

RESEARCH ARTICLE

Open Access



The landscape of GWAS validation; systematic review identifying 309 validated non-coding variants across 130 human diseases

Ammar J. Alsheikh^{1*} , Sabrina Wollenhaupt², Emily A. King¹, Jonas Reeb², Sujana Ghosh¹, Lindsay R. Stolzenburg¹, Saleh Tamim¹, Jozef Lazar¹, J. Wade Davis¹ and Howard J. Jacob¹

Abstract

Background: The remarkable growth of genome-wide association studies (GWAS) has created a critical need to experimentally validate the disease-associated variants, 90% of which involve non-coding variants.

Methods: To determine how the field is addressing this urgent need, we performed a comprehensive literature review identifying 36,676 articles. These were reduced to 1454 articles through a set of filters using natural language processing and ontology-based text-mining. This was followed by manual curation and cross-referencing against the GWAS catalog, yielding a final set of 286 articles.

Results: We identified 309 experimentally validated non-coding GWAS variants, regulating 252 genes across 130 human disease traits. These variants covered a variety of regulatory mechanisms. Interestingly, 70% (215/309) acted through cis-regulatory elements, with the remaining through promoters (22%, 70/309) or non-coding RNAs (8%, 24/309). Several validation approaches were utilized in these studies, including gene expression (n = 272), transcription factor binding (n = 175), reporter assays (n = 171), in vivo models (n = 104), genome editing (n = 96) and chromatin interaction (n = 33).

Conclusions: This review of the literature is the first to systematically evaluate the status and the landscape of experimentation being used to validate non-coding GWAS-identified variants. Our results clearly underscore the multifaceted approach needed for experimental validation, have practical implications on variant prioritization and considerations of target gene nomination. While the field has a long way to go to validate the thousands of GWAS associations, we show that progress is being made and provide exemplars of validation studies covering a wide variety of mechanisms, target genes, and disease areas.

Keywords: GWAS, Experimental validation, Functional variant, Systematic review, Non-coding

Background

A central goal of genetics is to identify the genetic underpinnings of human diseases. Advancements in human genetics and its related fields and technologies over

the past decades have had a remarkable impact on our understanding of human disease pathophysiology, diagnosis and management [1]. In Mendelian disorders and rare genetic diseases this often takes the form of a loss-of-function mutation or genomic abnormality driving the disease phenotype. There are more than 5,000 diseases that belong to this category accounted for in the Online Mendelian Inheritance in Man (OMIM) database [2]. For complex diseases, there are multiple genetic and

*Correspondence: ammaralsheikh@gmail.com

¹ Genomics Research Center, AbbVie Inc, North Chicago, Illinois 60064, USA

Full list of author information is available at the end of the article



environmental factors contributing to disease risk and the identification of genetic risk factors associated with complex diseases has been rapidly accelerating with the utilization of next generation sequencing and dense array genotyping technologies in genome-wide association studies (GWAS). In a GWAS, thousands of genetic variants are genotyped in individuals which are then used to identify statistical associations between variants at certain genomic loci and a particular phenotype [3]. Since the first reported GWAS association for age-related macular degeneration [4] the use of these studies have grown exponentially, with over 200,000 genetic variants associated with more than 3000 human traits reported [5]. The remarkable growth of GWAS has created a critical need to experimentally identify and validate the disease-associated variants [6, 7]. This barrier has hindered the translation of GWAS findings to disease biology mechanisms and hence therapies. There are seemingly very few examples of GWAS-identified genetic loci at which the causal variant and molecular mechanisms driving the association have been experimentally determined, especially considering the sheer number of genotype–phenotype associations that have been reported passing the genome-wide significance threshold.

Dissecting GWAS loci to uncover the underlying biology is a complicated multi-step process. High linkage disequilibrium (LD) between many variants often necessitates utilizing statistical fine-mapping approaches and overlapping with functional genomic annotations for prioritization of variants before experimental validation [3, 8]. For coding variants, the target gene is identified directly from the genomic location of the variant [9]. As protein-coding regions represent only a small percentage of the human genome, more than 90% of GWAS associated variants are annotated to be within non-coding parts of the genome [5]. Experimental identification and validation of non-coding variants involves additional level of complexity as compared to coding variants requiring the application of additional approaches [10, 11]. Moreover, the functionality of regulatory elements is often cell-type specific, which necessitates studying the mechanism in disease-relevant cell types [12].

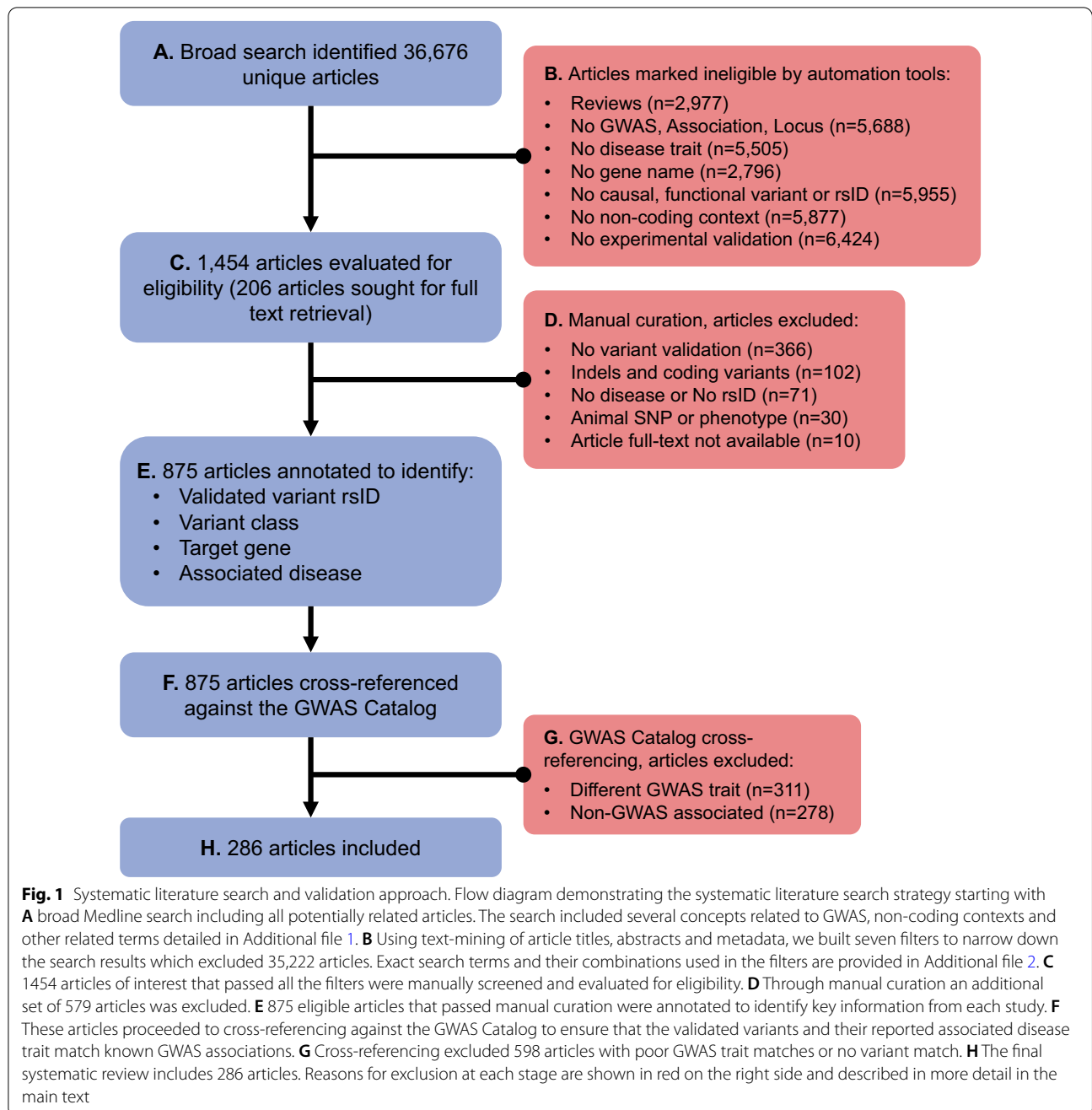
Experimental identification and validation are critical elements in translating GWAS findings. To date there has been limited study of the number of GWAS-identified loci that have been experimentally validated. A systematic literature review of 36,676 published articles identified 309 experimentally validated non-coding GWAS variants, regulating 252 genes across 130 human disease traits. This review of the literature is the first to systematically evaluate the status and the landscape of experimentation being used to validate non-coding GWAS-identified variants. We additionally curated key

information from all included studies such as validated variant class, distance-to-target gene, and experimental validation methods. Our findings have value for future experimental validation studies, target gene prioritization and functional variant prediction. The approaches utilized to validate coding variants as well as current methods used to nominate candidate functional variants for functional studies are outside the scope of this manuscript and have been reviewed previously [8, 9].

Methods

We conducted a systematic literature search and report it in compliance with the standards set forth by the 2020 PRISMA statement on the reporting of systematic reviews [13]. As a traditional keyword-based search approach would not enable us to thoroughly search for all relevant concepts and combinations, we leveraged natural language processing (NLP) and ontology-based text mining to ensure a systematic identification of relevant validation articles [14, 15]. We defined the scope to include studies that perform validation of GWAS associated non-coding variants at least at a molecular level.

In order to build a comprehensive literature search strategy, we first identified 28 validation studies from recent reviews and published resources [6, 7, 16]. These index studies were evaluated to identify the optimal keywords and concepts that would be used in the systematic literature search. Figure 1 shows a flow diagram summarizing the systematic literature search approach that was employed. The systematic literature search was conducted using search and filter concepts identified by thorough manual and text mining-supported concept analysis of index articles. The initial broad search was based on four different sub-queries aimed at identifying any articles that might include experimental validation of GWAS variants. We included explicit mention of GWAS, non-coding, functional or causal variant as well as contextual mentions of non-coding concepts such as enhancers and promoters (Additional file 1). Queries were run on MEDLINE Full Index [17] (all MEDLINE content until February 19, 2021) using IQVIA/Linguamatics I2E KNIME nodes [18]. Concepts and various combinations were searched in title, abstract and meta-data (author keywords, Medical Subject Headings (MeSH) terms and substances) leveraging public standard life science ontologies (such as MeSH [19], NCI Thesaurus [20] or Entrez Gene [21]), custom vocabularies and syntactical rules, grammatical pattern and linguistic entity classes allowing to build more generalized (comprehensive) queries, but at the same time more precise queries than standard key word search engines. The PMIDs identified by each query were combined and filtered for publication year ≥ 2007 (using “PubMed Publication Data (entrez”).



After removing duplicates, we arrived at 36,676 unique articles (Fig. 1A). We built seven filters reflecting our key inclusion criteria to narrow down the search results: (1) filter for primary research articles and exclude other article types, (2) GWAS and/or association filter, (3) filter for any human disease, (4) filter for any human gene (RefSeq), (5) filter for explicit mention of “non-coding” or non-coding context (enhancers, intron, non-coding, microRNA, etc.), (6) filter for functional, causal, or

regulatory variant or specific rsID, and (7) wet-lab experimental validation techniques (Fig. 1B, Additional file 2). Filters were built using an in-house entity extraction and literature classification pipeline combining SciBite’s TER-Mite (TERM identification, tagging & extraction) API coupled with SciBite’s VOCabs [22] and IQVIA/Linguamatics I2E Software.

In total 1454 articles passed all filter criteria and were then manually reviewed by three curators (Fig. 1C). All

articles had to meet the following criteria to be considered for inclusion: (1) investigate variants associated with a human disease, (2) include experimental wet-lab molecular validation of one or more variants, (3) include putative validation of at least one non-coding variant, and (4) investigate single nucleotide polymorphisms (SNPs), excluding indels, purely coding, somatic, or rare variants. Abstracts and full texts were reviewed resulting in the exclusion of 579 articles (Fig. 1D). Overall, this manual review identified 875 potentially relevant articles. All these articles were manually curated to confirm the rsID of the reportedly validated variants, variant class, the reported regulated gene, and the associated disease (Fig. 1E).

We then used the information on the validated variant's rsID and disease trait to cross validate our data with the GWAS Catalog [5] (accessed Mar 25, 2021) to confirm that each curated variant-disease association is reported in a GWAS (Fig. 1F). Corresponding associations were identified through LD between the curated SNP and the reported GWAS Catalog SNP, and similarity between the reported GWAS trait and the traits extracted from the PubMed abstract as detailed below. Because the GWAS Catalog only reports the lead variant for each locus, and this variant is not necessarily identical to the causal variant for the association, we performed an LD expansion from each top SNP to identify additional possible causal variants. Broad ancestry as reported in the GWAS Catalog was mapped to a 1000 Genomes superpopulation following methods we described recently [23]. For each associated SNP in the GWAS Catalog, an LD expansion was performed to identify SNPs within 1 Mb with LD $r^2 \geq 0.5$ in the corresponding 1000 Genomes superpopulation. A minor allele count threshold of 5 within the corresponding superpopulation was applied to reduce the impact of high variance LD estimates for rare variants. If it was not possible to map to a single superpopulation, LD expansion was performed using the full 1000 Genomes Phase 3 GRCh38 liftover to match the build used in the GWAS Catalog [24]. When the GWAS Catalog reported a specific risk allele, our LD expansion took this into account, such that for multiallelic SNPs we would only identify variants correlated with the reported allele. The choice of LD threshold is motivated by the goal to capture GWAS associations that could plausibly be explained by the cataloged variant and has been used elsewhere [25]. Using this methodology, it was possible to perform LD expansion for 91% of variants in the GWAS Catalog. GWAS Catalog variants for which an LD expansion was not possible were still included in the analysis but could only be matched to

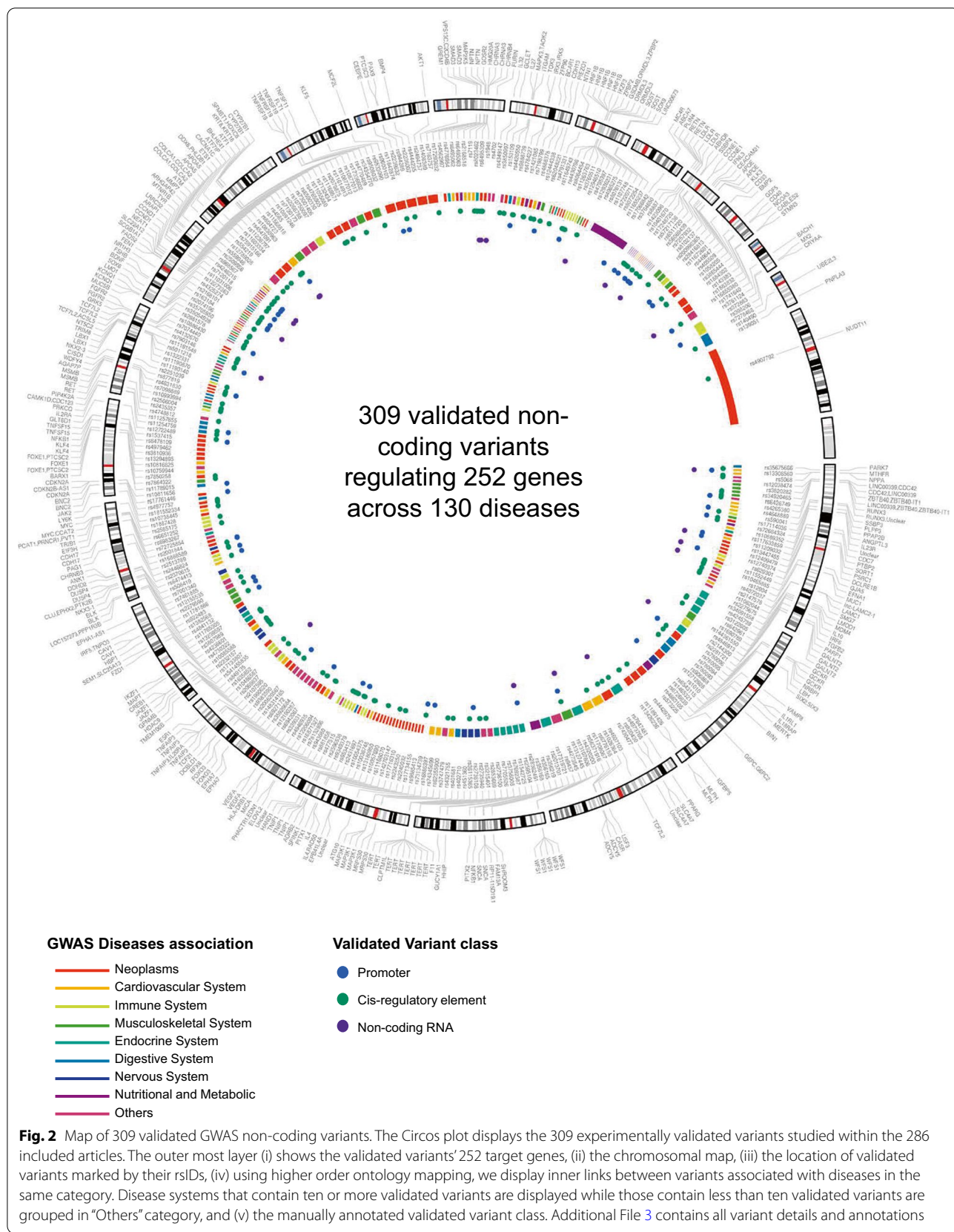
the reported variant rather than other possible causal variants.

