Journal of Genetics (2023)102:16
https://doi.org/10.1007/s12041-022-01413-7

RESEARCH ARTICLE

Celiac disease-associated loci show considerable genetic overlap with neuropsychiatric diseases but with limited transethnic applicability

NIDHI SHARMA[1], PRATIBHA BANERJEE[1], AJIT SOOD[2], VANDANA MIDHA[3], B. K. THELMA[4] and SABYASACHI SENAPATI[1]*

[1]*Immunogenomics Laboratory, Department of Human Genetics and Molecular Medicine, Central University of Punjab, Bathinda 151 401, India*
[2]*Department of Gastroenterology, Dayanand Medical College and Hospital, Ludhiana 141 001, India*
[3]*Department of Medicine, Dayanand Medical College and Hospital, Ludhiana 141 001, India*
[4]*Department of Genetics, University of Delhi South Campus, New Delhi 110 021, India*
*For correspondence. E-mail: sabyasachi1012@gmail.com.

Received 13 September 2022; revised 28 October 2022; accepted 9 November 2022

Abstract. Clinical and public health research has revealed the co-occurrence of several neuropsychiatric diseases among patients with celiac disease (CD). The significant presence of CD-specific autoantibodies in patients with neuropsychiatric diseases and vice versa are often reported. To explain the genetic basis of such frequent disease co-occurrence and investigate the underlying common pathways/processes, we performed an extensive cross-disease association study followed by supporting *in silico* functional validation of the leads. Genomewide association study (GWAS) data for CD and eight commonly co-occurring neuropsychiatric diseases from Caucasian populations were analysed, and the shared loci were determined. We performed Immunochip-based fine mapping of these overlapping association signals in an independent European CD data and tested their cross-ethnic transferability using CD association data from the genetically distinct north Indian population. This study identified 12 shared loci between the two diseases with genomewide significance ($P \leq$ 5e-8). Of these five loci, namely *NFIA*, *KIA1109*, *NOTCH4-TSBP1-PBX2*, *HLA-DQA1* and *CSK* replicated in an independent Dutch cohort representing European ancestry. Three of these loci, namely *NFIA*, *NOTCH4-TSBP1-PBX2* and *HLA-DQA1* that are common between CD, anxiety, migraine and schizophrenia respectively withstood locus transferability test in north Indians. Tissue-specific eQTL analysis of SNPs from transferable loci revealed expression QTL effects in brain tissue besides the small intestine and whole blood. Pathway analysis and evidence of epigenetic regulation highlighted the potential contribution of these SNPs to disease pathology. The replicable and transferable association of genetic variants from MHC locus and their functional implications suggest the process of antigen presentation and adaptive/innate immune response regulated by non-HLA genes in the locus may dominate the shared pathogenesis of CD and neuropsychiatric diseases. Functional validation of the shared candidate genes is warranted to unravel the molecular mechanism for the co-occurrence of CD and specific neuropsychiatric diseases.

Keywords. celiac disease; anxiety; migraine; schizophrenia; immunochip association study; genomewide association.

Introduction

Celiac disease (CD) is an autoimmune gastrointestinal disease triggered by dietary gluten protein in genetically predisposed individuals. Most CD patients (95%) are carriers of either HLA-DQ2 or DQ8 haplotypes. Several non-HLA loci were identified to provide genetic susceptibility for CD. Onsets of co-morbid conditions, such as neuropsychiatric diseases are quite often associated with adult (> 50%) onset of nonclassical CD (Lebwohl *et al.* 2015). Most common and first line of treatment includes a gluten-free diet (GFD), which frequently improves the disease symptoms but fails to

Nidhi Sharma and Pratibha Banerjee contributed equally to this work.

Supplementary Information: The online version contains supplementary material available at https://doi.org/10.1007/s12041-022-01413-7.

Published online: 10 January 2023

subside comorbid conditions (Ludvigsson *et al.* 2015). CD susceptibility genes have often been shown associated with multiple extra-intestinal manifestations. Neuropsychiatric diseases, schizophrenia, panic disorder, depression, migraine, epilepsy, mental retardation, headache, balance disorder, bipolar disorder etc. are commonly occurring neuropsychiatric disorders in CD patients besides cerebral ataxia and peripheral neuropathy (Sharma *et al.* 2021).

Concurrence of CD and neuropsychiatric diseases are known and could be due to the contribution of shared molecular process/pathways in genetically susceptible subjects. A number of studies conducted in the recent past have found significant associations between various neuropsychiatric complications and CD. Nearly 15% of all known CD-associated genes have a prominent role in neuronal health or nervous system functioning. Cooke and Smith (1966) were the first to publish an extensive report on the relatedness of neuropsychiatric disorders with gastrointestinal problems including CD. Around 22% of CD patients with neuropsychiatric disorders, and 57% of patients with neuropsychiatric issues test positive for anti-gliadin antibodies (AGAs) (Jackson *et al.* 2012). It has been evaluated that one-fifth of the adult CD patients suffer from various neuropsychiatric disorders, of which gluten ataxia and peripheral neuropathy were found to be most common (Durazzo *et al.* 2022). UK-based cohort study by Hadjivassiliou *et al.* (2019) has reported that at the time of diagnosis 67% of CD cases were having signs of neuropsychiatric disorders including gait instability loss of balance, persisting sensory disturbance, headache, gait ataxia, nystagmus etc. (Hadjivassiliou *et al.* 2019). A number of mechanisms are put forward to explain the deleterious effects of gluten-related neuropsychiatric disorders, including increased levels of inflammatory cytokines, low brain serotonin levels, or nutritional deficiencies. There is strong evidence of involvement of AGA in brain pathology either by pro-inflammatory induction or by cross-reactivity. Besides this anti-tTG6 antibodies are the pathogenic trigger for neurological manifestation in CD (Jackson *et al.* 2012; Losurdo *et al.* 2018). Other antibodies, including Anti-GAD (glutamic acid decarboxylase) and anti-ganglioside are associated with CD and various neurological complications. Sixty per cent of the patients with gluten allergy were having high concentrations of anti-GAD antibodies in the Purkinje cells and peripheral nerves and this prevalence is much higher in patients diagnosed with both CD and neuropsychiatric disorders (Jackson *et al.* 2012).

An immunochip association study identified a novel genetic association for ankyrin-3 (*ank3*) among the north Indian CD cohort. *Ank-3* is also known as a major susceptibility gene for bipolar disorder, schizophrenia and amyotrophic lateral sclerosis (Senapati *et al.* 2016). On this background, combined analysis of genomewide association studies (GWAS) reported on CD and relevant neuropsychiatric diseases were reanalysed to identify shared genetic variants and loci. To determine the significance of shared loci in ethically distinct north Indian population, locus transferability was also performed using immunochip genotype data detailed elsewhere (Senapati *et al.* 2015). *In silico* functional validation was performed to justify the role of shared loci.

