A GUIDE TO DRUG DISCOVERY

Validating therapeutic targets through human genetics

Robert M. Plenge[1,2], Edward M. Scolnick[2,3] and David Altshuler[2,4,5]

Abstract | More than 90% of the compounds that enter clinical trials fail to demonstrate sufficient safety and efficacy to gain regulatory approval. Most of this failure is due to the limited predictive value of preclinical models of disease, and our continued ignorance regarding the consequences of perturbing specific targets over long periods of time in humans. 'Experiments of nature' — naturally occurring mutations in humans that affect the activity of a particular protein target or targets — can be used to estimate the probable efficacy and toxicity of a drug targeting such proteins, as well as to establish causal rather than reactive relationships between targets and outcomes. Here, we describe the concept of dose–response curves derived from experiments of nature, with an emphasis on human genetics as a valuable tool to prioritize molecular targets in drug development. We discuss empirical examples of drug–gene pairs that support the role of human genetics in testing therapeutic hypotheses at the stage of target validation, provide objective criteria to prioritize genetic findings for future drug discovery efforts and highlight the limitations of a target validation approach that is anchored in human genetics.

[1]*Division of Rheumatology, Immunology and Allergy, Brigham And Women's Hospital, Boston, Massachusetts 02115, USA.*
[2]*Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA.*
[3]*Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA.*
[4]*Department of Molecular Biology and Diabetes Unit, Massachusetts General Hospital, Boston, Massachusetts 02114, USA.*
[5]*Department of Genetics and Medicine, Harvard Medical School, Boston, Massachusetts 02115, USA.*
Correspondence to R.M.P.
e-mail: rplenge@partners.org
doi:10.1038/nrd4051
Published online 19 July 2013

There has been a steady decline in the number of new drugs developed per US dollar spent on research and development (R&D) in the pharmaceutical industry[1]. Investment has grown from $10 billion to $60 billion per year, with the number of new molecular entities remaining steady at ~20 per year. In trying to understand why the cost per successful drug has risen dramatically, perhaps the most important observation is that less than 5% of the molecules that enter Phase I clinical trials are eventually approved as safe and effective therapeutics by the US Food and Drug Administration (FDA)[2,3]. That is, the cost of drug development is not dominated by the cost of the few programmes that succeed, but instead by the amortized cost of the other programmes that fail during clinical trials[3].

Thus, perhaps the most crucial question is: why do drugs fail? Analyses have shown that most failures are in Phase II trials, and at least 50% of these are due to lack of efficacy and 25% due to toxicity[2,4]. These failures occur despite the fact that the initiation of clinical trials is essentially always preceded by evidence that the drug candidate engages its target *in vitro* and is safe and effective in preclinical models. It follows that high failure rates indicate a key issue in drug discovery: the limited ability of preclinical disease models to predict benefit in patients[3].

In this Review, we highlight the crucial importance of the therapeutic hypothesis at the stage when a protein or biomolecule is nominated as a potential drug target (often referred to as target validation). In this context, 'therapeutic hypothesis' refers to the hypothesis that perturbing a target in a given manner will benefit patients and have minimal (or at least acceptable) toxicity (FIG. 1). Ideally, data for validating a therapeutic hypothesis would be derived from the patient population of interest and would involve direct perturbation of a target with a known function in a known direction. The result of the perturbation would be followed in many patients for many years, leading to the accumulation of all possible clinical outcomes. Finally, it would be ideal to obtain all of this information before a clinical trial is initiated. Strictly speaking, the only truly validated targets are those that are already successfully modulated by a safe and effective therapeutic. But for many diseases there is a lack of highly effective approaches for prevention and treatment, and so new mechanisms of action are needed.

Preclinical dose–response curves

The central feature of the therapeutic hypothesis is predicting a dose–response relationship between target perturbation and efficacy (or toxicity) in humans (FIG. 2a). Therefore, we argue that a primary goal of any

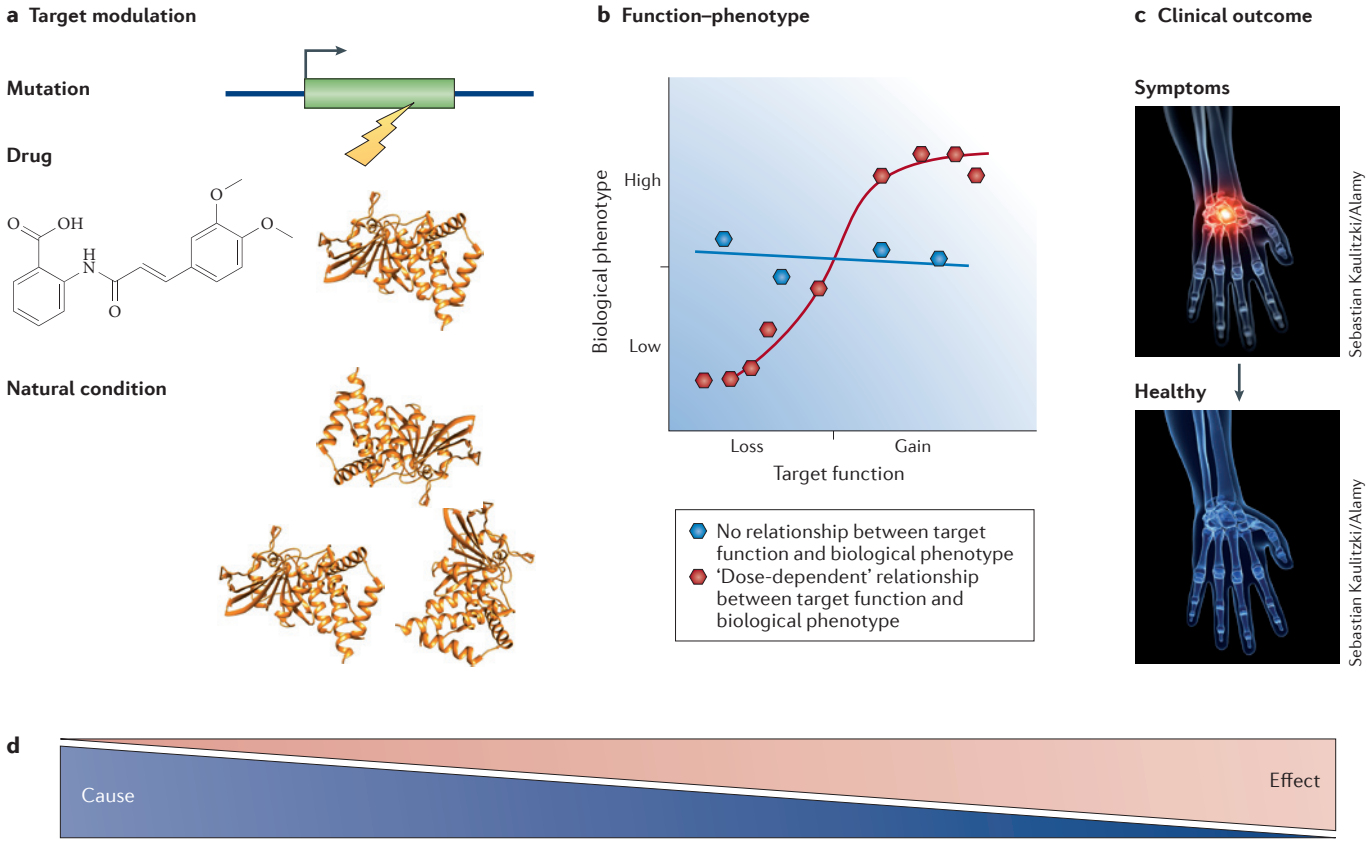

Figure 1 | **The therapeutic hypothesis.** **a** | There are three different ways to modulate a target: human mutations can increase or decrease the function of a gene through gain-of-function or loss-of-function alleles; drugs can pharmacologically increase or decrease target function; and naturally occurring conditions may increase or decrease the amount of a target, thereby increasing or decreasing its function. **b** | By modulating the function of a target (*x* axis), it is possible to assess its effect on a biological phenotype (*y* axis) such as cellular signalling or receptor levels. The red points on the graph indicate a dose-dependent relationship between target function and biological phenotype, as loss of function of a target leads to reduced (low) biological activity (phenotype), whereas gain of function leads to increased (high) biological activity. By contrast, the blue points indicate that modulating target function has no effect on biological phenotype or activity. **c** | Target modulation can be correlated with clinical outcomes in patients to assess for efficacy and toxicity. For example, if increased target function (represented by the red points on the graph in panel **b**) is associated with clinical symptoms, it follows that decreased target function should be an effective treatment to restore health. Ideally, the results of target modulation would be monitored in many patients for many years, leading to the accumulation of all possible clinical outcomes. **d** | Early events are more likely to be causal than events that are observed only after the onset of disease symptoms and sequelae. If genetic mutations, drug perturbations and natural conditions precede clinical outcome, then it is possible that a 'cause and effect' relationship can be established. By contrast, observations that are only made in individuals with a disease (for example, through *in vivo* expression or epidemiology studies) may be the cause or the effect of the underlying disease process.

Preclinical models
Any of a broad range of approaches to support the therapeutic hypothesis before a drug is tested in a clinical trial.

Therapeutic hypothesis
The hypothesis that perturbing a target in a given manner leads to patient benefit (efficacy with minimal toxicity).

Target validation
The process of gathering information about a potential drug target prior to initiating a screen to find biological or chemical modulators of the target of interest.

First-in-class drug
A drug that is the first to target a new biological mechanism of action.

Alleles
DNA sequence variations between two chromosomes (for example, one maternal chromosome and one paternal chromosome).

preclinical model should be to generate sufficient data to mimic a dose–response curve as early as possible in drug development.

Such complete dose–response data are generally only known for drugs with molecular structures or mechanisms of action that are very similar to approved drugs (often dubbed 'me too' drugs). Because a similar approved drug is known to be safe and effective, there is very strong support for the therapeutic hypothesis for 'me too' drugs (which may be the result of parallel competition between companies or follow-on products developed after a first-in-class drug has made it to market)[5]. Of course, adopting a 'follow-on' strategy will not lead to the development of new molecular entities that act on novel biological targets.

Fortunately, there are alternative data sources to identify novel drug targets[6], each within a hierarchy of evidence that approaches the ideal circumstance of a target that is already validated by a therapeutic. Such data may be derived from cellular or animal model systems, human epidemiology (for example, cholesterol in heart disease), *in vivo* expression studies in disease tissues (for example, inflammatory cytokines in autoimmune disease), natural conditions that alter human physiology (for example, using thyroid replacement to treat patients with hypothyroidism) or human genetics (for example, alleles that raise or lower low-density lipoprotein (LDL) cholesterol levels influence the risk of heart disease).

