The international journal of science / 11 February 2021

nature

# Twenty years of the human genome

**To fulfil the promises of the Human Genome Project, researchers, journals and funders must re-commit to equity and open data sharing.**

The first drafts of the human genome, published in *Nature* and *Science* 20 years ago, flung open the doors for what some predicted would be 'biology's century'. In just one-fifth of the century, the corpus of information has grown from two gappy and error-filled genome sequences to a full account of the genetic variation of hundreds of thousands of individuals around the world, and an increasing number of tools to study it (see page 212). This special issue of *Nature* examines how far the human genome sequence has taken us, and how far we have to go. But some aspects of the research ecosystem around the human genome have hardly changed, and that remains a concern.

Many of the ethical, legal and social implications of genome research – including questions of privacy, informed consent and equitable representation of researchers and participants – remain unresolved. Moreover, free and open access to genome data remains unevenly implemented. Just this week, researchers pointed out the problems caused by lack of accessibility to coronavirus genomes in the middle of a pandemic (see page 195). Researchers, funders and journals will need to address these issues if they are to fulfil the promises of the Human Genome Project and to better understand diseases and improve diagnoses and treatments.

The draft genome sequence published in *Nature* was immediately free to access – in fact, the initial assembly was posted online some seven months beforehand. This was in accordance with the Bermuda Principles, an agreement on data sharing signed by members of the international consortium that made the Human Genome Project possible.

*Nature* committed to open-data principles for genomics research back in 1996. By publishing the Human Genome Project's first paper, we worked with a publicly funded initiative that was committed to data sharing. But the journal acknowledged there would be challenges to maintaining the free, open flow of information, and that the research community might need to make compromises to these principles, for example when the data came from private companies. Indeed, in 2001, colleagues at *Science* negotiated publishing the draft genome generated by Celera Corporation in Rockville, Maryland. The research paper was immediately free to access, but there were some restrictions on access to the full data.

Twenty years later, compromises and delays are becoming the norm in three domains of genome research: data collection from participants; deposition in approved, publicly accessible databases; and access for research and health care. The promise of a fully open data-sharing environment has not yet been realized.

> "Rapid data sharing can provide massive benefits to science and, through science, to all of society."

For genomics to truly revolutionize medicine, it needs to be combined with phenotypic data – physical characteristics, medical histories and other identifiable traits that can be linked to variants in the genome. But collecting such data increases privacy risks for research participants, who are now rightly being given more control, such as choosing how their data will be used. Moreover, scientists involved need to be vetted to ensure that participants have given the appropriate consent and that their interests are protected.

The next step is to deposit the collected genome sequences and the accompanying data into approved international databases that can continue to protect those interests. But researchers regularly report being unable to deposit their data quickly, citing privacy and consent concerns, or agreements with companies that have contributed data. Technological limitations mean that the process of depositing data can also be extremely time-consuming. Scientists are producing increasing volumes of ever more complex data – and this is overwhelming under-resourced repositories.

Finally, researchers struggle to track down data that should be available as soon as the accompanying research is published. And even after locating the data, they can find it hard to access them (see page 198).

## Diversity deficit

In the years since the Human Genome Project published its first draft sequence, researchers have recognized that genome databases over-represent DNA from people of European descent who live in high-income nations.

Truly global databases and repositories need data that properly represent humanity's vast genetic diversity. That this has not been achieved in two decades is a reminder of science's history of mistreatment and neglect, particularly of African people and Indigenous populations. Many people from these communities are understandably wary of participating in research that they regard as having little chance of benefiting them, and even some chance of causing harm. For example, when diseases are associated with a particular population, it can result in stigma and discrimination.

A committee of researchers convened by the African Academy of Sciences is urging international funders to take more account of the needs and wishes of those who contribute their data to genomics. That includes informed-consent agreements that are better tailored to specific research purposes, instead of the broad consent that is often requested. Ultimately, the best way forward is for this research to be performed by teams with people from many communities, all with an equal share in the process and an equal stake in the outcomes (see page 209).

At this milestone anniversary, the genomics community – including funders, journals, researchers and participants from around the world – needs to recommit to open data sharing. At the same time, researchers must work in closer partnership with participants – devoting more time to

nature

engaging, building trust, listening and acting on concerns. This must be seen as a necessary part of genomics research, and will be key to its future.

