



Advancing the use of genome-wide association studies for drug repurposing

William R. Reay^{1,2} and Murray J. Cairns^{1,2}

Abstract | Genome-wide association studies (GWAS) have revealed important biological insights into complex diseases, which are broadly expected to lead to the identification of new drug targets and opportunities for treatment. Drug development, however, remains hampered by the time taken and costs expended to achieve regulatory approval, leading many clinicians and researchers to consider alternative paths to more immediate clinical outcomes. In this Review, we explore approaches that leverage common variant genetics to identify opportunities for repurposing existing drugs, also known as drug repositioning. These approaches include the identification of compounds by linking individual loci to genes and pathways that can be pharmacologically modulated, transcriptome-wide association studies, gene-set association, causal inference by Mendelian randomization, and polygenic scoring.

Polygenic

A term that denotes the contribution of many genes to the genetic component of a trait.

Genome-wide association studies

(GWAS). Studies using a design that tests the association (relationship) between sequence nucleotide alterations (genetic variants) throughout the genome with a trait of interest, such as a disease phenotype.

Complex disorders constitute the most substantial contribution to the healthcare burden and are aetiologically influenced by a polygenic genetic architecture, along with a range of environmental risk factors¹. The pathophysiology of these phenotypes cumulatively affects practically all organs and physiological systems within the human body. Although drug discovery is viewed as a vital endeavour to improve the treatment and management of these disorders, the difficulty of achieving novel drug development and regulatory approval has increased considerably owing to both the duration and the costs involved, which have led to a decrease in the overall success rate of translating discoveries to clinical practice^{2,3}.

The challenges facing the pharmaceutical industry have prompted an increasing focus on drug repurposing, also known as drug repositioning. Drug repurposing involves the use of existing compounds with approval for use in a particular disorder for new clinical indications, such that these drugs can be deployed clinically with greater speed and less expense compared with the de novo drug discovery pipeline. While the rationale and regulatory implications of drug repurposing have been discussed extensively elsewhere^{2,4,5}, the concept broadly encompasses both ‘on label’ repurposing, in which repositioned compounds gain regulatory approval for their new indication, along with ‘off label’ usage, whereby an approved compound is implemented for a use not explicitly defined in its marketing authorization⁶. Currently marketed drugs have already shown sufficient safety and efficacy for a particular phenotype by virtue of their regulatory approval; however, drug repurposing necessitates

that candidate compounds demonstrate clinical benefit for a new indication. As a result, improving our understanding of the pathophysiology of complex human disorders will expedite prioritizing suitable drugs, which is often the most onerous component of the repurposing pipeline.

Advances in genome-wide association studies (GWAS) have proven useful for drug repurposing as GWAS reveal important biological insights into complex traits that can assist to identify compounds suitable for repurposing. Specifically, GWAS have identified thousands of variants rigorously associated with clinically relevant phenotypes, yielding insight into genes and pathways in human health and disease, as well as the pleiotropy and genetic overlap displayed between different traits^{1,7,8}. Traditionally, GWAS have focused on determining the role of common variants in the population due to the greater power afforded by common loci and relative ease of their imputation. Whereas common variants tend to mostly have relatively small individual effects on traits, the combined polygenic effect of many such variants is considerably larger and constitutes a significant portion of overall trait heritability^{9–13}. This progress has been aided by the immense efforts of the statistical genetics community through the development of sophisticated methods to integrate GWAS with other ‘omics’ data sets and relate association signals to genes and biological networks. As a result, the unprecedented availability of GWAS data and post-GWAS analysis techniques provides a valuable opportunity to utilize common variant genetics for drug repurposing.

In recent years, an increasing number of methods for systematically prioritizing drug repurposing opportunities

¹School of Biomedical Sciences and Pharmacy, The University of Newcastle, Callaghan, New South Wales, Australia.

²Centre for Brain and Mental Health Research, Hunter Medical Research Institute, Newcastle, New South Wales, Australia.

✉e-mail: murray.cairns@newcastle.edu.au
<https://doi.org/10.1038/s41576-021-00387-z>

Pleiotropy

A term to denote the influence of a gene or genetic variant on multiple different biological traits.

Imputation

Using genetic variants to predict (impute) a particular variable.

Heritability

The proportion of variance in a phenotype in a population that is explained by genetic variation.

Gene-set association

A technique that examines whether a set of genes is associated with a trait by combining the association of individual genetic variants within the set.

have been proposed, suggesting growing interest in more immediate clinical translation. In this Review, we discuss various current approaches that could be leveraged to allow GWAS to inform drug repurposing, including the mapping of genome-wide significant loci to drug targets, transcriptomic imputation, gene-set association, causal inference and polygenic scoring. We outline strengths and limitations of these methods, and highlight the potential for exploiting GWAS variants to identify causal relationships between druggable exposures and complex disorders. Finally, we discuss how pharmacologically orientated polygenic scoring may aid in both the discovery and the precision use of both currently used and repurposed compounds, and provide future directions for the field.

GWAS signals implicating drug targets

The most common interpretation of GWAS is to identify genetic variants associated with the phenotype in question after the application of multiple-testing correction.

Traditionally, variants are designated ‘genome-wide significant’ if the association *P* value is below 5×10^{-8} (REFS^{14,15}). From the outset of large-scale GWAS, it was suggested that genes affected by trait-associated variation may provide opportunities for drug development or drug repurposing. Specifically, if a genome-wide significant variant is plausibly mapped to a gene modulated by a known pharmacological agent, this signal could be leveraged to repurpose a drug that is known to target the gene. In the following section, we describe a notable example of this methodology whereby a GWAS signal has informed drug repurposing for Crohn’s disease (FIG. 1). Crohn’s disease is thus far the most complete case study of drug repurposing informed by a single genome-wide significant variant, as the candidate compound has now been officially approved for this new clinical indication. However, there are other examples of single GWAS locus-driven drug repurposing candidates at earlier stages; for instance, genome-wide significant

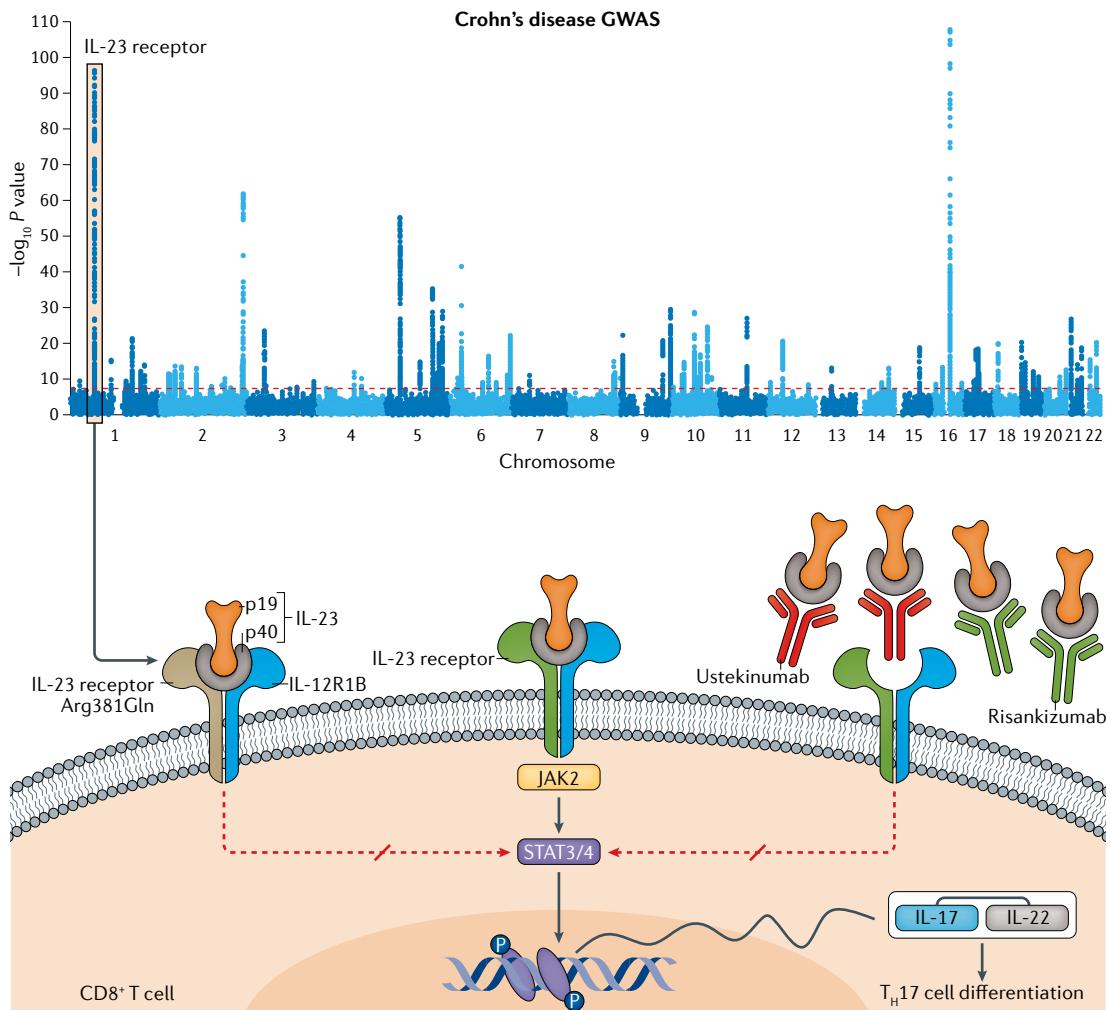


Fig. 1 | Genome-wide significant variants associated with Crohn’s disease spanning the IL-23 receptor provide drug repurposing opportunities. A missense variant in *IL23R* (rs11209026, p.Arg381Gln) that disrupts its function demonstrates rigorous association with Crohn’s disease and confers a protective effect (left pathway) by reducing T helper 17 cell (T_H17 cell) lymphocyte subset-associated autoimmunity (central pathway). As a result, monoclonal antibodies that target the interleukin-23 (IL-23) receptor (right pathway) were repositioned as Crohn’s disease interventions from their original indication for psoriasis. The signalling consequences of genomic and pharmacological IL-23 suppression on T_H17 cell differentiation of a CD8⁺ T cell population is visualized. GWAS, genome-wide association studies.

Table 1 | Examples of GWAS signals mapped to the target of a currently indicated pharmacological agent for the disease trait

Trait	Gene ^a	Drug or drug class	Ref.
Schizophrenia	DRD2	Antipsychotic drugs	¹²⁴
Type 2 diabetes mellitus	PPARG	Thiazolidinediones	¹⁰⁷
Breast cancer	ESR1	Tamoxifen	¹²⁵
Bone mineral density	RANKL	Denosumab (RANKL inhibitor)	¹²⁶
Rheumatoid arthritis	IL6R	Tocilizumab (IL-6 inhibitor)	¹²⁷
Blood pressure	SLC12A2	Loop diuretics	¹²⁸
Psoriasis	IL23R	Ustekinumab	¹²⁹
Low-density lipoprotein	HMGCR	Statins	¹³⁰
Generalized epilepsy	GRIK1	Carbonic anhydrase inhibitors	⁵⁹
Atrial fibrillation	SCN5A	Antiarrhythmics, class 1a	¹³¹

GWAS, genome-wide association studies; IL-6, interleukin-6. ^aThe named genes were implicated based on a genome-wide significant single-nucleotide polymorphism (SNP) spanning the region of that gene or proximally to it. This does not necessarily imply that this gene harbours causal variation without further follow-up analyses to that effect.

variation spanning the aromatase gene *CYP19A1* in endometrial cancer has suggested that aromatase inhibitors can be repurposed for its treatment^{16–18}.

