



OPEN

Understanding the impact of SNPs associated with autism spectrum disorder on biological pathways in the human fetal and adult cortex

E. Golovina¹, T. Fadason^{1,2}, T. J. Lints³, C. Walker⁴, M. H. Vickers^{1,2} & J. M. O'Sullivan^{1,2,5,6,7}✉

Autism spectrum disorder (ASD) is a neurodevelopmental disorder characterized by significant and complex genetic etiology. GWAS studies have identified genetic variants associated with ASD, but the functional impacts of these variants remain unknown. Here, we integrated four distinct levels of biological information (GWAS, eQTL, spatial genome organization and protein–protein interactions) to identify potential regulatory impacts of ASD-associated SNPs ($p < 5 \times 10^{-8}$) on biological pathways within fetal and adult cortical tissues. We found 80 and 58 SNPs that mark regulatory regions (i.e. expression quantitative trait loci or eQTLs) in the fetal and adult cortex, respectively. These eQTLs were also linked to other psychiatric disorders (e.g. schizophrenia, ADHD, bipolar disorder). Functional annotation of ASD-associated eQTLs revealed that they are involved in diverse regulatory processes. In particular, we found significant enrichment of eQTLs within regions repressed by Polycomb proteins in the fetal cortex compared to the adult cortex. Furthermore, we constructed fetal and adult cortex-specific protein–protein interaction networks and identified that ASD-associated regulatory SNPs impact on immune pathways, fatty acid metabolism, ribosome biogenesis, aminoacyl-tRNA biosynthesis and spliceosome in the fetal cortex. By contrast, in the adult cortex they largely affect immune pathways. Overall, our findings highlight potential regulatory mechanisms and pathways important for the etiology of ASD in early brain development and adulthood. This approach, in combination with clinical studies on ASD, will contribute to individualized mechanistic understanding of ASD development.

Autism spectrum disorder (ASD) represents a heterogeneous group of closely related conditions that are characterized by early-appearing social communication deficits and restricted, repetitive or unusual sensory-motor behaviours¹. Epidemiological studies estimate that approximately 1% of people worldwide have ASD².

Over the past decade, genome-wide association (GWAS) and genetic studies have identified increasing numbers of single nucleotide polymorphisms (SNPs)^{3,4} and other forms of variation (e.g., copy number variants, rare structural variants)^{5,6} that are associated with ASD. The proportion of ASD explained by SNPs has been estimated to be between 17 and 60%^{7,8}, thus their contribution should not be neglected. However, the functions of the genetic variants that are responsible for the association with ASD remain poorly characterized. As such, we do not yet fully understand how to translate information on ASD-associated SNPs into specific biological mechanisms that can be therapeutically targeted to alleviate the symptoms and complications of ASD.

The majority of ASD-associated SNPs are located within the non-coding components of the genome. Studies of non-coding disease-associated SNPs have demonstrated that they can mark regulatory elements that alter gene expression^{9,10}. Notably, these regulatory elements are only associated with the expression (eQTL or expression quantitative trait locus) of the adjacent gene in ~40% of cases¹¹. The remaining 60% of the identified eQTLs involve interactions with non-adjacent genes that can be > 1 Mb away in the linear DNA sequence or even on a different chromosome. The regulatory effects can occur in trans (e.g. miRNA) or by spatial associations

¹Liggins Institute, University of Auckland, Auckland, New Zealand. ²Maurice Wilkins Centre, University of Auckland, Auckland, New Zealand. ³School of Medical Science, University of Auckland, Auckland, New Zealand. ⁴School of Population Health, University of Auckland, Auckland, New Zealand. ⁵Brain Research New Zealand, University of Auckland, Auckland, New Zealand. ⁶MRC Lifecourse Epidemiology Unit, University of Southampton, Southampton, UK. ⁷Garvan Institute of Medical Research, Sydney, Australia. ✉email: justin.osullivan@auckland.ac.nz

of the regulatory element and target gene. As such, the three-dimensional (3D) genome organization, which emerges from the sum of the biophysical interactions within the nucleus, includes tissue-specific spatial interactions between eQTL regions and the genes that they control (hereafter eGenes)¹². These spatial interactions are dynamic, developmentally and temporally dependent¹³. Thus, integrating biological measurements on developmental and tissue-specific spatial chromatin interactions with eQTL information could inform our understanding of the regulatory impacts of ASD-associated SNPs.

ASD is widely considered to be a neurodevelopmental disorder resulting from functional changes within the brain. There are studies connecting cortical dysfunctions and ASD using imaging¹⁴, cortical architecture¹⁵, or gene expression¹⁶. Therefore, characterizing the functional impacts (i.e. on gene regulation) of the ASD-associated SNPs and translating them to the affected biological pathways in fetal and adult cortical tissues may provide mechanistic insights into the etiology of ASD during neurodevelopment.

Using protein–protein interaction (PPI) networks to explore interactions between proteins encoded by known disease-associated genes is a powerful approach to study the etiology of complex diseases, including psychiatric disorders^{17,18}. PPI network analyses have been used to discover essential proteins, clusters of proteins with similar, overlapping or combinatorial functions, and associated pathways involved in tissue-specific contributions to ASD etiology^{17,18}. However, the potential contributions of cortex-specific developmental changes to these networks in ASD development have yet to be investigated.

Here, we integrated ASD-associated GWAS SNPs with cortex-specific 3D genome structure and eQTL information to identify genes that are spatially regulated in fetal (14–21 postconceptional weeks) and adult (21–70 years of age) cortical tissues. We incorporated cortex-specific expression patterns and PPI networks to identify candidate genes and pathways that have putative roles in the etiology of ASD-associated changes in the cortex. The identified gene sets were enriched for immune pathways, fatty acid metabolism, ribosome biogenesis, aminoacyl-tRNA biosynthesis and the spliceosome in the fetal cortex. By contrast, the adult cortical gene set was largely related to immune pathways. Collectively, our results provide insight into potential cortex-specific regulatory mechanisms and pathways through which ASD-associated SNPs can contribute to the development and maintenance of ASD.

Results

ASD-associated SNPs mark putative regulatory regions shared between or specific to adult and fetal cortical tissues. ASD-associated SNPs ($p < 5 \times 10^{-8}$, $n = 454$) were downloaded from the GWAS Catalog (Supplementary Tables 1 and 2). Of 454 ASD-associated SNPs, 344 SNPs were represented in both fetal and adult cortex eQTL databases, and were run through the CoDeS3D pipeline (Fig. 1, Supplementary Fig. 1, Supplementary Table 3).

We identified 80 eQTLs that are involved in 131 significant spatial eQTL–eGene interactions in fetal cortex; and 58 eQTLs that are associated with 67 significant spatial eQTL–eGene interactions in adult cortex (Fig. 2a, Supplementary Table 3).

Of the 80 fetal and 58 adult eQTLs, 30 were observed in both fetal and adult cortical tissues (Fig. 2b). Fifteen of these shared eQTLs control the same eGenes in fetal and adult cortex (e.g. rs10791097–*SNX19*, rs11191419–*AS3MT*, rs7085104–*AS3MT*, rs174592–*FADS1*; Supplementary Table 4). However, the remaining 15 eQTLs are associated with different eGenes in adult and fetal cortical tissues (Supplementary Table 4). For example, rs35828350: (1) upregulates *NMB* in fetal cortex. *NMB* encodes the neuromedin B peptide that regulates physiological processes including cell growth, exocrine and endocrine secretion²²; and (2) downregulates *WDR73* in adult cortex. *WDR73* encodes the WD Repeat-containing protein 73 that is linked to microtubule organization and dynamics. Notably, eQTLs involving rs13218591 and rs2237234 regulate different butyrophilin alleles (i.e. *BTN2A2* and *BTN3A1* in the adult and fetal cortex, respectively). The butyrophilin genes encode proteins that belong to the immunoglobulin superfamily and help modulate the immune system²³. Finally a number of fetal ($n = 50$) and adult ($n = 28$) cortex-specific eQTLs were observed (Supplementary Table 3). Collectively, these observations are consistent with changes in the regulation of subsets of stable and remodelled spatial eQTLs, over the course of brain development, being associated with a predisposition to ASD.

ASD-associated eQTLs are linked to onset of psychiatric disorders. Previous research has reported shared neurobiological and cellular processes associated with differences in cortical thickness across six psychiatric disorders (i.e. ASD, attention-deficit/hyperactivity disorder (ADHD), bipolar disorder, unipolar depression, obsessive–compulsive disorder and schizophrenia), implicating common mechanisms underlying cortical development²⁴. To evaluate possible commonalities among ASD and other phenotypes at the eQTL level in the fetal and adult cortex, we intersected the identified ASD-associated eQTLs with SNPs associated with other traits in the GWAS catalog ($p < 5 \times 10^{-8}$, assessed on 26/08/2020). Fetal and adult ASD-associated eQTLs were also associated with schizophrenia, unipolar depression, ADHD, bipolar disorder, anorexia nervosa and obsessive–compulsive disorder (Supplementary Fig. 2, Supplementary Table 5). We observed that schizophrenia has the largest significant overlap with ASD-associated eQTLs both in fetal (78 out of 80, bootstrapping $p < 0.01$, $n = 10,000$) and adult (57 out of 58, bootstrapping $p < 0.01$, $n = 10,000$) cortical tissues (Supplementary Fig. 2, Supplementary Table 5). This observation is consistent with: (1) a comorbid association between ASD and schizophrenia²⁵; or (2) a lack of resolution and precision in defining the ASD and schizophrenia phenotypes, and thus possible false positives in GWASs²⁶.

