[23]

We found that 11.9% of patients (12 of 101) carried a pathogenic CNV, a frequency similar to the 9 to 11% observed for brain malformations or immunodeficiencies.[24,25] In this and other studies, selection bias may have contributed to an underestimation of CNV contribution to disease, because arrays are often clinically indicated and ordered before whole-exome sequencing is considered; thus, the diagnostic evaluation may end with discovery of a pathogenic CNV, which precludes the identification of a second mendelian disorder.

In contrast to digenic inheritance, which requires contributing pathogenic variation at two specific loci for the manifestation of a single disease,[2–4,26] dual molecular diagnoses represent an aggregation of independent diagnoses. The phenotypic complexity of genetic disease in persons with multiple molecular diagnoses may present a challenge to the physician. The blending of two distinct disease phenotypes in a single patient may suggest an apparently new clinical phenotype.[10] Alternatively, molecular diagnoses with two overlapping disease phenotypes may be incorrectly interpreted as the phenotypic expansion of a single disease.

Complementary to the idea of trait manifestations of more than one molecular diagnosis is the concept of mutational burden, in which variants at more than one locus associated with a

particular disease result in a modified (typically more severe) disease phenotype.[8] Further support for such a mutational burden hypothesis can be observed at a single locus for dosage-sensitive genes such as *PMP22*, which causes more severe polyneuropathy when in quadruplicate than when in triplicate.[27]

Our data indicate that bioinformatic tools and a structured ontology may be used to objectively assess complex phenotypes and that overlapping phenotypes may involve protein pairs that interact closely at the molecular level (for example, ARID1B and ANKRD11) or more distantly at the level of a functional unit or organ system, such as the eye (CRYGD and COL4A1) or brain (PLA2G6 and BCAP31). The mutational burden of the neuron as a functional unit has recently been shown to contribute to disease severity.[8] Using this same interaction analysis, we found in silico–predicted physical interactions between protein products of genes in patients with distinct phenotypes, although three of the disease pairs had phenotypic similarity scores above the mean similarity score observed for distinct cases.

Our data challenge the notion that a diagnostic investigation is necessarily complete after a single genetic diagnosis has been obtained. The phenotype of a patient with two genetic diagnoses may be influenced by the extent to which the phenotype associated with each individual disease overlaps that of the other. Our bioinformatic analysis of phenotypes focused on patients for whom only two molecular diagnoses were reported. However, as additional disease genes are defined and technologies for rare variant detection continue to improve, more cases of multiple molecular diagnoses are likely to be identified and, in turn, to improve knowledge about the effect of multiple rare variants at more than one locus on biology and human disease.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Knudson AG Jr. Mutation and cancer: statistical study of retinoblastoma. Proc Natl Acad Sci U S A. 1971; 68:820–823. [PubMed: 5279523]

2. Kajiwara K, Berson EL, Dryja TP. Digenic retinitis pigmentosa due to mutations at the unlinked peripherin/RDS and ROM1 loci. Science. 1994; 264:1604–1608. [PubMed: 8202715]

3. Lemmers RJ, Tawil R, Petek LM, et al. Digenic inheritance of an *SMCHD1* mutation and an FSHD-permissive D4Z4 allele causes facioscapulohumeral muscular dystrophy type 2. Nat Genet. 2012; 44:1370–1374. [PubMed: 23143600]

4. Timberlake AT, Choi J, Zaidi S, et al. Two locus inheritance of non-syndromic midline craniosynostosis via rare *SMAD6* and common *BMP2* alleles. Elife. 2016; 5:e20125. [PubMed: 27606499]

5. Katsanis N, Ansley SJ, Badano JL, et al. Triallelic inheritance in Bardet-Biedl syndrome, a Mendelian recessive disorder. Science. 2001; 293:2256–2259. [PubMed: 11567139]

6. Bougeard G, Baert-Desurmont S, Tournier I, et al. Impact of the *MDM2* SNP309 and *p53* Arg72Pro polymorphism on age of tumour onset in Li-Fraumeni syndrome. J Med Genet. 2006; 43:531–533. [PubMed: 16258005]

7. Emond MJ, Louie T, Emerson J, et al. Exome sequencing of extreme phenotypes identifies *DCTN4* as a modifier of chronic *Pseudomonas aeruginosa* infection in cystic fibrosis. Nat Genet. 2012; 44:886–889. [PubMed: 22772370]

8. Gonzaga-Jauregui C, Harel T, Gambin T, et al. Exome sequence analysis suggests that genetic burden contributes to phenotypic variability and complex neuropathy. Cell Rep. 2015; 12:1169–1183. [PubMed: 26257172]

9. Yang Y, Muzny DM, Reid JG, et al. Clinical whole-exome sequencing for the diagnosis of mendelian disorders. N Engl J Med. 2013; 369:1502–1511. [PubMed: 24088041]

