INNSPUB

International Journal of Biosciences | IJB |
ISSN: 2220-6655 (Print) 2222-5234 (Online)
http://www.innspub.net
Vol. 10, No. 3, p. 388-398, 2017

RESEARCH PAPER OPEN ACCESS

Identification of biological functions of risk loci associated with complex epilepsy: An in-silico approach for data analysis

Rubina Dad[1], Muhammad Ikram Ullah[2], Amjad Ali[1], Muhammad Jawad Hassan[*1]

[1]Atta-ur-Rahman School of Applied Biosciences (ASAB), National University of Sciences & Technology (NUST), Islamabad, Pakistan.

[2]Department of Biochemistry, University of Health Sciences, Lahore, Pakistan

Key words: Epilepsy, GWAS, SNPs, Risk loci, SNAP tool, Regulome DB

http://dx.doi.org/10.12692/ijb/10.3.388-398 Article published on March 31, 2017

Abstract

Complex epilepsy is caused by the interaction of multiple genes and environmental factors with number of susceptibility loci in human genome. In this study, we selected 32 epilepsy associated risk Single Nucleotide Polymorphisms (SNPs) from six published Genome Wide Association Studies (GWAS) and used online SNAP tool to deduce 288 proxy SNPs based on linkage disequilibrium. These results were then used as input data for Regulome DB; software for interpretation of regulatory variants in the human genome to predict their potential functions. After investigating these 288 SNPs, 157 SNPs returned back with a score indicative of no potential regulatory function. Only 10 SNPs returned with significant scores, indicating the regulatory function and only 3 out of them showed highly significant score. Our results illustrate the future interpretation of GWAS data to include LD structure and different loci association for epilepsy risk.

* **Corresponding Author:** Muhammad Jawad Hassan ✉ mjh@asab.nust.edu.pk

Introduction

Inherited epilepsy disorders can be divided into different types based on the types of seizures, age of onset and path physiology of the disease (Fisher *et al.* 2014). Young age incidence of epilepsy during first months of life is very high worldwide. Somewhat higher prevalence (ranging from 14 to 57 per 1,000) has been reported in pilot studies using a standardized World Health Organization (WHO) protocol in Panama (Sillanpää *et al.* 2016). Complex epilepsy is a common neurological disorder caused by both genetic and environmental factors hence does not follow Mendelian patterns. For complex epilepsy, examination of regulatory functions of associated SNPs may help in better understanding of the disease. Genome wide association studies are playing promising roles in identification of the novel genes associated with complex epilepsy. Since 2005, more than 1200 such studies have been reported. These GWAS along with similar large data sets from recent years have been utilized to determine significant number of rare and heterogeneous genetic diseases (Welter *et al.* 2014). It has also been reported that significant variants are located in non-coding and or intronic regions suggesting disease occurrence not only because of alteration in protein structure or function but also because of protein regulatory factors (Hindorff *et al.* 2009). These findings have open up new era of research for scientific community to find out the implications of these non-coding variants in different diseases. And also point out the need to decipher the new definition for junk DNA comprising of such intronic regions, promoters, splice site variants and intergenic regions within the genome. These regions must have some regulatory functions that lead towards the modification of gene expression. It is also a well-known fact that only 2% of the human genome is protein coding while remaining 98% is although not protein coding and consist of introns, transposons and repeat regions but might have important regulatory functions (Birney *et al.* 2007) (Boyle *et al.* 2012).

Several high throughput molecular technologies have been used to find out the impact of non-coding variants on the regulation of the transcription by determining the structure of chromatin, protein motifs and methylation sites located in these regions by Encyclopedia of DNA Elements (ENCODE) projects (Birney *et al.* 2007) (E. P. Consortium 2012).

To find out such data and further assess the role of specific variants on binding of transcription factors specific databases like Regulome DB have been developed and utilized in some studies to find out the association of genes with rare and complex diseases. GWAS of Epilepsy have identified 34 significantly associated risk SNPs, common SNPs were selected once so total 32 SNPs found (Guo *et al.* 2011). We have selected mostly GWAS with search word epilepsy only. In this study, we have utilized the search (SNAP) tool (www.broadinstitute.org/mpg/snap/) (A. D. Johnson *et al.* 2008) and RegulomeDB (www.regulomedb.org) (Boyle et al. 2012) to find the potential regulatory functions of recently identified, SNPs variants (index and proxy SNPs) for known and suggestive loci associated with risk of epilepsy.

Methods

SNP selection

A literature search was made to find out the single-nucleotide polymorphisms (SNPs) from genome wide association studies with significant or indicative risk for epilepsy. Six studies with 34 associated two publicly available bioinformatics tools, Broad Institute's SNP Annotation and Proxy

SNPs were selected, SCN1A was being one of the most significant SNP reported in 3 studies, while PTPRD reported in two studies, other GWAS significant SNPs include, (SNX11, SKAP1, CDK5RAP3, PNPO, ATAD4, COPZ2, has-mir-152, NFE2L1, CBX1, ZEB2, VRK2, FANCL, MMP8, CHRM3, GSTA4, GABRA2, PCDH7, AL132875.2, SH3BGRL2, ELOVL4, MAST4, GOLIM4, C1orf94, ADCY9, LPHN3, ARHGAP11B, AL354897.1, ZNF385D and PLA2G4A) (DOHERTY 2010) (Guo *et al.* 2011), (Kasperavičiūtė *et al.* 2013), (T. I. L. A. E. Consortium 2014), (Speed *et al.* 2014), (Steffens *et al.* 2012) and. All associated SNPs (given in the Table 1) are selected from already online published data so Internal Review Board approval and informed consent measures were delineated in those selected publications.

