Genetic instrumental variable framework for assessing relevance of in-vitro cellular phenotypes to organism-level phenotypes

Laurence J Howe[1,*], Yury Aulchenko[1], Adrian Cortes[2], Jorge Esparza Gordillo[1], Toby Johnson[1], Tom G Richardson[1], Philippe Sanseau[1], Robert A Scott[1], Daniel D Seaton[1]

[1] GSK, Gunnels Wood Road, Stevenage, SG1 2NFX, United Kingdom

[2] GSK, Meyerhofstrasse 1, 69117, Heidelberg, Germany

* correspondence to laurence.x.howe@gsk.com

authorship alphabetical by surname

Abstract

Gene perturbation screens (e.g. CRISPR-Cas9) assess the impact of gene disruption on cellular phenotypes (e.g., proliferation, anti-viral response, transcriptomics). Cellular phenotypes can be useful in-vitro models for understanding the aetiology of organism-level (in-vivo) phenotypes (e.g., anti-viral response in immune cells and infectious disease outcomes). However, demonstrating relevance between cellular and organism-level phenotypes is non-trivial.

If a cellular phenotype is relevant, then we should observe that perturbed genes impacting the cellular phenotype also impact the organism-level phenotype with mechanism-consistent directionality. We propose a framework, loss-of-function instrumental variable analysis (LoF-IV), to evaluate relevance between cellular and organism-level phenotypes using estimates of gene effects on cellular phenotypes from perturbation screens and estimates of gene effects on human-level phenotypes from loss-of-function burden tests.

Using data from four published genome-wide CRISPR-Cas9 screens and UK Biobank whole-genome sequence burden tests, we evaluated relevance of four in-vitro cellular phenotypes (chondrocyte proliferation, insulin content, adipocyte differentiation, lysosomal cholesterol accumulation) to human-level phenotypes. We found evidence that higher lysosomal cholesterol accumulation lowers LDL-cholesterol measured in human blood plasma ($P = 0.0006$), consistent with the known role of lysosomes in lipid biosynthesis. However, we did not find strong evidence for phenotypic relevance of the other cellular phenotypes, illustrating the complexities of developing disease-relevant in-vitro models. LoF-IV enables systematic evaluation of the relevance of cellular phenotypes to the human phenome informing assay prioritization and supporting novel hypothesis generation.

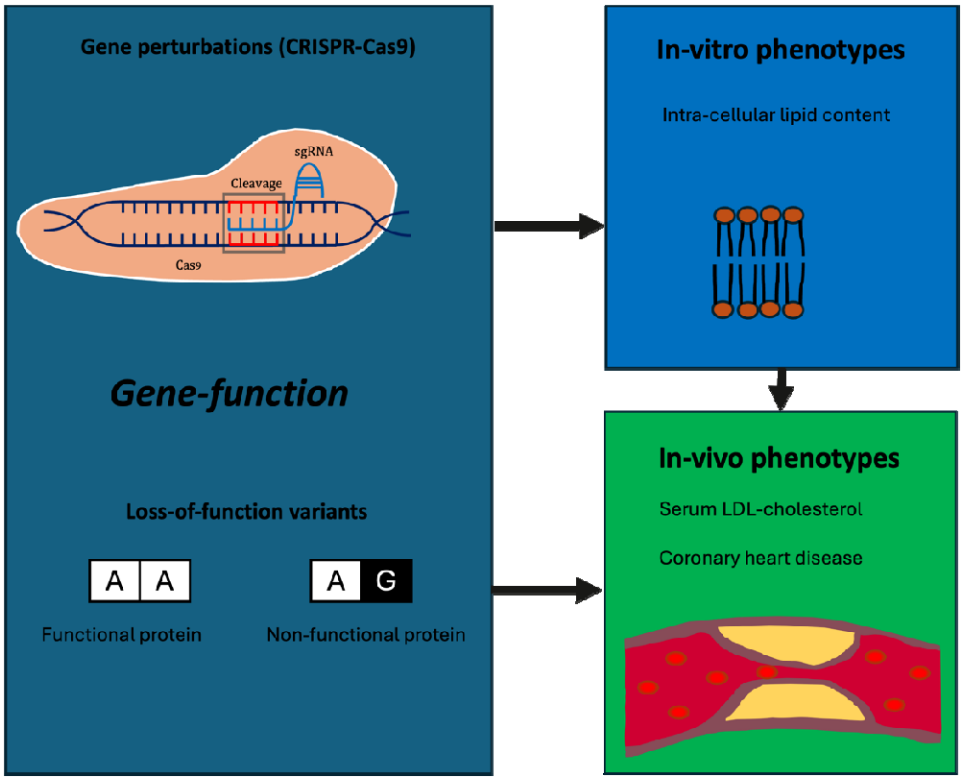
Main

Cellular phenotypes can be useful in-vitro models for organism-level (in-vivo) phenotypes. For example, anti-viral response in immune cells can be used to understand influences on organism-level response to the virus, such as Covid19 symptoms after Sars-Cov2 infection [1,2]. The application of gene editing technologies such as CRISPR-Cas9 in functional screens has enabled at-scale investigation of how gene knockdowns influence cellular phenotypes [1-10]. If the cellular phenotype is on the causal pathway to development or progression of a disease, then genes impacting the cellular phenotype can be putative drug targets for the disease. However, the relevance of a cellular phenotype to an organism-level phenotype is often untested and therefore often unclear. A further complication is that in-vitro experiments may not adequately capture the required biological context such as cell-type, cell-state and physiological state.

Genome-wide association study (GWAS) data have been used extensively within causal inference frameworks such as Mendelian randomization (MR) to evaluate causality between human-level phenotypes (e.g., LDL-cholesterol measured in plasma and coronary heart disease) [11-13]. The intuition is that if higher LDL-cholesterol increases coronary heart disease risk, then genetic variants associated with higher LDL-cholesterol should also be associated with higher coronary heart disease risk. This method can be extended to cellular and human phenotypes. If perturbing a gene increases intracellular lipid content and higher content increases coronary heart disease risk in humans, then loss-of-function (LoF) variants in that gene should be associated with higher risk of coronary heart disease in humans (**Figure 1**).

