OXFORD

Briefings in Functional Genomics, 2023, 1–12

https://doi.org/10.1093/bfgp/elad020

Protocol Article

Integrating functional scoring and regulatory data to predict the effect of non-coding SNPs in a complex neurological disease

Daniela Felício, Miguel Alves-Ferreira*, Mariana Santos, Marlene Quintas, Alexandra M. Lopes, Carolina Lemos, Nádia Pinto and Sandra Martins

*Corresponding author: Miguel Alves-Ferreira. The Population Studies Department, ICBAS School of Medicine and Biomedical Sciences, Universidade do Porto, Porto 4050-313, Portugal. Tel: (+351) 220428181. E-mail: miguel.ferreira@ibmc.up.pt

Abstract

Most SNPs associated with complex diseases seem to lie in non-coding regions of the genome; however, their contribution to gene expression and disease phenotype remains poorly understood. Here, we established a workflow to provide assistance in prioritising the functional relevance of non-coding SNPs of candidate genes as susceptibility *loci* in polygenic neurological disorders. To illustrate the applicability of our workflow, we considered the multifactorial disorder migraine as a model to follow our step-by-step approach. We annotated the overlap of selected SNPs with regulatory elements and assessed their potential impact on gene expression based on publicly available prediction algorithms and functional genomics information.

Some migraine risk *loci* have been hypothesised to reside in non-coding regions and to be implicated in the neurotransmission pathway. In this study, we used a set of 22 non-coding SNPs from neurotransmission and synaptic machinery-related genes previously suggested to be involved in migraine susceptibility based on our candidate gene association studies. After prioritising these SNPs, we focused on non-reported ones that demonstrated high regulatory potential: (1) *VAMP2*_rs1150 (3′ UTR) was predicted as a target of hsa-mir-5010-3p miRNA, possibly disrupting its own gene expression; (2) *STX1A*_rs6951030 (proximal enhancer) may affect the binding affinity of zinc-finger transcription factors (namely ZNF423) and disturb *TBL2* gene expression; and (3) *SNAP25*_rs2327264 (distal enhancer) expected to be in a binding site of ONECUT2 transcription factor.

This study demonstrated the applicability of our practical workflow to facilitate the prioritisation of potentially relevant non-coding SNPs and predict their functional impact in multifactorial neurological diseases.

Keywords: migraine susceptibility; non-coding SNPs; regulatory annotation; epigenomic databases; prediction algorithms

Introduction

In complex diseases, the identification of susceptibility *loci* and genes usually implies detecting differences in single nucleotide polymorphisms (SNPs) allele frequencies between cases and controls. Genome-wide association studies and candidate gene association studies, as well as linkage studies, have contributed to the identification of many risk genes and related pathways; however, the disease-associated SNP may not necessarily be functionally relevant but be in linkage disequilibrium with the (still unidentified) deleterious SNP [1–3]. An additional problem underlies non-coding SNPs since they are often excluded from these analyses or classified as variants of uncertain significance, mainly due to difficulties in predicting or defining their functional impact [4].

Most SNPs associated with complex diseases lie in intronic or intergenic regions, which makes them more likely to confer disease susceptibility by altering gene regulation rather than affecting protein function [2, 5]. Recently, regulatory data related to chromatin accessibility and state became publicly available [6–8], proposing that SNPs overlapping regulatory elements might alter the binding sites of gene expression regulators such as transcription factors (TFs) and micro RNAs (miRNAs) [2, 5]. These

Daniela Felício is a doctoral student in the Molecular and Cell Biology doctoral program at ICBAS, Univ Porto. Her research interest involves using genomic data, bioinformatics analyses, and molecular biology techniques to understand neurological disorders.

Miguel Alves-Ferreira is a clinical researcher at CGPP-IBMC/i3S. His research focuses on applying genomics, medical genetics and bioinformatics approaches to understand human diseases.

Mariana Santos is a researcher at IBMC/i3S, Univ Porto. Her current research covers the genetics and pathophysiology of neurological disorders, focusing on hereditary cerebellar ataxias and spastic paraplegias.

Marlene Quintas is a doctoral candidate at IBMC/i3S, Univ Porto. Her main research focus is on the genetics of migraine while simultaneously being interested in exploring the study of RNA dysfunction in neurological disorders.

Alexandra M. Lopes is a clinical researcher at CGPP-IMBC/i3S. Her research interests are focused on genomic medicine.

Carolina Lemos is a professor in the Institute of Biomedical Sciences, Univ Porto. Her research focus on genetic epidemiology and genetic modifiers of monogenic and complex traits.

Nádia Pinto is a researcher at IPATIMUP/i3S. Her research has been focused in the development of mathematical models and software for a wide range of research problems in the field of population genetics.

Sandra Martins is a researcher at IPATIMUP/i3S, mainly focused on the evolution of highly conserved paralogs, genetic epidemiology and identification of modifiers in neurological diseases.

Received: January 26, 2023. **Revised:** March 13, 2023. **Accepted:** May 9, 2023

© The Author(s) 2023. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oup.com

Downloaded from https://academic.oup.com/bfg/advance-article/doi/10.1093/bfgp/elad020/7186543 by University of Porto user on 20 June 2023

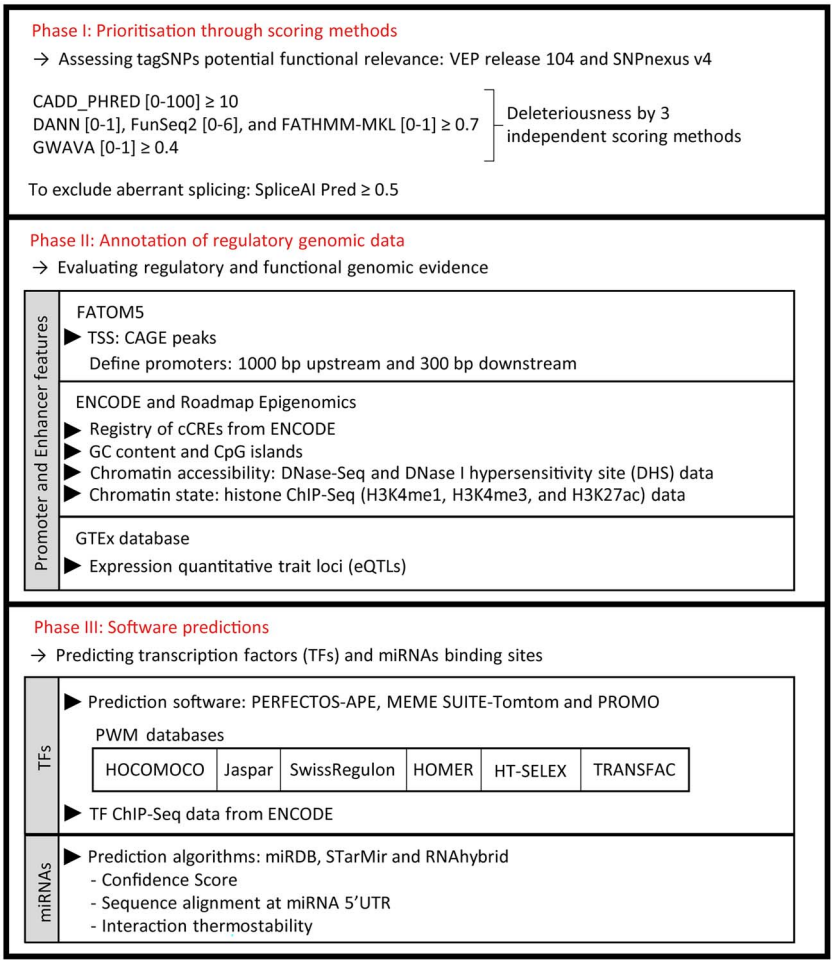

Figure 1. Workflow to prioritise regulatory SNPs possibly associated with complex neurological disorders through feature annotation with Ensembl VEP tool and SNPnexus (IW Scoring), and functional genomics analyses with FANTOM, ENCODE and Roadmap Epigenomic databases (Phases I and II). To assess if candidate regulatory SNPs may alter important motifs for gene expression, it is crucial to predict TFs (PERFECTOS-APE, MEME SUITE—Tomtom and PROMO) and miRNAs (miRDB, STarMir and RNAhybrid) binding sites (Phase III).

data sources offer an opportunity to interpret non-coding SNPs and find common regulatory features necessary to discriminate functionally relevant SNPs from benign ones [1]. A regulatory SNP located in an enhancer/promoter may increase or decrease the expression of a given gene by affecting the binding affinity or creating a novel binding site for TFs or miRNAs [4]. A bioinformatics approach can be used to predict the effect of SNPs by recurring to epigenomic and genomic data, and to determine probable transcription start sites (TSSs), splice sites, TFs and miRNAs binding sites [2, 4, 9]. Yet, the impact of regulatory SNPs is difficult to interpret since (1) quantitative rather than qualitative effects on gene expression may be present, and (2) distinct tissues, developmental stages and even individuals may be affected differently [4, 10].

