today[7]. This period is advantageous for the authors' purposes in that the seasonal difference of incoming solar radiation (insolation) was greater than during the Holocene, whereas the effects of other factors that alter climate, such as greenhouse gases and ice, were weaker, making it easier to identify seasonal biases.

More specifically, the authors' method involves identifying seasonal bias in the portion of an SST record that corresponds to the last interglacial, and in which there was a stronger correlation of SST with seasonal insolation than with mean annual insolation. The sensitivity of the SST record to seasonal insolation during this period is then calculated, and used as a benchmark to remove seasonal bias from the entire record, thereby allowing mean annual SST to be determined from that record. The authors first applied their method to an SST reconstruction based on a proxy taken from a marine site located off the northeast coast of Papua New Guinea. The transformed mean annual SST record was independently validated by applying the new method to SST data for the same geographical region produced in computational simulations of the past 300,000 years — the transformed data matched the mean annual SST output from the simulations.

Bova *et al*. went on to create a synthesis of previously published SST records spanning the last interglacial and the Holocene periods. These records are based on two common proxies used for reconstructing SST: the chemical composition of the fossilized calcium carbonate shells of surface-dwelling unicellular marine organisms known as foraminifera; and organic biomarkers known as alkenones, which are synthesized by marine algae and settle into marine sediments. The authors found that the majority of the examined SST records are indeed seasonally biased.

After converting the seasonally biased SST records into mean annual SST records, Bova and colleagues infer that the climate has been warming since the early Holocene — that is, there is no evidence for a Holocene thermal maximum in mean annual temperatures (Fig. 1). They suggest instead that the Holocene thermal maximum is a seasonal feature driven by a peak in summer insolation in the Northern Hemisphere that occurred during the early Holocene.

The reconstruction of mean annual temperatures produced from the authors' synthesis of proxy records strongly resembles a computational simulation[3] of Holocene climate that also reflects mean annual temperatures — effectively solving the Holocene temperature conundrum. This enabled Bova and colleagues to shed new light on the drivers of Holocene climate change. They find that the increase in global mean annual temperatures that occurred during the early Holocene (12,000–6,500 years ago) was a response to retreating ice sheets, whereas the continued increase in temperatures over the past 6,500 years is due to rising greenhouse-gas concentrations.

The authors also show that mean annual temperatures during the last interglacial period were more stable and higher than their estimates of Holocene temperatures. They attribute this to the near-constant greenhouse-gas concentrations and reduced extent of ice sheets during the last interglacial. Crucially, the researchers find that the current mean annual temperature exceeds those of the past 12,000 years, and probably approaches the warmth of the last interglacial period.

Bova and colleagues' method to identify and correct for seasonal biases in proxy SST reconstructions can now be applied to other temperature records on different timescales. This is a key benefit of their study, because palaeoclimate scientists have long known that temperature reconstructions are probably seasonally biased, but did not have a method for addressing the problem.

One limitation of the findings is that the new synthesis of proxy SST records is limited to the global region between 40°N and 40°S. Proxy records from higher latitudes were deliberately excluded because of the scarcity of such records for the last interglacial, and because of the proximity of those regions to ocean fronts, where ocean dynamics can affect SST. However, the inclusion of these regions might be needed in the future, given that processes at high latitudes have a substantial role in many climate feedback processes. Moreover, the new synthesis examines records based on only two SST proxies. Future work should include more records based on other temperature proxies. Nevertheless, by solving a conundrum that has puzzled climate scientists for years, Bova and colleagues' study is a major step forward for the field.