IL-23 as a repurposing candidate in Crohn's disease.

Crohn's disease is an inflammatory bowel disease that exhibits a strong contribution of common variation to its genomic architecture, with a single-nucleotide polymorphism (SNP)-based heritability estimate in the region of 25% and more than 200 genome-wide significant variants identified to date^{7,19}. One of the first genomic regions associated with Crohn's disease was the *IL23R* gene, which encodes the interleukin-23 (IL-23) receptor, with this signal replicated in larger subsequent GWAS^{19,20}. This locus contains two independent signals: a relatively uncommon missense variant (rs11209026, p.Arg381Gln, population frequency ~5%), which is protective for the disorder, and higher-frequency regulatory variation that spans intronic regions. IL-23 is a pro-inflammatory cytokine essential for the differentiation of the T helper 17 cell ($T_{H}17$ cell) lymphocyte subset^{21,22}. GWAS data were one of the central impetuses to explore the targeting of IL-23 as a drug repurposing candidate for Crohn's disease, particularly as the putatively disruptive rs11209026 missense variant in the IL-23 receptor has a protective effect, which would correspond to pharmacological inhibition of this gene being of clinical value^{23–26}. Two central monoclonal antibodies modulating IL-23 signalling were proposed as repurposing candidates for Crohn's disease: ustekinumab and risankizumab. Both these treatments were first used in the treatment of psoriasis: risankizumab targets IL-23 specifically via its p19 subunit, whereas ustekinumab targets the p40 subunit shared with IL-12 (REFS^{27–29}). At the time of writing, ustekinumab has been approved for the treatment of moderate to severe Crohn's disease in several jurisdictions including the United States, Europe and Australia. This regulatory approval was underpinned by a series of successful trials of the compound demonstrating benefit in Crohn's disease^{23,24,26}, with trials citing genetic

evidence from GWAS as a motivating factor for repurposing this psoriasis treatment. Data for risankizumab are also promising; however, trials are at a more preliminary stage. Thus far, both a randomized, double-blind, placebo-controlled phase II study and an open-label extension study have supported its utility^{25,30}. This case study of genetic variation physically mapped to *IL23R* and Crohn's disease provides an interesting framework for how individual GWAS loci can be leveraged to reposition existing therapeutic strategies. Moreover, GWAS signals have been mapped to genes that represent currently indicated drugs for numerous complex traits, providing additional support that genome-wide significant variants may be therapeutically actionable. Illustrating the range and diversity of phenotypes affected, we highlight ten examples for which this phenomenon has been observed (TABLE 1); however, many other such examples exist³¹. It should also be noted that support from GWAS does not always indicate that a drug targeting a specific gene will be efficacious, and limitations to this approach are discussed in the next section. For instance, SNPs mapped to the vitamin D receptor-encoding gene *VDR* have been associated with hypertension, but vitamin D supplementation has been shown to have inconclusive to null effects on blood pressure in randomized controlled trials^{7,32,33}.

Limitations of single GWAS locus-informed drug repurposing.

There are numerous limitations to this 'single locus' approach to inform drug repurposing. The vast majority of common variant GWAS signals have small individual effect sizes, which may not be substantial enough for therapeutic intervention. Notable exceptions include the association between the ε4 haplotype in *APOE* and late-onset Alzheimer disease and some GWAS signals for biologically simpler traits such as circulating metabolites, for which the magnitudes of SNP-trait associations are considerably larger³⁴. However, it has been argued that whilst the effect size of GWAS SNPs on the phenotype of interest may be low in most cases, their impact on molecular phenotypes, such as gene expression, may be large enough to be a relevant proxy for treatment¹. For instance, many common variants exert a regulatory effect on gene expression, acting as expression quantitative trait loci (eQTLs); thus, whereas the effect allele may only slightly elevate or decrease the disease risk, a drug that strongly affects the expression of that gene may be a suitable repurposing candidate. An important consideration when studying trait-associated eQTLs is the coherence between the sign of the eQTL and the GWAS effect, whereby genes may be upregulated in diseased cases despite the GWAS risk allele acting as an eQTL that downregulates expression, or vice versa (that is, incoherence)³⁵. This highlights the complexity of the relationship between regulatory variation and disease, as GWAS risk alleles may mechanistically exert their role in disease pathogenesis through impairment of a protective response by modulating gene expression in the opposite direction to that of the protective mechanism. In addition, small SNP-trait effect sizes of GWAS signals do not necessarily predict the efficacy of a compound targeting a prioritized gene. The monoclonal

Single-nucleotide polymorphism (SNP). A single-nucleotide alteration in the genomic sequence at any given position (locus).

Quantitative trait loci
Genetic variants or intervals that are linked to or associated with a quantitative trait (measurable continuous phenotype); for example, expression quantitative trait loci (eQTLs) are variants associated with mRNA expression for a given gene.

Linkage disequilibrium
Genetic variants that are inherited together at a higher rate than by chance alone are said to exhibit linkage disequilibrium.

Fine-mapping

Investigating which genetic variant or variants within a region of the genome significantly associated with a trait (genome-wide association study (GWAS) locus) causally influence the trait in question, rather than merely being inherited with the causal variant(s) through linkage disequilibrium.

Transcriptome-wide association study

(TWAS). A technique that tests the association between the predicted expression of a gene based on genetic variants from expression quantitative trait loci (eQTLs) analysis in an independent cohort and a trait of interest.

antibody denosumab, which is indicated for osteoporosis, is an example of this phenomenon. Denosumab was biologically supported after its approval by the discovery of genome-wide significant variants mapping to its target gene, *RANKL*, in a GWAS of estimated bone mineral density³⁶; the effect size of common *RANKL* SNPs on bone mineral density is small, but denosumab is nonetheless an effective osteoporosis treatment³⁷. Nevertheless, it is likely that additional biological evidence for the drug target in question would be required to initiate drug repurposing, as was the case for IL-23 and Crohn's disease given that dysregulated IL-23 signalling had supporting data from *in vivo* and *in vitro* studies^{38,39}. This additional evidence is particularly important as hundreds, if not thousands, of genes can be plausibly implicated in highly polygenic traits, and it remains unknown how many of these associations would be clinically actionable, with sufficient efficacy.

Furthermore, the haplotypes encompassing genome-wide significantly associated variants usually only span intronic or intergenic regions, meaning the interpretation of these variants is often inherently more difficult than that of a locus with non-synonymous variation. The assumption that the gene closest to the lead SNP is causally affected by trait-associated variation is also not necessarily valid, although it has been shown for some phenotypes that the closest gene is likely causal^{40,41}. In any given trait-associated locus, the patterns of

linkage disequilibrium are often complex, with true causal variants difficult to distinguish from the set of variants in high linkage disequilibrium with one another⁴². These haplotypes can extend across large genomic distances and, thus, may encompass several genes besides the putative drug target⁴³. Moreover, variants that modulate regulatory elements may exert a distal effect due to chromatin interactions in 3D space, as has been revealed by techniques such as Hi-C (high-throughput chromosome conformation capture)⁴⁴. The true causal variants in GWAS loci physically mapped to targets of drug repurposing candidates would ideally need to be refined in order for this approach to be effective, particularly in 'complex' loci where large haplotypes implicate numerous proximal genes and intergenic regions. Statistical fine-mapping of GWAS signals utilizing both frequentist and Bayesian approaches has been proposed as a framework to identify putative causal SNPs⁴². Further work is required to integrate functional annotation with credible causal variants to identify potential drug repurposing opportunities from single GWAS loci, particularly as the mechanism of variant effect is helpful to inform whether pharmacological activation or inhibition of a gene would be clinically beneficial. However, given that most drugs target genes rather than variants themselves, identifying causal genes without a precise understanding of the underlying mechanism of variant effect can still be useful. Biological support from functional genomics techniques along with loss-of-function interventions, such as gene knockout models, can support GWAS-prioritized genes for repurposing opportunities, although we acknowledge that some complex disease phenotypes are not recapitulated well by *in vitro* or *in vivo* models.

Box 1 | Transcriptome-wide association studies

A transcriptome-wide association study (TWAS) seeks to identify genes for which expression is associated with a trait, without directly measuring expression in the genome-wide association studies (GWAS) sample through approaches such as RNA sequencing. TWAS leverage the fact that genetic variants tend to explain a notable proportion of the variance in mRNA expression, and thus genetic variants can be used to construct models of genetically regulated expression (GReX; transcriptomic imputation). Specifically, this can be expressed as a linear model where gene expression (y) is associated with the effect (β) of single-nucleotide polymorphism (SNP) i covaried for k technical confounders (C), along with the conventional random error term (ϵ):

$$y \sim \beta_0 + \beta_k C_k + \beta_i \text{SNP}_i + \epsilon.$$

The estimated SNP effects on expression are usually optimized in a multivariate GReX prediction model. For illustration, we outline the FUSION TWAS framework as one example, but the underlying principles are fairly applicable to other methods⁴⁶. FUSION constructs GReX using four competing model types (least absolute shrinkage and selection operator (LASSO), elastic net, best linear unbiased prediction (BLUP) and Bayesian sparse linear mixed models (BSLMM)), and selects the SNP weight configuration that explains the most variance in gene expression. The most predictive GReX model is then utilized to correlate predicted (imputed) expression with the GWAS trait of interest. Consider the SNP weights (effects) from the GReX model on the expression of gene G as ω_G (ω_G' being the transpose; these weights are also adjusted for linkage disequilibrium) and the corresponding marginal effects of the SNP weights on the GWAS trait ($Z_{\text{GWAS}} = \beta_{\text{SNP}} / SE_{\text{SNP}}$). Therefore, the association between GReX and the GWAS trait (Z_{TWAS}) can be defined as:

$$Z_{\text{TWAS}} = \frac{\omega_G' Z_{\text{GWAS}}}{\sqrt{\text{var}(\omega_G' Z_{\text{GWAS}})}}$$

The TWAS test statistic can be thought of as a measure of genetic covariance and the trait of interest, and under conventional circumstances this informs the expression–trait relationship. In other words, a negative TWAS test statistic implies negative genetic covariance between expression and the trait of interest, meaning that downregulated expression is associated with an increase in the GWAS trait, and vice versa.

Transcriptomic imputation for repurposing

Interpreting the biological consequences of variants tested in GWAS has largely focused on their relationship with cellular read-outs that influence gene function, such as mRNA expression, protein abundance and epigenetic modifications. An important method developed to integrate functional genomics with GWAS that is increasing in popularity is the concept of transcriptome-wide association studies (TWAS)^{45,46}. TWAS leverage the *cis*-acting genetic component that influences gene expression by constructing multivariate models that seek to predict mRNA expression with genetic variants (BOX 1). SNP weights from these models of genetically regulated expression (GReX) are correlated with the corresponding effects of these SNPs in the GWAS of interest in order to calculate a measure of genetic covariance between the GReX model (gene expression) and the GWAS phenotype. As a result, the correlation between increased or decreased genetically predicted expression of a gene and the GWAS trait can be estimated. This approach can be exploited to inform drug repurposing as it gives insight into a putative relationship between expression and the trait; for instance, if increased genetically predicted expression of a particular gene is correlated with a disorder, then an antagonist of that gene may be useful, and vice versa. We recently used this approach for GWAS of lung function to propose compounds for

which the drug–gene interaction may enhance lung function based on the direction of the TWAS signal⁴⁷. Interestingly, the underlying principle of the TWAS has been extended to construct genetically regulated models of splicing⁴⁸, methylation⁴⁹ and protein expression^{50,51}, as well as the inclusion of *trans*-acting variants⁵². GReX weights for the TWAS have been derived from various tissues throughout the body, which has facilitated single-tissue and multi-tissue approaches for estimating these models^{53,54}. Indeed, weights are publicly available for each tissue from the Genotype–Tissue Expression (GTEx) Consortium^{46,55}, along with larger data sets specific to certain tissues such as blood⁴⁶, brain⁵⁶, retina⁵⁷ and adipose tissue⁴⁶. Additional methods such as Mendelian randomization can be deployed to assess whether a causal relationship between expression and trait exists depending on more onerous assumptions being met, as described below.