Loss-of-function instrumental variable analysis (LoF-IV) is a generalisable framework based on the principles of MR to evaluate causal relationships between cellular and organism-level phenotypes using gene-level effect estimates from gene-perturbation screens and aggregated rare variant tests.

Figure 1 Gene function, in-vitro and in-vivo phenotypes



In-vitro cellular phenotypes can impact organism-level phenotypes. For example, higher intra-cellular lipid content may increase serum LDL-cholesterol which then increases risk of coronary heart disease.

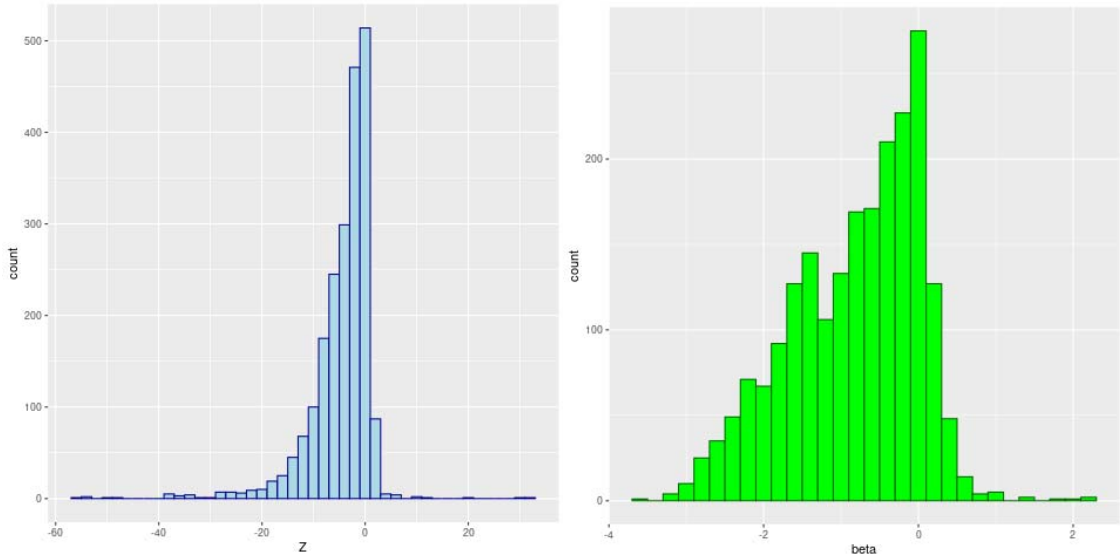
If there is a causal relationship, then all genes which influence the cellular phenotype should also influence the organism-level phenotype. For example, if loss-of-function of LDLR increases intra-cellular lipid content and intra-cellular lipid content increases serum LDL-cholesterol then loss-of-function of LDLR should increase serum LDL-cholesterol.

To evaluate potential causality we can use data from experimental in-vitro gene perturbation studies and from in-vivo population studies. In-vitro gene perturbation studies identify genes where reduced (or increased) function affects a cellular phenotype. We can then evaluate if genes from the in-vitro perturbation study are associated with human phenotypes using loss-of-function variant associations (e.g., in UK Biobank).

pLoF burden tests and plasma protein-levels

LoF-IV assumes that predicted loss-of-function (pLoF) variants substantially reduce gene function and so are comparable to in-vitro gene perturbations. We evaluated the effect of pLoF variants on levels of gene-encoded plasma protein for 2121 gene–protein pairs using whole-genome sequence (WGS) LoF gene burden data from UK Biobank (N ~ 40K). We found evidence at nominal significance ($P < 0.05$) that pLoF burden on 59.0% (1252) of the tested genes reduces plasma levels of the corresponding protein with large effect sizes (median: 1.3 SD reduction; interquartile range 0.8, 1.8). In contrast, pLoF burden increased plasma protein levels for 42 genes (1.9%) suggestive of incorrectly inferred directionality (**Figure 2**). These results provide evidence that pLoF burden tests can be an effective proxy for reduced gene function although we were unable to directly compare with effects of screen perturbations on plasma protein levels. A notable caveat is that inference was based on a subset of protein-coding genes where the encoded protein was measured in UK Biobank and which had observed pLoF variants.

Figure 2 Histogram of distribution of test statistics and effect sizes for associations between pLoF burden of gene and encoded protein (plasma-levels) pairs