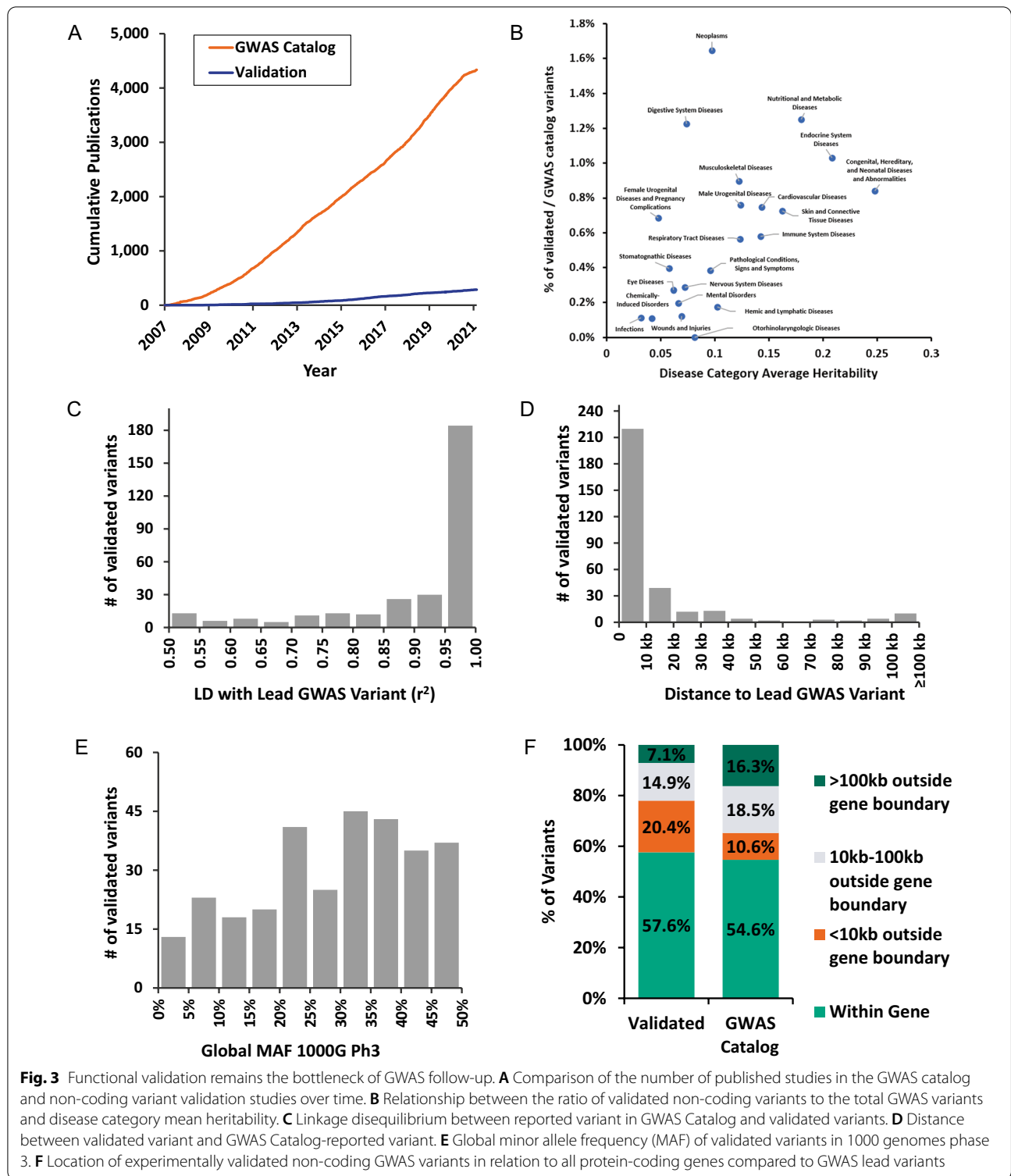
GWAS Catalog Experimental Factor Ontology (EFO) terms and disease terms curated from the literature were mapped to the 2020 MeSH thesaurus vocabulary using the approach outlined previously [26]. To allow for inexact matches in MeSH terms (e.g., hypertension and systolic blood pressure), we use two similarity metrics: Lin-Resnik average similarity with a cutoff value of 0.75 [26, 27] and odds ratio of MeSH term co-occurrence in the same PubMed article with a cutoff of 20 [23]. We count a match between an article identified in our systematic review and a GWAS study if any GWAS Catalog association satisfies the following criteria: (1) The reported variant in the GWAS Catalog has LD $R^2 \geq 0.5$ to at least one curated variant, and (2) the reported trait in the GWAS Catalog has similarity to a main or manually curated disease from the PubMed abstract, meeting or exceeding the cutoff value. We excluded 347 SNPs in 311 articles from the analysis due to not being linked to a GWAS Catalog SNP. A further 292 SNPs contained within 278 articles were excluded due to a poor match between the reported GWAS trait and the trait reported in the abstract (Fig. 1G). The final curated catalog includes 286 articles (Fig. 1H) [28–313].

Results

Curated catalog of 309 validated GWAS non-coding variants

Several prior studies have emphasized the importance of experimental validations to uncover the biological processes underlying the statistical GWAS associations [3, 6, 7, 314, 315]. The final list of 286 articles reports 309 experimentally validated functional non-coding variants regulating 252 genes across 130 human-diseases (Additional file 3 and Fig. 2). Additional File 3 includes several important aspects about the included articles and variants including PubMed identifiers (PMID), variant rsID, location, class, target gene as well as disease associations and experimental validation approaches. We examined several characteristics of the validated non-coding variants in relation to GWAS catalog studies and variants. Between 2007 and 2020 there is a steady increase in the number of validation articles over time up to the 286 we report here. In contrast, the total number of published GWAS articles is 4342 versus 286 validation articles for non-coding variants (Fig. 3A). Next, we evaluated the relationship between disease heritability explained by common SNPs and the ratio of validated variants to the total number of lead-GWAS variants. We mapped disease associations for all variants to the higher order disease categories in the MeSH terms tree structure. For heritability estimates, we considered liability scale h^2





for UK Biobank phenotypes estimated using LD Score Regression[316, 317] which (1) mapped to a MeSH disease (2) were considered high or medium confidence and averaged the heritability across higher level MeSH to

get average heritability per disease category. Using this approach, we find a statistically significant ($p=0.01$; correlation coefficient 0.51) positive relationship between mean heritability and the ratio of validated/lead GWAS

variants per disease category (Fig. 3B). Examination of individual validated variants showed the majority of validated variants are in strong LD with and in close proximity to the GWAS variant (Fig. 3C, D). Allele frequencies of validated variants have slightly skewed distribution with fewer validated variants having lower allele frequencies (Fig. 3E). Comparing the location of experimentally validated non-coding GWAS variants to GWAS lead variants, we found that validated variants are about equally likely to be located within a protein-coding gene (58% for functional variants versus 55% for GWAS lead variants). However, they are much more likely to be within 10 kb of a gene boundary (20% versus 11%) and much less likely to be more than 100 kb from the nearest gene (7% versus 16%) (Fig. 3F). Overall, these findings quantify the persistent need for more experimental validation studies to bridge the gap between association and biology. These findings also suggest that focusing experimental validation efforts to variants in close proximity and strong LD to the lead GWAS variant would lead to the identification of a causal variant in the majority of genetic loci.

Validated variants regulate 252 target genes through a variety of mechanisms

Non-coding genetic variants can exert their effect on target genes through a variety of mechanisms [318–320]. We divided variants into three broad categories based on their mechanism of regulation: cis-regulatory element (CRE) variants, promoter variants and variants acting through non-coding RNAs (Fig. 4A). Promoter variants were grouped separately from other CREs because they are functionally distinct and in addition the methods utilized for their validation are different from other CREs. Below we highlight several exemplar studies validating variants across all these mechanisms and many diseases. Interestingly, the majority of non-coding variants identified in our catalog regulate genes through CREs (n=215). These include variants in enhancers such as rs4420550-*MAPK3-TAOK2* in schizophrenia [168], rs11236797-*LRRC32* in inflammatory bowel disease [40], and rs9349379-*EDNI* in vascular diseases [49]. Some variants exerted their effect through silencers such as rs12038474-*CDC42* in endometriosis [130], rs2494737-*AKT1* in endometrial carcinoma [37] and rs9508032-*FLT1* in acute respiratory distress syndrome [267]. Additionally, rs12936231-*GSDMB-ORMDL3-ZPBP2*

seems to function through an insulator in an asthma and autoimmune disease risk locus [71].

Variants in gene promoters can alter transcription factor binding and promoter activity. For example, rs1887428-*JAK2* in inflammatory bowel disease [256], rs11789015-*BARX1* in esophageal adenocarcinoma [88], rs4065275-*ORMDL3* and rs8076131-*ORMDL3* in asthma, [248] and rs11603334-*ARAPI* in type 2 diabetes mellitus [34]. DNA methylation is an important epigenetic mechanism of gene regulation and increased DNA methylation at gene promoters can repress gene transcription [321, 322]. We identified several validated variants that appear to alter promoter methylation including rs780093-*NRBPI* in gout [127], rs143383-*GDF5* in osteoarthritis [119], and rs35705950-*MUC5B* in idiopathic pulmonary fibrosis [258]. Alternatively, variants could alter promoter and transcription start site usage. Examples for these mechanisms in our catalog include rs922483-*BLK* in systemic lupus erythematosus [302] and rs10465885-*GJA5* in atrial fibrillation [32].

The third broad category by which variants from our catalog exert their regulatory effect is through non-coding RNAs [323]. microRNAs are a major and well-studied class of regulatory small non-coding RNAs. Variants in microRNAs are known to impact disease biology through post-transcriptional regulation of their target genes, primarily via 3' untranslated region (UTR) binding [324–326]. GWAS variants located within microRNAs can alter their biogenesis, expression levels and/or target specificity, while variants located in target genes are capable of altering microRNA binding sites [326]. Examples of validated variants within microRNAs included in this catalog are miR-196a2 variant rs11614913 regulating *SFMBT1* and *HOXC8* in metabolic syndrome [277], and miR-4513 variant rs2168518 regulating *GOSR2* in cardio-metabolic diseases [51]. Given that microRNAs typically target hundreds to thousands of genes, it is very difficult to confidently assign target genes that are mediating the effect of a microRNA variant. On the other hand, studying variants located within microRNA-binding sites of target genes may yield more success in assigning underlying mechanisms [326, 327]. There are numerous examples of such variants reported in this catalog, such as rs5068 altering regulation of *NPPA* by miR-425 in hypertension [96], rs1058205 altering regulation of *KLK3* by miR-3162-5p and rs1010 altering regulation of *VAMP8*

(See figure on next page.)

Fig. 4 Non-coding variants regulate 252 target genes through diverse mechanisms. **A** Illustration of some of the diverse mechanisms of regulation within each variant category. Examples of each mechanism from included studies are discussed in the text. **B** Cumulative number of validated variants grouped by non-coding variant categories over time. **C** We used Ensembl's Biomart and hg38 to calculate the distance (in kb) between validated variants and their target gene's closest transcription start site (TSS). Graph plots the number of variant-gene pairs grouped by variant class. Variants more than 200 kb away are plotted at 200 kb. **D** Distribution of CRE variants relative to their target gene. CRE = Cis-Regulatory Element, ncRNA = non-coding RNA

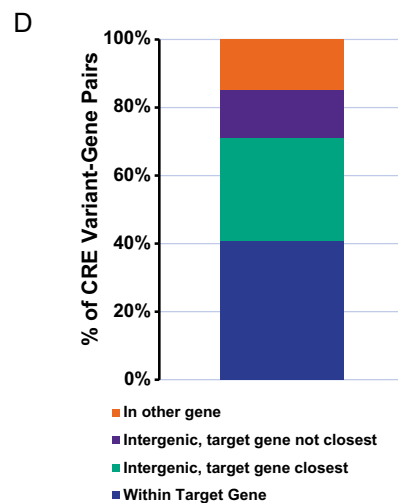
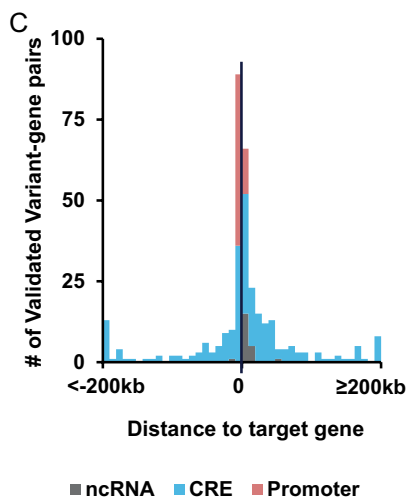
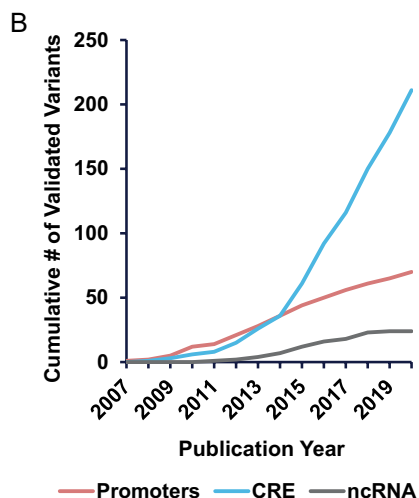
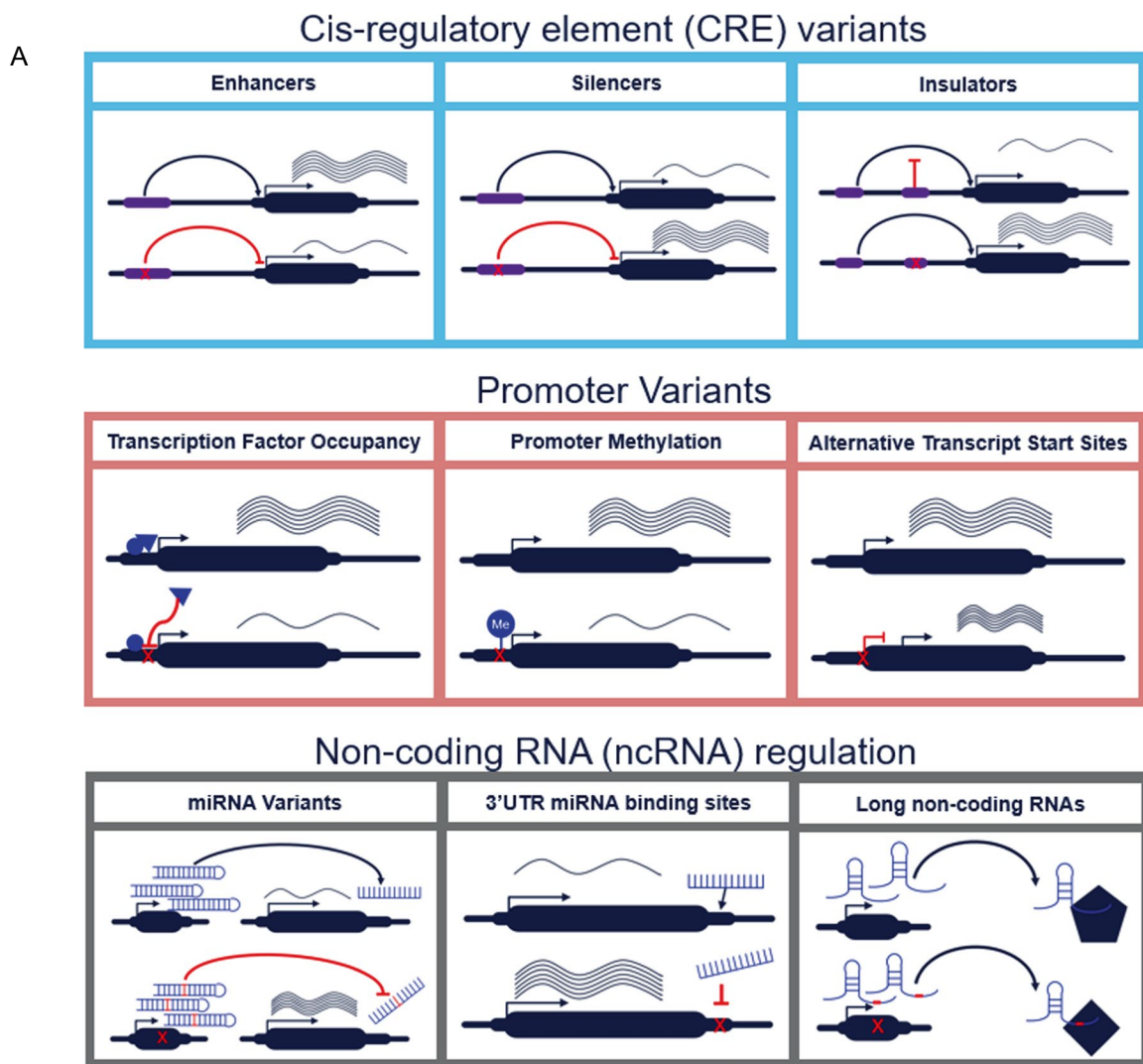


Fig. 4 (See legend on previous page.)

by miR-370 in prostate cancer [54], and rs372883 altering *BACH1* regulation by miR-1257 in pancreatic ductal adenocarcinoma [174]. Another important class of non-coding RNAs is long non-coding RNAs that are recognized to play an important role in biology and disease [328, 329]. Some examples of long non-coding RNA variants in this catalog include rs6983267 in *CCAT2* regulating cancer metabolism through allele-specific binding of *CPSF7* [76] and rs2147578 in *LAMC2-1* modulating microRNA binding to it in colorectal cancer [43]. We examined the distribution of these three broad categories of validated variants across publication dates. We observed a steady increase in the validation of promoter variants ($n=70$) and variants acting through non-coding RNAs ($n=24$) since 2007, but a sharp increase in the number of studies validating CRE variants around 2015. This trend persisted through 2020 to reach a total of 215 variants representing 70% of this catalog (Fig. 4B). We also characterized the distance between each validated variant and its target gene's closest transcription start site according to variant category. As expected, promoter variants clustered immediately upstream or downstream of their target's transcription start site. CRE variants were more widely distributed, but nevertheless, 157 (66%) of these fell within 50 kb from their target gene TSS. A notable example of a distally acting enhancer variant >50 kb, is the obesity *FTO* locus variant rs1421085 regulating *IRX3* and *IRX5*, which are 500 kb and 1,163 kb away respectively [147]. Since the majority of variants acting through non-coding RNAs identified in our catalog were located within 3' UTRs, this group of variants tended to cluster within 100 kb downstream of gene transcript start sites (Fig. 4C). The dataset gave us the opportunity to examine the relationship between CRE variants and their target genes ($n=235$ CRE variant-target gene pairs). Plotting the distribution of CRE variants based on their location relative to the target gene indicated that 41% of CRE variants are located within their target gene, and an additional 30% are intergenic and their target gene is the closest gene to the variant. 14% of CRE variants were intergenic and their target gene is not the closest gene, and the remaining 15% are located within a different gene than their target gene. (Fig. 4D). These results are interesting and provide greater support for consideration of same gene and nearby genes as candidate targets for CREs. These findings are also in agreement with recent empirical data [330, 331].

