ISSN 0006-2979, Biochemistry (Moscow), 2024, Vol. 89, No. 6, pp. 1002-1013 © The Author(s) 2024. This article is an open access publication.

REVIEW

Methods for Functional Characterization of Genetic Polymorphisms of Non-Coding Regulatory Regions of the Human Genome

Aksinya N. Uvarova[1,a]*, Elena A. Tkachenko[2,3], Ekaterina M. Stasevich[1,4], Elina A. Zheremyan[1], Kirill V. Korneev[1], and Dmitry V. Kuprash[1,3]

[1]*Center for Precision Genome Editing and Genetic Technologies for Biomedicine, Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, 119991 Moscow, Russia*

[2]*Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, 119991 Moscow, Russia*

[3]*Faculty of Biology, Lomonosov Moscow State University, 119234 Moscow, Russia*

[4]*Moscow Institute of Physics and Technology, 141700 Dolgoprudny, Moscow Region, Russia*

[a]*e-mail: uvarowww@gmail.com*

Received November 20, 2023
Revised March 27, 2024
Accepted April 11, 2024

Abstract—Currently, numerous associations between genetic polymorphisms and various diseases have been characterized through the Genome-Wide Association Studies. Majority of the clinically significant polymorphisms are localized in non-coding regions of the genome. While modern bioinformatic resources make it possible to predict molecular mechanisms that explain influence of the non-coding polymorphisms on gene expression, such hypotheses require experimental verification. This review discusses the methods for elucidating molecular mechanisms underlying dependence of the disease pathogenesis on specific genetic variants within the non-coding sequences. A particular focus is on the methods for identification of transcription factors with binding efficiency dependent on polymorphic variations. Despite remarkable progress in bioinformatic resources enabling prediction of the impact of polymorphisms on the disease pathogenesis, there is still the need for experimental approaches to investigate this issue.

DOI: 10.1134/S0006297924060026

Keywords: functional polymorphisms, regulatory genomic regions, transcription factors, reporter analysis, CRISPR-Cas

INTRODUCTION

In spite of the fact that human genomes are identical by 99.9%, it is precisely the remaining 0.1% of genetic variants that underlie phenotypic differences, including susceptibility to diseases [1]. These genetic variations are Single Nucleotide Variation (SNV) or Single Nucleotide Polymorphism (SNP), insertion/deletion (indel), and Structural Variation of more than 50 b.p. in length (SV) [2]. The most widespread genetic variation is SNP, i.e., a DNA sequence variation (a variant allele) of one nucleotide in size in the members of the same species, which occurs within a population at a frequency of at least 1% [3]. SNPs occur every 200-300 b.p. in the genome, being localized in its coding and regulatory parts (promoters, enhancers, introns, and untranslated regions) [4, 5]. Importance of studying SNP lies in the fact that such genetic variants are often associated with different diseases, as it has been shown by numerous Genome-Wide Association Studies (GWAS). About 95% of the clinically significant SNPs are localized in non-coding genome regions [6],

Abbreviations: ChIP, chromatin immunoprecipitation; CRISPR, clustered regularly interspaced short palindromic repeats; HDR, homology-directed repair; MPRA, massively parallel reporter assay; QTL, quantitative trait locus; raQTL, reporter assay quantitative trait locus; SNP, single nucleotide polymorphism; TF, transcription factor; UTR, untranslated region.

* To whom correspondence should be addressed.

and their functional significance is probably associated with the changes in the regulatory characteristics of the regions surrounding the polymorphism [7]. Such regulatory regions of the eukaryotic genome may be promoters, enhancers, 5′- and 3′-untranslated regions (UTR) of protein-coding genes, gene regions of non-coding RNA (ncRNA), and splicing regulatory elements (SRE) [5, 8]. Promoters initiate gene transcription and enhancer elements increase the rate of this initiation [9]. Promoters are preferred sites for binding transcription factors (TFs) and RNA polymerase II to DNA and include the region of the first transcribed nucleotide of the transcript (transcription start site, TSS) [10]. Enhancers, which have been identified for the first time with the help of reporter analysis as elements capable of enhancing the reporter gene expression [11], are the platforms for TF binding that can act irrespective of orientation, distance, and localization relative to the target gene [12]. The 5′- and 3′-UTRs play an important role in post-transcriptional regulation of gene expression and are part of mature coding mRNA. For example, 5′-UTRs contain different regulatory components influencing translation initiation, and 3′-UTRs comprise the sequences that bind microRNA and lead to transcript degradation [5]. In addition, it should be noted that the non-coding polymorphisms within UTR could also be involved in transcription regulation, because the 5′-UTR sequence usually overlaps with the promoter regions of the genes, while the 3′-UTR sequence could overlap with other regulatory elements of the genes, e.g., enhancers [13]. Non-coding polymorphisms are also localized in ncRNA; in recent years, a lot of information has been obtained about their effects on RNA maturation, transcription regulation, chromatin remodeling, and post-transcriptional modifications of RNA [14].

Being the most frequently occurring class of genetic variants, SNPs are the major genetic marker for Quantitative Trait Loci (QTL) mapping; they further could be conditionally divided into those regulating gene expression directly at the transcriptional and chromatin levels, exerting effect on the mRNA level (eQTL – expression QTL regulating gene expression at the transcriptional level), and those influencing post-transcriptional processes (sQTL – splicing QTL regulating alternative splicing of pre-mRNA; pQTL – protein QTL regulating protein expression) [15]. The following mechanism of functional effects of polymorphisms at the genomic level could be suggested: functions of the regulatory elements are impaired due to the change in the sequence of the sites for TF–DNA interaction (both decrease and increase in binding efficiency) [16]. At the post-transcriptional level, non-coding polymorphisms could affect activity of the 5′- and 3′-UTR mRNA, which play a key role in translation regulation and mRNA stability, including due to the change in the regulatory microRNA binding [17-19]. In addition, SNPs in the sequence of immature microRNAs could affect efficiency of microRNA maturation and change efficiency of mRNA binding [20, 21], and allele variants within the lnRNA (long non-coding RNA) could modulate, with different efficiency, concentration of the complementary microRNA [22]. Considerable number of functional genetic variants classified as sQTL is localized in the splicing regulatory elements, directly changing sequence of the splicing sites or modifying binding sites for the RNA-binding proteins [23]. The main mechanisms of effects of non-coding polymorphisms on gene expression regulation are shown in Fig. 1.

The present review describes main experimental approaches to the analysis of functional non-coding allele variations, including the methods for determination of TF with binding efficiency dependent on the allele variation.

ANALYSIS OF THE EFFECTS OF GENETIC POLYMORPHISMS ON GENE EXPRESSION USING REPORTER GENES

Experimental methods used for studying the effects of polymorphisms on gene expression can be divided into two large groups: the studies with involvement of genetic reporter constructs and the studies of polymorphisms directly in the native genomic context.

The former group of methods involves reporter genetic constructs, where the effect of genetic variant on a regulatory element is determined by the reporter gene activity (reporter assay QTL, raQTL). They also involve classical technique of luciferase reporter analysis, where the allele variants of the regulatory sequence under study (a promoter or an enhancer) are integrated into the reporter construct, and activities of the reporter genes in the resultant constructs are compared after their transfection into a physiologically relevant cell types [24]. The method of dual-luciferase assay was used to describe numerous raQTL in different types of cells and regulatory elements. For example, the previously proposed molecular mechanisms explain relationship between the development of diseases and genetic polymorphisms localized in the regulatory regions of different genes: promoters [25-28], closely located enhancer regions [29, 30], and enhancers located at distant intergenic loci [31, 32].

