

Comment

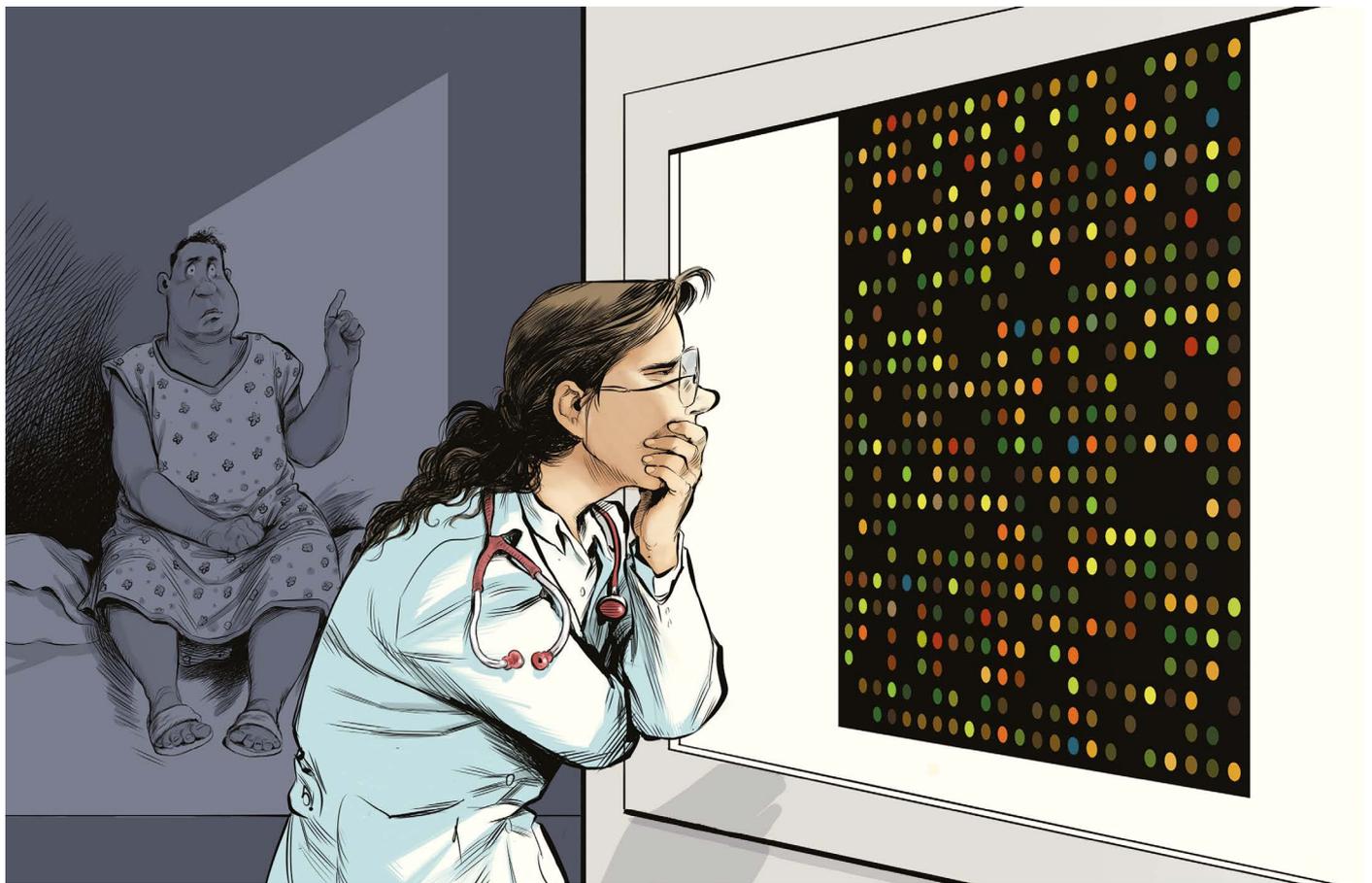


ILLUSTRATION BY DAVID PARKINS

Personalized profiles for disease risk must capture all facets of health

Mark McCarthy & Ewan Birney

To provide individual care and prevent disease, we need to go beyond genetics in risk scores and include metrics that follow a person's changing environment and health.

Providing the best possible care for an individual means having a better understanding of their risks of developing disease. The goal is to have personalized answers when people need to know whether, for instance, preventive surgery makes sense, a given medicine is likely to be risky or a certain diet should be recommended.

Information on genetic risk represents one promising approach to providing these answers. Genomic data, gathered across millions of individuals, have revealed thousands of DNA sequence variants associated with common diseases such as diabetes, heart disease, schizophrenia and cancer. These clues to disease risk can be combined to generate

'polygenic scores', which provide a measure of the degree to which an individual is genetically predisposed to developing each such disease¹.

