

A brief history of human disease genetics

<https://doi.org/10.1038/s41586-019-1879-7>

Received: 16 July 2019

Accepted: 13 November 2019

Published online: 8 January 2020

Melina Claussnitzer^{1,2,3}, Judy H. Cho^{4,5,6}, Rory Collins^{7,8}, Nancy J. Cox⁹, Emmanouil T. Dermizakis^{10,11}, Matthew E. Hurles¹², Sekar Kathiresan^{2,13,14}, Eimear E. Kenny^{4,6,15}, Cecilia M. Lindgren^{2,16,17}, Daniel G. MacArthur^{2,13,18}, Kathryn N. North^{19,20}, Sharon E. Plon^{21,22}, Heidi L. Rehm^{2,13,18,23}, Neil Risch²⁴, Charles N. Rotimi²⁵, Jay Shendure^{26,27,28}, Nicole Soranzo^{12,29} & Mark I. McCarthy^{17,30,31,32*}

A primary goal of human genetics is to identify DNA sequence variants that influence biomedical traits, particularly those related to the onset and progression of human disease. Over the past 25 years, progress in realizing this objective has been transformed by advances in technology, foundational genomic resources and analytical tools, and by access to vast amounts of genotype and phenotype data. Genetic discoveries have substantially improved our understanding of the mechanisms responsible for many rare and common diseases and driven development of novel preventative and therapeutic strategies. Medical innovation will increasingly focus on delivering care tailored to individual patterns of genetic predisposition.



**Anniversary
collection:**
[go.nature.com/
nature150](http://go.nature.com/nature150)

For almost all human diseases, individual susceptibility is, to some degree, influenced by genetic variation. Consequently, characterizing the relationship between sequence variation and disease predisposition provides a powerful tool for identifying processes fundamental to disease pathogenesis and highlighting novel strategies for prevention and treatment.

Over the past 25 years, advances in technology and analytical approaches, often building on major community projects—such as those that generated the human genome sequence¹ and elaborated on that reference to capture sites of genetic variation^{2–6}—have enabled many of the genes and variants that are causal for rare diseases to be identified and enabled a systematic dissection of the genetic basis of common multifactorial traits. There is growing momentum behind the application of this knowledge to drive innovation in clinical care, most obviously through developments in precision medicine. Genomic

medicine, which was previously restricted to a few specific clinical indications, is poised to go mainstream.

This Review charts recent milestones in the history of human disease genetics and provides an opportunity to reflect on lessons learned by the human genetics community. We focus first on the long-standing division between genetic discovery efforts targeting rare variants with large effects and those seeking alleles that influence predisposition to common diseases. We describe how this division, with its echoes of the century-old debate between Mendelian and biometric views of human genetics, has obscured the continuous spectrum of disease risk alleles—across the range of frequencies and effect sizes—observed in the population, and outline how genome-wide analyses in large biobanks are transforming genetic research by enabling a comprehensive perspective on genotype–phenotype relationships. We describe how the expansion in the scale and scope of strategies for enumerating the functional consequences of genetic variation is transforming the torrent of genetic discoveries from the past decade into mechanistic insights, and the ways in which this knowledge increasingly underpins advances in clinical care. Finally, we reflect on some of the challenges and opportunities that confront the field, and the principles that will, over the coming decade, drive the application of human genetics to enhance understanding of health and disease and maximize clinical benefit.

¹Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, USA. ²Broad Institute of MIT and Harvard Cambridge, Cambridge, MA, USA. ³Institute of Nutritional Science, University of Hohenheim, Stuttgart, Germany. ⁴Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA. ⁵Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA. ⁶Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA. ⁷Nuffield Department of Population Health (NDPH), University of Oxford, Oxford, UK. ⁸UK Biobank, Stockport, UK. ⁹Vanderbilt Genetics Institute and Division of Genetic Medicine, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA. ¹⁰Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva, Switzerland. ¹¹Health 2030 Genome Center, Geneva, Switzerland. ¹²Wellcome Sanger Institute, Hinxton, UK. ¹³Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA. ¹⁴Verve Therapeutics, Cambridge, MA, USA. ¹⁵Center for Genomic Health, Icahn School of Medicine at Mount Sinai, New York, NY, USA. ¹⁶Big Data Institute at the Li Ka Shing Centre for Health Information and Discovery, University of Oxford, Oxford, UK. ¹⁷Wellcome Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, Oxford, UK. ¹⁸Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA. ¹⁹Murdoch Children's Research Institute, Parkville, Victoria, Australia. ²⁰University of Melbourne, Parkville, Victoria, Australia. ²¹Departments of Pediatrics and Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA. ²²Texas Children's Cancer Center, Texas Children's Hospital, Houston, TX, USA. ²³Department of Pathology, Harvard Medical School, Boston, MA, USA. ²⁴Institute for Human Genetics, University of California San Francisco, San Francisco, CA, USA. ²⁵Center for Research on Genomics and Global Health, National Human Genome Research Institute, Bethesda, MD, USA. ²⁶Department of Genome Sciences, University of Washington, Seattle, WA, USA. ²⁷Brotman Baty Institute for Precision Medicine, Magnuson Health Sciences Building, Seattle, WA, USA. ²⁸Howard Hughes Medical Institute, Seattle, WA, USA. ²⁹Department of Haematology, University of Cambridge, Cambridge, UK. ³⁰Oxford Centre for Diabetes, Endocrinology and Metabolism, Oxford, UK. ³¹Oxford NIHR Biomedical Research Centre, Oxford University Hospitals NHS Foundation Trust, John Radcliffe Hospital, Oxford, UK. ³²Human Genetics, Genentech, South San Francisco, CA, USA. *e-mail: mccarthy.mark@gene.com

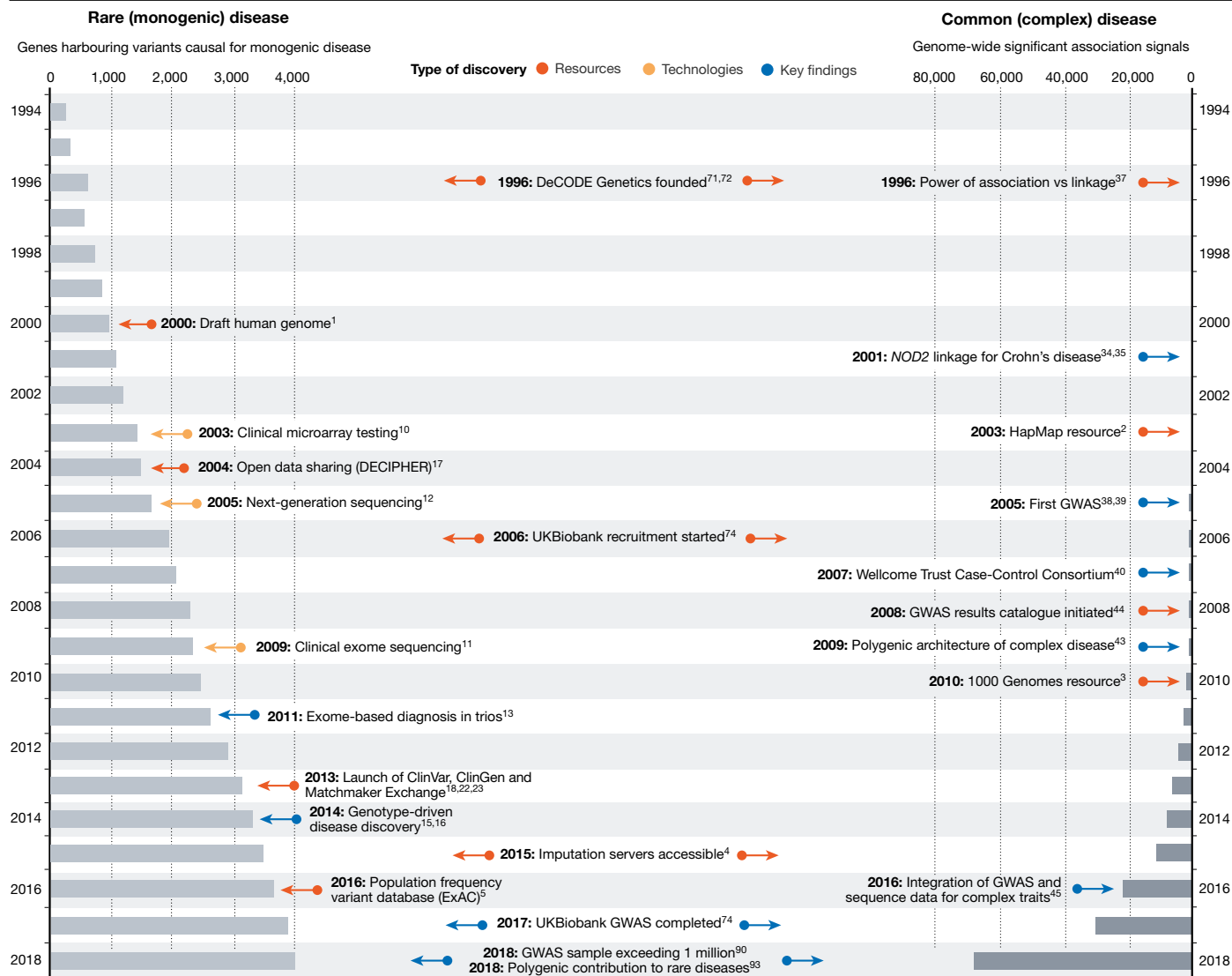


Fig. 1 | Growth in the discovery of disease-associated genetic variation. The cumulative numbers of genes harbouring variants causal for rare, monogenic diseases and traits and of significant GWAS associations implicated in common, complex diseases and traits are shown. Left, the advent of high-throughput sequencing technologies and availability of reference genomes from diverse populations has supported a fourfold increase in the discovery of rare disease-causing genes between 1999 and 2019. Right, international efforts such as the Human Genome Project and the HapMap Project, combined with

GWAS and sequencing studies, have supported identification of more than 60,000 genetic associations across thousands of human diseases and traits. Centre, more recent developments have brought a synthesis of the rare- and common-variant approaches based around the combination of sequence-informed analyses in large cohorts. Key events contributing to these themes are depicted in the timeline. GA4GH, Global Alliance for Genomics and Health¹⁶⁰; ExAC, Exome Aggregation Consortium⁵.

