Review

# From target discovery to clinical drug development with human genetics

https://doi.org/10.1038/s41586-023-06388-8

Received: 17 December 2021

Accepted: 29 June 2023

Published online: 23 August 2023

Check for updates

Katerina Trajanoska[1], Claude Bhérer[1], Daniel Taliun[1], Sirui Zhou[1], J. Brent Richards[2,3] & Vincent Mooser[1]✉

The substantial investments in human genetics and genomics made over the past three decades were anticipated to result in many innovative therapies. Here we investigate the extent to which these expectations have been met, excluding cancer treatments. In our search, we identified 40 germline genetic observations that led directly to new targets and subsequently to novel approved therapies for 36 rare and 4 common conditions. The median time between genetic target discovery and drug approval was 25 years. Most of the genetically driven therapies for rare diseases compensate for disease-causing loss-of-function mutations. The therapies approved for common conditions are all inhibitors designed to pharmacologically mimic the natural, disease-protective effects of rare loss-of-function variants. Large biobank-based genetic studies have the power to identify and validate a large number of new drug targets. Genetics can also assist in the clinical development phase of drugs—for example, by selecting individuals who are most likely to respond to investigational therapies. This approach to drug development requires investments into large, diverse cohorts of deeply phenotyped individuals with appropriate consent for genetically assisted trials. A robust framework that facilitates responsible, sustainable benefit sharing will be required to capture the full potential of human genetics and genomics and bring effective and safe innovative therapies to patients quickly.

Despite the great advances in the development of therapeutic agents that have transformed modern medicine, more therapies are needed to prevent and treat many common and rare diseases. Drug discovery and development is a lengthy, risky and expensive undertaking (Fig. 1). Nearly nine out of ten of the drugs that are tested in humans do not gain approval, mostly because of their insufficient efficacy[1,2]. Considering the expense of bringing new drugs to the market—research and development expenditures for the top 15 pharmaceutical companies amounted to US$571 billion over the past 5 years for 329 new active substances launched during this period[3]—this low success rate incurs yearly societal costs of dozens of billions of dollars.

Drug development failures can be attributed to gaps in our understanding of the biology underlying human disease, excessive reliance on non-human models, and prioritization of targets or agents with low specificity, poor pharmacology or a propensity to produce unwanted adverse reactions[4,5]. Some of these limitations were anticipated to be overcome by using human molecular genetics[6]. More explicitly, the discovery of disease-causing variants was expected to lead to novel targets and new indications for existing drugs. To that end, public agencies and private enterprises have made substantial investments over the past three decades into large genomic programmes involving millions of participants.

Since the publication of the first human genome sequences[7,8], there have been extraordinary breakthroughs in high-throughput sequencing and information technologies. Beyond considerably expanding our understanding of the genetic abnormalities leading to cancer[9], these advances have been remarkably successful in elucidating the molecular basis of more than 4,000 rare, monogenic diseases—that is, diseases due to defects in a single gene; these diseases may run in families and have a Mendelian mode of inheritance[10]. In addition, these technologies have revealed hundreds of thousands of associations between genetic variants and complex, polygenic, common diseases—that is, conditions mostly due to the combined effects of low-impact variants found in multiple genes and environment[11]. In this Review, we quantify the discoveries that have been translated into approved therapies. While acknowledging the decades-long lag time between target identification and regulatory approval of derived therapeutics, we reasoned that addressing this question now is both legitimate and essential to guide future investments in the field. We intentionally restricted our search to non-cancer drugs, as the success of genomics in cancer treatment has been reviewed elsewhere[12,13].

Examples such as proprotein convertase subtilisin kexin 9 (PCSK9) inhibitors[14] illustrate the power of molecular genetics in bringing new medicines to the market[15]. In this particular case, gain-of-function and loss-of-function variants were originally identified in the *PCSK9* gene, leading to increased[16] or reduced[17] blood levels of low-density lipoprotein (LDL) cholesterol, respectively. Moreover, loss-of-function variants of *PCSK9* were associated with a decreased risk of coronary

[1]Canada Excellence Research Chair in Genomic Medicine, Department of Human Genetics, Faculty of Medicine and Health Sciences, Victor Phillip Dahdaleh Institute of Genomic Medicine, McGill University, Montreal, Quebec, Canada. [2]Lady Davis Institute for Medical Research, Jewish General Hospital, McGill University, Montreal, Quebec, Canada. [3]Department of Epidemiology and Biostatistics and Occupational Health, McGill University, Montreal, Quebec, Canada. ✉e-mail: vincent.mooser@mcgill.ca

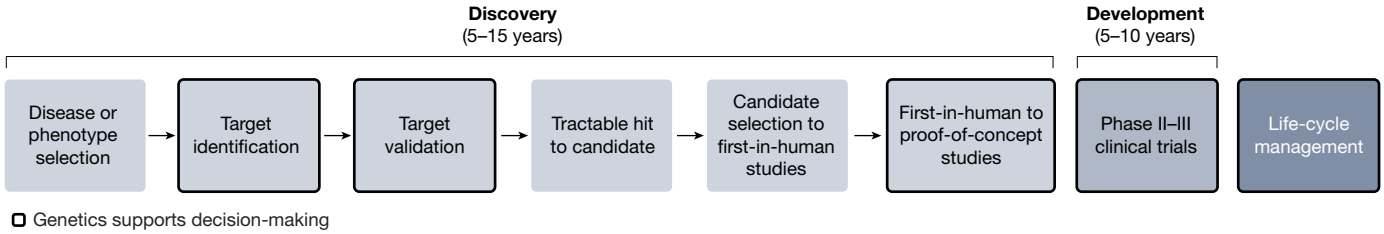


**Fig. 1 | The various steps along the drug discovery and clinical development pipeline.** Steps where genetics has been shown to bring critical or supportive information to decision-making along the pipeline are outlined.

artery disease[18]. Notably, complete loss-of-function mutations were not associated with any observed adverse outcomes[17]. Together, these observations were essential for the discovery and validation of this therapeutic target as they pointed to pharmacological inhibition of PCSK9 as a safe strategy to prevent or treat coronary artery disease. The US Food and Drug Administration (FDA) approved the first PCSK9 inhibitor for this indication in 2015 (ref. 19), 12 years after the publication of the original genetic observations.

In parallel, detailed retrospective analyses of historic drug development programmes provided undisputed evidence that drugs whose targets harbour genetic variants associated with the drug's indication are twice as likely to reach regulatory approval as those with no such variants[5,20–22]. Nevertheless, a comprehensive analysis of the prospective contribution of human genetics to the development of novel approved drugs such as PCSK9 inhibitors, or to the successful repositioning of existing drugs, is still lacking. Here we address this deficiency by synthesizing the information collected from the published literature and public databases to help guide the future direction of using human genetics and genomics in supporting the discovery and clinical development of novel therapies.

## Genetically driven therapies

Quantifying the full impact of genetics on decision making in the biotech and pharmaceutical industries is difficult, as such decisions are rarely communicated publicly. We therefore limited our analysis to approved drugs, and defined as 'genetically driven' those therapies (or repurposing opportunities) for which the original human genetic associations were reported to be sufficiently informative to support a new drug discovery campaign (or to develop a repositioning programme) that eventually resulted in their regulatory approval by the FDA and/or the European Medical Agency (EMA) (Supplementary Note). Given our scope, we excluded hormonal therapies, antimicrobial agents, vitamins and minerals. We restricted our analysis to approved drugs for which direct genetic evidence had been established between the target gene and the indication of interest. Accordingly, we excluded examples for which genetics had provided indirect evidence through associations with endophenotypes (such as blood glucose levels for diabetes), signs or symptoms associated with the condition or related gene pathways, although we acknowledge the value of such data for decisions to progress certain investigational therapies along the pipeline[20–22]. Considering that a minimum of five years separates an original genetic observation from the approval of a derived therapy[23], we also excluded drugs for which relevant genetic associations were first reported five years or less before the approval year.

We applied these criteria to data from OpenTargets[24], DrugBank[25] and the FDA (Supplementary Fig. 1). In total, we identified 2,832 FDA- or EMA-approved therapies. After exclusion of antineoplastic drugs ($n$ = 277), anti-infectives and antiparasitics ($n$ = 392), hormonal preparations ($n$ = 154), vitamins and analogues ($n$ = 90) and drugs whose target is unknown ($n$ = 902), our search resulted in 1,031 drugs with matching protein-coding target genes. Most drugs (766 out of 1,031 (74%)) acted through multiple targets; in addition, many drugs had more than one indication (794 out of 1,031 (77%)), resulting in 6,690 drug–target–gene indication triplets (Supplementary Table 1). Some direct genetic evidence derived from 8 data sources was reported for 619 triplets (corresponding to 346 different drugs) (Supplementary Note 1). For 98 triplets (corresponding to 80 drugs), such evidence had been reported more than five years before drug approval. Manual, literature-based curation documented that genetic information had been essential for 60 drugs—that is, these drugs would probably not have been developed had the genetic association not been discovered. After grouping the drugs by class (small molecule, biological or gene therapy), our analyses identified 47 first-in-class (and 13 follower) therapies for 40 targets that met our definition of being genetically driven (Supplementary Tables 1–3).

