Genes regulated by BCL11B during T# cell development are enriched for de novo mutations found in schizophrenia patients

Genes regulated by BCL11B during T cell development are enriched for *de novo* mutations found in schizophrenia patients

Laura Fahey[1,2], Gary Donohoe[1], Pilib Ó Broin[2]*, Derek W. Morris[1]*

1 Cognitive Genetics and Cognitive Therapy Group, Centre for Neuroimaging & Cognitive Genomics, School of Psychology and Discipline of Biochemistry, National University of Ireland Galway, Galway, Ireland.

2 School of Mathematics, Statistics and Applied Mathematics, National University of Ireland Galway, Galway, Ireland.

* Authors jointly directed this work.

Corresponding author:

Dr Derek Morris, Room 106, Discipline of Biochemistry, National University of Ireland Galway,

University Road, Galway, Ireland.

Tel: + 353 91 494439; Email: derek.morris@nuigalway.ie

Grant numbers:

Irish Research Council grant (GOIPG/2018/1501)

Running Title: BCL11B and immune function in schizophrenia

Abstract

While abnormal neurodevelopment contributes to schizophrenia (SCZ) risk, there is also evidence to support a role for immune dysfunction in SCZ. *BCL11B,* associated with SCZ in GWAS, is a transcription factor that regulates the differentiation and development of cells in the central nervous and immune systems. Here, we use functional genomics data from studies of *BCL11B* to investigate the contribution of neuronal and immune processes to SCZ pathophysiology. We identified the gene targets of BCL11B in brain striatal cells (n=223 genes), double negative 4 (DN4) developing T cells (n=114 genes) and double positive (DP) developing T cells (n=518 genes) using an integrated analysis of RNA-seq and ChIP-seq data. No gene-set was enriched for genes containing common variants associated with SCZ but the DP gene-set was enriched for genes containing missense *de novo* mutations (DNMs; P=0.001) using data from 3,447 SCZ trios. Post-hoc analysis revealed the enrichment to be stronger for DP genes negatively-regulated by BCL11B. Biological processes enriched for genes negatively-regulated by BCL11B in DP gene-set included immune system development and cytokine signaling. These analyses, leveraging a GWAS-identified SCZ risk gene and data on gene expression and transcription factor binding, indicate that DNMs in immune pathways contribute to SCZ risk.

Keywords: schizophrenia, de novo mutation, RNA-seq, immune system, ChIP-seq

Introduction

Schizophrenia (SCZ) is a chronic, severe mental disorder affecting ~1% of the population worldwide. It is ranked within the top 10 disabling conditions for young adults worldwide (Świtaj et al., 2012), affecting general health, ability to work and formation of personal relationships. A core feature of SCZ is cognitive dysfunction in the form of impairments in memory, attention and IQ. This decline in cognitive performance occurs years prior to onset of illness as the brain is still developing (Kahn & Keefe, 2013) and its severity is a strong predictor of functional outcome (Leeson et al., 2009).

SCZ has a complex aetiology influenced by both genetic and environmental factors. The heritability of SCZ is estimated from twin studies to be 80% (Hilker et al., 2018). The genetic architecture of SCZ is highly polygenic. GWAS has identified 179 independent SCZ associated single nucleotide polymorphisms (SNPs) across the genome (Pardiñas et al., 2018). Analysis of rare structural and sequence variants, including *de novo* events, indicates a role for rare mutations in SCZ (Kirov et al., 2012; Fromer, et al. 2014; Singh et al., 2016). The “neurodevelopmental hypothesis” is the most widely accepted explanation of SCZ pathology. Here, genetic and environmental insults are hypothesized to disturb prenatal neurodevelopment, making a person more vulnerable to developing SCZ upon secondary environmental insults in adolescence or young adulthood (Fatemi & Folsom, 2009).

Immune defects have also been hypothesized to contribute to SCZ risk. Interest in an immune hypothesis of SCZ has stemmed in part from the imbalance of pro- and anti-inflammatory cytokines repeatedly observed in SCZ (Bergink et al., 2014). It is hypothesized that an interaction between genetic and environmental influences leads to pathological microglial activation, resulting in excessive synaptic pruning and loss of cortical grey matter, particularly in stress-sensitive regions such as the prefrontal cortex and hippocampus, which ultimately results in cognitive dysfunction and increased illness risk (Howes & McCutcheon, 2017). Imbalances in the monoamine pathway and autoimmune processes relevant to synaptic refinement have also been hypothesized to be part of SCZ pathophysiology. In SCZ, inflammatory markers (including C-reactive protein and various interleukins, e.g. IL-6) are also associated with cognitive deficits (Misiak et al., 2018). Strong genetic evidence now exists for the role of immune-related processes in increasing SCZ risk where the most significant genetic association maps to the major histocompatibility complex (MHC) region on chromosome 6 (Sekar et al., 2016). This has identified one potential mechanism by which the MHC region may increase risk, involving a locus

containing the complement component 4 (C4) gene, an isotope of which (C4A) was demonstrated to have a role in cortical development of the visual system in animal models. Almost 40 non-MHC genes with known roles in immune function have been associated with SCZ risk (reviewed by Pouget, 2018). Associations with altered expression of immune gene networks (Birnbaum et al., 2018) have also been reported, as well as altered expression of immune proteins (Föcking et al., 2019) and observed dysregulation of immune complement proteins in SCZ (Benros and Mortensen, 2019).

B-cell chronic lymphocytic leukemia/lymphoma 11B (BCL11B), also named Coup-TF interacting protein 2 (CTIP2), is a zinc finger protein transcription factor. It was identified as a SCZ risk gene through genome-wide association studies (GWAS; Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014; Whitton et al., 2016). The association signal is located over the *BCL11B* gene and there are no other protein-coding genes in the region (Supplementary Figure 1). The most associated or index SNP and those SNPs in high linkage disequilibrium (LD) with it are located within the introns of the gene. Review of quantitative trait loci (QTL) datasets for gene expression, chromatin, isoform percentage, transcript expression, cell fraction or methylation does not identify how the SCZ risk SNP affects the function of the gene (Wu et al., 2018; Wang et al., 2018). BCL11B is required for neurogenesis both during postnatal development (Simon et al., 2012) and during adulthood (Simon et al., 2016). Conditional forebrain-specific *Bcl11b* knockout (KO) mice suffered impaired spatial learning and memory (Simon et al., 2012). Induction of the BCL11B KO selectively in adulthood also resulted in impaired memory, but to a lesser extent (Simon et al., 2016).