Materials and methods

***Selection of the study*:** All the published GWAS reports on CD, and eight neuropsychiatric conditions, namely anxiety, depression, headache, epilepsy, migraine, bipolar disorder, panic disorder and schizophrenia that were enlisted till June 2022 in GWAS catalog in https://www.ebi.ac.uk/gwas were considered. Disease names as mentioned above were used as keywords to search all the reports documented in GWAS catalog. Only those relevant neuropsychiatric diseases that commonly co-occur among CD patients and studied in our previous report were included in this study (Sharma *et al.* 2021).

***Retrieval of association summary statistics*:** Association statistics of each of the selected GWAS were retrieved from the available repositories. The following statistics were taken into consideration: disease/trait, publication details, sample size, population name, reported genes, associated SNP, risk allele, physical location (BP), chromosome number, reported *P* value and odds ratio (95%CI). The position of SNPs was expressed as per hg38 built and previous built were converted using UCSC hgLiftOver browser. All the associated SNPs with $P \leq$ 5e-08 were considered for the study.

***Immunochip genotyping data*:** Immunochip association data on CD was also retrieved for a representative European population (Dutch cohort) previously reported elsewhere (Trynka *et al.* 2010; Senapati *et al.* 2015). Comparable immunochip association summary statistics for a north Indian association study on CD were present in-house and published elsewhere (Senapati *et al.* 2015, 2016). A total of ~200 immune loci were covered in this SNP genotyping panel.

***Cross-disease association*:** Each of the reported SNPs was mapped within a 200-kb locus window (± 100 kb). Mapping was done as per hg38 built. Overlapping windows between GWAS reports of CD loci and neuropsychiatric conditions were considered as shared loci and taken forward for further investigation. Cross-disease association among published GWAS was confirmed at Bonferroni corrected $P \leq$ 5e-08. Replication of the association of the common and shared loci was further evaluated in the Celiac disease immunochip association data from the Dutch study (representative data for the European population) mentioned above. The appropriate level of significance following Bonferonni correction was used to determine replication. Per locus transferability of these replicated loci to genetically distinct north Indian population was also checked using Immunochip genotype data. Transferability was evaluated using *P*-value cut-off 0.05.

Plink-1.07 tool (Purcell *et al.* 2007) was used to reanalyse and manage the Immunochip genotype data. Calculation of pair-wise linkage disequilibrium (LD), tagSNPs, haplotypes and LD were generated using Haploview (Barrett *et al.* 2005).

In silico functional and regulatory validations

To examine the expression quantitative trait locus (eQTL) properties of the transferable SNPs in north Indian CD study GTEx portal v8 (https://gtexportal.org) was used (Kim-Hellmueth *et al.* 2020). To investigate the functional implications of all the markers from each loci, single-tissue eQTL was evaluated. Whole blood, small intestine and whole brain were taken into consideration because of their relevance in the pathogenesis of CD and neuropsychiatric diseases.

The regulatory significance of the variants from transferable loci was checked on ENCODE database (https://www.encodeproject.org/) (ENCODE Project Consortium 2012). Epigenetic modifications and transcription factor binding sites of each transferable SNP were evaluated. These include sites for DNase1 hypersensitivity, (H3K4me1/3), acetylation (H3K27ac), histone methylation, chromatin remodelling, histone modification, and CTCF binding. From ENCODE, tissue-specific epigenetic regulation data were evaluated for blood components T and B-cell, small intestine and brain, which directly contribute to CD and different neuropsychiatric conditions.

Evaluation of molecular networks and interactions

Pathway enrichment analysis was performed using NetworkAnalyst 3.0 (Zhou *et al.* 2019). To investigate the cumulative functional implication of the shared genes, tissue-specific protein–protein interactions were evaluated. Gene set enrichment analysis (GSEA) was performed to emphasize on the molecular pathways, which are overrepresented by the common genes identified for cross-disease association in this study. Curated databases such as KEGG, Reactome, gene ontology (biological process) and PANTHER (biological process) were used for this enrichment analysis. Significant molecular interactions were identified based on the adjusted *P* value. All the GWAS data is available in the GWAS catalog-EMBL-EBI. Raw genotype data for north Indians and Dutch can be obtained from Sabyasachi Senapati and Cisca Wijmenga, respectively.

Results

Eligible studies

A total of 192 eligible studies for eight neuropsychiatric conditions were considered for the present study, which include two studies each on anxiety and headache, four studies each on epilepsy and panic disorder, seven on migraine, 43 on depression, 52 on bipolar disorder and 78 on schizophrenia. A total of 13 studies were available on CD.

Selected markers

From the retrieved summary statistics, 147/158 variants for anxiety, 28/30 for headache, 26/154 for epilepsy, 2/22 for panic disorder, 82/219 for migraine, 954/2125 for depression, 283/858 for bipolar disorder and 1723/2729 for schizophrenia were found with genomewide significant $P \leq$ 5e-08. While a total of 56 variants of the 250 SNPs of CD were included with significant $P \leq$ 5e-08.

Locus sharing and transferability

Overall, 12 loci were identified to be shared between at least one of the eight neuropsychiatric conditions (one each from anxiety, bipolar disorder, epilepsy and migraine; three from depression, six from schizophrenia) and CD. Extended MHC region were found to have most significantly associated variants, which is shared between these diseases. A locus (*NOTCH4-TSBP1-TNXB*) was found shared between two neuropsychiatric diseases, namely migraine and schizophrenia. From this locus only one SNP each, rs424232 (P = 5.00e−21) and rs1400029 (P = 4.00e−08) were reported for CD and migraine, respectively, whereas 28 SNPs with higher significance levels were identified for schizophrenia (table 1). CD-specific risk variant rs2187668 (P = 1.00e−50, OR = 6.23) from *HLA-DQA1* was found to be independent marker within the loci except for rs9275602 ($r^2 > 0.9$) located at 3-kb 5′ of *XXbac-BPG254F23.7*, a noncoding transcript-coding region localized in between *HLA-DQB3* and *HLA-DQB1*.

Notably, same variant rs4625 from *DAG1* locus were reported for both CD and depression with similar strength of association. Variant rs188839109 from *IGF2-C11orf21-TSPAN32* locus was reportedly associated with schizophrenia and BD with same degree of association. Another variant rs17885785 from this locus was reported associated with CD (P = 8.00e-11). Both these SNPs are LD independent in Europeans and Asians. SNP rs188839109 is a missense variation located in the second exon of the uncharacterized protein-coding gene *C11orf21*. These two SNPs from *IGF2-C11orf21-TSPAN32* are not in LD ($r^2 < 0.2$). Further details of the shared loci are given in the table 1.