The utility of TWAS for genetics-informed drug repurposing goes beyond that of single drug–gene interactions through the comparison of drug-induced gene-expression profiles with the direction of predicted trait-associated variation (TWAS Z score). This approach is often termed ‘signature matching’ and involves correlating the expression of genes associated with a disorder with the effect of a pharmacological compound on those same genes. For example, if a compound was risk-decreasing for a disease, then genes whose expression is increased by the drug would exhibit a negative correlation with genes genetically predicted to be upregulated in the disorder (TWAS Z score significantly greater than 0). In other words, drugs can be identified that may reverse or enhance disease-associated predicted expression. Signature mapping has been applied to several disorders and complex traits^{58–61}. For instance, Gerring et al. applied a signature mapping approach to late-onset Alzheimer disease and prioritized several repurposing candidates, including cyclooxygenase inhibitors and PPAR receptor agonists⁶⁰, both of which have previously been proposed as potential repurposing candidates for Alzheimer disease based on data including cohort studies that showed a putative protective effect of using these compounds on the developing disorder^{62,63}. Studies that employ signature mapping commonly leverage drug-induced expression profiles from databases such as the Connectivity Map (CMap)⁶⁴, Crowd Extracted Expression of Differential Signatures (CREEDS)⁶⁵, DrugMatrix⁶⁶ and PharmOomics⁶⁷. Future efforts to expand the repertoire of tissues and cell types in which perturbagen-associated expression profiles are available would enhance signature mapping, given that CMap data, for example, are restricted to a selection of immortalized cell lines.

Limitations of transcriptome-wide association studies. There are a number important limitations in use of the TWAS to inform trait-associated expression changes, which have been comprehensively outlined elsewhere⁶⁸; we briefly highlight some salient points. For instance, a gene’s predicted level of expression displaying association does not necessarily inform which is the causal gene at any given locus. Such false positives can arise from co-regulation exhibited with an actual causal gene, or

from linkage disequilibrium between SNP weights, which may complicate the interpretation of the association. This factor may particularly confound signature mapping, as it is difficult to quantify the proportion of the TWAS Z score that may reflect causal genes for the trait of interest. One solution proposed to address this issue is to implement fine-mapping in a manner somewhat analogous to GWAS-based SNP fine-mapping and derive credible sets of proximally located genes identified as TWAS signals that contain a causal gene with a probability specified a priori, as implemented by the FOCUS methodology⁶⁹. One could potentially use this approach to reduce or weight a TWAS Z score for signature matching, such that genes with a higher posterior probability for inclusion in the credible set are upweighted; however, the choice of statistical prior for Bayesian inference would have to be carefully considered, particularly for less statistically significant signals that do not survive conventional multiple-testing correction. It is also important to consider whether cohorts used to train GReX models contain disease cases for the phenotype to be investigated with TWAS as this could lead to biased estimates; that said, it has been shown previously that *cis*-genetic effects are highly consistent between cases and controls in data sets used to derive TWAS weights, which suggests that TWAS are unlikely to be biased by the inclusion of disease cases when estimating GReX⁴⁸. GReX training data sets are also usually limited in sample size, and an effort to collect more samples with matched genotype and expression data, along with cell type-specific data, is warranted. Moreover, whilst GReX offers an attractive opportunity to evaluate the association between gene expression and a phenotype, these genetic proxies of expression are, in some cases, only modestly predictive and do not account fully for the effect of other factors, such as environmental perturbagens, on expression. Direct measurement of expression through RNA sequencing in diseased cases and healthy controls can, therefore, complement genetic approaches in terms of identifying differentially expressed genes that could inform repurposing opportunities and enable signature matching⁷⁰.

Repurposing using biological pathways

Testing the combined association of variants within sets of genes that comprise pathways and biological networks (gene-set association) is an attractive prospect for GWAS given that the impact of trait-associated variation on a pathway is likely to be larger than individual loci. Biological pathways may thus represent higher-order drug targets, in the sense that compounds can be selected that modulate sentinel proteins in the network or multiple proteins throughout it. For instance, a recent gene-set association analysis of a GWAS of bipolar disorder revealed an association with a pathway involving genes that regulate insulin secretion, which might support the utility of glycaemic interventions in psychiatry⁷¹.

Gene-set association techniques primarily test genes, rather than variants, as their unit of effect⁷². Specifically, SNP-level *P* values are combined for each gene to compute a genic *P* value, usually either through methods that account for linkage disequilibrium between SNPs^{73–76}

Mendelian randomization
Randomization using single-nucleotide polymorphisms (SNPs) as instrumental variables (IVs) (proxies of a trait, termed the exposure) to test the causal effect of that trait on another (termed the outcome).

Biological pathways
Genes whose products exert biologically related functions or interact together.

or by leveraging distributions that are not unduly biased by dependencies amongst *P* values due to linkage disequilibrium⁷⁷. This approach has previously been applied to sets of genes that represent the molecular targets of different drugs^{78,79} or genic targets of drug classes as defined by annotation categories, such as Anatomical Therapeutic Chemical (ATC) classification codes⁸⁰. For example, the schizophrenia common variant signal was enriched within the target genes of anti-epileptic drugs and calcium channel blockers via gene-set association⁸⁰. An enrichment was also observed amongst the targets of antipsychotic drugs, the current main indicated pharmacotherapy for schizophrenia. The primary limitation of current methods for gene-set association is that they are not informative as to the direction of effect, that is, whether a gene set exerts a risk-increasing or risk-decreasing effect on a disorder; however, the increased biological complexity of these signals relative to a single variant renders them useful. These approaches are also limited by our capacity to specifically map the genic influence of association signals obscured by linkage disequilibrium at gene-dense loci, particularly for non-coding variation in the absence of eQTL or TWAS fine-mapping. Moreover, our knowledge of biological pathways is often incomplete, meaning further work is required to exhaustively identify all genes involved in key molecular processes.

Genetics-informed causal inference

The utility of GWAS extends beyond variant discovery as these data sets can be leveraged to identify putative causal relationships between sets of traits. The most widely implemented method for GWAS-based causal inference is Mendelian randomization^{81,82} (BOX 2). Mendelian randomization is underpinned by the use of genetic variants as instrumental variables (IVs) to estimate an effect of an exposure variable on an outcome phenotype (FIG. 2a). Specifically, the genetic IV effect size associated with the exposure and the outcome, respectively, is utilized to compute an exposure–outcome relationship. These IVs are said to proxy the measured exposure, and as genetic IVs will be randomized in the population due to Mendel's laws, this provides a valid mechanism for causal inference provided several assumptions are met, as described elsewhere^{82–84}. In usual practice, IVs are genome-wide significant SNPs for a particular phenotype. Two-sample Mendelian randomization is particularly advantageous over one-sample methodologies that require access to individual-level genotype data given that GWAS summary statistics are used as the IV–SNP effect on the exposure and outcome. Mendelian randomization can be implemented to propose drug repurposing opportunities as it facilitates testing the relationship between phenotypes that can be modulated by drugs, or are proxies of pharmacological interventions themselves, in the case of exposures such as protein abundance of target receptors and circulating micronutrient levels.

Primarily, two-sample Mendelian randomization using summary data for both the exposure and the outcome can be categorized as ‘*cis*-acting’ or ‘polygenic’ based on the number and nature of IVs implemented^{81,85–87}. *Cis*-acting constructs often select only a single SNP that acts directly in *cis* on the exposure of interest and

usually have a more readily understandable putative biological impact (FIG. 2b). Multiple SNP–IV models (polygenic) for an exposure represent a more powerful approach, but these models have varying underlying assumptions that must be taken into account⁸². For instance, when examining the relationship between C-reactive protein (CRP) and an outcome phenotype, a single IV that acts in *cis* on the CRP gene could be selected or all genome-wide significant SNPs could be selected, encompassing *cis* and *trans* effects on circulating CRP abundance. The single IV approach is usually only applicable to comparatively simple molecular phenotypes, such as circulating proteins; for more complex traits, a single *cis*-acting SNP is not readily selectable. Statistical co-localization is often implemented as a complementary method for putative causal trait pairs in the *cis*-acting approach to establish whether the IV–exposure association and IV–outcome effect are driven by the same association signal^{88,89}. Several multi-SNP Mendelian randomization approaches have been developed that make different assumptions about the validity of each IV SNP in the model (BOX 2). For instance, the Mendelian randomization–Egger regression approach includes a non-zero intercept term to test for evidence of confounding pleiotropy and remains unbiased with invalid IVs if there is no significant correlation between direct IV effects on the outcome and genetic association of IVs with the exposure⁹⁰.

Causal inference using protein abundance of existing drug targets. One interesting example of Mendelian randomization-based drug repurposing is the utilization of variants associated with protein abundance, termed protein quantitative trait loci (pQTLs), as IVs. Until recently, these pQTLs were only derived from cohorts with limited sample sizes; however, sample availability is now increasing in concert with technological advances in assaying protein abundance⁹¹. The underlying concept of pQTL-based drug repurposing posits that if a protein target of a known drug causally influences an outcome in the same direction as the mode of pharmacological effect, then this compound may be therapeutically useful. This approach is applicable to the *cis*-acting and polygenic Mendelian randomization definitions discussed previously, as these studies can use *cis*-acting variants in the gene encoding the protein of interest or all variants associated with protein abundance, including *trans*-acting variants. It should be noted that multiple conditionally independent signals may at times be observed acting in *cis*⁹². Several studies have implemented this approach for various complex disorders and traits throughout the human genome^{93–95}. Encouragingly, causal relationships between the targets of approved drugs and their current clinical indication were supported by these analyses, suggesting that repurposing prioritized candidates may be therapeutically useful. For instance, a causal effect of increased PCSK9 protein expression was observed on hypercholesterolemia, which corresponds to the therapeutic action of the approved PCSK9 inhibitor alirocumab⁹³. There are some limitations to this pQTL-centric approach; firstly, whole blood is still the primary source of pQTL studies, which

Instrumental variables (IVs). Independent variables that are used to evaluate whether an exposure causes an outcome or is simply correlated with it.