In our study, we developed a detailed protocol to prioritise common regulatory non-coding SNPs from synaptic machinery-related genes, previously found to be associated with migraine risk or protection in the Portuguese population [11–13]. Recent studies suggested that exonic SNPs do not seem to explain disease in the majority of migraineurs, which stresses the importance of studying the role of functionally relevant non-coding SNPs [14, 15]. Migraine is a common multifactorial neurological disease with a strong genetic component, which affects about 15% of the population [16, 17]. This type of primary headache disorder includes several symptoms, such as unilateral throbbing, photophobia, nausea and/or vomiting [18]. Studies show an increased familial risk for common migraine, with heritability estimated between 30 and 60% that most likely results from the contribution of SNPs at several *loci*, each with a small effect [19–21]. Some rare autosomal dominant forms of migraine are caused by exonic variants in genes related to neurotransmission (e.g. *CACNA1A*, *ATP1A2*, and *SCN1A*) [22, 23], still the molecular mechanisms underpinning this disease are poorly understood. Thus, we developed a step-by-step bioinformatics protocol to select common non-coding SNPs within putative regulatory regions of genes likely associated with complex neurological diseases.

Methods

We designed a practical workflow to prioritise functionally relevant non-coding SNPs as potential susceptibility *loci* in complex neurological disorders. The overview of the protocol is shown in Figure 1. Details of the protocol are described below.

Phase I: prioritisation through scoring methods

To evaluate and prioritise the selected non-coding SNPs according to their functional potential, we used Variant Effect Prediction (VEP) tool [24] (Ensembl, https://www.ensembl.org/Tools/VEP; release 104; MANE transcripts) and SNPnexus v4—Integrative

Downloaded from https://academic.oup.com/bfg/advance-article/doi/10.1093/bfgp/elad020/7186543 by University of Porto user on 20 June 2023

Weighted (IW) Scoring model [25, 26] (Barts Cancer Institute, https://snp-nexus.org/IW-Scoring/). We performed an integrative assessment of the available scoring methods to evaluate non-coding variants [27]. A higher score in Combined Annotation Dependent Depletion (CADD), Deleterious Annotation of genetic variants using Neural Networks (DANN) (from VEP), Genome Wide Annotation of VAriants (GWAVA), Functional Analysis Through Hidden Markov Models (FATHMM) and FunSeq2 (from SNPnexus-IW) indicates a higher probability for the SNP to be functionally relevant (or deleterious). First, we selected SNPs that reached at least three scoring annotation methods indicating deleteriousness [referred as integration score (counts)], with cut-off values set at ≥ 10 for CADD_PHRED [range 0–100], ≥ 0.7 for FunSeq2 [range 0–6], DANN [range 0–1] and FATHMM-MKL (non-coding score; [range 0–1]), and ≥ 0.4 for GWAVA (Region, TSS, Unmatched; [range 0–1]), as recommended in software guidelines and previous publications [28, 29]. In addition, to detect and rank the SNPs with the highest potential, we normalised CADD_PHRED and FunSeq2 scores to 0–1 range to have all algorithms contributing equally for the integration score (sum).

Next, we used SpliceAI Pred (VEP tool) to predict if the non-coding SNPs disrupted mRNA splicing. A SpliceAI Pred score (DS_AG—acceptor gain, DS_AL—acceptor loss, DS_DG—donor gain, and DS_DL—donor loss; DS—delta score) ≥ 0.5 indicated aberrant splicing.

In this process, features such as genomic localisation, SNP allele frequencies (gnomAD and 1000 Genomes), GC content and evolutionary conservation (GERP++, PhastCons100way and PhyloP100way) were annotated with the VEP tool. Also, the Ensembl Regulatory Build data (VEP tool) were used for the feature annotation step to assess regulatory elements such as promoters, proximal or distal enhancers and CTCF binding sites, where SNPs under analysis may be located. The VEP tool includes a search for PubMed identifiers of SNPs previously associated with a phenotype or disease, which allowed us to explore the literature and avert non-coding SNPs already experimentally confirmed to regulate gene expression.

Phase II: annotation of regulatory genomic data

To assess the characteristics of the region where selected SNPs are located, we analysed relevant epigenomic and regulatory data from FANTOM (v5, https://fantom.gsc.riken.jp/5/) [8, 30], ENCODE (v127.2, https://www.encodeproject.org/) [6] and Roadmap Epigenomics Mapping Consortium (http://www.roadmapepigenomics.org/) [7, 31]. These databases allowed the visualisation of epigenomic data and concomitant selection of genomic regions and tissues for an integrative analysis (UCSC Genome Browser [32, 33] and WashU Epigenome Browser v46.2 [34]). Brief guidelines for the use of these databases and visualisation tools can be found in Supplementary Table 1.

The chromatin status (DNase-seq or ATAC-seq data), active chromatin histone marks (H3K4me1, H3K4me3, H3K27ac and H3K27me3 ChIP-seq data) and DNase I hypersensitive sites (DHSs) were assessed from ENCODE and Roadmap Epigenomics. Whenever available, these annotations were taken from experiments in human brain tissues or neuronal cell lines. In addition, the distance of the SNP to the nearest TSS, as well as GC content and the presence of CpG islands (clusters of CpG dinucleotides in GC-rich regions) were assessed in the VEP tool and FANTOM database.

After identifying possible regulatory SNPs that overlap with enhancers, promoters or 3′ untranslated regions (UTRs), we searched for potential target genes by looking for expression quantitative trait *loci* (eQTL) evidence ($P \leq 0.05$) in human brain tissue from the GTEx database (https://gtexportal.org/home/) [35, 36]. SNPs were sorted according to eQTL data since its presence increases the chance for a regulatory SNP to perturb gene expression.

Phase III: prediction of TFs and miRNAs binding sites

Prediction of TFs binding to promoter and enhancer elements

To predict the potential of selected SNPs to alter the binding of key regulatory molecules, we started by retrieving the respective flanking sequences (1000 bp, up and downstream) in FASTA format from Ensembl (v107, https://www.ensembl.org/index.html) [37]. Next, three prediction algorithms were used to identify and predict putative TFs whose binding sites might be altered by SNPs in promoter and enhancer elements. PERFECTOS-APE (v3.0.3, https://opera.autosome.org/perfectosape/scan) [38] required the submission of the dbSNP IDs and allowed to analyse position weight matrix (PWM) data from different public databases (e.g. HOCOMOCO, JASPAR, and SwissRegulon) by estimating the impact of SNPs within the TF binding motif occurrences. MEME SUITE motif (V5.4.1)—comparison tool Tomtom (https://meme-suite.org/meme/tools/tomtom) [39] and PROMO (v3, http://alggen.lsi.upc.es/cgi-bin/promo_v3/promo/promoinit.cgi?dirDB=TF_8.3) [40] required the submission of nucleotide sequences (∼30 bp) for comparison with PWM data from public motif databases like JASPAR and a commercial database (TRANSFAC v8.3), respectively.

When compiling a list of potential TFs from the different software, TF ChIP-Seq data from ENCODE were used as a selection strategy, since it shows the ability of these putative TFs to bind specific DNA sequences or chromatin configurations in human brain tissues or neuronal cell lines (if available). Whenever TF ChIP-seq data from ENCODE was lacking, we selected TFs that appeared in at least two of the three prediction tools to filter the most relevant. In addition, as a sorting approach, TFs expression data (nTPM, transcripts *per* million) were assessed from Protein Atlas database (v15, https://www.proteinatlas.org/) [41] to analyse tissue specificity.