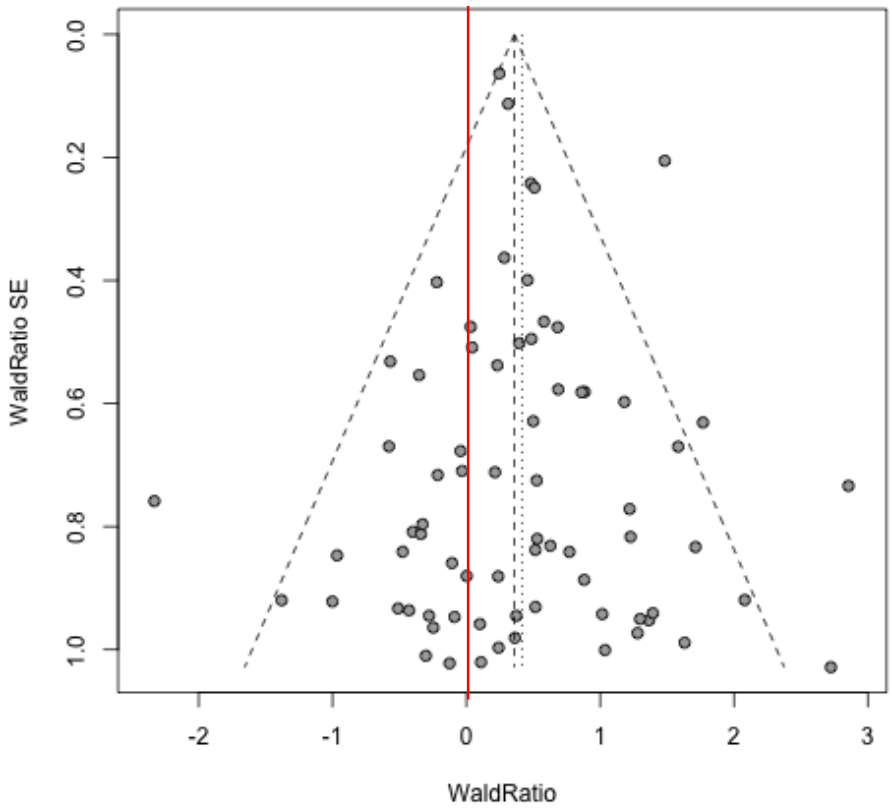


LoF-IV LDL-cholesterol and coronary heart disease

We first validated the use of LoF-IV to confirm established causal relationships in humans. Using WGS LoF burden test data from UK Biobank, we applied LoF-IV to estimate the effect of serum LDL-cholesterol on coronary heart disease (CHD) (both phenotypes measured in humans). LoF-IV using 72 genes with burden evidence for LDL-cholesterol (LoF burden test $P < 0.0005$) provided strong evidence that higher LDL-cholesterol increases odds of coronary heart disease with the inverse-variance weighted (IVW) fixed-effects estimate indicating that a SD increase in LDL increases odds of CHD (OR 1.43; 95% C.I. 1.31, 1.56; P-

value = 2.1×10^{-16}) (**Figure 3 / Supplementary Table 1).** These results illustrate that LoF-IV using UK Biobank scale sequence data has sufficient power to detect large effects for common binary disease phenotypes.

Figure 3 Funnel plot for LDL-cholesterol and coronary heart disease (LoF burden)



Each data point is the Wald ratio for a gene associated with LDL-cholesterol which is an estimate of the effect of LDL-cholesterol on coronary heart disease with the corresponding standard error. For example, if pLoF gene burden is associated with 0.5 unit increase in LDL-cholesterol and a 0.2 unit increase in CHD risk, then the Wald ratio for the gene is 0.4. The dashed lines represent the fixed-effects and random-effects meta-analysis (IVW) estimates and the 95% confidence intervals. Estimates are presented on the linear scale.

LoF-IV in-vitro cellular and human phenotypes

We next applied LoF-IV to evaluate the effects of cellular phenotypes on relevant phenotypes measured in humans using data from four previously published genome-wide CRISPR-Cas9 screens. These screens were selected based on putative relevance to human

phenotypes and data accessibility (**Methods**) and notably constitute only a small subset of published gene perturbation screens.

The identification of genes influencing a cellular phenotype (instruments) involved selecting genes that, when perturbed, led to robust phenotypic changes in the cellular phenotype of interest. Gene perturbation screens can be affected by numerous technical issues such as off-target effects and variable perturbation efficacy [14,15]. We assumed that the study output was not impacted by or had appropriately controlled for these factors. The impact of genes on human phenotypes was estimated from LoF burden tests based on whole genome sequencing of ~500,000 UK Biobank participants (**Methods**).

1) Chondrocyte proliferation and adulthood height.

Growth-plate chondrocytes play an aetiological role in bone development with dysregulated proliferation linked to height-related clinical phenotypes. A previous genome-wide CRISPR screen of chondrocyte proliferation in murine cells demonstrated enrichment of genes involved in height-related phenotypes (e.g., skeletal growth) [8].

2) Intracellular insulin content and type II diabetes (Hba1c).

Pancreatic beta cells play an important role in the aetiology of type II diabetes. A previous screen evaluated the impact of knockdown on intracellular insulin content in a human pancreatic beta cell line [9].

3) Adipocyte differentiation and adiposity (waist circumference, body-mass index, body fat).

Adipogenesis is the process via which adipocytes accumulate adipose tissue. A previous screen evaluated the impact of gene perturbations on adipocyte differentiation in a murine preadipocyte model [16].

4) Lysosomal cholesterol accumulation and plasma LDL-cholesterol.

Lysosomes are involved in lipid biosynthesis. A previous screen evaluated the impact of gene perturbations on lysosomal cholesterol accumulation in k562 cells [17].

Gene-set enrichment analyses were used to evaluate whether cellular phenotype gene-sets were enriched (relative to other protein-coding genes) for pLoF evidence for related phenotypes. This analysis does not consider directionality or model continuous effect sizes. We found some weak evidence that the chondrocyte proliferation gene-set was positively enriched for genes with pLoF burden evidence for adulthood height (16.5% compared to 10.3%, chi squared P = 0.049). We found limited evidence of enrichment of relevant pLoF-evidence for other cellular gene-sets but were not sufficiently well-powered to detect modest enrichments (e.g. <2-fold) (**Supplementary Table 2**).