Search for sQTL in the systems using reporter genes is usually limited by the size of the gene under study, if it exceeds capacity of the reporter plasmid. In such situation, the so-called reporter minigenes are used. A minigene construct includes a fragment of the studied locus containing a polymorphism and sufficient for reproduction of the natural pattern of splicing between the splicing reporters (as a rule, between two

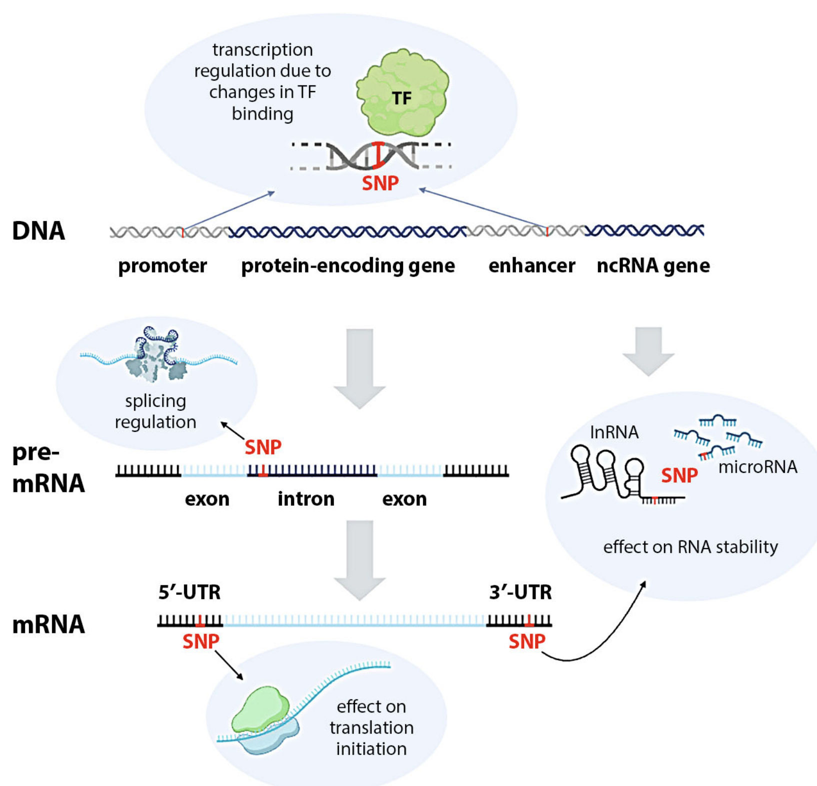

Fig. 1. Main mechanisms of the effects of non-coding polymorphisms on regulation of gene expression (the image was produced using BioRender.com).

exons). Ability of the region under study to influence splicing efficiency is measured based on expression of the target transcript, or, in a nuclear extract, by quantitative PCR analysis (qPCR), or, in live cells, if the reporter protein encoded by the minigene allows it [33]. For example, the reporter analysis of minigenes was used to characterize polymorphisms regulating gene splicing of the *SCN1A* subunit of the calcium channel associated with epilepsy [34], the *RAD51C* component of DNA repair system acting as a tumor suppressor [35], etc.

HIGH-THROUGHPUT REPORTER ASSAYS

Over the past decade, numerous modifications of high-throughput reporter assays have been developed. They can be categorized according to the regulatory regions studying of which they allow, as well as according to technological features used. For example, the Massively Parallel Reporter Assay (MPRA) protocol involves synthesis of DNA sequences (potential enhancers/promoters, 5′-UTR or 3′-UTR) with addition of unique barcodes and cloning of these sequences into reporter plasmids, which are next transfected into the cell types of interest. Activity of the regulatory regions is analyzed using high-throughput sequencing and quantification of barcodes, which unambiguously determine a particular regulatory sequence and correlate with the RNA level of the reporter gene [36, 37]. MPRA is obviously suitable not only for studying functionality of the regulatory elements but also for assessing functional effects of their genetic variants [38]. For example, MPRA was used for screening polymorphisms located in the non-coding regions of the genomes and associated with schizophrenia and Alzheimer's disease. It is interesting that only 9 out of 148 SNPs with the allele differences in the K562 and 53 cells have exhibited allele differences in both SK-SY5Y cell lines, clearly demonstrating that genetic variants usually exert their regulatory effects only in certain types of cells [39]. MPRA applied to the library of human gene 5′-UTRs made it possible to reveal 45 disease-associated allele variations exerting significant effects on the process of loading mRNAs onto ribosomes; however, it is interesting that the data on most of the revealed variants proved to be insufficient to change classification of pathogenicity in the Clinvar database, and the most striking effect was demonstrated by 3 polymorphisms generating a new start codon, i.e., affecting protein structure [40]. In another work, Griesemer studied more than 12,000 3′-UTR variants from 6 human cell lines, which were associated with human diseases and/or were under positive pressure in the human population [41]. It turned out that several hundreds of them had significant effects on the level of reporter transcript in at least one cell line, and several tens

of them coincided with the previously characterized variants with any level of clinical significance. Interestingly, only for two SNPs present in the *TRIM14* gene of viral defense and in the *PILRB* gene associated with age-related macular degeneration, combination of novelty and degree of influence on the level of the transcript proved to be sufficient to verify the hypothesis using the Cas9-mediated allele substitution in the genomic context [41]. Technological limitations of MPRA also include length of the tested DNA fragments (up to 130-230 b.p.) and number of the tested constructs (up to 100-200 thousand sequences) [42].

There are high-throughput approaches involving sequences obtained from the genomic DNA. For example, SuRE (Survey of Regulatory Elements) was developed with the involvement of sequencing data for the genomes of cell lines originating from four different ethnic groups and optimized to study potential effect of a single-nucleotide substitution on activity of the regulatory elements [43]. Random fragments of genomic DNA of several hundred b.p. in length are cloned in the reporter plasmid without a promoter, which, when transfected into the cultured cells, produce a transcript only if the inserted fragment carries a functional transcription start site. Since transcripts can produce both active promoters and enhancers, this method makes it possible to analyze activities of both types of regulatory elements. Like in the MPRA technique, the transcripts are analyzed by high-throughput sequencing and quantified with involvement of a barcode unique for each genomic fragment under study. This approach makes it possible to test activity of the regulatory elements containing alternative alleles of several millions of different SNPs (i.e., most of the known ones).

One another method that allows raQTL identification – High-resolution Dissection of Regulatory Activity (HiDRA) [42] – also involves fragmentation of genomic DNA and is a combination of the ATAC-seq and STARR-seq techniques. ATAC-seq (Assay for Transposase-Accessible Chromatin using sequencing) makes it possible to enrich samples with the transposase-accessible, i.e., open, chromatin, while STARR-seq (Self-Transcribing Active Regulatory Region sequencing) is a reporter assay, where putative regulatory elements (which are able to enhance transcriptional activity of the reporter) are cloned in 3′-UTR of the reporter gene and thereby promote their own transcription. Next, active DNA sequences are identified and quantified by high-throughput RNA sequencing [44]. For example, HiDRA made it possible to identify a 76-b.p. driver element in the *IKZF3* gene locus, which included rs12946510 associated with multiple sclerosis; hence, this SNP can be identified as potentially functional [42]. Indeed, further functional testing showed that presence of the risk allele rs12946510 reduced activation of T helpers and expression of the *IKZF3* and *ORMDL3* genes [45]. An important stage of processing the results of each of the high-throughput methods described above is using probabilistic mathematical models such as, e.g., the SHARPR-RE algorithm (Systematic High-resolution Activation and Repression Prediction from Reporter assays with Random Endpoints) [46], for analysis of the sequence overlap and assessment of the effects of particular nucleotides on activity of these sequences.

High-throughput reporter assays of polymorphic variants include Massively Parallel Splicing Assay (MaPSY) [47], which was used to study impaired splicing in the case of autism spectrum disorders. The screening results were used to characterize genetic variants in the *TNRC6C*, *MAPK8IP1*, and *USP45* genes, and it has been shown that the proteins of TNRC6 family could increase the risk of autism development [48]. Recently, the method of Cre-dependent MPRA *in vivo* has been proposed for functional analysis of the library of 3′-UTRs with genetic variants associated with autism. Quantification of the transcripts depending on activity of the regulatory element was performed in particular types of neurons by transduction of the libraries into the brain tissues of mice with tissue-specific expression of Cre recombinase. This method makes it possible to study regulatory effect in a more relevant cellular context, because neurons have an absolutely different expression profile of trans-acting factors (e.g., TF and microRNA) compared to other cell lines [49].