These particular criteria captured only a fraction of the impact of genetics on drug discovery and development. We acknowledge, for instance, that we missed drugs such as mavacamten (approved by the FDA in 2022), which was absent from the databases that we used despite the fact that its development depended in part on the discovery of mutations in several genes responsible for hypertrophic cardiomyopathy[26–28]. Nonetheless, we make several observations based on our analysis. The 40 targets correspond to around 6% of the approximately 500 drug target genes of FDA-approved non-cancer drugs[29]; the remaining 94% were probably identified using conventional pharmacology, biochemistry or molecular biology approaches. The therapies targeting these 40 gene products were approved for chronic, as opposed to acute, conditions. Most of them (26 out of 40) were targeted for treatments of metabolic disorders, whereas the remaining 14 genes were targeted to treat diseases in 7 other therapeutic areas (Fig. 2a).

The genetic evidence for the 40 targets is based on associations between low-frequency, protein-disruptive variants in the target gene and the corresponding disease. In most of these cases, the target genes had been cloned (38 out of 40) or the genetic associations had been reported (35 out of 40) before the publication of the first human genomes[7,8,30]. The median gap between the reporting of the original genetic observation and approval of the derived, first-in-class therapeutic agent was 25 years (range 4 to 38 years) (Fig. 2b and Supplementary Table 3). Although the genetic identification of a new potential therapeutic target does not necessarily prompt the immediate launch of a drug discovery campaign, this gap still reflects the decades-long lag time between the discovery of a new target and the eventual regulatory approval of a derived therapeutic.

Half of the first-in-class therapies (23 out of 47) were biologics (that is, monoclonal antibodies ($n$ = 4), enzymes ($n$ = 15) or proteins ($n$ = 4)), 36% (17 out of 47) were gene or RNA therapies and 15% (7 out of 47) were small molecules (Fig. 2c). Most targets (34 out of 40) led to approved therapies belonging to a single class, 5 targets led to approved therapies belonging to 2 classes and 1 led to approved therapies belonging to 3 classes (Supplementary Table 2). Finally, the vast majority of indications (36 out of 40) represented rare, monogenic conditions (Supplementary Tables 2 and 3). PCSK9 inhibitors were approved for both a Mendelian disorder (familial hypercholesterolaemia) as well as for the more common, difficult-to-treat hypercholesterolaemia[19], whereas CCR5 inhibitors, SOST inhibitors and SGLT2 inhibitors were

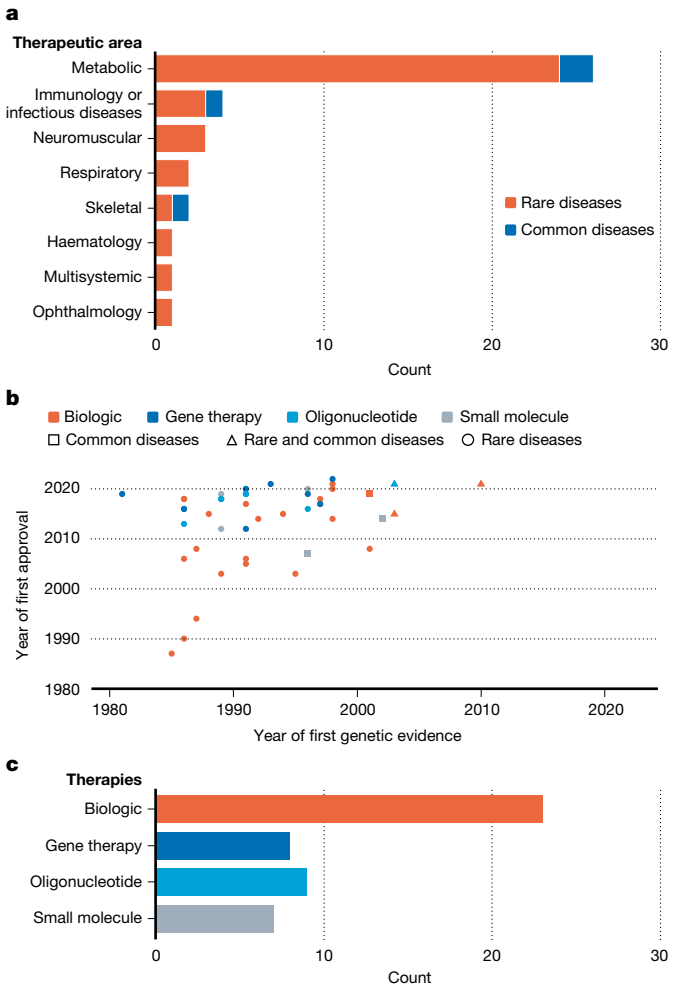


**Fig. 2 | Overview of 40 targets for 47 approved, first-in-class, genetically driven non-cancer therapies.** Targets and therapeutics are classified according to therapeutic area and prevalence of disease for approved indication (**a**), year of discovery of molecular genetic association and corresponding drug approval (**b**) or class of therapeutic agent (**c**).

approved for common diseases only—human immunodeficiency virus (HIV) infection, osteoporosis and type 2 diabetes, respectively.

## Correction of monogenic diseases

In our search, most first-in-class therapeutic agents approved for rare diseases (31 out of 36) were designed to compensate for disease-causing loss-of-function mutations in the target gene (Table 1). Compensation for genetic loss of function can be accomplished through pharmacological administration of the gene product—this is the case for certain genetic metabolic diseases[31], which can be partially corrected using enzyme replacement therapy (Supplementary Table 2). In cystic fibrosis, small molecules correct the malfunctioning protein encoded by specific mutations in the *CFTR* gene[32,33]. These revolutionary agents have transformed the lives of patients.

Gene and RNA therapies represent a rapidly growing category of therapies that show considerable promise for treating a variety of diseases. These modalities operate via distinct mechanisms, including gene silencing, replacement or editing. Gene silencing is achieved by administering RNA to suppress gene expression. There are two major classes of RNA therapeutic agents: double-stranded RNA-mediated interference (RNAi) and antisense oligonucleotides[34] (ASOs). Milasen was the first successful personalized ASO drug; it was developed to

**Table 1 | Classification of genetically driven therapies according to the underlying genotype–phenotype association**

| Genotype | Phenotype | |
|---|---|---|
| | **Deleterious** | **Beneficial** |
| Loss-of-function | Substitutive therapy, agonists and correctors: 31 rare diseases and 1 common disease | Mimicking inhibitors: 3 rare and 4 common diseases |
| Gain-of-function | Correcting inhibitors: 2 rare diseases | |

Variants in a target gene (genotypes) may lead to a decrease or an increase in the gene product or its function (loss-of-function or gain-of-function variants). These functional changes may have deleterious or beneficial, protective effects on a given phenotype. Derived drugs have been designed as agonists to substitute for deleterious effects of loss-of-function variants, or as inhibitors to mimic beneficial effects of loss-of-function variants or deleterious effects of gain-of-function variants. Our search did not identify any agonists that mimic the disease-protective effects of gain-of-function variants in target genes that met our definition of being genetically driven.

treat a patient with an ultra-rare mutation in the *CLN7* gene, which causes Batten disease[35]. So far, 13 RNAi and ASO therapies have received regulatory approval (Supplementary Tables 2 and 3).

Gene replacement therapy—that is, the insertion of a functional gene copy into the cell—enables transient or persistent production of a protein when there is insufficient or abnormal protein production. In 2012, alipogene tiparvovec (Glybera), a treatment for familial lipoprotein lipase deficiency[36], became the first gene replacement therapy to gain approval in Europe. Our search identified seven additional gene therapies (Supplementary Table 2). Recent trials have shown promising results for the treatment of sickle-cell disease[37,38] and transfusion-dependent beta-thalassaemia[39].

Gene editing is performed using programmable nucleases such as zinc-finger nucleases, clustered regularly interspaced short tandem repeats (CRISPR) or transcription activators such as effector nucleases[40]. Several gene-editing modalities are in development and carry high expectations. EDIT-101 is a CRISPR-based investigational treatment for Leber congenital amaurosis type 10 (ClinicalTrials.gov identifier: NCT03872479). EDIT-301 and CTX001 are currently being tested for the treatment of sickle-cell disease and beta-thalassaemia (ClinicalTrials.gov identifiers: NCT04853576, NCT05444894, NCT05329649 and NCT05356195), two conditions caused by a defective β-chain in tetrameric haemoglobin. These therapies—designed as autologous transplantation of gene-edited $CD34^+$ haematopoietic stem cells—disrupt the erythroid lineage-specific enhancer of the *BCL11A* gene and restore the production of fetal haemoglobin as a way to partly replace dysfunctional haemoglobin or compensate for low levels of adult haemoglobin, respectively[41].