Rare diseases, rare variants

During the 1980s and 1990s, efforts to map disease genes were focused on rare, monogenic and syndromic diseases and were mostly driven by linkage analysis and fine mapping within large multiplex pedigrees. Localization of genetic signals was typically followed by Sanger sequencing of the genes found to map within the linked locus to identify disease-causing alleles. Assessments of pathogenicity, based on segregation of a putatively causal variant with disease across multiple families and evidence that the risk genotype was absent in healthy individuals, were typically followed by confirmatory functional studies in cellular and animal models. This path to gene identification was laborious; nevertheless, by 2000, around 1,000 of the estimated 7,000 single-gene inherited diseases had been characterized, including many with substantial biomedical impact, such as Huntington's disease and cystic fibrosis^{7–9}.

Completion of the draft human genome sequence¹ reduced many of the obstacles to disease-gene mapping and propelled a fourfold increase in the genes implicated as causal for rare, single-gene disorders (Fig. 1). Microarray-based detection of structural variation¹⁰ and exome- and genome-wide sequencing^{11,12} have been pivotal, bolstered by in silico analysis and prioritization of the discovered genetic variants. Increasing availability of reference datasets cataloguing population genetic variation across diverse ethnic backgrounds has supported robust causal inference^{2,3,5,6}. More recently, the adoption of high-throughput sequencing technologies has enabled the full range of causal genetic variation, from single mutations to large structural rearrangements, to be identified in a single assay. These technologies have extended from research into clinical usage, driving earlier and faster diagnosis for genetic disorders.

Reduced reliance on multiplex pedigrees in favour of collections of affected cases, often with parents¹³, has proven decisive in identifying

new dominant disorders, many of which were previously considered recessive¹⁴. Increasingly, discovery of rare disease genes has transitioned from genetic characterization of small numbers of individuals with similar clinical presentations to genome-wide sequencing of larger cohorts of phenotypically diverse patients. This genotype-driven approach has revealed new disorders associated with more variable clinical presentation^{15,16}.

A more systematic approach to data sharing has been critical, both for the characterization of new disorders and diagnostic interpretation of potential causal alleles. The value of sharing genetic and phenotypic data from those thought to harbour rare undiagnosed genetic diseases has fostered global collaborative networks (for example, Matchmaker Exchange, DECIPHER and GeneMatcher) designed to match patients with similar genetic variants and/or phenotypic manifestations, even across continents^{17–19}. Interactions between researchers and families with rare disease have enabled natural history studies to be driven by family support groups positioned to initiate data collection from patient cohorts once a causal gene is discovered²⁰.

Clinical translation of these technologies has benefited from a series of information resources, including open databases of genes associated with rare disorders (for example, OMIM and ORPHANET)²¹, clinically interpreted variants (for example, ClinVar and ClinGen)^{22,23} and patient records (for example, DECIPHER and MyGene2 (<https://mygene2.org/MyGene2>))¹⁷. Access to resources that catalogue genetic variation across populations (such as ExAC and its successor gnomAD)^{5,6} has enabled the confident exclusion of genetic variants too common in population-level data to be plausible causes of rare, penetrant early-onset genetic diseases²⁴. These analyses have reduced the contamination of databases with variants erroneously interpreted as causal for disease, and are addressing the overestimation of disease penetrance arising from the historical focus on multiplex pedigrees²⁵. Improved recognition of the variable penetrance of many ‘monogenic’ disease alleles has invigorated efforts to identify the genetic and environmental modifiers responsible^{26,27}.

Although huge strides have been made in associating specific genes with particular disorders, establishing the causal role of individual variants within those genes remains problematic, and many patients with suspected rare genetic diseases are left without a definitive diagnosis²⁸. Even for variants with established causality, the penetrance is often unclear. Resolving these uncertainties represents the central challenge for the field. Aggregation of sequencing data from large numbers of affected cases and population reference samples will provide the evidence base required for robust interpretation of variants. Highly parallelized *in vitro* cellular assays that allow assessment of the functional effects of all variants in a disease-associated gene can transform interpretation of novel variants²⁹, although developing well-calibrated functional assays predictive of pathogenicity for all disease genes represents a daunting prospect. Direct functional genomic exploration of accessible and disease-relevant tissues from patients using RNA sequencing and DNA methylation assays^{30,31} can identify previously cryptic causal genetic variants, particularly in under-explored regions outside protein-coding genes^{32,33}. Developments in each of these areas will extend the range of variants and genes for which diagnostic and prognostic clinical information can be provided to patients and their families.

Common diseases, common variants

Efforts to apply the approach—linkage analysis in multiplex pedigrees—that had been so successful for the high-penetrance variants responsible for Mendelian disease were, with notable exceptions^{34–36}, largely unsuccessful for common, later-onset traits with more complex multifactorial aetiologies, such as asthma, diabetes and depression. Recognition that association-based methods, focused on detecting phenotype-related differences in variant allele frequencies might have

greater traction for identifying less penetrant common alleles redirected attention to analysis of case–control samples³⁷. However, initial efforts targeting variants within ‘candidate’ genes were plagued by inadequate power, unduly liberal thresholds for declaring significance and scant attention to sources of bias and confounding, resulting in overblown claims and failed replication.

Systematic efforts to characterize genome-wide patterns of genomic variation, initially through the HapMap Consortium², proved catalytic, demonstrating that the allelic structure of the genome was segmented into haplotype blocks, each containing sets of correlated variants. Recognition that this configuration could support genome-wide surveys of association energized the technological innovation—in the form of massively parallel genotyping arrays—to make such studies possible (Fig. 1). Early wins in acute macular degeneration³⁸ and inflammatory bowel disease³⁹ were encouraging, and progress on several fronts—expansion of study size, denser genotyping arrays, novel strategies for imputation, attention to biases and appropriate significance thresholds—delivered robust associations across a range of diseases⁴⁰. Most variants uncovered by these early genome-wide association studies (GWAS) were common, with more subtle effects than many had anticipated. A host of trait-specific consortia formed, covering diverse dichotomous and quantitative phenotypes, to accelerate genetic discovery through the aggregation and meta-analysis of data from multiple GWAS studies^{41–43}. Many tens of thousands of robust associations were identified⁴⁴. Recently, increased access to exome and whole-genome sequence data has, through both direct association analysis^{45,46} and imputation^{3,4}, extended discovery to low-frequency and rare alleles previously inaccessible to GWAS.

In the decade since the first GWAS, understanding of the genetic basis of common human disease has been transformed. The disparity between the observed effects of the variants first identified by GWAS and estimates of overall trait heritability (the ‘missing heritability’ conundrum) is now largely resolved⁴⁷. Common diseases are not simply aggregations of related Mendelian conditions: for most complex traits, genetic predisposition is shared across thousands of mostly common variants with individually modest effects on population risk^{41,43}.

Although the collective contribution of low-frequency and rare risk alleles to overall trait variability appears modest compared with that attributable to common variants^{45,48}, the rare risk alleles detected in current sample sizes necessarily have large phenotypic effects and are proportionately more likely to be coding, enhancing their value for biological inference. Founder populations (such as those from Finland and Iceland) have provided multiple examples of otherwise rare risk alleles driven to higher frequency locally through drift and/or selection^{49–52}. In addition, studies in populations with high rates of consanguinity make it possible to identify individuals homozygous for otherwise rare loss-of-function alleles, the basis for a ‘human knockout’ project to systematically investigate the phenotypic consequences of gene disruption in humans^{53,54}.

For most diseases, large-scale GWAS-aggregation efforts have been disproportionately powered by information from individuals of European descent⁵⁵. Whereas patterns of genetic predisposition appear broadly similar across major population groups and many common risk alleles discovered in one population group are detectable in others, allele frequencies can vary substantially; extending GWAS and sequencing studies to diverse populations will surely generate a rich harvest of novel risk alleles.

The relative contributions of common and rare variants indicate that, for many traits, particularly those with post-reproductive onset, purifying selection has had only limited effect^{45,56}. For a few risk alleles, hallmarks of balancing selection reflect increased carrier survival, usually through protection from infectious diseases. This includes well-known examples of alleles maintained at high frequency in populations of African descent^{57,58}.

While the extensive linkage disequilibrium within human populations has been essential to discovery in GWAS, high correlation between adjacent variants frustrates mapping of the specific variants responsible for these associations. Increasing sample size, improved access to trans-ethnic data, and more representative imputation reference panels³ provide a path to improved resolution of the causal variants⁵⁹ and clues to the molecular mechanisms through which they operate. Functional interpretation is easiest for causal variants within coding sequences; however, most common disease-risk variants map to noncoding sequences, and are presumed to influence predisposition through effects on transcriptional regulation. In these cases, mechanistic inference depends on connecting association signals to their downstream targets (see below). For many traits, there is clear convergence between common-variant association signals and genes implicated in monogenic forms of the same disease, as well as enrichment of GWAS signals in regulatory elements specifically active in cell types consistent with known disease biology^{60,61}. This provides reassurance that, even as the number of association signals for a given disease proliferates, the genetic associations uncovered will coalesce around molecular and cellular processes with a core role in pathogenesis^{62,63}.

Importantly, the signals discovered by GWAS have revealed many unexpected insights into the biological basis of complex disease. Examples include the role of complement in the pathogenesis of acute macular degeneration³⁸, synaptic pruning in schizophrenia⁶⁴ and autophagy in inflammatory bowel disease⁶⁵. In addition, as inherited sequence variation is a prominent cause of phenotypic variation (but the reverse is not true), risk variants identified by GWAS have value as genetic instruments, mapping causal relationships between traits and inferring contributions made by circulating biomarkers and environmental exposures to disease development⁶⁶.

As described below, findings from GWAS have increasing translational impact through identification of novel therapeutic targets⁶⁷, prioritization (and deprioritization) of existing ones⁶⁸ and development of polygenic scores that quantify individual genetic risk⁶⁹.

Comprehensive genotype–phenotype maps

The historical division of disease-gene discovery into monogenic and polygenic strands arose from development and implementation of analytical approaches—family-based linkage and case–control association³⁷—that are best-suited for detecting particular subsets of causal alleles. This obscured the true state of nature, with disease-risk alleles being distributed across a continuous spectrum of frequencies and effect sizes. In addition, the trait- and disease-specific perspective of early GWAS discovery (mostly reliant on case–control studies) was poorly equipped to investigate the contribution of genetic variants to phenotypic effects that are nested within or spread across classical disease definitions. Recent developments have enabled a more holistic perspective on genotype–phenotype relationships (Fig. 1).