Main limitation of the methods based on reporter assays is absence of the relevant chromatin context, which accompanies the regulatory element in the native genome. This limitation is partially eliminated in the lentiMPRA technique, when library with the regulatory elements under study is created in a lentiviral vector, which is integrated into the genome, facilitating analysis of transcription within the chromatin context [50].

FUNCTIONAL ANALYSIS OF GENETIC POLYMORPHISMS IN THE NATIVE GENOMIC CONTEXT

With regard to the effects of genetic variants on pathogenesis of a disease, it is important to take into account chromatin context which, in turn, varies between the different types and functional states of the cells. The eQTL mapping *per se* makes it possible to relate a particular genotype to the changes in mRNA levels of potential target genes in the native genomic context, including tissue specificity [51, 52]. Functional relationship between the genes and distant regulatory loci can be found by determining 3D chromatin organization using methods such as Hi-C (high-throughput chromosome conformation capture), ChIA-PET (chromatin interaction analysis with paired-

end tag sequencing), and their modifications [53, 54]. Comparison of the 3D tissue-specific genomic maps with disease-associated regulatory SNPs makes it possible to identify the most probable genes involved in pathogenesis. Hence, the most accurate method for verification of hypotheses constructed is genome editing and producing of cells with the desired combinations of variants. Precise and efficient editing of particular nucleotides in the human genome has become a daunting but realistic challenge due to the RNA-programmable bacterial nucleases found in the CRISPR (clustered regularly interspaced short palindromic repeats)-Cas system [55]. The double-strand break (DSB) in DNA induced in the target site by the Cas9 nuclease from *Streptococcus pyogenes* (currently, the most popular genome editor) triggers cellular mechanisms of DNA repair, including homology-directed repair (HDR) [56], which is used in the CRISPR-HDR methods, when the target region is repaired in the presence of a homologous DNA sequence containing the necessary allele variant.

This method used in many polymorphism studies [57, 58] has a significant limitation with respect to efficiency, because DSB repair in mammals occurs mainly with involvement of nonhomologous end joining (NHEJ) [59]. Due to these peculiarities, the CRISPR-HDR editing takes a lot of efforts and could lead to the impaired expression of the neighboring genes [45]. Another approach to precise genome editing, which performs well in a proper nucleotide context, is base editing (BE) with involvement of the catalytically inactive dCas9 (dead Cas9) or Cas9 with nickase activity (nCas9) fused with deaminase. With respect to enzyme specificity, there are cytosine (converting C•G into T•A) and adenine (converting A•T into G•C) editors, as well as an editor based on cytidine deaminase and uracil-DNA glycosylase (converting C•G into G•C) [60-63]. For example, the cytosine base editing system was used to study polymorphism rs12603332 associated with the risk of asthma and demonstrated its effect on expression of the genes of sphingolipid biosynthesis regulator *ORMDL3* and cellular stress response modulator *ATF6α* in the Jurkat T-cell line [64]. Due to the absence of the stage of DSB formation, the BE technique is safer for the cells than CRISPR-HDR; however, it has limitations with respect to enzyme activities and off-target editing of the neighboring nucleotides [60, 61]. Another recently developed and promising approach to genome editing is prime editing. The editor is based on the mutant nuclease Cas9 inserting single-strand breaks (nCas9) fused with the reverse transcriptase (MMLV RT) and uses a modified guide RNA (pegRNA), which simultaneously determines the target site for nCas9, acts as a primer for MMLV RT, and is an RNA template for the synthesis of a new DNA sequence. The edited DNA strand is then included into the genome by endogenous cellular processes [65].

Due to the high-precision of editing and wider area of its application compared to the standard base editors, prime editing has a great potential for working with single-nucleotide polymorphisms. Single-nucleotide substitutions are used for directed evolution in the selection of agricultural crops [66, 67]. Mouse models were used to demonstrate low off-target activity of prime editing when changing the variant of non-coding polymorphism compared to CRISPR-HDR [68]. In addition, prime editing in human myoblasts was used to correct mutation in the protein-coding region of the calcium channel gene *RYR1* associated with motor impairments [69]. Major fundamental limitation of prime editing is large size of the editor and difficulties with its delivery into the cells [70]. In the case of high-throughput screening systems development, the problem of delivery could be solved using lentiviral transduction of the target cells by the constructs encoding the editor and pegRNAs. Subsequent cultivation of the cells for several weeks makes it possible to achieve editing efficiency sufficient for studying functional effects of hundreds and even thousands of single-nucleotide substitutions in a single experiment [71].

IDENTIFICATION OF TRANSCRIPTION FACTORS MEDIATING THE ALLELE-SPECIFIC DIFFERENCE IN THE ACTIVITIES OF REGULATORY ELEMENTS

Identification of different types of QTL (eQTL, raQTL, etc.) does not provide information about particular molecular mechanism affected by a particular genetic variant; therefore, further functional annotation remains relevant. As mentioned before, the mechanisms of effects of polymorphisms on the functions of regulatory element include changes in the properties of promoter and enhancer regions, 5′-UTR and 3′-UTR, ncRNA, as well as impaired splicing. The best studied cause of the dependence of the properties of regulatory elements on SNPs localized in them is capability of the single-nucleotide substitution to influence affinity to the functional transcription factor.

There are various *in silico* approaches to predict the preferred TF binding motifs, in most cases based on Positional Weight Matrices (PWM) formed by the multiple alignment of TF-binding sequences [72, 73]. In turn, information about the particular TF-binding sequences can be obtained by high-throughput genome-wide mapping of binding sites *in vivo*, e.g., methods based on Chromatin Immunoprecipitation (ChIP) or on high-throughput systemic evolution of ligands by exponential enrichment (HT-SELEX) for the selection of TF-binding sequences *in vitro* [74]. Genome sequences associated with the specific proteins in their native chromatin context are identified by the ChIP-seq tech-

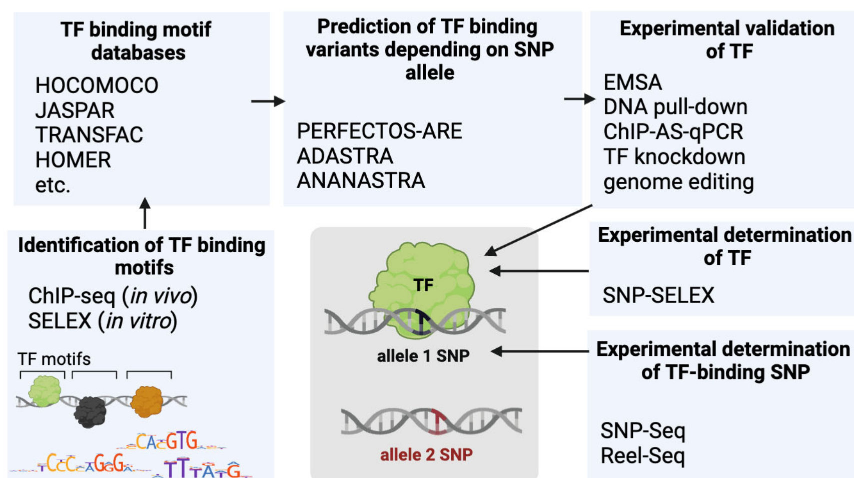

Fig. 2. Methods for identification of functional transcription factors with allele-specific binding to the region of polymorphism (the image was produced using BioRender.com).

