2021

Expression of autism spectrum disorder associated genes in non-diseased fetal brain and thymus

BOSTON UNIVERSITY

SCHOOL OF MEDICINE

Thesis

EXPRESSION OF AUTISM SPECTRUM DISORDER ASSOCIATED GENES IN

NON-DISEASED FETAL BRAIN AND THYMUS

by

Chi Vicky Cheng

B.S., University of Washington, 2015
M.S., Boston University, 2021

Submitted in partial fulfillment of the

requirements for the degree of

Master of Science

2021

Approved by

First Reader   _____
                  Maria Isabel Dominguez, Ph.D.
                  Assistant Professor of Medicine

Second Reader   _____
                  Kimberly Aldinger, Ph.D.
                  Senior Research Scientist
                  Seattle Children's Research Institute,
                  Center for Integrative Brain Research

Third Reader   _____
                  Jasmine Plummer, Ph.D.
                  Associate Director
                  Cedars Sinai Cancer,
                  Applied Genomics, Computational & Translational Core

EXPRESSION OF AUTISM SPECTRUM DISORDER ASSOCIATED GENES IN NON-DISEASED FETAL CORTEX AND THYMUS

CHI VICKY CHENG

ABSTRACT

Autism spectrum disorder (ASD) is a highly variable neurodevelopmental disorder. The main hallmarks of individuals with ASD are social communication impairments and repetitive sensory-motor behaviors, and they may present with additional comorbidities such as intellectual disability, epilepsy, anxiety, and/or attention-deficit/hyperactivity disorder.

The underlying cause of ASD is similarly heterogeneous. More than a thousand associated genetic variants including chromosomal abnormalities and *de novo* rare genetic variants have been identified to be associated with ASD. However, there is a general lack of understanding of how genetic variants contribute to the fetal development of ASD.

One interesting idea between the biological mechanism and ASD centers around immune associations and neurodevelopment. Many studies have found that a subset of genes are connected to immune pathways that converge on associated mechanisms of ASD. This work interrogates this idea by examining single cell expression of highly associated ASD genes in non-diseased human fetal thymus and brain. The high resolution of single cell expression highlights potential

associations between specific cell types or pathways to ASD at a time point that is critical for neurodevelopment. Additionally, gene expression analysis has largely been focused on the brain, and this work investigates the thymus, a transient organ responsible for T cell development and a central component of the immune system.

By analyzing highly associated ASD genes in non-diseased tissues of the fetal brain and thymus, the finding that a subset of genes are enriched in thymus tissue substantiates the reason for further interrogation of the possible associations of the thymus and ASD. This analysis also offers a baseline to compare to upon similar analyses of affected tissues of fetal brain and thymus from individuals with ASD.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

ASD…………………………..….…………………………Autism spectrum disorder

cTEC………………………..….…………………………Cortical thymic epithelial cell

CP………………………………………….……………...……Cortical plate

DCN…………………………………….……………………….Deep cerebellar nuclei

DSM-5………………………..Diagnostic and Statistical Manual of Mental Disorders

FGF……………………………………………..………Fibroblast growth factor

GTEx……………………………..….…………………Genotype-Tissue Expression

GW…………………………………………………….Gestational week

IL-7……………………..….…………………………………..Interleukin-7

IP………………………………………..….……………Intermediate progenitor

mTEC…………………………..…….…..……………Medullary thymic epithelial cell

NCC……………………..….……………………..…………Neural crest cell

NKT……………………..….………………………………..Natural killer T

OPC………………………..…….…………………Oligodendrocyte progenitor cell

oRG……………………………………………….……..……Outer radial glia

Ptf1a……………..…………...……..……...Pancreas specific transcription factor, 1a

PCW...………………………………………….……..………Postconception weeks

RG……………………………………………………….…..Radial glia

SATB1………………………..….……………..….……Special AT-rich binding

scRNA…………………..….……………………………...…Single cell RNA

SHH……………………………………………….…………Sonic hedgehog

INTRODUCTION

Autism Spectrum Disorder

Autism spectrum disorder (ASD) affects 1-2% of the global population and is a lifelong disorder of varying prognoses (Vijayakumar & Judy, 2016). According to Masi et al., ASD refers to a heterogeneous group of disorders characterized by two fundamental components -- persistent impairments in social communication and interaction, and restricted, repetitive behavior, interests, or activities. The severity to which these characteristics present varies significantly, and is often interpreted by the degree of specialized medical support an individual diagnosed with ASD requires (Masi et al., 2017). As described in the Diagnostic and Statistical Manual of Mental Disorders (DSM-5), additional psychiatric and cognitive comorbidities commonly accompany ASD, such as intellectual disability, attention-deficit/hyperactivity disorder, social anxiety, and epilepsy (Lord et al., 2018; Masi et al., 2017).

The precise etiology of ASD remains elusive and is as complex as its clinical presentation. Our current understanding of the etiopathology of ASD is multifactorial -- it encompasses environmental and genetic factors that either contribute independently or intersect to result in the observed phenotype.

Environmental Risk Factors

According to a recent study by Lord et al., prenatal, postnatal, and maternal lifestyle factors have been extensively studied as risk factors for ASD. These risk factors are associated with an increased risk of ASD, but it is unclear whether these are direct causes or associations.

Prenatal factors include advanced maternal and paternal age (≥40 years, and ≥50 years, respectively), family history with at least one individual diagnosed with ASD, and maternal diabetes.

Postnatal factors such as preterm birth (<32 weeks), duration and type of delivery, and low birthweight (<1500 g) may also be risk factors for ASD.

Maternal factors such as medication during pregnancy, infection, hypertension, diabetes, and other medical conditions are independently associated with an increased susceptibility to ASD and general developmental delay (Lord et al., 2018).

Genetic Risk Factors

ASD is three to four times more likely to affect males than females, and one of the most heritable neurodevelopmental disorders with estimates ranging from 50-

95% (Vijayakumar & Judy, 2016). Families with one child diagnosed with ASD have an increased recurrence risk of 7-20% in subsequent children, although the precise genetic association may differ between siblings as rare *de novo* variants can occur (Lord et al., 2018; Vijayakumar & Judy, 2016).

In the last decade, the advent of high-throughput genetic testing such as whole exome sequencing, and the development of mouse models to recapitulate ASD, have resulted in the identification of hundreds of ASD-associated risk genes (Grove et al., 2019). Rare *de novo* variants, common allelic variants, and chromosomal rearrangements and deletions have been implicated in both syndromic and non-syndromic ASD (Aldinger et al., 2011). It is theorized that these ASD-associated genetic variants may influence epigenetic regulation, gene-gene and gene-environment interactions, and genetic expression (Masi et al., 2017; Vijayakumar & Judy, 2016). Resources such as SFARI Gene (https://gene.sfari.org) maintain an updated database of reported low to high risk protein disrupting ASD-associated genes, and differentiate between non-syndromic and syndromic ASD-associated variants.

Convergent pathways: The Immune System and ASD

Immunological abnormalities are commonly reported in patients with ASD, suggesting the intersection of pathways between the immune and nervous systems. It is not well known if the immunological dysfunctions reported are a

primary causal mechanism or a secondary symptom of the pathogenesis of ASD. Ziats et al. examined expression of highly associated-ASD genes in unaffected human developing brain, and implicated the connection of NF-κB, Tnf, and Jnk immune signaling pathways, but the data could not substantiate either a primary or secondary relationship (Ziats & Rennert, 2011). Further, transcriptomic analysis from Voineagu et al. on affected post-mortem brain tissue suggest an immune-glial connection of non-genetic etiology to ASD (Voineagu et al., 2011). It is important to note, however, that this analysis was performed on post-mortem tissue, raising the possibility that the inflammation may have been a separate manifestation from ASD.

