



Review

Regulatory SNPs: Altered Transcription Factor Binding Sites Implicated in Complex Traits and Diseases

Arina O. Degtyareva ¹, Elena V. Antontseva ¹ and Tatiana I. Merkulova ^{1,2,*}

¹ Department of Molecular Genetic, Institute of Cytology and Genetics, 630090 Novosibirsk, Russia; degtyareva_rso@mail.ru (A.O.D.); antontseva@bionet.nsc.ru (E.V.A.)

² Department of Natural Sciences, Novosibirsk State University, 630090 Novosibirsk, Russia

* Correspondence: merkulova@bionet.nsc.ru

Abstract: The vast majority of the genetic variants (mainly SNPs) associated with various human traits and diseases map to a noncoding part of the genome and are enriched in its regulatory compartment, suggesting that many causal variants may affect gene expression. The leading mechanism of action of these SNPs consists in the alterations in the transcription factor binding via creation or disruption of transcription factor binding sites (TFBSs) or some change in the affinity of these regulatory proteins to their cognate sites. In this review, we first focus on the history of the discovery of regulatory SNPs (rSNPs) and systematized description of the existing methodical approaches to their study. Then, we brief the recent comprehensive examples of rSNPs studied from the discovery of the changes in the TFBS sequence as a result of a nucleotide substitution to identification of its effect on the target gene expression and, eventually, to phenotype. We also describe state-of-the-art genome-wide approaches to identification of regulatory variants, including both making molecular sense of genome-wide association studies (GWAS) and the alternative approaches the primary goal of which is to determine the functionality of genetic variants. Among these approaches, special attention is paid to expression quantitative trait loci (eQTLs) analysis and the search for allele-specific events in RNA-seq (ASE events) as well as in ChIP-seq, DNase-seq, and ATAC-seq (ASB events) data.

Keywords: regulatory SNPs; transcription factor binding sites; gene expression; gene by gene studies; genome wide approaches



Citation: Degtyareva, A.O.; Antontseva, E.V.; Merkulova, T.I. Regulatory SNPs: Altered Transcription Factor Binding Sites Implicated in Complex Traits and Diseases. *Int. J. Mol. Sci.* **2021**, *22*, 6454. <https://doi.org/10.3390/ijms22126454>

Academic Editor: Gabriel Santpere

Received: 5 May 2021
Accepted: 15 June 2021
Published: 16 June 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

A central goal of human genetics is to understand how genetic variation leads to phenotypic differences and complex diseases. Recently, genome-wide association studies (GWAS) have detected over 70 thousand variants (mainly, single nucleotide polymorphisms, SNPs) associated with various human traits and diseases [1,2]. The vast majority of the genetic variants identified from GWAS map to the noncoding part of the genome and are enriched in regulatory regions (promoters, enhancers, etc.), suggesting that many causal variants may affect gene expression [3–6].

As is known, the regulatory regions of the genome represent clusters of the binding sites for sequence-specific transcription factors (TFs). There, the interplay between these TFs and their binding sites (*cis*-regulatory elements) as well as the interaction of TFs with one another and the coactivator and chromatin remodeling complexes orchestrate the dynamic and diverse genetic programs, thereby determining the tissue-specific gene expression, spatiotemporal specificity of gene activities during development, and the ability of genes to respond to different external signals [7–12]. Thus, thanks to the binding to their specific sites on DNA (transcription factor binding sites, TFBSs), TFs directly interpret the regulatory part of the genome, performing the first step in deciphering the DNA sequence [13–15]. Consequently, regulatory SNPs (rSNPs), that is, genetic variation within TFBSs that alters expression, play a central role in the phenotypic variation in complex traits, including the risk of developing a disease.

Starting from the 1990s, numerous studies have been performed focusing on the non-coding SNPs that perturb the TF binding and are associated with various pathologies. As has been shown, risk alleles can (i) destroy a binding site for a TF [16–19]; (ii) create a binding site for a TF [20–22]; or alter the binding affinities towards an increase [23–25] or a decrease [25–28]. In addition, several cases have been observed when a damage/destruction of a binding site for a TF leads to a concurrent formation of another/other TFBS(s) [19,29,30].

The advent of the NGS technologies gave a strong impetus to the development of functional genomics and application of its methods to the genome-wide search for rSNPs. Currently, various methods of functional genomics are used for both mass interpretation of GWAS data and independent genome-wide identification of regulatory variants. So far, expression quantitative trait locus (eQTL) mapping and identification of allele-specific expression (ASE) events utilizing analysis of RNA-seq data (actually, the largest available genome-wide dataset) are the major relevant methods. The search for allele-specific binding (ASB) events in the data of DNase-seq, ChIP-seq, ATAC-seq (assay for transposase-accessible chromatin with high-throughput sequencing), and so on becomes ever more important. In addition, the approaches not directly associated with obtaining genome-wide data are actively used, including massively parallel reporter assay (MPRA), SNPs-seq, and SNPs-SELEX.

In this review, we brief the history of rSNP discovery, systematize and discuss the methods used in the studies of individual rSNPs, illustrate the narration with the case studies of several best-characterized rSNPs associated with different pathologies, and summarize the recent published data on the genome-wide approaches to the discovery and study of rSNPs.

2. Brief History of rSNP Discovery

The history of the research into the polymorphisms residing in noncoding gene regions and potentially able to influence the level of gene expression commenced as early as the 1990s. The SNPs associated with various pathologies were the main objects in this area. The medical genetic research at that time mainly focused on the variants localized to the gene coding regions [31]. Correspondingly, these studies were rather few [32,33] and the search for noncoding variants was frequently initiated by the absence of any SNPs associated with a disease in the coding part of candidate genes [30,34,35].

In particular, Comings et al. when studying the *TDO2* gene, a candidate gene in psychiatric genetics, failed to find any polymorphisms associated with psychiatric disorders in its coding part [36]. However, such variants were detected in intron 6, where a binding site for the receptor of glucocorticoids, the hormones stimulating *TDO2* expression, had been earlier identified [37]. According to Comings et al., both G→A and G→T substitutions, located 2 bp apart in the middle of intron 6, showed a significant positive association with drug dependence, Tourette syndrome, and attention deficit hyperactivity disorder [36]. Computer analysis emanating from conformational and physicochemical properties of TFBSs [38] predicted that both substitutions damage the binding site for Yin Yang 1 (YY1)—a transcription factor ubiquitously expressed throughout mammalian cells, regulating both transcriptional activation and repression and having a role in 3D chromatin organization [39]. For experimental confirmation the electrophoretic mobility shift assay (EMSA)—the method based on slower migration of protein–DNA complexes than free DNA fragments in gel electrophoresis—was used. EMSA with anti-YY1 antibodies confirmed the predictions made by showing the disappearance of the corresponding band in electropherogram due to preventing YY1 binding to DNA [30]. EMSA with specific antibodies demonstrated that both substitutions damaged the YY1 binding site and concurrently formed the binding sites for other unidentified TFs [30]. In another case, no mutations were detectable in the coding part of the candidate *CFTR* gene of several cystic fibrosis patients; correspondingly, its promoter region was examined and a T to G substitution was found at position –741 bp from the cap site, residing within a potential AP-1 binding site. Competitive EMSA demonstrated a change in the binding pattern of

nuclear proteins resulting from this substitution but did not confirm any presence of AP-1 site [34]. In addition, no mutations in the coding part of the *Gplbβ* candidate gene were found in a patient with Bernard–Soulier syndrome; however, a C to G transversion at position –133 bp was detected in the 5′-upstream region of this gene. This changed a GATA consensus binding site, disrupted GATA1 binding (EMSA + antibody to GATA1), and decreased the promoter activity by 84% (CAT reporter assay) [35].

Other important examples of the rSNPs described at this time include a G to A substitution at –376 bp with respect to the *TNF* transcriptional start site; this substitution causes transcription factor OCT-1 binding (EMSA, ultraviolet crosslinking experiments, and specific antibodies) and alters the gene expression in human monocytes. As has been shown, the OCT-1 binding genotype is associated with a fourfold increased susceptibility to cerebral malaria in West and East African populations [40]; an A/G base transition within the Alu element preceding the *MPO* gene, associated with acute myelocytic leukemias, which creates a strong SP1 binding site (similarity to the consensus and EMSA with purified human SP1) and activates *MPO* transcription (CAT reporter assay) [41]; a G to A substitution detected at position 69 bp downstream of the polyadenylation site of delta-globin gene in a Northern Sardinian family affected by thalassemia. This substitution increases the GATA1 binding (EMSA + antibody to GATA1) and the authors believe it is responsible for a deficient function of the gene in question [42]. See the review by Deplancke et al. [32] for several other relevant examples.

Analysis of these papers demonstrates that the toolkit for rSNP studies was rather poor at that time, with gel shift experiments and transient transfection assays being the main used experimental approaches. As for bioinformatics search for the TFBSs with the structure changed by a nucleotide substitution, the consensus sequences deduced by that time or position weight matrices (PWMs) from TRANSFAC [43] were the main available approaches and an rSNP study in most cases ended in detection of a change in the sequence of a putative site.

However, the situation has radically changed since then. First, it has become clear that rSNPs play a leading role in the phenotypic diversity, in particular, to a considerable degree determining the susceptibility/resistance to diseases and individual sensitivity to various environmental factors, drugs included [4,5,32,44–50]. Second, the methodology for studying individual rSNPs has considerably expanded. Third, the advance in NGS technologies has formed the background for the approaches that allow the regulatory polymorphisms to be searched for on a genome-wide scale.

3. Modern Array of Methods for Studying Individual rSNPs

The two main methods mentioned above dating back to the beginning of the history of rSNP research—EMSA and reporter gene assay—still remain a golden standard in this area and are widely used in the state-of-the-art studies at the first stage of analysis because they allow the presence of a regulatory potential of a nucleotide substitution to be asserted. However, the expressed sets of TFs in different tissues are significantly different [11]; correspondingly, it is most desirable in such experiments to use several cell lines [3,23,51]. It is especially important that the EMSA with specific antibodies or purified TFs is able to reliably identify the TF with its binding site affected by a nucleotide substitution [23,28,30,40,41] and numerous other papers (Table 1). In a similar manner, such TF can be identified, although somewhat less unambiguously, in the reporter assays with cotransfection by the plasmids expressing suspected TFs [52–54].

However, EMSA, the most popular approach, is a strictly in vitro technique. As for verification of an in vivo TF binding to a region, ChIP-PCR is currently used [16,23,26,55,56], as well as its modification, ChIP-AS-qPCR. This modification allows the effect of a nucleotide substitution on a TF binding efficiency in a living cell to be demonstrated [20,24,57–59]. It is noteworthy that the identification of a TF with the help of listed methods requires the prior knowledge about the TFBSs harboring SNPs and this knowledge is usually acquired by bioinformatics analysis of the corresponding DNA sequence. Currently, different

models of TFBSs, specialized databases, and the related tools are widely used for TFBS prediction, functional annotation of sequence variants, and prediction of the SNP impact on TF binding [60–66] and others. Moreover, the closer the result of bioinformatics analysis to the truth, the fewer labor and funds spent on the corresponding experiments. However, the tools currently used for TFBS discovery mainly rely on the recognition model of a traditional PWM [67]; this matrix is based on the hypothesis of additivity of different positions within TFBSs. This leads to considerable oversimplification of the mechanisms underlying the TF–DNA interaction and worsens the TFBS recognition efficiency [66,68–70].

Table 1. Experimental methods for analysis of rSNPs.

| Aim | Method | Advantages | Shortcomings | Comments |
|--|--|--|---|---|
| Registration of the fact of an effect of nucleotide substitution on TF binding | EMSA with nuclear extract (cross-competition assay when necessary) | Simple procedure | In vitro; tissue-specific effects | Testing of several cell lines is desirable |
| Identification of TF the binding site of which is disrupted by a nucleotide substitution | EMSA with purified TF or specific antibody | Unambiguous result | In vitro; requires prior knowledge about TFBS, purified TF, specific antibody | Prescreening in competition assay with unlabeled oligonucleotides may be helpful |
| Confirmation of TF binding in vivo | ChIP-PCR | In vivo | Requires prior knowledge about TFBS and specific antibody | |
| Identification of TF the binding site of which is disrupted by a nucleotide substitution | ChIP-AS-qPCR | In vivo; unambiguous result | Requires prior knowledge about TFBS and specific antibody | Copy number variation must be taken into account when using cell lines |
| Identification of TF the binding site of which is disrupted by a nucleotide substitution | Pull-down assay followed by mass spectrometry analysis | Requires no prior knowledge about TFBS | In vitro | Confirmation by EMSA with purified TF or specific antibody is necessary in some cases |
| Registration of the fact of an effect of nucleotide substitution on the activity of regulatory element | Reporter assays | Simple procedure | Out of genome context | Testing of several cell lines is desirable |
| Registration of the fact of an effect of nucleotide substitution on the activity of regulatory element | CRISPR/Cas9-mediated single nucleotide editing | In genome context | | Testing of several cell lines is desirable |

Correspondingly, unbiased approaches to identification of the TFs with the binding sites affected by a nucleotide substitution are now developed. The most popular of them are (i) proteome-wide analysis based on the interaction of oligonucleotides corresponding to the alternative alleles with metabolically labeled nuclear factors followed by quantitative mass spectrometry [24,26]; (ii) oligonucleotide pull-down assay with subsequent mass spectrometric analysis [71,72]; (iii) mass spectrometric analysis of EMSA protein–DNA complex bands [73]. However, the same peptides can be present in different proteins (especially, in the TFs of the same family); thus, it is necessary to supplement this approach with additional experiments providing more precise data (for example, EMSA with specific antibodies [23,24] or immunoblotting [26]).

The effect of the already identified TFs with their binding sites affected by a nucleotide substitution on the expression of putative target genes is confirmed by siRNA-induced knockdown of the TFs [23,26,59] and/or their overexpression [16,20,26,74]. The same approaches are applied to detect widespread effects of rSNPs at the level of transcriptome [20,75].

In order to clarify whether the intergenic region carrying the target rSNP is potentially regulatory, the data of ENCODE projects [76] are usually assayed for the presence of DNase I hypersensitive sites (DHSs) and ChIP-seq peaks for active histone marks and transcription factors. If ChIP-seq peaks are numerous, this suggests a potential enhancer function, which is further verified by an increase in the luciferase reporter activity with recording of allele-specific effects [23,26,77,78]. To find out which particular genes are influenced by the studied region, it is inactivated using siRNA-mediated transcriptional gene silencing [23] or CRISPR interference (CRISPRi) involving recruitment of a KRAB repressor domain fused to catalytically dead Cas9 [18,79] or just via deletion of this region with CRISPR/Cas9 technology [28,78,80]. Then the transcription levels of the selected genes are assayed [23,28,79,80] or a whole transcriptome analysis is performed [78,80]. A physical interaction between the region carrying an SNP and the potential target genes is usually confirmed with the help of Hi-C methods [26,28,59,80], including allele-specific chromosome conformation capture assays [74], or using available Hi-C data [59,79,81].

CRISPR/Cas9-mediated single nucleotide editing becomes ever more popular when it is necessary to find out a direct effect of single base substitutions on the target gene expression [16,20,58,82,83] as well as widespread transcriptomic changes [80].