BCL11B is essential in the immune system, where it functions in T cell differentiation in the thymus (Fang et al., 2018). T cell differentiation is divided into discrete stages based on expression of certain markers. The first stage is defined by the absence of markers cluster of differentiation (CD) 4 and CD8, and therefore referred to as double negative (DN). This stage can be further subdivided into DN1 to DN4 (Famili et al., 2017). DN cells further differentiate into $CD4^+CD8^+$ double positive (DP) cells and finally into either $CD4^+$ helper or $CD8^+$ cytotoxic single positive (SP) cells. Upon activation, SP cells differentiate into T effector subsets characterized by different cytokine profiles. $CD4^+$ subsets include T helper (Th) 1, Th2, Th9, Th17, Th22, regulatory T cells (Tregs), and follicular helper T cells (Tfh), while $CD8^+$ cell subsets include stem cell memory T cells, T central memory cells, T effector memory cells and T effector cells (Golubovskaya and Wu, 2016).

BCL11B is first expressed in DN2 cells, progressively increasing in expression during the successive progenitor stages and is maintained in all mature T cell subsets (Avram & Califano, 2014). BCL11B acts to lock cells into the T cell fate through induction of a dramatic reorganization of chromatin at the DN2 to DN3 checkpoint, and to a lesser extent at the DN4 to DP checkpoint (Hu et al., 2018). This is necessary to promote T cell lineage commitment and suppress alternative lineages (natural killer (NK), myeloid and B lymphocytes; Ha et al., 2017). Loss of BCL11B in murine thymocyte progenitors leads to developmental arrest and acquisition of natural killer (NK) cell features (Li et al., 2010). Loss of BCL11B in mature T cell subsets also leads to various defects including upregulation of type2 cytokines in Th2 cells (Fang et al., 2018), derepression of T helper 2 (Th2) genes in Th17 cells (Califano et al., 2014) and reduced functional immune suppression molecules in Treg cells (Drashansky et al., 2019).

A role for T cells in the immune-mediated etiology of SCZ was first proposed 40 years ago (Nyland et al., 1980) and many studies since have observed abnormal T-cell function or densities in SCZ. These include a Th2 shift, where over-production of Th2 cells in comparison to Th1 cells has been observed in SCZ (Avgustin et al., 2005; Chiang et al., 2013). Abnormal distribution of TCR variable genes have been reported in SCZ (Li et al., 2018). Higher levels of Tregs were found in the blood of stable medicated SCZ patients compared to controls and this was associated with decreased negative symptoms and increased cognitive performance (Kelly et al., 2018). Tregs act to maintain immune homeostasis and suppress aberrant immune responses. The higher levels Tregs had a positive effect, being associated with decreased negative symptoms and increased cognitive performance (Kelly et al., 2018)). The authors of the study suggest that the increased levels of Tregs might be a response to elevated inflammation in SCZ or may possibly be caused by medication (Kelly et al., 2018). Another study observed increased densities of Th17 cells in peripheral blood mononuclear cells derived from non-medicated SCZ patients compared to healthy controls (Varun et al., 2019).

A recent study identified 13 patients with *de novo* mutations (DNMs) in *BCL11B*, all of whom were affected by developmental delay, speech impairment and intellectual disability (Lessel et al., 2018). Immune phenotyping of these patients revealed alterations in the T cell compartment and lack of peripheral type 2 innate lymphoid cells (Ha et al., 2017; Lessel et al., 2018). Here, we use functional genomics data from studies of BCL11B to delineate sets of genes influenced by BCL11B during different neuronal and immune processes. We then test if these gene-sets are

enriched for genes that make a contribution to SCZ risk in order to determine if these BCL11B-influenced neuronal and/or immune processes represent potential molecular mechanisms for SCZ. We also test if these gene-sets are enriched for genes associated with cognitive traits given the importance of cognitive dysfunction to SCZ, the role of BCL11B in learning and memory (Simon et al., 2012), and evidence that the immune system influences cognition in psychiatric disorders (Khandaker and Dantzer, 2016).

Methods

Generation of gene-sets

A literature search was performed to identify studies that performed experiments enabling the genome-wide identification of direct targets of BCL11B in brain or immune cell types. We identified three studies that either provided the list of direct BCL11B target genes or provided access to the raw experimental data enabling us to perform an integrated differential expression and binding analysis.

The first gene-set, called Striatal, was obtained from a study that investigated the targets of BCL11B in mouse brain striatal cells (Tang et al., 2011). This study performed an integrated differential expression analysis of cells over-expressing BCL11B versus controls combined with a BCL11B binding analysis using ChIP-seq to identify 248 direct targets of BCL11B in striatal cells. This list of genes was supplied to us by the authors of that study. MGI symbols were converted to Entrez human IDs (for use with MAGMA) and HGNC symbols (for use with denovolyzeR) using the Bioconductor package biomaRt (https://bioconductor.org/packages/release/bioc/html/biomaRt.html) (Smedley et al., 2015), giving us the final Striatal gene-set of 223 genes (Supplementary Table 1).

The second gene-set, called DN4, was generated using data from a study that performed loss of function studies and DNA binding studies to delineate the function of BCL11B during the initial stages of human T cell differentiation (Ha et al., 2017). Direct BCLL1B target genes were not supplied with the manuscript. Therefore, we used raw RNA-seq and ChIP-seq data (GEO Accession: GSE84678) from this study to perform an integrated RNA-seq and ChIP-seq analysis (described below) to generate this gene-set. Data used for the differential gene expression analysis was from a whole transcriptome RNA-seq analysis of BCL11B KO and wild-type (WT) *in vitro* human CD34+CD7+CD1a+ cells (markers of DN4 thymocytes). Data used for the BCL11B binding analysis was from a whole genome BCL11B ChIP-seq of human CD34+ cells (marker of DN2, DN3, and DN4 thymocytes). This DN4 gene-set contains 114 genes (Supplementary Table 1).

The third gene-set, called DP, was generated using data from a study that performed genome-wide RNA-seq expression profiling of BCL11B KO and WT mouse DP cells and genome-wide BCL11B ChIP-seq in mouse total thymocytes (>80% DP cells) (Hu et al., 2018). BCL11B target

genes were not identified in this manuscript. Therefore, we used raw data from these studies (GEO Accession: GSE79875) to perform an integrated differential expression and BCL11B binding analysis to identify direct targets of BCL11B in mouse DP cells (described below). BioMart was again used for conversion of MGI symbols to Entrez human IDs and HGNC symbols. Major histocompatibility complex (*MHC*) genes do not have direct human orthologues and were therefore removed. The final DP gene-set contains 518 genes (Supplementary Table 1).

