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

Ammar J. Alsheikh1*, 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
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 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 super-population. 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 $h^2$
309 validated non-coding variants regulating 252 genes across 130 diseases

**GWAS Diseases association**
- Neoplasms
- Cardiovascular System
- Immune System
- Musculoskeletal System
- Endocrine System
- Digestive System
- Nervous System
- Nutritional and Metabolic
- Others

**Validated Variant class**
- Promoter
- Cis-regulatory element
- Non-coding RNA

**Fig. 2** Map of 309 validated GWAS non-coding variants. The Circos plot displays the 309 experimentally validated variants studied within the 286 included articles. The outer most layer (i) shows the validated variants’ 252 target genes, (ii) the chromosomal map, (iii) the location of validated variants marked by their rsIDs, (iv) using higher order ontology mapping, we display inner links between variants associated with diseases in the same category. Disease systems that contain ten or more validated variants are displayed while those contain less than ten validated variants are grouped in “Others” category, and (v) the manually annotated validated variant class. Additional File 3 contains all variant details and annotations.
for UK Biobank phenotypes estimated using LD Score Regression\cite{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

---

**Fig. 3** Functional validation remains the bottleneck of GWAS follow-up. A Comparison of the number of published studies in the GWAS catalog and non-coding variant validation studies over time. B Relationship between the ratio of validated non-coding variants to the total GWAS variants and disease category mean heritability. C Linkage disequilibrium between reported variant in GWAS Catalog and validated variants. D Distance between validated variant and GWAS Catalog-reported variant. E Global minor allele frequency (MAF) of validated variants in 1000 genomes phase 3. F Location of experimentally validated non-coding GWAS variants in relation to all protein-coding genes compared to GWAS lead variants.
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 exemplary 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-MAKP3-TAOK2 in schizophrenia [168], rs11236797-LRRC32 in inflammatory bowel disease [40], and rs9349379-EDN1 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 insulin 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-BARXI in esophageal adenocarcinoma [88], rs4065275-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-NRBP1 in gout [127], rs143383-GDFS in osteoarthritis [119], and rs35705950-MIICS 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 HOX8 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...
**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 ~70% of validated variants fall within 10 kb and r² ≥ 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

---

**Fig. 5** Studies utilize multiple avenues in validating non-coding variants. Using text-mining of abstracts and metadata, we examined the utilization of different avenues for non-coding variant validation across 286 included articles. The six broad categories were gene expression, reporter assays, transcription factor binding, in vivo or animal models, genome editing, and chromatin interaction. The intersection size denotes the number of articles that have the combination of validation categories below it. The color denotes the number of avenues used; pink—6, orange—5, green—4, black—3, blue—2, red—1. The upset plot shows the overlap of the variant validation avenues and the number of articles. The Set size bars on the right reflect the total number of studies that used/employed each of the categories.
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.

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, Rotmni 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/ [cited 2021 Apr 11]; Available from: https://www.omim.org/
3. Tam V, Patel N, Turcotte M, Bosse 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 JY, 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, Cezero 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. Genetic and epigenetic fine mapping of causal autoimmune disease variants. Nature. 2015;518:337–43.
7. Gallagher MD, Chen-Piotkin 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 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, Buiton I, Hoffmann TC, Mullow 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 J, Reed JZ, Milward D, Xu J, 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. Ghousaini 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. [cited 2021 Jun 15]; Available from: http://wayback.archive-it.org-350/20180312141554/https://www.nlm.nih.gov/pubs/factsheets/medline.html.
18. Linguamatics. https://www.linguamatics.com/. [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. NCBI Thesaurus. https://ncit.nci.nih.gov/ncitbrowser/, 2021;
21. NCBI Gene Database. https://www.ncbi.nlm.nih.gov/gene/. [cited 2021 Jun 15]; Available from: https://www.ncbi.nlm.nih.gov/gene/.
22. TERMathe-SciBite. https://www.scibite.com/platform/termite/. SciBite [Internet]. [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 RB, 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, Tinney 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. Almestachini NAM, Antone D, Zhou X, Vilmundarson RO, Zhang SX, Hao KKN, 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. Common atopy-associated variant in the Th2 cytokine locus control region impacts transcriptional regulation and alters SMAD3 and SPI1 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, Huynh JE, 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. Cho J, Xu M, Makowski MM, Zhang T, Law MH, Kostacs 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. painter JH, 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.