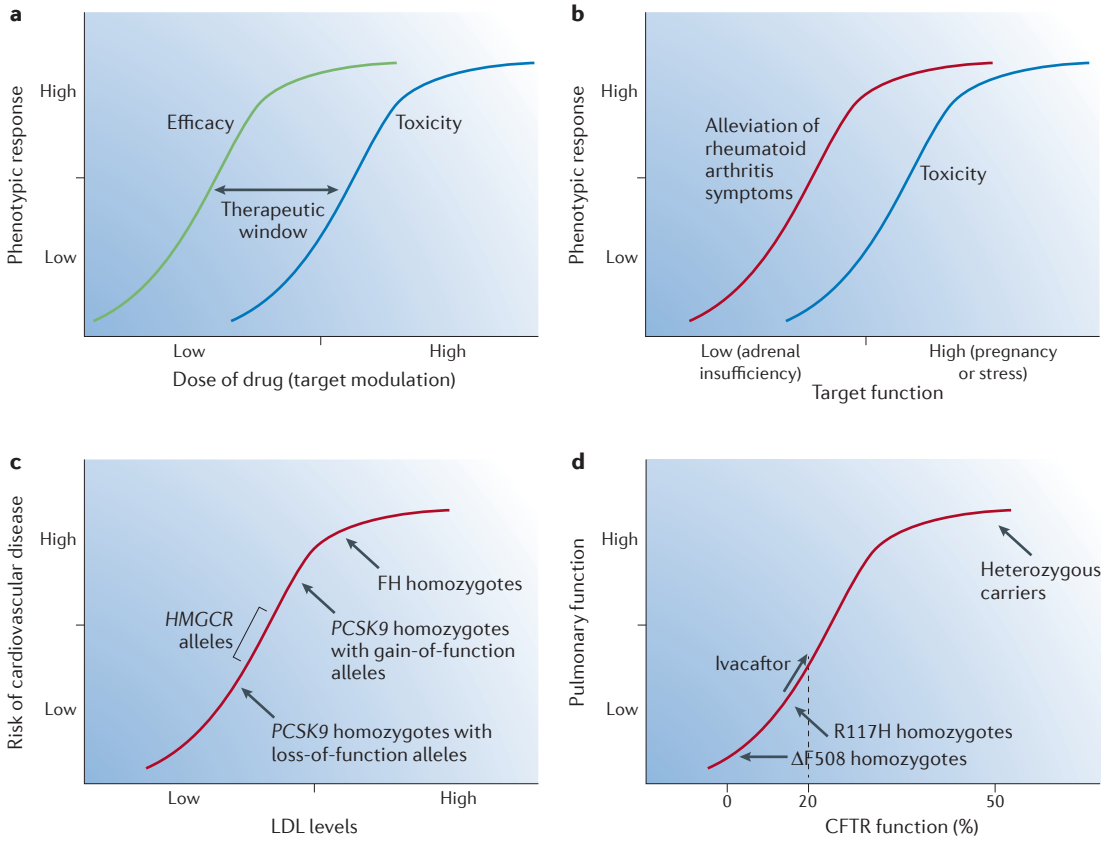

Figure 2 | **Dose–response curves derived from experiments of nature.** **a** | A basic dose–response curve is shown, in which the *x* axis represents the dose of a drug required to modulate a target, and the *y* axis represents the phenotype that is related to target modulation. **b** | Steroids and rheumatoid arthritis. Naturally occurring conditions such as pregnancy or stress increase the amount of endogenous corticosteroids, whereas other conditions such as adrenal insufficiency decrease the amount of endogenous corticosteroids. These natural conditions influence disease activity in patients with rheumatoid arthritis (disease activity represents efficacy; a high phenotypic response corresponds to low disease activity and few rheumatoid arthritis symptoms). They also provide an estimate of potential side effects, which lead to toxicity (for example, steroid-induced elevated blood glucose levels). For simplicity, adverse events associated with low cortisol levels are not shown. **c** | Low-density lipoprotein (LDL) levels and cardiovascular disease. Variants in different genes can lead to variations in the levels of LDL cholesterol, which can have a predictable effect on the risk of cardiovascular disease. Rare loss-of-function mutations in the LDL receptor (*LDLR*) gene lead to familial hypercholesterolaemia (FH) in homozygotes; gain-of-function mutations in the proprotein convertase subtilisin kexin 9 (*PCSK9*) gene increase LDL levels and the risk of cardiovascular disease, whereas *PCSK9* loss-of-function mutations have the opposite effect. Furthermore, a common DNA variant in the HMG-CoA reductase (*HMGCR*) gene, as well common variants in other gene loci discovered through genome-wide association studies (GWASs), have shown that there is a small but statistically robust association between LDL levels and the risk of cardiovascular disease. **d** | Cystic fibrosis transmembrane conductance regulator (*CFTR*) mutations and cystic fibrosis. A series of causal alleles that alter the function of the CFTR protein demonstrate a dose–response relationship. A drug, ivacaftor, can increase the function of the CFTR protein in patients with a specific genotype, thereby improving clinical symptoms.

Experiments of nature at the top of the hierarchy

'Experiments of nature', which represent naturally occurring human conditions or states that modulate a biological target with a reproducible effect on human physiology, occupy a prominent position in the hierarchy of evidence to support the therapeutic hypothesis. In the context of drug discovery, these natural experiments mimic the effect of therapeutically modulating the target and provide a mechanism to estimate dose–response curves before a clinical trial is initiated. In essence, they are nature's equivalent of clinical trials with an established therapeutic.

'Experiments of nature'
Naturally occurring human conditions or states that modulate a biological target with a reproducible effect on human physiology; in the context of drug discovery, these experiments mimic the effect of therapeutic modulation of the target.

This concept is well illustrated by the historical example of human conditions that alter the amount of cortisol, which is a naturally occurring steroid secreted by the adrenal gland that is under the control of the hypothalamic–pituitary axis in the brain. Today, steroid derivatives (for example, hydrocortisone) are routinely used as anti-inflammatory drugs for several clinical conditions, including the autoimmune disease rheumatoid arthritis.

In the 1930s, however, the hormones secreted by the adrenal cortex were unknown, and the effect of these hormones on human physiology and disease was also

unknown. A confluence of events at the Mayo Clinic, led by Dr Phillip Hench (a rheumatologist) and Dr Edward Kendall (a chemist studying hormones secreted by the adrenal gland), resulted in a series of studies culminating in a Nobel Prize[7]. Hench observed that the symptoms of patients with rheumatoid arthritis improved during pregnancy and following temporary stress brought upon by surgery — both clinical conditions in which levels of endogenous steroid hormones were known to be elevated. Hench was also aware of the clinical features shared by patients with active rheumatoid arthritis and those with Addison's disease, a form of adrenal insufficiency in which levels of endogenous steroids were known to be decreased. Finally, both Hench and Kendall were aware of the reported anti-inflammatory activity of corticosteroids in animal models. Together, they developed a therapeutic hypothesis that cortisol would suppress the clinical symptoms of rheumatoid arthritis. On 21 September 1948, Hench teamed up with Kendall to perform the first administration of cortisone, a metabolite of cortisol, to patients with rheumatoid arthritis. They observed an immediate and substantial improvement in symptoms, referring to cortisol as "Nature's dramatic antidote"[7].

In this example, there were several features that enabled an estimate of the dose–response curve for the efficacy and safety of corticosteroids in patients with rheumatoid arthritis (FIG. 2b). Naturally occurring conditions resulted in higher levels (for example, in pregnancy and stress) or lower levels (for example, in adrenal insufficiency) of endogenous steroids in patients with rheumatoid arthritis, thereby providing an estimate of the effects of modulating the 'target' (in this case, cortisol itself) on the symptoms of patients with rheumatoid arthritis. Furthermore, these conditions provided an estimate of the adverse events associated with excess steroids (for example, diabetes, weight gain, hypertension and osteoporosis). The clinical conditions represented perturbations in humans, thereby providing a direct link with human disease (rheumatoid arthritis). And the perturbations occurred in a temporal sequence, which helped to differentiate between cause and consequence.

There are other examples of experiments of nature that led to drug discovery: the development of HMG-CoA reductase inhibitors (statins) is a noteworthy success story[8]. In the 1950s, a biological link between cholesterol and heart disease was established, following epidemiological studies examining the relationship between blood cholesterol (and other potential risk factors) and death from coronary disease. Rare families with familial hypercholesterolaemia provided further support for a causal link between LDL cholesterol and heart disease. These patients have mutations in the LDL receptor (*LDLR*) gene, leading to high levels of LDL cholesterol and an increased risk of heart disease[9,10]. Furthermore, a dose–response relationship was observed between function (the number and type of *LDLR* mutations) and phenotype (LDL cholesterol levels and risk of heart disease), as shown in FIG. 2c. Individuals with two mutated *LDLR* alleles (familial hypercholesterolaemia homozygotes) are more severely affected than those with one mutant allele (familial hypercholesterolaemia heterozygotes), and familial hypercholesterolaemia homozygotes with a null allele (no LDLR activity) are more severely affected than familial hypercholesterolaemia homozygotes with a defective allele (these individuals have LDLR activity, but it is reduced relative to wild-type individuals).

As HMG-CoA reductase was known to be the rate-limiting enzyme in the cholesterol biosynthetic pathway, it represented a compelling drug target. Natural products found in the fermentation broth of *Penicillium citrinum* (compactin) and *Aspergillus terreus* (lovastatin) inhibited HMG-CoA reductase activity and lowered levels of LDL cholesterol in animal models. Clinical trials that were initially carried out in selected small groups of patients with severe heterozygous familial hypercholesterolaemia, and then in the general population or in patients at a very high risk of myocardial infarction, demonstrated the safety and efficacy of lovastatin[11]. Ultimately, treatment with statins proved the correlation between LDL levels and an increased risk of heart disease.