Upon examination of the literature, it is clear a diverse and vast number of immune impairments in individuals with ASD have been reported. When compared to controls, the ASD group was associated with an imbalance of cytokines in helper CD4+ and cytotoxic CD8+ T cells (Gupta et al., 1998). Other cytokine abnormalities include reports of elevated levels of specific cytokines during pregnancy. Lymphocyte-associated anomalies such as depressed T cell levels including the regulatory T cell (Treg) subtype, incomplete T cell activation, and abnormal natural killer T (NKT) cell activity are notable, and more widely reported than B lymphocyte irregularities. Additionally, anomalies associated with monocytes and microglia support this idea of immune involvement in ASD (Gładysz et al., 2018).

The Cerebral Cortex

The human brain is the most complex organ as the organizer of higher cognitive processes. It is composed of two main anatomical compartments, the cerebrum and cerebellum, and functionally distinct by the billions of neurons and glial cells that span fetal development to adulthood (Hodge et al., 2019; Polioudakis et al., 2019). This cellular diversity is a unique feature of human evolution, and we have only begun to characterize the vast cellular landscape, and delineate the molecular mechanisms that drive neurodevelopment and specialized functions.

The cerebral cortex is composed of sheets of neurons and non-neuronal glial cells that are organized into distinct hierarchical layers. Cortical tissue is further held together by glial cells including astrocytes, oligodendrocytes, and microglia that maintain the hierarchical and specialized nature of cortical tissue, but do not elicit electrical activity. Excitatory projection neurons send their axons away from the cell body into other distinct locations and are the major neuronal cell type, while inhibitory interneurons project their axons in the same structure as the soma (Fan et al., 2018; Masland, 2004; Nowakowski et al., 2017). These neuronal cell types can be further classified by their specialized functions related to individual gene expression signatures. These classifications are not restricted to the layer, but likely influenced by spatiotemporal variations and differences in local microenvironments (Nowakowski et al., 2017; Polioudakis et al., 2019).

Cerebral Development

Cerebral development begins around gestational week (GW) 3 and is established by the end of fetal development, with the frontal lobe as the latest developing area of the neocortex (van Essen et al., 2020). In the embryo, neuroepithelial tissue forms the neural tube that develops into the prosencephalon (forebrain), mesencephalon (midbrain), and rhombencephalon (hindbrain). Neurons originate from the neuroepithelium, differentiate into radial glia (RG), and form the first germinal zone known as the ventricular zone (VZ). As RG divide asymmetrically to self-renew, they form the scaffold upon which newly formed neurons migrate. RG also divide into other progenitor populations known as the outer RG (oRG) and intermediate progenitors (IP), cells that make up the subventricular zone (SVZ). The third major germinal zone, consisting of the subplate (SP) and cortical plate (CP) forms last and the most superficially. Earlier born neurons tend to stay in the deeper laminae, while later born neurons migrate to the superficial layers of the cortex. Each layer is thus composed of a mixture of neural progenitor cells, glial cells, and migratory neurons (Cadwell et al., 2019; Greig et al., 2013; Polioudakis et al., 2019).

As neurogenesis proceeds, the diversity of cell types and functional areas expands, giving rise to the arealization seen in the mature cerebral cortex. This mechanism in which the homogeneous neuroepithelium develops into the lobes of the brain is not well understood. According to Cadwell et al., two main theories

of how neurons become fate-restricted to these area-specific lineages are supported. One hypothesis centers around the idea that progenitors are intrinsically fate-restricted from inception; each progenitor cell is pre-programmed to a committed fate. A second hypothesis proposes that the microenvironment, or extrinsic factors, is what determines cell type and function. Cadwell theorizes that it may be an integration of the two -- a combination of a broadly determined fate at the progenitor stage that is later influenced by extrinsic factors during corticogenesis (Cadwell et al., 2019).

The Cerebellar Cortex

The cerebellar cortex is pivotal for regulating motor control and balance, and more recently, is being recognized for its roles in higher cognitive processes such as emotion and social behavior (Beckinghausen & Sillitoe, 2019; van Essen et al., 2020). The cerebellum is composed of neuronal and non-neuronal glial cell types that reside in three distinct and highly folded layers in the adult brain (Beckinghausen & Sillitoe, 2019). Excitatory cerebellar neurons are different from excitatory cortical neurons, as are cerebellar inhibitory neurons compared to cortical interneurons, in both specialized function and gene expression (van Essen et al., 2020). Glial cells such as the Bergmann glia are also specialized astrocytes unique to the cerebellum (Beckinghausen & Sillitoe, 2019).

The mature cerebellum consists of three superficial layers: the deepest layer is the granular layer, the intermediate layer is the Purkinje cell layer, and the most superficial is the outer molecular layer. These laminae are segregated from the deep cerebellar nuclei in the white matter (Beckinghausen & Sillitoe, 2019; van Essen et al., 2020).

Cerebellar Development

Human cerebellum is compartmentalized to the hindbrain, spanning fetal development and through the second postnatal period. In contrast, cerebral cortical development is formed by the third trimester. The VZ and the upper rhombic lip (uRL) serve as the main germinal centers of stem cells critical for cerebellar neurogenesis and gliogenesis. GABAergic progenitors of the VZ produce Purkinje cells, inhibitory interneurons, and glial subtypes such as astrocytes and Bergmann glia. Glutamergic progenitors in the uRL form granule cells, unipolar brush cell (UBC), and deep cerebellar nuclei (DCN) (van Essen et al., 2020; Vladoiu et al., 2018).

The developmental trajectories of these cerebellar subtypes are thought to be coordinated by a combination of regulatory gene expression and extrinsic signaling factors unique to the region and cell type. For example, in the mouse, sonic hedgehog (Shh) signaling in the VZ is critical for the proliferation of GABAergic progenitors, while pancreas specific transcription factor, 1a (*Ptf1a*)

likely represses the fate switch of these progenitors to granule cells (van Essen et al., 2020). While there are limitations in comparing mice cerebellar development to human cerebellar development, there is evidence in humans of similar spatiotemporal determinants in driving cell type diversity, leading to the complex patterning of the mature cerebellum (Beckinghausen & Sillitoe, 2019; van Essen et al., 2020; Vladoiu et al., 2018).

Neurodevelopment and ASD

Both the cortex and cerebellum are associated with neurodevelopment, psychiatric, and cognitive pathologies. Neurological features in individuals with ASD may include varying degrees of intellectual disability, speech, motor, and developmental delay, epilepsy, hypotonia, spasticity, ataxia, and abnormal brain imaging features (den Hoed et al., 2021; Lord et al., 2018; Masi et al., 2017). Syndromic ASD is associated with other neurological disorders, while nonsyndromic ASD is not.

Thymus

The thymus is a primary lymphoid organ that is established by birth, and involutes at puberty and is replaced by fatty tissue. This transient organ is an integral component of the immune system as it aids in T cell differentiation and maturation. When infants are born with an improperly formed thymus, there is an increased risk of life-threatening infections in infancy and adolescence due to the

reduced numbers of functional T cells (Thapa & Farber, 2019). Mature T cells vary in function and expression of the T cell receptor (TCR) that allows them to recognize specific antigens as part of the adaptive immune system (Gordon & Manley, 2011; Kernfeld et al., 2018; Thapa & Farber, 2019). The thymus also supports the innate immune system with the development of NKT cells. Although NKT cells lack TCR diversity, they respond more quickly than other mature T cells to invading pathogens.