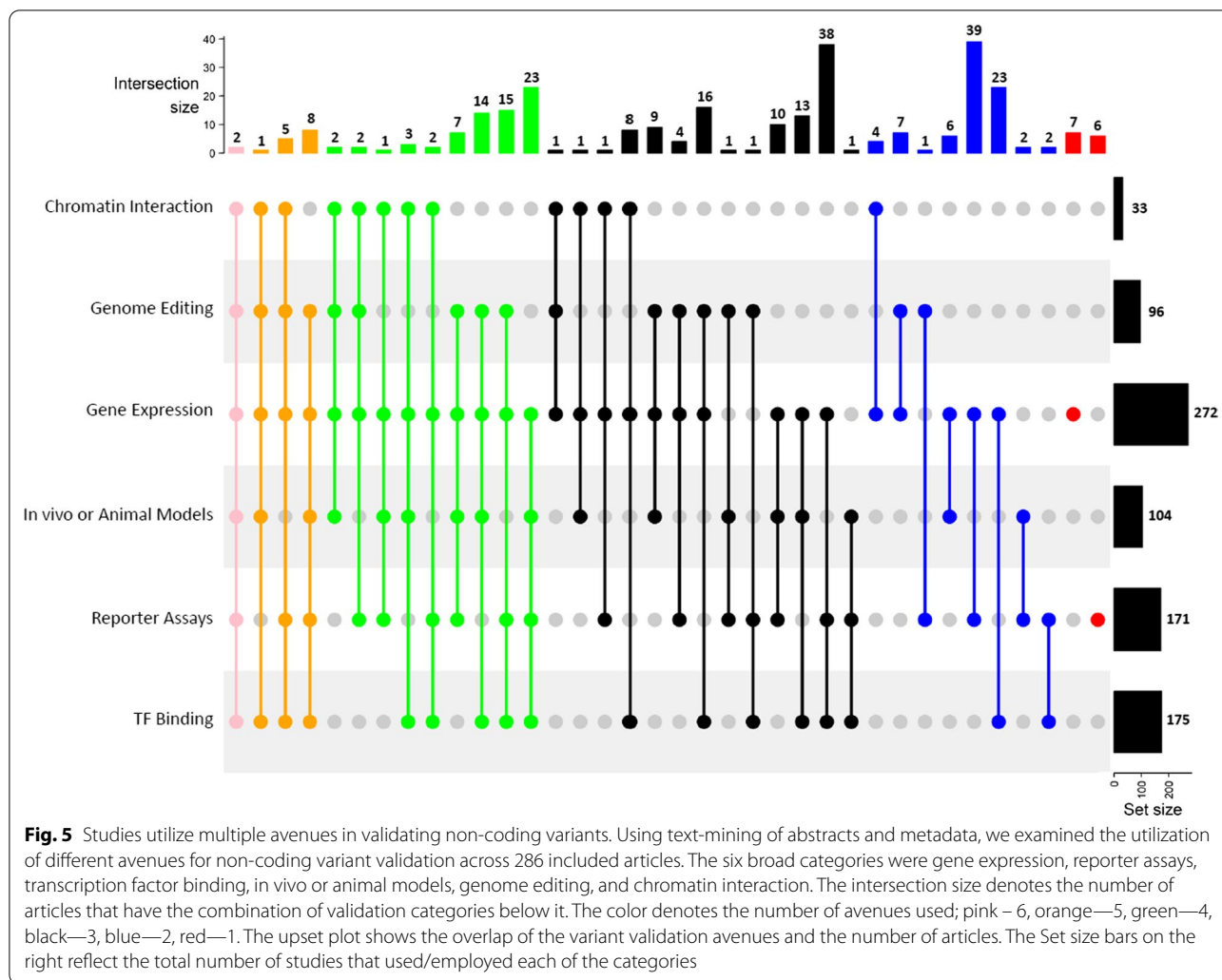
Next, using text mining, we extracted and analyzed the experimental methods that were used in each study to validate variants. We broadly classified them under six broad categories covering different types of established validation techniques and related terms: (1) gene expression, including eQTL and molecular assessment

of target gene expression and allele specific regulation ($n=272$ articles), (2) reporter assays, including luciferase and massively parallel reporter assays ($n=171$ articles), (3) transcription factor binding, including chromatin immunoprecipitation and electrophoretic mobility shift assays ($n=175$ articles), (4) in vivo or animal models ($n=104$ articles), (5) genome editing, including CRISPR and TALEN ($n=96$ articles), and (6) chromatin interaction, including chromosome conformation capture ($n=33$ articles) [11]. We examined the number of these approaches that were utilized by the included studies and found that 189 (66%) of all articles utilized three or more approaches (Fig. 5). These results demonstrate the multifaceted approach needed for validation of non-coding variants [11].

Discussion

GWAS have seen a remarkable growth in the past decade. The impact of GWAS on human healthcare is severely limited by the bottle neck of experimental validation of disease-associated variants. Here, we report the first systematic approach to curate all experimental validation studies of non-coding GWAS variants. While there is general recognition that experimental validation of GWAS are seriously lacking [7], this systematic assessment of (1) the number of published experimentally validated non-coding variants is quantified, (2) cataloged, and (3) methods used in identified studies analyzed.

Using a comprehensive approach, we employed natural-language processing-based text mining, manual curation and GWAS catalog cross validation. We have curated 286 validation studies that include 309 putatively validated variants regulating 252 genes across 130 diseases. We then evaluated several important characteristics of the identified variants and their relation to GWAS lead variants. The ratio of validated non-coding variants to total GWAS lead variants showed a positive correlation to the mean heritability of disease groups. This relationship could indicate greater success in validating variants in diseases with higher heritability perhaps because of greater individual contribution of these variants to the overall disease susceptibility. This could also potentially represent a greater interest of scientists to pursue validation of variants in more heritable diseases and with larger effect sizes, thus leading to greater proportion of variants being validated. However, we do not have enough data to directly address this possibility. We also evaluated the relationship in LD and distance between validated variants and GWAS lead variants. We find that $\sim 70\%$ of validated variants fall within 10 kb and $r^2 \geq 0.9$ with the lead GWAS variant. On one hand, this could reflect underlying genetics that most validated variants are in strong LD with lead GWAS variants and suggests that more



productive research should be limited to SNPs in high LD and closer distance to lead GWAS variants. On the other hand, the status quo might be reflective of prior limits in search space already considered by scientists who performed validation studies, however we do not have data to support this possibility[8].

Next, we annotated variants into broad classes based on the mechanisms by which these non-coding variants acted. This identified several interesting patterns, such as an increase in the number of variants functioning through cis-regulatory elements over time. One explanation for this increase could be the growing awareness of the importance of these regulatory elements in human biology and disease which has led to the initiation of large projects aimed at identification, annotation and prioritization of non-coding regulatory elements [10, 320, 332]. Additionally, several SNP-enrichment analyses have demonstrated that GWAS variants are significantly enriched in active regulatory regions [314]. We expect

this trend to continue with publications by larger consortia and projects that investigate regulatory elements in different life stages, tissues and biological conditions [332]. Interestingly, the majority of cis-regulatory element variants that we found appeared to act through transcriptional enhancers. This dominance of enhancer variants over other regulatory elements might be a result of enhancer elements having more clearly defined functions and biochemical markers (i.e., histone modification signatures) [333, 334]. This highlights the potential for increased discovery of GWAS variants acting through silencers and insulators as our understanding of their distinct biochemical signatures is refined and assayed in disease relevant cell types [333, 335].

Our comprehensive search and filter strategy enabled us to identify validated variants across a large number of complex human diseases and those that act through a myriad of mechanisms. Nevertheless, the systematic search was limited to the MEDLINE database. Relevant

articles published in journals not indexed in this standard database for biomedical literature will be missing in our data set [336, 337]. For quality control and to identify limitations of our search and filter approach, we analyzed the recall of our index studies throughout the entire process (Fig. 1A–H). It is important to highlight that broadening the initial search to include non-coding contexts and association/locus instead of limiting to explicit mentions of non-coding and GWAS terms ensured identification of relevant studies that we had otherwise missed. A significant number of index articles did not explicitly mention these terms [48, 78, 134, 143, 147, 171, 178, 210, 230, 256, 302]. Our final broad search covered 27 out of the 28 index studies which demonstrates good search coverage. Through an iterative process, we narrowed down these results, trying to maximize the recall of index studies while maintaining a manageable number of articles for manual review. We are aware that the implemented stringent criteria bias the search to exclude true validation articles that did not mention any disease, protein or specific experimental validation terms [338–345]. Additionally, the tagging of the articles and normalization of concepts for filtering relies on accurate named entity recognition (NER) and ontologies. Even when using highly curated, enriched vocabularies and state-of-the-art NER routines, recall rates of at maximum 80–95% are assumed (depending on entity type). Overall, a total of 19 index studies passed all filtering stages and were included in the final catalog. Finally, the data of our curated catalog is mainly based on the publications' abstract information. Only in cases where information was missing or unclear in the abstract did we gather data from the full text. Therefore, it is possible that information gathered from the final set of articles may be incomplete. This would have affected the experimental validation techniques analysis in particular, which was based only on abstract mining.

Construction of the catalog using controlled vocabularies for diseases, variants, genes, variant classes, and functional follow up methods is aimed to facilitate use in bioinformatics follow up analyses. We expect this resource to be useful in evaluating the performance of computational fine mapping and target prioritization methods. Quantifying the performance of these methods on real datasets has previously been hindered by a lack of true positive examples. A large dataset of true positive examples would allow researchers to computationally identify features associated with functional variation. Recent efforts to compile such true positive datasets and use them to train target prioritization methods have come with concerns about bias towards coding variation [16] or are aimed at a specific trait subset such as molecular phenotypes [346] or immune disease [347].

We expect this catalog to contribute a large number of much needed examples of functional noncoding variants in human disease and the genes on which they act. Despite this important contribution, bias towards nearby genes and variants to the top GWAS SNP is still a concern for our catalog due to the limited number of variants and genes evaluated in the cataloged studies. To generate an unbiased training set for computational methods, an ideal functional study following up on a GWAS association would consider all credible causal SNPs and their nearby genes, but studies in our catalog typically consider a more limited set of genes and SNPs. For example, eQTL variants may be shared among multiple transcripts [348], and in this scenario functional studies considering only a single gene could be misleading about the causal gene.

Conclusions

This review is the first to systematically evaluate the status and the landscape of experimentation being used to validate non-coding GWAS-identified variants. Our results clearly underscore the multifaceted approach needed for experimental validation. The findings of validated variants relationship to lead GWAS variants as well as to their target genes provide practical insights for future validation studies. Finally, we aim for the catalog to be a useful resource aiding in the development of prediction tools by providing a truth set of experimentally validated variants. Collectively this contributes to the overall effort to bridge the gap between genetic association and function in complex diseases.

Abbreviations

CRE: Cis-regulatory element; GWAS: Genome-Wide Association Study; LD: Linkage disequilibrium; MeSH: Medical subject headings; PRISMA: Preferred reporting items for systematic reviews and meta-analyses; SNP: Single nucleotide polymorphism.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12920-022-01216-w>.

Additional file 1 contains exact search terms and criteria used for creating the initial broad literature search.

Additional file 2 contains exact terms and phrases used to setup the seven filters that were used to narrow down the broad search results.

Additional file 3 contains all the validated variants and their details. The file is formatted to include separate rows for unique PMID-variant-gene triples, therefore variants that regulate multiple genes and variants that have been validated in more than one publication have more than one row in the file.

Acknowledgements

The authors would like to thank Rainer Winnenburg of AbbVie for his assistance with higher order disease mapping and Mark Reppell of AbbVie for helpful advice pertaining to the GWAS Catalog portion of the analysis and for suggesting heritability datasets.

Authors' contributions

AA, SW, JL designed the study. SW, JR collected the data. AA, SW, EK, JR, SG, ST, LS analyzed the results. AA, SW, EK, JR, SG, ST, LS, JL, JWD and HJ interpreted results. AA, SW, EK wrote the manuscript. JL, JWD and HJ supervised the project. All authors read and approved the final manuscript.

Funding

The design, study conduct, and financial support for this research were provided by AbbVie. AbbVie participated in the interpretation of data, review, and approval of the publication.

Availability of data and materials

The data supporting the conclusions of this article is included within the article (and its additional files).

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

AA, SW, EK, JR, SG, ST and HJ are employees of AbbVie. LS, JWD and JL were employees of AbbVie at the time of the study.

Author details

¹Genomics Research Center, AbbVie Inc, North Chicago, Illinois 60064, USA.

²Information Research, AbbVie Deutschland GmbH & Co. KG, 67061 Knollstrasse, Ludwigshafen, Germany.

Received: 13 October 2021 Accepted: 17 March 2022

Published online: 01 April 2022

References

1. Collins FS, Doudna JA, Lander ES, Rotimi CN. Human molecular genetics and genomics—Important advances and exciting possibilities. *N Engl J Med*. 2021;384:1–4.
2. OMIM - Online Mendelian Inheritance in Man. <https://www.omim.org/>. 2021 [cited 2021 Apr 11]; Available from: <https://www.omim.org/>
3. Tam V, Patel N, Turcotte M, Bossé Y, Paré G, Meyre D. Benefits and limitations of genome-wide association studies. *Nat Rev Genet*. 2019;20:467–84.
4. Klein RJ, Zeiss C, Chew EY, Tsai J-Y, Sackler RS, Haynes C, et al. Complement factor H polymorphism in age-related macular degeneration. *Science*. 2005;308:385–9.
5. Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res*. 2019;47:D1005–12.
6. Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, et al. 10 Years of GWAS discovery: biology, function, and translation. *Am J Hum Genet*. 2017;101:5–22.
7. Gallagher MD, Chen-Plotkin AS. The Post-GWAS Era: From Association to Function. *Am J Hum Genet*. 2018;102:717–30.
8. Schaid DJ, Chen W, Larson NB. From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nat Rev Genet*. 2018;19:491–504.
9. Cai M, Ran D, Zhang X. Advances in identifying coding variants of common complex diseases. *J Bio-X Res*. 2019;2:153–8.
10. Tak YG, Farnham PJ. Making sense of GWAS: using epigenomics and genome engineering to understand the functional relevance of SNPs in non-coding regions of the human genome. *Epigenetics Chromatin*. 2015;8:57.