An emerging story that further supports the therapeutic hypothesis for LDL cholesterol levels and the risk of heart disease relates to proprotein convertase subtilisin kexin 9 (PCSK9). In 2003, two families with autosomal dominant high LDL levels and an increased incidence of coronary heart disease were found to have rare gain-of-function mutations in the *PCSK9* gene[12]. Subsequent candidate gene association studies revealed that *PCSK9* loss-of-function mutations observed at a low frequency in the general population (~1%) correlated with reduced levels of LDL cholesterol and a reduced incidence of coronary heart disease[13–15]. Animal models revealed that PCSK9 is involved in the post-translational regulation of LDLR activity, thereby providing a mechanistic link between PCSK9 and LDL cholesterol levels[16,17]. Then, in 2012, randomized control trials were published that demonstrated that PCSK9-specific monoclonal antibodies significantly reduced LDL cholesterol levels in healthy volunteers as well as in individuals with hypercholesterolaemia[18–20].

Even genetic variants with a subtle effect on LDL cholesterol and myocardial infarction can point to successful targets for cardiac prevention. For example, a common, non-coding genetic polymorphism (rs3846663) in the gene that encodes HMG-CoA reductase (*HMGCR*) has a small influence on LDL cholesterol levels and on the risk of cardiovascular disease in the general population[21]. Furthermore, an aggregate genetic risk score, which is the sum total of the effect of all alleles that influence LDL cholesterol levels, directly correlates with the risk of cardiovascular disease in the general population (FIG. 2c). This is in contrast to individual alleles or a genetic risk score for HDL cholesterol, for which there is no obvious correlation with the risk of cardiovascular disease, as described in more detail below.

Thus, as with rheumatoid arthritis and cortisol, the example of LDL cholesterol and heart disease represents an experiment of nature (FIG. 2c), where naturally occurring conditions (genetic variations in the *LDLR*, *PCSK9* and *HMGCR* genes) modulate a target in a dose-dependent manner in humans, thereby providing

Inherited DNA variation
A variation in DNA sequence that is passed from the parent to the offspring according to the rules of Mendelian segregation.

Causal alleles
DNA variants that are responsible for influencing a clinical phenotype.

Complex traits
Diseases that do not segregate within families according to obvious rules; the underlying genetic cause is often highly polygenic and substantially influenced by environmental and stochastic factors.

a causal link between function and phenotype in a temporal sequence that precedes the clinical outcome of interest (such as heart disease).

Incomplete supporting packages

The examples of cortisol and LDL cholesterol represent relatively complete packages that relied not only on naturally occurring conditions in humans but also on strong supporting evidence from biology, epidemiology and animal models. Even with such strong supporting evidence, the development of steroids and statins was not without uncertainty and risk. However, packages to support novel therapeutic hypotheses can often be substantially less complete.

TABLE 1 lists various preclinical models for target validation[6]. In general, each model on its own is insufficient to support the therapeutic hypothesis, as each one has limitations for providing evidence to support or refute a therapeutic hypothesis for a given drug target. These limitations relate to four characteristics: target modulation (the ability to modulate a target of interest to achieve a desired effect on a biological pathway); human relevance (the ability to demonstrate the relevance of a target to a human disease process); causality in humans (the ability to determine whether a target perturbation is a cause or consequence of a human disease process); and mechanism of action (the ability to understand the relationship between the biological mechanism of the underlying model and the human disease state).

A target that emerges from an animal model has the great advantage of being tractable. Controlled experiments can establish a dose–response relationship between function and phenotype. That is, a target can be modulated through genetics or pharmacology, and the animal model can be studied to determine how a biological process is altered. However, the major limitation of an animal model is determining the relevance of the target to human disease. In addition, animal models cannot establish whether target modulation is a cause or a consequence of human disease.

Human epidemiology is highly relevant to human disease, but on its own it cannot be used to prove causality. One example is the relationship between high-density lipoprotein (HDL) cholesterol and heart disease[22]. Epidemiological studies suggested that pharmacological manipulation to raise HDL levels would lower the risk of myocardial infarction. Based on this theory, drugs that inhibit cholesteryl ester transfer protein (CETP), which promotes the transfer of cholesterol from HDL to LDL, thereby raising HDL levels, should protect against heart disease[23]. However, the clinical trial data on CETP inhibitors do not yet support the epidemiological data[24]. Furthermore, a missense N396S mutation in the endothelial lipase (*LIPG*) gene raises HDL cholesterol levels but does not lower the risk of myocardial infarction[25]. It remains to be determined whether other CETP inhibitors have a different efficacy profile or whether drugs that raise HDL levels through other mechanisms will lower the risk of myocardial infarction.

The main advantages of human genetics for validating therapeutic targets are that human genetics is highly relevant to human disease and can differentiate between cause and consequence. However, there are also several limitations. First, human genetics relies on DNA mutations and human evolution for the introduction of inherited DNA variation (alleles) into a gene target, and consequently not all gene targets will have disease-causing alleles. Once identified, causal alleles represent a natural perturbation of a potential therapeutic target; see BOX 1 for approaches to establish a causal link between a target and a clinical phenotype for Mendelian and complex traits. Furthermore, those genes that do harbour causal alleles might not have multiple alleles to allow the establishment of a genotype–phenotype dose–response curve in the same way as for LDL cholesterol levels (FIG. 2c).

Second, although human genetics provides a link between a natural perturbation and a physiological process of interest, it can be quite challenging to understand the mechanistic implications of the causal allele. Similarly, although human genetics can differentiate cause from

Table 1 | **Characteristics of preclinical models for target validation***

	Target modulation	Human relevance	Causality in humans	Mechanism of action
Cellular models	Highly effective	Ineffective	Ineffective	Effective, but with some limitations
Animal models	Highly effective	Effective, but with some limitations	Ineffective	Highly effective
Human epidemiology	Effective, but with some limitations	Highly effective	Ineffective	Effective, but with some limitations
***In vivo* expression studies**	Effective, but with some limitations	Highly effective	Ineffective	Effective, but with some limitations
Natural conditions	Effective, but with some limitations	Highly effective	Highly effective	Effective, but with some limitations
Human genetics	Effective, but with some limitations	Highly effective	Effective, but with some limitations	Effective, but with some limitations

*Target modulation is the ability to modulate a target of interest to achieve a desired effect on a biological pathway; human relevance is the ability to demonstrate the relevance of a target to a human disease process; causality in humans refers to the ability to determine whether a target perturbation is a cause or consequence of a human disease process; and the mechanism of action is the ability to understand the relationship between the biological mechanism of the underlying model and the human disease state.

Box 1 | **Genetic architecture of Mendelian and complex diseases**

Genetic architecture refers to the number, effect size and population frequency of causal alleles. Here, we compare and contrast the genetic architecture of Mendelian diseases and complex traits, and briefly describe statistical approaches to identify causal alleles and causal genes. We also describe how causal alleles from both disease categories provide information on target modulation.

Mendelian diseases segregate faithfully within a family according to Mendel's laws. For a given family, the underlying genetic architecture is generally a single mutation (that is, the causal allele) in one gene that is rare in the general population and highly penetrant in family members who inherit the mutation. Often, the causal mutation disrupts the protein-coding structure of a gene, thereby pinpointing the causal gene. Examples of Mendelian diseases include cystic fibrosis and Marfan's syndrome. The cystic fibrosis gene, cystic fibrosis transmembrane conductance regulator (*CFTR*)[27], was identified in 1989 and the Marfan's syndrome gene, fibrillin 1 (*FBN1*)[102], was identified in 1991.

By contrast, complex diseases do not segregate within families according to Mendel's rules. Examples include rheumatoid arthritis, type 2 diabetes and myocardial infarction. In a population of affected individuals, the underlying genetic architecture for a given disease is often highly polygenic and substantially influenced by environmental and stochastic factors. Advances in genomic technology have facilitated the identification of loci for complex traits; these advances include a draft sequence of the human genome, a catalogue of common DNA polymorphisms[103], high-throughput methods to genotype hundreds of thousands of single-nucleotide polymorphisms (SNPs) and statistical methods to analyse extremely large data sets[104]. These advances led to the first generation of genome-wide association studies (GWASs), which identified alleles that are associated with a variety of complex traits[104]. To date, GWASs and related methods have identified nearly 3,000 loci for approximately 300 complex human traits, as reported in the US National Human Genome Research Institute (NHGRI) GWAS catalogue[105] (see the 'Catalogue of Published Genome-Wide Association Studies' for further information).

Several themes have emerged from GWASs that shed light on the genetic architecture of complex traits: hundreds (if not thousands) of alleles contribute to the risk of developing any given complex disease[101,106]; each allele has a small effect on risk; and most alleles discovered to date are common in the general population (but this is a biased estimate, as only common alleles have been tested by contemporary GWASs).

In contrast to Mendelian diseases, it is more challenging to identify causal mutations and genes in complex disease. This is due to a number of factors: the alleles associated with the risk of a complex disease are often outside the coding regions; there are often many SNPs that are highly correlated with the top SNP (known as linkage disequilibrium); there is no obvious causal allele that can be identified from the SNPs that are in linkage disequilibrium with each other; and there are often many genes in the region (or genetic locus). A few themes have emerged, however. For example, the majority of causal alleles associated with complex traits are likely to influence gene expression rather than protein sequence[42,107]; occasionally one allele is an obvious functional allele (for example, one that changes the protein-coding structure of a gene), which helps to pinpoint the causal allele and causal gene; by comparing genes across multiple risk loci for a given disease, it is often possible to select the most likely causal gene[108,109]; and some loci may contain independent variants that are associated with disease, providing an allelic series that helps to identify the causal gene and enables the exploration of disease biology[46,110].

For target validation, complete loss-of-function mutations (usually observed in Mendelian diseases) provide different information compared with common alleles that have modest effects (observed in complex traits). If a gene is completely knocked out (a homozygous loss-of-function mutation), this provides the maximal phenotypic effect on target modulation. By contrast, alleles with a subtle effect on function indicate that modulation of the target influences clinical outcome; however, these alleles do not easily provide a broad range of biological or clinical effects on target modulation. In an ideal situation, a gene would harbour a series of causal alleles with a broad range of biological effects (from gain-of-function alleles to loss-of-function alleles) to generate function–phenotype dose–response curves.

Genetic architecture
The underlying genetic basis for a phenotypic trait; variables include: the number of causal genes (monogenic, oligogenic or polygenic); the population frequency of causal alleles (common, low-frequency or rare); and the effect size of the causal alleles (small effect reflecting low penetrance, or large effect reflecting high penetrance).

Genetic locus
A location or region of the genome; the boundaries of a locus can be defined by linkage disequilibrium blocks or other factors.