10. Yang Y, Muzny DM, Xia F, et al. Molecular findings among patients referred for clinical whole-exome sequencing. JAMA. 2014; 312:1870–1879. [PubMed: 25326635]

11. Farwell KD, Shahmirzadi L, El-Khechen D, et al. Enhanced utility of family-centered diagnostic exome sequencing with inheritance model-based analysis: results from 500 unselected families with undiagnosed genetic conditions. Genet Med. 2015; 17:578–586. [PubMed: 25356970]

12. Posey JE, Rosenfeld JA, James RA, et al. Molecular diagnostic experience of whole-exome sequencing in adult patients. Genet Med. 2016; 18:678–685. [PubMed: 26633545]

13. Retterer K, Juusola J, Cho MT, et al. Clinical application of whole-exome sequencing across clinical indications. Genet Med. 2016; 18:696–704. [PubMed: 26633542]

14. Bainbridge MN, Wang M, Wu Y, et al. Targeted enrichment beyond the consensus coding DNA sequence exome reveals exons with higher variant densities. Genome Biol. 2011; 12:R68. [PubMed: 21787409]

15. Lupski JR, Gonzaga-Jauregui C, Yang Y, et al. Exome sequencing resolves apparent incidental findings and reveals further complexity of *SH3TC2* variant alleles causing Charcot-Marie-Tooth neuropathy. Genome Med. 2013; 5:57. [PubMed: 23806086]

16. Green RC, Berg JS, Grody WW, et al. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. Genet Med. 2013; 15:565–574. [PubMed: 23788249]

17. Köhler S, Doelken SC, Mungall CJ, et al. The Human Phenotype Ontology project: linking molecular biology and disease through phenotype data. Nucleic Acids Res. 2014; 42:D966–D974. [PubMed: 24217912]

18. Robinson PN, Köhler S, Bauer S, See-low D, Horn D, Mundlos S. The Human Phenotype Ontology: a tool for annotating and analyzing human hereditary disease. Am J Hum Genet. 2008; 83:610–615. [PubMed: 18950739]

19. James RA, Campbell IM, Chen ES, et al. A visual and curatorial approach to clinical variant prioritization and disease gene discovery in genome-wide diagnostics. Genome Med. 2016; 8:13. [PubMed: 26838676]

20. Ratbi I, Elalaoui CS, Dastot-Le MF, Goossens M, Giurgea I, Sefiani A. Mowat-Wilson syndrome in a Moroccan consanguineous family. Indian J Hum Genet. 2007; 13:122–124. [PubMed: 21957361]

21. Fahiminiya S, Almuriekhi M, Nawaz Z, et al. Whole exome sequencing unravels disease-causing genes in consanguineous families in Qatar. Clin Genet. 2014; 86:134–141. [PubMed: 24102521]

22. Al-Qattan SM, Wakil SM, Anazi S, et al. The clinical utility of molecular karyotyping for neurocognitive phenotypes in a consanguineous population. Genet Med. 2015; 17:719–725. [PubMed: 25503496]

23. Lupski JR, Belmont JW, Boerwinkle E, Gibbs RA. Clan genomics and the complex architecture of human disease. Cell. 2011; 147:32–43. [PubMed: 21962505]

24. Karaca E, Harel T, Pehlivan D, et al. Genes that affect brain structure and function identified by rare variant analyses of mendelian neurologic disease. Neuron. 2015; 88:499–513. [PubMed: 26539891]

25. Stray-Pedersen A, Sorte HS, Samarakoon P, et al. Primary immunodeficiency diseases: genomic approaches delineate heterogeneous Mendelian disorders. J Allergy Clin Immunol. 2016 Jul 16. (Epub ahead of print).

26. Goldberg AF, Molday RS. Defective subunit assembly underlies a digenic form of retinitis pigmentosa linked to mutations in peripherin/rds and rom-1. Proc Natl Acad Sci U S A. 1996; 93:13726–13730. [PubMed: 8943002]

