Rare diseases: human genome research is coming home

Hans-Hilger Ropers[1] and Clara D. van Karnebeek[2]

[1]Max Planck Institute for Molecular Genetics, Berlin D-14195, and Institute of Human Genetics, University Medicine, Mainz D-55131, Germany; [2]Departments of Pediatrics and Human Genetics, Emma Children's Hospital, Amsterdam University Medical Centers, NL-1105 AZ Amsterdam, The Netherlands

Abstract After a long and largely disappointing detour, Genome Research has reidentified Rare Diseases as a major opportunity for improving health care and a clue to understanding gene and genome function. In this Special Issue of *CSH Molecular Case Studies* on Rare Diseases, several invited Perspectives, numerous Case Reports, and this Editorial itself address recent breakthroughs as well as unsolved problems in this wide field. These range from exciting prospects for gap-free diagnostic whole-genome sequencing to persisting problems related to identifying and distinguishing pathogenic and benign variants; and from the good news that soon, the United Kingdom will no longer be the only country to have introduced whole-genome sequencing into health care to the sobering conclusion that in many countries the clinical infrastructure for bringing Genome Medicine to the patient is still lacking. With less than 5000 genes firmly implicated in disease, the identification of at least twice as many disease genes is a major challenge, and the elucidation of their function is an even larger task. But given the renewed interest in rare diseases, their importance for health care, and the vast and growing spectrum of concepts and methods for studying them, the future of Human Genome Research is bright.

THE HUMAN GENOME PROJECT: FROM RARE TO COMMON DISEASES

That genome sequencing would be key to the diagnosis, prevention, and even cure for many of the genetic disorders listed in Victor McKusick's catalog of Mendelian phenotypes (McKusick 2007)[3] was an important driving force for establishing the Human Genome Project (HGP) in 1990 (McKusick 1982, 1992). About halfway to its completion, however, the orientation of this project changed. Paradoxically, this happened when several dozens of well-known Mendelian disorders, including Duchenne muscular dystrophy, cystic fibrosis, fragile X mental retardation, myotonic dystrophy, and Huntington's chorea, had already been elucidated by positional cloning strategies, maybe partly because of the belief that most low-hanging fruit had already been reaped. More importantly, however, common disorders such as dementia, schizophrenia, diabetes, autism, cardiovascular disease, and hypertension were more appealing and easier to "sell" to the public and policy makers alike, which is why they were soon promoted as the new target of human genome research (Risch and Merikangas 1996; Collins et al. 1998). Although it secured public support and continued funding for the HGP, the scientific underpinning of this move was rather weak, as it was based on the catchy but unproven common disease–common variant (CD/CV) hypothesis—that is, the assumption that for common diseases, common genetic risk factors must exist. And a long-lasting legacy of this decision was that the conceptually sound and

Corresponding author:
ropers@molgen.mpg.de

---

[3]Referred to here as "rare diseases," although collectively they are all but rare.

successful research into Mendelian disorders remained unpopular and vastly underfunded until well into the second decade of this millennium.

Through the years, billions were spent in search of common genetic variants and on genome-wide association studies (GWASs) to identify diagnostic and predictive markers of common diseases, despite growing concern about the doubtful validity of this concept and the meager yield of these efforts. However, because of the large number of prominent proponents of this research and their overly optimistic predictions (Bell 1998; Chakravarti 1999; Goldstein et al. 2003; Collins 2005), those critical voices were rarely heard. By 2003, several hundred thousand single-nucleotide polymorphisms (SNPs) had been identified (International HapMap Consortium 2003) and 5 years later, GWASs in large cohorts of patients and healthy individuals had indeed revealed significantly disease-associated genetic markers for a whole range of common diseases (Manolio et al. 2008). However, because of their low frequency and/or too low odds ratios, the vast majority of these SNPs turned out to be unsuitable as diagnostic or predictive markers. Also, the hope of the GWAS community that these markers would be a rich source of information about the molecular basis and the pathogenesis of the associated diseases has not materialized, except for a handful of common disorders (Ropers 2007; Maher 2008).

IT IS TIME TO TURN THE PAGE

In hindsight, there are plausible explanations for this sobering outcome. First, factors such as dietary habits, but also epigenetic factors and de novo mutations, seem to play a more significant role in the etiology of complex diseases than previously assumed.