37. 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.

38. Gallagher MD, Posvari 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.

39. Nasrallah R, Imianowski CJ, Bossini-Castillo L, Grant FM, Dogan M, Díaz-Jiménez D, Núñez L, De la Fuente M, Dubois-Camacho K, 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.

40. Dowson MC, 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.

41. Chinnaswamy S, Chatterjee S, Boopathi R, Mukherjee S, Bhattacharyya S, Khandpur S, et al. A common functional regulatory variant on 20q13.33 related to glioma risk alters enhancer activity and modulates expression of multiple genes. Hum Mutat. 2016;37:1821–26.

42. Chang J, Tian J, Yang Y, Zhong R, Li J, Zhai K, et al. A Rare Missense variant in TFAP2a. Am J Hum Genet. 2016;99:1281–91.

43. Ye J, Tucker NR, Weng L-C, Clauss S, Lubitz SA, Ellinor PT. A functional variant affecting the APOA5 gene expression is causally associated with plasma triglyceride levels conferring coronary atherosclerosis risk in Han Chinese Population. Biochim Biophys Acta. 2014;1842:2147–54.

44. Gong J, Tian J, Lou J, Ke J, Li L, Li J, et al. A functional polymorphism in Inc-LAMC2-1.1 confers risk of colorectal cancer by affecting miRNA binding. Carcinogenesis England. 2016;37:443–51.

45. Saeki N, Saito A, Choi UJ, Matsu 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. Gastroenterology USA. 2011;140:892–902.

46. Ogura Y, Kour 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.

47. Ye J, Tucker NR, Weng L-C, Clauss S, Lubitz SA, Ellinor PT. A functional variant associated with atrial fibrillation regulates PITX2c expression through TPATPA2. Am J Hum Genet. 2015;99:1281–91.

48. Akamatsu S, Takata R, Ashikawa K, Hosono N, Kamatani N, Fujioka T, et al. A functional variant associated with sagittal nonsyndromic craniosynostosis alters the activity of an methylatable Foxa2 binding site of the G6PC2 promoter is associ-
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. 2017;100:545–52.

71. Verlaan DJ, Berlivert S, Hunninghake GM, Madore A-M, Larivière M, Moussette S, et al. Allele-specific chromatin remodeling in the ZFPBP2/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;317:456–63.

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, et al. Redis RS, Vela LE, Lu W, Ferreira de Oliveira J, Ivan C, Rodriguez-Aguayo C, et al. Allele-specific transcriptional regulation at type 2 diabetes-associated single nucleotide polymorphisms in regions of pancreatic islet open chromatin at the JAZF1 locus. Diabetes. 2013;62:1756–62.

75. 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.

76. Pittman AM, Naranjo S, Dalenova S, Pohjalainen S, Poulton RJ, Oudit GY, et al. Characterization of coding/noncoding variants for SHROOM3 in patients of African descent. Am J Hum Genet. 2013;9:e1003750.

77. Thynn NN, Chen X-F, Zhang H, Wright JB, Bailey JD, Cole MD, Eeckhoute J, Bailey SD, Cole MD, Eeckhoute 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.

78. Shah MY, Ferracin M, Piecikowski D, Chen R, 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.

79. Glubb DM, Shi W, Beesley J, Fodde R, Pritchard J-L, McCue K, et al. Characterization of coding/noncoding variants for SHROOM3 in patients of African descent. Cancers (Basel). 2020;12.

80. McGovern A, Schoenfelder S, Martin P, Massey J, 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.

81. Alsheikh T, Alarcon C, Hernandez W, Liko I, Cavallini LH, Duarte JD, et al. Association of genetic variants with warfarin-associated bleeding among patients of African descent. Jama. 2018;320:1670–7.

82. 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.

83. Wu C, Khan AM, Bloch DB, Davis-Dunseben BN, Ghobari A, et al. Attrial natriuretic peptide is negatively regulated by microRNA-425. J Clin Invest. 2013;123:3378–82.

84. Zhao H, Yang W, Qiu R, Li J, Xin Q, Wang X, et al. An intronic variant of the transcriptional machinery bound across the widely presumed squamous cell carcinoma by regulating the function of BARX1. Cancer Lett. Ireland. 2018;421:103–11.

85. 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.

86. 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.