Box 2 | Methodological considerations in Mendelian randomization analyses

Mendelian randomization is a powerful epidemiological technique that is now extensively used in complex trait genetics. The key utility of Mendelian randomization is that it allows the estimation of a causal effect of an exposure trait on an outcome trait by leveraging genetic variants as instrumental variables (IVs). IVs are genetic variants that are said to ‘proxy’ the exposure and, due to Mendel’s laws of random segregation and independent assortment, will be randomly inherited within a population. As a result, given a series of assumptions, the effect size of the exposure–IV relationship can be used to estimate a causal effect on an outcome. Genome-wide association studies (GWAS) summary statistics can be utilized for this purpose in what is termed ‘two-sample Mendelian randomization’ such that the IV–exposure and IV–outcome estimates are obtained from the single-nucleotide polymorphism (SNP) effect sizes in each of the respective GWAS. The central assumptions of a valid IV are as follows:

1. The IV should be strongly associated with the exposure ($\hat{\gamma}$) — this is usually achieved by selecting SNPs that are genome-wide significant ($P < 5 \times 10^{-8}$).
2. The IV should be independent of all confounders (U) of the exposure–outcome relationship.
3. The IV can only be directly associated with the outcome (\hat{Y}) by acting through the exposure; in other words, the IV SNP should be independent of the outcome conditioned on the exposure.

To estimate the exposure–outcome relationship (β), these assumptions can be concisely expressed for a valid IV SNP j as:

$$\hat{\beta}_{\text{Exposure} \rightarrow \text{Outcome}} = \frac{\hat{\Gamma}_j}{\hat{\gamma}_j} \epsilon \{ \hat{\gamma} \neq 0, U = 0, \hat{\Gamma}|\hat{\gamma} = 0 \}$$

It is biologically naïve in practice that these assumptions will be satisfied for many SNPs, and, as a result, Mendelian randomization methods have been developed to account for this, along with sensitivity analyses to signal potential violations of the central IV assumptions. Usually, the inverse-variance weighted effect estimator is considered the most well-powered multi-SNP Mendelian randomization model; however, it has a 0% breakdown point, meaning it is likely biased if any of the IVs are invalid. Models robust to IV invalidity have therefore been developed under different assumptions, such as the ‘majority valid’ assumption in median-based estimators, which are robust to less than 50% of the exposure effects (or weight in the model) arising from invalid IVs¹³³. Furthermore, the ‘plurality valid’ assumption is that out of groups of IVs having the same asymptotic estimate, the largest group will comprise valid IVs, for example, modal-based estimators and the contamination mixture model^{134,135}. The ‘InSIDE assumption’ is implemented for Mendelian randomization–Egger regression, which includes a non-zero intercept as a test of the average pleiotropic effect and assumes that there is no significant correlation between direct IV effects on the outcome and genetic association of IVs with the exposure (Instrument Strength Independent of Direct Effect (InSIDE) assumption)⁹⁰. The Egger intercept in Mendelian randomization–Egger regression is a useful sensitivity analysis as a non-zero intercept value may indicate unbalanced pleiotropy and potentially imply that $U \neq 0$. Heterogeneity amongst the IV exposure–outcome estimates and outlier IVs that materially alter the Mendelian randomization estimate when removed may also be indicative of model violation and warrant follow-up.

In summary, methodological recommendations for Mendelian randomization have been discussed at length elsewhere^{82,86}; however, we recommend the four following brief guidelines for studies with a putative exposure–outcome causal relationship:

1. IVs should be sufficiently powerful, for instance, F statistic > 10 , and the biology/pleiotropy of individual SNPs should be considered relative to the trait where possible^{47,136,137}.
2. Polygenic Mendelian randomization models should be computed that represent at least three different assumptions about IV validity and their estimate compared^{81,90,133,138,139}, although this is not feasible for *cis*-acting models with two IVs or fewer.
3. Statistical estimates of heterogeneity and pleiotropy should be obtained.
4. Exposure–outcome genetic correlation should be considered^{83,104}.

Non-zero genetic correlation between an exposure and an outcome has been shown in simulations to upwardly bias Mendelian randomization due to pleiotropy between SNP effects and potential direct effects on the outcome that may violate the third IV assumption¹⁰⁴. As a result, when there is significant genetic correlation between the exposure and outcome GWAS we recommend also using a method such as the latent causal variable model, which explicitly models the genetic correlation estimate, to triangulate evidence for an exposure–outcome causal relationship. Importantly, it should be noted that genetic correlation does not preclude a causal relationship, whereas causal relationships can exist without genetic correlation.

is not the direct tissue of relevance for many human disorders. Furthermore, epitope binding artefacts through linkage with non-synonymous variants are a key analytical consideration, as outlined elsewhere^{91,96}. In addition, individual proteins may be a simplistic deconvolution of phenotypes that can be modulated by drugs, and, as a result, drug repurposing opportunities may be missed relative to more complex molecular factors, such as circulating metabolites and hormones.

Causal relationships with complex molecular traits may inform repurposing candidates. Drug repurposing candidates can also be derived using Mendelian randomization exposures for GWAS of more complex molecular traits. In this instance, there is no readily applicable single gene that represents a *cis*-acting variant effect on the trait of interest and, thus, composite multi-SNP Mendelian randomization approaches are required with different underlying assumptions about

IV validity and pleiotropy. For example, a putative causal relationship between vitamin D and multiple sclerosis has been supported through Mendelian randomization, which might therefore suggest therapeutical relevance of vitamin D supplementation⁹⁷. Several other Mendelian randomization constructs have been reported whereby a therapeutically actionable trait was suggested to have a causal effect on a complex trait, including branched-chain amino acids and type 2 diabetes mellitus⁹⁸, fasting glucose and lung function⁴⁷, systolic blood pressure and type 2 diabetes mellitus⁹⁹, serum calcium and migraine¹⁰⁰, and fasting insulin and anorexia nervosa¹⁰¹. The advantage of a Mendelian randomization-based drug repurposing approach is that it may help prioritize drug-trait pairs for which randomized controlled trials should be initiated. It is important to note that each of these studies performed Mendelian randomization using different combinations of models and sensitivity analyses, and, thus, their rigour must be considered by

investigators on a case by case basis. This is particularly important as Mendelian randomization methods can be biased estimators if the varying model-specific assumptions regarding IV validity are not met, with many individual SNPs likely invalid IVs due to linkage and the immense pleiotropy displayed amongst complex traits. Furthermore, confounding factors that affect the SNP effect sizes estimated by the exposure and/or outcome GWAS may also be problematic, including phenomena such as population stratification and selection bias^{102,103}. In other words, the ability of Mendelian randomization to draw causal inference is only facilitated when its underlying assumptions are satisfied. Methodological considerations for Mendelian randomization to overcome some of these issues have been discussed at length elsewhere, and we make some further recommendations in BOX 2 (REFS^{82,86}). Mendelian randomization has also been demonstrated to be potentially susceptible to bias arising from genetic correlation amongst the exposure

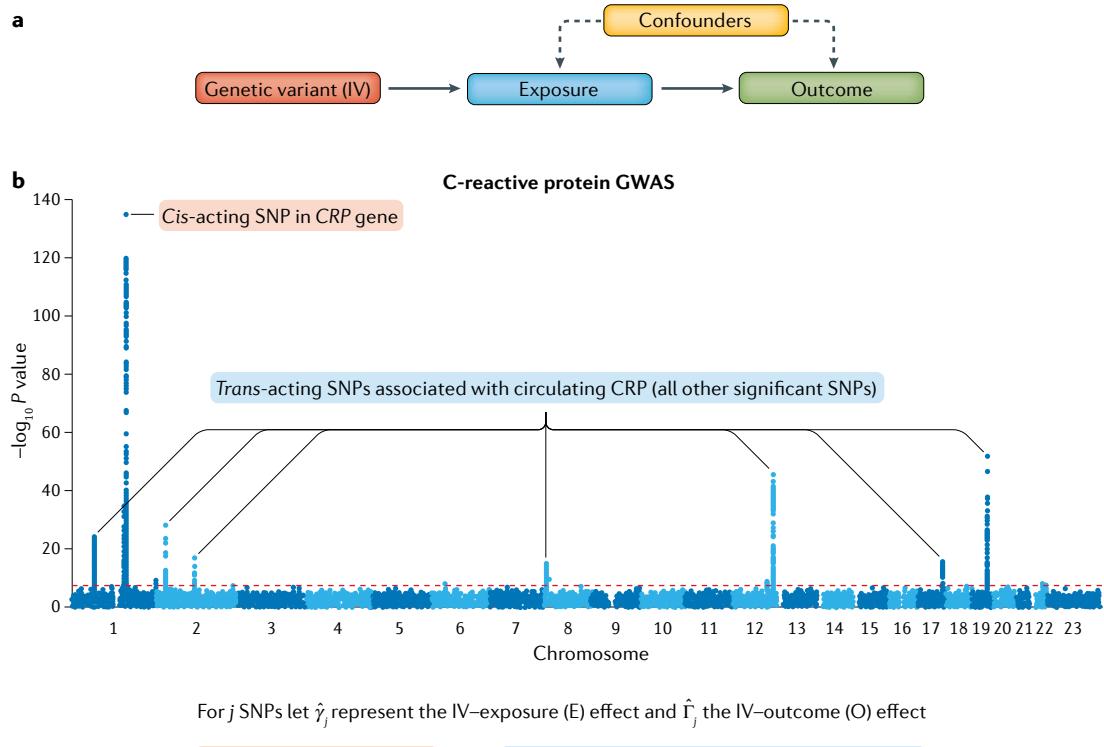


Fig. 2 | Mendelian randomization approach for causal inference leveraging GWAS data. **a** | Directed acyclic graph-like representation of the Mendelian randomization method. Genetic variants are selected as instrumental variables (IVs) to proxy a measured exposure, that is, variants significantly associated with the exposure via genome-wide association studies (GWAS). In the presence of a causal relationship, these variants proxy an exposure with a causal effect on an outcome. If the variants act through a confounder of the exposure–outcome relationship, then this may violate certain assumptions of Mendelian randomization, depending on the model implemented. In addition, IV variants should only be associated with the outcome by acting through the exposure. **b** | Example of cis-acting and polygenic (multi-SNP) Mendelian randomization approaches using either a single cis-acting variant as an IV or considering all genome-wide significant SNPs as an IV. In this example, a GWAS of circulating C-reactive protein (CRP), a cis-acting SNP within the CRP gene that encodes the CRP protein, could be utilized (highlighted in red). The remaining genome-wide significant SNPs act in trans on CRP levels, and thus all SNPs (cis and trans) could be utilized in a more liberal but powerful approach such as an inverse-variance weighted effect model (pictured). β denotes the estimated exposure–outcome effect, whilst SE refers to the standard error of the SNP effect on the outcome. SNP, single-nucleotide polymorphism.

Box 3 | Polygenic scoring for risk stratification in complex disorders

A polygenic score is the weighted effect size of each variant for a trait derived from genome-wide association studies (GWAS) and summed for an individual (usually assuming an additive model)^{105,106}. Traditionally, this comprises the following model, which sums the statistical effect size of each variant multiplied by the allele count (dosage) for the said variant. For the polygenic score for individual i (PGS $_i$), let $\hat{\beta}_j$ denote the statistical effect size from the GWAS for each variant j , multiplied by the dosage (G) of j in i :

$$\text{PGS}_i = \sum_{j=1}^M (\hat{\beta}_j G_{ij}).$$

It has been previously demonstrated the combined impact of variants that do not exceed genome-wide significance for association with a trait is also biologically salient, including variants above even nominal statistical significance ($P > 0.05$)¹⁰⁵. This phenomenon has been manifested as 'P value thresholding', whereby variants below a specified significance threshold for association are considered. For instance, a P value threshold of $P < 0.05$ indicates that all linkage disequilibrium pruned variants at this level of statistical significance or below will be utilized. In polygenic scores, different P value threshold values are used to construct the most parsimonious model, which likely represents different underlying biology captured by these variants. Newer approaches have also been developed to account for linkage disequilibrium rather than clumping correlated variants, such as shrinking estimates with a continuous weight through Bayesian methodologies^{140,141}. Polygenic scores applied to follow-up cohorts have demonstrated that these scores are associated with the phenotype of interest in many disorders^{11,47,105,107}. The effect size of this variant burden for some phenotypes is also now understood to be quite large for those individuals who harbour an elevated score relative to a population. For example, individuals in the top percentile for the breast cancer polygenic risk score have been shown to have a more than fourfold risk increase for developing breast cancer relative to the rest of that cohort, with the estimated lifetime oestrogen receptor-positive breast cancer absolute risk by age 80 years ranging from 2% to 31% for the lowest and highest PRS centile, respectively¹⁴².

and outcome pairs¹⁰⁴; thus, methods that account for this correlation, such as the latent causal variable approach, can be useful to complement Mendelian randomization, as demonstrated previously^{47,104}.