ASD-associated eQTLs are involved in diverse regulatory processes. We functionally annotated the ASD-associated eQTLs to understand the potential regulatory mechanisms of the regions they tagged (Supplementary Table 5). As expected, the majority of identified eQTLs were located within intronic and intergenic

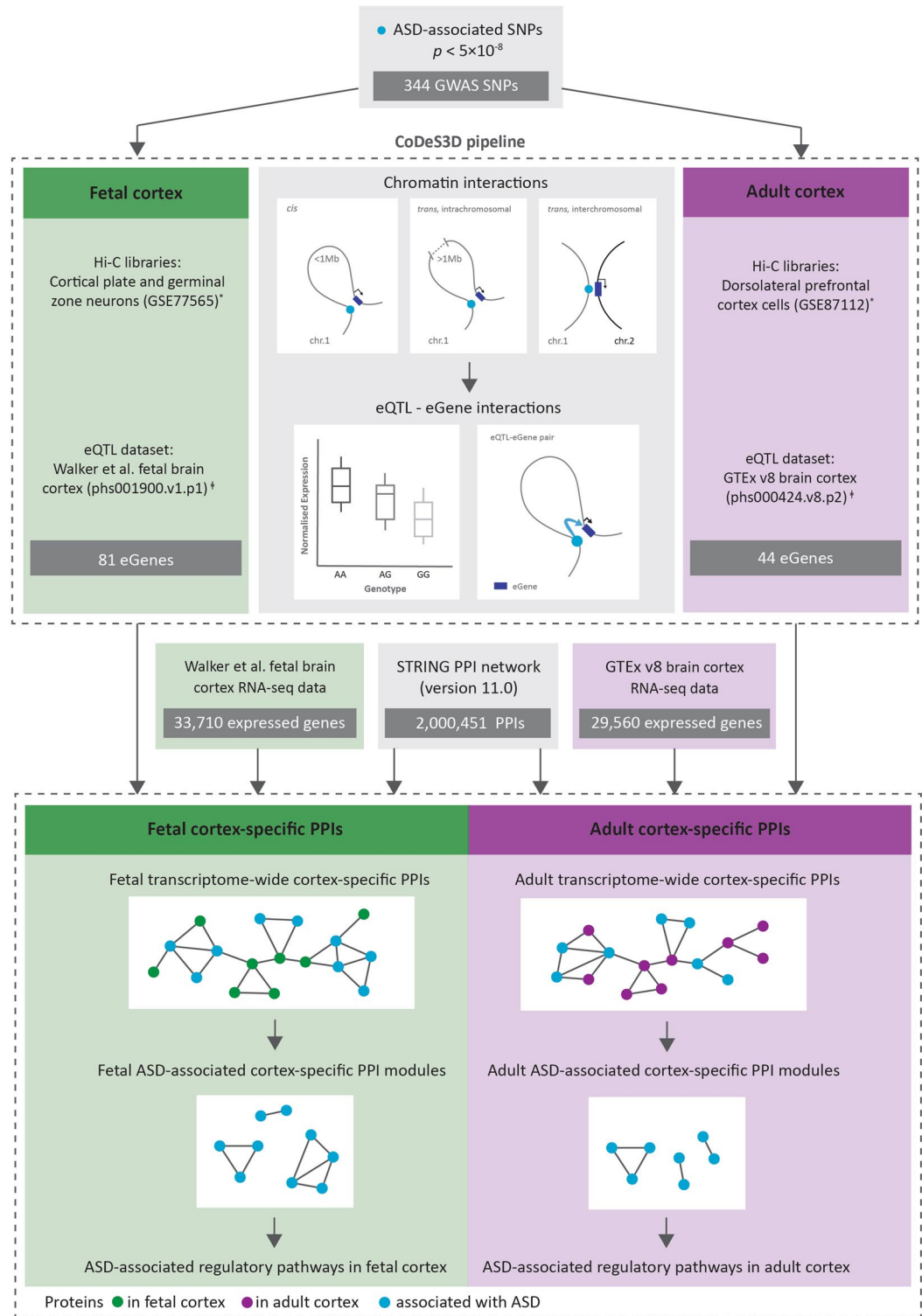
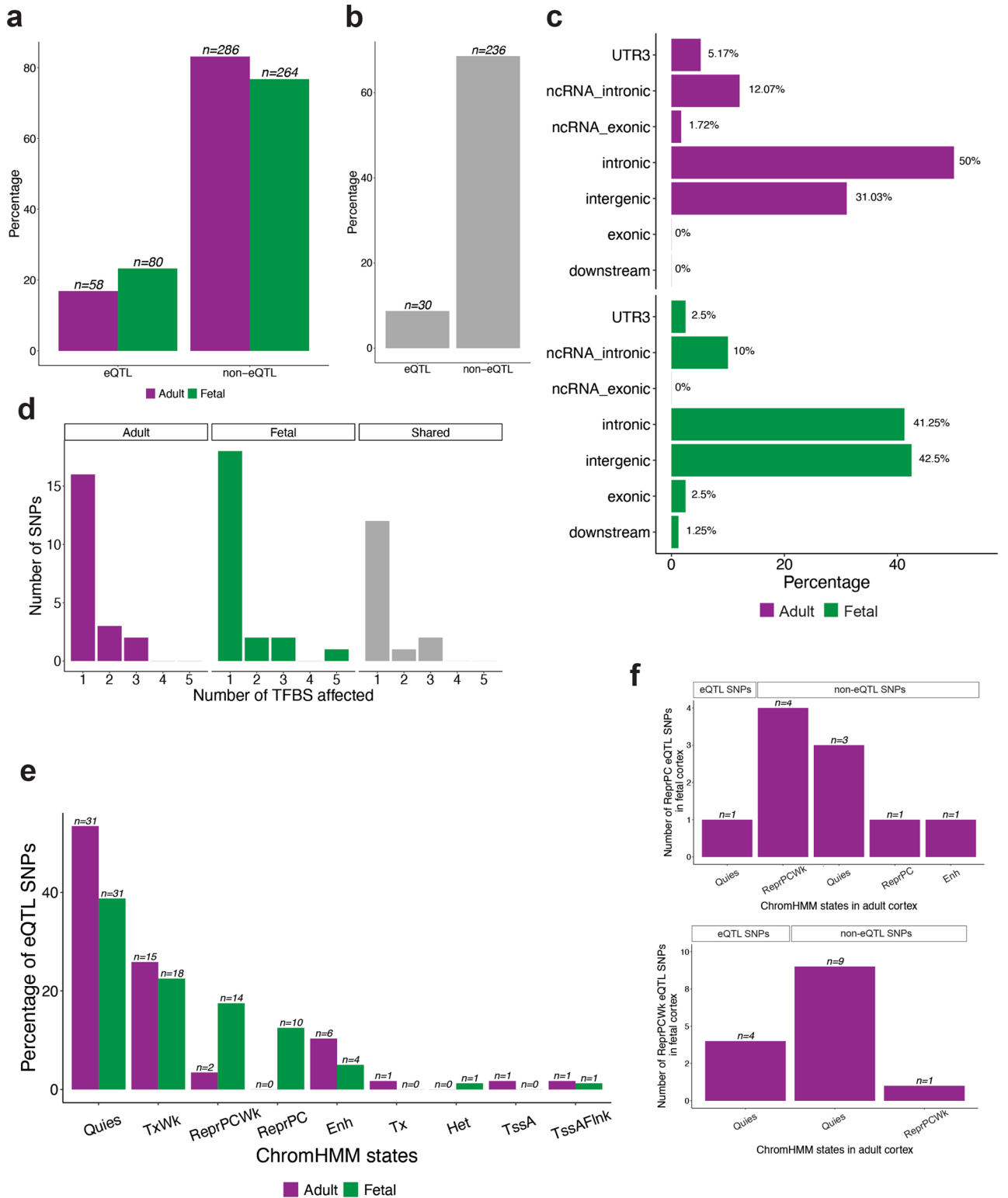


Figure 1. Overview of the analysis pipeline used in this study. 344 ASD-associated SNPs ($p < 5 \times 10^{-8}$) represented in both fetal and adult cortex-specific eQTL datasets were run through the CoDeS3D pipeline to identify 81 and 44 spatially regulated genes in fetal and adult cortical tissues, correspondingly. Over 11 million protein–protein interactions (PPIs) were downloaded from STRING database (version 11.0) and combined with cortex-specific expression data (GTEX v8 or Walker et al. datasets) to construct tissue-specific transcriptome-wide PPI networks. Modules that were enriched with ASD-eQTL associated genes were identified in the fetal and adult cortical tissues. *Hi-C datasets for cortical plate and germinal zone neurons (phs001190.v1.p1) were obtained from Won et al.¹⁹, Hi-C datasets for adult dorsolateral prefrontal cortex cells were obtained from Schmitt et al.¹². †eQTL datasets for fetal and adult cortex were obtained from Walker et al.²⁰ and GTEX v8²¹, correspondingly.



◀ **Figure 2.** ASD-associated SNPs are enriched within non-coding putative regulatory regions. (a) Of 344 ASD-associated SNPs represented in both fetal and adult cortex-specific eQTL databases, more SNPs ($n = 80$) are involved in spatial eQTL-gene interactions in the fetal cortex than in the adult cortex ($n = 58$). The proportions of eQTL and non-eQTL SNPs are significantly different in fetal and adult cortical tissues (Fisher's exact test, $p = 0.04531$). (b) Thirty ASD-associated SNPs are eQTLs in both fetal and adult cortical tissues. (c) All ASD-associated eQTLs in adult cortex ($n = 58$) and approximately 78 (97.5%) of the ASD-associated eQTLs within the fetal cortex are located within non-coding genomic regions (Supplementary Table 5). (d) 15 and 18 ASD-associated eQTLs affect at least one transcription factor binding sites within the fetal and adult cortical tissues, respectively. (e) Most of the fetal ASD-associated eQTLs are located within quiescent/low transcribed, weak repressed PolyComb, repressed PolyComb and weak transcription regions. By contrast, the majority of ASD-associated eQTLs, that were identified in the adult cortex, are located within quiescent/low transcribed and weak transcription regions. *Enh* enhancers, *Het* heterochromatin, *Quies* quiescent/low, *ReprPC* repressed PolyComb, *ReprPCWk* weak repressed PolyComb, *TssA* active TSS, *TssAFlnk* flanking active TSS, *Tx* strong transcription, *TxWk* weak transcription. (f) The majority of the fetal ASD-associated eQTLs that are located within weakly repressed PolyComb (ReprPCWk) and repressed PolyComb (ReprPC) regions were not identified as being eQTLs within the adult cortex.

regions (Fig. 2c). The SNP2TFBS²⁷ database was queried to identify eQTLs that are predicted to alter the affinity of transcription factor binding sites (TFBSs). We identified 21 and 23 eQTLs that reduce the affinity of at least one TFBS in fetal and adult cortical tissues, respectively (Fig. 2d). Finally, we tested for enrichment of ASD-associated eQTLs within active regulatory elements and histone modification marks, using ChromHMM²⁸ 15-state models for adult dorsolateral prefrontal cortex and fetal brain. Fetal ASD-associated eQTLs were located within quiescent/low transcribed ($n = 31$), weak transcription ($n = 18$), weak repressed Polycomb ($n = 14$) and repressed Polycomb ($n = 10$) regions (Fig. 2e). By contrast, adult ASD-associated eQTLs were located within quiescent/low transcribed ($n = 31$) and weak transcription ($n = 15$) regions (Fig. 2e). There was significant enrichment ($p < 0.01$, permutation test: 1000) of ASD-associated eQTLs within loci repressed by Polycomb proteins in the fetal cortex when compared to the adult cortex. Most of these fetal ASD-associated eQTLs located within the Polycomb-repressed eQTLs were not identified as eQTLs in the adult cortex (Fig. 2f). Notably, Polycomb repressive complexes have distinct regulatory roles in identity, proliferation and differentiation of neuronal progenitor cells during development^{29,30}.

