living polymer to produce a block copolymer: molecules that contain long, uniform runs of different monomers.

Block copolymers have become major commercial successes — for example, the whole field of thermoplastic elastomers is based on this technology. Thermoplastic elastomers are rubbery solids that, unlike conventional rubbers, can be reused by heating them to temperatures above their glass transition temperature, remoulding them and then rapidly cooling them (the glass transition temperature is the range of temperatures in which amorphous materials pass from a liquid state to a hard, glassy substance). Apart from block copolymers, a dizzying number of other polymeric molecular structures engineered by living polymerization are also now available (Fig. 1). Szwarc received international recognition for the synthetic aspect of his work when he was awarded the Kyoto Prize for advanced technology in 1991.

A development that was greatly aided by the routine availability of polymers with a narrow molecular-weight distribution was the scaling theory that allows many polymer properties to be expressed in terms of molecular weight. For example, in 1950, Flory and Thomas Fox determined an equation[6] that accurately expressed the glass transition temperature as a function of molecular weight. The improved polystyrene samples available after 1956 confirmed this prediction[7].

A crucial property of pure liquid polymers is their viscosity. Flory and Fox discovered[8] that, for high-molecular-weight polymers, the viscosity increases in proportion to the molecular weight raised to the power of 3.4, and they proposed a theory to explain this finding. This means that, even well above the glass transition temperature, such polymers can have a high viscosity and behave like a soft solid. That may seem an obscure finding, but it has practical applications — such as the polymeric 'solvent' used in advanced batteries that do not leak. Again, Szwarc's discovery allowed Flory and Fox's theory to be validated.

One of the theoretically most challenging issues in scaling theory was the molecular-weight dependence of the osmotic pressure of polymer solutions. This is of interest because many industrial polymers are used in solution, and because biologists require an understanding of naturally occurring polymer solutions. The physicist Pierre-Gilles de Gennes correctly intuited[9] that, because linear polymer chains in solution are 'swollen' by the solvent, the osmotic pressure will have a different molecular-weight dependence from that predicted by classical theory. Measurements[10] of osmotic pressure for solutions encompassing wide ranges of concentration and molecular weight confirmed de Gennes' predictions. Both Flory and de Gennes received a Nobel prize for their work in polymer science and condensed-matter science, respectively.

Many theoretically challenging issues remain to be solved in polymer science, and the synergistic relationship between theory and the availability of well-defined polymer samples will greatly aid this effort. For instance, rubbery materials are widespread in industry and in biology, yet the theory of rubber elasticity is yet to be fully validated. The chemistry of living polymers also remains a highly active area[11], with hundreds of investigators worldwide. Many synthetic routes to living polymers have been developed, and a wide range of monomers can now be used in this approach. The concept of living polymers has truly revolutionized the practice of polymer science. ■

Gary Patterson *is in the Department of Chemistry, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, USA.*
e-mail: gp9a@andrew.cmu.edu

1. Matyjaszewski, K. & Müller, A. H. E. *Prog. Polym. Sci.* **31,** 1039–1040 (2006).
2. Szwarc, M. *Nature* **178,** 1168–1169 (1956).
3. Franta, E. *et al. J. Polym. Sci. A* **45,** 2576–2579 (2007).
4. Flory, P. J. *J. Am. Chem. Soc.* **62,** 1561–1565 (1940).
5. Szwarc, M. *Living Polymers and Mechanisms of Anionic Polymerization*; *Adv. Polym. Sci.* **49** (Springer, 1983).
6. Fox, T. G. & Flory, P. J. *J. Appl. Phys.* **21,** 581–591 (1950).
7. Ferry, J. D. *Viscoelastic Properties of Polymers* 3rd edn (Wiley, 1980).
8. Fox, T. G. Jr & Flory, P. J. *J. Phys. Chem.* **55,** 221–234 (1951).
9. de Gennes, P.-G. *Scaling Concepts in Polymer Physics* (Cornell Univ. Press, 1979).
10. Patterson, G. *Physical Chemistry of Macromolecules* Ch. 5 (CRC Press, 2007) .
11. Matyjaszewski, K. *Macromolecules* **45,** 4015–4039 (2012).