11. Rao S, Yao Y, Bauer DE. Editing GWAS: experimental approaches to dissect and exploit disease-associated genetic variation. *Genome Med*. 2021;13:41.
12. Liu B, Montgomery SB. Identifying causal variants and genes using functional genomics in specialized cell types and contexts. *Hum Genet*. 2020;139:95–102.
13. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ: Br Med J Publ Group*. 2021;372:71.
14. Chang M, Chang M, Reed JZ, Milward D, Xu JJ, Cornell WD. Developing timely insights into comparative effectiveness research with a text-mining pipeline. *Drug Discov Today*. 2016;21:473–80.
15. McEntire R, Szalkowski D, Butler J, Kuo MS, Chang M, Chang M, et al. Application of an automated natural language processing (NLP) workflow to enable federated search of external biomedical content in drug discovery and development. *Drug Discov Today*. 2016;21:826–35.
16. Ghossaini M, Mountjoy E, Carmona M, Peat G, Schmidt EM, Hercules A, et al. Open Targets Genetics: systematic identification of trait-associated genes using large-scale genetics and functional genomics. *Nucleic Acids Res*. 2021;49:D1311–20.
17. MEDLINE. <http://wayback.archive-it.org/org-350/20180312141554/https://www.nlm.nih.gov/pubs/factsheets/medline.html>. 2021 [cited 2021 Jun 15]; Available from: <http://wayback.archive-it.org/org-350/20180312141554/https://www.nlm.nih.gov/pubs/factsheets/medline.html>
18. Linguamatics. <https://www.linguamatics.com/>. 2021 [cited 2021 Jun 15]; Available from: <https://www.linguamatics.com/>
19. Medical Subject Headings - Home Page [Internet]. U.S. National Library of Medicine; [cited 2021 Jun 15]. Available from: <https://www.nlm.nih.gov/mesh/meshhome.html>
20. NCI Thesaurus. <https://ncit.nci.nih.gov/ncitbrowser/>. 2021;
21. NCBI Gene Database. <https://www.ncbi.nlm.nih.gov/gene/>. 2021 [cited 2021 Jun 15]; Available from: <https://www.ncbi.nlm.nih.gov/gene/>
22. TERMite - SciBite. <https://www.scibite.com/platform/termite/>. SciBite [Internet]. 2021 [cited 2021 Jun 15]; Available from: <https://www.scibite.com/platform/termite/>
23. King EA, Dunbar F, Davis JW, Degner JF. Estimating colocalization probability from limited summary statistics. *BMC Bioinform*. 2021;22:254.
24. Auton A, Abecasis GR, Altshuler DM, Durbin RM, Abecasis GR, Bentley DR, et al. A global reference for human genetic variation. *Nature*. 2015;526:68–74.
25. Farh KK-H, Marson A, Zhu J, Kleinewietfeld M, Housley WJ, Beik S, et al. Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature*. 2015;518:337–43.
26. King EA, Davis JW, Degner JF. 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*. 2019;15:e1008489.
27. Nelson MR, Tipney H, Painter JL, Shen J, Nicoletti P, Shen Y, et al. The support of human genetic evidence for approved drug indications. *Nat Genet*. 2015;47:856–60.
28. Almontashiri NAM, Antoine D, Zhou X, Vilmundarson RO, Zhang SX, Hao KN, et al. 9p21.3 coronary artery disease risk variants disrupt TEAD transcription factor-dependent transforming growth factor β regulation of p16 expression in human aortic smooth muscle cells. *Circulation*. 2015;132:1969–78.
29. Yu C-Y, Han J-X, Zhang J, Jiang P, Shen C, Guo F, et al. A 16q22.1 variant confers susceptibility to colorectal cancer as a distal regulator of ZFP90. *Oncogene*. 2020;39:1347–60.
30. Piao X, Yahagi N, Takeuchi Y, Aita Y, Murayama Y, Sawada Y, et al. A candidate functional SNP rs7074440 in TCF7L2 alters gene expression through C-FOS in hepatocytes. *FEBS Lett England*. 2018;592:422–33.
31. Kretschmer A, Möller G, Lee H, Laumen H, von Toerne C, Schramm K, et al. A common atopy-associated variant in the Th2 cytokine locus control region impacts transcriptional regulation and alters SMAD3 and SP1 binding. *Allergy Denmark*. 2014;69:632–42.
32. Wirka RC, Gore S, Van Wagoner DR, Arking DE, Lubitz SA, Lunetta KL, et al. A common connexin-40 gene promoter variant affects connexin-40 expression in human atria and is associated with atrial fibrillation. *Circ Arrhythm Electrophysiol*. 2011;4:87–93.

33. Lattka E, Eggers S, Moeller G, Heim K, Weber M, Mehta D, et al. A common FADS2 promoter polymorphism increases promoter activity and facilitates binding of transcription factor ELK1. *J Lipid Res.* 2010;51:182–91.
34. Kulzer JR, Stitzel ML, Morken MA, Huyghe JR, Fuchsberger C, Kuusisto J, et al. A common functional regulatory variant at a type 2 diabetes locus upregulates ARAP1 expression in the pancreatic beta cell. *Am J Hum Genet.* 2014;94:186–97.
35. Choi J, Xu M, Makowski MM, Zhang T, Law MH, Kovacs MA, et al. A common intronic variant of PARP1 confers melanoma risk and mediates melanocyte growth via regulation of MITF. *Nat Genet United States.* 2017;49:1326–35.
36. Kycia I, Wolford BN, Huyghe JR, Fuchsberger C, Vadlamudi S, Kursawe R, et al. A common Type 2 diabetes risk variant potentiates activity of an evolutionarily conserved islet stretch enhancer and increases C2CD4A and C2CD4B expression. *Am J Hum Genet.* 2018;102:620–35.
37. Painter JN, Kaufmann S, O'Mara TA, Hillman KM, Sivakumaran H, Darabi H, et al. A common variant at the 14q32 endometrial cancer risk locus activates AKT1 through YY1 binding. *Am J Hum Genet.* 2016;98:1159–69.
38. Guo X, Lin W, Bao J, Cai Q, Pan X, Bai M, et al. A Comprehensive cis-eQTL analysis revealed target genes in breast cancer susceptibility loci identified in genome-wide association studies. *Am J Hum Genet.* 2018;102:890–903.
39. Gallagher MD, Posavi M, Huang P, Unger TL, Berlyand Y, Gruenewald AL, et al. A dementia-associated risk variant near TMEM106B alters chromatin architecture and gene expression. *Am J Hum Genet.* 2017;101:643–63.
40. Nasrallah R, Imianowski CJ, Bossini-Castillo L, Grant FM, Dogan M, Placek L, et al. A distal enhancer at risk locus 11q13.5 promotes suppression of colitis by T(reg) cells. *Nature.* 2020;583:447–52.
41. Díaz-Jiménez D, Núñez L, De la Fuente M, Dubois-Camacho K, Sepúlveda H, Montecino M, et al. A functional IL1RL1 variant regulates corticosteroid-induced sST2 expression in ulcerative colitis. *Sci Rep.* 2017;7:10180.
42. Shou W, Wang Y, Xie F, Wang B, Yang L, Wu H, et al. A functional polymorphism affecting the APOA5 gene expression is causally associated with plasma triglyceride levels conferring coronary atherosclerosis risk in Han Chinese Population. *Biochim Biophys Acta Netherlands.* 2014;1842:2147–54.
43. Gong J, Tian J, Lou J, Ke J, Li L, Li J, et al. A functional polymorphism in lnc-LAMC2-1:1 confers risk of colorectal cancer by affecting miRNA binding. *Carcinogenesis England.* 2016;37:443–51.
44. Saeki N, Saito A, Choi IJ, Matsuo K, Ohnami S, Totsuka H, et al. A functional single nucleotide polymorphism in mucin 1, at chromosome 1q22, determines susceptibility to diffuse-type gastric cancer. *Gastroenterol USA.* 2011;140:892–902.
45. Ogura Y, Kou I, Miura S, Takahashi A, Xu L, Takeda K, et al. A functional SNP in BNC2 is associated with adolescent idiopathic scoliosis. *Am J Hum Genet.* 2015;97:337–42.
46. Ye J, Tucker NR, Weng L-C, Clauss S, Lubitz SA, Ellinor PT. A functional variant associated with atrial fibrillation regulates PITX2c expression through TFAP2a. *Am J Hum Genet.* 2016;99:1281–91.
47. Akamatsu S, Takata R, Ashikawa K, Hosono N, Kamatani N, Fujioka T, et al. A functional variant in NKX3.1 associated with prostate cancer susceptibility down-regulates NKX3.1 expression. *Hum Mol Genet.* 2010;19:4265–72.
48. Ali MW, Patro CPK, Zhu JJ, Dampier CH, Plummer SJ, Kuscic C, et al. A functional variant on 20q13.33 related to glioma risk alters enhancer activity and modulates expression of multiple genes. *Hum Mutat.* 2021;42:77–88.
49. Gupta RM, Hadaya J, Trehan A, Zekavat SM, Roselli C, Klarin D, et al. A genetic variant associated with five vascular diseases is a distal regulator of endothelin-1 gene expression. *Cell.* 2017;170:522–533.e15.
50. De Castro-Orós I, Pérez-López J, Mateo-Gallego R, Rebollar S, Ledesma M, León M, et al. A genetic variant in the LDLR promoter is responsible for part of the LDL-cholesterol variability in primary hypercholesterolemia. *BMC Med Genomics.* 2014;7:17.
51. Ghanbari M, de Vries PS, de Looper H, Peters MJ, Schurmann C, Yaghothkar H, et al. A genetic variant in the seed region of miR-4513 shows pleiotropic effects on lipid and glucose homeostasis, blood pressure, and coronary artery disease. *Hum Mutat USA.* 2014;35:1524–31.
52. Stegeman S, Moya L, Selth LA, Spurdle AB, Clements JA, Batra J. A genetic variant of MDM4 influences regulation by multiple microRNAs in prostate cancer. *Endocr Relat Cancer England.* 2015;22:265–76.
53. Schaefer AS, Richter GM, Nothnagel M, Manke T, Dommisch H, Jacobs G, et al. A genome-wide association study identifies GLT6D1 as a susceptibility locus for periodontitis. *Hum Mol Genet England.* 2010;19:553–62.
54. Stegeman S, Amankwah E, Klein K, O'Mara TA, Kim D, Lin H-Y, et al. A large-scale analysis of genetic variants within putative miRNA binding sites in prostate cancer. *Cancer Discov.* 2015;5:368–79.
55. Kahali B, Chen Y, Feitosa MF, Bielak LF, O'Connell JR, Musani SK, et al. A noncoding variant near PPP1R3B promotes liver glycogen storage and MetS, but protects against myocardial infarction. *J Clin Endocrinol Metab.* 2021;106:372–87.
56. Yan R, Lai S, Yang Y, Shi H, Cai Z, Sorrentino V, et al. A novel type 2 diabetes risk allele increases the promoter activity of the muscle-specific small ankyrin 1 gene. *Sci Rep.* 2016;6:25105.
57. Rodriguez BAT, Bhan A, Beswick A, Elwood PC, Niiranen TJ, Salomaa V, et al. A platelet function modulator of thrombin activation is causally linked to cardiovascular disease and affects PAR4 receptor signaling. *Am J Hum Genet.* 2020;107:211–21.
58. Hing B, Davidson S, Lear M, Breen G, Quinn J, McGuffin P, et al. A polymorphism associated with depressive disorders differentially regulates brain derived neurotrophic factor promoter IV activity. *Biol Psychiatry.* 2012;71:618–26.
59. Schieck M, Sharma V, Michel S, Toncheva AA, Worth L, Potaczek DP, et al. A polymorphism in the TH2 locus control region is associated with changes in DNA methylation and gene expression. *Allergy Denmark.* 2014;69:1171–80.
60. Huang Q, Whittington T, Gao P, Lindberg JF, Yang Y, Sun J, et al. A prostate cancer susceptibility allele at 6q22 increases RFX6 expression by modulating HOXB13 chromatin binding. *Nat Genet United States.* 2014;46:126–35.
61. Chang J, Tian J, Yang Y, Zhong R, Li J, Zhai K, et al. A Rare Missense variant in TCF7L2 associates with colorectal cancer risk by interacting with a GWAS-identified regulatory variant in the MYC enhancer. *Cancer Res United States.* 2018;78:5164–72.
62. Walavalkar K, Saravanan B, Singh AK, Jayani RS, Nair A, Farooq U, et al. A rare variant of African ancestry activates Bq24 lncRNA hub by modulating cancer associated enhancer. *Nat Commun.* 2020;11:3598.
63. Sinnott-Armstrong N, Sousa IS, Laber S, Rendina-Ruedy E, Nitter Dankel SE, Ferreira T, et al. A regulatory variant at 3q21.1 confers an increased pleiotropic risk for hyperglycemia and altered bone mineral density. *Cell Metab.* 2021;33:615–628.e13.
64. Chinnaswamy S, Chatterjee S, Boopathi R, Mukherjee S, Bhattacharjee S, Kundu TK. A single nucleotide polymorphism associated with hepatitis C virus infections located in the distal region of the IL28B promoter influences NF-κB-mediated gene transcription. *PLoS ONE.* 2013;8:e75495.
65. Lidral AC, Liu H, Bullard SA, Bonde G, Machida J, Visel A, et al. A single nucleotide polymorphism associated with isolated cleft lip and palate, thyroid cancer and hypothyroidism alters the activity of an oral epithelium and thyroid enhancer near FOXE1. *Hum Mol Genet.* 2015;24:3895–907.
66. Dos Santos C, Bougnères P, Fradin D. A single-nucleotide polymorphism in a methylatable Foxa2 binding site of the G6PC2 promoter is associated with insulin secretion in vivo and increased promoter activity in vitro. *Diabetes.* 2009;58:489–92.
67. Roman TS, Cannon ME, Vadlamudi S, Buchkovich ML, Wolford BN, Welch RP, et al. A Type 2 diabetes-associated functional regulatory variant in a pancreatic islet enhancer at the ADCY5 locus. *Diabetes.* 2017;66:2521–30.
68. Hiramoto M, Udagawa H, Ishibashi N, Takahashi E, Kaburagi Y, Miyazawa K, et al. A type 2 diabetes-associated SNP in KCNQ1 (rs163184) modulates the binding activity of the locus for Sp3 and Lsd1/Kdm1a, potentially affecting CDKN1C expression. *Int J Mol Med.* 2018;41:717–28.
69. Justice CM, Kim J, Kim S-D, Kim K, Yagnik G, Cuellar A, et al. A variant associated with sagittal nonsyndromic craniosynostosis alters the

regulatory function of a non-coding element. *Am J Med Genet A*. 2017;173:2893–7.

70. Jee SH, Sull JW, Lee J-E, Shin C, Park J, Kimm H, et al. Adiponectin concentrations: a genome-wide association study. *Am J Hum Genet*. 2010;87:545–52.

71. Verlaan DJ, Berlivet S, Hunninghake GM, Madore A-M, Larivière M, Moussette S, et al. Allele-specific chromatin remodeling in the ZBP2/GSDMB/ORMDL3 locus associated with the risk of asthma and autoimmune disease. *Am J Hum Genet*. 2009;85:377–93.

72. Li X-X, Peng T, Gao J, Feng J-G, Wu D-D, Yang T, et al. Allele-specific expression identified rs2509956 as a novel long-distance cis-regulatory SNP for SCGB1A1, an important gene for multiple pulmonary diseases. *Am J Physiol Lung Cell Mol Physiol*. 2019;317:L456–63.

73. Palstra R-J, de Crignis E, Röling MD, van Staveren T, Kan TW, van Ijcken W, et al. Allele-specific long-distance regulation dictates IL-32 isoform switching and mediates susceptibility to HIV-1. *Sci Adv*. 2018;4:e1701729.

74. Benaglio P, D'Antonio-Chronowska A, Ma W, Yang F, Young Greenwald WW, Donovan MKR, et al. Allele-specific NKX2-5 binding underlies multiple genetic associations with human electrocardiographic traits. *Nat Genet*. 2019;51:1506–17.

75. Lee H, Qian K, von Toerne C, Hoerburger L, Clausnitzer M, Hoffmann C, et al. Allele-specific quantitative proteomics unravels molecular mechanisms modulated by cis-regulatory PPARG locus variation. *Nucleic Acids Res*. 2017;45:3266–79.

76. Redis RS, Vela LE, Lu W, Ferreira de Oliveira J, Ivan C, Rodriguez-Aguayo C, et al. Allele-specific reprogramming of cancer metabolism by the long non-coding RNA CCAT2. *Mol Cell*. 2016;61:520–34.

77. Richards TJ, Park C, Chen Y, Gibson KF, Di Peter Y, Pardo A, et al. Allele-specific transactivation of matrix metalloproteinase 7 by FOXA2 and correlation with plasma levels in idiopathic pulmonary fibrosis. *Am J Physiol Lung Cell Mol Physiol*. 2012;302:L746–754.

78. Fogarty MP, Panhuis TM, Vadlamudi S, Buchkovich ML, Mohlke KL. Allele-specific transcriptional activity at type 2 diabetes-associated single nucleotide polymorphisms in regions of pancreatic islet open chromatin at the JAZF1 locus. *Diabetes*. 2013;62:1756–62.

79. Nakaoka H, Gurumurthy A, Hayano T, Ahmadloo S, Omer WH, Yoshihara K, et al. Allelic imbalance in regulation of ANRIL through chromatin interaction at 9p21 endometriosis risk locus. *PLoS Genet*. 2016;12:e1005893.

80. Pittman AM, Naranjo S, Jalava SE, Twiss P, Ma Y, Olver B, et al. Allelic variation at the 8q23.3 colorectal cancer risk locus functions as a cis-acting regulator of ELF3H. *PLoS Genet*. 2010;6:e1001126.

81. Barrie ES, Lee S-H, Frater JT, Katakai M, Scharre DW, Sadee W. Alpha-synuclein mRNA isoform formation and translation affected by polymorphism in the human SNCA 3'UTR. *Mol Genet Genomic Med*. 2018;6:565–74.

82. Gallego X, Cox RJ, Laughlin JR, Stitzel JA, Ehringer MA. Alternative CHRN4 3'-UTRs mediate the allelic effects of SNP rs1948 on gene expression. *PLoS ONE*. 2013;8:e63699.