Precision drug repurposing

A limitation of the aforementioned drug repurposing strategies is that they make the inherent assumption that a candidate drug will be efficacious for all individuals with the relevant phenotype. In reality, patient response to the spectrum of pharmacotherapies used in clinical practice will always vary to some degree, with this heterogeneity adversely affecting patient outcomes for disorders that are difficult to treat. The polygenic nature of complex traits is an important consideration in drug repurposing as the underlying trait-associated biology for an individual will partially be attributed to their unique genotype. Specifically, a different profile of risk and protective alleles will be carried throughout the population, and thus distinct biological systems may be encompassed by each individual's genotypes. This is particularly true for highly polygenic traits with a large number of genome-wide significant loci, as well as a strong polygenic signal throughout the rest of the genome. Polygenic scoring has been implemented as a method to capture the association signal for complex traits conferred by common variation and quantify it on an individual level. In a polygenic score, the weighted effect size of each variant for a trait is derived from GWAS and summed for an individual^{105,106} (BOX 3). Polygenic scores may be efficacious at identifying

individuals at high or low risk for a trait beyond traditional risk factors, although further investigation of clinical utility is required^{11,47,105,107}.

Nevertheless, genome-wide polygenic scores are likely to be less useful in terms of formulating treatment and proposing drug repurposing candidates. This is largely due to the expansive nature of these scores, which encompass numerous genes and biological pathways throughout the genome. Therefore, in the case of an individual, the score itself does not provide any direct insights into specific biology that may be involved in the manifestation of the trait of interest. A systems-level approach, however, can be implemented to address this, whereby the polygenic score is deconvolved into distinct sets of genes, such as those that comprise biological pathways, to improve the biological salience of the score. Methods for deconvolving the polygenic signal at an individual level have been explored in only a few studies thus far. For example, the PGMRA (polygenic many-to-many relations analysis) approach clusters SNPs and scores for brain disorders based specifically on variants mapped to genes in co-expression networks with the insulin receptor^{108,109}. A pathway-based polygenic scoring framework to specifically inform drug repurposing, termed the pharmagenic enrichment score (PES), has recently been proposed by our group^{47,110}. In the PES framework, pathways with known drug targets are selected as candidates from a GWAS using gene-set association, whereby the pathways selected display at

Fig. 3 | Triangulating causal inference with the PES method to inform drug repurposing. **a** | Forced vital capacity (a measure of lung function) is a heritable quantitative trait with several genome-wide significant loci uncovered. Forced vital capacity genome-wide association studies (GWAS) can be leveraged for precision drug repurposing by constructing a forced vital capacity polygenic score within druggable pathways that are enriched with the common variant signal for the trait (pharmagenic enrichment score (PES)). For example, the class b/2 secretin signalling gene set is targeted by several approved drugs such as anti-hyperglycaemic agents. Class b/2 secretin signalling refers to a family of G-protein-coupled receptors which are ligands for several physiologically important hormones such as calcitonin, glucagon and the glucose-dependent insulinotropic polypeptide. The class b/2 secretin PES by itself does not necessarily inform whether an anti-hyperglycaemic compound would be beneficial for lung function; however, this can be coupled with GWAS-based causal inference that demonstrated evidence for a causal effect of elevated blood glucose on diminished forced vital capacity. As a result, one could potentially use anti-hyperglycaemic drugs for precision medicine in respiratory disorders by identifying individuals with low genetically predicted lung function in the class b/2 secretin signalling pathway (low PES) and elevated blood glucose. In this schematic, the class b/2 secretin pathway refers to the PES and fasting glucose refers to measured glucose. **b** | Schematic example of a class b/2 secretin signalling event whereby glucagon-like peptide (GLP) assists to regulate insulin efflux from pancreatic β cells in concert with glucose. Black arrows denote a stimulatory effect, red line-headed arrows denote an inhibitory relationship. VDCC, voltage-dependent calcium channel.

Polygenic score

A sum of the effect sizes of genetic variants throughout the genome for a particular trait.

Pharmagenic enrichment score

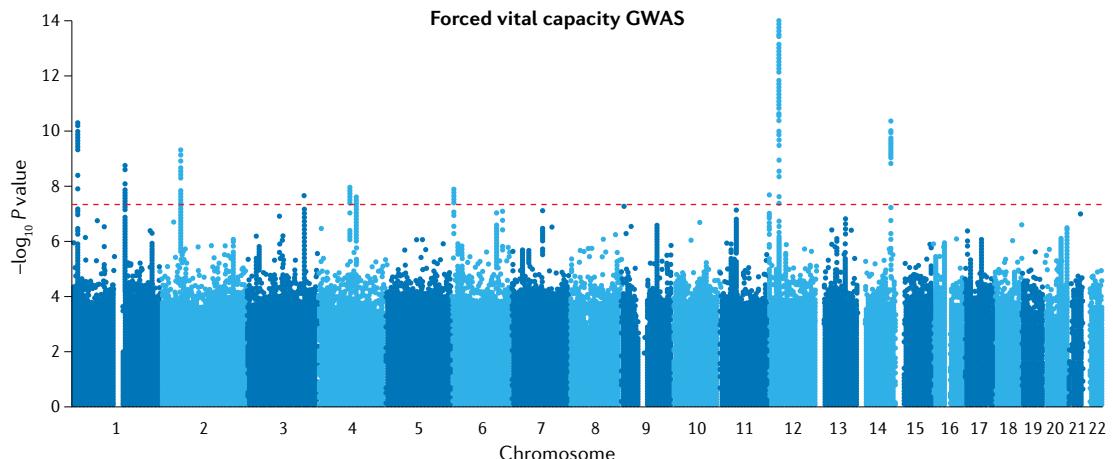
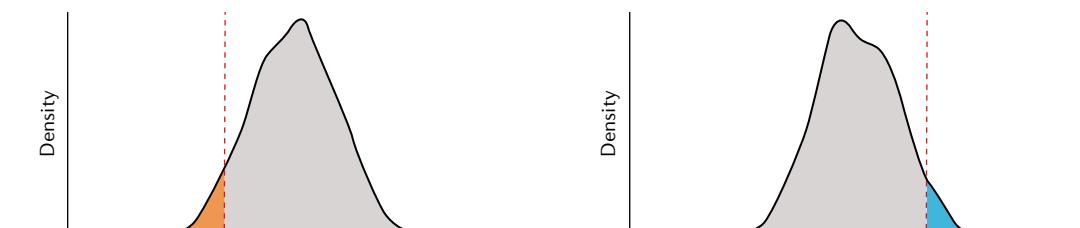
(PES). A polygenic score that is constructed from variants specifically within a biological pathway that is targeted by an approved drug, rather than genome-wide like a traditional polygenic score.

least some evidence of an enrichment of the association signal relative to all other genes. Pathway-specific risk scores for these druggable pathways can then be constructed, such that individuals with extreme scores, depending on the phenotype, may benefit from a drug which modulates that pathway. In other words, analogous to a traditional genome-wide score, the effect sizes (β) of i variants specifically within a druggable

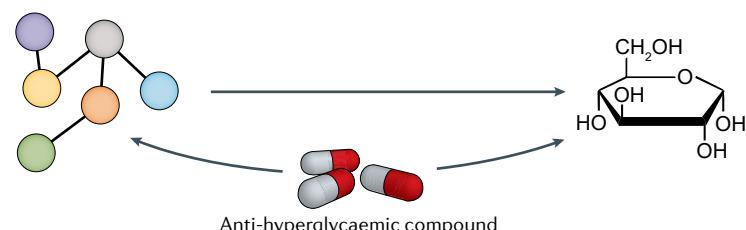
pathway of interest are multiplied by the dosage (G) in individual j :

$$\text{PES}_j = \sum_{i=1} \hat{\beta}_i G_{ij}.$$

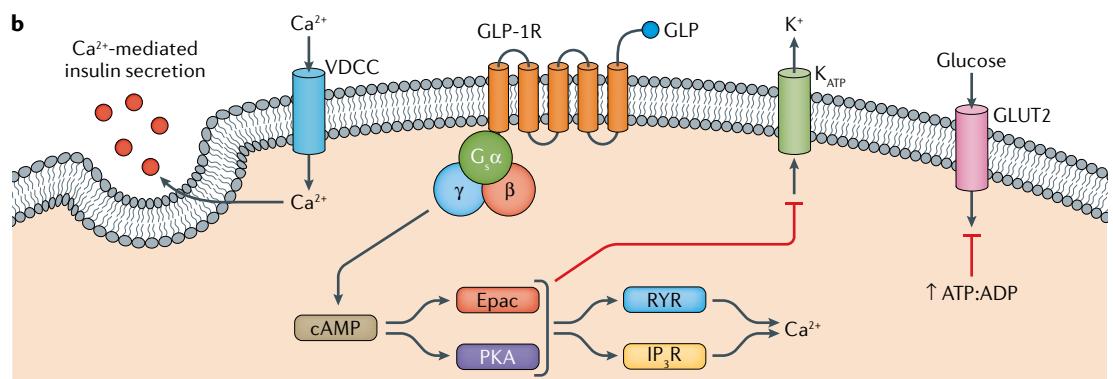
The interpretation of PES for a complex disorder will be that a higher score is of clinical interest; however, for a

a**Population****Individual**

Class b/2 secretin pathway



Fasting glucose

b

quantitative trait, the direction of clinical utility depends on the trait distribution itself. For instance, we applied the PES method to a GWAS of quantitative spirometry measures of lung function and found that a PES for lung function based on the class b/2 secretin pathway was associated with forced vital capacity in an independent cohort, even after adjusting for a genome-wide forced vital capacity polygenic score⁴⁷; coupling genetic evidence for a causal effect of fasting glucose on lung function with a PES in a glycaemia-relevant pathway (class b/2 secretin) emphasized the utility of anti-hyperglycaemic drugs in respiratory medicine (FIG. 3). This exemplifies the potential value of a pathway-specific approach beyond genome-wide scores, particularly when the said pathway of interest can be modulated by drugs. In addition, we explored the utility of the PES for the complex psychiatric disorder schizophrenia and observed further evidence to support the distinct biological information captured by a pathway-specific score such as the PES¹⁰. In that study, a notable proportion of the schizophrenia cohort examined had a low genome-wide polygenic risk score for schizophrenia but at least one elevated PES profile.

