Title: Phenome-wide Mendelian randomization mapping the influence of the plasma proteome on complex diseases

Jie Zheng*+[1], Valeriia Haberland*[1], Denis Baird*[1], Venexia Walker*[1], Philip Haycock*[1], Alex Gutteridge[2], Tom G. Richardson[1], James Staley[1,3], Benjamin Elsworth[1], Stephen Burgess[3], Benjamin B. Sun[3], John Danesh[3], Heiko Runz[4], Joseph C. Maranville[5], Hannah M. Martin[6], James Yarmolinsky[1], Charles Laurin[1], Michael V. Holmes[1,7,8,9], Jimmy Liu[4], Karol Estrada[4], Linda McCarthy[2], Mark Hurle[2], Dawn Waterworth[10], Matthew R. Nelson[10], Adam S. Butterworth*[3], George Davey Smith*[1,11], Gibran Hemani*[1], Robert A. Scott*[2], Tom R. Gaunt*[1,11]

1 MRC Integrative Epidemiology Unit (IEU), Bristol Medical School, University of Bristol, Oakfield House, Oakfield Grove, Bristol, BS8 2BN, UK.
2 Glaxosmithkline, Gunnels Wood Road, Stevenage, Hertfordshire, SG1 2NY, UK.
3 MRC/BHF Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge CB1 8RN, UK.
4 Translational Genome Sciences, Biogen, 225 Binney Street, Cambridge, MA 02142, USA
5 Informatics and Predictive Sciences, Celgene Corporation, 200 Cambridgepark Dr, Cambridge, MA, USA.
6 School of Biological Sciences, University of Edinburgh, Edinburgh, EH9 3FF, UK.
7 Medical Research Council Population Health Research Unit, University of Oxford, Oxford, UK
8 Clinical Trial Service Unit & Epidemiological Studies Unit, Nuffield Department of Population Health, University of Oxford, Oxford, UK
9 National Institute for Health Research, Oxford Biomedical Research Centre, Oxford University Hospital, Oxford, UK
10 Genetics, GlaxoSmithKline, Collegeville, PA, USA
11 NIHR Bristol Biomedical Research Centre, Bristol, UK.

* Proteome MR writing group
+ Corresponding author

Abstract:

The human proteome is a major source of therapeutic targets. Recent genetic association analyses of the plasma proteome enable systematic evaluation of the causal consequences of variation in protein levels. Here, we estimated the effects of 1002 proteins on 225 phenotypes using two-sample Mendelian randomization (MR) and colocalization. Of 413 associations supported by evidence from MR, 139 (34%) were not supported by results of colocalization analyses, suggesting that genetic confounding may be widespread in naive phenome-wide association studies of proteins. Combining MR and colocalization evidence in cis-only analyses, we identified 105 putatively causal effects between 64 proteins and 51 downstream phenotypes (www.epigraphdb.org/pqtl). Evaluation of historic data from 268 drug development programmes showed that target-indication pairs with MR and colocalization support were considerably more likely to succeed, evidencing the value of this approach in identifying and prioritising potential therapeutic targets.

Despite increasing investment in research and development (R&D) in the pharmaceutical industry [1], the rate of success for novel drugs continues to fall [2]. Lower success rates make new therapeutics more expensive, reducing availability and increasing healthcare costs. Indeed, only one in ten targets taken into clinical trials reaches approval [2], with many showing lack of efficacy (~50%) or adverse safety profiles (~25%) in late stage clinical trials after many years of development [3,4]. For some diseases such as Alzheimer's disease, the failure rates are even higher [5].

To reduce the costs of drug development, approaches to prioritize target-indication pairs that are more likely to be successful at an early stage are much needed. It has previously been shown that genetic associations of phenotypes at loci similar to target-encoding genes of indications increase the likelihood of targets being efficacious [6]. Thus, systematically evaluating the genetic evidence in support of potential target-indication pairs is a potential strategy to prioritise development programmes. However, whilst systematic genetic studies have evaluated the putative causal role of both methylome and transcriptome on diseases [7] [8], no systematic causal inference study has yet been conducted to evaluate the role of the proteome in disease.

Plasma proteins play key roles in a range of biological processes, are frequently dysregulated in disease, and represent a major source of druggable targets [9] [10] [11]. Recently published genome-wide association studies (GWAS) of plasma protein levels have identified 3606 conditionally independent single nucleotide polymorphisms (SNPs) associated with 2656 proteins ('protein quantitative trait loci', pQTLs) in more than 1000 participants [12] [13] [14] [15] [16]. While conventional randomized controlled trials are time-consuming and costly, these genetic associations offer the potential to systematically test the causal effects of a large number of potential drug targets on the human disease phenome through Mendelian randomization (MR) [17]. In essence, MR exploits the random allocation of genetic variants at conception and their associations with disease risk factors to uncover causal relationships between human phenotypes [18], and has been described in detail previously [19] [20].

Here, we pool and cross-validate pQTLs from five recently published GWASs and systematically evaluate the causal effects of 968 plasma proteins on the human disease phenome, including 153 diseases and 72 risk factors available in the MR-Base database [21]. Results of all analyses are available in an open online database (www.epigraphdb.org/pqtl), with graphical and programmatic interfaces to enable rapid and systematic queries.

Results

Characterising genetic instruments for protein levels

Figure 1A summarises the genetic instrument selection and validation process, with expanded detail in **Supplementary Figure 1**. We curated 3606 SNPs associated with 2656 proteins from five studies [12 13 14 15 16]. After removing proteins and SNPs in the major histocompatibility complex (MHC) region and performing strict LD clumping we retained 2113 pQTLs associated with 1699 proteins (**Online Method**: *Instruments selection*; instruments listed in **Supplementary Table 1**). We then conducted a validation process in which we categorised the instruments into three tiers based on their utility for MR analysis (**Online Methods**: *Instrument validation*, **Supplementary Figure 2 and 3**). In summary we curated 1064 pQTLs for 955 proteins with highest relative level of reliability (tier 1, **Supplementary Table 1**), 62 pQTLs which exhibited effect heterogeneity across studies (where we could test it), indicating uncertainty in the reliability of one or all pQTLs (tier 2, **Supplementary Table 2**), and 987 non-specific pQTLs that were associated with more than five proteins (tier 3, **Supplementary Table 1**). Amongst the tier 1 pQTLs, 738 (69.4%) were acting in cis (within 500kb of the protein coding gene) and 326 were trans-acting. 66 proteins were influenced by both cis and one or more trans SNPs (**Supplementary Table 3**), and 149 proteins had more than one conditionally independent cis instrument[12], involving 368 cis SNPs (**Supplementary Table 4**).

Estimated effects of plasma proteins on human phenotypes

We undertook two-sample MR to systematically evaluate evidence for causal effects of 1002 plasma proteins (with tier 1 and tier 2 instruments) on 153 diseases and 72 risk factors (**Supplementary Table 5**, **Online Methods**). As cis-pQTLs were considered to have a higher biological prior for a direct and specific impact of the SNP upon the protein (compared to trans pQTLs), we grouped the MR analyses based on whether the instruments were acting in cis or trans. Genetically predicted associations between protein levels and other phenotypes may indicate causality (the protein causally influences the phenotype); reverse causality (genetic liability to an outcome influences the protein level); linkage disequilibrium between leading SNPs within protein and human phenotypes, or horizontal pleiotropy (the protein-phenotype association is not mediated by the target protein, but the dual associations are as a result of two distinct biological phenomena) (**Supplementary Figure 4**). To address these alternative explanations we conducted a set of sensitivity analyses designed to increase confidence in the MR association reflecting a causal effect of the protein on the phenotype: colocalization analysis [22] to investigate whether the genetic associations with both protein and phenotype shared the same causal variants; tests of reverse causality using bi-directional MR [23] and MR Steiger filtering [24 25]; and heterogeneity analyses for proteins with multiple instruments [26] (**Figure 1B**). In general, we observed 413 protein-trait associations with MR evidence using either cis or trans instruments. For these associations, 274 (66.3%) also showed evidence of colocalization, suggesting that one third of the MR findings could be driven by genetic confounding by LD between pQTLs and other causal SNPs.

Estimating protein effects on human phenotypes using cis pQTLs

In cis pQTL MR analyses, we identified 105 putatively causal effects of 64 proteins with 51 phenotypes (**Figure 2**, **Supplementary Table 6**, $P < 3.5 \times 10^{-7}$), with evidence of MR and

colocalization (posterior probability>80%) between the protein- and phenotype-associated signals. 10 of the 105 associations showed evidence of colocalization only after conducting conditional analysis [27] for proteins and human phenotypes in the cis region (**Supplementary Table 6**). For example, Haptoglobin (HP) did not show strong evidence of colocalization on LDL cholesterol (colocalization probability ≅ 0; data from GLGC consortium) due to multiple association peaks for LDL cholesterol in the HP region (**Supplementary Figure 5**). After applying conditional analysis to LDL cholesterol, we observed strong evidence of colocalization (colocalization probability = 0.997).

Evidence of potentially causal effects with colocalization was identified across a range of disease categories including anthropometric and respiratory phenotypes as well as cardiovascular and autoimmune diseases (**Supplementary Table 6**; see **Supplementary Note 1 and 2**). For example, higher levels of Cerebellin 4 Precursor (CBLN4) showed evidence of association with higher body mass. We also identified an association of lower levels of AGT with lower diastolic blood pressure, lower levels of ADAM Metallopeptidase Domain 19 (ADAM19) with lower forced expiratory volume in 1-second and lower Intercellular Adhesion Molecule 5 (ICAM5) level with lower risk of Crohn's disease (**Supplementary Table 6**).

There was less evidence of colocalization for 75 associations involving 50 proteins and 29 phenotypes (**Supplementary Table 7**). Most of these findings provide evidence that the genetic signals for protein levels and human phenotypes represent linkage disequilibrium (LD) between the pQTL and other SNPs that causally influence the phenotype within the region and highlight the importance of colocalization in such MR analyses. We found some examples showing modest evidence of colocalization, including some where the probability of colocalization were close to our threshold (posterior probability > 80%). In cases that the colocalization values were close to our threshold, power may be an issue; for example, the association between nephroblastoma Overexpressed (NOV) and fibroblastic disorders showed a colocalization probability of 75.6%.

Where pQTL studies identified multiple conditionally independent SNPs, we applied a MR model which takes into account the LD structure between conditionally independent SNPs in the cis region [28 29]. In this analysis, 21 of the 29 top associations identified using a single cis instrument had consistent evidence of association (P value cut off = 3.5×10^{-7}) and directions of effect using multiple cis instruments (**Supplementary Table 8A**), which enhanced the reliability of these MR findings. Although the heterogeneity analysis suggested that 9 of 21 top associations show some evidence of inconsistent SNP effects across the multiple cis instruments (Cochrane Q P value < 0.05; **Supplementary Table 8A**), we found that the direction of causal effects across the multiple cis instruments was always consistent using either single cis or multiple cis pQTLs as instruments for MR.

An illustration of the value of using multiple cis SNPs is seen in our analysis of microseminoprotein beta (MSMB) and prostate cancer risk. Serum levels of MSMB is a clinical biomarker of prostate cancer risk, diagnosis and disease monitoring and Sun *et al* reported that a cis pQTL associated with lower MSMB plasma levels is the lead prostate cancer susceptibility variant (rs10993994) [12], supporting a protective role for MSMB in prostate cancer. In our study, we used the same single cis instrument (rs10993994) and

confirmed this association in an independent sample from UK Biobank (OR = 0.758, for prostate cancer per SD change in MSMB levels, 95%CI= 0.703 to 0.818, P= 7.81 x 10^{-13}; **Supplementary Table 6**). We further applied MR using two conditional independent cis instruments for MSMB (rs10993994 and rs61847070) and found a similar result (OR= 0.766, 95%CI= 0.713 to 0.819, P= 4.48x10^{-14}; **Supplementary Table 8A**). Furthermore, bi-directional MR suggested little evidence of reverse causality for the genetic liability of prostate cancer on MSMB (**Supplementary Data 1**).