nique combining chromatin immunoprecipitation with subsequent high-throughput DNA sequencing [75]. Sequences optimal for binding of a particular TF (probably not existing in nature) are found with the involvement of SELEX methods for enrichment of the libraries of randomly generated oligonucleotides with specific sequences exhibiting high affinity to a given TF [76]. The are well-known PWM motif databases including TRANSFAC [77], HOCOMOCO [78], JASPAR [79], HOMER [80], iRegulon [81], etc. Application of bioinformatics makes it possible to assess potential changes in the strength of TF binding depending on the variant of polymorphism. Efficiency of the allele-specific TF binding can be estimated directly by the ChIP-Seq data, if sequencing depth allows detection of the statistically significant deviations in the frequencies of alternative SNP alleles in the binding site [82, 83]. Combination of ChIP with quantification of alleles, ChIP-AS-qPCR (ChIP-based allele-specific quantitative PCR), makes it possible to measure effects of the allele variants on efficiency of TF binding in a living cell [57]. A high-throughput variant of the analysis of TF binding with polymorphisms in the regulatory regions, SNP-SELEX, based on the HT-SELEX has been proposed. This method allows analysis of the effects of about 100,000 allele variants of the potentially regulatory (GWAS-annotated) SNPs on binding of several hundreds of TFs [84]. Classical method of analysis of DNA–protein interactions based on the shifts in electrophoretic mobility (electrophoretic mobility shift assay, EMSA) can also be considered as an experimental approach to TF identification. During EMSA, proteins under study specifically bind to the labeled oligonucleotide probes, which is followed by analysis of mobility of such fragments using electrophoresis in polyacrylamide gel under native conditions; relative strength of the binding could be assessed based on the amount of the formed complex [85]. Specificity of determination of protein components in the complexes is achieved by adding antibodies against a specific protein in the reaction: EMSA–supershift [86]. There are also high-throughput methods for analysis of large amounts of SNP allowing to find out effects of the allele variants on TF binding based on incubation of the SNP-containing oligonucleotides with a nuclear extract from the particular cell type, followed by sequencing of the enriched libraries; such methods are SNPs-Seq [57] and Reel-Seq [87]. Neither of these methods *per se* makes it possible to establish, which TF binds to a particular allele variant; however, such information could be obtained by mass spectrometry and/or using a purified TF instead of the nuclear extract [24, 88].

Bioinformatics databases suitable for analysis of SNP of interest include the on-line resource PERFECTOS-ARE https://opera.autosome.org/perfectosape [76], where the predicted TF binding motifs are collected from various databases: HOCOMOCO [78], JASPAR [79], HT-SELEX [89], etc. Another bioinformatics resource, ADASTRA [82], that provides comprehensive data on the allele-specific TF binding with allele variants in different types of cells, is based on the HOCOMOCO and SPRy-SARUS data [90], as well as on the allele-specific data of the DNase footprinting assay [91]. The ANANASTRA resource [92] based on the systematic analysis of allelic imbalance in the ChIP-Seq experiments, makes it possible to annotate a great number of genetic variants in parallel.

One of the examples of using such annotation could be functional characterization of the SNPs rs7873784 and rs71327024 localized in the regulatory regions of the *TLR4* and *CXCR6* genes, respectively [13, 31]. According to the results of GWAS, both SNPs are disease-associated: the minor C allele of rs7873784 is associated with rheumatoid arthritis and the minor T allele of rs71327024 is associated with

severe COVID-19. The reporter assays have shown that both SNPs are raQTL; therefore, bioinformatics analysis was used to find TFs PU.1 (rs7873784) and c-Myb (rs71327024) relevant for the respective types of cells characterized by the allele-dependent binding to SNP-containing sites. This hypothesis was verified using the genetic knockdown of TF with involvement of small interfering RNA (siRNA), as well as the DNA pull-down immunoprecipitation technique [93]. The latter includes incubation of oligonucleotides containing alternative SNP variants with the nuclear extract from the relevant cells and immunoprecipitation with the specific antibodies against the predicted TF, followed by quantification of the enriched oligonucleotides. The described methods for identification of transcription factors with binding efficiency depending on the allele of polymorphism are shown in Fig. 2.

Due to continuously increasing amounts of data and modern machine learning models, bioinformatic computations provide a more precise annotation of the candidate TFs with allele-specific binding to the SNP region [94-96]. However, clinical validation and *a fortiori* application of these data in diagnostics and probably treatment of the diseases are possible only after experimental validation in different types of cells in the relevant functional context.

CONCLUSIONS

To date, meta-analysis of large amounts of experimental data makes it possible to develop bioinformatics tools for searching for the most probable functional genetic variants, as well as for prediction of particular mechanisms of their effects on pathogenesis of the diseases. Overwhelming majority of the genetic variants are localized in the non-coding regions of the genome; they affect functions of the genes by regulating their expression. Such regulation could vary widely depending on the type and functional state of cells, which is not always taken into consideration in the case of *in silico* methods involving statistical generalizations. In view of the above, it is still relevant to use versatile experimental techniques for characterization of particular genetic variants. The most informative method for studying effects of the genetic variants on phenotype is development of precise genetic models using genome editing techniques. However, due to the difficult procedure of precise genome editing, preliminary characterization of allele variants under study by the reporter assays remains relevant.

Contributions. A.N.U. concept and supervision of the work; A.N.U., E.A.T., E.M.S., and E.A.Zh. writing the manuscript; K.V.K. and D.V.K. editing the manuscript.

Funding. The work was financially supported by the Russian Science Foundation (project no. 22-24-00987).

Ethics declarations. This work does not contain any studies involving human and animal subjects. The authors of this work declare that they have no conflicts of interest.

Open access. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution, and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third-party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit https://creativecommons.org/licenses/by/4.0/.

REFERENCES

1. Ahmed, Z., Zeeshan, S., Mendhe, D., and Dong, X. (2020) Human gene and disease associations for clinical-genomics and precision medicine research, *Clin. Transl. Med.*, **10**, 297-318, doi: 10.1002/ctm2.28.
2. Lappalainen, T., Scott, A. J., Brandt, M., and Hall, I. M. (2019) Genomic analysis in the age of human genome sequencing, *Cell*, **177**, 70-84, doi: 10.1016/j.cell.2019.02.032.
3. Wright, A. F. (2005) Genetic variation: polymorphisms and mutations, in *eLS*, doi: 10.1038/npg.els.0005005.
4. Salisbury, B. A., Pungliya, M., Choi, J. Y., Jiang, R., Sun, X. J., and Stephens, J. C. (2003) SNP and haplotype variation in the human genome, *Mutat. Res.*, **526**, 53-61, doi: 10.1016/S0027-5107(03)00014-9.
5. Fabo, T., and Khavari, P. (2023) Functional characterization of human genomic variation linked to polygenic diseases, *Trends Genet.*, **39**, 462-490, doi: 10.1016/j.tig.2023.02.014.
6. Orozco, G., Schoenfelder, S., Walker, N., Eyre, S., and Fraser, P. (2022) 3D genome organization links non-coding disease-associated variants to genes, *Front. Cell Dev. Biol.*, **10**, 995388, doi: 10.3389/FCELL.2022.995388/BIBTEX.
7. Johnston, A. D., Simões-Pires, C. A., Thompson, T. V., Suzuki, M., and Greally, J. M. (2019) Functional genetic variants can mediate their regulatory effects through alteration of transcription factor binding, *Nat. Commun.*, **10**, 3472, doi: 10.1038/s41467-019-11412-5.
8. Grodecká, L., Buratti, E., and Freiberger, T. (2017) Mutations of pre-mRNA splicing regulatory elements: Are

predictions moving forward to clinical diagnostics? *Int. J. Mol. Sci.*, **18**, 1668, doi: 10.3390/ijms18081668.