## Mimicry of loss-of-function variants

In our search, two of the therapies approved for rare conditions and four of those approved for common conditions are inhibitors designed to pharmacologically mimic the disease-protective effects of loss-of-function variants within their target gene (Table 1). CCR5 inhibitors are an emblematic example. Their development was prompted by the observation that individuals who are genetically deficient in CCR5 are protected from AIDS when infected with HIV[42,43]. The FDA approved the first CCR5 inhibitor for HIV treatment in 2007 (ref. 44). Similarly, observations that carriers of disease-causing loss-of-function mutations in the *SOST* gene had high bone mass across the skeleton[45,46] were essential for the development of SOST inhibitors for osteoporosis, a common condition characterized by low bone mass leading to fractures. The SOST inhibitor romosozumab was approved for osteoporosis treatment in 2019. Similarly, the observation that carriers of loss-of-function

mutations in the *ANGPTL3* gene had low plasma lipid levels[47–49] prompted the development of ANGPTL3 inhibitors. In February 2021, the FDA approved evinacumab, a biological inhibitor of ANGPTL3, as an add-on treatment for homozygous familial hypercholesterolaemia[50]. Along the same lines, gene-editing or silencing technologies have been designed to block the production of PCSK9. Inclisiran was approved in 2022, and several similar modalities are in pre-clinical or early clinical development for permanent reduction of blood LDL cholesterol concentration and prevention of coronary artery disease[51–53]. Also in cardiovascular medicine, people with genetically low blood levels of lipoprotein(a) have a lower risk of developing coronary artery disease than those with high levels of this atherogenic lipoprotein[54]. Currently under development are the ASO pelacarsen and the small interfering RNA olpasiran that target apolipoprotein(a) mRNA, which encodes the main constituent of lipoprotein(a). Both treatments result in pronounced and sustained reductions in lipoprotein(a) levels[55–57].

## Genetically driven drug repositioning

A large body of work has underlined the pervasiveness of pleiotropy—the association of multiple traits or conditions with the same gene. Accordingly, one may assume that an existing drug approved for the treatment of a particular disease could promptly be repurposed to treat other conditions genetically associated with the same target. An example of this type of drug repositioning is 5α reductase inhibitors. Individuals deficient in 5α reductase have a small prostate and rarely develop male pattern baldness[58]. These drugs were initially developed to treat benign prostatic hyperplasia[59], but they were subsequently reformulated to prevent hair loss[60]. Sulfonylureas, first used to treat type 2 diabetes, are now also indicated for a rare form of neonatal diabetes caused by mutations in the sulfonylurea receptor gene[61]. Similarly, the observation that gain-of-function mutations in the *FGFR3* gene that predispose to cholangiocarcinoma and bladder cancer are also responsible for skeletal diseases prompted the repositioning of FGFR tyrosine kinase inhibitors—originally developed as anti-cancer drugs—for achondroplasia, a short-limbed skeletal dysplasia[62].

These examples are based on Mendelian genetics. More recently, inhibitors of interleukin-23 (IL-23) signalling, originally developed for psoriasis, have been repurposed for Crohn's disease on the basis of associations between common variants in the *IL23R* gene and Crohn's disease detected through genome-wide association studies[63,64] (GWAS). Similarly, inhibitors of IL-17A signalling initially developed for psoriasis, rheumatoid arthritis and uveitis have been tested and approved for ankylosing spondylitis on the basis of findings from GWAS (Table 2). On the development side, relying on GWAS-based evidence[65], GlaxoSmithKline is testing an inhibitor of IL-18 that originally failed to demonstrate efficacy in treating diabetes for the treatment of Crohn's disease (ClinicalTrials.gov identifier: NCT03681067) and atopic dermatitis (ClinicalTrials.gov identifier: NCT04975438). Just one year after the publication of associations between variants mapping to *TYK2* (part of the JAK gene family) and COVID-19 severity[66], a trial was launched to test the use of JAK inhibitors for the treatment of severe COVID-19. The trial showed that JAK inhibitors reduced mortality in these patients[67], illustrating the speed at which new and robust genetic observations can lead to successful repositioning of existing drugs (Table 2).

## Informativeness of rare high-impact variants

Several additional lessons can be drawn from the above examples to guide the exploitation of genetics and genomics to support the discovery and development of new drugs or indications. First, the genetic observations that led directly to approved drugs (Supplementary Table 3) and the initial examples of successful genetics-based drug repositioning (Table 2) relied on associations between rare coding variants in target genes (defined here as being found in fewer than 1 in 1,000 people) with large effects on specific phenotypes. These observations per se—although they illustrate the high information content of rare coding variants—have several implications. One is statistical power, meaning that a sufficiently large number of well-phenotyped carriers needs be identified to draw robust genetic associations. This also means that hundreds of thousands of individuals may need to have their exome (the coding portion of their genome) sequenced to achieve sufficient statistical support for discovering potential new drug targets. One strategy to overcome this challenge is to study diverse ancestries. The frequency of variants may vary considerably between ancestry groups, and variants that are rare in one ancestry may be more common in others owing to their distinct evolutionary histories[68]. Another approach is to capitalize on founder populations in which rare variants can be found at higher prevalence as a result of genetic drift—that is, changes in allele frequencies in a population due to random chance[69]. The latter approach has been particularly effective in revealing disease-causing and disease-protecting coding variants in Finland[70,71], Iceland[72], Sicily[73], among Ashkenazi Jews[74] and among French Canadians[75,76]. A recent study in the FinnGen cohort identified 26 potentially harmful variants relevant to cardiovascular and metabolic health[77]. Of these variants, 19 were unique to Finns or more than 20 times more common in Finland than elsewhere in Europe.

Most Mendelian diseases are clinically apparent when both parental copies are mutated—that is, they are recessive. Thus, another approach to identify high-impact variants is to examine consanguineous families—those in whom the parents share the same ancestors—as these families are markedly enriched in homozygous carriers of loss-of-function mutations[78–80]. Another possibility for identifying rare, disease-causing or disease-protective variants is to select participants on the basis of polygenic risk scores (PRSs) (Box 1). A PRS aggregates into a single score the effects of many genetic variants associated with a particular condition, most of which are commonly found in populations and have, individually, a small clinical effect. The larger the number of risk alleles a person carries, the higher their PRS. Individuals who present a certain condition but have a very low PRS for that particular trait tend to have a higher proportion of rare disease-causing variants[81,82]; whereas, reciprocally, disease-protecting variants are expected to be found among unaffected individuals with very high PRS for a given trait.

Finally, it must be acknowledged that the strongest genetic evidence does not necessarily translate into successful development of investigational therapies[83]. Several attempts to treat Huntington's disease (caused by a dominant mutation in the huntingtin (*HTT*) gene) using splice modulators, ASO therapies and gene therapy have failed owing to safety issues or a lack of target engagement[84]. Similarly, thousands of molecularly characterized rare Mendelian diseases remain untreatable.

## The emerging role of common variants

So far, common variants have contributed less than rarer variants to the discovery of new approved therapies (Supplementary Table 2), but they have already had a tangible impact on drug repurposing (Table 2). As the GWAS technology that enables the detection of associations between diseases and common variants emerged only in 2007, this is likely to be a result of the median 25-year lag time between genetic discovery and approval of drugs (Fig. 2b) and the much shorter time needed to demonstrate successful repurposing of an existing drug with a known safety profile. Nevertheless, some new drugs are emerging based on original observations of associations between common variants and diseases. For example, the observation that a common coding variant (I148M) in *PNPLA3* is associated with non-alcoholic fatty liver[85] and increased liver enzyme levels in the blood[86] prompted the development of PNPLA3 inhibitors, which are now being tested as potential treatments for liver disease (ClinicalTrials.gov identifiers: NCT05395481 and NCT04483947). Thanks to the commoditization of GWAS studies and their applications to large biobanks (discussed below), the genomics

