

Translational genomics: from genetic discovery to translational impact

Eleftheria Zeggini¹, Anna L Gloyn^{2,3,4}, Anne C Barton^{5,6}, Louise V Wain^{7,8}

¹Institute of Translational Genomics, Helmholtz Zentrum München – German Research Center for Environmental Health, Neuherberg, Germany

²Oxford Centre for Diabetes Endocrinology & Metabolism, Oxford University, Oxford, UK

³Wellcome Centre for Human Genetics, University of Oxford, Oxford, UK

³Oxford NIHR Biomedical Research Centre, Churchill Hospital, Oxford, UK

⁵Versus Arthritis Centre for Genetics and Genomics, Centre for Musculoskeletal Research, Manchester Academic Health Science Centre, University of Manchester, Manchester, UK

⁶NIHR Manchester Biomedical Research Centre, Manchester University NHS Foundation Trust, Manchester, UK

⁷Department of Health Sciences, University of Leicester, Leicester, UK

⁸National Institute for Health Research, Leicester Respiratory Biomedical Research Centre, Glenfield Hospital, Leicester, UK

ABSTRACT

Translational genomics can help catalyse the convergence of the digital and genomic revolutions with healthcare. Predicting disease risk by means of polygenic risk scores integrated with classical epidemiological risk factors can lead to better risk stratification and clinical decision-making. A deeper understanding of genome-wide sequence and association with well-characterised phenotypes can empower diagnosis. Better knowledge of disease progression trajectories can enable the development of prognostic markers and tailored interventions. Understanding modifiable and non-modifiable risk factors can lead to the design of prophylactic approaches. The identification of genetic factors important in disease development offers insights into aetiopathogenesis and can lead to therapeutic target discovery. Repositioning existing treatments that act on targets identified through genomics approaches but for different indications can accelerate the path to effective therapies. Finally, the discovery of markers for response to treatment can culminate in tailored disease management.

INTRODUCTION

Progress in the field of human genetics has been accelerated by a revolution in technological advances, enabling the genome-wide interrogation of sequence variation, and leading to the discovery of thousands of genetic risk loci. Methodological advances have also enabled deep molecular characterisation of disease-relevant primary tissues collected from patients, or studied in the context of cellular and organismal models of disease. Powerful integration of these approaches can enhance our understanding of the genetic and genomic aetiology of disease development and progression, moving from discovery to functional interpretation and ultimately to clinical application. The coming decade signposts a new era that is both exciting and promising, and its success will be underpinned by data generation at scale; the development of computational toolkits to process the wealth of information; addressing ethical, legal, social and economic considerations; and effective integration into routine clinical practice.

The field of cancer research has blazed a trail in translating insights from genomics into mechanisms of disease development and progression, shortening the path to translation and empowering precision medicine. Genome sequencing coupled to multi-omics-based molecular profiling have led to better patient stratification and prognostication, novel drug target identification, development of personalised interventions, enhanced understanding of disease biology, better prediction of response to treatment, and therefore improved standard of care in oncology practice. Precision medicine approaches are already possible for some rare monogenic diseases. For example, for cystic fibrosis, ivacaftor was the first drug licensed for use in patients with specific gating mutations (1). On the other end of the spectrum, genetically-driven precision medicine approaches for common complex diseases are now starting to emerge and have the potential to have a major impact on patient care. In this review, we focus on complex disease genetics and explore how the discoveries made possible by the convergence of the digital and genomics revolutions can fuel the continued achievement of translational milestones to improve human health [Figure 1].

APPLICATIONS OF GENETICS TO TRANSLATION

Novel target discovery

Genome-wide association studies (GWAS) in complex diseases and medically-relevant quantitative traits have been successful in identifying robustly-replicating associations between sequence variants and the phenotypes of interest. Individually, these associations characteristically confer modest to small effects, although there are notable exceptions beyond *HLA* alleles (2, 3). In addition, there is an increasing number of reported rare variants, or population-specific common variants with large effects on complex traits like osteoarthritis (4) and type 2 diabetes (T2D) (5). Generally, the modest-to-small effects typically detected can highlight biological pathways that, when targeted by drugs, can have large therapeutic effects. The potential for genetic studies to improve the time and cost-efficiency of drug development is well-established (6, 7). An efficient use of genetic association data in the drug development pipeline is to inform decision-making on further investment for potential targets at the preclinical stages. However, there is as-yet limited evidence of genetic signals leading directly to new treatments. The most widely-described examples of **drug targets** with genetic support arise from **Mendelian disease** studies (for example, *PCSK9*) (8). Nevertheless, complex trait signals frequently implicate genes already known to be important in disease, or genes which encode known drug targets [Figure 2], providing proof-of-concept of the ability to identify potentially new druggable targets.

The GWAS catalogue, one of the largest repositories of genetic association data, currently contains approximately 138,000 signals from 4000 studies. Whilst each of these signals could act as a signpost to genes and pathways important in the disease or trait being studied, making the link between the signal and the specific gene(s) underlying the association is a major bottleneck between signal

discovery and realisation of translational potential. Although some signals might very clearly implicate a particular gene, for example, via an experimentally-validated functional amino acid change or introduction of a stop codon, the majority of genetic signals are either very broad, encompassing many variants across multiple genes, or are outside of gene regions entirely.