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

88. Liu H, Duncan K, Helversen A, Kumar P, Mumm C, Xiao Y, et al. Analysis of zebrafish perilipin enhancers identifies regulation of a novel variant near human KRTB1/18. Elife. 2020;9.

89. Park JH, Chang HS, Park C-S, Jang A-S, Park YL, 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.
type 2 diabetes causal variant, rs7903146, within TCF7L2. Eur J Hum Genet. 2015;23:103–9.

109. Comisarow DF, He H, Lyamatarach 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;6:23302.

111. Matoba N, Liang D, Sun H, Aygün N, McMaele JC, Davis JE, et al. Common genetic risk variants identified in the SPARK cohort support DHD2 as a candidate risk gene for autism. Transl Psychiatry. 2020;10:265.

112. Hiramoto M, Udagawa H, Watanabe A, Miyazawa K, Ishibashi N, Clifton-Bligh RJ, Nguyen TV, Au A, Bullock M, Cameron I, Cumming CL, Haas U, Diaz R, Leeper NJ, Kundu RK, Patlolla B, et al. Gee F, Rushton MD, Loughlin J, Reynard LN. Correlation of the osteoarthritis genetic susceptibility locus with 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:e1004102.

118. 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;525:658–65.

119. Ye W, Wang Y, Mei H, 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.

120. 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.

126. Rahimov F, Marazita ML, Visel A, Cooper ME, Hitchler MJ, Rubini M, et al. Genet. 2015;23:103–9.

127. Hiramoto M, Udagawa H, Watanabe A, Miyazawa K, Ishibashi N, Clifton-Bligh RJ, Nguyen TV, Au A, Bullock M, Cameron I, Cumming CL, Haas U, Diaz R, Leeper NJ, Kundu RK, Patlolla B, et al. Gee F, Rushton MD, Loughlin J, Reynard LN. Correlation of the osteoarthritis genetic susceptibility locus with fracture risk in the elderly. Calcif Tissue Int. 2011;88:109–16.

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 SJ, Uhm TG, Kang JH, Park J-S, Jang A-S, et al. Effect of a single nucleotide polymorphism within the interleukin-4 promoter on aspin 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, Cioonan N, Henni G, et al. Endometriosis risk alleles at 1p36.12 act through inverse regulation of CDCA4 and LIN28359. 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:e024463.

132. Sio YY, Matta SA, Ng YT, Chew FT. Epistasis between phosphatidylinositol-3-kinase and β2-adrenergic receptor influences extracellular epinephrine level and associates with the susceptibility to allergic asthma. Clin Exp Allergy England. 2020;50:352–63.

133. Veccellio M, Cortes A, Roberts AR, Ellis J, Cohen CJ, Knight JC, et al. Evidence for a second ankylospondylitis-associated RUNX3 regulatory polymorphism. RMD Open. 2018;e000628.

134. Michailidou K, Edwards JL, Michailidou K, Nord S, Coop-Marini F, Desai K, et al. Evidence that breast cancer risk at the 2q35 locus is mediated through IGFBP5 regulation. Nat Commun. 2014;44:4999.

135. Shepherd C, Skeiton AJ, Rushton MN, Loughlin J. Expression analysis of the osteoarthritis genetic susceptibility locus mapping association study and functional analysis implicate a SNP in the MIF2L gene and marked by the polymorphism rs11842874. BMC Med Genet. 2015;16:108.
147. Claussnitzer M, Dankel SN, Kim K-H, Quon G, Meuleman W, Haugen C, et al. FTO obesity variant circuity 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, Hazeltine DL, et al. Functional analysis and fine mapping of the 9q22.2 ovarian cancer susceptibility locus. Cancer Res. 2019;79:467–81.

149. Boardman-Pretty F, Smith AJP, Cooper J, Palmen J, Folkesen 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, Martinkin A, Silva A, Lau P, Nikpay M, Eriksson P, et al. Functional analysis of a novel genome-wide association study signal in SNA3D 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, Baskin R, Woods NT, Mendoza-Fandiño G, Forsyth P, Egli RJ, Southam L, Wilkins JM, Lorenzen I, Pombo-Suarez M, Gonzalez Douvris A, Soubeyrand S, Naing T, Martinuk A, Nikpay M, Williams Zhang Y, Kuipers AL, Yerges-Armstrong LM, Nestlerode CS, Jin Z, Eckart N, Song Q, Yang R, Wang R, Zhu H, McCallion AS, et al. Functional evaluation of TERT-CLPTM1L genetic variants associated with susceptibility of papilloma virus infection spondylitis. Arthritis Rheumatol United States. 2021;73:980–90.