Functional alleles
Alleles to which a biological function can be ascribed; examples include differential gene expression or mRNA splicing, or differences in protein-coding sequence.

consequence because alleles are present from birth and thus before the onset of human disease, functional studies are required to understand the biological mechanisms involved. Last, human genetics might link a target perturbation to a disease trait, but the factors that lead to the disease might differ considerably from the factors that need to be modulated in order to treat the disease.

Building a complete package

In setting out to test the therapeutic hypothesis, a practical consideration is how to build a complete package that is based on preclinical models, each of which has its own limitations. We argue that it is better to first anchor target validation to a preclinical model that has relevance to human disease and can be used to differentiate between cause and consequence, and only then to try and understand the effect of target modulation and the biological mechanism of action. That is, we believe that there is great value in anchoring target validation to 'experiments of nature' such as naturally occurring conditions or human genetics. Below, we describe how to overcome the limitations of human genetics to build a complete package for testing the therapeutic hypothesis. In essence, the goal is to generate dose–response curves that are based on human genetics.

Target modulation. The underlying concept is that causal alleles represent natural perturbations of a drug target. In the ideal circumstance, a gene target would harbour a series of functional alleles that provide a range of perturbations, and these alleles would be correlated with function (see below) and clinical outcome. Some alleles would be

complete loss-of-function alleles, which — when inherited in the homozygous state — would mimic a state in which there is complete pharmacological inhibition of the target. Other alleles would be gain-of-function alleles, which would allow further examination of the relationship between function and phenotype in both the heterozygous and homozygous states. By combining all of these data, it should be possible to generate function–phenotype dose–response curves that share properties similar to those of drug dose–response curves.

A noteworthy example of function–phenotype dose–response curves comes from cystic fibrosis and mutations in the gene encoding cystic fibrosis transmembrane conductance regulator (CFTR)[26]; see FIG. 2d. Cystic fibrosis is an autosomal recessive disease that leads to pulmonary dysfunction. The causal gene, identified in 1989 through linkage analysis[27], is *CFTR*. To date, more than 1,800 independent alleles have been identified that cause cystic fibrosis[28]. Heterozygous carriers of null *CFTR* mutations, which include the most common causal allele ΔF508, are asymptomatic even though their cells only have 50% function of the CFTR protein. Homozygous carriers of loss-of-function alleles have no CFTR activity and a severe clinical phenotype. Patients who inherit *CFTR* alleles with 10–20% function have a mild cystic fibrosis phenotype, thereby indicating that restoration of CFTR function to this level should improve clinical symptoms in patients with severe disease. Indeed, ivacaftor (Kalydeco; Vertex Pharmaceuticals) — a drug that enhances CFTR function — improves clinical outcomes in patients with a specific genotype[29].

Another example of function–phenotype dose–response curves comes from rare mutations in the *SCN9A* gene, which encodes the voltage-gated sodium channel Nav1.7 (REF. 30). Gain-of-function mutations in *SCN9A* have been identified in rare families with primary erythermalgia (intermittent burning pain with redness and heat in the extremities)[31–35]. In addition, rare loss-of-function mutations in *SCN9A* have been identified in families with a congenital inability to perceive any form of pain. Based on these genetic data, drugs that block the Nav1.7 sodium channel are now under development to treat pain in the general population[36,37].

Biological mechanism. To generate function–phenotype dose–response curves, the biological effect of causal alleles on gene function must be experimentally determined. In particular, it is important to know whether causal alleles result in a gain of function or a loss of function, as this will help guide whether a therapy should inhibit or activate the target. In some instances, it may be easy to predict the biological function based on the mutations and phenotypes themselves. This is particularly true for mutations that dramatically change the protein-coding structure of a gene. For example, deletions and nonsense mutations in the Janus kinase 3 (*JAK3*) gene cause an autosomal recessive form of severe combined immunodeficiency (SCID)[38]. This observation was useful in the development of drugs to treat rheumatoid arthritis, in which JAK3 inhibition by the drug tofacitinib (Xeljanz; Pfizer) is effective in treating symptoms related to systemic inflammation[39,40].

In other instances, the functional consequences of causal alleles are less obvious. Functional studies in mice and humans demonstrated that for Marfan's syndrome the causal mutations in the gene fibrillin 1 (*FBN1*) result in loss of function of the fibrillin 1 protein. However, these mutations result in enhanced transforming growth factor-β (TGFβ) activation and signalling at the cellular level — a mechanism that was not previously appreciated in the pathophysiology of this disease[41].

Unravelling the biological mechanism for alleles that influence the risk of complex diseases, most of which have been identified by genome-wide association studies (GWASs), is especially challenging (BOX 1). Based on current knowledge, causal alleles that are responsible for most complex traits fall outside of protein-coding sequences[42]. For example, in a recent study of inflammatory bowel disease (IBD), 29 IBD-associated single nucleotide polymorphisms (SNPs) — out of a total of 193 SNPs from 163 loci — were in strong linkage disequilibrium with a protein-coding missense variant[43]. By contrast, 64 IBD-associated SNPs (33%) are in linkage disequilibrium with variants that are known to regulate gene expression. If a risk allele increases the expression of a gene that is a positive regulator of a pathway, then it follows that an effective drug might inhibit that particular gene or signalling pathway; this has been predicted for a non-coding variant in the *CD40* gene that increases the risk of rheumatoid arthritis[44,45,111]. For some GWAS loci that have been implicated by GWASs for influencing complex traits, independent and rare protein-coding variants can pinpoint the causal gene and provide further insight into its biological function, as observed for the caspase recruitment domain-containing protein 9 (*CARD9*) gene in IBD[46].

Biological pathways. If the indication for treatment is reduction of active disease (rather than prevention), and if human genetics is used to identify and validate targets, then it must be the case that the biological pathways that lead to disease are also relevant to symptoms in established disease. Two illustrative examples are the autoimmune diseases type 1 diabetes and rheumatoid arthritis. In type 1 diabetes the immune system destroys the pancreas, thereby preventing insulin secretion and the control of blood glucose levels. Once diagnosed, the primary treatment for type 1 diabetes is the administration of insulin to maintain glucose homeostasis. Human genetics has identified many alleles associated with the risk of type 1 diabetes, nearly all of which act on the immune system[47]. Thus, drugs that are developed based on the genetics of type 1 diabetes might be expected to prevent disease in susceptible individuals but not to treat the disease once the pancreas has been destroyed.

By contrast, in patients with rheumatoid arthritis the immunological pathways that lead to the disease also seem to be related to the immunological pathways that contribute to symptoms in patients with established disease. As direct proof of concept, several genes that are implicated in the pathogenesis of rheumatoid arthritis are the targets of drugs that are effective therapies for this disease; for example, cytotoxic T lymphocyte

Function–phenotype dose–response curves
An assessment of the effect of modulating the function of a target on a biological phenotype in a way that mirrors the traditional dose–response curves of drug efficacy and toxicity from clinical trials.

Causal gene
A gene that, when perturbed by a mutation, leads to a clinical phenotype.

Genome-wide association studies
(GWASs). Comprehensive testing of genetic variants in a collection of individuals to see whether any variant is associated with a trait; contemporary GWASs are limited to testing common variants, although newer technologies allow the testing of low-frequency variants.

Single nucleotide polymorphisms
(SNPs). DNA sequence variations that occur when a single nucleotide — A, T, C or G — differs between paired chromosomes.

Linkage disequilibrium
A non-random correlation of alleles at a locus (or region) of the genome, such that some combinations of alleles in a population are observed more frequently than would be expected by chance; the extent of linkage disequilibrium can be measured by the square of the correlation coefficient (r^2); non-random recombination across the genome during the course of human history results in blocks of linkage disequilibrium (often containing multiple genes).

antigen 4 (*CTLA4*) is targeted by abatacept (Orencia; Bristol-Myers Squibb)[48] and interleukin-6 receptor (*IL6R*) is targeted by tocilizumab (Actemra; Roche)[49].

Thus, to build a complete package that is based on human genetics, it is important to identify a series of causal alleles in a gene target of interest (known as target modulation) and to understand the functional consequences of causal alleles (that is, the biological mechanism) in order to generate function–phenotype dose–response curves. Moreover, there must be a connection between the disease state used in the genetic study and the disease state for the drug indication.

Historical support for genetics in target validation

The discussion above implies that identifying alleles that contribute to the risk of a disease or related medical traits (for example, LDL cholesterol, inflammation or pain) can be a productive strategy for identifying relevant drug targets for such diseases. An obvious question is whether there is historical precedence to support this view. Below, we provide examples of gene–drug pairs where a single gene is implicated by human genetics, and a drug directed against that gene is an effective therapeutic target. A more complete list of gene–drug pairs[50–55] is shown in TABLE 2.

It is useful to consider three categories of gene–drug pairs: drugs that are in development or have been approved for which human genetics had a major role in their development (referred to as prospective examples); approved drugs that were developed without strong human genetics data, but for which human genetics subsequently identified the drug target as being important (referred to as retrospective examples); and drugs that were developed for a particular indication, but human genetics data suggested another indication (referred to as repurposing examples).

In addition to the examples of *LDLR* (for which >1,000 pathogenic mutations have been reported)[56] and *PCSK9* discussed above, another prospective example is the development of 5-alpha-reductase inhibitors. Rare families with pseudohermaphroditism have mutations in the steroid-5-alpha-reductase α-polypeptide 2 (*SRD5A2*) gene, which leads to a deficiency of the male hormone dihydrotestosterone[57,58]. The finding that male patients with *SRD5A2* mutations have small prostates and lack male pattern baldness led to the development of 5-alpha-reductase inhibitors (for example, finasteride) for the treatment of benign prostatic hyperplasia and mild to moderate hair loss[57,59].

There are several examples of approved drugs that were developed without direct human genetics data, but for which human genetics subsequently identified the drug target as being important. A recent study systematically examined the US National Human Genome Research Institute (NHGRI) GWAS catalogue for links between gene–drug pairs[60]. Examples of gene–drug pairs (and their respective diseases) from this and other studies include: *HMGCR*–statins (for the treatment of hyperlipidaemia)[21,61]; peroxisome proliferator-activated receptor-γ (*PPARG*)–thiazolidinediones (for the treatment of type 2 diabetes)[62]; *CTLA4*–abatacept (for the treatment of rheumatoid arthritis)[48]; *IL12B*–ustekinumab (for the treatment of psoriasis and Crohn's disease)[43,63]; and receptor activator of NF-κB ligand (*RANKL*; also known as *TNFSF11*)–denosumab (for the treatment of osteoporosis)[64].