Of these 12 shared loci, only 10 loci represented by 33 SNPs could be mapped in the immunochip association summary statistics on CD for the Dutch population. A total of 17 SNPs localized within five loci were found to be replicated following bonferroni corrections ($P \leq 0.002$). These five loci were *NFIA* (rs6691768 with P = 0.002); *KIA1109* (rs13151961 with P = 6.51e-05); *NOTCH4-*

Table 1. Shared genetic loci between celiac disease and eight neuropsychiatric diseases. Summary statistics were collected from GWAS catalog documented in EBI-MBBL. SNPs with *P* value < 5e-08 were only considered. The position of the SNPs was noted from Haploreg (https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php).

	Celiac disease (GWAS reports)				
Locus	Chr: BP (hg.38)	Gene	SNP	*P*-value	OR
1	Chr1: 61326191	*NFIA*	rs6691768	1.00e-09	1.12
2	Chr1: 200906114	*INAVA*	rs12132349	8.00e-09	1.12
	Chr1: 200906769	*INAVA*	rs59655222	3.00e-08	0.11
	Chr1: 200912264	*INAVA*	rs10800746	3.00e-08	1.12
	Chr1: 200923009	7.3 kb 3′ of *C1orf106*	rs296547	4.00e-09	1.12
3	Chr2: 100142823	*AFF3*	rs11123810	7.00e-12	0.91
4	Chr3: 49534707	*DAG1*	rs4625	8.00e-11	
5	Chr4: 122194347	*KIA1109*	rs13151961	2.00e-27	1.35
	Chr4: 122117140	29 kb 5′ of *Metazoa SRP*	rs62323881	6.00e-09	1.16
	Chr4: 122340375	*KIA1109*	rs62321692	1.00e-08	0.19
	Chr4: 122345366	*KIA1109*	rs74643882	4.00e-11	1.62
	Chr4: 122451978	*IL2*	rs2069772	1.00e-08	0.12
6	Chr6: 32240547	15 kb 5′ of *XXbac-BPG154L12.4*	rs424232	5.00e-21	
7	Chr6: 32638107	HLA-DQA1, LOC107986589	rs2187668	1.00e-50	6.23
8	Chr11: 2146620	*IGF2,IGF2-AS, INS-IGF2*	rs17885785	8.00e-11	
	Chr11: 2146620	*IGF2,IGF2-AS, INS-IGF2*	rs17885785	8.00e-11	
9	Chr12: 56002984	*SUOX*	rs1689510	4.00e-09	
10	Chr15: 74804102	*CSK*	rs1378938	8.00e-09	1.13
	Chr15: 74839978	*ULK3*	rs936229	3.00e-08	0.11
11	Chr16: 28329624	5.8 kb 3′ of *SBK1*	rs12598357	4.00e-09	
12	Chr20: 45967568	*ZNF335*	rs6032606	5.00e-08	0.23

	Neuropsychiatric disease (GWAS reports)					
Locus	BP (hg.38)	Gene	SNP	*P*-value	OR	Reported association
1	Chr1: 61277021	*NFIA*	rs332828	4.00e-08	0.01	Anxiety
2	Chr1: 200857641	*CAMSAP2*	rs2292096	1.00e-08	1.59	Epilepsy
3	Chr2: 100190052	18 kb 5′ of *AC104782.3*	rs6712515	2.00e-15	0.02	Schizophrenia
4	Chr3: 49534707	*DAG1*	rs4625	4.00e-10	0.02	Depression
	Chr3: 49344465	3.8 kb 5′ of *USP4*	rs6774721	4.00e-08	0.02	
5	Chr4: 122201701	*KIA1109*	rs77087420	1.00e-10	0.03	Depression
	Chr4: 122201701	*KIA1109*	rs77087420	2.00e-10	1.02	
	Chr4: 122265238	*KIA1109*	rs45510091	8.00e-21	1.05	
6	Chr6: 32238272	*NOTCH4*	rs1400029	4.00e-08	1.10	Migraine
	Chr6: 32058086	*TNXB*	rs369637	2.00e-10	1.10	Schizophrenia
	Chr6: 32082981	*TNXB*	rs1150754	3.00e-12	1.12	
	Chr6: 32147408	950 bp 3′ of *PRRT1*	rs3134951	5.00e-13	1.10	
	Chr6: 32184665	*AGER, PBX2*	rs1800625	3.00e-09	3.78	
	Chr6: 32185629	*AGER, PBX2*	rs169504	2.00e-12	1.10	
	Chr6: 32185632	*AGER, PBX2*	rs1004095	3.00e-08	1.06	
	Chr6: 32186508	*PBX2*	rs204995	5.00e-10	1.08	
	Chr6: 32197667	*NOTCH4*	rs2071278	4.00e-14	1.12	
	Chr6: 32205216	*NOTCH4*	rs3131296	2.00e-10	1.19	
	Chr6: 32223264	*NOTCH4*	rs434841	3.00e-08	1.08	
	Chr6: 32228798	4.7 kb 5′ of *NOTCH4*	rs9267858	4.00e-11	1.1	
	Chr6: 32238466	14 kb 5′ of *NOTCH4*	rs9267920	3.00e-15	1.12	
	Chr6: 32239404	15 kb 5′ of *NOTCH4*	rs3130304	5.00e-11	1.10	
	Chr6: 32242482	13 kb 5′ of *XXbac-BPG154L12.4*	rs3130305	1.00e-12	1.12	
	Chr6: 32243308	12 kb 5′ of *XXbac-BPG154L12.4*	rs412657	2.00e-10	1.08	
	Chr6: 32368989	*TSBP1, TSBP1-AS1*	rs3129939	2.00e-13	1.12	
	Chr6: 32371870	*TSBP1, TSBP1-AS1*	rs2050189	4.00e-11	1.11	
	Chr6: 32374760	*TSBP1, TSBP1-AS1*	rs3129945	1.00e-12	1.11	
	Chr6: 32381780	*TSBP1, TSBP1-AS1*	rs3117103	1.00e-14	1.14	
	Chr6: 32410726	*BTNL2*	rs9268510	3.00e-11	1.11	
	Chr6: 32423415	12 kb 5′ of *BTNL2*	rs3135357	2.00e-16	1.13	
	Chr6: 32464185	19 kb 3′ of *HLA-DRA*	rs9268895	9.00e-14	1.67	
	Chr6: 32465853	21 kb 3′ of *HLA-DRA*	rs9268943	6.00e-13	1.15	
	Chr6: 32466042	21 kb 3′ of *HLA-DRA*	rs9268950	4.00e-13	1.14	
	Chr6: 32468178	23 kb 3′ of *HLA-DRA*	rs9269000	5.00e-12	1.15	
	Chr6: 32469754	16 kb 5′ of *SRY*	rs9269028	7.00e-17	1.14	
	Chr6: 32470272	25 kb 3′ of *HLA-DRA*	rs9269039	3.00e-09	1.12	
	Chr6: 32487060	30 kb 3′ of *HLA-DRB5*	rs142972412	1.00e-12	1.09	