152. 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 n4420550 in the 16p11.2 Schizophrenia-Associated Locus. Biol Psychiatry United States. 2021;89:246–55.

153. 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 Homebox Gene Causes Body Axis Deformation. PLoS Genet. 2016;12:e1005802.

154. 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.

155. Lawrenson K, Kar S, McCue K, Kuchenbaeker K, Michalidou K, Tyrer J, et al. Functional mechanisms underlying pleiotropic risk alleles at the 19p13.11 breast-ovarian cancer susceptibility locus. Nat Commun. 2016;7:2675.

156. Pérez-Razo JC, Cano-Martínez LJ, Vargas Alarcón G, Canizales-Quinteros S, Martínez-Rodríguez N, Canto P, et al. Functional polymorphism rs13306560 of the MTIFR gene is associated with essential hypertension in a Mexican-Mestizo Population. Circ Cardiovasc Genet United States. 2015;8:603–9.

157. Nanda V, Wang T, Planic M, Liu B, Nguyen T, Matic LR, et al. Functional regulatory dissection of smooth muscle cell-restricted LMOD1 coronary artery disease locus. PLoS Genet. 2018;14:e1007755.

158. 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.

159. Ustiuogova AS, Korneev KV, Kupershmidt D, Tan W, Che X, Tan W, et al. Functional role of SLC47A1 gene in neuronal cells and on DNA-protein interactions. Am J Med Genet B Neuropsychiatr Genet. 2010;153B:185–201.

160. 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.

161. Yu W, Zhang K, Wang Z, Zhang J, Chen T, Jin L. Functional variant in the pro-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.

162. French JD, Choussaini M, Edwards SL, Meyer KB, Michailidou K, Ahmed S, et al. Functional characterization of SNPs in the Human Autoimmunity-Associated Locus 17q12–21. Genes (Basel). 2019;10:2434.

163. Andiappan AK, Sio YY, Lee B, Suri BK, Matta SA, Lum J, et al. Functional characterization of three single-nucleotide polymorphisms (SNPs) in the Human Autoimmunity-Associated Locus 17q12–21. Genes (Basel). 2019;10:2434.

164. Li Y, Nie Y, Cao J, Tu S, Lin Y, Du Y, et al. G-A variant in miR-200c binding site of promoter region associated with drug behaviors. Brain Res. 2013;1529:1–15.

165. Oldridge DA, Wood AC, Weichert-Leahey N, Crimmins I, Sussman R, Cavalli M, Pan G, Nord H, Wallén Arzt E, Wallerman O, Wadelius C. Genetic prevention of hepatitis C virus-induced liver fibrosis by allele-specific downregulation of MERTK. Hepatol Res Netherlands. 2017;51:826–30.

166. Krause MD, Huang R-T, Wu D, Shentu T-P, Harrison DL, Whalen MB, et al. Genetic variant at the 17q12-21 is associated with allergic asthma but not allergic rhinitis. J Allergy Clin Immunol United States. 2016;137:758–766.e3.

167. Yu W, Zhang K, Zeng L, Liu M, Chen T. Functional association of a RUNX3 polymorphism associated with ankylosing spondylitis. Arthritis Rheumatol United States. 2021;73:980–90.

168. Cavalli M, Pan G, Nord H, Wallén Arzt E, Wallerman O, Wadelius C. Genetic prevention of hepatitis C virus-induced liver fibrosis by allele-specific downregulation of MERTK. Hepatol Res Netherlands. 2017;51:826–30.

169. 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:E13149–58.

170. Soderquest K, Hertweck A, Giambartolomei C, Henderson S, Mohamed R, Goldberg R, et al. Genetic variants alter T-bet binding
191. Hou S, Du L, Lei B, Pang CP, Zhang M, Zhuang W, et al. Genome-wide

192. Kawamura R, Tabara Y, Tsukada A, Igase M, Ohashi J, Yamada R, et al.

194. Kalita CA, Brown CD, Freiman A, Isherwood J, Wen X, Pique-Regi

197. Smith EN, D’Antonio-Chronowska A, Greenwald WW, Borja V, Aguiar

198. Hitomi Y, Kawashima M, Aiba Y, Nishida N, Matsuhashi M, Okazaki H,

199. López Rodríguez M, Kaminska D, Lappalainen K, Pihlajamäki J, Kaik -

202. Spracklen CN, Shi J, Vadlamudi S, Gaulton KJ, Mohlke KL.

203. Liu L, Pei Y-F, Liu T-L, Hu W-Z, Yang X-L, Li S-C, et al. Identification of

204. Zhou X, Baron RM, Hardin M, Cho MH, Zielinski J, Hawrylkiewicz I, et al.

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 KSFA-CYP27B1-METTL1- FAM19B1 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.

