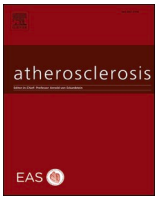
Atherosclerosis 390 (2024) 117462

Contents lists available at ScienceDirect

Atherosclerosis

journal homepage: www.elsevier.com/locate/atherosclerosis



Prioritising genetic findings for drug target identification and validation

Nikita Hukerikar [a,*], Aroon D. Hingorani [b,c], Folkert W. Asselbergs [a,b,d,e], Chris Finan [b,c,d], Amand F. Schmidt [b,c,d,e]

[a] Institute of Health Informatics, Faculty of Population Health Sciences, University College London, London, UK
[b] Institute of Cardiovascular Science, Faculty of Population Health Sciences, University College London, London, UK
[c] The National Institute for Health Research University College London Hospitals Biomedical Research Centre, University College London, London, UK
[d] Department of Cardiology, Division Heart and Lungs, University Medical Centre Utrecht, Utrecht University, Utrecht, the Netherlands
[e] Department of Cardiology, Amsterdam Cardiovascular Sciences, Amsterdam University Medical, Centre, University of Amsterdam, Amsterdam, the Netherlands

ARTICLE INFO

Keywords:
Human genetics
Mendelian randomisation
Drug development
Bioinformatics
NAFLD
Drug target validation
Colocalization
Loss-of-function

ABSTRACT

The decreasing costs of high-throughput genetic sequencing and increasing abundance of sequenced genome data have paved the way for the use of genetic data in identifying and validating potential drug targets. However, the number of identified potential drug targets is often prohibitively large to experimentally evaluate in wet lab experiments, highlighting the need for systematic approaches for target prioritisation.

In this review, we discuss principles of genetically guided drug development, specifically addressing loss-of-function analysis, colocalization and Mendelian randomisation (MR), and the contexts in which each may be most suitable. We subsequently present a range of biomedical resources which can be used to annotate and prioritise disease-associated proteins identified by these studies including 1) ontologies to map genes, proteins, and disease, 2) resources for determining the druggability of a potential target, 3) tissue and cell expression of the gene encoding the potential target, and 4) key biological pathways involving the potential target.

We illustrate these concepts through a worked example, identifying a prioritised set of plasma proteins associated with non-alcoholic fatty liver disease (NAFLD). We identified five proteins with strong genetic support for involvement with NAFLD: CYB5A, NT5C, NCAN, TGFBI and DAPK2. All of the identified proteins were expressed in both liver and adipose tissues, with TGFBI and DAPK2 being potentially druggable.

In conclusion, the current review provides an overview of genetic evidence for drug target identification, and how biomedical databases can be used to provide actionable prioritisation, fully informing downstream experimental validation.

1. Introduction

Target-based drug development is a paradigm that aims to identify druggable targets whose function or concentration can be modified by compounds (drugs) to mitigate the effects of a disease. While around 90 % of current drugs target a protein [1], drug targets may include other biomolecules, such as nucleic acids and RNA [2]. Historically, 90 % of clinical drug development programs fail [3], with the majority of late-stage clinical development stage failures driven by compound related toxicity or by lack of efficacy of the target protein, that is, the drug target is not observed to be causally related to the disease. This high rate of failure indicates the poor ability of pre-clinical experiments, conducted in animals, cell lines, and tissues, to appropriately anticipate effects of target perturbation in human diseases [4]. One promising approach that could help in lowering rates of clinical trial failure is the use of genetic data in drug target identification and validation.

The sequencing of the human genome in 2003 paved the way for human genetic evidence to be used in the drug development process. However, until relatively recently, only *ad-hoc* drug development has been initiated directly from human genomic evidence, largely through the discovery of rare Mendelian variation to causally model the effects of the drug target. For example, candidate gene studies conducted in the mid-1990s identified the central role of the *CCR5* gene in HIV progression [5], and family- and population-based genetic studies in the early 2000s found associations between the *PCSK9* gene and LDL-C concentration, subsequently leading to the successful development of PCSK9 inhibiting drugs for the treatment of hypercholesteremia [6]. More recently, with the increase in large cohort studies covering multiple

* Corresponding author.
E-mail address: nikita.hukerikar.21@ucl.ac.uk (N. Hukerikar).

https://doi.org/10.1016/j.atherosclerosis.2024.117462
Received 12 December 2023; Accepted 25 January 2024
Available online 26 January 2024
0021-9150/© 2024 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

diseases and biomarkers, and the fall in costs of whole genome sequencing and genotype arrays, there is a growing body of publicly accessible genomic data that can be used in drug target identification. In particular, genome-wide association studies (GWAS) present an opportunity to exploit genomics for drug target identification and validation. The use of GWAS has proven successful in identifying *de novo* targets in well-studied diseases. For example, analyses of the *IL6R* locus have anticipated that interleukin 6 receptor (IL-6R) inhibition through tocilizumab (originally indicated for rheumatoid arthritis) might be repurposed for treatment of coronary heart disease (CHD) [7–9]. This hypothesis is now supported by the CANTOS trial, which confirmed that using a monoclonal antibody to target interleukin 1-beta (IL-1β), a protein that drives the IL-6 signalling pathway, reduces the rate of cardiovascular events in patients with a history of CHD [10]. Additionally, trials of IL-6R blockade with tocilizumab are currently underway in patients with myocardial infarction (MI), with some positive results at early trial stages [11–13]. GWAS have additionally proven successful in identifying established drug-target/disease combinations (GWAS rediscoveries) [7,14–16], and systematic evaluation of historical drug development programs has found that compounds related to disease-target combinations with genetic support were substantially more likely to receive regulatory approval than those without (odds ratio (OR) 2.0, 95 % confidence interval (CI) 1.60, 2.40), clearly highlighting the potential of genetic evidence in drug development [17].

As showcased by industry investment into projects such as FinnGen [18], UK Biobank (UKB) [19] and other large population-scale genomic resources, pharmaceutical companies are increasingly considering genomics-first approaches. Here, information from the human genome is leveraged to identify and validate potential new drug targets. In modern, high powered genetic studies, leveraging large scale biobanks such as UKB [19], FinnGen [18], Estonia biobank [20], Biobank Japan [21] and China Kadoorie biobank [22], it is not uncommon to identify between 10 and 100+ genetic loci for a given disease. This number of loci is typically too large, and mechanistically too diverse, to systematically evaluate in wet lab experiments. Hence, further prioritisation is required to identify a subset of tractable targets for confirmatory analyses. For this purpose, many biomedical resources are available, providing orthogonal evidence to aid in prioritising genomics findings for drug development.

