HHS Public Access

Author manuscript

Trends Genet. Author manuscript; available in PMC 2019 July 01.

Published in final edited form as:

Trends Genet. 2018 July ; 34(7): 558–570. doi:10.1016/j.tig.2018.04.004.

Genetic-Driven Druggable Target Identification and Validation

Matteo Floris[1,2], **Stefania Olla**[2], **David Schlessinger**[3], and **Francesco Cucca**[1,2,*]

[1]Dipartimento di Scienze Biomediche, Universita degli Studi di Sassari, Sassari, Italy

[2]IRGB-CNR, Istituto di Ricerca Genetica e Biomedica, Consiglio Nazionale delle Ricerche (CNR), Monserrato, Cagliari, Italy

[3]Laboratory of Genetics, National Institute on Aging, National Institutes of Health, Baltimore, MD, USA

Abstract

Choosing the right biological target is the critical primary decision for the development of new drugs. Systematic genetic association testing of both human diseases and quantitative traits, along with resultant findings of coincident associations between them, is becoming a powerful approach to infer drug targetable candidates and generate *in vitro* tests to identify compounds that can modulate them therapeutically. Here, we discuss opportunities and challenges, and infer criteria for the optimal use of genetic findings in the drug discovery pipeline.

The Impact of Human Genetics in Drug Discovery

Identifying the molecules and mechanisms primarily involved in disease pathophysiology is an essential first step in the discovery of targets that can be modulated therapeutically. However, the search for new pharmacological targets is often limited by partial information about disease pathogenesis, and is typically based on indirect evidence from animal, cell, and *ex vivo* models and human epidemiological studies [1,2]. For instance, as summarized in a recent review [3], therapies based on animal model biology often fail to replicate when tested in human trials. Likewise, human epidemiological studies, such as those comparing blood levels of candidate disease-related molecules or exposures in cases versus controls, are subject to the bias of **reverse causality** (see Glossary).

Thus, there is a great need for more robust information about targets and their potential clinical efficacy. Relevant information should be based on disease-related causal human biology and more adequate preclinical assays to measure truly disease-related quantitative parameters [4].

To enhance the chances of successful drug discovery programs, human genetics is increasingly revealing its power. An early successful and paradigmatic history of translation from genetics to clinical practice started with the discovery, in 2003, that rare gain-of-

*Correspondence: fcucca@uniss.it (F. Cucca).

Appendix A Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.tig.2018.04.004.

function mutations (GoF) in the *PCSK9* gene (encoding pro-protein convertase subtilisin/kexin type 9) cause autosomal dominant hypercholesterolemia [5]. A few years later, targeted sequencing of this gene revealed that the reciprocal was also true: loss-of-function mutations (LoF) in the same gene were associated with substantial decreases in plasma levels of low-density lipoprotein (LDL) cholesterol [6], as well as with a significant reduction in the incidence of coronary heart disease (CHD) [7]. Such experiments of nature suggested that the therapeutic inhibition of *PCSK9* could decrease LDL cholesterol concentration and help prevent CHD, and resulted in approval, in 2015, of a series of *PCSK9* inhibitors as LDL cholesterol-lowering treatment for CHD [8].

The role of genetics in drug development is further supported by more recent independent analyses of multiple pharmaceutical pipelines [9,10] revealing that the proportion of drug mechanisms with direct genetic support [{i.e., drug targets with at least one genetic association with the disease in Online Mendelian Inheritance in Man (OMIM)[i] [11] or **genome-wide association study** (GWAS) catalogs[ii] [12]]} significantly increases along the course of drug development (from 2.0% at the preclinical stage to 8.2% among approved drugs). Overall, 73% of projects to discover therapeutic drugs with a genetic link between target and disease are active or successful in Phase II trials, compared with 43% of projects not having such support [9]. The potential of human genetics in drug development is also suggested by the increasing involvement of pharmaceutical companies in genomic research projects [13].

Here, we discuss the increasing benefits for drug discovery gained from recent advances in genomic and large-scale genetic studies of traits and diseases.

Selection and Prioritization of Druggable Targets Using Genetic Discoveries

The enormous potential of human genetic analysis for drug development is only beginning to be realized [1,14]. The underlying idea is to use the powerful tool of human genetic association to identify key drug targets that are involved in the etiology and pathogenesis of disease and that can be therapeutically modulated; Figure 1 (Key Figure) shows a schematic of the main areas in which human genetics can help in the selection and prioritization of drug targets, as well as in the generation of *in vitro* assays to assess the efficacy of modulating compounds. We place a strong emphasis on the detection of disease-related **intermediate phenotypes**. As described in the following sections, this can be optimally achieved through a combination of sequential approaches, including (i) finding coincident genetic associations of disease risk and quantitative traits; (ii) determining the causal genes underlying such coincident associations and establishing the direction of effects; and (iii) refining the causality link and obtaining further evidence of a substantive role for the therapeutic target in the disease process, based on available biological information and new focused experiments. It is salient that, because targets are easier to inhibit than to activate, the highest priority should be given to genes that are downregulated by variants protective against disease or upregulated by predisposing variants. In a further step toward assessing

[i] www.omim.org
[ii] www.ebi.ac.uk/gwas/

utility, any naturally occurring LoF mutations in the human population can often provide relevant information about the extent of possible serious clinical adverse effects.

Identifying Coincident Associations

The increasing availability of multi-trait GWAS summary statistics offers a tool to identify shared associations between disease or quantitative traits; to highlight etiological and mechanistic similarities and differences between diseases; and, in a broader sense, to understand better the phenotypic consequences of human variation and thereby help in the selection of optimal therapeutic targets. The GWAS catalog [12] is the indispensable compendium for such initial searches; for example, in the presence of an association with a metabolite [15] or an immune cell type [16], the catalog can be used to verify whether there are overlaps with known associations with disease. Currently, approximately 29 500 recorded associations with diseases and traits are **genome-wide significant** ($P \leq 5E^{-8}$) [17–19]. Most of the disease associations are with immune system disorders (~33%) (see Table 1 for categorization).

Detecting overlaps between GWAS associations can be confidently achieved using co-localization methods [20–22] that, in this context, aid in verifying whether a shared causal variant between a quantitative trait and a disease association signal is plausible. In recent years, several formal statistical tests have been proposed to resolve issues about co-localization, including overestimation of effect sizes due to the 'Winner's curse', how to treat multiple association signals in the same locus, differences in map density in different studies, and differing linkage disequilibrium patterns in different populations [21,23]. With the help of the refined methods, assessments of co-localization can be now systematically carried out with unprecedented statistical power using genetic data from mega-repositories, such as the UK Biobank[iii] [24] and large case–control consortium-type studies [19,25,26]; for most tests, summary statistics, such as allelic frequencies and *P* values, are sufficient.