Jennifer Hertzberg is in the International Ocean Discovery Program, Texas A&M University, College Station, Texas 77845, USA.
e-mail: hertzberg@iodp.tamu.edu

1. Kaufman, D. *et al*. *Sci. Data* **7**, 201 (2020).
2. Marcott, S. A., Shakun, J. D., Clark, P. U. & Mix, A. C. *Science* **339**, 1198–1201 (2013).
3. Liu, Z. *et al*. *Proc. Natl Acad. Sci. USA* **111**, E3501–E3505 (2014).
4. Bova, S., Rosenthal, Y., Liu, Z., Godad, S. P. & Yan, M. *Nature* **589**, 548–553 (2021).
5. Marsicek, J., Shuman, B. N., Bartlein, P. J., Shafer, S. L. & Brewer, S. *Nature* **554**, 92–96 (2018).
6. Liu, Y. *et al*. *Sci. Rep.* **8**, 4434 (2018).
7. Hoffman, J. S., Clark, P. U., Parnell, A. C. & He, F. *Science* **355**, 276–279 (2017).

Biotechnology

Base editor repairs gene of premature-ageing disease

Wilbert P. Vermeij & Jan H. J. Hoeijmakers

No cure exists for the lethal premature-ageing condition Hutchinson–Gilford progeria. A gene-editing tool — adenine base editors — offers a way to treat the condition in mice. Could this approach lead to an effective therapy? **See p.608**

Gene-editing technologies raise the possibility of tackling the fundamental cause of certain inherited human diseases. On page 608, Koblan *et al*.[1] report their use of such a technology in mice that provide a model for a human accelerated-ageing disorder.

Ageing is influenced by numerous factors — some external, some organ-specific and others systemic, affecting the entire body. It is one of the main biological processes for which the chief cause or causes are not fully known. Often, the mechanisms underlying a biological process are revealed by an analysis of genetic mutants, and mutations in systemic ageing factors are associated with processes of accelerated ageing that affect multiple organs. Most premature-ageing disorders point to problems in DNA maintenance and integrity as the underlying cause. People with Werner or Cockayne syndrome, for example, have defects in different mechanisms that affect genome stability[2].

Generally, premature-ageing conditions exhibit accelerated ageing of a subset of tissues. The best known of such disorders is Hutchinson–Gilford progeria syndrome, which is often referred to just as progeria. Children with this condition look healthy at birth. But, from around one year of age, symptoms begin to emerge, such as growth failure, skin abnormalities and hearing loss. The features of premature ageing increase over time, resulting in striking hallmarks of old age that include wrinkles, loss of fat under the skin, joint stiffness and musculoskeletal abnormalities. However, these children retain a normally functioning

nervous system, underscoring the syndrome's organ-specific nature. No cure exists for progeria, and affected individuals usually die aged around 14 or 15 as a consequence of conditions such as atherosclerosis, severe cardiovascular complications or stroke[3].

Progeria is caused[4] by the mutation of a single base (a cytosine mutated to a thymine) in one of the two copies of the gene encoding the protein lamin A (Fig. 1). This protein is a structural component of the outer rim of the cell nucleus. The mutated version of the gene leads to abnormalities during the splicing process that occurs during gene transcription. As a result, a shorter-than-normal version of lamin A, termed progerin, is produced. Progerin and maturing lamin A undergo a modification termed farnesylation, in which a specific lipid group called farnesyl is attached to the protein. As lamin A matures, this lipid-modified region of the protein is removed in an enzyme-mediated cleavage event. However, progerin remains farnesylated because it lacks the amino-acid residues that provide the usual cleavage site. Farnesylated progerin accumulates and hampers the normal role of lamin A, thereby perturbing nuclear shape, rigidity and function.

Nuclear malformations arising from progerin are particularly apparent in organs and tissues that are subject to mechanical stress, including the skin and the cardiovascular system. Parts of the body that are subject to high mechanical stress correspond to the most-affected organs in progeria, and the nuclear deformations that arise from stress activate the DNA-damage response[5,6]. Progerin-induced nuclear weakening might result in genome instability under mechanical stress if various delicate processes in the nucleus are disturbed. This would be consistent with the higher-than-normal levels of DNA damage and chromosome aberrations found in cells from people who have progeria, and in mice that provide a model for the syndrome[7], ranking progeria among other premature-ageing disorders associated with genomic instability.