Commitments are also needed to improve the standards for data repositories. The repositories must be made more accessible and less onerous to contribute to. Moreover, their governance needs to better reflect diverse perspectives, not only of the global genomics research community, but also of those whose data are being accessed.

As has been seen repeatedly during the pandemic, rapid data sharing can provide massive benefits to science and, through science, to all of society. It's time to shore up that foundation and improve sharing practices — but always with equity and respect.

# India must protect its landmark science agency

**The new funding agency will allow thousands of researchers to develop their talents. Its architects must ensure it is independent.**

In a groundbreaking change, India's neglected university and college researchers are getting a new funding agency. The National Research Foundation (NRF), announced in last week's budget, will distribute 100 billion rupees (US$1.37 billion) annually for its first five years, starting this year. It will have a particular focus on interdisciplinary work, and research in colleges and universities. The launch of the NRF comes on top of plans to invest more than 40 billion rupees over 5 years for deep-ocean research; a pledge to set up 4 new virology institutes; and a commitment to developing hydrogen energy (see page 189).

These funding boosts will hopefully start to reverse the steady decline that has blighted the country's investment in research and innovation as a percentage of national income. In 2018, India spent 0.69% of its gross domestic product on research and development, compared with 0.84% a decade earlier. This compares with China's 2018 spending of 2.1% and South Korea's of 4.2%. But the stellar budget news for India's researchers comes as academics continue to voice concerns about government interference in their affairs. The benefits of the changes will be fully realized only if the NRF is allowed to function independently.

The importance of creating the NRF cannot be overstated — it is the most significant development in India's research-funding policy in at least a decade. For more than 70 years, researchers at India's many thousands of colleges and close to 1,000 universities have had few sources of large grants. Most of India's research and development funding has been concentrated in government laboratories and a network of prestigious institutes of science and technology, whereas the focus of universities has been on teaching. As a consequence, India had just 255 researchers per million people in 2017 — a fraction of that in many other countries. For example, Israel had 8,342 per million, Sweden 7,597 and South Korea 7,498 in the same year.

> **"The NRF is the most significant development in India's research-funding policy in at least a decade."**

Generations of university-based researchers have wanted access to resources on a par with those provided to their better-funded colleagues, and there have previously been discussions in government about creating an agency like the NRF. That this ambition is now being realized is, in part, down to the foresight and diplomatic skills of biologist Krishnaswamy VijayRaghavan, who is the principal scientific adviser to the Indian government. Challenges such as eliminating poverty and providing clean drinking water, sanitation, quality education and health care will need a "deep understanding of the social sciences and humanities and the various socio-cultural dimensions of the nation", VijayRaghavan told *Nature*.

The government has not yet provided full details on where the NRF will sit in the nation's public administration. It could be attached to a government ministry — as the United Kingdom's largest science-funding body is — or it could report to parliament, in a model closer to that seen in the United States. The Indian government has pledged that the NRF will operate autonomously, regardless of where it finds a home. This will be crucial. VijayRaghavan and his colleagues need to work with the government to ensure that both grant recipients and those who run the agency can make decisions — such as appointing staff or peer reviewers — without interference from government officials, as is the convention in international science-funding policy.

India's researchers have been voicing concerns over the state's undermining of research autonomy for some time. In 2017, around 12,000 researchers participated in a march for science across 40 cities. In 2019, more than 100 economists wrote to Prime Minister Narendra Modi, urging an end to political influence over official statistics — particularly economic data. And just last month, the Ministry of Education told universities they must obtain permission from the government when organizing online events with international speakers on topics that relate to the country's security or internal affairs.

The government says this last move is no different from universities seeking permission when inviting international scholars to visit India for certain academic purposes, but researchers have told *Nature* that it is unnecessary because universities already have procedures in place for vetting conference speakers. They say that adding another bureaucratic hurdle will simply result in fewer international speakers attending online events being hosted in India.

India's NRF has been a long time coming. Its creation is an achievement of great vision. Enabling new generations of scholars to realize their potential will be its architects' most important legacy to their nation. But it needs to start with the right foundations. That means giving it protection from undue influence — not only from the current government, but from its successors, too.