9. Andersson, R., and Sandelin, A. (2020) Determinants of enhancer and promoter activities of regulatory elements, *Nat. Rev. Genet.*, **21**, 71-87, doi: 10.1038/S41576-019-0173-8.
10. Carninci, P., Sandelin, A., Lenhard, B., Katayama, S., Shimokawa, K., Ponjavic, J., Semple, C. A. M., Taylor, M. S., Engström, P. G., Frith, M. C., Forrest, A. R. R., Alkema, W. B., Tan, S. L., Plessy, C., Kodzius, R., Ravasi, T., Kasukawa, T., Fukuda, S., Kanamori-Katayama, M., Kitazume, Y., Kawaji, H., Kai, C., Nakamura, M., Konno, H., Nakano, K., Mottagui-Tabar, S., Arner, P., Chesi, A., Gustincich, S., Persichetti, F., Suzuki, H., Grimmond, S. M., Wells, C. A., Orlando, V., Wahlestedt, C., Liu, E. T., Harbers, M., Kawai, J., Bajic, V. B., Hume, D. A., and Hayashizaki, Y. (2006) Genome-wide analysis of mammalian promoter architecture and evolution, *Nat. Genet.*, **38**, 626-635, doi: 10.1038/NG1789.
11. Banerji, J., Rusconi, S., and Schaffner, W. (1981) Expression of a β-globin gene is enhanced by remote SV40 DNA sequences, *Cell*, **27**, 299-308, doi: 10.1016/0092-8674(81)90413-X.
12. Krivega, I., and Dean, A. (2012) Enhancer and promoter interactions-long distance calls, *Curr. Opin. Genet. Dev.*, **22**, 79-85, doi: 10.1016/j.gde.2011.11.001.
13. 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., and Kuprash, D. V. (2020) 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.*, **1866**, 165626, doi: 10.1016/j.bbadis.2019.165626.
14. Panni, S., Lovering, R. C., Porras, P., and Orchard, S. (2020) Non-coding RNA regulatory networks, *Biochim. Biophys. Acta Gene Regul. Mech.*, **1863**, 194417, doi: 10.1016/j.bbagrm.2019.194417.
15. Lappalainen, T., and MacArthur, D. G. (2021) From variant to function in human disease genetics, *Science*, **373**, 1464-1468, doi: 10.1126/science.abi8207.
16. Tseng, C. C., Wong, M. C., Liao, W. T., Chen, C. J., Lee, S. C., Yen, J. H., and Chang, S. J. (2021) Genetic variants in transcription factor binding sites in humans: triggered by natural selection and triggers of diseases, *Int. J. Mol. Sci.*, **22**, 4187, doi: 10.3390/ijms22084187.
17. Pan, X., Zhao, J., Zhou, Z., Chen, J., Yang, Z., Wu, Y., Bai, M., Jiao, Y., Yang, Y., Hu, X., Cheng, T., Lu, Q., Wang, B., Li, C. L., Lu, Y. J., Diao, L., Zhong, Y. Q., Pan, J., Zhu, J., Xiao, H. S., Qiu, Z. L., Li, J., Wang, Z., Hui, J., Bao, L., and Zhang, X. (2021) 5′-UTR SNP of *FGF13* causes translational defect and intellectual disability, *eLife*, **10**, e63021, doi: 10.7554/eLife.63021.
18. Cui, Y., Peng, F., Wang, D., Li, Y., Li, J. S., Li, L., and Li, W. (2022) 3′aQTL-atlas: An atlas of 3′UTR alternative polyadenylation quantitative trait loci across human normal tissues, *Nucleic Acids Res.*, **50**, D39-D45, doi: 10.1093/nar/gkab740.
19. Chhichholiya, Y., Suryan, A. K., Suman, P., Munshi, A., and Singh, S. (2021) SNPs in miRNAs and target sequences: role in cancer and diabetes, *Front. Genet.*, **12**, 793523, doi: 10.3389/fgene.2021.793523.
20. Hrdlickova, B., de Almeida, R. C., Borek, Z., and Withoff, S. (2014) Genetic variation in the non-coding genome: Involvement of micro-RNAs and long non-coding RNAs in disease, *Biochim. Biophys. Acta Mol. Basis Dis.*, **1842**, 1910-1922, doi: 10.1016/j.bbadis.2014.03.011.
21. Rykova, E., Ershov, N., Damarov, I., and Merkulova, T. (2022) SNPs in 3′UTR miRNA target sequences associated with individual drug susceptibility, *Int. J. Mol. Sci.*, **23**, 13725, doi: 10.3390/ijms232213725.
22. Feng, T., Feng, N., Zhu, T., Li, Q., Zhang, Q., Wang, Y., Gao, M., Zhou, B., Yu, H., Zheng, M., and Qian, B. (2020) A SNP-mediated lncRNA (LOC146880) and microRNA (miR-539-5p) interaction and its potential impact on the NSCLC risk, *J. Exp. Clin. Cancer Res.*, **39**, 157, doi: 10.1186/s13046-020-01652-5.
23. Garrido-Martín, D., Borsari, B., Calvo, M., Reverter, F., and Guigó, R. (2021) Identification and analysis of splicing quantitative trait loci across multiple tissues in the human genome, *Nat. Commun.*, **12**, 727, doi: 10.1038/s41467-020-20578-2.
24. Degtyareva, A. O., Antontseva, E. V., and Merkulova, T. I. (2021) Regulatory snps: Altered transcription factor binding sites implicated in complex traits and diseases, *Int. J. Mol. Sci.*, **22**, 6454, doi: 10.3390/ijms22126454.
25. Gorbacheva, A. M., Korneev, K. V., Kuprash, D. V., and Mitkin, N. A. (2018) The risk G allele of the single-nucleotide polymorphism rs928413 creates a CREB1-binding site that activates IL33 promoter in lung epithelial cells, *Int. J. Mol. Sci.*, **19**, 2911, doi: 10.3390/ijms19102911.
26. Putlyaeva, L. V., Demin, D. E., Korneev, K. V., Kasyanov, A. S., Tatosyan, K. A., Kulakovskiy, I. V., Kuprash, D. V., and Schwartz, A. M. (2018) Potential markers of autoimmune diseases, alleles rs115662534(T) and rs548231435(C), disrupt the binding of transcription factors STAT1 and EBF1 to the regulatory elements of human CD40 gene, *Biochemistry (Moscow)*, **83**, 1534-1542, doi: 10.1134/S0006297918120118.
27. Zhou, J., To, K. K. W., Dong, H., Cheng, Z. S., Lau, C. C. Y., Poon, V. K. M., Fan, Y. H., Song, Y. Q., Tse, H., Chan, K. H., Zheng, B. J., Zhao, G. P., and Yuen, K. Y. (2012) A functional variation in CD55 increases the severity of 2009 pandemic H1N1 influenza a virus infection, *J. Infect. Dis.*, **206**, 495-503, doi: 10.1093/infdis/jis378.