Prediction of miRNAs binding to 3′UTR and enhancer elements

The primary prediction of miRNAs binding was carried out in miRDB (http://www.mirdb.org/, custom prediction) [42, 43] using a converted mRNA target sequence around 200 bp (∼100 bp up and downstream the SNP). The overlap of miRNAs seed sequence with each SNP was analysed. The miRDB guidelines suggest selecting miRNA with a target prediction score [range 50–100] above 60, but we kept all predictions to avoid being too stringent. Therefore, to confirm the miRDB predictions, we performed analysis in two other software: STarMir (https://sfold.wadsworth.org/cgi-bin/starmir.pl) [44] and RNAhybrid (v2.2, https://bibiserv.cebitec.uni-bielefeld.de/rnahybrid) [45]. For both software, we submitted the mRNA nucleotide sequence and miRNA name or mature sequence from miRBase (v22.1, https://www.mirbase.org/) [46]. STarMir logistic probability (logitProb) was used to measure miRNA prediction confidence. A probability above 0.5 and 0.75 indicates a good and an excellent confidence of miRNA binding. Then, to evaluate the binding potential of predicted miRNAs, the minimal free energy's (MFE, from RNAhybrid) threshold value was set up at −10 kcal/mol to ensure minimum thermodynamic stability. In addition to the MFE value, we analysed the binding pattern at the seed sequence of the miRNA (2–8 nucleotides from the 5′ end of the miRNA; Seed_Type from STarMir). If there was no gap in the alignment within the Watson and Crick matching, the binding of the miRNA on its target mRNA was considered

Downloaded from https://academic.oup.com/bfg/advance-article/doi/10.1093/bfgp/elad020/7186543 by University of Porto user on 20 June 2023

ideal [47, 48]. Since gene regulation is tissue-specific, miRNAs expression (log2RPM, reads *per* million) in the brain was used as a sorting criterion (microRNA Tissue Expression Database—miTED, https://dianalab.e-ce.uth.gr/mited/#/expressions) [49].

Results

We applied our protocol to migraine and described the obtained results to show how this step-by-step workflow may be followed while interpreting the potential functional relevance of non-coding SNPs in other neurological diseases.

Prioritisation of SNPs according to their potential regulatory role

Previously, we identified a set of 76 tagSNPs that were analysed through candidate gene association studies under the hypothesis that SNPs from neurotransmission and vesicle machinery-related genes are involved in migraine susceptibility, including: *synapsin I and II* (*SYN1*, *SYN2*), *synaptosome-associated protein* 25 (*SNAP25*), *vesicle-associated membrane protein 2* (*VAMP2*), *syntaxin 1A* (*STX1A*), *syntaxin binding protein 1 and* 5 (*STXBP1*, *STXBP5*), *unc-13 homolog B* (*UNC13B*), *gamma-aminobutyric acid type A receptor subunit alpha3* (*GABRA3*) and *gamma-aminobutyric acid type A receptor subunit theta* (*GABRQ*) [11–13]. A total of 22 non-coding tagSNPs were suggested to be relevant for migraine risk or protection, and therefore selected for our analysis [11–13]. From this set of non-coding SNPs, nine were previously associated with other neurological diseases such as epilepsy, Asperger's syndrome, autism spectrum disorder (ASD) and attention-deficit hyperactivity disorder (ADHD) (Table 1; Supplementary Table 2). The annotation process has shown that 8 out of the 22 non-coding SNPs were within regulatory regions, based on the Ensembl Regulatory Build dataset (Table 1; Supplementary Table 2). However, considering the scoring prioritising scheme, only six SNPs met our criteria of a minimum of three deleterious scores (integration score (counts), Table 2): *VAMP2*_rs1150, *SNAP25*_rs362990, *SYN1*_rs723556, *SYN1*_rs5906437, *STX1A*_rs6951030 and *SNAP25*_rs2327264. Noteworthy, we did not select *SNAP25*_rs362990 for further analysis since it has already been experimentally validated and no additional studies are needed to confirm its functional impact. No splicing defects were predicted by SpliceAI Pred for any of the 22 non-coding SNPs (Table 2).

Thus, from this group of six non-coding SNPs, we restricted to the non-reported ones with the highest potential to exert regulatory activity according to the integration score (sum) values (Table 2): *VAMP2*_rs1150 is within a CTCF-binding site (ENSR00000548703) in the 3′UTR; *STX1A*_rs6951030 is near the promoter region (ENSR00000213499) in intron 1; and *SNAP25*_rs2327264 locates in an enhancer (ENSR00000645066) and CTCF-binding site (ENSR00000645067) in intron 1 (Ensembl Regulatory Build, Table 1; Supplementary Table 2). In accordance, the ENCODE candidate *cis*-regulatory elements (cCREs) registry reports that *VAMP2*_rs1150 locates within a 3′UTR element (E1845038, Figure 2A), whereas *STX1A*_rs6951030 and *SNAP25*_rs2327264 flank a proximal enhancer element (E2563221, Figure 2B) near the promoter and a distal enhancer element (E2097223, Figure 2C), respectively.

Regulatory marks overlapping the selected SNPs

To confirm the regulatory potential of our selected SNPs, we analysed functional genomic data from ENCODE and Roadmap Epigenomics within their flanking regions. ENCODE and Roadmap Epigenomics data were shown to be in accordance with the Ensembl regulatory registry mentioned before. DNA accessibility data (i.e. DNase-seq data/DHSs cluster peaks) showed an open chromatin region for the binding of regulatory elements in *STX1A*_rs6951030 (Figure 2B), whereas *VAMP2*_rs1150 (Figure 2A) and *SNAP25*_rs2327264 (Figure 2C) were solely in the proximity of DHSs clusters. The overlapping of our selected SNPs with H3K27ac and H3K4me1 peaks suggested that they are in transcriptionally active elements (promoter or enhancer), particularly *STX1A*_rs6951030 (Figure 2B). *STX1A*_rs6951030 also presented an H3K4me3 broad peak indicating a promoter-like element (Figure 2B). In addition, DNase-seq and histone ChIP-seq data from human brain tissues or neuronal cell lines in ENCODE and Roadmap Epigenomics databases presented similar patterns (Supplementary Figures 1 and 2) as described above. *STX1A*_rs6951030 was shown to have the most promoter-like marks (Supplementary Figures 1B and 2B; flanking active TSS chromatin state), whereas *VAMP2*_rs1150 (Supplementary Figures 1A and 2A; transcription at 5′ and 3′ chromatin state) and *SNAP25*_rs2327264 (Supplementary Figures 1C and 2C; weak transcription/weak repressed Polycomb chromatin state) presented slight signals of enhancer elements through DNase-seq and H3K4me1/H3K27ac peaks. Still, *SNAP25*_rs2327264 presented less evidence of high regulatory activity compared with *VAMP2*_rs1150.

Prediction of TFs able to bind to enhancer SNPs

Considering that the selected SNPs were embedded in regulatory elements, it was important to identify DNA-binding motifs in these regions by matching PWM data from TFs with prediction software and assessing the presence of strong nearby TF ChIP-seq peaks with ENCODE (Figure 1). Unfortunately, TF ChIP-seq information was insufficient to allow a prioritisation of the predictions due to the lack of experimental data of these particular TFs, especially in neuronal cell lines. Some TFs were not in the collection since these assays are typically performed on known putative TF or in major regulators such as CCCTC-binding factor (CTCF) and RNA polymerase II subunit A (POLR2A), and on cell lines such as HepG2, K562 and HEK293. Nevertheless, Zinc Finger protein 263 (ZNF263), predicted from PERFECTOS-APE to be downregulated by *STX1A*_rs6951030 (Supplementary Table 3A, HOMER database), presented TF ChIP-seq peaks in the SNP region on K562, HEK293 and HepG2 cell lines (Figure 3).

Also, based on predictions from PERFECTOS-APE and MEME SUITE—Tomtom (with different PWM databases), Zinc Finger protein 423 (ZNF423) and HIC ZBTB transcriptional repressor 1 (HIC1; also known as ZNF901) were the two most likely TFs to be affected (through upregulation) by *STX1A*_rs6951036 (Supplementary Table 3D). However, when analysing TFs tissue expression in Protein Atlas database, ZNF423 (midbrain, 30.4 nTPM) and ZNF263 from ChIP-seq data (cerebral cortex, 22.5 nTPM) have shown the highest potential to perturb gene expression in brain.

Fewer TFs were predicted to be affected by *SNAP25*_rs2327264, but One Cut Homeobox 2 (ONECUT2) was suggested to be downregulated by this SNP, simultaneously by PERFECTOS-APE and MEME SUITE—Tomtom (Supplementary Table 3D, HOCOMOCO database). In the Protein Atlas database, ONECUT2 expression is enhanced in the brain–basal ganglia (9.1 nTPM), intestine-duodenum (10.1 nTPM) and gallbladder (10.1 nTPM).