Integrated RNA-seq and ChIP-seq analysis for DN4 and DP gene-sets

RNA-seq analysis

Data for two replicate samples were available for the DN4 gene-set (Ha et al., 2017) and data for three replicate samples were available for the DP gene-set (Hu et al., 2018). FASTQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc) (Andrews 2010) was used to assess the quality of reads. Low quality reads (defined as FASTQC per base quality score < 28) and adaptors were trimmed using Trimmomatic (Bolger et al., 2014). HISAT2 (https://ccb.jhu.edu/software/hisat2/index.shtml) (Kim et al., 2015) was used to align reads to the human genome (hg19; for the DN4 gene-set) or mouse genome (mm10; for the DP gene-set). STRINGTIE (https://ccb.jhu.edu/software/stringtie/) (Pertea et al., 2015) was used to count reads that aligned back to each gene and the Bioconductor package EdgeR (https://bioconductor.org/packages/release/bioc/html/edgeR.html) et al., 2010) was used for statistical analysis.

ChIP-seq analysis

FASTQC and Trimmomatic were used for quality control. Bowtie2 (http://bowtie-bio.sourceforge.net/bowtie2/index.shtml) (Langmead and Salzberg, 2012) was used to align reads to the genome. Post-processing was performed using SAMTOOLS. Two biological replicates for treatment and control were available for each study (Ha et al., 2017; Hu et al., 2018). Model-based Analysis of ChIP-Seq data 2 (MACS2) (https://github.com/taoliu/MACS) (Zhang et al., 2008) was used to call peaks for each replicate. The more liberal P value cutoff of <0.001 was used so that peaks could then be assessed for reproducibility across replicates. The irreproducibility discovery rate (IDR) was calculated for each set of peaks using the IDR tool https://github.com/nboley/idr (Li et al. 2011). Peaks with IDR < 0.05 were retained.

Identification of target genes

Binding and Expression Target Analysis (BETA) (http://cistrome.org/BETA/) (Wang et al., 2013) was used to infer direct BCL11B target genes. Differential expression data (EdgeR statistical output) and binding data (narrowPeak output file from IDR) were used as input to BETA. A rank product (RP) was calculated for each gene based on 1) differential expression upon BCL11B KO (only genes with a false discovery rate (FDR) $P < 0.1$ were considered) and 2) binding potential based on the number of BCL11B binding sites within 100kb and the distance of each binding site from the transcription start site. Genes with a $RP < 0.005$ were considered to be direct targets of BCL11B.

Gene-set analysis

A gene-set analysis (GSA) is a statistical method for simultaneously analyzing multiple common genetic markers in order to determine their joint effect. We performed GSA using MAGMA (de Leeuw et al., 2015) (http://ctg.cncr.nl/software/magma) and summary statistics from published GWAS on SCZ ((Pardiñas et al., 2018), 40,675 cases and 64,643 controls), intelligence (IQ (Savage et al., 2018), 269,867 individuals) and educational attainment (EA (Lee et al., 2018), ~1.1 million individuals). An analysis involved three steps. First, in the annotation step we mapped SNPs with available GWAS results to genes (GRCh37/hg19 start-stop coordinates +/-20kb). Second, in the gene analysis step we computed gene P values for each GWAS dataset. This gene analysis is based on a multiple linear principal components regression model that accounts for LD between SNPs. The European panel of the 1000 Genomes data was used as a reference panel for LD. Third, a competitive GSA based on the gene P values, also using a regression structure, was used to test if the genes in a gene-set were more strongly associated with either phenotype than other genes in the genome. MAGMA was chosen because it corrects for LD, gene size and gene density (potential confounders) and has significantly more power than other GSA tools (de Leeuw et al., 2016).

Analysis of *de novo* mutations

Enrichment of DNMs in our gene-sets was tested using the R package, denovolyzeR (http://denovolyzer.org/) (Ware et al., 2015). We tested for enrichment of synonymous, missense and loss of function (included nonsense, frameshift and splice) DNMs. The expected number of each mutational class of DNM for each gene is predicted using the denovolyzeR internal mutation probability table. Enrichment of DNMs in individual genes is investigated using a Poisson test that

compares the observed number of DNMs to the expected. Enrichment of DNMs in gene-sets is tested using a two-sample Poisson rate ratio test, using the number of observed to expected DNMs in genes outside of the gene-set as a background model.

The major source of SCZ trio data was from (Howrigan et al., 2020), who combined DNMs from a new sample of Taiwanese SCZ trios with DNMs from previously published SCZ trio studies (Girard, et al. 2011; Xu, et al. 2011; Xu, et al. 2012; Gulsuner, et al. 2013; Fromer, et al. 2014; Guipponi, et al. 2014; McCarthy, et al. 2014) to give a total of 2,772 trios. We combined these data with SCZ DNMs from (Rees et al., 2020; number of trios (N) = 613), Wang et al., 2015; N = 45) and Ambalavanan et al., 2016; N = 17). In total, we had data on 3,447 SCZ trios. There are no reported DNMs within *BCL11B* in SCZ patients. Control data was also obtained from (Howrigan et al., 2020), they gathered previously published data on 2,216 unaffected sibling and control trios (Iossifovet al., 2012; Iossifov et al., 2014; DeRubeis et al., 2014). DNM annotations were re-coded to match the variant classes in the denovolyzeR built-in probability table (syn (synonymous), mis (missense), non (nonsense), frameshift and splice (canonical splice site)). DNM annotations that did not fit into any of these classes were removed (in-frame, start-loss, and stop-loss).

Overrepresentation analysis

ConsensusPathDB is a meta-database currently integrating data from 32 resources on protein-protein, genetic, metabolic, signaling, gene regulatory and drug-target interactions et al., 2016). The ConsensusPathDB web interface (http://cpdb.molgen.mpg.de/) offers a tool for gene-set over-representation analysis, where user defined gene-sets are tested for overrepresentation of genes from sets divided into four categories 1) network neighborhood-based sets, 2) pathway-based sets, 3) Gene Ontology (GO)-based sets and 4) protein complex-based gene-sets. A hypergeometric test is used to calculate a P value for each set within these four categories based on the number of genes in that set and the number of genes in the user defined gene-sets. FDR adjusted P value of 0.05 was used as cut-off.

Results

We had three BCL11B gene-sets: (i) The Striatal gene-set (n=223 genes) contained genes that are regulated by BCL11B in the brain striatum. (ii) The DN4 gene-set (n=114 genes) contained genes regulated by BCL11B during the DN4 stage of T cell development. (iii) The DP (n=518 genes) gene-set contained genes regulated by BCL11B during the later DP stage of T cell development. There were no genes common to all three gene-sets but there was overlap between DN4 and DP (8 genes), Striatal and DN4 (2 genes) and Striatal and DP (4 genes; Supplementary Figure 2). The little overlap between the DN4 and DP gene-sets may be due to many of BCL11B's target genes being differentiation stage specific (Ha et al. 2017).