As discussed early on [27], founder populations can be especially advantageous to search for coincident associations between quantitative traits and clinical endpoints [16]. Such populations, which are enriched in variants that are rarer or even absent in more mixed populations [18,28–32], may reveal associations and potential therapeutic targets that are otherwise missed. For instance, in a recent example, a variant in the *TNFSF13B* gene, common on the island of Sardinia although rarer elsewhere, was shown to increase blood levels of the cytokine and drug target BAFF, leading to a cascade of immune changes that augment the risk of multiple sclerosis and systemic lupus erythematosus [33].

Identifying the Genes and/or Variants Underlying Overlapping Associations

After establishing co-localized association signals that are likely to share a causal variant, a critical step toward the identification of the right therapeutic targets is to identify the gene and, preferably, variant underpinning such overlapping associations. This requires a combination of sequential approaches to reduce the association signal to primary element(s).

[iii] http://ukbiobank.ac.uk

An initial strategy encompassing different methods [34] collectively known as ‘fine mapping’ is aimed at statistically excluding all but one or a few polymorphisms as causal variants in GWAS-associated regions. This strategy requires comprehensive and unbiased ascertainment of genetic variation, through large-scale DNA sequencing and the use of informative imputation panels, to split the genetic contributions of individual variants in an associated region, allowing the prioritization of those with the highest probability of being causal. Coincident association with a quantitative trait, in addition to providing relevant functional clues, may also simplify the fine mapping of the variant involved in disease predisposition and/or protection. In fact, associations with intermediate phenotypes are inherently less complex and typically show effect sizes much larger than those observed for clinical endpoints, so that they are more potent in pinpointing statistically convincing causal variants [33]. Further benefits can come from cross-population comparisons that can help break down correlation or linkage disequilibrium (LD) between genetic variants because of differences in the LD structure in different ethnic backgrounds [31,33–36]. The most plausible causal polymorphism(s) are then ranked with different metrics, based on cross-species sequence conservation, functional genomic data (e.g., transcription factor binding) and transcript information that quantitatively predict functional relevance [37,38], helping to connect them with a causal gene and its products. Unfortunately, even after these methods have been used, the genetic resolution of the association signal to a single variant and gene may still be limited by the strong LD between different candidate variants, which, in extreme cases, may be so closely correlated that they are genetically indistinguishable (because they always co-occur). Further difficulty arises because most (>90%) lead variants of association signals are located in ‘noncoding regions’ of the genome [12] with only 10–20% altering known transcription factor sequence motifs [39]. Furthermore, even the statistical refinement of the association signal to a putative causal DNA variant within gene-flanking and intronic regions does not indicate *per se* that the gene harboring them is causal. Indeed, there are multiple examples of the long-range control of gene expression from variants located in nearby genes [40–42], detected through technologies such as promoter capture with ‘Hi-C’[iv] [43].

A useful approach to identify causal genes underlying the associations of interest and to pinpoint their products as therapeutic targets assesses whether the prioritized variant(s) affect the expression of a gene, through **expression quantitative trait locus** (eQTL) and/or **protein QTL** (pQTL) analyses. In addition, these analyses, by revealing *trans*-eQTL and/or *trans*-pQTL associations, may also reveal suitable protein targets for therapeutic intervention even encoded by genes localized on different chromosomes but the expression of which is influenced by the causal variant and/or gene [44,45].

The utility of pQTL and eQTL analyses extends to the determination of the effective direction of the association. This is inferred from the direction of change in levels of the product of a gene associated with disease risk; for instance, seeing whether an allele protective against disease (the effect of which we wish to reproduce therapeutically) decreases or increases levels of a gene transcript or encoded protein. Thus, this is a critical

[iv] http://promoter.bx.psu.edu/hi-c/

step because it informs the direction (inhibition or stimulation) of therapeutic modulation of the target.

Such analyses are facilitated by the rapidly growing number of large data sets annotating information that can systematically help to bridge GWAS associations with expression levels. One fundamental resource is the GTEX catalog[v] (release V7, accessed April 2018), providing eQTL analysis for 48 distinct tissues in 620 individuals [46]. Additional sources to help assess the impact of regulatory variants include databases, such as HipSci[vi] [47], reporting mutations in reprogrammed induced pluripotent stem cells (iPSCs) [48]. This information can be merged with integrated chromatin profiles and chromatin contact maps [49,50] as well as with data on DNA methylation profiles and methylation QTL [51].

Regarding pQTL databases, a few surveys have been reported, mostly focusing on the plasma proteome. For example, the effects of 10.6 million DNA variants on the levels of 2994 proteins in 3301 individuals were assessed, and 1927 genetic associations with 1478 proteins were identified [52].

Additional useful annotation is available regarding the genetic control of immune cells and molecules that are directly relevant in the context of immunological and autoimmune diseases[vii] [16,53–56]. Several data sets are also available for genetic variant effects on serum and/or plasma levels of a large number of metabolites [15], relevant to a range of hematological and metabolic diseases. Such data sets will become increasingly valuable as QTL analyses are extended to more subtypes of cells in the body.

Further Resolving Causal Relationships

To select the optimal therapeutic targets, pinpointing the causal gene encoding or directly regulating a putative intermediate phenotype and obtaining information about its transcript and protein product(s) must be followed by further studies to establish its precise role in disease etiopathogenesis.

An emerging feature of current GWAS results that can complicate the resolution of the causal relationships to a true intermediate phenotype and, hence, the identification of the right therapeutic target, is the finding of numerous pleiotropic effects, in which one gene influences two or more phenotypic traits [57]. For example, a variant affecting the immune system may change the levels of several types of immune cell, but this does not imply that all of these are involved in the disease process. Furthermore, the true disease-related cell type may have not even been assessed in the study. In the presence of strong pleiotropy, approaches exploiting Mendel’s second law of inheritance to look for multiple independent genetic variants associated with both the same intermediate phenotype and disease outcome provide an incisive route to refine the putative targets to the one(s) most likely involved in disease pathogenesis and, thus, susceptible to therapeutic modulation. The intermediate

[v] www.gtexportal.org
[vi] www.hipsci.org
[vii] http://facsdataexplorer.irgb.cnr.it/

phenotype can also be used to implement *in vitro* assays to test potentially modulating compounds.

Likewise, **Mendelian randomization** (MR) approaches are especially useful to establish causal relationships between modifiable risk factors and clinical endpoints [58]. The approach involves using genetic polymorphisms as proxies, or 'instruments', for measures (e.g., lipid and vitamin levels) related to a target exposure (e.g., diet) to test the association of the genetic instrument with the outcome of interest [59]. Therefore, it is more robust than observational studies because genotypes are not prone to **confounders** and **reverse causation** effects. Furthermore, genotypes, in contrast with epidemiological variables, can be measured with high precision and mirror long-term patterns of exposure beginning early in life. The limitations of observational epidemiology and benefits of MR approaches are exemplified by the large meta-analyses of epidemiological studies that initially supported an inverse association between high-density lipoprotein (HDL) cholesterol and the risk of coronary artery disease (CAD) [60], whereas, consistent with MR analyses [61,62], interventions aimed at increasing HDL cholesterol did not lead to reductions in CAD incidence [63].