In this review, we will provide an overview of a subset of biomedical databases, and integrated software tools, relevant to prioritising findings from genetically guided drug development for subsequent wet-lab validation and eventually clinical testing. We will first discuss the most common types of genetic studies relevant for drug development: loss of function analyses, GWAS, colocalization and Mendelian randomisation. Subsequently, we will discuss the utility of biomedical databases in 1) ontologies mapping genes, proteins and disease 2) identifying druggable proteins 3) clinical effects profiles to identify repurposing candidates 4) tissue and cell expression 5) and key biological pathways including protein-protein interactions. We will illustrate the utility of human genetics and integration of biomedical databases by identifying and prioritising plasma proteins with an anticipated effect on non-alcoholic fatty liver disease (NAFLD).

When reviewing the examples illustrated below, it is worth considering the distinction between a geneticist's definition of a drug target, and the types of targets employed in clinical practice. Of the protein targets, it is important to note the type of target that we are interrogating when using genetic data. Almost exclusively, genetic data assays single protein targets, which make up 80 % of known clinically used target types in ChEMBL (ChEMBL v33). However, other highly represented target types include protein families which represent ~10 % of all protein targets, and protein complexes which make up ~7 % of targets [23]. Whilst individual proteins within these groups are tested using genetic data, entire protein complexes or families are not, and it could be the case that the efficacy of some drugs relies on multiple protein target engagement.

2. Genetic studies supporting drug development

2.1. Loss of function analysis

Predicted loss-of-function (pLoF) variants are rare genetic variants that are predicted to severely disrupt or inactivate the function of a protein, based on the changes that they encode in the protein sequence. Identifying pLoF variants and associating them with disease provides a 'natural' experiment for drug target discovery. pLoF variants usually occur within coding regions of a gene, are very rare and not typically correlated with other variants in the genome, and so are assumed to be causal. This has an advantage over other study designs evaluating common variation, as traversing from the associated genetic variant to the causal gene and protein is implicit in the study design. Given the known functional consequences of pLoF variants, their effect direction on disease provides valuable indication on the mechanism of action a drug compound should have. If a pLoF variant is associated with reduced disease risk, it suggests that the encoded protein should be the target of an inhibitor. Conversely, if the variant is associated with increased disease risk, the encoded protein should be targeted by an activator. Historically, the identification of these variants focussed on rare, monogenic diseases, driven by family-based linkage analyses, with genetic signals determined in families with the disease, and subsequently confirmed by sequencing to identify disease-causing alleles. More recently, the development of high-throughput sequencing technologies has allowed larger, more phenotypically diverse cohorts to be genotyped, and more computational techniques to be used to predict LoF variants [24]. The products of genes harbouring disease-associated LoF variants are now the targets of drugs, both approved and in ongoing trials. For example, genetic studies finding that LoF variants in angiopoietin-like 3 gene (*ANGPTL3*) were associated with decreased concentrations of triglycerides and cholesterol, led to clinical trials, and subsequently the approval of the drug evinacumab targeting ANGPTL3 as a lipid-lowering therapy [25]. The increase in DNA sequence data from large populations has also given rise to numerous computational models predicting the consequences of altered protein function, including LoF as well as missense variants. Examples include SIFT [26], PolyPhen [27], and most recently, AlphaMissense [28].

As discussed by Minikel et al., the sample size required to identify the LoF using whole genome or whole exome sequencing (WGS/WES) in unselected populations is generally prohibitively large. For example, they estimate this would require up to 1,000 times the worldwide available number of genotyped individuals [29]. Genetic studies of isolated populations, where the frequency of rare alleles has genetically drifted upwards, or populations that have a historic propensity for consanguineous children, may provide an opportunity to identify LoF in a more realistic sample size setting. However, this also has significant cost and ethical concerns that may limit routine use. For example, the discovery of a deleterious genetic variant can have implications for participants as well as their family members, especially if the latter did not provide consent for the study. In addition, even if LoF variants are found, their importance may not be clear. Not all pLoF variants occur in disease-coding regions, and, even amongst those that do, many of them are in fact 'benign' and have no clear association with a disease or phenotype [30].

2.2. Genome-wide association studies

The common disease–common variant hypothesis proposes that for common diseases in a population, genetic variations associated with the disease will also be widespread within the population [24]. One study type that supports and exploits this hypothesis is the genome-wide association study (GWAS). GWAS are high-throughput techniques which genotype large numbers of common genetic markers across the genome of a population and test for the association of each one with a phenotype of interest. GWAS can be used to study dichotomous traits, such as the

Table 1
Access details of bioinformatics resources for annotation of genes and proteins.
Summary and URLs for data sources described in all sections of this review.

Resource	Access via	URL
Mapping gene, protein and disease identifiers		
Ensembl genome browser	Web interface	https://www.ensembl.org/index.html
UniProt KnowledgeBase	Web interface	https://www.uniprot.org/uniprotkb
UniProt ID Mapper	Web interface	http://www.uniprot.org/uploadlists/
Medical Subject Headings (MeSH)	Web interface	https://www.ncbi.nlm.nih.gov/mesh/
ChEMBL	Web interface	https://www.ebi.ac.uk/chembl/
	REST API	https://www.ebi.ac.uk/chembl/api/data/docs
	Python web client	https://github.com/chembl/chembl_webresource_client
	Flat file download	https://chembl.gitbook.io/chembl-interface-documentation/downloads
Unified Medical Language System (UMLS) thesaurus	Web interface	https://uts.nlm.nih.gov/uts/umls/home
	REST API	https://documentation.uts.nlm.nih.gov/rest/home.html
	Flat file download	https://www.nlm.nih.gov/research/umls/licensedcontent/downloads.html
HUGO Gene Nomenclature Committee (HGNC)	Web interface	https://www.genenames.org/
	Flat file download	https://www.genenames.org/download/custom/
	REST API	https://www.genenames.org/help/rest/
Entrez Gene	Web interface	https://www.ncbi.nlm.nih.gov/gene
Determining the druggability of potential drug targets		
OpenTargets	Web interface	https://platform.opentargets.org/
	Flat file download	https://www.proteinatlas.org/about/download
	GraphQL API	https://www.proteinatlas.org/api/search_download.php
DrugBank	Web interface	https://go.drugbank.com/
Drug-gene interaction database (DGIdb)	Web interface	https://dgidb.org/
	Flat file download	https://dgidb.org/downloads
	GraphQL API	https://dgidb.org/api
Tissue and cell-specific expression of drug targets		
GTex	Flat file download	https://gtexportal.org/home/downloads/
	REST API	https://gtexportal.org/home/apiPage
Human Protein Atlas (HPA)	Flat file download	https://www.proteinatlas.org/about/download
Biological pathways and protein-protein interaction		
Gene Ontology (GO)	Web interface	https://geneontology.org/
	Flat file download	https://geneontology.org/docs/downloads/
	REST API	https://api.geneontology.org/
Reactome pathway browser	Web interface	https://reactome.org/PathwayBrowser/
	REST API	https://reactome.org/ContentService/
	Flat file download	https://reactome.org/download-data/
	Graph database download	https://reactome.org/dev/graph-database

diagnosis of coronary artery disease, as well as quantitative traits, such as body mass index (BMI) or metabolite or protein concentrations. Due to the relatively low costs of GWAS chips, GWAS currently cover a vast range of phenotypes, in large sample sizes, relevant for drug development. These include analyses of disease onset (e.g., CHD), biomedical traits (e.g., glucose or lipid concentrations), imaging traits (e.g., abdominal MRIs), as well as consideration of high throughput proteomics. Analyses are typically conducted by considering biallelic variants, and comparing difference in average phenotype across both alleles [4,31].

