THE CONTROVERSIAL EMBRYO TESTS THAT PROMISE A HEALTHIER BABY

Some companies offer tests that rank embryos based on their risk of developing complex diseases such as schizophrenia or heart disease. Are the tests accurate — or ethical? **By Max Kozlov**

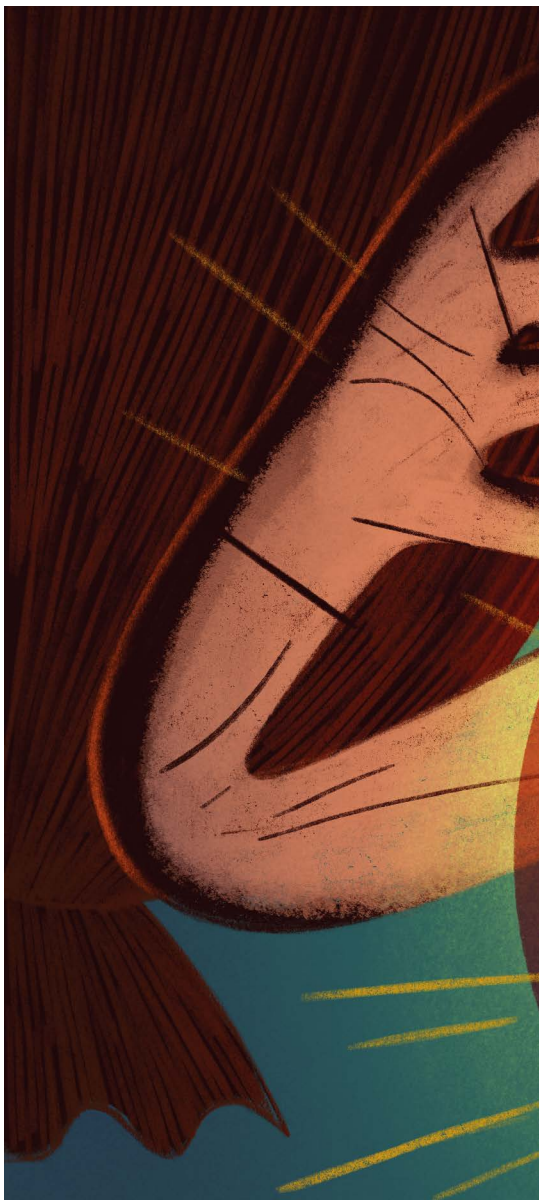


ILLUSTRATION BY PAWEŁ JONCA

"She has her mother's eyes," begins the advertisement, "but will she also inherit her breast cancer diagnosis?" The smooth voice in the video is promoting the services of Genomic Prediction, a US company that says it can help prospective parents to answer this question by testing the genetics of embryos during fertility treatment.

For Nathan Treff, the company's chief scientific officer, this mission is personal. At 24, he was diagnosed with type 1 diabetes — a disease that cost his grandfather his leg. If Treff had it his way, no child would be born with a high risk for the condition.

His company, in North Brunswick, New Jersey, offers tests based on a decade of research into 'polygenic risk scores', which calculate someone's likelihood of getting a disease on the basis of the genetic contributions of hundreds, thousands or even millions of single DNA letter changes in the genome.

Genomic Prediction and some other companies have been using these scores to test embryos generated by *in vitro* fertilization (IVF), allowing prospective parents to choose those with the lowest risk for diseases such as diabetes or certain cancers. A co-founder of Genomic Prediction has said, controversially, that people might eventually be able to select for traits that are unrelated to disease, such as intelligence.