211. Parker MM, Hao Y, Guo F, Pham B, Chase R, Platig J, et al. Identification of an enhysmema-associated genetic variant near TGFβ2 with regulatory effects in lung fibroblasts. Elife. 2019;8.

212. Ryoo H, Kong M, Kim Y, Lee C. Identification of functional nucleotide and haplotype variants in the promoter of the CEBFβ gene. J Hum Genet. 2013;58:500–3.

213. van Ooijenfert AK, 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, Yu F, Naing ZCC, Li Y, et al. Identification of function variants in the FAM13A chronic obstructive pulmonary disease genome-wide Association Study locus by Massive 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. Fairoozy 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 rs617172 association with lower LDL-C and risk of CHD. PLoS ONE. 2016;11:e0167676.

217. Guo X, Lin W, Wen W, Huyghe J, Bens 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. Amiel-Wolf A, Tang M, Way J, Dombroski B, Jiang M, Vrettos N, et al. Inhibiting 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, Hammad R, 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, Boocock J, Nikkola 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-Sal lari B, Bailey SD, Moore JH, Lupien M. Integrative functional genomics identifies an enhancer looping to the SOX9 gene disrupted by the 17q24.5 prostate cancer risk locus. Genotype. 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, Champain D, Roy S, Song JS. Integrative genomic analysis predicts causative cis-regulatory mechanisms of the Breast Cancer-Associated Genetic Variant rs4415084. Cancer Research. 2018;78:1579–91.

224. Berlivi S, Moussette S, Oumet M, Verlaan DJ, Koka A, Ji Tuwaijari A, et al. Integration 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 (Solitin 1) locus with primary human hepatocytes, induced pluripotent stem
et al. BMC Medical Genomics          (2022) 15:74

227. Reschen ME, Gaulton KJ, Lin D, Soilleux EJ, Morris AJ, Smyth SS, Mei B, Wang Y, Ye W, Huang H, Zhou Q, Chen Y, et al. LncRNA ZBTB40-IT1

229. Vicente CT, Edwards SL, Hillman KM, Kaufmann S, Mitchell H, Bain L, Cavalli M, Pan G, Nord H, Wadelius C. Looking beyond GWAS: allele-

231. Lu X, Zoller EE, Weirauch MT, Wu Z, Namjou B, Williams AH, et al. Lupus clastogenesis via sponging miR-7. J Bone Miner Res. 2015;30:105063.

232. Choi J, Zhang T, Vu A, Ablain J, Makowski MM, Colli LM, et al. Massively functional variants at 13q14 risk locus for osteoporosis regulate RANKL expression through long-range super-enhancer. J Bone Miner Res. 2018;33:1335–46.

233. He H, Li W, Liyanarachchi S, Sinivas 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;112:6218–33.

234. Roman TS, Marville AF, Fogarty MP Vadlamudi S, Gonzalez AJ, Buchko-vich 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.

235. Bojesen SE, Pooley KA, Johnnathy SE, Beesley J, Michailidou K, Tyer 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.

236. 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.

237. John G, Hegarty JP, Yu W, Berg A, Pastor DM, Kelly AA, et al. NKX2-3 variations converge to promote ESR1 expression in breast cancer. Mol Genet Metab United States. 2011;104:174–9.

238. Bailey SD, Desai K, Kron KI, Mrazekki P, Sinnott-Armstrong NA, Treloar AE, et al. Noncoding standing DNA and inherited single-nucleotide variants converge to promote ESR1 expression in breast cancer. Nat Genet. 2016;48:1260–6.

239. Gorbatenko A, Olesen CW, Loelbl N, Sigurdsson HH, Bianchi C, Pedraz-Cuesta E, et al. Oncogenic p53HER2 regulates Na+–HCO3- cotransporter NBCn1 mRNA stability in breast cancer cells via 3'UTR-dependent posttranslational processes. Mol Chem J England. 2016;47:4027–44.