HUMAN GENOMICS

A deep dive into genetic variation

The exome is the portion of the genome that encodes proteins. Aggregation of 60,706 human exome sequences from 14 studies provides in-depth insight into genetic variation in humans. SEE ARTICLE P.285

JAY SHENDURE

Just seven years ago, my colleagues and I reported the protein-coding DNA sequences, called exomes, of 12 individuals[1] — among the first to be produced with a new generation of sequencing technologies[2]. Exome sequencing is much less expensive than whole-genome sequencing and, for cancers and Mendelian disorders (the latter caused by mutations in single genes), there is much more disease-associated genetic variation in the exome than in the rest of the genome. On page 285, the Exome Aggregation Consortium (ExAC) and collaborators[3] report the exome sequences of 60,706 individuals, collected from diverse studies: a venture 5,000 times larger than our initial study.

The current work highlights the pace at which human genetics is being scaled up. The project is almost ten times bigger than the Exome Sequencing Project (ESP) reported in 2013 (ref. 4), which was an important forerunner of ExAC. Indeed, this may be the deepest dive into the well of human genetic variation so far.

The study and accompanying database are noteworthy on several counts. First, for the sheer number of individuals sequenced and the depth of coverage — that is, the number of times each nucleotide in each individual's exome was sequenced. In the recently completed 1000 Genomes Project, 2,504 genomes were shallowly sequenced[5], a cost-saving strategy that favours the discovery of common over rare genetic variation. By contrast, each exome in ExAC has been sequenced deeply. Consequently, even genetic variants observed in just one individual can be confidently considered to be real (Fig. 1).

More than half of the approximately 7.5 million variants found by ExAC are seen only once. But collectively, they occur at a remarkably high density — at one out of every eight sites in the exome. For each gene, the authors contrasted the expected and observed numbers of variants that cause the production of truncated proteins, to search for regions containing lower-than-predicted levels of protein-truncating variants. This allowed them to identify several thousand genes that are highly sensitive to such variants — that is, unable to function normally after loss of one copy of the gene, even if the other copy is intact. Most of these genes have not yet been associated with disease, but mutation probably leads to embryonic death or strongly affects fitness in some other way. These genes are also intolerant of variants in regulatory DNA sequences that markedly alter levels of RNA synthesis from the gene[6], and are more likely than other genes to be implicated in genome-wide association studies of common disease.

The second noteworthy achievement of the research is that it provides a glimpse

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

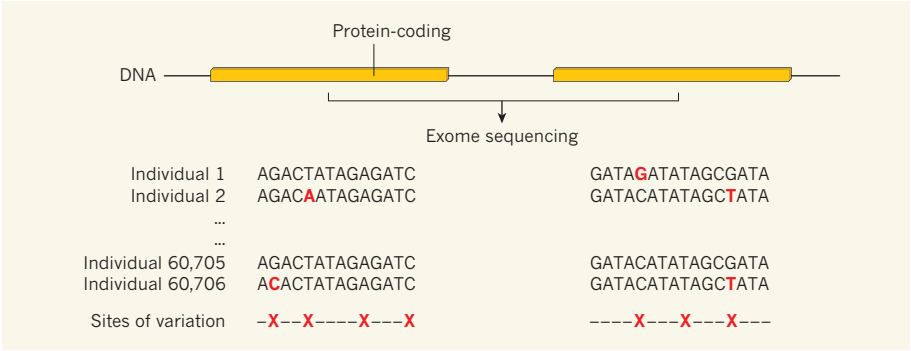

Figure 1 | Exome aggregation. The Exome Aggregation Consortium (ExAC)[3] reanalysed the raw DNA-sequencing data from the protein-coding part of the genome, known as the exome, of 60,706 individuals, aggregated from 14 distinct studies. Genetic variants (red) are compared to produce a database of all sites of variation between the individuals.

of the bottom of the well of genetic variation in humans. In human genetics, it is generally assumed that when the same variant is found in more than one individual, it arose once in an ancestor shared by those individuals, rather than through independent mutations of the same site. However, at a particular class of site, called CpG dinucleotides, the researchers make a convincing case that variants observed in multiple individuals often reflect mutational recurrence.

In support of their assertion, the researchers find that discovery rates for new CpG dinucleotide mutations decrease in samples larger than 20,000 individuals. This provides further evidence that the size of the ExAC cohort is sufficiently large that we are beginning to saturate this class of human genetic variation, at least within the exome. It is worth noting, however, that CpG dinucleotides have a highly elevated mutation rate in human genomes, making the number of samples needed to observe such saturation much lower than for other kinds of variants. Nonetheless, this exciting finding presages what lies ahead, as larger aggregate analyses of exomes and genomes are performed.