Recently, the toolkit for rSNP studies has been supplemented with assessment of the allelic expression imbalance (AEI or ASE) of the transcribed SNP, which either is regarded as a regulatory polymorphism [72,84] or is a marker for the rSNPs located beyond the transcribed genome part [23,55,56]. Either available heterozygous cell lines [23,72,84] or the heterozygous cells generated via CRISPR editing [84] are commonly used. Other recently used options are cells of healthy volunteers [24], biopsy specimens, or samples derived from patients during surgery [53,55,56,84]. This is an important advantage because a studied rSNP in this case exhibits its functionality under the conditions most close to the body's natural context. Blood cells of healthy volunteers [85] and clinical samples [18,28,77,86] are also used to analyze the allele-dependent expression of individual genes by comparing the expression levels observed in the carriers of different genotypes. However, this kind of study requires a considerably larger number of participants as compared with ASE analysis.

4. Recent Comprehensive Examples

Here, we will describe the recent comprehensive examples of rSNPs associated with diseases (Table 2).

Table 2. Examples of rSNPs altering TFBSs and affecting gene expression.

| ID | Location | Risk Allele | TFs with ASB | Genes with ASE | Risk Disease According to GWAS | Ref |
|------------|--|-------------|--|--|---|------|
| rs36115365 | chr5p15.33 intergenic region, putative enhancer | C | ZNF148 (EMSA+AB, EMSA+purified ZNF148) | <i>TERT</i> (ASE, siRNA-mediated knockdown of ZNF148) | Increased pancreatic and testicular cancer risk but a decreased lung cancer and melanoma risk | [23] |
| rs11672691 | Chr19q13.2 Intron 2 of lncRNA <i>PCAT19</i> | G | HOXA2 (ChIP-AS-qPCR) | <i>PCAT19</i> <i>CEACAM21</i> (ASE, HOXA2 knockdown CRISPR/Cas9) | Aggressive prostate cancer | [20] |
| rs2107595 | Chr7p21 noncoding DNA 3' to the <i>HDAC</i> , DHSs | A | E2F3 (ChIP-PCR) | <i>HDAC9</i> (ASE) | Atherosclerosis, coronary artery disease, stroke | [26] |
| rs12411216 | Chr1q22 DHSs | A | E2F4 (EMSA+AB) | <i>GBA</i> (ASE, CRISPR/Cas9) | Parkinson's disease, cognitive damage | [28] |
| rs13239597 | Chr7q32.1 <i>TNPO3</i> promoter | A | EVI1 (ChIP-AS-qPCR) | <i>IRF5</i> (ASE, shRNA-mediated knockdown of EVI1) | Systemic lupus erythematosus and systemic sclerosis | [59] |
| rs17079281 | Chr6q22.2 <i>DCBLD1</i> promoter | C | YY1 (ChIP-qPCR) | <i>DCBLD1</i> (ASE, CRISPR/Cas9) | Lung cancer | [16] |

Notes: allele-specific binding (ASB), allele-specific expression (ASE), transcription factors (TFs), DNase I hypersensitive sites (DHSs).

4.1. Allele C of rs36115365 from chr5p15.33 Multi-Cancer Risk Locus Enhances ZNF148 Binding and Telomerase Reverse Transcriptase (*TERT*) Expression

The rs36115365 polymorphism is one of the nine highly correlated SNPs residing in chr5p15.33 region 2 of the GWAS mapped multi-cancer risk locus. The performed EMSA functional analysis of these nine SNPs has shown that only the rs36115365 polymorphism displays the changes in protein binding pattern in EMSA with nuclear extracts of eight human cell lines [23]. (Note that its minor C allele of which is associated with an increased pancreatic and testicular cancer risk but a decreased lung cancer and melanoma risk.) This polymorphism is located in the region between the 5' end of *TERT* (~18 kb upstream) and 3' end of *CLPTM1L* (~5 kb downstream) genes. According to ENCODE data, this region is a putative enhancer since it overlaps with the multiple active histone modifications and TF ChIP-seq peaks. When transfecting the same eight cell lines, a 240-bp fragment carrying rs36115365 displayed an increase in luciferase reporter activity. In addition, allele C exhibited both preferred protein binding in EMSA and enhanced regulatory activity in reporter assay [23].

In order to clarify which of the neighboring genes are affected by the found enhancer, it was inactivated using siRNA-mediated transcriptional gene silencing [87]. This decreased the expression of *TERT* gene alone. The product of this gene, telomerase reverse transcriptase, in combination with an RNA template adds nucleotide repeats to chromosome ends,

which is important to viability of cancer cells. A higher level of *TERT* expression from allele C was demonstrated using a marker SNP in the transcribed *TERT* gene part [23].

The TF the binding site of which changes as a result of a G to C substitution (rs36115365) was identified using the pull-down of nuclear proteins with the oligonucleotides corresponding to G or C allele followed by quantitative mass spectrometry [88]. Since four proteins that preferred allele C (ZNF148, VEZF1/ZNF161, ZNF281, and ZNF740) were detected, their binding was tested with EMSA: only the antibody to ZNF148 consistently caused a loss in the C allele binding. Moreover, the EMSA with recombinant purified ZNF148 confirmed specific binding to the C allele. Finally, siRNA-mediated knockdown of ZNF148 mRNA resulted in reduced *TERT* expression, telomerase activity, and telomere length. Thus, the C allele improves ZNF148 binding site, which elevates the *TERT* expression level and, as a consequence, increases the risk of multiple cancer types [23].

4.2. Allele G of rs11672691 from Chr19q13.2, Associated with Aggressive Prostate Cancer, Creates a HOXA2 Binding Site and Raises the Transcription Levels of PCAT19 and CEACAM21 Genes, Implicated in Prostate Cancer Cell Growth and Tumor Progression

The G allele of rs11672691 was identified by GWAS and additionally confirmed as being associated with aggressive prostate cancer by meta-analysis and genotyping of cancer cases and controls from 26 studies from European populations [20,89]. An eQTL analysis of The Cancer Genome Atlas (TCGA; Cancer Genome Atlas Research Network [90]) data, comprising about 1000 prostate tissue sample, has shown that the presence of allele G correlates with elevated transcription levels of the *PCAT19* and *CEACAM21* genes, both involved in the prostate cancer cell growth and tumor progression [20].

The rs11672691 polymorphism resides in the intron 2 of long noncoding RNA (lncRNA) *PCAT19* and is 100 kb away from *CEACAM21*. The region housing this polymorphism is enriched in active enhancer marks, contains several TF peaks (ENCODE), and exhibits an enhancer activity in the luciferase reporter assay [20]. As is demonstrated using PWM, the rs11672691 polymorphism maps within the binding motifs of homeodomain transcription factors, including NKX3.1 and *HOXA2*; further, ChIP-AS-qPCR has shown that *HOXA2* prevalently binds to allele G. *HOXA2* knockdown decreases both *PCAT19* and *CEACAM21* expression. CRISPR/Cas9-mediated introduction of single nucleotide mutation was used to directly demonstrate that the presence of allele G led to higher levels of *PCAT19* and *CEACAM21* transcripts [91]; the genotype of rs11672691 was successfully converted from G/A to G/G or A/A in prostate cancer cell line 22Rv1. A comparison of the mutated and parental cells suggested that the G/G genotype was associated with higher transcriptional levels of *PCAT19* and *CEACAM21* as compared with the G/A and A/A genotypes; note that the transcriptional levels of these genes were the lowest for the A/A variant [20].

4.3. Atherosclerosis Risk Variant A of rs2107595 from Chr7p21.1 Interferes with E2F3 in Putative Enhancer Region, Which Leads to HDAC9 Activation

The rs2107595 polymorphism was identified by recent GWAS as the lead SNP for stroke and coronary artery disease (CAD) [92,93]. There are also numerous data indicating its involvement in the control of systolic blood pressure [94–96]. It resides in noncoding DNA 3' to the *HDAC9* gene in a region overlapping with DHSs and the histone activation marks H3K27ac and H3K4me1 (ENCODE). The search for the TF with ASB at rs2107595 commenced with the proteome-wide analysis of the interaction between the oligonucleotides that carried either risk (A) or normal (G) allele with labeled nuclear factors and subsequent quantitative mass spectrometry [88]. All constituents of the E2F3/TFDP1 (transcription factor Dp-1)/Rb1 complex were identified among the factors that prevalently bound to the G-centered oligonucleotide. The subsequent oligonucleotide pull-down assay followed by immunoblotting confirmed enriched binding of E2F3 to the common allele [26]. These findings agreed well with the presence of an E2F3 consensus sequence in the allele G region disrupted by the risk allele A [62]. ChIP-PCR in the HeLa cells homozygous for the G allele confirmed E2F3 binding in vivo. As was demonstrated, the rs2107595 risk A allele displayed a higher transcriptional capacity in luciferase assays as compared

with common G allele and caused an increase in the *HDAC9* mRNA in genome-edited Jurkat cells [26]. Analysis of the allele-dependent expression of *HDAC9* in peripheral blood mononuclear cells of healthy donors also demonstrated increased mRNA levels of *HDAC9* only in risk allele carriers [85]. Since rs2107595 is located at a considerable distance from *HDAC9*, Prestel et al. [26] performed circularized chromosome conformation capture experiment [97] and discovered a physical interaction between rs2107595 and the *HDAC9* promoter in common allele (GG) but not in the risk allele cells (AA). This demonstrates the role of E2F3 in allele specific differences in the chromatin organization. These results suggest that an elevated *HDAC9* expression is involved in the etiology of stroke and CAD and that *HDAC9* targeted inhibition is one of the strategies to prevent atherosclerosis although the mechanism underlying a promoting effect of increased *HDAC9* expression on atherogenesis and vascular risk is vague [26].

4.4. Allele A of rs12411216 from Chr1q22 Decreases E2F4 Binding, Which Results in a Decreased GBA Expression and an Increased Cognitive Damage in Parkinson's Disease

As is known, a decrease in the glucocerebrosidase (*GBA*) gene expression in the brain promotes a prion-like spread of α -Syn interpolymer complexes and progression of Parkinson's disease (PD) as well as increases the cognitive damage [98,99]. Jiang et al. [28] were the first to identify the rs12411216 polymorphism as an eQTL that influences the *GBA* gene expression. Genotyping of 122 PD patients with mild cognitive impairment (PD-MCI) and 184 PD patients who had PD but no cognitive impairment suggested a statistically significant correlation between the A allele and PD-MCI. In addition, the *GBA* mRNA level was significantly decreased in clinical samples of the patients with AA genotype as compared with the patients with CC. The rs12411216 polymorphism is located at a distance of ~50 kb from the *GBA* transcriptional start site and falls into one of the DHSs that contacts the *GBA* promoter as is suggested by Hi-C data. According to the TFBS database (wgEncodeRegTfbsClustered), rs12411216 overlaps with the core motif of E2F4 TF, which was confirmed by EMSA with specific antibodies. EMSA has also shown a drastic decrease in the E2F4 binding in the case of the risk A allele. A CRISPR/Cas9-mediated deletion of the DHS housing rs12411216 decreased *GBA* expression, weakened the enzyme activity, and enhanced an abnormal aggregation of α -Syn in SH-SY5Y cells [28]. Interestingly, a little earlier a significant association of rs12411216 with occipital lobe volume in the European ancestry-only meta-analysis was found [100]. The authors also showed that the most significant genetic correlation with brain lobar volume and diseases was observed between occipital lobe volume and Parkinson's disease ($rg = 0.18, p = 0.03$). However, this finding was not significant after multiple testing correction, which makes the authors consider it as a preliminary result [100].

4.5. Allele A of rs13239597, Associated with Two Systemic Autoimmune Diseases, Enhances the Binding of EVI1, Which Promotes Formation of a Long-Range Chromatin Loop and an Increased Expression of IRF5, Located 118 kb Away

GWAS identified genetic variants conferring the risks of autoimmune diseases systemic lupus erythematosus (SLE) and systemic sclerosis (SSc) at 7q32.1, harboring *IRF5* and *TNPO3* genes. The rs13239597 polymorphism is located in the *TNPO3* promoter 118 kb away from *IRF5*. According to eQTL analysis involving 373 unrelated European samples of lymphoblastoid cell lines, the minor allele A of rs13239597 was significantly associated with an increased *IRF5* expression; this was also confirmed with GTEx data [6]. On the other hand, any significant association between rs13239597 and *TNPO3* was unobservable. Analysis of the available Hi-C data demonstrates that *IRF5* is among the 12 genes that interact with rs13239597. A motif analysis predicted four potential TFs with ASB affinity to rs13239597, namely, EVI1, ERF, GATA1, and TAL1. In order to find out which particular TF influences the *IRF5* expression, their shRNA-mediated knockdown in U2OS cell line was performed. A significant decline in the *IRF5* expression was detected only in EVI1 knockdown U2OS cells. Moreover, a 3C assay showed that EVI1 knockdown significantly decreased the interaction between rs13239597 and *IRF5* promoter. Then, ChIP-AS-qPCR

demonstrated that EVI1 was preferentially recruited to the rs13239597 A allele as compared with its C allele.

Finally, analysis of three SLE genome-wide gene expression datasets revealed a significantly higher *IRF5* expression in the SLE patients as compared with healthy subjects [59].

4.6. Allele T of rs17079281 Decreases Lung Cancer Risk through Creating an YY1 Binding Site to Suppress Proto-Oncogene *DCBLD1* Expression

According to GWAS, rs9387478 in 6q22.2 is associated with lung cancer risk in both Asian [101] and European populations [102]. Linkage disequilibrium (LD) analysis, meta-analysis involving 4403 cases and 5336 controls, and two additional case–control studies have discovered a novel SNP, rs17079281, in the *DCBLD1* promoter, which is associated with lung cancer risk in Chinese populations [16]. As is shown, the patients with T allele have a lower risk of adenocarcinoma as compared with the carriers of C allele (adjusted OR = 0.86; 95% CI: 0.80–0.92) and that the subjects with the C/T or T/T genotype have lower levels of *DCBLD1* expression than those with C/C genotype in lung adenocarcinoma tissues [16]. According to TRANSFAC data [43], a C→T substitution in this region may create an YY1 binding site. This is confirmed with the help of ChIP-qPCR analysis in wild-type Beas2B cells (C/T at rs17079281) and the CRISPR/Cas9 modified cells with C/C knockin, which demonstrated that T allele was necessary for binding YY1. Transfection of these lines with a plasmid expressing this TF showed a decrease in the *DCBLD1* expression only in wild-type Beas2B cells. Thus, the YY1 transcription repressor has a higher binding affinity for the T allele of rs17079281, which results in suppression of *DCBLD1* proto-oncogene expression and, consequently, in a decreased lung adenocarcinoma risk [16].

5. rSNPs on a Genome-Wide Scale

Genome-wide approaches to the search for rSNPs fall into two large groups. The first group comprises GWAS mass data analysis utilizing manifold methods of functional genomics, while the second group uses the same methods but independently without any prior knowledge about trait associations (Figure 1, Table 3). The latter group includes eQTL analysis, identification of allele-specific events, and some other genome-wide approaches. As for the rSNPs discovered by the approaches of the second group, it is necessary to additionally determine their association with a certain trait (most frequently, via comparison with GWAS data or by analysis of rSNPs as an eQTL in transcriptome data and reconstruction of the gene networks and molecular pathways).