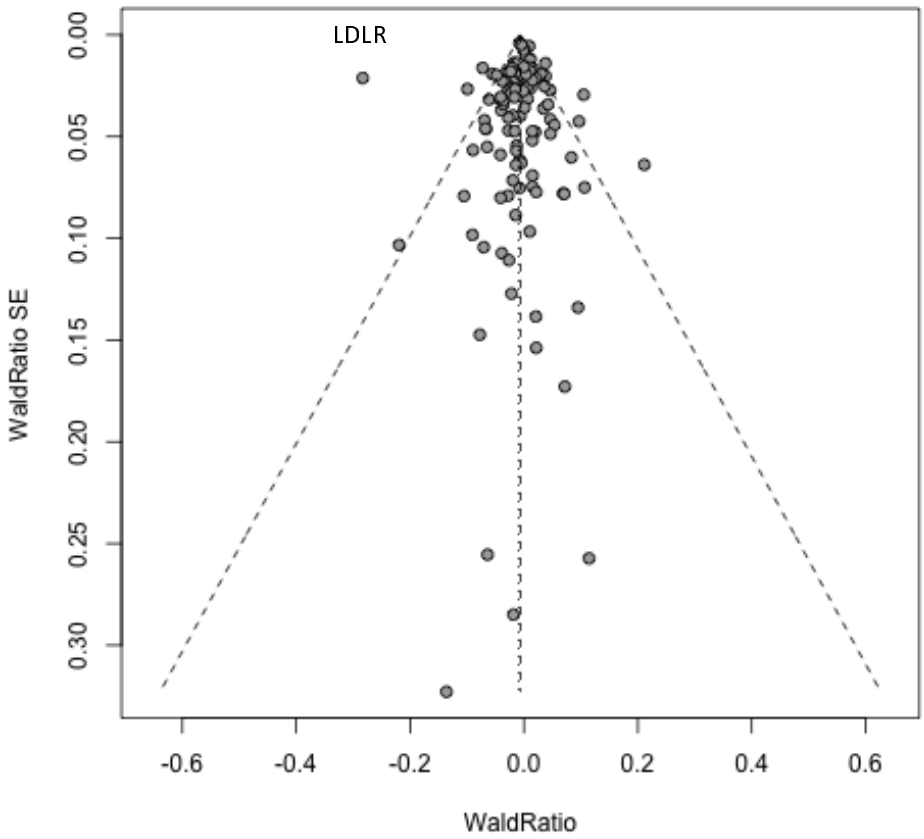
Next, we performed LoF-IV analyses to estimate effects of cellular phenotypes on human phenotypes. The IVW estimate provided evidence that higher lysosomal cholesterol

accumulation lowers levels of LDL-cholesterol in human blood serum with a one-unit increase in lysosomal cholesterol accumulation estimated to reduce LDL-cholesterol by 0.006 SD units (95% C.I. -0.010, -0.003; P = 0.006) (**Table 1** / **Figure 4**).

The random effects IVW estimate modelling heterogeneity had a larger coefficient but was less precisely estimated (beta: -0.010; 95% C.I. -0.020, 0.001; P = 0.06). Estimates from Egger, weighted median and weighted mode sensitivity analyses were consistent with the IV estimates but more imprecisely estimated. There was no strong evidence that the IVW estimate was biased by directional pleiotropy (P = 0.64) (**Supplementary Table 3**).

Gene perturbation of *LDLR* in the screen reduced lysosomal cholesterol accumulation by 2.8 units while loss-of-function of *LDLR* in UK Biobank was associated with an increase in serum LDL-C of 0.79 SD units ($P = 1.7 \times 10^{-40}$). The estimated causal effect ratio for *LDLR* was a visual outlier (annotated in **Figure 4**) so we repeated LoF-IV analyses excluding *LDLR*. The fixed-effects IVW effect estimate slightly attenuated but still provided consistent evidence of the inverse relationship between lysosomal cholesterol accumulation and circulating LDL-cholesterol (beta: -0.004; 95% C.I. -0.008, -0.001; P = 0.02).

Figure 4 Funnel plot for lysosomal cholesterol accumulation and LDL-cholesterol



Each data point is the Wald ratio for a gene associated with lysosomal cholesterol accumulation which is an estimate of the effect of lysosomal cholesterol accumulation on LDL-cholesterol with the corresponding standard error. The dashed lines represent the fixed-effects and random-effects meta-analysis (IVW) estimates and the 95% confidence intervals. Estimates are presented on the linear scale. The LDLR gene is annotated.

In contrast, we found limited evidence for relevance of the other cellular phenotypes to human phenotypes (**Table 1, Supplementary Tables 5-7**). For example, we observed some evidence (albeit borderline significant) enrichment for the chondrocyte-gene set for height LoF evidence but found limited evidence when modelling directionality and continuous effect sizes in LoF-IV that chondrocyte proliferation affects height (fixed-effects IVW: $\log_2$ fold unit change decreases height by 0.007 cm; P-value = 0.42). Sensitivity analysis estimates were consistent with the fixed-effects IVW estimate (**Supplementary Table 5**). Genes where perturbation influenced chondrocyte proliferation were positively enriched for evidence that germline pLoF in humans impacts adulthood height. However, when orientating the

directions of effect of the genes on chondrocyte proliferation and on height, around half of the genes implied a positive relationship while the other half implied the opposite resulting in an overall null estimate in the meta-analysis of Wald ratios. These results suggest that the observed height pLoF evidence enrichment for chondrocyte proliferation genes is likely to reflect gene-level pleiotropy where common genes influence both chondrocyte proliferation and adulthood height but through distinct mechanisms (**Figure 5**).