Computational (*in silico*) approaches can be used to begin to map signals to effector genes, empowered by increasing sample sizes, deeply sequenced large imputation haplotype reference panels and the availability of genetic data across diverse populations. The increasing availability of multi-omics data across cell types and tissues has further facilitated this. Statistical **fine-mapping**, followed by interrogation of molecular quantitative trait loci (e.g. **eQTL** or **pQTL** data), annotation of putative regulatory regions (e.g. through **epigenetics**) and **chromosome interactions**, have allowed identification of potentially causal variants and genes at a large number of **GWAS signals** [Figure 3]. For example, in T2D, incorporation of functional annotation from human pancreatic islets has enabled fine-mapping of 20% of T2D loci to fewer than 5 variants (9). Whilst these give valuable early insight into the underlying mechanisms that lead to the genetic signal and inform the design of targeted molecular experiments, they are often limited by the range of molecular data and disease-relevant tissues that are available, as well as by reproducibility and **statistical power**. Where multiple lines of *in silico* evidence do point towards a causative role for a specific variant, gene or pathway, the next crucial step is to undertake laboratory experiments to confirm (or refute) the computational evidence, and to identify the mechanism through which the causal variant(s) affect the gene (giving rise to the observed association), and the role of the gene in the disease [Figure 2].

Functional characterisation of arising genetic signals can take many years. Indeed, although the first GWAS signals began to emerge around 15 years ago, very few have yet been characterised in terms of mechanism with any certainty. *FTO* harbours the strongest signal of association for obesity, and was first reported in 2007 (10). Subsequent studies attempted to understand the role of *FTO* and other nearby genes in obesity but it was not until 2015 that a connection between the original signal and genes located over 1Mb away was identified (11). This was made possible by advances in the availability of epigenomic data and development of revolutionary techniques in gene editing, including **CRISPR-Cas9**. Though not without its limitations, CRISPR-Cas9 gene editing has now become relatively common-place and is an exciting new addition to the toolbox for linking genetic signals to their effector transcript or gene, bringing promise that the next ten years will yield new drug targets as a direct result of genetic association findings. One example where genome-editing has helped bring resolution is in the T2D field, where uncertainty over the impact of T2D-protective alleles in *SLC30A8* on transcript levels had left ambiguity over the pursuit of agonists or antagonists. The recent demonstration in human IPS cell-derived pancreatic beta-cell models that the alleles lead to reduced levels of both the transcript and protein now provides reassurance that antagonists are the route forward (12).

Prediction

As it has become apparent that common complex disease risk is conferred potentially by many thousands of variants, development of polygenic risk scores to capture risk of disease has received considerable attention. Genetic risk scores (GRS) have been developed to comprise signals that pass stringent criteria (genome-wide significant, robustly-replicating), or signals that capture all genome-wide information. These provide greatest promise to improve on the predictive value of models that incorporate clinical data only, but have demonstrated substantial gradients of genetic risk in the general population (13). Such scores present an appealing metric upon which decisions about targeting of disease screening might be based, but require careful consideration in terms of risk thresholds and added value beyond tried-and-tested clinical information (14). A recently published genetic risk score for COPD demonstrated a 60% difference in absolute risk for smokers when

comparing high vs low genetic risk score deciles, with a risk of COPD for some non-smokers equivalent to that of many smokers (15). For T2D, the field has been exploring how GRS can be improved, refined and implemented in different ways by dissecting the genetic risk into different components of disease pathology. Using disease-relevant quantitative traits to define the underlying biological processes could assist with patient stratification and inform on therapy and progression (16). GRS have numerous applications in prediction, but care in their design and implementation is warranted, for example with respect to transferability across populations and within-population strata (17). Misconceptions about genetic risk, potential for additional burden on clinical services, and unforeseen consequences of lifestyle adaptations in response to risk of individual diseases mean that further multi-disciplinary research is needed to evaluate how genetic risk can bring about positive changes in clinical practice. For example, results from an implementation study to introduce genetic testing for variants known to affect Warfarin metabolism in routine clinical care reported equivalent benefits to those seen in previous randomised controlled trials; however, the time to obtain the genetic test results (45 mins) affected the clinic workflow and initially met with some resistance, potentially requiring a re-organisation of service delivery (18).

Prevention

The utility of preventative strategies like population-based genetic screening depends on the background frequency of the disease being investigated as well as the genetic risk conferred by the associated variants. It is likely to be useful for prevalent conditions such as coronary artery disease but, for diseases with a low background prevalence, even those at the highest genetic risk are more likely not to develop disease than to develop it. Additionally, as those at lower genetic risk are much more frequent in the population, the majority of cases within the population will come from people at lower genetic risk for disease. Age-related macular degeneration is a case in point: it was one of the first conditions for which robust evidence for genetic associations was produced, and carriage of two genetic susceptibility variants confers a 50-fold risk of disease development; however, only 20% of patients carry high-risk genetic variants (19). Environmental factors have been identified and include age, education, body mass index and smoking but even when combined with genetic risk loci, progression in patients with the same baseline macular grade varied widely with respect to subsequent visual loss (20). Nonetheless, genetic information has been reported to add value over and above clinical characteristics alone in predicting disease progression, particularly in those with a less severe disease stage (21) and may be useful in identifying high-risk groups for clinical trials of preventative therapy. Even where population prevalence is low, the fact that modifiable environmental risk factors exist means that identifying those at high risk could still have utility in effecting behaviour change. For example, in a randomised control trial in asymptomatic first degree relatives of patients with rheumatoid arthritis (RA), disclosure of genetic risk information led to positive behaviour modifications including increased smoking cessation and better dental hygiene (22).