Second, there are apparently no common genetic risk factors for the vast majority of common diseases—the CD/CV hypothesis is therefore misguided. For intellectual disability and autism, but also for schizophrenia and probably other common conditions, the lack of common associated markers this can be explained by the reduced fecundity of affected individuals (Power et al. 2013). Therefore, under equilibrium conditions, mutations causing these disorders will be rapidly lost from populations and replaced by others, which is why attempts to infer genetic relationships between psychiatric disorders from shared associated markers were bound to fail (Lee et al. 2013).

Moreover, as pointed out cogently and indeed irrefutably, even the concept underlying the HapMap project (Lander 1996; Terwilliger and Weiss 1998)—that is, to search for conserved haplotype blocs in the genome to reduce the number of markers needed for GWASs—turned out to have serious limitations (Weiss and Terwilliger 2000; Terwilliger and Hiekkalinna 2006).

Finally, the complexity of common diseases is often not due to multifactorial inheritance, but to genetic heterogeneity, because many clinically uniform disorders result from defects in different genes, with intellectual disability (ID) being the most heterogeneous example by far. Extrapolation from the number of known X-linked, autosomal dominant and recessive forms of ID suggests that the total number of genes implicated in ID will run into the thousands (Ropers 2010).

Because several of these facts and arguments are not new, one cannot help wondering why they have not been taken into consideration before reorienting human genome research on common disorders, and why it took almost 15 years before the most fervent supporters of this decision admitted the failure of this strategy (Zuk et al. 2012). Already in 2004, the estimated number of protein-coding human genes had dropped to 25,000 or less (International Human Genome Sequencing Consortium 2004), and soon afterward, high-throughput next-generation sequencing techniques had set the stage for the serial elucidation of single-gene disorders.

RARE DISEASES REENTER THE STAGE

For obvious reasons, X-chromosomal disorders were among the first to be studied, but after the cloning of the fragile X gene in 1991 (Verkerk et al. 1991), several years elapsed before the systematic search for other causes of X-linked ID picked up speed (Ropers and Hamel 2005). A little later, systematic research into recessive ID was triggered by homozygosity mapping in consanguineous families showing that this condition is extremely heterogeneous (Najmabadi et al. 2007; Kuss et al. 2011). The elucidation of dominant forms of ID received a massive boost when next-generation trio sequencing revealed that, in outbred populations, dominant de novo mutations are the predominant cause of ID (Vissers et al. 2010).

During the past decade, numerous large studies were performed to identify molecular defects underlying ID and related disorders (Najmabadi et al. 2011; Alazami et al. 2015; Karaca et al. 2015; Scott et al. 2016; Riazuddin et al. 2017; Harripaul et al. 2018; Hu et al. 2019). For X-linked (Hu et al. 2016; Martin et al. 2021) and autosomal dominant forms of ID, the yield of novel genes has been decreasing (https://www.sanger.ac.uk/collaboration/deciphering-developmental-disorders-ddd/) suggesting that the size of the relevant gene pools is limited and that most of the respective genes have already been identified. In contrast, despite many large studies performed in populations from the so-called "consanguinity belt" (Bittles 2001), the relative yield of novel genes for recessive forms of ID has been remarkably stable over time, and the overlap between subsequent samples from the same population or between unrelated ones is still small (Harripaul et al. 2018; Hu et al. 2019). In line with theoretical considerations (Ropers 2007), these observations indicate that most Mendelian disorders are recessive, and that their molecular causes are still largely unknown.

According to recent estimates, based on counts involving the entire human genome except the gene-poor Y chromosome (Nurk et al. 2021), there are almost 20,000 protein-coding genes in the human genome—that is, fivefold less than estimated at the start of the HGP. Because almost all are shared with other mammals and even vertebrates, it is very likely that at least "certain mutations in every non-redundant gene will have [negative] phenotypic consequences, either constitutively or in response to specific environmental challenges" (Chong et al. 2015). In other words, nearly all of these should be (at least conditional) disease genes. Because so far, less than 5000 human genes with disease-causing mutations have been identified, this suggests that there are at least twice as many waiting to be detected. Given that in the United Kingdom whole-genome sequencing (WGS) has been adopted as routine diagnostic measure by the NHS (Turnbull et al. 2018) and other countries are about to follow suit, it is a safe bet that henceforth the bulk of relevant genetic and clinical data needed to identify novel genes and their functions will be generated through health care and that the time for large gene hunting expeditions is over.