In regions with multiple signals, we performed colocalization analysis of conditionally distinct signals of proteins against disease traits. For example, Interleukin 23 Receptor (IL23R) levels showed two association peaks within the cis region (**Figure 3A** showing the two conditionally independent cis instruments, rs11581607 and rs3762318 identified by Sun et al [12]). MR analyses combining both instruments showed a strong association of IL23R with Crohn's disease (OR= 3.22, OR in Crohn's disease risk per SD change in IL23R, 95%CI= 2.93 to 3.53, P=6.93x10^{-131}; **Supplementary Table 8A**). In addition, there were 4 possible conditionally independent signals (conditional P value<1x10^{-7}) predicted for Crohn's disease in the same region (**Figure 3B**; Crohn's disease data from de Lange et al [30]). After adjusting for other distinct signals in the region, we observed colocalized association peaks between IL23R and Crohn's disease for the top IL23R signal (rs11581607) (**Figure 3C-D,** colocalization probability = 99.3%) and observed limited evidence for the second independent IL23R hit (rs7528804) (**Figure 3E-F;** colocalization probability = 62.9%). Given that naïve colocalization of the marginal associations in the region showed no evidence of shared signals (colocalization probability = 0), this example demonstrates both the complexity of the associations in this region and the importance of applying conditional analysis before colocalization when multiple distinct signals exist in the region.

In addition, we found 6 associations identified in the 'single cis' analysis that had weaker associations when analysed with multiple cis instruments (multiple-cis P value > 1x10^{-5}), representing less reliable MR findings (**Supplementary Table 6 and 7**), which reflected heterogeneity across instruments. However, in many cases multiple cis SNPs resulted in increased power and precision of the causal estimates and identified 10 new associations, which were not shown in the single cis analysis (**Supplementary Table 8B**). For example, using three conditionally independent cis instruments, we identified a potential novel association between higher chymotrypsinogen B1 (CTRB1) levels and lower risk of Crohn's disease (odds ratio [OR]= 0.928, for Crohn's disease per SD change in CTRB1 levels, 95%CI=0.903 to 0.954, P = 1.18 x 10^{-7}) (**Supplementary Table 8B**).

Due to epitope-binding artefacts driven by coding variants, some of the cis instruments could be artefactual [31], we therefore conducted a sensitivity MR analysis that excluded 123 Tier 1 cis instruments which are in the coding region. After this additional exclusion, we saw no compelling evidence for a large proportion of coding variant-driven pQTL associations being artefactual, at least on the basis that a comparable proportion showed strong evidence (>80%) of colocalization with an outcome phenotype association (70/164 compared to 75/180 for those pQTLs not led by coding variants). (**Supplementary Table 6 and 7**, filtered by column "VEP").

Using trans-pQTLs as additional instrument sources

Trans pQTLs are more likely to influence targets though pleiotropic pathways. For example, among the 1316 trans instruments we identified from 5 studies 73.5% were associated with more than 5 proteins, compared to 1.8 % of cis instruments (**Supplementary Table 1**). An illustration of applying colocalization analysis to identify potential mediators for association of Alpha 1-3-N-Acetylgalactosaminyltransferase And Alpha 1-3-Galactosyltransferase (ABO) protein on ovarian cancer is shown in **Supplementary Note 3**. However, trans pQTLs that overlap disease associations can highlight previously unsuspected candidate proteins through which genetic loci may influence disease risk [12], so we did extend MR analyses to include trans instruments, where they were associated with fewer than 5 proteins.

First, we aimed to further boost power to identify causal links by combining cis and trans instruments for 66 proteins that had both cis and trans pQTLs (noted as cis + trans analysis). However, none reached our pre-defined Bonferroni-corrected threshold, although two protein-phenotype associations provided some evidence at $P<1x10^{-5}$ (**Supplementary table 9**).

Next, we performed trans-only analyses of 293 proteins, and identified 158 associations with 44 phenotypes, all with evidence of colocalization (**Supplementary Table 10**). Some of those are consistent with established causal evidence from drug trials. For example, Protein C, Inactivator Of Coagulation Factors Va And VIIIa (PROC) is a target for approved drugs, such as activated protein C and warfarin, to treat venous thrombosis. There were no pQTLs found to be associated with PROC in the cis region (top associated SNP, rs6755028, with P value = $1.23x10^{-4}$; **Supplementary Figure 6**). However, in MR and colocalization analysis using PROC trans pQTL (rs867186), we found a strong association between PROC and deep venous thrombosis (DVT) (OR=1.21, OR of DVT risk per SD change in PROC level, 95%CI= 1.42 to 1.28, P= $1.27x10^{-10}$; probability of colocalization >0.9). In addition, the 44 associations involving 25 proteins and 25 phenotypes with less evidence of colocalization are listed in **Supplementary Table 11**.

Estimating protein effects on human phenotypes using pQTLs with heterogeneous effects across studies

As we were able to access full genome-wide results (rather than just "top" results) for the pQTLs identified in three of the pQTL GWA studies [12,13,14], we could check whether the same pQTL was observed in other studies. We examined any differences in effect size between studies using the pair-wise Z test (where we defined a Z statistic greater than 5 (equal to a P value of 0.001) as indicating strong evidence for heterogeneity). Of the 494 pQTLs where we could test for heterogeneity across studies, we found that 144 (29.1%) showed evidence of difference in effect size across studies (so called Tier 2 instruments). Recognising that lack of replication and effect heterogeneity does not preclude at least one of these effects being genuine, we performed MR analyses using the most significant SNP across studies and report the findings with caution. Some proteins that are targets of approved drugs were found in this analysis, such as IL6R (targeted by tocilizumab) on rheumatoid arthritis [32], and CHD, the latter being supported by treatment trials of therapies targeting IL1B [33], which is upstream of IL6R (**Supplementary Table 12**).

As another test of heterogeneity across studies, where the same protein was measured in two or more studies, we performed colocalization analysis of each protein (in one study) against the same protein (in another study) for all studies in which we had access to full summary results. Of 41 pQTLs where we could test for colocalization of the same protein in another pQTL study, we found 25 had little evidence of colocalization, which suggested either two different signals within the test region or the protein have pQTL in one study but not in the other study. For the first case, since one of the two distinct signals may be a genuine pQTL, we therefore performed MR analysis of these 25 pQTLs using instruments from each study separately. The findings of this analysis were reported in **Supplementary Table 13**.

Orienting causal direction in protein-disease associations

For potential associations between proteins and phenotypes identified in the previous analyses (single cis, cis + trans and trans-only analysis), we undertook two sensitivity analyses to eliminate spurious results due to reverse causation — bi-directional MR [23] and Steiger filtering [24 25] (details of the difference of the two methods and when to apply them can be found in the Online Methods section: Distinguishing causal effects from reverse causality). We found no strong evidence of reverse causality for diseases on protein level changes, in either bi-directional MR analyses or Steiger filtering. More details of the causal direction results can be found in **Supplementary Note 4** and **Supplementary Data 1**.

Drug target prioritisation and repositioning using phenome-wide MR

Recent MR studies highlight the value of hypothesis-free ("phenome-wide") MR in building a comprehensive picture of the causal effects of risk factors on disease outcomes [8 34 35]. Given that human proteins represent the major source of therapeutic targets, we sought to mine our results for targets of molecules already approved as treatments or in ongoing clinical development, and which might represent promising candidates for repositioning. We compared MR findings for 1002 proteins against 225 phenotypes with historic data on clinical trials for target-indication pairs in Citeline's PharmaProjects. Of 27064 target-indication pairs that had gone into clinical or pre-clinical development for indications, for 2024 pairs we had a pQTL for the protein and captured a disease phenotype similar to the indication for which the drug had been trialled (or under pre-clinical development). 268 of the 2024 pairs were for approved (73) or failed (195) drugs (**Supplementary Table 14**), where the remaining 1756 pairs were for drug under development or for new targets. Of the 73 target indication pairs for approved drugs, we observed positive MR and colocalization evidence for 8 of them (**Supplementary Table 14**). Of 195 target indication pairs that had failed to gain approval after clinical trials, none of them had MR and colocalization evidence. A Fisher's exact test indicated that protein-phenotype associations with MR and colocalization evidence were more likely to be successful drugs (Odds ratio = 26.4; 95%CI: 4.0 to 580; P value = 4.95×10^{-5}) (**Table 1**). Although we acknowledge the limited sample size of the test set, this does support the utility of pQTL MR analyses as a source of target identification and validation.

For approved drugs, 8 protein-phenotype associations with robust MR and colocalization evidence were established target-indication pairs, including proprotein convertase

subtilisin/kexin type 9 (PCSK9) inhibitor (target for evolocumab) for hypercholesterolemia and hyperlipidemia, angiotensinogen (AGT) for hypertension, interleukin 6 receptor (IL6R) for rheumatoid arthritis, PROC (which is modified by warfarin) for deep venous thrombosis, IL12B for psoriatic arthritis and psoriasis and TNF Receptor Superfamily Member 11a (TNFRSF11A) for osteoporosis (**Supplementary Table 15**). We further predicted the potential target-mediated repositioning opportunities of a few marketed drugs. For example, our phenome-wide MR and colocalization analysis further support the effect of inhibition of IL6R as a valid therapeutic strategy for lowering risk of coronary heart disease (monoclonal antibody inhibitors of IL6R are already licensed for the treatment of rheumatoid arthritis) and tumor necrosis factor receptor superfamily member 11a (TNFRSF11A, also termed 'RANK' which is involved in osteoclast differentiation) for treatment of Paget's disease [36] (**Supplementary Table 16**).

We also evaluated drugs in current clinical trials and identified 8 protein-phenotype associations with MR and colocalization evidence that corresponded to target-indication pairs with therapeutics in clinical trials or in preclinical experiments. Examples include lipoprotein(a) (LPA) for blood lipids and angiopoietin like 3 (ANGPTL3) for plasma lipids (**Supplementary Table 17**).

Our results also offer the potential to identify drug repositioning opportunities for drugs under investigation within current clinical trials. We identified 40 existing drug targets associated with 51 phenotypes other than the primary indication (**Supplementary Table 18**). Our phenome-wide MR analysis suggests that lifelong higher urokinase-type plasminogen activator (PLAU) levels are associated with lower inflammatory bowel disease (IBD) risk (OR=0.75, 95%CI= 0.69 to 0.83, P= 1.28×10^{-9}; **Supplementary Figure 7**). Urokinase was initially developed for use as a thrombolytic in the treatment of acute myocardial infarction and ischaemic stroke, and thus a target-mediated adverse effect is an increase in bleeding and potential haemorrhage. While our data suggest that urokinase might be protective in the aetiology of IBD, a careful risk benefit assessment would be required as part of an investigation on whether drugs targeting urokinase might be repurposed for the treatment of IBD.

Discussion

MR analysis of molecular phenotypes against disease phenotypes provides a promising opportunity to validate or prioritise novel or existing drug targets through prediction of efficacy and potential on-target beneficial or adverse side-effects [37 38 39]. Our phenome-wide MR study of the plasma proteome employed five well-powered pQTL studies to robustly identify and validate genetic instruments for thousands of proteins. We used these instruments to evaluate the potential effects of modifying protein levels on hundreds of complex phenotypes available in MR-Base (www.mrbase.org) [21] in a hypothesis free approach [17], and confirmed that protein-phenotype associations with both MR and colocalization evidence predicted the likelihood of a particular target-indication pair being successful than protein-phenotype pairs without such evidence. Collectively, we underline the important role for MR combined with colocalization as an evidence source to support drug discovery and development.