Attempts to find treatments for progeria initially focused on trying to reduce the accumulation of farnesylated progerin[8]. Compounds that generated interest included those with reported anti-ageing activities, such as metformin and rapamycin — which might affect splicing and turnover of progerin, respectively. The most-advanced drugs, in terms of clinical use, are inhibitors of the enzyme farnesyl transferase, which reduce the accumulation of farnesylated progerin. One of these, lonafarnib, was approved in November 2020 by the US Food and Drug Administration. This is the first licensed therapy for progeria. However, the treatment only partially alleviates the syndrome[8].

The most rigorous approach for tackling progeria would be to directly correct the genetic defect. In 2019, two teams reported using the gene-editing method CRISPR–Cas9 to repair the associated mutation in the gene encoding lamin A in mice[9,10]. This treatment alleviated the decline in health normally expected in such animals and extended their lifespan, compared with those whose mutation was not corrected. CRISPR–Cas9 targets a specific genome sequence through a process aided by a guide RNA sequence that helps to ensure that editing occurs at the desired location. However, unwanted alterations could arise — either from the formation of the double-stranded DNA breaks that occur during this editing process, or from off-target edits of sequences that are similar to the sequence of interest. The possibility of such unwanted events would demand extreme caution in any potential clinical roll-out.

Koblan *et al*. present their use of an editing approach that could offer a way forwards. The authors harnessed tools known as base editors[11]. Like CRISPR–Cas9, these can alter a single base at a specific genomic location, such as where the mutant thymine is paired with adenine in the gene encoding lamin A. However, one difference is that base editors do not cleave DNA's phosphate backbone when targeting the nucleotide, so double-stranded DNA breaks are not generated. Instead, the approach chemically modifies the targeted nucleotide[12]. Adenine base editors convert adenine, through an intermediate called inosine, to guanine during DNA replication. The need for DNA replication could pose a problem if trying to use this method to target mutations in other diseases in which non-dividing cells, such as those of the nervous system, are the main target of interest.

Koblan and colleagues used an adenine base editor, transferred into cells by means of a virus called a lentivirus, to target the mutation in the gene encoding lamin A in cells from people with progeria. Repair occurred in 90% of all cells. Correction of the mutation resulted in the normal splicing of lamin A, reduced the expression of progerin and corrected abnormalities in nuclear shape. Minimal off-target editing was observed.

Using a mouse model of progeria, the authors delivered the base editor using adeno-associated virus. Following a single injection of the editor near the eye socket, or in the abdominal cavity, of mice that were up to two weeks old, the authors observed targeted repair of the gene encoding lamin A in many organs. This took place mainly in the liver and heart, but also occurred, to a lesser extent, in muscle, in the aortic artery and in bone tissue.

Although most of the key organs affected in progeria showed only a modest level of genetic correction and reduction in progerin,