27. Liu P, Gelowani V, Zhang F, et al. Mechanism, prevalence, and more severe neuropathy phenotype of the Charcot-Marie-Tooth type 1A triplication. Am J Hum Genet. 2014; 94:462–469. [PubMed: 24530202]
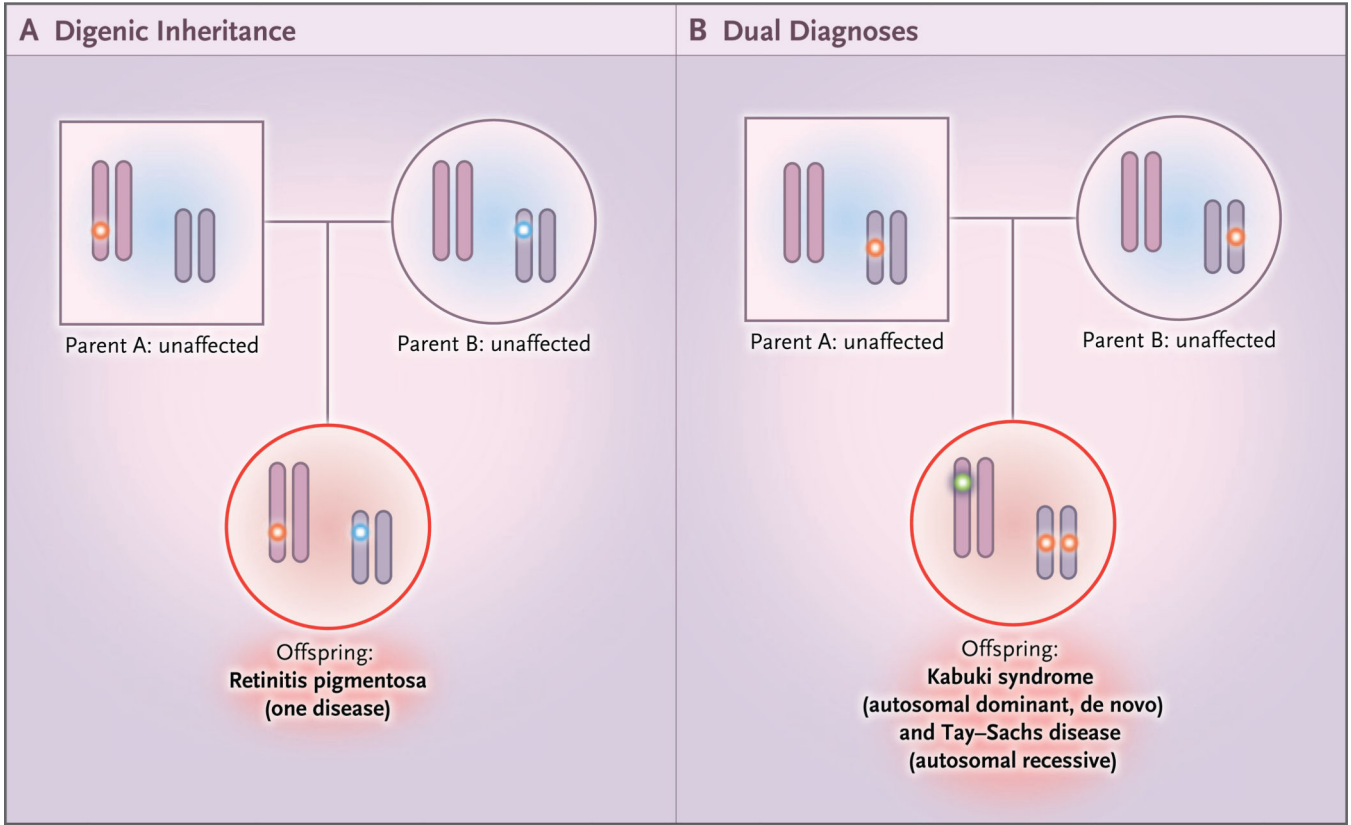
Figure 1. Models of Nonmendelian Inheritance