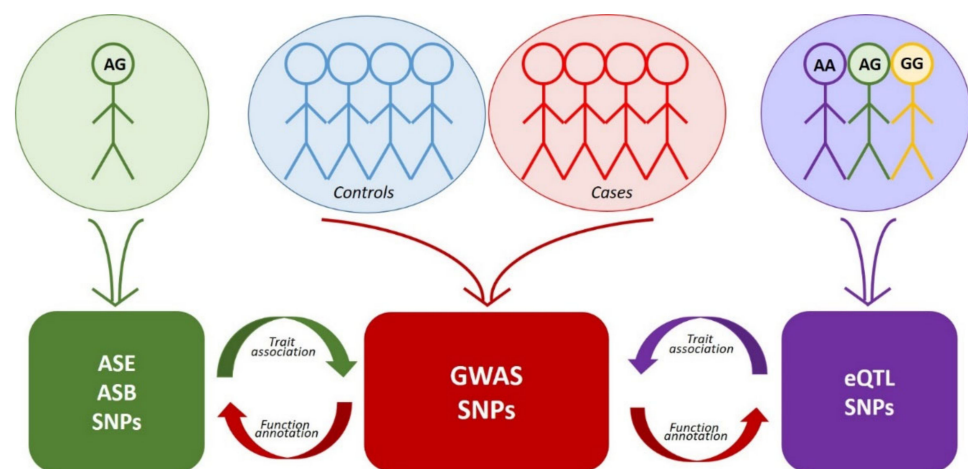


Figure 1. Interplay between the approaches to the search for functional SNPs. Colored blocks denote arrays of corresponding data. Red arrows indicate functional annotation of GWAS data using eQTL, ASE or ASB analysis. Purple arrow indicates the search for association of eQTL SNPs with traits via comparison with GWAS data. Green arrow—the same for SNPs detected by ASE or ASB analysis.

Table 3. Main features of the most widespread approaches to the search for functional SNPs.

| | Approach | GWAS | eQTL Analysis | ASE | ASB |
|---|-----------------------------------|--|-----------------------------------|-----|-----|
| 1 | Initial association with trait | + | – | – | – |
| 2 | Initial association with function | – | + | + | + |
| 3 | Causal or in LD | Both | + | ++ | +++ |
| 4 | Number of participants | Tens and hundreds of thousands (large cohorts) | Hundreds (modestly sized cohorts) | Few | Few |

In row 3, +/++/+++ shows an increase in the bias towards causal.

5.1. Making Molecular Sense of GWAS

Historically, GWAS is the first genome-wide approach to identification of the genetic variants (mainly SNPs) associated with traits. Having appeared in the mid-2000s, GWAS have so far detected over 70 thousand loci associated with various human traits and diseases [1,2]. However, this technology is unable to give any information about the functionality of discovered variants, making it very difficult to translate GWAS data into biological insights, which is necessary to reveal the molecular mechanisms underlying diseases [25,103]. In addition, GWAS cannot distinguish between causal polymorphisms and numerous marker SNPs detected due to LD. Thus, considerable efforts have been recently focused on the subsequent functional analysis of the SNPs with disease/trait associations revealed with the help of GWAS. Both individual SNPs (mapped by GWAS and according to LD) [16,18,20,23,26,53,74,78,82,104] and others and large arrays of polymorphisms are analyzed in this way; manifold methods of the state-of-the-art functional genomics are used for this purpose.

One of the approaches in functional genomics frequently applied to mass functional interpretation of the SNPs detected by GWAS is MPRA in different variants [24,79,84,105,106]. MPRA is an upscaled version of gene reporter assay allowing the effect of an allele on the expression of a reporter construct to be determined with concurrent testing of several hundred to several thousand DNA fragments [107]. In particular, this method was used to test 1605 SNPs, 35 of which were associated with osteoarthritis in Europeans via GWAS and the remaining ones were in LD with them [84]. Six of these polymorphisms displayed differential regulatory activity between the major and minor alleles in the STARR-seq MPRA in Saos-2 osteosarcoma cell line and for three of them, this activity was confirmed by conventional luciferase reporter system. A more detailed study of the most significant SNP, rs4730222, showed differential nuclear protein binding in EMSA as well as the effect of alleles on the expression level of HBP1 isoform, transcribed from an alternative promoter containing rs4730222 at position +80 bp relative to its transcriptional start site [84]. Analogously, MPRA was used to study 832 variants associated with melanoma risk; 30 of them displayed significant difference between two alleles in UACC903 melanoma cells [24]. The rs398206 polymorphism, located in the first intron of *MX2* gene, was studied in detail; a most pronounced allelic difference was observed. As was shown, the risk-associated A allele significantly increases the YY1 binding to a DNA region carrying rs398206 in vitro (EMSA) and in vivo (ChIP-AS-qPCR), leading to an increase in *MX2* expression, which accelerates melanoma formation [24]. Analysis of these data suggests that the number of MPRA-revealed rSNPs is relatively small (0.1–4.7% of the tested GWAS variants). Perhaps, this is explainable in part with the use of only one cell line in each case. Due to tissue-specific effects, the use of several cell lines could increase the number of SNPs displaying significant difference between two alleles.

Similar to MPRA as an upscaled variant of gene reporter assay, an upscaled EMSA variant—Reel-seq (Regulatory element-sequencing)—was designed [25]. For this approach, a sequence library containing disease-associated SNP constructs with both the risk and non-

risk alleles were generated by massive parallel oligonucleotide synthesis. After binding to nuclear proteins from MDA-MB-468 cell line and several EMSA rounds, the SNPs that demonstrated allele-imbalanced gel shift pattern between the risk and non-risk allele rSNPs were selected from this library [25]. Thus, 521 (12%) potential rSNPs were selected out of 4316 breast cancer-associated (GWAS) SNPs. Allele-specific effects were confirmed for 12 of the selected polymorphisms by conventional EMSA and luciferase assay. For three SNPs from breast cancer-associated *FGFR2* locus, the TFs with the binding altered as a result of an SNP were identified using the approach devised by the authors: SNP-specific DNA competition pulldown-mass spectrometry (SDCP-MS) and allele-imbalanced DNA pulldown-Western blot (AIDP-Wb). Thus, the authors succeeded in demonstrating that the TFs PARP-2 and TFAM bound to rs7895676 with less binding of risk allele C; TFs TEAD1 and TEAD3 bound to rs2981578 with more binding of risk allele G; and NFIB bound to rs2981584 with less binding of risk allele G.

In its essence, Reel-seq is similar to the earlier described SNPs-seq [108], which differs from it only by the method used to distinguish between the protein-bound DNA oligonucleotides from free oligonucleotides (for this purpose, a protein purification column is used in SNPs-seq). SNPs-seq has been used to study allele-dependent protein binding at 903 SNPs identified by GWAS as variants that increase prostate cancer risk. Using the nuclear extract of prostate adenocarcinoma cell line LNCaP, 403 SNPs (45%) that showed protein-binding differences (>1.5-fold) between the reference and variant alleles were found. Of interest is that the rate (percentage) of the detected functional SNPs in GWAS data using Reel-seq and SNPs-seq methods was by an order of magnitude higher as compared with MPRA despite that only one cell line was used. This fact is explainable with a very high regulatory potential of naked DNA, discovered in our earlier studies on computational recognition of TFBS and experimental verification of the predicted sites by EMSA [109]. In reporter studies, several factors can conceal this effect, such as the TFBS position relative to promoter and the need in target TF interaction with both close and remote partner TFs.

The year of 2021 brought about another high-throughput method for assessment of the direct effect of a nucleotide substitution on TF binding—SNP-SELEX—an ultra-high-throughput multiplex protein–DNA binding assay [66]. SNP-SELEX utilizes a library of 40-bp DNA sequences matching the reference human genomic sequence in which the tested SNP permuted to all four bases located in the center. In the Yan et al. [66] study, the library consisted of 383544 distinct oligonucleotides corresponding to 95886 SNPs, including those linked to T2D susceptibility via GWAS and those located in putative *cis*-regulatory sequences 500 kb of T2D-tagging SNPs. Using 270 recombinant human TFs, the authors performed 828 million measurements of transcription factor–DNA interactions and succeeded in discovering 11079 SNPs (11.5% of the analyzed ones) that exhibited significantly differential binding to at least one TF [66].

Functional genomics data available in the current databases are also widely used in a high-throughput interpretation of GWAS data. In particular, Li et al. [110] used the ENCODE ChIP-seq data for 34 TFs obtained using human brain tissues or neuronal cells and bioinformatics search (using PWM) for TFBSs in ChIP-seq peaks. Analysis of the 8005 SNPs (including 40 index SNPs and those that were in LD with index SNPs) associated with major depressive disorder (MDD) [111] detected 34 MDD risk SNPs that disrupted the binding sites for 15 TFs. The allelic effect on reporter gene expression was confirmed for 29 of the analyzed SNPs, one of these polymorphisms, rs3101339, appeared to be associated with the *NEGR1* gene in qQTL analysis as well as affected its expression in the experiments on knockout of the region containing rs2050033 using CRISPR-Cas9-mediated genome editing. This suggests that rs3101339 may confer MDD risk by affecting *NEGR1* expression [110]. The ChIP-seq datasets for various histone modifications were used to construct a comprehensive list of super enhancers in T2D [112] and CAD [113]. The rVarBase [114] was used in further functional annotation of super enhancer SNPs. This gave 286 T2D- and 366 CAD-associated super enhancer SNPs, part of which was annotated as being involved the regulation of chromatin structure and in the effects on TF

binding [112,113]. Similarly, the own ChIP-seq data for the histone modifications marking the active regulatory elements of the genome were used for analyzing the GWAS SNPs associated with risk of epithelial ovarian cancer [115]. H3K27Ac ChIP-seq were generated for 26 ovarian cancer and precursor-related cell and tissue types and in combination with motifbreakR tool allowed for the discovery of 469 candidate causal risk variants in H3K27Ac peaks that were predicted to significantly break TF binding motifs [115].

There are many other examples of a global functional interpretation of the SNPs from the GWAS Catalog with the use of the data on TF-based motifs, promoters, enhancers, chromatin accessibility landscapes, three-dimensional chromatin interactions, and, especially, eQTL analysis [19,116–122].

5.2. eQTL Analysis

eQTL mapping is used to identify the association of a genetic variant with gene expression level based on transcriptome analysis. The term eQTL either means the presence of such association between a variant (eVariant) and the expression level(s) of gene/genes (eGene/eGenes) [6,123,124] or refers to the variant itself that displays such association [28,125–128]. When searching for an eQTL, differential gene expression in the transcriptomes of the subjects with different genotypes is determined for each SNP. Unlike GWAS, requiring tens and hundreds of thousands of participants, eQTL mapping requires just several hundreds of samples [6,124,125]. Transcriptome data alone are sufficient to detect the eVariants located in the transcribed region [129], while detection of all eQTLs also demands whole genome sequencing data [124].

Initially, eQTL analysis was conducted using microarrays and later, RNA-seq was used. Note that the very first studies demonstrated that eQTL effects considerably varied between the examined cell types or tissues (see review [130] for numerous examples). That is why the Genotype-Tissue Expression (GTEx) project was initiated in 2010 with the goal to create the catalog of events of the effects of genetic variants on gene expression determined in the maximally possible number of human tissues. The goal was to eventually detect the association of such variants with complex diseases and traits and to get a deeper insight into the molecular mechanisms underlying their action [131].

Currently, the GTEx Consortium has at its disposal the results of analysis of 15,201 RNA-seq samples from 49 tissues of 838 postmortem donors. In total, 4,278,636 genetic variants (cis-eQTLs) associated with a change in expression level of 18,262 protein coding and 5006 lincRNA genes have been found; each of them manifested itself at least in one tissue. All this suggests the presence of regulatory associations for almost all genes in the human genome. The genes lacking a cis-eQTL have emerged to be mainly the genes that lack any expression in the analyzed tissues, in particular, the genes that are active only in the early development. In addition, the genome regulatory regions and GWAS loci have been shown enriched in eQTLs [124]. However, detection of the causal variants remains challenging in both eQTL analysis and GWAS because of the presence of multiple variants in LD [132,133].

At present, eQTL analysis is actively used to identify trait-associated genes, especially susceptibility genes from GWAS loci, since this informs on the genes for which expression levels correlate with trait-associated variants [134–143]. Although eQTL analysis is not initially associated with any prior knowledge about traits, the data on differential gene expression in the individuals with different genotypes obtained with this method are also helpful in detection of the functional associations between these genes and construction of gene networks. This makes it possible to determine putative phenotypic outcomes for at least part of the detected eSNPs [71,134,135,144–146].

However, many eQTLs (eVariants) map to genome regulatory regions; correspondingly, the results of eQTL analysis are frequently the starting point in identification of rSNPs. In particular, rs12411216, the A allele of which decreases E2F4 binding and *GBA* gene transcription and thus increases the cognitive damage in PD, was first identified as an eQTL affecting *GBA* gene expression [28]. In addition, eQTL analysis has been used for

prioritization of rs13239597, linked to lupus systemic erythematosus and systemic sclerosis via GWAS. Further studies of rs13239597 demonstrated that its risk A allele increased EVI1 binding and acted as an allele-specific enhancer regulating *IRF5* expression [59]. Both examples are detailed in Section 4 of this review. In a similar manner, the rs10085588 polymorphism, associated with bone mineral density and osteoporosis, was initially identified as an eQTL for *SLC25A13*. Its minor allele A displayed a decreased gene expression both in vivo in human primary osteoblasts and in vitro in luciferase reporter assay [86]. In addition, eQTL analysis detected six potential functional SNPs (rs9533090, rs9594738, rs8001611, rs9533094, rs9533095, and rs9594759) exclusively correlated with the *RANKL* gene expression. They all belong to the group of multiple intergenic SNPs located over 100 kb upstream of the *RANKL* gene, associated with osteoporosis via GWAS. Later, one of these polymorphisms, rs9533090, was identified as an allele-specific regulatory SNP. The variant of the C allele resulted in the binding of TF NFIC, which led to the activation of enhancer and an increase in the expression of *RANKL*, a key regulator of bone metabolism [81]. Manifold methods of functional genomics, such as MPRA and identification of active chromatin modifications and open chromatin regions, are used in the mass search for rSNPs among eQTLs variants [147–149].

5.3. Allele-Specific Expression (ASE) Analysis

RNA-seq technology gives a brilliant opportunity of quantifying the expression of two alleles of any polymorphic sites in a diploid individual and of detecting allelic imbalance of transcription or an ASE event. In turn, ASE mapping is a useful instrument making it possible to identify variations in gene expression underlying phenotypic differences among individuals [126,133,150].

Typically, ASE events are detectable by joint analysis of transcriptome data and the WGS (whole genome sequencing) data for the same individuals. For example, Kang et al. [151] used RNA-seq data for lymphoblastoid cell lines derived from 77 unrelated European subjects (their genomic data are available through the 1000 Genomes Project) and discovered 2309 SNPs associated with ASE patterns. These SNPs were enriched in promoter regions and 108 of them had been earlier associated with human immune diseases [151]. Liu et al. [152] utilized RNA-seq and WGS data from a single cancer sample for each of the 13 pediatric T-lineage acute lymphoblastic leukemias (T-ALLs) and found dozens of somatic noncoding regulatory variants able to cause cis-activation of 222 candidate genes. These variants comprised both known noncoding mutations activating T-ALL oncogenes (*TAL1/2*, *LMO1/2*, and *TLX3*) and the new ones, including a C to T substitution in the *TAL1* intron 1, which created an YY1 binding site and, as a consequence, activated an enhancer residing in the same region [152]. RNA-seq and WGS data from the GTEx v8 release [124] allowed Castel et al. [150] to generate an ASE resource containing in total 431 million ASE events at an SNP level and 153 million measurements at a haplotype level. However, when genotype information is not available, it could be derived from RNA-seq reads directly via their sophisticated allele-specific analysis [46,144,153].