At present, even long-read WGS techniques have not revealed causative molecular defects in more than, say, two-thirds of diagnostic cases. But because of the unexpected frequency and complexity of large genome rearrangements revealed by optical mapping (Mantere et al. 2021) and the significantly reduced error rate and improved throughput of very-long-read nanopore sequencing (Miller et al. 2021), this could change soon.

Still, it is quite possible that a substantial portion of the missing genetic factors will not cause disease on their own but only in concert with others, resulting in di-, tri-, or even more complex polygenic disorders. Since 2018, when so-called polygenic risk scores (PRSs), calculated by integration of odds ratios for hundreds or thousands of disease-associated SNPs, were presented as markers of complex diseases (Khera et al. 2018), it has become clear that in general, their diagnostic and predictive value is much lower than that of mutations causing monogenic disorders. Like GWASs, PRSs are also plagued by stratification, and

they are particularly unsuitable as starting points for unraveling the pathogenesis of complex diseases. For the vast majority of the variants that are associated with complex disease, the risk conferred by individual markers is small, and they are difficult to identify—two compelling reasons for not losing interest in the elucidation of single-gene disorders.

WHAT THIS ISSUE CONTAINS—AND WHAT NOT YET

Celebrating this year's Rare Disease Day, this Special Issue of *CSH Molecular Case Studies* comprises four invited Commentaries or Perspectives on important developments in this field as well as more than a dozen short publications dealing with rare diseases.

Rachel Alvarez's "Patient perspective: my rare disease journey" sets the stage with a deeply moving account of her own disease history, her indefatigable fight for a correct diagnosis, and her determination to stay alive and to make the best of it for herself and others. As the Executive Director of Cure CMD, an international registry pushing research and therapy of congenital muscle disease, she reminds us that many "rare disorders" are severe, difficult to diagnose clinically, and often without a known molecular cause, and that despite recent progress in disease management (see van Karnebeek et al., this issue), there is presently no cure for most of these disorders. Therefore she emphasizes the importance of defining standards of care for rare diseases and their manifestations.

Although whole-exome sequencing (WES) and, more recently, WGS have added a new dimension to the diagnosis and elucidation of rare diseases, on their own these methods do not reliably detect mutations interfering with normal splicing nor regulatory mutations altering gene expression. In their Perspective entitled "Toward transcriptomics as a primary tool for rare disease investigation," Stephen Montgomery et al. review the recent literature and the impacts of sequencing RNA in a diagnostic context, concluding that "interrogation of multiple cell types through transcriptomic profiling may be an effective approach to enhancing diagnostic yield in rare disease."

In the Perspective entitled "Personalized medicine for rare genetic disorders: can we make it happen?," van Eeghen et al. provide an overview of personalized medicine for rare diseases, with a focus on inherited disorders of metabolism (IDMs). Because numerous IDMs cause neurodevelopmental disorders, they argue that recent progress in this field will also improve the chances for prevention, early detection, and treatment of early-onset brain dysfunction. Three separate articles in this issue (Coughlin et al.; Hauth et al.; Nguyen et al.) deal with monogenic disorders of metabolism that lead to neurodevelopment disorders, thereby supporting the validity of this concept.

Our present inability to reliably identify pathogenic variants is one of the central problems hampering the elucidation of rare diseases and genetic diagnosis. Approaching the problem from the other end, databases listing variants from exomes and genomes from very large cohorts of apparently healthy individuals have been generated to identify, or at least enrich, nondamaging variants (see Genome Aggregation Database, https://gnomad.broadinstitute.org; or the TOPMed Bravo database, https://bravo.sph.umich.edu/). Variants that are conserved between humans and other primates have been exposed to a far more stringent purifying selection because of the evolutionary distance separating these species; therefore, such shared SNVs (Sundaram et al. 2018), but also conserved structural variants, are likely benign (Kleinert and Kircher 2021).

