

decades). These are based on, for example, silicon- or germanium-based materials^{10,11}, silicon nanowires¹² and III–V–silicon interfaces (heterojunctions)¹³.

For TFETs to have steep SS values, devices must be designed with precise control of doping — the substitution of a small percentage of one type of atom for another, to change the amount of charge available to conduct current through the semiconductor. This property allows an internal electrical field to be effectively induced at the tunnelling junction. Also, established architectures such as the gate stacks of conventional MOSFETs need improvement.

Sarkar *et al.* propose a TFET made from a bilayer of molybdenum disulfide (MoS₂) and bulk germanium that overcomes many of the above challenges. The bilayer 2D MoS₂ crystal is placed on a germanium substrate, forming an ultra-thin junction through van der Waals bonding (see Fig. 1a of the paper¹). This custom-built heterojunction shows negative differential resistance (that is, the amount of current through the junction tends to decrease with increasing voltage across it), confirming the occurrence of tunnelling transport across the van der Waals bonds. Because it is not easy to make gate stacks from 2D MoS₂ materials, the authors use a solid polymer electrolyte as the device's gate electrode.

At room temperature, Sarkar and colleagues' TFET achieves a very steep SS (31.1 mV per decade of current, averaged over 4 decades); the minimum value reached is 3.9 mV per decade of current. This 2D semiconducting heterojunction ranks as one of the most promising materials for the fabrication of switches that could operate with supply voltages as low as 0.1 V, resulting in a reduction in power consumption of more than 90% compared with conventional FETs.

However, challenges remain. They include enhancing the robustness of the 2D crystal, controlling the properties of the gate's solid-state insulator, tuning the threshold voltage (above which the current increases nonlinearly with the supplied gate voltage), finessing the doped channel's structure and increasing the tunnelling current. Assuming that these issues can be addressed, Sarkar and co-workers' findings could lead to improved CMOS applications and, as the authors suggest, even to highly effective sensors for biological applications based on TFETs¹⁴. Junctions and materials such as those reported by the authors could contribute to the development of ultra-low-power and energy-efficient integrated circuits, which would find their way into mainstream electronic devices. ■

Katsuhiro Tomioka is in the Graduate School of Information Science and Technology, and the Research Center for Integrated Quantum Electronics (RCIQE), Hokkaido University, 060–0814 Sapporo, Japan.
e-mail: tomioka@rciqe.hokudai.ac.jp

1. Sarkar, D. *et al.* *Nature* **526**, 91–95 (2015).
2. del Alamo, J. A. *Nature* **479**, 317–323 (2011).
3. Pillarisetty, R. *Nature* **479**, 324–328 (2011).
4. Radisavljevic, B., Radenovic, A., Brivio, J., Giacometti, V. & Kis, A. *Nature Nanotechnol.* **6**, 147–150 (2011).
5. Ferrain, I., Collinge, C. A. & Colinge, J.-P. *Nature* **479**, 310–316 (2011).
6. Seabaugh, A. C. & Zhang, Q. *Proc. IEEE* **98**, 2095–2110 (2010).
7. Gopalakrishnan, K., Griffin, P. B. & Plummer, J. D. *IEEE Int. Electron Devices Meet.* 289–292 (2002).
8. Salahuddin, S. & Datta, S. *Nano Lett.* **8**, 405–410 (2008).
9. Pott, V. *et al.* *Proc. IEEE* **98**, 2076–2094 (2010).
10. Jeon, K. *et al.* *IEEE VLSI Technol. Symp.* 121–122 (2010).
11. Kim, S. H., Kam, H., Hu, C. & Liu, T.-J. K. *IEEE VLSI Technol. Symp.* 178–179 (2009).
12. Gandhi, R., Chen, Z., Singh, N., Banerjee, K. & Lee, S. *IEEE Electron Device Lett.* **32**, 437–439 (2011).
13. Tomioka, K., Yoshimura, M. & Takashi, F. *IEEE VLSI Technol. Symp.* 47–48 (2012).
14. Sarkar, D. & Banerjee, K. *Appl. Phys. Lett.* **100**, 143108 (2012).

HUMAN GENOMICS

The end of the start for population sequencing

In the final phase of a seven-year project, the genomes of 2,504 people across five continental regions have been sequenced. The result is a compendium of in-depth data on variation in human populations. SEE ARTICLES P.68 & P.75

EWAN BIRNEY & NICOLE SORANZO

The human genome comprises three billion bases, of which millions differ between any two genomes from different people. The 1000 Genomes Project, which was set up in 2008, aimed to study this variation in at least this many people, and in doing so to provide a solid foundation on which to build an understanding of genetic variation in the human population. Early work from the pilot phase of the project provided preliminary data sets and gave the first details about methods for analysing these data. In this issue of *Nature*, the 1000 Genomes Project publishes its final two papers^{1,2}, which analyse 2,504 genomes from 26 populations, and provide the most comprehensive view of global human variation so far.

In the first paper (page 68), the 1000 Genomes Project Consortium¹ focuses on relatively simple, short variations that affect up to 500 bases. As expected, the vast majority of variants affect only one base. However, the paper also outlines small but complex changes that were not studied in previous analyses. In the second paper (page 75), Sudmant *et al.*² explore more-complex changes that affect larger portions of the chromosome. These structural variants, which can be up to 500,000 bases long, are analysed in much greater detail than had been possible previously, thanks to improvements in genome sequencing over the past decade.

