the same direction when they aligned with the field. If they were pointing in opposite directions, they stayed that way after the pulse. Sailer *et al*. detected this difference by moving the ions back into the analysis trap. By repeating this cycle and varying the delay between pulses, the difference in Larmor frequencies and hence the *g* factors could be directly mapped out. Amazingly, the authors were able to follow the process over several seconds, during which the moments revolved 100 billion times.

Assuming that no exotic new physics was involved, the measurement yielded the difference in the nuclear size of the two isotopes with an uncertainty of 0.0001 of the radius. This agrees with existing data, but is ten times more accurate. Importantly, the approach is radically different from other methods, and can provide powerful systematic checks. Because the effect of nuclear size on the *g* factor grows rapidly with nuclear charge, there is considerable potential for applying the method to elements that are heavier than neon.

The authors arrived at an intriguing interpretation by incorporating an independently measured value for the nuclear size difference in their analysis[8]. By doing so, they ensured that all known physics was accounted for, which allowed them to set upper bounds on unknown physics, such as particles that might mediate a hypothetical interaction between electrons and nucleons. Neon-22 has two more neutrons than neon-20, so an interaction between the electron and these two neutrons would be evident in the team's difference measurements.

One such interaction might arise through the exchange of a massive particle, dubbed the relaxion, that was proposed in 2015 as a possible solution to a puzzle known as the hierarchy problem[9]. This puzzle essentially concerns the fact that the observed mass of the Higgs boson is considered unexpectedly low in the context of the standard model of particle physics, and evidence for new physics that could explain the situation has not yet been discovered in experiments at the Large Hadron Collider at CERN, Europe's particle-physics laboratory near Geneva, Switzerland. But the relaxion couples to the Higgs boson, influencing its mass, and as the early Universe expanded, the expansion would have weakened this coupling, thereby 'relaxing' the Higgs mass down to its currently observed value.

Constraints have been placed on the parameter values that could feasibly verify the existence of this particle in the past five years[10]. But in some of the methods used previously to pin down these values, the signature of the relaxion competed directly with effects caused by features of the nuclear structure, which are difficult to track[11,12]. Sailer and colleagues' interpretation is quite straightforward, and a welcome addition to the tool chest for searches for new physics. Ultimately, combined use of these diverse methods will ensure that a robust and coherent picture emerges. There are already plans to take this to the next level by evaluating combinations of *g* factors and the lowest (ground-state) energies of electrons for pairs of isotopes[13].

Finally, one cannot discuss methods for precise measurements of ions without considering how they can be applied to antimatter. This question looms large in particle-physics research, because the Universe seems to be made almost entirely of matter, but the Big Bang is thought to have created just as much antimatter. And, indeed, Sailer *et al*. conclude their report by outlining ideas for comparing the *g* factors of protons and antiprotons using their method in search of an answer. Relaxion or not, there is work to be done; no time to relax.

Gerald Gwinner is in the Department of Physics and Astronomy, University of Manitoba, Winnipeg, Manitoba R3T 2N2, Canada. **Roshani Silwal** is in the Department of Physics and Astronomy, Appalachian State University, Boone, North Carolina 28608, USA.
e-mails: gerald.gwinner@umanitoba.ca; silwalr@appstate.edu

1. Sailer, T. *et al*. *Nature* **606**, 479–483 (2022).
2. Dirac, P. A. M. *Proc. R. Soc. Lond. A* **117**, 610–624 (1928).
3. Schwinger, J. *Phys. Rev.* **73**, 416–417 (1948).
4. Odom, B., Hanneke, D., D'Urso, B. & Gabrielse, G. *Phys. Rev. Lett.* **97**, 030801-4 (2006).
5. Aoyama, T., Hayakawa, M., Kinoshita, T. & Nio, M. *Phys. Rev. Lett.* **109**, 111807 (2012).
6. Sturm, S. *et al*. *Eur. Phys. J. Spec. Top.* **227**, 1425–1491 (2019).
7. Rainville, S., Thompson, J. K. & Pritchard, D. E. *Science* **303**, 334–338 (2004).
8. Angeli, I. & Marinova, K. P. *Atomic Data Nucl. Data Tables* **99**, 69–95 (2013).
9. Graham, P. W., Kaplan, D. E. & Rajendran, S. *Phys. Rev. Lett.* **115**, 221801-9 (2015).
10. Delaunay, C., Frugiuele, C., Fuchs, E. & Soreq, Y. *Phys. Rev. D* **96**, 115002 (2017).
11. Counts, I. *et al*. *Phys. Rev. Lett.* **125**, 123002 (2020).
12. Solaro, C. *et al*. *Phys. Rev. Lett.* **125**, 123003 (2020).
13. Debierre, V., Keitel, C. H. & Harman, Z. Preprint at https://arxiv.org/abs/2202.01668 (2022).

