These large swathes of repetitive DNA come with different sets of rules, in terms of their genomic organization and evolution. They are also subject to different epigenetic regulation (molecular modifications to DNA and associated proteins that do not alter the underlying DNA sequence), which leads repetitive DNA to differ from euchromatin in terms of its organization, replication timing and transcriptional activity[17–19]. Many genome-wide tools and data sets cannot yet fully capture all this information from extremely repetitive DNA regions, and so scientists do not yet have a complete picture of what transcription factors bind to them, how these regions are spatially organized in the nucleus, or how regulation of these parts of our genome changes during development and in disease states. Now, much like the initial release of the genome decades ago, researchers are faced with a new, unexplored functional landscape in the human genome. Access to this information will drive technology and innovation to be inclusive of these repeat regions, once again broadening our understanding of genome biology.

In the past year, scientists have used extremely long and highly accurate sequence reads to reconstruct entire human chromosomes from telomere to telomere[4,5]. Last year also saw the release of a near-complete human reference genome from an effectively 'haploid' human cell line, with only five remaining gaps that mark the sites of rDNA arrays (go.nature.com/3rgz93y). In this line, cells have two identical pairs of chromosomes, simplifying the challenge of repeat assembly compared with typical human cells (which are diploid, with different chromosomes inherited from the mother and father). These maps together offer the first high-resolution glimpse of centromeric regions, segmental duplications, subtelomeric repeats and each of the five acrocentric chromosomes, which have very short arms made up almost entirely of highly repetitive DNA at one end.

It is tempting to think scientists are finally approaching the finish line. However, a single genome assembly, even if complete with near-perfect sequence accuracy, is an insufficient reference from which to study the sequence variation that exists across the human population. Existing maps that chart the diversity across the euchromatic parts of the genome must be extended to fully capture repetitive regions, where copy number and repeat organization vary between individuals. Doing so will require the development of strategies for routine production and analysis of complete human diploid genomes. The aspirational goal of reaching a more-complete and comprehensive reference of humanity will undoubtedly improve our understanding of genome structure and its role in human disease, and align with the promise and legacy of the Human Genome Project.

Karen H. Miga is at the UC Santa Cruz Genomics Institute, University of California, Santa Cruz, California 95064, USA.
e-mail: khmiga@ucsc.edu

1. International Human Genome Sequencing Consortium. *Nature* **409**, 860–921 (2001).
2. Venter, J. C. *et al. Science* **291**, 1304–1351 (2001).
3. Schneider, V. A. *et al. Genome Res.* **27**, 849–864 (2017).
4. Miga, K. H. *et al. Nature* **585**, 79–84 (2020).
5. Logsdon, G. A. *et al.* Preprint at bioRxiv https://doi.org/10.1101/2020.09.08.285395 (2020).
6. International Human Genome Sequencing Consortium. *Nature* **431**, 931–945 (2004).
7. Frankish, A. *et al. Nucleic Acids Res.* **47**, D766–D773 (2019).
8. Koren, S. *et al. Nature Biotechnol.* **30**, 693–700 (2012).
9. Jain, M. *et al. Nature Biotechnol.* **36**, 338–345 (2018).
10. Jain, M. *et al. Nature Biotechnol.* **36**, 321–323 (2018).
11. Chaisson, M. J. P. *et al. Nature* **517**, 608–611 (2015).
12. Wenger, A. M. *et al. Nature Biotechnol.* **37**, 1155–1162 (2019).
13. Nurk, S. *et al. Genome Res.* **30**, 1291–1305 (2020).
14. Bzikadze, A. V. & Pevzner, P. A. *Nature Biotechnol.* **38**, 1309–1316 (2020).
15. Zentner, G. E., Saiakhova, A., Manaenkov, P., Adams, M. D. & Scacheri, P. C. *Nucleic Acids Res.* **39**, 4949–4960 (2011).
16. Schueler, M. G., Higgins, A. W., Rudd, M. K., Gustashaw, K. & Willard, H. F. *Science* **294**, 109–115 (2001).
17. Sullivan, B. A. & Karpen, G. H. *Nature Struct. Mol. Biol.* **11**, 1076–1083 (2004).
18. Gilbert, D. M. *Curr. Opin. Cell Biol.* **14**, 377–383 (2002).
19. Prior, C. P., Cantor, C. R., Johnson, E. M., Littau, V. C. & Allfrey, V. G. *Cell* **34**, 1033–1042 (1983).