Table 1. (*contd*)

	Neuropsychiatric disease (GWAS reports)					
Locus	BP (hg.38)	Gene	SNP	*P*-value	OR	Reported association
7	Chr6: 32510807	6.5 kb 3′ of *HLA-DRB5*	rs188190243	3.00e-08	1.11	Schizophrenia
	Chr6: 32511744	5.6 kb 3′ of *HLA-DRB5*	rs116182620	3.00e-11	1.19	
	Chr6: 32514520	2.8 kb 3′ of *HLA-DRB5*	rs184153866	2.00e-13	1.1	
	Chr6: 32517285	*HLA-DRB5*	rs115641444	3.00e-09	1.1	
	Chr6: 32518167	*HLA-DRB5*	rs139547629	1.00e-09	1.09	
	Chr6: 32519189	*HLA-DRB5*	rs184981897	2.00e-08	1.09	
	Chr6: 32520273	*HLA-DRB5*	rs113397282	6.00e-12	1.11	
	Chr6: 32520992	*HLA-DRB5*	rs149961934	3.00e-08	1.08	
	Chr6: 32521255	*HLA-DRB5*	rs114875775	4.00e-08	1.13	
	Chr6: 32521819	*HLA-DRB5*	rs191269336	5.00e-09	1.09	
	Chr6: 32522018	*HLA-DRB5*	rs144532965	1.00e-11	1.11	
	Chr6: 32522469	*HLA-DRB5*	rs139480376	6.00e-09	1.09	
	Chr6: 32522589	*HLA-DRB5*	rs184538485	7.00e-09	1.08	
	Chr6: 32522830	*HLA-DRB5*	rs140849564	4.00e-08	1.07	
	Chr6: 32524630	*HLA-DRB5*	rs112209031	2.00e-08	1.09	
	Chr6: 32525190	*HLA-DRB5*	rs191843781	3.00e-11	1.14	
	Chr6: 32525469	*HLA-DRB5*	rs145470632	5.00e-10	1.12	
	Chr6: 32526115	*HLA-DRB5*	rs147793969	1.00e-09	1.11	
	Chr6: 32526320	*HLA-DRB5*	rs189600472	8.00e-11	1.12	
	Chr6: 32526927	*HLA-DRB5*	rs117616320	1.00e-10	1.10	
	Chr6: 32528783	*HLA-DRB5*	rs193267147	2.00e-11	1.22	
	Chr6: 32531497	*HLA-DRB5*	rs114812317	3.00e-10	1.14	
	Chr6: 32534622	4.3 kb 5′ of *HLA-DRB5*	rs180778602	4.00e-12	1.15	
	Chr6: 32542425	7.5 kb 3′ of *U1*	rs182908437	3.00e-11	1.15	
	Chr6: 32557141	*HLA-DRB6*	rs9469174	3.00e-10	1.10	
	Chr6: 32558215	*HLA-DRB6*	rs191239160	7.00e-09	1.08	
	Chr6: 32559333	*HLA-DRB6*	rs78110044	3.00e-14	1.22	
	Chr6: 32566867	12 kb 3′ of *HLA-DRB1*	rs147976543	4.00e-11	1.1	
	Chr6: 32568724	10 kb 3′ of *HLA-DRB1*	rs185717927	7.00e-13	1.23	
	Chr6: 32569865	8.9 kb 3′ of *HLA-DRB1*	rs184123737	3.00e-09	1.13	
	Chr6: 32570699	8.1 kb 3′ of *HLA-DRB1*	rs111639056	2.00e-15	1.15	
	Chr6: 32571330	7.4 kb 3′ of *HLA-DRB1*	rs9269271	4.00e-15	1.11	
	Chr6: 32574236	4.5 kb 3′ of *HLA-DRB1*	rs41293330	2.00e-09	1.09	
	Chr6: 32575798	3 kb 3′ of *HLA-DRB1*	rs41294271	3.00e-08	1.11	
	Chr6 32582101	*HLA-DRB1*	rs41293179	2.00e-11	1.11	
	Chr6: 32582137	*HLA-DRB1*	rs142790902	1.00e-11	1.22	
	Chr6: 32583926	*HLA-DRB1*	rs144660248	6.00e-13	1.18	
	Chr6: 32584671	*HLA-DRB1*	rs11753207	3.00e-11	1.1	
	Chr6: 32586352	*HLA-DRB1*	rs9270074	1.00e-09	1.08	
	Chr6: 32588824	*HLA-DRB1*	rs28724212	1.00e-09	1.14	
	Chr6: 32597688	7.8 kb 5′ of *HLA-DRB1*	rs2760981	4.00e-11	1.08	
	Chr6: 32627606	572 bp 5′ of *HLA-DQA1*	rs9271871	2.00e-12	1.1	
	Chr6: 32631878	*HLA-DQA1*	rs9272081	1.00e-10	1.09	
	Chr6: 32635879	*HLA-DQA1*	rs116139966	1.00e-09	1.08	
	Chr6: 32645524	*HLA-DQA1*	rs9273177	1.00e-13	1.14	
	Chr6: 32646583	*HLA-DQA1*	rs115443066	4.00e-09	1.08	
	Chr6 32654605	*HLA-DQA1*	rs17843707	7.00e-11	1.14	
	Chr6 32656240	3.2 kb 3′ of *HLA-DQB*	rs145607970	1.00e-08	1.14	
	Chr6: 32664181	*HLA-DQB1*	rs9274299	4.00e-13	1.11	
	Chr6: 32664882	*HLA-DQB1*	rs9274390	2.00e-14	1.13	
	Chr6: 32668221	*HLA-DQB1*	rs9274623	6.00e-19	1.14	
	Chr6: 32668587	*HLA-DQB1*	rs9274657	1.00e-08	1.07	
	Chr6: 32668836	452 bp 5′ of *HLA-DQB1*	rs9274675	3.00e-10	1.08	
	Chr6: 32700133	18 kb 5′ of *XXbac-BPG254F23.7*	rs2647044	6.00e-18	1.16	
	Chr6: 32706552	11 kb 5′ of *XXbac-BPG254F23.7*	rs9275511	5.00e-10	1.07	
	Chr6: 32715035	3 kb 5′ of *XXbac-BPG254F23.7*	rs9275602	3.00e-08	1.09	
	Chr6: 32725517	6.7 kb 3′ of *XXbac-BPG254F23.7*	rs9275957	7.00e-14	1.11	
	Chr6: 32742213	*HLA-DQA2*	rs116593970	7.00e-10	1.08	
	Chr6: 32793729	19 kb 3′ of *HLA-DOB*	rs4318925	4.00e-12	1.11	
8	Chr11: 2301859	*TSPAN32, C11orf21*	rs188839109	2.00e-08		Schizophrenia
	Chr11: 2301859	*TSPAN32, C11orf21*	rs188839109	2.00e-08		BD
9	Chr12: 56055651	11 kb 3′ of *RPS26*	rs7302200	1.00e-14	0.02	Schizophrenia
10	Chr12: 74735539	9.9 kb 5′ of *CYP1A1*	rs2472297	4.00e-10		Schizophrenia
11	Chr16: 28479196	*CLN3*	rs151181	3.00e-19		Bipolar disorder