Despite the advantages of using GWAS, several limitations mean that further downstream analyses and drug target validation must be carried out on identified targets. Unlike LoF variants, GWAS associations tend to be common, and the identified genetic variants often reflect non-coding, predicted mutations, located near protein-coding genes. In addition, common variants tend to occur in groups of highly correlated alleles at a population level, a concept referred to as linkage disequilibrium (LD). Therefore, it is often difficult to discern the precise causal variant and gene driving the association signal. This issue has been pervasive in GWAS since the first studies conducted in 2007, however, in the latest causal gene prediction models, distance appears to be a key feature in identifying the causal gene, suggesting that in many cases it is the closest gene that drives the association [32,33]. In addition, most GWAS variants are thought to be acting in a regulatory capacity, affecting either transcript or protein concentration rather than protein function itself. Therefore, even though most GWAS identify protein-coding genes, there can be no a-priori assumption that this is the case for a causal gene driving a GWAS signal. Additionally, given that the effect direction of a GWAS reflects the arbitrary choice of effect allele, inference on the required mechanism of a developed drug is not immediate and requires additional information, for example anchoring genetic associations on CHD by their LDL-C effect. The relevance of GWAS findings for drug development is often improved by conducting additional analyses such as Mendelian randomisation (MR), which natively account for these two sources of information.

2.3. *Mendelian randomisation*

Mendelian randomisation (MR) is a type of instrumental variable (IV) analysis which leverages genetic variants as instruments to identify causal associations between (modifiable) exposures and outcomes. For this purpose, MR leverages genetic variants associated with an exposure and subsequently determines whether there is a dose-response relationship with the genetic variant effect on an outcome (Fig. 2), where the estimated slope provides an indication of anticipated effect direction of the exposure-outcome association. The original IV methodology used in MR has been adapted to a 'two-sample' paradigm, where GWAS summary statistics are used rather than individual-level data, allowing the use of non-identifiable genetic data from different exposure and outcome datasets, maximising the available sample size compared to traditional cohort studies.

MR is based on three key principles:1) that genetic variants are strongly associated with the potential drug target 2) the genetic variant does not share any common causes with the exposure and/or outcome and 3) that there are no horizontal pleiotropy pathways where the genetic variants might affect disease risk without influencing the exposure of interest [4]. By selecting GWAS hits as the variants to study, we can have confidence that the first assumption is met. While the second assumption is hard to formally prove it largely holds true by nature of the experiment. As genetic variation in the population is fixed at gamete

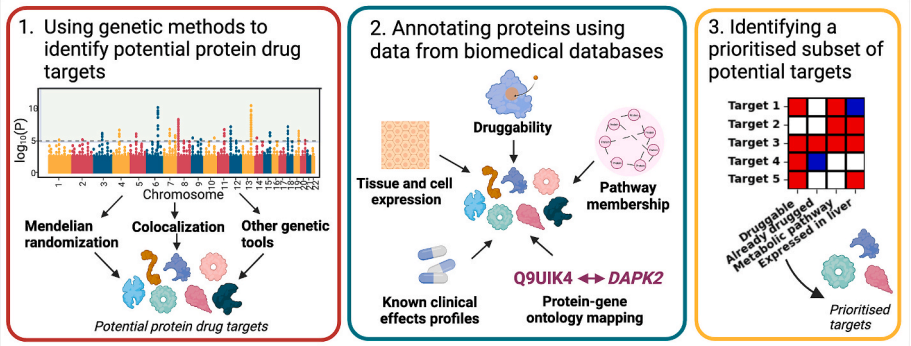


Fig. 1. Graphical abstract.
Stages of genetically guided drug development as explained in this review: 1) identifying potential protein drug targets from genetic data using appropriate methods, 2) annotating potential targets using available biomedical datasets, 3) prioritising a subset of the annotated drug targets.

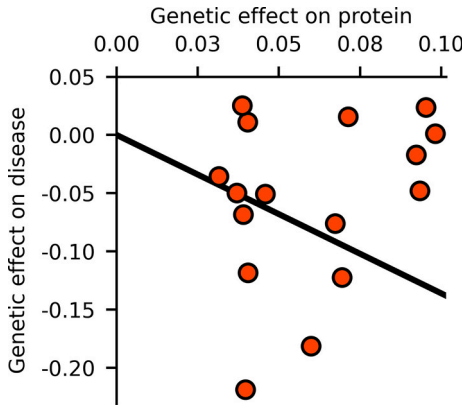


Fig. 2. Dose-response curve between genetic variants associating with plasma concentration of CYB5A, and their effects on non-alcoholic fatty liver disease. Effect sizes represent mean differences in standard deviation change of protein CYB5A (x-axis), and the log(odds ratio) on non-alcoholic fatty liver disease (y-axis). Each point represents a variant effect, and the gradient of the line is the estimated beta coefficient effect size of the protein on the outcome, weighted by the precision of the y-axis estimates (using an inverse variance weighted Mendelian randomisation estimator [61]). The underlying data are available from Supplementary Table S1.

formation, the probability of confounding, whilst not zero, is greatly reduced. The validity of the third assumption is more difficult to ascertain, but the influence of potential horizontal pleiotropy can be reduced analytically, for which a myriad of pleiotropy robust estimators have been derived [4,34–36].

MR has predominantly been used to establish the causal effects of 'traditional' biomarkers such as blood pressure, LDL-C and BMI, using GWAS associations from throughout the genome as instrumental variables. However, due to the increasing abundance of available protein-quantitative trait loci (pQTLs), MR has been adapted to validate potential protein drug targets. MR studies on proteins typically only leverage proteins from in, or very close to, the encoding gene and are termed *cis*-MR or drug target MR.

MR for drug target identification and evaluation typically, but not exclusively, sources genetic instruments from within and around a small *cis* region of the protein encoding gene. MR has produced successful results in a range of settings in drug target validation for cardiovascular disease (CVD), CHD, and multiple other disease groups. In CHD and CVD, for example, MR studies have shown that on-target inhibition of cholesteryl ester transfer protein (CETP) is likely to reduce the risk of CHD and heart failure [15], and that previous failed trials were likely compound related rather than target related [15]; an MR study of HMG-coenzyme A reductase (HMGCR) [14], a licensed drug target for statins, has shown that inhibition of the protein may also have off-target effects such as an increased risk of Type 2 Diabetes [14]; another MR study showed an association between Interleukin 6 receptor (IL-6R) and the risk of ischemic stroke and coronary artery disease (CAD), presenting the protein as a viable therapeutic target for these diseases [37]. Aside from this, MR studies found increased interleukin 18 (IL18) to be associated with a decreased risk of inflammatory bowel disease (IBD) [38]. This highlighted the potential to repurpose IL18 inhibitors which were previously evaluated in clinical trials for treatment of diabetes.