83. Wasserman NF, Aneas I, Nobrega MA. An 8q24 gene desert variant associated with prostate cancer risk confers differential in vivo activity to a MYC enhancer. *Genome Res*. 2010;20:1191–7.

84. Thynn HN, Chen X-F, Hu W-X, Duan Y-Y, Zhu D-L, Chen H, et al. An allele-specific functional SNP associated with two systemic autoimmune diseases modulates IRF5 expression by long-range chromatin loop formation. *J Invest Dermatol United States*. 2020;140:348–360.e11.

85. Roberts AR, Vecellio M, Chen L, Ridley A, Cortes A, Knight JC, et al. An ankylosing spondylitis-associated genetic variant in the IL23R-IL12RB2 intergenic region modulates enhancer activity and is associated with increased Th1-cell differentiation. *Ann Rheum Dis*. 2016;75:2150–6.

86. Caussy C, Charrière S, Marçais C, Di Filippo M, Sassolas A, Delay M, et al. An APOA5 3' UTR variant associated with plasma triglycerides triggers APOA5 downregulation by creating a functional miR-485-5p binding site. *Am J Hum Genet*. 2014;94:129–34.

87. Wang S, Wen F, Wiley GB, Kinter MT, Gaffney PM. An enhancer element harboring variants associated with systemic lupus erythematosus engages the TNFAIP3 promoter to influence A20 expression. *PLoS Genet*. 2013;9:e1003750.

88. Yan C, Ji Y, Huang T, Yu F, Gao Y, Gu Y, et al. An esophageal adenocarcinoma susceptibility locus at 9q22 also confers risk to esophageal squamous cell carcinoma by regulating the function of BARX1. *Cancer Lett Ireland*. 2018;421:103–11.

89. Savic D, Bell GI, Nobrega MA. An in vivo cis-regulatory screen at the type 2 diabetes associated TCF7L2 locus identifies multiple tissue-specific enhancers. *PLoS ONE*. 2012;7:e36501.

90. Zhao H, Yang W, Qiu R, Li J, Xin Q, Wang X, et al. An intronic variant associated with systemic lupus erythematosus changes the binding affinity of YinYang1 to downregulate WDFY4. *Genes Immun England*. 2012;13:536–42.

91. Chen X-F, Zhu D-L, Yang M, Hu W-X, Duan Y-Y, Lu B-J, et al. An osteoporosis risk SNP at 1p36.12 acts as an allele-specific enhancer to modulate LINC00339 expression via long-range loop formation. *Am J Hum Genet*. 2018;102:776–93.

92. Liu H, Duncan K, Helverson A, Kumari P, Mumm C, Xiao Y, et al. Analysis of zebrafish periderm enhancers facilitates identification of a regulatory variant near human KRT8/18. *Elife*. 2020;9.

93. Park JH, Chang HS, Park C-S, Jang A-S, Park BL, Rhim TY, et al. Association analysis of CD40 polymorphisms with asthma and the level of serum total IgE. *Am J Respir Crit Care Med*. 2007;175:775–82.

94. Zhao Z, Fan Q, Zhou P, Ye H, Cai L, Lu Y. Association of alpha A-crystallin polymorphisms with susceptibility to nuclear age-related cataract in a Han Chinese population. *BMC Ophthalmol*. 2017;17:133.

95. De T, Alarcon C, Hernandez W, Liko I, Cavallari LH, Duarte JD, et al. Association of genetic variants with warfarin-associated bleeding among patients of African descent. *JAMA*. 2018;320:1670–7.

96. Arora P, Wu C, Khan AM, Bloch DB, Davis-Dusenbery BN, Ghorbani A, et al. Atrial natriuretic peptide is negatively regulated by microRNA-425. *J Clin Invest*. 2013;123:3378–82.

97. Gao P, Xia J-H, Sipeky C, Dong X-M, Zhang Q, Yang Y, et al. Biology and clinical implications of the 19q13 aggressive prostate cancer susceptibility locus. *Cell*. 2018;174:576–589.e18.

98. Bai X, Mangum KD, Dee RA, Stouffer GA, Lee CR, Oni-Orisan A, et al. Blood pressure-associated polymorphism controls ARHGAP42 expression via serum response factor DNA binding. *J Clin Invest*. 2017;127:670–80.

99. de Smith AJ, Walsh KM, Francis SS, Zhang C, Hansen HM, Smirnov I, et al. BMI1 enhancer polymorphism underlies chromosome 10p12.31 association with childhood acute lymphoblastic leukemia. *Int J Cancer*. 2018;143:2647–58.

100. Cowper-Salari R, Zhang X, Wright JB, Bailey SD, Cole MD, Eeckhoutte J, et al. Breast cancer risk-associated SNPs modulate the affinity of chromatin for FOXA1 and alter gene expression. *Nat Genet*. 2012;44:1191–8.

101. Shah MY, Ferracin M, Pilecki V, Chen B, Redis R, Fabris L, et al. Cancer-associated rs6983267 SNP and its accompanying long noncoding RNA CCAT2 induce myeloid malignancies via unique SNP-specific RNA mutations. *Genome Res*. 2018;28:432–47.

102. Glubb DM, Shi W, Beesley J, Fachal L, Pritchard J-L, McCue K, et al. Candidate Causal Variants at the 8p12 Breast Cancer Risk Locus Regulate DUSP4. *Cancers (Basel)*. 2020;12.

103. McGovern A, Schoenfelder S, Martin P, Massey J, Duffus K, Plant D, et al. Capture Hi-C identifies a novel causal gene, IL20RA, in the pan-autoimmune genetic susceptibility region 6q23. *Genome Biol*. 2016;17:212.

104. Ahluwalia TS, Troelsen JT, Balslev-Harder M, Bork-Jensen J, Thuesen BH, Cerqueira C, et al. Carriers of a VEGFA enhancer polymorphism selectively binding CHOP/DDIT3 are predisposed to increased circulating levels of thyroid-stimulating hormone. *J Med Genet England*. 2017;54:166–75.

105. Spisák S, Lawrenson K, Fu Y, Csabai I, Cottman RT, Seo J-H, et al. CAUSEL: an epigenome- and genome-editing pipeline for establishing function of noncoding GWAS variants. *Nat Med*. 2015;21:1357–63.

106. Mehta ZB, Fine N, Pullen TJ, Cane MC, Hu M, Chabosseau P, et al. Changes in the expression of the type 2 diabetes-associated gene VPS13C in the β -cell are associated with glucose intolerance in humans and mice. *Am J Physiol Endocrinol Metab*. 2016;311:E488–507.

107. Prokop JW, Yeo NC, Ottmann C, Chhetri SB, Florus KL, Ross EJ, et al. Characterization of coding/noncoding variants for SHROOM3 in patients with CKD. *J Am Soc Nephrol*. 2018;29:1525–35.

108. Xia Q, Deliard S, Yuan C-X, Johnson ME, Grant SFA. Characterization of the transcriptional machinery bound across the widely presumed

- type 2 diabetes causal variant, rs7903146, within TCF7L2. *Eur J Hum Genet.* 2015;23:103–9.
109. Comiskey DJF, He H, Liyanarachchi S, Sheikh MS, Hendrickson IV, Yu L, et al. Characterizing the function of EPB41L4A in the predisposition to papillary thyroid carcinoma. *Sci Rep.* 2020;10:19984.
 110. Du M, Tillmans L, Gao J, Gao P, Yuan T, Dittmar RL, et al. Chromatin interactions and candidate genes at ten prostate cancer risk loci. *Sci Rep.* 2016;6:23202.
 111. Matoba N, Liang D, Sun H, Aygün N, McAfee JC, Davis JE, et al. Common genetic risk variants identified in the SPARK cohort support DDHD2 as a candidate risk gene for autism. *Transl Psychiatry.* 2020;10:265.
 112. Hiramoto M, Udagawa H, Watanabe A, Miyazawa K, Ishibashi N, Kawaguchi M, et al. Comparative analysis of type 2 diabetes-associated SNP alleles identifies allele-specific DNA-binding proteins for the KCNQ1 locus. *Int J Mol Med Greece.* 2015;36:222–30.
 113. Hazelett DJ, Rhie SK, Gaddis M, Yan C, Lakeland DL, Coetzee SG, et al. Comprehensive functional annotation of 77 prostate cancer risk loci. *PLoS Genet.* 2014;10:e1004102.
 114. Cheng M, Huang X, Zhang M, Huang Q. Computational and functional analyses of T2D GWAS SNPs for transcription factor binding. *Biochem Biophys Res Commun United States.* 2020;523:658–65.
 115. Ye W, Wang Y, Mei B, Hou S, Liu X, Wu G, et al. Computational and functional characterization of four SNPs in the SOST locus associated with osteoporosis. *Bone United States.* 2018;108:132–44.
 116. Clifton-Bligh RJ, Nguyen TV, Au A, Bullock M, Cameron I, Cumming R, et al. Contribution of a common variant in the promoter of the 1- α -hydroxylase gene (CYP27B1) to fracture risk in the elderly. *Calcif Tissue Int.* 2011;88:109–16.
 117. Miller CL, Haas U, Diaz R, Leeper NJ, Kundu RK, Patlolla B, et al. Coronary heart disease-associated variation in TCF21 disrupts a miR-224 binding site and miRNA-mediated regulation. *PLoS Genet.* 2014;10:e1004263.
 118. Gee F, Rushton MD, Loughlin J, Reynard LN. Correlation of the osteoarthritis susceptibility variants that map to chromosome 20q13 with an expression quantitative trait locus operating on NCOA3 and with functional variation at the polymorphism rs116855380. *Arthritis Rheumatol.* 2015;67:2923–32.
 119. Reynard LN, Bui C, Syddall CM, Loughlin J. CpG methylation regulates allelic expression of GDF5 by modulating binding of SP1 and SP3 repressor proteins to the osteoarthritis susceptibility SNP rs143383. *Hum Genet.* 2014;133:1059–73.
 120. Wu J, Yang S, Yu D, Gao W, Liu X, Zhang K, et al. CRISPR/cas9 mediated knockout of an intergenic variant rs6927172 identified IL-20RA as a new risk gene for multiple autoimmune diseases. *Genes Immunol England.* 2019;20:103–11.
 121. Deng Y, Zhao J, Sakurai D, Sestak AL, Osadchiv V, Langefeld CD, et al. Decreased SMG7 expression associates with lupus-risk variants and elevated antinuclear antibody production. *Ann Rheum Dis.* 2016;75:2007–13.
 122. Vezzoli G, Terranegra A, Aloia A, Arcidiacono T, Milanese L, Mosca E, et al. Decreased transcriptional activity of calcium-sensing receptor gene promoter 1 is associated with calcium nephrolithiasis. *J Clin Endocrinol Metab.* 2013;98:3839–47.
 123. Ryu J, Lee C. Differential promoter activity by nucleotide substitution at a type 2 diabetes genome-wide association study signal upstream of the wolframin gene. *J Diabetes Australia.* 2016;8:253–9.
 124. Smith JG, Felix JF, Morrison AC, Kalogeropoulos A, Trompet S, Wilk JB, et al. Discovery of genetic variation on chromosome 5q22 associated with mortality in heart failure. *PLoS Genet.* 2016;12:e1006034.
 125. Miller CL, Anderson DR, Kundu RK, Raiesdana A, Nürnberg ST, Diaz R, et al. Disease-related growth factor and embryonic signaling pathways modulate an enhancer of TCF21 expression at the 6q23.2 coronary heart disease locus. *PLoS Genet.* 2013;9:e1003652.
 126. Rahimov F, Marazita ML, Visel A, Cooper ME, Hitchler MJ, Rubini M, et al. Disruption of an AP-2alpha binding site in an IRF6 enhancer is associated with cleft lip. *Nat Genet.* 2008;40:1341–7.
 127. Zhu Z, Meng W, Liu P, Zhu X, Liu Y, Zou H. DNA hypomethylation of a transcription factor binding site within the promoter of a gout risk gene NRBP1 upregulates its expression by inhibition of TFAP2A binding. *Clin Epigenetics.* 2017;9:99.
 128. Wang X, Srivastava Y, Jankowski A, Malik V, Wei Y, Del Rosario RC, et al. DNA-mediated dimerization on a compact sequence signature controls enhancer engagement and regulation by FOXA1. *Nucleic Acids Res.* 2018;46:5470–86.
 129. Kim BS, Park S-M, Uhm TG, Kang JH, Park J-S, Jang A-S, et al. Effect of single nucleotide polymorphisms within the interleukin-4 promoter on aspirin intolerance in asthmatics and interleukin-4 promoter activity. *Pharmacogenet Genomics United States.* 2010;20:748–58.
 130. Powell JE, Fung JN, Shakhbazov K, Sapkota Y, Cloonan N, Hemani G, et al. Endometriosis risk alleles at 1p36.12 act through inverse regulation of CDC42 and LINC00339. *Hum Mol Genet.* 2016;25:5046–58.
 131. Gant VU, Junco JJ, Terrell M, Rashid R, Rabin KR. Enhancer polymorphisms at the IKZF1 susceptibility locus for acute lymphoblastic leukemia impact B-cell proliferation and differentiation in both Down syndrome and non-Down syndrome genetic backgrounds. *PLoS ONE.* 2021;16:e0244863.
 132. Sio YY, Matta SA, Ng YT, Chew FT. Epistasis between phenylethanolamine N-methyltransferase and β 2-adrenergic receptor influences extracellular epinephrine level and associates with the susceptibility to allergic asthma. *Clin Exp Allergy England.* 2020;50:352–63.
 133. Vecellio M, Cortes A, Roberts AR, Ellis J, Cohen CJ, Knight JC, et al. Evidence for a second ankylosing spondylitis-associated RUNX3 regulatory polymorphism. *RMD Open.* 2018;4:e000628.
 134. Ghousaini M, Edwards SL, Michailidou K, Nord S, Cowper-Sal Lari R, Desai K, et al. Evidence that breast cancer risk at the 2q35 locus is mediated through IGFBP5 regulation. *Nat Commun.* 2014;4:4999.
 135. Shepherd C, Skelton AJ, Rushton MD, Reynard LN, Loughlin J. Expression analysis of the osteoarthritis genetic susceptibility locus mapping to an intron of the MCF2L gene and marked by the polymorphism rs11842874. *BMC Med Genet.* 2015;16:108.
 136. Surgucheva I, Surguchov A. Expression of caveolin in trabecular meshwork cells and its possible implication in pathogenesis of primary open angle glaucoma. *Mol Vis.* 2011;17:2878–88.
 137. Lou H, Yeager M, Li H, Bosquet JG, Hayes RB, Orr N, et al. Fine mapping and functional analysis of a common variant in MSMB on chromosome 10q11.2 associated with prostate cancer susceptibility. *Proc Natl Acad Sci USA.* 2009;106:7933–8.
 138. Chang B-L, Cramer SD, Wiklund F, Isaacs SD, Stevens VL, Sun J, et al. Fine mapping association study and functional analysis implicate a SNP in MSMB at 10q11 as a causal variant for prostate cancer risk. *Hum Mol Genet.* 2009;18:1368–75.
 139. Westra H-J, Martínez-Bonet M, Onengut-Gumuscu S, Lee A, Luo Y, Teslovich N, et al. Fine-mapping and functional studies highlight potential causal variants for rheumatoid arthritis and type 1 diabetes. *Nat Genet.* 2018;50:1366–74.
 140. Orr N, Dudbridge F, Dryden N, Maguire S, Novo D, Perrakis E, et al. Fine-mapping identifies two additional breast cancer susceptibility loci at 9q31.2. *Hum Mol Genet.* 2015;24:2966–84.
 141. Painter JN, O'Mara TA, Batra J, Cheng T, Lose FA, Dennis J, et al. Fine-mapping of the HNF1B multicancer locus identifies candidate variants that mediate endometrial cancer risk. *Hum Mol Genet.* 2015;24:1478–92.
 142. Pan Y, Tian R, Lee C, Bao G, Gibson G. Fine-mapping within eQTL credible intervals by expression CROP-seq. *Biol Methods Protoc.* 2020;5:bpaa008.
 143. Glubb DM, Maranian MJ, Michailidou K, Pooley KA, Meyer KB, Kar S, et al. Fine-scale mapping of the 5q11.2 breast cancer locus reveals at least three independent risk variants regulating MAP3K1. *Am J Hum Genet.* 2015;96:5–20.
 144. Meyer KB, O'Reilly M, Michailidou K, Carlebur S, Edwards SL, French JD, et al. Fine-scale mapping of the FGFR2 breast cancer risk locus: putative functional variants differentially bind FOXA1 and E2F1. *Am J Hum Genet.* 2013;93:1046–60.
 145. Cheng TH, Thompson DJ, O'Mara TA, Painter JN, Glubb DM, Flach S, et al. Five endometrial cancer risk loci identified through genome-wide association analysis. *Nat Genet.* 2016;48:667–74.
 146. Bohaczk SC, Thackray VG, Shen J, Skowronska-Krawczyk D, Mellon PL. FSHB Transcription is Regulated by a Novel 5' Distal Enhancer With a Fertility-Associated Single Nucleotide Polymorphism. *Endocrinology.* 2021;162.