The thymus is anatomically organized into cortical and medullary regions, and each compartment houses specific cell types such as thymic epithelial cells (TEC), endothelial cells, macrophages, and dendritic cells. Cortical TEC (cTEC) aid in the positive selection process in which naive T cells are selected for the ability to recognize major histocompatibility complexes (MHC) I and II, and medullary TEC (mTEC) regulate the negative selection of T lymphocytes that are autoreactive. Thus, this process results in the production of the majority of the population of thymus-dependent T cells by birth (Gordon & Manley, 2011; Kernfeld et al., 2018; Thapa & Farber, 2019).

Thymus Development
The thymic epithelium is derived from endodermal origins in the pharyngeal pouches. It is thought that ectodermal-originating neural crest cells (NCC) have a role in the detachment from the pharynx, subsequent migration of the thymus to

its final location in the body, and interact with thymic epithelial cells in early organogenesis (Gordon & Manley, 2011). Additionally, NCCs contribute to the formation of mesenchymal capsule and thymic vasculature as a precursor to pericytes and smooth muscle cells (Figueiredo et al., 2020). As thymus organogenesis continues, this complex network of vasculature interwoven with the stroma serves as the pathway in which lymphocytes migrating from bone-marrow precursors enter the cortex and undergo positive selection, pass through the medulla for negative selection, and exit upon differentiation (Thapa & Farber, 2019).

A number of signaling and regulatory pathways have been implicated in thymus organogenesis. The role of signaling pathways such as fibroblast growth factor (FGF), Notch, and Wingless-Int (Wnt) in morphogenesis have been studied. Transcription factors such as the TEC-specific Foxn1 are thought to contribute to organogenesis because mice lacking Foxn1 form a thymic primordium, but not a mature functional thymus. Finally, cytokines such as interleukin-7 (IL-7) are another early marker of thymic-restricted cells; IL-7 is produced by stromal cells and necessary for thymocyte development and survival (Figueiredo et al., 2020).

Thymus and ASD

A definitive relationship between the thymus and ASD is yet to be established. The thymus itself is vastly understudied as well as the thorough systematic

review of organs other than the brain in individuals with ASD. However,

immunological dysfunction associated with ASD is well documented and

constitutes a significant portion of abnormalities reported in individuals (Gładysz

et al., 2018). This substantiates the need for interrogation of the possible

relationship between the thymus and ASD.

SATB1 expression in the thymus

A potential link between the thymus and ASD is the well-known chromatin

organizer and transcription factor special AT-rich binding protein *(SATB1)*.

SATB1 encodes a nuclear protein essential for thymocyte development

(Balamotis et al., 2012; Kakugawa et al., 2017; Nüssing et al., 2019). SATB1

regulates expression of lineage-specific genes driving T lymphocyte

differentiation primarily in positive selection, including CD4, CD8, and Treg factor

Foxp3, a repressor of SATB1. This function aligns with the BBI fetal data set as

thymocytes have the highest SATB1 expression (Figure 2) and with the Park et

al. data set as SATB1 is enriched in aBT(entry) cells, the transient stage

preceding commitment to either CD4+ or CD8+ lineage (Figure 7). The role of

the *SATB1* in T cell development is further supported by the finding that *Satb1*-

mutant mice have abnormal T cell populations, lacking NKT and Treg subtypes

at specific developmental time points (Kakugawa et al., 2017). Interestingly, an

imbalance of helper T cells to Treg cells was described in a study of 44 children

diagnosed with varying severities of ASD. The ratio of helper T cells to Treg cells was significantly higher in children with ASD than controls (Moaaz et al., 2019).

Postnatally, the expression of SATB1 is attenuated in peripheral mature T cell subtypes when compared to the higher expression levels in immature T lymphocytes (Nüssing et al., 2019). This further supports the spatiotemporal role of *SATB1* in T cell development.

Single Cell RNA Sequencing

Single cell RNA (scRNA) sequencing is a robust and high-throughput approach that contrasts gene sequencing of bulk tissues. Sequencing of tissues offers a high level overview of genetic information, while scRNA sequencing renders genomics to a finer granularity that better captures biological complexity. Single cell transcriptomics offers an opportunity to trace developmental trajectories, capture transient expression, and gain insights into gene regulation and function. However, single cell transcriptomics can be limited by sparse data as it is highly dependent on each cell expressing a specific gene at a given time.

In scRNA studies, cell-specific barcodes bind to each cell or nuclei for reverse transcription and cDNA fragments are amplified to create libraries. These libraries are sequenced and data analysis is performed a number of ways. Generally, the data is mapped to the genome, filtered, and quality-controlled

before normalization and further analysis (Griffiths et al., 2018; Paolillo et al., 2019). R packages such as Seurat (satijalab.org/seurat) and Monocle (cole-trapnell-lab.github.io/monocle3/) are widely used toolkits to perform cluster, differential gene, and pseudotime analysis.

METHODS

BBI Single Cell Atlas

The BBI Single Cell Atlas is a publicly available reference of single cell sequencing of 4 million cells from 15 non-diseased human fetal organs, spanning about 10 to 18 postconceptional weeks (PCW). Cell subtypes were broadly classified based on cross-examination with mouse cell atlases. Approximately 1.7 million cells from the cerebrum, 1.09 cells from the cerebellum, and 8,700 cells from the thymus were sequenced (Cao et al., 2020). Single cell expression analysis was performed on the cerebrum, cerebellum, and thymus.

Three normalized gene expression data sets of transcripts per million (TPM) were accessed. TPM across all 15 fetal tissues, thymus-, and brain-specific cell types were each separately downloaded (descartes.brotman bay.org, accessed 2020-11-14).

SFARI Gene

The SFARI Gene database is a public and regularly updated resource of genetic risk factors of ASD (gene.sfari.org/). Genes are scored as syndromic (S), high confidence (1), strong candidate (2), or suggestive evidence (3) based on a set of criteria that examines the strength of supporting evidence (gene.sfari.org/about-gene-scoring/).

SFARI Gene was accessed to generate a list of monogenic and highly associated ASD genes. Genes classified as "Rare Single Gene Mutation" with a gene score of 1 were identified, for a total of 190 genes (Table 1, accessed 2020-11-14).

Genotype-Tissue Expression (GTEx)

Genotype-Tissue Expression (GTEx) was accessed for *SATB1* expression across tissues, on a logarithmic scale and set to highest to lowest expression (accessed 2021-01-29, Figure 6).

GTEx is a public resource of gene expression data collected from 838 individuals across 49 different tissues. These individuals, ranging 20 to 70 years old, experienced varying causes of death, from traumatic injury to heart disease. The tissues were collected from non-diseased sites and RNA sequencing was performed using two methods: Illumina TrueSeq RNA sequencing and Affymetrix Human Gene 1.1 ST Expression Array (gtexportal.org).

Gene expression

Several expression heatmaps were constructed using the R package pheatmap (version 1.0.12) using two main approaches. One approach involved comparing the filtered list of monogenic and highly associated ASD genes to all 15 fetal tissues, thymus-, and brain-specific cell types. The second approach interrogated

the expression of the highest expressing genes based on normalized TPM

counts in the thymic- and brain-cell specific data sets.

RESULTS

Expression of highly associated ASD genes

The expression of highly associated and rare genetic variants of ASD across 15 fetal tissues with a minimum of 40 TPM clusters broadly in the brain and the thymus (Figure 1). Interestingly, three distinct groups of genes with varying expression levels separately cluster in the thymus. Genes expressed in the cerebrum, cerebellum, and eye also cluster together. Across tissues, thymus and liver cluster with placenta, and the group of highest expressing genes in the thymus group cluster with the cerebrum, cerebellum, and eye group.