Prediction of miRNAs able to bind 3′UTR and enhancer SNPs

To further investigate how these regulatory SNPs may affect gene expression, we predicted miRNAs binding to the SNP regions (Figure 1). Considering the target scores of the resulting four

Downloaded from https://academic.oup.com/bfg/advance-article/doi/10.1093/bfgp/elad020/7186543 by University of Porto user on 20 June 2023

Table 1. Summary of the most relevant features of our candidate non-coding SNPs from the Ensembl-VEP tool

Gene[a]	dbSNP ID	Ref Allele Freq	Alt Allele Freq	Ensembl Regulatory Build	Associated Disease	PubMed identifiers
GABRA3	rs2131190	C=0.88	T=0.12		Migraine	24040174
	rs3902802	A=0.84	G=0.16		Migraine	24040174
SNAP25	rs2327264	T=0.87	C=0.13	ENSR00000645066 (enhancer); ENSR00000645067 (CTCF binding site)		
	rs3787303	G=0.29	A=0.71			21651830,18821566,19806613,23593184
	rs363039	G=0.63	A=0.37	ENSR00001233331 (promoter flanking region)	Cognitive ability (IQ), working memory, and ADHD	21302343,18821566,17908175,25629685,27047369, 25445064,25222038,26779543,23593184,23497716, 22193912,27075583,30334187,30958380,20950795, 32979221
	rs363050	G=0.43	A=0.57		Cognitive ability (IQ) and ASDs	21302343,23927501,18821566,17908175,25629685, 27047369,26779543,28783930,23497716,27075583, 30334187,30958380,16801949,30914946,31192914, 30992661,23012269,25606466,31723209,19713048, 18030083
	rs362990	A=0.77	T=0.23		ADHD	25017045,21302343,27047369,31426340,23593184
	rs6108463	T=0.77[b]	C=0.23[b]	ENSR00001048462 (promoter flanking region)		18821566
STX1A	rs941298	G=0.73	A=0.27		Migraine and Asperger syndrome	19368856,20385907,22250207,24548729,27191890
	rs6951030	T=0.79	G=0.21	ENSR00000213499 (promoter)	Migraine	26322220,19368856,20385907,22250207,24548729, 27191890,23185642
STXBP1	rs3780658	A=0.76	G=0.24			
STXBP5	rs1765028	A=0.62	G=0.38		Migraine	32979221
SYN1	rs723556	A=0.52	G=0.48			20371206
	rs5906435	C=0.73	T=0.27		Migraine	22671570,32979221
	rs5906437	A=0.55[b]	G=0.45[b]	ENSR00000906589 (promoter flanking region)		
	rs2239459	G=0.63	A=0.37	ENSR00001160196 (promoter flanking region)		
SYN2	rs217049	C=0.12	T=0.88			
	rs307574	G=0.14[b]	T=0.86[b]			
	rs3773364	A=0.86	G=0.14	ENSR00000677450 (promoter flanking region)	Migraine and epilepsy	25595263,20034013,21465568,27458546,17913586, 22384280,19201561,32979221
	rs310763	T=0.16	C=0.84			25743335,30643256
UNC13B	rs7851161	A=0.69[b]	T=0.31[b]			18633107
VAMP2	rs1150	A=0.33	G=0.67	ENSR00000548703 (CTCF binding site)	Visual working memory and ADHD	25445064

[a]Non-coding SNPs of interest were not found in *GABRQ*
[b]Data retrieved from 1000 Genomes Other features such as chromosomal localisation, SNP position, GC content (%), TSS distance and conservation scores can be found in Supplementary Table 2.

Downloaded from https://academic.oup.com/bfg/advance-article/doi/10.1093/bfgp/elad020/7186543 by University of Porto user on 20 June 2023

Table 2. Prioritisation strategy of candidate SNPs based on an integrative analysis of the functional annotation software

		CADD		DANN	FunSeq2		FATHMM-MKL	GWAVA			Integration Score	
Gene	dbSNP ID	PHRED [0–100]	Normalised to [0–1]	Score [0–1]	Score [0–6]	Normalised to [0–1]	Non-coding Score [0–1]	Region Score [0–1]	TSS Score [0–1]	Unmatched Score [0–1]	Counts	Sum
VAMP2	rs1150	15.93	0.16	0.83	1.63	0.27	0.88	0.46	0.48	0.61	7.00	3.70
SNAP25	rs362990	7.03	0.07	0.77	1.37	0.23	0.24	0.56	0.35	0.50	4.00	2.73
SYN1	rs723556	2.36	0.02	0.77	0.78	0.13	0.11	0.34	0.29	0.47	4.00	2.13
SYN1	rs5906437	3.63	0.04	0.85	0.78	0.13	0.11	0.43	0.29	0.28	4.00	2.12
STX1A	rs6951030	9.96	0.10	0.66	1.65	0.28	0.09	0.36	0.46	0.70	3.00	2.65
SNAP25	rs2327264	10.90	0.11	0.84	1.56	0.26	0.27	0.36	0.22	0.17	3.00	2.23
SYN2	rs3773364	4.64	0.05	0.60	0.62	0.10	0.20	0.59	0.32	0.54	2.00	2.40
SNAP25	rs363039	5.58	0.06	0.74	0.75	0.12	0.11	0.33	0.30	0.09	2.00	1.75
STXBP1	rs3780658	0.87	0.01	0.83	0.76	0.13	0.07	0.36	0.11	0.23	2.00	1.73
STXBP5	rs1765028	3.45	0.03	0.76	0.13	0.02	0.18	0.38	0.23	0.09	1.00	1.70
SYN1	rs5906435	8.80	0.09	0.80	0.59	0.10	0.13	0.28	0.17	0.07	1.00	1.64
SYN2	rs310763	0.17	0.00	0.31	0.00	0.00	0.13	0.35	0.52	0.26	1.00	1.58
SNAP25	rs3787303	1.20	0.01	0.38	1.37	0.23	0.07	0.34	0.31	0.22	1.00	1.56
SYN1	rs2239459	0.16	0.00	0.56	0.78	0.13	0.12	0.22	0.29	0.24	1.00	1.56
STX1A	rs941298	4.57	0.05	0.58	0.84	0.14	0.08	0.27	0.20	0.21	1.00	1.54
SNAP25	rs6108463	5.71	0.06	0.55	0.75	0.12	0.17	0.20	0.13	0.17	1.00	1.40
SNAP25	rs363050	2.11	0.02	0.32	0.75	0.12	0.15	0.29	0.27	0.04	1.00	1.21
SYN2	rs307574	0.60	0.01	0.70	0.00	0.00	0.15	0.24	0.03	0.01	1.00	1.14
GABRA3	rs3902802	4.43	0.04	0.61	0.43	0.07	0.64	0.39	0.37	0.11	0.00	2.24
UNC13B	rs7851161	7.21	0.07	0.59	0.39	0.06	0.11	0.26	0.16	0.05	0.00	1.31
GABRA3	rs2131190	3.31	0.03	0.46	0.24	0.04	0.15	0.39	0.09	0.14	0.00	1.30
SYN2	rs217049	0.35	0.00	0.41	0.00	0.00	0.07	0.30	0.15	0.06	0.00	1.00

SpliceAI Pred (DS_AG, DS_AL, DS_DG and DS_DL) scores are between 0.00 and 0.03 for all SNPs Non-coding SNPs were defined as likely functionally relevant when at least three scoring annotation methods indicated deleteriousness [referred as integration score (counts); highlighted in green], according to stipulated cut-off values: ≥10 for CADD_PHRED, ≥0.7 for FunSeq2, DANN and FATHMM-MKL (non-coding score), and ≥ 0.4 for GWAVA (Region, TSS, Unmatched). In addition, to detect SNPs with the highest potential to be functionally relevant, we normalised CADD_PHRED and FunSeq2 scores to 0–1 range to calculate the integration score (sum). SNPs selected for this study are highlighted in dark green.

Downloaded from https://academic.oup.com/bfg/advance-article/doi/10.1093/bfgp/elad020/7186543 by University of Porto user on 20 June 2023