The authors declare no competing interests.

Genetics

The road to accurate and complete human genomes

John T. Lovell & Jane Grimwood

Sequences of the human genome have typically included gaps in repetitive regions of DNA. A combination of state-of-the-art technologies has now enabled researchers to generate the first complete human genome sequence.

The telomere-to-telomere (T2T) consortium has just taken a massive step forwards in human genomics. Writing in *Science*[1], the group describes the first essentially complete sequence of a human genome, called the T2T-CHM13 assembly. This landmark resource, which contains regions of the genome that had not previously been represented in a reference assembly, provides an alternative sequence to the existing human reference genome[2–4]. Along with the sophisticated methods that enabled its construction, the assembly opens up a path to the generation of many diverse human genomes.

The Human Genome Project was launched in 1990 with the goal of obtaining a highly accurate sequence of most of the human genome. A draft was published[2] in 2001, and a finished genome three years later[3]. This reference sequence was a momentous achievement by scientists from 20 institutions in 6 countries, at an estimated cost of more than US$2 billion (see go.nature.com/3se0wie). It has enabled some of the most important biological discoveries of the past two decades. Nonetheless, technological limitations left some regions of the genome unresolved.

The strategy for sequencing the existing human reference genome was a clone-based approach, in which individual genomic regions (clones) were ordered and oriented across each chromosome, then sequenced. Each clone was 'finished' by an experienced scientist who manually fixed any sequence misassemblies and carried out post-hoc sequencing reactions to fill in gaps. The finished clones were pieced together to construct the final genome assembly. But accurately placing clones was not always possible in repetitive regions of the genome, resulting in some large gaps in regions around these areas (Fig. 1), including around structures called centromeres at the chromosome centre and the sub-telomeric regions close to the chromosome ends.

These limitations were further exacerbated by the fact that the original human genome

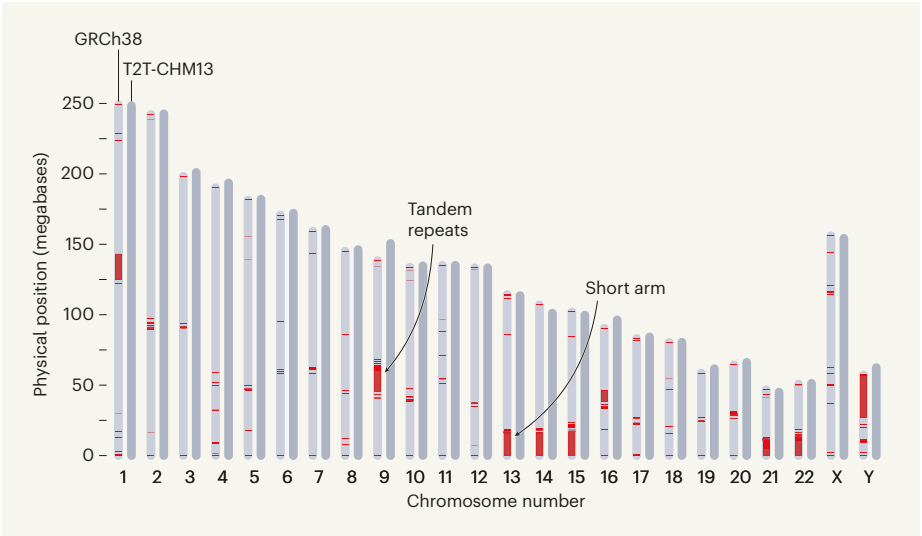

Figure 1 | Filling in the gaps in human genome assemblies. Structure of the GRCh38 and T2T-CHM13 human genome assemblies. The telomere-to-telomere (T2T) consortium[1] has generated an almost-complete human genome, dubbed T2T-CHM13. Comparison of the chromosomes in this assembly with an earlier human genome, GRCh38, reveals regions in which previous gaps (red) have now been filled. Gaps in GRCh38 arose owing to the complexity of certain regions, including areas rich in tandem repeats, such as that in chromosome 9, and short chromosome arms such as that in chromosome 13.

was constructed from the DNA of several people, leading to a mosaic genome. The gene models and sequencing gaps in the finished genome have been continuously improved over the past two decades by the Genome Reference Consortium (GRC), with the 2013 'GRCh38' assembly and its 2019 update (GRCh.38.p13) forming the source of comparison for most modern genome analyses[4].