28. Matveeva, M. Y., Kashina, E. V., Reshetnikov, V. V., Bryzgalov, L. O., Antontseva, E. V., Bondar, N. P., and Merkulova, T. I. (2016) Regulatory single nucleotide polymorphisms (rSNPs) at the promoters 1A and 1B of the human APC gene, *BMC Genet.*, **17**, 127-135, doi: 10.1186/s12863-016-0460-8.
29. Mitkin, N. A., Muratova, A. M., Korneev, K. V., Pavshintsev, V. V., Rumyantsev, K. A., Vagida, M. S., Uvarova, A. N., Afanasyeva, M. A., Schwartz, A. M., and Kuprash, D. V. (2018) Protective C allele of the single-nucleotide polymorphism rs1335532 is associated with strong binding of Ascl2 transcription factor and elevated CD58 expression in B-cells, *Biochim. Biophys. Acta Mol. Basis Dis.*, **1864**, 3211-3220, doi: 10.1016/j.bbadis.2018.07.008.
30. Uvarova, A. N., Ustiugova, A. S., Mitkin, N. A., Schwartz, A. M., Korneev, K. V., and Kuprash, D. V. (2022) The minor T allele of the single nucleotide polymorphism rs13360222 decreases the activity of the HAVCR2 gene enhancer in a cell model of human macrophages, *Mol. Biol.*, **56**, 90-96, doi: 10.1134/S0026893322010095.
31. Uvarova, A. N., Stasevich, E. M., Ustiugova, A. S., Mitkin, N. A., Zheremyan, E. A., Sheetikov, S. A., Zornikova, K. V., Bogolyubova, A. V., Rubtsov, M. A., Kulakovskiy, I. V., Kuprash, D. V., Korneev, K. V., and Schwartz, A. M. (2023) rs71327024 Associated with COVID-19 hospitalization reduces CXCR6 promoter activity in human CD4$^+$ T cells via disruption of c-Myb binding, *Int. J. Mol. Sci.*, **24**, 13790, doi: 10.3390/IJMS241813790.
32. Ustiugova, A. S., Korneev, K. V., Kuprash, D. V., and Afanasyeva, M. A. (2019) Functional SNPs in the human autoimmunity-associated locus 17q12-21, *Genes*, **10**, 77, doi: 10.3390/GENES10020077.
33. Cooper, T. A. (2005) Use of minigene systems to dissect alternative splicing elements, *Methods*, **37**, 331-340, doi: 10.1016/J.YMETH.2005.07.015.
34. Sparber, P., Sharova, M., Davydenko, K., Pyankov, D., Filatova, A., and Skoblov, M. (2023) Deciphering the impact of coding and non-coding SCN1A gene variants on RNA splicing, *Brain*, **147**, 1278-1293, doi: 10.1093/BRAIN/AWAD383.
35. Sanoguera-Miralles, L., Bueno-Martínez, E., Valenzuela-Palomo, A., Esteban-Sánchez, A., Llinares-Burguet, I., Pérez-Segura, P., García-Álvarez, A., de la Hoya, M., and Velasco-Sampedro, E. A. (2022) Minigene splicing assays identify 20 spliceogenic variants of the breast/ovarian cancer susceptibility gene RAD51C, *Cancers*, **14**, 2960, doi: 10.3390/CANCERS14122960.
36. Nguyen, T. A., Jones, R. D., Snavely, A. R., Pfenning, A. R., Kirchner, R., Hemberg, M., and Gray, J. M. (2016) High-throughput functional comparison of promoter and enhancer activities, *Genome Res.*, **26**, 1023-1033, doi: 10.1101/GR.204834.116.
37. Melnikov, A., Murugan, A., Zhang, X., Tesileanu, T., Wang, L., Rogov, P., Feizi, S., Gnirke, A., Callan, C. G., Kinney, J. B., Kellis, M., Lander, E. S., and Mikkelsen, T. S. (2012) Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay, *Nat. Biotechnol.*, **30**, 271-277, doi: 10.1038/nbt.2137.
38. 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., and Sabeti, P. C. (2016) Direct identification of hundreds of expression-modulating variants using a multiplexed reporter assay, *Cell*, **172**, 1519-1529, doi: 10.1016/j.cell.2018.02.021.
39. Myint, L., Wang, R., Boukas, L., Hansen, K. D., Goff, L. A., and Avramopoulos, D. (2020) A screen of 1,049 schizophrenia and 30 Alzheimer's-associated variants for regulatory potential, *Am. J. Med. Genet. Part B Neuropsychiatr. Genet.*, **183**, 61-73, doi: 10.1002/AJMG.B.32761.
40. Sample, P. J., Wang, B., Reid, D. W., Presnyak, V., McFadyen, I. J., Morris, D. R., and Seelig, G. (2019) Human 5′ UTR design and variant effect prediction from a massively parallel translation assay, *Nat. Biotechnol.*, **37**, 803-809, doi: 10.1038/s41587-019-0164-5.
41. Griesemer, D., Xue, J. R., Reilly, S. K., Ulirsch, J. C., Kukreja, K., Davis, J. R., Kanai, M., Yang, D. K., Butts, J. C., Guney, M. H., Luban, J., Montgomery, S. B., Finucane, H. K., Novina, C. D., Tewhey, R., and Sabeti, P. C. (2021) Genome-wide functional screen of 3′UTR variants uncovers causal variants for human disease and evolution, *Cell*, **184**, 5247-5260, doi: 10.1016/j.cell.2021.08.025.
42. Wang, X., He, L., Goggin, S. M., Saadat, A., Wang, L., Sinnott-Armstrong, N., Claussnitzer, M., and Kellis, M. (2018) High-resolution genome-wide functional dissection of transcriptional regulatory regions and nucleotides in human, *Nat. Commun.*, **9**, 5380, doi: 10.1038/s41467-018-07746-1.
43. van Arensbergen, J., Pagie, L., FitzPatrick, V. D., de Haas, M., Baltissen, M. P., Comoglio, F., van der Weide, R. H., Teunissen, H., Võsa, U., Franke, L., de Wit, E., Vermeulen, M., Bussemaker, H. J., and van Steensel, B. (2019) High-throughput identification of human SNPs affecting regulatory element activity, *Nat. Genet.*, **51**, 1160-1169, doi: 10.1038/s41588-019-0455-2.
44. Arnold, C. D., Gerlach, D., Stelzer, C., Boryń, Ł. M., Rath, M., and Stark, A. (2013) Genome-wide quantitative enhancer activity maps identified by STARR-seq, *Science*, **339**, 1074-1077, doi: 10.1126/SCIENCE.1232542.
45. Ustiugova, A. S., Dvorianinova, E. M., Melnikova, N. V., Dmitriev, A. A., Kuprash, D. V., and Afanasyeva, M. A. (2023) CRISPR/Cas9 genome editing demonstrates functionality of the autoimmunity-associated SNP rs12946510, *Biochim. Biophys. Acta Mol. Basis Dis.*, **1869**, 166599, doi: 10.1016/j.bbadis.2022.166599.