**Table 2 | Nine drugs with new indications driven by genetics; four are approved and five are under development**

| Primary therapeutic area | Extended therapeutic area | Gene: drug target | Drug generic name | Drug class (example) | Primary indication | Genetic observation leading to repositioning | Extended indication |
|---|---|---|---|---|---|---|---|
| **Approved** | | | | | | | |
| Urology | Dermatology | *SRD5A2*: 5α reductase 2 | Finasteride | Inhibitor | Benign prostatic hyperplasia | Absence of baldness in patients deficient in 5α reductase | Male pattern baldness |
| Metabolic | Metabolic | *SUR*: sulfonylurea receptor | Sulfonylureas | Positive modulator | Type 2 diabetes | Neonatal diabetes due to mutations in *SUR* | Certain forms of neonatal diabetes |
| Immunology | Immunology | *IL23*: IL-23 | Ustekinumab | Signalling inhibitor | Psoriasis | *IL23R* variants have been associated with Crohn's disease in GWAS | Crohn's disease |
| | | *IL17A*: IL-17A | Secukinumab | Signalling inhibitor | Psoriasis, rheumatoid arthritis and uveitis | IL-17A forms an immune axis with IL-23, and the association of a variant downstream of *IL23R* with ankylosing spondylitis provided a rationale for repositioning | Ankylosing spondylitis |
| **Under development** | | | | | | | |
| Musculoskeletal | Immunology | *TNFRSF11A*: TNF receptor superfamily member 11a (also known as RANK) | Denosumab | Inhibitor | Postmenopausal women at high risk of fracture with osteoporosis | *TNFRSF11A* genetic variants have been associated with Crohn's disease in GWAS | Crohn's disease |
| Oncology | Musculoskeletal | *FGFR1–4*: fibroblast growth factor receptors 1, 2, 3 and 4 | Infigratinib | Inhibitor | FGFR2-dependent cholangiocarcinoma | *FGFR3* genetic mutations are causing achondroplasia | Achondroplasia |
| Immunology | Cardiovascular | *IL6R*: IL-6R | Tocilizumab | Inhibitor | Rheumatoid arthritis | GWAS associations and evidence from Mendelian randomization studies showing causal role of IL-6R signalling in development of coronary artery disease | Coronary artery disease |
| Immunology | Infectious diseases | *JAK*: Janus kinase | Baricitinib | Inhibitor | Rheumatoid arthritis | Baricitinib also has moderate inhibitory activity against TYK2 and genetic data support a causal link between high TYK2 expression and life-threatening COVID-19 | COVID-19 |
| Haematology | Ophthalmology | *CFH*: complement factor H | Eculizumab | Inhibitor of signalling | Sepsis and paroxysmal nocturnal haemoglobinuria | GWAS identified a missense mutation (Y402H) in *CFH* as an indicator of an increased risk for age-related macular degeneration | Age-related macular degeneration |

field has expanded rapidly, and important benefits from common variants genetics are expected in the next few decades.

Common variants are currently extensively exploited for investigations of potential causal relationships between risk factors and disease using Mendelian randomization analyses. Since alleles are randomly distributed at conception, this approach provides evidence of causal associations between an exposure (such as high blood levels of LDL cholesterol) and an outcome (such as coronary heart disease) by minimizing the bias from confounds (the effect of a separate variable affecting both the exposure and the outcome) and reverse causation (where the outcome causally affects the risk factor), two major limitations that affect evidence from other types of observational studies in humans[87–89]. Since most drug targets are proteins, proteome-wide Mendelian randomization analyses are used to identify new potentially disease-causing proteins and to prioritize potential drug targets[90–92].

Getting from a locus identified in a GWAS to a drug target remains a complex endeavour. Extensive post-GWAS studies are necessary to identify the causal variants underlying the genetic association and to obtain a mechanistic understanding of their functional effects. Combining genetics and genomics with other omics approaches such as proteomics, epigenomics, metabolomics or transcriptomics analyses has been a useful approach[93–98] and is recommended for future development of drugs based on genetic studies (Box 1).

Analyses based on common and rare variants are also used to anticipate potential safety issues associated with a particular drug target and to de-risk clinical development[99]. Phenome-wide association studies (PheWAS) enable hypothesis-free testing of associations between variants in a given gene and a broad range of disease outcomes. For instance, a PheWAS assessing more than 500 phenotypes using electronic medical records (EMRs) in more than 29,000 subjects found no evidence that rare, functional variants mapping to *TYK2* that are associated with rheumatoid arthritis, systemic lupus erythematosus, inflammatory bowel disease and COVID-19 are associated with any other phenotypes[100]. On the basis of this genetic information, adverse on-target effects are unlikely to occur during clinical testing (Box 1). One limitation of this approach is that it does not account for possible off-target adverse effects.

## The power and limitations of biobanks

Several large-scale initiatives integrating genomic and dense phenotypic data have been launched to capture the information that

## Box 1

# Priorities for future investments in human genetics and genomics to support the discovery and development of innovative, safe and efficacious non-cancer drugs

**Drug discovery**

- Further mining of clinical, genomic and other omics data from large existing public initiatives (such as UK Biobank, All of Us and FinnGen) to find and validate new drug targets.
- Investment in more diverse populations.
- Focusing on the identification of protective loss-of-function variants as potential targets of new inhibitors to prevent and/or treat common diseases.
- Generation of PRSs and use extremes to enrich the pool of carriers of disease-predisposing or disease-protecting coding variants.
- Expansion of Mendelian randomization and PheWAS studies to corroborate primary indication and identify early potential on-target safety signals for investigational medicines.
- Development of new tools to identify causal genes from large GWAS.
- Expansion of functional genomics studies to validate potential new drug targets.
- Investment in more studies to identify environmental and genetic determinants of penetrance of rare disease-causing variants for better risk stratification of patient populations and anticipation of clinical benefits.
- Consideration of duration, cost and risks of drug development and expectation management regarding the delivery of new genetically driven therapeutics.

**Drug development**

- Establishment of an ethical, legal and societal framework for ethnically diverse recall-by-genotype studies.
- Development of large, diverse, disease-based, deeply phenotyped, EMR-linked cohorts with consent in place for future recall-by-genotype studies.
- Pilot proof-of-concept trials across various therapeutic areas to demonstrate the feasibility and efficiency of genetically assisted clinical development trials.
- Upfront integration of highest standards in equity–diversity–inclusion in the design and execution of these studies.
- Promotion of global and equitable access to novel therapeutics to improve health.

can be gleaned from natural variations in the human genome. Most of these initiatives are public–private partnerships, including the 100,000 Genomes Project and the UK Biobank in the UK[101–105], All of Us in the USA[106], deCODE in Iceland[107], BioBank Japan[108,109], the Estonian Biobank[110] and FinnGen in Finland[77,111,112]. The UK Biobank, for instance, has collected genetic data and health information from more than 500,000 volunteers and is continuously adding data. These data, which include genomic, metabolomics and proteomic profiling have been made fully available to the research community. The impact of these studies on our understanding of the genetic and environmental determinants and the trajectories of diseases has already been immense.

Beyond enabling the discovery of disease-causing variants, current large-scale population-based biobanks are particularly well-suited to revealing rare, disease-protecting, loss-of-function variants that, as discussed above, could lead to novel therapies for common conditions. For example, exome sequencing of 645,626 individuals from the UK, USA and Mexico revealed that 4 in every 10,000 people were heterozygote carriers of loss-of-function variants in *GPR75* and these carriers had a 54% lower risk of being obese[113], suggesting that pharmacological inhibition of GPR75 may prevent or treat obesity. These findings prompted Regeneron to launch discovery campaigns for small molecules, ASOs and monoclonal antibodies that block this receptor[114]. Similarly, genetics coupled with EMR data from 46,544 participants in the DiscovEHR study identified loss-of-function variants in *HSD17B13* that were associated with slower progression of liver steatosis to steatohepatitis[115]. On the basis of these observations, Regeneron recently registered a clinical trial testing the use of HSD17B13 inhibitors to treat liver steatosis (ClinicalTrials.gov identifier: NCT04565717), exemplifying the power of longitudinal, EMR-linked biobanks.

Participants of European descent are highly over-represented in most existing studies[116]. Genetic findings arising from one ancestry group are not always generalizable to other ancestries owing to clustering of disease-associated variants in specific ancestry groups. This may in turn affect our understanding of the disease burden, the accuracy of the genetic risk model and disease prediction[117] or the responses to drugs[118]. Accordingly, the lack of participants from diverse populations[119–122] in genomic studies represents a major ethical and scientific challenge that can exacerbate existing healthcare inequalities. Thus, advancing diversity in genetic studies is now a priority for the research community[119,120]. A step forward was made by Human Heredity & Health in Africa[123–125] (H3Africa), with the goal of bringing deeper characterization of the genomic diversity of African individuals and identifying novel high-impact variants that are relevant for drug discovery.