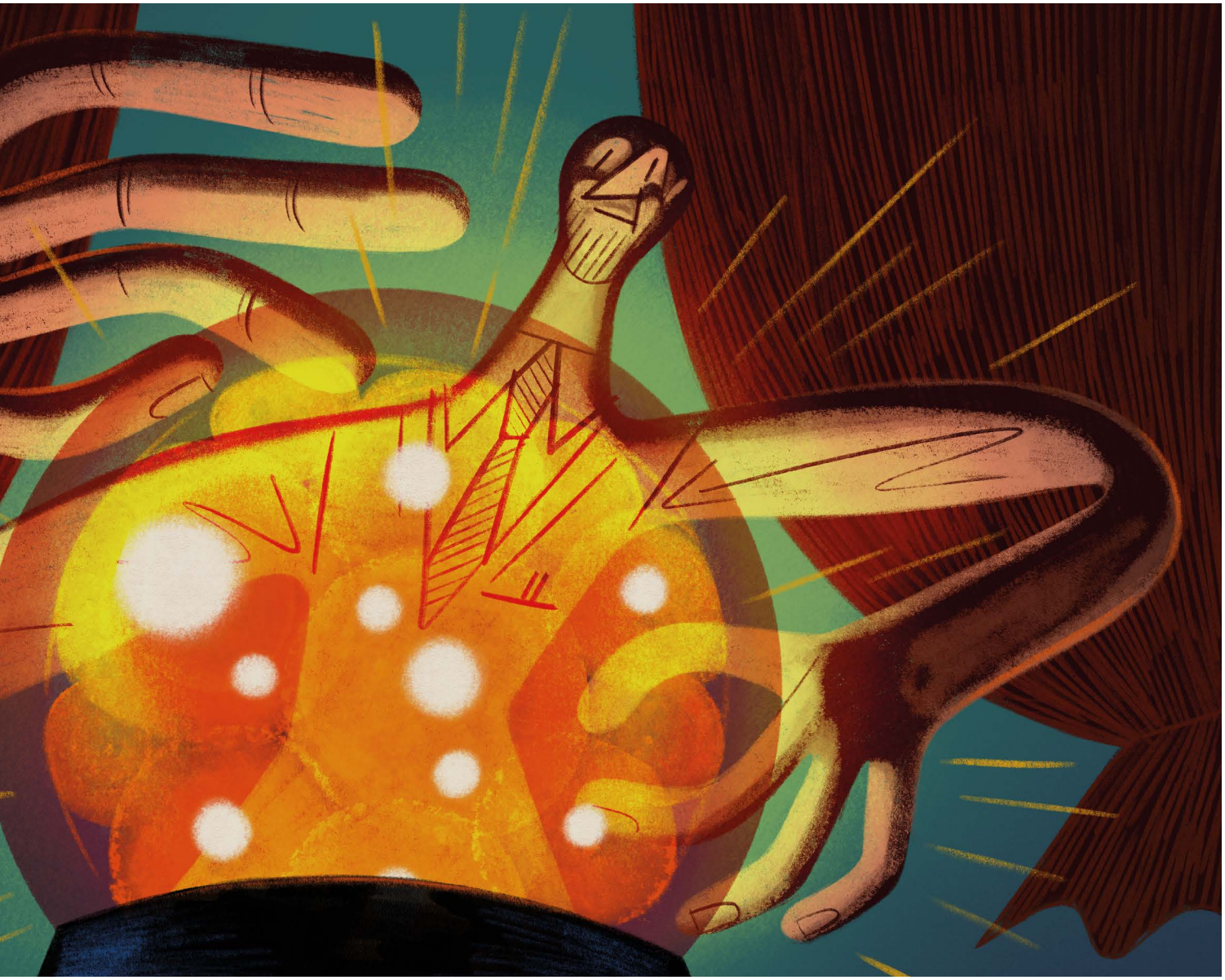
Pre-implantation genetic testing (PGT) for rare genetic disorders and chromosomal abnormalities has become common practice in the US$14-billion IVF industry. But testing for polygenic conditions (often referred to as PGT-P) is much newer, with only a small handful of companies selling it in a few countries, including the United States and Brazil, where it is largely unregulated.

In the United States, people undergoing IVF can request that their clinicians order PGT-P, which promises screening for various conditions, including some cancers, heart disorders, diabetes and schizophrenia. Only a few hundred people have done so, according to Treff. But if experience with other forms of PGT is any indication, the use of PGT-P could skyrocket: the proportion of IVF cycles that included more-established forms of PGT in the United States increased from 13% in 2014 to 27% in 2016.

Many are troubled by the possibilities that PGT-P presents: bioethicists have long been wary of attempting to select disease and disability out of the human gene pool, and the high cost of testing could further entrench health inequities.