46. Ernst, J., Melnikov, A., Zhang, X., Wang, L., Rogov, P., Mikkelsen, T. S., and Kellis, M. (2016) Genome-scale high-resolution mapping of activating and repressive nucleotides in regulatory regions, *Nat. Biotechnol.*, **34**, 1180-1190, doi: 10.1038/nbt.3678.
47. Soemedi, R., Cygan, K. J., Rhine, C. L., Wang, J., Bulacan, C., Yang, J., Bayrak-Toydemir, P., McDonald, J., and Fairbrother, W. G. (2017) Pathogenic variants that alter protein code often disrupt splicing, *Nat. Genet.*, **49**, 848-855, doi: 10.1038/ng.3837.
48. Rhine, C. L., Neil, C., Wang, J., Maguire, S., Buerer, L., Salomon, M., Meremikwu, I. C., Kim, J., Strande, N. T., and Fairbrother, W. G. (2022) Massively parallel reporter assays discover *de novo* exonic splicing mutants in paralogs of Autism genes, *PLoS Genet.*, **18**, e1009884, doi: 10.1371/journal.pgen.1009884.
49. Lagunas, T., Plassmeyer, S. P., Fischer, A. D., Friedman, R. Z., Rieger, M. A., Selmanovic, D., Sarafinovska, S., Sol, Y. K., Kasper, M. J., Fass, S. B., Aguilar Lucero, A. F., An, J. Y., Sanders, S. J., Cohen, B. A., and Dougherty, J. D. (2023) A Cre-dependent massively parallel reporter assay allows for cell-type specific assessment of the functional effects of non-coding elements *in vivo*, *Commun. Biol.*, **6**, 1151, doi: 10.1038/s42003-023-05483-w.
50. Gordon, M. G., Inoue, F., Martin, B., Schubach, M., Agarwal, V., Whalen, S., Feng, S., Zhao, J., Ashuach, T., Ziffra, R., Kreimer, A., Georgakopoulous-Soares, I., Yosef, N., Ye, C. J., Pollard, K. S., Shendure, J., Kircher, M., and Ahituv, N. (2020) lentiMPRA and MPRAflow for high-throughput functional characterization of gene regulatory elements, *Nat. Protoc.*, **15**, 2387-2412, doi: 10.1038/s41596-020-0333-5.
51. GTEx Consortium (2020) The GTEx Consortium atlas of genetic regulatory effects across human tissues, *Science*, **369**, 1318-1330, doi: 10.1126/science.aaz1776.
52. Bryois, J., Calini, D., Macnair, W., Foo, L., Urich, E., Ortmann, W., Iglesias, V. A., Selvaraj, S., Nutma, E., Marzin, M., Amor, S., Williams, A., Castelo-Branco, G., Menon, V., De Jager, P., and Malhotra, D. (2022) Cell-type-specific cis-eQTLs in eight human brain cell types identify novel risk genes for psychiatric and neurological disorders, *Nat. Neurosci.*, **25**, 1104-1112, doi: 10.1038/s41593-022-01128-z.
53. Capurso, D., Tang, Z., and Ruan, Y. (2020) Methods for comparative ChIA-PET and Hi-C data analysis, *Methods*, **170**, 69-74, doi: 10.1016/J.YMETH.2019.09.019.
54. Huang, L., Yang, Y., Li, G., Jiang, M., Wen, J., Abnousi, A., Rosen, J. D., Hu, M., and Li, Y. (2022) A systematic evaluation of Hi-C data enhancement methods for enhancing PLAC-seq and HiChIP data, *Brief. Bioinform.*, **23**, bbac145, doi: 10.1093/BIB/BBAC145.
55. Khalil, A. M. (2020) The genome editing revolution: review, *J. Genet. Eng. Biotechnol.*, **18**, 68, doi: 10.1186/S43141-020-00078-Y.
56. Moon, S. B., Kim, D. Y., Ko, J. H., and Kim, Y. S. (2019) Recent advances in the CRISPR genome editing tool set, *Exp. Mol. Med.*, **51**, 1-11, doi: 10.1038/s12276-019-0339-7.
57. Zhang, P., Xia, J. H., Zhu, J., Gao, P., Tian, Y. J., Du, M., Guo, Y. C., Suleman, S., Zhang, Q., Kohli, M., Tillmans, L. S., Thibodeau, S. N., French, A. J., Cerhan, J. R., Wang, L. D., Wei, G. H., and Wang, L. (2018) High-throughput screening of prostate cancer risk loci by single nucleotide polymorphisms sequencing, *Nat. Commun.*, **9**, 2022, doi: 10.1038/s41467-018-04451-x.
58. Rodríguez-Rodríguez, D. R., Ramírez-Solís, R., Garza-Elizondo, M. A., Garza-Rodríguez, M. D. L., and Barrera-Saldaña, H. A. (2019) Genome editing: a perspective on the application of CRISPR/Cas9 to study human diseases (Review), *Int. J. Mol. Med.*, **43**, 1559-1574, doi: 10.3892/ijmm.2019.4112.
59. Yang, H., Ren, S., Yu, S., Pan, H., Li, T., Ge, S., Zhang, J., and Xia, N. (2020) Methods favoring homology-directed repair choice in response to CRISPR/Cas9 Induced-double strand breaks, *Int. J. Mol. Sci.*, **21**, 6461, doi: 10.3390/IJMS21186461.
60. Rees, H. A., and Liu, D. R. (2018) Base editing: precision chemistry on the genome and transcriptome of living cells, *Nat. Rev. Genet.*, **19**, 770-788, doi: 10.1038/s41576-018-0059-1.
61. Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A., and Liu, D. R. (2016) Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage, *Nature*, **533**, 420-424, doi: 10.1038/nature17946.
62. Gaudelli, N. M., Komor, A. C., Rees, H. A., Packer, M. S., Badran, A. H., Bryson, D. I., and Liu, D. R. (2017) Programmable base editing of A·T to G·C in genomic DNA without DNA cleavage, *Nature*, **551**, 464-471, doi: 10.1038/nature24644.
63. Zhao, D., Li, J., Li, S., Xin, X., Hu, M., Price, M. A., Rosser, S. J., Bi, C., and Zhang, X. (2021) Glycosylase base editors enable C-to-A and C-to-G base changes, *Nat. Biotechnol.*, **39**, 35-40, doi: 10.1038/s41587-020-0592-2.
64. Weng, N., Miller, M., Pham, A. K., Komor, A. C., and Broide, D. H. (2022) Single-base editing of rs12603332 on chromosome 17q21 with a cytosine base editor regulates ORMDL3 and ATF6α expression, *Allergy*, **77**, 1139-1149, doi: 10.1111/ALL.15092.
65. Anzalone, A. V., Randolph, P. B., Davis, J. R., Sousa, A. A., Koblan, L. W., Levy, J. M., Chen, P. J., Wilson, C., Newby, G. A., Raguram, A., and Liu, D. R. (2019) Search-and-replace genome editing without double-strand breaks or donor DNA, *Nature*, **576**, 149-157, doi: 10.1038/s41586-019-1711-4.
66. Jiang, Y., Chai, Y., Qiao, D., Wang, J., Xin, C., Sun, W., Cao, Z., Zhang, Y., Zhou, Y., Wang, X. C., and Chen, Q. J. (2022) Optimized prime editing efficiently generates glyphosate-resistant rice plants carrying homozygous TAP-IVS mutation in EPSPS, *Mol. Plant*, **15**, 1646-1649, doi: 10.1016/j.molp.2022.09.006.
67. Hassan, M. M., Yuan, G., Chen, J. G., Tuskan, G. A., and Yang, X. (2020) Prime editing technology and

its prospects for future applications in plant biology research, *BioDes. Res.*, **2020**, 9350905, doi: 10.34133/2020/9350905.