Table 1 LoF-IV: effects of cellular phenotypes on human phenotypes

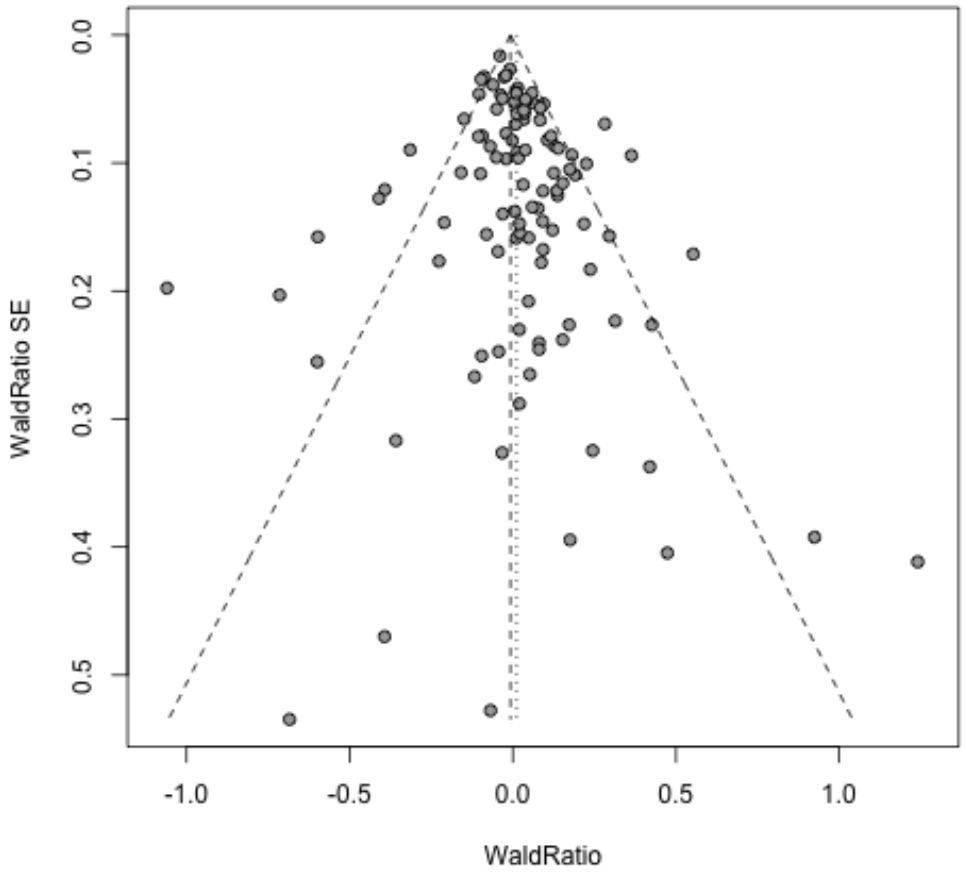
Cellular phenotype	**Number of genes**[1]	**Human phenotype**	**Enrichment P-value**[2]	**LoF-IV estimate**[3] **(95% C.I.)**	**LoF-IV P-value**
Chondrocyte proliferation	109	Adulthood height	0.049	-0.007 (-0.020, 0.006)	0.42
Insulin content	522	Hba1c	0.31	0.003 (-0.012, 0.019)	0.48
		Type II diabetes (OR)	0.96	1.00 (0.99, 1.00)	0.68
Adipocyte differentiation	27	Body mass index	0.06	0.007 (-0.006, 0.021)	0.30
		Waist circumference	0.27	0.013 (-0.001, 0.027)	0.07
		Body fat percentage	0.68	0.009 (-0.005, 0.023)	0.20
Lysosomal cholesterol accumulation	112	LDL-cholesterol	0.23	-0.006 (-0.010, -0.003)	0.0006

[1] Number of instrument genes found in screen which were matched with genes with pLoF burden test data

[2] Gene-set enrichment P-value for pLoF evidence (see Supplementary Table 2)

[3] Inverse-variance weighted fixed-effects LoF-IV estimate

Figure 5 Funnel plot for Chondrocyte proliferation genes and height pLoF burden tests



Each data point is the Wald ratio for the effect of chondrocyte proliferation on height and corresponding standard error for a gene modelling directionality and effect estimates. The dashed lines represent the meta-analysis (IVW) estimate (graphically indistinguishable from zero) and the 95% confidence intervals. For example, if gene knockdown is associated with a 0.5 $\log_2$ fold increase in the functional screen and a 0.25 cm decrease in height from the burden test then the Wald ratio is -0.5. Higher variance in Wald ratios is expected at the bottom of the plots where the standard errors are larger.

We found limited evidence for causal relevance of the intracellular insulin content and adipocyte differentiation cellular phenotypes to human phenotypes although there was borderline evidence at nominal significance that higher adipocyte differentiation increases waist circumference (**Table 1** / **Supplementary Figures 1-5**). These results are consistent with limited relevance of the cellular phenotypes to the respective human phenotypes.

Discussion

Here we present LoF-IV a novel causal inference framework using data from genome-wide perturbation screens and human genetic studies. We have applied LoF-IV to evaluate relevance between cellular and human phenotypes for a few hypothesis-driven examples. Hypothesis-free applications across the human phenome are also possible with the caveat of multiple testing. Improved understanding of the interplay between in-vitro and organism-level phenotypes can enable the assessment of assay relevance and predictability and provide opportunities to discover novel targets for the treatment and prevention of disease.

LoF-IV is a gene-level application of MR and so benefits from a suite of existing instrumental variable methods. LoF-IV has advantages over commonly applied enrichment analyses by using information from both directionality and magnitude of effects, and so can better determine if a cellular phenotype is aetiologically relevant to a disease or simply shares aetiology. In a positive control analysis, we demonstrated that LoF-IV using UK Biobank burden test data confirms established causal relationships between organism-level phenotypes (LDL-cholesterol and coronary heart disease) indicating that LoF-IV is sufficiently well-powered to detect large effects of in-vitro cellular phenotypes.

We found some evidence that higher lysosomal cholesterol accumulation has a lowering effect on circulating serum LDL-cholesterol, consistent with the role of lysosomes in lipid biosynthesis [17,18]. LoF-IV results remained consistent after removing the *LDLR* gene and from other sensitivity analyses. The *LDLR* gene was outlying with the large Wald ratio suggesting that the *LDLR* gene affects serum LDL-cholesterol through additional mechanisms distinct from lysosomal cholesterol accumulation. Although the in-vitro model seems to be relevant to serum LDL-C there may be more relevant models (e.g. different cell-types, cellular conditions) with larger correlations. For growth-plate chondrocyte proliferation and adulthood height, our results indicated that the observed enrichment of genes with LoF height evidence is more likely to be explained by gene-level pleiotropy. We found no strong evidence of causality for adipocyte differentiation and insulin content with adiposity and type II diabetes-related human phenotypes respectively. A potential explanation for null findings is that the cellular phenotype in the functional screen has limited shared aetiology with the cellular phenotype in developing humans, indicating low relevance of the in-vitro model. For example, height is a complex phenotype which is fixed from early adulthood and so mirroring *in-vivo* conditions is likely to be challenging. Another explanation is that it can be difficult to measure the most relevant cellular phenotypes. For example, the study authors acknowledged that insulin content may be less disease-relevant than insulin secretion and that the conditions of the screen were under basal glucose conditions [9].