2.4. Colocalization

Colocalization is a method which estimates if two or more distinct GWAS signals are in fact reflecting the same underlying causal variant. Colocalization of GWAS disease and biomarker associations with expression-quantitative trait loci (eQTL) and pQTL signals has been used to attempt to locate the causal gene for a GWAS, where co-located variants are taken as indicators that the gene encoding a pQTL protein is also responsible for the GWAS association.

For drug target validation, colocalization has generally been used post-MR as a prioritisation step to ensure that the identified signal is attributed to the correct exposure. In this context, if it is found that an exposure and the outcome are in fact associated with distinct, causal variants, then it is possible that the GWAS associations are distinct from those in the pQTL, and a pleiotropic pathway to the outcome may exist through, for example, a neighbouring gene [39].

3. Biomedical databases to prioritise genetic findings for drug development

Methods using genomic data, including MR and colocalization, can provide robust evidence for associations between numerous potential drug targets, and clinically relevant outcomes. However, as previously mentioned, the number of proteins will often be too large, and mechanistically too diverse, to evaluate each finding in confirmatory wet lab experiments. Enriching the results of genetically-based drug target identification and validation studies with a range of additional data sources, incorporating important biomedical context, can help in reducing this set of proteins, and prioritising those which are more likely to be clinically relevant.

3.1. Mapping gene, protein and disease identifiers

When using genetic data to prioritise protein drug targets, we assume a trivial one-to-one mapping of gene to protein. However, genes are not labelled with the same unique identifiers as their encoded proteins across datasets, where notably both proteins and genes may have more than one abbreviation or name. For example, the gene *PCSK9* (written in

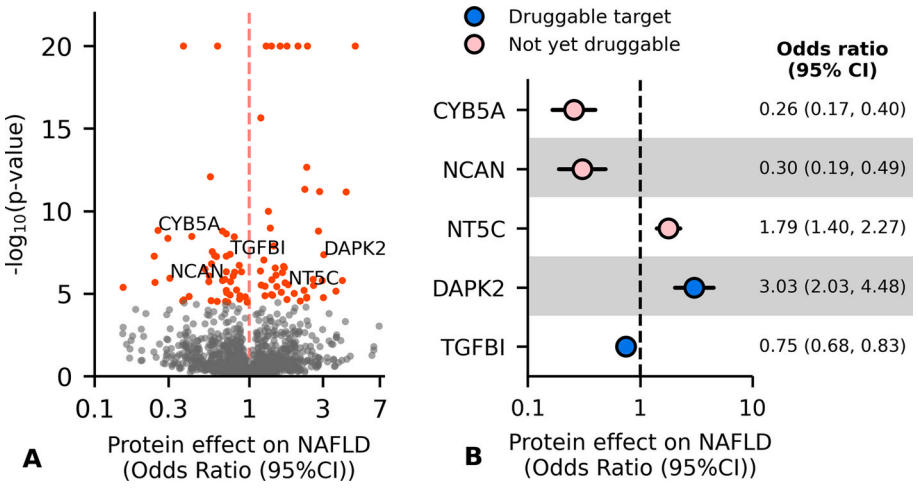


Fig. 3. The Mendelian randomisation estimates of proteins on non-alcoholic fatty liver disease.
(A) (left panel) Effect sizes and statistical significance of each protein on non-alcoholic fatty liver disease. Each point represents a protein, effect estimates are represented in log(odds ratio) (x-axis) and statistical significance in log(*p*-value) (y-axis). Coloured points represent proteins passing the Bonferroni multiplicity-corrected p-value of 3.20×10^{-5} based on the number of proteins (1,978). (B) (right panel) Mendelian randomisation estimates of five prioritised proteins with an effect on non-alcoholic fatty liver disease. Effect estimates are reported as odds ratios with 95 % confidence intervals (95 %CI). Proteins are annotated according to their druggability based on information from the British National Formulary and ChEMBL. Proteins are referred to by their Ensembl gene names. The underlying data are available from Supplementary Table S2.

italic font) has 3 gene synonyms (*FH3, HCHOLA3, NARC-1*), whereas the protein PCSK9 (written in roman font), has a single synonym NARC1.

A widely-used identification system for genes is Ensembl [40], a genome browser in which each gene is assigned a unique identifier. Ensembl incorporates gene annotations from a range of different sources such as the dbSNP [41] for variant information, and the Database of Genotypes and Phenotypes (dbGaP) [42] for phenotype data. Other gene identification systems include the Entrez Gene [43] database for gene-specific information, and the HUGO Gene Nomenclature Committee (HGNC) [44] which maintains unique symbols and names for human loci. An analogue for proteins is the UniProt Knowledgebase (UniProtKB) [45], which contains data on protein sequences and function, and each protein in the database is assigned a unique UniProt accession ID. UniProt provides functionality to map between different identifiers, including Ensembl IDs and UniProt accession IDs.

A common naming convention is also required to identify the diseases associated with the drug targets. Medical Subject Headings (MeSH) [46] are terms defined by the National Library of Medicine, and act as a standardised thesaurus for diseases and medical conditions which can be used to index PubMed. In some data sources, such as the Chemical Biology Database (ChEMBL) [23], diseases and outcomes will be identified by MeSH terms. However, in other cases, this will not be the case, and a metathesaurus such as the Unified Medical Language System (UMLS) [47] can be used to map synonymous disease terms.

3.2. Determining the druggability of potential drug targets

Even if a protein is causally related to a disease, for it to be modifiable, it must be a viable drug target, or 'druggable'. Not all genes encode druggable proteins and as such it is important to determine if this is the case for any of the candidate drug targets. By definition, targets of existing drugs must be druggable, however these represent less than 1,000 proteins out of the entire proteome, estimated to cover over 20,000 proteins [1,48]. The question arises, how do we determine if a currently undrugged protein is indeed druggable?

To first identify disease-associated proteins which are already targeted by a drug compound, databases such as ChEMBL [23] can be consulted. ChEMBL is an open-source database which provides information on bioactive molecules and their interactions with biological targets, and contains data on over 2.4 million drug compounds and their effects on biological systems. ChEMBL data is manually retrieved from a variety of sources, including drug product labels for marketed drugs, published literature, and ClinicalTrials.gov, which publishes information from clinical trials around the world. From ChEMBL, a range of data can be extracted, including the clinical trial phase of a drug (i.e., was the drug licensed or did it fail at an earlier trial stage), the disease indication, the mechanism of action of the drug, and potential adverse effects. Pre-clinical compounds, that is compounds that are bioactive but have not yet been clinically trialled, are also included in the database [23].