The expression of highly associated ASD genes across thymic-specific cells with a minimum of 50 TPM depicts expression at a finer resolution than when examined across all 15 fetal tissue types (Figure 2). A threshold of 50 TPM was arbitrarily selected to better capture the top expressing genes. Across cell types, thymocytes and antigen presenting cells cluster, while vascular endothelial cells and thymic epithelial cells are grouped together.

The expression of highly associated ASD genes across cerebral- and cerebellar-specific cells with a minimum of 60 TPM depicts expression at a finer resolution than when examined across all 15 fetal tissue types (Figure 3). A threshold of 60 TPM was arbitrarily selected to better capture the top expressing genes, especially as these ASD genes are highly associated and known to be reported

in brain-specific cell types. In the cerebral-specific heatmap, microglia and endothelial cells are grouped, megakaryocytes are not grouped with another specific cell type, glial cells (astrocytes and oligodendrocytes) cluster with neuronal cell types (excitatory, inhibitory, limbic, and SKOR2+ NPRS1+ cells). In the cerebellar-specific heatmap, microglia and endothelial cells are similarly grouped, and no megakaryocytes are annotated as a cell type. Astrocytes cluster with granule neurons and UBC, while oligodendrocytes cluster with Purkinje cells and SLC24A4+ PEX5L+ cells.
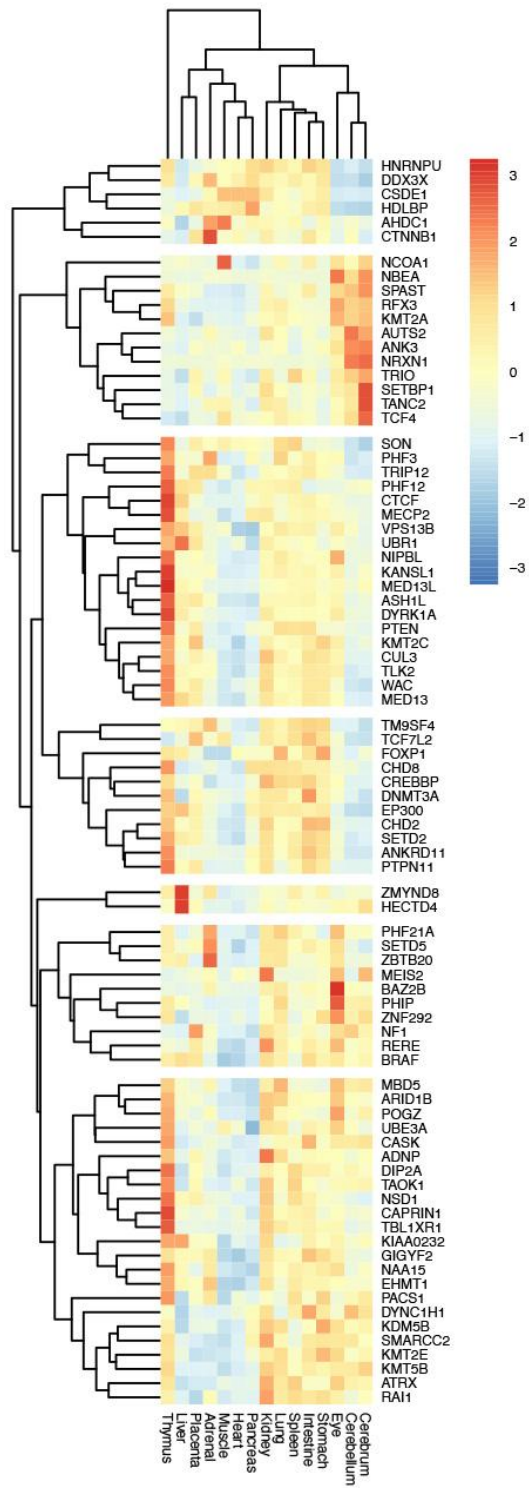
Figure 1. ASD associated gene expression across 15 fetal tissues.
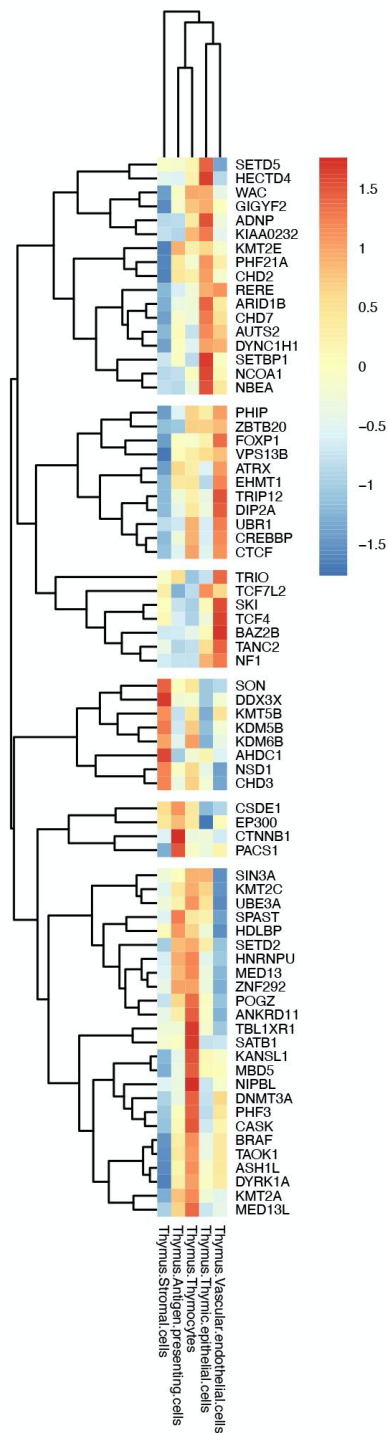Expression of highly associated rare genetic variants with at least 40 TPM.