There are also examples of the third category: drugs that were developed for a particular indication but have been 'repurposed' for another indication. For Marfan's syndrome, mechanistic studies of *FBN1* were integrated with data demonstrating that angiotensin II receptor blockers decreased TGFβ signalling, which allowed these drugs to be repurposed from an existing indication (hypertension) to improve outcomes for patients with Marfan's syndrome who have aortic root dilation[65].

Another repurposing example is that of complement inhibitors for the treatment of age-related macular degeneration (AMD). Before 2005, the complement pathway had not been widely implicated in the pathogenesis of AMD. One of the first GWASs in any complex trait identified a common, missense mutation (Y402H) in the complement factor H (*CFH*) gene as an indicator of an increased risk of AMD[66]. Subsequent genetic studies confirmed the role of the complement pathway in AMD, including the discovery that multiple independent alleles in *CFH* influence the risk of AMD[67–69]. As complement inhibitors had been developed for the treatment of other diseases (for example, sepsis and paroxysmal nocturnal haemoglobinuria)[70], they have since been repurposed for the treatment of AMD, and several clinical trials are underway in this setting[71]. Other complement inhibitors are also under development for the treatment of AMD (for example, inhibitors of complement factor D and of complement factor C3)[72], which indicates the overlap between developing new compounds and repurposing existing compounds. Other repurposing examples[73–82] are shown in TABLE 2.

Criteria for gene–drug pairs in target validation

Based on a conceptual framework for the role of preclinical models in target validation (FIG. 1; TABLE 1) and historical examples of gene–drug pairs (TABLE 2), we propose a set of criteria for the application of genetic findings to target validation (BOX 2). The criteria are agnostic to frequency, penetrance or the effect size of the associated alleles. That is, these criteria can be applied to genetic discoveries made from Mendelian diseases as well as complex traits. The goal is to apply these criteria, which have been ordered by importance below, to help prioritize research on the most promising targets and ultimately nominate a gene product as the target for a drug development programme.

The gene harbours a causal variant that is unequivocally associated with a medical trait of interest. It is crucial that the genetic finding is robust. We do not provide strict guidelines for statistical significance, as these issues have been discussed exclusively elsewhere in the literature[83–86]. The bottom line is that one must be convinced, beyond any doubt, that the genetic variant influences the trait of interest. Consistent replication of the genetic finding is one of the most important measures of significance. Furthermore, the variant must be the causal allele (that is, not a proxy or marker SNP). This criterion

Mendelian diseases
Diseases that segregate faithfully within a family according to Mendel's laws; for a given family, the underlying genetic cause is generally a single mutation that is rare in the general population and highly penetrant in family members who inherit the mutation.

Table 2 | **Gene–drug pairs**

Gene	Allele (or alleles)	Drugs	Disease or indication	Genetic approach	Comments	Refs
Prospective examples						
LDLR	Many	Statins	Hyperlipidaemia	Biochemical	*LDLR* mutations indicated that the LDL cholesterol pathway is critical in the risk of heart disease	9,10
SRD5A2	Many	Finasteride	Benign prostate hyperplasia	Biochemical	Rare *SRD5A2* mutations lead to pseudohermaphroditism	57–59
PCSK9	Many	Compounds in clinical trials	Hyperlipidaemia	Linkage and family-based sequencing; candidate gene sequencing	S127R and F216L were the first gain-of-function mutations; Y142X and C679X were the first nonsense mutations	12–15
SCN9A	Many	Compounds in development	Pain	Linkage and family-based sequencing	Loss-of-function nonsense mutations include S459X, I767X and W897X	30–32
BCL11A	rs4671393	Compounds in clinical trials	Sickle cell anaemia	GWAS	Non-coding allele; *BCL11A* repressors increase fetal haemoglobin levels in sickle cell anaemia	50–52
CFTR	Many	Ivacaftor; compounds in clinical trials	Cystic fibrosis	Linkage and family-based sequencing	The first mutation identified was ΔF508; the CFTR potentiator ivacaftor was developed for a specific genotype (G551D)	27,28
LMNA	Many	Compounds in clinical trials	Hutchinson–Gilford progeria syndrome (HGPS)	Linkage and family-based sequencing	Mutations in *LMNA* cause a broad range of human diseases, including the premature aging seen in HGPS; the most common mutation is a point mutation in exon 11 that does not alter an amino acid (G608G)	53–55
Retrospective examples						
HMGCR	rs3846663	Statins	Hyperlipidaemia	GWAS	A non-coding allele discovered by GWASs may affect the alternative splicing of exon 13	21,61
PPARG	rs1801282	Thiazolidinediones	Type 2 diabetes	Candidate gene study	The more common allele encodes the amino acid proline and contributes to the risk of diabetes	62
CTLA4	rs3087243	Abatacept	Rheumatoid arthritis	Candidate gene study	A non-coding allele may alter the expression of the ratio of soluble to full-length CTLA4 isoforms	48
IL12B	rs12188300	Ustekinumab	Psoriasis	GWAS	Non-coding allele; a different allele (rs6871626) is associated with Crohn's disease	43,63
RANKL	rs9533090	Denosumab	Osteoporosis	GWAS	Also known as *TNFSF11*; a non-coding allele has been discovered by GWASs	64
Repurposing examples						
CFH	Several	Eculizumab	AMD	GWAS	Missense mutations include Y402H and A69S; complement inhibitors are under investigation for AMD	66–69
IL6R	D358A	Tocilizumab	Coronary artery disease	GWAS-related approach using custom bead chip	An IL-6R-targeted therapy is approved for rheumatoid arthritis and under investigation for coronary artery disease	73
IL1	Many	Anakinra	Autoinflammatory disease	Linkage and family-based sequencing	Mutations in *NLRP3*, *TNFR1*, *IL1RN* and *MEFV* lead to elevated IL-1 levels	74
FBN1	Many	Angiotensin II receptor blockers	Marfan's syndrome	Linkage and family-based sequencing	*FBN1* mutations lead to elevated TGFβ levels, and angiotensin II receptor blockers inhibit TGFβ signalling	79,102
SMN1	Many	Riluzole	Spinal muscular atrophy	Linkage and family-based sequencing	The first mutations were gene deletions; based on phenotypic screening, riluzole is in clinical trials for the treatment of spinal muscular atrophy	80–82

AMD, age-related macular degeneration; *BCL11A*, B cell lymphoma 11A; *CFH*, complement factor H; *CFTR*, cystic fibrosis transmembrane conductance regulator; *CTLA4*, cytotoxic T lymphocyte antigen 4; *FBN1*, fibrillin 1; GWAS, genome-wide association study; *IL1*, interleukin-1; *IL6R*, IL-6 receptor; *LDLR*, low-density lipoprotein receptor; *LMNA*, lamin A/C; *PCSK9*, proprotein convertase subtilisin kexin 9; PNH, paroxysmal nocturnal haemoglobinuria; *PPARG*, peroxisome proliferator-activated receptor-γ; *RANKL*, receptor activator of NF-κB ligand (also known as *TNFSF11*); *SCN9A*, voltage-gated sodium channel Nav1.7; *SMN1*, survival of motor neuron 1; *SRD5A2*, steroid-5-α-reductase α-polypeptide 2.

Box 2 | **Criteria for gene–drug pairs in drug discovery**

- The gene harbours a causal variant that is unequivocally associated with a medical trait of interest
- The biological function of the causal gene and causal variant are known
- The gene harbours multiple causal variants of known biological function, thereby enabling the generation of genotype–phenotype dose–response curves
- The gene harbours a loss-of-function allele that protects against disease, or a gain-of-function allele that increases the risk of disease
- The genetic trait is related to the clinical indication targeted for treatment
- The causal variant is associated with an intermediate phenotype that can be used as a biomarker
- The gene target is druggable
- The causal variant is not associated with other adverse event phenotypes
- Corroborating biological data support genetic findings

is especially important for variants that have been discovered by GWASs, as the associated SNP is likely to be a proxy for the true causal allele owing to patterns of linkage disequilibrium.

The biological function of the causal gene and causal variant are known. It is important to know the biological effect of the associated variant, especially whether the variant results in a gain or loss of function. Studies in human tissues are invaluable for understanding the effects of individual alleles, and animal models can be very helpful in understanding the function of the gene itself.

The gene harbours multiple causal variants of known biological function. The observation that multiple alleles of the gene influence the trait, or a related trait, provides evidence for genotype–phenotype dose–response curves (as discussed above for *LDLR*, *PCSK9* and *CFTR*). Ideally, the causal alleles would be in the same gene (for example, in *CFTR*). Alternatively, the causal alleles might reside in different genes (for example, in *LDLR*, *PCSK9* and *HMGCR*) that converge on a common biological pathway (for example, LDL cholesterol levels). These alleles might be common or rare; coding or non-coding; gain-of-function or loss-of-function. The important point is that multiple causal alleles of known function help to calibrate the phenotypic consequences of target modulation over a range (FIG. 1). For Mendelian diseases, multiple unrelated families are required to find independent alleles; for complex traits, deep sequencing in large case–control populations — or in families with highly penetrant forms of the disease related to the complex trait — is required to find independent alleles.

The gene harbours a loss-of-function allele that protects against disease, or a gain-of-function allele that increases the risk of disease. The rationale behind this criterion is that it is easier to develop drugs that are inhibitors rather than activators of protein targets. The loss-of-function *PCSK9* variants that protect from coronary heart disease, and the gain-of-function *PCSK9* mutations that increase the risk of coronary heart disease, represent excellent examples. Moreover, if a gene is completely knocked out (as in homozygous loss-of-function mutations), this provides the maximal phenotypic effect on target modulation. Indeed, there is great interest in annotating all variants that are predicted to result in loss of function in the human genome in order to prioritize drug targets[87]. Mutations that introduce premature stop codons into genes often result in truncated proteins that have completely lost their function. Mutations that change a conserved amino acid from one polarity group to another can be predicted to be damaging by computational algorithms such as PolyPhen-2 or SIFT[88,89]. Gain-of-function mutations are more difficult to predict based on computational methods alone. For both gain-of-function and loss-of-function mutations, direct experimentation is required to demonstrate function.