Gene-set analysis was performed on these three gene-sets to test for enrichment of common genetic variation associated with SCZ, IQ or EA. The 9 independent tests performed were corrected for multiple testing by calculating FDR adjusted P values (Q values) (Benjamini & Hochberg, 1995). None of our gene-sets were significantly enriched for genes associated with SCZ, IQ or EA (Table 1).

Our gene-sets were next investigated for the enrichment of rare genetic variants reported in SCZ patients in the form of DNMs. We tested each gene-set for enrichment of synonymous, missense and loss of function SCZ DNMs. The DP gene-set was significantly enriched for SCZ missense DNMs (two-sample rate ratio = 1.43, P = 0.001, Q = 0.013; Table 2). In comparison, this gene-set was not significantly enriched for SCZ synonymous DNMs (two-sample rate ratio = 1.14, P = 0.47, Q = 1; Table 2) or for missense DNMs in control trios (two-sample rate ratio = 0.92, P = 0.69, Q = 1; Supplementary Table 2).

To investigate if the significant enrichment of SCZ missense DNMs in the DP gene-set was largely from genes negatively or positively regulated by BCL11B, we split this gene-set into genes up-regulated (DP_upreg; n=256 genes) or down-regulated (DP_downreg; n=262 genes) upon BCL11B KO. Testing of these two gene subsets for enrichment of missense DNMs revealed the signal to be stronger for genes negatively regulated by BCL11B. DP_upreg was significantly enriched for missense DNMs (two-sample rate ratio = 1.48, P = 0.009), whereas DP_downreg was not significantly enriched for missense DNMs (two-sample rate ratio = 1.37, P = 0.053; Table 2). The DP_upreg gene-set was not significantly enriched for missense DNMs in control trios (two-sample rate ratio = 1.12, P = 0.578; Supplementary Table 2).

The DP gene-set contained 78 genes with a missense or loss of function SCZ DNM and 15 of these contained more than one such DNM. Sixteen genes in the DP gene-set were individually nominally significant for an excess of SCZ DNMs (Supplementary Table 3). The most significant gene is the NLR family CARD domain containing 5 (*NLRC5*) gene, which contained four protein altering DNMs but this gene result did not survive correction for all 507 genes in the DP gene-set tested (P = 0.0009, Q = 0.45).

GO and pathway analysis revealed the DP_upreg gene subset to be enriched for genes with immune function. The most significant GO term for DP_upreg gene subset was immune system development (Q = 1.68e-12) and a number of the other significant GO terms represented daughter terms of this, including leukocyte activation and lymphocyte activation. The top 15 GO terms were visualized as a GO ancestral tree, which enabled the identification of clusters of enriched GO terms involving immune cell signaling, activation and differentiation (Supplementary Figure 3). The most significantly enriched pathway was IL-6-mediated signaling events (Q = 1.79e-05) and other cytokine signaling pathways featured among the top pathways, including transforming growth factor beta (TGFβ), IL-2 and tumour necrosis factor (TNF; (Supplementary Tables 4 and 5)).

Discussion

We have found that target genes of BCL11B, primarily those genes that are negatively regulated during the DP stage of T cell development, are enriched for genes harbouring missense DNMs that have been reported in SCZ patients. Negatively regulated means that these genes ordinarily have their expression lowered or blocked by BCL11B binding during the DP stage of T cell development but these genes show an increase in expression when BCL11B is knocked out as in the model that we have used, i.e. these genes are upregulated upon BCL11B KO. During the DP stage, thymocytes undergo T cell receptor-α (TCRα) gene locus rearrangement and expression, and a mature TCR expressing the protein chains β and α is presented at the cell surface. DP cells expressing a functional TCRαβ will receive a positive-selection signal and differentiate to the $CD4^+$ or $CD8^+$ SP stage (Wang and Bosselut, 2009). BCL11B-deficient DP thymocytes rearrange and express TCRα, however, TCR signaling is impaired and positive selection into SP thymocytes is blocked (Albu et al., 2007). BCL11B-deficient DP thymocytes also undergo increased spontaneous apoptosis, indicating that BCL11B is required for DP thymocyte survival (Albu et al., 2007). BCL11B acts as a master transcriptional regulator in DP thymocytes, promoting expression of genes required for DP thymocyte survival and progression of T cell development, while repressing genes belonging to progenitor lineages and mature T cells (Kastner et al., 2010).

Aligning with BCL11B's repressive role, our functional enrichment analysis revealed that genes negatively regulated by BCL11B are mostly enriched in biological processes and pathways for immune system development and cytokine signaling. These cytokine signaling pathways include IL-6, TNF, TGFβ and IL-2, which have all been reported to be dysregulated in SCZ. Cytokines are a family of proteins that function as chemical messengers in cell-cell communication. There are two types of cytokines, pro-inflammatory cytokines, which signal inflammation upon recognition of infection and anti-inflammatory cytokines that downregulate the immune response. Pro-inflammatory cytokines, IL-6 and TNF, are increased in the blood of SCZ patients and higher levels are associated with worsening symptoms (Frydecka et al., 2015), et al., 2017). Whereas IL-2, which is primarily an anti-inflammatory cytokine (Boerrigter et al., 2017), is reduced in blood plasma SCZ patients in comparison to controls (Asevedo et al., 2014). This is representative of the low grade inflammation observed in some SCZ patients (Upthegrove and Khandaker, 2019). The Avon Longitudinal Study of Parents and Children found that children with higher IL-6 levels were twice as likely to develop psychosis as young adults (Khandaker et al., 2014), indicating that

inflammation is a risk factor for psychosis and not just a symptom of the disorder (Upthegrove and Khandaker, 2019).

Our findings that genes negatively regulated by BCL11B during the DP stage of T cell development are enriched for genes harbouring missense DNMs in SCZ patients suggests that reduced expression of BCL11B could result in increased expression of genes where alterations to the coding sequence are increasing SCZ risk via the immune system development and cytokine signaling pathways highlighted above. An important issue that remains is the effect of the SCZ risk SNP at *BCL11B* on gene function. The lead risk SNP at *BCL11B* in the SCZ GWAS, rs35604463, sits in intron 2 of the gene within a gene regulatory element mapped by H3K27ac ChIP-seq in human stem cell derived glutamatergic neurons, 35 day differentiation (Matthew Hill, personal communication). However, despite multiple expression QTL databases now being available, including data on both adult (Wang et al., 2018; GTEx Consortium, 2013) and fetal brain (O'Brien et al., 2018) gene expression, none yet provide evidence that the SCZ risk SNP affects BCL11B function via altered gene expression. The PsychENCODE study does report that the expression of *BCL11B* is reduced in the post mortem brains of SCZ cases compared to controls, although this result did not survive experiment-wide correction (Gandal et al., 2018). This requires further investigation in different cell and tissue types and at different stages of development to understand how BCL11B gene function is affected by the risk SNP and if the downstream effect is to express genes carrying SCZ risk mutations that further disrupt normal immune function.