240. 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.

241. Zheng J, Huang X, Tan W, Yu D, Du Z, Chang J, et al. Pancreatic cancer risk variant in LIN00673 creates a miR-1231 binding site and interferes with PTTPN11 degradation. Nat Genet. 2016;48:47–57.

242. Soldner F, Stelzer Y, Shivaliya CS, Abraham BJ, Latourelle JC, Barsana MI, et al. Parkinson-associated risk variant in distal enhancer of o-synuclein modulates target gene expression. Nature. 2016;533:95–9.

243. Schedel M, Michel S, Gaertner FD, 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;136:989–903.e14.

244. Yang C, Stueve TR, Yan C, Rho S, Liu D, Luo J, et al. Positional inte- pr.}{cfrac{\text{gene expression}}{\text{RNA}}}_{\text{o}}\text{RNA}

245. Padhy B, Hatay 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.

246. Carvalho MA, March ME, Lin X, Liu Y, Spruce LA, Bradford JP, et al. Regulation of Junas kinase 2 by an inflammatory bowel disease causal non-coding single nucleotide polymorphism. J Crohns Colitis England. 2020;14:646–53.

247. 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.

248. 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.

249. Reinisalo M, Putula J, Mannermmaa E, Linti A, Honkakoski P. Regulation of the human tyrosinase gene in retinal pigment epithelium cells: the significance of transcription factor orthodenticle homebox 2 and its polymorphic binding site. Mol Vis. 2012;18:38–54.

250. Du M, Zheng R, Ma G, Chu H, Liu L, Li S, et al. Remote modulation of IncRNA GCLET by risk variant at 16p13 underlying genetic susceptibility to gastric cancer. Sci Adv. 2020;6:eaya5525.

251. Pasula S, Tessneer KL, Yu F, Gopalakrishnan J, Pelikan RC, Kelly JA, et al. Role of systemic lupus erythematosus risk variant with opposing functional effects as a driver of hypomorphic expression of TNIP1 and other genes within a three-dimension chromatin network. Arthritis Rheumatol. 2020;72:780–90.

252. Yang Y-C, Wu F-W, 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. Carcogenes England. 2018;39:1216–21.

253. Pan G, Cavalli M, Carlsson B, Skrtic S, Kumar C, Wadelius C. rs553413 Regulates polysaturated fatty acid metabolism by modulating ELOVL2 expression. iScience. 2020;23:100808.

254. Nanda GG, Kumar MV, Pradhan L, Padhy B, Sundaraj S, Das S, et al. rs466215 is targeted by hsa-miR1236 to regulate FEN1 expression but is not associated with Fuchs’ endothelial corneal dystrophy. PLoS ONE. 2018;13:e0204278.
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.

Guillen-Guido B, Lorenzo-Salazar JM, Ma S-F, Hou P-C, Hernandez-Beftink 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.

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 MLPH splicing variant 4. Mol Carcinog. 2020;59:45–55.

Hou G, Harley ITW, Lu X, Zhou T, Xu N, Yao C, et al. SLE non-coding genetic variant determines the epigenetic dysfunction of an immune cell specific enhancer that controls disease-critical miRNA expression. Nat Commun. 2021;12:1235.

Fortini BK, Tring S, Devall MA, Ali MW, Plummer SJ, Case G. SNPs associated with colorectal cancer at 15q13.3 affect risk enhancers that modulate GREG1 gene expression. Hum Mutat. 2021;42:237–45.

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. 2018;17:194.

Kong X, Sawalha AH. Takayasu arteritis risk locus in IL6 represses the anti-inflammatory gene GPNMB through chromatin looping and recruiting MEF2-HDAC complex. Ann Rheum Dis. 2019;78:1388–97.

Wang S, Wen F, Tesseneur KL, Gaffney PM. TALEN-mediated enhancer knockout influences TNAFAP3 gene expression and mimics a molecular phenotype associated with systemic lupus erythematosus. Genes Immun. 2016;17:165–70.

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.

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.

Lubbe SJ, Pittman AM, Oliver B, Lloyd A, Vijayakrishnan J, Naranjo S, et al. The 1q42q22 colorectal cancer variant rs4444235 shows cis-acting regulation of BMP4. Oncogene. 2012;31:1777–84.