In particular, we noted the important role of a number of sensitivity analyses following the initial MR in order to distinguish causal effects of proteins from associations driven by horizontal pleiotropy, genetic confounding through linkage disequilibrium [20] and reverse causation [23 24]. Of note, of 413 observed associations with MR evidence, only 274 (66.3%) showed strong evidence of colocalization, suggesting that at least part of the initial findings could be driven by genetic confounding through LD between pQTLs and other disease-causal SNPs. Thus, we suggest that investigation of colocalization is vital for improved causal inference when conducting the MR analyses of molecular traits (such as DNA methylation, gene expression and proteins). One caveat is that reliance on such colocalization may underestimate the number of 'true' MR findings due to 1) lack of power, e.g. owing to limited sample size for the protein and/or phenotype GWASs; and/or, 2) lack of mature colocalization pipeline to deal with multiple association peaks in a region, which collectively may lead to false negatives on colocalization. Recent MR studies suggested an increased precision and power of MR estimates using multiple independent signals from a single gene region [28 29], which aligned with our findings, e.g. the MSMB and prostate cancer association. To deal with the challenge of applying colocalization analysis to these regions, we applied conditional analysis to GWASs of protein and human phenotypes and used the conditionally independent SNP effects for colocalization analysis, which identified further evidence of colocalization for 10 additional protein-phenotype associations, further strengthening the evidence-base underpinning our claims of causality. It is important to consider applying the integrative conditional and colocalization analysis to regions with multiple cis pQTLs, which we demonstrated a case of IL23R and Crohn's disease.

In addition, our study improved upon some previous MR studies of omics [7 8] by utilising both cis and trans instruments. Several of the approaches (single-cis, multi-cis and trans-only) yielded informative results. For example, the PROC and deep venous thrombosis example using trans pQTLs (see Results) as well as the identification of a potential mediator for the well-known effect of ABO on ovarian cancer (see **Supplementary Note 3**). Trans-pQTL findings should be interpreted with more caution unless a link can be drawn between the loci and the target protein, or there are a good number of trans pQTLs the estimated effects of which show little evidence of heterogeneity. As a well-known example, using a trans instrument within IL6R as an instrument to test the association for CRP on cardiovascular disease yields incorrect causal interpretation [37].

In an evaluation of the potential of MR to inform drug target prioritisation, we demonstrated that pQTL MR and colocalization evidence for a target-indication pair predicts a positive clinical outcome and higher likelihood of approval (OR = 26.4, P value =4.95×10^{-5}). Applying this approach to drug targets in development, we highlighted 8 examples where we predict a successful outcome of ongoing clinical trials. One of the limitations of our approach is the lack of comprehensive coverage of genetic data for all outcomes for which drugs are in development, as well as in our inability to instrument the entire genome through pQTLs. As such, ongoing expansions in the scale and diversity of GWAS will greatly enable the vision enabled by the current study, as will availability of summary statistics upon publication.

An important limitation of this work is that protein levels are known to differ between tissues. In this study, we have estimated the role of protein levels measured in plasma on a range of complex diseases but are unable to assess the relevance of protein levels in other tissue. Whilst expression QTL (eQTL) studies highlight a major proportion of eQTLs being shared across tissues [37], pQTL studies have not yet been performed as systematically across tissues. However, it is encouraging that we identify associations across a range of disease categories, including for psychiatric diseases for which we may expect key proteins to function primarily in the brain. Another potential limitation is the limited coverage of the proteome afforded by current technologies, leaving the possibility of undetected pleiotropy of instruments. While cis pQTLs are less likely to be prone to horizontal pleiotropy than trans pQTLs, it is well known from study of gene expression that cis variants can influence levels of multiple neighbouring genes and hence the same is likely to be true for protein levels [41]. Similarly, future larger GWAS of the plasma proteome are likely to uncover many more variant-protein associations, increasing the apparent pleiotropy of many pQTLs.

Conclusion

In conclusion, this study systematically identified 105 putatively causal effects between the plasma proteome and the human phenome using the principles of Mendelian randomization and colocalization. These observations support, but do not prove, causality, as potential horizontal pleiotropy remains an alternative explanation. Our study provides an open resource to prioritise potential new targets on the basis of MR evidence and a valuable resource for evaluation of both efficacy and repurposing opportunities by phenome-wide evaluation of putative on-target associations.

Author contribution

JZ, VH and DB performed the Mendelian randomization analysis; JZ and DB performed the colocalization analysis; JZ performed the conditional analysis; VH, BE and TRG developed the web browser; JZ and VW performed the drug target prioritisation and enrichment analysis. AG, TGR, BE, HM, JY, CL conducted supporting analyses; JS, BBS, JD, HR, JCM provided key data and supported the MR analysis; JL, KE, LM, MVH, MH, DW, MRN reviewed the paper and provided key comments. JZ, VH, DB, VW, PH, AB, GDS, GH, RAS and TRG wrote the manuscript. JZ, TRG, RAS, GH, GDS and PH conceived and designed the study and oversaw all analyses.

Acknowledgements

We are extremely grateful to all the families who took part in the ALSPAC study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses.

The UK Medical Research Council and Wellcome (Grant ref: 102215/2/13/2) and the University of Bristol provide core support for ALSPAC. This publication is the work of the authors and Jie Zheng will serve as guarantor for the contents of this paper. A comprehensive list of grants funding is available on the ALSPAC website (http://www.bristol.ac.uk/alspac/external/documents/grant-acknowledgements.pdf); This research was specifically funded by MC_UU_00011/4. The work is supported by Cancer Research UK grant, Integrative Cancer Epidemiology Programme (C18281/A19169). The work/Tom Gaunt is supported by CRUK (C18281/A19169). GH is funded by the Wellcome Trust and the Royal Society [208806/Z/17/Z]. MVH is supported by a British Heart Foundation Intermediate Clinical Research Fellowship (FS/18/23/33512) and the National Institute for Health Research Oxford Biomedical Research Centre. This study was funded/supported by the NIHR Biomedical Research Centre at University Hospitals Bristol NHS Foundation Trust and the University of Bristol (GDS and TRG). The views expressed in this publication are those of the author(s) and not necessarily those of the NHS, the National Institute for Health Research or the Department of Health and Social Care. This work was supported by the Elizabeth Blackwell Institute for Health Research, University of Bristol and the Medical Research Council Proximity to Discovery Award.

We gratefully acknowledge all studies and databases that have made their GWAS summary data available for this study: arcOGEN (Arthritis Research UK Osteoarthritis Genetics), BCAC (the Breast Cancer Association Consortium), C4D (Coronary Artery Disease Genetics Consortium), CARDIoGRAM (Coronary ARtery DIsease Genome wide Replication and Meta-analysis), CKDGen (Chronic Kidney Disease Genetics consortium), DIAGRAM (DIAbetes Genetics Replication And Meta-analysis), EAGLE (EArly Genetics and Lifecourse Epidemiology Consortium), EAGLE Eczema (EArly Genetics and Lifecourse Epidemiology Eczema Consortium), EGG (Early Growth Genetics Consortium), ENIGMA (Enhancing Neuro Imaging Genetics through Meta Analysis), GCAN (Genetic Consortium for Anorexia Nervosa), GEFOS (GEnetic Factors for OSteoporosis Consortium), GIANT (Genetic Investigation of ANthropometric Traits), GIS (Genetics of Iron Status consortium), GLGC (Global Lipids Genetics Consortium), GliomaScan (cohort-based genome-wide association study of glioma), GPC (Genetics of Personality Consortium), GUGC (Global Urate and Gout consortium),

HaemGen (haemotological and platelet traits genetics consortium), IGAP (International Genomics of Alzheimer's Project), IIBDGC (International Inflammatory Bowel Disease Genetics Consortium), ILCCO (International Lung Cancer Consortium), IMSGC (International Multiple Sclerosis Genetic Consortium), ISGC (International Stroke Genetics Consortium), MAGIC (Meta-Analyses of Glucose and Insulin-related traits Consortium), MDACC (MD Anderson Cancer Center), MESA (Multi-Ethnic Study of Atherosclerosis), Neale's lab (a team of researchers from Dr Benjamin Neale's group, who made the UK Biobank GWAS summary statistics publically available), OCAC (Ovarian Cancer Association Consortium), IPSCSG (the International PSC study group), NHGRI-EBI GWAS catalog (National Human Genome Research Institute and European Bioinformatics Institute Catalog of published genome-wide association studies), PanScan (Pancreatic Cancer Cohort Consortium), PGC (Psychiatric Genomics Consortium), Project MinE consortium, ReproGen (Reproductive ageing Genetics consortium), SSGAC (Social Science Genetics Association Consortium), TAG (Tobacco and Genetics Consortium), TRICL (Transdisciplinary Research in Cancer of the Lung consortium) and UK Biobank.

JZ acknowledges his grandmother ChenZhu for all her support, may she rest in peace.

Conflict of interests

AG, LM, MH, DW, MN and RAS are employees and shareholders in GlaxoSmithKline. HR, JL and KE are employees and shareholders in by Biogen. VH is employed on a grant funded by GlaxoSmithKline. DB is employed on a grant funded by Biogen. TG, GH and GDS receive funding from GlaxoSmithKline and Biogen for the work described here. AB has received grants from Merck, Novartis, Biogen, Pfizer and AstraZeneca.

Online methods

Instrument selection

pQTLs from five GWAS (Sun *et al*, Emilsson *et al,* Suhre *et al*, Folkersen *et al* and Yao *et al*) [12 13 14 15 16] were used as genetic instruments to estimate the causal effect of plasma protein levels on human diseases and other phenotypes (**Supplementary Figure 1**).

We used the following criteria to select pQTL instruments:

- We selected SNPs that were associated with any protein (using a P value $\leq 5x10^{-8}$) in at least one of the five studies, including both cis and trans pQTLs.
- Due to the complex LD structure of SNPs within the human Major Histocompatibility Complex (MHC) region, we removed SNPs and proteins coded for by genes within the MHC region (chr6: from 26Mb to 34Mb).
- We then conducted linkage disequilibrium (LD) clumping for the instruments with the TwoSampleMR R package to identify independent pQTLs for each protein. We used $r^2 < 0.001$ as the threshold to exclude dependent pQTLs in the cis (or trans) gene region.

After instrument selection, we mapped SNPs to genome build GRCh37.p13 coordinates (**Supplementary Table 1**). The instrument selection process, and the number of instruments for proteins at each step in the process, is illustrated in **Supplementary Figure 1**.

Instrument validation

For the 2113 instruments we selected in the *Instrument selection* section, we further classified them into three groups (noted as tier 1, tier 2 and tier 3 instruments) using two major instrument filtering steps: a pleiotropy test and a consistency test. More details of instrument validation, including harmonization of proteins and instruments and statistical tests for consistency can be found in **Supplementary Note 5**.

Test estimating instrument specificity

Absence of horizontal pleiotropy is one of the core assumptions for MR. This assumes that the genetic variant should only be related to the outcome of interest through the instrumented exposure. We noted that some SNPs were associated with more than one protein. For example, *APOE* SNP rs7412 is associated with a set of proteins such as ADAM11, APBB2 and APOB. We plotted a histogram of the number of proteins each pQTL was associated with (**Supplementary Figure 8**) and considered instruments associated with more than 5 proteins as non-specific for any particular protein and highly pleiotropic and assigned them as Tier 3 instruments (which were excluded from all analyses). For instruments associated with fewer than (or equal to) 5 proteins, we reported the number of proteins each of them (and their proxies with LD $r^2>0.5$) was associated with to indicate the level of pleiotropy.

Consistency test estimating instrument heterogeneity across studies

We noted some examples where SNPs were reported to be associated with a protein in one study but did not reach the genome-wide p-value threshold for statistical significance in other studies including the same protein. In these instances, we investigated whether this reflected no statistical evidence of association (in which case, this inconsistency may indicate potentially artefactual associations) or simply fluctuation of association strength with directionally consistent signals in both studies (which would provide supporting

evidence for an instrument). Among the 2113 pQTLs selected as instruments, we looked up available protein GWAS results (Sun *et al*, Suhre *et al* and Folkersen *et al* with full GWAS summary statistics; Yao *et al* and Emilsson *et al* with pQTLs only) and found 1062 pQTLs (or proxies with r2>0.8) with association information in at least two studies (**Supplementary Table 19**). We then tested the beta-beta correlation using Pearson correlation function in R. The results of the beta-beta correlations of SNP effects for each pair of studies and the number of SNPs included in each correlation analysis can be found in **Supplementary Table 20**. More details of the consistency test can be found in **Supplementary Note 5**.