A growing chorus of scientists and clinicians emphasize the value of such genetic profiling as an integral part of a person's medical record². Others argue that the clinical benefits have been massively overstated³. This debate often fails to recognize that the challenge is not merely to improve understanding of genetic risk, but to capture more about the interwoven, multifaceted factors that play into disease risk (see 'Path to personalization').

Here, we argue that clinical medicine must learn to develop more-holistic measures of individual risk, both genetic and non-genetic,

Comment

and to combine these with clinical data over time to deliver better care.

Limited measures

Although current polygenic scores hold clinical promise, they come with several limitations. They leave out many sources of relevant data, and work best for the predominantly white, wealthy populations in which most genetic studies have been performed. The emphasis on genetic risk diverts attention away from non-genetic factors that might be equally important for disease risk and progression. Risk estimation on the basis of polygenic scores alone also fails to incorporate real-time measurements of clinical state that are especially important in diseases linked to ageing.

Both authors are strongly invested in the value of human genetics as a tool for understanding disease mechanisms, and are enthusiastic about the contribution that genetic profiling will make to personalizing care. M.M. is an endocrinologist who has focused on understanding the genetics of type 2 diabetes, and leads human genetic research at the biotechnology firm Genentech in South San Francisco, California. E.B. is the deputy director-general of the European Molecular Biology Laboratory (EMBL) and director of the EMBL European Bioinformatics Institute near Cambridge, UK, and has played a pivotal part in the design and analysis of multiple genome projects.

To gain a more accurate assessment of individual health risks (that is, to make medicine truly personalized), researchers and clinicians must integrate disparate types of data from a wider diversity of populations. First, researchers need to expand measures of genetic risk by embracing more-diverse populations, cataloguing the full spectrum of variants, and understanding the environmental context in which these variants act. Second, researchers and clinicians need to be able to consider both genetic and non-genetic risk factors (for type 2 diabetes, for example, these would encompass hundreds of genetic markers and measures of diet, exercise and socio-economic status alongside measures of current clinical state, such as glucose levels). Finally, the field needs to move away from its tendency to collapse all these rich, individual-level data into rigid clinical categories. Rather than classifying an individual as simply being at average or high risk for a condition such as coronary artery disease, researchers and clinicians should consider a gradation of risk. And instead of trying to categorize people into discrete subtypes of disease, we should appreciate that common disease typically involves several processes running in parallel⁴.

Inclusive genetics

Polygenic scores for late-onset diseases are mostly built around the common risk variants that have emerged from large-scale genetic studies. In contrast to the rare, high-impact

genetic variants that underlie diseases such as cystic fibrosis and sickle-cell anaemia, these generally have subtle effects that limit their clinical value when considered one at a time. However, when information from hundreds or thousands of relevant disease-risk variants is combined, we can capture a substantial slice of individual variation in disease risk^{1,5}. In European populations, for example, someone in the highest 1% of polygenic risk for coronary artery disease is at least ten times more likely to develop the disease than is someone in the lowest 1% (ref. 5).

These polygenic scores have the potential to inform individual decisions about screening, lifestyle interventions and therapeutic choices. For example, rather than all women starting to have annual mammography screening at 45 years old (as currently recommended by the American Cancer Society), polygenic scores for breast cancer risk could be used to tailor schedules so that women with the highest genetic risk are screened earlier and more intensively than are those with below-average risk⁶.

“Researchers should commit to adopting a more-holistic perspective in their work.”

The reliability of these scores depends on the accuracy and inclusivity of the genetic information that goes into them. Most data currently used to construct polygenic scores come disproportionately from individuals of recent European descent. However, scores generated in one population typically perform poorly when deployed in another: a polygenic score for body mass index (BMI) constructed from European individuals loses more than 60% of its predictive power when applied to individuals of more-recent African descent, for example⁷.

Another concern is that common genetic variants tell only part of the story of genetic risk. For many diseases, rare variants also contribute, often having a much greater impact on risk than any single common variant. Notable examples include effects of rare variants in the genes *BRCA1* and *BRCA2* on breast and ovarian cancer risk, and of those in *LDLR*, *APOB* and *PCSK9* on coronary artery disease (mediated through the effects of these variants on lipid levels). Polygenic scores that do not incorporate these rare, ‘high penetrance’ variants will provide misleading estimates of overall genetic risk for those who carry the high-impact version (or allele) of the genes responsible. Equally, the clinical consequences of inheriting a high-impact allele are modulated by an individual’s polygenic background: in some diseases, carriers of high-impact alleles who have a favourable polygenic background have a disease risk that is at or below the population average^{8,9}.