One major advance has been the increasing availability of large prospective population-based cohorts. These biobank efforts, pioneered in studies such as the Framingham Cohort⁷⁰ and the efforts of DeCODE in Iceland^{71,72}, now encompass a growing inventory of national cohorts in North America, Europe, Asia and beyond^{73–76}. The UK Biobank study, including 500,000 largely healthy, middle-aged participants has been particularly influential, transforming human genetic research in part through permissive data-sharing policies that have allowed multiple research groups to analyse the data⁷⁴. Efforts to make clinical data embedded in electronic health records and registries available for research^{77,78} mean that biobanks increasingly provide access to a wide range of demographic, clinical and lifestyle data, captured in harmonized, systematic fashion from large, often multi-ethnic collections of individuals. For millions of biobank participants, this rich phenotypic information has been combined with genome-wide genetic data. There are nascent efforts to capture transcriptomic, proteomic and

metabolomic phenotypes, although these are not yet at equivalent scale to the genetic data^{79,80}. Biobank analyses have provided more generalizable estimates of the relevance of genetic risk factors in the context of the separate and joint effects of non-genetic factors⁸¹. Increasingly, integration with healthcare data brings a longitudinal dimension to phenotypic characterization, which facilitates analyses of disease progression and lifelong disease risk⁸².

The rich phenotypic scope of these cohorts has enabled variants of interest to be interrogated for associations across the gamut of available phenotypes. These phenome-wide association studies (PheWAS) have revealed the extent to which many variants have pleiotropic effects across multiple traits⁸³. Some of these relationships are expected, such as the impact of obesity variants on risk of hepatic steatosis and type 2 diabetes⁸⁴ or variants that influence multiple autoimmune conditions⁸⁵. Others connect diseases and traits in surprising ways, highlighting shared polygenic, pleiotropic effects and cell-type specificity, and delivering insights into shared biology and overlapping mechanisms^{86,87}. These findings inform the prioritization of therapeutic targets, providing clues to potential on-target side effects and opportunities for drug repurposing^{87–89}.

The second enabler of inclusive, systematic analysis of genotype–phenotype relationships has been access to whole-genome sequence data. The scale of genetic analysis based on sequence data still lags behind that of genome-wide genotyping data (the largest sequence-based datasets are one tenth the size of the largest GWAS^{90–92}), although reductions in sequencing costs are decreasing the differential. Most direct analysis of high-throughput sequence data has focused on the coding regions. Strategies for assigning variant function and jointly analysing sets of variants of similar functional effect have enabled aggregate, gene-level tests of rare functional-variant association that are often better powered than single-variant tests^{91,92}. However, the principal benefit to date of whole-genome sequence data to genetic discovery has been to bolster array-based access to lower-frequency alleles, either directly, through their inclusion on genotyping platforms, or indirectly, through imputation from sequence-based reference samples^{3,4}.

These developments have enabled researchers to bridge the gap between the monogenic and polygenic realms, identifying common variant modifiers of monogenic phenotypes contributing to the variable expression of rare, large-effect alleles^{26,93}, and low-frequency and rare variants that influence common multifactorial traits^{94,95}. This enables more rigorous evaluation of the contribution of rare and common variants to trait susceptibility⁴⁸ and supports the enumeration of ‘allelic series’ (sets of alleles of varying frequency, effect size and direction that disrupt the same gene) critical for studies of disease mechanism and therapeutic target optimization^{89,96}. These developments are rapidly converging towards the ultimate destination: a comprehensive matrix of the effect of all observable genetic variants across the widest possible range of cross-sectional and longitudinal biomedical phenotypes. Success in this endeavour depends on ever greater harmonization between, and integration of results from, individual studies through sustained investments in data sharing.

Adding function

From the first linkage maps to whole-genome sequencing of large cohorts, human genetics has deployed increasingly sophisticated and inherently systematic approaches for mapping the genetic factors that underlie traits and diseases. However, progress in determining how these variants influence disease, through systematic interrogation of their functional effects on molecular, cellular and physiological processes, has been far slower.

For monogenic diseases, for which the alleles responsible are typically rare, penetrant and coding, genetic approaches have generally been both necessary and sufficient to implicate a gene as causal²⁸.

However, as efforts to elucidate the genetic basis of Mendelian disorders progress towards completion⁹⁷, functional studies remain important to understand the mechanisms by which disruptive variation within a causal gene leads to disease phenotypes. Unlike common diseases, the clarity of causation for Mendelian disorders usually simplifies the task of generating models (including human cells and organoids or rodents) to connect genotype to organismal phenotype; these have led to many critical insights into the biology of health and disease in humans^{98,99}. In addition, for genes harbouring variants with medically actionable consequences (as with the *BRCA1* and *BRCA2* mutations that are causal for early-onset breast and ovarian cancer), functional studies can support the translational interpretation of novel alleles identified by medical sequencing²⁹.

For common diseases, functional studies have a more fundamental role. Although tens of thousands of associations have been discovered across thousands of common human diseases and traits⁴⁴, multiple factors have frustrated efforts to convert these genetic signals to knowledge about causal variants, genes and mechanisms. For the common variants that underlie the bulk of complex-disease risk, the resolution of association mapping is often limited by the haplotype structure of the human genome²⁻⁴. Furthermore, most GWAS associations map to the noncoding genome and thus lack a direct address to the gene that mediates their effects. Growing appreciation of the pervasive role of pleiotropy complicates matters: many variants identified by GWAS are associated with multiple traits and exert diverse effects across multiple cell types¹⁰⁰.

To date, relatively few studies have achieved the goal of connecting variants causal for complex traits to the molecular and cellular functions that mediate that predisposition. One early success described how regulatory variants that modulate *SORT1* expression influence low-density lipoprotein cholesterol and myocardial infarction risk¹⁰¹. More recent examples have focused on the relationship between obesity-associated variants intronic to *FTO*, altered expression of *IRX3* and *IRX5*, and adipocyte¹⁰² and hypothalamic¹⁰³ function. Similar functional descriptions have been reported for individual loci implicated in schizophrenia⁶⁴, cardiovascular disease¹⁰⁴, type 2 diabetes¹⁰⁵ and Alzheimer's disease¹⁰⁶, among others.

Over the past decade, the challenge for the functional genomics community has been to convert this 'one-locus-at-a-time' workflow to a systematic, multidimensional, integrative approach able to deliver genome-scale functional analyses to match genome-wide variant discovery (Fig. 2). At the molecular level, one cornerstone has been generation of genome-wide catalogues of functional activity. For example, the ENCODE and Roadmap Epigenomics projects have generated maps of histone modifications, transcription-factor binding, chromatin accessibility, three-dimensional genome structure and other regulatory annotations across hundreds of cell types and tissues^{107,108}. The patterns of genomic overlap between these data and GWAS results enable the functional inference of risk variants, deliver clues to the specific cell types driving disease pathogenesis^{60,109} and accelerate locus-specific mechanistic insights.

In parallel, there has been a scaling of efforts to connect trait-associated regulatory variants to the genes and processes that they regulate in cell types relevant to the disease of interest^{110,111}. For example, the GTEx (Genotype-Tissue Expression) consortium has mapped thousands of expression quantitative trait loci (QTLs) across hundreds of individuals and dozens of tissues¹¹². Further clues to the relationships between regulatory variants and their effector genes can be gathered from DNA proximity assays (such as Hi-C) and single-cell data¹¹³ (Fig. 2). Programs such as HubMAP¹¹⁴ and the Human Cell Atlas¹¹⁵ are set to deliver comprehensive, high-resolution reference maps of individual human cell types across diverse developmental stages, providing new opportunities to understand how regulatory genetic variation results in cellular and organismal phenotypes.

Efforts to probe the clinical consequences of coding alleles with large phenotypic effects (particularly null alleles) in humans^{33,34} and across

diverse animal models¹¹⁶ represent powerful strategies for extending functional analyses to the whole-body level. Connections between genetic variation and circulating proteomic and metabolomic data provide additional mechanistic links between cellular events and whole-body physiology^{79,80}. These efforts are paralleled by PheWAS approaches⁸³, which, by mapping variant effects across the range of traits available in biobanks and EMRs, can inform priors for cell types and pathways at individual loci. Importantly, whereas early studies typically linked GWAS risk alleles to data from a single functional assay, the focus is increasingly on maximizing biological insight through the multi-dimensional integration of multiple genome-wide data types using approaches such as heritability partitioning¹¹⁷, functional enrichment analyses^{60,109}, integration of the three-dimensional genome structure¹¹⁸ and deep convolutional neural networks^{119,120}.

Although QTL analyses can implicate a haplotype in a molecular, cellular or organismal phenotype, they are, in isolation, insufficient to define the specific causal variants responsible. To address this, there has been rapid maturation of technologies, such as massively parallel reporter assays¹²¹⁻¹²³ and CRISPR genome editing, to support functional characterization of targeted sequence perturbations at scale. Variations on these methods enable the functional evaluation of genes (via knockout screens¹²⁴), regulatory elements (using CRISPR interference and CRISPR activation screens^{125,126}), and genetic variants (base editors¹²⁷) at increasing scale and resolution²⁹. Combined with complex readouts—including high-content imaging¹²⁸ and single-cell transcriptomics and epigenomics^{129,130}—these methods can generate empirical 'truth' data, supporting the development of *in silico* models to predict causal variants, effector transcripts¹²⁶ and cellular effects. In due course, such models should reduce the need for exhaustive experimental characterization of function for all variants across all cell types.

The goal of such efforts is to enumerate the cascade of molecular events that underlie observed genotype-phenotype associations using physiologically relevant cellular systems (from primary cells to organoids and 'organ-on-chip' designs) and whole-body assays appropriate to the disease of interest. Collectively, strategies that offer large-scale functional evaluation of variants and genes of interest will reduce (but probably not eliminate) the intensive effort required for 'final mile' validation of disease mechanisms in dedicated systems, thereby accelerating downstream translational application.

Clinical implementation

Medical genetics, as applied to rare diseases, has been characterized by the rapid application in the clinic of the transformative genomic technologies that drove initial research discoveries. There are now targeted genetic tests for nearly all clinical presentations attributable to large-impact alleles, alongside more extensive genome-sequencing assays that, when necessary, enable interrogation of a longer list of relevant genes. Genetic testing for symptomatic individuals and at-risk relatives occurs routinely in many medical specialties. In parallel, the use of somatic cancer testing has increased as therapies targeted to specific mutational events have entered clinical practice (these developments are reviewed elsewhere^{131,132}).