Studies in which participants consent to being re-contacted for subsequent trials derived from the original cohort (nested studies) come with an additional advantage, although there is a limited number of such studies so far. Recontact based on genetic information can be used to perform in-depth phenotypic characterization of carriers of specific genetic variants and expand our understanding of the physiological role of these genes, including those encoding drug targets[126–129]. Best practices and policies for this type of recall-by-genotype studies are discussed in depth elsewhere[130]. Having the possibility to recontact participants on the basis of their genetic make-up can also be used to support the development of investigational therapies (discussed below). However, this option is only possible if participants are consented for this approach from the outset, and if biobanks have enough participants diagnosed with the disease selected as a drug indication. A few cohorts already exist that incorporate these requirements in their design, such as FinnGen[77,111,112], the 100,000 Genomes Project (focusing on rare diseases[102], cancer[103] and COVID-19 (ref. 104)) and—at a smaller scale—the Quebec COVID-19 Biobank[97,131] in Canada and the Lausanne Institutional Biobank in Switzerland[132–134].

## Embedding genetics in clinical drug development

In light of the emerging abundance of new, robustly genetically validated potential drug targets, innovative approaches are needed to prioritize investigational compounds to reduce the costs and duration of their clinical development and increase the probability of approval

(Fig. 1). Embedding genetics and genomics into the clinical development process has the potential to address some of these challenges. More explicitly, genetics can be used to increase the number of patients in clinical trials who are predisposed to respond optimally to the investigational therapy, thus enabling smaller trials with shorter durations and lower costs, and facilitating early attrition (that is, a drug that fails in a patient population enriched to optimally respond to it has a very low chance of being efficacious in non-selected patients with the same disease indication).

In early clinical development, results from phase IIa trials (proof-of-concept studies usually performed to demonstrate whether the investigational therapy is efficacious on a small number of patients) are carefully scrutinized before the decision is made to progress to clinical development and/or to optimize the parameters of subsequent, larger scale phase IIb/III trials (Fig. 1). Genetics has already contributed to proof-of-concept trials. Forty years ago, statins, which increase the clearance of blood-borne LDL particles by upregulating the LDL receptor, were originally tested on patients who were genetically deficient in one copy of the LDL receptor gene. Only seven patients[135] were needed to show that the investigational drug was highly effective at lowering blood levels of LDL cholesterol. Subsequent statin trials were expanded to include patients with more modestly increased LDL cholesterol levels and showed that the drug could prevent heart attacks in these populations.

Conceptually, a similar strategy using genetic enrichment of expected drug responders could be used to show proof of concept for any investigational agonists or antagonists. Agonists can be tested in individuals with a partial or complete deficiency in the target, or genes in the target pathway, to demonstrate correction or improvement of the associated phenotype. This approach has been successful for CFTR modulators in cystic fibrosis[136] and cancer drugs[137]. Conversely, inhibitors can be tested in patients with gain-of-function mutations in the targeted gene or pathway. Inhibitors of LRRK2 and PNPLA3 are being tested in individuals with gain-of-function variants in these genes, which cause Parkinson's disease (ClinicalTrials.gov identifier: NCT03976349) and fatty liver disease (ClinicalTrials.gov identifier: NCT03976349), respectively. Information gleaned from these genetically enriched trials can be critical for showing target engagement and supporting internal decision making, more than for final approval by regulatory agencies.

Genetics and genomics can also be used to enrich phase IIb/III trials with high-risk patients to reduce costs and diminish the size and duration of these usually very expensive trials[138]. For example, high-risk carriers of the *APOE* E4 risk variant have been selected for Alzheimer prevention trials[139,140]. Alternatively, PRSs can be used to identify individuals at greatest risk of disease[141]. The power of this approach was supported by retrospective analyses of two independent phase III trials, which demonstrated that patients who had a high PRS for coronary artery disease received substantially greater benefit from PCSK9 inhibitors than participants with only non-genetic risk factors[142,143].

Individuals absorb, distribute, metabolize and eliminate medicines in a variable manner. This holds particularly true for small molecules[144]. These parameters are partly under genetic control and have been extensively studied in pharmacogenetics. CYP2D6 for instance, is a key enzyme that is responsible for the metabolism and elimination of around 25% of drugs used in the clinic, such as antidepressants, beta blockers, opioids and antiarrhythmics[145]. Carriers of variants that increase or decrease CYP2D6 activity metabolize its substrates more or less rapidly. To take this inter-individual variability into account, systematic pharmacogenetic assessment has been broadly advocated for premarket evaluation in early phase clinical studies of new therapies[146].

## The tip of the iceberg

In our effort to quantify the returns of investments made in genetics in terms of the discovery and development of new therapies, we identified 47 first-in-class approved therapies that had been directly driven by original human genetic observations. We also identified several examples of successful genetically driven repurposing of existing drugs, and have highlighted examples that demonstrate the value of genetically assisted design of clinical trials to accelerate development and reduce costs. Given the conservative definition that we used in our search, the large body of genetic knowledge that is accumulating as a result of the investments in large biobanks and technologies, and the decades-long lag time between the discovery of a genetic association and approval of derived therapies, results from our search represent the tip of the iceberg of the full impact of genetics and genomics on the discovery and validation of new drugs.

Capturing the full benefits of genetics and genomics while ensuring equitable access to these benefits requires the establishment of additional, large, ethnically diverse cohorts of participants, including their medical information. Proper consent is critical, as participants must specifically agree to have detailed analyses performed on their biobanked samples such as genomics, transcriptomics, proteomics or metabolomics, to have their EMR data available to investigators and potentially to be re-contacted for additional nested studies. In parallel, harmonization to enable inter-institutional or international collaborations between recruitment centres may be required to reach a sufficiently large number of participants to draw adequate statistical power. Finally, equitable and fair access to benefits—including new drugs driven by genetics—will require concerted efforts to address existing disparities in health-related knowledge, scientific capacity, workforce and clinical translation practices. More specifically, ensuring that the populations whose genetic information has been used to identify or validate their targets, or to accelerate their clinical development, are actively involved in the research and decision-making activities and can access these therapies should be a fundamental objective for companies manufacturing and marketing them. Increased investments into large and diverse cohort studies and in a robust ethical, legal, societal and financial framework[147] that facilitates responsible, sustainable benefit sharing need to be planned now.


1. Arrowsmith, J. & Miller, P. Trial watch: phase II and phase III attrition rates 2011-2012. *Nat. Rev. Drug Discov.* **12**, 569 (2013).
2. Harrison, R. K. Phase II and phase III failures: 2013-2015. *Nat. Rev. Drug Discov.* **15**, 817–818 (2016).
3. Wouters, O. J., McKee, M. & Luyten, J. Estimated research and development investment needed to bring a new medicine to market, 2009-2018. *JAMA* **323**, 844–853 (2020).
4. Cook, D. et al. Lessons learned from the fate of AstraZeneca's drug pipeline: a five-dimensional framework. *Nat. Rev. Drug Discov.* **13**, 419–431 (2014).
5. Razuvayevskaya, O., Lopez, I., Dunham, I. & Ochoa, D. Why clinical trials stop: the role of genetics. Preprint at *medRxiv* https://doi.org/10.1101/2023.02.07.23285407 (2023).
6. Collins, F. S. Shattuck lecture—Medical and societal consequences of the Human Genome Project. *N. Engl. J. Med.* **341**, 28–37 (1999).
7. Lander, E. S. et al. Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
8. Venter, J. C. et al. The sequence of the human genome. *Science* **291**, 1304–1351 (2001).
9. Spotlight on cancer genomics. *Nat. Cancer* **1**, 265–266 (2020).
10. Nguengang Wakap, S. et al. Estimating cumulative point prevalence of rare diseases: analysis of the Orphanet database. *Eur. J. Hum. Genet.* **28**, 165–173 (2020).
11. Abdellaoui, A., Yengo, L., Verweij, K. J. H. & Visscher, P. M. 15 years of GWAS discovery: realizing the promise. *Am. J. Hum. Genet.* **110**, 179–194 (2023).
12. Malone, E. R., Oliva, M., Sabatini, P. J. B., Stockley, T. L. & Siu, L. L. Molecular profiling for precision cancer therapies. *Genome Med.* **12**, 8 (2020).
13. Waarts, M. R., Stonestrom, A. J., Park, Y. C. & Levine, R. L. Targeting mutations in cancer. *J. Clin. Invest.* **132**, e154943 (2022).
14. Sabatine, M. S. et al. Efficacy and safety of evolocumab in reducing lipids and cardiovascular events. *N. Engl. J. Med.* **372**, 1500–1509 (2015).
15. Hall, S. S. Genetics: a gene of rare effect. *Nature* **496**, 152–155 (2013).
16. Abifadel, M. et al. Mutations in *PCSK9* cause autosomal dominant hypercholesterolemia. *Nat. Genet.* **34**, 154–156 (2003).
17. Zhao, Z. et al. Molecular characterization of loss-of-function mutations in PCSK9 and identification of a compound heterozygote. *Am. J. Hum. Genet.* **79**, 514–523 (2006).
18. 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).
**To our knowledge, refs. 16–18 provided the first genetic rationale for the discovery and development of PCKS9 inhibitors as a safe way to treat familial hypercholesterolemia and coronary artery disease.**