The solution is to integrate both common and rare variants into a single genetic risk score. Historically, research at the common and rare ends of the allele-frequency spectrum has involved different groups of researchers deploying distinct techniques (genotyping arrays and targeted sequencing, respectively). However, whole-genome sequencing is swiftly becoming the default genetic assay. This shift is eroding the artificial distinction between ‘rare’ and ‘common’ variants, and is making it much easier to consider the entire spectrum of genetic risk at once. This will, for example, allow carriers of high-risk alleles for breast cancer to make better decisions about screening and prophylactic surgery. Crucially, however, rare variants vary more between ancestries than do common variants, and the pursuit of equitable genetic information will depend even more on collecting inclusive global data on genetic variation and disease risk in diverse populations.

Getting holistic

There is more to disease risk than genetics. For most common, late-onset diseases, individual risk is heavily influenced by non-genetic factors. Often collectively labelled as environmental, these might include factors as varied as diet, socio-economic status, access to health care, the status of personal relationships and gut-microbiome diversity.

It is not straightforward to measure and integrate these factors into risk estimates. Even for well-understood factors, such as smoking, diet and exercise, the lifelong impact on disease risk cannot easily be assembled from ‘snapshot’ measurements, such as steps walked or estimated calories consumed in the past week. What’s more, even when epidemiological associations are strong, it can be challenging to pin down the factors that are causal: consider ongoing debates about how dietary components, such as carbohydrate and fat intake, influence disease risk. Many exposures that might be relevant to disease are simply hard to reconstruct, for example prenatal nutrition and exposure to pathogens or antibiotics during infancy.

Complex societal factors, such as access to health care, education, effective sanitation or housing, have a profound impact on individual patterns of disease. As with genetic risk, data gathered from wealthier populations can translate poorly into disease prediction in disadvantaged communities¹⁰. Unless scientific leaders, funders, industry and societies work together to rebalance the populations involved in data generation and clinical validation, existing health disparities will be perpetuated and perhaps even amplified.

Genetic and non-genetic risk factors often interact in ways that can be hard to disentangle. For example, genetic variants that alter the function of nicotinic receptors influence smoking behaviour, and, as a consequence, are associated with individual risk

of smoking-related diseases. The metabolic disease phenylketonuria is a striking example of how modifying the environment can modulate the consequence of genetic variation: the devastating consequences of inherited defects in the causative *PAH* gene can be mitigated by adopting a diet low in phenylalanine.

Clinical measurements, particularly when gathered over time, represent another route for improving risk estimation. Consider two people aged 50, both with polygenic scores in the top 10% of genetic risk for type 2 diabetes. One is sedentary and overweight, the other active and slim. One might reasonably expect the former to be at greater risk of diabetes than the latter. But now assume that yearly measurements of glycosylated haemoglobin (which reflect a person's glucose levels over the previous two to three months) have remained firmly in the normal range for the first individual for more than a decade, but show a steady increase towards the diabetic range for the second. That makes the second individual much more likely to become diabetic.

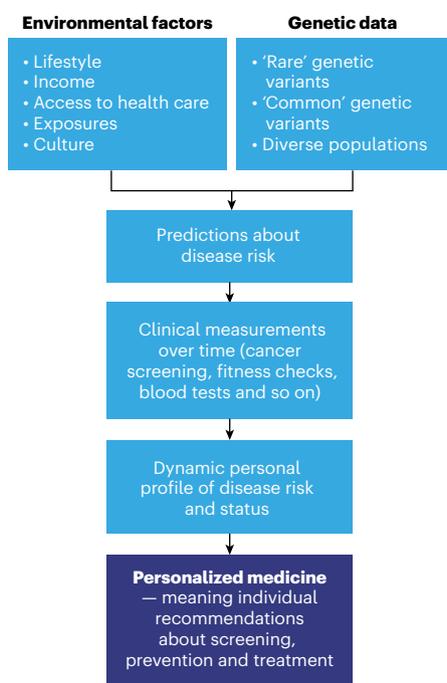
In general, clinical data collected repeatedly over time – from sources such as blood tests, imaging and wearable devices – reveal how broad-brush predictions derived from genetic and non-genetic risk factors are actually playing out in a given individual, and make it possible to chart personal trajectories from health to disease. The inclusion of real-time clinical data also helps to counter the fatalism that can seep into the interpretation of genetic risk. It emphasizes how, even in those with the highest genetic risk, interventions can mitigate disease progression. Such integrated assessments are also readily incorporated into clinical practice. For example, cholesterol measurements are already widely used to stratify cardiovascular risk precisely because they integrate diverse genetic and environmental factors, as well as dynamic measurements of current clinical state.

