Perinatal Epigenetic Signature of Autism Spectrum Disorder

By

YIHUI ZHU

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Integrative Genetics and Genomics

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

_____

Janine M. LaSalle, Chair

_____

Rebecca J. Schmidt

_____

Fereydoun Hormozdiari

Committee in Charge

2021

This work is dedicated to my family.

Thank you to my parents who taught me to be kind, persevere, and work hard.

Your consistent support, love, encouragement, and always believing I can do it are invaluable to my success.

Acknowledgements

I would like to sincerely thank everyone during my graduate school journey for both my research and training experiences which are extremely valuable and irreplaceable. First, I would like to thank my dissertation advisor, Janine LaSalle, for her consistent support, trust, enthusiasm, and mentorship. Her immense knowledge and plentiful experience have encouraged me to climb multiple mountains and spark me to see the larger and brighter pictures in research.

I would like to thank all the LaSalle lab members who provided numerous supports in my projects and career during my entire PhD time. Charles Mordaunt, for his incredibly supportive, collaborative, passionate, generous with his knowledge, and having all those great discussions with me about research, graduate school, and career trajectories. Antonio Gomez, for his numerous supports, collaboration, erudite on designing creative experiments starting with small initial ideas. Ben Laufer, for his enthusiasm, collaboration, and deep knowledge. Dag Yasui, for his interest, generosity, support, and expertise in multi-dimensional fields. Rochelle Coulson, for her insights, consistent support, hard work, kindness, and having all those great discussions with me about science and life. Kari Neier, for her kindness, advice, and endless motivation. Ria Marathe, for her persistent work, excitement, and curiosity. Julia Jianu, for her consistent persistence, reliability, and insightful about science. Annie Vogel Ciernia, for her wisdom, excellent advice, and inspiration. Keith Dunaway, for his insights, kindness, and advice. I would also like to thank Florence Crary-Dooley, Jesse Lopez, Oran Guiterrez, Theresa Tatah, Aron Mendiola, Osman Sharifi, Julia Mouat, Viktoria Haghani, Logan Williams, Rebecca Palmer, Hyeyeon Hwang, and Sophia Hakam for their support and kindness assistance in the LaSalle lab.

I would also like to thank my dissertation committee. Thank you to Rebecca Schmidt, for her continuous support, guidance, mentorship, enthusiasm, and great care. Her leadership and epidemiological insights give me lots of opportunities for collaborations, and extending the horizon of my knowledge. Thank you to Fereydoun Hormozdiari, for his guidance and support, especially his expertise in genomics extend my interdisciplinary research.

UC Davis has been a wonderful place to learn, develop, and grow in science, and I'm glad to have the opportunities to work with so many excellent faculty and colleagues in a wide range of disciplines. I would like to thank everyone in the MARBLES study, including Irva Hertz-Picciotto, Pamela Lein, Anthony Valenzuela, Birgit Puschner, Cheryl Walker, Sally Ozonoff, Bill Elms, Daniel Young, Elizabeth Angel, Melissa Rose, Kristina Forward, and families participated in this study. I would like to thank everyone in IGG graduate group, including Sean Burgess, Richard Michelmore, David Segal, Frederic Chedin, Ian Korf, Megan Dennis, Michele La Merrill, Nobuko Hagiwara, Jan Dvorak, Mona Ereiqat, and other faculties and colleagues. Thank you to staff in the DEB (Designated Emphasis in Biotechnology): Denneal Jamison-McClung, Abhaya Dandekar, Karen McDonald, Judith Kjelstrom and FUTURE: Daniel Moglen, Rachel Reeves, Millie Copara for career exploration. I would also like to thank my collaborators in other institutions, Kelly Bakulski, Danielle Fallin, and Jason Feinberg.

Finally, thank you to my parents, Zhongwu Zhu and Xinghua Lu for consistent support and encouragement to keep me moving and become successful all the time. Thank you to Haotian Li, for his love, kind and support to encourage me in achieving my goals. Thank you all, I will always strive to make you proud.

Preface

- Chapter 2 is published in Human Molecular Genetics. Yihui Zhu is the lead author of this manuscript.
    - Zhu, Y., Mordaunt, C. E., Yasui, D. H., Marathe, R., Coulson, R. L., Dunaway, K. W., ... & LaSalle, J. M. (2019). Placental DNA methylation levels at *CYP2E1* and *IRS2* are associated with child outcome in a prospective autism study. *Human molecular genetics*, 28(16), 2659-2674.
    - Additional files with supplemental figures and tables are available as part of the online version at https://doi.org/10.1093/hmg/ddz084.