ASD-associated eQTLs are developmental stage-specific. There is no fundamental reason why the same regulatory elements must or must not impact on the same gene in different tissues, or at different stages of development. ASD-associated eQTLs regulate 81 genes in fetal and 44 genes in adult cortical tissues (Fig. 3). Of these genes, 15 are spatially regulated in both fetal and adult cortical tissues (Fig. 3, Supplementary Table 4). Different eQTLs are often associated with the gene transcript levels, although the effects of the minor allele are typically similar for the gene in question (i.e. associated with an increase or reduction in transcript levels; Supplementary Fig. 3). For example, rs4647903, rs2535629, rs221902, rs7743252, rs832190 eQTLs were associated with increased transcript levels of *DDHD2*, *ITIH4*, *PCNX1*, *TAP2*, *THOC7* genes both in fetal and adult cortical tissues. However, despite having the same effects in fetal and adult tissues, some eQTLs had opposite direction of effects on the gene of interest (e.g. the effects of rs7432375 and rs7618871 on *PCCB* gene transcript levels are not collinear; Supplementary Fig. 3). Moreover, we identified development stage-specific eQTLs that were associated with gene transcript levels in either fetal, or adult cortical tissue (Supplementary Fig. 3). For example, rs4481150 eQTL is present in both fetal and adult eQTL databases. However, rs4481150 is only associated with increased transcript levels for *ITIH4* in adult cortex. By contrast, the rs1518367 eQTL is associated with reduced transcript levels for *SF3B1* only in fetal cortex. *HLA-DMA* and *BAG6* gene transcript levels are associated with distinct sets of eQTLs in both fetal and adult cortical tissues (Supplementary Fig. 3).

Despite the existence of 15 genes in eQTL with ASD-associated SNPs in both fetal and adult cortical tissues, the majority of the changes in gene transcript levels were specific to either fetal ($n = 66$), or adult ($n = 29$) cortical tissue (Fig. 3). Transcript levels for these genes are associated with 65 eQTLs in fetal cortex and 39 eQTLs in adult cortex (Supplementary Table 3).

Of the 66 fetal cortex genes, transcript levels for 36 are upregulated in association with a change of the eQTL SNP to the minor allele, while 24 are reduced (Fig. 3). Notably, 6 genes are associated with multiple eQTLs which exhibit opposing effects on transcript levels (Fig. 3). Nineteen eGenes are regulated by Polycomb-repressed eQTL SNPs, and seven eGenes—by trans-acting eQTL SNPs in fetal cortex (Fig. 3).

Similar changes occur in adult cortex where transcript levels for 15 genes are upregulated in association with a change to the minor allele at the eQTL SNP (Fig. 3). Again transcript levels for 12 genes are reduced. We also identified two genes that had multiple eQTLs where substitution of the SNP with the minor allele had opposing associations with the gene's transcript levels (Fig. 3). Collectively, these findings are consistent with a subset of ASD-associated eQTLs acting in a combinatorial and development stage specific manner to affect the risk of developing ASD.

Eleven genes associated with ASD-eQTLs have previously been linked to ASD risk. To identify existing and novel gene associations, we intersected our lists of genes, from fetal and adult cortical tissues, with a curated list of 1237 genes that had been previously implicated in autism development (AutDB³¹, accessed on 16/11/2020). Eleven genes (i.e. 8 from fetal cortex: *DDHD2*, *HLA-DRB1*, *PCCB*, *PCDH15*, *SF3B1*, *SNX19*, *TBL1X*, *VWA7*; and 7 from adult cortex: *DDHD2*, *GALNT10*, *PBX1*, *PCCB*, *REPER*, *SF3B1*, *SNX19*) had been previously linked to ASD (Fig. 3). Bootstrapping analysis revealed that these overlaps are significant ($p < 0.01$,

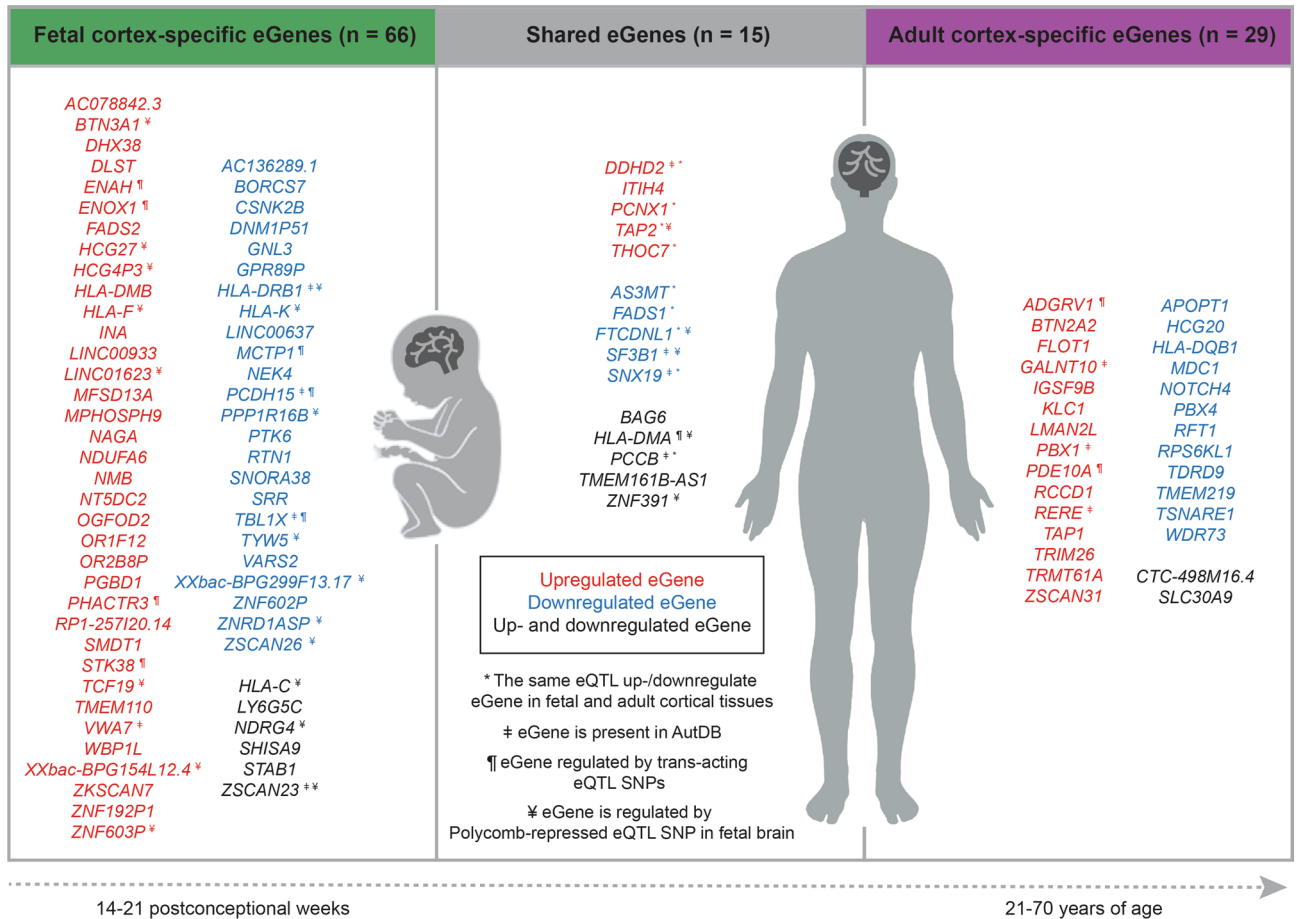


Figure 3. ASD-associated eQTLs mark loci that spatially regulate the expression of genes involved in the fetal brain, in the adult brain, or both. Transcript levels for 15 spatially regulated genes were altered by ASD-associated eQTLs in both the fetal and adult cortical tissues, 66 genes were specific to fetal cortex, and 29 eGenes were specific to the adult cortex. *The same eQTL is associated with transcription levels for the gene in both the fetal and adult cortical tissues. †Genes that have been curated as being involved in ASD (AutDB³¹; <http://autism.mindspec.org/autdb/Welcome.do>, assessed on 16/11/2020). ‡Genes whose transcript levels are associated with a trans-acting ASD-associated eQTL. §Genes regulated by Polycomb-repressed ASD-associated eQTLs in the fetal cortex.

n = 10,000). However, more than 84% of the identified spatially regulated genes were ‘novel’ and have not previously been linked to autism or curated in AutDB.

ASD-eQTL associated gene set is enriched for loss-of-function tolerant genes. Genes that have essential functions show a decreased tolerance for loss-of-function (LoF) mutations³². LoF analysis revealed that 59% (fetal) and 77% (adult) of the eGenes are tolerant to variation that alters the gene sequence. By contrast, 9 fetal cortex-specific genes (i.e. *PHACTR3*, *BAG6*, *CSNK2B*, *SF3B1*, *PPP1R16B*, *FADS2*, *RTN1*, *TBL1X* and *ENAH*) and 5 adult cortex-specific genes (i.e. *PDE10A*, *PBX1*, *SF3B1*, *BAG6* and *RERE*) were LoF intolerant (Supplementary Table 6).