Human genome

A genetic revolution in rare-disease medicine

Fowzan S. Alkuraya

Mendelian diseases are caused by mutations in a single gene. The first draft of the human genome, published in 2001, had broad implications for how these diseases are diagnosed, managed and prevented.

When the first draft of the human genome was published[1,2], it was expected to have a transformative impact on medicine. Bold predictions were made about a paradigm shift in which medicine became personalized, predictive and preventive[3]. To many, no such transformation materialized, probably because of a focus on common diseases such as diabetes and coronary artery disease. But the predictions were right on target for Mendelian diseases — those caused by mutations in single genes — such as hereditary cancers and many forms of developmental delay in children.

"The true game-changer came when the draft genome was used in combination with 'next-generation' sequencing technologies."

Before the draft genome, basic information about the sequence and genomic location of a mutated gene had to be worked out through a process called cloning, in which short chromosomal segments were cut from human DNA using enzymes, and replicated in bacteria to produce sufficient quantities for analysis. Cloning was a stupendously laborious exercise that often took years and could be performed by only a few laboratories. The genetic underpinnings of most Mendelian diseases were therefore unknown, making diagnosis extremely difficult. Even for the few that did have a known underlying genetic basis (such as fragile X syndrome), a specialist was still likely to fail to make a diagnosis, because of the remarkable variability of the diseases' clinical presentation and their rarity[4].

In the 1990s, the development of 'positional mapping' methods made it easier to identify genes associated with Mendelian diseases. Early positional-mapping efforts involved comparing the DNA of several people who had the same disease, using a primitive genome map containing a few known sequences that vary between individuals; these acted as location markers to help researchers zero in on a candidate disease-causing region[5]. The primitive map, which dates back to 1987, was essential to early gene-discovery efforts. Nonetheless, its low resolution was a major obstacle to gene-discovery efforts.

It is hard, then, to overstate just how influential the human genome draft was for people with Mendelian diseases and their families. The draft did not directly link individual genes to diseases, but it did provide the necessary elements for a revolution in diagnosis. Initially, it provided a rich map of markers that permitted a much higher resolution in positional mapping. However, the true game-changer

came when the draft genome was used in combination with 'next-generation' sequencing technologies, which read entire genomes, rather than individual genes[4]. This gave researchers the ability to identify potential disease-causing variants across the genome much more rapidly than had previously been possible.

Thanks to this technological advance, the number of Mendelian diseases that have a known genetic cause went from 1,257 in 2001 to 4,377 at the time of writing (according to the OMIM database, an online catalogue of human genes and disorders; go.nature.com/omimdb). Patients are now increasingly liberated from the long-standing diagnostic bottleneck. Many can receive a diagnosis in hours when urgently needed, with a precision that remains unparalleled in medicine. This has opened the door for a true personalization of disease management. For instance, therapies are available for some specific disease-causing genetic variants, such as those in the gene *CFTR* that cause cystic fibrosis. We can also avoid futile interventions such as growth-hormone therapy that would be ineffective in children who have a Mendelian condition called Seckel syndrome, a type of dwarfism.

Once an association between a Mendelian disease and a gene is established, the disease is highly predictable, meaning that prevention is possible. For example, the American College of Medical Genetics and Genomics recommends that people who have their genomes sequenced for any diagnostic purpose be informed if they carry disease-causing variants in any of 59 genes[6] that are linked to potentially life-threatening Mendelian conditions for which pre-emptive management is available. A recent sequencing study[7] in the United Kingdom, involving some 50,000 volunteers aged between 40 and 69, showed that 2% harbour such actionable variants. And early data[8] show that population-based screening for these variants leads to a high rate of take-up for risk-management procedures. The ability to detect these variants and many more that will follow — as well as variants that influence responses to drugs — offers a glimpse of the potential medical benefits of a future in which genome sequencing is universal.

Another benefit of large-scale genome sequencing is reproductive empowerment. Carrier screening determines whether a person carries one copy of a 'recessive' genetic variant, which would cause disease if present in both copies of the gene in question — usually when both parents carry and pass the variant on to a child (Fig. 1). Empowered by that knowledge, carriers can make informed reproductive choices. Tay–Sachs disease and thalassaemia, two life-threatening conditions caused by recessive variants, have been largely eliminated in high-risk communities in New York and Cyprus, respectively, by carrier screening[9]. A future in which this model is expanded to target all severe or lethal recessive Mendelian disease genes is within reach, and is being championed by the private sector and publicly funded initiatives. However, it is important to note that there is much ethical debate around the use of genetic screening for reproductive choices, with concerns about 'screening out' certain groups, as well as other societal risks. In addition, the use of genetic screening for traits not related to health is not considered ethical.