For patients with symptoms that indicate a probable monogenic aetiology (such as retinal degeneration, hearing loss or cardiomyopathy), targeted panels are typically the platform of choice¹³³, although they are increasingly performed on a more extensive sequence backbone. For more complex phenotypes—those without a clear match to a specific syndrome, such as neurodevelopmental disorders and multiple congenital anomalies—testing has gravitated towards early deployment of exome and genome-sequencing platforms that offer speedy resolution of what has historically often been a traumatic diagnostic odyssey^{15,134}. The power of genomic diagnosis is especially clear for those presenting with monogenic neurodevelopmental disorders and critically ill infants^{135,136}. Sequencing of the parent-offspring trio can

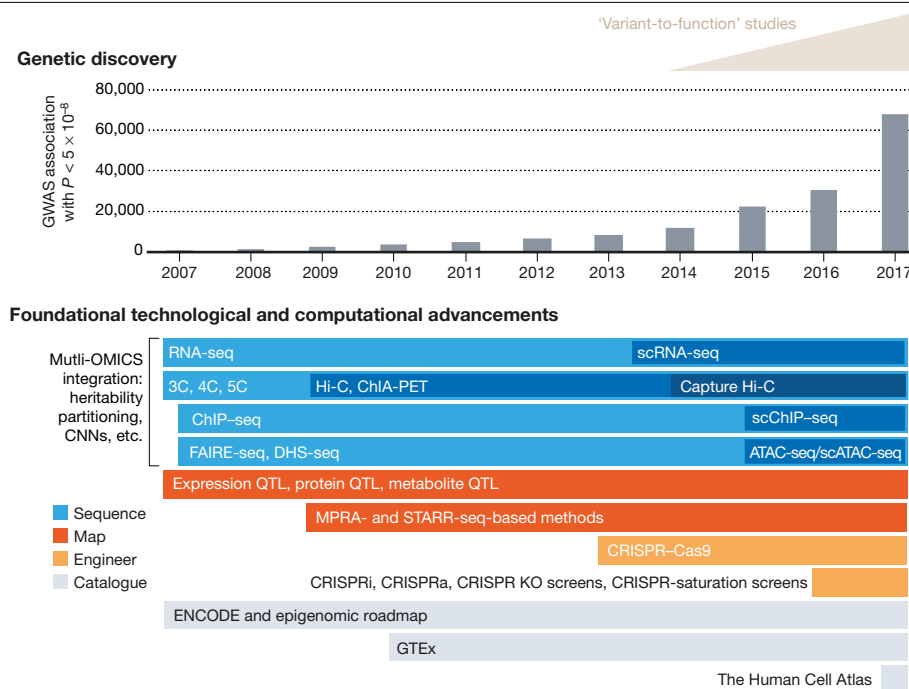


Fig. 2 | Genetic discovery is paralleled by advances in functional genomics technologies. Top, the growth in the number of genetic loci associated by GWAS with human traits and diseases (bars) and of variant-to-function studies (area under line, not to scale). Bottom, foundational technological and computational advances over the last decade that enabled (1) development of systematic, genome-wide catalogues of functional elements across multiple cell types and tissues (blue); (2) mapping of QTLs in the context of gene expression, metabolites, proteins and regulatory elements (red); (3) engineering of genes, genetic elements and genetic variation at increasing scale (orange); and (4) systematic tissue-specific surveys of regulatory

elements and transcription (grey). scRNA-seq, single-cell RNA-sequencing analysis; ChIA-PET, chromatin interaction analysis by paired-end tag sequencing; ChIP-seq, chromatin immunoprecipitation followed by sequencing; FAIRE-seq, formaldehyde-assisted isolation of regulatory elements with sequencing; DHS-seq, DNase I-hypersensitive sites sequencing; ATAC-seq, assay for transposase-accessible chromatin using sequencing; MPRA, massively parallel reporter assay; STARR-seq, self-transcribing active regulatory region sequencing; CNN: convolutional neural networks. For further details and primary literature on many of these assays, see ref. ¹⁷³.

detect *de novo* variation in dominant disorders and phase biallelic rare variants in recessive disease¹³.

The transition from targeted gene tests to genomic sequencing enables recursive reanalysis, including reinterpretation of individual sequences on the basis of subsequent discoveries regarding causal disease alleles and their phenotypic consequences¹³⁷. However, improved molecular diagnostics are required to ensure reliable detection of a subset of genetic disorders, including those arising from triplet repeats and complex rearrangements¹³⁸. Deep sequencing of affected tissues for mosaic variants and the use of RNA sequencing to detect noncoding variants that drive early-onset disease (for example, through effects on splicing) represent new fronts for clinical diagnostics³⁰.

Other examples of the rapid adoption of new genomic technologies include noninvasive prenatal testing (more than ten million tests by 2018 across multiple countries^{139–141}) and the use of recessive carrier panels for couples planning pregnancies. Newborn screening is now universal in many countries, although it is limited to disorders combining high-throughput low-cost detection with effective early interventions (such as diet restrictions or enzyme replacement)¹⁴². Genetic diagnostics are also increasingly applied to newborn screening as a reflex test following an abnormal (for example, metabolic) screening test¹⁴³. Over the next decade, the repertoire of disorders captured by neonatal screening and prenatal testing is likely to expand markedly. Whereas prenatal testing may be more effective at avoiding disease, the associated ethical issues are more complex¹⁴⁴.

Although genetic testing for rare disease and cancer has exploded, there has been more limited uptake of genetic information in other aspects of healthcare. For example, despite multiple examples of clinically important genetic markers related to drug efficacy and side-effect

profile¹⁴⁵, the roll-out of pharmacogenetics has been hampered by a range of factors, including lack of clinical decision support in electronic medical systems to guide the drug choice or dosing by the physician. This has been compounded by challenges in diagnostic testing: complex haplotype structures and structural variants at some key drug metabolism loci necessitate genome sequencing or specific targeted panels to detect all clinically relevant variants.

For common diseases, translational attention is currently focused on the clinical potential of polygenic risk scores. The development of robust polygenic scores for several common diseases has been catalysed by more precise per-variant effect estimates from larger GWAS datasets, improved algorithms for combining information across millions of single-nucleotide polymorphisms, and large-scale biobanks that support score validation^{69,146,147}. For example, a genome-wide polygenic score for heart attack, incorporating 6.6 million variants, indicates that 5% of European-descent individuals have a risk of future cardiac events equivalent to that seen in those with less frequent monogenic forms of hypercholesterolaemia⁶⁹. Increasingly, the shift from array-based genotyping to sequence-based analysis is facilitating risk prediction, which integrates information from rare, large-effect alleles with that from polygenic scores⁹³. By improving the capture of genetic risk, particularly in non-European populations, and integrating environmental and biomarker data to quantify aspects of non-genetic risk, it should be possible to achieve increasingly accurate prediction of individual disease risk, and to use this information to tailor screening, prevention and treatment. Success will depend on developing models of risk that robustly integrate these diverse data types and on optimizing the strategies deployed to ensure effective implementation.

Table 1 | Examples showing progress in the clinical utility of genetic testing

Mode of clinical utility	Disorder	Gene(s)	Treatment or action
Targeted therapeutics	Cystic fibrosis	<i>CFTR</i> (p.G551D)	Ivacaftor
	Retinal dystrophy	<i>RPE65</i>	Voretigene neparvovec-rzyl
	Fabry disease	<i>GLA</i> (specific variants)	Migalastat
	Spinal muscular atrophy	<i>SMN1</i>	Nusinersen
	Familial hypercholesterolaemia	<i>PCSK9</i>	Evolocumab and alirocumab
Altering diet or behaviour	Phenylketonuria	<i>PAH</i>	Phenylalanine-free diet
	Hereditary fructose intolerance	<i>ALDOB</i>	Elimination of fructose and sucrose from diet
	Biotinidase deficiency	<i>BTBD</i>	Biotin vitamin supplements
	Usher syndrome	<i>MYO7A</i> and others	Avoid cycling and other activities that depend on balance and peripheral vision
Prophylactic interventions	Sickle-cell disease	<i>HBB</i> ⁸⁸	Antibiotics and pneumococcal vaccine
	Brugada syndrome	<i>SCN5A</i>	Implantable cardioverter–defibrillator
	Hereditary breast or ovarian cancer	<i>BRCA1</i> or <i>BRCA2</i>	Bilateral salpingo-oophorectomy
	Multiple endocrine neoplasia 2	<i>RET</i>	Thyroidectomy
Increased surveillance	Lynch syndrome	<i>MSH2</i> and others	Early initiation of colonoscopy
	Marfan syndrome	<i>FBN1</i>	Echocardiography
	Haemochromatosis	<i>HFE</i>	Iron screening
	Alpha-1 antitrypsin deficiency	<i>AAT</i>	Spirometry, liver function tests

There is an increasing number of examples in which a specific molecular diagnosis results in downstream actions to improve patient outcomes. These actions can range from a change in diet or behaviour to prophylactic surgery. For each example shown, we provide the disorder, the gene and corresponding actions. The GeneReviews compendium (<https://www.ncbi.nlm.nih.gov/books/NBK1116/>) provides up-to-date information on the management of these and other Mendelian disorders.

The absence of evidence-based guidelines to support healthcare recommendations continues to hinder the clinical applications of genetic data. In some countries, this is compounded by confusion over reimbursement and disparities in testing across society¹⁴⁸. Many healthcare professionals lack experience in genomic medicine and need education and guidance to practice in the rapidly evolving space of genetic and genomic testing¹⁴⁹. One consequence of these difficulties has been an expanding direct-to-consumer testing market, variably controlled by country-specific regulations¹⁵⁰, which is moving beyond a focus on ancestry and personal traits, towards models in which individuals have direct access to ordering physicians and genetic counselors¹⁵¹. The risk of commercial influence in this model remains high. There are concerns about the consequences of unfettered release of genetic data of dubious or inflated clinical relevance, and limited infrastructure to pull these results into mainstream medical systems.

These advances have fostered debate about the value of genetics for population screening, for both monogenic and complex disorders. Population screening for monogenic disorders is most likely to be initiated for conditions for which risk estimates are well-understood and there are actionable interventions (for example, Lynch syndrome and familial hypercholesterolaemia). Expansion to other disorders requires better understanding of the penetrance of pathogenic alleles in unselected populations¹⁵² and caution before extending screening to longer lists of genes that are less securely implicated in disease causation¹⁵³. As certain countries consider universal capture of genome-wide genetic data at birth or later in life, key questions concern the strategies for releasing this information to citizens and their medical teams to support individual healthcare.