Sudmant *et al.* show that structural variants are abundant in human genomes and arise from a series of evolutionary processes that are more complex than had been thought³. The authors compare the effect of these variants on gene expression with the effect of variants

in which one base is substituted for another, known as single nucleotide polymorphisms (SNPs). They report that the structural variants have a disproportionate impact on gene expression, given their relatively low numbers in the genome compared with SNPs.

Both studies make huge strides in increasing the accuracy and sensitivity of DNA sequencing, particularly in identifying mutations that result in the insertion or deletion of bases (known as indels) and so shift the position of each subsequent base up or down the sequence. This advance allowed the consortium to correlate the presence of indels at specific positions with the more-humdrum SNPs.

These correlative data sets have benefits for genome-wide association studies (GWAS), which compare genomes from large cohorts of people to identify variants that are associated with disease or other traits. The correlated data sets from the 1000 Genomes Project Consortium¹ allow researchers to infer a large complement of variants, including indels and structural variants, in panels of people for whom only a small subset of SNPs have been analysed, using partial sequencing techniques such as genotyping arrays. Because genotyping arrays are cheap, the ability to infer variation allows researchers to focus on increasing sample sizes — a crucial next step in improving our understanding of the genetics of disease. Furthermore, panels inferred in this way should enable the identification of disease-associated variants that occur at substantially lower frequencies than can be identified by GWAS alone (the consortium identified variants that are present in as little as some 0.5% of the population with European ancestry, whereas these rare variants are found only sparsely in direct genotyping studies). The



Figure 1 | Genetic variation in human populations. Two studies^{1,2} have sequenced the genomes of 2,504 people from 26 populations across 5 continental regions — the broadest swathe of populations examined so far.

use of such panels will be particularly effective when combined with other whole-genome-sequencing data sets, such as that of the UK10K Consortium⁴, which is studying the genetic code of 10,000 people in fine detail.

Sudmant and colleagues also find that some structural variants occur in genomic regions that have been previously associated with complex traits or disease. The mechanisms by which these variants underpin disease risk remain largely unexplored, but the authors' data set provides the starting point for further mechanistic studies. Thus, the new data are expected to facilitate future exploration of rare structural variants that have previously been largely inaccessible.

Another major advance on previous phases of the project is the broad sampling of populations (Fig. 1). Sequences have been obtained from people in five continental regions (East and South Asia, Europe, Africa and the Americas). As expected, given the sub-Saharan origin of modern humans, the 1000 Genomes Project Consortium finds that most of the world's variation between humans occurs in sub-Saharan populations. The papers' repository of variants thus provides a much richer view than previous, Euro-centric data sets⁵ as to what constitutes normal variation in humans, and will enable cost-effective genetic studies in sub-Saharan populations. Indeed, such studies are already under way, designed to make use of data from the 1000 Genomes Project. Understanding how genetic variation can differ between people from different continents also affects our understanding both of recent human evolution and of medicine. The latter is particularly pertinent in cosmopolitan cities, where clinicians are increasingly assessing people from many ethnic backgrounds.

The papers also sequence admixed

populations, in which two previously separate populations have become mixed — for example, African American populations, which have African, European and Native American genetic heritage. Sequencing admixed populations is important because, for instance, it can help us to understand genetic variation in populations for whom few genomes have been sequenced, such as Native Americans. There are many admixed populations worldwide, including African American, Afro-Caribbean, Hispanic and North African populations, and the current studies lay the foundation for analysing and using genetic information from these groups.

The current data sets and analyses have been openly released and have already been used in thousands of publications.

Consistent with the aims of the 1000 Genomes Project, the current data sets and analyses have been openly released and have already been used in thousands of publications, ensuring that they will have a lasting impact. It is to the credit of the project that the people who donated DNA consented to the full release of their genetic data with the understanding that no other associated information, for example about health problems, would be collected. However, such completely accessible genome data sets are likely to become a minority, because there is a growing shift towards genomic data sets that are clinically annotated and so cannot normally be freely distributed. National laws and ethical standards mean that these data sets sometimes come with complex restrictions, even for use in research.

In this new world of genome sharing, baseline genetic data on human populations will

still be needed, and such data will be much more useful when openly released. The International Genome Sample Resource⁶, which was created earlier this year, provides a coordination centre for open data sets. However, given that controlled-access data are likely to become the norm, strategies that make it easier to use and reuse clinical data sets will be hugely beneficial. The nascent Global Alliance for Genomics and Health⁷ provides an international framework for discussion about these complex issues.

The future of human population genetics is both rosy, with many more data being produced, and complex, with more-involved ethics and more-strictly controlled access likely to be required. The 1000 Genomes Project has delivered the data and the methodological foundation for this future. ■

Ewan Birney is at the European Molecular Biology Laboratory, European Bioinformatics Institute, Cambridge CB10 1SD, UK.

Nicole Soranzo is at the Wellcome Trust Sanger Institute, Hinxton CB10 1HH, UK, and at the School of Clinical Medicine, University of Cambridge, UK.
e-mails: birney@ebi.ac.uk; ns6@sanger.ac.uk

1. The 1000 Genomes Project Consortium. *Nature* **526**, 68–74 (2015).
2. Sudmant, P. H. *et al.* *Nature* **526**, 75–81 (2015).
3. Alkan, C., Coe, B. P. & Eichler, E. E. *Nature Rev. Genet.* **12**, 363–376 (2011).
4. UK10K Consortium. *Nature* <http://dx.doi.org/10.1038/nature14962> (2015).
5. The 1000 Genomes Project Consortium. *Nature* **491**, 56–65 (2012).
6. www.internationalgenome.org
7. www.genomicsandhealth.org

E.B. declares competing financial interests. See go.nature.com/vy2wgw for details.