Table 1. GWAS significant SNPs selected for analysis of regulatory function.

SNP ID	Locus	Region	Location	Gene	Risk Allele	Reference	p-value
rs6732655	2q24.3	chr2:166038556	intron variant	*SCN1A*	T	Consortium TILAE 2014.	9 x10^{-10} (All epilepsy)
rs28498976	4p15.1	chr4:31149735	downstream_gene_variant	*PCDH7*	A	Consortium TILAE 2014.	5 x10^{-9} (All epilepsy)
rs2947349	2p16.1	chr2:57832668	intergenic_variant	*VRK2,FANCL*	C	Consortium TILAE 2014.	1 x10^{-8} (GGE)
rs1939012	11q22.2	chr11:102724404	intron_variant	*MMP8*	T	Consortium TILAE 2014.	2 x10^{-8} (GGE)
rs55670112	5q22.3	chr5:114932773	intergenic_variant	intergenic	C	Consortium TILAE 2014.	6 x10^{-8} (GGE)
rs12987787	2q24.3	chr2:166001881	intron_variant	*SCN1A*	T	Consortium TILAE 2014.	1 x10^{-7} (Focal epilepsy)
rs535066	4p12	chr4:46238270	intergenic_variant	*GABRA2*	G	Consortium TILAE 2014.	2 x10^{-7} (All epilepsy)
rs1044352	4p15.1	chr4:31146252	3_prime_UTR_variant	*PCDH7*	T	Consortium TILAE 2014.	2 x10^{-7} (GGE)
rs111577701	3q26.2	chr3:168143620	intergenic_variant	*GOLIM4*	C	Consortium TILAE 2014.	4 x10^{-7} (All epilepsy)
rs13026414	2p16.1	chr2:57706920	intergenic_variant	intergenic	C	EPICURE Consortium, 2012.	2 x10^{-9} (All GGE)
rs72823592	17q21.32	chr17:48045642	non_coding_transcript_exon_variant	*SNX11,SKAP1,CDK5RAP3, PNPO,ATAD4,COPZ2*,hsa-mir-152,*NFE2L1,CBX1*	G	EPICURE Consortium, 2012.	9 x10^{-9} (All GGE)
rs10496964	2q22.3	chr2:144602342	intergenic_variant	*ZEB2*	C	EPICURE Consortium, 2012.	9 x10^{-9} (GAE)
rs12059546	1q43	chr1:239806797	intron_variant	*CHRM3*	G	EPICURE Consortium, 2012.	4 x10^{-8} (JME)
rs2717068	2p16.1	chr2:57867738	intergenic_variant	intergenic	T	EPICURE Consortium, 2012.	4 x10^{-7} (GAE)
rs771390	1p35.1	chr1:34285335	intergenic_variant	C1orf94	C	EPICURE Consortium, 2012.	6 x10^{-7} (All GGE)
rs10030601	4q31.23	chr4:149804060	intron_variant	intergenic	C	EPICURE Consortium, 2012.	1 x10^{-6} (GAE)
rs11890028	2q24.3	chr2:166086767	intron_variant	*SCN1A*	T	EPICURE Consortium, 2012.	4 x10^{-6} (All GGE)
rs12720541	1q31.1	chr1:186900940	intron_variant	*PLA2G4A*	T	EPICURE Consortium, 2012.	9 x10^{-6} (GAE)
rs7587026	2q24.3	chr2:166122240	intron_variant	*SCN1A*	A	Kasperaviciute *et al.*, 2013.	3 x10^{-9} (MTLEHS+FS vs. Controls)
rs2292096	1q32.1	chr1:200857641	3_prime_UTR_variant	*CAMSAP1L1*	NK	Guo Y, *et al.*, 2011	1 x10^{-8}
rs492146	6p12.2	chr6:52971097	intergenic_variant	*GSTA4*	A	Speed D *et al.*, 2013.	2 x10^{-7}
rs72700966	9p23	chr9:10505224	intron_variant	*PTPRD*	C	Speed D *et al.*, 2013.	3 x10-7
rs61670327	5p15.33	chr5:3276582	regulatory_region_variant	*NR*	A	Speed D *et al.*, 2013.	7 x10-7
rs143536437	15q13.2	chr15:30606129	intron_variant	*ARHGAP11B*	C	Speed D *et al.*, 2013.	3 x10-6
rs11861787	16p13.2	chr16:8467743	intergenic_variant	*NR*	C	Speed D *et al.*, 2013.	3 x10-6
rs72698613	4q34.1	chr4:175087067	intergenic_variant	*NR*	A	Speed D *et al.*, 2013.	4 x10-6 6 x10^{-6}
rs12744221	1q21.1	chr1:145789475	intron_variant	*NR*	C	Speed D *et al.*, 2013.	6 x10-6
rs346291	6q14.1	chr6:79855119	non_coding_transcript_exon_variant	*AL132875.2,SH3BGRL2,ELOVL4*	NK	Kasperaviciute D *et al.*, 2010.	3 x10^{-7}
rs2601828	16p13.3	chr16:4053870	intron_variant	*ADCY9*	A	Kasperaviciute D *et al.*, 2010.	1 x10^{-6}
rs2172802	4q13.1	chr4:61587491	intron_variant	*LPHN3*	NK	Kasperaviciute D *et al.*, 2010.	3 x10^{-6}
rs2841498	9q21.33	chr9:85315130	intergenic_variant	*AL354897.1*	NK	Kasperaviciute D *et al.*, 2010.	4 x10^{-6}
rs1490157	3p24.3	chr3:21677754	intron_variant	*ZNF385D*	NK	Kasperaviciute D *et al.*, 2010.	5 x10^{-6}
rs2475335	9p23	chr9:10260263	intron_variant	*PTPRD*	NK	Kasperaviciute D *et al.*, 2010.	9 x10^{-6}