Researchers are also concerned that, in most cases, the genomic models behind these tests are too weak to predict disease risk in a meaningful way for a developing embryo. Polygenic risk scores are ripe for misinterpretation, and people might be misled by the information they receive. Genomic Prediction says that it offers genetic counselling to clients.

There are already indications that those who



aren't infertile might turn to IVF to take advantage of the testing, subjecting themselves to health risks for a reward that is speculative at best, says Laura Hercher, a genetic counsellor at Sarah Lawrence College in Bronxville, New York. Treff doesn't see a problem with otherwise-healthy people opting for the tests, but Hercher says people "shouldn't be hyper-stimulating their ovaries so they can use mediocre tests to pick between embryos".

Jared Robins, executive director of the American Society for Reproductive Medicine (ASRM) in Washington DC, which represents fertility clinics and researchers, agrees. The ASRM is reviewing the technology and has yet to take an official stance on the tests. But, he says, "it's a technology not quite ready for prime time".

Reading the risk

The first baby conceived by IVF was born in 1978, and researchers have since developed tests that screen embryos for chromosomal abnormalities and some monogenic conditions — those caused by a single defective gene, such as cystic fibrosis. Although these tests come with their own ethical and practical concerns, they are not nearly as controversial as PGT-P.

PGT-P takes advantage of a decades-long effort in genomics to identify the genetic contributors to many common diseases. Pinpointing the precise roots of diabetes, schizophrenia, heart disease and a host of other conditions has proved to be a fiendishly difficult task, however. Most diseases are considered polygenic — often linked to many different genes interacting with each other and their environment in complex ways.

Amassing genome and health data on hundreds to thousands of people, typically as part of biobanking projects, has enabled researchers to compare subtle differences in the DNA of people with or without a certain condition. They use artificial-intelligence models to detect DNA-letter differences, called single nucleotide polymorphisms (SNPs), that are more widespread in the group with the condition.

One SNP alone might make a negligible difference to a person's risk of developing diabetes, but adding up the effects from tens to millions of these variants can produce a model for scoring that risk. Researchers can then see how predictive their models are by looking at the genomes and health conditions of people who weren't included in the original population, to see whether high scorers are indeed more likely to have a given disease.

For example, a 2018 study[1] trained a model to detect SNPs for coronary artery disease using the genomic data of almost 61,000 people with the condition and around 123,000 individuals without it. After testing the predictor on a separate group of nearly 290,000 people in the UK Biobank, they found that those scoring in the highest few percentiles had, on average, a risk of developing the disease that was more than three times

higher than in the remaining population. The UK National Health Service is piloting these scores as a way to identify adults who have a high risk of developing heart disease. The idea is that physicians can recommend lifestyle changes and routine screening to those with high risk scores.

But researchers are still grappling with the limitations of these scores, says Peter Visscher, a quantitative geneticist at the University of Queensland in Brisbane, Australia, who pioneered the methods that underlie polygenic risk scores. One issue is that genetic variation can explain only a proportion of the total risk — environmental factors such as diet or air quality, for example, are also important contributors. And because polygenic scores simply correlate with the presence of a condition, it is difficult to discern whether they truly reflect genes that contribute to the disease, or whether they reveal something broader about the populations that have that condition.

The scores might also be misleading because the underlying data lack ethnic and geographical diversity. Scores are typically generated and validated using biomedical information from people with European ancestry, in data sets such as the UK Biobank, limiting their applicability to people of other ethnicities.

For all of these reasons, the scores are not yet ready for widespread use in clinics for any purpose, says Visscher — let alone as a basis for selecting an embryo.

Testing boundaries

The PGT-P process starts off the same as any other pre-implantation genetic test: clinicians take a small sample of a days'-old embryo and sequence its DNA. Then, using information from studies of polygenic risks, they find the variants that correlate with the likelihood of developing a certain condition later in life.