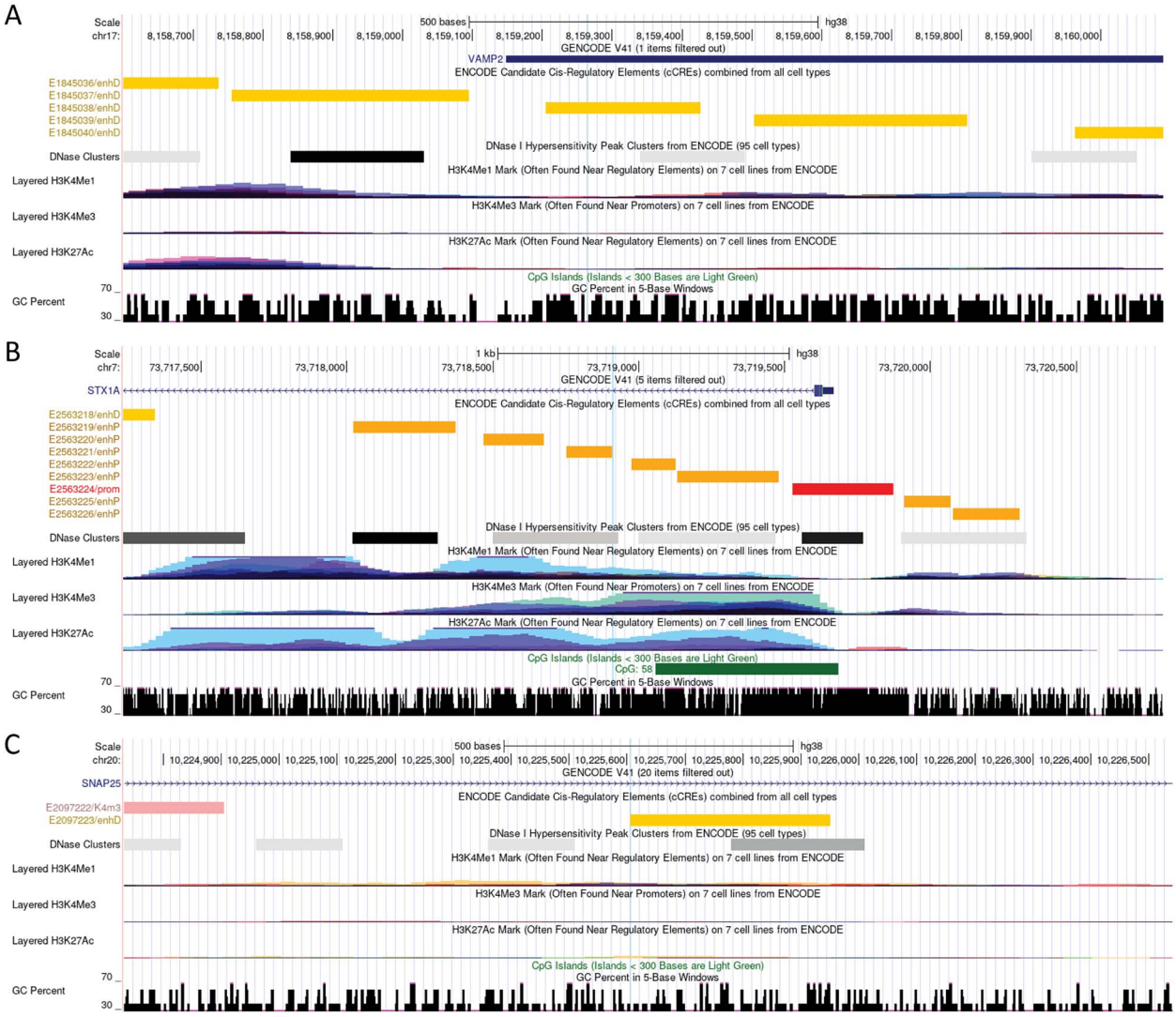

Figure 2. UCSC Genome Browser view displaying ENCODE summarised chromatin-related features of (**A**) *VAMP2*_rs1150, (**B**) *STX1A*_rs6951030 and (**C**) *SNAP25*_rs2327264. The following regulatory tracks are shown (up to bottom): genomic localisation and gene structure (from GENECODE V41), ENCODE data of cCREs, DNase I hypersensitivity peak clusters (grayscale indicates signal strength) and active histone (H3K4me1, H3K4me3 and HEK27ac) marks. cCREs are labelled as follows: red—promoter, orange—proximal enhancer, yellow—distal enhancer, pink—DNase-H3K4me3 (promoter-like signatures that are not within 200 bp from a TSS) and blue—CTCF-only. A light blue vertical line denotes the SNP position. USCS Genome Browser: http://genome.ucsc.edu.

miRNAs (Supplementary Table 4A; highlighted in green), the most confident predictions were hsa-miR-4528 for *SNAP25*_rs2327264 (miRDB—75 target score; STarMir—0.62 logitProb; Supplementary Table 4B and C) and hsa-miR-5010-3p for *VAMP2*_rs1150 (miRDB—55 target score; STarMir—0.91 logitProb; Supplementary Table 4B and C). No miRNAs were predicted to overlap their seed sequence to the region encompassing *STX1A*_rs6951030.

All miRNAs showed a MFE below −10 kcal/mol, indicating a minimal thermodynamic stability (Figure 4; Supplementary Table 4D). However, has-miR-5010-3p predicted to bind *VAMP2*_rs1150 stands out with a MFE of −29.8 kcal/mol (Figure 4D; Supplementary Table 4D), which indicates a stronger stability of the interaction. Also, the hsa-miR-5010-3p (*VAMP2*_rs1150) and hsa-miR-4528 (*SNAP25*_rs2327264) miRNAs presented a complete alignment from nucleotides 2–8 at the 5′ end of the seed sequence (Figure 4A and D; Supplementary Table 4C, Seed_Type: 8mer), contrarily to the other miRNAs predicted for *SNAP25*_rs2327264 (Figure 4B and C). The predictions described above together with the tissue expression data mostly supported the hsa-miR-5010-3p binding to the *VAMP2* 3′UTR in brain tissues (Figure 4E).

Curiously, in the miRDB prediction from *VAMP2*_rs1150, there was a highly concomitant motif of about 25–30 bp before the SNP for several miRNAs with high target scores (miRDB—92, 91, 73, 73 and 65 target scores; Supplementary Table 4A), suggesting the existence of a deeply important motif for gene expression regulation.

Search for eQTLs related to the selected SNPs

Next, to identify the potential target genes affected by these regulatory SNPs, we examined associations between these SNPs and gene expression data (eQTLs) in human brain tissues by using the GTEx database (Figure 1). Only *VAMP2*_rs1150 was significantly associated with its expression in human brain tissues (Table 3). On the other hand, *STX1A*_rs6951030 was significantly associated with *transducin beta-like 2* (*TBL2*) expression in the

Downloaded from https://academic.oup.com/bfg/advance-article/doi/10.1093/bfgp/elad020/7186543 by University of Porto user on 20 June 2023

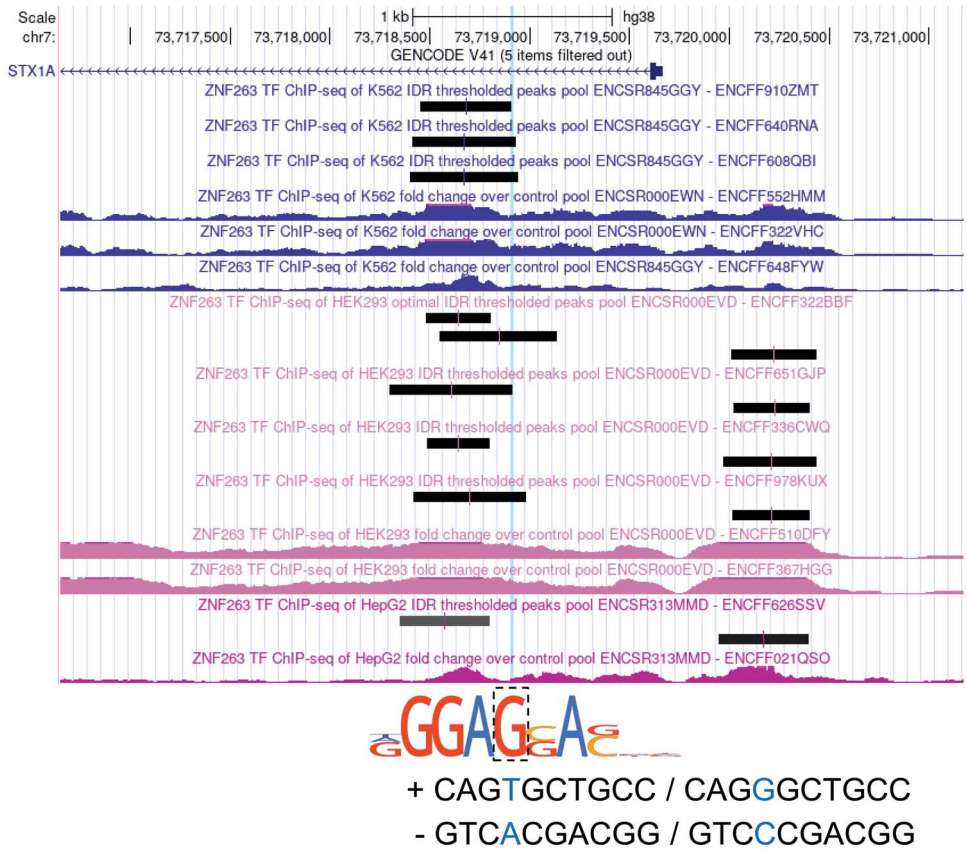

Figure 3. UCSC genome browser view of *STX1A_rs6951030 locus* displays the ENCODE TF ChIP-seq peaks of ZNF263 from K562, HEK293 and HepG2 cell lines (fold change over control and IDR thresholded peaks; track height: 64 pixels). PWM from PERFECTOS-APE (HOMER database) is represented under the UCSC genome browser view. A light blue vertical line denotes the SNP position. USCS Genome Browser: http://genome.ucsc.edu.