Figure 2. ASD associated gene expression across thymic-specific cell types. Expression of highly associated rare genetic variants with at least 50 TPM.
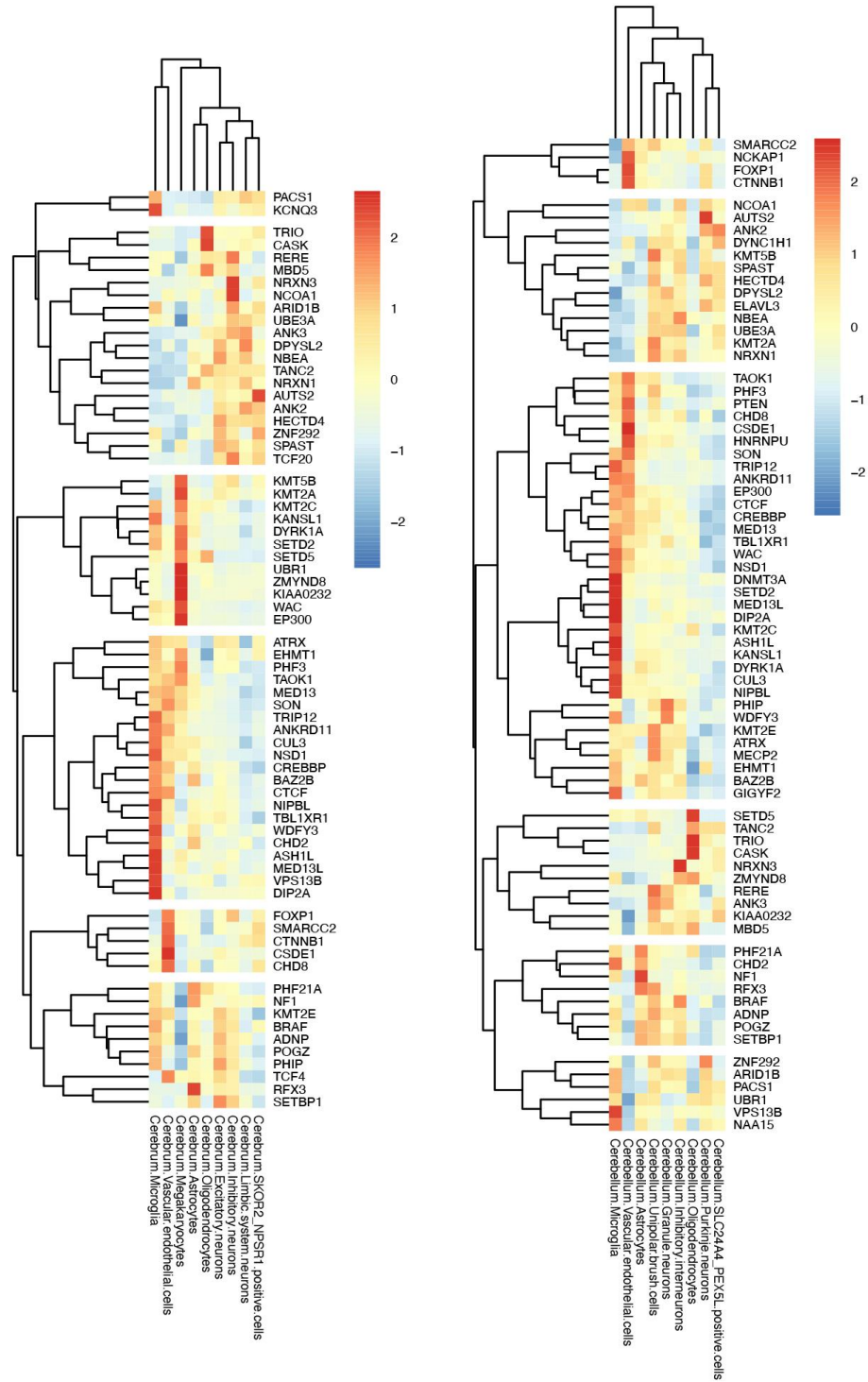
Figure 3. ASD associated gene expression in cerebral and cerebellar-specific cell types. Expression of highly associated rare genetic variants with at

least 60 TPM in cerebral- (3a, left) and cerebellar-specific cell subtypes (3b, right).

Highest expressing genes in thymic- and cerebral-specific cell types

An arbitrary threshold of 350 TPM per gene was set to examine the expression of

top genes in thymic- (Figure 4a) and cerebral-specific data sets (Figure 4b). In

the thymus, stromal cells separately cluster with the thymic epithelial and

vascular endothelial group, and the antigen presenting and thymocyte group.

Broadly speaking, the cell types with the highest expressing genes are stromal

cells and thymocytes. In the cerebrum, a slightly different clustering pattern

arises when compared to the expression of ASD genes in the cerebrum (Figure

4a). Vascular endothelial cells are initially grouped with astrocytes and then with

megakaryocytes. The neuronal subtypes cluster amongst one another, and then

with microglia, followed by oligodendrocytes.

Genes such as *AUTS2*, *MACF1*, and *FTX* are highly expressed in both thymus

and cerebral tissues (Figure 4a and b).
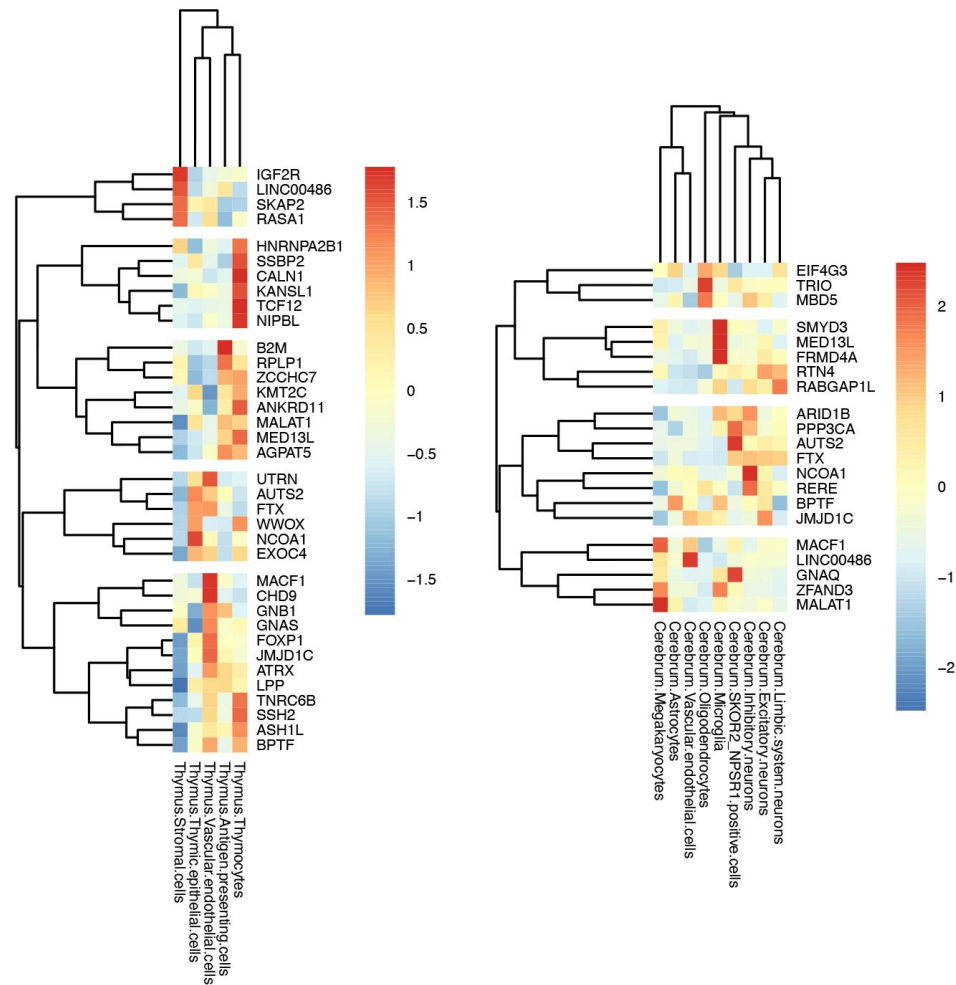
Figure 4a and b. Highly expressed genes of the thymus and cerebrum.
Expression of highest expressing genes (>350 TPM) in thymic- (4a, left) and
cerebral-specific (4b, right) cell subtypes.

SATB1 expression in the fetal brain

During fetal brain development, *SATB1* is a critical chromatin regulator of genes

involved in neurodevelopment (Satterstrom et al., 2020). In the BBI fetal data set,

oligodendrocytes are observed to have the highest SATB1 expression in both

cerebral and cerebellar tissues, with enrichment also observed in

megakaryocytes, astrocytes, inhibitory neurons, and excitatory neurons (descartes.brotman bay.org). In the Nowakowski et al. scRNA sequencing study performed on normal fetal human cortical and medial ganglionic eminence samples spanning 5 to 37 PCW, SATB1 is highly expressed in oligodendrocyte progenitor cells (OPC), newborn to mature excitatory and inhibitory neurons during the critical period for neurogenesis (Figure 5).