Prognosis

Genetic data can be used to identify those most at risk of severe disease, in order to target more aggressive therapies earlier in the disease process. Studies of disease prognosis require larger collections of patient samples with high-quality phenotype data to define disease trajectories or classify disease subtypes, and generate additional methodological considerations. For example, the *MUC5B* promoter variant risk allele accounts for over 35% of idiopathic pulmonary fibrosis risk and is associated with improved survival, leading to speculation it might distinguish a milder subtype of disease. However, this paradox has recently been demonstrated to be the result of a form of selection bias (index event bias) (23). Further applications of genetic data toward improved prognosis include integrating genotypes with eQTL and transcriptional data to determine transcriptional risk

scores (TRS) associated with disease progression, for example as demonstrated in inflammatory bowel disease (24). GWAS is opening up opportunities for biomarker discovery. One of the challenges in clinical practice is differential diagnosis and getting this right when it influences treatment. Increasingly, proteomic screens of blood (and further tissue) biomarkers are linked to genome-wide genotype or sequence data (25). The identification of pQTL can be coupled to causal inference analyses, to provide direct clues for measurable biomarkers on the causal path to clinical endpoints.

Drug repurposing

Genetic data can be leveraged to inform **drug repurposing** opportunities (26, 27). Data about which gene products have been successfully targeted by small molecule or biologic drug candidates can be easily interrogated to identify whether the putative causal genes or pathways implicated by a genetic association signal have already been shown to be both druggable and tolerated in patient populations. This shortcut, whereby a drug already in clinical trials for one indication might additionally be investigated for an effect on an alternative indication on the basis of genetic findings, represents a means of reducing overall costs and also an attractive option to recover losses should a drug prove inefficacious for the indication against which it was originally designed. For example, two genes implicated in a recent GWAS of osteoarthritis encode targets for which there are already approved drugs for other indications (27). Genetic association signals that mimic the effect of a drug (for example, through association with expression levels of the drug target) might be informative about likely unintended effects of that drug. For example, atacicept, which blocks B cell stimulation by TNFSF13, is at an advanced stage of development for treating systemic lupus erythematosus. A recent GWAS showed that an allele associated with decreased lung function is also associated with decreased expression of *TNFSF13I*, suggesting that monitoring of respiratory health whilst on atacicept might be recommended (15).

Precision medicine

One of the immediate clinical applications arising from our greater understanding of complex disease genetics could be the application to better targeting of available therapies to those most likely to respond, or avoidance of therapy in those likely to develop adverse events. This may arise due to the identification of subgroups of patients with apparently the same clinical disease but with a different underlying genetic architecture determining response to therapy, or because of the identification of genetic variants affecting drug metabolism. Whilst all of the currently FDA-approved companion diagnostics pertain to oncology, over 350 drugs labels contain information on genetic-drug interactions informing guidance on efficacy, adverse events, or dosing (28). These include HLA-B*5701 interactions with abacavir, which can lead to extreme toxicity, and TPMT testing for azathioprine (28).

Most common complex diseases exhibit a variable disease course and response to therapy; of the 10 best-selling drugs in the US, numbers needed to treat to achieve one good responder were estimated between 4 and 25. This illustrates the need to develop better ways to target therapies (29). National initiatives in the US (30) and Europe (31) have been established to drive forward the stratified medicine agenda, but progress has been relatively slow. One of the major limiting factors is the measurement of progression or treatment response, often based on clinical end-points rather than intermediate biomarkers. For example, response to treatment of asthma often uses the exacerbation rate as a primary outcome measure, whilst in RA, response is assessed using physical measures and an inflammatory marker. Genetic studies can identify which components are heritable and, therefore, predictable (32). This can inform new outcome measures, such as in the case of RA, where composite scores were re-weighted to include only subcomponents showing heritability and better correlation with joint inflammation and subsequent joint damage (33). A second challenge is that

adherence is often not considered: a drug is unlikely to be effective if it is not taken, and non-adherence is widespread in chronic diseases with up to 1 in 4 patients not taking medication as prescribed (34). This degree of misclassification has a severe impact on the power of genetic and other biomarker studies to identify predictors of response. Despite these challenges, some response biomarkers are emerging, such as the associations of HLA-C*06:02 with biologic therapy response in psoriasis (35), and HLA-DRB1 with severity, mortality and treatment response to biologic drugs in RA (36).

ACCELERATING THE DISCOVERY PIPELINE

Genomics in diverse populations

The majority of genetic association studies have focused on individuals from the general European population. There is an imperative to fill this fundamental gap, leveraging different population characteristics to identify and fine-map causal variants. At one end of the spectrum, founder populations demonstrate high levels of homogeneity (37), while at the other end African populations are characterised by high levels of genetic diversity and low levels of linkage disequilibrium (38). In isolated populations, protective or deleterious rare variants may have increased in frequency while neutral rare variation is lost, providing advantages for the discovery of rare variant signals. Genetic studies in African populations can provide unprecedented insights into the architecture of complex disease, and reveal associations that have been missed due to differences in allele frequency between populations (38). Studying diverse populations can lead to novel locus identification, fine-mapping of effector genes, and can have translational ramifications. For example, a *G6PD* variant present in high frequency in African populations affects HbA1c levels, a diagnostic marker for T2D, leading to potentially widespread patient misclassification (39). Importantly, the potential for targeted intervention and disease prevention afforded by polygenic risk scores can only be realised if the genetic determinants of disease have been studied in diverse populations to enable applicability of the risk scores (17).