Sequencing the DNA of an embryo when it is composed of only a few hundred cells is no easy feat. Researchers at MyOme, a company in Menlo Park, California, developed a technique[2] that can reconstruct the full genome of such embryos with nearly complete accuracy — with the help of genome sequences from both parents. MyOme is creating its own embryo-screening test for polygenic conditions, and is studying how clients and clinicians process the results, according to Matthew Rabinowitz, a co-founder of the company.

Genomic Prediction already offers screening for schizophrenia, four heart conditions, five cancers and type 1 and type 2 diabetes. IVF clinics try to generate multiple embryos to increase the chance of successful pregnancy, and Genomic Prediction calculates an overall health score for each embryo (see 'Scoring embryos'). The score incorporates the risk of each disease, weighted on the basis of how many years a disease could take off that embryo's future life. It also includes a breakdown of individual risks for each disease, says Treff.

In its simulations[3], the company says that its tests can reduce the risk of choosing an embryo that later develops type 1 diabetes by 72%. But that figure assumes that the simulated couple is selecting between five viable embryos, is of European ancestry and has a family history of the disease. For most people, who don't have a genetic predisposition to type 1 diabetes, the chances that their child will develop the condition are exceedingly rare. In opting for this test, they would only slightly reduce the risk of something that is already very unlikely, says Hercher.

Other geneticists argue that PGT-P doesn't have adequate predictive power to significantly reduce disease risk for many conditions. One reason, says Visscher, who co-authored a 2021 paper[4] that outlined problems with PGT-P, is that sibling embryos are much more similar to each other than are non-related embryos, so the difference in risk will never be as noticeable as for two unrelated adults. And because a single cycle of IVF typically yields only three to four viable embryos, the tests are less useful than if people had, say, 20 embryos to choose from, says Gabriel Lázaro-Muñoz, a bioethicist at Harvard Medical School in Boston, Massachusetts. Treff acknowledges these concerns, but he notes that any risk reduction — however small — is significant enough to warrant using PGT-P, and that it would be unethical for providers not to offer the test.

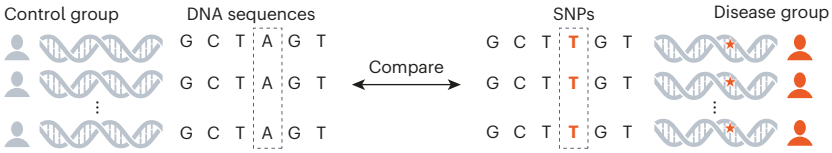
Francesca Forzano, a clinical geneticist at Guy's and St. Thomas' NHS Foundation Trust in London, counters that it would be unethical to offer someone a test when you don't have clear, real-world evidence that it's beneficial.

SCORING EMBRYOS

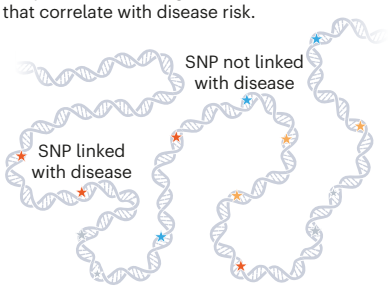
Some companies are selling genetic tests for embryos generated by *in vitro* fertilization that they say can identify an embryo with the lowest risk of developing common diseases, such as diabetes and some cancers. The method relies on sequencing the embryo's genome and comparing it with existing data on genetic risk factors in adult populations.

Genetics and risk

Researchers have discovered links between individual DNA letter changes, called single nucleotide polymorphisms (SNPs), and the risk of having a disease. To do this, they compare genome and health data from thousands of people with or without various diseases.

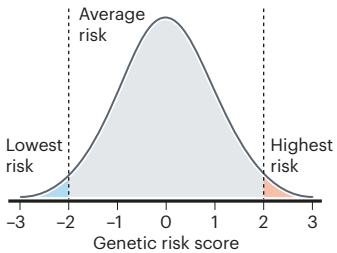


They look across the genome for SNPs that correlate with disease risk.