BCL11B could also offer a possible explanation for the aberrant T cell mediated immunity in SCZ. BCL11B has been shown to suppress Th2 differentiation (Fang et al., 2018), therefore lower levels of BCL11B could result in the observed SCZ Th2 shift. The increase of Treg cells in stable medicated SCZ could also possibly be explained by aberrant BCL11B expression as BCL11B is essential for promoting the expression of Treg genes and inhibiting innate lineages in Treg cells (Drashansky et al., 2019). BCL11B is a negative regulator of Th17 cytokines (Chen, Han, Chen, Wu, & Zhang, 2018), therefore reduced expression of BCL11B could be an explanation for the increased densities of Th17 cells in SCZ.

The strengths of the study include use of the largest available datasets from GWAS for analysis of common variants associated with SCZ and cognition, the study of both brain and immune cell

types to investigate BCL11B's different functions, and the strategy of an integrated analysis of RNA-seq and ChIP-seq data to infer direct target genes of BCL11B. There were limitations also. For the DN4 gene-set, differential expression analysis was performed on CD34+CD7+CD1a+ cells (markers of DN4 thymocytes *in vitro*) and BCL11B binding analysis on BCL11B ChIP-seq of human CD34+ cells (marker of DN2-4 thymocytes). It would have been preferential to perform the integrative analysis on the same cell types identified using the same markers. Although the SCZ trios dataset was the largest available, our analysis of *NLRC5* did not survive multiple test correction here, possibly due to the SCZ trios dataset being underpowered to identify single gene associations. In addition, for the DNMs in this gene (one LoF and three missense), the putative functionality of these mutations is based on *in silico* analysis and the true effect of DNMs on protein function is yet to be determined. NLRC5 has multiple proposed roles in the immune system. It is a transcriptional activator of MHC class I genes. MHC molecules present peptide antigens to T cells, the class I type being responsible for presenting antigens to $CD8^+$ cytotoxic T lymphocytes. NLRC5 was found to increase immune response to tumours by enhancing processing and presentation of tumor antigens to $CD8^+$ cytotoxic T lymphocytes (Rodriguez et al., 2016). NLRC5 has also been found to negatively regulate innate immunity through suppression of NF-kappaB and type I interferon signaling pathways (Cui et al., 2010). Larger datasets on rare variants will be required to determine if *NLRC5* represents a risk gene for SCZ.

BCL11B is an example of a SCZ risk gene with dual function in both the central nervous system and the immune system. Investigations of other such risk genes with dual functions, for example C4 (Sekar et al., 2016), have provided new insights into SCZ biology. In summary, our study suggests that genes negatively regulated by BCL11B during the DP stage of T cell development are contributing to the aetiology of SCZ. These analyses, leveraging data on gene expression and transcription factor binding in immune cell models, indicate that DNMs in genes that encode immune system development and cytokine signalling contribute to SCZ risk.

Acknowledgements

We wish to thank Matthew Hill (Cardiff University) for information on mapping of gene regulatory elements at BCL11B.

Conflict of Interest

The authors have no conflict of interest to declare.

Author Contributions

Genes regulated by BCL11B during T cell development are enriched for de novo mutations found in schizophrenia patients. Laura Fahey, Gary Donohoe, Pilib Ó Broin, Derek W. Morris.

D.W.M, P.Ó.B and G.D. conceived the project. L.F. analyzed the data with support from P.Ó.B. L.F. and D.W.M. wrote the draft manuscript, with subsequent revisions based on involvement from P.Ó.B and G.D.

Bibliography

Albu, D. I., Feng, D., Bhattacharya, D., Jenkins, N. A., Copeland, N. G., Liu, P., & Avram, D. (2007). BCL11B is required for positive selection and survival of double-positive thymocytes. *The Journal of Experimental Medicine*, *204*(12), 3003–3015. doi:10.1084/jem.20070863

Asevedo, E., Rizzo, L. B., Gadelha, A., Mansur, R. B., Ota, V. K., Berberian, A. A., … Brietzke, E. (2014). Peripheral interleukin-2 level is associated with negative symptoms and cognitive performance in schizophrenia. *Physiology & Behavior*, *129*, 194–198. doi:10.1016/j.physbeh.2014.02.032

Avgustin, B., Wraber, B., & Tavcar, R. (2005). Increased Th1 and Th2 immune reactivity with relative Th2 dominance in patients with acute exacerbation of schizophrenia. *Croatian Medical Journal*, *46*(2), 268–274.

Avram, D., & Califano, D. (2014). The multifaceted roles of Bcl11b in thymic and peripheral T cells: impact on immune diseases. *Journal of Immunology*, *193*(5), 2059–2065. doi:10.4049/jimmunol.1400930

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300. doi:10.1111/j.2517-6161.1995.tb02031.x

Benros, M. E., & Mortensen, P. B. (2019). Role of Infection, Autoimmunity, Atopic Disorders, and the Immune System in Schizophrenia: Evidence from Epidemiological and Genetic Studies. *Current Topics in Behavioral Neurosciences*. doi:10.1007/7854_2019_93

Bergink, V., Gibney, S. M., & Drexhage, H. A. (2014). Autoimmunity, inflammation, and psychosis: a search for peripheral markers. *Biological Psychiatry*, *75*(4), 324–331. doi:10.1016/j.biopsych.2013.09.037

Birnbaum, R., Jaffe, A. E., Chen, Q., Shin, J. H., BrainSeq Consortium, Kleinman, J. E., … Weinberger, D. R. (2018). Investigating the neuroimmunogenic architecture of schizophrenia. *Molecular Psychiatry*, *23*(5), 1251–1260. doi:10.1038/mp.2017.89

Boerrigter, D., Weickert, T. W., Lenroot, R., O'Donnell, M., Galletly, C., Liu, D., … Weickert, C. S. (2017). Using blood cytokine measures to define high inflammatory biotype of schizophrenia and schizoaffective disorder. *Journal of Neuroinflammation*, *14*(1), 188. doi:10.1186/s12974-017-0962-y