Assaying the full allelic spectrum

The next frontier in genomic medicine entails embracing whole genome sequencing and integrating it as a standard tool into medical care. Several large-scale initiatives such as the 100,000 Genomes Project in the UK have spearheaded proof of concept studies, primarily with an emphasis on diagnosis for rare diseases and characterising mutational profiles in cancer for targeted treatment. Genomics England is now preparing to fulfil the aspiration of 5 Million Genome Analyses in the next quinquennium, including full sequencing of the UK Biobank. Deep whole genome sequencing provides access to variation across the full allele frequency spectrum, and to structural variation that is not probed through genome-wide genotyping even if coupled to dense imputation. For common complex diseases, whole genome sequencing-based studies will identify novel targets for drug development, and will detect rare and structural variants that can contribute to comprehensive, truly genome-wide risk scores for disease prediction and patient stratification. The ethical, legal and societal aspects of integrating whole genome sequencing into patient care warrant in-depth consideration, and are recognised by the field as inextricably linked components of paramount importance.

Deep phenotypic characterisation

Power to discover novel associations can not only be catalysed through comprehensively studying sequence variation, but also by enhancing the phenotypic space examined. Longitudinal linkage to electronic health records (EHRs), for example, can substantially bolster the depth of real-world phenotyping available, leading to better clinical characterisation of the study population, and to the

availability of endophenotype-associated traits for study in conjunction with genomics. For instance, imaging-derived traits can serve as powerful markers of biological processes with a genetic underpinning and boost discovery power in genetic association studies (40). Several large-scale biobanks and population-based cohorts, such as the All of US initiative and the Million Vets programme, include linkage to EHRs (41). The next generation of GWAS will traverse a wide phenotypic space, cover ethnically diverse populations, and benefit from the increasingly longitudinal nature of measurements and sampling. The establishment of large-scale biobanks and registries geared towards genomics (as elegantly exemplified by the Nordic countries (42)) also means that previously difficult-to-reach genetic effects, such as gene-environment and gene-gene interactions (e.g. (43)) can be further explored. Successful translational examples of coupling genomics to EHRs have started to emerge, as illustrated by the Regeneron-Geisinger collaboration that led to the identification of loss-of-function mutations in *ANGPTL3* associated with the development of coronary artery disease, and the subsequent development of an inactivating monoclonal antibody offering promise for disease prevention (44).

Next-generation analysis approaches

As big data become even bigger, and as whole genome sequencing assumes its position as an emerging standard in healthcare, we need computational tools for efficient processing and analysis, and for clinical decision making support based on real-world data. Applied artificial intelligence (AI) offers the computational arsenal required to navigate through the wealth of data generated, and is being widely adopted for big data crunching, with early implementation of machine learning and deep learning approaches to analyse genomic data (45). Explainable AI carries particular importance in translational genomics, so that the processes leading to decisions made can be transparent and interpretable.

Multimorbidity

A growing number of individuals suffer from more than one serious chronic condition, such as obesity, osteoarthritis and T2D. Multimorbidity has been identified as a priority for global health research, with at least 50 million people suffering from multimorbidity in Europe alone (46). Richly-phenotyped cohort resources can be mined to identify pleiotropic loci that affect risk of developing multiple diseases. Novel method development to jointly analyse association between sequence variation and epidemiologically-linked traits, including blood biomarkers, is a flourishing and fast-moving field (47). Mendelian randomisation approaches can be employed to infer causality and disentangle directions of effect. For example, obesity-associated variants in *FTO* are also robustly associated with the risk of developing osteoarthritis, and causal inference analysis unveiled a causal relationship (48). Patients with schizophrenia are at increased risk of developing T2D, and the observed comorbidity has been demonstrated to be at least in part due to shared genetic mechanisms (49). Understanding pleiotropic effects can have important consequences for disease management, including prognosis, drug repositioning and adverse effect prevention.

ACCELERATING THE PIPELINE

Functional genomic follow-up and mechanistic modelling

Unravelling disease biology and translational insight has not kept pace with the success of genetic discovery efforts. The inability to link genetic signals to the effector transcripts has been a major obstacle in these endeavours. The power of working at scale to decipher causal variants and molecular mechanisms has been illustrated in human blood cells (50). The challenge now is to extend these efforts to less accessible disease-relevant cell types, at different developmental stages and under disease-relevant conditions to allow integration with large-scale genetic discovery efforts.

There continues to be debate around disease models and the balance between throughput and authenticity. Human IPS cells are attractive due to opportunities for genome-editing and their differentiation into otherwise inaccessible cell types, but the infancy of the field, poor recapitulation of certain cell types, expense of culture and high levels of variability between protocols indicates that they do not yet offer a perfect solution. Considerable progress is anticipated in approaches coupling high-throughput screening and genetic manipulation with cellular and transcriptomic phenotyping. This will enable working at genome-scale to generate datasets which can be harnessed across multiple diseases in order to prioritise transcripts at genetic loci for detailed mechanistic study. Efforts from the community to bring such datasets together will help accelerate biological inference.