Ultimately, barriers to genomic medicine are most directly overcome by demonstrating clinical utility in disease management and therapeutic decision-making, with evidence for improved patient outcomes. Hereditary cancers provide multiple examples, such as the use of *BRCA1/BRCA2* testing to inform PARP inhibitor treatment in patients with cancer¹⁵⁴. There is a growing list of diseases for which a molecular diagnosis results in specific interventions designed to improve patient outcomes (<https://www.ncbi.nlm.nih.gov/books/NBK1116/>) (some

examples are listed in Table 1), and there are currently more than 50 FDA-approved drugs for genetic disorders¹⁵⁵. Although gene therapy has been slow to evolve since its early introduction, recent advances in gene editing are reinvigorating approaches to treat disorders by manipulation of the underlying genetic defects¹⁵⁶.

Looking forward

Over the coming decade, the challenge will be to optimize and to implement at scale, strategies that use human genetics to further the understanding of health and disease, and to maximize the clinical benefit of those discoveries. Realizing these goals will require the concerted effort of researchers in academia and industry to bring about transformational change across a range of highly interconnected domains, for example, through the auspices of the recently established International Common Disease Alliance (<https://www.icda.bio>). Such efforts will be directed towards establishing: (a) comprehensive inventories of genotype–phenotype relationships across populations and environments; (b) systematic assays of variant- and gene-level function across cell types, states and exposures; (c) improved scalable strategies for turning this basic knowledge into fully developed molecular, cellular and physiological models of disease pathogenesis; and (d) application of those biological insights to drive novel preventative and therapeutic options.

The first of these will involve documenting the full spectrum of natural genetic variation across all human populations, including capture of structural variants, and somatic mutations that accumulate with aging^{157,158}, and associating these variations with the ever-richer disease-related intermediate and clinical traits available through biobanks and electronic health records. It will be particularly important to include populations historically under-represented in genomic research, following the pioneering work of the H3Africa consortium¹⁵⁹. As over time, clinically sequenced genomes will outnumber those collected in academia, research and healthcare communities will need to develop a harmonized approach to genomics to transcend historical boundaries. Progress will be critically dependent on platforms and governance that lower barriers to the integration of genetic and phenotypic data across

Box 1

Global genomics

Present and future advances in genetics and genomics have the potential to provide benefits to individuals and societies across the world, but equitable and fair access to those benefits will require proactive measures to address entrenched disparities in scientific capacity and clinical opportunities. This includes:

(1) Global characterization of genetic variation. Systematic catalogues of human genome variation from a broad range of global populations will maximize the genetic diversity available for genetic discovery and clinical implementation. This will support more accurate imputation, enable more effective use of polygenic scores, and improve detection and interpretation of rare large-effect alleles.

(2) Genetic discovery for diseases with restricted geographical coverage. Genetic studies conducted in diverse populations will support genetic discovery as well as functional characterization of hitherto-neglected diseases with disproportionate regional impact and may lead to novel preventative and therapeutic strategies.

(3) An ethical framework for global research. Genetic research needs to proceed on the expectation that the benefits of research will be available to those who have participated, and that relationships between researchers and patients are based on robust expressions of accountability and governance. Participation should be based on appropriate informed individual consent in the context of community involvement, with mechanisms to support equitable data sharing with researchers in participating countries.

(4) Support for local research capacity. Collection of biological specimens and data are dependent on adequate support for data collection, generation and analysis in participating countries, with infrastructure, training and career structures in place to support local researchers and equitable credit for shared research outputs.

(5) Equitable translation. Participants and their communities should have fair and equitable access to the biomedical information arising from the studies in which they are involved to support population-relevant clinical diagnostics and implementation (for example, of polygenic scores). There should be commitment to the development of technologies and strategies that enable the clinical benefits of genomic medicine to be accessible to those in less-developed healthcare environments.

studies and countries, along with technical standards that are reliable, secure and compatible with the international regulatory landscape¹⁶⁰.

Mechanistic interpretation of genetic associations, particularly those in regulatory regions, will be driven by the systematic annotation of sequence variants and genes for functional impact across disease-relevant cell types, enabling mapping of processes contributing to disease development with respect to place (tissue and cell type), time (developmental stage) and context (external influences)¹⁶¹. Accelerating efforts to characterize the cellular composition of tissues through single-cell assays¹¹⁵ will increase the granularity of these observations. Large-scale perturbation studies across diverse cellular and animal models will, together with analyses of coding variants in humans^{53,54}, provide confidence in causal inference. Large-scale proteomic and metabolomic analyses (in tissues and biological fluids) will provide a bridge to downstream pathways^{79,80}. Research access to such functional data, generated at scale, should lower the barriers to mechanistic inference, provide system-wide context and enable researchers to focus wet-laboratory validation on the most critical experiments. Collectively, these efforts will support compilation of

a systematic catalogue of key networks and processes that influence normal physiology and disease development and inform a revised molecular taxonomy of disease.

This knowledge will reinforce the essential contribution of human genetics to the identification and prioritization of targets for therapeutic development^{89,162}. Insights into the efficacy of target perturbation and potential for adverse events, allied to characterization of translatable biomarkers, provide ways to boost the efficiency of drug-development pipelines¹⁶². Given the clinical importance of slowing disease progression¹⁶³, target-discovery efforts will increasingly need to embrace the genetics of disease progression and treatment response, as these may involve processes distinct from those captured by studies of disease onset.

In parallel, the clinical use of human genetics will benefit from progress towards universal determination of individual genome sequences built through a combination of biobank expansion and direct access within healthcare systems. This will power clinical applications that extend beyond the current focus on neonatal sequencing, Mendelian diagnostics and somatic tumour sequencing¹⁶⁴. In particular, improvements in polygenic score derivation will boost risk prediction for multifactorial traits, provide a molecular basis for disease classification, support biomarker discovery and therapeutic optimization and contribute to understanding of the variable penetrance of monogenic conditions⁶⁹. Implementing genomic medicine as a routine component of clinical care across diverse healthcare environments will inevitably require investment in the training of healthcare professionals and attention to optimal strategies for returning genetic findings to patients.

The limited heritability of many multifactorial traits constrains the clinical precision available from genetic data alone. This will drive efforts to integrate information on personal environment, lifestyle and behaviour, and to combine prognostic, predictive information on disease risk with longitudinal measures of molecular and clinical state that track an individual's journey from health to disease. Human genetics will also, given its unique potential for causal inference, support identification of the non-genetic risk factors (often modifiable) that directly contribute to disease predisposition and development¹⁶⁵. As polygenic score performance improves, analysis of individuals who show marked divergence between genetic predisposition and real-world clinical outcomes should define exposures (such as lifestyle choices or gut microbiome) the contribution of which to disease causation remains unclear¹⁶⁶.

Collectively, these developments can be expected to accelerate personalization of healthcare delivery. Provided costs are sustainable, a more preventative perspective on health could emerge, managed through proactive genomic, clinical and lifestyle surveillance using risk scores, complex biomarkers, liquid biopsies and wearables. Improved understanding of aetiological heterogeneity, patterns of sharing of genetic risk across diseases, variation in therapeutic response and risk of adverse events will enhance targeting of preventative and therapeutic interventions¹⁶⁷. At the population level, intervention strategies will seek to combine population-wide and targeted strategies to best effect¹⁶⁸. It will be critical to ensure that these benefits are available to as many as possible, so that genomics reduces, rather than exacerbates, national and global health disparities^{55,169} (Box 1).

The developments described above, represent variations on the theme of 'reading' the genome. The emerging capacity to block this reading (for example, through siRNA therapies¹⁷⁰) or even to 'write' the genome (through CRISPR editing) promises to be equally transformative, providing new opportunities to correct, and even cure, Mendelian disease. Spectacular advances in developing novel therapeutic strategies are likely for many diseases, based, for example, on ex vivo cellular manipulation¹⁷¹ or in vivo somatic cell editing¹⁷².

Importantly, developments in genomic medicine need to proceed in a bioethical framework for research and clinical use that recognizes

the personal relevance of human genetics and the critical importance of autonomous consent and the protection of privacy, while minimizing the adverse consequences of genetic exceptionalism. Governance needs to reaffirm the rights of citizens to make individual contributions to scientific progress through research participation and encourage the responsible exchange of data for clinical and research purposes.

Future prospects

Over the past two decades, understanding of the genetic basis of human disease has been transformed by a combination of spectacular technological and analytical advances, collaborative commitment to the development of foundational resources and the collection and analysis of vast amounts of genetic, molecular and clinical data. The biological insights derived from these data are, increasingly, drivers of translational innovation, and widening personal access to large-scale genetic and molecular data promises to reshape medical care.

However, for the full potential of genomic medicine to be realized, there will need to be sustained collaborative endeavour on several fronts to ensure that the capacity to generate ever more detailed maps of the relationships between sequence variation and biomedical phenotypes delivers a comprehensive understanding of disease mechanisms that can be translated into the medicines of tomorrow.

1. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
This paper describes the first analyses from the draft human genome sequence assembled over the previous decade: it launched modern human genetics and represents a tribute to the power of collaborative science.
2. International HapMap Consortium. The International HapMap Project. *Nature* **426**, 789–796 (2003).
The HapMap Consortium developed the first genome-wide maps of common sequence variation, using this information to lay out the haplotypic structure of this variation across three major ancestral groupings (from Europe, East Asia and Africa).
3. 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
4. The Haplotype Reference Consortium. A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283 (2016).
5. Lek, M. et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
6. Karczewski, K. J. et al. Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. Preprint at [bioRxiv](https://doi.org/10.1101/531210) <https://doi.org/10.1101/531210> (2019).
The most recent report from the genome aggregation database (gnomAD) project combining human sequence data on an unprecedented scale to characterize rare, high-impact variants and their relationship to health and disease.
7. Kremer, B. et al. A worldwide study of the Huntington's disease mutation. The sensitivity and specificity of measuring CAG repeats. *N. Engl. J. Med.* **330**, 1401–1406 (1994).
8. Collins, F. S. Identifying human disease genes by positional cloning. *Harvey Lect.* **86**, 149–164 (1990–1991).
9. Gusella, J. F. & MacDonald, M. E. Huntington's disease and repeating trinucleotides. *N. Engl. J. Med.* **330**, 1450–1451 (1994).
10. Vissers, L. E. et al. Array-based comparative genomic hybridization for the genomewide detection of submicroscopic chromosomal abnormalities. *Am. J. Hum. Genet.* **73**, 1261–1270 (2003).
11. Ng, S. B. et al. Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* **461**, 272–276 (2009).
12. Wheeler, D. A. et al. The complete genome of an individual by massively parallel DNA sequencing. *Nature* **452**, 872–876 (2008).
13. Vissers, L. E. et al. A de novo paradigm for mental retardation. *Nat. Genet.* **42**, 1109–1112 (2010).
14. Deciphering Developmental Disorders Study. Prevalence and architecture of de novo mutations in developmental disorders. *Nature* **542**, 433–438 (2017).
15. Splinter, K. et al. Effect of genetic diagnosis on patients with previously undiagnosed disease. *N. Engl. J. Med.* **379**, 2131–2139 (2018).
16. Deciphering Developmental Disorders Study. Large-scale discovery of novel genetic causes of developmental disorders. *Nature* **519**, 223–228 (2015).
17. Firth, H. V. et al. DECIPHER: database of chromosomal imbalance and phenotype in humans using ensembl resources. *Am. J. Hum. Genet.* **84**, 524–533 (2009).
18. Philippakis, A. A. et al. The Matchmaker Exchange: a platform for rare disease gene discovery. *Hum. Mutat.* **36**, 915–921 (2015).
19. Boycott, K. M. et al. International cooperation to enable the diagnosis of all rare genetic diseases. *Am. J. Hum. Genet.* **100**, 695–705 (2017).
20. Kennedy, J. et al. KAT6A syndrome: genotype–phenotype correlation in 76 patients with pathogenic KAT6A variants. *Genet. Med.* **21**, 850–860 (2019).
21. Amberger, J. S., Bocchini, C. A., Schiettecatte, F., Scott, A. F. & Hamosh, A. OMIM.org: Online Mendelian Inheritance in Man (OMIM), an online catalog of human genes and genetic disorders. *Nucleic Acids Res.* **43**, D789–D798 (2015).
22. Landrum, M. J. et al. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res.* **42**, D980–D985 (2014).
23. Rehms, H. L. et al. ClinGen—the Clinical Genome Resource. *N. Engl. J. Med.* **372**, 2235–2242 (2015).
24. Minikel, E. V. et al. Quantifying prior disease penetrance using large population control cohorts. *Sci. Transl. Med.* **8**, 322ra9 (2016).
25. Flannick, J. et al. Assessing the phenotypic effects in the general population of rare variants in genes for a dominant Mendelian form of diabetes. *Nat. Genet.* **45**, 1380–1385 (2013).
26. Bećanović, K. et al. A SNP in the *HTT* promoter alters NF-κB binding and is a bidirectional genetic modifier of Huntington disease. *Nat. Neurosci.* **18**, 807–816 (2015).
27. Castel, S. E. et al. Modified penetrance of coding variants by *cis*-regulatory variation contributes to disease risk. *Nat. Genet.* **50**, 1327–1334 (2018).
28. MacArthur, D. G. et al. Guidelines for investigating causality of sequence variants in human disease. *Nature* **508**, 469–476 (2014).
29. Findlay, G. M. et al. Accurate classification of *BRCA1* variants with saturation genome editing. *Nature* **562**, 217–222 (2018).
This paper describes an effort to obtain massively parallel functional assessments of all potential sequence variants in a gene causal for familial breast cancer, and thereby provide more confident predictions of clinical significance for those found to carry a previously unseen mutation.
30. Cummings, B. B. et al. Improving genetic diagnosis in Mendelian disease with transcriptome sequencing. *Sci. Transl. Med.* **9**, eaal5209 (2017).
31. Butcher, D. T. et al. CHARGE and Kabuki syndromes: gene-specific DNA methylation signatures identify epigenetic mechanisms linking these clinically overlapping conditions. *Am. J. Hum. Genet.* **100**, 773–788 (2017).
32. Short, P. J. et al. De novo mutations in regulatory elements in neurodevelopmental disorders. *Nature* **555**, 611–616 (2018).
33. Jaganathan, K. et al. Predicting splicing from primary sequence with deep learning. *Cell* **176**, 535–548 (2019).
34. Hugot, J. P. et al. Association of *NOD2* leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* **411**, 599–603 (2001).
35. Ogura, Y. et al. A frameshift mutation in *NOD2* associated with susceptibility to Crohn's disease. *Nature* **411**, 603–606 (2001).
36. Barbosa, J., Chern, M. M., Noreen, H. & Anderson, V. E. Analysis of linkage between the major histocompatibility system and juvenile, insulin-dependent diabetes in multiplex families. Reanalysis of data. *J. Clin. Invest.* **62**, 492–495 (1978).
37. Risch, N. & Merikangas, K. The future of genetic studies of complex human diseases. *Science* **273**, 1516–1517 (1996).
A highly influential commentary that switched much of the complex trait genetics field from linkage to association and imagined GWAS several years before the technology made them a reality.
38. Klein, R. J. et al. Complement factor H polymorphism in age-related macular degeneration. *Science* **308**, 385–389 (2005).
Arguably the first full GWAS, this study demonstrated the potential of agnostic genomic surveys to highlight entirely novel biology—in this case, a role for complement in the pathogenesis of macular degeneration.
39. Duerr, R. H. et al. A genome-wide association study identifies *IL23R* as an inflammatory bowel disease gene. *Science* **314**, 1461–1463 (2006).
40. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661–678 (2007).
This paper, which describes what was, at the time, the largest GWAS yet conducted, demonstrated the broad applicability of the approach, and set the scene for the decade of GWAS discovery that followed.
41. Mahajan, A. et al. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat. Genet.* **50**, 1505–1513 (2018).
42. Astle, W. J. et al. The allelic landscape of human blood cell trait variation and links to common complex disease. *Cell* **167**, 1415–1429 (2016).
43. Lango Allen, H. et al. Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* **467**, 832–838 (2010).
44. Buñiello, A. et al. The NHGRI-EBI GWAS catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* **47**, D1005–D1012 (2019).
45. Fuchsberger, C. et al. The genetic architecture of type 2 diabetes. *Nature* **536**, 41–47 (2016).
46. UK10K Consortium. The UK10K project identifies rare variants in health and disease. *Nature* **526**, 82–90 (2015).
47. Manolio, T. A. et al. Finding the missing heritability of complex diseases. *Nature* **461**, 747–753 (2009).
48. Wainschtein, P. et al. Recovery of trait heritability from whole genome sequence data. Preprint at [bioRxiv](https://doi.org/10.1101/588020) <https://doi.org/10.1101/588020> (2019).
49. Flannick, J. et al. Loss-of-function mutations in *SLC30A8* protect against type 2 diabetes. *Nat. Genet.* **46**, 357–363 (2014).
50. Moltke, I. et al. A common Greenlandic *TBC1D4* variant confers muscle insulin resistance and type 2 diabetes. *Nature* **512**, 190–193 (2014).
51. Nioi, P. et al. Variant *ASGR1* associated with a reduced risk of coronary artery disease. *N. Engl. J. Med.* **374**, 2131–2141 (2016).
52. Pollin, T. I. et al. A null mutation in human *APOC3* confers a favorable plasma lipid profile and apparent cardioprotection. *Science* **322**, 1702–1705 (2008).
53. Narasimhan, V. M. et al. Health and population effects of rare gene knockouts in adult humans with related parents. *Science* **352**, 474–477 (2016).
A powerful illustration of the insights that can be gained from the study of individuals who are homozygous carriers of null alleles in genes of biomedical interest.
54. Saleheen, D. et al. Human knockouts and phenotypic analysis in a cohort with a high rate of consanguinity. *Nature* **544**, 235–239 (2017).