(P-value for most strongly associated SNPs from GWAS along with information describing context of p-value.).

Linkage Disequilibrium

After selection of Genome wide associated SNPs in epilepsy, SNAP tool was assessed on 7th May 2016 to recognize SNPs in linkage disequilibrium (LD) (A. D. Johnson *et al.* 2008). SNAP is user friendly, freely available online tool and permits users to find proxy SNPs based upon LD. We selected the parameters of CEU population from available genomes of 1000 Genome Pilot 1 and Hap Map (3) with queries SNPs as proxy SNPs of themselves and these SNPs were not limited by array.

Search was also done with repeated r^2 ≥0.90 and r^2 ≥1.00 parameters after obtaining with r^2 ≥0.80 in linkage disequilibrium with 34 already selected GWAS SNPs to better understand the LD among these SNPs. Total no. of identified SNPs decreases with increased r^2 threshold values (Supplementary Table S1).

Supplementary **Table S1.** Proxy SNPs for reported GWAS significant SNPs from SNAP web portal search (1000 Genomes and Hap Map3).

SNP	Proxy r^2 ≥ 0.8	Chr.
rs12720541	rs12720541	chr1
	rs6683515	chr1
	rs10911952	chr1
	rs1474590	chr1
	rs7519192	chr1
	rs4651343	chr1
	rs12128551	chr1
	rs12720662	chr1
	rs12144159	chr1
	rs4650708	chr1
	rs7555140	chr1
rs2475335	rs2475335	chr9
	rs2475339	chr9
	rs2498612	chr9
	rs2475349	chr9
	rs1322147	chr9
rs12744221		
rs1490157	rs1490157	chr3
	rs12491351	chr3
	rs9839129	chr3
	rs11710743	chr3

SNP	Proxy $r^2 \geq 0.8$	Chr.
	rs9881055	chr3
	rs986503	chr3
	rs4298061	chr3
	rs9828403	chr3
rs72698613		
rs11890028	rs11890028	chr2
	rs13004083	chr2
	rs10167228	chr2
	rs4667867	chr2
	rs10930201	chr2
	rs3812719	chr2
	rs1461193	chr2
	rs13421166	chr2
	rs7580482	chr2
	rs7574618	chr2
	rs6432860	chr2
	rs2126152	chr2
	rs7601520	Chr2
	rs2298771	chr2
	rs4667862	chr2
	rs2114760	chr2
	rs1841547	chr2
	rs2390322	chr2
	rs1834840	chr2
	rs536744	chr2
	rs565348	chr2
	rs577306	chr2
	rs545238	chr2
	rs567652	chr2
	rs10497280	chr2
	rs17744737	chr2
	rs11884723	chr2
rs2841498	rs2841498	chr9
	rs2841494	chr9
rs11861787	rs11861787	chr16
	rs11865847	chr16
	rs12934978	chr16
	rs11077309	chr16
	rs12919774	chr16
	rs12596798	chr16
	rs12918041	chr16
	rs12920180	chr16
rs143536437		
rs2172802	rs2172802	chr4
	rs35550109	chr4
	rs10018746	chr4
	rs1846161	chr4
rs10030601		
	rs11936885	chr4
	rs2884771	chr4
	rs7695783	chr4
	rs6852150	chr4
	rs17026528	chr4
rs2601828	rs2601828	chr16
rs61670327	rs10063962	chr5
rs771390	rs771390	chr1
	rs771400	chr1
	rs928693	chr1
	rs12028676	chr1
rs111577701		
rs2717068	rs717817	chr2
	rs3886275	chr2
	rs13002041	chr2
	rs2678910	chr2
	rs1518395	chr2
	rs6732310	chr2
	rs2312147	chr2
	rs2717054	chr2
	rs2717055	chr2
	rs1402398	chr2