Table 3. eQTL information of *STX1A_rs6951030* and *VAMP2_rs1150* (brain tissue; $P < 0.05$) from GTEx portal

Gencode ID	Gene	Variant ID	dbSNP ID	P-value	NES	Tissue
ENSG00000220205.8	*VAMP2*	chr17_8159265_A_G_b38	rs1150	1.4e-13	0.42	Brain—putamen (basal ganglia)
				4.0e-13	0.32	Brain—caudate (basal ganglia)
				4.5e-12	0.35	Brain—cerebellum
				5.4e-9	0.22	Brain—hippocampus
				4.0e-8	0.34	Brain—substantia nigra
				5.7e-8	0.26	Brain—cerebellar hemisphere
				3.1e-7	0.28	Brain—hypothalamus
				0.0000027	0.14	Brain—cortex
				0.0000088	0.14	Brain—frontal cortex (BA9)
				0.000026	0.41	Brain—spinal cord (cervical c-1)
ENSG00000106638.15	*TBL2*	chr7_73718911_T_G_b38	rs6951030	0.000099	0.31	Brain—cerebellar hemisphere

NES stands for normalised effect size

cerebellar hemisphere (Table 3). GTEx data also showed an association of *STX1A_rs6951030* with its expression; however it was a small effect, only observed in artery-tibial tissue ($P = 0.00011$; NES = −0.13). No eQTL data suggested *SNAP25_rs2327264* and *STX1A_rs6951030* targeting their own genes in brain.

Discussion

We are progressively starting to understand the major role of functional non-coding SNPs in modulating disease's susceptibility and phenotype. Simultaneously, the need to establish a prioritisation strategy to study these SNPs urges since intergenic and intronic regions are extensive and intrinsically less conserved. Currently, there are no consensual, focused and easy-friendly bioinformatics protocols to prioritise non-coding SNPs. Instead, only guidelines and recommendations regarding the different software and the clinical interpretation of non-coding SNPs were published. Here, we proposed a practical workflow to assist the analysis of the functional potential of non-coding SNPs in multifactorial neurological diseases, from the annotation of regulatory elements to the prediction of TFs and miRNAs, by combining publicly available bioinformatics tools and databases.

We considered migraine as a disease model, and started by analysing non-coding SNPs in candidate genes related to the synaptic vesicle machinery. SNPs were previously assessed by us in candidate gene association studies [11–13], but our

Downloaded from https://academic.oup.com/bfg/advance-article/doi/10.1093/bfgp/elad020/7186543 by University of Porto user on 20 June 2023

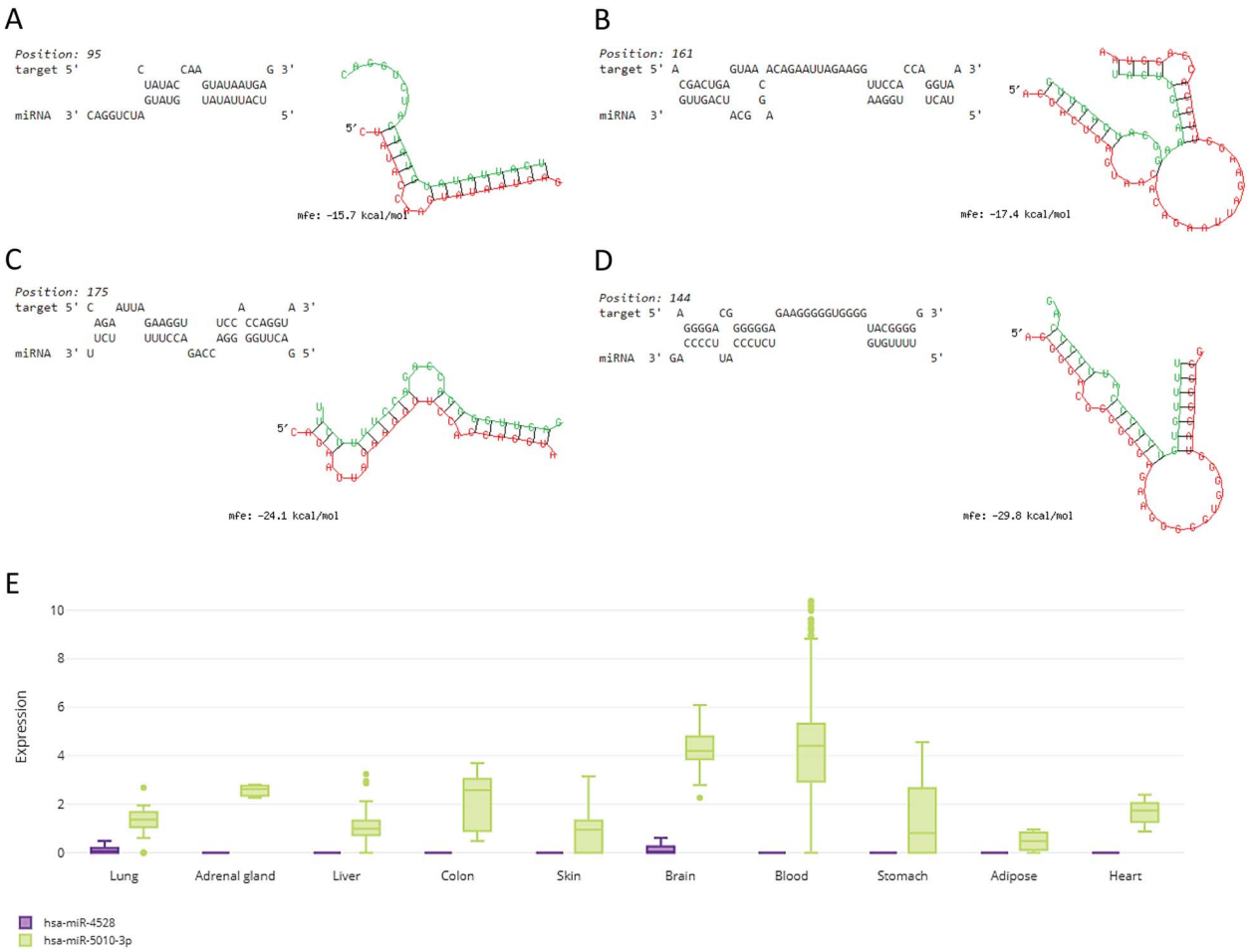

Figure 4. RNAhybrid algorithm showing MFE scores (hybridisation energy), and the alignment between mature miRNAs (seed sequence) and the mRNA sequence (predicted by miRDB). (**A-C**) *SNAP25_rs2327264* interaction with has-miR-4528, -890 and -12126, respectively. (**D**) *VAMP2_rs1150* interaction with has-miR-5010-3p. Also, miRNA expression levels from the best candidates were obtained from miTED database. (**E**) hsa-miR-4528 and hsa-miR-5010-3p expression data (Log2RPM, from miTED) are graphically represented in purple and light green, respectively. Several tissues were selected to compare with brain expression levels, namely lung, adrenal gland, liver, colon, skin, blood, stomach, adipose and heart.

workflow can be used to study candidate non-coding SNPs identified by other methods. If experimental studies have already shown that SNPs affect gene expression, Phase III of our workflow can ultimately be used to predict the binding of TFs or miRNAs to complement *in vitro* functional validation.

Interestingly, from the selected set of 22 relevant SNPs in migraine, some were also previously associated with neuropsychiatric disorders such as ADHD and epilepsy but the knowledge of the functional effect of these SNPs on gene expression is lacking for most of them. Two SNPs in *SNAP25* were described to affect TF binding sites [50, 51]: rs363039 (A-allele) that may introduce a DNA binding site for the glucocorticoid receptor, remove a tumor protein p53 half site and a binding site for Zinc Finger protein 589 (also known as SZF1) [51]; and rs363050 that encompasses a regulatory element that decreases SNAP25 protein expression in ASD [52]. Likewise, *SNAP25_rs362990* (A-allele) was shown to significantly decrease *SNAP25* expression in ADHD [53].

We evaluated the functional potential of candidate non-coding SNPs by an integrative analysis of some well-described scoring methods and prioritised these SNPs based on the number of algorithms with deleterious predictions and score values. Still, these methods have two limitations that must be taken into account: (1) the nucleotide conservation is assessed, but the relevance of this feature in non-coding SNPs is controversial since intergenic and intronic regions are intrinsically less conserved [10, 54]; and (2) the absence of a specific tissue analysis can be misleading since regulatory effects are strongly tissue-specific [4, 55]. It is noteworthy that *SNAP25_rs362990*, the second SNP in our scoring prioritisation strategy (with four scoring methods pointing to deleteriousness and an integration score sum of 2.73), has previously been associated with *SNAP25* expression regulation in ADHD [53], which reinforces the effectiveness of the herein presented scoring prioritisation strategy (Table 2) to select novel potential susceptibility SNPs in complex diseases.