One of the main advantages of ASE approaches as compared with eQTL analysis and the others, GWAS consists in that both eQTL and GWAS rely on the analysis of numerous samples from the subjects with different genetic backgrounds and conditions of individual life. As for ASE approach, it relies on comparison of allelic effects within subjects and thus controls genetic background and cell environment; this allows the sample to be significantly reduced even to a single individual [153]. Thus, ASE approach gives the opportunity to considerably increase the number of temporal and environmental conditions that can be analyzed in parallel, thereby providing unique possibilities, first and foremost, for the studies in pharmacogenetics and pharmacogenomics. A perfect example here is the study by Moyerbrailean et al. [46] on detection of allele-specific effects of 50 substances (steroid and peptide hormones, metal ions, dietary components, common drugs, and environmental contaminants) using five types of cells (LCLs, PBMCs, HUVECs, SMCs, and melanocytes) each derived from three individuals. Analysis of transcriptome data allowed the authors to

discover 1455 genes with ASE events and to identify 215 genes with gene-by-environment (G×E) interactions [46]. In a similar manner, condition-dependent ASE events in 19 genes related to the inflammatory response were detected via RNA-seq of primary white blood cells from eight human subjects before and after LPS treatment [154]. In addition, M0 and M1 macrophage states were compared using the samples of 48 healthy subjects. This gave 408 and 334 unique ASE events in M0 and M1 state, respectively, while 1280 genes showed evidence of ASE under both conditions [126].

Gutierrez-Arcelus et al. [83] studied the enrichment dynamics of the alleles of heterozygous SNPs in transcriptomes during development of the response to an external stimulus. The authors analyzed their RNA-seq data for eight time points (0, 2, 4, 8, 12, 24, 48, and 72 h) during memory CD4⁺ T cell (from 24 genotyped individuals) activation by anti-CD3/CD28 beads. The result was 561 dynamic ASE events where the reference and alternative alleles demonstrated different patterns in time, including 182 dynASE events in MHC locus and 15 events in *HLA-DQB1*. Using CRISPR/Cas9 editing in the HLA class II expressing T cell line (HH), they demonstrated that the allele G of rs71542466, located 39 bp upstream of the *HLA-DQB1* transcription start site, increased its expression [83].

5.4. Allele-Specific Binding (ASB) Analysis

Although ASE analysis is most efficient for identifying gene expression variations, it yet fails to answer the question on whether the observed effect is a direct result of a nucleotide substitution in the SNPs used to measure ASE or simply captures the effects of other cis-acting variation. Our present consensus is that most of the disease-associated SNPs are located in regulatory regions [5]; there, they can lead to ASB of TFs with subsequent differential expression of the target gene alleles. Several studies focused on a genome-wide detection of ASB events have been so far completed. This direction commences from the pioneering research aimed at identification of the sequence variants that influence TF occupancy in the accessible chromatin (DNase-seq) [155] and TF ChIP-seq data [156].

Maurano et al. [155] have analyzed 493 high-resolution DNase-seq profiles (both published and acquired by the authors) from diverse cultured primary cells, cultured multipotent and pluripotent progenitor cells, and fetal tissues of 166 individuals and 114 cell types. In total, they succeeded in detecting 64,599 SNPs that displayed allelic imbalance in chromatin accessibility. Using PWMs for 2203 TF motifs from TRANSFAC [157], JASPAR [158], UniPROBE [159], and a published SELEX dataset [160], the authors demonstrated that the majority of the found SNPs are able to directly influence the TF occupancy and, as a consequence, to change regulatory DNA accessibility in vivo [155].

A direct genome-wide search for asymmetric TF binding events with the help of public ChIP-seq data was for the first time performed in Claes Wadelius's laboratory. The authors have analyzed the TF ChIP-seq data available at the time of download for four cell lines—GM12878 (B cells), H1-hESC, K562, and SK-N-SH from ENCODE project—and discovered 9962 SNPs with biased TF allele binding. Their computations suggested that the most common polymorphisms could be tested for ASB via repeated ChIP-seq experiments with 20 selected TFs in 3–10 individuals [156]. Further analysis of their data showed that 141 of the detected AS-SNPs emerged to be associated with different GWAS traits (15 were listed in the GWAS catalog and 126 fell in a high-LD interval); 84 AS-SNPs detected in B cells coincided with eSNPs for B cells [161]; an additional 362 AS-SNPs were in LD with an eSNP [156]. The same approach allowed for detection of 3713 SNPs displaying significant difference in the binding between alleles in HepG2 and HeLa-S3 cell lines. The dual luciferase reporter assay of 39 of them confirmed ASE of 27 [162]. A comprehensive functional analysis of the rs953413 polymorphism, identified as an AS-SNP in human liver HepG2 cells and located in an evolutionarily conserved enhancer element in the first intron of *ELOVL2* gene [162], has shown that the A allele disrupts a FOXA binding site. This decreases the binding not only of this TF, but also of HNF4 α , which cooperatively interacts with it; correspondingly, this leads to a decrease in *ELOVL2* expression and impaired

hepatic docosahexaenoic acid synthesis, which may play a role in the pathogenesis of nonalcoholic fatty liver disease [58].

Recently, ATAC-seq [163] was applied in a genome-wide search for ASB events concurrently for all TFs functioning in the cell. ATAC-seq makes it possible not only to detect genomic footprints left by DNA-binding proteins, but also to determine the allelic bias in the binding of these proteins [164]. Using ATAC-seq, the authors succeeded in detecting 53 rSNPs in human MCF-7 breast cancer cells and 125 rSNPs in human mesenchymal stem cells (MSCs). Using their own RNA-seq data and publicly available chromatin interaction data for MCF-7 cells, they demonstrated that the detected 53 rSNPs were associated with 74 potential target genes. A comparison of rSNPs with the eQTLs from GTEx Project database demonstrates that 30% of the rSNPs from MCF-7 and 43% from MSC fall into eQTL regions, suggesting their role in the allelic differences in gene expression [164].

A genome-wide search for ASB events for the binding sites of an individual TF has been also performed. In particular, such search was conducted for NKX2-5, a cardiac-specific TF [165]; according to GWAS data, this TF can be regarded as a candidate gene associated with EKG phenotypes [166]. For this purpose, 15 ChIP-seq experiments with anti-NKX2-5 antibodies were performed in pluripotent stem cell-derived cardiomyocytes from seven related individuals. As a result, about 2000 SNPs with allele-specific effects on NKX2-5 binding were discovered; they were enriched for altered TF motifs, heart-specific eQTLs, and EKG GWAS signals [165]. Allele-specific effects of two of these SNPs (rs3807989 and rs590041) were confirmed by EMSA, luciferase assay, and analysis of their effect on target gene expression [165].

The ChIP-seq data for whole-genome histone modification profiles are also helpful in the search for ASB events on a genome-wide scale; these profiles characterize the energy landscape of chromatin, whereto the TF binding with regulatory regions considerably contributes [167–170]. In particular, this approach was implemented when searching for ASB events in K562, MCF-7, and HCT-116 human cell lines by analyzing the ENCODE ChIP-Seq data for histone epigenetic modifications (H3K27ac, H3K4me1, H3K4me2, H3K4me3, and H3K27me3) and 456 different chromatin-associated proteins, mainly transcriptional factors [171]. ASE events were also assessed in HCT-116, MCF-7, and K562 cells (ENCODE) using RNA-Seq and ChIA-PET data with an RNA pol II antibody. This allowed for detection of 1633 rSNPs simultaneously associated with both types of allele-specific events. According to GWAS data, 27 of them were associated with a risk of malignancy [171] and 14 with cognitive disorders [172]. In addition, an association with colorectal cancer (CRC) was suggested for 30 rSNPs based on a comparison of allele frequencies in the ICGC cohort [173] with the MAFs reported by dbSNP [171]. Genotyping of CRC patients and healthy controls according to six of these polymorphisms demonstrated that rs590352 of *ATXN7L3B* gene was associated with CRC in men and rs4796672 of *KRT15* gene, with CRC in women. In addition, the analysis of haplotypes shows that rs2072580, located in the promoter region common for the *ISCU* and *SART3* genes, can be also associated with CRC [174].

The allele-specific signals were also searched for in the ChIP-seq data for histone modifications in a study aimed to identify the regulatory variants involved in the development mechanisms of immune and B-cell related diseases [175]. The SNPs with allele-specific behavior in the available ChIP-seq datasets produced for the histone modifications defining promoters (H3K4me3) and enhancers (H3K4me1 and H3K27ac) and for domain boundary proteins (CTCF and SA.1) were the focus of this study. Thus, 17293 such SNPs (AS-SNPs) were found in seven lymphoblastoid cell lines; of them, 237 were associated with immune GWAS traits and 714 with gene expression in B cells.

6. Conclusions

Gene expression programs underlying development, differentiation, and environmental responses are guided by the regulatory DNA portion of the metazoan genomes. The corresponding information encoded in regulatory DNA is actuated via the combi-

natorial binding of sequence-specific TFs to regulatory regions (cis-regulatory modules, CRMs). CRMs switch on promoters and enhancers and are actually the assemblies of TFBSs arranged to provide particular functions [10,11,14,176–178].

The SNPs located in transcriptional regulatory regions can alter gene expression, which may be either adaptive or lead to a disease. The main mechanism underlying the action of these SNPs consists in changes of TF binding, which comprises creation or disruption of TFBSs (cis-regulatory elements) or alteration of the affinity of TFs for their cognate sites [32,155,179,180]. Although many SNPs with such properties have been so far discovered, their mass search in genomes remains challenging. This is mainly associated with the tissue, developmental, and environmental specificities in the effects of rSNPs, which is a direct consequence of the corresponding specificities of the harboring cis-regulatory elements [32,45,124]. Thus, myriads of omics experiments are necessary for this purpose; however, this is still too expensive and time-consuming. The computer methods for recognition of TFBSs in DNA sequences are free of this disadvantage but yet ineffective in detection of both TFBSs and the SNPs changing these sites without the cooperation with omics experiments. The objective reasons here are a high degeneracy of the regulatory DNA code [15,109,181,182]; high importance of low-affinity sites in gene regulation [183]; the presence of structural variants of the binding sites for the same TF [184–187]; and even nonconsensus TFBSs [188,189]. All these facts considerably decrease the efficacy of the available methods for TFBS recognition, most of which are based on the PWM model, which oversimplifies the mechanisms underlying TF–DNA interaction [66,68–70]. Development of new generation bioinformatics approaches relying on machine learning and neural networks raises the hope for more efficient and accurate recognition of both the TFBSs and rSNPs in the genomes [190–195].

Thus, despite the achieved progress, we are still at the beginning of the way to comprehensive annotation of the genome regulatory portion, full cataloging of rSNPs, and clarification of their association with molecular phenotypes and, eventually, with various complex traits, including diseases. The further advance requires improving the efficiency of the existing experimental and bioinformatics methods of systems biology and advent of the new relevant approaches.

Funding: The work was supported by the Grants 18-29-09041 from Russian Foundation for Basic Research and State Budget Project 0259-2021-0013.

Conflicts of Interest: The authors have no conflicts of interest to declare.

References

1. Buniello, A.; MacArthur, J.A.L.; Cerezo, M.; Harris, L.W.; Hayhurst, J.; Malangone, C.; McMahon, A.; Morales, J.; Mountjoy, E.; Sollis, E.; 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–D1012. [[CrossRef](#)] [[PubMed](#)]
2. Claussnitzer, M.; Cho, J.H.; Collins, R.; Cox, N.J.; Dermitzakis, E.T.; Hurles, M.E.; Kathiresan, S.; Kenny, E.E.; Lindgren, C.M.; MacArthur, D.G.; et al. A brief history of human disease genetics. *Nature* **2020**, *577*, 179–189. [[CrossRef](#)] [[PubMed](#)]
3. Bryzgalov, L.O.; Antontseva, E.V.; Matveeva, M.Y.; Shilov, A.G.; Kashina, E.V.; Mordvinov, V.A.; Merkulova, T.I. Detection of Regulatory SNPs in Human Genome Using ChIP-seq ENCODE Data. *PLoS ONE* **2013**, *8*, e78833. [[CrossRef](#)] [[PubMed](#)]
4. Farh, K.K.-H.; Marson, A.; Zhu, J.; Kleinewietfeld, M.; Housley, W.J.; Beik, S.; Shores, N.; Whitton, H.; Ryan, R.J.H.; Shishkin, A.A.; et al. Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* **2015**, *518*, 337–343. [[CrossRef](#)] [[PubMed](#)]
5. Maurano, M.T.; Humbert, R.; Rynes, E.; Thurman, R.E.; Haugen, E.; Wang, H.; Reynolds, A.P.; Sandstrom, R.; Qu, H.; Brody, J.; et al. Systematic Localization of Common Disease-Associated Variation in Regulatory DNA. *Science* **2012**, *337*, 1190–1195. [[CrossRef](#)] [[PubMed](#)]
6. GTEx Consortium. Genetic effects on gene expression across human tissues. *Nature* **2017**, *550*, 204–213. [[CrossRef](#)]
7. Levo, M.; Segal, E. In pursuit of design principles of regulatory sequences. *Nat. Rev. Genet.* **2014**, *15*, 453–468. [[CrossRef](#)]
8. Andersson, R. Promoter or enhancer, what's the difference? Deconstruction of established distinctions and presentation of a unifying model. *BioEssays* **2015**, *37*, 314–323. [[CrossRef](#)]
9. Erokhin, M.; Vassetzky, Y.; Georgiev, P.; Chetverina, D. Eukaryotic enhancers: Common features, regulation, and participation in diseases. *Cell. Mol. Life Sci.* **2015**, *72*, 2361–2375. [[CrossRef](#)]
10. Chen, H.; Pugh, B.F. What Do Transcription Factors Interact with? *J. Mol. Biol.* **2021**, 166883. [[CrossRef](#)]