SNP	Proxy $r^2 \geq 0.8$	Chr.
	rs2717018	chr2
	rs2683620	chr2
	rs1474215	chr2
	rs2717071	chr2
	rs12991325	chr2
	rs10166481	chr2
	rs13026830	chr2
	rs11898858	chr2
	rs10865304	chr2
	rs10496079	chr2
	rs7561842	chr2
	rs12990792	chr2
	rs2683634	chr2
	rs2678908	chr2
	rs2678905	chr2
	rs2717001	chr2
	rs2678903	chr2
	rs2717002	chr2
	rs2465804	chr2
	rs2678880	chr2
	rs6545677	chr2
	rs2717076	chr2
	rs2953439	chr2
	rs10496078	chr2
	rs1460255	chr2
	rs1402399	chr2
	rs2953441	chr2
	rs2717048	chr2
	rs6730037	chr2
	rs2290867	chr2
	rs10496076	chr2
	rs2678901	chr2
	rs2717063	chr2
	rs2717023	chr2
	rs2717024	chr2
	rs2678889	chr2
	rs2717031	chr2
	rs2717036	chr2
	rs1401100	chr2
	rs2717040	chr2
	rs1568253	chr2
	rs6722461	chr2
rs72700966		
	rs17203683	chr9
	rs17793027	chr9
rs39861	rs39861	chr5
	rs30731	chr5
	rs258090	chr5
	rs28636	chr5
rs346291	rs346291	chr6
rs1044352	rs1044352	chr4
	rs16884451	chr4
	rs4692500	chr4
	rs1463847	chr4
	rs16884471	chr4
	rs16884473	chr4
	rs7671046	chr4
	rs1827141	chr4
	rs3857021	chr4
rs535066	rs535066	chr4
	rs567926	chr4
	rs561779	chr4
	rs566776	chr4
	rs517350	chr4
	rs571576	chr4
	rs534787	chr4
	rs504696	chr4
	rs572227	chr4
	rs573400	chr4
	rs522636	chr4
	rs534459	chr4

SNP	Proxy $r^2 \geq 0.8$	Chr.
	rs554112	chr4
	rs526805	chr4
	rs537134	chr4
	rs540363	chr4
	rs1440133	chr4
	rs502038	chr4
	rs505474	chr4
	rs548583	chr4
	rs573935	chr4
	rs543809	chr4
	rs519869	chr4
	rs10010493	chr4
	rs279873	chr4
	rs279867	chr4
	rs279864	chr4
	rs450571	chr4
	rs1808851	chr4
	rs279858	chr4
	rs203656	chr4
	rs279857	chr4
	rs279856	chr4
	rs279852	chr4
	rs279851	chr4
	rs532363	chr4
	rs2083422	chr4
	rs183962	chr4
	rs279843	chr4
	rs279836	chr4
	rs279837	chr4
	rs279839	chr4
	rs279840	chr4
	rs279842	chr4
	rs279841	chr4
	rs279828	chr4
	rs189957	chr4
	rs10805145	chr4
	rs488447	chr4
rs492146		
	rs375872	chr6
	rs449690	chr6
	rs428957	chr6
	rs426169	chr6
	rs387853	chr6
	rs419129	chr6
	rs405729	chr6
	rs375887	chr6
	rs316135	chr6
	rs316128	chr6
	rs385636	chr6
	rs584124	chr6
	rs612483	chr6
	rs384505	chr6
rs12987787		
rs55670112		
	rs1453018	chr5
	rs4618419	chr5
	rs4292441	chr5
	rs11241298	chr5
	rs10478214	chr5
rs7587026	rs7587026	chr2
	rs10930205	chr2
	rs6731869	chr2
	rs580041	chr2
	rs6432879	chr2
	rs11896706	chr2
	rs498918	chr2
	rs13406905	chr2
	rs11692675	chr2
	rs557222	chr2
	rs10188577	chr2
	rs10168027	chr2

SNP	Proxy $r^2 \geq 0.8$	Chr.
	rs11691603	chr2
rs12059546	rs12059546	chr1
	rs1110615	chr1
rs1939012		
	rs1940475	chr11
	rs6590983	chr11
	rs3765620	chr11
	rs11225395	chr11
	rs1939013	chr11
	rs3758857	chr11
	rs10750653	chr11
	rs6590986	chr11
rs2947349		
rs2292096	rs2292096	chr1
	rs2292095	chr1
	rs12733378	chr1
	rs2275193	chr1
	rs16830092	chr1
	rs2275194	chr1
	rs16847251	chr1
	rs10494818	chr1
	rs6658596	chr1
	rs16847180	chr1
	rs6660197	chr1
	rs12742404	chr1
rs10496964	rs10496964	chr2
	rs17690892	chr2
	rs17741930	chr2
	rs1346343	chr2
	rs2162571	chr2
	rs4662386	chr2
	rs7565134	chr2
	rs1035837	chr2
	rs10496965	chr2
	rs767381	chr2
	rs1427500	chr2
	rs1427507	chr2
	rs12691696	chr2
	rs6731249	chr2
rs72823592		
	rs4794321	chr17
	rs12951323	chr17
	rs1986693	chr17
	rs16953461	chr17
	rs16954204	chr17
rs28498976		
	rs16884473	chr4
	rs1827141	chr4
	rs3857021	chr4
	rs16884451	chr4
	rs1044352	chr4
	rs4692500	chr4
	rs1463847	chr4
	rs1827140	chr4
	rs10003037	chr4
	rs11935515	chr4
	rs4580617	chr4
	rs16884471	chr4
	rs7587026	chr2
	rs10930205	chr2
	rs6731869	chr2
	rs580041	chr2
	rs6432879	chr2
	rs11896706	chr2
	rs498918	chr2
	rs13406905	chr2
	rs11692675	chr2
	rs557222	chr2
	rs10188577	chr2
	rs10168027	chr2
	rs11691603	chr2

SNP	Proxy $r^2 \geq 0.8$	Chr.
rs13026414	rs13026414	chr2
	rs13012916	chr2
rs6732655		

