increased fourfold. They used 3D scans of the lizard's skin and corrected for the curvature of its body and irregular surface texture to identify the centres of its scales. They then used a slight but sophisticated adjustment to map these points to the centres of hexagons; the resulting hexagonal array becomes a flattened 'tiling pattern', in which each tile represents one scale. The authors were able to track the skin scales because their number and relative position are maintained throughout the reptile's growth. The level of detail achieved by Manukyan *et al.* makes this study innovative in terms of providing empirical data with which to drive a theory of pattern formation.

At the macroscopic scale, pattern formation is usually described as a smooth process that is continuous in space and time, and it is modelled by a set of reaction–diffusion equations — mathematical equations that describe how chemicals redistribute over space and time. The authors follow this convention, but with a key difference: they show that the boundaries between *T. lepidus* scales constrict during morphogenesis, and argue that this creates partial barriers to the diffusion of cells and chemicals between adjoining scales. As a result, the scales form discrete spatial units that each take on a uniform colour (black or green on the lizard's back) in a way that depends on the states of their neighbours. Formally, then, the biological pattern-forming system resembles the output of a cellular automaton. The scale pattern evolves by obeying a set of rules that transform one configuration of scale colours into another, with certain probabilities. It is in this sense that the authors describe *T. lepidus* as a 'living' cellular automaton.

The authors determined empirically the probabilities of scales changing colour for distinct colour configurations of scales and their nearest neighbours. They then linked the reaction–diffusion and cellular-automaton approaches in a theoretical model. They found good agreement between the patterns that evolve over years on their reptilian subjects and the solutions to their model based on neighbourhood-dependent rules obtained from empirical data. This agreement is surprising and adds to the novelty of the authors' approach.

What underlying mechanism drives this pattern formation? According to the authors, a system consisting of pigment cells (melanophores and xanthophores) interacting with a long-range, rapidly diffusing chemical suffices to explain the pattern formation — taking into account the partial diffusion barriers between adjoining scales that form during morphogenesis. The authors' model, which is modified from a pre-existing zebrafish pigmentation model[13], reproduces the black and green labyrinthine pattern of the adult lizard's skin. Future work in which the chemicals and cellular interactions are identified in more detail (possibly in related but simpler *in vitro* experimental systems, such as cells and chemicals interacting in a tissue culture) would provide an opportunity for manipulating the patterns experimentally, and therefore allow us to learn more about the underlying cellular and molecular pattern-forming mechanisms.

As the authors conclude, a cellular automaton is not just an abstract concept, but corresponds to a process generated by biological evolution. Nearly 80 years after its conception, the cellular automaton has come of age — it has matured from an abstract concept in the 1940s to *in silico* realizations since the 1960s and, finally, to a pattern-forming mechanism that has biological relevance and can be observed *in reptilio*. ■

Leah Edelstein-Keshet *is in the Department of Mathematics, University of British Columbia, Vancouver, British Columbia, V6T 1Z2, Canada.*
e-mail: keshet@math.ubc.ca

1. Murray, J. D. *Sci. Am.* **258,** 80–87 (1988).
2. Manukyan, L., Montandon, S. A., Fofonjka, A., Smirnov, S. & Milinkovitch, M. C. *Nature* **544,** 173–179 (2017).
3. Ulam, S. *Bull. Am. Math. Soc.* **64,** 1–49 (1958).
4. Shannon, C. E. *Bull. Am. Math. Soc.* **64,** 123–129 (1958).
5. Cook, M. *Complex Syst.* **15,** 1–40 (2004).
6. Wolfram, S. *Nature* **311,** 419–424 (1984).
7. Gardner, M. *Sci. Am.* **224,** 112–117 (1971).
8. Ermentrout, G. B. & Edelstein-Keshet, L. *J. Theor. Biol.* **160,** 97–133 (1993).
9. Turing, A. M. *Phil. Trans. R. Soc. Lond. B* **237,** 37–72 (1952).
10. Gierer, A. & Meinhardt, H. *Biol. Cybern.* **12**(1), 30–39 (1972).
11. Keller, P. J., Schmidt, A. D., Wittbrodt, J. & Stelzer, E. H. K. *Science* **322,** 1065–1069 (2008).
12. Keller, P. J. *Science* **340,** 1234168 (2013).
13. Nakamasu, A., Takahashi, G., Kanbe, A. & Kondo, S. *Proc. Natl Acad. Sci. USA* **106,** 8429–8434 (2009).

BIOMEDICINE

Human genes lost and their functions found

Individuals who lack a functional copy of a gene — gene knockouts — can reveal the gene's role. Most knockout research has used model organisms, but now a comprehensive catalogue of human knockouts is in sight. SEE LETTER P.235

ROBERT M. PLENGE

In the nineteenth century, Charles Darwin and Gregor Mendel studied natural genetic variation. But by the twentieth century, scientists did not have to rely on natural variation to investigate gene function because they could delete genes in genetically tractable model organisms such as the fruit fly *Drosophila melanogaster* and inbred mouse strains. Since then, research using model organisms has provided many fundamental biological insights. On page 235, Saleheen *et al.*[1] describe approaches to identifying and studying people who lack functional copies of specific genes and who are therefore natural human 'knockouts' for those genes.