Panel A shows digenic inheritance of pathogenic alleles at different loci from each parent, resulting in an affected offspring. Panel B shows dual molecular diagnoses — in this case, one homozygous pathogenic variant in a recessive disease gene inherited from each carrier parent and one de novo variant that resulted in a second, independent autosomal dominant disorder. Many combinations are possible: autosomal dominant plus autosomal dominant, autosomal dominant plus autosomal recessive, autosomal dominant plus X-linked, autosomal recessive plus autosomal recessive, autosomal recessive plus X-linked, and X-linked plus X-linked.
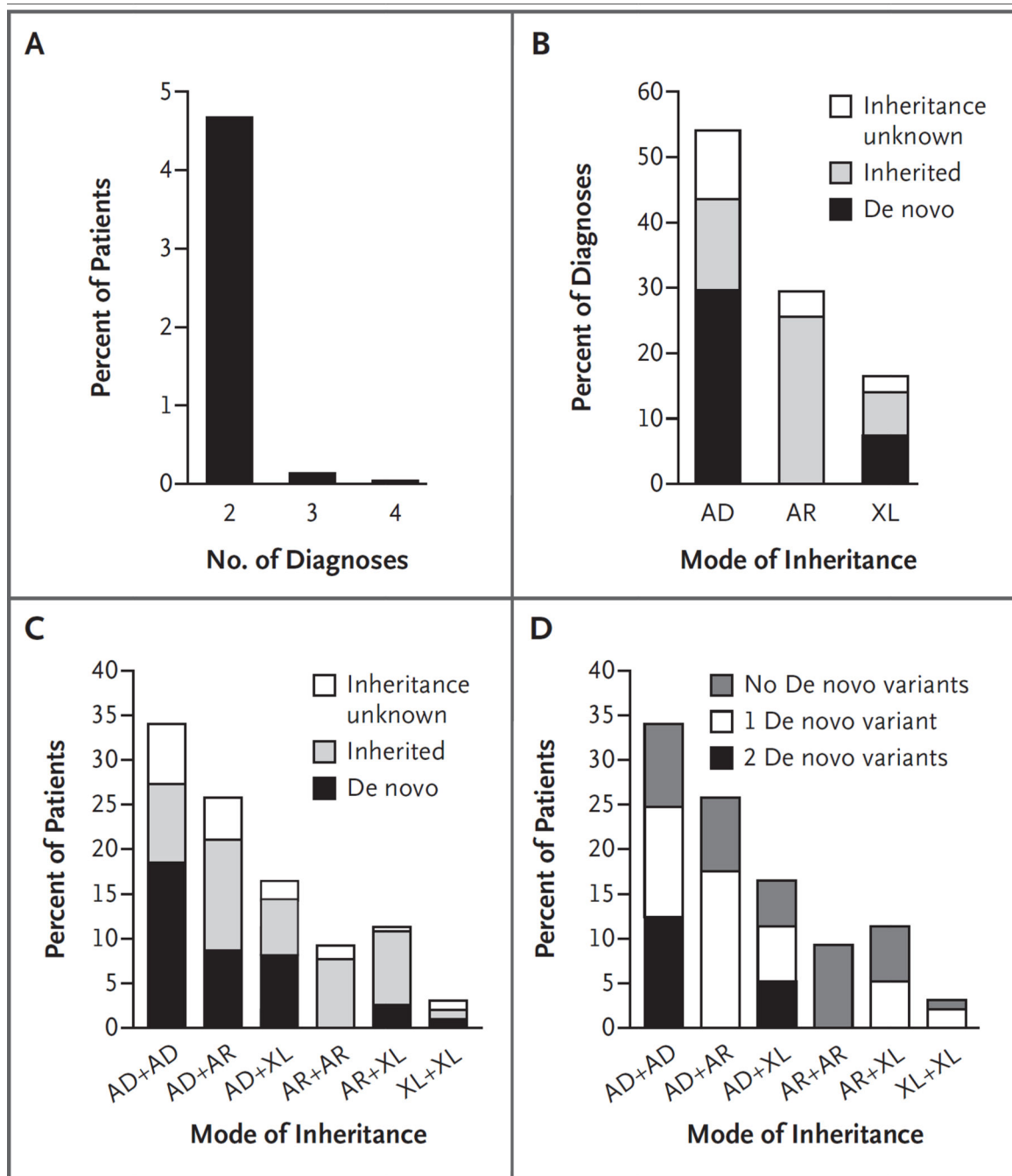
Figure 2. Frequencies and Modes of Inheritance in Multiple Molecular Diagnoses