The observed null results illustrate the challenges of developing disease-relevant in-vitro models. A class of impactful Mendelian randomization studies is those with negative findings consistent with randomised controlled trials such as limited evidence of a causal relationship between HDL-cholesterol and coronary heart disease [19,20]. LoF-IV can potentially have similar value in demonstrating absence of evidence that an in-vitro cellular phenotype is disease-relevant enabling greater prioritisation of disease assays.

LoF-IV is reliant on the three instrumental variable assumptions (relevance, independence, and exclusion-restriction) and so can provide biased causal estimates if these assumptions are violated. Gene perturbations being robustly associated with the cellular phenotype (IV 1: relevance) is unlikely to be a major concern statistically because genome-wide screens are often followed-up by validation screens. However, for interpretation it is important to consider the details of the in-vitro experiment such as organism (e.g., murine) and state, because this will impact relevance to human phenotypes. No common causes (IV 2: independence) of loss-of-function variants and human phenotypes is likely a reasonable assumption with population stratification biasing the pLoF burden tests a possible violation. Violations of the exclusion-restriction assumption (IV 3) include unbalanced horizontal pleiotropy where genes influence the cellular phenotype and the human phenotype via different pathways, and this biases the overall estimate. Loss-of-gene function could have many downstream consequences and so pleiotropy is a major concern for LoF-IV. Indeed, we observed evidence of heterogeneity in the majority of the LoF-IV analyses which could be a consequence of pleiotropy. There are many MR sensitivity analyses which can be used to evaluate unbalanced horizontal pleiotropy to detect and quantify bias. We have applied some of these methods here (Egger, weighted median, weighted mode) finding limited evidence of directional pleiotropy [21,22].

A further assumption of LoF-IV is that gene perturbations (e.g., from CRISPR-Cas9) and from pLoF variants present in humans in biobanks have consistent downstream effects. This is unlikely to hold in practice as most naturally occurring pLoF carriers are heterozygous inducing partial loss of function while gene editing knockouts are likely to have larger (although often variable) effects on gene function. If the relationship between gene function and in-vitro/organism-level phenotypes is linear then this will not impact statistical evidence of causality but will make it difficult to derive meaningful effect estimates (e.g. two-fold increase in cell proliferation increases height by 1cm). Non-linear effects of gene function (e.g. partial loss of function does not affect a phenotype while full loss of function does) could bias LoF-IV estimates. Using plasma proteomics data, we provided evidence that LoF burden can be an effective instrument for lower gene function although with caveats including possible misannotations and inference being limited to a subset of proteins measured in plasma. Future work could attempt to harmonise the relative effects of LoF variants and gene perturbations on gene function as these effects are measurable.

LoF-IV is sensitive to problems with input data such as technical effects in functional screens such as off-target effects [15] or errors with prediction of LoF variants in humans. pLoF variants in highly constrained and/or short in base-pair length genes can be very rare in population biobanks. This will impact statistical power but not necessarily induce bias unless there is a relationship between gene-constraint and phenotypic relevance. For lower frequency disease phenotypes, the power to detect protective effects is often much lower than to detect adverse effects because of the low frequency of pLoF variants. Therefore, LoF-IV is likely to be most effective when applied to continuous or common binary human phenotypes.

LoF-IV enables hypothesis-free and systematic evaluation of causality between cellular and organism-level phenotypes. The method is sensitive to certain assumptions and so we recommend its application in a triangulation of evidence framework alongside complementary in-vitro methods and genomic approaches (e.g., differential expression analysis). The potential applications of LoF-IV are likely to accelerate over the next decade with more functional genomics data, larger datasets of sequencing data and better variant predictions.

Methods

LoF-IV overview

From an analytical perspective, LoF-IV is MR using loss of gene function as an instrumental variable instead of genetic variants. Therefore, many MR estimators (e.g., MR Egger [21]) can be applied with minor differences in interpretation and the three MR instrumental variable assumptions can be adapted as follows. First, the gene perturbations used as instrumental variables from the functional screen are robustly associated with the cellular phenotype in humans (relevance). Second, the gene perturbations from the human genetics study (pLoF variants) share no common causes with the human phenotype (independence). For example, ancestry could influence LoF variant frequency and human phenotypes. Third, the gene perturbations influence the human phenotype only via the cellular phenotype (exclusion-restriction). An additional assumption specific to LoF-IV is that the gene perturbations from the functional screen have comparable effects on gene function with gene perturbations from LoF variants occurring naturally in humans (consistency). If these assumptions are satisfied, then LoF-IV can provide unbiased evaluation of causal relationships. However, effect sizes are likely to be difficult to interpret because of differences in the magnitude of gene perturbation and pLoF variants.

The statistical aspects of LoF-IV are identical to that of conventional Mendelian randomization using gene-level estimates instead of genetic-variants. In brief, Wald ratios and their standard errors are calculated from the effect estimate of each gene on the exposure (cellular phenotype) β_E as well as the effect estimate of the gene on the outcome (human phenotype) β_O and the corresponding standard error σ_O.

$$Wald_G = \frac{\beta_O}{\beta_E}$$

$$SE(Wald_G) = abs(\frac{\sigma_O}{\beta_E})$$

The inverse-variance weighted (IVW) estimator is a fixed-effects meta-analysis of these Wald ratios weighted by the inverse of the variance. Other models can be used to meta-analyse the Wald ratios such as a random-effects model or MR-Egger.

Genome-wide perturbation screen data

There is a substantial literature of published gene perturbation screens on a range of cellular phenotypes. For the purposes of this manuscript, we selected four studies based on the following criteria:

a) Putative relevance to human phenotypes measured in UK Biobank.
b) Genome-wide screen rather than targeted screen (to increase the number of genes with data).
c) Accessibility of output data – level of output data varied substantially across screens in terms of the information provided and the units presented. The heterogeneity of output data made consistent data analysis across different screens challenging.

The studies analysed here are a hand-picked small subset of the available literature.