Functional Annotation

All the selected SNPs and the proxy SNPs in LD obtained from SNAP (Table S1) tool results using $r^2 \geq 0.80$ after omitting repeated SNPs were considered further to explore their possible regulatory functions using another online tool, Regulome DB (www.regulomedb.org), accessed [8th May 2016] (Guo *et al.* 2011). Regulome DB is freely accessible database, (www.regulomedb.org) used to get functional annotation of SNPs from the non-coding region of the genome by utilizing high-throughput data from the ENCODE Project (ENCODE Project Consortium, 2012) and from eQTL and NCBI Sequence Read Archive. It comprises of a 962 sets of complete data and covers about 100 different types of cells and tissues. It has more than 60 million annotations so regarded as a treasured tool to explore possible regulatory functions of the selected SNPs in disease gene expression and implementation in the phenotype of the disease (Boyle *et al.* 2012). Regulome DB utilized already assigned scores ranging from 1-6 to compare annotation and show the potential roles of the SNPs, provided in the Table 2. Methodology followed for biological functional annotation of epilepsy associated risk loci is also shown in Fig. 1.

Table 2. List of the Regulome DB scores and their relevant functional annotation.

Score	Description
Possible affect binding and linked to expression of a gene target	
1a	eQTL + TF binding + any motif + DNase Footprint + DNase peak
1b	eQTL + TF binding + any motif + DNase Footprint + DNase peak
1c	eQTL + TF binding + matched TF motif + DNase peak
1d	eQTL + TF binding + any motif + DNase peak
1e	eQTL + TF binding + matched TF motif
1f	eQTL + TF binding/DNase peak Possible to affect binding
2a	TF binding + matched TF motif + matched DNase Footprint + DNase peak
2b	TF binding + any motif + DNase Footprint + DNase peak
2c	TF binding + matched TF motif + DNase peak Less possible to affect binding
3a	TF binding + any motif + DNase peak
3b	TF binding + matched TF motif Minimal binding proof
4	TF binding + DNase peak
5	TF binding or DNase peak
6	Other

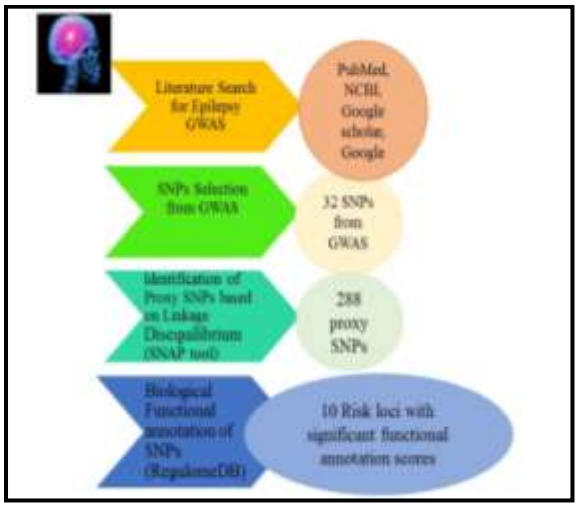

Fig. 1. Illustration of the methodology used for biological functional annotation of epilepsy associated risk loci.

Results

We analyzed 32 SNPs at $r^2$0.80 in SNAP web portal, we found 288 proxy SNPs in LD with GWAS significant SNPs. For the better understanding of association between proxy SNPs and query SNPs, SNAP proxy search was done again with thresholds of r^2 (0.90 and 1.0).

The results obtained were 162 and 101 SNPs with thresholds of r^2 0.90 and r^2 1.0, respectively. With the increase of threshold value, the number of proxy SNPs in LD with 32 GWAS significant SNPs decreased. For rs12744221, rs72698613, rs143536437, rs111577701, rs12987787, rs2947349 and rs6732655 no proxy SNPs were found in 1000 genome pilot 1.

For 288 SNPs analyzed in RegulomeDB, no data was obtained for 131 and these had score of "7" while remaining 157 SNPs had scores of 1–6. Ten SNPs from these remaining data SNPs have RegulomeDB score 1-3. Table 3 shows SNPs and their relative potential regulatory functions. Interestingly, none these 10 SNPs were the selected reported genome-wide significant SNPs. Only 3% of the all 288 annotated SNPs showed regulatory function that is limitation of RegulomeDB as it covers a limited no. of regulatory features.

There are also only six Genome Wide Association Studies with epilepsy associated risk loci that came up in our search because of available small no. of sample size for epilepsy.

We found total of 6 GWAS significant loci that had proxy SNPs with a RegulomeDB score 1-3. Three SNPs, rs375872, rs612483 and rs316128 had strong evidence of regulatory function having a score of 1f. All these three SNPs are in intronic region of *GSTA4* gene and are predicted SNPs having regulatory function in disease pathogenesis and outcome of epilepsy.

Two out of 288 SNPs evaluated has score of 2b and 3a and predicted associated SNPs with epilepsy with *VRK2* gene, rs2678905 and rs2678903 also intronic SNPs and having associated protein motifs Nkx6-1, IPF1, Pou4f3 and Evi-1.