For cases where proteins have not yet been targeted by approved drugs, there are various definitions of 'druggable' which can be consulted. The work by Finan et al. [49] combines protein data from both the British National Formulary (BNF) [50] and ChEMBL, in addition to proteins encoding secreted or plasma membrane proteins that are not included in these databases, to produce a list of 4,479 druggable proteins. These additional proteins, whilst not already targeted by compounds, possess biological characteristics such as location, size and membership in 'highly druggable' protein families which provide strong evidence that they could be targeted by monoclonal antibodies. Open Targets [51], a platform developed specifically for the identification and prioritisation of drug targets, integrates data from Finan et al. with a range of resources including ChEMBL, UniProt and the Human Protein Atlas (HPA) [48], to provide details on the tractability of a protein based on its structure, existing clinical trials, and other relevant features.

It is important to note that definitions of druggability are not static and are constantly evolving. Traditionally these definitions focus on proteins that can be activated or inhibited by small molecules. However, with the development of new targeting modalities such as Proteolysis-Targeting Chimeras (PROTACs) [52], which target specific proteins for degradation, or the targeting of peptides rather than small molecules themselves [53], it is likely that the number of druggable proteins will continue to increase.

3.3. Tissue and cell-specific expression of drug targets

Most diseases, at least initially, affect a single or a limited number of tissues. For example, asthma specifically affects the lung, and neurological diseases such as schizophrenia affect the tissue in the brain. It is therefore important to consider in which tissue a genetically identified drug target is expressed, and how likely tissue expression is related to

disease onset. Furthermore, tissue expression already provides some indication on which therapeutic modality might need to be pursued to ensure the drug can access the target [4]. For example, targets expressed in tissues of privileged organs such as in the brain or eye require considerations on how a drug might traverse the blood brain barrier, which is designed to regulate and limit movement between plasma and the brain. Or alternatively, whether drugs acting in tissues such as blood plasma may indirectly affect processes in more privileged areas, for example through active or passive transport. Further insight can be gained by considering single-cell expression data, which can be used to measure the differential expression of a gene across specific cell types. Taking this into consideration can aid in anticipating the efficacy of modulating a potential drug target. For example, if the gene encoding a protein associated with a cancer is found to be expressed in healthy cells but not cancer cells, this could be an indication that targeting the protein may not be effective against the disease.

The Genotype-Tissue Expression (GTEx) [54] project aims to provide tissue-level information on how genetic variation influences gene expression across different tissues. The tissue data is obtained from donors, either post-mortem, or during organ and tissue transplantation surgery, and RNA-sequencing is conducted on the samples. GTEx publishes a range of data based on these analyses, including gene expression at the tissue-level across 54 tissues in the human body and expression-quantitative trait loci (eQTL) which capture genetic associations with gene expression levels across many tissues [55].

The Human Protein Atlas (HPA) [56] is a fully open-access resource which aims to map all human proteins in cells, tissues and organs by integrating results from a range of different technologies including RNA sequencing and tissue imaging. The HPA publishes a breadth of data. The integral part of the HPA is the tissue-level data which focusses on the expression of genes on the mRNA and protein level in human tissues. Here, data from GTEx is combined with internal HPA data and data from the FANTOM5 consortium [57] to provide a consensus classification of gene specificity (a measure of whether a gene is broadly expressed or tissue-specific) and details on gene expression profiles across tissues. The HPA additionally collates single-cell data which measures the expression profile of genes across cell types, and tissue cell-type data which measures cell type specificity of genes within given tissues.

3.4. Biological pathways and protein-protein interaction

In almost all cases, candidate drug targets will not act independently in determining disease onset but will rather form part of a complex network of interrelated pathways. Often, the failure of a drug trial is due to lack of efficacy, or adverse side effects of modulating the drug target. Adopting a more systems-based approach to drug target prioritisation, and identifying pathways that are implicated in disease onset and progression has multiple benefits in this regard. Understanding pathways affected by protein perturbation could help in identifying downstream effects, both beneficial and potentially adverse effects of a drug compound. This can be investigated on a more granular level, by observing the direct interactions between the candidate target and other proteins in either the same or different pathways. Furthermore, if a protein identified by GWAS is not druggable, it is possible that other proteins in shared pathways may be, and could be alternative candidates for targeting.

The Gene Ontology (GO) [58] project is a standardised model which organises and classifies gene products to annotate and analyse the role of different genes in biological processes. GO describes the gene products in three distinct domains: *Molecular Function* which describes activity solely at the molecular level, *Biological Processes* which describes larger processes accomplished by multiple molecular activities, and *Cellular Component* which describes locations in which gene products perform functions. These human curated annotations cover over 20,000 individual genes, as well as providing the ontology itself, allowing analysis of gene function at different granularities. A key use of GO is enrichment analysis; given a set of genes, the set of GO terms that are over- or under-represented can be ascertained.

The Reactome knowledgebase [59] is a comprehensive human pathway database where data is obtained from literature, verified manually by biological experts before being published, and is cross-referenced to other sources including GO, Ensembl and UniProt. Reactome is built as a network of reactions, defined as any molecular event, between molecules, including proteins and small molecules, where pathways are built as a series of connected reactions and are organised hierarchically [59]. Alongside publishing these curated pathways, Reactome provides a number of tools for subsequent analysis including analysing gene lists for over-represented pathways. In addition, Reactome can query IntAct [60], a database of protein-protein interactions, to obtain lists of protein-protein interactions.

A summary of all mentioned data sources and how they may be accessed can be found in Table 1, and a graphical representation of the approach described in this review can be found in Fig. 1.

4. Illustrative example: identifying and prioritising proteins associated with non-alcoholic fatty liver disease

As an illustrative example, we identify and prioritise plasma proteins for involvement with non-alcoholic fatty liver disease (NAFLD), representing a range of conditions caused by the build-up of fat in the liver. NAFLD is the most common form of chronic liver disease, with an estimated prevalence of ~25 % globally [62], which is associated with an increased risk of all-cause mortality, predominantly through an increased risk of CVD [63]. The aetiology of NAFLD is not yet clearly understood, and currently, no drugs exist for the treatment of NAFLD. Therapeutic strategies are instead aimed at symptom management, focusing on interventions such as improved diet and weight loss, and controlling the cardiometabolic risk factors associated with the disease [64].

In this illustrative example (see Supplementary Methods), we carry out Mendelian randomisation and colocalization analyses to identify a subset of plasma proteins associated with NAFLD. For this, we use the deCODE plasma pQTL (sample size 35,559) [65] and the Anstee et al. GWAS of NAFLD (with 1, 483 biopsy confirmed cases and 17,781 controls). We subsequently demonstrate how a subset of biomedical data resources can be leveraged to validate and prioritise these proteins as targets for drug development.