Once a potential causal link is established between a gene and/or variant and a disease using all the approaches summarized above, functional validation may begin. This step is a major research challenge that frequently requires years of focused experimentation *in vitro* and *in vivo* using animal models based on human biology. In this respect, the generation of humanized knock-in mice, in which a mouse gene is replaced with the human ortholog carrying the genotype of interest, appears particularly promising [64–66]. Such models take advantage of recent advances in gene-editing strategies and could then be used, through appropriate crossbreeding, to combine several knock-in alleles in individual mice. This will allow the generation of more multigene humanized biological systems both to study intermediate phenotypes and pathways genetically linked to the disease of interest and to carry out *in vivo* tests of potential therapeutic agents.

Predicting A Therapeutic Window

In principle, human genetic variation can also help to define the optimal 'dosage' for therapeutic inhibition and/or stimulation of a given target through the presence of different alleles, a so-called **'allelic series'**, with corresponding effects on levels of intermediate phenotypes producing a 'dose–response profile'. However, given our current knowledge of human genetic variation, it is often not possible to detect allelic series. By contrast, we can anticipate that immense QTL databases including millions of individuals from diverse and thus far underrepresented populations (especially from Africa) [67] may provide a broader spectrum of the consequences of naturally occurring mutations on therapeutic targets. More accurate prediction of the therapeutic window for their pharmacological modulation then becomes possible, as demonstrated by the exemplary case of a G-protein-coupled receptor [68]. In addition, gene-editing methodologies can also be used to generate true allelic series in cellular and animal models. This can be especially important to evaluate targets that have a narrow therapeutic index and, thus, require fine modulation to be effective without adverse effects.

Finally, naturally occurring human knockout mutations that completely ablate the function of a gene provide a lower bound for the inhibition of candidate targets. These variants, categorized through large human compilations, such as the ‘Human knockout project’ [69], integrate well with information from mouse knockout studies to establish whether LoF of a particular gene can be tolerated. They can thereby help to predict well-tolerated targets for extreme therapeutic inhibition [70]. Likewise, they help establish which ‘genes are ‘essential’ (i.e., intolerant to LoF), helping to predict severe adverse effects of the therapeutic inhibition of their protein products [71]. For instance, one study [72] used exome sequence data from 60 706 individuals of diverse ancestries to highlight genes subjected to purifying selection against various classes of mutation, identifying 3230 genes highly intolerant to LoF (i.e., genes that are essentially depleted of predicted protein-truncating variants), 72% of which have no currently reported loss associated with a human disease. These findings suggest caution in the extreme inhibition of proteins corresponding to those genes.

Likewise, extensive phenotypic assessment of multiple traits provides direct information about the functional impact of observed LoF mutations. With this in mind, the exons of 10 503 adult participants from a highly consanguineous cohort from Pakistan who are more likely to inherit two mutant copies of the same gene were sequenced [70]. Indeed, homozygote knockouts in 1317 genes were predicted. This was followed by detailed cardiometabolic phenotypic profiling of more than 200 biochemical and disease traits in these individuals. As a paradigmatic example, the authors found that individuals predicted to be homozygote knockout for the *APOC3* gene, encoding apolipoprotein C3, a component of LDL, had significantly lower levels of blood triglycerides after a meal than did individuals lacking the specific LoF mutation. These findings support apolipoprotein C3 as a therapeutic target for cardiovascular disease, with a low probability of adverse effects even at high levels of inhibition.

Another study within the UKBiobank cohort [24,73] characterized the effect of 18 228 protein-truncating variants across 135 phenotypes and found 28 associations between medical phenotypes and protein-truncating variants in genes that were encoded outside of the major histocompatibility complex.

As discussed above, although pleiotropy represents a complication in resolving causal relationships to the truly disease-related targets, it may offer an ‘early warning system’ for off-target and adverse effects. These are especially relevant because safety issues represent the leading cause for 82% of the terminations of preclinical drug discovery projects [9]. Therefore, a promising approach to predict adverse drug events is to test whether a genetic variant with the desired therapeutic effects is associated with multiple phenotypes across a broad range of different health-related traits [74].

Overall, genetic findings and genetic-driven analyses can thus help to better define the hypothetical therapeutic index for the modulation of any human gene while also pointing toward possible adverse effects.

The Special Value of Positive Selection in Identifying Therapeutic Targets

Studies of natural selection provide further utility for human genetics in defining candidate targets. In this context, positive selection provides an ‘experiment of nature’, typically across thousands of years, in which genetic variation provided a benefit, for example by conferring resistance against a serious infectious disease. In principle, such an effect could be mimicked therapeutically. Positive selection tests are able to identify single nucleotide polymorphisms (SNPs) under selection by looking for frequency differences in different populations [75] and the presence of a long core haplotype (i.e., a segment of DNA that tends to be transmitted as a unit across generations in a population) around the selected variant [76,77]. A signal for selection does not directly identify the selective pressure that has maintained the variant during evolution; but hints are afforded by associations of a positively selected variant with a quantitative trait at a checkpoint controlling the biology of the selective agent. In the case of some infectious diseases, including malaria, the frequency of variants positively selected in different populations can be correlated with the prevalence of the infection now and in the past. Since a reduction in function is customarily easier to be reproduced therapeutically, in this special case, the ideal target is highlighted by a beneficial (positively selected) variant reducing the function of the product of the gene; for instance, a protein the reduced levels or impaired activity of which may protect against a pathogen. Thus, inhibition of the protein product of the affected gene by a drug might produce a comparable defense. Such targets can have advantages over approaches directly targeting an infectious agent, because the candidate target is less prone to develop resistance to therapy than the pathogen: and host-directed therapies are an emerging route to fight TB, HIV, and influenza, as well as malaria [78]. However, as a caution, a drug effective against a particular agent or host-related mechanism could be detrimental for other disease conditions [16,33].

Assessing Gene Druggability

The obvious critical step in the road toward the therapeutic exploitation of a protein target identified with genetic approaches is the assessment of its **druggability**; that is, its susceptibility to be potentially modulated in its effects by drug-like small molecules (typically targeting hydrophobic pockets) or by so-called ‘biologicals’ (more commonly targeting extracellular domains, such as those of receptor proteins or soluble molecules).