We performed two consistency tests on the instruments which were present across studies: 1) a *heterogeneity test* using a pair-wise Z statistic to investigate whether there was statistical evidence of heterogeneity between effect sizes in different studies (for all pQTL studies included in our analysis: 1) effect sizes were always in SD unit; 2) using similar sets of covariates). If the Z score was greater than 5 (equal to a P value of 0.001), we considered the instrument to have strong evidence of heterogeneity indicating inconsistency of effect sizes between studies; 2) the *colocalization analysis* estimates the posterior probability (PP) of the same protein measured in different studies sharing the same causal pQTL within a 2Mb window around the pQTL with the smallest P value. The default priors of colocalization analysis were used here. A lack of evidence (i.e. PP<80%) in this analysis would suggest that the pQTL reported in the two studies do not share the same causal signals within the region, therefore are not consistent between the studies. The colocalization analysis was conducted using the "coloc" R package [22]. For instruments with SNP association information in both Sun *et al* and Folkersen *et al*, we were able to conduct colocalization analysis. However, due to lack of sufficient SNP coverage it was not possible to conduct colocalization analysis to compare the pQTLs from the Emilsson *et al*, Suhre *et al* and Yao *et al* studies. We therefore conducted a linkage disequilibrium (LD) check for these pQTLs instead. For proteins measured in multiple studies, we estimated the LD between the sentinel variant for each pQTL from one study and the top 30 associated SNPs of the other study in the same region. For pQTLs that showed only weak LD ($r^2 < 0.8$) with any of the top 30 associated SNPs in the other study, we considered these pQTLs did not share the same causal SNP in the region and therefore had inconsistent instruments.

Instruments showing evidence of high heterogeneity across studies using either the pair-wise Z test (pair-wise Z > 5) or colocalization analysis (PP<80%), were flagged as Tier 2 instruments.
Recognising that lack of replication and effect heterogeneity does not preclude at least one of these effects being genuine, we therefore used these instruments separately for the follow-up genetic analyses (**Supplementary Table 2**) and reported the findings with caution. We designated instruments passing both pleiotropy and consistency tests as Tier 1 instruments and used them as primary instruments for the MR analysis.

Identifying cis and trans instruments

We further split Tier 1 instruments into two groups: 1) *cis-acting pQTLs* within a 500Kb window from each side of the protein coding genes were used for the initial MR analysis (defined as the cis-only analysis) [20]; (2) *trans-acting pQTLs* outside the 500Kb window of the protein coding gene were designated as trans instruments. Whilst trans instruments may be more prone to pleiotropy, their inclusion could increase statistical power as well as the

scope of downstream sensitivity analyses. (e.g. tests for heterogeneity between instruments). Therefore, for the proteins with cis instruments, we also looked for additional trans instruments and if these were available, we conducted further MR analyses using both sets of instruments (defined as the "cis + trans" analysis).
For cis instruments, we looked up their predicted consequence via Variant Effect Predictor [42] (VEP: https://www.ensembl.org/info/docs/tools/vep/index.html) hosted by Ensembl. We designated coding variants, as epitope-binding artefacts driven by coding variants may yield artefactual cis pQTLs [31]. We also conducted a sensitivity MR analysis that excluded cis instruments which are in the coding region to further avoid the potential issue of epitope-binding artefacts driven by coding variants.

Human phenotype selection

We obtained effect estimates for the association of the pQTLs with complex human phenotypes using GWAS summary statistics which were included in the MR-Base database (http://www.mrbase.org). We used the following inclusion criteria to select complex phenotypes to be analysed:

- The GWAS with the greatest expected statistical power (e.g. largest sample size / number of cases) when multiple GWAS records of the same disease / risk factor were available in MR-Base.
- GWAS with betas, standard errors and effect alleles for all tested variants (i.e. full GWAS summary statistics available)

Diseases were defined as primary outcomes. Risk factors were defined as secondary outcomes. After selection, 153 diseases and 72 risk factors (such as lipids and glucose phenotypes) were included as outcomes for the MR analyses (**Supplementary Table 5**).

Causal inference and sensitivity analyses

The following sections describe two-sample MR analyses using single or small numbers of instruments on 153 diseases and 72 risk factors (**Supplementary Table 5**). Positive associations between genetic instruments and phenotypes may indicate a number of potential scenarios: 1) the protein has a causal effect on the phenotype (the scenario of causality we wish to identify), 2) that the phenotype has a causal effect on protein (the reverse causality scenario), 3) confounding through linkage disequilibrium between pQTLs and variants associated with the phenotype (for simplicity we refer to this as the 'linkage disequilibrium scenario') or 4) that the pQTL shares causal variants with the phenotype, but the association of the pQTL with the phenotype is not mediated by the hypothesised protein target (the 'horizontal pleiotropy" scenario) (see **Supplementary Figure 4**). Most of the current sensitivity analysis methods such as MR Egger regression [43] and Weighted Median [42] need a large number of independent instrumental SNPs in order to test for pleiotropy. Due to the small number of independent pQTLs available per protein we were therefore unable to implement these sensitivity analyses. To identify possible violations of assumptions of MR and to distinguish between the aforementioned scenarios, we therefore conducted the following sensitivity analyses: colocalization analysis [22], tests for heterogeneity between instrumental SNPs [26], bi-directional MR [23] and Steiger filtering [24] [25] (Methods section – Causal inference and sensitivity analyses) (**Figure 1B**).

Estimating the causal effects of protein levels on human phenotypes using MR

In the initial MR analysis, proteins were treated as the exposures and 225 complex human phenotypes as the outcomes (**Supplementary Figure 1 –** Estimate putative causal relationship). Due to high correlation amongst some of the tested phenotypes (e.g. coronary heart disease (CHD) and myocardial infarction), we used the PhenoSpD method [45 46 47] to provide a more appropriate estimate of the number of independent tests. We selected a p-value threshold of 0.05, corrected for the number of independent tests, as our threshold for prioritising MR results for follow up analyses (number of tests= 142,857; $P< 3.5x10^{-7}$).

MR analysis using single locus instruments

Firstly, the strongest cis pQTL variants for each protein was used as the instrumental variable (described as 'single cis' analysis). The Wald ratio [48] method was used to obtain MR effect estimates. In this analysis, the MR effect estimates were sensitive to the particular choice of pQTLs, since only the most strongly associated SNPs within each genomic region were used as instruments. Burgess *et al* recently suggested that more precise causal estimates can be obtained using multiple genetic variants from a single gene region, even if the variants are correlated [29 28] . Sun *et al* reported proteins with multiple cis instruments[12], so after quality checking and LD clumping (r^2<0.6), we used the remaining cis SNPs against all 225 phenotypes to further evaluate the MR findings from our initial MR analysis and identify potential novel associations (described as 'multiple cis' analysis) (**Supplementary Table 4**). A generalised inverse variance weighted (IVW) model considering the LD pattern between the multiple cis SNPs was used to estimate the MR effects. In this analysis, weights for the contribution of each SNP were obtained using pairwise LD (r^2) calculations obtained from the 1000 Genomes European ancestry reference samples.

MR analysis using multi-locus instruments

Among the measured proteins reported in Sun *et al*, 34% had both cis and trans pQTLs and 30% had only trans pQTLs [12]. Trans pQTLs that overlap disease association loci can provide information about previously unsuspected candidate proteins [12]. Also, using both cis and trans instruments can provide additional accuracy and statistical power to detect causal effects [49]. Therefore, as well as MR using only cis pQTLs, we also conducted MR on proteins with both cis and trans pQTLs (noted as the cis + trans MR analysis) and proteins with only trans pQTLs (noted as trans-only analysis). In the cis + trans MR analysis, we tested the protein-phenotype associations of 66 proteins with both cis and trans instruments. The IVW method [50] was used to obtain MR effect estimates. In the trans-only MR analysis, we used 351 trans instruments for 298 proteins. The IVW method was used when two or more trans instruments were included in the analysis, whereas the Wald ratio method was used when only one trans instrument was included in the analysis.

MR analysis software

The majority of MR analyses (including Wald ratio, IVW, single SNP MR, bi-directional MR, MR Steiger filtering and heterogeneity test across multiple instruments) were conducted using the MR-Base TwoSampleMR R package (github.com/MRCIEU/TwoSampleMR [21]). The IVW analysis considering LD pattern was conducted using the MendelianRandomization R package (https://cran.r-project.org/web/packages/MendelianRandomization/index.html [51]). The MR results were plotted as forest plots and Miami plots using code derived from the ggplot2 package in R (https://cran.r-project.org/web/packages/ggplot2/index.html).

Distinguishing causal effects from genomic confounding due to linkage disequilibrium

Results that survived the multiple testing threshold in the MR analysis were evaluated using a stringent Bayesian model (colocalization analysis) to estimate the posterior probability (PP) of each genomic locus containing a single variant affecting both the protein and the phenotype [22] (**Supplementary Figure 1 –** Distinguishing causal effects from confounding due to linkage disequilibrium). The default priors were used for the analysis. A PP > 80% in this analysis would suggest that the two association signals are likely to colocalize within the test region. For protein and phenotype GWAS lacking sufficient SNP coverage or missing key information (e.g. allele frequency or effect size) in the test region, we conducted a LD check for the sentinel variant for each pQTL against the 30 strongest SNPs in the region associated with the phenotype as an approximate colocalization analysis. r^2 of 0.8 between the sentinel pQTL variant and any of the 30 strongest SNPs associated with the phenotype was used as evidence for approximate colocalization. For all MR top findings, we treated colocalised findings (PP>=80%) as "Colocalised" and LD checked findings (r^2>=0.8) as "LD checked"; other findings that did not pass the colocalization or LD check analysis were annotated as "Not colocalised". For findings given a "Not colocalised" flag, we further controlled the possible influence of multiple conditionally independent signals within the genomic region. A two-step conditional analysis was applied using the GCTA-COJO package [27], with genotype data from mothers in the Avon Longitudinal Study of Parents and Children (ALSPAC) as the LD reference panel [52,53] (a description of the ALSPAC cohort can be found in **Supplementary Note 6**). We conducted colocalization analysis using the joint SNP effects for the phenotype (e.g. CHD) conditioned on either the top phenotype-associated SNP within 1MB window around the sentinel pQTL SNP (noted as "On top hit" in **Supplementary Table 6 and 10**) or on the second strongest phenotype-associated SNP (noted as "On second hit"). For MR findings using multiple instruments (e.g. cis + trans analysis), we tested each pQTL with the phenotype separately. Only if all pQTLs colocalised with the phenotype at r^2>=0.8 did we treat this finding as colocalised.

Heterogeneity test of MR findings

For MR analyses using two or more instruments, we conducted heterogeneity tests to estimate the variability in the causal estimates obtained for each SNP (i.e. how consistent is the causal estimate across all SNPs used as separate instruments) (**Supplementary Figure 1 —** Consistency of the causal estimate across all SNPs). The Cochran's Q test statistic was calculated for the IVW analyses, which is expected to be chi-squared distributed with number of SNPs minus one degrees of freedom [26]. Lower heterogeneity suggests a lower chance of violations of assumptions in MR estimates, such as the presence of confounding through horizontal pleiotropy [54].

Distinguishing causal effects for proteins on phenotypes from reverse causality

With sufficiently large sample sizes, a SNP associated with an outcome through a mediating exposure could reach the conventional threshold for statistical significance in both the outcome and exposure GWAS, for example lipid on bmd . Therefore, using such thresholds to define instruments could lead to situations where the instrumental SNP influences the hypothesised exposure via the hypothesised outcome (i.e. the hypothesised outcome actually has a causal effect on the hypothesised exposure and not vice versa). In order to

mitigate the potential impact of this limitation, we used two approaches to identify directions of causality: bi-directional MR and Steiger filtering.

Reverse Mendelian randomization

For associations between proteins and phenotypes identified in the MR analysis, we applied bi-directional MR to evaluate evidence for causal effects in the reverse direction by modelling complex phenotypes as our exposure and plasma protein level as our outcome [23]. Instruments for complex phenotypes were selected based on a threshold of $P < 5 \times 10^{-8}$ from GWAS after LD clumping to identify independent variants. The IVW method was applied to estimate the causal effects of phenotypes on proteins where more than one instrument was available, otherwise the Wald ratio was used. MR-Egger [43] was used as a sensitivity analysis to test for potential pleiotropic effects.