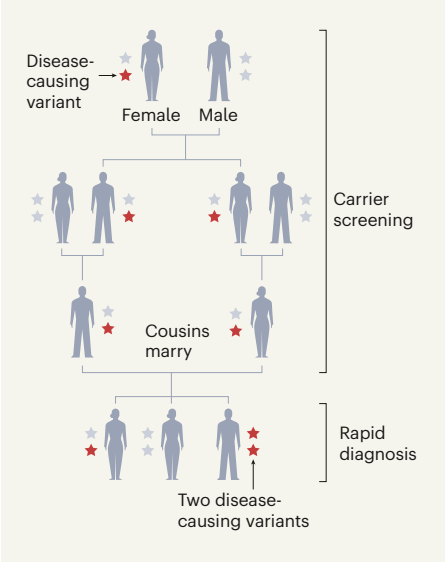

Figure 1 | Screening for Mendelian diseases. Recessive Mendelian diseases are those that arise when a person carries two copies of a disease-causing gene variant. In this hypothetical family tree, two children inherit a copy of a disease-causing variant from their mother, and a copy of a non-damaging variant from their father. In turn, they each pass one copy of the disease-causing variant on to their children. If these cousins (or any two people carrying the variant) were to have children, there is a chance that each offspring would inherit two copies, and so develop the disease. The human genome sequence[1,2] transformed our ability to identify disease-causing variants. Today, people can be screened to determine whether they carry such a variant, and people who have the disease can be rapidly diagnosed through genome sequencing.

Nowhere will carrier screening have a greater impact than in countries where unions between cousins are common. Because cousins share more variants than do unrelated individuals, it is more likely that they will share, and pass on, harmful recessive variants, causing recessive diseases. Saudi Arabia is a case in point. When the initial draft of the human genome was published, Saudi Arabia had the highest documented rates of recessive diseases in the world[10]. Twenty years later, nearly all major recessive diseases in the country have been characterized at the gene level[11]. Countless couples have pursued reproductive options informed by variant identification, and the country is on the cusp of rolling out an expanded screening programme.

Our improved understanding of Mendelian diseases has also begun to benefit people with common diseases that have more-complex genetic underpinnings. For example, a 2020 sequencing study[7] revealed that, for a small but significant fraction of people who have common diseases, a single genetic variant is the cause — that is, they have a Mendelian form of the disease. And aside from causation, genes associated with Mendelian disorders have been found to be risk factors for many common diseases[12]. New therapeutics for common diseases are being informed purely by human genomics, and Mendelian genes play no small part in that[13].

The medical genetics community has often been accused of making empty promises[14]. But genomics is now truly improving people's health. This is not just a vindication but also an inspiration to continue rewriting medicine using our DNA.

Fowzan S. Alkuraya is in the Department of Translational Genomics, Center for Genomic Medicine, King Faisal Specialist Hospital and Research Center, Riyadh 11211, Saudi Arabia.
e-mail: falkuraya@kfshrc.edu.sa

1. International Human Genome Sequencing Consortium. *Nature* **409**, 860–921 (2001).
2. Venter, J. C. *et al. Science* **291**, 1304–1351 (2001).
3. Collins, F. S. & McKusick, V. A. *J. Am. Med. Assoc.* **285**, 540–544 (2001).
4. Chong, J. X. *et al. Am. J. Hum. Genet.* **97**, 199–215 (2015).
5. Donis-Keller, H. *et al. Cell* **51**, 319–337 (1987).
6. Kalia, S. S. *et al. Genet. Med.* **19**, 249–255 (2017).
7. Van Hout, C. V. *et al. Nature* **586**, 749–756 (2020).
8. Buchanan, A. H. *et al. Genet. Med.* **22**, 1874–1882 (2020).
9. Antonarakis, S. E. *Nature Rev. Genet.* **20**, 549–561 (2019).
10. Abouelhoda, M. *et al. Genet. Med.* **18**, 1244–1249 (2016).
11. Monies, D. *et al. Am. J. Hum. Genet.* **104**, 1182–1201 (2019).
12. Freund, M. K. *et al. Am. J. Hum. Genet.* **103**, 535–552 (2018).
13. Szustakowski, J. D. *et al.* Preprint at medRxiv https://doi.org/10.1101/2020.11.02.20222232 (2020).
14. Joyner, M. J. & Paneth, N. *J. Clin. Invest.* **129**, 946–948 (2019).