Bench to bedside pipelines and loops

There is a well-recognised need to accelerate the development pathway in order to reduce the current estimated 15 years it takes for a drug to reach market. Genetic evidence doubles the chance of a drug being successful (7), and there have been examples from the field of rare disease genetics, notably the development of the lipid-lowering drug Evolocumab (51) within 10 years of the identification of the association of *PCSK9* with cholesterol levels (8). Whilst these examples illustrate how basic science discoveries can translate to the bedside, it is also important that the reverse pipeline be operational to close the experimental medicine loop. Targeted genetic studies identified the *IL23R* gene to be associated with psoriasis in 2007, and a drug acting on this pathway, Ustekinumab, was licensed shortly thereafter. Subsequently, GWAS detected associations that informed the development of more selective IL23 inhibitors, which are now showing efficacy in phase III clinical trials, even in patients who do not respond to Ustekinumab (52). Identification of early response biomarkers can accelerate progress further and avoid the need for large and costly phase III studies where there is a low likelihood of success. Trial designs such as futility trials allow abandoning of studies where the intervention is unlikely to be superior to conventional therapy, enabling resources to be diverted to more promising treatments (53).

CONCLUSION

A better understanding of the genetic aetiology of complex diseases can provide novel insights into fundamental biology and translational opportunities. In recognition of this translational potential, there is a rising number of high-profile, large-investment initiatives focused on genomics in medicine. These traverse public and private funding mechanisms to generate the large-scale clinical and biodata resources needed to spur innovation for personalised medicine and population health. This new, eagerly-awaited digital health era is now becoming a tangible prospect.

ACKNOWLEDGEMENTS

Funding: ALG is a Wellcome Trust Senior Fellow Basic Biomedical Science. ALG is funded by the Wellcome Trust (095101, 200837, 106130, 203141), Medical Research Council (MR/L020149/1), European Union Horizon 2020 Programme (T2D Systems), and NIH (U01-DK105535; U01-DK085545) and by the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre (BRC). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. ACB is an NIHR Senior Investigator and supported by the NIHR Manchester Biomedical Research Centre. LVW holds a GSK/British Lung Foundation Chair in Respiratory Research. The authors are grateful to Dr. Iris Fischer for helpful edits.

Competing interests: The authors declare no competing interests.

FIGURE LEGENDS

Figure 1: The translational potential of complex disease genomics.

Improvements to human health (inner circle) are achieved through various enabling milestones at different translational axes (outer circle).

Figure 2: Identifying therapeutic targets.

(A) Effector transcripts identified at genetic signals are genetically manipulated to recapitulate *in vivo* effects on gene expression (e.g. CRISPR knockdown, over-expression) in human cell lines (e.g. IPS cell-derived models) and in animal models, which can be phenotyped. (B) Additional alleles are identified using sequence data and assessed for their relationship to disease risk or related-traits. To provide insight on the therapeutic window, *in vitro* functional severity and clinical severity are explored to establish the relationship between target perturbation and outcome. Potential adverse on-target effects are investigated using genome-wide datasets for other disorders (PheWAS). (C) Examples of therapeutic targets confirmed or identified by human GWAS.

Figure 3: Pathway to identifying causal variants.

Trans-ethnic fine-mapping is used to refine genetic association signals and genomic annotation to prioritize likely causal variants. Variants in regulatory elements (e.g. enhancers) are mapped to the promoters and transcripts they regulate through conformation capture approaches (e.g. promoter Hi-C). Expression QTL mapping can provide additional evidence for the effector transcripts and the direction of effect (e.g. increase or decrease in transcript levels).

REFERENCES

1. J. P. Clancy, M. Jain, Personalized medicine in Cystic Fibrosis. *Am. J. Respir. Crit. Care Med.* **186**, 593-597 (2012).
2. R. J. Allen *et al.*, Genome-wide association study of susceptibility to idiopathic pulmonary fibrosis. *bioRxiv*, 636761 (2019).
3. M. A. Seibold *et al.*, A common *MUC5B* promoter polymorphism and Pulmonary Fibrosis. *N. Engl. J. Med.* **364**, 1503-1512 (2011).
4. U. Styrkarsdottir *et al.*, Whole-genome sequencing identifies rare genotypes in *COMP* and *CHADL* associated with high risk of hip osteoarthritis. *Nature Genet.* **49**, 801-805 (2017).
5. I. Moltke *et al.*, A common Greenlandic *TBC1D4* variant confers muscle insulin resistance and type 2 diabetes. *Nature* **512**, 190-193 (2014).
6. E. A. King, J. W. Davis, J. F. Degner, Are drug targets with genetic support twice as likely to be approved? Revised estimates of the impact of genetic support for drug mechanisms on the probability of drug approval. *bioRxiv*, 513945 (2019).
7. M. R. Nelson *et al.*, The support of human genetic evidence for approved drug indications. *Nature Genet.* **47**, 856-860 (2015).
8. I. K. Kotowski *et al.*, A spectrum of *PCSK9* alleles contributes to plasma levels of low-density lipoprotein cholesterol. *Am. J. Hum. Genet.* **78**, 410-422 (2006).
9. A. Mahajan *et al.*, Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nature Genet.* **50**, 1505-1513 (2018).
10. T. M. Frayling *et al.*, A common variant in the *FTO* gene is associated with Body Mass Index and predisposes to childhood and adult obesity. *Science* **316**, 889-894 (2007).
11. M. Claussnitzer *et al.*, *FTO* obesity variant circuitry and adipocyte browning in humans. *N. Engl. J. Med.* **373**, 895-907 (2015).
12. O. P. Dwivedi *et al.*, Loss of ZnT8 function protects against diabetes by enhanced insulin secretion. *bioRxiv*, 436030 (2018).