ASD-eQTL associated genes are enriched for immune-related processes. Functional gene ontology enrichment analysis identified immune-related processes (e.g. antigen processing and presentation) as being enriched in the ASD-eQTL associated gene sets for both fetal and adult cortical tissues (Supplementary Fig. 4). Changes to genes within the immune-related processes within adult cortex mostly affect the processing of exogenous antigen. By contrast, immune-related genes that are associated with ASD-eQTLs within the fetal cortex have been implicated in the processing of both endogenous and exogenous antigens (Supplementary Table 7). Removal of all *HLA* genes from the analysis, identified enrichments for genes involved in fatty acid metabolism and processes related to the endoplasmic-reticulum-associated protein degradation (ERAD) pathway within fetal cortex. Removal of *HLA* genes from analyses of the adult cortex gene set identified a retained enrichment for immune-related processes (e.g. antigen processing and presentation), protein kinase C signalling and regulation of cell-cell adhesion processes (Supplementary Table 8).

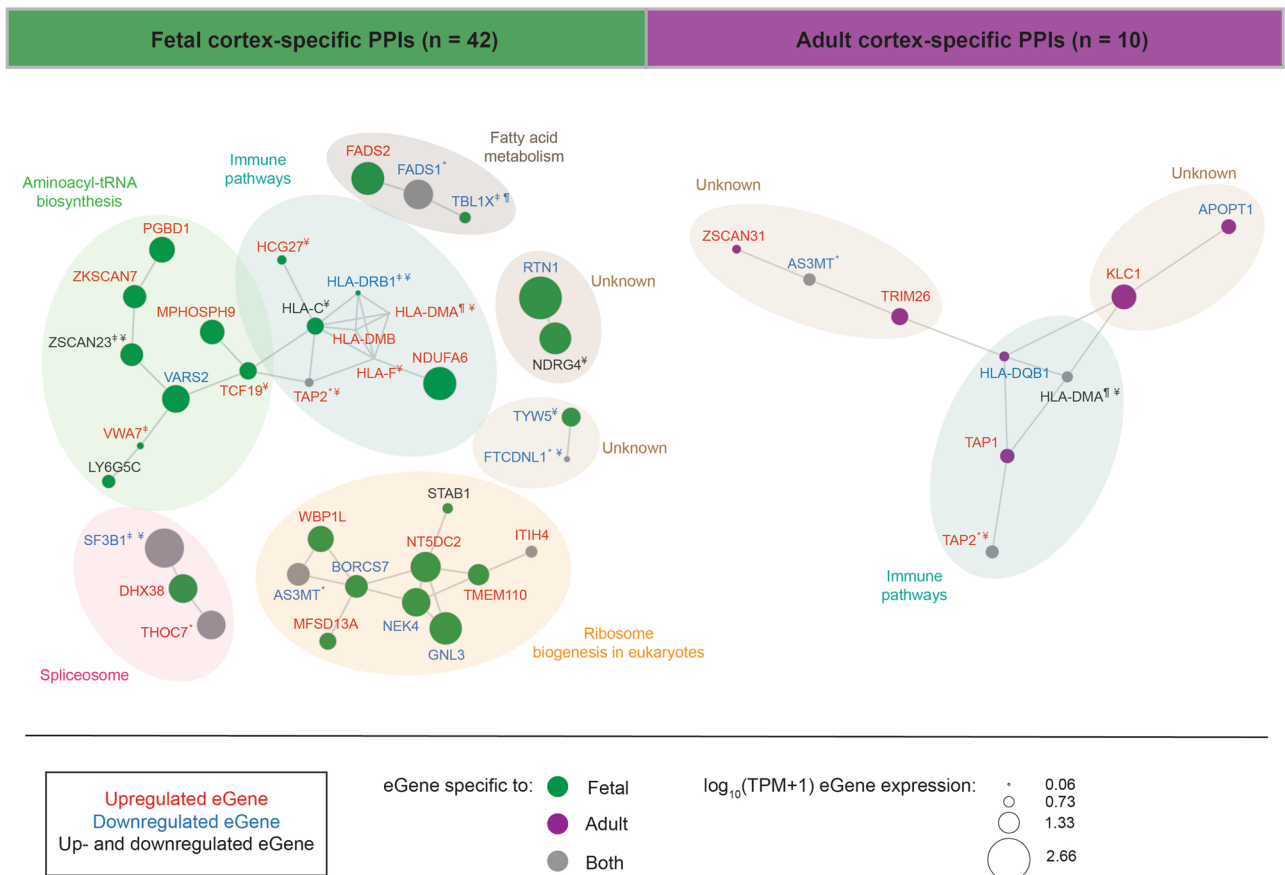


Figure 4. Fetal and adult cortical tissue-specific PPI networks with immune and growth related phenotypes are affected by ASD-associated eQTLs. We identified 42 PPIs in the fetal and 10 PPIs in the adult cortical tissues. Louvain clustering identified seven PPI modules within the fetal cortex that were enriched for immune pathways, fatty acid metabolism, aminoacyl-tRNA biosynthesis, spliceosome, ribosome biogenesis in eukaryotes and two unknown modules. Adult PPIs form three modules that were enriched in immune pathways and two unknown modules. *Gene transcript levels are associated with the same eQTL in both the fetal and adult cortical tissues. †Genes that have been curated as being involved in ASD (AutDB³¹; <http://autism.mindspec.org/autdb/Welcome.do>, assessed on 16/11/2020). ‡Genes whose transcript levels are associated with a trans-acting ASD-associated eQTL. §Genes regulated by Polycomb-repressed ASD-associated eQTLs in the fetal cortex.

ASD-specific protein–protein interaction networks in adult and fetal cortical tissues. The protein–protein interaction (PPIs) network serves as a foundation for cellular signalling circuitry, which mediates cellular responses to environmental and genetic cues. Understanding how ASD-eQTLs affect fetal and adult cortex PPIs could lead to the identification of the pathways that affect cortical development and ASD susceptibility.

PPI data was retrieved from STRING³³ (version 11.0; 24/09/2020). Adult and fetal cortex-specific PPIs (CSPPi) were generated. From these CSPPi networks we identified 42 ASD-associated PPIs within the fetal gene set, and 10 ASD-associated PPIs from the adult cortical tissue gene set (Fig. 4, Supplementary Table 9). Louvain clustering analysis identified seven highly connected PPI modules within fetal cortex. KEGG pathway analysis³⁴ of these modules revealed that they are associated with immune pathways, fatty acid metabolism, aminoacyl-tRNA biosynthesis, spliceosome, ribosome biogenesis in eukaryotes and two modules were not enriched for specific pathway (Fig. 4). By contrast, the adult PPI gene set contained three highly connected modules, two of which were not associated with specific pathway, and one of which was enriched for immune pathways.

The fetal immune PPI cluster contains both MHC class I (i.e. *HLA-C* and *HLA-F*; associated with endogenous antigen processing) and MHC class II (i.e. *HLA-DRB1*, *HLA-DMB* and *HLA-DMA*; associated with exogenous peptide processing) genes. Five genes within this cluster (i.e. *HCG27*, *TAP2*, *HLA-F*, *HLA-DMA* and *HLA-DMB*) are not highly expressed (TPM < 3) within fetal cortical tissue. However, switching the ASD-eQTL to the minor allele is associated with an increase in transcript levels for these genes within the fetal cortex (Fig. 4). By contrast, the rs3129968 minor allele is associated with a reduction in transcript levels for *HLA-DRB1*, which is expressed at low levels (TPM < 1.25) within fetal cortex. ASD-eQTLs were also associated with increases in *HCG27* (TPM = 3) and *NDUFA6* (TPM = 116.31) transcript levels within fetal cortex. Notably, the transcript levels for 6 genes in the fetal immune cluster (*HCG27*, *TAP2*, *HLA-F*, *HLA-DMA*, *HLA-C* and *HLA-DMB*) are associated with regulation by Polycomb-repressed ASD-eQTLs.

ASD-eQTLs within the fetal cortex PPI aminoacyl-tRNA biosynthesis cluster are associated with: (a) increases in transcript levels for five genes (i.e. *PGBD1*, *ZKSCAN7*, *MPHOSPH9*, *TCF19* and *VWA7*); (b) decreases in transcript levels for *VARS2*; and (c) two genes (i.e. *ZSCAN23* and *LY6G5C*) whose transcript levels increase or decrease, dependent upon specific eQTL. Similarly, the “ribosome biogenesis in eukaryotes” cluster contained ASD-eQTL associated genes whose transcripts were increased ($n = 5$), decreased ($n = 4$), and one that was subject to increases or decreases in transcript levels depending on specific ASD-eQTL. By contrast, the ASD-eQTL associations within the fetal cortex PPI fatty acid metabolism and spliceosome clusters are less complex (decreases: *FADS1*, *TBLIX*, *SF3B1*, or increases: *FADS2*, *DHX38* and *THOC7*; Fig. 4). *SF3B1* and *THOC7* were also associated with ASD-eQTLs within the adult cortex. The PPI clusters with unknown functions (Fig. 4), contained genes (i.e. *TYW5*, *FTCDNL1*, and *RTN1*) whose transcript levels decrease with the ASD-eQTL and *NDRG4*, whose transcript levels increase or decrease dependent upon the ASD-eQTL.

The patterns of transcript changes in the adult cortex PPI network were similar to those observed in the fetal cortex. Transcript levels for four genes: (a) increased (i.e. *TAP1* and *TAP2*); (b) decreased (*HLA-DQB1*); or (c) both increased and decreased depending on specific ASD-eQTLs (*HLA-DMA*; Fig. 4). Both *HLA-DQB1* and *HLA-DMA* are examples of the MHC class II genes that are associated with processing of exogenous antigen. The transcript levels of the genes within the two unknown adult cortex PPI clusters increased (i.e. *KLC1*, *ZSCAN31* and *TRIM26*), or decreased (i.e. *APOPT1*, *AS3MT*) with the ASD-eQTL.

Discussion

In this study, we integrated four distinct levels of biological information (GWAS, eQTL, genome organization [Hi-C] and protein–protein interactions [PPI] networks) to translate genetic variation associated with ASD to the biological pathways that are affected—through alterations to the transcription levels of their component proteins in fetal and adult cortical tissues. We identified shared and development-specific changes to transcript levels for spatially regulated genes within immune pathways. Interestingly, most of the genes within immune-related pathways in fetal cortex are associated with Polycomb-repressed ASD-eQTLs. At the same time, ASD-eQTLs are also associated with regulation of fatty acid metabolism, ribosome biogenesis, aminoacyl-tRNA biosynthesis and spliceosome pathways in fetal cortex. Notably, the significant difference ($p = 0.04531$) in numbers of fetal cortical eQTLs, when compared to adult cortical eQTLs is consistent with a developmental origin for ASD risk. Our findings highlight potential mechanisms through which ASD-associated variants potentially contribute to ASD development (fetal) and onset/maintenance (adult).