The genetic trait is related to the clinical indication targeted for treatment. As described for type 1 diabetes and rheumatoid arthritis, the biological pathways that lead to disease might be different from the biological pathways that cause symptoms. Accordingly, the clinical indication for drug development must be precisely defined, and supporting evidence must link the biological pathways underlying the genetic trait to the biological pathways related to the clinical indication being targeted for treatment. As an example, a loss-of-function mutation in the amyloid precursor protein (*APP*) gene protects against Alzheimer's disease and cognitive decline[90]. If this finding is replicated, as suggested by a small follow-up study[91], it offers hope that pharmacological blockade of this gene or pathway will be an effective therapy to prevent Alzheimer's disease. Whether an APP inhibitor or drugs that act through a related mechanism (for example, β- and γ-secretase inhibitors) are effective at improving cognition in patients with established disease will be dependent on whether the biological pathways that lead to Alzheimer's disease are the same as those that cause impaired cognition in patients with established disease.

The variant is also associated with an intermediate phenotype that can be used as a biomarker. *PCSK9* serves as a good example of a variant that can also be used as a biomarker: loss-of-function alleles are associated with lower LDL cholesterol levels (and protect against coronary heart disease), whereas gain-of-function alleles are associated with higher LDL cholesterol levels (and increase the risk of coronary heart disease). As a consequence, LDL cholesterol levels can be used as a biomarker in clinical trials for the development of PCSK9 inhibitors[18,19]. For some alleles, a relevant biomarker may be developed during the course of functional studies, which can then be used during clinical trials.

The variant is within a gene that is 'druggable'. One of the challenges for human genetics is that only a subset of potential drug targets are 'druggable' using standard chemistry and assays. Thus, human genetics may uncover exciting new targets, but if these are not druggable then little is gained. However, what is considered druggable at present is likely to change in the future[92]. For example, kinases used to be considered

'undruggable' but now are druggable. New chemical approaches and assay development are needed to make it possible to pursue those targets with the strongest evidence from human biology.

The variant is not associated with other phenotypes that might be considered adverse events. An interesting aspect of human genetics that can be used to predict on-target side effects is whether the variant is associated with other phenotypes that could be considered adverse events. This serves as a form of Mendelian randomization[93,94]. If a drug inhibits the function of a gene product, then it would be useful to know whether there are any adverse clinical consequences of an allele that knocks out the function of the same gene. For example, it is possible to evaluate clinical phenotypes of complete PCSK9 inhibition in the general population from a handful of individuals who are homozygous null for *PCSK9* loss-of-function mutations. In this regard, genetic data in patients who are followed for long periods of time — such as prospective cohorts or patients with clinical data from electronic medical records — serve as a valuable resource for estimating potential adverse events.

Corroborating biological data support genetic findings. Genetic data should be integrated with other aspects of disease biology, including animal models, epidemiological studies and *in vivo* expression studies. If non-genetic data support the implicated role of the associated gene, then this substantially strengthens the relevance of the gene to disease. For instance, if the associated gene (such as *PCSK9*) has an orthologue with supporting data from animal models for a related phenotype, or if the associated gene is part of a family of genes (that is, a paralogue) for which there are validated therapeutic targets, then this strengthens its prioritization as a drug target.

From GWASs in complex diseases to drug target

Given the wealth of data emerging on the genetics of complex diseases from GWASs, how might these genetic data be used to select drug targets? Although most alleles associated with complex diseases (approximately 85%) fall outside the protein-coding sequence, each disease-associated allele should be evaluated to see whether it is in linkage disequilibrium with a variant that changes the protein structure (for example, a non-synonymous mutation or truncating mutations that introduce a premature stop codon). If it is, then these findings should be fast tracked for functional studies in human cells and animal models to assess gain of function or loss of function. For non-coding risk alleles, the effect on gene expression (expression quantitative trait loci) should be evaluated in a relevant human cell type. If a risk allele is associated with higher gene expression, then pharmacological inhibition may be effective in treating the disease.

Ultimately, however, we believe that an allelic series will be most valuable for prioritizing which genes implicated by GWASs for complex diseases should be followed up for drug discovery. That is, if multiple alleles modulate gene function in a way that can be linked to a phenotype that is a good surrogate for drug efficacy, then this provides strong evidence that pharmacological modulation of the same target will also be effective at treating the disease. To find an allelic series, large-scale genetic studies, including whole-genome sequencing studies in large patient cohorts, are required to define the complete spectrum of alleles (from common to rare alleles). Although these studies are expensive, the cost is modest when compared to the cost of the entire drug discovery process, which has recently been estimated to approach ~$2 billion when failures are taken into account[3]. Indeed, a drug discovery programme that is anchored in human genetics many actually lower costs, as discussed briefly below.

Limitations of genetics-based target validation

Although some limitations of target validation based on human genetics have been described above, several important limitations are revisited again here. First, not all genes in the human genome will have an allelic series to derive function–phenotype dose–response curves. Many safe and effective drugs have been developed without any direct genetic evidence, and there is little direct evidence to date that genetic data would have identified the target (or targets) of these drugs. As one example, biologics that target the inflammatory cytokine tumour necrosis factor (TNF) are remarkably effective at treating rheumatoid arthritis, but genetics alone has not yet identified TNF as a drug target.

Second, the complexity between genetic diathesis and disease pathogenesis should not be underestimated. We have emphasized that human genetics represents the first step towards a complete package for drug development. Substantial investments in functional follow-up studies — in humans, animal models and cellular models — will be crucial for realizing the potential of human genetics in drug discovery. In some instances, an approach that is anchored in human genetics may slow down a drug discovery programme, especially if human genetics identifies a drug target for which the biology is not well understood or that does not conform to the existing model of disease pathogenesis.

Third, disease-associated alleles, especially those discovered by GWASs, often have a very small effect on the overall risk of disease. Direct testing is required to determine whether exaggerated pharmacological modulation of the same target will have an effect beyond that observed from human genetics. For example, a common polymorphism in *HMGCR*, which has a very small effect on variation in LDL cholesterol levels in the general population[95], highlights that the relationship between genetic perturbation and pharmacological modulation is not a one-to-one relationship. In fact, based on *HMGCR* and other examples cited above, we believe that a key feature of human genetics is to identify which targets — when perturbed — will lead to safe and effective therapies; human genetics may not directly indicate how much target modulation is optimal to treat disease. An allelic series with a range of effects may help to overcome this limitation, if such gain-of-function and loss-of-function alleles can be identified.

Spectrum of alleles
Somewhat arbitrary thresholds for the frequency of alleles observed in the general population; 'common alleles' are those that are observed in >5% of the general population; 'low-frequency alleles' are those that are observed in 0.1–5% of the general population; and 'rare alleles' are private to families; in practical terms, alleles that are common or low-frequency can be catalogued in a reference population (for example, the International HapMap Project) to facilitate testing in another population (for example, patients), whereas rare alleles must be discovered and tested in the same individuals.

Potential for reduced attrition and lower costs

At the beginning of this article, we highlighted the issue of the increasing costs of drug development, which are driven primarily by drug failures in Phase II and Phase III clinical trials[3]. Despite the limitations of human genetics cited above, it does have the potential to have a major impact on the cost of drug development. It is estimated that a reduction in Phase II attrition from 66% to 50% would decrease the cost per new molecular entity by ~$0.5 billion, and a reduction in Phase III attrition from 30% to 20% would decrease costs by ~$0.3 billion[3]. Accordingly, the most obvious practical application of human genetics in drug development is to increase the probability that therapeutic modulation of a target will yield a drug that is safe and effective in humans (that is, decrease the rate of attrition).

During the course of functional studies to understand the biological consequences of disease-associated alleles, it is likely that biomarkers will be developed that can serve as surrogate end points for early proof-of-concept studies. An appealing strategy is a 'quick win, fast fail' paradigm[3], in which proof-of-concept mechanistic studies are filled with drugs that emerge from human genetics. Only those molecules that engage their target (or targets) and have a desired pharmacological activity in humans — a stringent test of the therapeutic hypothesis — would be advanced into Phase II studies.

Human genetics may also help to deprioritize drug development programmes that were started without the benefit of human genetic data, if genetic data do not support the therapeutic hypothesis. One example, as discussed above, is alleles that are associated with HDL cholesterol and the development of drugs to raise HDL cholesterol and prevent cardiovascular disease.

Thus, we argue that an increased investment in R&D — and, specifically, in large-scale human genetics studies and functional follow-up studies to estimate dose–response curves at the stage of target validation — will result in an overall decrease in the cost of drug development.

Pathway-based approach

In this article, we have focused almost exclusively on an approach that uses human genetics to identify a series of alleles that are associated with a human trait and that could be used to derive dose–response curves at the time of target validation. However, a complementary approach is to use human genetics to uncover biological pathways that are important in human disease, and then to use a pathway-based approach to conduct high-throughput screens[96]. A pathway-based approach is appealing because it attempts to model the complex relationships between human genetic perturbations and disease.

There are an increasing number of computational strategies to derive biological insight from human genetics data[97,98]. When coupled with high-throughput biological strategies to interrogate networks[99,100], a pathway-based approach may prove to be quite powerful. For example, genes that are involved in bone mineral density are mapped in or near genes encoding proteins that are involved in pharmacological pathways related to osteoporosis: for example, *TNFSF11* encodes RANKL, *TNFRSF11B* encodes osteoprotegerin, *TNFRSF11A* encodes RANK, parathyroid hormone-like hormone (*PTHLH*) encodes the parathyroid hormone-related protein (PTHRP), *LRP5* encodes LDLR-related protein 5, *SOST* encodes sclerostin and *DKK1* encodes Dickkopf-related protein 1 (REF. 64). The strengths and limitations of the pathway-based approach are of great interest but beyond the scope of this Review.

Conclusions

The ideal preclinical model would provide a reliable estimate of the dose–response relationships between target perturbation and efficacy or safety in humans. In theory, experiments of nature that are based on human genetic variation can be used to generate dose–response curves at the time of target validation, and there are compelling examples that demonstrate the utility of such knowledge in drug discovery. The ultimate success, however, will depend on whether the criteria outlined in BOX 2 can be fulfilled for novel drug targets. To accomplish this vision, there is a pressing need to continue and expand large-scale disease consortia to discover the complete spectrum of alleles (from common to rare alleles) associated with complex traits. Given the underlying architecture of complex traits[101], this is likely to require genome-wide sequencing in large patient collections. Furthermore, collaborations between geneticists and biologists will be required to link mutations with function in cells derived from humans. If genetics can unlock novel genotype–phenotype relationships, then this will provide substantial new therapeutic opportunities for many diseases that are currently inadequately treated.