147. Claussnitzer M, Dankel SN, Kim K-H, Quon G, Meuleman W, Haugen C, et al. FTO obesity variant circuitry and adipocyte browning in humans. *N Engl J Med*. 2015;373:895–907.
148. Buckley MA, Woods NT, Tyrer JP, Mendoza-Fandiño G, Lawrenson K, Hazelett DJ, et al. Functional analysis and fine mapping of the 9p22.2 ovarian cancer susceptibility locus. *Cancer Res*. 2019;79:467–81.
149. Boardman-Pretty F, Smith AJP, Cooper J, Palmen J, Folkersen L, Hamsten A, et al. Functional analysis of a carotid intima-media thickness locus implicates BCAR1 and suggests a causal variant. *Circ Cardiovasc Genet United States*. 2015;8:696–706.
150. Turner AW, Martinuk A, Silva A, Lau P, Nikpay M, Eriksson P, et al. Functional analysis of a novel genome-wide association study signal in SMAD3 that confers protection from coronary artery disease. *Arterioscler Thromb Vasc Biol United States*. 2016;36:972–83.
151. Hamdi Y, Leclerc M, Dumont M, Dubois S, Tranchant M, Reimnitz G, et al. Functional analysis of promoter variants in genes involved in sex steroid action, DNA repair and cell cycle control. *Genes (Basel)*. 2019;10.
152. Pang DX, Smith AJP, Humphries SE. Functional analysis of TCF7L2 genetic variants associated with type 2 diabetes. *Nutr Metab Cardiovasc Dis*. 2013;23:550–6.
153. Baskin R, Woods NT, Mendoza-Fandiño G, Forsyth P, Egan KM, Monteiro ANA. Functional analysis of the 11q23.3 glioma susceptibility locus implicates PHLDB1 and DDX6 in glioma susceptibility. *Sci Rep*. 2015;5:17367.
154. Egli RJ, Southam L, Wilkins JM, Lorenzen I, Pombo-Suarez M, Gonzalez A, et al. Functional analysis of the osteoarthritis susceptibility-associated GDF5 regulatory polymorphism. *Arthritis Rheum*. 2009;60:2055–64.
155. Douvris A, Soubeyrand S, Naing T, Martinuk A, Nikpay M, Williams A, et al. Functional analysis of the TRIB1 associated locus linked to plasma triglycerides and coronary artery disease. *J Am Heart Assoc*. 2014;3:e000884.
156. Zhang Y, Kuipers AL, Yerges-Armstrong LM, Nestlerode CS, Jin Z, Wheeler VW, et al. Functional and association analysis of frizzled 1 (FZD1) promoter haplotypes with femoral neck geometry. *Bone*. 2010;46:1131–7.
157. Fang J, Jia J, Makowski M, Xu M, Wang Z, Zhang T, et al. Functional characterization of a multi-cancer risk locus on chr5p15.33 reveals regulation of TERT by ZNF148. *Nat Commun*. 2017;8:15034.
158. Eckart N, Song Q, Yang R, Wang R, Zhu H, McCallion AS, et al. Functional characterization of schizophrenia-associated variation in CACNA1C. *PLoS ONE*. 2016;11:e0157086.
159. Flora AV, Zambrano CA, Gallego X, Miyamoto JH, Johnson KA, Cowan KA, et al. Functional characterization of SNPs in CHRNA3/B4 intergenic region associated with drug behaviors. *Brain Res*. 2013;1529:1–15.
160. Bigot P, Colli LM, Machiela MJ, Jessop L, Myers TA, Carrouget J, et al. Functional characterization of the 12p12.1 renal cancer-susceptibility locus implicates BHLHE41. *Nat Commun*. 2016;7:12098.
161. Roca-Ayats N, Martínez-Gil N, Cozar M, Gerousi M, Garcia-Giral N, Ovejero D, et al. Functional characterization of the C7ORF76 genomic region, a prominent GWAS signal for osteoporosis in 7q21.3. *Bone*. 2019;123:39–47.
162. Kessler T, Wobst J, Wolf B, Eckhold J, Vilne B, Hollstein R, et al. Functional characterization of the GUCY1A3 coronary artery disease risk locus. *Circulation*. 2017;136:476–89.
163. Maloney B, Ge Y-W, Petersen RC, Hardy J, Rogers JT, Pérez-Tur J, et al. Functional characterization of three single-nucleotide polymorphisms present in the human APOE promoter sequence: Differential effects in neuronal cells and on DNA-protein interactions. *Am J Med Genet B Neuropsychiatr Genet*. 2010;153B:185–201.
164. Helbig S, Wockner L, Bouendou A, Hille-Betz U, McCue K, French JD, et al. Functional dissection of breast cancer risk-associated TERT promoter variants. *Oncotarget*. 2017;8:67203–17.
165. Ge M, Shi M, An C, Yang W, Nie X, Zhang J, et al. Functional evaluation of TERT-CLPTM1L genetic variants associated with susceptibility of papillary thyroid carcinoma. *Sci Rep*. 2016;6:26037.
166. Elsby LM, Orozco G, Denton J, Worthington J, Ray DW, Donn RP. Functional evaluation of TNFAIP3 (A20) in rheumatoid arthritis. *Clin Exp Rheumatol*. 2010;28:708–14.
167. Vecellio M, Chen L, Cohen CJ, Cortes A, Li Y, Bonham S, et al. Functional genomic analysis of a RUNX3 polymorphism associated with ankylosing spondylitis. *Arthritis Rheumatol United States*. 2021;73:980–90.
168. Chang H, Cai X, Li H-J, Liu W-P, Zhao L-J, Zhang C-Y, et al. Functional genomics identify a regulatory risk variation rs4420550 in the 16p11.2 Schizophrenia-Associated Locus. *Biol Psychiatry United States*. 2021;89:246–55.
169. Guo L, Yamashita H, Kou I, Takimoto A, Meguro-Horike M, Horike S, et al. Functional Investigation of a Non-coding Variant Associated with Adolescent Idiopathic Scoliosis in Zebrafish: Elevated Expression of the Ladybird Homeobox Gene Causes Body Axis Deformation. *PLoS Genet*. 2016;12:e1005802.
170. Kong M, Kim Y, Lee C. Functional investigation of a venous thromboembolism GWAS signal in a promoter region of coagulation factor XI gene. *Mol Biol Rep Netherlands*. 2014;41:2015–9.
171. Lawrenson K, Kar S, McCue K, Kuchenbaecker K, Michailidou K, Tyrer J, et al. Functional mechanisms underlying pleiotropic risk alleles at the 19p13.1 breast-ovarian cancer susceptibility locus. *Nat Commun*. 2016;7:12675.
172. Pérez-Razo JC, Cano-Martínez LJ, Vargas Alarcón G, Canizales-Quinteros S, Martínez-Rodríguez N, Canto P, et al. Functional polymorphism rs13306560 of the MTHFR gene is associated with essential hypertension in a Mexican-Mestizo Population. *Circ Cardiovasc Genet United States*. 2015;8:603–9.
173. Nanda V, Wang T, Pjanic M, Liu B, Nguyen T, Matic LP, et al. Functional regulatory mechanism of smooth muscle cell-restricted LMOD1 coronary artery disease locus. *PLoS Genet*. 2018;14:e1007755.
174. Huang X, Zheng J, Li J, Che X, Tan W, Tan W, et al. Functional role of BTB and CNC Homology 1 gene in pancreatic cancer and its association with survival in patients treated with gemcitabine. *Theranostics*. 2018;8:3366–79.
175. Ustiugova AS, Korneev KV, Kuprash DV, Afanasyeva AMA. Functional SNPs in the Human Autoimmunity-Associated Locus 17q12–21. *Genes (Basel)*. 2019;10.
176. Klein JC, Keith A, Rice SJ, Shepherd C, Agarwal V, Loughlin J, et al. Functional testing of thousands of osteoarthritis-associated variants for regulatory activity. *Nat Commun*. 2019;10:2434.
177. Yu W, Zhang K, Wang Z, Zhang J, Chen T, Jin L. Functional variant in the promoter region of IL-27 alters gene transcription and confers a risk for ulcerative colitis in northern Chinese Han. *Hum Immunol United States*. 2017;78:287–93.
178. French JD, Ghossaini M, Edwards SL, Meyer KB, Michailidou K, Ahmed S, et al. Functional variants at the 11q13 risk locus for breast cancer regulate cyclin D1 expression through long-range enhancers. *Am J Hum Genet*. 2013;92:489–503.
179. Andiappan AK, Sio YY, Lee B, Suri BK, Matta SA, Lum J, et al. Functional variants of 17q12-21 are associated with allergic asthma but not allergic rhinitis. *J Allergy Clin Immunol United States*. 2016;137:758–766.e3.
180. Li Y, Nie Y, Cao J, Tu S, Lin Y, Du Y, et al. G-A variant in miR-200c binding site of EFNA1 alters susceptibility to gastric cancer. *Mol Carcinog United States*. 2014;53:219–29.
181. Gaulton KJ, Ferreira T, Lee Y, Raimondo A, Mägi R, Reschen ME, et al. Genetic fine mapping and genomic annotation defines causal mechanisms at type 2 diabetes susceptibility loci. *Nat Genet*. 2015;47:1415–25.
182. Liu S, Wu N, Zuo Y, Zhou Y, Liu J, Liu Z, et al. Genetic Polymorphism of LBX1 Is Associated With Adolescent Idiopathic Scoliosis in Northern Chinese Han Population. *Spine (Phila Pa 1976)*. United States; 2017;42:1125–9.
183. Oldridge DA, Wood AC, Weichert-Leahey N, Crimmins I, Sussman R, Winter C, et al. Genetic predisposition to neuroblastoma mediated by a LMO1 super-enhancer polymorphism. *Nature*. 2015;528:418–21.
184. Cavalli M, Pan G, Nord H, Wallén Artzt E, Wallerman O, Wadelius C. Genetic prevention of hepatitis C virus-induced liver fibrosis by allele-specific downregulation of MERTK. *Hepatol Res Netherlands*. 2017;47:826–30.
185. Krause MD, Huang R-T, Wu D, Shentu T-P, Harrison DL, Whalen MB, et al. Genetic variant at coronary artery disease and ischemic stroke locus 1p32.2 regulates endothelial responses to hemodynamics. *Proc Natl Acad Sci USA*. 2018;115:E11349–58.
186. Soderquest K, Hertweck A, Giambartolomei C, Henderson S, Mohamed R, Goldberg R, et al. Genetic variants alter Tbet binding

and gene expression in mucosal inflammatory disease. *PLoS Genet.* 2017;13:e1006587.

187. Wu C, Hu Z, Yu D, Huang L, Jin G, Liang J, et al. Genetic variants on chromosome 15q25 associated with lung cancer risk in Chinese populations. *Cancer Res United States.* 2009;69:5065–72.
188. Bernstein DI, Lummus ZL, Kesavalu B, Yao J, Kottyan L, Miller D, et al. Genetic variants with gene regulatory effects are associated with diisocyanate-induced asthma. *J Allergy Clin Immunol United States.* 2018;142:959–69.
189. Bamji-Mirza M, Li Y, Najem D, Liu QY, Walker D, Lue L-F, et al. Genetic Variations in ABCA7 Can Increase Secreted Levels of Amyloid-β40 and Amyloid-β42 Peptides and ABCA7 Transcription in Cell Culture Models. *J Alzheimers Dis Netherlands.* 2016;53:875–92.
190. Keller M, Gebhardt C, Huth S, Schleinitz D, Heyne H, Scholz M, et al. Genetically programmed changes in transcription of the novel progranulin regulator. *J Mol Med (Berl).* 2020;98:1139–48.
191. Hou S, Du L, Lei B, Pang CP, Zhang M, Zhuang W, et al. Genome-wide association analysis of Vogt-Koyanagi-Harada syndrome identifies two new susceptibility loci at 1p31.2 and 10q21.3. *Nat Genet.* 2014;46:1007–11.
192. Kawamura R, Tabara Y, Tsukada A, Igase M, Ohashi J, Yamada R, et al. Genome-wide association study of plasma resistin levels identified rs1423096 and rs10401670 as possible functional variants in the Japanese population. *Physiol Genomics United States.* 2016;48:874–81.
193. Stitzel ML, Sethupathy P, Pearson DS, Chines PS, Song L, Erdos MR, et al. Global epigenomic analysis of primary human pancreatic islets provides insights into type 2 diabetes susceptibility loci. *Cell Metab.* 2010;12:443–55.
194. Kalita CA, Brown CD, Freiman A, Isherwood J, Wen X, Pique-Regi R, et al. High-throughput characterization of genetic effects on DNA-protein binding and gene transcription. *Genome Res.* 2018;28:1701–8.
195. Zhou Y, Oskolkov N, Shcherbina L, Ratti J, Kock K-H, Su J, et al. HMGB1 binds to the rs7903146 locus in TCF7L2 in human pancreatic islets. *Mol Cell Endocrinol Ireland.* 2016;430:138–45.
196. Ross-Adams H, Ball S, Lawrenson K, Halim S, Russell R, Wells C, et al. HNF1B variants associate with promoter methylation and regulate gene networks activated in prostate and ovarian cancer. *Oncotarget.* 2016;7:74734–46.
197. Smith EN, D'Antonio-Chronowska A, Greenwald WW, Borja V, Aguiar LR, Pogue R, et al. Human iPSC-derived retinal pigment epithelium: a model system for prioritizing and functionally characterizing causal variants at AMD risk loci. *Stem Cell Reports.* 2019;12:1342–53.
198. Hitomi Y, Kawashima M, Aiba Y, Nishida N, Matsushashi M, Okazaki H, et al. Human primary biliary cirrhosis-susceptible allele of rs4979462 enhances TNFSF15 expression by binding NF-1. *Hum Genet Germany.* 2015;134:737–47.
199. López Rodríguez M, Kaminska D, Lappalainen K, Pihlajamäki J, Kaikonen MU, Laakso M. Identification and characterization of a FOXA2-regulated transcriptional enhancer at a type 2 diabetes intronic locus that controls GSKR expression in liver cells. *Genome Med.* 2017;9:63.
200. Biancolella M, Fortini BK, Tring S, Plummer SJ, Mendoza-Fandino GA, Hartiala J, et al. Identification and characterization of functional risk variants for colorectal cancer mapping to chromosome 11q23.1. *Hum Mol Genet.* 2014;23:2198–209.
201. Flachsbart F, Dose J, Gentschew L, Geismann C, Caliebe A, Knecht C, et al. Identification and characterization of two functional variants in the human longevity gene FOXO3. *Nat Commun.* 2017;8:2063.
202. Spracklen CN, Shi J, Vadlamudi S, Wu Y, Zou M, Raulerson CK, et al. Identification and functional analysis of glycemic trait loci in the China Health and Nutrition Survey. *PLoS Genet.* 2018;14:e1007275.
203. Liu L, Pei Y-F, Liu T-L, Hu W-Z, Yang X-L, Li S-C, et al. Identification of a 1p21 independent functional variant for abdominal obesity. *Int J Obes (Lond).* 2019;43:2480–90.
204. Zhou X, Baron RM, Hardin M, Cho MH, Zielinski J, Hawrylkiewicz I, et al. Identification of a chronic obstructive pulmonary disease genetic determinant that regulates HHIP. *Hum Mol Genet.* 2012;21:1325–35.
205. Boulling A, Masson E, Zou W-B, Paliwal S, Wu H, Issarapu P, et al. Identification of a functional enhancer variant within the chronic pancreatitis-associated SPINK1 c.101A>G (p.Asn34Ser)-containing haplotype. *Hum Mutat.* 2017;38:1014–24.
206. Ke J, Tian J, Li J, Gong Y, Yang Y, Zhu Y, et al. Identification of a functional polymorphism affecting microRNA binding in the susceptibility locus 1q25.3 for colorectal cancer. *Mol Carcinog.* 2017;56:2014–21.
207. Alcina A, Fedetz M, Fernández O, Saiz A, Izquierdo G, Lucas M, et al. Identification of a functional variant in the KIF5A-CYP27B1-METTL1-FAM119B locus associated with multiple sclerosis. *J Med Genet.* 2013;50:25–33.
208. Lo PHY, Urabe Y, Kumar V, Tanikawa C, Koike K, Kato N, et al. Identification of a functional variant in the MICA promoter which regulates MICA expression and increases HCV-related hepatocellular carcinoma risk. *PLoS ONE.* 2013;8:e61279.
209. Ke J, Lou J, Chen X, Li J, Liu C, Gong Y, et al. Identification of a potential regulatory variant for colorectal cancer risk mapping to chromosome 5q31.1: A Post-GWAS Study. *PLoS ONE.* 2015;10:e0138478.
210. Fogarty MP, Cannon ME, Vadlamudi S, Gaulton KJ, Mohlke KL. Identification of a regulatory variant that binds FOXA1 and FOXA2 at the CDC123/CAMK1D type 2 diabetes GWAS locus. *PLoS Genet.* 2014;10:e1004633.
211. Parker MM, Hao Y, Guo F, Pham B, Chase R, Platig J, et al. Identification of an emphysema-associated genetic variant near TGFβ2 with regulatory effects in lung fibroblasts. *Elife.* 2019;8.
212. Ryo H, Kong M, Kim Y, Lee C. Identification of functional nucleotide and haplotype variants in the promoter of the CEBPE gene. *J Hum Genet England.* 2013;58:600–3.
213. van Ouwkerk AF, Bosada FM, Liu J, Zhang J, van Duijvenboden K, Chaffin M, et al. Identification of functional variant enhancers associated with atrial fibrillation. *Circ Res United States.* 2020;127:229–43.
214. Castaldi PJ, Guo F, Qiao D, Du F, Naing ZC, Li Y, et al. Identification of functional variants in the FAM13A chronic obstructive pulmonary disease genome-wide Association Study Locus by Massively Parallel Reporter Assays. *Am J Respir Crit Care Med.* 2019;199:52–61.
215. Bai W-Y, Wang L, Ying Z-M, Hu B, Xu L, Zhang G-Q, et al. Identification of PIEZO1 polymorphisms for human bone mineral density. *Bone.* 2020;133:115247.
216. Fairouz RH, White J, Palmen J, Kalea AZ, Humphries SE. Identification of the functional variant(s) that explain the low-density lipoprotein receptor (LDLR) GWAS SNP rs6511720 association with lower LDL-C and risk of CHD. *PLoS ONE.* 2016;11:e0167676.
217. Guo X, Lin W, Wen W, Huyghe J, Bien S, Cai Q, et al. Identifying novel susceptibility genes for colorectal cancer risk from a transcriptome-wide association study of 125,478 subjects. *Gastroenterology.* 2021;160:1164–1178.e6.
218. Amlie-Wolf A, Tang M, Way J, Dombroski B, Jiang M, Vrettos N, et al. Inferring the molecular mechanisms of noncoding Alzheimer's disease-associated genetic variants. *J Alzheimers Dis.* 2019;72:301–18.
219. Hamadou I, Garritano S, Romanel A, Naimi D, Hammada T, Demichelis F. Inherited variant in NFκB-1 promoter is associated with increased risk of IBD in an Algerian population and modulates SOX9 binding. *Cancer Rep (Hoboken).* 2020;3:e1240.
220. Pan DZ, Garske KM, Alvarez M, Bhagat YV, Boockchay J, Nikkila E, et al. Integration of human adipocyte chromosomal interactions with adipose gene expression prioritizes obesity-related genes from GWAS. *Nat Commun.* 2018;9:1512.
221. Zhang X, Cowper-Salari R, Bailey SD, Moore JH, Lupien M. Integrative functional genomics identifies an enhancer looping to the SOX9 gene disrupted by the 17q24.3 prostate cancer risk locus. *Genome Res.* 2012;22:1437–46.
222. Miller CL, Pjanic M, Wang T, Nguyen T, Cohain A, Lee JD, et al. Integrative functional genomics identifies regulatory mechanisms at coronary artery disease loci. *Nat Commun.* 2016;7:12092.
223. Zhang Y, Manjunath M, Zhang S, Chasman D, Roy S, Song JS. Integrative genomic analysis predicts causative cis-regulatory mechanisms of the Breast Cancer-Associated Genetic Variant rs4415084. *Cancer Res.* 2018;78:1579–91.
224. Berlivet S, Moussette S, Ouimet M, Verlaan DJ, Koka V, Al Tuwaijri A, et al. Interaction between genetic and epigenetic variation defines gene expression patterns at the asthma-associated locus 17q12-q21 in lymphoblastoid cell lines. *Hum Genet.* 2012;131:1161–71.
225. Wang X, Raghavan A, Peters DT, Pashos EE, Rader DJ, Musunuru K. Interrogation of the atherosclerosis-associated SORT1 (Sortilin 1) locus with primary human hepatocytes, induced pluripotent stem