In particular, the potential of protein targets to be modulated by drug-like small molecules can be predicted on the basis of sequence and any structural similarity to the targets of approved drugs. Several reports have estimated the set of druggable genes across the genome [79–81]; an exhaustive recent re-examination [82] looked at the druggability potential of all human proteins, estimating 4479 (22% of protein-coding genes) as druggable, and used the information to design an Illumina DrugDev array. In that initiative, the druggable gene set is stratified into three tiers corresponding to the position of the gene product in the drug development pipeline: **Tier 1 genes** include targets of approved small molecules and biotherapeutic drugs (as well as clinical-phase drug candidates); **Tier 2 genes** encode targets with known bioactive drug-like small-molecule binding partners as well as those with high sequence similarity with approved drug targets; and **Tier 3 genes** encode secreted or extracellular proteins that have only distant similarity to approved drug targets, as well as

other members of key druggable gene families not already included in Tier 1 or 2. A limitation of this resource is that only restricted information is provided for each gene, and many genes and/or proteins remain unstudied. One might expect that more detailed information about gene function and pharmaceutical potential could optimize drug repositioning and provide alerts to any adverse effects of the pharmacological modulation of the target. Thus, a more complete public resource based on the procedures used in [82] would have exceptional value. A resource of this kind, PHAROS[viii] [a user interface to the Knowledge Management Center (KMC) for the Illuminating the Druggable Genome (IDG) program funded by the National Institutes of Health] has been developed [83]. The Target Validation Platform[ix] is another source of genetic and high-throughput genomics data for drug target selection, developed by the public–private partnership OpenTargets [84].

In addition to the druggability of a target through ‘small molecules’, it is expected that some will be best suited to therapeutic modulation through ‘biological therapeutics’: drug products manufactured in, extracted from, or semisynthesized from biological sources. For example, associations protective against autoimmune disease that reduce levels of cellular and soluble receptors can be optimally mimicked therapeutically through the generation of monoclonal antibodies.

Even if a potential target is not assessed as druggable by small molecules and biologicals, it can reveal pathways and mechanisms that expand candidate targets. Detailed pathway analysis, through the integration of association signals with annotated pathways (such as Kegg[x] and Reactome[xi]), and focused biological studies, thereby adds candidate molecules that had not previously been discovered by coincident genetic associations. For example, upstream and downstream molecules in a pathway involving a protein associated with a disease can extend potential targets to interacting protein partners [4].

Finally, while we have emphasized traditional protein targets amenable to modulation by small molecules and monoclonal antibodies, pathway analysis may also reveal unconventional targets open to therapeutic intervention with new molecular approaches, such as those based on small interfering RNA and antisense oligonucleotides, mRNA delivery, gene editing with CRISPR-Cas9, and targeting for proteolysis (PROTAC) [85–87].

Concluding Remarks and Emerging Challenges

Recent advances in the assessment of human genetic variation and its phenotypic consequences at the molecular, cellular, and population levels increasingly contribute to the discovery of new drug targets and the generation of assays to test modulating compounds.

A set of ten steps can help to define optimal candidates for pharmaceutical development based on human genetic association: (i) identify coincident associations between clinical endpoints and quantitative variables; (ii) pinpoint the causal gene and related target protein

[viii] https://pharos.nih.gov/idg/index
[ix] http://targetvalidation.org
[x] www.genome.jp/kegg/pathway.html
[xi] https://reactome.org/

through eQTL, pQTL, or DNA conformational data; (iii) obtain a clear direction of change in the level of the gene product (protein) associated with disease risk; (iv) reinforce causal relationships and resolve pleiotropic effects by finding multiple independent associations with the same target and the same disease; (v) provide additional support for the therapeutic target with functional evidence; (vi) take advantage of naturally occurring human variation that produces effects similar to the desired therapeutic intervention; (vii) give higher priority to targets that are downregulated by variants protective against disease; (viii) prioritize products of genes that are not 'essential' and that can be inhibited safely based on human and animal model knockout findings; (ix) obtain information about the optimal therapeutic window from human genetic variation or through *in vitro* and *in vivo* engineered models; and (x) assess the target for druggability and perform detailed pathway analysis to assess additional possible druggable targets.

The most favorable scenario in exploiting genetic data toward target selection through the above steps and using available resources is exemplified by naturally occurring mutations in the *PCSK9* gene as an experiment of nature that has been mimicked therapeutically [8]. However, some emerging challenges provide a more complex scenario than we have discussed thus far and may complicate the use of genetic data in drug development (Table 2). In particular, it is relevant that most GWAS thus far performed looked for associations with the genetic risk of disease occurrence and, therefore, focused on mechanisms of the underlying causes of disease. However, that approach may not identify factors that influence disease progression. Unfortunately, collecting measures of progression and severity in case-only studies is inherently more difficult than acquiring yes–no answers about disease status in case–control studies. Indeed, only a small proportion of GWAS, typically with smaller sample sizes and, hence, reduced statistical power compared with standard GWAS, were conducted to identify variants associated with disease progression, which could be more informative to identify ameliorating therapies [88,89]. Such case-only GWAS may become more extensive as longitudinal and retrospective clinical data are accumulated for large disease cohorts and will further extend the impact of genetics on the development of new drugs.

With available data, the selection of therapeutic targets should try to infer the phase of the disease process at which they are likely to act. For instance, intermediate phenotypes revealing targets that likely affect early phases of a disease process may more likely inform possible prevention strategies, whereas overt disease therapies would aim to select targets that could influence both disease onset and disease progression. For example, autoimmune attack typically begins up to decades before overt clinical onset, by which time non-regenerating target organs, such as the pancreatic beta cells in type 1 diabetes mellitus, have already been extensively destroyed. Instead, therapy directed against targets involved in the initiation of disease could be of special value for affected individuals diagnosed preclinically with genetic and early biomarker profiling. Therapy against such targets could also be useful in preventing recurrence after cell and replacement therapy (e.g., pancreatic beta cell transplant in the case of type 1 diabetes mellitus).

Furthermore, an additional challenge comes from the striking prevalence of gene regulatory mechanisms underlying multifactorial trait associations suggesting cellular and temporal

effects on target expression that can constrain potential therapeutic value. The generation of new modalities to ensure targeted, cell-specific therapeutic delivery, such as those based on polymeric nanoparticles or capsid proteins as molecular recognition units [90–92], are needed and can help to overcome some of these constraints. Resolving these challenges and answering other open questions (see Outstanding Questions) will help to fully exploit the enormous potential of human genetic analysis for drug development.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

he authors thank Maristella Steri, Carlo Sidore, Mauro Pala, Edoardo Fiorillo, and Valeria Orru for helpful suggestions and discussions. We also would like to thank the anonymous reviewers for constructive criticisms and useful comments and suggestions. The authors gratefully acknowledge the support of the Italian Foundation for Multiple Sclerosis (FISM 2015/R/09), the Intramural Research Program of the National Institute on Aging, NIH, with contracts N01-AG-1-2109 and HHSN271201100005C and of the European Union's Horizon 2020 Research and Innovation Program (under grant agreement 633964, ImmunoAgeing).

Glossary

Allelic series:
a set of distinct mutant alleles within different regions of the same gene that affect a range of phenotypes, including different effect sizes on the same trait.

Clinical trial phases:
research studies performed in three phases to collect data on the safety and efficacy of new drugs. Phase I studies test the safety, adverse effects, best dose, and timing of a new treatment in a small number of patients and healthy volunteers. Phase II studies are conducted in a small number of patients at different doses, to preliminarily explore the effectiveness of a drug while seeking guidance on appropriate dosage to see whether the treatment merits further investigation. In Phase III studies, the drug is compared with inactive drug (placebo) or with another drug standard, if available, on a larger number of patients, encoding the classification of patients and record of administration of drug versus placebo or best previous treatment ('double blinded') to ensure better statistical power while guarding against bias.