MR identified 91 plasma proteins which significantly associated with NAFLD after accounting for multiple testing (Fig. 3, See Supplementary Methods and Supplementary Table S2). Colocalization analysis between the plasma protein expression and NAFLD GWAS found evidence for shared variants at 40 loci (See Supplementary Methods and Supplementary Table S3), including five proteins with MR association for NAFLD: NCAN, DAPK2, CYB5A, TGFBI, NT5C; See Fig. 2 for the individual instruments for CYB5A.

The HPA database was used to identify the tissues these five proteins were expressed in, particularly focusing on any potential over-expression (i.e., above averagely expressed) in liver, adipose, or granulocyte tissue, which are of particular relevance to NAFLD [66] (Supplementary Methods). Each of the five proteins were found to be expressed in both liver and adipose tissue, with CYB5A over-expressed in the liver (Supplementary Table S4).

We next consulted the ChEMBL and druggable genome definition to determine whether any of these proteins have been drugged by existing compounds, by a developmental compound, or required completely *de novo* drug development. According to ChEMBL, none of the five proteins have been targeted by a compound or drug in clinical phase testing. ChEMBL included compounds with activity against DAPK2 and TGFBI, indicating these proteins are druggable and may be considered for NAFLD drug development.

Finally, we queried the Reactome pathway knowledgebase to identify any pathways which were enriched for the five NAFLD associated

proteins in comparison to all proteins available in the Decode GWAS. Enrichment analysis identified Reactome pathway R-HSA-1430728 reflecting cellular energy metabolism, including mitochondrial lipid metabolism, which is strongly implicated with NAFLD [67,68] (Supplementary Table S5).

5. Conclusion

In this review, we have discussed the benefits of using genetic data to guide drug target validation, discussing common methods used to identify drug targets associated with disease endpoints. We particularly focussed on leveraging information from biomedical datasets to annotate and prioritise candidate drug targets based on information on compound affinity, tissue expression, and biological pathway membership. Finally, we demonstrated how a combination of these datasets could be used to prioritise proteins associated with NAFLD.

Financial support

AFS is supported by BHF grant PG/22/10989, the UCL BHF Research Accelerator AA/18/6/34223, and MR/V033867/1, and the National Institute for Health and Care Research University College London Hospitals Biomedical Research Centre. CF is supported by the UCL BHF Research Accelerator AA/18/6/34223, and MR/V033867/1. FA received grant funding from the EU Horizon scheme (AI4HF 101080430 and DataTools4Heart 101057849) and the Dutch Research Council (MyDigiTwin 628.011.213). AH is supported by the UCL British Heart Foundation Accelerator (AA/18/6/34223), the UCL NIHR Biomedical Research Centre (NIHR203328), and the UKRI/NIHR funded Multimorbidity Mechanism and Therapeutics Research Collaborative (MR/V033867/1). NH is supported by UKRI Training Grant EP/S021612/1, the Centre for Doctoral Training in AI-enabled Healthcare Systems.

Author contributions

NH, AFS, CF, ADH, FWA designed the study. NH performed the analyses and drafted the manuscript. NH, AFS, CF, ADH, FWA provided critical input on the analysis, as well as the drafted manuscript.

Code availability

Analyses were conducted using Python v3.9.18, Pandas v1.3.5, Numpy v1.21.6, bio-misc v0.1.4, plot-misc v0.2.0 hpatools v0.1 and matplotlib v3.8.0.

Data availability

The genetic data used for this analysis can be found in Supplementary Table S6.

The individual GWAS data leveraged in the illustrative example can be accessed as follows: NAFLD data is available from Anstee et al. (cases: 1,483, total n: 19,264 https://www.sciencedirect.com/science/article/pii/S0168827820302130), and deCODE plasma pQTL data is available from Ferkingstad, Sulem, Atlason et al. (n: 35,559 https://www.nature.com/articles/s41588-021-00978-w). Details on accessing each described dataset can be found in Table 1.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: AFS and CF have received funding from New Amsterdam Pharma for an unrelated project.

Acknowledgments

This research has been conducted using the UK Biobank Resource under Application Number 12113. The authors are grateful to UK Biobank participants.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.atherosclerosis.2024.117462.