In spite of the numerous published algorithms to predict the pathogenicity of genetic variants (e.g., see Niroula and Vihinen 2019) and the ongoing clinical implementation of WGS, not even 100 thousand of the nearly 10 billion theoretically possible SNVs in the human genome have been listed to date as pathogenic or probably pathogenic in the ClinVar database (Landrum et al. 2018), the most important diagnostic reference used by geneticists

world-wide. Since the development of CRISPR–Cas9-mediated saturation genome editing, the functional consequences of all possible variants in genes or given genomic segments can also be tested experimentally (Findlay et al. 2014, 2018), and although it is not a panacea, this potent approach is rapidly gaining popularity (for review, see Findlay 2021). Finally, thanks to the recent success of artificial intelligence-based algorithms to reliably deduce the three-dimensional structure of proteins from their amino acid (and, hence, DNA) sequence (e.g., see Jumper et al. 2021), there is reason to believe that this breakthrough will also further our ability to distinguish pathogenic from neutral variants and will prioritize them for functional studies. Given the importance and complexity of this theme, which merits a more extensive presentation, it will be the topic of a separate review, to be published in the next issue of this journal.

In view of these promising developments Stephen Kingsmore has appropriately called his Perspective "2022: a pivotal year for diagnosis and treatment of rare genetic diseases." Apart from addressing the highly successful introduction of ultra-rapid WGS for critically ill neonates and infants and decreasing sequencing costs, his central themes are the equitable implementation of WGS, realized only in the United Kingdom so far, but soon in Germany, too, and the urgent need to educate medical personnel for the age of Genomic Medicine that has already begun.

All case studies presented in this issue are related to rare diseases. For each of these, there is a special reason why it has been included, often because it is educative in one way or the other, such as the detection of a recessive founder mutation that appears to be specific to patients with Syrian ancestry (Hassan et al.), the variable intrafamilial expressivity of a heterozygous *TAOK1* frameshift variant (Hunter et al.), ethylmalonic encephalopathy masquerading as meningococcemia (Horton et al.), the unexpected detection of a somatic *FGFR1* mutation in cultured fibroblasts of a patient with closed spinal dysraphism (Kautto et al.), or just the description of a patient with a novel, not previously reported mutation. *CSH Molecular Case Studies* is one of the few journals accepting plausible associations between novel DNA variants and well-described disease phenotypes, even if only supported by a single case. The reason is that is takes such first observations to trigger the search for matching cases, as amply documented by our years of contact with readers of a pioneering paper reporting all sequence variants identified by next-generation sequencing in a large cohort of families with autosomal recessive ID (Najmabadi et al. 2011).

CONCLUSIONS AND OUTLOOK

"Tackling rare diseases through research and to enable all people living with a rare disease to receive an accurate diagnosis, care, and available therapy within 1 yr of coming to medical attention" was the vision of the International Rare Diseases Research Consortium (IRDiRC), a global collaborative initiative launched in 2011 by the European Commission and the U.S. National Institutes of Health (https://irdirc.org/who-we-are/). Today, a decade later, rare diseases are finally getting the attention they always deserved (Antonarakis and Beckmann 2006) but had not received for many years. Even the German government, one of the last in Europe with a blind eye for rare diseases, has recently adopted a law to implement WGS into health care (see Kingsmore, this issue). At long last, it has also joined the European "1+ Million Genomes" project (https://digital-strategy.ec.europa.eu/en/policies/1-million-genomes) realizing that for the future of genome medicine and research and their successful interplay, international exchange and data sharing is essential.

Technological innovations and novel algorithms, but also databases for pathogenic or benign variants, have drastically reduced the time to diagnosis, and thanks to affordable and fast very-long-read sequencing, genetic diagnoses based on the sequence of the

complete, truly personal genome have appeared on the horizon (Nurk et al. 2021). Although at present, the vision of IRDiRC to reduce the time to diagnosis and treatment to less than 1 year is mostly met for patients and families contacting academic or other specialized centers for rare diseases, some patients arrive there only after long diagnostic odysseys. This illustrates the urgent need to establish Genome Medicine as important aspect of medical education and training (see Kingsmore, in this issue), because the difficult truth is that in many countries, the infrastructure and training for referral to such centers is simply not there. Finally, by far not all individuals undergoing state-of-the-art diagnostics will receive a clear diagnosis, because either the underlying gene defect or the significance of identified variants may still be unknown.

This reminds us that despite the remarkable progress and the growing recognition of rare diseases as key to the understanding of gene and genome function, even its original aims defined a decade ago have not been achieved yet, and not only the identification of hitherto unknown disease genes, but above all the elucidation of their function, remains an enormous challenge.