Coincident genetic associations between disease risk and quantitative traits:
represent full co-localization and/or juxtaposition at a site in DNA of associations between clinical endpoints and one or more variables that can be measured on a continuous scale.

Confounder:
a third variable that can distort the association between two phenomena, such as an exposure and an outcome, making it difficult to establish a clear causal relationship between them. Confounders can be positive or negative; that is, introducing a tendency to overestimate or underestimate a relationship between the two phenomena. Confounders should be prevented as much as possible by the design of a study or controlled for by adjusting for them after its completion, using information about confounding variables gathered during the study.

Druggability:
describes the extent to which a biological target (typically a protein) is known or predicted to bind to a drug with high affinity.

Expression quantitative trait locus (eQTL):
a locus containing DNA variation associated with variation in gene expression measured by mRNA levels.

Genome-wide association study (GWAS):
a hypothesis-free study design based on the scan of a dense genome-wide set of genetic variants (at present typically millions of polymorphisms) in large and statistically well-powered sample sets of individuals, aimed at finding statistically robust genetic associations with qualitative and/or discrete traits (e.g., disease risk) or quantitative traits (e.g., height or levels of specific molecules and cells in the blood).

Genome-wide significance:
a statistical threshold to reduce the number of false positive findings in GWAS studies. In particular, a significance threshold of 5×10^{-8} was proposed by the International HapMap Consortium in 2005 regardless of the actual variant density of the study. More stringent empirical thresholds have been proposed more recently and used to adjust for multiple testing of the actual number of independent variants assessed in sequencing-based GWAS.

Intermediate phenotype:
a measurable phenotype that controls a key pathogenic checkpoint along the chain of molecular events leading to disease. Here, we apply it to quantitative variable; for example, levels of immune system cells or of a metabolite that are genetically correlated with disease.

Mendelian randomization (MR):
a class of statistical methods ascertaining unbiased estimates of causal associations between exposures and outcomes based on measured variation in genes. The approach exploits the principle that genotypes are unchanged during life, unlike confounders, which can seriously bias standard observational epidemiological studies. Hence, MR was originally applied to assess the etiological role of exposures that are mirrored by variables under genetic control. In a broad sense, it represents any approach that uses genetic information to make inferences about the causal relation between traits.

Protein quantitative trait locus (pQTL):
a locus containing DNA variation associated with variation in protein levels. The regulatory DNA variants underlying eQTL and pQTL are defined as acting either in *cis* or in *trans*, depending on the physical distance from the gene they regulate (also referred to as local or distant eQTLs and pQTLs) mapping respectively near or far (‘far’ meaning typically on different chromosomes) compared with the location of the association signal.

Quantitative trait locus (QTL):
a segment (or ‘locus’) of the genome containing DNA variation associated with a quantitative trait (a phenotype, e.g., height, varying on a continuous scale).

Reverse causality:
refers to the direction of cause and effect, which can be contrary to a supposed direction in an observational study. For example, during the early 1980s, based on the observation that individuals with many forms of cancer have low serum levels of cholesterol, it was proposed that low cholesterol levels and diets and/or treatments designed to reduce cholesterol levels, could increase risk of cancer. Later, it became clear that the notion was spurious, because individuals with DNA changes leading to genetically inherited reduction of serum cholesterol have no increased risk of cancer compared with matched controls, a result also consistent with subsequent evidence from clinical trials with statins. Thus, lower cholesterol was a result of cancer (likely because large amounts of cholesterol are used to form cell membranes and intracellular structures of cancer cells) rather than increasing cancer risk. The rejection of the initial false supposition using genetic data forms the basis of MR.

References

1. Plenge RM et al. (2013) Validating therapeutic targets through human genetics. Nat. Rev. Drug Discov 12, 581–59423868113
2. Folkersen L et al. (2015) Applying genetics in inflammatory disease drug discovery. Drug Discov. Today 20, 1176–118126050580
3. Hackam DG et al. (2006) Translation of research evidence from animals to humans. JAMA 296, 1731–173217032985
4. Morgan P et al. (2018) Impact of a five-dimensional framework on R&D productivity at AstraZeneca. Nat. Rev. Drug Discov 17, 167–18129348681
5. Abifadel M et al. (2003) Mutations in PCSK9 cause autosomal dominant hypercholesterolemia. Nat. Genet 34, 154–15612730697
6. Cohen J et al. (2005) Low LDL cholesterol in individuals of African descent resulting from frequent nonsense mutations in PCSK9. Nat. Genet 35, 161–165
7. Cohen J et al. (2006) Sequence variations in PCSK9, low LDL, and protection against coronary heart disease. N. Engl. J. Med 354, 1264–127216554528
8. Sabatine MS et al. (2015) Efficacy and safety of evolocumab in reducing lipids and cardiovascular events. N. Engl. J. Med 372, 1500–150925773607
9. Cook D et al. (2014) Lessons learned from the fate of AstraZeneca's drug pipeline: a five-dimensional framework. Nat. Rev. Drug Discov 13, 419–43124833294
10. Nelson MR et al. (2015) The support of human genetic evidence for approved drug indications. Nat. Genet 47, 856–86026121088
11. Anon (2017) Online Mendelian Inheritance in Man, OMIM®, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University
12. MacArthur J et al. (2017) The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). Nucleic Acids Res. 45, D896–D90127899670
13. Ledford H (2016) AstraZeneca launches project to sequence 2 million genomes. Nature 532, 427
14. Plenge RM (2016) Disciplined approach to drug discovery and early development. Sci. Transl. Med 8, 349ps15
15. Shin SY et al. (2014) An atlas of genetic influences on human blood metabolites. Nat. Genet 46, 543–55024816252
16. Orrù V et al. (2013) Genetic variants regulating immune cell levels in health and disease. Cell 155, 242–25624074872
17. International HapMap Consortium (2005) A haplotype map of the human genome. Nature 437, 1299–132016255080
18. Sidore C et al. (2015) Genome sequencing elucidates Sardinian genetic architecture and augments association analyses for lipid and blood inflammatory markers. Nat. Genet 47, 1272–128126366554