References

[1] R. Santos, et al., A comprehensive map of molecular drug targets, Nat. Rev. Drug Discov. 16 (1) (Jan. 2017) 1, https://doi.org/10.1038/nrd.2016.230.
[2] It's all druggable, Nat. Genet. 49 (2) (Feb. 2017) 2, https://doi.org/10.1038/ng.3788.
[3] D. Sun, W. Gao, H. Hu, S. Zhou, Why 90% of clinical drug development fails and how to improve it? Acta Pharm. Sin. B 12 (7) (Jul. 2022) 3049–3062, https://doi.org/10.1016/j.apsb.2022.02.002.
[4] A.F. Schmidt, A.D. Hingorani, C. Finan, Human genomics and drug development, Cold Spring Harb Perspect Med 12 (2) (Jan. 2022) a039230, https://doi.org/10.1101/cshperspect.a039230.
[5] P.J. McLaren, J. Fellay, HIV-1 and human genetic variation, Nat. Rev. Genet. 22 (10) (Oct. 2021) 10, https://doi.org/10.1038/s41576-021-00378-0.
[6] M.D. Shapiro, H. Tavori, S. Fazio, PCSK9: from basic science discoveries to clinical trials, Circ. Res. 122 (10) (May 2018) 1420–1438, https://doi.org/10.1161/CIRCRESAHA.118.311227.
[7] A.J. Cupido, et al., Dissecting the IL-6 pathway in cardiometabolic disease: a Mendelian randomization study on both IL6 and IL6R, Br. J. Clin. Pharmacol. 88 (6) (2022) 2875–2884, https://doi.org/10.1111/bcp.15191.
[8] The interleukin-6 receptor as a target for prevention of coronary heart disease: a mendelian randomisation analysis, Lancet 379 (9822) (Mar. 2012) 1214–1224, https://doi.org/10.1016/S0140-6736(12)60110-X.
[9] M. Mihara, M. Hashizume, H. Yoshida, M. Suzuki, M. Shiina, IL-6/IL-6 receptor system and its role in physiological and pathological conditions, Clin. Sci. 122 (4) (Oct. 2011) 143–159, https://doi.org/10.1042/CS20110340.
[10] P.M. Ridker, et al., Antiinflammatory therapy with canakinumab for atherosclerotic disease, N. Engl. J. Med. 377 (12) (Sep. 2017) 1119–1131, https://doi.org/10.1056/NEJMoa1707914.
[11] J. Bo Kunkel, et al., Low-dose dobutamine infusion and single-dose tocilizumab in acute myocardial infarction patients with high risk of cardiogenic shock development - rationale and design of the DOBERMANN trial, European Heart Journal. Acute Cardiovascular Care 12 (Supplement_1) (May 2023) zuad036, https://doi.org/10.1093/ehjacc/zuad036.131, 131.
[12] O. Kleveland, et al., Effect of a single dose of the interleukin-6 receptor antagonist tocilizumab on inflammation and troponin T release in patients with non-ST-elevation myocardial infarction: a double-blind, randomized, placebo-controlled phase 2 trial, Eur. Heart J. 37 (30) (Aug. 2016) 2406–2413, https://doi.org/10.1093/eurheartj/ehw171.
[13] K. Broch, et al., Randomized trial of interleukin-6 receptor inhibition in patients with acute ST-segment elevation myocardial infarction, J. Am. Coll. Cardiol. 77 (15) (Apr. 2021) 1845–1855, https://doi.org/10.1016/j.jacc.2021.02.049.
[14] D.I. Swerdlow, et al., HMG-coenzyme A reductase inhibition, type 2 diabetes, and bodyweight: evidence from genetic analysis and randomised trials, Lancet 385 (9965) (Jan. 2015) 351–361, https://doi.org/10.1016/S0140-6736(14)61183-1.
[15] A.F. Schmidt, et al., Cholesteryl ester transfer protein (CETP) as a drug target for cardiovascular disease, Nat. Commun. 12 (1) (Sep. 2021) 1, https://doi.org/10.1038/s41467-021-25703-3.
[16] A.F. Schmidt, et al., Phenome-wide association analysis of LDL-cholesterol lowering genetic variants in PCSK9, BMC Cardiovasc. Disord. 19 (1) (Oct. 2019) 240, https://doi.org/10.1186/s12872-019-1187-z.
[17] E.A. King, J.W. Davis, J.F. Degner, Are drug targets with genetic support twice as likely to be approved? Revised estimates of the impact of genetic support for drug mechanisms on the probability of drug approval, PLoS Genet. 15 (12) (Dec. 2019) e1008489, https://doi.org/10.1371/journal.pgen.1008489.
[18] FinnGen: Unique genetic insights from combining isolated population and national health register data | medRxiv'. Accessed: Sep. 01, 2023. [Online]. Available: https://www.medrxiv.org/content/10.1101/2022.03.03.22271360v1.
[19] C. Sudlow, et al., UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age, PLoS Med. 12 (3) (Mar. 2015) e1001779, https://doi.org/10.1371/journal.pmed.1001779.
[20] The Estonian Genome Project - Metspalu - 2004 - Drug Development Research - Wiley Online Library'. Accessed: Sep. 01, 2023. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/ddr.10371?casa_token=7u-bQKOCsGwAAAAA:W-DVnk6-oeHJKn5nYrXfRQe4q7fPUtayCzaNZCa9jRShD7q2prNcdnIyHQFEB6oYB9NjW1tlqmRK0H4.
[21] A. Nagai, et al., Overview of the BioBank Japan project: study design and profile, J. Epidemiol. 27 (Supplement_III) (2017) S2–S8, https://doi.org/10.1016/j.je.2016.12.005.