We acknowledge that considerable future work is required to establish whether a pathway-specific or gene set-specific score will have sufficient clinical utility in terms of precision drug repositioning. One limitation of the current approach is that the pathways used are selected based on their enrichment for the common variant association signal related to the trait of interest, but the PES does not inherently predict whether a compound that modulates the gene set will have a beneficial impact, or whether an agonist or antagonist compound would be most appropriate. Triangulation with other methods, such as causal inference in the class b/2 secretin example above, can be employed to overcome this uncertainty of effect direction. Moreover, given that genome-wide polygenic scores often explain only a small amount of phenotypic variance, the effect sizes of PES will most likely be modest, as they represent a deconvolution of the overall signal. Analogous to genome-wide significant variants, the utility of these scores will likely be determined relative to their effect on molecular traits, such as gene expression, that influence the pathway of interest. Nevertheless, initial data suggest that the PES demonstrates distinct associations with gene expression not captured by a genome-wide score alone^{47,10}.

Software applications for GWAS-informed drug repurposing

Several bioinformatic tools have been developed to specifically implement some, or a combination of, the methods for GWAS-informed drug repurposing (TABLE 2). For instance, the Open Targets Genetics resource facilitates the identification of drug repurposing candidates from single GWAS loci by implementing a suite of fine-mapping and functional annotation methods to predict genes implicated by the variant of interest, along with information about drugs targeting genes identified in this manner¹¹. Functional annotation is also incorporated into the priority index pipeline developed by Fang et al. for immune phenotypes that scores genes based on the following: genomic distance to trait-associated

variation; evidence of physical interaction with GWAS SNPs via chromatin conformation data from immune cells; evidence of the GWAS signal modulating gene expression through eQTLs (eGene); functional annotations related to involvement in immune-relevant gene ontologies; and evidence of involvement in rare immune disorders¹². These genes then act as ‘seed genes’ to define networks of biological interaction, and therefore drugable genes that interact with highly ranked seed genes can be prioritized. Moreover, the GREP (Genome for REPositioning drugs) framework is applicable to both sets of candidate functional genes from GWAS loci as well as pathways discovered through gene-set association, as it tests for an enrichment of drugs in specific categories of clinical indications derived from classification systems such as the ATC classification system¹³. The Drug Targetor platform represents another integrative approach to rank potential drug repurposing opportunities from GWAS; this tool combines gene-set association of the targets of a particular drug and the association of target genes through TWAS accompanied by their direction of effect relative to a said drug¹⁴. Packages have also been developed that automate the TWAS-based signature matching approach such as Trans-Phar (integration of TWAS and pharmacological database)¹⁵, whereas the PS4DR framework performs signature matching without explicitly modelling predicted expression and instead uses genetic overlap between genes perturbed by drugs and in disease¹⁶. Software that can automate the detection of candidate gene sets that could be used to calculate PES is also publicly available⁴⁷. A selection of bioinformatic tools that are not explicitly designed for drug repurposing but implement methods discussed in this Review are also outlined in TABLE 2. Accurately matching drugs to the genes they target is another critical component of the drug repurposing process, with databases including DrugBank, the ChEMBL bioactivity database and DGIdb (Drug–Gene Interaction Database) freely available for this purpose^{17,18}. DGIdb is likely to be particularly useful to researchers as it collates drug–gene interaction evidence from several different databases along with literature evidence to facilitate the identification of high-confidence interaction pairs.

Conclusions

The clinical imperative driving drug repurposing will likely intensify as the sample sizes and discovery power of GWAS continue to increase. Numerous mechanisms exist by which GWAS could be leveraged to inform drug repurposing. The complexity of these approaches ranges from the attribution of single GWAS signals to drug targets to polygenic scores that leverage variation from multiple biologically related genes throughout the genome. The polygenic approach may be particularly promising as it provides a mechanism to capture the heterogeneity between individuals in their genetic architecture of common disorders and use it as a target for treatment.

Future work in this field should be focused on triangulating evidence from multiple GWAS-informed methods to select the most promising candidates for further study in clinical trials. In addition, clinical trials themselves should be designed in the context

Table 2 | Selected bioinformatic tools that facilitate GWAS-informed drug repurposing

Software	Description	Code or website link	Refs
Open Targets Genetics	Single GWAS variant–gene drug repurposing	https://genetics.opentargets.org/ https://github.com/opentargets	¹¹¹
Priority index	Single GWAS variant–gene drug repurposing integrated with network-based prioritization using genes from GWAS as seed genes	http://pi.well.ox.ac.uk:3010/	¹¹²
GREP	Identifies drug classes enriched for genes prioritized from GWAS or pathways revealed by gene-set association	https://github.com/saorisakaue/GREP	¹¹³
Drug Targetor	Ranks drug repurposing candidates based on gene-set association of genes targeted by a compound and the direction of effect from a TWAS of its target genes relative to the drug's mechanism of action	https://drugtargetor.com/index_v1.21.html	¹¹⁴
Trans-Phar	Signature matching using TWAS results and drug perturbagen expression data	https://github.com/konumat/Trans-Phar	¹¹⁵
PS4DR	Signature matching integrating GWAS with disease-associated and drug-induced expression profiles	https://github.com/ps4dr/ps4dr	¹¹⁶
FUSION	TWAS software—results derived can inform drug repurposing opportunities	https://github.com/gusevlab/fusion_twasi	⁴⁶
S-PrediXcan	TWAS software—results derived can inform drug repurposing opportunities	https://github.com/hakyimlab/PrediXcan	⁴⁵
MAGMA	Gene-based and gene-set association—results derived can inform drug repurposing opportunities	https://ctg.cncr.nl/software/magma	⁷⁶
TwoSampleMR	Mendelian randomization methods—results derived can inform drug repurposing opportunities	https://github.com/MRCIEU/TwoSampleMR	¹³²
PES	Package for identifying candidate pathways that could be used to construct the PES	https://github.com/Williamreay/Pharmagenic_enrichment_score	^{47,110}

GREP, Genome for REPositioning drugs; GWAS, genome-wide association study; PES, pharmagenic enrichment score; TWAS, transcriptome-wide association study.

of expansive genetic heterogeneity of complex traits, which may render certain individuals more susceptible to particular interventions. Although we focused on the utility of common frequency variants for drug repurposing as these are the most widely available and well-powered analyses, they only partially explain the heritable component of these traits¹¹⁹. The importance of rare variants to drug repurposing will likely be enhanced as larger cohorts of sequenced individuals become available, along with reference panels of improved quality such that the imputation of rare variants can be more accurately performed. Gene-based approaches for rare variant association, such as the sequence kernel association test (SKAT), are likely to be critical to effectively discover rare variant-informed drug repurposing opportunities as power is an ongoing issue to detect individual rare variant signals after multiple-testing correction¹²⁰. Candidate loss-of-function variants are another example of the utility of rare variants, in that they may provide phenotypic insights into pharmacological inhibition of the gene in question. Future rare variant-orientated GWAS will continue to discover these associations with

greater statistical confidence, along with opportunities to integrate rare variants into polygenic scores. In addition, analyses of additive effects encompassed by variants in GWAS do not take into account potential gene by environment or gene by gene (epistasis) interactions, which may be relevant to drug repurposing, although statistical challenges remain in the efficient detection of these phenomena^{121,122}. Analogous to the progress in rare variants, larger sample sizes and method development will be critical to further identify non-additive effects. Moreover, network pharmacology methods, which seek to combine drugs in a polypharmacological approach based on their roles in specific biological networks, should be further explored in relation to GWAS¹²³.

In summary, we believe that there are great prospects to advance drug repurposing using GWAS. An interdisciplinary approach that integrates multiple lines of genomic evidence to triangulate the suitability of proposed drugs will likely be the most direct path to clinical translation.

Published online 23 July 2021

- Visscher, P. M. et al. 10 years of GWAS discovery: biology, function, and translation. *Am. J. Hum. Genet.* **101**, 5–22 (2017).
- Pushpakom, S. et al. Drug repurposing: progress, challenges and recommendations. *Nat. Rev. Drug Discov.* **18**, 41–58 (2019).
- This review provides a comprehensive overview of the rationale for drug repurposing.
- Pammolli, F., Magazzini, L. & Riccaboni, M. The productivity crisis in pharmaceutical R&D. *Nat. Rev. Drug Discov.* **10**, 428–438 (2011).
- Nosengo, N. Can you teach old drugs new tricks? *Nature* **534**, 314–316 (2016).
- Ashburn, T. T. & Thor, K. B. Drug repositioning: identifying and developing new uses for existing drugs. *Nat. Rev. Drug Discov.* **3**, 673–683 (2004).
- Verbaanderd, C., Rooman, I., Meheus, L. & Huys, I. On-label or off-label? Overcoming regulatory and financial barriers to bring repurposed medicines to cancer patients. *Front. Pharmacol.* **10**, 1664 (2019).
- Watanabe, K. et al. A global overview of pleiotropy and genetic architecture in complex traits. *Nat. Genet.* **51**, 1339–1348 (2019).