Table 1. (*contd*)

	Neuropsychiatric disease (GWAS reports)					
Locus	BP (hg.38)	Gene	SNP	*P*-value	OR	Reported association
12	Chr20: 46052214	*SLC12A5*	rs12624433	2.00e-14	1.02	Depression
	Chr20: 46080900	*NCOA5*	rs6074013	1.00e-09	0.01	
	Chr20: 46093017	3.1 kb 5′ of *NCOA5*	rs4578918	1.00e-11	0.02	

TSBP1-TNXB (rs3131296 with $P = 1.79e\text{-}106$); *HLA-DQA1* (rs2187668 with $P = 8.52e\text{-}121$); and *CSK* (rs1378938 with $P = 0.0004$) (table 2). Locus transferability ($P \leq 0.05$) in north India was observed only for three loci, namely *NFIA*, *NOTCH4-TSBP1-TNXB* and *HLA-DQA1* (table 2).

NFIA (rs6691768; $P_{\text{Dutch}} = 0.002$; $P_{\text{NI}} = 0.04$) is the only nonMHC locus that withstood transferability test. *HLA-DQA1* is well investigated in CD that defines the susceptibility HLA alleles HLA-DQ2 and DQ8. Shared variants rs2187668 and rs9275602 was found to be in moderate LD ($r^2 = 0.5$; $d' = 0.8$) among Dutch but are having absolutely no LD in north Indians ($r^2 = 0.01$; $d' = 1$). Further investigation of the long distance LD in *NOTCH4-TSBP1-TNXB* locus revealed significant decay in LD in north Indian population, where long distance LD was found to be broken when both r^2 and d' were considered separately. In both Dutch and north Indian populations, three LD blocks were identified such as block 1 (rs1800625 and rs204995) and block 2 (rs2071278 and rs3131296) and block 3 (rs3130304, rs4242232 and rs412657). Higher LD was observed in Dutch as compared to north Indian population. rs3130304 and rs424232 showed highest degree of LD in Dutch but showed little or low LD in Indians (figure 1).

Functional implications

eQTL analysis identified to have strong regulatory effects of six SNPs from the *NOTCH4-TSBP1-TNXB* locus. eQTLs were observed for all the three tissue types, namely small intestine, whole blood and brain (table 1 in electronic supplementary material at http://www.ias.ac.in/jgenet/). rs424232 that was originally reported in CD have strong expression effects on *CYP21A1* and C4A in brain tissue. Similar trends were observed for other five variants, which were originally identified, in neuropsychiatric conditions. It is noteworthy that *HLA-DQB1*, *C4A* and *AGER* are the three common genes that were significantly regulated to this set of six variants in all the three relevant tissues.

Strong epigenetic signatures were noted for two SNPs rs204995 of *PBX2* and rs2187668 of *HLA DQA1* in more than one tissue but moderate level of signatures were found for two other variants rs3130304 of *NOTCH4* and rs3129939 of *TSBP1-TSBP1AS1* provided in table 2 in electronic supplementary.

Enriched pathways

Major pathways that were highlighted in whole brain tissue-specific PPI were protein poly-ubiquitination ($P = 1.03e\text{-}10$), ubiquitin-mediated proteolysis ($P = 4.61e\text{-}07$), protein modification by small protein conjugation ($P = 8.5e\text{-}06$), antigen processing ubiquitination and proteasome degradation ($P = 4.2e\text{-}05$), immune system ($P = 0.007$) are listed in table 3 in electronic supplementary material.

Discussion

This study comprehensively analysed the summary statistics of well-powered GWAS on CD and eight neuropsychiatric diseases published until June 2022. Most of the GWAS were reportedly included study participants of European ancestry. Considerable genetic sharing was observed between CD and neuropsychiatric diseases that include anxiety, epilepsy, schizophrenia, depression, migraine and bipolar disease. Each of these loci (defined based on the physical distance in this study) harbours multiple SNPs as tabulated in table 1. Reported SNPs from most of these loci are not in LD and thus arises a possibility of independent association signals from each of these loci. Multiple associations were also observed from two of these shared loci (*NOTCH4-TSBP1-TNXB* and *IGF2-C11orf21-TSPAN32*) where LD independent SNPs from the loci were found to be associated with schizophrenia, bipolar disease and migraine (table 1). Such notable sharing of common loci supports our published systematic review-based meta-analysis where significant concurrence of CD was reported with anxiety, epilepsy, depression, headache, panic disorder and dysthymia (Sharma *et al.* 2021). A strong pathological overlap is evident between CD and neuropsychiatric diseases by the presence of circulating anti-neural antibodies such as anti-ganglioside IgG and anti-neural antibodies to central nervous system in the sera of CD patients (Volta *et al.* 2002, 2006; Briani *et al.* 2005; Mata *et al.* 2006). Further, several studies reported the presence of CD-specific serological markers, such as IgG anti-gliadin, anti-tTG, IgA anti-gliadin and endomysial antibodies in individuals with neuropsychiatric diseases.

Replication of the 12 shared disease associated loci was attempted on a European (represented by a Dutch study) immunochip CD association data to fine map the association signal. Only five loci represented by 17 SNPs could be

Table 2. Results of the replication of shared loci in a Dutch Immunochip-CD association dataset followed by the transferability of the replicated loci in a north Indian Immunochip-CD association dataset.