Panel A shows the percentage of all patients with two or more molecular diagnoses. Panel B shows the percentage of all diagnoses in patients with two or more molecular diagnoses, according to mode of inheritance and de novo status. AD denotes autosomal dominant, AR autosomal recessive, and XL X-linked. Panel C shows the percentage of patients with two molecular diagnoses represented by each combination of modes of inheritance. The contribution of de novo versus inherited variants is represented within each bar as percentages of diagnoses (not patients). Panel D shows the percentage of patients with two
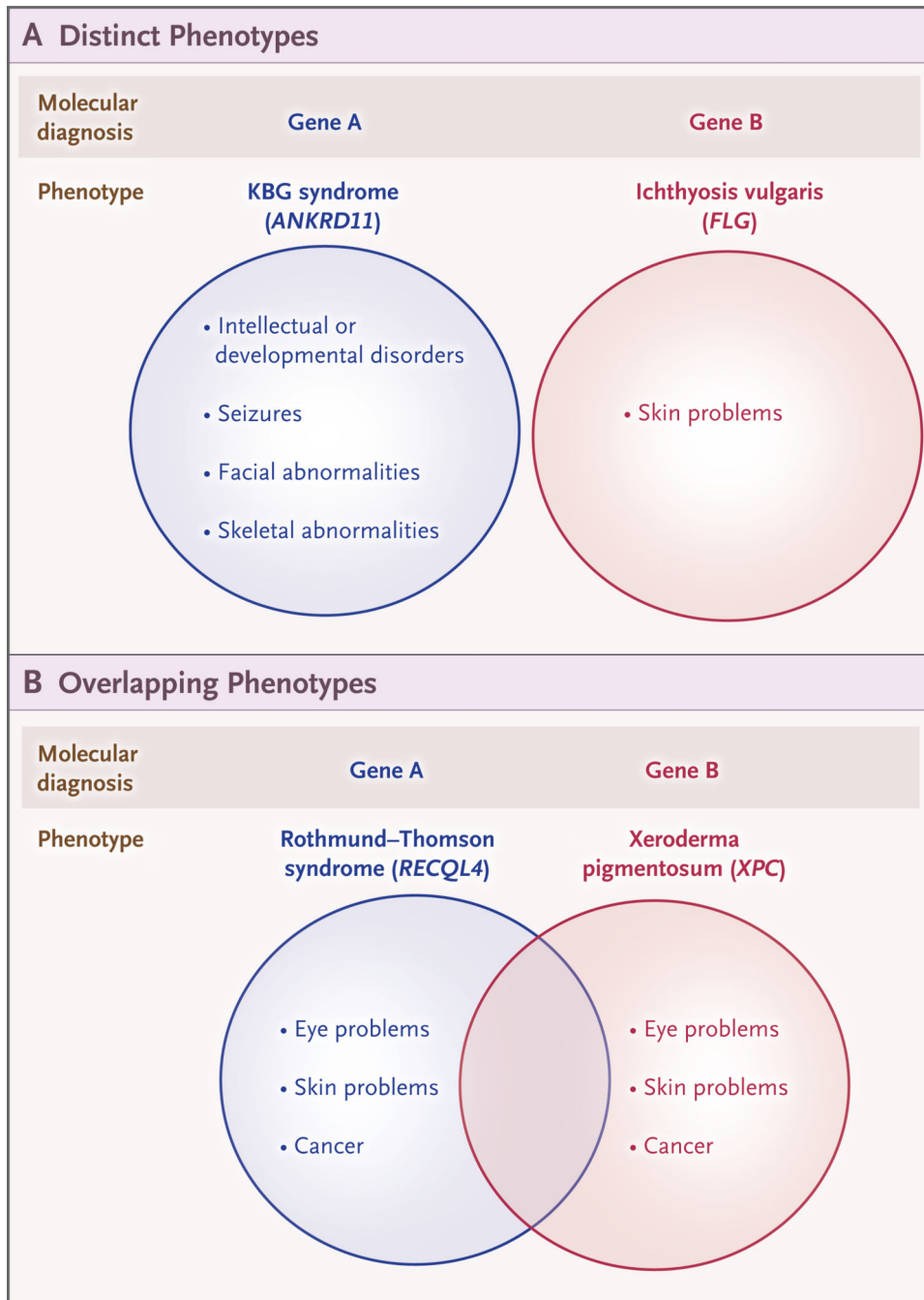
molecular diagnoses represented by each combination of modes of inheritance. The percentage of patients with one or two de novo variants contributing to diagnoses is indicated within each bar.