13. A. V. Khera *et al.*, Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nature Genet.* **50**, 1219-1224 (2018).
14. F. Dudbridge, Power and predictive accuracy of polygenic risk scores. *PLOS Genet.* **9**, e1003348 (2013).
15. N. Shrine *et al.*, New genetic signals for lung function highlight pathways and chronic obstructive pulmonary disease associations across multiple ancestries. *Nature Genet.* **51**, 481-493 (2019).
16. M. I. McCarthy, Painting a new picture of personalised medicine for diabetes. *Diabetologia* **60**, 793-799 (2017).
17. A. R. Martin *et al.*, Clinical use of current polygenic risk scores may exacerbate health disparities. *Nature Genet.* **51**, 584-591 (2019).
18. A. L. Jorgensen *et al.*, Implementation of genotype-guided dosing of warfarin with point-of-care genetic testing in three UK clinics: a matched cohort study. *BMC Med.* **17**, 76 (2019).
19. B. Feigl, C. P. Morris, The challenge of predicting macular degeneration. *Curr. Med. Res. Opin.* **27**, 1745-1748 (2011).
20. J. M. Seddon, B. Rosner, Validated prediction models for macular degeneration progression and predictors of visual acuity loss identify high-risk individuals. *Am. J. Ophthalmol.* **198**, 223-261 (2019).
21. T. J. Heesterbeek *et al.*, Genetic risk score has added value over initial clinical grading stage in predicting disease progression in age-related macular degeneration. *Sci. Rep.* **9**, 6611 (2019).
22. J. A. Sparks *et al.*, Disclosure of personalized rheumatoid arthritis risk using genetics, biomarkers, and lifestyle factors to motivate health behavior improvements: A randomized controlled trial. *Arthrit. Care Res.* **70**, 823-833 (2018).
23. F. Dudbridge *et al.*, Adjustment for index event bias in genome-wide association studies of subsequent events. *Nature Commun.* **10**, 1561 (2019).
24. U. M. Marigorta *et al.*, Transcriptional risk scores link GWAS to eQTLs and predict complications in Crohn's disease. *Nature Genet.* **49**, 1517-1521 (2017).
25. B. B. Sun *et al.*, Genomic atlas of the human plasma proteome. *Nature* **558**, 73-79 (2018).
26. S. Pushpakom *et al.*, Drug repurposing: progress, challenges and recommendations. *Nat. Rev. Drug Discov.* **18**, 41-58 (2018).
27. I. Tachmazidou *et al.*, Identification of new therapeutic targets for osteoarthritis through genome-wide analyses of UK Biobank data. *Nature Genet.* **51**, 230-236 (2019).
28. U.S. Food and Drug Administration (FDA), Table of Pharmacogenomic Biomarkers in Drug Labeling (available at <https://www.fda.gov/drugs/science-research-drugs/table-pharmacogenomic-biomarkers-drug-labeling>).
29. N. J. Schork, Personalized medicine: Time for one-person trials. *Nature* **520**, 609-611 (2015).
30. National Institutes of Health (NIH), Overview of the Accelerating Medicines Partnership (AMP) Rheumatoid Arthritis and Lupus Network (available at <https://www.niams.nih.gov/grants-funding/funded-research/accelerating-medicines/overview>).
31. Medical Research Council (MRC), Precision Medicine (available at <https://mrc.ukri.org/research/initiatives/precision-medicine/research/>).
32. J. Massey *et al.*, Genome-wide association study of response to tumour necrosis factor inhibitor therapy in rheumatoid arthritis. *Pharmacogenomics J.* **18**, 657-664 (2018).
33. E. M. A. Hensor *et al.*, Validity of a two-component imaging-derived disease activity score for improved assessment of synovitis in early rheumatoid arthritis. *Rheumatology*, doi:10.1093/rheumatology/kez1049 (2019).
34. J. Bluett *et al.*, Impact of inadequate adherence on response to subcutaneously administered anti-tumour necrosis factor drugs: results from the Biologics in Rheumatoid Arthritis Genetics and Genomics Study Syndicate cohort. *Rheumatology* **54**, 494-499 (2014).