Identifying the direction of effects for instruments using Steiger filtering

Due to lack of sufficient SNP association information (e.g. allele information, effect size, standard error) for some pQTL studies, it was not possible to conduct bi-directional MR using all proteins as outcomes. Therefore, we conducted Steiger filtering as an alternative method to test the directionality of protein-phenotype associations. The Steiger method [55] has been implemented in the TwoSampleMR R package [21] to assess directionality of instrument-outcome associations [24 25]. This approach infers the causal direction between two phenotypes using a very simple inequality. Given phenotype A causes phenotype B then we would expect that:

$$\sum_{i=1}^{M} cor(g_i, A)^2 > \sum_{i=1}^{M} cor(g_i, B)^2$$

because $cor(gi, B)^2 = cor(A, B)^2 * cor(gi, A)^2$, where "cor" denotes correlation, and the vector g_i is the ith of M SNPs that associated with phenotype A.

The process of choosing valid instruments using Steiger filtering follows these steps:

1. Select the top findings from all five studies using a p-value threshold of 3.5×10^{-7} (which is the Bonferroni P value threshold of the MR analysis).
2. Classify instruments in each MR analysis based on Steiger filtering:
   - 'TRUE': evidence for causality in the expected direction i.e. protein precedes phenotype.
   - 'FALSE': evidence for causality in the reverse direction i.e. phenotype precedes protein. Instruments with 'FALSE' were removed from the sensitivity analysis.
   - 'NA': no result (due to insufficient summary data from the study to estimate the SNP-trait correlation, e.g. missing effect allele frequencies in the outcome data or missing numbers of cases and controls for binary phenotypes).

For disease phenotypes, we estimated the variance explained on the liability scale. Based on step 2, we set up a flag (categorical variable) to record the direction of the effects of the SNPs using Steiger filtering.

Drug target validation and repositioning

Approved drug targets have previously been shown to be enriched for gene-phenotype associations [6]. We therefore wished to assess whether approved drug targets were enriched

for protein-phenotype associations, as obtained in the present study using MR. We assessed the support for approved drug targets among our MR findings using Fisher's exact test. Target-indication pairs for successful and failed drugs were identified using a manually annotated version of PharmaProjects database from Citeline (https://pharmaintelligence.informa.com/). The phenotypes used in the MR analyses and the indications listed in Citeline's PharmaProjects were then manually mapped to MeSH headings as a common ontology. This allowed us to match the protein-phenotype associations with corresponding target-indication pairs. To improve this matching, we implemented a similarity matrix, derived from all MeSH headings in the manual mapping, and retained matches with a relative similarity greater than 0.7 for our analyses (the similarity matrix was described in Nelson *et al* [6]). We then conducted Fisher's exact test to compare whether the target-indication pair represented a successful or failed drug against whether there was a signal or not for the corresponding protein-phenotype pair among our MR findings. For the purposes of this test, a signal was defined as an MR result with a p-value less than 3.5×10^{-7} (which is the Bonferroni P value threshold of the MR analysis) and supported by evidence from colocalization analysis. For cells in the 2x2 contingency table containing zero(s), we added 1 to each cell so that Fisher's exact test would return an odds ratio for the enrichment analysis. Fisher's exact test was implemented using the R package 'exact2x2' version 1.6.2 [56].

Phenome-wide MR has demonstrated the potential to validate, repurpose and predict on-target side effects of drug targets. Of the protein-phenotype associations that showed evidence of colocalization identified in the cis-only, cis+trans, trans-only or MR analyses using pQTLs with heterogeneous effects across studies (noted as Tier 2 instruments), we first looked up how many proteins with MR evidence were established drug targets in the Informa PharmaProjects database. We then looked up how many of the associations were established target-indication pairs in the PharmaProjects database. More importantly, we predicted the potential adverse effects and repositioning opportunities of all marketed drugs and drugs under development using phenome-wide MR. The forest plots illustrating phenome-wide MR results were drawn using the R package "ggplot2" (https://ggplot2.tidyverse.org/).

Data availability

The data (GWAS summary statistics) used in the analyses described here are freely accessible in the MR-Base platform (www.mrbase.org). All our analysis results for 1684 proteins against 225 human phenotypes are freely available to browse, query and download in EpiGraphDB (http://www.epigraphdb.org/pqtl/). An application programming interface (API) documented on the site enables users to programmatically access data from the database.

Code availability

The code used in the Mendelian randomization analyses described here are freely accessible in the TwoSampleMR R package via GitHub (https://github.com/MRCIEU/TwoSampleMR). Full documentations of the R package were provided (https://mrcieu.github.io/TwoSampleMR/). We implemented the colocalization analysis using the coloc R package (created by Chris Wallace *et al.*), which can be downloaded here (https://cran.r-project.org/web/packages/coloc/index.html).

References

1. Plenge, R. M., Scolnick, E. M. & Altshuler, D. Validating therapeutic targets through human genetics. *Nat Rev Drug Discov* **12**, 581–594 (2013).
2. Hay, M., Thomas, D. W., Craighead, J. L., Economides, C. & Rosenthal, J. Clinical development success rates for investigational drugs. *Nat Biotechnol* **32**, 40–51 (2014).
3. Arrowsmith, J. & Miller, P. Phase II and Phase III attrition rates 2011–2012. *Nat Rev Drug Discov* **12**, 569 (2013).
4. Harrison, R. K. Phase II and phase III failures: 2013–2015. *Nat Rev Drug Discov* **15**, 817 (2016).
5. Cummings, J. L., Morstorf, T. & Zhong, K. Alzheimer's disease drug-development pipeline: few candidates, frequent failures. *Alzheimers Res Ther* **6**, 37 (2014).
6. Nelson, M. R. *et al.* The support of human genetic evidence for approved drug indications. *Nat Genet* **47**, 856–860 (2015).
7. Zhu, Z. *et al.* Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet* **48**, 481–487 (2016).
8. Richardson, T. G. *et al.* Systematic Mendelian randomization framework elucidates hundreds of CpG sites which may mediate the influence of genetic variants on disease. *Hum Mol Genet* (2018). doi:10.1093/hmg/ddy210
9. Santos, R. *et al.* A comprehensive map of molecular drug targets. *Nat Rev Drug Discov* **16**, 19–34 (2017).
10. Würtz, P. *et al.* Metabolomic Profiling of Statin Use and Genetic Inhibition of HMG-CoA Reductase. *J Am Coll Cardiol* **67**, 1200–1210 (2016).
11. Ference, B. A., Majeed, F., Penumetcha, R., Flack, J. M. & Brook, R. D. Effect of naturally random allocation to lower low-density lipoprotein cholesterol on the risk of

coronary heart disease mediated by polymorphisms in NPC1L1, HMGCR, or both: a 2 × 2 factorial Mendelian randomization study. *J Am Coll Cardiol* **65**, 1552–1561 (2015).

12. Sun, B. B. *et al.* Genomic atlas of the human plasma proteome. *Nature* **558**, 73–79 (2018).
13. Suhre, K. *et al.* Connecting genetic risk to disease end points through the human blood plasma proteome. *Nat Commun* **8**, 14357 (2017).
14. Folkersen, L. *et al.* Mapping of 79 loci for 83 plasma protein biomarkers in cardiovascular disease. *PLoS Genet* **13**, e1006706 (2017).
15. Yao, C. *et al.* Genome-wide Association Study Of Plasma Proteins Identifies Putatively Causal Genes, Proteins, And Pathways For Cardiovascular Disease. (2017). doi:10.1101/136523
16. Emilsson, V. *et al.* Co-regulatory networks of human serum proteins link genetics to disease. *Science* (2018). doi:10.1126/science.aaq1327
17. Evans, D. M. & Davey Smith, G. Mendelian Randomization: New Applications in the Coming Age of Hypothesis-Free Causality. *Annu Rev Genomics Hum Genet* **16**, 327–350 (2015).
18. Davey Smith, G. & Ebrahim, S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol* **32**, 1–22 (2003).
19. Zheng, J. *et al.* Recent Developments in Mendelian Randomization Studies. *Curr Epidemiol Rep* **4**, 330–345 (2017).
20. Hemani, G., Bowden, J. & Davey Smith, G. Evaluating the potential role of pleiotropy in Mendelian randomization studies. *Hum Mol Genet* **27**, R195–R208 (2018).
21. Hemani, G. *et al.* The MR-Base platform supports systematic causal inference across the human phenome. *Elife* **7**, (2018).

22. Giambartolomei, C. *et al.* Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet* **10**, e1004383 (2014).
23. Timpson, N. J. *et al.* C-reactive protein levels and body mass index: elucidating direction of causation through reciprocal Mendelian randomization. *Int J Obes* **35**, 300–308 (2011).
24. Hemani, G., Tilling, K. & Davey Smith, G. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. *PLoS Genet* **13**, e1007081 (2017).
25. Hemani, G. *et al.* Automating Mendelian randomization through machine learning to construct a putative causal map of the human phenome. (2017). doi:10.1101/173682
26. Bowden, J. *et al.* A framework for the investigation of pleiotropy in two-sample summary data Mendelian randomization. *Stat Med* **36**, 1783–1802 (2017).
27. Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet* **44**, 369–75, S1-3 (2012).
28. Burgess, S., Dudbridge, F. & Thompson, S. G. Combining information on multiple instrumental variables in Mendelian randomization: comparison of allele score and summarized data methods. *Stat Med* **35**, 1880–1906 (2016).
29. Burgess, S., Zuber, V., Valdes-Marquez, E., Sun, B. B. & Hopewell, J. C. Mendelian randomization with fine-mapped genetic data: Choosing from large numbers of correlated instrumental variables. *Genet Epidemiol* **41**, 714–725 (2017).
30. de Lange, K. M. *et al.* Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. *Nat Genet* **49**, 256–261 (2017).
31. Solomon, T. *et al.* Identification of Common and Rare Genetic Variation Associated With Plasma Protein Levels Using Whole-Exome Sequencing and Mass Spectrometry. *Circ Genom Precis Med* **11**, e002170 (2018).

32. Hashizume, M. *et al.* Tocilizumab, a humanized anti-IL-6R antibody, as an emerging therapeutic option for rheumatoid arthritis: molecular and cellular mechanistic insights. *Int Rev Immunol* **34**, 265–279 (2015).

33. Ridker, P. M. *et al.* Antiinflammatory Therapy with Canakinumab for Atherosclerotic Disease. *N Engl J Med* **377**, 1119–1131 (2017).

34. Haycock, P. C. *et al.* Association Between Telomere Length and Risk of Cancer and Non-Neoplastic Diseases: A Mendelian Randomization Study. *JAMA Oncol* **3**, 636–651 (2017).

35. Dorothée Diogo, C. T., Christopher S. Franklin, Mervi Alanne-Kinnunen, Michael March, Chris C. A. Spencer, Ciara Vangjeli, Michael E. Weale, Hannele Mattsson, Elina Kilpeläinen, Patrick M. A. Sleiman, Dermot F. Reilly, Joshua McElwee, Joseph C. Maranville, Arnaub K. Chatterjee, Aman Bhandari, the 23andMe Research Team, Mary-Pat Reeve, Janna Hutz, Nan Bing, Sally John, Daniel MacArthur, Veikko Salomaa, Samuli Ripatti, Hakon Hakonarson, Mark J. Daly, Aarno Palotie, David Hinds, Peter Donnelly, Caroline S. Fox, Aaron Day-Williams, Robert M. Plenge, Heiko Runz. Phenome-wide association studies (PheWAS) across large “real-world data” population cohorts support drug target validation. *BioRxiv* doi:10.1101/218875

36. Guerrini, M. M. *et al*. Human osteoclast-poor osteopetrosis with hypogammaglobulinemia due to TNFRSF11A (RANK) mutations. *Am J Hum Genet* **83**, 64–76 (2008).

37. Holmes, M. V., Ala-Korpela, M. & Smith, G. D. Mendelian randomization in cardiometabolic disease: challenges in evaluating causality. *Nat Rev Cardiol* **14**, 577–590 (2017).