A Distinct Phenotypes

Molecular diagnosis	Gene A	Gene B
Phenotype	KBG syndrome (ANKRD11)	Ichthyosis vulgaris (FLG)

- Intellectual or developmental disorders
- Seizures
- Facial abnormalities
- Skeletal abnormalities

- Skin problems

B Overlapping Phenotypes

Molecular diagnosis	Gene A	Gene B
Phenotype	Rothmund–Thomson syndrome (RECQL4)	Xeroderma pigmentosum (XPC)

- Eye problems
- Skin problems
- Cancer

- Eye problems
- Skin problems
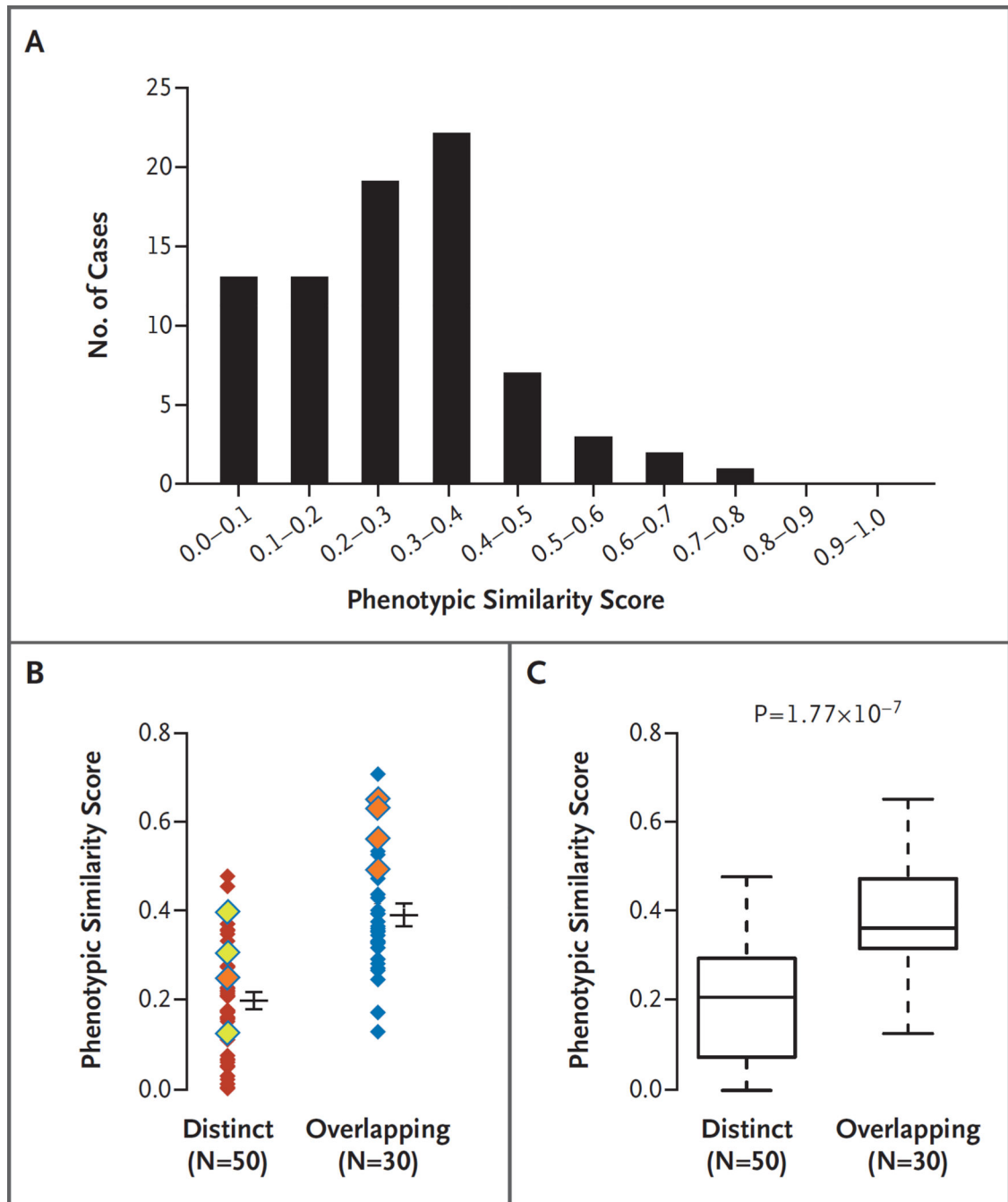- Cancer

Figure 3. Blended Phenotypes among Patients with Dual Diagnoses
Dual molecular diagnoses may result from variants at two loci (Gene A and Gene B) and can result in distinct clinical features (Panel A) or similar phenotypic features (Panel B).

Figure 4. Phenotypic Similarity of 80 Disease Pairs with Blended Phenotypes Categorized as Overlapping or Distinct

Panel A shows the distribution of cases plotted across the range of phenotypic similarity scores. Panel B shows the relative disease-pair similarity scores calculated by the symmetric Resnik method with the use of Human Phenotype Ontology (HPO) terms assigned to disease identification numbers in the Online Mendelian Inheritance in Man database. Each disease pair is represented by a single diamond; cases classified by two physician scientists as distinct or overlapping are indicated. Increasing phenotypic similarity (determined bioinformatically) corresponds to higher similarity scores. Eight pairs of diseases associated

with proteins that have physical or pathway interactions with one another are indicated by yellow diamonds (secondary interaction) or orange diamonds (tertiary interaction); the ninth patient (Patient 12) is not shown here, because neither disease in the pair was mapped to the HPO. Horizontal bars indicate means, and I bars indicate standard errors. Panel C shows box plots of the degree of phenotypic similarity for dual diagnoses that were clinically assigned to distinct or overlapping phenotype categories. The lower and upper limits of the box indicate the first and third quartiles, the bar inside the box the median, and the T bars the minimum and maximum or 1.5 times the interquartile range (whichever value was smaller).