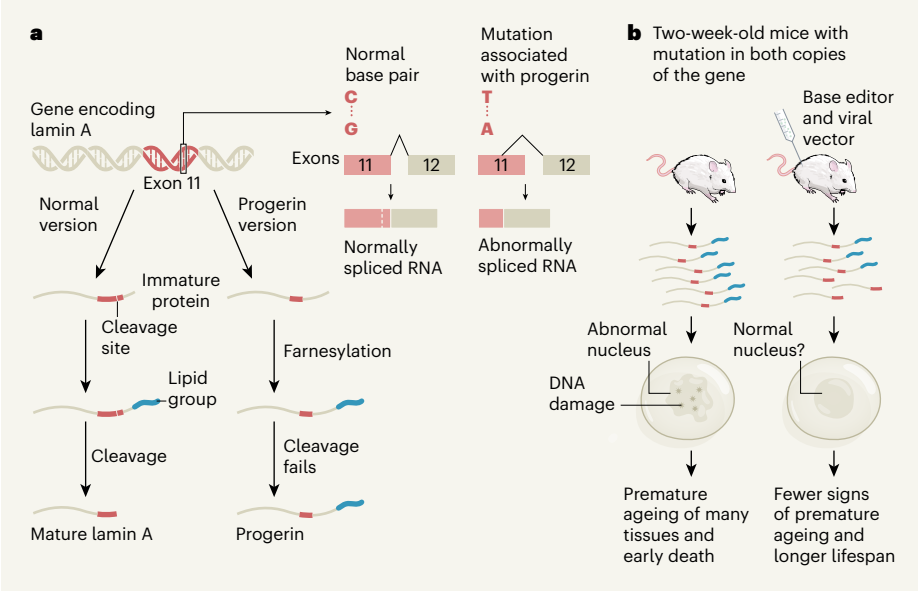

Figure 1 | Repair of a mutation that causes a premature-ageing syndrome. a, The lethal disease Hutchinson–Gilford progeria syndrome, also known as progeria, arises from a mutation in the gene that encodes the protein lamin A. This mutation, in a region of the gene called exon 11, changes a cytosine (C) base that is paired with a guanine (G) to a thymine (T) paired with adenine (A). This causes an abnormal splicing event in which a section of exon 11, including a region that encodes a cleavage site (dotted white line), is absent from the RNA sequence that encodes protein. As lamin A matures, a lipid group is added in an event called farnesylation. Cleavage of the protein at the site encoded by exon 11 produces mature lamin A. The mutated version of the protein, termed progerin, lacks this cleavage site and retains the lipid group. Progerin causes nuclear abnormalities and DNA damage, resulting in accelerated ageing[5,6]. **b**, Mice with two mutant copies of the gene encoding lamin A are a model for human progeria. Koblan *et al*.[1] used gene-editing tools called base editors. When the mice were given a base editor and a viral vector to enable the editing machinery to enter cells, the mutation was corrected in many, but not all, tissues. The reduction in progerin improved the health and lifespan of treated mice, and presumably corrected nuclear abnormalities in their cells.

a striking increase in the level of lamin A was observed as a consequence of the treatment. Crucially, compared with the model animals that did not undergo gene editing, those that received the base editor aged with remarkably fewer abnormalities in the usually lifespan-limiting cardiovascular system. These animals also had greater vitality (a better ability to move and a better overall appearance) and a statistically significant lifespan extension.

By directly addressing the root cause of the disease, base editing could offer great advantages over current drug-based therapeutic strategies. Many key questions remain to be answered, however, before people might benefit from the introduction of this technology. For example, what is the optimal distribution of base editor mediated by adeno-associated virus or by other delivery methods? And which organs can be targeted? The adeno-associated virus injection strategy was less efficient in targeting the skin than in targeting other mouse organs.

To what extent can the genetic defect be corrected? High efficiency of editing might be crucial, particularly for efforts to treat other diseases. Previous attempts[11,12] to use gene editing to address the defects underlying Duchenne muscular dystrophy and rare liver diseases met with only limited success. However, Koblan and colleagues' work indicates that correction does not need to reach 100% efficiency to provide positive benefits, opening the possibility of reconsidering this approach for some other diseases, too.

Another important question is whether an immune response might develop that would target components of the editing system. Such a response might result in inefficient treatment if cells harbouring editing components were selectively eliminated[11].

What about long-term considerations? For example, would a single administration of the editors be sufficient? And what would be the best age for treatment to be administered? Progeria is diagnosed relatively early in life compared with many other diseases for which base editing is a possibility. A treatment age of two weeks, for mice, is therefore much lower than the equivalent age, for humans, at which many diseases are diagnosed. Moreover, with animal testing, it is difficult to benchmark the equivalent human age that corresponds to a mouse of a given age. Finally, how would the current drug therapy available for progeria fit with the potential of repair by base editing in a treatment plan?

If base editing is to be used to treat human disease, the safety of such an intervention must be ensured. If it can be, and if this method successfully repairs the progeria-causing alteration in the crucial tissues, such an approach holds tremendous promise as a way of prolonging health, extending lifespan and improving the quality of life of those who have this mutation.