Three SNPs out of 10 SNPs with Regulome DB score of 4 are GWAS associated, rs2601828/ADCY9, rs72823592/COPZ2 and rs12720541/PLA2G4A. Among 48 proxy SNPs with Regulome DB score of 5, and 89 SNPs with score 6, 4 are GWAS loci rs1044352/PCDH7, rs39861/MAST4, rs61670327/LINCO1377 and rs2841498/LOC105376119 and 6 rs28498976/PCDH7, Ars2292096/CAMSAP2, rs2947349/LOC105377632, rs346291/C6orf7, rs11861787/ LOC105371072 and rs1490157/ZNF385D that were in LD with selected GWAS SNP but less likely to have regulatory function according to Regulome DB score. While 13 SNPs with Regulome DB score of 7 were also in LD with GWAS SNP but might have function other than regulation as no data retrieved and these include, rs7587026/SCN1A, rs10496964/linco1412, rs1939012/MMP8, rs12059546/CHRM3, rs55670112/LOC101927078, rs492146/GSTA3, rs72700966/PTPRD, rs2717068/LOC105377632, rs111577701/GOLIM4, rs10030601/LOC285423, rs2172802/ADGRL3, rs12744221/RNF115 and rs2475335/ PTPRD (DOHERTY 2010) (Guo *et al.* 2011), (Steffens *et al.* 2012), (Kasperavičiūtė *et al.* 2013), (T. I. L. A. E. Consortium 2014).

Three SNPs with Regulome DB score of 1f were in the region of *GSTA4* on chromosome 6 (rs375872, rs612483 and rs316128). rs375872 has eQTL and has functional consequences of irtronic variation, with global MAF score of G=0.4499/2253 while remaining two has functional consequences of intergenic variation without eQTL details are provided in the Table 3. SNP rs6590986 located within chromosome 11 and has intergenic variation with RegulomeDB score of 2a and affect protein motifs FOS and Jundm2 and rs2678905 with RegulomeDB score of 2b and rs2678903 with RegulomeDB score of 3a are on chromosome 11 with the region of *VRK2* gene and their functional consequences include intronic variation. These SNPs also affect protein motifs Nkx6-1, IPF1, Pou4f3 and Evi-1 respectively. Another important SNP is rs12596798 has RegulomeDB score of 3a and have location on chromosome 16 affecting binding of motifs FOXC1, FOXC2, FOXJ2, FOXJ3 and Brn2. rs2292095 with Regulome DB score of 3a located on chromosome 1 with the intronic region of CAMSAP2 and also affect binding motif FOXD3. rs986503 has 3a score of Regulome DB and located on chromosome 3 with intonic region of *ZNF38D* having effect on binding of NFATC1 while rs572227 also had Regulome DB score of 3a with location on chromosome 4 within the intergenic region of *GABRA2* gene also affecting binding motif DMRT7.

All these 10 SNPs with RegulomeDB score of 1-3 are predictions by SNAP tool and indicate that these might considered as important SNPs according to regulatory function prediction and may implicated in future studies.

Table 3. List of Proxy SNPs, returned back from RegulomeDB, with most significant regulatory also with some details drom NCBI-SNP (http://www.ncbi.nlm.nih.gov/snp).

Chr.	Co-ordinate (o based)	dbSNP ID	Regulome DB score	Gene/Locus Per dbSNP	Position	eQTL	Motifs	Validated	Global Minor Allele Freq.
Chr6	52834916	rs375872	1f	*GSTA4*	Intergenic	Yes		By 1000G, by 2hit 2allele, by cluster, by frequency, by hapmap	C=0.4794/2401
chr6	52839957	rs612483	1f	*GSTA4*	Intergenic			By 1000G, by 2hit 2allele, by cluster, by frequency, by hapmap, by submitter	T=0.4335/2171
chr6	52849145	rs316128	1f	*GSTA4*	Intronic			by 1000G, by 2hit 2allele, by cluster, by frequency, by hapmap, by submitter	G=0.4499/2253
Chr11	102598858	rs6590986	2a		Intergenic		1: FOS 2:Jundm2	by 1000G, by 2hit 2allele, by cluster, by frequency, by hapmap	C=0.3460/1733
Chr2	58135871	rs2678905	2b	*VRK2*	Intronic		Nkx6-1 IPF1 Pou4f3	by 1000G, by 2hit 2allele, by cluster, by frequency, by hapmap, by submitter	A=0.3413/1709
Chr16	8513276	rs12596798	3a		Intergenic		FOXC1 FOXC2 FOXJ2 FOXJ3 Brn2	by 1000G, by 2hit 2allele, by cluster, by frequency, by hapmap	A=0.2242/1123
Chr1	200826394	rs2292095	3a	*CAMSAP2*	Intronic		FOXD3	by 1000G, by cluster, by frequency, by hapmap	T=0.1538/770
Chr2	58137929	rs2678903	3a	*VRK2*	Intronic		Evi-1	1000G, by 2hit 2allele, by cluster, by frequency, by hapmap, by submitter	A=0.3415/1710
Chr3	21714102	rs986503	3a	*ZNF385D*	Intronic		NFATC1	by 1000G, by cluster, by frequency, by hapmap, by submitter	A=0.2841/1423
Chr4	46251392	rs572227	3a	*GABRA2*	Intergenic		DMRT7	By 1000G, by 2hit 2allele, by cluster, by frequency, by hapmap, by submitter	T=0.3530/1768

Discussion

Total 288 proxy SNPs were returned from SNAP tool and evaluated in Regulome DB. For 5 SNPs RegulomeDB result score was less than 3 that suggests a higher degree association of these SNPs with possible regulatory roles and affect binding. Only 3 of these 5 SNPs have score 1f and one, SNP has score of 2a and one has score of 2b. Other five SNPs have score of 3a. Three proxy SNPs rs375872/GSTA4, rs612483/GSTA4, and rs316128/*GSTA4* with Regulome DB score of 1f are likely to affect binding and linked to expression of *GSTA4*. One of these variant SNP rs316128, is located in intronic regions while two others rs375872 and rs612483 have intergenic localization indicating their association with epilepsy through regulatory network. One SNP including the GWAS significant SNPs rs375872 is eQTL for *GSTA4*. The *GSTA4* gene in human is about 18 kb, have seven exons, located on the chromosome 6p12 (DESMOTS *et al.* 1998). *GSTA4* mRNA was detected in heart, brain, placenta, lung, liver, skeletal muscle, kidney and pancreas with highest expression in brain, heart, placenta and pancreas (Desmots *et al.* 2001).