Next, our prioritisation strategy followed a straightforward analysis of public epigenetic data in brain tissue and/or neuronal cell lines to assess the chromatin status and accessibility in the SNP region and ultimately determine their regulatory signature (as promoters, enhancers or 3′UTRs). In addition to the typical epigenetic marks (DNase-Seq and histone ChIP-seq data) here analysed, other features can be further evaluated to support the presence of enhancers in the SNP region. ChIP-seq data from the histone acetyltransferase EP300 have a strong association with active enhancers because of its ability to acetylate H3K27 [56, 57]. Also, ChIP-seq peaks from chromatin factors such as the insulator factor CTCF and the cohesin complex (also called SMC3-RAD21) can imply enhancer–promoter functional interaction [57, 58].

Downloaded from https://academic.oup.com/bfg/advance-article/doi/10.1093/bfgp/elad020/7186543 by University of Porto user on 20 June 2023

The next step in our workflow was to predict if these SNPs could regulate gene expression in an allele-specific manner through different binding affinity to TFs or miRNAs. For that, we recurred to different algorithms, sequencing-based data (TF ChIP-Seq) and tissue expression information. Still, one should bear in mind that these predictions do not necessarily reflect a 'biological state'. *In vivo* and *in vitro*, the RNA might form an alternative secondary structure different from the ones predicted by interacting with long non-coding RNAs or RNA-binding proteins [59]. Also, the degenerate nature of TFs binding motifs along with motif-independent effects can confound the scenario and invalidate predictions [60]. Moreover, a subset of miRNAs was recently reported to unconventionally activate enhancers targets' transcription, alter chromatin status and increase E1A binding protein p300 (EP300), argonaute RISC catalytic component 2 (AGO2) and RNA Polymerase II at the enhancer region [61, 62]. Thus, *in vitro* studies are necessary to prove their presence in regulatory elements, as well as the allele-specific binding to the TFs and/or miRNAs, which may increase or decrease the gene expression of their targets.

Another important factor to consider is the evidence for eQTLs, which connects SNPs to gene regulatory mechanisms and increases the chance for a candidate non-coding SNP to perturb gene expression. eQTL data are particularly critical for *trans*-acting regulatory elements, where the target gene is not the closest gene to the non-coding SNP under study. In this case, new biological mechanisms may underlie a complex disorder. However, for many SNPs we did not detect effects on gene expression (i.e. eQTLs) even if located in putatively regulatory regions, due to the lack of data from cell types or conditions that are most relevant for the disease context (since these large-scale studies are performed across human tissues obtained opportunistically or *post-mortem*) [63].

In conclusion, our workflow was designed to assist in the prioritisation of non-coding SNPs with a high regulatory potential. From our list of candidate genes and non-coding SNPs, our step-by-step analysis suggested that besides *SNAP25*_rs362990 (previously experimentally validated), *VAMP2*_rs1150, *STX1A*_rs6951030 and *SNAP25*_rs2327264 have the highest potential to exert a regulatory function in the context of migraine. *VAMP2*_rs1150 (3′ UTR) is a predicted target of hsa-mir-5010-3p that likely disrupts its gene expression. *STX1A*_rs6951030 (proximal enhancer) may affect the binding affinity of TFs from the zinc-finger protein family and disrupt *TBL2* gene expression (a new gene, potentially interesting to be studied in migraine). *SNAP25*_rs2327264 (distal enhancer) is a possible target for ONECUT2 binding. Together, these findings provide insight into the impact of non-coding SNPs and gene regulation of SNARE (soluble *N*-ethylmaleimide-sensitive factor activating protein receptor) complex proteins (involved in neurotransmitter release).

Key Points

- This protocol is useful to prioritise potential susceptibility genes and variants in complex disorders.
- Non-coding SNPs in SNARE complex genes potentially impact gene regulation mechanisms in migraine.
- *VAMP2*_rs1150 may be particularly relevant in migraine since it was predicted to affect hsa-mir-5010-3p binding to the 3′UTR and dysregulate its own expression.

Acknowledgments

The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. The data used for the analyses described in this manuscript were obtained from the GTEx Portal on 04/10/2022.

Supplementary data

Supplementary data are available online at https://academic.oup.com/bfg.

Funding

This work was supported by Fundo Europeu de Desenvolvimento Regional (FEDER) funds through the COMPETE 2020—Operational Programme for Competititveness and Internationalisation (POCI), Portugal, 2020; by Programa de Cooperação Transfronteiriça Interreg VA España-Portugal (POCTEP 2014–2020) under the project 'Análisis y correlación entre la epigenética y la actividad cerebral para evaluar el riesgo de migraña crónica y episódica en mujeres' (0702_MIGRAINEE_2_E) and by Portuguese funds through Fundação para a Ciência e a Tecnologia (FCT)/Ministério da Ciência, Tecnologia e Ensino Superior in the framework of the project POCI-01-0145-FEDER-029486 (PTDC/MEC-NEU/29486/2017). This research was also funded by Sociedade Portuguesa de Cefaleias (SPC). S.M. (CEECIND/00684/2017), N.P. and M.S. (Decreto Lei n°57/2016 de 29 de Agosto - Norma Transitória) are funded by FCT. M.Q. was the recipient of a fellowship (SFRH/BD/96253/2013) funded by FCT.

Conflict of interest

The authors declare no conflict of interest.

Authors' contributions

M.A.-F. and N.P conceived the study and were in charge of overall administration and planning of the project; D.F. and M.Q. collected the data with support from M.A.-F. and C.L.; D.F. developed and performed the protocol with input from all authors; D.F. analysed and interpreted the data with support from all authors; M.S., S.M., M.Q., A.M.L., M.A.-F. and C.L. verified and discussed the methodology and results; D.F. wrote the original draft; all authors critically revised and edited the manuscript; S.M., M.S., C.L., M.A.-F. and N.P. supervised the work; M.A.-F. and N.P. contributed with resources and funding. All authors have read and agreed to the published version of the manuscript.

Data availability

All data generated or analysed during this study are included in this published article and its supplementary data. Also, public websites used in this study are described in the methods section.

References

1. Zhang F, Lupski JR. Non-coding genetic variants in human disease. *Hum Mol Genet* 2015;**24**:R102–10.
2. Rojano E, Seoane P, Ranea JAG, *et al*. Regulatory variants: from detection to predicting impact. *Brief Bioinform* 2019;**20**:1639–54.

Downloaded from https://academic.oup.com/bfg/advance-article/doi/10.1093/bfgp/elad020/7186543 by University of Porto user on 20 June 2023