We identified ASD-associated eQTLs that mark putative regulatory regions in fetal ($n = 80$) and/or adult ($n = 80$) cortical tissues. It was expected that we would not identify eQTLs for all 344 of the tested ASD-associated SNPs. There are several reasons for this: (1) there are multiple potential mechanisms through which a genetic variant can impact on a phenotype. These mechanisms are not limited to impacts on gene regulation and can affect the splicing activity (so called sQTL SNPs³⁵), or trans-acting factors (e.g. non-coding RNAs); (2) ASD is a spectrum disorder not a single highly characterised phenotype; and (3) like all polygenic disorders, ASD is likely to be a whole of body disorder with important etiological contributions from discrete and distant organs within the body.

Studies have previously reported associations between ASD and: (1) schizophrenia³⁶; (2) depression³⁷; (3) ADHD^{38,39}; (4) bipolar disorder⁴⁰; and other co- and multimorbidities^{24,41,42}. Our finding that a subset of the identified ASD-associated eQTLs were linked to psychiatric phenotypes highlights the existence of potential shared regulatory mechanisms contributing to the risk of developing these multimorbid conditions. An alternative explanation is that the existence of the shared eQTLs between the multimorbid conditions is due to ambiguity in the phenotyping that was used in the GWAS studies that characterised the phenotype associated-SNPs. However, we contend that these results are consistent with the growing evidence that the vertical approach to connecting genetic variation to phenotype does not adequately account for the multimorbid nature of conditions within the typical variation that is present in humans. Future analyses that incorporate horizontal analyses of all genetic variants associated with ASD and its high-frequency multimorbid conditions will improve our ability to stratify autistic individuals and manage their complications.

Polycomb proteins are known to be involved in transcriptional silencing^{43,44}. However, studies have shown that Polycomb repressive complexes (PRC) can have a dual role in gene regulation during development^{45,46}. For example, PRC1.5 is recruited to activate genes⁴⁷ and in combination with AUTS2 (autism susceptibility candidate 2) activates gene expression in neurons⁴⁷. Our results support a dual role for Polycomb—as both a repressor and enhancer of transcription—in the development of ASD risk. Firstly, there was a significant enrichment of ASD-eQTLs within loci that are annotated as being regulated by PolyComb within the fetal, but not adult, cortex. Secondly, the finding that these ASD-eQTLs are associated with both increases and decreases in transcript levels is consistent with the up- and downregulation of the target genes. We contend that empirical studies are required to (a) confirm the regulatory activity of the sites (e.g. enhancer reporter assays); (b) confirm that the Polycomb complexes are responsible for the observed activity (e.g. by chromatin immunoprecipitation); and (c) identify the ‘Polycomb’ subunits that differentiate those sites that enhance or repress transcription within the developing cortex. The results of these experiments would be valuable in identifying novel therapeutic approaches to reduce the risk of full-blown ASD development, particularly given the strength of the evidence for Polycomb roles transcription control⁴⁶ and increasing evidence for links to neuronal development (e.g. reviewed in²⁹).

Dysregulation of fatty acid metabolism in early brain development may be a risk factor or marker for early-onset of ASD⁴⁸. Abnormalities in lipid metabolism may affect the proper functioning of the nervous system and, thus can contribute to ASD etiology^{48–50}. Consistent with this, we identified that ASD-associated genetic variants impact transcript levels for genes involved in fatty acid metabolism in the developing fetal cortex (14–21 postconceptional weeks). Notably, transcript levels for genes within this pathway were not significantly affected

by ASD-eQTLs within adult cortex tissues (21–70 years of age). Therefore, with appropriate pre-natal genetic diagnosis of risk and patient stratification, it remains possible that targeted lipid supplementation could reduce the risk of ASD. However, this would require randomised control trials in animals prior to testing in humans. Randomised controlled trials involving pre- and post-conception interventions with different lipids are currently being undertaken or followed up (e.g.⁵¹)—opening the possibility of post-hoc analyses for ASD risk.

Gene expression is the outcome of numerous processes including transcription, co-transcriptional splicing, mRNA export, and translation. We identified changes within multiple key component pathways of gene expression (i.e. spliceosome and splicing, aminoacyl-tRNA biosynthesis, and ribosome biogenesis) in the fetal cortex. Collectively these results could indicate the existence of a window of tolerable variation within gene expression—outside of which there is risk of developing ASD through changes in global gene expression. Roles for these component gene expression processes in ASD are supported by existing studies (e.g. aberrant splicing and ASD^{52,53}; upregulation of ribosomal protein genes and a higher ribosomal gene dosage can be linked to ASD risk and severity^{54,55}). Moreover, Tärnlungeanu et al. identified a form of ASD resulting from a failure of the brain to properly import amino acids—a critical substrate in gene expression⁵⁶. The significance of these associations in the developing cortex and how they contribute to the etiology of ASD remains to be empirically determined.

Roles for immune dysfunction in ASD etiology are increasingly being identified^{57–59}. We also found that immune pathways were affected both in fetal and adult cortical tissues. As we observed for other pathways, there was a developmental separation in the immune pathways that were affected in the fetal and adult cortical tissues. Notably, we found that ASD-eQTL associated changes in transcript levels occurred in adult cortex immune pathways that were associated with processing of exogenous antigen. By contrast, in the fetal cortex changes in gene transcript levels occurred in immune pathways that were related to the processing of both endogenous and exogenous antigens. These findings indicate that there may be developmental stage-specific differences in the impact of the immune system on ASD risk and on-going severity. The existence of changes within fetal cortex pathways for endogenous antigens (e.g. viral) lends some support the hypothesis that the risk of ASD may be greater among children whose mothers suffered from infectious or immune-related diseases during pregnancy—when the infant brain is developing^{60–63}.

The results of our study should be interpreted in view of its strengths and limitations. The main strength of this study is the integration of independent data sets, across distinct biological levels, that include cortex-specific 3D genome structure, eQTL and PPI data with ASD-associated GWAS SNPs. Integrating datasets across biological levels enables us to predict how genetic variation impacts on biological pathways and their involvement in the etiology of ASD. However, our study also has several limitations. Firstly, there is a great phenotypic heterogeneity within autism spectrum disorders, which has led to question their classification into a single diagnostic category⁶⁴. Secondly, common SNPs only account for ~20% to the ASD risk⁶, suggesting that other genetic (e.g., rare variants, structural variation) and environmental factors also contribute to ASD etiology^{65,66}. Increasing the number and sample sizes of the ASD GWAS studies will identify additional genetic variants which may help explain some of this missing heritability⁶⁷. Thirdly, the brain regions that are involved in the etiology of ASD remain difficult to determine. Here, we focused on roles for changes within cortical tissue. However, it is likely that additional regions of the brain (e.g. cerebellum⁶⁸) or other organs are important in the pathophysiology of ASD. As eQTL patterns are tissue-specific, we are unable to extrapolate the results of this study to these tissues. Fourthly, the human brain takes over two decades to build via precisely regulated cell type-specific molecular processes governed by both genetic blueprint and environmental factors. In our study eQTL data represent composite datasets across critical periods of development (e.g. fetal samples were aged from 14 to 21 post-conceptual weeks and the adult samples were from individuals aged 21–70 years). As such, our results only represent snap-shots within the plastic neurodevelopmental trajectory^{69,70}. This is further complicated by tissue averaging across the complex cellular organization and composition which is also different in early development and adulthood. Fifthly, we are aware that the tools and datasets used in this study are potentially biased. For example, identical samples were not used in the ChromHMM, eQTL and Hi-C analyses of the fetal and adult cortical tissues. Furthermore, the Hi-C dataset used to inform the adult cortex analysis consisted of one sample (with one replicate), while two samples (each with three replicates) were used for the fetal cortex analysis (Supplementary Table 1). Finally, protein identifiers (STRING)³³ and transcript identifiers (GTEx and Walker et al. RNA-seq data)^{20,21} were mapped to gene identifiers, thus there was a potential loss of data specificity, since genes typically produce multiple transcripts and protein variants due to alternative splicing.

84% of the genes we identified were not present in AutDB. The Gene module within AutDB is a manually curated reference set of ASD-linked genes that was first released in 2007. There are several reasons for this apparent discrepancy. Firstly, SNPs can mark regulatory regions that affect distant genes (i.e. trans-acting eQTLs)⁷¹. In this manuscript we integrated data on cortex-specific spatial chromatin interactions (i.e. Hi-C libraries) and cortex-specific eQTLs to identify how ASD-associated SNPs impact cortex-specific gene expression. The integration of spatial and functional data represents a significant difference from other approaches where functional assignments for phenotype-associated SNPs are typically made to the gene that is closest to the phenotype-associated SNP⁷². Simply assigning function to the adjacent gene increases the number of false positive gene associations⁷³. By contrast, incorporating data on spatial chromatin organization (i.e. Hi-C data) enables the identification of novel local and distal eQTL-gene connections. Secondly, in our manuscript we focused on cortex-specific eQTL-gene connections. Repeating similar analysis in other brain (e.g. cerebellum⁶⁸) or non-brain regions implicated in ASD may identify additional candidate genes. Finally, our approach focuses on identifying changes in gene regulation that are associated with and may contribute to ASD. We contend that changing the transcription levels of these genes, alters biological pathways that modify an individual's predisposition to develop ASD. However, it is not the only mechanism by which changes in the DNA sequence can contribute to ASD. As such, the genes we identified need not directly overlap those that have been previously curated as being involved in ASD through deletion or mutation studies.