8. Bulik-Sullivan, B. et al. An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
9. Park, J.-H. et al. Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. *Nat. Genet.* **42**, 570–575 (2010).
10. Dudbridge, F. Power and predictive accuracy of polygenic risk scores. *PLoS Genet.* **9**, e1003348 (2013).
11. Khera, A. V. et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat. Genet.* **50**, 1219–1224 (2018).
12. Yang, J., Zeng, J., Goddard, M. E., Wray, N. R. & Visscher, P. M. Concepts, estimation and interpretation of SNP-based heritability. *Nat. Genet.* **49**, 1304–1310 (2017).
13. Speed, D. et al. Reevaluation of SNP heritability in complex human traits. *Nat. Genet.* **49**, 986–992 (2017).
14. Pe'er, I., Yelensky, R., Altshuler, D. & Daly, M. J. Estimation of the multiple testing burden for genome-wide association studies of nearly all common variants. *Genet. Epidemiol.* **32**, 381–385 (2008).
15. Fadista, J., Manning, A. K., Florez, J. C. & Groop, L. The (in)famous GWAS *P*-value threshold revisited and updated for low-frequency variants. *Eur. J. Hum. Genet.* **24**, 1202–1205 (2016).
16. Paleari, L. et al. Aromatase inhibitors as adjuvant treatment for ER/PgR positive stage I endometrial carcinoma: a retrospective cohort study. *Int. J. Mol. Sci.* **21**, 2227 (2020).
17. van Weelden, W. J., Massuger, L. F. A. G., ENITEC, Pijnenborg, J. M. A. & Romano, A. Anti-estrogen treatment in endometrial cancer: a systematic review. *Front. Oncol.* **9**, 359 (2019).
18. O'Mara, T. A. et al. Identification of nine new susceptibility loci for endometrial cancer. *Nat. Commun.* **9**, 3166 (2018).
19. Liu, J. Z. et al. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat. Genet.* **47**, 979–986 (2015).
20. Duerr, R. H. et al. A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. *Science* **314**, 1461–1463 (2006).
21. Stritesky, G. L., Yeh, N. & Kaplan, M. H. IL-23 promotes maintenance but not commitment to the T_H17 lineage. *J. Immunol.* **181**, 5948–5955 (2008).
22. Bunte, K. & Beikler, T. T_H17 cells and the IL-23/IL-17 axis in the pathogenesis of periodontitis and immune-mediated inflammatory diseases. *Int. J. Mol. Sci.* **20**, 3394 (2019).
23. Sandborn, W. J. et al. A randomized trial of ustekinumab, a human interleukin-12/23 monoclonal antibody, in patients with moderate-to-severe Crohn's disease. *Gastroenterology* **135**, 1130–1141 (2008).
24. Sandborn, W. J. et al. Ustekinumab induction and maintenance therapy in refractory Crohn's disease. *N. Engl. J. Med.* **367**, 1519–1528 (2012).
25. Feagan, B. G. et al. Induction therapy with the selective interleukin-23 inhibitor risankizumab in patients with moderate-to-severe Crohn's disease: a randomised, double-blind, placebo-controlled phase 2 study. *Lancet* **389**, 1699–1709 (2017).
26. Feagan, B. G. et al. Ustekinumab as induction and maintenance therapy for Crohn's disease. *N. Engl. J. Med.* **375**, 1946–1960 (2016).
27. Savage, L. J., Wittmann, M., McGonagle, D. & Helliwell, P. S. Ustekinumab in the treatment of psoriasis and psoriatic arthritis. *Rheumatol. Ther.* **2**, 1–16 (2015).
28. Banaszczyk, K. Risankizumab in the treatment of psoriasis — literature review. *Reumatologia* **57**, 158–162 (2019).
29. Singh, S. et al. Selective targeting of the IL23 pathway: generation and characterization of a novel high-affinity humanized anti-IL23A antibody. *MAbs* **7**, 778–791 (2015).
30. Feagan, B. G. et al. Risankizumab in patients with moderate to severe Crohn's disease: an open-label extension study. *Lancet Gastroenterol. Hepatol.* **3**, 671–680 (2018).
31. Nelson, M. R. et al. The support of human genetic evidence for approved drug indications. *Nat. Genet.* **47**, 856–860 (2015). **This paper demonstrates the extent to which currently indicated drugs are supported by findings from GWAS.**
32. Beveridge, L. A. et al. Effect of vitamin D supplementation on blood pressure: a systematic review and meta-analysis incorporating individual patient data. *JAMA Intern. Med.* **175**, 745 (2015).
33. Zhang, D. et al. Effect of vitamin D on blood pressure and hypertension in the general population: an update meta-analysis of cohort studies and randomized controlled trials. *Prev. Chronic Dis.* **17**, E03 (2020).
34. Jansen, I. E. et al. Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nat. Genet.* **51**, 404–413 (2019).
35. Marigorta, U. M. et al. Transcriptional risk scores link GWAS to eQTLs and predict complications in Crohn's disease. *Nat. Genet.* **49**, 1517–1521 (2017).
36. Morris, J. A. et al. An atlas of genetic influences on osteoporosis in humans and mice. *Nat. Genet.* **51**, 258–266 (2019).
37. Chandran, T. & Venkatachalam, I. Efficacy and safety of denosumab compared to bisphosphonates in improving bone strength in postmenopausal osteoporosis: a systematic review. *Singap. Med. J.* **60**, 364–378 (2019).
38. McGovern, D. & Powrie, F. The IL23 axis plays a key role in the pathogenesis of IBD. *Gut* **56**, 1333–1336 (2007).
39. Hue, S. et al. Interleukin-23 drives innate and T cell-mediated intestinal inflammation. *J. Exp. Med.* **203**, 2473–2483 (2006).
40. Morris, J. A. et al. Discovery of target genes and pathways of blood trait loci using pooled CRISPR screens and single cell RNA sequencing. Preprint at *bioRxiv* <https://doi.org/10.1101/2021.04.07.458882> (2021).
41. Nasser, J. et al. Genome-wide enhancer maps link risk variants to disease genes. *Nature* **593**, 238–243 (2021).
42. Schaid, D. J., Chen, W. & Larson, N. B. From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nat. Rev. Genet.* **19**, 491–504 (2018). **This review provides a comprehensive description of fine-mapping techniques for GWAS signals.**
43. Conrad, D. F. et al. A worldwide survey of haplotype variation and linkage disequilibrium in the human genome. *Nat. Genet.* **38**, 1251–1260 (2006).
44. Lieberman-Aiden, E. et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
45. GTEx Consortium et al. A gene-based association method for mapping traits using reference transcriptome data. *Nat. Genet.* **47**, 1091–1098 (2015).
46. Gusev, A. et al. Integrative approaches for large-scale transcriptome-wide association studies. *Nat. Genet.* **48**, 245–252 (2016).
47. Reay, W. R. et al. Genetic association and causal inference converge on hyperglycaemia as a modifiable factor to improve lung function. *eLife* **10**, e63115 (2021). **This study demonstrates how causal inference can be integrated with the PES approach to support specific repurposing opportunities.**
48. Schizophrenia Working Group of the Psychiatric Genomics Consortium et al. Transcriptome-wide association study of schizophrenia and chromatin activity yields mechanistic disease insights. *Nat. Genet.* **50**, 538–548 (2018).
49. Han, S. et al. Integrating brain methylome with GWAS for psychiatric risk gene discovery. Preprint at *bioRxiv* <https://doi.org/10.1101/440206> (2018).
50. Zhang, J. et al. Large Bi-ethnic study of plasma proteome leads to comprehensive mapping of *cis*-pQTL and models for proteome-wide association studies. Preprint at *bioRxiv* <https://doi.org/10.1101/2021.03.15.435533> (2021). **This Perspective comprehensively describes the utility and limitations of the TWAS methodology.**
51. Wingo, A. P. et al. Integrating human brain proteomes with genome-wide association data implicates new proteins in Alzheimer's disease pathogenesis. *Nat. Genet.* **53**, 143–146 (2021).
52. Luningham, J. M. et al. Bayesian genome-wide TWAS method to leverage both *cis*- and *trans*-eQTL information through summary statistics. *Am. J. Hum. Genet.* **107**, 714–726 (2020).
53. Hu, Y. et al. A statistical framework for cross-tissue transcriptome-wide association analysis. *Nat. Genet.* **51**, 568–576 (2019).
54. Feng, H. et al. Leveraging expression from multiple tissues using sparse canonical correlation analysis and aggregate tests improves the power of transcriptome-wide association studies. *PLoS Genet.* **17**, e1008973 (2021).
55. GTEx Consortium et al. Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *Nat. Commun.* **9**, 1825 (2018).
56. Gandal, M. J. et al. Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science* **362**, eaat8127 (2018).
57. Ratnapriya, R. et al. Retinal transcriptome and eQTL analyses identify genes associated with age-related macular degeneration. *Nat. Genet.* **51**, 606–610 (2019).
58. Wright, G. E. B. et al. Gene expression profiles complement the analysis of genomic modifiers of the clinical onset of Huntington disease. *Hum. Mol. Genet.* **29**, 2788–2802 (2020).
59. International League Against Epilepsies Consortium on Complex Epilepsies. Genome-wide mega-analysis identifies 16 loci and highlights diverse biological mechanisms in the common epilepsies. *Nat. Commun.* **9**, 5269 (2018).
60. Gerrings, Z. F., Gamazon, E. R., White, A. & Derkx, E. M. An integrative network-based analysis reveals gene networks, biological mechanisms, and novel drug targets in Alzheimer's disease. Preprint at *bioRxiv* <https://doi.org/10.1101/853580> (2019).
61. Zhang, W. et al. Integrative transcriptome imputation reveals tissue-specific and shared biological mechanisms mediating susceptibility to complex traits. *Nat. Commun.* **10**, 3834 (2019).
62. Zhang, C., Wang, Y., Wang, D., Zhang, J. & Zhang, F. NSAID exposure and risk of Alzheimer's disease: an updated meta-analysis from cohort studies. *Front. Aging Neurosci.* **10**, 83 (2018).
63. Heneka, M. T., Reyes-Irisarri, E., Hüll, M. & Kummer, M. P. Impact and therapeutic potential of PPARs in Alzheimer's disease. *Curr. Neuropharmacol.* **9**, 643–650 (2011).
64. Musa, A. et al. A review of Connectivity Map and computational approaches in pharmacogenomics. *Brief. Bioinform.* **19**, 506–523 (2018).
65. Wang, Z. et al. Extraction and analysis of signatures from the Gene Expression Omnibus by the crowd. *Nat. Commun.* **7**, 12846 (2016).
66. Svoboda, D. L., Saddler, T. & Auerbach, S. S. in *Advances in Computational Toxicology* Vol. 30 (ed. Hong, H.) 141–157 (Springer, 2019).
67. Chen, Y.-W. et al. PharmOomics: a species- and tissue-specific drug signature database and online tool for drug repurposing. Preprint at *bioRxiv* <https://doi.org/10.1101/837773> (2019).
68. Wainberg, M. et al. Opportunities and challenges for transcriptome-wide association studies. *Nat. Genet.* **51**, 592–599 (2019). **This Perspective comprehensively describes the utility and limitations of the TWAS methodology.**
69. Mancuso, N. et al. Probabilistic fine-mapping of transcriptome-wide association studies. *Nat. Genet.* **51**, 675–682 (2019).
70. Subramanian, A. et al. A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. *Cell* **171**, 1437–1452.e17 (2017).
71. eQTLGen, Consortium et al. Genome-wide association study identifies 50 loci associated with bipolar disorder. *Nat. Genet.* **51**, 793–803 (2019).
72. de Leeuw, C. A., Neale, B. M., Heskes, T. & Posthuma, D. The statistical properties of gene-set analysis. *Nat. Rev. Genet.* **17**, 353–364 (2016). **This review summarizes the different approaches and statistical considerations for performing gene-set association.**
73. Liu, J. Z. et al. A versatile gene-based test for genome-wide association studies. *Am. J. Hum. Genet.* **87**, 139–145 (2010).
74. Li, M.-X., Gui, H.-S., Kwan, J. S. H. & Sham, P. C. GATES: a rapid and powerful gene-based association test using extended Simes procedure. *Am. J. Hum. Genet.* **88**, 283–293 (2011).
75. Lamparter, D., Marbach, D., Rueedi, R., Katalik, Z. & Bergmann, S. Fast and rigorous computation of gene and pathway scores from SNP-based summary statistics. *PLoS Comput. Biol.* **12**, e1004714 (2016).
76. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.* **11**, e1004219 (2015).
77. Liu, Y. & Xie, J. Cauchy combination test: a powerful test with analytic *P*-value calculation under arbitrary dependency structures. *J. Am. Stat. Assoc.* **115**, 393–402 (2020).
78. de Jong, S., Vidler, L. R., Mokrab, Y., Collier, D. A. & Breen, G. Gene-set analysis based on the pharmacological profiles of drugs to identify repurposing opportunities in schizophrenia. *J. Psychopharmacol.* **30**, 826–830 (2016).