11. Tobias, I.C.; Abatti, L.E.; Moorthy, S.D.; Mullany, S.; Taylor, T.; Khader, N.; Filice, M.A.; Mitchell, J.A. Transcriptional enhancers: From prediction to functional assessment on a genome-wide scale. *Genome* **2021**, *64*, 426–448. [[CrossRef](#)]
12. Singh, G.; Mullany, S.; Moorthy, S.D.; Zhang, R.; Mehdi, T.; Tian, R.; Duncan, A.G.; Moses, A.M.; Mitchell, J.A. A flexible repertoire of transcription factor binding sites and a diversity threshold determines enhancer activity in embryonic stem cells. *Genome Res.* **2021**, *31*, 564–575. [[CrossRef](#)]
13. Lambert, S.A.; Jolma, A.; Campitelli, L.F.; Das, P.K.; Yin, Y.; Albu, M.; Chen, X.; Taipale, J.; Hughes, T.R.; Weirauch, M.T. The Human Transcription Factors. *Cell* **2018**, *175*, 598–599. [[CrossRef](#)]
14. Lelli, K.M.; Slattery, M.; Mann, R.S. Disentangling the Many Layers of Eukaryotic Transcriptional Regulation. *Annu. Rev. Genet.* **2012**, *46*, 43–68. [[CrossRef](#)]
15. Merkulova, T.I.; Ananko, E.A.; Ignat'eva, E.V.; Kolchanov, N.A. Regulatory transcription codes in eukaryotic genomes. *Genetika* **2013**, *49*, 37–54. [[CrossRef](#)]
16. Wang, Y.; Ma, R.; Liu, B.; Kong, J.; Lin, H.; Yu, X.; Wang, R.; Li, L.; Gao, M.; Zhou, B.; et al. SNP rs17079281 decreases lung cancer risk through creating an YY1-binding site to suppress DCBLD1 expression. *Oncogene* **2020**, *39*, 4092–4102. [[CrossRef](#)]
17. Padhy, B.; Hayat, B.; Nanda, G.G.; Mohanty, P.P.; Alone, D.P. 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*, 4519–4529. [[CrossRef](#)]
18. Krause, M.D.; Huang, R.-T.; Wu, D.; Shentu, T.-P.; Harrison, D.L.; Whalen, M.B.; Stolze, L.K.; Di Rienzo, A.; Moskowitz, I.P.; Civelek, M.; 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–e11358. [[CrossRef](#)]
19. Hazelett, D.J.; Rhie, S.K.; Gaddis, M.; Yan, C.; Lakeland, D.L.; Coetzee, S.G.; Henderson, B.E.; Noushmehr, H.; Cozen, W.; Kote-Jarai, Z.; et al. Comprehensive Functional Annotation of 77 Prostate Cancer Risk Loci. *PLoS Genet.* **2014**, *10*, e1004102. [[CrossRef](#)]
20. Gao, P.; Xia, J.-H.; Sipeky, C.; Dong, X.-M.; Zhang, Q.; Yang, Y.; Zhang, P.; Cruz, S.P.; Zhang, K.; Zhu, J.; et al. Biology and Clinical Implications of the 19q13 Aggressive Prostate Cancer Susceptibility Locus. *Cell* **2018**, *174*, 576–589.e18. [[CrossRef](#)]
21. Afanasyeva, M.A.; Putlyayeva, L.V.; Demin, D.E.; Kulakovskiy, I.V.; Vorontsov, I.E.; Fridman, M.V.; Makeev, V.J.; Kuprash, D.V.; Schwartz, A.M. The single nucleotide variant rs12722489 determines differential estrogen receptor binding and enhancer properties of an IL2RA intronic region. *PLoS ONE* **2017**, *12*, e0172681. [[CrossRef](#)]
22. Korneev, K.V.; Sviriaeva, E.N.; Mitkin, N.A.; Gorbacheva, A.M.; Uvarova, A.N.; Ustiugova, A.S.; Polanovsky, O.L.; Kulakovskiy, I.V.; Afanasyeva, M.A.; Schwartz, A.M.; et al. Minor C allele of the SNP rs7873784 associated with rheumatoid arthritis and type-2 diabetes mellitus binds PU.1 and enhances TLR4 expression. *Biochim. Biophys. Acta Mol. Basis Dis.* **2020**, *1866*, 165626. [[CrossRef](#)]
23. Fang, J.; Jia, J.; Makowski, M.; Xu, M.; Wang, Z.; Zhang, T.; Hoskins, J.W.; Choi, J.; Han, Y.; Zhang, M.; et al. Functional characterization of a multi-cancer risk locus on chr5p15.33 reveals regulation of TERT by ZNF148. *Nat. Commun.* **2017**, *8*, 15034. [[CrossRef](#)]
24. Choi, J.; Zhang, T.; Vu, A.; Ablain, J.; Makowski, M.M.; Colli, L.M.; Xu, M.; Hennessey, R.C.; Yin, J.; Rothschild, H.; et al. Massively parallel reporter assays of melanoma risk variants identify MX2 as a gene promoting melanoma. *Nat. Commun.* **2020**, *11*, 2718. [[CrossRef](#)]
25. Zhao, Y.; Wu, D.; Jiang, D.; Zhang, X.; Wu, T.; Cui, J.; Qian, M.; Zhao, J.; Oesterreich, S.; Sun, W.; et al. A sequential methodology for the rapid identification and characterization of breast cancer-associated functional SNPs. *Nat. Commun.* **2020**, *11*, 3340. [[CrossRef](#)]
26. Prestel, M.; Prell-Schicker, C.; Webb, T.; Malik, R.; Lindner, B.; Ziesch, N.; Rex-Haffner, M.; Röh, S.; Viturawong, T.; Lehm, M.; et al. The Atherosclerosis Risk Variant rs2107595 Mediates Allele-Specific Transcriptional Regulation of HDAC9 via E2F3 and Rb1. *Stroke* **2019**, *50*, 2651–2660. [[CrossRef](#)] [[PubMed](#)]
27. Thomas, R.; Trapani, D.; Goodyer-Sait, L.; Tomkova, M.; Fernandez-Rozadilla, C.; Sahnane, N.; Woolley, C.; Davis, H.; Chegwid-den, L.; Kriaucionis, S.; et al. The polymorphic variant rs1800734 influences methylation acquisition and allele-specific TFAP4 binding in the MLH1 promoter leading to differential mRNA expression. *Sci. Rep.* **2019**, *9*, 13463. [[CrossRef](#)] [[PubMed](#)]
28. Jiang, Z.; Huang, Y.; Zhang, P.; Han, C.; Lu, Y.; Mo, Z.; Zhang, Z.; Li, X.; Zhao, S.; Cai, F.; et al. Characterization of a pathogenic variant in GBA for Parkinson's disease with mild cognitive impairment patients. *Mol. Brain* **2020**, *13*, 102. [[CrossRef](#)] [[PubMed](#)]
29. Allen, E.K.; Randolph, A.G.; Bhangale, T.; Dogra, P.; Ohlson, M.; Oshansky, C.M.; Zamora, A.E.; Shannon, J.P.; Finkelstein, D.; Dressen, A.; et al. SNP-mediated disruption of CTCF binding at the IFITM3 promoter is associated with risk of severe influenza in humans. *Nat. Med.* **2017**, *23*, 975–983. [[CrossRef](#)]
30. Vasiliev, G.V.; Merkulov, V.M.; Kobzev, V.F.; Merkulova, T.I.; Ponomarenko, M.P.; Kolchanov, N.A. Point mutations within 663–666 bp of intron 6 of the human TDO2 gene, associated with a number of psychiatric disorders, damage the YY-1 transcription factor binding site. *FEBS Lett.* **1999**, *462*, 85–88. [[CrossRef](#)]
31. Cooper, D. The human gene mutation database. *Nucleic Acids Res.* **1998**, *26*, 285–287. [[CrossRef](#)]
32. Deplancke, B.; Alpern, D.; Gardeux, V. The Genetics of Transcription Factor DNA Binding Variation. *Cell* **2016**, *166*, 538–554. [[CrossRef](#)]
33. Ponomarenko, J.V. rSNP_Guide, a database system for analysis of transcription factor binding to target sequences: Application to SNPs and site-directed mutations. *Nucleic Acids Res.* **2001**, *29*, 312–316. [[CrossRef](#)]

34. Bienvenu, T.; Lacronique, V.; Raymondjean, M.; Cazeneuve, C.; Hubert, D.; Kaplan, J.-C.; Beldjord, C. Three novel sequence variations in the 5' upstream region of the cystic fibrosis transmembrane conductance regulator (CFTR) gene: Two polymorphisms and one putative molecular defect. *Hum. Genet.* **1995**, *95*, 698–702. [[CrossRef](#)]
35. Ludlow, L.B.; Schick, B.P.; Budarf, M.L.; Driscoll, D.A.; Zackai, E.H.; Cohen, A.; Konkle, B.A. Identification of a Mutation in a GATA Binding Site of the Platelet Glycoprotein I β Promoter Resulting in the Bernard-Soulier Syndrome. *J. Biol. Chem.* **1996**, *271*, 22076–22080. [[CrossRef](#)]
36. Comings, D.E.; Gade, R.; Muhleman, D.; Chiu, C.; Wu, S.; To, M.; Spence, M.; Dietz, G.; Winn-Deen, E.; Rosenthal, R.J.; et al. Exon and intron variants in the human tryptophan 2,3-dioxygenase gene: Potential association with Tourette syndrome, substance abuse and other disorders. *Pharmacogenetics* **1996**, *6*, 307–318. [[CrossRef](#)]
37. Merkulov, V.M.; Merkulova, T.I. Nucleotide sequence of a fragment of the rat tryptophan oxygenase gene showing high affinity to glucocorticoid receptor in vitro. *Biochim. Biophys. Acta Gene Struct. Expr.* **1992**, *1132*, 100–102. [[CrossRef](#)]
38. Ponomarenko, J.V.; Ponomarenko, M.P.; Frolov, A.S.; Vorobyev, D.G.; Overton, G.C.; Kolchanov, N.A. Conformational and physicochemical DNA features specific for transcription factor binding sites. *Bioinformatics* **1999**, *15*, 654–668. [[CrossRef](#)]
39. Verheul, T.C.J.; van Hijfte, L.; Perenthaler, E.; Barakat, T.S. The Why of YY1: Mechanisms of Transcriptional Regulation by Yin Yang 1. *Front. Cell Dev. Biol.* **2020**, *8*. [[CrossRef](#)]
40. Knight, J.C.; Udalova, I.; Hill, A.V.S.; Greenwood, B.M.; Peshu, N.; Marsh, K.; Kwiatkowski, D. A polymorphism that affects OCT-1 binding to the TNF promoter region is associated with severe malaria. *Nat. Genet.* **1999**, *22*, 145–150. [[CrossRef](#)]
41. Piedrafita, F.J.; Molander, R.B.; Vansant, G.; Orlova, E.A.; Pfahl, M.; Reynolds, W.F. An Alu Element in the Myeloperoxidase Promoter Contains a Composite SP1-Thyroid Hormone-Retinoic Acid Response Element. *J. Biol. Chem.* **1996**, *271*, 14412–14420. [[CrossRef](#)] [[PubMed](#)]
42. Moi, P.; Loudianos, G.; Lavinha, J.; Murru, S.; Cossu, P.; Casu, R.; Oggiano, L.; Longinotti, M.; Cao, A.; Pirastu, M. Delta-thalassemia due to a mutation in an erythroid-specific binding protein sequence 3' to the delta-globin gene. *Blood* **1992**, *79*, 512–516. [[CrossRef](#)] [[PubMed](#)]
43. Wingender, E. TRANSFAC: A database on transcription factors and their DNA binding sites. *Nucleic Acids Res.* **1996**, *24*, 238–241. [[CrossRef](#)]
44. Nishizaki, S.S.; Boyle, A.P. Mining the Unknown: Assigning Function to Noncoding Single Nucleotide Polymorphisms. *Trends Genet.* **2017**, *33*, 34–45. [[CrossRef](#)]
45. Liu, B.; Montgomery, S.B. Identifying causal variants and genes using functional genomics in specialized cell types and contexts. *Hum. Genet.* **2020**, *139*, 95–102. [[CrossRef](#)]
46. Moyerbrailean, G.A.; Richards, A.L.; Kurtz, D.; Kalita, C.A.; Davis, G.O.; Harvey, C.T.; Alazizi, A.; Watzka, D.; Sorokin, Y.; Hauff, N.; et al. High-throughput allele-specific expression across 250 environmental conditions. *Genome Res.* **2016**, *26*, 1627–1638. [[CrossRef](#)]
47. Chen, Q.; Deng, X.; Hu, X.; Guan, S.; He, M.; Wang, Y.; Wei, B.; Zhang, J.; Zhao, H.; Yao, W.; et al. Breast Cancer Risk-Associated SNPs in the mTOR Promoter Form De Novo KLF5- and ZEB1-Binding Sites that Influence the Cellular Response to Paclitaxel. *Mol. Cancer Res.* **2019**, *17*, 2244–2256. [[CrossRef](#)] [[PubMed](#)]
48. Matana, A.; Ziros, P.G.; Chartoumpakis, D.V.; Renaud, C.O.; Polašek, O.; Hayward, C.; Zemunik, T.; Sykiotis, G.P. Rare and common genetic variations in the Keap1/Nrf2 antioxidant response pathway impact thyroglobulin gene expression and circulating levels, respectively. *Biochem. Pharmacol.* **2020**, *173*, 113605. [[CrossRef](#)]
49. Levings, D.; Shaw, K.E.; Lacher, S.E. Genomic resources for dissecting the role of non-protein coding variation in gene-environment interactions. *Toxicology* **2020**, *441*, 152505. [[CrossRef](#)] [[PubMed](#)]
50. Wei, Y.B.; McCarthy, M.; Ren, H.; Carrillo-Roa, T.; Shekhtman, T.; DeModena, A.; Liu, J.J.; Leckband, S.G.; Mors, O.; Rietschel, M.; et al. A functional variant in the serotonin receptor 7 gene (HTR7), rs7905446, is associated with good response to SSRIs in bipolar and unipolar depression. *Mol. Psychiatry* **2020**, *25*, 1312–1322. [[CrossRef](#)] [[PubMed](#)]
51. Boldes, T.; Merenbakh-Lamin, K.; Journo, S.; Shachar, E.; Lipson, D.; Yeheskel, A.; Pasmanik-Chor, M.; Rubinek, T.; Wolf, I. R269C variant of ESR1: High prevalence and differential function in a subset of pancreatic cancers. *BMC Cancer* **2020**, *20*, 531. [[CrossRef](#)]
52. Zhao, Y.Y.; Zhou, J.; Narayanan, C.S.; Cui, Y.; Kumar, A. Role of C/A Polymorphism at -20 on the Expression of Human Angiotensinogen Gene. *Hypertension* **1999**, *33*, 108–115. [[CrossRef](#)]
53. López Rodríguez, M.; Kaminska, D.; Lappalainen, K.; Pihlajamäki, J.; Kaikkonen, M.U.; Laakso, M. Identification and characterization of a FOXA2-regulated transcriptional enhancer at a type 2 diabetes intronic locus that controls GCKR expression in liver cells. *Genome Med.* **2017**, *9*, 63. [[CrossRef](#)]
54. Boulling, A.; Masson, E.; Zou, W.; Paliwal, S.; Wu, H.; Issarapu, P.; Bhaskar, S.; Génin, E.; Cooper, D.N.; Li, Z.; 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*, 1014–1024. [[CrossRef](#)]
55. Li, X.-X.; Peng, T.; Gao, J.; Feng, J.-G.; Wu, D.-D.; Yang, T.; Zhong, L.; Fu, W.-P.; Sun, C. 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. Cell. Mol. Physiol.* **2019**, *317*, L456–L463. [[CrossRef](#)]
56. Peng, T.; Zhong, L.; Gao, J.; Wan, Z.; Fu, W.-P.; Sun, C. Identification of rs11615992 as a novel regulatory SNP for human P2RX7 by allele-specific expression. *Mol. Genet. Genom.* **2020**, *295*, 23–30. [[CrossRef](#)]