						Dutch		Indian		
Locus	Chr.	Location	SNP	Context	Gene	*P*-value	OR	P-value	OR	Reported association
1	1	61326191	rs6691768*	Intron	*NFIA*	**0.002**	0.8	***0.04***	0.8	Anxiety
2	1	200906114	rs12132349*	Intergenic	*INAVA*	0.1	0.9	0.2	0.8	Epilepsy
	1	200906769	rs59655222*	Intron	*INAVA*	0.1	0.9	0.2	0.8	Epilepsy
	1	200912264	rs10800746*	Intron	*INAVA*	0.07	0.9	0.6	1.0	Epilepsy
	1	200923009	rs296547*	Intron	7.3 kb 3' of *C1orf106*	0.3	0.9	0.9	1.0	Epilepsy
3	2	100190052	rs6712515#	Intron	18 kb 5' of *AC104782.3*	0.2	1.1	9.5e−03	0.8	Schizophrenia
	2	100142823	rs11123810*	Intron	*AFF3*	0.2	0.9	2.0e−03	0.7	Schizophrenia
4	3	49534707	rs4625*#	3′UTR	*DAG1*	0.6	1.0	0.3	1.1	Depression
5	4	122194347	rs13151961*	Intron	*KIA1109*	**6.513e−05**	0.7	0.1	0.8	Depression
	4	122117140	rs62323881*	Intron	*KIA1109*	0.07	1.2	0.1	1.4	Depression
	4	122340375	rs62321692*	Intron	*KIA1109*	0.2	1.1	0.1	1.4	Depression
	4	122451978	rs2069772*	Intron	*IL2*	0.003	1.2	0.5	1.1	Depression
	4	122201701	rs77087420#	Intron	*KIA1109*	0.5	0.9	0.2	1.6	Depression
6	6	32240547	rs424232*	Intergenic	*NOTCH4 - TSBP1-AS1*	**1.42e−81**	4.6	***1.857e−08***	0.6	Migraine Schizophrenia
	6	32082981	rs1150754#	Intron	*TNXB*	**7.068e−91**	5.6	0.2	0.6	Schizophrenia
	6	32184665	rs1800625#	Intergenic	*AGER, PBX2*	**3.126e−92**	5.7	***0.03***	0.7	Schizophrenia
	6	32205216	rs3131296#	Intergenic	*NOTCH4*	**1.794e−106**	7.5	0.8	1.1	Schizophrenia
	6	32186508	rs204995#	Non_coding_transcript exon_variant	*PBX2*	**2.022e−96**	5.7	***0.01***	0.7	Schizophrenia
	6	32197667	rs2071278#	Intron	*NOTCH4*	**5.599e−99**	6.4	0.2	0.8	Schizophrenia
	6	32223264	rs434841#	Intron	*NOTCH4*	**6.056e−47**	2.7	***2.12e−08***	0.5	Schizophrenia
	6	32239404	rs3130304#	Intergenic	15 kb 5′ of *NOTCH4*	**1.675e−82**	4.7	***0.0004***	0.7	Schizophrenia
	6	32243308	rs412657#	Intergenic	*NOTCH4 - TSBP1-AS1*	**5.265e−53**	0.3	***1.62e−06***	0.6	Schizophrenia
	6	32368989	rs3129939#	Intron	*TSBP1, TSBP1-AS1*	**1.167e−89**	5.5	***2.01e−07***	0.4	Schizophrenia
	6	32371870	rs2050189#	5′ UTR	*TSBP1, TSBP1-AS1*	**9.114e−77**	4.3	***5.10e−14***	2.1	Schizophrenia
	6	32381780	rs3117103#	Intron	*TSBP1, TSBP1-AS1*	**7.884e−112**	8.5	0.8	0.9	Schizophrenia
7	6	32638107	rs2187668*	Intron	*HLA-DQA1*, LOC107986589	**8.515e−121**	10.0	***1.373e−47***	6.7	Schizophrenia
	6	32715035	rs9275602#	Intergenic	3 kb 5′ of *XXbac-BPG254F23.7*	**2.366e−79**	4.8	0.1	0.7	Schizophrenia
10	15	74804102	rs1378938*	Intergenic	*CSK*	**0.0004**	1.3	0.5	1.1	Schizophrenia
	15	74839978	rs936229*	Intron	*ULK3*	0.004	1.2	0.5	1.1	Schizophrenia
11	16	28329624	rs12598357*	Intergenic	5.8 kb 3′ of *SBK1*	0.3	1.1	0.7	1.0	Bipolar Disorder
	16	28479196	rs151181#	Intron	*CLN3*	1.0	1.0	0.7	1.05	Bipolar Disorder
12	20	46052214	rs12624433#	Intron	*SLC12A5*	0.3	1.1	0.7	1.0	Depression
	20	46080900	rs6074013#	Intron	*NCOA5*	0.6	1.0	0.2	0.9	Depression
	20	45967568	rs6032606*	Missense	3.1 kb 5′ of *NCOA5*	0.9	1.0	0.04	0.8	Depression

*GWAS reported markers on CD.
#GWAS reported markers on neuropsychiatric disorder.
P-values of the replicated variants in Dutch dataset are written in bold fronts.
P-values of the transferable varinats in Indian data are bold and italicized