35. N. Dand *et al.*, HLA-C*06:02 genotype is a predictive biomarker of biologic treatment response in psoriasis. *J. Allergy Clin. Immunol.*, doi:10.1016/j.jaci.2018.1011.1038 (2018).
36. S. Viatte *et al.*, Association of HLA-DRB1 haplotypes with rheumatoid arthritis severity, mortality, and treatment response. *JAMA* **313**, 1645-1656 (2015).
37. E. Zeggini, A. L. Gloyn, T. Hansen, Insights into metabolic disease from studying genetics in isolated populations: stories from Greece to Greenland. *Diabetologia* **59**, 938-941 (2016).
38. D. Gurdasani *et al.*, The African Genome Variation Project shapes medical genetics in Africa. *Nature* **517**, 327-332 (2014).
39. E. Wheeler *et al.*, Impact of common genetic determinants of Hemoglobin A1c on type 2 diabetes risk and diagnosis in ancestrally diverse populations: A transethnic genome-wide meta-analysis. *PLOS Med.* **14**, e1002383 (2017).
40. L. T. Elliott *et al.*, Genome-wide association studies of brain imaging phenotypes in UK Biobank. *Nature* **562**, 210-216 (2018).
41. B. S. Glicksberg, K. W. Johnson, J. T. Dudley, The next generation of precision medicine: observational studies, electronic health records, biobanks and continuous monitoring. *Hum. Mol. Genet.* **27**, R56-R62 (2018).
42. P. R. Njølstad *et al.*, Roadmap for a precision-medicine initiative in the Nordic region. *Nature Genet.* **51**, 924-930 (2019).
43. A. R. Bentley *et al.*, Multi-ancestry genome-wide gene-smoking interaction study of 387,272 individuals identifies new loci associated with serum lipids. *Nature Genet.* **51**, 636-648 (2019).
44. F. E. Dewey *et al.*, Genetic and pharmacologic inactivation of ANGPTL3 and cardiovascular disease. *N. Engl. J. Med.* **377**, 211-221 (2017).
45. G. Eraslan, Ž. Avsec, J. Gagneur, F. J. Theis, Deep learning: new computational modelling techniques for genomics. *Nat. Rev. Genet.*, doi:10.1038/s41576-41019-40122-41576 (2019).
46. M. Rijken *et al.*, ICARE4EU: improving care for people with multiple chronic conditions in Europe. *Eurohealth* **19**, 29-31 (2013).
47. S. Hackinger, E. Zeggini, Statistical methods to detect pleiotropy in human complex traits. *Open Biol* **7**, 170125 (2017).
48. E. Zengini *et al.*, Genome-wide analyses using UK Biobank data provide insights into the genetic architecture of osteoarthritis. *Nature Genet.* **50**, 549-558 (2018).
49. S. Hackinger *et al.*, Evidence for genetic contribution to the increased risk of type 2 diabetes in schizophrenia. *Transl. Psychiatry* **8**, 252 (2018).
50. H. G. Stunnenberg *et al.*, The International Human Epigenome Consortium: A blueprint for scientific collaboration and discovery. *Cell* **167**, 1145-1149 (2016).
51. M. J. Koren *et al.*, Anti-PCSK9 monotherapy for hypercholesterolemia: The MENDEL-2 randomized, controlled phase III clinical trial of evolocumab. *J. Am. Coll. Cardiol.* **63**, 2531-2540 (2014).
52. R. G. Langley *et al.*, Efficacy and safety of guselkumab in patients with psoriasis who have an inadequate response to ustekinumab: results of the randomized, double-blind, phase III NAVIGATE trial. *Br. J. Dermatol.* **178**, 114-123 (2018).
53. B. Levin, The futility study—Progress over the last decade. *Contemp. Clin. Trials* **45**, 69-75 (2015).

BOX

Genome-wide association study (GWAS)

An analysis that tests each genetic variant measured across the genome for statistical association with a chosen disease or trait.

GWAS signal

A set of correlated genetic variants that are associated with a disease or trait at a genome-wide significant statistical significance threshold.

Fine-mapping

An analysis that attempts to identify which of the correlated genetic variants within GWAS signal are most likely to be the true underlying causal variant.

eQTL and pQTL resources

Data sets for which both genome-wide genetic variation and gene (eQTL) or protein (pQTL) expression variation is available and that can be analysed to identify genetic variants that are associated with gene or protein expression levels. Simply, these can be used to identify whether a GWAS signal is also associated with expression of a particular gene thereby highlighting that gene as a potential effector of the disease or trait.

Epigenetics

The study of DNA variation that is not encoded in the DNA but which can be heritable and influence how genes are expressed in different cell types. This includes DNA modification (for example, methylation of CpG sites) and histone modifications (methylation, acetylation).

Chromosome interactions

Gene expression is often controlled via 3-dimensional interaction with distal regulatory motifs. The presence of an interaction between a locus containing a GWAS signal and a gene promoter some distance away is one line of evidence in GWAS signal to gene mapping. Chromosome conformation capture (3C) based techniques that use high throughput sequencing, such as HiC, can be used to generate genome-wide maps of chromosome interactions in different cell types.

Drug target

A molecule (usually a protein) that a drug (compound or biologic) binds to, or in some other way interacts with, to alter a biochemical process in the body and illicit a therapeutic effect.

Monogenic disease

Also referred to as Mendelian disease, a usually rare disease (or trait) that results from genetic variation effecting a specific gene.

CRISPR-Cas9

A method of genome editing used to introduce sequence changes into DNA to enable experimental investigation of the functional consequences of the DNA change. This method can introduce single base changes.

Drug repurposing,

Also referred to as drug repositioning, the re-direction of a drug already at some stage of clinical development for treatment of one disease into trials for an alternative disease.

Polygenic risk scores

Also referred to as genetic risk scores, measures of genetic risk comprising effects of multiple variants. They can include all variants genome-wide or be restricted to those that meet some significance threshold for association with the disease or trait being investigated. Simply, the number of risk alleles that a person carries at each variant is counted up to construct a score which is then tested for association with disease. The score may be weighted by incorporating the effect sizes of each variant into the score.

Sequencing

The measurement of all variant and non-variant base positions of an individual's DNA or RNA. Has applications in studies of rare genomic variation, gene expression and epigenetic modifications.