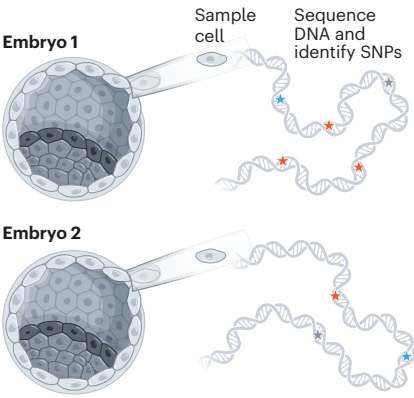
Control for age, sex and ancestry

Combining information on all SNPs yields a prediction for how likely someone is to develop a disease.



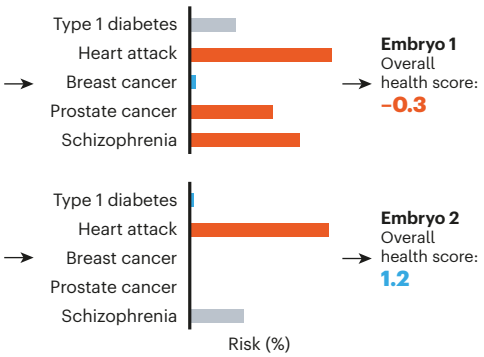
Embryo sampling

To apply this information to an embryo, clinicians take a few cells from embryos that are about 5 days old. They extract and sequence the DNA and look for SNPs.



Calculate scores

One company, Genomic Prediction, uses this method to help clients to identify embryos with a low risk of developing 12 disorders, including diabetes and some cancers. Each embryo's overall score is calculated by combining the risk of each disease and weighting them by their effect on life expectancy. A higher health score suggests a lower overall risk.



SOURCE FOR 'CALCULATE SCORES': GENOMIC PREDICTION

"It's one thing when you have no other options, like when you offer a very sick child something off-label, but here there is no label to begin with," says Forzano, who chairs the Public and Professional Policy Committee of the European Society of Human Genetics in Vienna, Austria, and who co-authored a commentary[5] arguing against PGT-P last year. Treff says that he has published evidence for the tests' efficacy[6]; MyOme and Orchid in San Francisco, California, a third provider of PGT-P, did not respond to requests for comment on the ethics of their tests.

Long road ahead

Proving that these tests work as intended will be a long and extremely difficult task, says Hercher. To properly validate them, researchers would need to follow people born as a result of these tests to see how many develop the conditions that PGT-P screens for, and compare them with people who were not screened in this way as embryos. But because many of these conditions arise later in life and are relatively rare, such trials would span an entire lifetime and involve thousands of people, says Hercher.

In the absence of data from any longitudinal trials, researchers have been modelling the potential benefits of PGT-P screening by retrospectively scoring groups of adults and seeing whether the tests would have made an impact. For example, in 2019, Shai Carmi, a statistical geneticist at the Hebrew University of Jerusalem in Israel, and his colleagues created virtual genomes for simulated embryos by blending together the DNA sequences of randomly generated pairs of men and women. They then predicted how tall the simulated embryos would be in adulthood, using polygenic risk scores. If the hypothetical parents had five embryos to choose from, they could expect to gain only about 2.5 cm in height and 2.5 IQ points over the average, the researchers found[7].

Using a similar method, researchers at Genomic Prediction applied their polygenic risk scores to 40,000 late-life individuals in the UK Biobank whose medical and genetic history was available, and put them in groups of up to 10 people. Comparing the highest-scoring individual in each group with the rest of the data set, they found that their test would have reduced the risk of choosing an embryo with a high chance of developing almost any of the 20 diseases they screened for, and would have improved life expectancy by between a day and a month for most conditions tested[8]. (The study was posted to a preprint server and has not yet been peer reviewed.) The researchers found no evidence that selecting against any of the 20 diseases would increase the likelihood of developing another – a phenomenon known as antagonistic pleiotropy, in which one gene controls more than one trait. However, they didn't search for effects on diseases outside the group of 20. Carmi finds the results promising, but says he'd like to see the analysis replicated using other, more diverse genomic databases.

One significant weakness of using such retrospective analyses, says Leila Jamal, a bioethicist at the US National Cancer Institute in Bethesda, Maryland, is that it is difficult to predict which conditions will be problematic by the time babies screened with PGT-P reach old age. "The treatment landscape is changing so rapidly," she says, adding that both medicine and people's environments looked very different only 50 years ago.