55. Martin, A. R. et al. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat. Genet.* **51**, 584–591 (2019).
56. Zeng, J. et al. Signatures of negative selection in the genetic architecture of human complex traits. *Nat. Genet.* **50**, 746–753 (2018).
57. Genovese, G. et al. Association of trypanolytic ApoL1 variants with kidney disease in African Americans. *Science* **329**, 841–845 (2010).
58. Luzzatto, L., Usanga, F. A. & Reddy, S. Glucose-6-phosphate dehydrogenase deficient red cells: resistance to infection by malarial parasites. *Science* **164**, 839–842 (1969).
59. Wojcik, G. L. et al. Genetic analyses of diverse populations improves discovery for complex traits. *Nature* **570**, 514–518 (2019).
60. Farh, K. K. et al. Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* **518**, 337–343 (2015).
- An early demonstration of the value of overlapping genome-wide association data with tissue-specific regulatory maps to define the cell-types and tissues likely to be driving disease pathology.**
61. Freund, M. K. et al. Phenotype-specific enrichment of Mendelian disorder genes near GWAS regions across 62 complex traits. *Am. J. Hum. Genet.* **103**, 535–552 (2018).
62. Boyle, E. A., Li, Y. I. & Pritchard, J. K. An expanded view of complex traits: from polygenic to omnigenic. *Cell* **169**, 1177–1186 (2017).
63. de Lange, K. M. et al. Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. *Nat. Genet.* **49**, 256–261 (2017).
64. Sekar, A. et al. Schizophrenia risk from complex variation of complement component 4. *Nature* **530**, 177–183 (2016).
65. Jostins, L. et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* **491**, 119–124 (2012).
66. Richardson, T. G., Harrison, S., Hemani, G. & Davey Smith, G. An atlas of polygenic risk score associations to highlight putative causal relationships across the human phenome. *eLife* **8**, e43657 (2019).
67. Dendrou, C. A. et al. Resolving TYK2 locus genotype-to-phenotype differences in autoimmunity. *Sci. Transl. Med.* **8**, 363ra149 (2016).
68. Voight, B. F. et al. Plasma HDL cholesterol and risk of myocardial infarction: a Mendelian randomisation study. *Lancet* **380**, 572–580 (2012).
69. Khera, A. V. et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat. Genet.* **50**, 1219–1224 (2018).
- This paper used large-scale genetic data to demonstrate the clinical potential of the polygenic scores that can be constructed for many common diseases, emphasizing that, in some situations, the lifetime risk of disease for those with the highest scores approaches that of established monogenic disease.**
70. Mahmood, S. S., Levy, D., Vasan, R. S. & Wang, T. J. The Framingham Heart Study and the epidemiology of cardiovascular disease: a historical perspective. *Lancet* **383**, 999–1008 (2014).
71. Gudbjartsson, D. F. et al. Variants conferring risk of atrial fibrillation on chromosome 4q25. *Nature* **448**, 353–357 (2007).
72. Thorgeirsson, T. E. et al. A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. *Nature* **452**, 638–642 (2008).
73. Collins, F. S. & Varmus, H. A new initiative on precision medicine. *N. Engl. J. Med.* **372**, 793–795 (2015).
74. Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
75. Nagai, A. et al. Overview of the BioBank Japan Project: study design and profile. *J. Epidemiol.* **27**, S2–S8 (2017).
76. Chen, Z. et al. China Kadoorie Biobank of 0.5 million people: survey methods, baseline characteristics and long-term follow-up. *Int. J. Epidemiol.* **40**, 1652–1666 (2011).
77. Kohane, I. S. Using electronic health records to drive discovery in disease genomics. *Nat. Rev. Genet.* **12**, 417–428 (2011).
78. Abul-Husn, N. S. & Kenny, E. E. Personalized medicine and the power of electronic health records. *Cell* **177**, 58–69 (2019).
79. Sun, B. B. et al. Genomic atlas of the human plasma proteome. *Nature* **558**, 73–79 (2018).
80. Shin, S. Y. et al. An atlas of genetic influences on human blood metabolites. *Nat. Genet.* **46**, 543–550 (2014).
81. Sung, Y. J. et al. A multi-ancestry genome-wide study incorporating gene-smoking interactions identifies multiple new loci for pulse pressure and mean arterial pressure. *Hum. Mol. Genet.* **28**, ddz070 (2019).
82. Pendergrass, S. A. & Crawford, D. C. Using electronic health records to generate phenotypes for research. *Curr. Protoc. Hum. Genet.* **100**, e80 (2019).
83. Denny, J. C. et al. PheWAS: demonstrating the feasibility of a genome-wide scan to discover gene-disease associations. *Bioinformatics* **26**, 1205–1210 (2010).
84. Liu, Z. et al. Mendelian randomization analysis dissects the relationship between NAFLD, T2D and obesity and provides implications for precision medicine. Preprint at *bioRxiv* <https://doi.org/10.1101/657734> (2019).
85. Cotsapas, C. et al. Pervasive sharing of genetic effects in autoimmune disease. *PLoS Genet.* **7**, e1002254 (2011).
86. Kanai, M. et al. Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. *Nat. Genet.* **50**, 390–400 (2018).
87. Gudmundsson, J. et al. Two variants on chromosome 17 confer prostate cancer risk, and the one in *TCF2* protects against type 2 diabetes. *Nat. Genet.* **39**, 977–983 (2007).
88. Abul-Husn, N. S. et al. A protein-truncating *HSD17B13* variant and protection from chronic liver disease. *N. Engl. J. Med.* **378**, 1096–1106 (2018).
89. Plenge, R. M., Scolnick, E. M. & Altshuler, D. Validating therapeutic targets through human genetics. *Nat. Rev. Drug Discov.* **12**, 581–594 (2013).
90. Nielsen, J. B. et al. Biobank-driven genomic discovery yields new insight into atrial fibrillation biology. *Nat. Genet.* **50**, 1234–1239 (2018).
91. Flannick, J. et al. Exome sequencing of 20,791 cases of type 2 diabetes and 24,440 controls. *Nature* **570**, 71–76 (2019).
92. Van Hout, C. V. et al. Whole exome sequencing and characterization of coding variation in 49,960 individuals in the UK Biobank. Preprint at *bioRxiv* <https://doi.org/10.1101/572347> (2019).
93. Khera, A. V. et al. Whole-genome sequencing to characterize monogenic and polygenic contributions in patients hospitalized with early-onset myocardial infarction. *Circulation* **139**, 1593–1602 (2019).
94. Grarup, N. et al. Loss-of-function variants in *ADCY3* increase risk of obesity and type 2 diabetes. *Nat. Genet.* **50**, 172–174 (2018).
95. Cohen, J. C., Boerwinkle, E., Mosley, T. H. Jr & Hobbs, H. H. Sequence variations in *PCSK9*, low LDL, and protection against coronary heart disease. *N. Engl. J. Med.* **354**, 1264–1272 (2006).
96. Rivas, M. A. et al. Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. *Nat. Genet.* **43**, 1066–1073 (2011).
97. Chong, J. X. et al. The genetic basis of Mendelian phenotypes: discoveries, challenges, and opportunities. *Am. J. Hum. Genet.* **97**, 199–215 (2015).
98. Goldstein, J. L. & Brown, M. S. A century of cholesterol and coronaries: from plaques to genes to statins. *Cell* **161**, 161–172 (2015).
99. Peltonen, L., Perola, M., Naukkarinen, J. & Palotie, A. Lessons from studying monogenic disease for common disease. *Hum. Mol. Genet.* **15**, R67–R74 (2006).
100. Bulik-Sullivan, B. et al. An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
101. Musunuru, K. et al. From noncoding variant to phenotype via *SORT1* at the 1p13 cholesterol locus. *Nature* **466**, 714–719 (2010).
- Widely accepted as the first paper to describe the detailed mechanistic dissection of a complex trait locus and still one of relatively few examples of success in this endeavour.**
102. Claussnitzer, M. et al. *FTO* obesity variant circuitry and adipocyte browning in humans. *N. Engl. J. Med.* **373**, 895–907 (2015).
103. Smemo, S. et al. Obesity-associated variants within *FTO* form long-range functional connections with *IRX3*. *Nature* **507**, 371–375 (2014).
104. Gupta, R. M. et al. A genetic variant associated with five vascular diseases is a distal regulator of endothelin-1 gene expression. *Cell* **170**, 522–533 (2017).
105. Small, K. S. et al. Regulatory variants at *KLF14* influence type 2 diabetes risk via a female-specific effect on adipocyte size and body composition. *Nat. Genet.* **50**, 572–580 (2018).
106. Lin, Y. T. et al. *APOE4* causes widespread molecular and cellular alterations associated with Alzheimer's disease phenotypes in human iPSC-derived brain cell types. *Neuron* **98**, 1294 (2018).
107. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
108. Roadmap Epigenomics Consortium et al. Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
- The flagship paper from the NIH Roadmap Epigenomics Consortium, describing a detailed analysis of the epigenomic profiles of more than 100 human cell types and characterizing how these regulatory patterns relate to gene regulation, cellular differentiation and human disease.**
109. Tansey, K. E., Cameron, D. & Hill, M. J. Genetic risk for Alzheimer's disease is concentrated in specific macrophage and microglial transcriptional networks. *Genome Med.* **10**, 14 (2018).
110. Gamazon, E. R. et al. Using an atlas of gene regulation across 44 human tissues to inform complex disease- and trait-associated variation. *Nat. Genet.* **50**, 956–967 (2018).
111. Ongen, H. et al. Estimating the causal tissues for complex traits and diseases. *Nat. Genet.* **49**, 1676–1683 (2017).
112. GTEx Consortium. Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).
- One of several papers describing the results obtained from the Genotype-Tissue Expression (GTEx) project, which used RNA-sequence data from more than 40 human tissues in several hundred individuals to explore the relationship between DNA sequence variation and tissue-specific expression.**
113. Chiou, J. et al. Single cell chromatin accessibility reveals pancreatic islet cell type- and state-specific regulatory programs of diabetes risk. Preprint at *bioRxiv* <https://doi.org/10.1101/693671> (2019).
114. HuBMAP Consortium. The human body at cellular resolution: the NIH Human Biomolecular Atlas Program. *Nature* **574**, 187–192 (2019).
115. Regev, A. et al. The Human Cell Atlas. *eLife* **6**, e27041 (2017).
116. Lloyd, K. C. A knockout mouse resource for the biomedical research community. *Ann. NY Acad. Sci.* **1245**, 24–26 (2011).
117. Finucane, H. K. et al. Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat. Genet.* **50**, 621–629 (2018).
118. Delaneau, O. et al. Chromatin three-dimensional interactions mediate genetic effects on gene expression. *Science* **364**, eaat8266 (2019).
119. Zhou, J. & Troyanskaya, O. G. Predicting effects of noncoding variants with deep learning-based sequence model. *Nat. Methods* **12**, 931–934 (2015).
120. Kelley, D. R., Snoek, J. & Rinn, J. L. Basset: learning the regulatory code of the accessible genome with deep convolutional neural networks. *Genome Res.* **26**, 990–999 (2016).
121. Patwardhan, R. P. et al. High-resolution analysis of DNA regulatory elements by synthetic saturation mutagenesis. *Nat. Biotechnol.* **27**, 1173–1175 (2009).
122. Melnikov, A. et al. Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. *Nat. Biotechnol.* **30**, 271–277 (2012).
123. Wang, X. et al. High-resolution genome-wide functional dissection of transcriptional regulatory regions and nucleotides in human. *Nat. Commun.* **9**, 5380 (2018).
124. Potting, C. et al. Genome-wide CRISPR screen for PARKIN regulators reveals transcriptional repression as a determinant of mitophagy. *Proc. Natl Acad. Sci. USA* **115**, E180–E189 (2018).
125. Gasperini, M. et al. A genome-wide framework for mapping gene regulation via cellular genetic screens. *Cell* **176**, 377–390 (2019).