Figure 5. Uniform Manifold Approximation and Projection (UMAP) of SATB1 expression in Nowakowski et al. scRNA sequencing data set of primary cortical and MGE tissue. Dark red denotes high expression, while light blue indicates low to zero expression of SATB1 (cells.ucsc.edu/?ds=cortex-dev). CGE, caudal ganglionic eminence; CTX, cortex; div, dividing; EN, excitatory neurons; nEN, newborn EN; IN, interneuron; IPC, intermediate progenitor cell;

RG, radial glia; oRG, outer radial glia; vRG, ventral radial glia, OPC, oligodendrocyte progenitor cell; PFC, prefrontal cortex; MGE, medial ganglionic eminence; V1, primary visual cortex

The importance of *SATB1* in neurodevelopment is further corroborated by findings in a mouse study conducted by Balamotis et al. In the postnatal brain of mice, they observed that Satb1 is highly expressed in mature neurons of the cerebral cortex, dentate hilus, and amygdala. *Satb1*-null postnatal mice had normal cortical patterning and cell morphology at each level, but cortical pyramidal neurons had significantly decreased dendritic density. Dendritic branching complexity and shape were unaffected, suggesting the role of Satb1 in postsynaptic transmission and plasticity (Balamotis et al., 2012).

SATB1 expression in the adult brain

In non-diseased tissues collected from adult tissues, SATB1 shows the highest expression in tissues collected from the frontal cortex, bladder, colon, and tibial artery (Figure 6). This bulk tissue observation aligns with the dynamic nature of SATB1 expression from fetal development to adulthood.
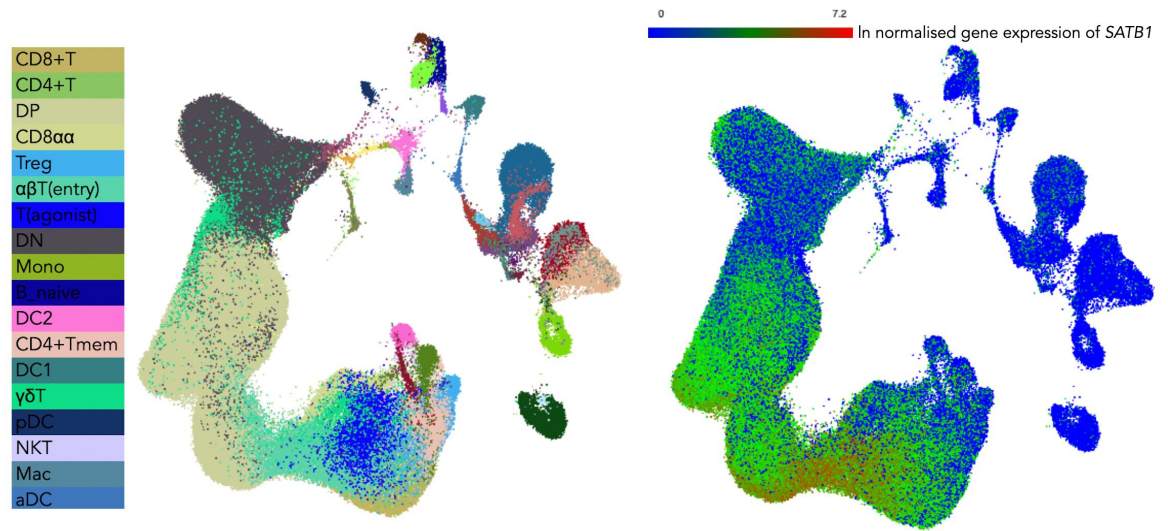
Gene expression for SATB1 (ENSG00000182568.16)

Figure 6. SATB1 expression evaluated in GTEx database. Log(TPM) vs tissue type sorted from highest (left) to lowest (right).

SATB1 expression in the thymus

In the fetal thymus cell atlas reported by Park et al., aBT(entry) cells have the highest SATB1 expression (Figure 7). aBT(entry) cells are defined as immature and transient T lymphocytes that have yet to commit to a cell lineage (Park et al., 2020)**.**

Figure 7. UMAP of SATB1 expression evaluated in Park et al. scRNA sequencing data set (developmentcellatlas.ncl.ac.uk/datasets/ HCA_thymus). Cell types are annotated on the left, while normalized SATB1 expression is depicted on the right.

DISCUSSION

ASD associated gene expression in unaffected fetal tissues

The expression of highly associated and rare genetic variants of ASD in 15 unaffected fetal tissues from different organs broadly clustered in two main groups (Figure 1). Group 1 included the cerebrum, cerebellum, and eye, while group 2 represents the thymus. Clustering suggests a relationship between the subset of genes and the tissue type, and highly expressed genes in the cluster are potential genes of interest for interrogation of how and why they are enriched. The enrichment in Group 1 is consistent with the understanding of ASD as a neurodevelopmental disorder that affects the brain. The enrichment in Group 2, however, is surprising as there is limited knowledge as to the potential relationship between the thymus and ASD.

ASD associated gene expression in unaffected fetal thymic cell types

The expression of highly associated genes of ASD in thymic stromal cells, antigen presenting cells, thymocytes, epithelial cells, and vascular endothelial cells are not enriched in one specific cell type (Figure 2).

The relationship of *SATB1* and ASD

Large genetic studies of neurodevelopmental disorders implicate rare de novo *SATB1* variants as likely pathogenic in the development of ASD (den Hoed et al., 2021; Kaplanis et al., 2020; Satterstrom et al., 2020). The mechanism of variants

in *SATB1* leading to ASD focuses on, but is not limited to, improper

synaptogenesis, abnormal cellular migration, or dysregulation of other neuronal

genes (Balamotis et al., 2012; Satterstrom et al., 2020; Voineagu et al., 2011)**.**

Despite the critical role of *SATB1* in T cell development, however, the

dysregulation of this pathway as a contributor to ASD development has yet to be

interrogated.

Highest expressing genes in thymic- and cerebral-specific cell types

The agnostic evaluation of the highest expressing genes in the thymus and

cerebral tissue warrants further investigation. Activator of transcription and

developmental regulator, *AUTS2*, associated with intellectual disability, and

Microtubule-actin cross-linking factor 1, *MACF1*, associated with lissencephaly

with complex brainstem malformation (omim.org), are both highly expressed in

thymus tissue and genes that are known to be associated with

neurodevelopmental disorders (Figure 4a).

Of the thymocyte group, *KANSL1* and *NIPBL* are both classified as rare and

high-confidence ASD genes (sfari.org). Variants in *HNRNPA2B1* are associated

with Inclusion body myopathy with early-onset Paget disease with or without

frontotemporal dementia, and variants in *TCF12* are associated with

Craniosynostosis type 3 (omim.org). Lastly, genes *SSBP2* and *CALN1* are not

currently known to be associated with clinical phenotypes, but the high

expression signifies a point of future interrogation.

CONCLUSION

Implications

ASD is a complex and heterogeneous neurodevelopmental disorder wherein the brain has been the main tissue of research in understanding the mechanism driving the disease course. However, individuals diagnosed with ASD experience also have non-neurological phenotypic features such as cardiac, skeletal, gastrointestinal, and immune abnormalities (den Hoed et al., 2021; Masi et al., 2017; Vijayakumar & Judy, 2016)**.** Whether these are primary causal or secondary phenotypic features of ASD remains unclear, but the neurodevelopmental features of ASD, and thus the brain, have been the main focus of research in order to elucidate the neurobiological mechanism. This has allowed for a deeper interrogation of neurodevelopment in ASD, but has also limited the scope of understanding ASD as a systematic disorder. This work expands the scope by evaluating the expression of highly associated ASD genes across a variety of fetal tissues. As a result, a group of ASD-associated genes clustered in the thymus, an understudied tissue in ASD. The thymus warrants further interrogation as many studies support a convergence of pathways leading to ASD, including those that are immune-associated.