The science behind polygenic scores has also progressed rapidly in the past five years, which could soon render current methods out of date or, worse, incorrect, says Forzano.

Despite the limitations, researchers say there are ways in which PGT-P might be useful. Screening clients with a family history of conditions that tend to be diagnosed early in life, such as type 1 diabetes, might be beneficial, says Hercher. That's because environmental factors are thought to have less of an influence on these diseases than on conditions that typically arise later in life, such as heart disease. "It's this unique moment where you can stop this paradigm of transferring diseases and disabilities from one generation to the next," says Rabinowitz.

> "THE TREATMENT LANDSCAPE IS CHANGING SO RAPIDLY."

Still, some specialists worry that thinking of embryos in terms of their health scores could increase the stigma around some conditions, especially those affecting mental health. For now, Genomic Prediction offers only one mental-health-related test – for schizophrenia – but the company's co-founder Stephen Hsu has hinted that he'd like to see tests for intelligence.

Lázaro-Muñoz points out that for some people with a mental-health disorder, their diagnosis forms a much more crucial part of their identity than for people with cancer or heart disease, for instance. That means that selecting against embryos with a risk of mental-health conditions could add further stigma.

"This is taking us a step closer where people are narrowing their version of what is acceptable in a child," says Hercher.

And the search for the 'perfect baby' could lead some people to put themselves through unnecessary IVF cycles, says Norbert Gleicher, founder and head of the Center for Human Reproduction, an IVF clinic in New York City. IVF success rates are already low – about 25–40% of IVF cycles lead to a live birth, depending on the pregnant person's age. Gleicher worries that PGT-P will inevitably lead to otherwise healthy embryos being discarded, because people won't be satisfied with the scores and will opt for additional cycles. "The first rule in medicine is 'do no harm', not 'get a small benefit if there is one'," he says.

Loopholes and bans

Countries vary in how stringently they regulate PGT. In the United States, PGT – as with most other direct-to-consumer genetic tests – is not subject to close scrutiny and evaluation by the US Food and Drug Administration. Such tests are much more strictly regulated in the United Kingdom, where the Human Fertilisation and Embryology Authority (HFEA) has said they can be used only to avoid "serious inherited illnesses". PGT-P is illegal in the United Kingdom, and the HFEA has noted that there is no scientific consensus on the validity of the tests.

For the United States and other countries with a permissive approach, Lázaro-Muñoz says it will be important to work with patient advocacy groups and companies to set standards about the information that consumers and clinicians receive about these tests.

Hercher says she would be more comfortable with an approach directed at people who have a high risk for a particular disease. "A healthier conversation would feel a lot less like you're building a better baby, and a lot more targeted and specific." She says that professional organizations such as the ASRM can play an important part in setting expectations for companies and clinicians offering PGT-P.

Some power in is the hands of consumers of these tests. Hercher implores those debating using PGT-P to thoughtfully consider their intentions for using the test. Pregnancy is already a very fraught time, she says. If access to PGT-P continues to expand, Hercher wonders whether its existence will change people's perception of parenthood. She worries that people will effectively be able to 'shop' for desirable traits, "taking us away from a place of being unconditional in our regard for our children and instead toward a consumerist mentality".

Max Kozlov reports for *Nature* from Washington DC.

1. Khera, A. V. *et al. Nature Genet.* **50**, 1219–1224 (2018).
2. Kumar, A. *et al Nature Med.* **28**, 513–516 (2022).
3. Treff, N. R. *et al. Front. Endocrinol.* **10**, 845 (2019).
4. Turley, P. *et al. N. Engl. J. Med.* **385**, 78–86 (2021).
5. Forzano, F. *et al. Eur. J. Human Genet.* **30**, 493–495 (2021).
6. Treff, N. R. *et al. Genes* **11**, 648 (2020).
7. Karavani, E. *et al. Cell* **179**, 1424–1435 (2019).
8. Widen, E., Lello, L., Raben, T. G., Tellier, L. C. A. M. & Hsu, S. D. H. Preprint at medRxiv https://doi.org/10.1101/2022.06.15.22276102 (2022).