3. Rao S, Yao Y, Bauer DE. Editing GWAS: experimental approaches to dissect and exploit disease-associated genetic variation. *Genome Med* 2021;**13**:1–20.
4. Ellingford JM, Ahn JW, Bagnall RD, *et al.* Recommendations for clinical interpretation of variants found in non-coding regions of the genome. *Genome Med* 2022;**14**:1–19.
5. Maurano MT, Humbert R, Rynes E, *et al.* Systematic localization of common disease-associated variation in regulatory DNA. *Science* 2012;**337**:1190–5.
6. Abascal F, Acosta R, Addleman NJ, *et al.* Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* 2020;**583**:699–710.
7. Bernstein BE, Stamatoyannopoulos JA, Costello JF, *et al.* The NIH roadmap Epigenomics mapping consortium. *Nat Biotechnol* 2010;**28**:1045–8.
8. Lizio M, Abugessaisa I, Noguchi S, *et al.* Update of the FANTOM web resource: expansion to provide additional transcriptome atlases. *Nucleic Acids Res* 2019;**47**:D752–8.
9. Bocher O, Génin E. Rare variant association testing in the non-coding genome. *Hum Genet* 2020;**139**:1345–62.
10. Ritchie GRS, Dunham I, Zeggini E, *et al.* Functional annotation of noncoding sequence variants. *Nat Methods* 2014;**11**:294–6.
11. Lemos C, Pereira-Monteiro J, Mendonça D, *et al.* Evidence of syntaxin 1A involvement in migraine susceptibility: a Portuguese study. *Arch Neurol* 2010;**67**:422–7.
12. Quintas M, Neto JL, Pereira-Monteiro J, *et al.* Interaction between γ-aminobutyric acid a receptor genes: new evidence in migraine susceptibility. *PLoS One* 2013;**8**:e74087.
13. Quintas M, Neto JL, Sequeiros J, *et al.* Going deep into synaptic vesicle machinery genes and migraine susceptibility—a case-control association study. *Headache* 2020;**60**:2152–65.
14. Gormley P, Anttila V, Winsvold BS, *et al.* Meta-analysis of 375,000 individuals identifies 38 susceptibility loci for migraine. *Nat Genet* 2016;**48**:856–66.
15. Techlo TR, Rasmussen AH, Møller PL, *et al.* Familial analysis reveals rare risk variants for migraine in regulatory regions. *Neurogenetics* 2020;**21**:149–57.
16. Lipton RB, Stewart WF, Diamond S, *et al.* Prevalence and burden of migraine in the United States: data from the American migraine study II. *Headache* 2001;**41**:646–57.
17. Lipton RB, Bigal ME, Diamond M, *et al.* Migraine prevalence, disease burden, and the need for preventive therapy. *Neurology* 2007;**68**:343–9.
18. Olesen J. Headache classification Committee of the International Headache Society (IHS) the international classification of headache disorders3rd edition. *Cephalalgia* 2018;**38**:1–211.
19. Mulder EJ, Van Baal C, Gaist D, *et al.* Genetic and environmental influences on migraine: a twin study across six countries. *Twin Res* 2003;**6**:422–31.
20. Hansen RD, Christensen AF, Olesen J. Family studies to find rare high risk variants in migraine. *J Headache Pain* 2017;**18**:32.
21. Polderman TJC, Benyamin B, De Leeuw CA, *et al.* Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat Genet* 2015;**47**:702–9.
22. Sutherland HG, Albury CL, Griffiths LR. Advances in genetics of migraine. *J Headache Pain* 2019;**20**:72.
23. Bron C, Sutherland HG, Griffiths LR. Exploring the hereditary nature of migraine. *Neuropsychiatr Dis Treat* 2021;**17**:1183–94.
24. McLaren W, Gil L, Hunt SE, *et al.* The Ensembl variant effect predictor. *Genome Biol* 2016;**17**:1–14.
25. Oscanoa J, Sivapalan L, Gadaleta E, *et al.* SNPnexus: a web server for functional annotation of human genome sequence variation (2020 update). *Nucleic Acids Res* 2020;**48**:W185–92.
26. Wang J, Ullah AZD, Chelala C. IW-scoring: an integrative weighted scoring framework for annotating and prioritizing genetic variations in the noncoding genome. *Nucleic Acids Res* 2018;**46**:E47.
27. Li MJ, Pan Z, Liu Z, *et al.* Predicting regulatory variants with composite statistic. *Bioinformatics* 2016;**32**:2729–36.
28. Zhang J, Wu D, Dai Y, *et al.* Functional relevance for central cornea thickness-associated genetic variants by using integrative analyses. *BioData Mining* 2018;**11**:19.
29. Rogers MF, Shihab HA, Gaunt TR, *et al.* CScape: a tool for predicting oncogenic single-point mutations in the cancer genome. *Sci Rep* 2017;**7**:11597.
30. Lizio M, Harshbarger J, Shimoji H, *et al.* Gateways to the FANTOM5 promoter level mammalian expression atlas. *Genome Biol* 2015;**16**:1–14.
31. Consortium RE, Kundaje A, Meuleman W, *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* 2015;**518**:317–30.
32. Lee BT, Barber GP, Benet-Pagès A, *et al.* The UCSC Genome Browser database: 2022 update. *Nucleic Acids Res* 2022;**50**:D1115–D1122.
33. Rosenbloom KR, Sloan CA, Malladi VS, *et al.* ENCODE Data in the UCSC Genome Browser: year 5 update. *Nucleic Acids Res* 2012;**41**:D56–D63.
34. Daofeng Li, Purushotham D, Harrison JK, *et al.* WashU Epigenome Browser update 2022. *Nucleic Acids Res* 2022;**50**: W774–W781.
35. GTEx Consortium. The genotype-tissue expression (GTEx) project. *Nat Genet* 2013;**45**:580–5.
36. GTEx Consortium. The GTEx consortium atlas of genetic regulatory effects across human tissues. *Science* 2020;**369**:1318–30.
37. Cunningham F, Allen JE, Allen J, *et al.* Ensembl 2022. *Nucleic Acids Res* 2022;**50**:D988–95.
38. Kulakovskiy IV, Vorontsov IE, Yevshin IS, *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**:D252–9.
39. Bailey TL, Boden M, Buske FA, *et al.* MEME suite: tools for motif discovery and searching. *Nucleic Acids Res* 2009;**37**:202–8.
40. Messeguer X, Escudero R, Farré D, *et al.* PROMO: detection of known transcription regulatory elements using species-tailored searches. *Bioinformatics* 2002;**18**:333–4.
41. Thul PJ, Lindskog C. The human protein atlas: a spatial map of the human proteome. *Protein Sci* 2018;**27**:233–44.
42. Liu W, Wang X. Prediction of functional microRNA targets by integrative modeling of microRNA binding and target expression data. *Genome Biol* 2019;**20**:1–10.
43. Chen Y, Wang X. MiRDB: an online database for prediction of functional microRNA targets. *Nucleic Acids Res* 2020;**48**:D127–31.
44. Kanoria S, Rennie W, Liu C, *et al.* STarMir tools for prediction of microRNA binding sites. *Methods Mol Biol* 2016;**1490**:73–83.
45. Krüger J, Rehmsmeier M. RNAhybrid: MicroRNA target prediction easy, fast and flexible. *Nucleic Acids Res* 2006;**34**:451–4.
46. Kozomara A, Griffiths-Jones S. MiRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res* 2011;**39**:152–7.
47. Riffo-Campos ÁL, Riquelme I, Brebi-Mieville P. Tools for sequence-based miRNA target prediction: what to choose? *Int J Mol Sci* 2016;**17**:1–18.
48. Peterson SM, Thompson JA, Ufkin ML, *et al.* Common features of microRNA target prediction tools. *Front Genet* 2014;**5**:1–10.

Downloaded from https://academic.oup.com/bfg/advance-article/doi/10.1093/bfgp/elad020/7186543 by University of Porto user on 20 June 2023

49. Kavakiotis I, Alexiou A, Tastsoglou S, *et al.* DIANA-miTED: a microRNA tissue expression database. *Nucleic Acids Res* 2022;**50**:D1055–61.
50. Gosso MF, De Geus EJC, Polderman TJC, *et al.* Common variants underlying cognitive ability: further evidence for association between the SNAP-25 gene and cognition using a family-based study in two independent Dutch cohorts. *Genes Brain Behav* 2008;**7**:355–64.
51. Söderqvist S, McNab F, Peyrard-Janvid M, *et al.* The SNAP25 gene is linked to working memory capacity and maturation of the posterior cingulate cortex during childhood. *Biol Psychiatry* 2010;**68**:1120–5.
52. Braida D, Guerini FR, Ponzoni L, *et al.* Association between SNAP-25 gene polymorphisms and cognition in autism: functional consequences and potential therapeutic strategies. *Transl Psychiatry* 2015;**5**:e500–11.
53. Hawi Z, Matthews N, Wagner J, *et al.* DNA variation in the SNAP25 gene confers risk to ADHD and is associated with reduced expression in prefrontal cortex. *PLoS One* 2013;**8**:1–8.
54. Kircher M, Witten DM, Jain P, *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* 2014;**46**:310–5.
55. Boyle AP, Hong EL, Hariharan M, *et al.* Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res* 2012;**22**:1790–7.
56. Visel A, Blow MJ, Li Z, *et al.* ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* 2009;**457**:854–8.
57. Whalen S, Truty RM, Pollard KS. Enhancer-promoter interactions are encoded by complex genomic signatures on looping chromatin. *Nat Genet* 2016;**48**:488–96.
58. Oti M, Falck J, Huynen MA, *et al.* CTCF-mediated chromatin loops enclose inducible gene regulatory domains. *BMC Genomics* 2016;**17**:1–16.
59. Riolo G, Cantara S, Marzocchi C, *et al.* miRNA targets: from prediction tools to experimental validation. *Methods Protoc* 2021;**4**:1–20.
60. Deplancke B, Alpern D, Gardeux V. The genetics of transcription factor DNA binding variation. *Cell* 2016;**166**:538–54.
61. Xiao M, Li J, Li W, *et al.* MicroRNAs activate gene transcription epigenetically as an enhancer trigger. *RNA Biol* 2017;**14**:1326–34.
62. Odame E, Chen Y, Zheng S, *et al.* Enhancer RNAs: transcriptional regulators and workmates of NamiRNAs in myogenesis. *Cell Mol Biol Lett* 2021;**26**:1–20.
63. Umans BD, Battle A, Gilad Y. Where are the disease-associated eQTLs? *Trends Genet* 2021;**37**:109–24.

Downloaded from https://academic.oup.com/bfg/advance-article/doi/10.1093/bfgp/elad020/7186543 by University of Porto user on 20 June 2023