- cell-hepatocytes, and locus-humanized mice. *Arterioscler Thromb Vasc Biol.* 2018;38:76–82.
226. Hammaker D, Whitaker JW, Maeshima K, Boyle DL, Ekwall A-KH, Wang W, et al. LBH Gene Transcription Regulation by the Interplay of an Enhancer Risk Allele and DNA Methylation in Rheumatoid Arthritis. *Arthritis Rheumatol.* 2016;68:2637–45.
 227. Reschen ME, Gaulton KJ, Lin D, Soilleux EJ, Morris AJ, Smyth SS, et al. Lipid-induced epigenomic changes in human macrophages identify a coronary artery disease-associated variant that regulates PPAP2B Expression through Altered C/EBP-beta binding. *PLoS Genet.* 2015;11:e1005061.
 228. Zhang Y, Chen X-F, Li J, He F, Li X, Guo Y. lncRNA Neat1 stimulates osteoclastogenesis via sponging miR-7. *J Bone Miner Res.* 2020;35:1772–81.
 229. Mei B, Wang Y, Ye W, Huang H, Zhou Q, Chen Y, et al. lncRNA ZBTB40-IT1 modulated by osteoporosis GWAS risk SNPs suppresses osteogenesis. *Hum Genet.* 2019;138:151–66.
 230. Vicente CT, Edwards SL, Hillman KM, Kaufmann S, Mitchell H, Bain L, et al. Long-range modulation of PAG1 expression by 8q21 allergy risk variants. *Am J Hum Genet.* 2015;97:329–36.
 231. Cavalli M, Pan G, Nord H, Wadelius C. Looking beyond GWAS: allele-specific transcription factor binding drives the association of GALNT2 to HDL-C plasma levels. *Lipids Health Dis.* 2016;15:18.
 232. Lu X, Zoller EE, Weirauch MT, Wu Z, Namjou B, Williams AH, et al. Lupus risk variant increases pSTAT1 binding and decreases ETS1 expression. *Am J Hum Genet.* 2015;96:731–9.
 233. Choi J, Zhang T, Vu A, Ablain J, Makowski MM, Colli LM, et al. Massively parallel reporter assays of melanoma risk variants identify MX2 as a gene promoting melanoma. *Nat Commun.* 2020;11:2718.
 234. Elek Z, Németh N, Nagy G, Németh H, Somogyi A, Hosszufalusi N, et al. Micro-RNA binding site polymorphisms in the WFS1 gene are risk factors of diabetes mellitus. *PLoS ONE.* 2015;10:e0139519.
 235. Rong H, Gu S, Zhang G, Kang L, Yang M, Zhang J, et al. MiR-2964a-5p binding site SNP regulates ATM expression contributing to age-related cataract risk. *Oncotarget.* 2017;8:84945–57.
 236. Elek Z, Dénes R, Prokop S, Somogyi A, Yowanto H, Luo J, et al. Multicapillary gel electrophoresis based analysis of genetic variants in the WFS1 gene. *Electrophoresis.* 2016;37:2313–21.
 237. Zhu D-L, Chen X-F, Hu W-X, Dong S-S, Lu B-J, Rong Y, et al. Multiple functional variants at 13q14 risk locus for osteoporosis regulate RANKL expression through long-range super-enhancer. *J Bone Miner Res.* 2018;33:1335–46.
 238. He H, Li W, Liyanarachchi S, Srinivas M, Wang Y, Akagi K, et al. Multiple functional variants in long-range enhancer elements contribute to the risk of SNP rs965513 in thyroid cancer. *Proc Natl Acad Sci U S A.* 2015;112:6128–33.
 239. Roman TS, Marvelle AF, Fogarty MP, Vadlamudi S, Gonzalez AJ, Buchkovich ML, et al. Multiple hepatic regulatory variants at the GALNT2 GWAS locus associated with high-density lipoprotein cholesterol. *Am J Hum Genet.* 2015;97:801–15.
 240. Bojesen SE, Pooley KA, Johnatty SE, Beesley J, Michailidou K, Tyrer JP, et al. Multiple independent variants at the TERT locus are associated with telomere length and risks of breast and ovarian cancer. *Nat Genet.* 2013;45:371–84, 384e1–2.
 241. Beaudoin M, Gupta RM, Won H-H, Lo KS, Do R, Henderson CA, et al. Myocardial infarction-associated SNP at 6p24 interferes with MEF2 binding and associates with PHACTR1 expression levels in human coronary arteries. *Arterioscler Thromb Vasc Biol.* 2015;35:1472–9.
 242. John G, Hegarty JP, Yu W, Berg A, Pastor DM, Kelly AA, et al. NKK2-3 variant rs11190140 is associated with IBD and alters binding of NFAT. *Mol Genet Metab United States.* 2011;104:174–9.
 243. Bailey SD, Desai K, Kron KJ, Mazrooei P, Sinnott-Armstrong NA, Treloar AE, et al. Noncoding somatic and inherited single-nucleotide variants converge to promote ESR1 expression in breast cancer. *Nat Genet.* 2016;48:1260–6.
 244. Gorbatenko A, Olesen CW, Loebl N, Sigurdsson HH, Bianchi C, Pedraz-Cuesta E, et al. Oncogenic p95HER2 regulates Na⁺-HCO₃⁻ cotransporter NBCn1 mRNA stability in breast cancer cells via 3'UTR-dependent processes. *Biochem J England.* 2016;473:4027–44.
 245. Wang Y, Ye W, Liu Y, Mei B, Liu X, Huang Q. Osteoporosis genome-wide association study variant c.3781 C>A is regulated by a novel anti-osteogenic factor miR-345–5p. *Hum Mutat.* 2020;41:709–18.
 246. Zheng J, Huang X, Tan W, Yu D, Du Z, Chang J, et al. Pancreatic cancer risk variant in LINC00673 creates a miR-1231 binding site and interferes with PTPN11 degradation. *Nat Genet.* 2016;48:747–57.
 247. Soldner F, Stelzer Y, Shivalila CS, Abraham BJ, Latourelle JC, Barrasa MI, et al. Parkinson-associated risk variant in distal enhancer of α -synuclein modulates target gene expression. *Nature.* 2016;533:95–9.
 248. Schedel M, Michel S, Gaertner VD, Toncheva AA, Depner M, Binia A, et al. Polymorphisms related to ORMDL3 are associated with asthma susceptibility, alterations in transcriptional regulation of ORMDL3, and changes in TH2 cytokine levels. *J Allergy Clin Immunol.* 2015;136:893–903.e14.
 249. Yang C, Stueve TR, Yan C, Rhie SK, Mullen DJ, Luo J, et al. Positional integration of lung adenocarcinoma susceptibility loci with primary human alveolar epithelial cell epigenomes. *Epigenomics.* 2018;10:1167–87.
 250. Oldoni F, Palmieri J, Giambartolomei C, Howard P, Drenos F, Plagnol V, et al. Post-GWAS methodologies for localisation of functional non-coding variants: ANGPTL3. *Atherosclerosis.* 2016;246:193–201.
 251. Sakurai D, Zhao J, Deng Y, Kelly JA, Brown EE, Harley JB, et al. Preferential binding to Elk-1 by SLE-associated IL10 risk allele upregulates IL10 expression. *PLoS Genet.* 2013;9:e1003870.
 252. Padhy B, Hayat B, Nanda GG, Mohanty PP, Alone DP. Pseudoexfoliation and Alzheimer's associated CLU risk variant, rs2279590, lies within an enhancer element and regulates CLU, EPHX2 and PTK2B gene expression. *Hum Mol Genet.* 2017;26:4519–29.
 253. Bu H, Narisu N, Schlick B, Rainer J, Manke T, Schäfer G, et al. Putative prostate cancer risk SNP in an androgen receptor-binding site of the melanophilin gene illustrates enrichment of risk snps in androgen receptor target sites. *Hum Mutat.* 2016;37:52–64.
 254. Jones SA, Cantsilieris S, Fan H, Cheng Q, Russ BE, Tucker EJ, et al. Rare variants in non-coding regulatory regions of the genome that affect gene expression in systemic lupus erythematosus. *Sci Rep.* 2019;9:15433.
 255. Richard AC, Peters JE, Savinykh N, Lee JC, Hawley ET, Meylan F, et al. Reduced monocyte and macrophage TNFSF15/TL1A expression is associated with susceptibility to inflammatory bowel disease. *PLoS Genet.* 2018;14:e1007458.
 256. Cardinale CJ, March ME, Lin X, Liu Y, Spruce LA, Bradfield JP, et al. Regulation of Janus kinase 2 by an inflammatory bowel disease causal non-coding single nucleotide polymorphism. *J Crohns Colitis England.* 2020;14:646–53.
 257. Qin L, Tiwari AK, Zai CC, Freeman N, Zhai D, Liu F, et al. Regulation of melanocortin-4-receptor (MC4R) expression by SNP rs17066842 is dependent on glucose concentration. *Eur Neuropsychopharmacol Netherlands.* 2020;37:39–48.
 258. Helling BA, Gerber AN, Kadiyala V, Sasse SK, Pedersen BS, Sparks L, et al. Regulation of MUC5B expression in idiopathic pulmonary fibrosis. *Am J Respir Cell Mol Biol.* 2017;57:91–9.
 259. Reinisalo M, Putula J, Mannermaa E, Urtti A, Honkakoski P. Regulation of the human tyrosinase gene in retinal pigment epithelium cells: the significance of transcription factor orthodenticle homeobox 2 and its polymorphic binding site. *Mol Vis.* 2012;18:38–54.
 260. Du M, Zheng R, Ma G, Chu H, Lu J, Li S, et al. Remote modulation of lncRNA GCLET by risk variant at 16p13 underlying genetic susceptibility to gastric cancer. *Sci Adv.* 2020;6:eaa5525.
 261. Pasula S, Tesson KL, Fu Y, Gopalakrishnan J, Pelikan RC, Kelly JA, et al. Role of systemic lupus erythematosus risk variants with opposing functional effects as a driver of hypomorphic expression of TNIP1 and other genes within a three-dimensional chromatin network. *Arthritis Rheumatol.* 2020;72:780–90.
 262. Yang Y-C, Fu W-P, Zhang J, Zhong L, Cai S-X, Sun C. rs401681 and rs402710 confer lung cancer susceptibility by regulating TERT expression instead of CLPTM1L in East Asian populations. *Carcinogenesis England.* 2018;39:1216–21.
 263. Pan G, Cavalli M, Carlsson B, Skrtic S, Kumar C, Wadelius C. rs953413 Regulates polyunsaturated fatty acid metabolism by modulating ELOVL2 expression. *iScience.* 2020;23:100808.
 264. Nanda GG, Kumar MV, Pradhan L, Padhy B, Sundaray S, Das S, et al. rs4246215 is targeted by hsa-miR1236 to regulate FEN1 expression but is not associated with Fuchs' endothelial corneal dystrophy. *PLoS ONE.* 2018;13:e0204278.