79. So, H.-C., Chau, C. K.-L., Lau, A., Wong, S.-Y. & Zhao, K. Translating GWAS findings into therapies for depression and anxiety disorders: gene-set analyses reveal enrichment of psychiatric drug classes and implications for drug repositioning. *Psychol. Med.* **49**, 2692–2708 (2019).
80. Gaspar, H. A. & Breen, G. Drug enrichment and discovery from schizophrenia genome-wide association results: an analysis and visualisation approach. *Sci. Rep.* **7**, 12460 (2017).
81. Burgess, S., Butterworth, A. & Thompson, S. G. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet. Epidemiol.* **37**, 658–665 (2013).
82. Slob, E. A. W. & Burgess, S. A comparison of robust Mendelian randomization methods using summary data. *Genet. Epidemiol.* **44**, 313–329 (2020). **This study compares different Mendelian randomization methods and their underlying assumptions.**
83. VanderWeele, T. J., Tchetgen Tchetgen, E. J., Cornelis, M. & Kraft, P. Methodological challenges in mendelian randomization. *Epidemiology* **25**, 427–435 (2014).
84. Burgess, S., Bowden, J., Fall, T., Ingelsson, E. & Thompson, S. G. Sensitivity analyses for robust causal inference from Mendelian randomization analyses with multiple genetic variants. *Epidemiology* **28**, 30–42 (2017).
85. Burgess, S. et al. Using published data in Mendelian randomization: a blueprint for efficient identification of causal risk factors. *Eur. J. Epidemiol.* **30**, 543–552 (2015).
86. Burgess, S. et al. Guidelines for performing Mendelian randomization investigations. *Wellcome Open Res.* **4**, 186 (2019).
87. McGowan, L. M., Davey Smith, G., Gaunt, T. R. & Richardson, T. G. Integrating Mendelian randomization and multiple-trait colocalization to uncover cell-specific inflammatory drivers of autoimmune and atopic disease. *Hum. Mol. Genet.* **28**, 3293–3300 (2019).
88. Giambalozzi, C. et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* **10**, e1004383 (2014).
89. Hormozdiari, F. et al. Colocalization of GWAS and eQTL signals detects target genes. *Am. J. Hum. Genet.* **99**, 1245–1260 (2016).
90. Bowden, J., Davey Smith, G. & Burgess, S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int. J. Epidemiol.* **44**, 512–525 (2015).
91. Suhre, K., McCarthy, M. I. & Schwenk, J. M. Genetics meets proteomics: perspectives for large population-based studies. *Nat. Rev. Genet.* <https://doi.org/10.1038/s41576-020-0268-2> (2020). **This study demonstrates how pQTLs could be utilized through Mendelian randomization to inform drug repurposing.**
92. Sun, B. B. et al. Genomic atlas of the human plasma proteome. *Nature* **558**, 73–79 (2018).
93. Zheng, J. et al. Phenome-wide Mendelian randomization mapping the influence of the plasma proteome on complex diseases. *Nat. Genet.* <https://doi.org/10.1038/s41588-020-0682-6> (2020).
94. Schmidt, A. F. et al. Genetic drug target validation using Mendelian randomisation. *Nat. Commun.* **11**, 3255 (2020).
95. Folkersen, L. et al. Genomic evaluation of circulating proteins for drug target characterisation and precision medicine. Preprint at *bioRxiv* <https://doi.org/10.1101/2020.04.03.203804> (2020).
96. Suhre, K. et al. Connecting genetic risk to disease end points through the human blood plasma proteome. *Nat. Commun.* **8**, 14357 (2017).
97. Mokry, L. E. et al. Vitamin D and risk of multiple sclerosis: a mendelian randomization study. *PLoS Med.* **12**, e1001866 (2015).
98. Lotta, L. A. et al. Genetic predisposition to an impaired metabolism of the branched-chain amino acids and risk of type 2 diabetes: a Mendelian randomisation analysis. *PLoS Med.* **13**, e1002179 (2016).
99. Aikens, R. C. et al. Systolic blood pressure and risk of type 2 diabetes: a Mendelian randomization study. *Diabetes* **66**, 543–550 (2017).
100. Yin, P. et al. Serum calcium and risk of migraine: a Mendelian randomization study. *Hum. Mol. Genet.* **26**, 820–828 (2016).
101. Adams, D. M., Reay, W. R., Geaghan, M. P. & Cairns, M. J. Investigation of glycaemic traits in psychiatric disorders using Mendelian randomisation revealed a causal relationship with anorexia nervosa. *Neuropsychopharmacology* **46**, 1093–1102 (2020).
102. Koellinger, P. D. & de Vlaming, R. Mendelian randomization: the challenge of unobserved environmental confounds. *Int. J. Epidemiol.* **48**, 665–671 (2019).
103. Gkatzionis, A. & Burgess, S. Contextualizing selection bias in Mendelian randomization: how bad is it likely to be? *Int. J. Epidemiol.* **48**, 691–701 (2019).
104. O'Connor, L. J. & Price, A. L. Distinguishing genetic correlation from causation across 52 diseases and complex traits. *Nat. Genet.* **50**, 1728–1734 (2018). **This study reveals that genetic correlation can bias Mendelian randomization and provides a novel causal inference method, which explicitly models genetic correlation, to overcome this.**
105. International Schizophrenia Consortium et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748–752 (2009).
106. Wray, N. R., Goddard, M. E. & Visscher, P. M. Prediction of individual genetic risk to disease from genome-wide association studies. *Genome Res.* **17**, 1520–1528 (2007).
107. Xue, A. et al. Genome-wide association analyses identify 143 risk variants and putative regulatory mechanisms for type 2 diabetes. *Nat. Commun.* **9**, 2941 (2018).
108. Arnedo, J. et al. PGMR: a web server for (phenotype×genotype) many-to-many relation analysis in GWAS. *Nucleic Acids Res.* **41**, W142–W149 (2013).
109. Hari Dass, S. A. et al. A biologically-informed polygenic score identifies endophenotypes and clinical conditions associated with the insulin receptor function on specific brain regions. *EBioMedicine* **42**, 188–202 (2019).
110. Reay, W. R., Atkins, J. R., Carr, V. J., Green, M. J. & Cairns, M. J. Pharmacological enrichment of polygenic risk for precision medicine in complex disorders. *Sci. Rep.* **10**, 879 (2020). **This study describes the rationale for the PES approach.**
111. Ghousaini, M. et al. Open Targets Genetics: systematic identification of trait-associated genes using large-scale genetics and functional genomics. *Nucleic Acids Res.* **49**, D1311–D1320 (2021).
112. Fang, H. et al. A genetics-led approach defines the drug target landscape of 30 immune-related traits. *Nat. Genet.* **51**, 1082–1091 (2019). **This study demonstrates how individual GWAS loci can be integrated with systems biology to repurpose drugs for immunological disorders.**
113. Sakaue, S. & Okada, Y. GREP: Genome for REPositioning drugs. *Bioinformatics* **35**, 3821–3823 (2019).
114. Gaspar, H. A., Hübel, C. & Breen, G. Drug Targetor: a web interface to investigate the human druggome for over 500 phenotypes. *Bioinformatics* **35**, 2515–2517 (2019).
115. Konuma, T., Ogawa, K. & Okada, Y. Integration of genetically regulated gene expression and pharmacological library provides therapeutic drug candidates. *Hum. Mol. Genet.* **30**, 294–304 (2021).
116. Emon, M. A., Domingo-Fernández, D., Hoyt, C. T. & Hofmann-Apitius, M. PS4DR: a multimodal workflow for identification and prioritization of drugs based on pathway signatures. *BMC Bioinforma.* **21**, 231 (2020).
117. Wishart, D. S. et al. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* **46**, D1074–D1082 (2018).
118. Freshour, S. et al. Integration of the Drug–Gene Interaction Database (DGIdb) with open crowdsourcing efforts. Preprint at *biorxiv* <https://doi.org/10.1101/2020.09.18.301721> (2020).
119. Manolio, T. A. et al. Finding the missing heritability of complex diseases. *Nature* **461**, 747–753 (2009).
120. Wu, M. C. et al. Rare-variant association testing for sequencing data with the sequence kernel association test. *Am. J. Hum. Genet.* **89**, 82–93 (2011).
121. Cordell, H. J. Detecting gene–gene interactions that underlie human diseases. *Nat. Rev. Genet.* **10**, 392–404 (2009).
122. McAllister, K. et al. Current challenges and new opportunities for gene–environment interaction studies of complex diseases. *Am. J. Epidemiol.* **186**, 753–761 (2017).
123. Hopkins, A. L. Network pharmacology: the next paradigm in drug discovery. *Nat. Chem. Biol.* **4**, 682–690 (2008).
124. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014).
125. Michailidou, K. et al. Association analysis identifies 65 new breast cancer risk loci. *Nature* **551**, 92–94 (2017).
126. Estrada, K. et al. Genome-wide meta-analysis identifies 56 bone mineral density loci and reveals 14 loci associated with risk of fracture. *Nat. Genet.* **44**, 491–501 (2012).
127. Eyre, S. et al. High-density genetic mapping identifies new susceptibility loci for rheumatoid arthritis. *Nat. Genet.* **44**, 1336–1340 (2012).
128. Evangelou, E. et al. Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. *Nat. Genet.* **50**, 1412–1425 (2018).
129. Nair, R. P. et al. Genome-wide scan reveals association of psoriasis with IL-23 and NF- κ B pathways. *Nat. Genet.* **41**, 199–204 (2009).
130. Teslovich, T. M. et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**, 707–713 (2010).
131. Nielsen, J. B. et al. Biobank-driven genomic discovery yields new insight into atrial fibrillation biology. *Nat. Genet.* **50**, 1234–1239 (2018).
132. Hemani, G. et al. The MR-Base platform supports systematic causal inference across the human genome. *eLife* **7**, e34408 (2018).
133. Bowden, J., Davey Smith, G., Haycock, P. C. & Burgess, S. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genet. Epidemiol.* **40**, 304–314 (2016).
134. Hartwig, F. P., Davey Smith, G. & Bowden, J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. *Int. J. Epidemiol.* **46**, 1985–1998 (2017).
135. Burgess, S., Foley, C. N., Allara, E., Staley, J. R. & Howson, J. M. A robust and efficient method for Mendelian randomization with hundreds of genetic variants. *Nat. Commun.* **11**, 376 (2020).
136. Pierce, B. L., Ahsan, H. & VanderWeele, T. J. Power and instrument strength requirements for Mendelian randomization studies using multiple genetic variants. *Int. J. Epidemiol.* **40**, 740–752 (2011).
137. Bowden, J. et al. Assessing the suitability of summary data for two-sample Mendelian randomization analyses using MR-Egger regression: the role of the *P* statistic. *Int. J. Epidemiol.* **45**, 1961–1974 (2016).
138. Burgess, S., Zuber, V., Gkatzionis, A. & Foley, C. N. Modal-based estimation via heterogeneity-penalized weighting: model averaging for consistent and efficient estimation in Mendelian randomization when a plurality of candidate instruments are valid. *Int. J. Epidemiol.* **47**, 1242–1254 (2018).
139. Verbanck, M., Chen, C.-Y., Neale, B. & Do, R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat. Genet.* **50**, 693–698 (2018).
140. Vilhjálmsson, B. J. et al. Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *Am. J. Hum. Genet.* **97**, 576–592 (2015).
141. Ge, T., Chen, C.-Y., Ni, Y., Feng, Y.-C. A. & Smoller, J. W. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat. Commun.* **10**, 1776 (2019).
142. Mavaddat, N. et al. Polygenic risk scores for prediction of breast cancer and breast cancer subtypes. *Am. J. Hum. Genet.* **104**, 21–34 (2019).

Acknowledgements

M.J.C. is supported by a National Health and Medical Research Council (NHMRC) Senior Research Fellowship (1121474) and a University of Newcastle Faculty of Health and Medicine Gladys M Brown Senior Fellowship. W.R.R. is supported by an Australian government research training programme stipend.

Author contributions

W.R.R. researched the literature. The authors contributed equally to all other aspects of the article.

Competing interests

W.R.R. and M.J.C. have filed a patent related to the use of the pharmagenetic enrichment score (PES) framework in complex disorders (WIPO Patent Application WO/2020/237314).

Peer review information

Nature Reviews Genetics thanks S. Burgess and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.