Table 1

Molecular Diagnoses According to Expected Mode of Inheritance for 97 Patients with Dual Molecular Diagnoses.[*]

Patient No.	Genes with Pathogenic Variants		Modes of Inheritance (Gene A + Gene B)	Class	Similarity Score[†]
	Gene A	Gene B			
1	ADAR[‡]	APOB	AD + AD	Distinct	0.30
2[§]	ANKRD11[‡]	FLG	AD + AD	Distinct	0.14
3[§]	ANKRD11[‡]	ARID1B[‡]	AD + AD	Overlapping	0.49
4	ARID1B[‡]	KMT2A[‡]	AD + AD	Overlapping	0.65
5[§]	ASXL3	ENG	AD + AD	Distinct	NA
6	CACNA1A[‡]	SLC26A1[‡]	AD + AD	Overlapping	NA
7[§]	CHD2[‡]	PRRT2	AD + AD	Overlapping	0.43
8	CHD8	COL5A1	AD + AD	Distinct	NA
9	COL4A1[‡]	CRYGD	AD + AD	Overlapping	0.71
10[§]	CREBBP[‡]	PRICKLE2[‡]	AD + AD	Overlapping	0.36
11	DNM1[‡]	PTEN[‡]	AD + AD	Overlapping	NA
12	FBN1	MYO1F[‡]	AD + AD	Distinct	NA
13[§]	FLG	MEF2C[‡]	AD + AD	Distinct	0.08
14	FLG	PACS1[‡]	AD + AD	Distinct	NA
15	GDF6	SOX10[‡]	AD + AD	Overlapping	0.24
16[§]	GLI2	IRF6	AD + AD	Overlapping	0.33

Patient No.	Genes with Pathogenic Variants		Modes of Inheritance (Gene A + Gene B)	Class	Similarity Score[†]
	Gene A	Gene B			
17	*GLI2*[‡]	*SCN2A*[‡]	AD + AD	Overlapping	0.29
18	*HBB*	*KANSL1*	AD + AD	Distinct	0.37
19	*PTPN11*[‡]	*SHH*[‡]	AD + AD	Distinct	0.22
20[§]	*SCN1A*[‡]	*SMARCA2*[‡]	AD + AD	Overlapping	0.28
21	*KAT6A*	16p11.2 del	AD + AD	NA	NA
22	*NOTCH1*	*TTN*	AD + AD	Distinct	0.45
23	*KCNQ2*[‡]	*PRRT2*	AD + AD	Overlapping	0.47
24	*ANKRD11*	*SLC6A1*	AD + AD	Overlapping	NA
25	*PTCH1*[‡]	*TCF12*	AD + AD	Distinct	0.48
26	*COL11A1*	*KRIT1*	AD + AD	Distinct	0.16
27	*KCNQ2*[‡]	*SCN8A*	AD + AD	Overlapping	0.63
28	*NF1*[‡]	*SOX9*	AD + AD	Distinct	0.35
29	*MYH2*[‡]	*SMC1A*	AD + AD	Distinct	0.02
30	*SCN1A*[‡]	16p13.11 del	AD + AD	NA	NA
31	*CTNNB1*[‡]	1q21.1q21.2 del	AD + AD	NA	NA
32	*SOX11*[‡]	17q11.2 dup	AD + AD	NA	NA
33[§]	*SETBP1*[‡]	*CLCN1*	AD + AD	Distinct	0.23
34[§]	*KCNT1*[‡]	*TTN*	AD + AR	Distinct	0.18

Patient No.	Genes with Pathogenic Variants		Modes of Inheritance (Gene A + Gene B)	Class	Similarity Score[†]
	Gene A	Gene B			
35[§]	*ABCC9*	*RAPSN*	AD + AR	Distinct	0.01
36	*ACTG1[‡]*	*WFS1*	AD + AR	Overlapping	0.35
37[§]	*DES*	*CLCN1*	AD + AR	Overlapping	0.33
38	*GATAD2B[‡]*	*WWOX*	AD + AR	Overlapping	0.37
39	*GNAO1[‡]*	*ACADM*	AD + AR	Distinct	0.21
40	*HBB*	*TJP2*	AD + AR	Distinct	0.27
41[§]	*KIF5C[‡]*	*NRXN1*	AD + AR	Overlapping	0.17
42	*KMT2D[‡]*	*HEXA*	AD + AR	Overlapping	0.27
43[§]	*NF1[‡]*	*MEGF8*	AD + AR	Distinct	0.26
44[§]	*NF1[‡]*	*GALNT3*	AD + AR	Distinct	0.17
45	*PUF60[‡]*	*LOXHD1*	AD + AR	Distinct	0.05
46	*SCN8A[‡]*	*MAN2B1*	AD + AR	Overlapping	0.32
47	*SLC2A9*	*ETHE1*	AD + AR	Distinct	0.28
48	*SPRED1[‡]*	*MEGF10*	AD + AR	Distinct	0.30
49[§]	*SYNGAP1[‡]*	*MTFMT*	AD + AR	Overlapping	0.39
50	*TGFB2[‡]*	*TYR1*	AD + AR	Distinct	0.17
51	*GFAP[‡]*	*MPZ*	AD + AR	Distinct	0.35
52	*ARHGEF1*	*ECEL1*	AD + AR	Distinct	0.01