19. Astle WJ et al. (2016) The allelic landscape of human blood cell trait variation and links to common complex disease. Cell 167, 1415–142927863252
20. Nica AC et al. (2010) Candidate causal regulatory effects by integration of expression QTLs with complex trait genetic associations. PLoS Genet. 6, e100089520369022
21. Giambartolomei C et al. (2014) Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet. 10, e100438324830394
22. Hormozdiari F et al. (2016) Colocalization of GWAS and eQTL Signals Detects Target Genes. Am. J. Hum. Genet 99, 1245–126027866706
23. Plagnol V et al. (2009) Statistical independence of the colocalized association signals for type 1 diabetes and RPS26 gene expression on chromosome 12q13. Biostatistics 10, 327–33419039033
24. Sudlow C et al. (2015) UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med. 12, e100177925826379
25. Huang H et al. (2017) Fine-mapping inflammatory bowel disease loci to single-variant resolution. Nature 547, 173–17828658209
26. Patsopoulos N et al. (2017) The Multiple Sclerosis Genomic Map: role of peripheral immune cells and resident microglia in susceptibility. bioRxiv 2017, 143933
27. Zavattari P et al. (2000) Major factors influencing linkage disequilibrium by analysis of different chromosome regions in distinct populations: demography, chromosome recombination frequency and selection. Hum. Mol. Genet 9, 2947–295711115838
28. Lim ET et al. (2014) Distribution and medical impact of loss-of-function variants in the Finnish founder population. PLoS Genet. 10, e100449425078778
29. Gudbjartsson DF et al. (2015) Large-scale whole-genome sequencing of the Icelandic population. Nat. Genet 47, 435–44425807286
30. Danjou F et al. (2015) Genome-wide association analyses based on whole-genome sequencing in Sardinia provide insights into regulation of hemoglobin levels. Nat. Genet 47, 1264–127126366553
31. Zoledziewska M et al. (2015) Height-reducing variants and selection for short stature in Sardinia. Nat. Genet 47,1352–135626366551
32. Low-Kam C et al. (2016) Whole-genome sequencing in French Canadians from Quebec. Hum. Genet 135, 1213–122127376640
33. Steri M et al. (2017) Overexpression of the cytokine BAFF and autoimmunity risk. N. Engl. J. Med 376, 1615–162628445677
34. Spain SL et al. (2015) Strategies for fine-mapping complex traits. Hum. Mol. Genet 24, R111–R11926157023
35. Zaitlen N et al. (2010) Leveraging genetic variability across populations for the identification of causal variants. Am. J. Hum. Genet 86, 23–3320085711
36. Asimit JL et al. (2016) Trans-ethnic study design approaches for fine-mapping. Eur. J. Hum. Genet 24, 1330–133626839038
37. Kircher M et al. (2014) A general framework for estimating the relative pathogenicity of human genetic variants. Nat. Genet 46, 310–31524487276
38. lonita-Laza I et al. (2016) A spectral approach integrating functional genomic annotations for coding and noncoding variants. Nat. Genet 48, 214–22026727659
39. Farh KKH et al. (2015) Genetic and epigenetic fine mapping of causal autoimmune disease variants. Nature 518, 337–34325363779
40. Davison LJ et al. (2012) Long-range DNA looping and gene expression analyses identify DEXI as an autoimmune disease candidate gene. Hum. Mol. Genet 21, 322–33321989056
41. Leikfoss IS et al. (2013) Multiple sclerosis-associated single-nucleotide polymorphisms in CLEC16A correlate with reduced SOCS1 and DEXI expression in the thymus. Genes Immun. 14, 62–6623151489
42. Smemo S et al. (2014) Obesity-associated variants within FTO form long-range functional connections with IRX3. Nature 507, 371–37524646999
43. Javierre BM et al. (2016) Lineage-specific genome architecture links enhancers and non-coding disease variants to target gene promoters. Cell 167, 1369–138427863249

44. Bryoist J et al. (2014) Cis and trans effects of human genomic variants on gene expression. PLoS Genet. 10, e100446125010687
45. Yao C et al. (2017) Dynamic role of trans regulation of gene expression in relation to complex traits. Am. J. Hum. Genet 100, 571–58028285768
46. The GTEx Consortium (2017) Genetic effects on gene expression across human tissues. Nature 550, 204–21329022597
47. Streeter I et al. (2017) The human-induced pluripotent stem cell initiative-data resources for cellular genetics. Nucleic Acids Res. 45, D691–9727733501
48. Kilpinen H et al. (2017) Common genetic variation drives molecular heterogeneity in human iPSCs. Nature 546, 370–37528489815
49. Grubert F et al. (2015) Genetic control of chromatin states in humans involves local and distal chromosomal interactions. Cell 162, 1051–106526300125
50. Wang Y et al. (2017) The 3D Genome Browser: a web-based browser for visualizing 3D genome organization and long-range chromatin interactions. bioRxiv 2017, 112268
51. Gao X et al. (2017) The impact of methylation quantitative trait loci (mQTLs) on active smoking-related DNA methylation changes. Clin. Epigenetics 9, 8728824732
52. Sun BB et al. (2017) Consequences of natural perturbations in the human plasma proteome. bioRxiv 2017, 134551
53. Fairfax BP et al. (2012) Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles. Nat. Genet 44, 502–51022446964
54. Ye CJ et al. (2014) Intersection of population variation and autoimmunity genetics in human T cell activation. Science 345, 125466525214635
55. Roederer M et al. (2015) The genetic architecture of the human immune system: a bioresource for autoimmunity and disease pathogenesis. Cell 161, 387–40325772697
56. Patin E et al. (2018) Natural variation in the parameters of innate immune cells is preferentially driven by genetic factors. Nat. Immunol 19, 302–31429476184
57. Hackinger S et al. (2017) Statistical methods to detect pleiotropy in human complex traits. Open Biol. 7, 17012529093210
58. Katan MB et al. (1986) Apolipoprotein E isoforms, serum cholesterol, and cancer. Lancet 8479, 507–508
59. Haycock PC et al. (2016) Best (but oft-forgotten) practices: the design, analysis, and interpretation of Mendelian randomization studies. Am. J. Clin. Nutr 103, 965–97826961927
60. Di Angelantonio E et al. (2009) Major lipids, apolipoproteins, and risk of vascular disease. J. Am. Med. Assoc 302, 1993–2000
61. Voight BF et al. (2012) Plasma HDL cholesterol and risk of myocardial infarction: a mendelian randomisation study. Lancet 380, 572–58022607825
62. Holmes MV et al. (2015) Mendelian randomization of blood lipids for coronary heart disease. Eur. Heart J 36, 539–55024474739
63. Keene D et al. (2014) Effect on cardiovascular risk of high density lipoprotein targeted drug treatments niacin, fibrates, and CETP inhibitors: meta-analysis of randomised controlled trials including 117,411 patients. Br. Med. J 349, g437925038074
64. Yoshimini K et al. (2016) ssODN-mediated knock-in with CRISPR-Cas for large genomic regions in zygotes. Nat. Commun 7, 1043126786405
65. Mir SA et al. (2016) Analysis and validation of traits associated with a single nucleotide polymorphism Gly364Ser in catestatin using humanized chromogranin A mouse models. J. Hypertens 34, 68–7826556564
66. Suzuki K et al. (2016) In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration. Nature 540, 144–14927851729
67. Hindorff LA et al. (2018) Prioritizing diversity in human genomics research. Nat. Rev. Genet 19, 175–18529151588
68. Hauser AS et al. (2018) Pharmacogenomics of GPCR Drug Targets. Cell 172, 41–5429249361
69. Perdigoto C (2017) Mutations: dawn of the Human Knockout Project. Nat. Rev. Genet 18, 328–329