Many sophisticated tools are available for investigating human biology. Genomic engineering of human cells can be performed *in vitro* using the gene-editing approach known as CRISPR–Cas9. The biology of complex tissues can also be studied *in vitro* using populations of human cells. In biobanks, clinical samples can be linked to data from electronic health records, enabling investigation of the relationship between genetics and disease. Saleheen and colleagues' work adds to the growing set of experimental resources for understanding human biology.

The authors studied mutations — known as nonsense, frameshift and splice-site mutations — that cause a copy of a gene to be non-functional because it encodes a highly abnormal or truncated protein. As a result of selective evolutionary pressure, these mutations are extremely rare in human populations[2], or if present, usually occur on only one of the two parental chromosomes; such mutations are said to be heterozygous. The functional copy of the gene on the other parental chromosome is often sufficient to fulfil the gene's normal function. If this is the case, an individual who is heterozygous for the gene is unaffected by the mutant copy.

When, as happens occasionally, both inherited parental chromosomes contain non-functional copies of a specific gene, this results in a gene knockout called a homozygous null mutation. If a gene has a key role in human physiology, its loss might cause disease or abnormal function. But if the gene is not essential, an individual lacking it will have normal function. Knowing the effects of not having a functioning copy of a specific gene provides useful information about the gene's role.

The frequency with which homozygous null mutations occur depends on the relatedness of an individual's parents. In outbred populations, where the genetic relationship between the two parents will be slight, it is unlikely that both will have a mutation in the same gene.

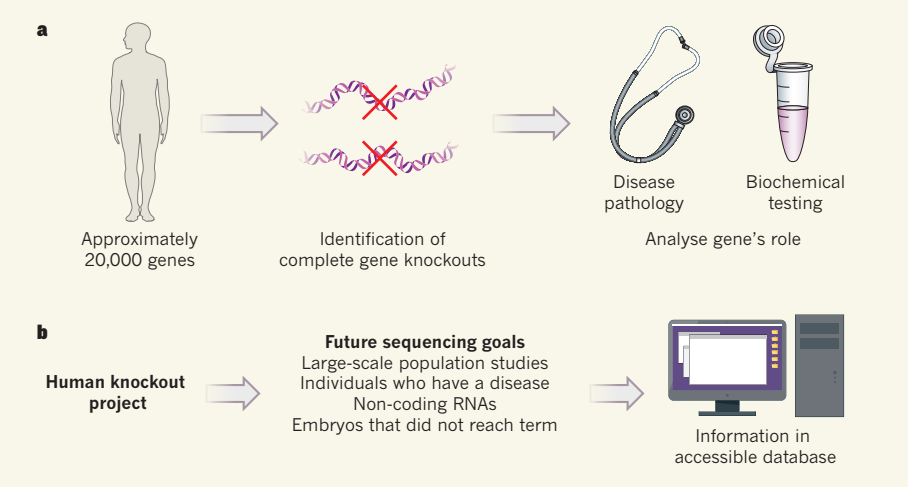

Figure 1 | A road map for identifying and studying the absence of human genes in individuals. **a**, A human has approximately 20,000 genes. One way of investigating the role of a gene is to use DNA sequencing to identify individuals who lack a functional copy of that gene, known as gene knockouts. Saleheen *et al.*[1] identified knockouts of 1,317 human genes. For some gene knockouts, the authors analysed biochemical and clinical data to try to determine the gene's role. **b**, A project to identify knockouts in all human genes is now within reach. In addition to other large-scale population studies similar to that carried out by Saleheen and colleagues, future sequencing goals for the project should include targeting individuals who have certain diseases, or investigating non-protein-coding genes. Identification of genes that are essential for life might require sequencing the genomes of embryos that do not develop to term. All of these data should be assembled into an easily accessible database.

Outbred offspring are therefore unlikely to inherit two mutant copies of the same gene. By contrast, closely related parents, such as first cousins, have a higher probability of carrying mutations for the same gene, and therefore their offspring will be more likely to inherit two mutant copies of it.

Saleheen *et al.* analysed 10,503 adult participants in the Pakistan Risk of Myocardial Infarction Study (PROMIS)[3]. Approximately 40% of the members of this group married their first cousins. The rate of inbreeding in the PROMIS group is nearly four times that of outbred populations of European or African-American ancestry, as measured by analysing identical DNA stretches in both parental copies of an individual's chromosomes. To identify gene knockouts, the authors sequenced the protein-coding regions (the exome) of an individual's genome; they identified 1,843 individuals who carried at least 1 gene that was a predicted homozygous-knockout mutation. The researchers found 1,317 different gene knockouts, representing around 7% of the known protein-coding genes.