57. Kuang, X.; Zhou, Q.; Li, Z.; Hu, Y.; Kang, Y.; Huang, W. -254C>G SNP in the TRPC6 Gene Promoter Influences Its Expression via Interaction with the NF- κ B Subunit RELA in Steroid-Resistant Nephrotic Syndrome Children. *Int. J. Genom.* **2019**, *2019*, 1–7. [[CrossRef](#)]
58. 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. [[CrossRef](#)]
59. Thynn, H.N.; Chen, X.-F.; Hu, W.-X.; Duan, Y.-Y.; Zhu, D.-L.; Chen, H.; Wang, N.-N.; Chen, H.-H.; Rong, Y.; Lu, B.-J.; et al. An Allele-Specific Functional SNP Associated with Two Systemic Autoimmune Diseases Modulates IRF5 Expression by Long-Range Chromatin Loop Formation. *J. Investig. Dermatol.* **2020**, *140*, 348–360.e11. [[CrossRef](#)]
60. Coetzee, S.G.; Coetzee, G.A.; Hazelett, D.J. motifbreakR: An R/Bioconductor package for predicting variant effects at transcription factor binding sites: Fig. 1. *Bioinformatics* **2015**, *btv470*. [[CrossRef](#)]
61. Kumar, S.; Ambrosini, G.; Bucher, P. SNP2TFBS—A database of regulatory SNPs affecting predicted transcription factor binding site affinity. *Nucleic Acids Res.* **2017**, *45*, D139–D144. [[CrossRef](#)]
62. Kulakovskiy, I.V.; Vorontsov, I.E.; Yevshin, I.S.; Sharipov, R.N.; Fedorova, A.D.; Rumynskiy, E.I.; Medvedeva, Y.A.; Magana-Mora, A.; Bajic, V.B.; Papatsenko, D.A.; et al. HOCOMOCO: Towards a complete collection of transcription factor binding models for human and mouse via large-scale ChIP-Seq analysis. *Nucleic Acids Res.* **2018**, *46*, D252–D259. [[CrossRef](#)]
63. Fornes, O.; Castro-Mondragon, J.A.; Khan, A.; van der Lee, R.; Zhang, X.; Richmond, P.A.; Modi, B.P.; Correard, S.; Gheorghe, M.; Baranašić, D.; et al. JASPAR 2020: Update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **2019**. [[CrossRef](#)]
64. Nishizaki, S.S.; Ng, N.; Dong, S.; Porter, R.S.; Morterud, C.; Williams, C.; Asman, C.; Switzenberg, J.A.; Boyle, A.P. Predicting the effects of SNPs on transcription factor binding affinity. *Bioinformatics* **2020**, *36*, 364–372. [[CrossRef](#)]
65. Shin, S.; Hudson, R.; Harrison, C.; Craven, M.; Keleş, S. atSNP Search: A web resource for statistically evaluating influence of human genetic variation on transcription factor binding. *Bioinformatics* **2019**, *35*, 2657–2659. [[CrossRef](#)]
66. Yan, J.; Qiu, Y.; Ribeiro dos Santos, A.M.; Yin, Y.; Li, Y.E.; Vinckier, N.; Nariari, N.; Benaglio, P.; Raman, A.; Li, X.; et al. Systematic analysis of binding of transcription factors to noncoding variants. *Nature* **2021**, *591*, 147–151. [[CrossRef](#)]
67. Stormo, G.D. DNA binding sites: Representation and discovery. *Bioinformatics* **2000**, *16*, 16–23. [[CrossRef](#)]
68. Slattery, M.; Zhou, T.; Yang, L.; Dantas Machado, A.C.; Gordán, R.; Rohs, R. Absence of a simple code: How transcription factors read the genome. *Trends Biochem. Sci.* **2014**, *39*, 381–399. [[CrossRef](#)]
69. Srivastava, D.; Mahony, S. Sequence and chromatin determinants of transcription factor binding and the establishment of cell type-specific binding patterns. *Biochim. Biophys. Acta Gene Regul. Mech.* **2020**, *1863*, 194443. [[CrossRef](#)]
70. Inukai, S.; Kock, K.H.; Bulyk, M.L. Transcription factor–DNA binding: Beyond binding site motifs. *Curr. Opin. Genet. Dev.* **2017**, *43*, 110–119. [[CrossRef](#)] [[PubMed](#)]
71. Fagny, M.; Paulson, J.N.; Kuijjer, M.L.; Sonawane, A.R.; Chen, C.-Y.; Lopes-Ramos, C.M.; Glass, K.; Quackenbush, J.; Platig, J. Exploring regulation in tissues with eQTL networks. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, e7841–e7850. [[CrossRef](#)] [[PubMed](#)]
72. Syddall, C.M.; Reynard, L.N.; Young, D.A.; 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. [[CrossRef](#)] [[PubMed](#)]
73. Liu, D.; Qin, S.; Ray, B.; Kalari, K.R.; Wang, L.; Weinshilboum, R.M. Single Nucleotide Polymorphisms (SNPs) Distant from Xenobiotic Response Elements Can Modulate Aryl Hydrocarbon Receptor Function: SNP-Dependent CYP1A1 Induction. *Drug Metab. Dispos.* **2018**, *46*, 1372–1381. [[CrossRef](#)] [[PubMed](#)]
74. Tian, J.; Lou, J.; Cai, Y.; Rao, M.; Lu, Z.; Zhu, Y.; Zou, D.; Peng, X.; Wang, H.; Zhang, M.; et al. Risk SNP-Mediated Enhancer–Promoter Interaction Drives Colorectal Cancer through Both FADS2 and AP002754.2. *Cancer Res.* **2020**, *80*, 1804–1818. [[CrossRef](#)]
75. Merkulov, V.M.; Leberfarb, E.Y.; Merkulova, T.I. Regulatory SNPs and their widespread effects on the transcriptome. *J. Biosci.* **2018**, *43*, 1069–1075. [[CrossRef](#)]
76. ENCODE Project Consortium. Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* **2020**, *583*, 699–710. [[CrossRef](#)]
77. Berner, D.; Hoja, U.; Zenkel, M.; Ross, J.J.; Uebe, S.; Paoli, D.; Frezzotti, P.; Rautenbach, R.M.; Ziskind, A.; Williams, S.E.; et al. The protective variant rs7173049 at LOXL1 locus impacts on retinoic acid signaling pathway in pseudoexfoliation syndrome. *Hum. Mol. Genet.* **2019**, *28*, 2531–2548. [[CrossRef](#)]
78. Ali, M.W.; Patro, C.P.K.; Zhu, J.J.; Dampier, C.H.; Plummer, S.J.; Kuscus, C.; Adli, M.; Lau, C.; Lai, R.K.; Casey, G. 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. [[CrossRef](#)]
79. Liu, S.; Liu, Y.; Zhang, Q.; Wu, J.; Liang, J.; Yu, S.; Wei, G.-H.; White, K.P.; Wang, X. Systematic identification of regulatory variants associated with cancer risk. *Genome Biol.* **2017**, *18*, 194. [[CrossRef](#)]
80. Gupta, R.M.; Hadaya, J.; Trehan, A.; Zekavat, S.M.; Roselli, C.; Klarin, D.; Emdin, C.A.; Hilvering, C.R.E.; Bianchi, V.; Mueller, C.; et al. A Genetic Variant Associated with Five Vascular Diseases Is a Distal Regulator of Endothelin-1 Gene Expression. *Cell* **2017**, *170*, 522–533. [[CrossRef](#)]
81. Zhu, D.-L.; Chen, X.-F.; Hu, W.-X.; Dong, S.-S.; Lu, B.-J.; Rong, Y.; Chen, Y.-X.; Chen, H.; Thynn, H.N.; Wang, N.-N.; 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*, 1335–1346. [[CrossRef](#)]

82. Wang, X.; Hayes, J.E.; Xu, X.; Gao, X.; Mehta, D.; Lilja, H.G.; Klein, R.J. Validation of prostate cancer risk variants rs10993994 and rs7098889 by CRISPR/Cas9 mediated genome editing. *Gene* **2021**, *768*, 145265. [[CrossRef](#)]
83. Gutierrez-Arcelus, M.; Baglaenko, Y.; Arora, J.; Hannes, S.; Luo, Y.; Amariuta, T.; Teslovich, N.; Rao, D.A.; Ermann, J.; Jonsson, A.H.; et al. Allele-specific expression changes dynamically during T cell activation in HLA and other autoimmune loci. *Nat. Genet.* **2020**, *52*, 247–253. [[CrossRef](#)]
84. Klein, J.C.; Keith, A.; Rice, S.J.; Shepherd, C.; Agarwal, V.; Loughlin, J.; Shendure, J. Functional testing of thousands of osteoarthritis-associated variants for regulatory activity. *Nat. Commun.* **2019**, *10*, 2434. [[CrossRef](#)]
85. Azghandi, S.; Prell, C.; van der Laan, S.W.; Schneider, M.; Malik, R.; Berer, K.; Gerdes, N.; Pasterkamp, G.; Weber, C.; Haffner, C.; et al. Deficiency of the Stroke Relevant HDAC9 Gene Attenuates Atherosclerosis in Accord with Allele-Specific Effects at 7p21.1. *Stroke* **2015**, *46*, 197–202. [[CrossRef](#)]
86. Roca-Ayats, N.; Martínez-Gil, N.; Cozar, M.; Gerousi, M.; Garcia-Giralt, N.; Ovejero, D.; Mellibovsky, L.; Nogués, X.; Díez-Pérez, A.; Grinberg, D.; et al. Functional characterization of the C7ORF76 genomic region, a prominent GWAS signal for osteoporosis in 7q21.3. *Bone* **2019**, *123*, 39–47. [[CrossRef](#)]
87. Malecová, B.; Morris, K.V. Transcriptional gene silencing through epigenetic changes mediated by non-coding RNAs. *Curr. Opin. Mol. Ther.* **2010**, *12*, 214–222.
88. Butter, F.; Davison, L.; Vitarawong, T.; Scheibe, M.; Vermeulen, M.; Todd, J.A.; Mann, M. Proteome-Wide Analysis of Disease-Associated SNPs That Show Allele-Specific Transcription Factor Binding. *PLoS Genet.* **2012**, *8*, e1002982. [[CrossRef](#)]
89. Amin Al Olama, A.; Kote-Jarai, Z.; Schumacher, F.R.; Wiklund, F.; Berndt, S.I.; Benlloch, S.; Giles, G.G.; Severi, G.; Neal, D.E.; Hamdy, F.C.; et al. A meta-analysis of genome-wide association studies to identify prostate cancer susceptibility loci associated with aggressive and non-aggressive disease. *Hum. Mol. Genet.* **2013**, *22*, 408–415. [[CrossRef](#)]
90. Cerami, E.; Gao, J.; Dogrusoz, U.; Gross, B.E.; Sumer, S.O.; Aksoy, B.A.; Jacobsen, A.; Byrne, C.J.; Heuer, M.L.; Larsson, E.; et al. The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data: Figure 1. *Cancer Discov.* **2012**, *2*, 401–404. [[CrossRef](#)]
91. Ran, F.A.; Hsu, P.D.; Wright, J.; Agarwala, V.; Scott, D.A.; Zhang, F. Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **2013**, *8*, 2281–2308. [[CrossRef](#)]
92. Malik, R.; Chauhan, G.; Traylor, M.; Sargurupremraj, M.; Okada, Y.; Mishra, A.; Rutten-Jacobs, L.; Giese, A.-K.; van der Laan, S.W.; Gretarsdottir, S.; et al. Multiethnic genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nat. Genet.* **2018**, *50*, 524–537. [[CrossRef](#)]
93. Nelson, C.P.; Goel, A.; Butterworth, A.S.; Kanoni, S.; Webb, T.R.; Marouli, E.; Zeng, L.; Ntalla, I.; Lai, F.Y.; Hopewell, J.C.; et al. Association analyses based on false discovery rate implicate new loci for coronary artery disease. *Nat. Genet.* **2017**, *49*, 1385–1391. [[CrossRef](#)]
94. Takeuchi, F.; Akiyama, M.; Matoba, N.; Katsuya, T.; Nakatochi, M.; Tabara, Y.; Narita, A.; Saw, W.-Y.; Moon, S.; Spracklen, C.N.; et al. Interethnic analyses of blood pressure loci in populations of East Asian and European descent. *Nat. Commun.* **2018**, *9*, 5052. [[CrossRef](#)]
95. Hoffmann, T.J.; Ehret, G.B.; Nandakumar, P.; Ranatunga, D.; Schaefer, C.; Kwok, P.-Y.; Iribarren, C.; Chakravarti, A.; Risch, N. Genome-wide association analyses using electronic health records identify new loci influencing blood pressure variation. *Nat. Genet.* **2017**, *49*, 54–64. [[CrossRef](#)]
96. Giri, A.; Hellwege, J.N.; Keaton, J.M.; Park, J.; Qiu, C.; Warren, H.R.; Torstenson, E.S.; Kovesdy, C.P.; Sun, Y.V.; Wilson, O.D.; et al. Trans-ethnic association study of blood pressure determinants in over 750,000 individuals. *Nat. Genet.* **2019**, *51*, 51–62. [[CrossRef](#)]
97. Karasu, N.; Sexton, T. 4C-Seq: Interrogating Chromatin Looping with Circular Chromosome Conformation Capture. In *Capturing Chromosome Conformation*; Humana: New York, NY, USA, 2021; pp. 19–34.
98. Sardi, S.P.; Clarke, J.; Viel, C.; Chan, M.; Tamssett, T.J.; Treleaven, C.M.; Bu, J.; Sweet, L.; Passini, M.A.; Dodge, J.C.; et al. Augmenting CNS glucocerebrosidase activity as a therapeutic strategy for parkinsonism and other Gaucher-related synucleinopathies. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 3537–3542. [[CrossRef](#)]
99. Mata, I.F.; Leverenz, J.B.; Weintraub, D.; Trojanowski, J.Q.; Chen-Plotkin, A.; Van Deerlin, V.M.; Ritz, B.; Rausch, R.; Factor, S.A.; Wood-Siverio, C.; et al. GBA Variants are associated with a distinct pattern of cognitive deficits in Parkinson’s disease. *Mov. Disord.* **2016**, *31*, 95–102. [[CrossRef](#)]
100. van der Lee, S.J.; Knol, M.J.; Chauhan, G.; Satizabal, C.L.; Smith, A.V.; Hofer, E.; Bis, J.C.; Hibar, D.P.; Hilal, S.; van den Akker, E.B.; et al. A genome-wide association study identifies genetic loci associated with specific lobar brain volumes. *Commun. Biol.* **2019**, *2*, 285. [[CrossRef](#)]
101. Lan, Q.; Hsiung, C.A.; Matsuo, K.; Hong, Y.-C.; Seow, A.; Wang, Z.; Hosgood, H.D.; Chen, K.; Wang, J.-C.; Chatterjee, N.; et al. Genome-wide association analysis identifies new lung cancer susceptibility loci in never-smoking women in Asia. *Nat. Genet.* **2012**, *44*, 1330–1335. [[CrossRef](#)]
102. McKay, J.D.; Hung, R.J.; Han, Y.; Zong, X.; Carreras-Torres, R.; Christiani, D.C.; Caporaso, N.E.; Johansson, M.; Xiao, X.; Li, Y.; et al. Large-scale association analysis identifies new lung cancer susceptibility loci and heterogeneity in genetic susceptibility across histological subtypes. *Nat. Genet.* **2017**, *49*, 1126–1132. [[CrossRef](#)] [[PubMed](#)]
103. Lappalainen, T. Functional genomics bridges the gap between quantitative genetics and molecular biology. *Genome Res.* **2015**, *25*, 1427–1431. [[CrossRef](#)] [[PubMed](#)]