Acknowledgments

REFERENCES

Alazami AM, Patel N, Shamseldin HE, Anazi S, Al-Dosari MS, Alzahrani F, Hijazi H, Alshammari M, Aldahmesh MA, Salih MA, et al. 2015. Accelerating novel candidate gene discovery in neurogenetic disorders via whole-exome sequencing of prescreened multiplex consanguineous families. *Cell Rep* **10:** 148–161. doi:10.1016/j.celrep.2014.12.015

Antonarakis SE, Beckmann JS. 2006. Mendelian disorders deserve more attention. *Nat Rev Genet* **7:** 277–282. doi:10.1038/nrg1826

Bell J. 1998. The new genetics in clinical practice. *Brit Med J* **316:** 618–620. doi:10.1136/bmj.316.7131.618

Bittles A. 2001. Consanguinity and its relevance to clinical genetics. *Clin Genet* **60:** 89–98. doi:10.1034/j.1399-0004.2001.600201.x

Chakravarti A. 1999. Population genetics—making sense out of sequence. *Nat Genet* **21:** 56–60. doi:10.1038/4482

Chong JX, Buckingham KJ, Jhangiani SN, Boehm C, Sobreira N, Smith JD, Harrell TM, McMillin MJ, Wiszniewski W, Gambin T, et al. 2015. The genetic basis of Mendelian phenotypes: discoveries, challenges, and opportunities. *Am J Hum Genet* **97:** 199–215. doi:10.1016/j.ajhg.2015.06.009

Collins FS. 2005. Personalized medicine: a new approach to staying well. Boston Globe July 17, 2005. http://archive.boston.com/news/globe/editorial_opinion/oped/articles/2005/07/17/personalized_medicine/

Collins FS, Patrinos A, Jordan E, Chakravarti A, Gesteland R, Walters L. 1998. New goals for the U.S. Human Genome Project: 1998–2003. *Science* **282:** 682–689. doi:10.1126/science.282.5389.682

Findlay GM. 2021. Linking genome variants to disease: scalable approaches to test the functional impact of human mutations. *Hum Mol Genet* **30:** R187–R197. doi:10.1093/hmg/ddab219

Findlay GM, Boyle EA, Hause RJ, Klein JC, Shendure J. 2014. Saturation editing of genomic regions by multiplex homology-directed repair. *Nature* **513:** 120–123. doi:10.1038/nature13695

Findlay GM, Daza RM, Martin B, Zhang MD, Leith AP, Gasperini M, Janizek JD, Huang X, Starita LM, Shendure J. 2018. Accurate classification of *BRCA1* variants with saturation genome editing. *Nature* **562:** 217–222. doi:10.1038/s41586-018-0461-z

Goldstein DB, Ahmadi KR, Weale ME, Wood NW. 2003. Genome scans and candidate gene approaches in the study of common diseases and variable drug responses. *Trends Genet* **19:** 615–622. doi:10.1016/j.tig.2003.09.006

Harripaul R, Vasli N, Mikhailov A, Rafiq MA, Mittal K, Windpassinger C, Sheikh TI, Noor A, Mahmood H, Downey S, et al. 2018. Mapping autosomal recessive intellectual disability: combined microarray and exome sequencing identifies 26 novel candidate genes in 192 consanguineous families. *Mol Psychiatry* **23:** 973–984. doi:10.1038/mp.2017.60

Hu H, Haas SA, Chelly J, Van Esch H, Raynaud M, de Brouwer AP, Weinert S, Froyen G, Frints SG, Laumonnier F, et al. 2016. X-exome sequencing of 405 unresolved families identifies seven novel intellectual disability genes. *Mol Psychiatry* **21:** 133–148. doi:10.1038/mp.2014.193

Hu H, Kahrizi K, Musante L, Fattahi Z, Herwig R, Hosseini M, Oppitz C, Abedini SS, Suckow V, Larti F, et al. 2019. Genetics of intellectual disability in consanguineous families. *Mol Psychiatry* **24:** 1027–1039. doi:10.1038/s41380-017-0012-2

International HapMap Consortium. 2003. The International HapMap Project. *Nature* **426:** 789–796. doi:10 .1038/nature02168

International Human Genome Sequencing Consortium. 2004. Finishing the euchromatic sequence of the human genome. *Nature* **431:** 931–945. doi:10.1038/nature03001

Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, et al. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature* **596:** 583– 589. doi:10.1038/s41586-021-03819-2

Karaca E, Harel T, Pehlivan D, Jhangiani SN, Gambin T, Coban Akdemir Z, Gonzaga-Jauregui C, Erdin S, Bayram Y, Campbell IM, et al. 2015. Genes that affect brain structure and function identified by rare variant analyses of Mendelian neurologic disease. *Neuron* **88:** 499–513. doi:10.1016/j.neuron.2015.09 .048