Chondrocyte proliferation

Growth-plate chondrocyte proliferation is a putative in-vitro model for height. For genetic instruments for chondrocyte proliferation, we extracted a list of 162 genes putatively associated with the cellular phenotype at 4 or 15 days from Table 1 in the original manuscript [8]. We then used the bulk download functions from the ShinyApp browser (https://chondrocyte.shinyapps.io/Live/) to download results (beta coefficients, p-values) from the primary screen for these genes. The authors noted in the manuscript that the day-4 phenotype was more strongly related to height GWAS then the day-15 phenotype, so we focused on the day-4 screen hits.

The chondrocyte gene-set composed of 126 genes associated with 4-day proliferation in the primary screen (P-value < 0.005) and in the validation screen (P-value < 0.05, consistent directionality). The units of chondrocyte proliferation were $\log_2$ fold changes in single-guide RNA (sgRNA) abundance before and after gene knockdowns. After merging on gene symbol with the human pLoF data 109/126 (86.5%) of genes had pLoF burden data.

Insulin content

Insulin content is a putative in-vitro model for diabetes-related phenotypes. For genetic instruments for intracellular insulin content, we used 580 genes contained in Supplementary Table 1 of the original manuscript [9]. These genes were selected if sgRNAs demonstrated consistent effects across replicates and if the association met an FDR threshold (< 0.1). The units of intracellular insulin content were based on sgRNA abundance in cells with low insulin content relative to those with high insulin content. The authors did not provide P-values in the supplementary tables, so it was not possible to further refine the provided gene-set on P-value. After merging on gene symbol with the human pLoF data 522/580 (90%) of genes had pLoF burden data.

Adipocyte differentiation

Adipocyte differentiation is a putative in-vitro model for adiposity related phenotypes. We extracted data from Supplementary Table 1 of the manuscript [16] which contained results for 4603 genes and three adipocyte phenotypes (differentiation, lipid accumulation and mitotic clonal expansion). For the purposes of analyses relating to human adiposity phenotypes we focused on adipocyte differentiation, selecting 32 genes as instruments where perturbation was strongly associated with differentiation ($P < 1x10^{-4}$). We decided on this threshold as a compromise between strength of instrument (P-value threshold) and the number of genetic instruments. After merging on gene symbol with the human pLoF data 27/32 genes (84.4%) had pLoF burden data.

Lysosomal cholesterol accumulation

Lysosomal cholesterol accumulation is a putative in-vitro model for cholesterol and lipid related phenotypes. We extracted data from Supplementary Table 1 of the manuscript [17] which contained data for 20,525 genes including annotations, scores and P-values. The table included both "minimum" and "maximum" effect estimates as well as P-values. To reduce possible bias from winner's curse, we used the minimum effect sizes in LoF-IV and selected 174 genetic instruments with $P < 0.0005$ (threshold compromise between strength of instrument and number of genetic instruments).

When merging with the pLoF data on ensemble gene ID, only 112 genes (64.4%) remained (103 if merging on gene symbol) so we investigated by merging the author gene list with our reference genome (ensemble v106). Merging on ensemble gene id we matched 19724/20525 genes (96.1%) of which 18820 (95.4%) were annotated as protein-coding genes. Specifically for the set of 174 genes, 163 (93.7%) were in the reference genome with 129 (79.1%) annotated as protein-coding suggesting that 17 protein-coding genes (129-112) did not have sufficient pLoF variants for burden data. The majority of "lost" genes were missing from our reference genome or annotated as non-protein-coding genes.

Aggregated rare variant test data in UK Biobank

For gene pLoF associations with human-level phenotypes, we used pLoF gene burden data from UK Biobank based on Whole-Genome Sequencing data from 450K participants [23]. REGENIE (v3.4.1) was used for all analyses [24]. Burden tests were performed using a mask of high-confidence LoF variants (based on variant-effect predictions: VEP) with MAF < 0.01.

For quantitative phenotypes, we analysed LDL-cholesterol and Hba1c from the blood and urine biochemistry (UK Biobank Showcase Category 17518) and BMI, body fat percentage and waist circumference from the baseline assessment data (UK Biobank Showcase Category 100010). Quantitative phenotypes were rank-based inverse-normal transformed prior to analysis. For disease phenotypes, we included coronary heart disease and type II diabetes. These phenotypes were defined based on a combination of electronic health record data sources from CALIBER atlas phenotypes [25]. For proteomics, 2923 unique proteins were measured in plasma from 54,219 UKB participants using the antibody-based

Olink Explore 3072 PEA. These proteins were then matched to the corresponding gene encoding the protein [26]. In analyses, we used burden data on 2,121 proteins which had 1 or more pLoF carriers in the study sample.

For the LDL-CHD positive control analysis, we selected 72 genes as genetic instruments for LDL-C based on pLoF burden test results (P-value < 0.0005 for test with MAF < 0.01).

Enrichment analysis

For enrichment analyses we created two-by-two tables based on the number of genes in a gene-set with pLoF evidence, number of genes without pLoF evidence, number of protein-coding genes (excluding the gene-set) with and without pLoF evidence. pLoF-evidence was defined as having a pLoF burden P-value < 0.05. A Chi squared test was used to assess gene-set enrichment.

LoF-IV analyses

Using the cellular phenotype gene-sets detailed above we extracted relevant estimates from the UK Biobank WGS pLoF burden test data. As detailed above some genes from the cellular datasets were lost when merging with the burden test data. In some instances, mismatch was because protein-coding genes did not have sufficient pLoF carriers for burden tests while in other instances this was because of differences in genome nomenclature / versions or because genes were non protein coding.

To reduce bias from misannotation of pLoF variants, we decided to flip (multiplied by -1) the pLoF burden effect-estimate directions for 42 genes where there was evidence (P < 0.05) that pLoF was associated with increased plasma protein-levels. This did not have a substantial impact on our estimates as only one of these 42 genes were in the instrument sets for the four cellular phenotypes (*NME3* and intracellular insulin content) and pLoF of this gene was not strongly associated with the relevant human phenotypes (T2D P = 0.09, Hba1c P = 0.10). Notably there was very strong statistical evidence that annotated pLoF variants in *NME3* were strongly associated with higher plasma levels of the relevant protein (Log_{10} P-value = 188.7) justifying the correction. We note that we were only able to identify misannotated pLoF variants using the plasma proteomics data for ~10% of protein-coding genes (N ~ 2K) with a further caveat of low statistical power of burden tests for some genes.