1. Scannell, J. W., Blanckley, A., Boldon, H. & Warrington, B. Diagnosing the decline in pharmaceutical R&D efficiency. *Nature Rev. Drug Discov.* **11**, 191–200 (2012).
2. Kola, I. & Landis, J. Can the pharmaceutical industry reduce attrition rates? *Nature Rev. Drug Discov.* **3**, 711–715 (2004).
3. Paul, S. M. *et al.* How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nature Rev. Drug Discov.* **9**, 203–214 (2010).
This article provides a good perspective on the challenges facing the pharmaceutical industry, including the need for better preclinical models to validate drug targets.
4. Arrowsmith, J. Trial watch: phase II failures: 2008–2010. *Nature Rev. Drug Discov.* **10**, 328–329 (2011).
5. DiMasi, J. A. & Faden, L. B. Competitiveness in follow-on drug R&D: a race or imitation? *Nature Rev. Drug Discov.* **10**, 23–27 (2011).
6. Wehling, M. Assessing the translatability of drug projects: what needs to be scored to predict success? *Nature Rev. Drug Discov.* **8**, 541–546 (2009).
7. Glyn, J. The discovery and early use of cortisone. *J. R. Soc. Med.* **91**, 513–517 (1998).
8. Tobert, J. A. Lovastatin and beyond: the history of the HMG-CoA reductase inhibitors. *Nature Rev. Drug Discov.* **2**, 517–526 (2003).
9. Brown, M. S. & Goldstein, J. L. Expression of the familial hypercholesterolemia gene in heterozygotes: mechanism for a dominant disorder in man. *Science* **185**, 61–63 (1974).
10. Rader, D. J., Cohen, J. & Hobbs, H. H. Monogenic hypercholesterolemia: new insights in pathogenesis and treatment. *J. Clin. Invest.* **111**, 1795–1803 (2003).
11. The Lovastatin Study Group II. Therapeutic response to lovastatin (mevinolin) in nonfamilial hypercholesterolemia. A multicenter study. *JAMA* **256**, 2829–2834 (1986).
12. Abifadel, M. *et al.* Mutations in PCSK9 cause autosomal dominant hypercholesterolemia. *Nature Genet.* **34**, 154–156 (2003).
This is the first study to describe a gain-of-function mutation in *PCSK9* that causes hypercholesterolaemia.
13. Cohen, J. *et al.* Low LDL cholesterol in individuals of African descent resulting from frequent nonsense mutations in PCSK9. *Nature Genet.* **37**, 161–165 (2005).

14. Kotowski, I. K. *et al.* A spectrum of PCSK9 alleles contributes to plasma levels of low-density lipoprotein cholesterol. *Am. J. Hum. Genet.* **78**, 410–422 (2006).
15. 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).
This is a landmark study that relates loss-of-function mutations in *PCSK9* to low LDL cholesterol levels and protection from heart disease.
16. Park, S. W., Moon, Y. A. & Horton, J. D. Post-transcriptional regulation of low density lipoprotein receptor protein by proprotein convertase subtilisin/kexin type 9a in mouse liver. *J. Biol. Chem.* **279**, 50630–50638 (2004).
17. Maxwell, K. N. & Breslow, J. L. Adenoviral-mediated expression of Pcsk9 in mice results in a low-density lipoprotein receptor knockout phenotype. *Proc. Natl Acad. Sci. USA* **101**, 7100–7105 (2004).
18. Stein, E. A. *et al.* Effect of a monoclonal antibody to PCSK9, REGN727/SAR236553, to reduce low-density lipoprotein cholesterol in patients with heterozygous familial hypercholesterolaemia on stable statin dose with or without ezetimibe therapy: a phase 2 randomised controlled trial. *Lancet* **380**, 29–36 (2012).
19. Stein, E. A. *et al.* Effect of a monoclonal antibody to PCSK9 on LDL cholesterol. *N. Engl. J. Med.* **366**, 1108–1118 (2012).
This paper describes one of the first clinical trials demonstrating that a drug that mimics the effect of *PCSK9* mutations is effective at lowering LDL cholesterol levels in patients.
20. Mullard, A. Cholesterol-lowering blockbuster candidates speed into Phase III trials. *Nature Rev. Drug Discov.* **11**, 817–819 (2012).
21. Kathiresan, S. *et al.* Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. *Nature Genet.* **40**, 189–197 (2008).
22. Kathiresan, S. Will cholesteryl ester transfer protein inhibition succeed primarily by lowering low-density lipoprotein cholesterol?: insights from human genetics and clinical trials. *J. Am. Coll. Cardiol.* **60**, 2049–2052 (2012).
23. Barter, P. & Rye, K. A. Cholesteryl ester transfer protein inhibition to reduce cardiovascular risk: where are we now? *Trends Pharmacol. Sci.* **32**, 694–699 (2011).
24. Bots, M. L. *et al.* Torcetrapib and carotid intima-media thickness in mixed dyslipidaemia (RADIANCE 2 study): a randomised, double-blind trial. *Lancet* **370**, 153–160 (2007).
25. Voight, B. F. *et al.* Plasma HDL cholesterol and risk of myocardial infarction: a mendelian randomisation study. *Lancet* **380**, 572–580 (2012).
This study is an example of how human genetics can be used to deprioritize therapeutic targets, arguing that drugs that raise HDL cholesterol levels will not be effective at lowering the risk of cardiovascular disease.
26. Mickle, J. E. & Cutting, G. R. Clinical implications of cystic fibrosis transmembrane conductance regulator mutations. *Clin. Chest Med.* **19**, 443–458 (1998).
27. Kerem, B. *et al.* Identification of the cystic fibrosis gene: genetic analysis. *Science* **245**, 1073–1080 (1989).
28. Salvatore, D. *et al.* An overview of international literature from cystic fibrosis registries. Part 3. Disease incidence, genotype/phenotype correlation, microbiology, pregnancy, clinical complications, lung transplantation, and miscellanea. *J. Cyst. Fibros.* **10**, 71–85 (2011).
29. 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).
This paper describes clinical trial data for ivacaftor, a drug that has been developed to increase CFTR potentiation and treat patients with cystic fibrosis.
30. Cox, J. J. *et al.* An SCN9A channelopathy causes congenital inability to experience pain. *Nature* **444**, 894–898 (2006).
31. Yang, Y. *et al.* Mutations in SCN9A, encoding a sodium channel α subunit, in patients with primary erythermalgia. *J. Med. Genet.* **41**, 171–174 (2004).
32. Drenth, J. P. *et al.* SCN9A mutations define primary erythermalgia as a neuropathic disorder of voltage gated sodium channels. *J. Invest. Dermatol.* **124**, 1333–1338 (2005).
33. Fertleman, C. R. *et al.* SCN9A mutations in paroxysmal extreme pain disorder: allelic variants underlie distinct channel defects and phenotypes. *Neuron* **52**, 767–774 (2006).
34. Estacion, M. *et al.* NaV1.7 gain-of-function mutations as a continuum: A1632E displays physiological changes associated with erythromelalgia and paroxysmal extreme pain disorder mutations and produces symptoms of both disorders. *J. Neurosci.* **28**, 11079–11088 (2008).
35. Drenth, J. P. & Waxman, S. G. Mutations in sodium-channel gene SCN9A cause a spectrum of human genetic pain disorders. *J. Clin. Invest.* **117**, 3603–3609 (2007).
36. Schmalhofer, W. A. *et al.* ProTx-II, a selective inhibitor of NaV1.7 sodium channels, blocks action potential propagation in nociceptors. *Mol. Pharmacol.* **74**, 1476–1484 (2008).
37. Muroi, Y. *et al.* Selective silencing of $Na_v1.7$ decreases excitability and conduction in vagal sensory neurons. *J. Physiol.* **589**, 5663–5676 (2011).
38. Notarangelo, L. D. *et al.* Mutations in severe combined immune deficiency (SCID) due to JAK3 deficiency. *Hum. Mutat.* **18**, 255–263 (2001).
39. van Vollenhoven, R. F. *et al.* Tofacitinib or adalimumab versus placebo in rheumatoid arthritis. *N. Engl. J. Med.* **367**, 508–519 (2012).
40. Fleischmann, R. *et al.* Placebo-controlled trial of tofacitinib monotherapy in rheumatoid arthritis. *N. Engl. J. Med.* **367**, 495–507 (2012).
41. Neptune, E. R. *et al.* Dysregulation of TGF-β activation contributes to pathogenesis in Marfan syndrome. *Nature Genet.* **33**, 407–411 (2003).
42. Stranger, B. E., Stahl, E. A. & Raj, T. Progress and promise of genome-wide association studies for human complex trait genetics. *Genetics* **187**, 367–383 (2011).
43. Jostins, L. *et al.* Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* **491**, 119–124 (2012).
44. Raychaudhuri, S. *et al.* Common variants at CD40 and other loci confer risk of rheumatoid arthritis. *Nature Genet.* **40**, 1216–1223 (2008).
45. Fairfax, B. P. *et al.* Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles. *Nature Genet.* **44**, 502–510 (2012).
46. Rivas, M. A. *et al.* Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. *Nature Genet.* **43**, 1066–1073 (2011).
47. Todd, J. A. Etiology of type 1 diabetes. *Immunity* **32**, 457–467 (2010).
48. Plenge, R. M. *et al.* Replication of putative candidate-gene associations with rheumatoid arthritis in >4,000 samples from North America and Sweden: association of susceptibility with PTPN22, CTLA4, and PADI4. *Am. J. Hum. Genet.* **77**, 1044–1060 (2005).
49. Eyre, S. *et al.* High-density genetic mapping identifies new susceptibility loci for rheumatoid arthritis. *Nature Genet.* **44**, 1336–1340 (2012).
50. Lettre, G. *et al.* DNA polymorphisms at the BCL11A, HBS1L-MYB, and β-globin loci associate with fetal hemoglobin levels and pain crises in sickle cell disease. *Proc. Natl Acad. Sci. USA* **105**, 11869–11874 (2008).
51. Uda, M. *et al.* Genome-wide association study shows BCL11A associated with persistent fetal hemoglobin and amelioration of the phenotype of β-thalassemia. *Proc. Natl Acad. Sci. USA* **105**, 1620–1625 (2008).
52. Menzel, S. *et al.* A QTL influencing F cell production maps to a gene encoding a zinc-finger protein on chromosome 2p15. *Nature Genet.* **39**, 1197–1199 (2007).
53. Eriksson, M. *et al.* Recurrent *de novo* point mutations in lamin A cause Hutchinson–Gilford progeria syndrome. *Nature* **423**, 293–298 (2003).
54. Worman, H. J., Fong, L. G., Muchir, A. & Young, S. G. Laminopathies and the long strange trip from basic cell biology to therapy. *J. Clin. Invest.* **119**, 1825–1836 (2009).
55. De Sandre-Giovannoli, A. *et al.* Lamin A truncation in Hutchinson-Gilford progeria. *Science* **300**, 2055 (2003).
56. Usifo, E. *et al.* Low-density lipoprotein receptor gene familial hypercholesterolemia variant database: update and pathological assessment. *Ann. Hum. Genet.* **76**, 387–401 (2012).
57. Imperato-McGinley, J., Guerrero, L., Gautier, T. & Peterson, R. E. Steroid 5α-reductase deficiency in man: an inherited form of male pseudohermaphroditism. *Science* **186**, 1213–1215 (1974).
58. Andersson, S., Berman, D. M., Jenkins, E. P. & Russell, D. W. Deletion of steroid 5α-reductase 2 gene in male pseudohermaphroditism. *Nature* **354**, 159–161 (1991).
59. Rittmaster, R. S. Finasteride. *N. Engl. J. Med.* **330**, 120–125 (1994).
60. Sanseau, P. *et al.* Use of genome-wide association studies for drug repositioning. *Nature Biotech.* **30**, 317–320 (2012).
This is a study that integrated GWAS data with drug databases, thereby showing that the genes targeted by many approved therapies have been implicated by human genetics.
61. Burkhardt, R. *et al.* Common SNPs in HMGCR in micronesians and whites associated with LDL-cholesterol levels affect alternative splicing of exon13. *Arterioscler. Thromb. Vasc. Biol.* **28**, 2078–2084 (2008).
62. Altshuler, D. *et al.* The common PPARγ Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. *Nature Genet.* **26**, 76–80 (2000).
63. Tsoi, L. C. *et al.* Identification of 15 new psoriasis susceptibility loci highlights the role of innate immunity. *Nature Genet.* **44**, 1341–1348 (2012).
64. Estrada, K. *et al.* Genome-wide meta-analysis identifies 56 bone mineral density loci and reveals 14 loci associated with risk of fracture. *Nature Genet.* **44**, 491–501 (2012).
65. Brooke, B. S. *et al.* Angiotensin II blockade and aortic-root dilation in Marfan's syndrome. *N. Engl. J. Med.* **358**, 2787–2795 (2008).
66. Klein, R. J. *et al.* Complement factor H polymorphism in age-related macular degeneration. *Science* **308**, 385–389 (2005).
This study represents one of the first GWASs. Based on findings from this and other studies, drugs targeting the complement pathway are under development for AMD.
67. Maller, J. *et al.* Common variation in three genes, including a noncoding variant in CFH, strongly influences risk of age-related macular degeneration. *Nature Genet.* **38**, 1055–1059 (2006).
68. Maller, J. B. *et al.* Variation in complement factor 3 is associated with risk of age-related macular degeneration. *Nature Genet.* **39**, 1200–1201 (2007).
69. Raychaudhuri, S. *et al.* A rare penetrant mutation in CFH confers high risk of age-related macular degeneration. *Nature Genet.* **43**, 1232–1236 (2011).
70. Hillmen, P. *et al.* Effect of eculizumab on hemolysis and transfusion requirements in patients with paroxysmal nocturnal hemoglobinuria. *N. Engl. J. Med.* **350**, 552–559 (2004).
71. Troutbeck, R., Al-Qureshi, S. & Guymer, R. H. Therapeutic targeting of the complement system in age-related macular degeneration: a review. *Clin. Experiment. Ophthalmol.* **40**, 18–26 (2012).
72. Katschke, K. J. Jr *et al.* Inhibiting alternative pathway complement activation by targeting the factor D exosite. *J. Biol. Chem.* **287**, 12886–12892 (2012).
73. Hingorani, A. D. & Casas, J. P. The interleukin-6 receptor as a target for prevention of coronary heart disease: a mendelian randomisation analysis. *Lancet* **379**, 1214–1224 (2012).
74. Park, H., Bourla, A. B., Kastner, D. L., Colbert, R. A. & Siegel, R. M. Lighting the fires within: the cell biology of autoinflammatory diseases. *Nature Rev. Immunol.* **12**, 570–580 (2012).
75. Lunn, M. R. & Stockwell, B. R. Chemical genetics and orphan genetic diseases. *Chem. Biol.* **12**, 1063–1073 (2005).
76. Russman, B. S., Iannaccone, S. T. & Samaha, F. J. A phase 1 trial of riluzole in spinal muscular atrophy. *Arch. Neurol.* **60**, 1601–1603 (2003).
77. Abbara, C. *et al.* Riluzole pharmacokinetics in young patients with spinal muscular atrophy. *Br. J. Clin. Pharmacol.* **71**, 403–410 (2011).
78. Wadman, R. I. *et al.* Drug treatment for spinal muscular atrophy type I. *Cochrane Database Syst. Rev.* **4**, CD006281 (2012).
79. Dietz, H. C. New therapeutic approaches to Mendelian disorders. *N. Engl. J. Med.* **363**, 852–863 (2010).
This is a good review on therapeutic approaches based on genetic findings from Mendelian diseases, including the example of Marfan's syndrome.
80. Lorson, C. L., Rindt, H. & Shababi, M. Spinal muscular atrophy: mechanisms and therapeutic strategies. *Hum. Mol. Genet.* **19**, R111–R118 (2010).
81. Melki, J. *et al. De novo* and inherited deletions of the 5q13 region in spinal muscular atrophies. *Science* **264**, 1474–1477 (1994).