Third, ExAC promotes the discovery of genes involved in rare diseases. In 2009, my group and others showed how exome sequencing could be used to identify Mendelian-disease genes or to diagnose Mendelian disease[1,7,8]. Because there are tens of thousands of genetic variants in an exome, these strategies depended on effectively filtering out common variants, which are not likely to cause Mendelian disorders. At that time, databases of common variants were uneven and of suspect quality. Although ESP greatly improved the situation by uniformly and systematically cataloguing both common and rare variants across the exome[4], ExAC is an order of magnitude larger, and so enables better filtering. This is especially relevant for exome sequencing of non-European, non-African-American individuals, because ExAC provides greater sampling of individuals from outside the United States than ESP does.

On a related point, the study finds that hundreds of variants previously claimed to cause Mendelian disorders occur at implausibly high frequencies. As such, the authors suggest that they be reclassified as benign. A related study[9] shows how ExAC may also force a reassessment of whether some genes are involved at all in particular rare disorders. There is little doubt that ExAC will both refine and accelerate Mendelian-gene discovery and clinical genetics.

Finally, the consortium's approach to data aggregation and sharing is admirable. ExAC is both a technical and political achievement, requiring wrangling not only of data but also of investigators, consents and more from 14 studies — most of which were directed at the genetics of various common diseases.

An ongoing challenge in genomics is balancing the privacy rights of human participants with a strong tradition of promptly and openly sharing data. Building on the precedent of ESP, ExAC hits this balance by publicly releasing aggregate analyses —a catalogue of variants and the frequencies at which they arise — but not data about associated traits or other individual-level information (although raw data for many studies in ExAC is theoretically accessible through restricted databases). In this way, the study maximizes benefit while minimizing harm. These data have already been available on a terrifically intuitive website for nearly two years (http://exac.broadinstitute.org/), and the site has accrued more than 4 million page views.

If there is one take-home message, it is that there is incredible value in aggregating sequencing data across genomic studies. As the exomes aggregated by ExAC represent just a small fraction of the human samples that have been subjected to exome or genome sequencing so far, we can and should do better. In the coming decade, the number of human genomes that will be sequenced in some manner will grow to at least tens of millions and, by the end of this century, perhaps even billions. The beginnings of saturation seen here with CpG dinucleotides may eventually be observed deeply and at every site, providing a nucleotide-level footprint of the human genome. ■

Jay Shendure *is in the Department of Genome Sciences, University of Washington, Seattle, Washington 98195, USA, and is an investigator of the Howard Hughes Medical Institute.*
e-mail: shendure@uw.edu

1. Ng, S. B. *et al. Nature* **461,** 272–276 (2009).
2. Shendure, J. & Ji, H. *Nature Biotechnol.* **26,** 1135–1145 (2008).
3. Lek, M. *et al. Nature* **536,** 285–291 (2016).
4. Fu, W. *et al. Nature* **493,** 216–220 (2013).
5. The 1000 Genomes Project Consortium *Nature* **526,** 68–74 (2015).
6. The GTEx Consortium. *Science* **348,** 648–660 (2015).
7. Ng, S. B. *et al. Nature Genet.* **42,** 30–35 (2010).
8. Choi, M. *et al. Proc. Natl Acad. Sci. USA* **106,** 19096–19101 (2009).
9. Walsh, R. *et al. Genet. Med.* http://dx.doi.org/10.1038/GIM.2016.90 (2016).

NEUROSCIENCE

Flipping the sleep switch

Inactivation of a group of sleep-promoting neurons through dopamine signalling can cause acute or chronic wakefulness in flies, depending on changes in two different potassium-channel proteins. SEE LETTER P.333

STEPHANE DISSEL & PAUL J. SHAW

Many people have nodded off during a long road trip, or lain in bed desperately trying to fall asleep. These experiences illustrate real-world consequences of an improperly maintained balance between sleep- and wake-promoting neural circuits. On page 333, Pimentel *et al.*[1] describe the identification of a bona fide molecular switch that allows wake-promoting signals to turn off individual sleep-promoting neurons to regulate waking. These findings open up avenues for understanding the complexity of sleep regulation in healthy individuals and during disease.

Multiple sleep and wake circuits are found throughout the mammalian central nervous system and are believed to interact in a mutually inhibitory manner[2,3]. A similar organization

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.