Patient No.	Genes with Pathogenic Variants		Modes of Inheritance (Gene A + Gene B)	Class	Similarity Score[†]
	Gene A	Gene B			
53	*FLG*	*TECPR2*	AD + AR	Distinct	0.15
54	*TUBB3[‡]*	*ISCA2*	AD + AR	Overlapping	NA
55	*NLRC4[‡]*	*SERPINA1*	AD + AR	Distinct	0.17
56	*CACNB4*	*TANGO2*	AD + AR	Overlapping	NA
57	*CPOX*	*ABCA1*	AD + AR	Overlapping	0.33
58	*GLI2[‡]*	*BLM*	AD + AR	Distinct	0.36
59	*CHD8[‡]*	*BRWD3[‡]*	AD + XL	Overlapping	NA
60[§]	*ARID1B[‡]*	*GRIA3[‡]*	AD + XL	Overlapping	0.39
61	*ARID1B[‡]*	*G6PD*	AD + XL	Distinct	0
62	*CHD7[‡]*	*DMD*	AD + XL	Distinct	0.27
63	*COL9A3*	*PLP1*	AD + XL	Distinct	0.12
64	*DNM1L[‡]*	*PDHA1[‡]*	AD + XL	Overlapping	0.37
65[§]	*EFHC1*	*SMC1A[‡]*	AD + XL	Overlapping	0.13
66	*GLI2*	*KDM5C[‡]*	AD + XL	Overlapping	0.44
67	*GLI3*	*LAS1L[‡]*	AD + XL	Distinct	NA
68	*GRIN2B[‡]*	*SHOXY*	AD + XL	Distinct	0.05
69	*PAFAH1B1[‡]*	*FGD1[‡]*	AD + XL	Distinct	0.16
70	*SCN1A*	*PDHA1*	AD + XL	Overlapping	0.35

Patient No.	Genes with Pathogenic Variants		Modes of Inheritance (Gene A + Gene B)	Class	Similarity Score[†]
	Gene A	Gene B			
71	*KIF5C*	*DMD*	AD + XL	Distinct	0.25
72	*MED13L[‡]*	*HUWE1[‡]*	AD + XL	Distinct	0.001
73	*COL7A1*	*DDX3X*	AD + XL	Distinct	NA
74	*COL9A1*	*ATRX*	AD + XL	Distinct	0.31
75	*AGL*	*PCCA*	AR + AR	Distinct	0.37
76	*HEXB*	*MCCC2*	AR + AR	Distinct	0.22
77	*MTPAP*	*NPC1*	AR + AR	Overlapping	0.40
78[§]	*RECQL4*	*XPC*	AR + AR	Overlapping	0.56
79	*TPO*	*MMP2*	AR + AR	Distinct	0.03
80	*AGL*	*LPAR6*	AR + AR	Distinct	0.11
81[§]	*PAPSS2*	*TRDN*	AR + AR	Distinct	0.06
82	*OTOF*	*SLC12A6*	AR + AR	Distinct	0.31
83	*AGPAT2*	*OBSL1*	AR + AR	Distinct	0.21
84[§]	*BBS10*	*PDHA1[‡]*	AR + XL	Distinct	0.28
85	*F7*	*MECP2[‡]*	AR + XL	Distinct	0.21
86	*FANCG*	*G6PD*	AR + XL	Distinct	0.40
87	*TCF12*	*SLC35A2*	AR + XL	Distinct	0.33
88[§]	*TREX1*	*PHEX*	AR + XL	Distinct	0.16
89	*PLA2G6*	*BCAP31*	AR + XL	Overlapping	0.53

Patient No.	Genes with Pathogenic Variants		Modes of Inheritance (Gene A + Gene B)	Class	Similarity Score[†]
	Gene A	Gene B			
90[§]	*CFTR*	*SMC1A[‡]*	AR + XL	Distinct	0.07
91	*SLC45A2*	*AVPR2*	AR + XL	Distinct	0.06
92	*ALG6*	*SHOX[‡]*	AR + XL	Distinct	0.03
93	*IRX5*	*HDAC8[‡]*	AR + XL	Overlapping	0.53
94	*BCS1L*	*NLGN4X*	AR + XL	Overlapping	0.27
95	*KIAA2022*	*PHF8*	XL + XL	Overlapping	0.34
96	*SMC1A*	*DMD*	XL + XL	Distinct	0.21
97	*KIAA2022*	Xp22.31 del	XL + XL	NA	NA

[*] AD denotes autosomal dominant, AR autosomal recessive, NA not available, and XL X-linked.

[†] The similarity score was calculated by the symmetric Resnik method (see Section S4 in the Supplementary Appendix), with higher scores indicating greater phenotypic similarity.

[‡] The variant in this gene was a de novo variant.

[§] Data for this patient have been reported previously.