68. Gao, P., Lyu, Q., Ghanam, A. R., Lazzarotto, C. R., Newby, G. A., Zhang, W., Choi, M., Slivano, O. J., Holden, K., Walker, J. A., Kadina, A. P., Munroe, R. J., Abratte, C. M., Schimenti, J. C., Liu, D. R., Tsai, S. Q., Long, X., and Miano, J. M. (2021) Prime editing in mice reveals the essentiality of a single base in driving tissue-specific gene expression, *Genome Biol.*, **22**, 83, doi: 10.1186/s13059-021-02304-3.
69. Godbout, K., Rousseau, J., and Tremblay, J. P. (2023) Successful correction by prime editing of a mutation in the *RYR1* gene responsible for a myopathy, *Cells*, **13**, 31, doi: 10.3390/CELLS13010031.
70. Petrova, I. O., and Smirnikhina, S. A. (2023) The development, optimization and future of prime editing, *Int. J. Mol. Sci.*, **24**, 17045, doi: 10.3390/IJMS242317045.
71. Ren, X., Yang, H., Nierenberg, J. L., Sun, Y., Chen, J., Beaman, C., Pham, T., Nobuhara, M., Takagi, M. A., Narayan, V., Li, Y., Ziv, E., and Shen, Y. (2023) High-throughput PRIME-editing screens identify functional DNA variants in the human genome, *Mol. Cell*, **83**, 4633-4645.e9, doi: 10.1016/J.MOLCEL.2023.11.021.
72. Ambrosini, G., Vorontsov, I., Penzar, D., Groux, R., Fornes, O., Nikolaeva, D. D., Ballester, B., Grau, J., Grosse, I., Makeev, V., Kulakovskiy, I., and Bucher, P. (2020) Insights gained from a comprehensive all-against-all transcription factor binding motif benchmarking study, *Genome Biol.*, **21**, 114, doi: 10.1186/s13059-020-01996-3.
73. Lambert, S. A., Jolma, A., Campitelli, L. F., Das, P. K., Yin, Y., Albu, M., Chen, X., Taipale, J., Hughes, T. R., and Weirauch, M. T. (2018) The human transcription factors, *Cell*, **172**, 650-665, doi: 10.1016/J.CELL.2018.01.029.
74. Tognon, M., Giugno, R., and Pinello, L. (2023) A survey on algorithms to characterize transcription factor binding sites, *Brief Bioinform.*, **24**, bbad156, doi: 10.1093/bib/bbad156.
75. Mundade, R., Ozer, H. G., Wei, H., Prabhu, L., and Lu, T. (2014) Role of ChIP-seq in the discovery of transcription factor binding sites, differential gene regulation mechanism, epigenetic marks and beyond, *Cell Cycle*, **13**, 2847-2852, doi: 10.4161/15384101.2014.949201.
76. Vorontsov, I. E., Kulakovskiy, I. V., Khimulya, G., Nikolaeva, D. D., and Makeev, V. J. (2015) PERFECTOS-APE: Predicting regulatory functional effect of SNPs by approximate P-value estimation, *Bioinforma. 2015 – 6th Int. Conf. Bioinforma. Model. Methods Algorithms, Proceedings; Part 8th Int. Jt. Conf. Biomed. Eng. Syst. Technol., BIOSTEC 2015*, **2**, 102-108, doi: 10.5220/0005189301020108.
77. Wingender, E., Chen, X., Fricke, E., Geffers, R., Hehl, R., Liebich, I., Krull, M., Matys, V., Michael, H., Ohnhäuser, R., Prüß, M., Schacherer, F., Thiele, S., and Urbach, S. (2001) The TRANSFAC system on gene expression regulation, *Nucleic Acids Res.*, **29**, 281-283, doi: 10.1093/nar/29.1.281.
78. Vorontsov, I. E., Eliseeva, I. A., Zinkevich, A., Nikonov, M., Abramov, S., Boytsov, A., Kamenets, V., Kasianova, A., Kolmykov, S., Yevshin, I. S., Favorov, A., Medvedeva, Y. A., Jolma, A., Kolpakov, F., Makeev, V. J., and Kulakovskiy, I. V. (2024) HOCOMOCO in 2024: a rebuild of the curated collection of binding models for human and mouse transcription factors, *Nucleic Acids Res.*, **52**, D154-D163, doi: 10.1093/NAR/GKAD1077.
79. Castro-Mondragon, J. A., Riudavets-Puig, R., Rauluseviciute, I., Berhanu Lemma, R., Turchi, L., Blanc-Mathieu, R., Lucas, J., Boddie, P., Khan, A., Perez, N. M., Fornes, O., Leung, T. Y., Aguirre, A., Hammal, F., Schmelter, D., Baranasic, D., Ballester, B., Sandelin, A., Lenhard, B., Vandepoele, K., Wasserman, W. W., Parcy, F., and Mathelier, A. (2022) JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles, *Nucleic Acids Res.*, **50**, D165-D173, doi: 10.1093/NAR/GKAB1113.
80. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., Cheng, J. X., Murre, C., Singh, H., and Glass, C. K. (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities, *Mol. Cell*, **38**, 576-589, doi: 10.1016/j.molcel.2010.05.004.
81. Janky, R., Verfaillie, A., Imrichová, H., van de Sande, B., Standaert, L., Christiaens, V., Hulselmans, G., Herten, K., Naval Sanchez, M., Potier, D., Svetlichnyy, D., Kalender Atak, Z., Fiers, M., Marine, J. C., and Aerts, S. (2014) iRegulon: from a gene list to a gene regulatory network using large motif and track collections, *PLoS Comput. Biol.*, **10**, e1003731, doi: 10.1371/journal.pcbi.1003731.
82. Abramov, S., Boytsov, A., Bykova, D., Penzar, D. D., Yevshin, I., Kolmykov, S. K., Fridman, M. V., Favorov, A. V., Vorontsov, I. E., Baulin, E., Kolpakov, F., Makeev, V. J., and Kulakovskiy, I. V. (2021) Landscape of allele-specific transcription factor binding in the human genome, *Nat. Commun.*, **12**, 2751, doi: 10.1038/s41467-021-23007-0.
83. Li, Y., Zhang, X. O., Liu, Y., and Lu, A. (2023) Allele-specific binding (ASB) analyzer for annotation of allele-specific binding SNPs, *BMC Bioinform.*, **24**, 464, doi: 10.1186/S12859-023-05604-6.
84. Yan, J., Qiu, Y., Ribeiro dos Santos, A. M., Yin, Y., Li, Y. E., Vinckier, N., Nariai, N., Benaglio, P., Raman, A., Li, X., Fan, S., Chiou, J., Chen, F., Frazer, K. A., Gaulton, K. J., Sander, M., Taipale, J., and Ren, B. (2021) Systematic analysis of binding of transcription factors to noncoding variants, *Nature*, **591**, 147-151, doi: 10.1038/s41586-021-03211-0.
85. Hellman, L. M., and Fried, M. G. (2007) Electrophoretic mobility shift assay (EMSA) for detecting protein–

nucleic acid interactions, *Nat. Protoc.*, **2**, 1849-1861, doi: 10.1038/nprot.2007.249.

86. Parés-Matos, E. I. (2013) Electrophoretic mobility-shift and super-shift assays for studies and characterization of protein-DNA complexes, *Methods Mol. Biol.*, **977**, 159-167, doi: 10.1007/978-1-62703-284-1_12.
87. Zhao, Y., Wu, D., Jiang, D., Zhang, X., Wu, T., Cui, J., Qian, M., Zhao, J., Oesterreich, S., Sun, W., Finkel, T., and Li, G. (2020) A sequential methodology for the rapid identification and characterization of breast cancer-associated functional SNPs, *Nat. Commun.*, **11**, 3340, doi: 10.1038/s41467-020-17159-8.
88. Butter, F., Davison, L., Viturawong, T., Scheibe, M., Vermeulen, M., Todd, J. A., and Mann, M. (2012) Proteome-wide analysis of disease-associated SNPs that show allele-specific transcription factor binding, *PLoS Genet.*, **8**, e1002982, doi: 10.1371/journal.pgen.1002982.
89. Jolma, A., Kivioja, T., Toivonen, J., Cheng, L., Wei, G., Enge, M., Taipale, M., Vaquerizas, J. M., Yan, J., Sillanpää, M. J., Bonke, M., Palin, K., Talukder, S., Hughes, T. R., Luscombe, N. M., Ukkonen, E., and Taipale, J. (2010) Multiplexed massively parallel SELEX for characterization of human transcription factor binding specificities, *Genome Res.*, **20**, 861-873, doi: 10.1101/gr.100552.109.
90. Mille, M., Ripoll, J., Cazaux, B., and Rivals, E. (2023) dipwmsearch: a Python package for searching di-PWM motifs, *Bioinformatics*, **39**, btad141, doi: 10.1093/BIOINFORMATICS/BTAD141.
91. Maurano, M. T., Haugen, E., Sandstrom, R., Vierstra, J., Shafer, A., Kaul, R., and Stamatoyannopoulos, J. A. (2015) Large-scale identification of sequence variants influencing human transcription factor occupancy *in vivo*, *Nat. Genet.*, **47**, 1393-1401, doi: 10.1038/ng.3432.
92. Boytsov, A., Abramov, S., Aiusheeva, A. Z., Kasianova, A. M., Baulin, E., Kuznetsov, I. A., Aulchenko, Y. S., Kolmykov, S., Yevshin, I., Kolpakov, F., Vorontsov, I. E., Makeev, V. J., and Kulakovskiy, I. V. (2022) ANANASTRA: annotation and enrichment analysis of allele-specific transcription factor binding at SNPs, *Nucleic Acids Res.*, **50**, W51-W56, doi: 10.1093/nar/gkac262.
93. Mitkin, N. A., Korneev, K. V., Gorbacheva, A. M., and Kuprash, D. V. (2019) Relative efficiency of transcription factor binding to allelic variants of regulatory regions of human genes in immunoprecipitation and real-time PCR, *Mol. Biol.*, **53**, 346-353, doi: 10.1134/S0026893319030117.
94. Yevshin, I., Sharipov, R., Valeev, T., Kel, A., and Kolpakov, F. (2017) GTRD: a database of transcription factor binding sites identified by ChIP-seq experiments, *Nucleic Acids Res.*, **45**, D61-D67, doi: 10.1093/nar/GKW951.
95. Zhang, Y., Mo, Q., Xue, L., and Luo, J. (2021) Evaluation of deep learning approaches for modeling transcription factor sequence specificity, *Genomics*, **113**, 3774-3781, doi: 10.1016/J.YGENO.2021.09.009.
96. Chen, C., Hou, J., Shi, X., Yang, H., Birchler, J. A., and Cheng, J. (2021) DeepGRN: prediction of transcription factor binding site across cell-types using attention-based deep neural networks, *BMC Bioinformatics*, **22**, 38, doi: 10.1186/S12859-020-03952-1.

Publisher's Note. Pleiades Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.