82. Lefebvre, S. *et al.* Identification and characterization of a spinal muscular atrophy-determining gene. *Cell* **80**, 155–165 (1995).
83. Hirschhorn, J. N. & Daly, M. J. Genome-wide association studies for common diseases and complex traits. *Nature Rev. Genet.* **6**, 95–108 (2005).
84. McCarthy, M. I. *et al.* Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nature Rev. Genet.* **9**, 356–369 (2008).
85. Cirulli, E. T. & Goldstein, D. B. Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nature Rev. Genet.* **11**, 415–425 (2010).
86. Lander, E. & Kruglyak, L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nature Genet.* **11**, 241–247 (1995).
87. MacArthur, D. G. *et al.* A systematic survey of loss-of-function variants in human protein-coding genes. *Science* **335**, 823–828 (2012).
88. Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nature Methods* **7**, 248–249 (2010).
89. Kumar, P., Henikoff, S. & Ng, P. C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature Protoc.* **4**, 1073–1081 (2009).
90. Jonsson, T. *et al.* A mutation in APP protects against Alzheimer's disease and age-related cognitive decline. *Nature* **488**, 96–99 (2012).
This study shows that a loss-of-function mutation in the *APP* gene protects against Alzheimer's disease.
91. Kero, M. *et al.* Amyloid precursor protein (APP) A673T mutation in the elderly Finnish population. *Neurobiol. Aging* **34**, 1518.e1–1518.e3 (2013).
92. Gashaw, I., Ellinghaus, P., Sommer, A. & Asadullah, K. What makes a good drug target? *Drug Discov. Today* **16**, 1037–1043 (2011).
93. Katan, M. B. Commentary: Mendelian randomization, 18 years on. *Int. J. Epidemiol.* **33**, 10–11 (2004).
94. Ebrahim, S. & Davey Smith, G. Mendelian randomization: can genetic epidemiology help redress the failures of observational epidemiology? *Hum. Genet.* **123**, 15–33 (2008).
95. Kathiresan, S. *et al.* Common variants at 30 loci contribute to polygenic dyslipidemia. *Nature Genet.* **41**, 56–65 (2009).
96. Swinney, D. C. Phenotypic versus target-based drug discovery for first-in-class medicines. *Clin. Pharmacol. Ther.* **93**, 299–301 (2013).
97. Rossin, E. J. *et al.* Proteins encoded in genomic regions associated with immune-mediated disease physically interact and suggest underlying biology. *PLoS Genet.* **7**, e1001273 (2011).
98. Trynka, G. *et al.* Chromatin marks identify critical cell types for fine mapping complex trait variants. *Nature Genet.* **45**, 124–130 (2013).
99. Barabasi, A. L., Gulbahce, N. & Loscalzo, J. Network medicine: a network-based approach to human disease. *Nature Rev. Genet.* **12**, 56–68 (2011).
100. Schadt, E. E., Friend, S. H. & Shaywitz, D. A. A network view of disease and compound screening. *Nature Rev. Drug Discov.* **8**, 286–295 (2009).
101. Stahl, E. A. *et al.* Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis. *Nature Genet.* **44**, 483–489 (2012).
102. Dietz, H. C. *et al.* Marfan syndrome caused by a recurrent *de novo* missense mutation in the fibrillin gene. *Nature* **352**, 337–339 (1991).
103. International HapMap Consortium. A haplotype map of the human genome. *Nature* **437**, 1299–1320 (2005).
104. Altshuler, D., Daly, M. J. & Lander, E. S. Genetic mapping in human disease. *Science* **322**, 881–888 (2008).
105. Hindorff, L. A. *et al.* Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl Acad. Sci. USA* **106**, 9362–9367 (2009).
106. Visscher, P. M., Brown, M. A., McCarthy, M. I. & Yang, J. Five years of GWAS discovery. *Am. J. Hum. Genet.* **90**, 7–24 (2012).
107. Nicolae, D. L. *et al.* Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS Genet.* **6**, e1000888 (2010).
108. Raychaudhuri, S. *et al.* Identifying relationships among genomic disease regions: predicting genes at pathogenic SNP associations and rare deletions. *PLoS Genet.* **5**, e1000534 (2009).
109. Hu, X. *et al.* Integrating autoimmune risk loci with gene-expression data identifies specific pathogenic immune cell subsets. *Am. J. Hum. Genet.* **89**, 496–506 (2011).
110. Nejentsev, S., Walker, N., Riches, D., Egholm, M. & Todd, J. A. Rare variants of IFIH1, a gene implicated in antiviral responses, protect against type 1 diabetes. *Science* **324**, 387–389 (2009).
111. Li, G. *et al.* Human genetics in rheumatoid arthritis guides a high-throughput drug screen of the CD40 signaling pathway. *PLoS Genet.* **9**, e1003487 (2013).

Acknowledgements
The authors thank S. Kathiresan for his assistance in providing critical comments on the manuscript.

Competing interests statement
The authors declare competing financial interests: see Web version for details.

FURTHER INFORMATION
Catalogue of Published Genome-Wide Association Studies: http://www.genome.gov/gwastudies

ALL LINKS ARE ACTIVE IN THE ONLINE PDF