[22] Z. Chen, et al., China Kadoorie Biobank of 0.5 million people: survey methods, baseline characteristics and long-term follow-up, Int. J. Epidemiol. 40 (6) (Dec. 2011) 1652–1666, https://doi.org/10.1093/ije/dyr120.
[23] A. Gaulton, et al., ChEMBL: a large-scale bioactivity database for drug discovery, Nucleic Acids Res. 40 (D1) (Jan. 2012) D1100–D1107, https://doi.org/10.1093/nar/gkr777.
[24] M. Claussnitzer, et al., A brief history of human disease genetics, Nature 577 (7789) (Jan. 2020) 179–189, https://doi.org/10.1038/s41586-019-1879-7.
[25] F.E. Dewey, et al., Genetic and pharmacologic inactivation of ANGPTL3 and cardiovascular disease, N. Engl. J. Med. 377 (3) (Jul. 2017) 211–221, https://doi.org/10.1056/NEJMoa1612790.
[26] R. Vaser, S. Adusumalli, S.N. Leng, M. Sikic, P.C. Ng, SIFT missense predictions for genomes, Nat. Protoc. 11 (1) (Jan. 2016) 1, https://doi.org/10.1038/nprot.2015.123.
[27] I.A. Adzhubei, et al., A method and server for predicting damaging missense mutations, Nat. Methods 7 (4) (Apr. 2010) 248–249, https://doi.org/10.1038/nmeth0410-248.
[28] J. Cheng, et al., Accurate proteome-wide missense variant effect prediction with AlphaMissense, Science 381 (6664) (Sep. 2023) eadg7492, https://doi.org/10.1126/science.adg7492.
[29] E.V. Minikel, et al., Evaluating drug targets through human loss-of-function genetic variation, Nature 581 (7809) (May 2020) 7809, https://doi.org/10.1038/s41586-020-2267-z.
[30] D.G. MacArthur, et al., A systematic survey of loss-of-function variants in human protein-coding genes, Science 335 (6070) (Feb. 2012) 823–828, https://doi.org/10.1126/science.1215040.
[31] A. Lau, H.-C. So, Turning genome-wide association study findings into opportunities for drug repositioning, Comput. Struct. Biotechnol. J. 18 (Jan. 2020) 1639–1650, https://doi.org/10.1016/j.csbj.2020.06.015.
[32] E. Mountjoy, et al., An open approach to systematically prioritize causal variants and genes at all published human GWAS trait-associated loci, Nat. Genet. 53 (11) (Nov. 2021) 11, https://doi.org/10.1038/s41588-021-00945-5.
[33] V. Forgetta, et al., An effector index to predict target genes at GWAS loci, Hum. Genet. 141 (8) (Aug. 2022) 1431–1447, https://doi.org/10.1007/s00439-022-02434-z.
[34] J. Bowden, G. Davey Smith, S. Burgess, Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression, Int. J. Epidemiol. 44 (2) (Apr. 2015) 512–525, https://doi.org/10.1093/ije/dyv080.
[35] F.P. Hartwig, G. Davey Smith, J. Bowden, Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption, Int. J. Epidemiol. 46 (6) (Dec. 2017) 1985–1998, https://doi.org/10.1093/ije/dyx102.
[36] J. Bowden, G. Davey Smith, P.C. Haycock, S. Burgess, Consistent estimation in mendelian randomization with some invalid instruments using a weighted median estimator, Genet. Epidemiol. 40 (4) (May 2016) 304–314, https://doi.org/10.1002/gepi.21965.
[37] M.K. Georgakis, R. Malik, D. Gill, N. Franceschini, C.L.M. Sudlow, M. Dichgans, Interleukin-6 signaling effects on ischemic stroke and other cardiovascular outcomes, Circ Genom Precis Med 13 (3) (May 2020) e002872, https://doi.org/10.1161/CIRCGEN.119.002872.
[38] L.E. Mokry, et al., Interleukin-18 as a drug repositioning opportunity for inflammatory bowel disease: a Mendelian randomization study, Sci. Rep. 9 (Jun. 2019) 9386, https://doi.org/10.1038/s41598-019-45747-2.
[39] V. Zuber, et al., Combining evidence from Mendelian randomization and colocalization: review and comparison of approaches, Am. J. Hum. Genet. 109 (5) (May 2022) 767–782, https://doi.org/10.1016/j.ajhg.2022.04.001.
[40] F.J. Martin, et al., Ensembl 2023, Nucleic Acids Res. 51 (D1) (Jan. 2023) D933–D941, https://doi.org/10.1093/nar/gkac958.
[41] S.T. Sherry, et al., dbSNP: the NCBI database of genetic variation, Nucleic Acids Res. 29 (1) (Jan. 2001) 308–311, https://doi.org/10.1093/nar/29.1.308.
[42] K.A. Tryka, et al., NCBI's database of genotypes and phenotypes: dbGaP, Nucleic Acids Res. 42 (D1) (Jan. 2014) D975–D979, https://doi.org/10.1093/nar/gkt1211.
[43] D. Maglott, J. Ostell, K.D. Pruitt, T. Tatusova, Entrez Gene: gene-centered information at NCBI, Nucleic Acids Res. 35 (Database issue) (Jan. 2007) D26–D31, https://doi.org/10.1093/nar/gkl993.
[44] R.L. Seal, et al., Genenames.org: the HGNC resources in 2023, Nucleic Acids Res. 51 (D1) (Jan. 2023) D1003–D1009, https://doi.org/10.1093/nar/gkac888.
[45] The UniProt Consortium, UniProt: the universal protein knowledgebase in 2021, Nucleic Acids Res. 49 (D1) (Jan. 2021) D480–D489, https://doi.org/10.1093/nar/gkaa1100.
[46] I.K. Dhammi, S. Kumar, Medical subject headings (MeSH) terms, Indian J. Orthop. 48 (5) (2014) 443–444, https://doi.org/10.4103/0019-5413.139827.
[47] O. Bodenreider, The unified Medical Language System (UMLS): integrating biomedical terminology, Nucleic Acids Res. 32 (Database issue) (Jan. 2004) D267–D270, https://doi.org/10.1093/nar/gkh061.
[48] E. Sjöstedt, et al., An atlas of the protein-coding genes in the human, pig, and mouse brain, Science 367 (6482) (Mar. 2020) eaay5947, https://doi.org/10.1126/science.aay5947.
[49] C. Finan, et al., The druggable genome and support for target identification and validation in drug development, Sci. Transl. Med. 9 (383) (Mar. 2017) eaag1166, https://doi.org/10.1126/scitranslmed.aag1166.
[50] J. F. Committee and R. P. S. of G. Britain, British National Formulary, vol. 64, Pharmaceutical Press, 2012.
[51] Open Targets Platform: supporting systematic drug–target identification and prioritisation | Nucleic Acids Research | Oxford Academic'. Accessed: Sep. 01, 2023. [Online]. Available: https://academic.oup.com/nar/article/49/D1/D1302/5983621.
[52] PROTAC targeted protein degraders: the past is prologue | Nature Reviews Drug Discovery'. Accessed: Oct. 25, 2023. [Online]. Available: https://www.nature.com/articles/s41573-021-00371-6.
[53] K. Fosgerau, T. Hoffmann, Peptide therapeutics: current status and future directions, Drug Discov. Today 20 (1) (Jan. 2015) 122–128, https://doi.org/10.1016/j.drudis.2014.10.003.
[54] The Genotype-Tissue Expression (GTEx) project | Nature Genetics'. Accessed: Aug. 31, 2023. [Online]. Available: https://www.nature.com/articles/ng.2653.
[55] F. Aguet, et al., Genetic effects on gene expression across human tissues, Nature 550 (7675) (Oct. 2017) 7675, https://doi.org/10.1038/nature24277.
[56] The Human Protein Atlas'. Accessed: Aug. 25, 2022. [Online]. Available: https://www.proteinatlas.org/.
[57] FANTOM5 CAGE profiles of human and mouse samples | Scientific Data'. Accessed: Oct. 21, 2023. [Online]. Available: https://www.nature.com/articles/sdata2017112.
[58] S.A. Aleksander, et al., The gene ontology knowledgebase in 2023, Genetics 224 (1) (Mar. 2023) iyad031, https://doi.org/10.1093/genetics/iyad031.
[59] B. Jassal, et al., The reactome pathway knowledgebase, Nucleic Acids Res. 48 (D1) (Jan. 2020) D498–D503, https://doi.org/10.1093/nar/gkz1031.
[60] N. del Toro, et al., The IntAct database: efficient access to fine-grained molecular interaction data, Nucleic Acids Res. 50 (D1) (Jan. 2022) D648–D653, https://doi.org/10.1093/nar/gkab1006.
[61] S. Burgess, A. Butterworth, S.G. Thompson, Mendelian randomization analysis with multiple genetic variants using summarized data, Genet. Epidemiol. 37 (7) (2013) 658–665, https://doi.org/10.1002/gepi.21758.
[62] Z.M. Younossi, A.B. Koenig, D. Abdelatif, Y. Fazel, L. Henry, M. Wymer, Global epidemiology of nonalcoholic fatty liver disease—meta-analytic assessment of prevalence, incidence, and outcomes, Hepatology 64 (1) (Jul. 2016) 73, https://doi.org/10.1002/hep.28431.
[63] G. Targher, C.D. Byrne, H. Tilg, NAFLD and increased risk of cardiovascular disease: clinical associations, pathophysiological mechanisms and pharmacological implications, Gut 69 (9) (Sep. 2020) 1691–1705, https://doi.org/10.1136/gutjnl-2020-320622.
[64] S.L. Friedman, B.A. Neuschwander-Tetri, M. Rinella, A.J. Sanyal, Mechanisms of NAFLD development and therapeutic strategies, Nat. Med. 24 (7) (Jul. 2018) 7, https://doi.org/10.1038/s41591-018-0104-9.
[65] E. Ferkingstad, et al., Large-scale integration of the plasma proteome with genetics and disease, Nat. Genet. 53 (12) (Dec. 2021) 12, https://doi.org/10.1038/s41588-021-00978-w.
[66] R. Parker, The role of adipose tissue in fatty liver diseases, Liver Research 2 (1) (Mar. 2018) 35–42, https://doi.org/10.1016/j.livres.2018.02.002.
[67] B. Fromenty, M. Roden, Mitochondrial alterations in fatty liver diseases, J. Hepatol. 78 (2) (Feb. 2023) 415–429, https://doi.org/10.1016/j.jhep.2022.09.020.
[68] Y. Zheng, S. Wang, J. Wu, Y. Wang, Mitochondrial metabolic dysfunction and non-alcoholic fatty liver disease: new insights from pathogenic mechanisms to clinically targeted therapy, J. Transl. Med. 21 (1) (Jul. 2023) 510, https://doi.org/10.1186/s12967-023-04367-1.