70. Saleheen D et al. (2017) Human knockouts and phenotypic analysis in a cohort with a high rate of consanguinity. Nature 544, 235–23928406212
71. Bartha I et al. (2018) Human gene essentiality. Nat. Rev. Genet 19, 51–6229082913
72. Lek M et al. (2016) Analysis of protein-coding genetic variation in 60,706 humans. Nature 536, 285–29127535533
73. De Boever C et al. (2017) Medical relevance of protein truncating variants across 337,208 individuals in the UK Biobank study. bioRxiv 2017, 179762
74. Diogo D et al. (2017) Phenome-wide association studies (Phe-WAS) across large 'real-world data' population cohorts support drug target validation. bioRxiv 2017, 218875
75. Yi X et al. (2010) Sequencing of 50 human exomes reveals adaptation to high altitude. Science 329, 75–7820595611
76. Voight BF et al. (2006) A map of recent positive selection in the human genome. PLoS Biol. 4, e7216494531
77. Nielsen R et al. (2007) Recent and ongoing selection in the human genome. Nat. Rev. Genet 8, 857–86817943193
78. Zumla A et al. (2016) Host-directed therapies for infectious diseases: current status, recent progress, and future prospects. Lancet Infect. Dis 16, e47–e6327036359
79. Hopkins AL et al. (2002) The druggable genome. Nat. Rev. Drug Discov 1, 727–73012209152
80. Russ AP et al. (2005) The druggable genome: an update. Drug Discov. Today 10, 1607–161016376820
81. Kumar RD et al. (2013) Prioritizing potentially druggable mutations with dGene: an annotation tool for cancer genome sequencing data. PLoS One 8, e6798023826350
82. Finan C et al. (2017) The druggable genome and support for target identification and validation in drug development. Sci. Transl. Med 9, eaag116628356508
83. Nguyen DT et al. (2017) Pharos: collating protein information to shed light on the druggable genoma. Nucleic Acids Res. 45, D995–D100227903890
84. Koscienlny G et al. (2017) Open Targets: a platform for therapeutic target identification and validation. Nucleic Acids Res. 45, D985–D99427899665
85. Liang F et al. (2017) Efficient targeting and activation of antigen-presenting cells in vivo after modified mRNA vaccine administration in rhesus macaques. Mol. Ther 25, 2635–264728958578
86. Bosley K (2017) Nat. Rev. Drug Discov 16, 672–67628935913
87. Lai AC et al. (2017) Induced protein degradation: an emerging drug discovery paradigm. Nat. Rev. Drug Discov 16, 101–11427885283
88. Joo YB et al. (2017) Biological function integrated prediction of severe radiographic progression in rheumatoid arthritis: a nested case control study. Arthritis. Res. Ther 19, 24429065906
89. Hensman Moss DJ et al. (2017) Identification of genetic variants associated with Huntington's disease progression: a genome-wide association study. Lancet Neurol. 16, 701–71128642124
90. Medina-Kauwe LK (2013) Development of adenovirus capsid proteins for targeted therapeutic delivery. Ther. Deliv 4, 267–27723343164
91. Casi G et al. (2015) Antibody-Drug Conjugates and Small Mole-cule-Drug Conjugates: Opportunities and Challenges for the Development of Selective Anticancer Cytotoxic Agents. J. Med. Chem 58, 8751–876126079148
92. Hall J et al. (2016) Delivery of therapeutic proteins via extracellular vesicles: Review and potential treatments for Parkinson's disease, glioma and schwannoma. Cell. Mol. Neurobiol 36, 417–42727017608

Highlights

Analysis of the phenotypic consequences of human genetic variation is an unbiased tool to reveal molecules primarily involved in disease occurrence.

A systematic search for coincident genetic associations between human multifactorial disease risk and quantitative phenotypes using genome-wide association studies (GWAS) findings is becoming an especially powerful approach to identify intermediate phenotypes controlling key checkpoints in disease pathogenesis that can be modulated therapeutically.

Criteria are suggested to select optimal targets taking advantage of GWAS results and of various sources of functional and gene expression annotation.

Main emerging challenges to the use of genetic data for therapeutic exploitation are also discussed.

Outstanding Questions

An increasing number of coincident genetic associations with disease risk and quantitative traits will become available in the coming years. Will the creation of megaclinical resources for genetic research and the extension of the current approaches to identify multiple independent associations (with the same trait and disease) make it possible to solve pleiotropic effects and restrict discoveries to real disease-related intermediate phenotypes and causal therapeutic targets?

What will be the final impact of the improved identification of therapeutic targets based on human genetically driven causal biology in the repurposing of existing drugs and development of new ones?

How can naturally occurring human genetic variation combined with engineered *in vitro* and *in vivo* models help estimate the optimal target dose and therapeutic window for a drug?

As observed in some effective oncological therapies, will we also see in other areas, such as autoimmunity, advances in therapeutic methods based on cell-specific delivery that will accommodate the observed cellular effects on target expression suggested by GWAS findings?

Will combination therapies strategically targeted at different disease-related intermediate phenotypes and pathways have a future for the treatment of multifactorial traits, such as in autoimmune diseases?

Will relevant genetic data that could be used to compare the amelioration of disease course (through longitudinal and retrospective case-only studies) to odds of developing disease in the first place (through case-control studies) become available and impact the development of new drugs?

Key Figure

Impact of a Genetic-Driven Approach on the Drug Research and Development Pipeline