38. Sekula, P., Del Greco M, F., Pattaro, C. & Köttgen, A. Mendelian Randomization as an Approach to Assess Causality Using Observational Data. *J Am Soc Nephrol* **27**, 3253–3265 (2016).

39. Walker, V. M., Davey Smith, G., Davies, N. M. & Martin, R. M. Mendelian randomization: a novel approach for the prediction of adverse drug events and drug repurposing opportunities. *Int J Epidemiol* **46**, 2078–2089 (2017).

40. GTEx Consortium *et al.* Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).

41. Smemo, S. *et al.* Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature* **507**, 371–375 (2014).

42. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol* **17**, 122 (2016).

43. 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).

44. Bowden, J. *et al.* Assessing the suitability of summary data for two-sample Mendelian randomization analyses using MR-Egger regression: the role of the I 2 statistic. *Int J Epidemiol* **45**, 1961–1974 (2016).

45. Jie Zheng, T. R., Louise Millard, Gibran Hemani, Chris Raistrick, Bjarni Vilhjalmsson, Philip Haycock, Tom Gaunt. PhenoSpD: an integrated toolkit for phenotypic correlation estimation and multiple testing correction using GWAS summary statistics. doi:10.1101/148627

46. Nyholt, D. R. A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. *Am J Hum Genet* **74**, 765–769 (2004).

47. Cichonska, A. *et al.* metaCCA: summary statistics-based multivariate meta-analysis of genome-wide association studies using canonical correlation analysis. *Bioinformatics* **32**, 1981–1989 (2016).

48. Lawlor, D. A., Harbord, R. M., Sterne, J. A. C., Timpson, N. & Davey Smith, G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Stat Med* **27**, 1133–1163 (2008).

49. Pierce, B. L. *et al.* Co-occurring expression and methylation QTLs allow detection of common causal variants and shared biological mechanisms. *Nat Commun* **9**, 804 (2018).

50. Burgess, S., Butterworth, A. & Thompson, S. G. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet Epidemiol* **37**, 658–665 (2013).

51. Yavorska, O. O. & Burgess, S. MendelianRandomization: an R package for performing Mendelian randomization analyses using summarized data. *Int J Epidemiol* (2017). doi:10.1093/ije/dyx034

52. Boyd, A. *et al.* Cohort Profile: the ’children of the 90s’--the index offspring of the Avon Longitudinal Study of Parents and Children. *Int J Epidemiol* **42**, 111–127 (2013).

53. Fraser, A. *et al.* Cohort Profile: the Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. *Int J Epidemiol* **42**, 97–110 (2013).

54. Haycock, P. C. *et al.* Best (but oft-forgotten) practices: the design, analysis, and interpretation of Mendelian randomization studies. *Am J Clin Nutr* **103**, 965–978 (2016).

55. Steiger, J. H. Tests for comparing elements of a correlation matrix. *Psychol Bull* **87**, 245 (1980).

56. Fay, M. P. Confidence intervals that match Fisher’s exact or Blaker’s exact tests. *Biostatistics* **11**, 373–374 (2010).

Table 1. Enrichment analysis comparing target-indication pairs with or without MR and colocalization evidence

		Mendelian randomization and colocalization evidence	
		YES	NO
Target-indication pair approved after clinical trials	YES	8	65
	NO	0	195

Odds ratio for approval given MR and colocalization evidence compared to not having MR or colocalization evidence = 26.39; 95%CI: 3.99 to 580.16; P value = 4.95×10^{-5}

Note: The protein-phenotype association pairs were grouped into four categories: 1) pairs with both MR/colocalization and drug trial evidence; 2) pairs with MR and colocalization evidence but no drug trial evidence; 3) pairs with no strong MR or colocalization evidence but with drug trial evidence; and 4) pairs with no strong MR, colocalization or drug trial evidence. The cut-off for MR evidence was $p < 3.5 \times 10^{-7}$; the cut off for colocalization evidence was posterior probability > 80%. The drug trial evidence was obtained from PharmaProjects database. A Fisher's exact test was then conducted in this enrichment analysis to test whether the target-indication pair represented a successful or failed drug against a signal or not for the corresponding protein-phenotype pair among the MR and colocalization findings. The MR and colocalization analysis results involved in this analysis including both Tier 1 and Tier 2 instruments in both cis and trans region. Since the cell representing failed drug with positive MR and colocalization evidence was null, we added one to each of the cells to estimate the odds ratio and confidence internal for the Fisher's test. More enrichment analysis results for cis-only and tier 1 instruments can be found in **Supplementary Table 21**.

Figure Legend

Figure 1. Flowchart of instrument selection and validation results, and summary of MR and colocalization approach. A. Selection and validation of Instruments. B. Schematic of our MR analysis of plasma proteins on human phenotypes (more detailed pipeline shown in **Supplementary Figure 1**)

Figure 2. Miami plot for the cis-only analysis, with circles representing MR results for specific proteins. The labels refer to top MR findings with colocalization evidence, with each region represented by one label. The colour refers to top MR findings with $P<3.09x10^{-7}$, where red refers to immune mediated phenotypes, blue refers to cancer outcomes, purple refers to cardiovascular outcomes, green refers to other outcomes and **grey** refers to outcomes that shown less evidence of colocalization. The X-axis is the chromosome and position of each MR finding in the cis region. The Y-axis is the -log10 P value of the MR findings, MR findings with positive effects are represented by filled circles on the upper side of the Miami plot, while MR findings with negative effects are on the lower side of the Miami plot.

Figure 3. Regional association plots for IL23R plasma protein level and Crohn's disease in the IL23R region. A. and B. the regional plots of IL23R protein level and Crohn's disease without conditional analysis, Plot B listed the sets of conditional independent signals for Crohn's disease in this region: rs7517847, rs7528924, rs183020189, rs7528804 (a proxy for the second IL23R hit rs3762318, r^2=0.42 in the 1000 Genome Europeans) and rs11209026 (a proxy for the top IL23R hit rs11581607, r^2=1 in the 1000 Genome European), conditional P value < $1x10^{-7}$; C. the regional plot of IL23R with the joint SNP effects conditioned on the second hit (rs3762318) for IL23R; D. the regional plot of Crohn's disease with the joint SNP effects adjusted for other independent signals except top IL23R signal rs11581607; E. the regional plot of IL23R with the joint SNP effects conditioned on the top hit (rs11581607) for IL23R; F. the regional plot of Crohn's disease with the joint SNP effects adjusted for other independent signals except second IL23R signal rs3762318.

Figure 1.

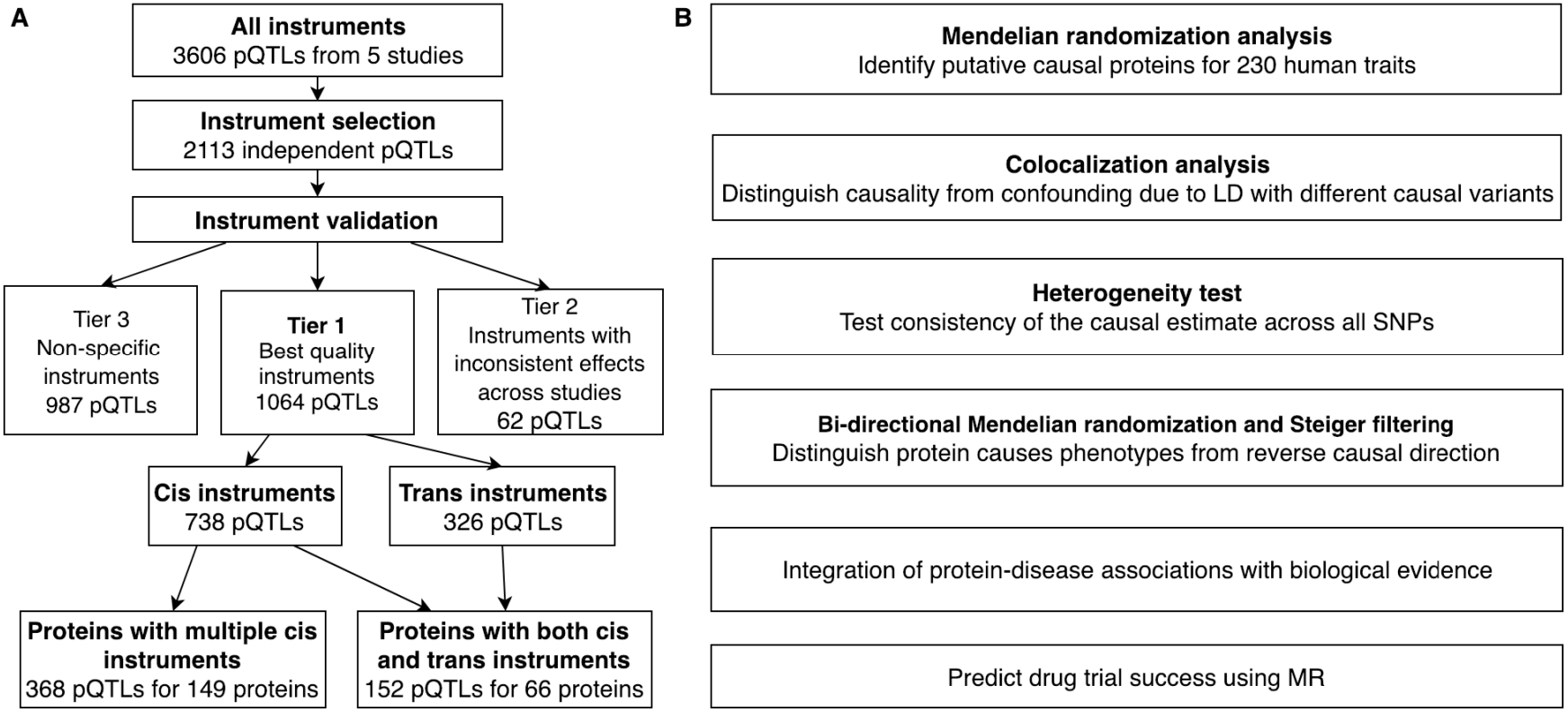

A
All instruments
3606 pQTLs from 5 studies
Instrument selection
2113 independent pQTLs
Instrument validation
Tier 3
Non-specific instruments
987 pQTLs
Tier 1
Best quality instruments
1064 pQTLs
Tier 2
Instruments with inconsistent effects across studies
62 pQTLs
Cis instruments
738 pQTLs
Trans instruments
326 pQTLs
Proteins with multiple cis instruments
368 pQTLs for 149 proteins
Proteins with both cis and trans instruments
152 pQTLs for 66 proteins
B
Mendelian randomization analysis
Identify putative causal proteins for 230 human traits
Colocalization analysis
Distinguish causality from confounding due to LD with different causal variants
Heterogeneity test
Test consistency of the causal estimate across all SNPs
Bi-directional Mendelian randomization and Steiger filtering
Distinguish protein causes phenotypes from reverse causal direction
Integration of protein-disease associations with biological evidence
Predict drug trial success using MR

Figure 2.