This enzyme has function for the detoxification of electrophiles and oxidative stress inducing factors in the neurodegenerative diseases, cancers and liver diseases. Glutathione *S*-transferase alpha 4 (*GSTA4*) is a major superfamily GSTs member that are phase-II drug-metabolizing enzymes. This alpha class encodes enzymes that have glutathione peroxidase activity and involved to detoxify lipid peroxidation products and protects neurons after injury (HUBATSCH *et al.* 1998). Calcium signaling pathway, zinc ion binding and phosphatidylinositol signaling pathway are associated with prognosis of epilepsy. These results may also illustrate need of further studies to find any association between non-coding SNPs and epilepsy development or prognosis.

Proxy SNP affecting binding of the FOS motif rs6590986 returned with RegulomeDB score 2a and rs2678905 has score of 2b. These both SNPs likely affect the binding of the different motifs. As rs6590086 has bound motif FOS and Jundm2. FOS is a transcription factor that regulates neuronal cell death and survival. Human neuronal cell death may occur because of excitotoxicity in which excitatory

amino acids like glutamate cause neuronal cell death (Zhang *et al.* 2002). It is also studied in the mice mutant model of FOS that these mice have seizures induced by kainic acid, leading towards excitability of the neurons that ultimately leads towards neuronal cell death. FOS is also involved in the regulation of kainic acid receptor and brain-derived neurotrophic factor expression (Rogaev *et al.* 1993). Moreover, it is also a genetic factor that regulates the mechanisms of the cells to mediate excitability and survival of neurons. It also has a critical role to regulate the development of cells that have fate to develop and maintain the skeleton. It is also a main component of the activator protein-1 (AP-1) transcription factor complex that also includes members of the JUN family. In the promoter region of the FOS there are two elements that control the induction of FOS by nerve growth factor. It is also involved in the fibrous dysplasia. These score of the RegulomeDB suggest this SNP affect binding of FOS motif that might regulate the epilepsy pathogenesis. Jundm2 also known as JDP2 and has N-terminal domain, a central basic region-leucine zipper domain, and a C-terminal domain. It also transcription factor and involved in the interaction of the AP-1 transcription factor complex (C. M. Johnson *et al.* 1997). While SNP rs2678905 affects binding of IPF1, Nkx6-1 and Pou4f3 motifs. IPF1transcription factor is involved in Diabetes mellitus and known risk factors loci for liver and pancreatic diseases and cancers (Fajans *et al.* 2001) (Ma *et al.* 2008). While Pou4f3 has been associated with dominant hearing loss (Weiss *et al.* 2003) and (Kim *et al.* 2013). Proxy SNP rs12596798 with regulomeDB score of 3a affects protein motifs FOXC1, FOXC2, FOXJ2, FOXJ3 and Brn2, rs2292095, rs2678903, rs986503 and rs572227 also came with regulomeDB score of 3a and affects bindings of different protein motifs. Different subtypes of forkhead proteins like FOXC1, FOXC2, FOXJ2, FOXJ3 and FOXD3 are already reported to be associated with complex genetically inherited diseases of eyes, brain malformations and diabetes. All these SNPs mentioned above with RegulomeDB scores of 1-3 predict their potential roles in epilepsy syndromes to be investigated in near future.

But the major genes in case of predicted regulomeDB score of 3a have yet to be known and also direct future goals for researchers to work on.

RegulomeDB is widely used database for variants annotation and their effects on expression of the gene. But it also gives results of selected DNA binding elements in some cells. 132 variants out of 288 examined variants came back with "No data," score. We might not reason their contribution to the expression of the gene as associated with Epilepsy pathogenesis. Similarly, some loci may have higher number of associated SNPs and might have been established to effect expression compare to our selected variants so we are not making assumptions for all putative regulatory variants for a specified locus. The main focus of this study was use of RegulomeDB and to predict regulatory effects on gene expression according to data found in this database. We have not covered the regulation of RNA splicing or predict the changes in protein structure or function. This might be the reason for large proportion of the variants are without any regulatory signal

Conclusions

In order to comprehend and resolve the pathogenesis of the disease completely, understanding of the molecular basis and cellular mechanisms involved in the disease is needed. Results in this study illustrate the future interpretation of GWAS data to include LD structure and different loci association for epilepsy risk. This *in-silico* approach to interpret the GWAS risk loci of epilepsy and identification of potent SNPs emphasizes that attention must be given to these risk variants to identify the link between epilepsy and these variants. This would increase understanding regarding mechanisms of the epilepsy and would help resolve this dilemma completely.

Declaration

All authors declare no conflict of interest.

Acknowledgements

Rubina Dad is partially supported by Higher Education Commission (HEC), Pakistan for their PhD research. We are also grateful to our staff for coffee/tea during the course of manuscript assembly.