Ghanbari M, Sedaghat S, de Looper HWJ, Hoffman 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.

Prestel M, Prell-Schicker C, Webb T, Malik R, Lindner B, Ziesch N, et al. The atherosclerosis risk variant rs2107595 mediates allelic-specific transcriptional regulation of HDAC9 via EZF3 and Rb1. Stroke United States. 2015;99:261–7.

Tuupanen 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.

Matthews SM, Eshelman MA, Berg AS, Koltun WA, Yochum GS. The Ccrh3 disease-associated SNP rs6651252 impacts MYC gene expression in human colonic epithelial cells. PLoS ONE. 2013;14:e0212850.

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.

Vecellio M, Roberts AR, Cohen CJ, Cortes A, Knight JC, Bowes P, et al. The genetic association of RUNX3 with ankylosing spondylitis can be explained by allele-specific effects on IF4F4 recruitment that alter gene expression. Ann Rheum Dis. 2016;75:1534–40.

Deng Y, Li Y, 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;27:150–61.

Yang S, Gao Y, Liu G, Li J, Shi K, Du B, et al. The human ATFI rs11169571 polymorphism increases essential hypertension risk through modifying miRNA binding. FEBS Lett England. 2015;589:2087–93.

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.

Wang L, Li H, Yang B, Guo L, Han X, Li L, et al. The hypertension risk variant Rs820430 functions as an enhancer of SLCA47. Am J Hypertens. 2017;30:202–8.

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.

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.

Shao L, Zuo X, Yang Y, Zhang Y, Yang N, Shen B, et al. The inherited variation of a p53-responsive enhancer in 13q12.12 confer lung cancer risk by attenuating TNFRSF19 expression. Genome Biol. 2019;20:103.

Tuo XM, Zhu DL, Chen XF, Rong Y, Guo Y, Yang TL. The osteosarcoma susceptible SNP rs4325747 remotely regulates the SOX5 gene through enhancers. Yi Chuan China. 2020;42:889–97.

Richardson K, Louie-Gao A, 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.

Jendrzejewski J, He H, Radomska HS, LW, Tomics, 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.

Kong HK, Yoon S, Park JH. The regulatory mechanism of the LYST gene expression in human breast cancer cells. J Biol Chem. 2012;287:38889–900.

Wang Y, He H, Liyanarachchi S, Genuitis 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.

Afansayeva MA, Putlayeva LV, Demin DE, Kulakovskiy IV, Vorontsov IE, Fridman MV, et al. The single nucleotide variant rs17272489 determines differential estrogen receptor binding and enhancer properties of an IL7RA intronic region. PLoS ONE. 2017;12:e0172681.

Xia Q, Chesi 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.

Mellado-Gil JM, Fuente-Martin E, Lorenzo PI, Cobo-Villelmeur N, Lopez-Noriega I, Martin-Antolavalo A, et al. The type 2 diabetes-associated HMV20A gene is mandatory for islet beta cell functional maturity. Cell Death Dis. 2018;9:279.

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.

Pattison JM, Posternak V, Cole MD. Transcription factor KLF5 binds a cyclin E1 polyomimic enhancer to confer increased bladder cancer risk. Mol Cancer Res. 2016;14:1078–86.

Ding C, Zhang C, Kopp R, Kuney L, Meng Q, Wang L, et al. Transcription factor POUSF2 regulates TRIM8 expression contributing to cellular functions implicated in schizophrenia. Mol Psychiatry. 2020.

Liu W, Anstee QM, Wang X, Gawrieh S, Gamazon ER, Lépez-Noireiga I, Martin-Antolavalo A, et al. The type 2 diabetes-associated HMG20A gene is mandatory for islet beta cell functional maturity. Cell Death Dis. 2018;9:279.

Lewis MJ, Vyse S, Shields AM, Boeltz S, Gordon PA, Spector TD, et al. The 14q22.2 colorectal cancer variant rs4444235 shows cis-acting regulation of HMG20A. Oncogene. 2012;31:1777–84.

Afanasayeva MA, Putlayeva LV, Demin DE, Kulakovskiy IV, Vorontsov IE, Fridman MV, et al. The single nucleotide variant rs17272489 determines differential estrogen receptor binding and enhancer properties of an IL7RA intronic region. PLoS ONE. 2017;12:e0172681.
Publisher’s Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.