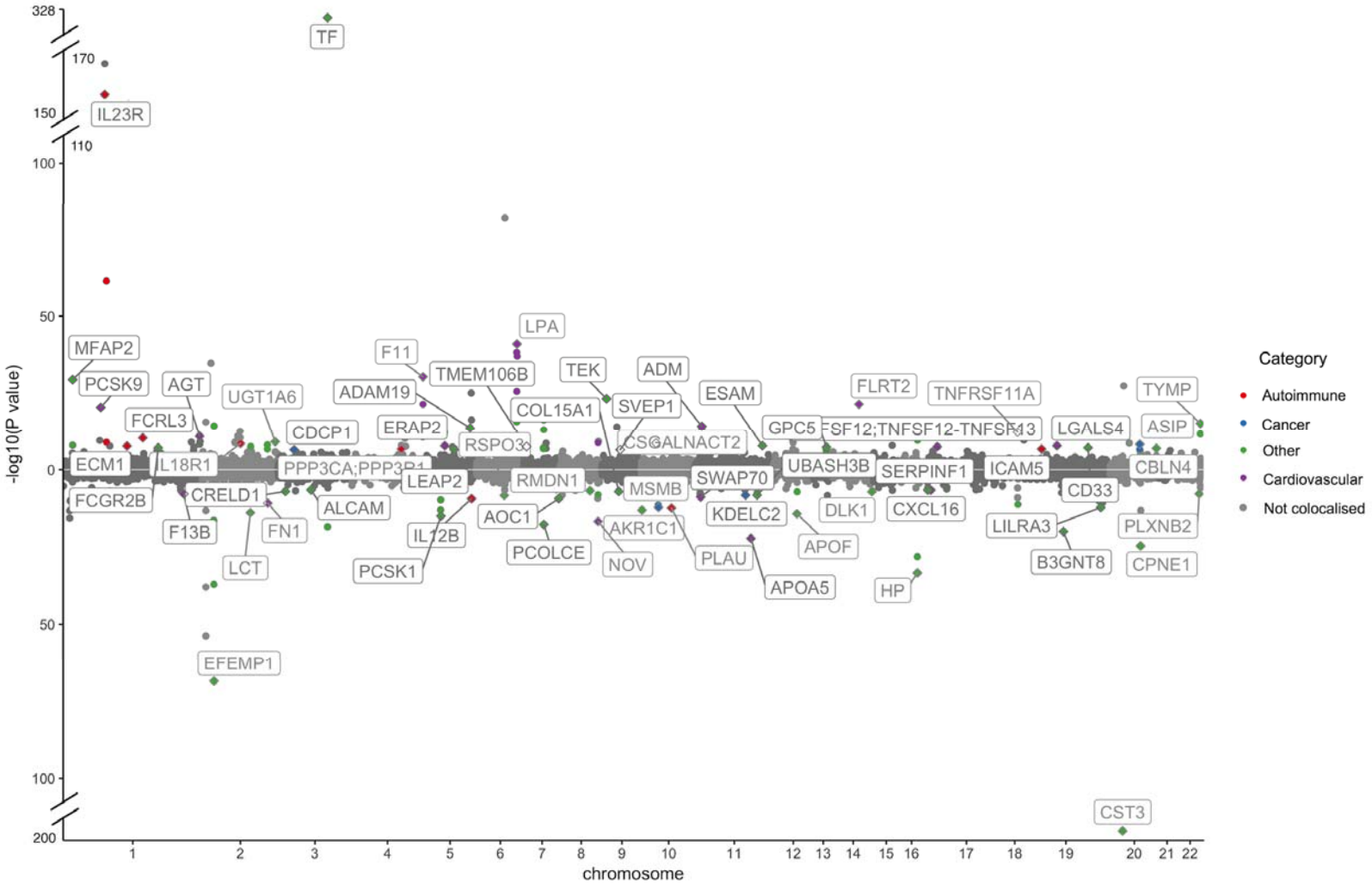
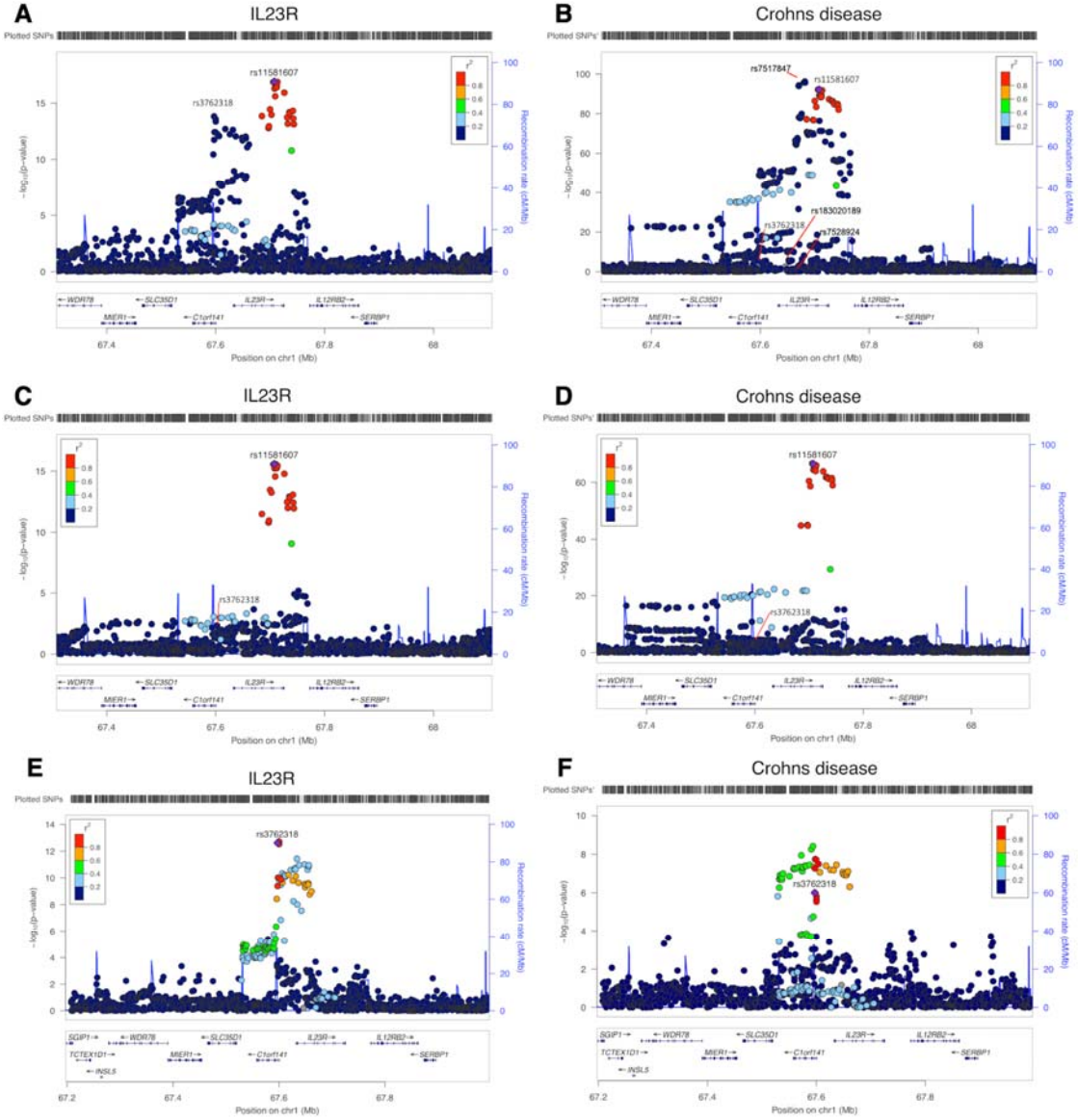
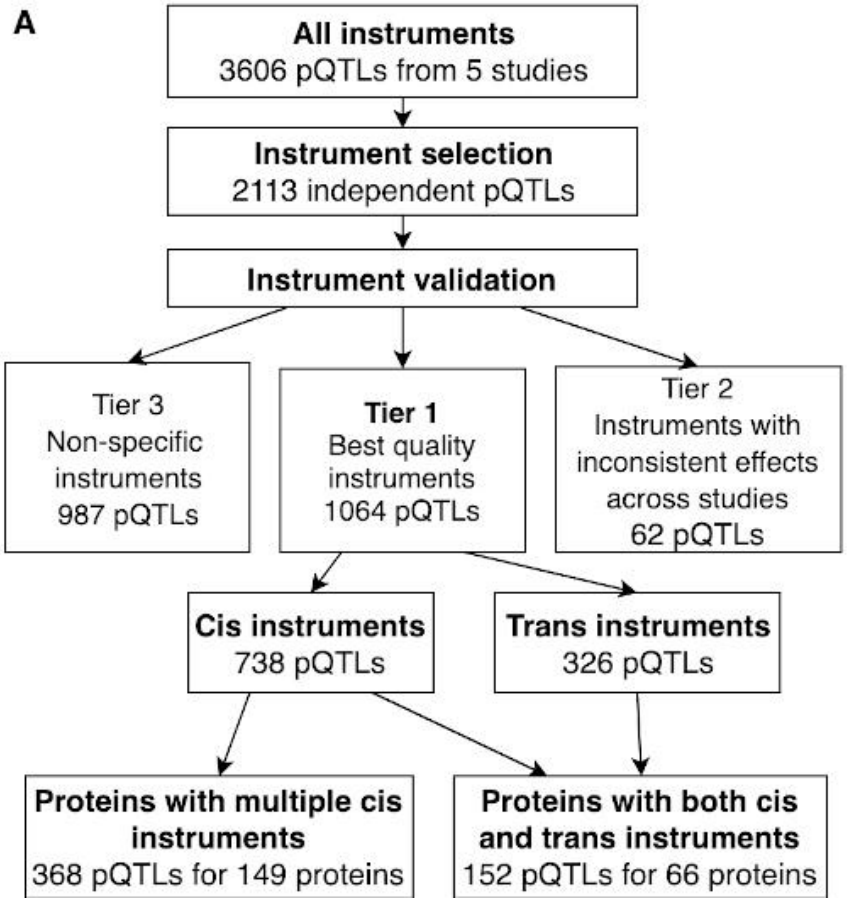
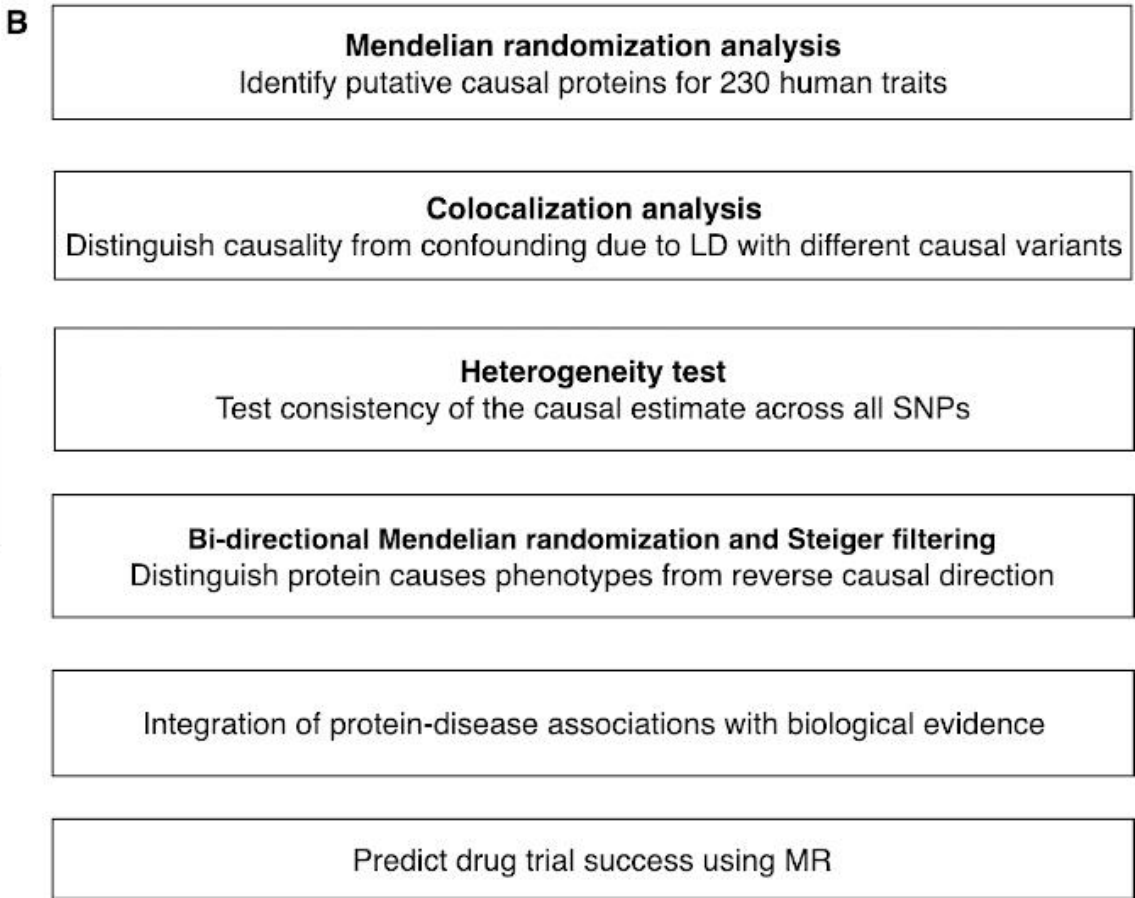
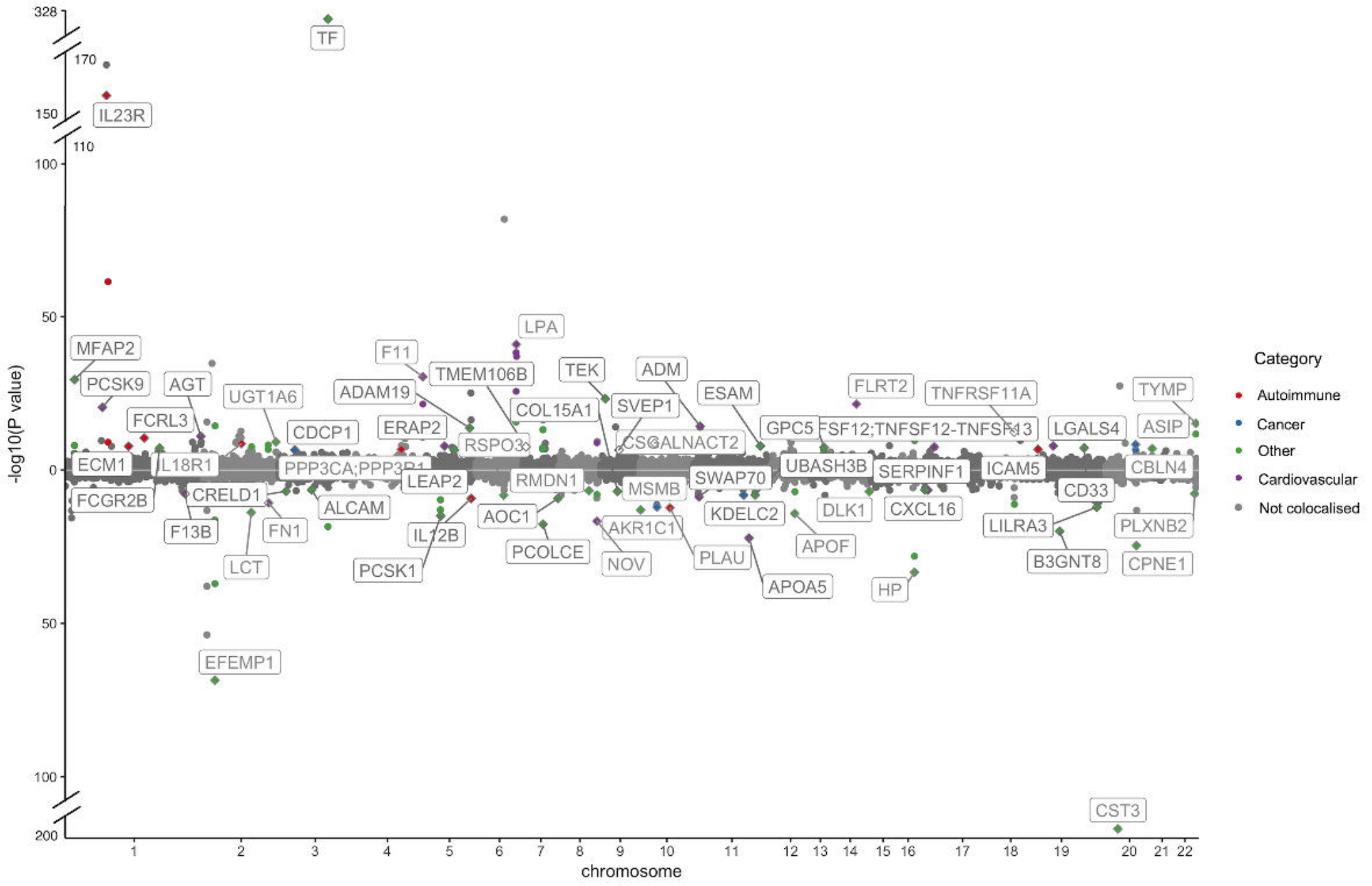