265. Hauberg ME, Holm-Nielsen MH, Mattheisen M, Askou AL, Grove J, Børglum AD, et al. Schizophrenia risk variants affecting microRNA function and site-specific regulation of NT5C2 by miR-206. *Eur Neuropsychopharmacol.* 2016;26:1522–6.
266. Hou Y, Liang W, Zhang J, Li Q, Ou H, Wang Z, et al. Schizophrenia-associated rs4702 G allele-specific downregulation of FURIN expression by miR-338-3p reduces BDNF production. *Schizophr Res.* 2018;199:176–80.
267. Guillen-Guio B, Lorenzo-Salazar JM, Ma S-F, Hou P-C, Hernandez-Beeftink T, Corrales A, et al. Sepsis-associated acute respiratory distress syndrome in individuals of European ancestry: a genome-wide association study. *Lancet Respir Med.* 2020;8:258–66.
268. Xiao F, Zhang P, Wang Y, Tian Y, James M, Huang C-C, et al. Single-nucleotide polymorphism rs13426236 contributes to an increased prostate cancer risk via regulating MLPB1 splicing variant 4. *Mol Carcinog.* 2020;59:45–55.
269. Hou G, Harley ITW, Lu X, Zhou T, Xu N, Yao C, et al. SLE non-coding genetic risk variant determines the epigenetic dysfunction of an immune cell specific enhancer that controls disease-critical microRNA expression. *Nat Commun.* 2021;12:135.
270. Fortini BK, Tring S, Devall MA, Ali MW, Plummer SJ, Casey G. SNPs associated with colorectal cancer at 15q13.3 affect risk enhancers that modulate GREM1 gene expression. *Hum Mutat.* 2021;42:237–45.
271. Liu S, Liu Y, Zhang Q, Wu J, Liang J, Yu S, et al. Systematic identification of regulatory variants associated with cancer risk. *Genome Biol.* 2017;18:194.
272. Kong X, Sawalha AH. Takayasu arteritis risk locus in IL6 represses the anti-inflammatory gene GPMB through chromatin looping and recruiting MEF2-HDAC complex. *Ann Rheum Dis.* 2019;78:1388–97.
273. Wang S, Wen F, Tessner KL, Gaffney PM. TALEN-mediated enhancer knockout influences TNFAIP3 gene expression and mimics a molecular phenotype associated with systemic lupus erythematosus. *Genes Immun.* 2016;17:165–70.
274. Wei R, Cao L, Pu H, Wang H, Zheng Y, Niu X, et al. TERT Polymorphism rs2736100-C is associated with EGFR mutation-positive non-small cell lung cancer. *Clin Cancer Res.* 2015;21:5173–80.
275. Sheng X, Tong N, Tao G, Luo D, Wang M, Fang Y, et al. TERT polymorphisms modify the risk of acute lymphoblastic leukemia in Chinese children. *Carcinogenesis England.* 2013;34:228–35.
276. Lubbe SJ, Pittman AM, Oliver B, Lloyd A, Vijayakrishnan J, Naranjo S, et al. The 14q22.2 colorectal cancer variant rs4444235 shows cis-acting regulation of BMP4. *Oncogene.* 2012;31:3777–84.
277. Ghanbari M, Sedaghat S, de Looper HWJ, Hofman A, Erkeland SJ, Franco OH, et al. The association of common polymorphisms in miR-196a2 with waist to hip ratio and miR-1908 with serum lipid and glucose. *Obesity (Silver Spring);* 2015;23:495–503.
278. Prestel M, Prell-Schicker C, Webb T, Malik R, Lindner B, Ziesch N, et al. The atherosclerosis risk variant rs2107595 mediates allele-specific transcriptional regulation of HDAC9 via E2F3 and Rb1. *Stroke United States.* 2019;50:2651–60.
279. Tuupainen S, Turunen M, Lehtonen R, Hallikas O, Vanharanta S, Kivioja T, et al. The common colorectal cancer predisposition SNP rs6983267 at chromosome 8q24 confers potential to enhanced Wnt signaling. *Nat Genet.* 2009;41:885–90.
280. Matthews SM, Eshelman MA, Berg AS, Koltun WA, Yochum GS. The Crohn's disease associated SNP rs6651252 impacts MYC gene expression in human colonic epithelial cells. *PLoS ONE.* 2019;14:e0212850.
281. Li D, Zhu G, Lou S, Ma L, Zhang C, Pan Y, et al. The functional variant of NTN1 contributes to the risk of nonsyndromic cleft lip with or without cleft palate. *Eur J Hum Genet.* 2020;28:453–60.
282. Vecellio M, Roberts AR, Cohen CJ, Cortes A, Knight JC, Bowness P, et al. The genetic association of RUNX3 with ankylosing spondylitis can be explained by allele-specific effects on IRF4 recruitment that alter gene expression. *Ann Rheum Dis.* 2016;75:1534–40.
283. Deng Y, Li P, Liu W, Pu R, Yang F, Song J, et al. The genetic polymorphism down-regulating HLA-DRB1 enhancer activity facilitates HBV persistence, evolution and hepatocarcinogenesis in the Chinese Han population. *J Viral Hepat England.* 2020;27:1150–61.
284. Yang S, Gao Y, Liu G, Li J, Shi K, Du B, et al. The human ATF1 rs11169571 polymorphism increases essential hypertension risk through modifying miRNA binding. *FEBS Lett England.* 2015;589:2087–93.
285. Li C, Yu Q, Han L, Wang C, Chu N, Liu S. The hURAT1 rs559946 polymorphism and the incidence of gout in Han Chinese men. *Scand J Rheumatol.* 2014;43:35–42.
286. Wang L, Li H, Yang B, Guo L, Han X, Li L, et al. The hypertension risk variant Rs820430 functions as an enhancer of SLC4A7. *Am J Hypertens.* 2017;30:202–8.
287. Syddall CM, Reynard LN, Young DA, Loughlin J. The identification of trans-acting factors that regulate the expression of GDF5 via the osteoarthritis susceptibility SNP rs143383. *PLoS Genet.* 2013;9:e1003557.
288. Zhou L, Fu G, Wei J, Shi J, Pan W, Ren Y, et al. The identification of two regulatory ESCC susceptibility genetic variants in the TERT-CLPTM1L loci. *Oncotarget.* 2016;7:5495–506.
289. Shao L, Zuo X, Yang Y, Zhang Y, Yang N, Shen B, et al. The inherited variations of a p53-responsive enhancer in 13q12.12 confer lung cancer risk by attenuating TNFRSF19 expression. *Genome Biol.* 2019;20:103.
290. Tuo XM, Zhu DL, Chen XF, Rong Y, Guo Y, Yang TL. The osteoporosis susceptible SNP rs4325274 remotely regulates the SOX6 gene through enhancers. *Yi Chuan China.* 2020;42:889–97.
291. Richardson K, Louie-Gao Q, Arnett DK, Parnell LD, Lai C-Q, Davalos A, et al. The PLIN4 variant rs8887 modulates obesity related phenotypes in humans through creation of a novel miR-522 seed site. *PLoS ONE.* 2011;6:e17944.
292. Jendrzewski J, He H, Radomska HS, Li W, Tomsic J, Liyanarachchi S, et al. The polymorphism rs944289 predisposes to papillary thyroid carcinoma through a large intergenic noncoding RNA gene of tumor suppressor type. *Proc Natl Acad Sci USA.* 2012;109:8646–51.
293. Kong HK, Yoon S, Park JH. The regulatory mechanism of the LY6K gene expression in human breast cancer cells. *J Biol Chem.* 2012;287:38889–900.
294. Wang Y, He H, Liyanarachchi S, Genutis LK, Li W, Yu L, et al. The role of SMAD3 in the genetic predisposition to papillary thyroid carcinoma. *Genet Med.* 2018;20:927–35.
295. Afanasyeva MA, Putlyaeva LV, Demin DE, Kulakovskiy IV, Vorontsov IE, Fridman MV, et al. The single nucleotide variant rs12722489 determines differential estrogen receptor binding and enhancer properties of an IL2RA intronic region. *PLoS ONE.* 2017;12:e0172681.
296. Xia Q, Chesni A, Manduchi E, Johnston BT, Lu S, Leonard ME, et al. The type 2 diabetes presumed causal variant within TCF7L2 resides in an element that controls the expression of ACSL5. *Diabetologia Germany.* 2016;59:2360–8.
297. Mellado-Gil JM, Fuente-Martín E, Lorenzo PI, Cobo-Vuilleumier N, López-Noriega L, Martín-Montalvo A, et al. The type 2 diabetes-associated HMG20A gene is mandatory for islet beta cell functional maturity. *Cell Death Dis.* 2018;9:279.
298. Kamens HM, Miyamoto J, Powers MS, Ro K, Soto M, Cox R, et al. The β3 subunit of the nicotinic acetylcholine receptor: modulation of gene expression and nicotine consumption. *Neuropharmacology.* 2015;99:639–49.
299. Pattison JM, Posternak V, Cole MD. Transcription factor KLF5 binds a cyclin E1 polymorphic intronic enhancer to confer increased bladder cancer risk. *Mol Cancer Res.* 2016;14:1078–86.
300. Ding C, Zhang C, Kopp R, Kunej L, Meng Q, Wang L, et al. Transcription factor POU3F2 regulates TRIM8 expression contributing to cellular functions implicated in schizophrenia. *Mol Psychiatry.* 2020;
301. Liu W, Anstee QM, Wang X, Gawrieh S, Gamazon ER, Athinayagan S, et al. Transcriptional regulation of PNPLA3 and its impact on susceptibility to nonalcoholic fatty liver Disease (NAFLD) in humans. *Aging (Albany NY).* 2016;9:26–40.
302. Guthridge JM, Lu R, Sun H, Sun C, Wiley GB, Dominguez N, et al. Two functional lupus-associated BLK promoter variants control cell-type- and developmental-stage-specific transcription. *Am J Hum Genet.* 2014;94:586–98.
303. Liu L, Yang X-L, Zhang H, Zhang Z-J, Wei X-T, Feng G-J, et al. Two novel pleiotropic loci associated with osteoporosis and abdominal obesity. *Hum Genet.* 2020;139:1023–35.
304. Lewis MJ, Vyse S, Shields AM, Boeltz S, Gordon PA, Spector TD, et al. UBE2L3 polymorphism amplifies NF-κB activation and promotes plasma cell development, linking linear ubiquitination to multiple autoimmune diseases. *Am J Hum Genet.* 2015;96:221–34.

305. Dryden NH, Broome LR, Dudbridge F, Johnson N, Orr N, Schoenfelder S, et al. Unbiased analysis of potential targets of breast cancer susceptibility loci by Capture Hi-C. *Genome Res.* 2014;24:1854–68.
306. Wright JB, Brown SJ, Cole MD. Upregulation of c-MYC in cis through a large chromatin loop linked to a cancer risk-associated single-nucleotide polymorphism in colorectal cancer cells. *Mol Cell Biol.* 2010;30:1411–20.
307. Smith AJP, Howard P, Shah S, Eriksson P, Stender S, Giambartolomei C, et al. Use of allele-specific FAIRE to determine functional regulatory polymorphism using large-scale genotyping arrays. *PLoS Genet.* 2012;8:e1002908.
308. Wang X, Hayes JE, Xu X, Gao X, Mehta D, Lilja HG, et al. Validation of prostate cancer risk variants rs10993994 and rs7098889 by CRISPR/Cas9 mediated genome editing. *Gene.* 2021;768:145265.
309. Sribudiani Y, Metzger M, Osinga J, Rey A, Burns AJ, Thapar N, et al. Variants in RET associated with Hirschsprung's disease affect binding of transcription factors and gene expression. *Gastroenterology.* 2011;140:572–582.e2.
310. Vincentz JW, Firulli BA, Toolan KP, Arking DE, Sotoodehnia N, Wan J, et al. Variation in a left ventricle-specific Hand1 enhancer impairs GATA transcription factor binding and disrupts conduction system development and function. *Circ Res.* 2019;125:575–89.
311. Shirts BH, Howard MT, Hasstedt SJ, Nanjee MN, Knight S, Carlquist JF, et al. Vitamin D dependent effects of APOA5 polymorphisms on HDL cholesterol. *Atherosclerosis.* 2012;222:167–74.
312. Chen G, Ribeiro CMP, Sun L, Okuda K, Kato T, Gilmore RC, et al. XBP1S regulates MUC5B in a promoter variant-dependent pathway in idiopathic pulmonary fibrosis airway epithelia. *Am J Respir Crit Care Med.* 2019;200:220–34.
313. Mizuta I, Takafuji K, Ando Y, Satake W, Kanagawa M, Kobayashi K, et al. YY1 binds to α -synuclein 3'-flanking region SNP and stimulates antisense noncoding RNA expression. *J Hum Genet England.* 2013;58:711–9.
314. Cano-Gamez E, Trynka G. From GWAS to Function: Using Functional Genomics to Identify the Mechanisms Underlying Complex Diseases. *Front Genet [Internet].* 2020 [cited 2020 Jun 8];11. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7237642/>
315. Edwards SL, Beesley J, French JD, Dunning AM. Beyond GWASs: illuminating the dark road from association to function. *Am J Human Genet.* 2013;93:779–97.
316. Bulik-Sullivan BK, Loh P-R, Finucane H, Ripke S, Yang J, Patterson N, et al. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet.* 2015;47:291–5.
317. Heritability of >4,000 traits & disorders in UK Biobank [Internet]. [cited 2022 Feb 4]. Available from: https://nealelab.github.io/UKBB_ldsc/
318. Perenthaler E, Yousefi S, Niggi E, Barakat TS. Beyond the Exome: The Non-coding Genome and Enhancers in Neurodevelopmental Disorders and Malformations of Cortical Development. *Front Cell Neurosci [Internet].* Frontiers; 2019 [cited 2021 Jun 10];13. Available from: <https://doi.org/10.3389/fncel.2019.00352/full>
319. French JD, Edwards SL. The role of noncoding variants in heritable disease. *Trends Genet.* 2020;36:880–91.
320. Rojano E, Seoane P, Ranea JAG, Perkins JR. Regulatory variants: from detection to predicting impact. *Brief Bioinform.* 2019;20:1639–54.
321. Moore LD, Le T, Fan G. DNA methylation and its basic function. *Neuropsychopharmacology.* 2013;38:23–38.
322. Lowdon RF, Jang HS, Wang T. Evolution of epigenetic regulation in vertebrate genomes. *Trends Genet.* 2016;32:269–83.
323. Zhang P, Wu W, Chen Q, Chen M. Non-Coding RNAs and their Integrated Networks. *J Integr Bioinform [Internet].* 2019 [cited 2021 May 31];16. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6798851/>
324. Cammaerts S, Strazisar M, De Rijk P, Del Favero J. Genetic variants in microRNA genes: impact on microRNA expression, function, and disease. *Front Genet.* 2015;6:186.
325. Felekis K, Touvana E, Stefanou C, Deltas C. microRNAs: a newly described class of encoded molecules that play a role in health and disease. *Hippokratia.* 2010;14:236–40.
326. Steri M, Idda ML, Whalen MB, Orrù V. Genetic variants in mRNA untranslated regions. *Wiley Interdiscip Rev RNA.* 2018;9:e1474.
327. A M, M G, Jf C, R B. SNPs in microRNA target sites and their potential role in human disease. *Open biology [Internet].* Open Biol; 2017 [cited 2021 May 31];7. Available from: <https://pubmed.ncbi.nlm.nih.gov/28381629/>
328. Statello L, Guo C-J, Chen L-L, Huarte M. Gene regulation by long non-coding RNAs and its biological functions. *Nature Rev Mol Cell Biol.* 2021;22:96–118.
329. Giral H, Landmesser U, Kratzer A. Into the Wild: GWAS Exploration of Non-coding RNAs. *Front Cardiovasc Med [Internet].* Frontiers; 2018 [cited 2021 Jun 10];5. Available from: <https://doi.org/10.3389/fcvm.2018.00181/full>
330. Gasperini M, Hill AJ, McFaline-Figueroa JL, Martin B, Kim S, Zhang MD, et al. A genome-wide framework for mapping gene regulation via cellular genetic screens. *Cell.* 2019;176:377–390.e19.
331. Schraivogel D, Gschwind AR, Milbank JH, Leonce DR, Jakob P, Mathur L, et al. Targeted Perturb-seq enables genome-scale genetic screens in single cells. *Nat Methods [Internet].* 2020 [cited 2020 Jun 2]; Available from: <http://www.nature.com/articles/s41592-020-0837-5>
332. Boix CA, James BT, Park YP, Meuleman W, Kellis M. Regulatory genomic circuitry of human disease loci by integrative epigenomics. *Nature.* 2021;590:300–7.
333. Doni Jayavelu N, Jajodia A, Mishra A, Hawkins RD. Candidate silencer elements for the human and mouse genomes. *Nat Commun [Internet].* 2020 [cited 2020 Apr 21];11. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7044160/>
334. Gasperini M, Tome JM, Shendure J. Towards a comprehensive catalogue of validated and target-linked human enhancers. *Nat Rev Genet.* 2020;21:292–310.
335. Pang B, Snyder MP. Systematic identification of silencers in human cells. *Nat Genet.* 2020;52:254–63.
336. Bramer WM, de Jonge GB, Rethlefsen ML, Mast F, Kleijnen J. A systematic approach to searching: an efficient and complete method to develop literature searches. *J Med Libr Assoc.* 2018;106:531–41.
337. Bramer WM, Giustini D, Kramer BMR. Comparing the coverage, recall, and precision of searches for 120 systematic reviews in Embase, MEDLINE, and Google Scholar: a prospective study. *Syst Rev.* 2016;5:39.
338. Wang X, Tucker NR, Rizki G, Mills R, Krijger PH, de Wit E, et al. Discovery and validation of sub-threshold genome-wide association study loci using epigenomic signatures. *Elife.* 2016;5.
339. Stadhouders R, Aktuna S, Thongjuea S, Aghajaniyeh A, Pourfarzad F, van Ijcken W, et al. HBS1L-MYB intergenic variants modulate fetal hemoglobin via long-range MYB enhancers. *J Clin Invest.* 2014;124:1699–710.
340. Pashos EE, Park Y, Wang X, Raghavan A, Yang W, Abbey D, et al. Large, diverse population cohorts of hiPSCs and derived hepatocyte-like cells reveal functional genetic variation at blood lipid-associated loci. *Cell Stem Cell.* 2017;20:558–570.e10.
341. Visser M, Kayser M, Palstra R-J. HERC2 rs12913832 modulates human pigmentation by attenuating chromatin-loop formation between a long-range enhancer and the OCA2 promoter. *Genome Res.* 2012;22:446–55.
342. Musunuru K, Strong A, Frank-Kamenetsky M, Lee NE, Ahfeldt T, Sachs KV, et al. From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. *Nature.* 2010;466:714–9.
343. Guo H, Ahmed M, Zhang F, Yao CQ, Li S, Liang Y, et al. Modulation of long noncoding RNAs by risk SNPs underlying genetic predispositions to prostate cancer. *Nat Genet.* 2016;48:1142–50.
344. Ghousaini M, French JD, Michailidou K, Nord S, Beesley J, Canisus S, et al. Evidence that the 5p12 variant rs10941679 confers susceptibility to estrogen-receptor-positive breast cancer through FGF10 and MRPS30 regulation. *Am J Hum Genet.* 2016;99:903–11.
345. Viñuela A, Varshney A, van de Bunt M, Prasad RB, Asplund O, Bennett A, et al. Genetic variant effects on gene expression in human pancreatic islets and their implications for T2D. *Nat Commun.* 2020;11:4912.
346. Stacey D, Fauman EB, Ziemek D, Sun BB, Harshfield EL, Wood AM, et al. ProGeM: a framework for the prioritization of candidate causal genes at molecular quantitative trait loci. *Nucleic Acids Res.* 2019;47:e3.
347. Fang H, ULTRA-DD Consortium, De Wolf H, Knezevic B, Burnham KL, Osgood J, et al. A genetics-led approach defines the drug target landscape of 30 immune-related traits. *Nat Genet.* 2019;51:1082–91.
348. Lukowski SW, Lloyd-Jones LR, Holloway A, Kirsten H, Hemani G, Yang J, et al. Genetic correlations reveal the shared genetic architecture of transcription in human peripheral blood. *Nat Commun.* 2017;8:483.

Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