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. doi:10.1093/bioinformatics/btu170

Califano, D., Sweeney, K. J., Le, H., VanValkenburgh, J., Yager, E., O'Connor, W., … Avram, D. (2014). Diverting T helper cell trafficking through increased plasticity attenuates autoimmune encephalomyelitis. *The Journal of Clinical Investigation*, *124*(1), 174–187. doi:10.1172/JCI70103

Chen, S., Han, Y., Chen, H., Wu, J., & Zhang, M. (2018). Bcl11b Regulates IL-17 Through the TGF-β/Smad Pathway in HDM-Induced Asthma. *Allergy, Asthma & Immunology Research*, *10*(5), 543–554. doi:10.4168/aair.2018.10.5.543

Chiang, S. S.-W., Riedel, M., Schwarz, M., & Mueller, N. (2013). Is T-helper type 2 shift schizophrenia-specific? Primary results from a comparison of related psychiatric disorders and healthy controls. *Psychiatry and Clinical Neurosciences*, *67*(4), 228–236. doi:10.1111/pcn.12040

Cui, J., Zhu, L., Xia, X., Wang, H. Y., Legras, X., Hong, J., … Wang, R.-F. (2010). NLRC5 negatively regulates the NF-kappaB and type I interferon signaling pathways. *Cell*, *141*(3), 483–496. doi:10.1016/j.cell.2010.03.040

de Leeuw, C. A., Mooij, J. M., Heskes, T., & Posthuma, D. (2015). MAGMA: generalized gene-set analysis of GWAS data. *PLoS Computational Biology*, *11*(4), e1004219. doi:10.1371/journal.pcbi.1004219

de Leeuw, C. A., Neale, B. M., Heskes, T., & Posthuma, D. (2016). The statistical properties of gene-set analysis. *Nature Reviews. Genetics*, *17*(6), 353–364. doi:10.1038/nrg.2016.29

Drashansky, T. T., Helm, E., Huo, Z., Curkovic, N., Kumar, P., Luo, X., … Avram, D. (2019). Bcl11b prevents fatal autoimmunity by promoting Treg cell program and constraining innate lineages in Treg cells. *Science Advances*, *5*(8), eaaw0480. doi:10.1126/sciadv.aaw0480

Famili, F., Wiekmeijer, A.-S., & Staal, F. J. (2017). The development of T cells from stem cells in mice and humans. *Future Science OA*, *3*(3), FSO186. doi:10.4155/fsoa-2016-0095

Fang, D., Cui, K., Hu, G., Gurram, R. K., Zhong, C., Oler, A. J., … Zhu, J. (2018). Bcl11b, a novel GATA3-interacting protein, suppresses Th1 while limiting Th2 cell differentiation. *The Journal of Experimental Medicine*, *215*(5), 1449–1462. doi:10.1084/jem.20171127

Fatemi, S. H., & Folsom, T. D. (2009). The neurodevelopmental hypothesis of schizophrenia, revisited.

Schizophrenia Bulletin, *35*(3), 528–548. doi:10.1093/schbul/sbn187

Föcking, M., Sabherwal, S., Cates, H. M., Scaife, C., Dicker, P., Hryniewiecka, M., … Cotter, D. R. (2019). Complement pathway changes at age 12 are associated with psychotic experiences at age 18 in a longitudinal population-based study: evidence for a role of stress. *Molecular Psychiatry*. doi:10.1038/s41380-018-0306-z

Frydecka, D., Misiak, B., Pawlak-Adamska, E., Karabon, L., Tomkiewicz, A., Sedlaczek, P., … Beszłej, J. A. (2015). Interleukin-6: the missing element of the neurocognitive deterioration in schizophrenia? The focus on genetic underpinnings, cognitive impairment and clinical manifestation. *European Archives of Psychiatry and Clinical Neuroscience*, *265*(6), 449–459. doi:10.1007/s00406-014-0533-5

Gandal, M. J., Zhang, P., Hadjimichael, E., Walker, R. L., Chen, C., Liu, S., … Geschwind, D. H. (2018). Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science*, *362*(6420). doi:10.1126/science.aat8127

Golubovskaya, V., & Wu, L. (2016). Different Subsets of T Cells, Memory, Effector Functions, and CAR-T Immunotherapy. *Cancers*, *8*(3). doi:10.3390/cancers8030036

GTEx Consortium. (2013). The Genotype-Tissue Expression (GTEx) project. *Nature Genetics*, *45*(6), 580–585. doi:10.1038/ng.2653

Ha, V. L., Luong, A., Li, F., Casero, D., Malvar, J., Kim, Y. M., … Parekh, C. (2017). The T-ALL related gene BCL11B regulates the initial stages of human T-cell differentiation. *Leukemia*, *31*(11), 2503–2514. doi:10.1038/leu.2017.70

Herwig, R., Hardt, C., Lienhard, M., & Kamburov, A. (2016). Analyzing and interpreting genome data at the network level with ConsensusPathDB. *Nature Protocols*, *11*(10), 1889–1907. doi:10.1038/nprot.2016.117

Hilker, R., Helenius, D., Fagerlund, B., Skytthe, A., Christensen, K., Werge, T. M., … Glenthøj, B. (2018). Heritability of schizophrenia and schizophrenia spectrum based on the nationwide danish twin register. *Biological Psychiatry*, *83*(6), 492–498. doi:10.1016/j.biopsych.2017.08.017

Howes, O. D., & McCutcheon, R. (2017). Inflammation and the neural diathesis-stress hypothesis of schizophrenia: a reconceptualization. *Translational Psychiatry*, *7*(2), e1024. doi:10.1038/tp.2016.278

Howrigan, D. P., Rose, S. A., Samocha, K. E., Fromer, M., Cerrato, F., Chen, W. J., … Neale, B. M.