104. Qian, Y.; Zhang, L.; Cai, M.; Li, H.; Xu, H.; Yang, H.; Zhao, Z.; Rhie, S.K.; Farnham, P.J.; Shi, J.; et al. The prostate cancer risk variant rs55958994 regulates multiple gene expression through extreme long-range chromatin interaction to control tumor progression. *Sci. Adv.* **2019**, *5*, eaaw6710. [[CrossRef](#)]
105. Ulirsch, J.C.; Nandakumar, S.K.; Wang, L.; Giani, F.C.; Zhang, X.; Rogov, P.; Melnikov, A.; McDonel, P.; Do, R.; Mikkelsen, T.S.; et al. Systematic Functional Dissection of Common Genetic Variation Affecting Red Blood Cell Traits. *Cell* **2016**, *165*, 1530–1545. [[CrossRef](#)] [[PubMed](#)]
106. Kalita, C.A.; Brown, C.D.; Freiman, A.; Isherwood, J.; Wen, X.; Pique-Regi, R.; Luca, F. High-throughput characterization of genetic effects on DNA–protein binding and gene transcription. *Genome Res.* **2018**, *28*, 1701–1708. [[CrossRef](#)] [[PubMed](#)]
107. Patwardhan, R.P.; Lee, C.; Litvin, O.; Young, D.L.; Pe'er, D.; Shendure, J. High-resolution analysis of DNA regulatory elements by synthetic saturation mutagenesis. *Nat. Biotechnol.* **2009**, *27*, 1173–1175. [[CrossRef](#)] [[PubMed](#)]
108. Zhang, P.; Xia, J.-H.; Zhu, J.; Gao, P.; Tian, Y.-J.; Du, M.; Guo, Y.-C.; Suleman, S.; Zhang, Q.; Kohli, M.; et al. High-throughput screening of prostate cancer risk loci by single nucleotide polymorphisms sequencing. *Nat. Commun.* **2018**, *9*, 2022. [[CrossRef](#)]
109. Kolchanov, N.A.; Merkulova, T.I.; Ignatieva, E.V.; Ananko, E.A.; Oshchepkov, D.Y.; Levitsky, V.G.; Vasiliev, G.V.; Klimova, N.V.; Merkulov, V.M.; Charles Hodgman, T. Combined experimental and computational approaches to study the regulatory elements in eukaryotic genes. *Brief. Bioinform.* **2007**, *8*, 266–274. [[CrossRef](#)]
110. Li, S.; Li, Y.; Li, X.; Liu, J.; Huo, Y.; Wang, J.; Liu, Z.; Li, M.; Luo, X.-J. Regulatory mechanisms of major depressive disorder risk variants. *Mol. Psychiatry* **2020**, *25*, 1926–1945. [[CrossRef](#)]
111. Wray, N.R.; Ripke, S.; Mattheisen, M.; Trzaskowski, M.; Byrne, E.M.; Abdellaoui, A.; Adams, M.J.; Agerbo, E.; Air, T.M.; Andlauer, T.M.F.; et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat. Genet.* **2018**, *50*, 668–681. [[CrossRef](#)]
112. Sun, W.; Yao, S.; Tang, J.; Liu, S.; Chen, J.; Deng, D.; Zeng, C. Integrative analysis of super enhancer SNPs for type 2 diabetes. *PLoS ONE* **2018**, *13*, e0192105. [[CrossRef](#)] [[PubMed](#)]
113. Gong, J.; Qiu, C.; Huang, D.; Zhang, Y.; Yu, S.; Zeng, C. Integrative functional analysis of super enhancer SNPs for coronary artery disease. *J. Hum. Genet.* **2018**, *63*, 627–638. [[CrossRef](#)] [[PubMed](#)]
114. Guo, L.; Du, Y.; Qu, S.; Wang, J. rVarBase: An updated database for regulatory features of human variants. *Nucleic Acids Res.* **2016**, *44*, D888–D893. [[CrossRef](#)] [[PubMed](#)]
115. Jones, M.R.; Peng, P.-C.; Coetzee, S.G.; Tyrer, J.; Reyes, A.L.P.; Corona, R.I.; Davis, B.; Chen, S.; Dezem, F.; Seo, J.-H.; et al. Ovarian Cancer Risk Variants Are Enriched in Histotype-Specific Enhancers and Disrupt Transcription Factor Binding Sites. *Am. J. Hum. Genet.* **2020**, *107*, 622–635. [[CrossRef](#)]
116. Guo, Y.; Perez, A.A.; Hazelett, D.J.; Coetzee, G.A.; Rhie, S.K.; Farnham, P.J. CRISPR-mediated deletion of prostate cancer risk-associated CTCF loop anchors identifies repressive chromatin loops. *Genome Biol.* **2018**, *19*, 160. [[CrossRef](#)]
117. Gamazon, E.R.; Segrè, A.V.; van de Bunt, M.; Wen, X.; Xi, H.S.; Hormozdiari, F.; Ongen, H.; Konkashbaev, A.; Derks, E.M.; Aguet, F.; et al. Using an atlas of gene regulation across 44 human tissues to inform complex disease- and trait-associated variation. *Nat. Genet.* **2018**, *50*, 956–967. [[CrossRef](#)]
118. Barbeira, A.N.; Bonazzola, R.; Gamazon, E.R.; Liang, Y.; Park, Y.; Kim-Hellmuth, S.; Wang, G.; Jiang, Z.; Zhou, D.; Hormozdiari, F.; et al. Exploiting the GTEx resources to decipher the mechanisms at GWAS loci. *Genome Biol.* **2021**, *22*, 49. [[CrossRef](#)]
119. Corces, M.R.; Shcherbina, A.; Kundu, S.; Gloudemans, M.J.; Frésard, L.; Granja, J.M.; Louie, B.H.; Eulalio, T.; Shams, S.; Bagdatli, S.T.; et al. Single-cell epigenomic analyses implicate candidate causal variants at inherited risk loci for Alzheimer’s and Parkinson’s diseases. *Nat. Genet.* **2020**, *52*, 1158–1168. [[CrossRef](#)]
120. Ray, J.P.; de Boer, C.G.; Fulco, C.P.; Lareau, C.A.; Kanai, M.; Ulirsch, J.C.; Tewhey, R.; Ludwig, L.S.; Reilly, S.K.; Bergman, D.T.; et al. Prioritizing disease and trait causal variants at the TNFAIP3 locus using functional and genomic features. *Nat. Commun.* **2020**, *11*, 1237. [[CrossRef](#)]
121. Zeng, B.; Lloyd-Jones, L.R.; Montgomery, G.W.; Metspalu, A.; Esko, T.; Franke, L.; Vosa, U.; Claringbould, A.; Brigham, K.L.; Quyyumi, A.A.; et al. Comprehensive Multiple eQTL Detection and Its Application to GWAS Interpretation. *Genetics* **2019**, *212*, 905–918. [[CrossRef](#)]
122. Zhao, J.; Cheng, F.; Jia, P.; Cox, N.; Denny, J.C.; Zhao, Z. An integrative functional genomics framework for effective identification of novel regulatory variants in genome–phenome studies. *Genome Med.* **2018**, *10*, 7. [[CrossRef](#)]
123. Gerring, Z.F.; Vargas, A.M.; Gamazon, E.R.; Derks, E.M. An integrative systems-based analysis of substance use: eQTL-informed gene-based tests, gene networks, and biological mechanisms. *Am. J. Med. Genet. Part B Neuropsychiatr. Genet.* **2020**. [[CrossRef](#)]
124. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **2020**, *369*, 1318–1330. [[CrossRef](#)]
125. Fairfax, B.P.; Humburg, P.; Makino, S.; Naranbhai, V.; Wong, D.; Lau, E.; Jostins, L.; Plant, K.; Andrews, R.; McGee, C.; et al. Innate Immune Activity Conditions the Effect of Regulatory Variants upon Monocyte Gene Expression. *Science* **2014**, *343*, 1246949. [[CrossRef](#)]
126. Fan, J.; Hu, J.; Xue, C.; Zhang, H.; Susztak, K.; Reilly, M.P.; Xiao, R.; Li, M. ASEP: Gene-based detection of allele-specific expression across individuals in a population by RNA sequencing. *PLoS Genet.* **2020**, *16*, e1008786. [[CrossRef](#)]

127. Kim-Hellmuth, S.; Bechheim, M.; Pütz, B.; Mohammadi, P.; Nédélec, Y.; Giangreco, N.; Becker, J.; Kaiser, V.; Fricker, N.; Beier, E.; et al. Genetic regulatory effects modified by immune activation contribute to autoimmune disease associations. *Nat. Commun.* **2017**, *8*, 266. [[CrossRef](#)]
128. Werling, D.M.; Pochareddy, S.; Choi, J.; An, J.-Y.; Sheppard, B.; Peng, M.; Li, Z.; Dastmalchi, C.; Santpere, G.; Sousa, A.M.M.; et al. Whole-Genome and RNA Sequencing Reveal Variation and Transcriptomic Coordination in the Developing Human Prefrontal Cortex. *Cell Rep.* **2020**, *31*, 107489. [[CrossRef](#)]
129. Göring, H.H.H.; Curran, J.E.; Johnson, M.P.; Dyer, T.D.; Charlesworth, J.; Cole, S.A.; Jowett, J.B.M.; Abraham, L.J.; Rainwater, D.L.; Comuzzie, A.G.; et al. Discovery of expression QTLs using large-scale transcriptional profiling in human lymphocytes. *Nat. Genet.* **2007**, *39*, 1208–1216. [[CrossRef](#)]
130. Westra, H.-J.; Franke, L. From genome to function by studying eQTLs. *Biochim. Biophys. Acta Mol. Basis Dis.* **2014**, *1842*, 1896–1902. [[CrossRef](#)]
131. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **2013**, *45*, 580–585. [[CrossRef](#)]
132. Umans, B.D.; Battle, A.; Gilad, Y. Where Are the Disease-Associated eQTLs? *Trends Genet.* **2021**, *37*, 109–124. [[CrossRef](#)] [[PubMed](#)]
133. Zou, J.; Hormozdiari, F.; Jew, B.; Castel, S.E.; Lappalainen, T.; Ernst, J.; Sul, J.H.; Eskin, E. Leveraging allelic imbalance to refine fine-mapping for eQTL studies. *PLoS Genet.* **2019**, *15*, e1008481. [[CrossRef](#)] [[PubMed](#)]
134. Gamazon, E.R.; Zwinderman, A.H.; Cox, N.J.; Denys, D.; Derks, E.M. Multi-tissue transcriptome analyses identify genetic mechanisms underlying neuropsychiatric traits. *Nat. Genet.* **2019**, *51*, 933–940. [[CrossRef](#)] [[PubMed](#)]
135. Gerring, Z.F.; Lupton, M.K.; Edey, D.; Gamazon, E.R.; Derks, E.M. An analysis of genetically regulated gene expression across multiple tissues implicates novel gene candidates in Alzheimer’s disease. *Alzheimers Res. Ther.* **2020**, *12*, 43. [[CrossRef](#)] [[PubMed](#)]
136. Hormozdiari, F.; van de Bunt, M.; Segrè, A.V.; Li, X.; Joo, J.W.J.; Bilow, M.; Sul, J.H.; Sankararaman, S.; Pasaniuc, B.; Eskin, E. Colocalization of GWAS and eQTL Signals Detects Target Genes. *Am. J. Hum. Genet.* **2016**, *99*, 1245–1260. [[CrossRef](#)]
137. Fadason, T.; Ekblad, C.; Ingram, J.R.; Schierding, W.S.; O’Sullivan, J.M. Physical Interactions and Expression Quantitative Traits Loci Identify Regulatory Connections for Obesity and Type 2 Diabetes Associated SNPs. *Front. Genet.* **2017**, *8*. [[CrossRef](#)] [[PubMed](#)]
138. Jaffe, A.E.; Hoepfner, D.J.; Saito, T.; Blanpain, L.; Ukaigwe, J.; Burke, E.E.; Collado-Torres, L.; Tao, R.; Tajinda, K.; Maynard, K.R.; et al. Profiling gene expression in the human dentate gyrus granule cell layer reveals insights into schizophrenia and its genetic risk. *Nat. Neurosci.* **2020**, *23*, 510–519. [[CrossRef](#)]
139. Morrow, J.D.; Cho, M.H.; Platig, J.; Zhou, X.; DeMeo, D.L.; Qiu, W.; Celli, B.; Marchetti, N.; Criner, G.J.; Bueno, R.; et al. Ensemble genomic analysis in human lung tissue identifies novel genes for chronic obstructive pulmonary disease. *Hum. Genom.* **2018**, *12*, 1. [[CrossRef](#)]
140. O’Brien, H.E.; Hannon, E.; Hill, M.J.; Toste, C.C.; Robertson, M.J.; Morgan, J.E.; McLaughlin, G.; Lewis, C.M.; Schalkwyk, L.C.; Hall, L.S.; et al. Expression quantitative trait loci in the developing human brain and their enrichment in neuropsychiatric disorders. *Genome Biol.* **2018**, *19*, 194. [[CrossRef](#)]
141. Ratnapriya, R.; Sosina, O.A.; Starostik, M.R.; Kwicklis, M.; Kapphahn, R.J.; Fritsche, L.G.; Walton, A.; Arvanitis, M.; Gieser, L.; Pietraszkiewicz, A.; et al. Retinal transcriptome and eQTL analyses identify genes associated with age-related macular degeneration. *Nat. Genet.* **2019**, *51*, 606–610. [[CrossRef](#)]
142. Zhang, T.; Choi, J.; Kovacs, M.A.; Shi, J.; Xu, M.; Goldstein, A.M.; Trower, A.J.; Bishop, D.T.; Iles, M.M.; Duffy, D.L.; et al. Cell-type-specific eQTL of primary melanocytes facilitates identification of melanoma susceptibility genes. *Genome Res.* **2018**, *28*, 1621–1635. [[CrossRef](#)]
143. Zhu, Z.; Zhang, F.; Hu, H.; Bakshi, A.; Robinson, M.R.; Powell, J.E.; Montgomery, G.W.; Goddard, M.E.; Wray, N.R.; Visscher, P.M.; et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* **2016**, *48*, 481–487. [[CrossRef](#)]
144. Korbolina, E.E.; Bryzgalov, L.O.; Ustrokhanova, D.Z.; Postovalov, S.N.; Poverin, D.V.; Damarov, I.S.; Merkulova, T.I. A panel of rSNPs demonstrating allelic asymmetry in both ChIP-seq and RNA-seq data and the search for their phenotypic outcomes through analysis of DEGs. *Int. J. Mol. Sci.* **2021**, (in press).
145. Saha, A.; Kim, Y.; Gewirtz, A.D.H.; Jo, B.; Gao, C.; McDowell, I.C.; Engelhardt, B.E.; Battle, A. Co-expression networks reveal the tissue-specific regulation of transcription and splicing. *Genome Res.* **2017**, *27*, 1843–1858. [[CrossRef](#)]
146. van der Wijst, M.; de Vries, D.; Groot, H.; Trynka, G.; Hon, C.; Bonder, M.; Stegle, O.; Nawijn, M.; Idaghdour, Y.; van der Harst, P.; et al. The single-cell eQTLGen consortium. *Elife* **2020**, *9*. [[CrossRef](#)]
147. Tewhey, R.; Kotliar, D.; Park, D.S.; Liu, B.; Winnicki, S.; Reilly, S.K.; Andersen, K.G.; Mikkelsen, T.S.; Lander, E.S.; Schaffner, S.F.; et al. Direct Identification of Hundreds of Expression-Modulating Variants using a Multiplexed Reporter Assay. *Cell* **2016**, *165*, 1519–1529. [[CrossRef](#)]
148. Richard, A.C.; Peters, J.E.; Lee, J.C.; Vahedi, G.; Schäffer, A.A.; Siegel, R.M.; Lyons, P.A.; Smith, K.G.C. Targeted genomic analysis reveals widespread autoimmune disease association with regulatory variants in the TNF superfamily cytokine signalling network. *Genome Med.* **2016**, *8*, 76. [[CrossRef](#)]
149. Beer, M.A. Predicting enhancer activity and variant impact using gkm-SVM. *Hum. Mutat.* **2017**, *38*, 1251–1258. [[CrossRef](#)]
150. Castel, S.E.; Aguet, F.; Mohammadi, P.; Ardlie, K.G.; Lappalainen, T. A vast resource of allelic expression data spanning human tissues. *Genome Biol.* **2020**, *21*, 234. [[CrossRef](#)]