For LoF-IV analyses, we calculated Wald ratios using the equations above and used the R package metagen to generate fixed-effects and random-effects inverse-variance weighted estimates as well as funnel plots. We reported the fixed-effects estimates as the primary analysis and the random-effects and heterogeneity estimates in the supplement.

For MR Egger, weighted median and weighted mode we used functions from the TwoSampleMR package [27]. Standard errors for the gene-cellular phenotype association are modelled (with a small impact) in these estimators. These data were not available for all cellular phenotypes, so we imputed the standard errors to be 0. This will lead to slightly underestimated standard errors for these estimators.

For the lysosomal analysis we reran the fixed-effects IVW analyses excluding *LDLR* to evaluate the impact of *LDLR* on the overall estimate.

References

1. Wei, J. *et al.* Genome-wide CRISPR Screens Reveal Host Factors Critical for SARS-CoV-2 Infection. *Cell* **184**, 76-91.e13 (2021).
2. Baggen, J. *et al.* Genome-wide CRISPR screening identifies TMEM106B as a proviral host factor for SARS-CoV-2. *Nature Genetics* **53**, 435-444 (2021).
3. Przybyla, L. & Gilbert, L.A. A new era in functional genomics screens. *Nature Reviews Genetics* **23**, 89-103 (2022).
4. Shi, H., Doench, J.G. & Chi, H. CRISPR screens for functional interrogation of immunity. *Nature Reviews Immunology* **23**, 363-380 (2023).
5. Haley, B. & Roudnicky, F. Functional Genomics for Cancer Drug Target Discovery. *Cancer Cell* **38**, 31-43 (2020).
6. Kendirli, A. *et al.* A genome-wide in vivo CRISPR screen identifies essential regulators of T cell migration to the CNS in a multiple sclerosis model. *Nature Neuroscience* **26**, 1713-1725 (2023).
7. Belk, J.A. *et al.* Genome-wide CRISPR screens of T cell exhaustion identify chromatin remodeling factors that limit T cell persistence. *Cancer Cell* **40**, 768-786.e7 (2022).
8. Baronas, J.M. *et al.* Genome-wide CRISPR screening of chondrocyte maturation newly implicates genes in skeletal growth and height-associated GWAS loci. *Cell Genom* **3**, 100299 (2023).
9. Rottner, A.K. *et al.* A genome-wide CRISPR screen identifies CALCOCO2 as a regulator of beta cell function influencing type 2 diabetes risk. *Nature Genetics* **55**, 54-65 (2023).
10. Dong, C. *et al.* A genome-wide CRISPR-Cas9 knockout screen identifies essential and growth-restricting genes in human trophoblast stem cells. *Nature Communications* **13**, 2548 (2022).
11. Davey Smith, G. & Ebrahim, S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol* **32**, 1-22 (2003).
12. Ference, B.A. *et al.* Effect of long-term exposure to lower low-density lipoprotein cholesterol beginning early in life on the risk of coronary heart disease: a Mendelian randomization analysis. *J Am Coll Cardiol* **60**, 2631-9 (2012).
13. Davey Smith, G. & Hemani, G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet* **23**, R89-98 (2014).
14. Bock, C. *et al.* High-content CRISPR screening. *Nature Reviews Methods Primers* **2**, 8 (2022).
15. Lazar, N.H. *et al.* High-resolution genome-wide mapping of chromosome-arm-scale truncations induced by CRISPR–Cas9 editing. *Nature Genetics* (2024).
16. Turn, R.E. *et al.* A CRISPR-based genome-wide screen for adipogenesis reveals new insights into mitotic expansion and lipogenesis. *bioRxiv*, 2020.07.13.201038 (2023).
17. Lu, A. *et al.* CRISPR screens for lipid regulators reveal a role for ER-bound SNX13 in lysosomal cholesterol export. *J Cell Biol* **221**(2022).
18. Thelen, A.M. & Zoncu, R. Emerging Roles for the Lysosome in Lipid Metabolism. *Trends Cell Biol* **27**, 833-850 (2017).
19. Voight, B.F. *et al.* Plasma HDL cholesterol and risk of myocardial infarction: a mendelian randomisation study. *Lancet* **380**, 572-80 (2012).
20. Holmes, M.V. *et al.* Mendelian randomization of blood lipids for coronary heart disease. *Eur Heart J* **36**, 539-50 (2015).

21. Bowden, J., Davey Smith, G. & Burgess, S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *International Journal of Epidemiology* **44**, 512-525 (2015).
22. 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**, 304-14 (2016).
23. Li, S., Carss, K.J., Halldorsson, B.V. & Cortes, A. Whole-genome sequencing of half-a-million UK Biobank participants. *medRxiv*, 2023.12.06.23299426 (2023).
24. Mbatchou, J. *et al.* Computationally efficient whole-genome regression for quantitative and binary traits. *Nature Genetics* **53**, 1097-1103 (2021).
25. Denaxas, S. *et al.* UK phenomics platform for developing and validating electronic health record phenotypes: CALIBER. *J Am Med Inform Assoc* **26**, 1545-1559 (2019).
26. Sun, B.B. *et al.* Plasma proteomic associations with genetics and health in the UK Biobank. *Nature* **622**, 329-338 (2023).
27. Hemani, G. *et al.* The MR-Base platform supports systematic causal inference across the human phenome. *Elife* **7** (2018).

Contributions

LJH conceptualised the project, performed LoF-IV analyses and drafted the manuscript. TGR and AC performed supporting WGS burden analyses. All authors reviewed and critically contributed to the manuscript.

Competing interests

All authors are employees and/or stockholders of GSK (GlaxoSmithKline).