-log10(P value)
chromosome
Category
Autoimmune
Cancer
Other
Cardiovascular
Not colocalised
TF
IL23R
MFAP2
PCSK9
AGT
UGT1A6
FCRL3
CDCP1
ADAM19
F11
TMEM106B
ERAP2
RSPO3
LPA
COL15A1
TEK
SVEP1
ADM
ESAM
FLRT2
TNFRSF11A
TYMP
GPC5
LGALS4
ASIP
ECM1
L18R1
PPP3CA;PPP3R1
LEAP2
RMDN1
MSMB
SWAP70
UBASH3B
SERPINF1
ICAM5
CBLN4
FCGR2B
CRELD1
ALCAM
AOC1
KDELC2
DLK1
CXCL16
CD33
F13B
FN1
IL12B
AKR1C1
PCOLCE
NOV
PLAU
APOF
LILRA3
PLXNB2
LCT
PCSK1
APOA5
HP
B3GNT8
CPNE1
EFEMP1
CST3

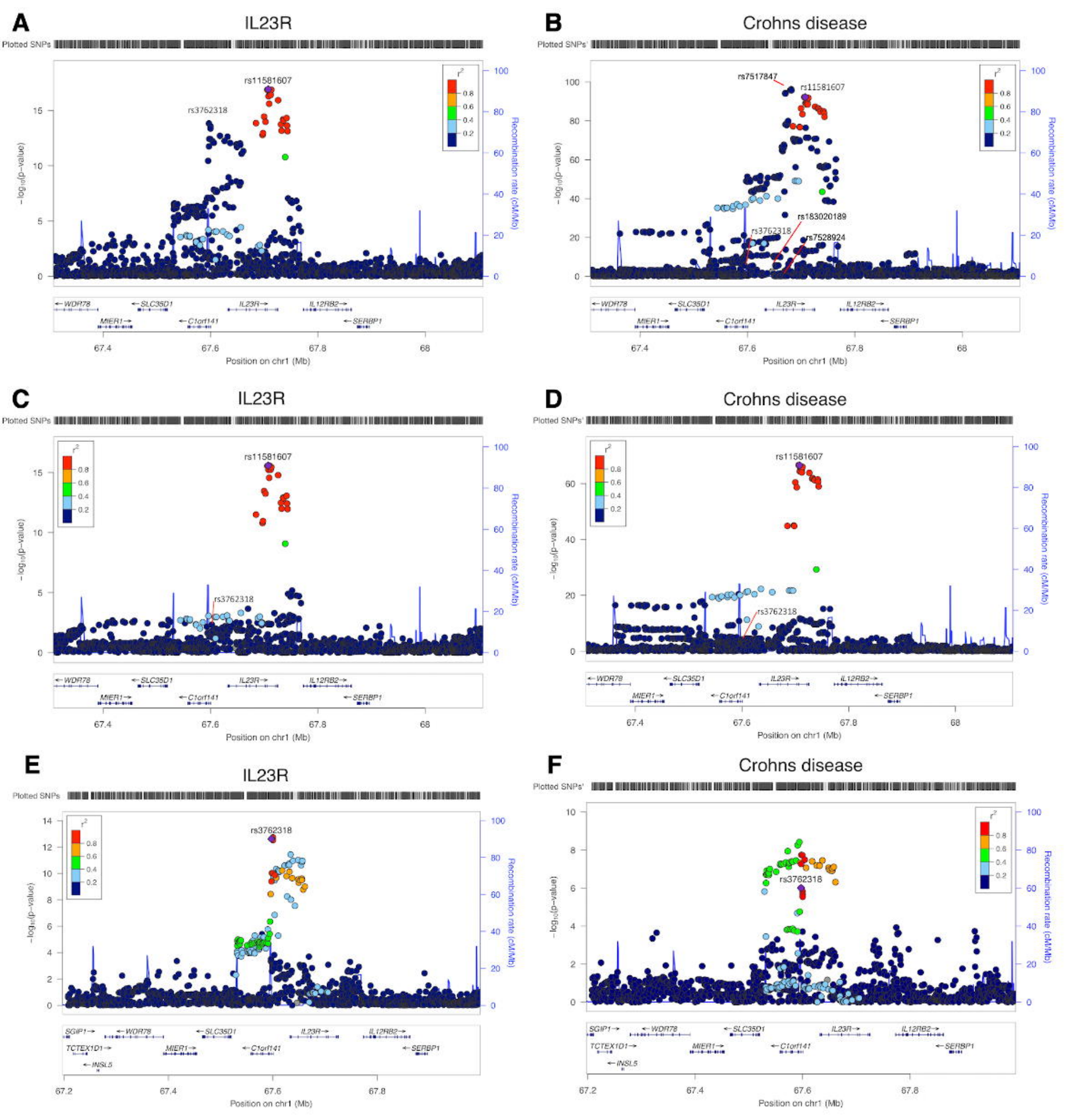
Figure 3.


A
IL23R
B
Crohns disease
C
IL23R
D
Crohns disease
E
IL23R
F
Crohns disease
Plotted SNPs
rs11581607
rs3762318
rs7517847
rs183020189
rs7528924
Position on chr1 (Mb)
Recombination rate (cM/Mb)


A
All instruments
3606 pQTLs from 5 studies
Instrument selection
2113 independent pQTLs
Instrument validation
Tier 3
Non-specific instruments
987 pQTLs
Tier 1
Best quality instruments
1064 pQTLs
Tier 2
Instruments with inconsistent effects across studies
62 pQTLs
Cis instruments
738 pQTLs
Trans instruments
326 pQTLs
Proteins with multiple cis instruments
368 pQTLs for 149 proteins
Proteins with both cis and trans instruments
152 pQTLs for 66 proteins


B
Mendelian randomization analysis
Identify putative causal proteins for 230 human traits
Colocalization analysis
Distinguish causality from confounding due to LD with different causal variants
Heterogeneity test
Test consistency of the causal estimate across all SNPs
Bi-directional Mendelian randomization and Steiger filtering
Distinguish protein causes phenotypes from reverse causal direction
Integration of protein-disease associations with biological evidence
Predict drug trial success using MR


-log10(P value)
chromosome
Category
Autoimmune
Cancer
Other
Cardiovascular
Not colocalised
TF
IL23R
MFAP2
PCSK9
AGT
FCRL3
UGT1A6
CDCP1
ECM1
L18R1
FCGR2B
CRELD1
F13B
FN1
LCT
EFEMP1
ALCAM
PPP3CA;PPP3R1
ADAM19
F11
ERAP2
LEAP2
IL12B
PCSK1
TMEM106B
RSPO3
LPA
COL15A1
RMDN1
AOC1
PCOLCE
TEK
SVEP1
ADM
CSGALNACT2
MSMB
AKR1C1
NOV
ESAM
SWAP70
KDELC2
PLAU
APOA5
GPC5
UBASH3B
APOF
FLRT2
TNFSF12;TNFSF12-TNFSF13
DLK1
SERPINF1
CXCL16
HP
TNFRSF11A
ICAM5
LILRA3
B3GNT8
LGALS4
CD33
TYMP
ASIP
CBLN4
PLXNB2
CPNE1
CST3


A
IL23R
Plotted SNPs
rs11581607
rs3762318
r^2
0.8
0.6
0.4
0.2
$-\log_{10}$(p-value)
Recombination rate (cM/Mb)
WDR78
SLC35D1
IL23R
IL12RB2
MIER1
C1orf141
SERBP1
67.4
67.6
67.8
68
Position on chr1 (Mb)
B
Crohns disease
Plotted SNPs'
rs7517847
rs11581607
rs183020189
rs3762318
rs7528924
C
IL23R
Plotted SNPs
rs11581607
rs3762318
D
Crohns disease
Plotted SNPs'
rs11581607
rs3762318
E
IL23R
Plotted SNPs
rs3762318
SGIP1
TCTEX1D1
INSL5
67.2
F
Crohns disease
Plotted SNPs'
rs3762318