19. Raedler, L. A. Praluent (alirocumab): first PCSK9 inhibitor approved by the FDA for hypercholesterolemia. *Am. Health Drug Benefits* **9**, 123–126 (2016).
20. King, E. A., Davis, J. W. & Degner, J. F. 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. *PLoS Genet.* **15**, e1008489 (2019).
21. Nelson, M. R. et al. The support of human genetic evidence for approved drug indications. *Nat. Genet.* **47**, 856–860 (2015).
**References 20,21 demonstrated retrospectively that genetically supported targets are more likely to lead to approved therapies**.
22. Ochoa, D. et al. Human genetics evidence supports two-thirds of the 2021 FDA-approved drugs. *Nat. Rev. Drug Discov.* **21**, 551 (2022).
23. Seidah, N. G. The PCSK9 revolution and the potential of PCSK9-based therapies to reduce LDL-cholesterol. *Glob. Cardiol. Sci. Pract.* **2017**, e201702 (2017).
24. Ochoa, D. et al. The next-generation Open Targets Platform: reimagined, redesigned, rebuilt. *Nucleic Acids Res.* **51**, D1353–D1359 (2022).
25. Wishart, D. S. et al. DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res.* **34**, D668–672 (2006).
26. Sparrow, A. J., Watkins, H., Daniels, M. J., Redwood, C. & Robinson, P. Mavacamten rescues increased myofilament calcium sensitivity and dysregulation of $Ca^{2+}$ flux caused by thin filament hypertrophic cardiomyopathy mutations. *Am. J. Physiol. Heart. Circ. Physiol.* **318**, H715–h722 (2020).
27. Watkins, H., Ashrafian, H. & Redwood, C. Inherited cardiomyopathies. *N. Engl. J. Med.* **364**, 1643–1656 (2011).
28. Toepfer, C. N. et al. Myosin sequestration regulates sarcomere function, cardiomyocyte energetics, and metabolism, informing the pathogenesis of hypertrophic cardiomyopathy. *Circulation* **141**, 828–842 (2020).
29. Santos, R. et al. A comprehensive map of molecular drug targets. *Nat. Rev. Drug Discov.* **16**, 19–34 (2017).
30. Brinkman, R. R., Dube, M. P., Rouleau, G. A., Orr, A. C. & Samuels, M. E. Human monogenic disorders—a source of novel drug targets. *Nat. Rev. Genet.* **7**, 249–260 (2006).
31. Desnick, R. J. & Schuchman, E. H. Enzyme replacement therapy for lysosomal diseases: lessons from 20 years of experience and remaining challenges. *Annu. Rev. Genomics Hum. Genet.* **13**, 307–335 (2012).
32. Yang, H. et al. Nanomolar affinity small molecule correctors of defective Delta F508-CFTR chloride channel gating. *J. Biol. Chem.* **278**, 35079–35085 (2003).
33. Van Goor, F. et al. Rescue of CF airway epithelial cell function in vitro by a CFTR potentiator, VX-770. *Proc. Natl Acad. Sci. USA* **106**, 18825–18830 (2009).
34. RNAi therapeutics market: growing investments in RNAi therapies. *BioSpace* https://www.biospace.com/article/rnai-therapeutics-market-growing-investments-in-rnai-therapies/ (2021).
35. Kim, J. et al. Patient-customized oligonucleotide therapy for a rare genetic disease. *N. Engl. J. Med.* **381**, 1644–1652 (2019).
36. European Medicines Agency recommends first gene therapy for approval. *European Medicines Agency* https://www.ema.europa.eu/en/news/european-medicines-agency-recommends-first-gene-therapy-approval (2012).
37. Ribeil, J. A. et al. Gene therapy in a patient with sickle cell disease. *N. Engl. J. Med.* **376**, 848–855 (2017).
38. Esrick, E. B. et al. Post-transcriptional genetic silencing of *BCL11A* to treat sickle cell disease. *N. Engl. J. Med.* **384**, 205–215 (2021).
39. Thompson, A. A. et al. Gene therapy in patients with transfusion-dependent beta-thalassemia. *N. Engl. J. Med.* **378**, 1479–1493 (2018).
40. Li, H. et al. Applications of genome editing technology in the targeted therapy of human diseases: mechanisms, advances and prospects. *Signal Transduct. Target. Ther.* **5**, 1 (2020).
41. Frangoul, H. et al. CRISPR–Cas9 gene editing for sickle cell disease and beta-thalassemia. *N. Engl. J. Med.* **384**, 252–260 (2021).
42. Dean, M. et al. Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the CKR5 structural gene. Hemophilia Growth and Development Study, Multicenter AIDS Cohort Study, Multicenter Hemophilia Cohort Study, San Francisco City Cohort, ALIVE Study. *Science* **273**, 1856–1862 (1996).
43. Samson, M. et al. Resistance to HIV-1 infection in Caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene. *Nature* **382**, 722–725 (1996).
44. FDA approves maraviroc tablets. *AIDS Patient Care STDs* **21**, 702 (2007).
45. Piters, E. et al. First missense mutation in the SOST gene causing sclerosteosis by loss of sclerostin function. *Hum. Mutat.* **31**, E1526–E1543 (2010).
46. Balemans, W. et al. Identification of a 52 kb deletion downstream of the SOST gene in patients with van Buchem disease. *J. Med. Genet.* **39**, 91–97 (2002).
47. Musunuru, K. et al. Exome sequencing, ANGPTL3 mutations, and familial combined hypolipidemia. *N. Engl. J. Med.* **363**, 2220–2227 (2010).
48. Martin-Campos, J. M. et al. Identification of a novel mutation in the *ANGPTL3* gene in two families diagnosed of familial hypobetalipoproteinemia without *APOB* mutation. *Clin. Chim. Acta* **413**, 552–555 (2012).
49. Romeo, S. et al. Rare loss-of-function mutations in ANGPTL family members contribute to plasma triglyceride levels in humans. *J. Clin. Invest.* **119**, 70–79 (2009).
50. Mullard, A. FDA approves first anti-ANGPTL3 antibody, for rare cardiovascular indication. *Nat. Rev. Drug Discov.* **20**, 251 (2021).
51. Banerjee, Y. et al. Inclisiran: a small interfering RNA strategy targeting PCSK9 to treat hypercholesterolemia. *Expert Opin. Drug Saf.* **21**, 9–20 (2022).
52. Wang, L. et al. Long-term stable reduction of low-density lipoprotein in nonhuman primates following in vivo genome editing of *PCSK9*. *Mol. Ther.* **29**, 2019–2029 (2021).
53. Musunuru, K. et al. In vivo CRISPR base editing of *PCSK9* durably lowers cholesterol in primates. *Nature* **593**, 429–434 (2021).
54. Reyes-Soffer, G. et al. Lipoprotein(a): a genetically determined, causal, and prevalent risk factor for atherosclerotic cardiovascular disease: a scientific statement from the American Heart Association. *Arterioscler. Thromb. Vasc. Biol.* **42**, e48–e60 (2022).
55. Tsimikas, S. et al. Lipoprotein(a) reduction in persons with cardiovascular disease. *N. Engl. J. Med.* **382**, 244–255 (2020).