Wilbert P. Vermeij is at the Princess Máxima Center for Pediatric Oncology, Oncode Institute, 3584 CS Utrecht, the Netherlands. **Jan H. J. Hoeijmakers** is in the Molecular Genetics Department, Erasmus MC, 3015 GD Rotterdam, the Netherlands, and also at the Princess Máxima Center for Pediatric Oncology, Oncode Institute, and at the Institute for Genome Stability in Ageing and Disease, CECAD Research Centre, Cologne, Germany.
e-mails: w.p.vermeij@prinsesmaximacentrum.nl; j.hoeijmakers@erasmusmc.nl

1. Koblan, L. W. *et al. Nature* **589**, 608–614 (2021).
2. Niedernhofer, L. J. *et al. Annu. Rev. Biochem.* **87**, 295–322 (2018).
3. Merideth, M. A. *et al. N. Engl. J. Med.* **358**, 592–604 (2008).
4. Eriksson, M. *et al. Nature* **423**, 293–298 (2003).
5. Kumar, A. *et al. Cell* **158**, 633–646 (2014).
6. Nava, M. M. *et al. Cell* **181**, 800–817 (2020).
7. Liu, B. *et al. Nature Med.* **11**, 780–785 (2005).
8. Harhouri, K. *et al. Nucleus* **9**, 265–276 (2018).
9. Beyret, E. *et al. Nature Med.* **25**, 419–422 (2019).
10. Santiago-Fernández, O. *et al. Nature Med.* **25**, 423–426 (2019).
11. Porto, E. M., Komor, A. C., Slaymaker, I. M. & Yeo, G. W. *Nature Rev. Drug. Discov.* **19**, 839–859 (2020).
12. Katrekar, D. *et al. Nature Methods* **16**, 239–242 (2019).

This article was published online on 6 January 2021.

Computational materials science

Machine learning from diverse data sources

Rohit Batra

A strategy for machine learning has been developed that exploits the fact that data are often collected in different ways with varying levels of accuracy. The approach was used to build a model that predicts a key property of materials.

Scientists are always hunting for materials that have superior properties. They therefore continually synthesize, characterize and measure the properties of new materials using a range of experimental techniques. Computational modelling is also used to estimate the properties of materials. However, there is usually a trade-off between the cost of the experiments (or simulations) and the accuracy of the measurements (or estimates), which has limited the number of materials that can be tested rigorously. Writing in *Nature Computational Science*, Chen *et al.*[1] report a machine-learning approach that combines data from multiple sources of measurements and simulations, all of which have different levels of approximation, to learn and predict materials' properties. Their method allows the construction of a more general and accurate model of such properties than was previously possible, thereby facilitating the screening of promising material candidates.

Materials scientists commonly supplement their own 'chemical intuition' with predictions from machine-learning models, to decide which experiments to conduct next[2,3]. For example, artificial intelligence has been used to identify candidate compounds that act as superconductors at high temperatures[4], electrolyte materials that conduct electric currents using lithium ions[5], and electrically insulating polymers that can withstand large electric fields without breaking down[6]. Artificial intelligence has also been used to work out ways of synthesizing materials — that is, to suggest which reagents, catalysts and experimental conditions to use[7].

Most of these studies involve supervised learning, in which a machine is exposed to a large volume of historical data about the chemical composition or atomic structure of materials, and their associated properties, to build a model that can predict the properties of other materials. More importantly, almost all of these studies use models built on data obtained from a single, consistent source. Such models are referred to as single-fidelity models.

However, for most real-world applications, measurements of materials' properties have varying levels of fidelity, depending on the resources available. For instance, the most accurate (high-fidelity) measurements of properties of crystalline materials are made using single crystals, which can be laborious to prepare. Approximate (low-fidelity) measurements are therefore often made using easily synthesizable polycrystalline samples. Similarly, a hierarchy of increasingly accurate, but progressively more expensive, computational modelling schemes are used to calculate materials' properties — high-fidelity modelling is therefore often restricted by its cost.

Overall, this variation in measurement and modelling techniques leads to a heterogeneous data structure — low-fidelity measurements are plentiful, whereas high-fidelity data are sparse. Notably, each type of fidelity has its own advantages: low-fidelity