151. Kang, E.Y.; Martin, L.J.; Mangul, S.; Isvilanonda, W.; Zou, J.; Ben-David, E.; Han, B.; Lusi, A.J.; Shifman, S.; Eskin, E. Discovering Single Nucleotide Polymorphisms Regulating Human Gene Expression Using Allele Specific Expression from RNA-seq Data. *Genetics* **2016**, *204*, 1057–1064. [[CrossRef](#)]
152. Liu, Y.; Li, C.; Shen, S.; Chen, X.; Szlachta, K.; Edmonson, M.N.; Shao, Y.; Ma, X.; Hyle, J.; Wright, S.; et al. Discovery of regulatory noncoding variants in individual cancer genomes by using cis-X. *Nat. Genet.* **2020**, *52*, 811–818. [[CrossRef](#)]
153. Harvey, C.T.; Moyerbrailean, G.A.; Davis, G.O.; Wen, X.; Luca, F.; Pique-Regi, R. QuASAR: Quantitative allele-specific analysis of reads. *Bioinformatics* **2015**, *31*, 1235–1242. [[CrossRef](#)]
154. Edsgård, D.; Iglesias, M.J.; Reilly, S.-J.; Hamsten, A.; Tornvall, P.; Odeberg, J.; Emanuelsson, O. GeneiASE: Detection of condition-dependent and static allele-specific expression from RNA-seq data without haplotype information. *Sci. Rep.* **2016**, *6*, 21134. [[CrossRef](#)]
155. Maurano, M.T.; Haugen, E.; Sandstrom, R.; Vierstra, J.; Shafer, A.; Kaul, R.; Stamatoyannopoulos, J.A. Large-scale identification of sequence variants influencing human transcription factor occupancy in vivo. *Nat. Genet.* **2015**, *47*, 1393–1401. [[CrossRef](#)]
156. Cavalli, M.; Pan, G.; Nord, H.; Wallerman, O.; Wallén Arzt, E.; Berggren, O.; Elvers, I.; Eloranta, M.-L.; Rönnblom, L.; Lindblad Toh, K.; et al. Allele-specific transcription factor binding to common and rare variants associated with disease and gene expression. *Hum. Genet.* **2016**, *135*, 485–497. [[CrossRef](#)]
157. Matys, V. TRANSFAC(R) and its module TRANSCOMP(R): Transcriptional gene regulation in eukaryotes. *Nucleic Acids Res.* **2006**, *34*, D108–D110. [[CrossRef](#)]
158. Portales-Casamar, E.; Thongjuea, S.; Kwon, A.T.; Arenillas, D.; Zhao, X.; Valen, E.; Yusuf, D.; Lenhard, B.; Wasserman, W.W.; Sandelin, A. JASPAR 2010: The greatly expanded open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **2010**, *38*, D105–D110. [[CrossRef](#)] [[PubMed](#)]
159. Newburger, D.E.; Bulyk, M.L. UniPROBE: An online database of protein binding microarray data on protein-DNA interactions. *Nucleic Acids Res.* **2009**, *37*, D77–D82. [[CrossRef](#)] [[PubMed](#)]
160. Jolma, A.; Yan, J.; Whittington, T.; Toivonen, J.; Nitta, K.R.; Rastas, P.; Morgunova, E.; Enge, M.; Taipale, M.; Wei, G.; et al. DNA-Binding Specificities of Human Transcription Factors. *Cell* **2013**, *152*, 327–339. [[CrossRef](#)] [[PubMed](#)]
161. Lappalainen, T.; Sammeth, M.; Friedländer, M.R.; 't Hoen, P.A.C.; Monlong, J.; Rivas, M.A.; González-Porta, M.; Kurbatova, N.; Griebel, T.; Ferreira, P.G.; et al. Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* **2013**, *501*, 506–511. [[CrossRef](#)] [[PubMed](#)]
162. Cavalli, M.; Pan, G.; Nord, H.; Wallén Arzt, E.; Wallerman, O.; Wadelius, C. Allele-specific transcription factor binding in liver and cervix cells unveils many likely drivers of GWAS signals. *Genomics* **2016**, *107*, 248–254. [[CrossRef](#)]
163. Marinov, G.K.; Shipony, Z. Interrogating the Accessible Chromatin Landscape of Eukaryote Genomes Using ATAC-seq. In *Deep Sequencing Data Analysis*; Humana: New York, NY, USA, 2021; pp. 183–226. [[CrossRef](#)]
164. Xu, S.; Feng, W.; Lu, Z.; Yu, C.Y.; Shao, W.; Nakshatri, H.; Reiter, J.L.; Gao, H.; Chu, X.; Wang, Y.; et al. regSNPs-ASB: A Computational Framework for Identifying Allele-Specific Transcription Factor Binding From ATAC-seq Data. *Front. Bioeng. Biotechnol.* **2020**, *8*. [[CrossRef](#)]
165. Benaglio, P.; D'Antonio-Chronowska, A.; Ma, W.; Yang, F.; Young Greenwald, W.W.; Donovan, M.K.R.; DeBoever, C.; Li, H.; Drees, F.; Singhal, S.; et al. Allele-specific NKX2-5 binding underlies multiple genetic associations with human electrocardiographic traits. *Nat. Genet.* **2019**, *51*, 1506–1517. [[CrossRef](#)]
166. van Setten, J.; Brody, J.A.; Jamshidi, Y.; Swenson, B.R.; Butler, A.M.; Campbell, H.; Del Greco, F.M.; Evans, D.S.; Gibson, Q.; Gudbjartsson, D.F.; et al. PR interval genome-wide association meta-analysis identifies 50 loci associated with atrial and atrioventricular electrical activity. *Nat. Commun.* **2018**, *9*, 2904. [[CrossRef](#)]
167. D'Oliveira Albanus, R.; Kyono, Y.; Hensley, J.; Varshney, A.; Orchard, P.; Kitzman, J.O.; Parker, S.C.J. Chromatin information content landscapes inform transcription factor and DNA interactions. *Nat. Commun.* **2021**, *12*, 1307. [[CrossRef](#)]
168. Li, M.; Huang, H.; Li, L.; He, C.; Zhu, L.; Guo, H.; Wang, L.; Liu, J.; Wu, S.; Liu, J.; et al. Core transcription regulatory circuitry orchestrates corneal epithelial homeostasis. *Nat. Commun.* **2021**, *12*, 420. [[CrossRef](#)]
169. Liu, Y.; Guo, B.; Aguilera-Jimenez, E.; Chu, V.S.; Zhou, J.; Wu, Z.; Francis, J.M.; Yang, X.; Choi, P.S.; Bailey, S.D.; et al. Chromatin Looping Shapes KLF5-Dependent Transcriptional Programs in Human Epithelial Cancers. *Cancer Res.* **2020**, *80*, 5464–5477. [[CrossRef](#)]
170. Sun, J.; Zhao, Y.; McGreal, R.; Cohen-Tayar, Y.; Rockowitz, S.; Wilczek, C.; Ashery-Padan, R.; Shechter, D.; Zheng, D.; Cvekl, A. Pax6 associates with H3K4-specific histone methyltransferases Mll1, Mll2, and Set1a and regulates H3K4 methylation at promoters and enhancers. *Epigenet. Chromatin* **2016**, *9*, 37. [[CrossRef](#)]
171. Korbolina, E.E.; Brusentsov, I.I.; Bryzgalov, L.O.; Leberfarb, E.Y.; Degtyareva, A.O.; Merkulova, T.I. Novel approach to functional SNPs discovery from genome-wide data reveals promising variants for colon cancer risk. *Hum. Mutat.* **2018**, *39*, 851–859. [[CrossRef](#)]
172. Bryzgalov, L.O.; Korbolina, E.E.; Brusentsov, I.I.; Leberfarb, E.Y.; Bondar, N.P.; Merkulova, T.I. Novel functional variants at the GWAS-implicated loci might confer risk to major depressive disorder, bipolar affective disorder and schizophrenia. *BMC Neurosci.* **2018**, *19*, 22. [[CrossRef](#)]
173. Seshagiri, S.; Stawiski, E.W.; Durinck, S.; Modrusan, Z.; Storm, E.E.; Conboy, C.B.; Chaudhuri, S.; Guan, Y.; Janakiraman, V.; Jaiswal, B.S.; et al. Recurrent R-spondin fusions in colon cancer. *Nature* **2012**, *488*, 660–664. [[CrossRef](#)] [[PubMed](#)]

174. Leberfarb, E.Y.; Degtyareva, A.O.; Brusentsov, I.I.; Maximov, V.N.; Voevoda, M.I.; Autenshlus, A.I.; Morozov, D.V.; Sokolov, A.V.; Merkulova, T.I. Potential regulatory SNPs in the ATXN7L3B and KRT15 genes are associated with gender-specific colorectal cancer risk. *Per. Med.* **2020**, *17*, 43–54. [[CrossRef](#)]
175. Cavalli, M.; Baltzer, N.; Umer, H.M.; Grau, J.; Lemnian, I.; Pan, G.; Wallerman, O.; Spalinskas, R.; Sahlén, P.; Grosse, I.; et al. Allele specific chromatin signals, 3D interactions, and motif predictions for immune and B cell related diseases. *Sci. Rep.* **2019**, *9*, 2695. [[CrossRef](#)] [[PubMed](#)]
176. Dubois-Chevalier, J.; Mazrooei, P.; Lupien, M.; Staels, B.; Lefebvre, P.; Eeckhoutte, J. Organizing combinatorial transcription factor recruitment at cis-regulatory modules. *Transcription* **2018**, *9*, 233–239. [[CrossRef](#)] [[PubMed](#)]
177. Gerstein, M.B.; Kundaje, A.; Hariharan, M.; Landt, S.G.; Yan, K.-K.; Cheng, C.; Mu, X.J.; Khurana, E.; Rozowsky, J.; Alexander, R.; et al. Architecture of the human regulatory network derived from ENCODE data. *Nature* **2012**, *489*, 91–100. [[CrossRef](#)]
178. Lan, X.; Farnham, P.J.; Jin, V.X. Uncovering Transcription Factor Modules Using One- and Three-dimensional Analyses. *J. Biol. Chem.* **2012**, *287*, 30914–30921. [[CrossRef](#)]
179. Gan, K.A.; Carrasco Pro, S.; Sewell, J.A.; Fuxman Bass, J.I. Identification of Single Nucleotide Non-coding Driver Mutations in Cancer. *Front. Genet.* **2018**, *9*. [[CrossRef](#)]
180. Carrasco Pro, S.; Bulekova, K.; Gregor, B.; Labadorf, A.; Fuxman Bass, J.I. Prediction of genome-wide effects of single nucleotide variants on transcription factor binding. *Sci. Rep.* **2020**, *10*, 17632. [[CrossRef](#)]
181. Badis, G.; Berger, M.F.; Philippakis, A.A.; Talukder, S.; Gehrke, A.R.; Jaeger, S.A.; Chan, E.T.; Metzler, G.; Vedenko, A.; Chen, X.; et al. Diversity and Complexity in DNA Recognition by Transcription Factors. *Science* **2009**, *324*, 1720–1723. [[CrossRef](#)]
182. Nagy, G.; Nagy, L. Motif grammar: The basis of the language of gene expression. *Comput. Struct. Biotechnol. J.* **2020**, *18*, 2026–2032. [[CrossRef](#)]
183. Crocker, J.; Preger-Ben Noon, E.; Stern, D.L. The Soft Touch: Low-affinity transcription factor binding sites in development and evolution. *Curr. Top. Dev. Biol.* **2016**, *117*, 455–469. [[CrossRef](#)]
184. Levitsky, V.G.; Kulakovskiy, I.V.; Ershov, N.I.; Oshchepkov, D.; Makeev, V.J.; Hodgman, T.C.; Merkulova, T.I. Application of experimentally verified transcription factor binding sites models for computational analysis of ChIP-Seq data. *BMC Genom.* **2014**, *15*, 80. [[CrossRef](#)]
185. Levitsky, V.G.; Oshchepkov, D.Y.; Klimova, N.V.; Ignatieva, E.; Vasiliev, G.V.; Merkulov, V.M.; Merkulova, T.I. Hidden heterogeneity of transcription factor binding sites: A case study of SF-1. *Comput. Biol. Chem.* **2016**, *64*, 19–32. [[CrossRef](#)]
186. Osz, J.; McEwen, A.G.; Bourguet, M.; Przybilla, F.; Peluso-Iltis, C.; Poussin-Courmontagne, P.; Mély, Y.; Cianférani, S.; Jeffries, C.M.; Svergun, D.I.; et al. Structural basis for DNA recognition and allosteric control of the retinoic acid receptors RAR-RXR. *Nucleic Acids Res.* **2020**, *48*, 9969–9985. [[CrossRef](#)]
187. Yin, M.; Wang, J.; Wang, M.; Li, X.; Zhang, M.; Wu, Q.; Wang, Y. Molecular mechanism of directional CTCF recognition of a diverse range of genomic sites. *Cell Res.* **2017**, *27*, 1365–1377. [[CrossRef](#)]
188. Afek, A.; Cohen, H.; Barber-Zucker, S.; Gordân, R.; Lukatsky, D.B. Nonconsensus Protein Binding to Repetitive DNA Sequence Elements Significantly Affects Eukaryotic Genomes. *PLoS Comput. Biol.* **2015**, *11*, e1004429. [[CrossRef](#)]
189. Teif, V.B. Soft Power of Nonconsensus Protein-DNA Binding. *Biophys. J.* **2020**, *118*, 1797–1798. [[CrossRef](#)]
190. Zheng, A.; Lamkin, M.; Zhao, H.; Wu, C.; Su, H.; Gymrek, M. Deep neural networks identify sequence context features predictive of transcription factor binding. *Nat. Mach. Intell.* **2021**, *3*, 172–180. [[CrossRef](#)]
191. Wang, S.; Zhang, Q.; Shen, Z.; He, Y.; Chen, Z.-H.; Li, J.; Huang, D.-S. Predicting transcription factor binding sites using DNA shape features based on shared hybrid deep learning architecture. *Mol. Ther. Nucleic Acids* **2021**, *24*, 154–163. [[CrossRef](#)]
192. Wada, K.; Wada, Y.; Ikemura, T. Mb-level CpG and TFBS islands visualized by AI and their roles in the nuclear organization of the human genome. *Genes Genet. Syst.* **2020**, *95*, 29–41. [[CrossRef](#)]
193. Pei, G.; Hu, R.; Dai, Y.; Manuel, A.M.; Zhao, Z.; Jia, P. Predicting regulatory variants using a dense epigenomic mapped CNN model elucidated the molecular basis of trait-tissue associations. *Nucleic Acids Res.* **2021**, *49*, 53–66. [[CrossRef](#)]
194. Jing, F.; Zhang, S.-W.; Cao, Z.; Zhang, S. An Integrative Framework for Combining Sequence and Epigenomic Data to Predict Transcription Factor Binding Sites Using Deep Learning. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2021**, *18*, 355–364. [[CrossRef](#)]
195. Chen, C.; Hou, J.; Shi, X.; Yang, H.; Birchler, J.A.; Cheng, J. DeepGRN: Prediction of transcription factor binding site across cell-types using attention-based deep neural networks. *BMC Bioinform.* **2021**, *22*, 38. [[CrossRef](#)] [[PubMed](#)]