Khera AV, Chaffin M, Aragam KG, Haas ME, Roselli C, Choi SH, Natarajan P, Lander ES, Lubitz SA, Ellinor PT, et al. 2018. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet* **50:** 1219–1224. doi:10.1038/s41588-018-0183-z

Kleinert P, Kircher M. 2021. CADD-SV—a framework to score the effects of structural variants in health and disease. bioRxiv doi:10.1101/2021.07.10.451798

Kuss AW, Garshasbi M, Kahrizi K, Tzschach A, Behjati F, Darvish H, Abbasi-Moheb L, Puettmann L, Zecha A, Weissmann R, et al. 2011. Autosomal recessive mental retardation: homozygosity mapping identifies 27 single linkage intervals, at least 14 novel loci and several mutation hotspots. *Hum Genet* **129:** 141–148. doi:10.1007/s00439-010-0907-3

Lander ES. 1996. The new genomics: global views of biology. *Science* **274:** 536–539. doi:10.1126/science.274 .5287.536

Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, Gu B, Hart J, Hoffman D, Jang W, et al. 2018. ClinVar: improving access to variant interpretations and supporting evidence. *Nucl Acids Res* **46:** D1062–D1067. doi:10.1093/nar/gkx1153

Lee SH, Ripke S, Neale BM, Faraone SV, Purcell SM, Perlis RH, Mowry BJ, Thapar A, Goddard ME, Witte JS, et al. 2013. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat Genet* **45:** 984–994. doi:10.1038/ng.2711

Maher B. 2008. Personal genomes: the case of the missing heritability. *Nature* **456:** 18–21. doi:10.1038/ 456018a

Manolio TA, Brooks LD, Collins FS. 2008. A HapMap harvest of insights into the genetics of common disease. *J Clin Invest* **118:** 1590–1605. doi:10.1172/jci34772

Mantere T, Neveling K, Pebrel-Richard C, Benoist M, van der Zande G, Kater-Baats E, Baatout I, van Beek R, Yammine T, Oorsprong M, et al. 2021. Optical genome mapping enables constitutional chromosomal aberration detection. *Am J Hum Genet* **108:** 1409–1422. doi:10.1016/j.ajhg.2021.05.012

Martin HC, Gardner EJ, Samocha KE, Kaplanis J, Akawi N, Sifrim A, Eberhardt RY, Tavares ALT, Neville MDC, Niemi MEK, et al. 2021. The contribution of X-linked coding variation to severe developmental disorders. *Nat Commun* **12:** 627. doi:10.1038/s41467-020-20852-3

McKusick VA. 1982. The human genome through the eyes of a clinical geneticist. *Cytogenet Cell Genet* **32:** 7– 23. doi:10.1159/000131682

McKusick VA. 1992. Human genetics: the last 35 years, the present, and the future. *Am J Hum Genet* **50:** 663– 670

McKusick VA. 2007. Mendelian Inheritance in Man and its online version, OMIM. *Am J Hum Genet* **80:** 588– 604. doi:10.1086/514346

Miller DE, Sulovari A, Wang T, Loucks H, Hoekzema K, Munson KM, Lewis AP, Fuerte EPA, Paschal CR, Walsh T, et al. 2021. Targeted long-read sequencing identifies missing disease-causing variation. *Am J Hum Genet* **108:** 1436–1449. doi:10.1016/j.ajhg.2021.06.006

Najmabadi H, Motazacker MM, Garshasbi M, Kahrizi K, Tzschach A, Chen W, Behjati F, Hadavi V, Nieh SE, Abedini SS, et al. 2007. Homozygosity mapping in consanguineous families reveals extreme heterogeneity of non-syndromic autosomal recessive mental retardation and identifies 8 novel gene loci. *Hum Genet* **121:** 43–48. doi:10.1007/s00439-006-0292-0

Najmabadi H, Hu H, Garshasbi M, Zemojtel T, Abedini SS, Chen W, Hosseini M, Behjati F, Haas S, Jamali P, et al. 2011. Deep sequencing reveals 50 novel genes for recessive cognitive disorders. *Nature* **478:** 57–63. doi:10.1038/nature10423