The T2T consortium took a different approach, known as shotgun sequencing, in which the genome is randomly broken up into millions of fragments for sequencing, and a computer algorithm finds overlaps between the sequences to build the fragments into chromosomes. In the past, genomes generated through shotgun sequencing have tended to be incomplete, but technological advances — including the ability to generate long and accurate sequences from sequence fragments — have now facilitated gap-free chromosome assemblies, even in the difficult repetitive regions of the genome.

Despite the major methodological differences used to construct all the genomes completed since 2003, the order of genes on the large chromosome arms are predominantly the same. The exceptions are small inversions of sequences between GRCh38 and T2T-CHM13. These occur on chromosome 1, in a stretch of 1.9 million bases (1.9 megabases), chromosome 8 (0.84 megabases) and chromosome 16 (3.8 megabases), and are probably caused by molecular variation between the genomes sequenced to construct GRCh38 and T2T-CHM13.

By contrast, the highly repetitive regions of the genome are vastly improved in T2T-CHM13 compared with previous assemblies[1,5,6], especially on chromosome 9 and the acrocentric chromosomes, in which the centromere is located near one end of the chromosome. In these cases, repetitive sequences in T2T-CHM13 largely fill in the previous gaps. These expansive non-protein-coding sequences were once thought of as 'junk', but are now known to underlie a variety of traits and disorders[7,8]. The improvements in repeat annotation are striking: there are 3.6% more telomere repeats and 254% more tandem repeated sequences in T2T-CHM13 than in GRCh.38.p13. These sequences generally fill the gaps that remained in GRCh38 (Fig. 1).

"The complete representation of repeats in the new genome will facilitate the diagnosis and potential treatment of genetic disorders."

The complete representation of repeats in T2T-CHM13 will facilitate the diagnosis and potential treatment of genetic disorders. This sequence will ultimately replace GRCh38 as the reference used for analysis, but genomic-medicine initiatives must first adapt by deploying the advanced technologies needed to sequence repeats in patients. This process will require new analytical and clinical methods[9]. Furthermore, accurate strategies for producing long sequences must be improved, because their production is currently much slower and lower-throughput than is standard for short-read sequencing.

Despite the success of the T2T-CHM13 assembly, the available human reference sequences do not capture the diversity of sequences across human populations. It is this molecular diversity that underlies genetic disorders, inherited traits and disease susceptibility. Genetic differences between unrelated people mean that some genomes will be less comparable with T2T-CHM13 than others. Going forward, the next step must be the generation of a collection of reference genomes from unrelated individuals — a challenge that is currently being addressed through pan-genome frameworks[10].

Furthermore, T2T-CHM13 was made possible by the use of a line of 'homozygous hydatidiform mole' cells, in which a single sperm-derived genome has been duplicated. In these cells, the two copies of DNA are identical, whereas the two copies of DNA present in most cells are derived from both egg and sperm and so are different (heterozygous). The use of this homozygous cell line greatly reduces the complexity of the human genome by providing a single target for assembly, thus increasing accuracy compared with the patchwork of many individuals whose heterozygous DNA made up the first reference genome. New algorithms, and possibly new sequencing technologies, will be required to generate multiple heterozygous genome assemblies, to fully represent the diversity of human DNA.

John T. Lovell and **Jane Grimwood** are at the Genome Sequencing Center, HudsonAlpha Institute for Biotechnology, Huntsville, Alabama 35806, USA, and at the Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, California, USA.
e-mails: jlovell@hudsonalpha.org; jgrimwood@hudsonalpha.org

1. Nurk, S. *et al. Science* **376**, 44–53 (2022).
2. Lander, E. S. *et al. Nature* **409**, 860–921 (2001).
3. International Human Genome Sequencing Consortium. *Nature* **431**, 931–945 (2004).
4. Schneider, V. A. *et al. Genome Res.* **27**, 849–864 (2017).
5. Hoyt, S. J. *et al. Science* **376**, eabk3112 (2022).
6. Altemose, N. *et al. Science* **376**, eabl4178 (2022).
7. Zhou, J. *et al. Nature Genet.* **51**, 973–980 (2019).
8. Rheinbay, E. *et al. Nature* **578**, 102–111 (2020).
9. Aganezov, S. *et al. Science* **376**, eabl3533 (2022).
10. Wang, T. *et al. Nature* **604**, 437–446 (2022).

The authors declare no competing interests.
This article was published online on 23 May 2022.