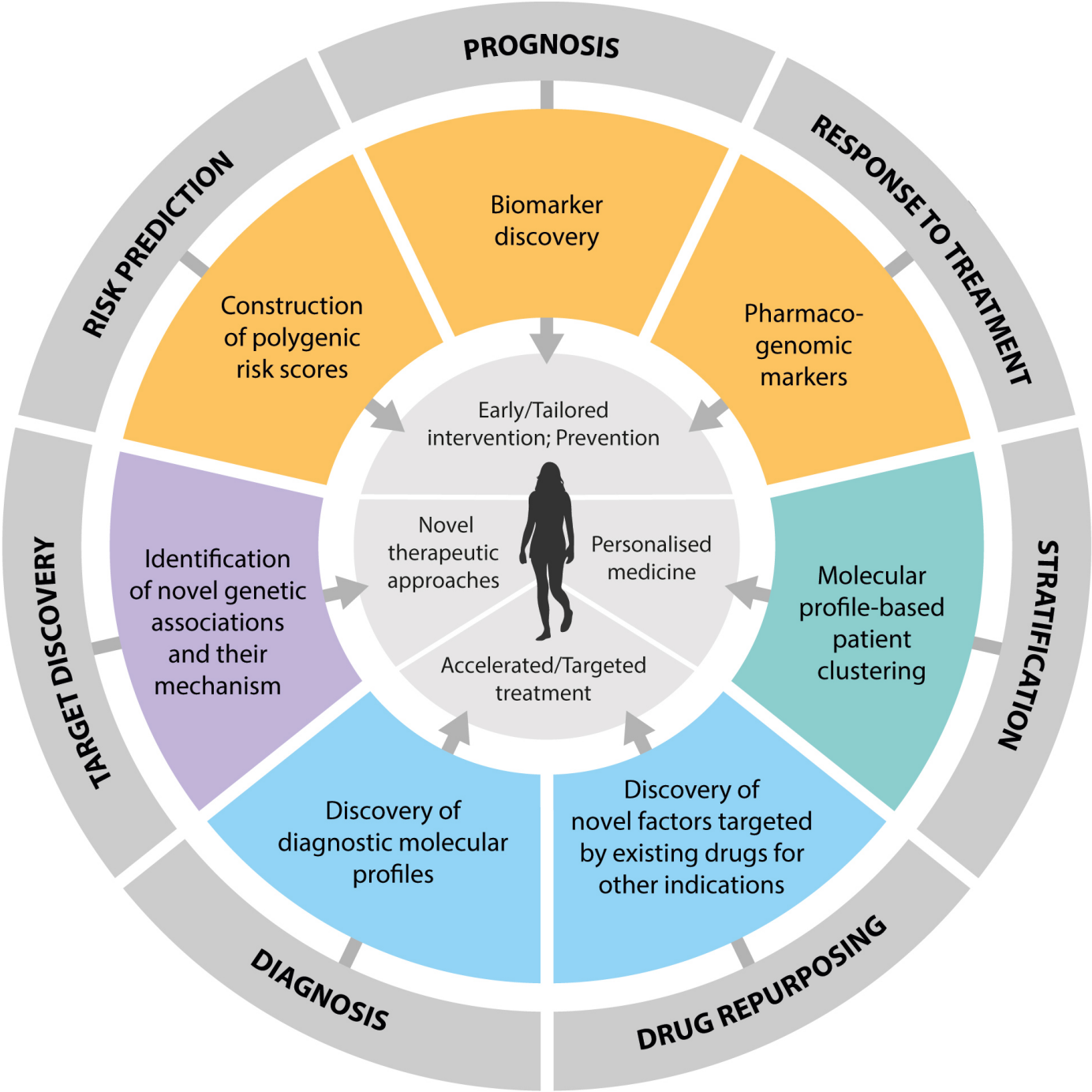
Precision medicine

The concept of providing the right medicine to the right patient at the right time. Also referred to as personalised medicine.

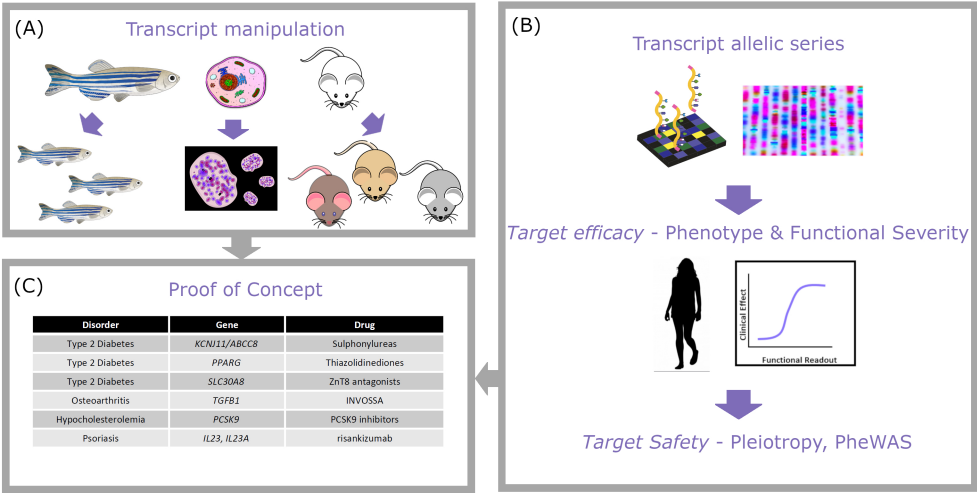
Complex disease

In contrast to monogenic disease, used to describe diseases, usually common in the population, where multiple genes each contribute a small proportion of the overall genetic risk.

"Figure #1"



"Figure #2"



"Figure #3"