In conclusion, we have identified clinically relevant putative functional impacts for ASD-associated genetic variants within fetal and adult cortical tissues. We have shown that the transcript levels of genes, whose encoded proteins are known to contribute to immune pathways, fatty acid metabolism, ribosome biogenesis, aminoacyl-tRNA biosynthesis and spliceosome are affected in the fetal cortex. In the adult cortex, the known functions of the impacted genes were enriched in immune pathways. Furthermore, despite not having discussed them in detail, there are number of genes whose transcript levels are affected by ASD-eQTLs whose functions were not enriched within known pathways. Future studies of the roles of these genes in ASD will be important for understanding the full impact of ASD-associated genetic variation in the cortex. We contend that our approach represents a valuable strategy to identify potential ASD candidate genes. Moreover, this approach is not tissue or disease specific and is capable of identifying previously unknown tissue-specific contributions to ASD etiology and its interactions with multimorbid conditions. Similar approaches, in combination with existing and future clinical studies of ASD will contribute to individualized mechanistic understanding of ASD etiology in early brain development and adulthood.

Methods

Data access approvals. Data access was approved by the dbGaP (<https://www.ncbi.nlm.nih.gov/gap/>) Data Access Committee(s) for: (1) cortical plate and germinal zone neuron Hi-C datasets (project #16489: "Finishing predictors of cognitive development", accession: phs001190.v1.p1)¹⁹; (2) total RNA-seq and WGS datasets across GTEx v8 tissues (project #22937: "Untangling the genetics of disease multimorbidity", accession: phs000424.v8.p2)²¹; and (3) total RNA-seq and genotyping datasets for fetal brain cortical tissue from 14 to 21 postconceptional weeks (PCWs) (project #25321: "Gene regulatory networks in Autism", accession: phs001900.v1.p1)²⁰ (Supplementary Table 1).

GTEx data processing. Genotypes (derived by Whole Genome Sequencing) were processed by the Genotype-Tissue Expression (GTEx) project, and filtered genotypes (with minor allele frequency ≥ 0.1) for 838 tissue donors were downloaded from dbGaP (<https://www.ncbi.nlm.nih.gov/gap/>, 01/05/2020). RNA-seq data were processed by GTEx using RNA-seq analysis (<https://github.com/broadinstitute/gtex-pipeline/tree/master/rnaseq>) and eQTL discovery (<https://github.com/broadinstitute/gtex-pipeline/tree/master/ctl>) pipelines to calculate normalized gene expression and covariates. The resulting expression (GTEx_Analysis_v8_eQTL_expression_matrices.tar) and covariates (GTEx_Analysis_v8_eQTL_covariates.tar.gz) data were downloaded from GTEx website (<https://www.gtexportal.org/home/datasets>, 01/05/2020).

Fetal RNA-seq data processing. Raw RNA-seq fastq files²⁰ were downloaded from dbGaP (05/06/2020), merged across lanes from the same sample (final dataset of 219 individuals) and analysed using FastQC (v0.11.9; default parameters). FastQC reports were visually inspected and there were no samples that did not pass the quality check (no failures for "Per base sequence quality", "Per sequence quality scores", "Per base N content" and "Sequence Length Distribution" metrics).

All RNA-seq data were processed according to the GTEx pipeline (<https://github.com/broadinstitute/gtex-pipeline/tree/master/rnaseq>) (Supplementary Fig. 1). The same reference genome and annotation files were used to calculate expression for fetal brain RNA-seq data. Briefly, merged fastq files were aligned to the GRCh38 reference genome (Homo_sapiens_assembly38_noALT_noHLA_noDecoy.fasta, <gs://gtex-resources>) using STAR (v2.5.3a). Duplicated mapped reads were marked using Picard MarkDuplicates module (v2.21.4). Quality control metrics and gene-level expression data were calculated using RNA-seQC (v2.3.6) on the basis of GENCODE v26 gene annotation (gencode.v26.GRCh38.genes.gtf, <gs://gtex-resources>). Sample-level gene read and TPM (Transcripts Per kilobase Million) counts were concatenated using combine_GCTs.py.

The GTEx eQTL discovery (<https://github.com/broadinstitute/gtex-pipeline/tree/master/ctl>) pipeline was further used to calculate normalized gene expression and covariates. Briefly, read counts were normalised using the TMM algorithm and genes were selected if they had counts of ≥ 0.1 TPM in $\geq 20\%$ samples and ≥ 6 unnormalized reads in $\geq 20\%$ samples. Genes were inverse normal transformed across samples. Top five genotype principal components (calculated using compute_genotype_pcs.py script from <https://github.com/broadinstitute/gtex-pipeline/tree/master/genotype>), 30 PEER factors, sex and genotyping platform were used as covariates in the eQTL analysis.

Fetal genotype data processing. Genotypes (derived by Array Genotyping) for 219 fetal brain donors²⁰ were downloaded from dbGaP (05/06/2020), processed and prepared in the GTEx format (Supplementary Fig. 1). Briefly, data were preprocessed to correct strand orientation and position of the variants on the GRCh37 reference genome (update_build.sh script). Variants that do not have strand information for HumanOmni25-8v1-2_A1 and HumanOmni2-5Exome-8-v1-1-A genotyping chips were excluded (strand files and update_build.sh script were downloaded from <https://www.well.ox.ac.uk/~wrayner/strand/>, 01/07/2020, Supplementary Table 1).

Genotype data quality control was performed using PLINK (v2.0). Genetic variants were filtered based on Hardy-Weinberg equilibrium $p < 1 \times 10^{-6}$, minor allele frequency 0.01 and variant missing genotype rate 0.05. Within-family IDs were used as sample IDs in the output vcf file (-recode vcf-iid bgz). In total, 663,956 variants passed QC filters. BCFtools (v1.10.2) was used to exclude genetic variants on chromosome 0 (omitted due to mapping to multiple locations) and chromosome 25 (XY pseudoautosomal region), to rename chromosomes 23 (X), 24 (Y) and 26 (MT), to fix REF allele, to check sample's sex and to normalize the output vcf file to the GRCh37 reference genome (human_g1k_v37.fasta.gz, downloaded from <ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/reference/>, 01/07/2020). Normalized vcf files were further validated using VCFtools (v0.1.15).

The Sanger Imputation Service (<https://imputation.sanger.ac.uk/>, 02/07/2020)⁷⁴ was used to: (1) pre-phase the validated genotypes with Eagle (v2.4.1); and (2) to impute them using the 1000 Genomes Phase 3 multi-ethnic reference panel and PBWT algorithm. Imputation accuracy was assessed by filtering variants by info score (INFO < 0.8). Imputed genotypes were filtered for Hardy–Weinberg equilibrium $p < 1 \times 10^{-6}$, variant missing genotype rate 0.05 and minor allele frequency 0.01.

CrossMap (v0.2.6) was used to convert coordinates of genetic variants from genome build hg19 to hg38, resulting in ~54.8 million genetic variants. BCFtools were used to set variant IDs according to the GTEX variant ID format (e.g. chr1_61170_C_T_b38 where chr1 is chromosome name, 61170 is variant position on the chromosome, C is reference allele, T is alternate allele and b38 is genome build 38). The resulting vcf file was converted to plink format and information on sample sex included. To create a lookup table, genetic variants were annotated with rsIDs from dbSNP build 151 database.

eQTL mapping. Genotypes, expression matrices and covariates for fetal and adult brain were integrated into CoDeS3D¹¹ (<https://github.com/Genome3d/codes3d-v2>) pipeline as two separate eQTL databases. Lastly, tensorQTL (<https://github.com/broadinstitute/tensorqtl>) algorithm was used to perform cis- and trans-QTL mapping.

Hi-C data processing. In order to study spatial regulatory interactions in fetal and adult cortical tissues, we analysed two fetal brain-specific (i.e. cortical plate and germinal zone neurons)¹⁹ and one adult cortex-specific (i.e. dorsolateral prefrontal cortex cells)¹² Hi-C chromatin interaction libraries (Supplementary Table 1). Raw Hi-C data were downloaded from dbGaP (accession: phs001190.v1.p1) and GEO (<https://www.ncbi.nlm.nih.gov/geo/>, accession: GSE87112) and analyzed using Juicer (v1.5)⁷⁵ (<https://github.com/aidenlab/juicer>) pipeline to generate Hi-C libraries. The pipeline included BWA (v0.7.15) alignment of paired-end reads onto the hg38 reference genome, merging paired-end read alignments and removing chimeric, unmapped and duplicated reads. The remaining read pairs we refer to as “contacts”. Only Hi-C libraries that contain > 90% alignable unique read pairs, and > 50% unique contacts (< 40% duplication rate) within the total sequenced read pairs were included in the analysis. Files containing cleaned Hi-C contacts locations (i.e. *_merged_nodups.txt files) were processed to obtain Hi-C chromatin interaction libraries in the following format: read name, str1, chr1, pos1, frag1 mapq1, str2, chr2, pos2, frag2, mapq2 (str = strand, chr = chromosome, pos = position, frag = restriction site fragment, mapq = mapping quality score, 1 and 2 correspond to read ends in a pair). Reads where both ends had a mapq ≥ 30 were included in the final library. Hi-C chromatin interactions represent all captured pairs of interacting restriction fragments in the genome. As such, restriction fragments were used to identify regulatory interactions between SNPs and genes (Fig. 1).

Identification of SNPs associated with ASD. Single-nucleotide polymorphisms (SNPs) associated with ASD (n = 454) were downloaded from the GWAS Catalog (www.ebi.ac.uk/gwas/; 05/04/2020; Supplementary Table 2). The GWAS Catalog is a publicly available database of all published GWAS studies⁷⁶. We obtained ASD-associated SNPs from nine GWAS studies that were performed in cohorts containing individuals of mostly European ancestry (Supplementary Table 2). Only SNPs associated with ASD with a $p < 5 \times 10^{-8}$ were included in downstream analyses. Phenotypes were defined as the mapped traits associated with the SNP in the GWAS Catalog.

Identification of spatial regulatory interactions using CoDeS3D. CoDeS3D¹¹ (<https://github.com/Genome3d/codes3d-v2>) was used to identify genes that spatially interact with putative regulatory regions tagged by ASD-associated SNPs (Fig. 1, Supplementary Table 3). Briefly, the human genome reference (hg38) was fragmented at HindIII sites (A/AGCTT), the restriction enzyme that was used in the preparation of the Hi-C libraries. The CoDeS3D algorithm then identified the restriction fragments containing the SNPs within fetal (i.e. cortical plate and germinal zone neurons), or adult (i.e. dorsolateral prefrontal cortex cells) cortex-specific Hi-C libraries. Next, the CoDeS3D algorithm identified the restriction fragments interacting with the SNP-containing restriction fragments. Interacting fragments that overlapped annotated genes (GENCODE transcript model version 26) were subsequently identified (hereafter referred to as SNP-gene pairs). The resulting SNP-gene pairs were then used to query adult cortex, or fetal cortex eQTL databases to identify cis- and trans-acting eQTL-eGene interactions (i.e. genes, whose expression levels are associated with a SNP). Finally, significant cortex-specific eQTL-eGene interactions were identified using the Benjamini-Hochberg (BH) FDR correction to adjust the eQTL p values (FDR < 0.05) (Supplementary Table 3).