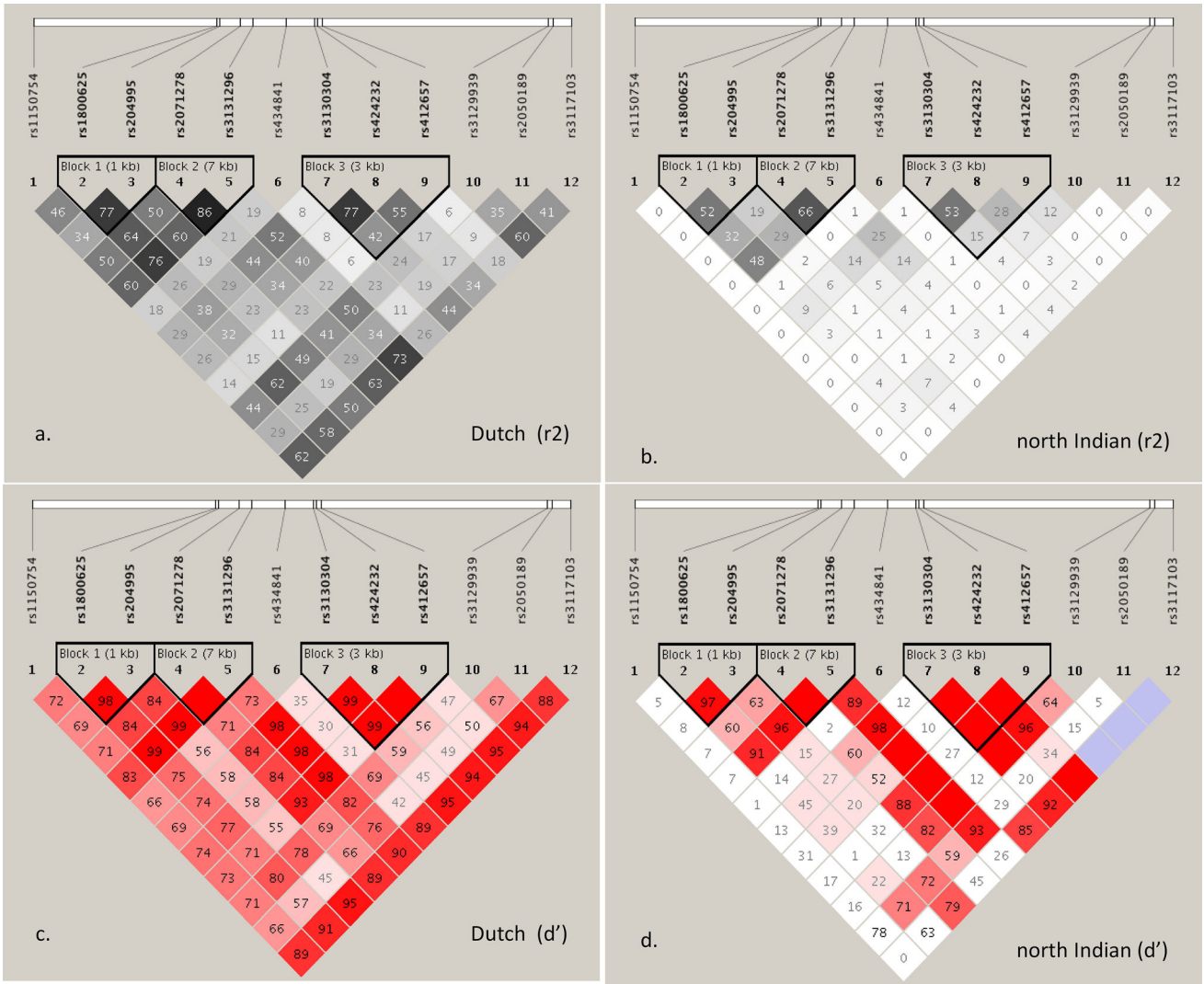

Figure 1. Comparative LD plots showing significantly different patterns of pair-wise LD (r^2 and d′) between 12 SNPs from *NOTCH4-TSBP1-TNXB* locus. Comparative LD plots are showing differences in r^2 (a and b), and d′ (c and d) for Dutch and north Indian controls respectively. LD plots were reconstructed using Immunochip genotype data.

replicated where most strong signals were restricted to MHC region (table 2). This signifies the possible contribution of the MHC driven common processes and pathways in pathogenesis of CD and these neuropsychiatric diseases.

Only three loci, one from nonHLA (*NFIA*) and two from extended HLA region (*NOTCH4-TSBP1-TNXB* and *HLA-DQA1*) were found transferable across ethnicities when locus transferability was tested using the north Indian immunochip CD association data. *HLA-DQA1* is directly implicated in CD pathogenesis by pathogenic HLA subunits molecules (HLA-DQ2 and HLA-DQ8). Limited replication of European CD associated signals in north Indians were previously explored and reported (Senapati *et al.* 2015). Both the loci from extended HLA locus are independent association signals, where significantly different LD was observed in the *NOTCH4-TSBP1-TNXB* locus when compared dense immunochip genotypes from a European (Dutch data) and Asian (north Indian data) populations (figure 1). NOTCH receptor 4 is a type I trans-membrane protein that regulates cell fate determination in neural, vascular, renal and hepatic development and in the adult brain predisposing neurons to apoptosis also by binding to ligands Jagged1, Jagged2 and Delta1 (Christopoulos *et al.* 2021). A British study on parent–offspring trios first reported an association of *NOTCH4* with schizophrenia, which was later confirmed by GWAS (Wei and Hemmings 2000; Gejman *et al.* 2010). Notably, Notch has been implicated in innate and adaptive immune responses as well. During ischemic stroke, Notch and its ligands (e.g. delta-like 1, DLL1) contribute to vascular inflammation, endothelial cell dysregulation, macrophage activation, and is involved in the interaction between immune cells and the brain (Vieceli Dalla Sega *et al.* 2019). Systemic inflammatory process including T cell activation is observed in schizophrenia patients as immune-pathogenic

mechanisms. Notch signalling mechanism in inflammatory cells would be dysregulated both as a surrogate marker for CNS tissues but also representing important cells that could alter several neural mechanisms (Hoseth *et al.* 2018). Ubiquitin-mediated proteolysis and immune system genes via Myd88 pathway activates TNF alpha and leads to inflammation (Hu and Sun 2016).

Another gene, *PBX2* (pre B cell leukaemia transcription factor 2) located in 6p21.3 region is a member of TALE/PBX homeobox family functions as a transcriptional activator that binds to TLX1 promoter. *PBX2* identified in schizophrenia is tightly linked to neighbouring gene *NOTCH4* (Avramopoulos 2018). TSBP1 is testis expressed basic protein 1 whose role has not been well established. Transferable SNPs from *NOTCH4* and *TSBP1* regulate expression of *C4A* and *HLA-DQB1* that are involved in innate and adaptive immune system, respectively. Notably, *HLA DQA1* is a well-known immune gene in CD that plays major role in antigen presentation by binding to *HLA DQB1* and elicits T cell responses and induces cytokine storm which leads to inflammation (Megiorni and Pizzuti 2012; Campagna *et al.* 2017).

This is the first report of shared genetics between CD and neuropsychiatric diseases, where considerable genetic overlaps were identified. Fine mapping and locus transferability analyses highlighted that each of the shared loci have more than one association signals corroborating with differential LD background. *NOTCH4-TSBP1-TNXB* locus can be further studied to evaluate their role in regulating gut–brain axis.

Acknowledgements

PB acknowledges DST-INSPIRE for fellowship (DST/INSPIRE-Fellowship/2019/IF190501). DST-FIST support (SR/FST/LS-I/2017/49-C) to the Department of Human Genetics and Molecular Medicine, Central University of Punjab, Bathinda is gratefully acknowledged. We acknowledge the contribution of Prof. Cisca Wijmenga, UMCG, Groningen for provided access to the Dutch Immunochip data. Early Career Research grant (ECR/2016/001660) to SS from Scientific Engineering Research Board (SERB), New Delhi, India. Women Scientist-A Fellowship (SR/WOS-A/LS-333/2018-G) to NS from Department of Science and Technology (DST), New Delhi, India.