Niroula A, Vihinen M. 2019. How good are pathogenicity predictors in detecting benign variants? *PLoS Comput Biol* **15:** e1006481. doi:10.1371/journal.pcbi.1006481

Nurk S, Koren S, Rhie A, Rautiainen M, Bzikadze AV, Mikheenko A, Vollger MR, Altemose N, Uralsky L, Gershman A. 2021. The complete sequence of a human genome. bioRxiv doi:10.1101/2021.05.26.445798

Power RA, Kyaga S, Uher R, MacCabe JH, Långström N, Landen M, McGuffin P, Lewis CM, Lichtenstein P, Svensson AC. 2013. Fecundity of patients with schizophrenia, autism, bipolar disorder, depression, anorexia nervosa, or substance abuse vs their unaffected siblings. *JAMA Psychiatry* **70:** 22–30. doi:10.1001/jamapsychiatry.2013.268

Riazuddin S, Hussain M, Razzaq A, Iqbal Z, Shahzad M, Polla DL, Song Y, van Beusekom E, Khan AA, Tomas-Roca L, et al. 2017. Exome sequencing of Pakistani consanguineous families identifies 30 novel candidate genes for recessive intellectual disability. *Mol Psychiatry* **22:** 1604–1614. doi:10.1038/mp.2016.109

Risch N, Merikangas K. 1996. The future of genetic studies of complex human diseases. *Science* **273:** 1516–1517. doi:10.1126/science.273.5281.1516

Ropers HH. 2007. New perspectives for the elucidation of genetic disorders. *Am J Hum Genet* **81:** 199–207. doi:10.1086/520679

Ropers HH. 2010. Genetics of early onset cognitive impairment. *Annu Rev Genomics Hum Genet* **11:** 161–187. doi:10.1146/annurev-genom-082509-141640

Ropers HH, Hamel BC. 2005. X-linked mental retardation. *Nat Rev Genet* **6:** 46–57. doi:10.1038/nrg1501

Scott EM, Halees A, Itan Y, Spencer EG, He Y, Azab MA, Gabriel SB, Belkadi A, Boisson B, Abel L, et al. 2016. Characterization of Greater Middle Eastern genetic variation for enhanced disease gene discovery. *Nat Genet* **48:** 1071–1076. doi:10.1038/ng.3592

Sundaram L, Gao H, Padigepati SR, McRae JF, Li Y, Kosmicki JA, Fritzilas N, Hakenberg J, Dutta A, Shon J, et al. 2018. Predicting the clinical impact of human mutation with deep neural networks. *Nat Genet* **50:** 1161–1170. doi:10.1038/s41588-018-0167-z

Terwilliger JD, Hiekkalinna T. 2006. An utter refutation of the "fundamental theorem of the HapMap". *Eur J Hum Genet* **14:** 426–437. doi:10.1038/sj.ejhg.5201583

Terwilliger JD, Weiss KM. 1998. Linkage disequilibrium mapping of complex disease: fantasy or reality? *Curr Opin Biotechnol* **9:** 578–594. doi:10.1016/s0958-1669(98)80135-3

Turnbull C, Scott RH, Thomas E, Jones L, Murugaesu N, Pretty FB, Halai D, Baple E, Craig C, Hamblin A, et al. 2018. The 100,000 Genomes Project: bringing whole genome sequencing to the NHS. *Brit Med J* **361:** k1952. doi:10.1136/bmj.k1952

Verkerk AJ, Pieretti M, Sutcliffe JS, Fu YH, Kuhl DP, Pizzuti A, Reiner O, Richards S, Victoria MF, Zhang FP, et al. 1991. Identification of a gene (*FMR*-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell* **65:** 905–914. doi:10.1016/0092-8674(91)90397-h

Vissers LE, de Ligt J, Gilissen C, Janssen I, Steehouwer M, de Vries P, van Lier B, Arts P, Wieskamp N, del Rosario M, et al. 2010. A de novo paradigm for mental retardation. *Nat Genet* **42:** 1109–1112. doi:10.1038/ng.712

Weiss KM, Terwilliger JD. 2000. How many diseases does it take to map a gene with SNPs? *Nat Genet* **26:** 151–157. doi:10.1038/79866

Zuk O, Hechter E, Sunyaev SR, Lander ES. 2012. The mystery of missing heritability: genetic interactions create phantom heritability. *Proc Natl Acad Sci* **109:** 1193–1198. doi:10.1073/pnas.1119675109