Functional annotation of eQTL SNPs associated with ASD. The identified ASD-associated eQTLs were annotated using wANNOVAR tool⁷⁷ (<http://wannovar.wglab.org/>, 10/08/2020) to obtain information about the locus they tagged (Supplementary Table 4). Enrichment of the eQTLs within transcription factor binding sites was determined using SNP2TFBS (<https://ccg.epfl.ch/snp2tfbs/>, 07/09/2020)²⁷. Enrichment within active regulatory elements and histone modification marks was identified using the Roadmap Epigenomics Project 15-state ChromHMM models^{28,78} for adult dorsolateral prefrontal cortex (E073_15_coreMarks_hg38lift_mnemonics.bed.gz) and fetal brain (E081_15_coreMarks_hg38lift_mnemonics.bed.gz) (downloaded from <https://egg2.wustl.edu/roadmap/data/byFileType/chromhmmSegmentations/ChmmModels/coreMarks/jointModel/final/>, 22/11/2020) (Supplementary Table 1). Enrichment analyses were performed using R package regioneR⁷⁹ (permutation test: 1000). Finally, we evaluated identified eQTL SNP associations with other phenotypes in the GWAS Catalog (downloaded on 26/08/2020) (Supplementary Table 4). Phenotypes were defined as the mapped traits associated with the SNP in the GWAS Catalog. Only SNP-phenotype associations with a $p < 5 \times 10^{-8}$ were

included in the analysis. All genomic positions and SNP annotations were obtained for human genome reference build hg38 (GRCh38) release 75.

Construction of ASD-associated PPI networks. The STRING³³ PPI network (version 11.0, protein.links.full.v11.0.txt.gz, <https://string-db.org/>) was downloaded on 24/09/2020. We extracted 2,000,451 protein–protein interactions (with a combined score ≥ 400) between a total of 19,258 unique human proteins (Fig. 1).

Transcriptome-wide fetal and adult cortex-specific PPIs (CSPPIs) were constructed by combining the STRING PPI network with cortex-specific expression data from GTEx v8 or fetal brain datasets (Fig. 1). Ensembl protein (STRING)³³ and transcript identifiers (GTEx and Walker et al. RNA-seq data) were mapped to Ensembl gene identifiers. The CSPPIs represents subnetworks of the STRING PPI network, in which a protein/node is only present if it is expressed in the cortical tissue (adult or fetal). The size of each node depends on the protein expression levels (no missing values and minimum expression level > 0 TPM) in the corresponding cortical tissue. Proteins that were not annotated in the expression datasets were also removed from the CSPPi network. Edges are only present if both interacting proteins are expressed in the cortical tissue. The resulting CSPPi networks contained 1,784,342 PPIs between 17,156 unique proteins in the adult brain, and 1,690,571 PPIs between 16,519 unique proteins in the fetal brain. To build ASD-specific fetal and adult CSPPIs, only interactions between ASD-associated genes we extracted from fetal and adult CSPPIs. The Louvain clustering algorithm⁸⁰ was further applied to identify ASD-specific clusters of functionally related proteins within the CSPPi networks.

Gene ontology enrichment and pathway analyses. Gene Ontology (GO) enrichment and pathways analyses for the eGenes within the ASD-specific CSPPi clusters were performed using the g:GOST module of the g:Profiler tool⁸¹. eGene enrichment was tested within the biological process, molecular function and cellular component GO terms. All annotated human genes were chosen as the statistical domain scope. The significance level was determined using the BH algorithm (FDR < 0.05). The Kyoto Encyclopedia of Genes and Genomes (KEGG) database³⁴ was used to query the most impacted biological pathways.

Loss-of-function analysis. Tolerance to loss-of-function (LoF) variants was measured using the probability-of-being-LoF-intolerant (pLI) method and gene LoF metrics were obtained from gnomAD (v2.1.1, <https://gnomad.broadinstitute.org/>)³². Genes depleted for null variants were defined as having pLI > 0.9 .

Bootstrapping analysis. Bootstrapping analysis ($n = 10,000$ iterations) was performed to test if observed overlaps were non-random. Each bootstrap iteration generated samples of the same size as in the tested sample for tested condition. The number of shared items (e.g. SNPs) among conditions was counted for each bootstrap iteration. After 10,000 iterations we counted those instances where the number of shared items in the bootstrapped overlap is greater than or equal to the number of shared items in the observed overlap. The p value was calculated as the sum of these instances divided by the total number of iterations n ,

$$p \text{ value} = \frac{\sum(\text{bootstrapped} \geq \text{observed})}{n} \quad (1)$$

If the $p < 0.01$ we assume that the observed relationship is non-random.

For SNPs analysis, we resampled SNPs from the list of all GWAS SNPs with $gwas\ p < 5 \times 10^{-8}$. For gene analysis, we resampled eGenes from the list of all genes in the genome (GENCODE transcript model version 26).

Data availability

Python (version 3.6.9), R (version 4.0.2) and RStudio (version 1.2.5033) were used for data analysis and visualisation. All datasets and software used in the analysis are listed in Supplementary Table 1.

Code availability

A Dockerfile (including the CoDeS3D pipeline and downstream analyses), all findings, scripts and reproducibility report are available on github at https://github.com/Genome3d/genetic_regulation_in_ASD.

Received: 13 April 2021; Accepted: 26 July 2021

Published online: 05 August 2021

References

1. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders (DSM-5 (R))* (American Psychiatric Association Publishing, 2013).
2. Lai, M.-C., Lombardo, M. V. & Baron-Cohen, S. Autism. *Lancet* **383**, 896–910 (2014).
3. Grove, J. *et al.* Identification of common genetic risk variants for autism spectrum disorder. *Nat. Genet.* **51**, 431–444 (2019).
4. Robinson, E. B. *et al.* Genetic risk for autism spectrum disorders and neuropsychiatric variation in the general population. *Nat. Genet.* **48**, 552–555 (2016).
5. de la Torre-Ubieta, L., Won, H., Stein, J. L. & Geschwind, D. H. Advancing the understanding of autism disease mechanisms through genetics. *Nat. Med.* **22**, 345–361 (2016).
6. Bourgeron, T. From the genetic architecture to synaptic plasticity in autism spectrum disorder. *Nat. Rev. Neurosci.* **16**, 551–563 (2015).
7. Klei, L. *et al.* Common genetic variants, acting additively, are a major source of risk for autism. *Mol. Autism* **3**, 9 (2012).
8. Gaugler, T. *et al.* Most genetic risk for autism resides with common variation. *Nat. Genet.* **46**, 881–885 (2014).
9. Maurano, M. T. *et al.* Systematic localization of common disease-associate variation in regulatory DNA. *Science* **337**, 1190–1195 (2012).