Saleheen and colleagues' work adds to other large-scale human-knockout investigations, including a study of 3,222 closely related British adults of Pakistani heritage, which identified 781 gene knockouts[4], and a study of 2,636 Icelanders, which found 1,171 knockouts[5]. These other studies linked genetic data with health records to investigate the relationship between gene knockouts and clinical characteristics related to rare inherited disorders such as Ehlers–Danlos syndrome[4] or more-common conditions such as Alzheimer's disease[6]. A third large-scale study was carried out by the Exome Aggregation Consortium (ExAC), which identified 3,230 predicted gene knockouts by exome sequencing of 60,706 individuals of diverse ancestries[2].

In a subset of PROMIS participants, Saleheen and colleagues also linked gene knockouts with clinical and physiological parameters by measuring more than 200 biochemical and disease traits, and analysing the blood levels of 1,310 proteins. One compelling example of this work was their investigation of the *PLA2G7* gene, which encodes the enzyme Lp-PLA2. Previous epidemiological studies indicated that a lower activity and amount of Lp-PLA2 protects against cardiovascular disease[7]. Unexpectedly, Saleheen *et al.* found no obvious clinical consequence in individuals who had 50% or 100% reduction in normal Lp-PLA2 activity, and no association between lower Lp-PLA2 activity and protection against cardiovascular disease.

PROMIS participants gave their consent to being contacted again if researchers wished to invite them to take part in more-detailed studies. Saleheen *et al.* contacted individuals who lacked the *APOC3* gene. This gene encodes the protein apolipoprotein C3, a component of a low-density lipoprotein complex that carries fats around the body, and which has been linked to increased risk of cardiovascular disease[8,9]. Compared with people who had normal apolipoprotein C3 function, the predicted *APOC3*-homozygote knockouts had significantly lower levels of triglyceride fats in their bloodstream after a meal. This finding supports the hypothesis that pharmacological lowering of apolipoprotein C3 will protect against cardiovascular disease. Moreover, complete pharmacological inhibition of apolipoprotein C3, which would mimic an *APOC3* knockout, should probably be safe in humans.

The authors briefly discussed the development of a road map for a human-knockout project — a systematic effort to understand the consequences of complete disruption of every human gene. They estimate that sequencing the genomes of approximately 200,000 individuals from the PROMIS cohort would identify homozygous knockouts for around 40% of the protein-coding genes, five times more knockouts than it is estimated would be found by sequencing an outbred population of the same size. Given the availability of large population-based biobanks and the continued decrease in sequencing costs, it seems probable that large biological databases of human knockouts will be available within the next five years. Integrated and widely accessible databases must be assembled to catalogue the knockouts being gathered and to document the consequences of these mutations. At the same time, these databases must protect the privacy of individuals whose DNA has been sequenced.

In this road-map project, sequencing the genomes of individuals who have inherited diseases should further aid the identification of knockouts. Sequencing goals should also include the identification of knockouts for non-protein-coding genes, such as long, non-coding RNAs and microRNAs, because such mutations can cause disease. Obviously, for any genes whose loss causes embryonic lethality, there won't be a person with that knockout. However, sequencing human embryos that have not survived to term would be a possible way of identifying such genes (Fig. 1).

As the human genetic toolbox expands, so will acceptance that humans can represent an ideal model organism for discovery research, within ethical limits. Conventional model organisms will still be needed, but the expansion of resources such as those described by Saleheen *et al.* will change the nature of the scientific investigation of the genetic basis of human disease. ■

Robert M. Plenge *is in the Translational Medicine Department, Merck, Boston, Massachusetts 02115, USA.*
e-mail: robert.plenge@gmail.com

1. Saleheen, D. *et al. Nature* **544,** 235–239 (2017).
2. Lek, M. *et al.* Nature **536,** 285–291 (2016).
3. Saleheen, D. *et al. Eur. J. Epidemiol.* **24,** 329–338 (2009).
4. Narasimhan, V. M. *et al. Science* **352,** 474–477 (2016).
5. Sulem, P. *et al. Nature Genet.* **47,** 448–452 (2015).
6. Steinberg, S. *et al. Nature Genet.* **47,** 445–447 (2015).
7. The Lp-PLA$_2$ Studies Collaboration. *Lancet* **375,** 1536–1544 (2010).
8. The TG and HDL Working Group of the Exome Sequencing Project. *N. Engl. J. Med.* **371,** 22–31 (2014).
9. Jørgensen, A. B., Frikke-Schmidt, R., Nordestgaard, B. G. & Tybjærg-Hansen, A. *N. Engl. J. Med.* **371,** 32–41 (2014).

The author declares competing financial interests.
See go.nature.com/2p1srjx for details.