Authors' contribution

SS conceptualized and designed the study. NS and PB performed the data analysis and statistical tests. AS and VM provided Indian celiac disease cohort clinical data. BKT provided the access to the Indian CD immunochip data and supervised the data analysis. SS, NS and PB wrote the manuscript. All the authors reviewed the final manuscript and approved for publication.

References

Avramopoulos D. 2018 Recent advances in the genetics of schizophrenia. *Complex Psychiatry* **4**, 35–51.

Barrett J. C., Fry B., Maller J. and Daly M. J. 2005 Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263–265.

Briani C., Zara G., Toffanin E., Ruggero S., Ferrarini A., De Lazzari F. *et al.* 2005 Neurological complications of celiac disease and autoimmune mechanisms: preliminary data of a prospective study in adult patients. *Ann. n. y. Acad. Sci.* **1051**, 148–155.

Campagna G., Pesce M., Tatangelo R., Rizzuto A., La Fratta I. and Grilli A. 2017 The progression of coeliac disease: its neurological and psychiatric implications. *Nutr. Res. Rev.* **30**, 25–35.

Christopoulos P. F., Gjølberg T. T., Krüger S., Haraldsen G., Andersen J. T. and Sundlisæter E. 2021 Targeting the notch signaling pathway in chronic inflammatory diseases. *Front. Immunol.* **12**, 668207.

Cooke W. T. and Smith W. T. 1966 Neurological disorders associated with adult celiac disease. *Brain* **89**, 683–722.

Durazzo M., Ferro A., Brascugli I., Mattivi S., Fagoonee S. and Pellicano R. 2022 Extra-intestinal manifestations of celiac disease: what should we know in 2022? *J. Clin. Med.* **11**, 258.

ENCODE Project Consortium 2012 An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74.

Gejman P. V., Sanders A. R. and Duan J. 2010 The role of genetics in the etiology of schizophrenia. *Psychiatr. Clin.* **33**, 35–66.

Hadjivassiliou M., Croall I. D., Zis P., Sarrigiannis P. G., Sanders D. S., Aeschlimann P. *et al.* 2019 Neurologic deficits in patients with newly diagnosed celiac disease are frequent and linked with autoimmunity to transglutaminase 6. *Clin. Gastroenterol. Hepatol.* **17**, 2678–2686.

Hoseth E. Z., Krull F., Dieset I., Mørch R. H., Hope S., Gardsjord E. S. *et al.* 2018 Attenuated Notch signaling in schizophrenia and bipolar disorder. *Sci. Rep.* **8**, 1–8.

Hu H. and Sun S. C. 2016 Ubiquitin signaling in immune responses. *Cell Res.* **26**, 457–483.

Jackson J. R., Eaton W. W., Cascella N. G., Fasano A. and Kelly D. L. 2012 Neurnologic and psychiatric manifestations of celiac disease and gluten sensitivity. *Psychiatr. q.* **83**, 91–102.

Kim-Hellmuth S., Aguet F., Oliva M., Muñoz-Aguirre M., Kasela S. *et al.* 2020 Cell type specific genetic regulation of gene expression across human tissues. *Science* **369**, eaaz8528, https://doi.org/10.1126/science.aaz8528.

Lebwohl B., Ludvigsson J. F. and Green P. H. 2015 Celiac disease and non-celiac gluten sensitivity. *BMJ.* **351**, h4347.

Losurdo G., Principi M., Iannone A., Amoruso A., Ierardi E., Di Leo A. *et al.* 2018 Extra-intestinal manifestations of non-celiac gluten sensitivity: An expanding paradigm. *World J. Gastroenterol.* **24**, 1521.

Ludvigsson J. F., Card T., Ciclitira P. J., Swift G. L., Nasr I., Sanders D. S. *et al.* 2015 Support for patients with celiac disease: A literature review. *United Eur. Gastroenterol. J.* **3**, 146–159.

Mata S., Renzi D., Pinto F. and Calabro A. 2006 Anti- tissue transglutaminase IgA antibodies inperipheral neuropathy and motor neuronopathy. *Acta. Neurol. Scand.* **114**, 54–58.

Megiorni F. and Pizzuti A. 2012 HLA-DQA1 and HLA-DQB1 in Celiac disease predisposition: practical implications of the HLA molecular typing. *J. Biomed. Sci.* **19**, 1–5.

Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M. A., Bender D. *et al.* 2007 PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575.

Senapati S., Gutierrez-Achury J., Sood A., Midha V., Szperl A., Romanos J. *et al.* 2015 Evaluation of European coeliac disease risk variants in a north Indian population. *Eur. J. Hum. Genet.* **23**, 530–535.

Senapati S., Sood A., Midha V., Sood N., Sharma S., Kumar L. *et al.* 2016 Shared andunique common genetic determinants between pediatric and adult celiac disease. *BMC Med. Genomics* **9**, 44.

Sharma N., Singh K. and Senapati S. 2021 Celiac disease poses significant risk in developing depression, anxiety, headache, epilepsy, panic disorder, dysthymia: A meta-analysis. *Indian J. Gastroenterol.* **40**, 1–10.

Trynka G., Wijmenga C. and van Heel D. A. 2010 A genetic perspective on coeliac disease. *Trends Mol. Med.* **16**, 537–550.

Vieceli Dalla Sega F., Fortini F., Aquila G., Campo G., Vaccarezza M. and Rizzo P. 2019 Notch signaling regulates immune responses in atherosclerosis. *Front. Immunol.* **10**, 1130.

Volta U., Giorgio R. D., Petrolini N., Stanghellini V., Barbara G., Granito A. *et al*. 2002 Clinical findings and anti-neuronal antibodies in coeliac disease with neurological disorders. *Scand. J. Gastroenterol.* **37**, 1276–1281.

Volta U., De Giorgio R., Granito A., Stanghellini V., Barbara G., Avoni P. *et al*. 2006 Anti-ganglioside antibodies in coeliac disease with neurological disorders. *Dig. Liver Dis.* **38**, 183–187.

Wei J. and Hemmings G. P. 2000 The NOTCH4 locus is associated with susceptibility to schizophrenia. *Nat. Genet.* **25**, 376–377.

Zhou G., Soufan O., Ewald J., Hancock R. E. W., Basu N. and Xia J. 2019 NetworkAnalyst 3.0: a visual analytics platform for comprehensive gene expression profiling and meta-analysis. *Nucleic Acids Res.* **47**, W234–W241.

Corresponding editor: Ashwin Dalal