10. van Arensbergen, J. *et al.* High-throughput identification of human SNPs affecting regulatory element activity. *Nat. Genet.* **51**, 1160–1169 (2019).
11. Fadason, T., Schierding, W., Lumley, T. & O’Sullivan, J. M. Chromatin interactions and expression quantitative trait loci reveal genetic drivers of multimorbidities. *Nat. Commun.* **9**, 5198 (2018).
12. Schmitt, A. D. *et al.* A compendium of chromatin contact maps reveals spatially active regions in the human genome. *Cell Rep.* **17**, 2042–2059 (2016).
13. Li, F., An, Z. & Zhang, Z. The dynamic 3D genome in gametogenesis and early embryonic development. *Cells* **8**, 788 (2019).
14. Van Rooij, D. *et al.* Cortical and subcortical brain morphometry differences between patients with autism spectrum disorder and healthy individuals across the lifespan: Results from the ENIGMA ASD Working Group. *Am. J. Psychiatry* **175**, 359–369 (2018).
15. Stoner, R. *et al.* Patches of disorganization in the neocortex of children with autism. *N. Engl. J. Med.* **370**, 1209–1219 (2014).
16. Willsey, A. J. *et al.* Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. *Cell* **155**, 997–1007 (2013).
17. Mahfouz, A., Ziats, M. N., Rennert, O. M., Lelieveldt, B. P. F. & Reinders, M. J. T. Shared pathways among autism candidate genes determined by co-expression network analysis of the developing human brain transcriptome. *J. Mol. Neurosci.* **57**, 580–594 (2015).
18. Hormozdiari, F., Penn, O., Borenstein, E. & Eichler, E. E. The discovery of integrated gene networks for autism and related disorders. *Genome Res.* **25**, 142–154 (2015).
19. Won, H. *et al.* Chromosome conformation elucidates regulatory relationships in developing human brain. *Nature* **538**, 523–527 (2016).
20. Walker, R. L. *et al.* Genetic control of expression and splicing in developing human brain informs disease mechanisms. *Cell* **179**, 750–771.e22 (2019).
21. Aguet, F. *et al.* The GTEx Consortium atlas of genetic regulatory effects across human tissues. *bioRxiv* <https://doi.org/10.1101/787903> (2019).
22. Ohki-Hamazaki, H. Subchapter 22B—neuromedin B. In *Handbook of Hormones* (eds Takei, Y. *et al.*) 193–e22B-3 (Academic Press, 2016).
23. Malinowska, M., Tokarz-Deptuła, B. & Deptuła, W. Butyrophilins: An important new element of resistance. *Cent. Eur. J. Immunol.* **42**, 399–403 (2017).
24. Writing Committee for the Attention-Deficit/Hyperactivity Disorder *et al.* Virtual histology of cortical thickness and shared neurobiology in 6 psychiatric disorders. *JAMA Psychiat.* **78**, 47–63 (2021).
25. Sunwoo, M. *et al.* Prevalence and outcomes of young people with concurrent autism spectrum disorder and first episode of psychosis. *Schizophr. Res.* **216**, 310–315 (2020).
26. MacRae, C. A. & Vasan, R. S. Next-generation genome-wide association studies: Time to focus on phenotype?. *Circ. Cardiovasc. Genet.* **4**, 334–336 (2011).
27. Kumar, S., Ambrosini, G. & Bucher, P. SNP2TFBS—A database of regulatory SNPs affecting predicted transcription factor binding site affinity. *Nucleic Acids Res.* **45**, D139–D144 (2017).
28. Ernst, J. & Kellis, M. Chromatin-state discovery and genome annotation with ChromHMM. *Nat. Protoc.* **12**, 2478–2492 (2017).
29. Corley, M. & Kroll, K. L. The roles and regulation of Polycomb complexes in neural development. *Cell Tissue Res.* **359**, 65–85 (2015).
30. Desai, D. & Pethe, P. Polycomb repressive complex 1: Regulators of neurogenesis from embryonic to adult stage. *J. Cell. Physiol.* **235**, 4031–4045 (2020).
31. Basu, S. N., Kollu, R. & Banerjee-Basu, S. AutDB: A gene reference resource for autism research. *Nucleic Acids Res.* **37**, D832–D836 (2009).
32. Karczewski, K. J. *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).
33. Szklarczyk, D. *et al.* STRING v11: Protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D607–D613 (2019).
34. Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
35. Chiang, H.-L., Wu, J.-Y. & Chen, Y.-T. Identification of functional single nucleotide polymorphisms in the branchpoint site. *Hum. Genom.* **11**, 27 (2017).
36. Pepper, K. L. *et al.* Autism, early psychosis, and social anxiety disorder: Understanding the role of social cognition and its relationship to disability in young adults with disorders characterized by social impairments. *Transl. Psychiatry* **8**, 233 (2018).
37. Hudson, C. C., Hall, L. & Harkness, K. L. Prevalence of depressive disorders in individuals with autism spectrum disorder: A meta-analysis. *J. Abnorm. Child Psychol.* **47**, 165–175 (2019).
38. Rau, S. *et al.* Identifying comorbid ADHD in autism: Attending to the inattentive presentation. *Res. Autism Spectr. Disord.* **69**, 101468 (2020).
39. Pedapati, E. V. *et al.* Motor cortex facilitation: A marker of attention deficit hyperactivity disorder co-occurrence in autism spectrum disorder. *Transl. Psychiatry* **9**, 298 (2019).
40. Vannucchi, G., Perugi, G. & Masi, G. Bipolar disorder and ASD. In *Psychopathology in Adolescents and Adults with Autism Spectrum Disorders* (ed. Keller, R.) 89–110 (Springer International Publishing, 2019).
41. Gandal, M. J. *et al.* Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science* **362**, eaat8127 (2018).
42. Cross-Disorder Group of the Psychiatric Genomics Consortium *et al.* Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat. Genet.* **45**, 984–994 (2013).
43. Grossniklaus, U. & Paro, R. Transcriptional silencing by polycomb-group proteins. *Cold Spring Harb. Perspect. Biol.* **6**, a019331 (2014).
44. von Schimmelmann, M. *et al.* Polycomb repressive complex 2 (PRC2) silences genes responsible for neurodegeneration. *Nat. Neurosci.* **19**, 1321–1330 (2016).
45. Gentile, C. *et al.* PRC2-associated chromatin contacts in the developing limb reveal a possible mechanism for the atypical role of PRC2 in HoxA gene expression. *Dev. Cell* **50**, 184–196.e4 (2019).
46. Aranda, S., Mas, G. & Di Croce, L. Regulation of gene transcription by Polycomb proteins. *Sci. Adv.* **1**, e1500737 (2015).
47. Gao, Z. *et al.* An AUTS2–Polycomb complex activates gene expression in the CNS. *Nature* **516**, 349–354 (2014).
48. Brown, C. M., Austin, D. W. & Busija, L. Observable essential fatty acid deficiency markers and autism spectrum disorder. *Breastfeed. Rev.* **22**, 21–26 (2014).
49. Usui, N. *et al.* VLDL-specific increases of fatty acids in autism spectrum disorder correlate with social interaction. *EBioMedicine* **58**, 102917 (2020).
50. Tamiji, J. & Crawford, D. A. The neurobiology of lipid metabolism in autism spectrum disorders. *Neurosignals* **18**, 98–112 (2010).
51. Godfrey, K. M. *et al.* Nutritional Intervention Preconception and During Pregnancy to Maintain Healthy Glucose Metabolism and Offspring Health (“NiPPeR”): Study protocol for a randomised controlled trial. *Trials* **18**, 131 (2017).
52. Smith, R. M. & Sadee, W. Synaptic signaling and aberrant RNA splicing in autism spectrum disorders. *Front. Synaptic Neurosci.* **3**, 1 (2011).
53. Thacker, S., Sefyi, M. & Eng, C. Alternative splicing landscape of the neural transcriptome in a cytoplasmic-predominant Pten expression murine model of autism-like Behavior. *Transl. Psychiatry* **10**, 380 (2020).

54. Porokhovnik, L. Individual copy number of ribosomal genes as a factor of mental retardation and autism risk and severity. *Cells* **8**, 1151 (2019).
55. Lombardo, M. V. Ribosomal protein genes in post-mortem cortical tissue and iPSC-derived neural progenitor cells are commonly upregulated in expression in autism. *Mol. Psychiatry* <https://doi.org/10.1038/s41380-020-0773-x> (2020).
56. Tärnlungeanu, D. C. *et al.* Impaired amino acid transport at the blood brain barrier is a cause of autism spectrum disorder. *Cell* **167**, 1481–1494.e18 (2016).
57. Estes, M. L. & McAllister, A. K. Immune mediators in the brain and peripheral tissues in autism spectrum disorder. *Nat. Rev. Neurosci.* **16**, 469–486 (2015).
58. Hughes, H. K., Mills Ko, E., Rose, D. & Ashwood, P. Immune dysfunction and autoimmunity as pathological mechanisms in autism spectrum disorders. *Front. Cell. Neurosci.* **12**, 405 (2018).
59. DiStasio, M. M., Nagakura, I., Nadler, M. J. & Anderson, M. P. T lymphocytes and cytotoxic astrocyte blebs correlate across autism brains. *Ann. Neurol.* **86**, 885–898 (2019).
60. Han, V. X. *et al.* Maternal acute and chronic inflammation in pregnancy is associated with common neurodevelopmental disorders: A systematic review. *Transl. Psychiatry* **11**, 71 (2021).
61. Knuesel, I. *et al.* Maternal immune activation and abnormal brain development across CNS disorders. *Nat. Rev. Neurol.* **10**, 643–660 (2014).
62. Estes, M. L. & McAllister, A. K. Maternal immune activation: Implications for neuropsychiatric disorders. *Science* **353**, 772–777 (2016).
63. Boulanger-Bertolus, J., Pancaro, C. & Mashour, G. A. Increasing role of maternal immune activation in neurodevelopmental disorders. *Front. Behav. Neurosci.* **12**, 230 (2018).
64. Mottron, L. & Bzdok, D. Autism spectrum heterogeneity: Fact or artifact?. *Mol. Psychiatry* **25**, 3178–3185 (2020).
65. Husson, T. *et al.* Rare genetic susceptibility variants assessment in autism spectrum disorder: Detection rate and practical use. *Transl. Psychiatry* **10**, 77 (2020).
66. Brandler, W. M. *et al.* Paternally inherited cis-regulatory structural variants are associated with autism. *Science* **360**, 327–331 (2018).
67. Tam, V. *et al.* Benefits and limitations of genome-wide association studies. *Nat. Rev. Genet.* **20**, 467–484 (2019).
68. Wang, S.S.-H., Kloth, A. D. & Badura, A. The cerebellum, sensitive periods, and autism. *Neuron* **83**, 518–532 (2014).
69. Stiles, J. & Jernigan, T. L. The basics of brain development. *Neuropsychol. Rev.* **20**, 327–348 (2010).
70. Silbereis, J. C., Pochareddy, S., Zhu, Y., Li, M. & Sestan, N. The cellular and molecular landscapes of the developing human central nervous system. *Neuron* **89**, 248–268 (2016).
71. Westra, H.-J. *et al.* Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat. Genet.* **45**, 1238–1243 (2013).
72. Brodie, A., Azaria, J. R. & Ofran, Y. How far from the SNP may the causative genes be?. *Nucleic Acids Res.* **44**, 6046–6054 (2016).
73. Nott, A. *et al.* Brain cell type-specific enhancer-promoter interactome maps and disease-risk association. *Science* **366**, 1134–1139 (2019).
74. McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283 (2016).
75. Durand, N. C. *et al.* Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* **3**, 95–98 (2016).
76. MacArthur, J. *et al.* The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res.* **45**, D896–D901 (2017).
77. Chang, X. & Wang, K. WANNVAR: Annotating genetic variants for personal genomes via the web. *J. Med. Genet.* **49**, 433–436 (2012).
78. Roadmap Epigenomics Consortium *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
79. Gel, B. *et al.* regioneR: An R/Bioconductor package for the association analysis of genomic regions based on permutation tests. *Bioinformatics* **32**, 289–291 (2016).
80. Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech.* **2008**, P10008 (2008).
81. Raudvere, U. *et al.* g:Profiler: A web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res.* **47**, W191–W198 (2019).

Acknowledgements

The authors would like to thank the Genomics and Systems Biology Group (Liggins Institute) for useful discussions. EG is the recipient of a Liggins Ph.D. scholarship and was supported by MBIE Catalyst Grant (The New Zealand–Australia LifeCourse Collaboration on Genes, Environment, Nutrition and Obesity (GENO); UOAX1611). This work was funded by a University of Auckland FRDF Grant (Confirming spatial connections to unravel how SNPs affect phenotype; 3714499) and a MBIE Catalyst grant (The New Zealand–Australia LifeCourse Collaboration on Genes, Environment, Nutrition and Obesity (GENO); UOAX1611) to JOS. JOS was also funded by a Royal Society of New Zealand Marsden Fund [Grant 16-UOO-072]. The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS.

Author contributions

E.G. performed the analyses and wrote the manuscript. T.F. contributed to the eQTL analysis and commented on the manuscript. M.V. and C.W. co-supervised E.G. and commented on the manuscript. T.L. contributed to discussions that aided results interpretation and commented on the manuscript. J.O.S. supervised E.G. and co-wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-95447-z>.

Correspondence and requests for materials should be addressed to J.M.O.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



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

© The Author(s) 2021