References

Birney E, Stamatoyannopoulos JA, Dutta A, Guigó R, Gingeras TR, Margulies EH, 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature **447(7146),** 799-816.

Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, 2012. Annotation of functional variation in personal genomes using RegulomeDB. Genome research **22(9),** 1790-1797.

Consortium EP. 2012. An integrated encyclopedia of DNA elements in the human genome. Nature **489(7414),** 57-74.

Consortium TILAE. 2014. Genetic determinants of common epilepsies: a meta-analysis of genome-wide association studies. The Lancet. Neurology **13(9),** 893.

Desmots F, Rauch C, Henry C, Guillouzo A, Morel F. 1998. Genomic organization, 5′-flanking region and chromosomal localization of the human glutathione transferase A4 gene. Biochemical Journal **336(2),** 437-442.

Desmots F, Rissel M, Loyer P, Turlin B, Guillouzo A. 2001. Immunohistological analysis of glutathione transferase A4 distribution in several human tissues using a specific polyclonal antibody. Journal of histochemistry & cytochemistry **49(12),** 1573-1579.

Doherty C. 2010. Common genetic variation and susceptibility to partial epilepsies: A genome-wide association study.

Fajans SS, Bell GI, Polonsky K. S. 2001. Molecular mechanisms and clinical pathophysiology of maturity-onset diabetes of the young. New England Journal of Medicine **345(13),** 971-980.

Fisher RS, Acevedo C, Arzimanoglou A, Bogacz A, Cross JH, Elger CE, 2014. ILAE official report: a practical clinical definition of epilepsy. Epilepsia **55(4),** 475-482.

Guo Y, Baum LW, Sham PC, Wong V, Ng PW, Lui CHT, 2011. Two-stage genome-wide association study identifies variants in CAMSAP1L1 as susceptibility loci for epilepsy in Chinese. Human molecular genetics ddr550.

Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, 2009. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proceedings of the National Academy of Sciences **106(23),** 9362-9367.

Hubatsch I, Ridderström M, Mannervik B. 1998. Human glutathione transferase A4-4: an alpha class enzyme with high catalytic efficiency in the conjugation of 4-hydroxynonenal and other genotoxic products of lipid peroxidation. Biochemical Journal **330(1),** 175-179.

Johnson AD, Handsaker RE, Pulit SL, Nizzari MM, O'Donnell CJ, De Bakker PI. 2008. SNAP: a web-based tool for identification and annotation of proxy SNPs using Hap Map. Bioinformatics **24(24),** 2938-2939.

Johnson CM, Hill CS, Chawla S, Treisman R, Bading H. 1997. Calcium controls gene expression via three distinct pathways that can function independently of the Ras/mitogen-activated protein kinases (ERKs) signaling cascade. The Journal of neuroscience **17(16),** 6189-6202.

Kasperavičiūtė D, Catarino CB, Matarin M, Leu C, Novy J, Tostevin A, 2013. Epilepsy, hippocampal sclerosis and febrile seizures linked by common genetic variation around SCN1A. Brain **136(10),** 3140-3150.

Kim HJ, Won HH, Park KJ, Hong SH, Ki CS, Cho SS, 2013. SNP linkage analysis and whole exome sequencing identify a novel POU4F3 mutation in autosomal dominant late-onset nonsyndromic hearing loss (DFNA15). PloS one **8(11),** e79063.

Ma J, Chen M, Wang J, Xia HH, Zhu S, Liang Y, 2008. Pancreatic duodenal homeobox-1 (PDX1) functions as a tumor suppressor in gastric cancer. Carcinogenesis **29(7),** 1327-1333.

Rogaev E, Lukiw W, Vaula G, Haines J, Rogaeva E, Tsuda T, 1993. Analysis of the c-FOS gene on chromosome 14 and the promoter of the amyloid precursor protein gene in familial Alzheimer's disease. Neurology **43(11),** 2275-2275.

Sillanpää M, Besag F, Aldenkamp A, Caplan R, Dunn DW, Gobbi G. 2016. Psychiatric and Behavioural Disorders in Children with Epilepsy (ILAE Task Force Report): Epidemiology of psychiatric/behavioural disorder in children with epilepsy. Epileptic Disorders **1(1).**

Speed D, Hoggart C, Petrovski S, Tachmazidou I, Coffey A, Jorgensen A, 2014. A genome-wide association study and biological pathway analysis of epilepsy prognosis in a prospective cohort of newly treated epilepsy. Human molecular genetics **23(1),** 247-258.

Steffens M, Leu C, Ruppert AK, Zara F, Striano P, Robbiano A, 2012. Genome-wide association analysis of genetic generalized epilepsies implicates susceptibility loci at 1q43, 2p16. 1, 2q22. 3 and 17q21. 32. Human molecular genetics **21(24),** 5359-5372.

Weiss S, Gottfried I, Mayrose I, Khare SL, Xiang M, Dawson SJ, 2003The DFNA15 deafness mutation affects POU4F3 protein stability, localization, and transcriptional activity. Molecular and cellular biology **23(22),** 7957-7964.

Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, 2014. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic acids research **42(D1),** D1001-D1006.

Zhang J, Zhang D, McQuade JS, Behbehani M, Tsien JZ, Xu M. 2002. c-fos Regulates neuronal excitability and survival. Nature genetics **30(4),** 416-420.