Preserving complexity

Medicine has historically focused on categorizing disease. Personalized medicine has often followed the same path, subdividing people into perceived disease subtypes, or establishing arbitrary divisions in continuous measurements (such as high and low risk). Such efforts assume that the highly variable manifestations of disease can best be explained by allocating individuals to distinct groups, and that each disease subtype has its own set of causes. However, most common diseases represent a confluence of disordered processes, several of which are likely to be at play in any given individual. For instance, premature coronary artery disease typically occurs amid a blend of abnormal processes, including disordered glucose metabolism, elevated lipids, high blood pressure and chronic inflammation. The precise mix will differ from one person to another, and even across a person's lifetime. Only in

PATH TO PERSONALIZATION

To tailor health care to individuals, information from various sources must be brought together. These data, both genetic and environmental, should be drawn from diverse populations.



relatively few individuals (for example, those with familial hypercholesterolaemia) can premature disease be attributed to a single cause.

When many causes contribute to disease in an individual, it makes more sense to track each process involved, rather than collapsing rich quantitative information into a set of rigid, often-arbitrary, disease or risk categories. Even though clinical decision-making often demands binary decisions (such as to treat or not at a particular time point), these might not map neatly onto categories defined years previously. There is the danger that these become 'once-and-for-all' labels in the medical record that define future health care for that individual and divert attention away from personal differences in disease trajectory. A more quantitative approach would, for example, render moot unproductive debates about the most appropriate definition of metabolic syndrome, or how best to use ancestry to define which BMI thresholds constitute overweight and obese.

Tracking multiple measurements reveals the ebb and flow of each individual's status with respect to health and disease. Then, when it becomes necessary to make a binary clinical decision – whether or not to operate, or whether to try drug A or B – both the individual and the physician can rely on much richer and more up-to-date information than on categories assigned years previously.

Moving forward

How do we get there? Researchers should commit to adopting a more-holistic perspective in their work. Researchers, funders and

industry need to embrace greater diversity in the design and implementation of studies, focusing not only on gender and ethnicity, but also on social, cultural and economic factors that influence disease risk and access to health care. Recent moves by major funders to encourage more-diverse participation in population cohorts and biobanks are welcome, but reducing the diversity of modern populations to census-defined categories does not do justice to the complex, admixed ancestries of so many.

Efforts to base personalized medicine on risk-factor prediction alone will fall short. All involved in this endeavour – researchers, industry, funders, governments and citizens – will need to come together to enable the collection of large, rich data sets that go beyond static one-time measurements and which capture individual health trajectories. Such efforts are, however, destined to fail unless the data are collected in standardized formats and shared in ways that allow information from different studies and populations to be combined and compared. This will inevitably bring the realms of research and clinical care together, and will require us to address fundamental questions about data ownership, privacy, equality of access, fairness and social responsibility. Global efforts to create such standards are in place, for example the Global Alliance for Genomics and Health (www.ga4gh.org).

Achieving this more-holistic mindset will take time and effort. But the resulting understanding of disease and framing of personal risk will be deeper, broader and much better equipped to bring the promise of personalized medicine into routine health care.

The authors

Mark McCarthy is executive director of Human Genetics at Genentech, South San Francisco, California, USA. **Ewan Birney** is deputy director-general of the European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Cambridge, UK. e-mails: mccarthy.mark@gene.com; birney@ebi.ac.uk

Disclaimer: M.M.'s views in this article are his own and not those of Genentech.

1. Torkamani, A., Wineinger, N. E. & Topol, E. J. *Nature Rev. Genet.* **19**, 581–590 (2018).
2. Lambert, S. A., Abraham, G. & Inouye, M. *Hum. Mol. Genet.* **28**, R133–R142 (2019).
3. Sud, A., Turnbull, C. & Houlston, R. *NPJ Precis. Oncol.* **5**, 40 (2021).
4. Erdmann, J., Kessler, T., Munoz Venegas, L. & Schunkert, H. *Cardiovasc. Res.* **114**, 1241–1257 (2018).
5. Khera, A. V. et al. *Nature Genet.* **50**, 1219–1224 (2018).
6. Mavaddat, N. et al. *Am. J. Hum. Genet.* **104**, 21–34 (2019).
7. Martin, A. R. et al. *Nature Genet.* **51**, 584–591 (2019).
8. Fahed, A. C. et al. *Nature Commun.* **11**, 3635 (2020).
9. Mars, N. et al. *Nature Commun.* **11**, 6383 (2020).
10. Mostafavi, H. et al. *eLife* **9**, e48376 (2020).

The authors declare competing interests; see go.nature.com/3msjwzr.