126. Fulco, C. P. et al. Activity-by-contact model of enhancer-promoter regulation from thousands of CRISPR perturbations. *Nat. Genet.* **51**, 1664–1669 (2019).
 127. Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420–424 (2016).
 128. Feldman, D. et al. Optical pooled screens in human cells. *Cell* **179**, 787–799 (2019).
 129. Dixit, A. et al. Perturb-Seq: dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. *Cell* **167**, 1853–1866 (2016).
 130. Datlinger, P. et al. Pooled CRISPR screening with single-cell transcriptome readout. *Nat. Methods* **14**, 297–301 (2017).
 131. Ding, L. et al. Perspective on oncogenic processes at the end of the beginning of cancer genomics. *Cell* **173**, 305–320 (2018).
 132. Chae, Y. K. et al. Path toward precision oncology: review of targeted therapy studies and tools to aid in defining “actionability” of a molecular lesion and patient management support. *Mol. Cancer Ther.* **16**, 2645–2655 (2017).
 133. Shearer, A. E. & Smith, R. J. Massively parallel sequencing for genetic diagnosis of hearing loss: the new standard of care. *Otolaryngol. Head Neck Surg.* **153**, 175–182 (2015).
 134. Yang, Y. et al. Molecular findings among patients referred for clinical whole-exome sequencing. *J. Am. Med. Assoc.* **312**, 1870–1879 (2014).
 135. Friedman, J. M. et al. Genome-wide sequencing in acutely ill infants: genomic medicine's critical application? *Genet. Med.* **21**, 498–504 (2019).
 136. Clark, M. M. et al. Diagnosis of genetic diseases in seriously ill children by rapid whole-genome sequencing and automated phenotyping and interpretation. *Sci. Transl. Med.* **11**, eaat6177 (2019).
 137. Eldomery, M. K. et al. Lessons learned from additional research analyses of unsolved clinical exome cases. *Genome Med.* **9**, 26 (2017).
 138. Boycott, K. M. et al. A diagnosis for all rare genetic diseases: the horizon and the next frontiers. *Cell* **177**, 32–37 (2019).
 139. Liu, S. et al. Genomic analyses from non-invasive prenatal testing reveal genetic associations, patterns of viral infections, and Chinese population history. *Cell* **175**, 347–359 (2018).
 140. Thung, D. T., Beulen, L., Hehir-Kwa, J. & Faas, B. H. Implementation of whole genome massively parallel sequencing for noninvasive prenatal testing in laboratories. *Expert Rev. Mol. Diagn.* **15**, 111–124 (2015).
 141. Lo, J. O., Cori, D. F., Norton, M. E. & Caughey, A. B. Noninvasive prenatal testing. *Obstet. Gynecol. Surv.* **69**, 89–99 (2014).
 142. Watson, M. S. et al. Newborn screening: toward a uniform screening panel and system—executive summary. *Pediatrics* **117**, S296–S307 (2006).
 143. Currier, R. J. et al. Genomic sequencing in cystic fibrosis newborn screening: what works best, two-tier predefined CFTR mutation panels or second-tier CFTR panel followed by third-tier sequencing? *Genet. Med.* **19**, 1159–1163 (2017).
 144. Bauer, P. E. “Tell them it's not so bad”: prenatal screening for Down syndrome and the bias toward abortion. *Intellect. Dev. Disabil.* **46**, 247–251 (2008).
 145. Volpi, S. et al. Research directions in the clinical implementation of pharmacogenomics: an overview of US Programs and Projects. *Clin. Pharmacol. Ther.* **103**, 778–786 (2018).
 146. International Schizophrenia Consortium. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748–752 (2009).
 147. Evans, D. M., Visscher, P. M. & Wray, N. R. Harnessing the information contained within genome-wide association studies to improve individual prediction of complex disease risk. *Hum. Mol. Genet.* **18**, 3525–3531 (2009).
 148. Kurian, A. W. et al. Genetic testing and results in a population-based cohort of breast cancer patients and ovarian cancer patients. *J. Clin. Oncol.* **37**, 1305–1315 (2019).
 149. Katz, S. J. et al. Association of attending surgeon with variation in the receipt of genetic testing after diagnosis of breast cancer. *JAMA Surg.* **153**, 909–916 (2018).
 150. Kalokairinou, L. et al. Legislation of direct-to-consumer genetic testing in Europe: a fragmented regulatory landscape. *J. Community Genet.* **9**, 117–132 (2018).
 151. Allyse, M. A., Robinson, D. H., Ferber, M. J. & Sharp, R. R. Direct-to-Consumer Testing 2.0: emerging models of direct-to-consumer genetic testing. *Mayo Clin. Proc.* **93**, 113–120 (2018).
 152. Wright, C. F. et al. Assessing the pathogenicity, penetrance, and expressivity of putative disease-causing variants in a population setting. *Am. J. Hum. Genet.* **104**, 275–286 (2019).
 153. ACMG Board of Directors. The use of ACMG secondary findings recommendations for general population screening: a policy statement of the American College of Medical Genetics and Genomics (ACMG). *Genet. Med.* **21**, 1467–1468 (2019).
 154. George, A., Kaye, S. & Banerjee, S. Delivering widespread BRCA testing and PARP inhibition to patients with ovarian cancer. *Nat. Rev. Clin. Oncol.* **14**, 284–296 (2017).
 155. CenterWatch. *FDA Approved Drugs for Genetic Disease* (CenterWatch); <https://www.centerwatch.com/drug-information/fda-approved-drugs/therapeutic-area/34/genetic-disease>
 156. Wu, Y. et al. Highly efficient therapeutic gene editing of human hematopoietic stem cells. *Nat. Med.* **25**, 776–783 (2019).
 157. Jaiswal, S. et al. Clonal hematopoiesis and risk of atherosclerotic cardiovascular disease. *N. Engl. J. Med.* **377**, 111–121 (2017).
 158. Lee, M. H. et al. Somatic APP gene recombination in Alzheimer's disease and normal neurons. *Nature* **563**, 639–645 (2018).
 159. H3Africa Consortium. Enabling the genomic revolution in Africa. *Science* **344**, 1346–1348 (2014).
 160. Dolman, L. et al. ClinGen advancing genomic data-sharing standards as a GA4GH driver project. *Hum. Mutat.* **39**, 1686–1689 (2018).
 161. Fairfax, B. P. et al. Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. *Science* **343**, 1246949 (2014).
 162. Nelson, M. R. et al. The support of human genetic evidence for approved drug indications. *Nat. Genet.* **47**, 856–860 (2015).
- An influential paper that demonstrates that the probability of therapeutic success is enhanced (by around a factor of 2) for drug targets with supportive evidence from human genetics.**
163. Veith, D. P. et al. Understanding disease progression and improving Alzheimer's disease clinical trials: recent highlights from the Alzheimer's Disease Neuroimaging Initiative. *Alzheimers Dement.* **15**, 106–152 (2019).
 164. Ashley, E. A. Towards precision medicine. *Nat. Rev. Genet.* **17**, 507–522 (2016).
 165. Franks, P. W. & McCarthy, M. I. Exposing the exposures responsible for type 2 diabetes and obesity. *Science* **354**, 69–73 (2016).
 166. Sanna, S. et al. Causal relationships among the gut microbiome, short-chain fatty acids and metabolic diseases. *Nat. Genet.* **51**, 600–605 (2019).
 167. Udler, M. S. et al. Type 2 diabetes genetic loci informed by multi-trait associations point to disease mechanisms and subtypes: a soft clustering analysis. *PLoS Med.* **15**, e1002654 (2018).
 168. Rose, G. Sick individuals and sick populations. *Int. J. Epidemiol.* **14**, 32–38 (1985).
 169. Hindorf, L. A. et al. Prioritizing diversity in human genomics research. *Nat. Rev. Genet.* **19**, 175–185 (2018).
 170. Wittrup, A. & Lieberman, J. Knocking down disease: a progress report on siRNA therapeutics. *Nat. Rev. Genet.* **16**, 543–552 (2015).
 171. Schuster, S. J. et al. Chimeric antigen receptor T cells in refractory B-cell lymphomas. *N. Engl. J. Med.* **377**, 2545–2554 (2017).
 172. Ding, Q. et al. Permanent alteration of PCSK9 with in vivo CRISPR–Cas9 genome editing. *Circ. Res.* **115**, 488–492 (2014).
 173. Risca, V. I. & Greenleaf, W. J. Unraveling the 3D genome: genomics tools for multiscale exploration. *Trends Genet.* **31**, 357–372 (2015).
- Acknowledgements** We acknowledge grant funding from the following funders. J.H.C.: NIH (U01DK062429, U01DK062422 and R01DK106593); N.J.C.: NIH (U01HG009086, U54MD010722 and R01MH113362); E.T.D.: Swiss National Science Foundation, Louis Jeantet Foundation; E.E.K.: NIH (R01HG010297, R01HL104608, U01HG009610, R01DK110113, U01HG009080, X01HL1345 and UM1HG0089001); C.M.L.: NIHR Oxford Biomedical Research Centre and NIH (5P50HD028138-27); K.N.N.: NHMRC (APP113531); S.E.P.: NIH (U01HG006485 and 1U41HG009649); C.N.R.: NIH Intramural Program at the Center for Research on Genomics and Global Health and National Human Genome Research Institute; and M.I.M.: Wellcome (090532, 098381, 106130, 203141 and 212259) and NIH (U01-DK105535, DK085545 and DK098032). Personal funding comes from the following sources. M.C.: Next Generation Fund at the Broad Institute of MIT and Harvard; J.H.C.: Sanford Grossman Charitable Trust and Helmsley Charitable Trust; R.C.: UKBiobank; C.M.L.: Li Ka Shing Foundation; J.S.: Howard Hughes Medical Institute; and M.I.M.: Wellcome and NIHR.
- Author contributions** M.C. and M.I.M. coordinated the drafting of the Review. All other authors contributed to drafting of the text and approved the final version.
- Competing interests** R.C. has received research grants from British Heart Foundation, Cancer Research UK, Medical Research Council, Merck & Co, UKBiobank, Wellcome and Medco; a Pfizer Prize Award (to NDPH) and is named on a patent for a statin-related myopathy genetic test. R.C. receives no personal remuneration from these: any share of royalties or other payments have been waived in favour of NDPH. E.T.D. is chairman and board member of Hybridstat and on the advisory board of DNAnexus. M.E.H. is a co-founder, shareholder and director of Congenica. S.K. has received research grants from Bayer and Novartis; is on Scientific Advisory Boards of Regeneron Genetics Center, Corvidia Therapeutics and Maze Therapeutics; has equity in San Therapeutics, Catabasis, Verve and Maze Therapeutics and is a consultant for Maze Therapeutics, Alynham, ExpertConnect, Leerink Partners, Noble Insights, Bayer and Novo Ventures. E.E.K. receives honoraria from Illumina and Regeneron Pharmaceuticals. C.M.L. has research collaborations with Novo Nordisk and Bayer, receiving no personal payment. D.G.M. is co-founder and shareholder of Goldfinch Bio and has received research funding from AbbVie, Biogen, BioMarin, Merck, Pfizer and Sanofi-Genzyme. S.E.P. is a member of the Scientific Advisory Board of Baylor Genomics. J.S. is a member of the Scientific Advisory Board for Maze Therapeutics, Camp4 Therapeutics, Nanostring, Phase Genomics, Adaptive Biotechnology and Stratos Genomics, a founder of Phase Genomics and a consultant for Guardant Health. M.I.M. was a member of advisory panels for Pfizer, NovoNordisk and Zee Global; received honoraria from Merck, Pfizer, NovoNordisk and Eli Lilly and research funding from AbbVie, Astra Zeneca, Boehringer Ingelheim, Eli Lilly, Janssen, Merck, NovoNordisk, Pfizer, Roche, Sanofi Aventis, Servier and Takeda. As of June 2019, M.I.M. is an employee of Genentech and a holder of Roche stock. The views expressed in this article are those of the author(s) and not necessarily those of the NHS, the NIHR, or the Department of Health.
- Additional information**
- Correspondence and requests for materials** should be addressed to M.I.M.
- Peer review information** Nature thanks Ewan Birney, Jeff Schloss, Richard Trembath and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.
- Reprints and permissions information** is available at <http://www.nature.com/reprints>.
- Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© Springer Nature Limited 2020