SATB1 is critical for both proper T cell differentiation and neurodevelopment. Expression of SATB1 varies temporally and spatially on a single cell level, a complexity that aligns with the heterogeneity of clinical features among

individuals with variants in *SATB1* (den Hoed et al., 2021)**.** In the BBI fetal data set of the thymus, thymocytes have the highest expression of *SATB1*. This observation is consistent with evaluations of *SATB1* in the Park et al. scRNA sequencing data set of human fetal thymic tissue.

In the BBI fetal data set of the cerebral cortex, megakaryocytes, oligodendrocytes, astrocytes, and excitatory and inhibitory cell types have the highest SATB1 expression. This resembles expression in the Nowakowski et al. single cell data set, although megakaryocytes or astrocytes are annotated in the study.

Next Steps

The BBI fetal data set can be reclassified by specifying the cell types of thymus, cerebral, and cerebellum based on annotations from data sets such as those from Park et al. and Nowakowski et al. In doing so, genes of interest can be interrogated at a higher spatial resolution. Data sets from Park et al. and Nowakowski et al. should also be accessed to increase the power of the BBI fetal data set. Aggregating multiple sources of non-diseased scRNA sequencing data will decrease variability of the expression of the genes of interest, and can also help to further define expression at specific time points.

The highest expressing genes of the thymus can undergo enrichment analysis in order to elucidate the relevant biological pathways. Little is known about the

expression of neurodevelopmental-associated genes such as *AUTS2* during fetal thymic development.

Lastly, investigating scRNA data collected from diseased tissue types from individuals diagnosed with ASD would provide the opportunity to assess the differences between control and affected groups. This analysis can elucidate the various biological pathways that are affected in the disease progression of ASD.

Bibliography

Aldinger, K. A., Plummer, J. T., Qiu, S., & Levitt, P. (2011). SnapShot: Genetics of

Autism. *Neuron*, *72*(2), 418-418.e1. https://doi.org/10.1016/j.neuron.2011.10.007

Balamotis, M. A., Tamberg, N., Woo, Y. J., Li, J., Davy, B., Kohwi-Shigematsu, T., &

Kohwi, Y. (2012). Satb1 Ablation Alters Temporal Expression of Immediate Early

Genes and Reduces Dendritic Spine Density during Postnatal Brain

Development. *Molecular and Cellular Biology*, *32*(2), 333–347.

https://doi.org/10.1128/MCB.05917-11

Beckinghausen, J., & Sillitoe, R. V. (2019). Insights into cerebellar development and

connectivity. *Neuroscience Letters*, *688*, 2–13.

https://doi.org/10.1016/j.neulet.2018.05.013

Cadwell, C. R., Bhaduri, A., Mostajo-Radji, M. A., Keefe, M. G., & Nowakowski, T. J.

(2019). Development and Arealization of the Cerebral Cortex. *Neuron*, *103*(6),

980–1004. https://doi.org/10.1016/j.neuron.2019.07.009

Cao, J., O'Day, D. R., Pliner, H. A., Kingsley, P. D., Deng, M., Daza, R. M., Zager, M. A.,

Aldinger, K. A., Blecher-Gonen, R., Zhang, F., Spielmann, M., Palis, J., Doherty,

D., Steemers, F. J., Glass, I. A., Trapnell, C., & Shendure, J. (2020). A human

cell atlas of fetal gene expression. *Science*, *370*(6518), eaba7721.

https://doi.org/10.1126/science.aba7721

den Hoed, J., de Boer, E., Voisin, N., Dingemans, A. J. M., Guex, N., Wiel, L., Nellaker,

C., Amudhavalli, S. M., Banka, S., Bena, F. S., Ben-Zeev, B., Bonagura, V. R.,

Bruel, A.-L., Brunet, T., Brunner, H. G., Chew, H. B., Chrast, J., Cimbalistienė, L.,

Coon, H., … Vissers, L. E. L. M. (2021). Mutation-specific pathophysiological

mechanisms define different neurodevelopmental disorders associated with

SATB1 dysfunction. *The American Journal of Human Genetics*, *108*(2), 346–356. https://doi.org/10.1016/j.ajhg.2021.01.007

Fan, X., Dong, J., Zhong, S., Wei, Y., Wu, Q., Yan, L., Yong, J., Sun, L., Wang, X., Zhao, Y., Wang, W., Yan, J., Wang, X., Qiao, J., & Tang, F. (2018). Spatial transcriptomic survey of human embryonic cerebral cortex by single-cell RNA-seq analysis. *Cell Research*, *28*(7), 730–745. https://doi.org/10.1038/s41422-018-0053-3

Figueiredo, M., Zilhão, R., & Neves, H. (2020). Thymus Inception: Molecular Network in the Early Stages of Thymus Organogenesis. *International Journal of Molecular Sciences*, *21*(16), 5765. https://doi.org/10.3390/ijms21165765

Gładysz, D., Krzywdzińska, A., & Hozyasz, K. K. (2018). Immune Abnormalities in Autism Spectrum Disorder—Could They Hold Promise for Causative Treatment? *Molecular Neurobiology*, *55*(8), 6387–6435. https://doi.org/10.1007/s12035-017-0822-x

Gordon, J., & Manley, N. R. (2011). Mechanisms of thymus organogenesis and morphogenesis. *Development*, *138*(18), 3865–3878. https://doi.org/10.1242/dev.059998

Greig, L. C., Woodworth, M. B., Galazo, M. J., Padmanabhan, H., & Macklis, J. D. (2013). Molecular logic of neocortical projection neuron specification, development and diversity. *Nature Reviews Neuroscience*, *14*(11), 755–769. https://doi.org/10.1038/nrn3586

Griffiths, J. A., Scialdone, A., & Marioni, J. C. (2018). Using single-cell genomics to understand developmental processes and cell fate decisions. *Molecular Systems Biology*, *14*(4). https://doi.org/10.15252/msb.20178046

Grove, J., Ripke, S., Als, T. D., Mattheisen, M., Walters, R. K., Won, H., Pallesen, J.,

    Agerbo, E., Andreassen, O. A., Anney, R., Awashti, S., Belliveau, R., Bettella, F.,

    Buxbaum, J. D., Bybjerg-Grauholm, J., Bækvad-Hansen, M., Cerrato, F.,

    Chambert, K., Christensen, J. H., … Børglum, A. D. (2019). Identification of

    common genetic risk variants for autism spectrum disorder. *Nature Genetics*,

    51(3), 431–444. https://doi.org/10.1038/s41588-019-0344-8

Gupta, S., Aggarwal, S., Rashanravan, B., & Lee, T. (1998). *Th1- and Th2-like cytokines*

    in CD4q and CD8q T cells in autism. 4.