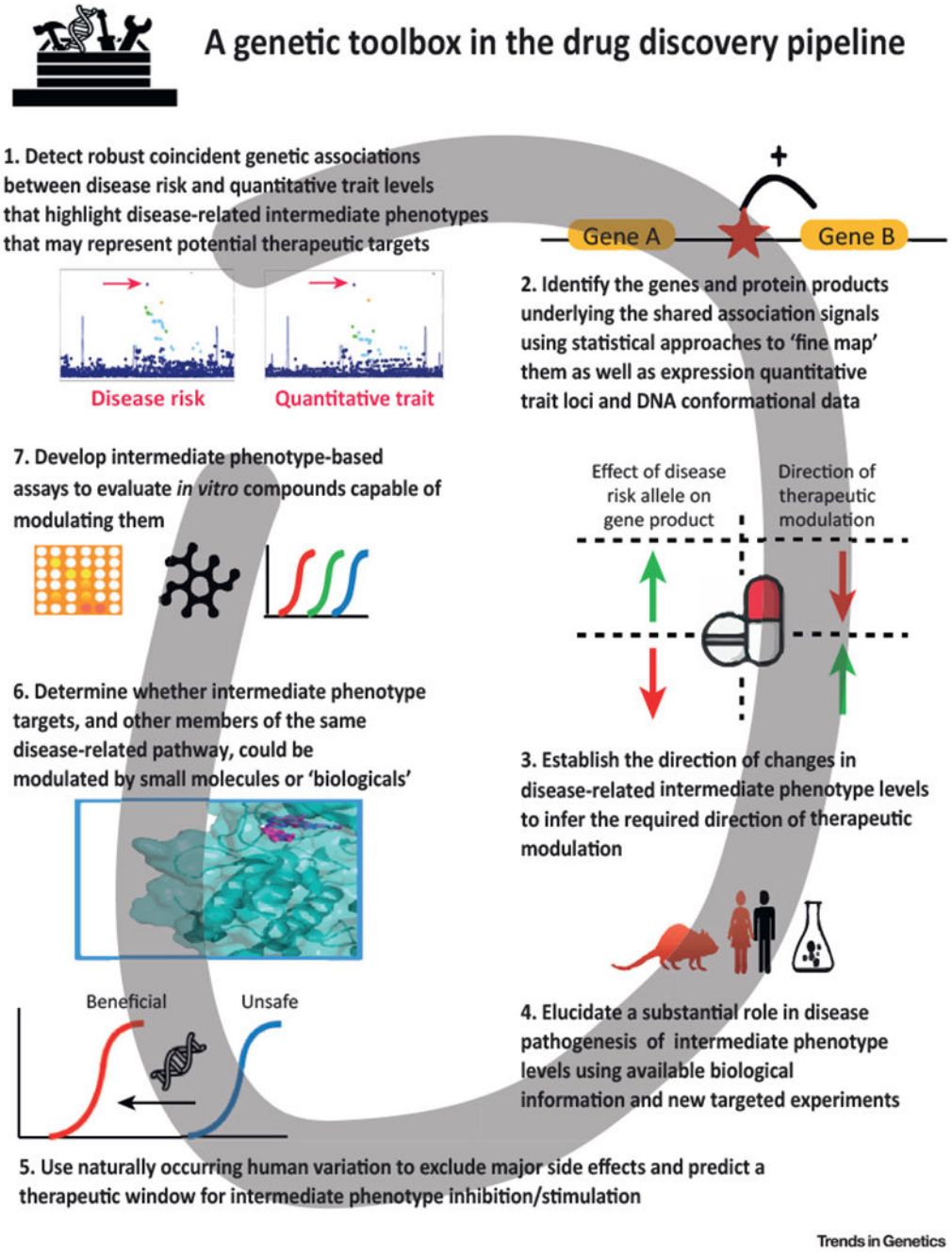

Figure 1.

A schematic roadmap for taking advantage of genetic discoveries to identify drug targets is shown.

Table 1.

Relative frequency of total unique associations ($P \leq 5E^{-8}$) for each disease or trait category in the GWAS catalog defined by the released experimental factor ontology (EFO) terms[a]

Category	%	Top Diseases/Traits in Category
Disease		
Immune system disorders	32.50	Crohn's disease 16.57%; inflammatory bowel disease 15.82%; ulcerative colitis 10.18%; systemic lupus erythematosus 9.51%; rheumatoid arthritis 9.43%; psoriasis 5.56%; atopic eczema 4.18%; multiple sclerosis 4.99%
Cancer	21.42	Prostate carcinoma 15.09%; breast carcinoma 12.4%; colorectal cancer 6.89%
Neurological disorders	19.48	Schizophrenia 41.28%; Alzheimer's disease 10.34%; Parkinson's disease 6.91%
Other diseases	11.34	Asthma 10.52%; influenza A H1N1 9.39%; chronic obstructive pulmonary disease 8.82%
Cardiovascular diseases	7.,56	Coronary artery disease 19.69%; atrial fibrillation 14.94%; coronary heart disease 13.75%
Digestive system disorders	3.50	Primary biliary cirrhosis 32.6%; Barrett's esophagus 8.79%; sclerosing cholangitis 7.33%
Metabolic disorders	4.,19	Type 2 diabetes mellitus 77.98%; obesity 6.17%; metabolic syndrome 6.17%
Traits		
Hematological measurements	32.69	Reticulocyte count 14.16%; platelet count 10.29%; hemoglobin measures 9%
Other measurements	30.73	Anthropometry 12.85%; SPARC-like protein 1 measurement 8.76%; FEV/FEC ratio 6.96%
Body measurements	12.54	Body mass index (BMI) 55.12%; waist circumference 15.5%; BMI-adjusted waist circumference 13.14%
Lipid or lipoprotein measurements	8.13	High-density lipoprotein cholesterol measurement 25.84%; total cholesterol measurement 20.1%; triglyceride measurement 16.87%
Biological process	3.75	Smoking behavior 62.07%; circadian rhythm 9.36%; intelligence 4.19%
Cardiovascular measurements	3.55	QT interval 13.67%; heart function measurement 13.15%; pulse pressure measurement 12.5%
Inflammatory measurements	3.01	Basophil count 67.69%; basophil percentage of white cells 9.95%; basophil percentage of granulocytes 9.49%
Other traits	2.93	Coronary artery calcification 14.51%; cleft palate 10.41%; myopia 8.04%
Response to drug	2.23	Response to bronchodilator 63.01%; response to vaccine 12.81%; response to anticoagulant 2.69%
Liver enzyme *measurements*	0.45	Serum gamma-glutamyl transferase measurement 38.14%; aspartate aminotransferase measurement 20.61%; serum alanine aminotransferase measurement 19.59%

[a]For example, associations explaining 'Response to drug' account for 2.23% of all unique SNP and/or trait pairs; in the context of this category, 'response to bronchodilator' is currently the most studied trait (63.01% of all 'Response to drug' associations).

Table 2.

Challenges to Genetic-Driven Target Discovery and Validation

Challenge	Practical Response
Resolving association signals of multifactorial traits to causal gene and variant and understanding underlying mechanism	Sequencing-based fine-mapping, including cross-population and cross-phenotype analyses; eQTL and pQTL analyses, along with examination of epigenetic and chromatin profiles and promoter capture Hi-C data; informative studies of *in vitro* and *in vivo* models, such as those based on genome-editing approaches, to better understand functional consequences of human variation in health and disease
Accounting for pleiotropic effects to restrict coincident genetic associations to truly disease-related intermediate phenotypes and therapeutic targets based on causal human biology	Assembling very large clinical resources comprehensively characterized for quantitative traits potentially relevant to disease of interest to identify multiple independent coincident associations with same specific trait
Estimating optimal target dose and therapeutic window for a drug	Expanding understanding of phenotypic consequences of naturally occurring genetic variation in diverse populations, including some poorly examined thus far (e.g., African populations); and constructing engineered informative *in vitro* and *in vivo* models to expand range of effects that can be assessed
Distinguishing targets relevant for disease prevention and disease treatment	Performing large powerful case-only GWAS to study disease progression and severity, and examining standard GWAS and available biological information to discriminate targets active in disease occurrence from those that could be effective in ameliorating therapies
Addressing cellular and temporal effects in therapeutic delivery	Generating new modalities to ensure targeted, cell-specific therapeutic delivery