56. Yeang, C. et al. Effect of pelacarsen on lipoprotein(a) cholesterol and corrected low-density lipoprotein cholesterol. *J. Am. Coll. Cardiol.* **79**, 1035–1046 (2022).
57. Koren, M. J. et al. Preclinical development and phase 1 trial of a novel siRNA targeting lipoprotein(a). *Nat. Med.* **28**, 96–103 (2022).
58. Mendonca, B. B. et al. Steroid 5α-reductase 2 deficiency. *J. Steroid Biochem. Mol. Biol.* **163**, 206–211 (2016).
59. Gormley, G. J. et al. The effect of finasteride in men with benign prostatic hyperplasia. The Finasteride Study Group. *N. Engl. J. Med.* **327**, 1185–1191 (1992).
60. Kaufman, K. D. et al. Finasteride in the treatment of men with androgenetic alopecia. Finasteride Male Pattern Hair Loss Study Group. *J. Am. Acad. Dermatol.* **39**, 578–589 (1998).
61. Rafiq, M. et al. Effective treatment with oral sulfonylureas in patients with diabetes due to sulfonylurea receptor 1 (*SUR1*) mutations. *Diabetes Care* **31**, 204–209 (2008).
62. Savarirayan, R. et al. Infigratinib in children with achondroplasia: the PROPEL and PROPEL 2 studies. *Ther. Adv. Musculoskelet. Dis.* **14**, 1759720x221084848 (2022).
63. Duerr, R. H. et al. A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. *Science* **314**, 1461–1463 (2006).
64. Wellcome Trust Case Control Consortium. Association scan of 14,500 nonsynonymous SNPs in four diseases identifies autoimmunity variants. *Nat. Genet.* **39**, 1329–1337 (2007).
65. Mokry, L. E. et al. Interleukin-18 as a drug repositioning opportunity for inflammatory bowel disease: a Mendelian randomization study. *Sci. Rep.* **9**, 9386 (2019).
66. Niemi, M. E. K. et al. Mapping the human genetic architecture of COVID-19. *Nature* **600**, 472–477 (2021).
67. RECOVERY Collaborative Group. Baricitinib in patients admitted to hospital with COVID-19 (RECOVERY): a randomised, controlled, open-label, platform trial and updated meta-analysis. *Lancet* **400**, 1102 (2022).
68. Tishkoff, S. A. & Verrelli, B. C. Patterns of human genetic diversity: implications for human evolutionary history and disease. *Annu. Rev. Genomics Hum. Genet.* **4**, 293–340 (2003).
69. Slatkin, M. A population-genetic test of founder effects and implications for Ashkenazi Jewish diseases. *Am. J. Hum. Genet.* **75**, 282–293 (2004).
70. de la Chapelle, A. Disease gene mapping in isolated human populations: the example of Finland. *J. Med. Genet.* **30**, 857–865 (1993).
71. Kurki, M. I. et al. FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature* **613**, 508–518 (2023).
72. Moises, H. W. et al. An international two-stage genome-wide search for schizophrenia susceptibility genes. *Nat. Genet.* **11**, 321–324 (1995).
73. Spedicati, B. et al. Natural human knockouts and Mendelian disorders: deep phenotyping in Italian isolates. *Eur. J. Hum. Genet.* **29**, 1272–1281 (2021).
74. Baskovich, B. et al. Expanded genetic screening panel for the Ashkenazi Jewish population. *Genet. Med.* **18**, 522–528 (2016).
75. Bherer, C. et al. Admixed ancestry and stratification of Quebec regional populations. *Am. J. Phys. Anthropol.* **144**, 432–441 (2011).
76. Laberge, A. M. et al. Population history and its impact on medical genetics in Quebec. *Clin. Genet.* **68**, 287–301 (2005).
77. Locke, A. E. et al. Exome sequencing of Finnish isolates enhances rare-variant association power. *Nature* **572**, 323–328 (2019).
78. Erzurumluoglu, A. M., Shihab, H. A., Rodriguez, S., Gaunt, T. R. & Day, I. N. Importance of genetic studies in consanguineous populations for the characterization of novel human gene functions. *Ann. Hum. Genet.* **80**, 187–196 (2016).
79. Minikel, E. V. et al. Evaluating drug targets through human loss-of-function genetic variation. *Nature* **581**, 459–464 (2020).
**This analysis provides a comprehensive roadmap for human knockout studies to discover relevant loss-of-function variants that can guide drug discovery and development**.
80. Saleheen, D. et al. Human knockouts and phenotypic analysis in a cohort with a high rate of consanguinity. *Nature* **544**, 235–239 (2017).
81. Lu, T. et al. Individuals with common diseases but with a low polygenic risk score could be prioritized for rare variant screening. *Genet. Med.* **23**, 508–515 (2021).
82. Wu, H. et al. Polygenic risk score for low-density lipoprotein cholesterol is associated with risk of ischemic heart disease and enriches for individuals with familial hypercholesterolemia. *Circ. Genom. Precis. Med.* **14**, e003106 (2021).
83. Pfizer and Ionis announce discontinuation of vupanorsen clinical development program. *CISION* https://www.prnewswire.com/news-releases/pfizer-and-ionis-announce-discontinuation-of-vupanorsen-clinical-development-program-301471041.html (2022).
84. Kingwell, K. Double setback for ASO trials in Huntington disease. *Nat. Rev. Drug Discov.* **20**, 412–413 (2021).
85. Romeo, S. et al. Genetic variation in PNPLA3 confers susceptibility to nonalcoholic fatty liver disease. *Nat. Genet.* **40**, 1461–1465 (2008).
86. Yuan, X. et al. Population-based genome-wide association studies reveal six loci influencing plasma levels of liver enzymes. *Am. J. Hum. Genet.* **83**, 520–528 (2008).
87. Davey Smith, G. & Hemani, G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum. Mol. Genet.* **23**, R89–98 (2014).
88. Gill, D. et al. Mendelian randomization for studying the effects of perturbing drug targets. *Wellcome Open Res.* **6**, 16 (2021).
89. Ference, B. A. Mendelian randomization studies: using naturally randomized genetic data to fill evidence gaps. *Curr. Opin. Lipidol.* **26**, 566–571 (2015).
90. Schmidt, A. F. et al. Genetic drug target validation using Mendelian randomisation. *Nat. Commun.* **11**, 3255 (2020).
**This study provides an analytical framework on how to utilize Mendelian randomization for drug target validation**.
91. Zheng, J. et al. Phenome-wide Mendelian randomization mapping the influence of the plasma proteome on complex diseases. *Nat. Genet.* **52**, 1122–1131 (2020).
92. Ferkingstad, E. et al. Large-scale integration of the plasma proteome with genetics and disease. *Nat. Genet.* **53**, 1712–1721 (2021).
**This article depicts in great depth the largest genome-wide association studies of plasma protein levels to date and describes 938 genes encoding potential drug targets with variants that influence levels of protein biomarkers**.