(2020). Exome sequencing in schizophrenia-affected parent-offspring trios reveals risk conferred by protein-coding de novo mutations. *Nature Neuroscience*, *23*(2), 185–193. doi:10.1038/s41593-019-0564-3

Hu, G., Cui, K., Fang, D., Hirose, S., Wang, X., Wangsa, D., … Zhao, K. (2018). Transformation of accessible chromatin and 3D nucleome underlies lineage commitment of early T cells. *Immunity*, *48*(2), 227-242.e8. doi:10.1016/j.immuni.2018.01.013

Kahn, R. S., & Keefe, R. S. E. (2013). Schizophrenia is a cognitive illness: time for a change in focus. *JAMA Psychiatry*, *70*(10), 1107–1112. doi:10.1001/jamapsychiatry.2013.155

Kastner, P., Chan, S., Vogel, W. K., Zhang, L.-J., Topark-Ngarm, A., Golonzhka, O., … Leid, M. (2010). Bcl11b represses a mature T-cell gene expression program in immature CD4(+)CD8(+) thymocytes. *European Journal of Immunology*, *40*(8), 2143–2154. doi:10.1002/eji.200940258

Kelly, D. L., Li, X., Kilday, C., Feldman, S., Clark, S., Liu, F., … Tonelli, L. H. (2018). Increased circulating regulatory T cells in medicated people with schizophrenia. *Psychiatry Research*, *269*, 517–523. doi:10.1016/j.psychres.2018.09.006

Khandaker, G. M., & Dantzer, R. (2016). Is there a role for immune-to-brain communication in schizophrenia? *Psychopharmacology*, *233*(9), 1559–1573. doi:10.1007/s00213-015-3975-1

Khandaker, G. M., Pearson, R. M., Zammit, S., Lewis, G., & Jones, P. B. (2014). Association of serum interleukin 6 and C-reactive protein in childhood with depression and psychosis in young adult life: a population-based longitudinal study. *JAMA Psychiatry*, *71*(10), 1121–1128. doi:10.1001/jamapsychiatry.2014.1332

Kim, D., Langmead, B., & Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nature Methods*, *12*(4), 357–360. doi:10.1038/nmeth.3317

Kirov, G., Pocklington, A. J., Holmans, P., Ivanov, D., Ikeda, M., Ruderfer, D., … Owen, M. J. (2012). De novo CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. *Molecular Psychiatry*, *17*(2), 142–153. doi:10.1038/mp.2011.154

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. doi:10.1038/nmeth.1923

Lee, E. E., Hong, S., Martin, A. S., Eyler, L. T., & Jeste, D. V. (2017). Inflammation in schizophrenia:

cytokine levels and their relationships to demographic and clinical variables. *The American Journal of Geriatric Psychiatry*, *25*(1), 50–61. doi:10.1016/j.jagp.2016.09.009

Lee, J. J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., … Cesarini, D. (2018). Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nature Genetics*, *50*(8), 1112–1121. doi:10.1038/s41588-018-0147-3

Leeson, V. C., Barnes, T. R. E., Hutton, S. B., Ron, M. A., & Joyce, E. M. (2009). IQ as a predictor of functional outcome in schizophrenia: a longitudinal, four-year study of first-episode psychosis. *Schizophrenia Research*, *107*(1), 55–60. doi:10.1016/j.schres.2008.08.014

Lessel, D., Gehbauer, C., Bramswig, N. C., Schluth-Bolard, C., Venkataramanappa, S., van Gassen, K. L. I., … Kubisch, C. (2018). BCL11B mutations in patients affected by a neurodevelopmental disorder with reduced type 2 innate lymphoid cells. *Brain: A Journal of Neurology*, *141*(8), 2299–2311. doi:10.1093/brain/awy173

Li, P., Burke, S., Wang, J., Chen, X., Ortiz, M., Lee, S.-C., … Liu, P. (2010). Reprogramming of T cells to natural killer-like cells upon Bcl11b deletion. *Science*, *329*(5987), 85–89. doi:10.1126/science.1188063

Li, Q., Zhou, J., Cao, X., Liu, Q., Li, Q., Li, W., & Wang, X. (2018). Clonal Characteristics of T-Cell Receptor Repertoires in Violent and Non-violent Patients With Schizophrenia. *Frontiers in Psychiatry*, *9*, 403. doi:10.3389/fpsyt.2018.00403

Misiak, B., Stańczykiewicz, B., Kotowicz, K., Rybakowski, J. K., Samochowiec, J., & Frydecka, D. (2018). Cytokines and C-reactive protein alterations with respect to cognitive impairment in schizophrenia and bipolar disorder: A systematic review. *Schizophrenia Research*, *192*, 16–29. doi:10.1016/j.schres.2017.04.015

Nyland, H., Naess, A., & Lunde, H. (1980). Lymphocyte subpopulations in peripheral blood from schizophrenic patients. *Acta Psychiatrica Scandinavica*, *61*(4), 313–318. doi:10.1111/j.1600-0447.1980.tb00584.x

O'Brien, H. E., Hannon, E., Hill, M. J., Toste, C. C., Robertson, M. J., Morgan, J. E., … Bray, N. J. (2018). Expression quantitative trait loci in the developing human brain and their enrichment in neuropsychiatric disorders. *Genome Biology*, *19*(1), 194. doi:10.1186/s13059-018-1567-1

Pardiñas, A. F., Holmans, P., Pocklington, A. J., Escott-Price, V., Ripke, S., Carrera, N., … Walters, J. T. R. (2018). Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nature Genetics*, *50*(3), 381–389. doi:10.1038/s41588-018-0059-2

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T.-C., Mendell, J. T., & Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology*, *33*(3), 290–295. doi:10.1038/nbt.3122

Pouget, J. G. (2018). The emerging immunogenetic architecture of schizophrenia. *Schizophrenia Bulletin*, *44*(5), 993–1004. doi:10.1093/schbul/sby038

Rees, E., Han, J., Morgan, J., Carrera, N., Escott-Price, V., Pocklington, A. J., … Owen, M. J. (2020). De novo mutations identified by exome sequencing implicate rare missense variants in SLC6A1 in schizophrenia. *Nature Neuroscience*, *23*(2), 179–184. doi:10.1038/s41593-019-0565-2

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, *26*(1), 139–140. doi:10.1093/bioinformatics/btp616

Rodriguez, G. M., Bobbala, D., Serrano, D., Mayhue, M., Champagne, A., Saucier, C., … Ilangumaran, S. (2016). NLRC5 elicits antitumor immunity by enhancing processing and presentation of tumor antigens to CD8(+) T lymphocytes. *Oncoimmunology*, *5*(6), e1151593. doi:10.1080/2162402X.2016.1151593

Savage, J. E., Jansen, P. R., Stringer, S., Watanabe, K., Bryois, J., de Leeuw, C. A., … Posthuma, D. (2018). Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nature Genetics*, *50*(7), 912–919. doi:10.1038/s41588-018-0152-6

Schizophrenia Working Group of the Psychiatric Genomics Consortium. (2014). Biological insights from 108 schizophrenia-associated genetic loci. *Nature*, *511*(7510), 421–427. doi:10.1038/nature13595

Sekar, A., Bialas, A. R., de Rivera, H., Davis, A., Hammond, T. R., Kamitaki, N., … McCarroll, S. A. (2016). Schizophrenia risk from complex variation of complement component 4. *Nature*, *530*(7589), 177–183. doi:10.1038/nature16549

Simon, R, Baumann, L., Fischer, J., Seigfried, F. A., De Bruyckere, E., Liu, P., … Britsch, S. (2016).