Hodge, R. D., Bakken, T. E., Miller, J. A., Smith, K. A., Barkan, E. R., Graybuck, L. T.,

    Close, J. L., Long, B., Johansen, N., Penn, O., Yao, Z., Eggermont, J., Höllt, T.,

    Levi, B. P., Shehata, S. I., Aevermann, B., Beller, A., Bertagnolli, D., Brouner,

    K., … Lein, E. S. (2019). Conserved cell types with divergent features in human

    versus mouse cortex. *Nature*, *573*(7772), 61–68. https://doi.org/10.1038/s41586-

    019-1506-7

Kakugawa, K., Kojo, S., Tanaka, H., Seo, W., Endo, T. A., Kitagawa, Y., Muroi, S.,

    Tenno, M., Yasmin, N., Kohwi, Y., Sakaguchi, S., Kowhi-Shigematsu, T., &

    Taniuchi, I. (2017). Essential Roles of SATB1 in Specifying T Lymphocyte

    Subsets. *Cell Reports*, *19*(6), 1176–1188.

    https://doi.org/10.1016/j.celrep.2017.04.038

Kaplanis, J., Samocha, K. E., Wiel, L., Zhang, Z., Arvai, K. J., Eberhardt, R. Y., Gallone,

    G., Lelieveld, S. H., Martin, H. C., McRae, J. F., Short, P. J., Torene, R. I., de

    Boer, E., Danecek, P., Gardner, E. J., Huang, N., Lord, J., Martincorena, I.,

    Pfundt, R., … Retterer, K. (2020). Evidence for 28 genetic disorders discovered

    by combining healthcare and research data. *Nature*, *586*(7831), 757–762.

https://doi.org/10.1038/s41586-020-2832-5

Kernfeld, E. M., Genga, R. M. J., Neherin, K., Magaletta, M. E., Xu, P., & Maehr, R.

(2018). A Single-Cell Transcriptomic Atlas of Thymus Organogenesis Resolves

Cell Types and Developmental Maturation. *Immunity*, *48*(6), 1258-1270.e6.

https://doi.org/10.1016/j.immuni.2018.04.015

Lord, C., Elsabbagh, M., Baird, G., & Veenstra-Vanderweele, J. (2018). Autism spectrum

disorder. *The Lancet*, *392*(10146), 508–520. https://doi.org/10.1016/S0140-

6736(18)31129-2

Masi, A., DeMayo, M. M., Glozier, N., & Guastella, A. J. (2017). An Overview of Autism

Spectrum Disorder, Heterogeneity and Treatment Options. *Neuroscience

Bulletin*, *33*(2), 183–193. https://doi.org/10.1007/s12264-017-0100-y

Masland, R. H. (2004). Neuronal cell types. *Current Biology*, *14*(13), R497–R500.

https://doi.org/10.1016/j.cub.2004.06.035

Moaaz, M., Youssry, S., Elfatatry, A., & El Rahman, M. A. (2019). Th17/Treg cells

imbalance and their related cytokines (IL-17, IL-10 and TGF-β) in children with

autism spectrum disorder. *Journal of Neuroimmunology*, *337*, 577071.

https://doi.org/10.1016/j.jneuroim.2019.577071

Nowakowski, T. J., Bhaduri, A., Pollen, A. A., Alvarado, B., Mostajo-Radji, M. A., Di

Lullo, E., Haeussler, M., Sandoval-Espinosa, C., Liu, S. J., Velmeshev, D.,

Ounadjela, J. R., Shuga, J., Wang, X., Lim, D. A., West, J. A., Leyrat, A. A., Kent,

W. J., & Kriegstein, A. R. (2017). Spatiotemporal gene expression trajectories

reveal developmental hierarchies of the human cortex. *Science*, *358*(6368),

1318–1323. https://doi.org/10.1126/science.aap8809

Nüssing, S., Koay, H., Sant, S., Loudovaris, T., Mannering, S. I., Lappas, M., d′Udekem,

Y., Konstantinov, I. E., Berzins, S. P., Rimmelzwaan, G. F., Turner, S. J.,

Clemens, E. B., Godfrey, D. I., Nguyen, T. H., & Kedzierska, K. (2019). Divergent

SATB1 expression across human life span and tissue compartments.

Immunology & Cell Biology, *97*(5), 498–511. https://doi.org/10.1111/imcb.12233

Paolillo, C., Londin, E., & Fortina, P. (2019). Single-Cell Genomics. *Clinical Chemistry*,

65(8), 972–985. https://doi.org/10.1373/clinchem.2017.283895

Park, J.-E., Botting, R. A., Domínguez Conde, C., Popescu, D.-M., Lavaert, M., Kunz, D.

J., Goh, I., Stephenson, E., Ragazzini, R., Tuck, E., Wilbrey-Clark, A., Roberts,

K., Kedlian, V. R., Ferdinand, J. R., He, X., Webb, S., Maunder, D., Vandamme,

N., Mahbubani, K. T., … Teichmann, S. A. (2020). A cell atlas of human thymic

development defines T cell repertoire formation. *Science*, *367*(6480), eaay3224.

https://doi.org/10.1126/science.aay3224

Polioudakis, D., de la Torre-Ubieta, L., Langerman, J., Elkins, A. G., Shi, X., Stein, J. L.,

Vuong, C. K., Nichterwitz, S., Gevorgian, M., Opland, C. K., Lu, D., Connell, W.,

Ruzzo, E. K., Lowe, J. K., Hadzic, T., Hinz, F. I., Sabri, S., Lowry, W. E.,

Gerstein, M. B., … Geschwind, D. H. (2019). A Single-Cell Transcriptomic Atlas

of Human Neocortical Development during Mid-gestation. *Neuron*, *103*(5), 785-

801.e8. https://doi.org/10.1016/j.neuron.2019.06.011

Satterstrom, F. K., Kosmicki, J. A., Wang, J., Breen, M. S., De Rubeis, S., An, J.-Y.,

Peng, M., Collins, R., Grove, J., Klei, L., Stevens, C., Reichert, J., Mulhern, M.

S., Artomov, M., Gerges, S., Sheppard, B., Xu, X., Bhaduri, A., Norman, U., …

Walters, R. K. (2020). Large-Scale Exome Sequencing Study Implicates Both

Developmental and Functional Changes in the Neurobiology of Autism. *Cell*,

180(3), 568-584.e23. https://doi.org/10.1016/j.cell.2019.12.036

Thapa, P., & Farber, D. L. (2019). The Role of the Thymus in the Immune Response.
Thoracic Surgery Clinics, *29*(2), 123–131.
https://doi.org/10.1016/j.thorsurg.2018.12.001

van Essen, M. J., Nayler, S., Becker, E. B. E., & Jacob, J. (2020). Deconstructing
cerebellar development cell by cell. *PLOS Genetics*, *16*(4), e1008630.
https://doi.org/10.1371/journal.pgen.1008630

Vijayakumar, N. T., & Judy, M. V. (2016). Autism spectrum disorders: Integration of the
genome, transcriptome and the environment. *Journal of the Neurological
Sciences*, *364*, 167–176. https://doi.org/10.1016/j.jns.2016.03.026

Vladoiu, M. C., El-Hamamy, I., Donovan, L. K., Farooq, H., Holgado, B. L., Ramaswamy,
V., Mack, S. C., Lee, J. J., Kumar, S., Przelicki, D., Michealraj, A., Juraschka, K.,
Skowron, P., Luu, B., Suzuki, H., Morrissy, A. S., Cavalli, F. M., Garzia, L.,
Daniels, C., … Taylor, M. D. (2018). *Childhood Cerebellar Tumors Mirror
Conserved Fetal Transcriptional Programs* [Preprint]. Cancer Biology.
https://doi.org/10.1101/350280

Voineagu, I., Wang, X., Johnston, P., Lowe, J. K., Tian, Y., Horvath, S., Mill, J., Cantor,
R. M., Blencowe, B. J., & Geschwind, D. H. (2011). Transcriptomic analysis of
autistic brain reveals convergent molecular pathology. *Nature*, *474*(7351), 380–
384. https://doi.org/10.1038/nature10110

Ziats, M. N., & Rennert, O. M. (2011). Expression Profiling of Autism Candidate Genes
during Human Brain Development Implicates Central Immune Signaling
Pathways. *PLoS ONE*, *6*(9), e24691.
https://doi.org/10.1371/journal.pone.0024691

CURRICULUM VITAE