93. Miller, C. L. et al. Integrative functional genomics identifies regulatory mechanisms at coronary artery disease loci. *Nat. Commun.* **7**, 12092 (2016).
94. Ganesan, A., Arimondo, P. B., Rots, M. G., Jeronimo, C. & Berdasco, M. The timeline of epigenetic drug discovery: from reality to dreams. *Clin. Epigenetics* **11**, 174 (2019).
95. Steinberg, J. et al. Integrative epigenomics, transcriptomics and proteomics of patient chondrocytes reveal genes and pathways involved in osteoarthritis. *Sci. Rep.* **7**, 8935 (2017).
96. Spreafico, R., Soriaga, L. B., Grosse, J., Virgin, H. W. & Telenti, A. Advances in genomics for drug development. *Genes* **11**, 942 (2020).
97. Zhou, S. et al. A Neanderthal OAS1 isoform protects individuals of European ancestry against COVID-19 susceptibility and severity. *Nat. Med.* **27**, 659–667 (2021).
98. Yoshiji, S. et al. Proteome-wide Mendelian randomization implicates nephronectin as an actionable mediator of the effect of obesity on COVID-19 severity. *Nat. Metab.* **5**, 248–264 (2023).
99. Carss, K. J. et al. Using human genetics to improve safety assessment of therapeutics. *Nat. Rev. Drug Discov.* **22**, 145–162 (2023).
**This review discusses how genetics can be used to anticipate potential safety issues associated with a particular drug target and to de-risk clinical development**.
100. Diogo, D. et al. TYK2 protein-coding variants protect against rheumatoid arthritis and autoimmunity, with no evidence of major pleiotropic effects on non-autoimmune complex traits. *PLoS ONE* **10**, e0122271 (2015).
101. Marx, V. The DNA of a nation. *Nature* **524**, 503–505 (2015).
102. 100,000 Genomes Project Pilot Investigators. 100,000 Genomes Pilot on Rare-Disease Diagnosis in Health Care—preliminary report. *N. Engl. J. Med.* **385**, 1868–1880 (2021).
103. Turnbull, C. Introducing whole-genome sequencing into routine cancer care: the Genomics England 100,000 Genomes Project. *Ann. Oncol.* **29**, 784–787 (2018).
104. Kousathanas, A. et al. Whole-genome sequencing reveals host factors underlying critical COVID-19. *Nature* **607**, 97–103 (2022).
105. Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
106. Sankar, P. L. & Parker, L. S. The Precision Medicine Initiative's All of Us research program: an agenda for research on its ethical, legal, and social issues. *Genet. Med.* **19**, 743–750 (2017).
107. Gudbjartsson, D. F. et al. Large-scale whole-genome sequencing of the Icelandic population. *Nat. Genet.* **47**, 435–444 (2015).
108. Sakaue, S. et al. A cross-population atlas of genetic associations for 220 human phenotypes. *Nat. Genet.* **53**, 1415–1424 (2021).
**This study provides an atlas of up to 220 genotype–phenotype associations in non-Europeans that enriches the field with data from diverse populations**.
109. Nagai, A. et al. Overview of the BioBank Japan Project: study design and profile. *J. Epidemiol.* **27**, S2–s8 (2017).
110. Leitsalu, L. et al. Cohort profile: Estonian Biobank of the Estonian Genome Center, University of Tartu. *Int. J. Epidemiol.* **44**, 1137–1147 (2014).
111. Gharahkhani, P. et al. Genome-wide meta-analysis identifies 127 open-angle glaucoma loci with consistent effect across ancestries. *Nat. Commun.* **12**, 1258 (2021).
112. Tanigawa, Y. et al. Rare protein-altering variants in *ANGPTL7* lower intraocular pressure and protect against glaucoma. *PLoS Genet.* **16**, e1008682 (2020).
113. & Akbari, P. et al. Sequencing of 640,000 exomes identifies GPR75 variants associated with protection from obesity. *Science* **373**, eab8683 (2021).
**This publication reports the discovery of loss-of-function variants in GPR75 and protection from obesity**.
114. Regeneron Genetics Center discovers GPR75 gene mutations that protect against obesity. *Regeneron* https://investor.regeneron.com/news-releases/news-release-details/regeneron-genetics-center-discovers-gpr75-gene-mutations-protect (2021).
115. 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).
116. Sirugo, G., Williams, S. M. & Tishkoff, S. A. The missing diversity in human genetic studies. *Cell* **177**, 26–31 (2019).
117. Martin, A. R. et al. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat. Genet.* **51**, 584–591 (2019).
118. Ramamoorthy, A., Pacanowski, M. A., Bull, J. & Zhang, L. Racial/ethnic differences in drug disposition and response: review of recently approved drugs. *Clin. Pharmacol. Ther.* **97**, 263–273 (2015).
119. Gurdasani, D., Barroso, I., Zeggini, E. & Sandhu, M. S. Genomics of disease risk in globally diverse populations. *Nat. Rev. Genet.* **20**, 520–535 (2019).
120. Fatumo, S. et al. A roadmap to increase diversity in genomic studies. *Nat. Med.* **28**, 243–250 (2022).
121. Popejoy, A. B. & Fullerton, S. M. Genomics is failing on diversity. *Nature* **538**, 161–164 (2016).
122. Bien, S. A. et al. The future of genomic studies must be globally representative: perspectives from PAGE. *Annu. Rev. Genomics Hum. Genet.* **20**, 181–200 (2019).
123. Wonkam, A. Sequence three million genomes across Africa. *Nature* **590**, 209–211 (2021).
**This commentary discusses the research benefits of analysing Africa's genetic variation andsets up a framework to sequence three million African genomes**.
124. H3Africa Consortium. Research capacity. Enabling the genomic revolution in Africa. *Science* **344**, 1346–1348 (2014).
125. Choudhury, A. et al. High-depth African genomes inform human migration and health. *Nature* **586**, 741–748 (2020).
126. Wade, K. H. et al. Assessing the causal role of body mass index on cardiovascular health in young adults: Mendelian randomization and recall-by-genotype analyses. *Circulation* **138**, 2187–2201 (2018).
127. Corbin, L. J. et al. Metabolic characterisation of disturbances in the APOC3/triglyceride-rich lipoprotein pathway through sample-based recall by genotype. *Metabolomics* **16**, 69 (2020).
128. Hellmich, C. et al. Genetics, sleep and memory: a recall-by-genotype study of ZNF804A variants and sleep neurophysiology. *BMC Med. Genet.* **16**, 96 (2015).
129. Alver, M. et al. Recall by genotype and cascade screening for familial hypercholesterolemia in a population-based biobank from Estonia. *Genet. Med.* **21**, 1173–1180 (2019).
130. Mascalzoni, D. et al. Balancing scientific interests and the rights of participants in designing a recall by genotype study. *Eur. J. Hum. Genet.* **29**, 1146–1157 (2021).
**This article describes the best practices and policies for recall-by-genotype studies**.
131. Tremblay, K. et al. The Biobanque quebecoise de la COVID-19 (BQC19)-A cohort to prospectively study the clinical and biological determinants of COVID-19 clinical trajectories. *PLoS ONE* **16**, e0245031 (2021).
132. Mooser, V. & Currat, C. The Lausanne Institutional Biobank: a new resource to catalyse research in personalised medicine and pharmaceutical sciences. *Swiss Med. Wkly* **144**, w14033 (2014).
133. Maurer, F. et al. Identification and molecular characterisation of Lausanne Institutional Biobank participants with familial hypercholesterolaemia—a proof-of-concept study. *Swiss Med. Wkly* **146**, w14326 (2016).
134. Bochud, M., Currat, C., Chapatte, L., Roth, C. & Mooser, V. High participation rate among 25,721 patients with broad age range in a hospital-based research project involving whole-genome sequencing—the Lausanne Institutional Biobank. *Swiss Med. Wkly* **147**, w14528 (2017).
135. Mabuchi, H. et al. Effect of an inhibitor of 3-hydroxy-3-methylglutaryl coenzyme A reductase on serum lipoproteins and ubiquinone-10-levels in patients with familial hypercholesterolemia. *N. Engl. J. Med.* **305**, 478–482 (1981).
136. Ramsey, B. W. et al. A CFTR potentiator in patients with cystic fibrosis and the G551D mutation. *N. Engl. J. Med.* **365**, 1663–1672 (2011).
137. Chapman, P. B. et al. Improved survival with vemurafenib in melanoma with BRAF V600E mutation. *N. Engl. J. Med.* **364**, 2507–2516 (2011).
138. Mead, S. et al. Clinical trial simulations based on genetic stratification and the natural history of a functional outcome measure in Creutzfeldt–Jakob disease. *JAMA Neurol.* **73**, 447–455 (2016).
139. van Bokhoven, P. et al. The Alzheimer's disease drug development landscape. *Alzheimers Res. Ther.* **13**, 186 (2021).
140. Lopes Alves, I. et al. Strategies to reduce sample sizes in Alzheimer's disease primary and secondary prevention trials using longitudinal amyloid PET imaging. *Alzheimers Res. Ther.* **13**, 82 (2021).
141. Fahed, A. C., Philippakis, A. A. & Khera, A. V. The potential of polygenic scores to improve cost and efficiency of clinical trials. *Nat. Commun.* **13**, 2922 (2022).
142. Damask, A. et al. Patients with high genome-wide polygenic risk scores for coronary artery disease may receive greater clinical benefit from alirocumab treatment in the ODYSSEY OUTCOMES trial. *Circulation* **141**, 624–636 (2020).
143. Marston, N. A. et al. Predicting benefit from evolocumab therapy in patients with atherosclerotic disease using a genetic risk score: results from the FOURIER trial. *Circulation* **141**, 616–623 (2020).
**These two independent articles (refs. 142,143) show that patients with a high polygenic risk score for coronary artery disease benefit most from PCSK9 inhibition, suggesting that the selection of participants based on their genomic profile may be useful in early trials**.
144. Roden, D. M. et al. Benefit of preemptive pharmacogenetic information on clinical outcome. *Clin. Pharmacol. Ther.* **103**, 787–794 (2018).
145. Petrović, J., Pešić, V. & Lauschke, V. M. Frequencies of clinically important *CYP2C19* and *CYP2D6* alleles are graded across Europe. *Eur J Hum Genet* **28**, 88–94 (2020).
146. US Department of Health and Human Services et al. *Clinical Pharmacogenomics: Premarket Evaluation in Early-Phase Clinical Studies and Recommendations for Labeling* https://www.fda.gov/files/drugs/published/Clinical-Pharmacogenomics--Premarket-Evaluation-in-Early-Phase-Clinical-Studies-and-Recommendations-for-Labeling.pdf (FDA, 2013).
147. Government Chief Scientific Adviser. *Genomics Beyond Health* https://www.gov.uk/government/publications/genomics-beyond-health (Government Office for Science, 2022).



**Acknowledgements The** authors thank B. Dalton, P. Gros, H. Hobbs, R. McInnes, D. Roden, R. Touyz and G. Waeber for their careful reading of the manuscript and their advice. This work was supported by the McGill Canada Excellence Research Chair Program in Genomic Medicine.


**Author contributions** K.T. and V.M. planned and co-wrote the manuscript and conceptualized and designed the figures and tables. C.B., D.T., S.Z. and J.B.R. provided conceptual input and edited and revised the manuscript.

**Competing interests** V.M. has received honoraria from DalGene and shares from MedeLoop. J.B.R. is the founder of 5 Prime Sciences, has served as a consultant to GlaxoSmithKline and Deerfield Capital for their genetics programmes and has received shares from MedeLoop. The other authors declare no competing interests.

**Additional information**


**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s41586-023-06388-8.



**Correspondence and requests for materials** should be addressed to Vincent Mooser.



**Peer review information** *Nature* thanks Edoardo Aromataris, Matthew Nelson and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.