Structure-function integrity of the adult hippocampus depends on the transcription factor Bcl11b/Ctip2. *Genes, Brain, and Behavior*, *15*(4), 405–419. doi:10.1111/gbb.12287

Simon, Ruth, Brylka, H., Schwegler, H., Venkataramanappa, S., Andratschke, J., Wiegreffe, C., … Britsch, S. (2012). A dual function of Bcl11b/Ctip2 in hippocampal neurogenesis. *The EMBO Journal*, *31*(13), 2922–2936. doi:10.1038/emboj.2012.142

Singh, T., Kurki, M. I., Curtis, D., Purcell, S. M., Crooks, L., McRae, J., … Barrett, J. C. (2016). Rare loss-of-function variants in SETD1A are associated with schizophrenia and developmental disorders. *Nature Neuroscience*, *19*(4), 571–577. doi:10.1038/nn.4267

Smedley, D., Haider, S., Durinck, S., Pandini, L., Provero, P., Allen, J., … Kasprzyk, A. (2015). The BioMart community portal: an innovative alternative to large, centralized data repositories. *Nucleic Acids Research*, *43*(W1), W589-98. doi:10.1093/nar/gkv350

Świtaj, P., Anczewska, M., Chrostek, A., Sabariego, C., Cieza, A., Bickenbach, J., & Chatterji, S. (2012). Disability and schizophrenia: a systematic review of experienced psychosocial difficulties. *BMC Psychiatry*, *12*, 193. doi:10.1186/1471-244X-12-193

Tang, B., Di Lena, P., Schaffer, L., Head, S. R., Baldi, P., & Thomas, E. A. (2011). Genome-wide identification of Bcl11b gene targets reveals role in brain-derived neurotrophic factor signaling. *Plos One*, *6*(9), e23691. doi:10.1371/journal.pone.0023691

Upthegrove, R., & Khandaker, G. M. (2019). Cytokines, oxidative stress and cellular markers of inflammation in schizophrenia. *Current Topics in Behavioral Neurosciences*. doi:10.1007/7854_2018_88

Varun, C. N., Venkataswamy, M. M., Ravikumar, R., Nagaraju, R., Debnath, M., Varambally, S., … Ravi, V. (2019). Th17 and MAIT cell mediated inflammation in antipsychotic free schizophrenia patients. *Schizophrenia Research*. doi:10.1016/j.schres.2019.08.013

Wang, D., Liu, S., Warrell, J., Won, H., Shi, X., Navarro, F. C. P., … Gerstein, M. B. (2018). Comprehensive functional genomic resource and integrative model for the human brain. *Science*, *362*(6420). doi:10.1126/science.aat8464

Wang, L., & Bosselut, R. (2009). CD4-CD8 lineage differentiation: Thpok-ing into the nucleus. *Journal of Immunology*, *183*(5), 2903–2910. doi:10.4049/jimmunol.0901041

Wang, S., Sun, H., Ma, J., Zang, C., Wang, C., Wang, J., … Liu, X. S. (2013). Target analysis by integration of transcriptome and ChIP-seq data with BETA. *Nature Protocols*, *8*(12), 2502–2515. doi:10.1038/nprot.2013.150

Ware, J. S., Samocha, K. E., Homsy, J., & Daly, M. J. (2015). Interpreting de novo Variation in Human Disease Using denovolyzeR. *Current Protocols in Human Genetics*, *87*, 7.25.1-7.25.15. doi:10.1002/0471142905.hg0725s87

Whitton, L., Cosgrove, D., Clarkson, C., Harold, D., Kendall, K., Richards, A., … Morris, D. W. (2016). Cognitive analysis of schizophrenia risk genes that function as epigenetic regulators of gene expression. *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics*, *171*(8), 1170–1179. doi:10.1002/ajmg.b.32503

Wu, Y., Zeng, J., Zhang, F., Zhu, Z., Qi, T., Zheng, Z., … Yang, J. (2018). Integrative analysis of omics summary data reveals putative mechanisms underlying complex traits. *Nature Communications*, *9*(1), 918. doi:10.1038/s41467-018-03371-0

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., … Liu, X. S. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biology*, *9*(9), R137. doi:10.1186/gb-2008-9-9-r137

Table 1: Results from MAGMA gene-set analysis of BCL11B gene-sets using GWAS data for EA, IQ and SCZ.

Gene-set	Phenotype	# Genes	Beta	P value	FDR Q value
Striatal	EA	215	-0.072	0.799	0.799
	IQ	219	0.059	0.213	0.639
	SCZ	218	0.116	0.059	0.516
DN4	EA	106	0.027	0.405	0.643
	IQ	106	0.047	0.313	0.643
	SCZ	104	0.013	0.446	0.643
DP	EA	502	0.067	0.115	0.516
	IQ	506	-0.025	0.709	0.799
	SCZ	495	1.43e-05	0.500	0.643

Table 2: Results from de novo mutation analysis of BCL11B gene-sets using data from 3,447 SCZ trios.

Gene set	# Genes inside set	# Genes outside set	DNM type	# DNMs inside set		# DNMs outside set		Two sample rate ratio (lower 95% bound; upper 95% bound)	P-Value	Q-Value
				Observed	Expected	Observed	Expected			
Striatal	220	19,398	syn	14	13	797	953.5	1.288 (0.701; 2.175)	0.356	1
			mis	31	29.7	2140	2141.4	1.044 (0.708; 1.487)	0.781	1
			lof	4	4.7	360	296.7	0.701 (0.19; 1.812)	0.67	1
DN4	112	19,506	syn	8	7.9	803	958.6	1.209 (0.52; 2.394)	0.554	1
			mis	17	17.6	2154	2153.6	0.966 (0.562; 1.55)	1	1
			lof	2	2.3	362	299.1	0.718 (0.087; 2.614)	1	1
DP	507	19,111	syn	27	28.4	784	938.1	1.138 (0.745; 1.668)	0.467	1
			mis	90	63.6	2081	2107.5	**1.433 (1.147; 1.77)**	**0.001**	**0.013**
			lof	8	8.1	356	293.3	0.814 (0.349; 1.621)	0.744	1
DP_upreg	251	19,367	mis	49	33.6	2122	2137.6	**1.469 (1.093; 1.967)**	**0.009**	post-hoc test
DP_down reg	256	19,362	mis	41	30.1	2130	2141	1.369 (0.98; 1.864)	0.053	post-hoc test