- Chapter 3 is published in Autism Research. Yihui Zhu is the lead author of this manuscript.
    - Zhu, Y., Mordaunt, C. E., Durbin-Johnson, B. P., Caudill, M. A., Malysheva, O. V., Miller, J. W., ... & LaSalle, J. M. (2021). Expression Changes in Epigenetic Gene Pathways Associated With One-Carbon Nutritional Metabolites in Maternal Blood From Pregnancies Resulting in Autism and Non-Typical Neurodevelopment. *Autism Research*, 14(1), 11-28.
    - Additional files with supplementary figures and tables are available as part of the online version at https://doi.org/10.1002/aur.2428.

- Chapter 4 is in preparation for submission. Yihui Zhu is the lead author of this manuscript.

Abstract

Autism Spectrum Disorder (ASD) is a polygenic neurodevelopmental disorder that includes deficits in social communication and repetitive behaviors. Most ASD cases are thought to arise from complex genetics interacting with perinatal environmental factors, complicating the discovery of common genetic risk. The prenatal period is a critical window for neurodevelopment. Environmental factors, especially prenatal nutrients in one-carbon metabolites are crucial for neurodevelopment and provide methyl donors for downstream epigenetic pathways. The epigenetic layer of DNA methylation in placenta reflects developmental and molecular memory of *in utero* experiences. Placenta, a fetal tissue usually discarded a birth, is a potentially rich source of DNA methylation patterns predictive of ASD in the child.

This dissertation details recent progress in understanding the interface of epigenetics, genetics, and environment in ASD etiology. The perinatal epigenomic signature of ASD was investigated by whole-genome-bisulfite sequencing (WGBS) on placenta. First, a pilot study of all male placenta samples from a prospective study MARBLES identified 400 differential methylated regions (DMRs) distinguishing ASD diagnosis. ASD DMRs were significantly enriched at promoters of genes with functions in neuronal development and that overlapped with brain ASD DMRs and known ASD risk genes. Two DMRs at *CYP2E1* and *IRS2* reached genome-wide significant with methylation level separately affected by genotype, or periconceptional maternal prenatal vitamin use. This pilot study demonstrated that DNA methylation changes in

placenta are relevant to brain development and may serve as early epigenetic markers for ASD at birth.

Second, maternal blood samples collected during pregnancy were examined for genome-wide transcriptional signatures of child outcome and nutritional metabolites of one-carbon metabolism. Six transcripts were associated with child outcomes at genome-wide significance, and 1,912 differentially expressed genes with nominal confidence overlapped with known ASD risk genes. A coexpression module enriched in genes with DNA methylation functions showed a suggestive protective association with folic acid and ASD risk. These results demonstrated that the prenatal maternal blood transcriptome is an indicator of gestational nutrition and child neurodevelopment.

Lastly, a multi-cohort and multi-tiered study was performed to integrate genomics, epigenomics, and transcriptomics in placenta and brain, leading to the discovery and functional characterization of a previously uncharacterized novel ASD risk gene *NHIP* located in a comethylated block at 22q13.33. *NHIP* was highly expressed in brain, and was induced following neuronal differentiation or hypoxia. Transient *NHIP* overexpression increased cellular proliferation and altered transcriptome levels in synapses, neurogenesis, and response to oxidative stress. A nearby structural variant was significantly associated with ASD, *NHIP* expression, and DNA hypomethylation at 22q13.33. Together, these studies identified a novel ASD risk gene related to oxidative stress during brain development that was altered by both common genetic and environmental factors.

Table of Contents

Chapter 1 – Introduction

Epigenetics

The term epigenetics refers to heritable alterations to nucleotides or chromatin without changes to DNA sequence. Epigenetic marks reflect a complex interaction between genome and environment that also dynamically changed during development and differentiation. Epigenetic mechanisms can help explain the dynamic phenotypic transitions that occur in the genome starting from a fertilized egg to a multicellular organism containing hundreds of specialized cell types. Epigenetic mechanisms play a crucial role in the maintenance of cell identities during development and throughout life through the expression of specific combinations of genes (1). Epigenetic modifications, including DNA methylation, histone post-translational modifications (PTMs), noncoding RNA, and chromatin architecture are dynamically variable and have essential roles in mammalian development (2,3). Different layers of the epigenome come together to influence chromatin structure and gene expression which are crucial for differentiation and cell type maintenance. Epigenetic layers are regulated by epigenetic players with the epigenetic modification functions as writer (depositing), reader (interpreting), and eraser (removing) (4).

Understanding the functional relevance of epigenetic layers and players is important to uncover the transcriptional regulation mechanisms in human health and disease. The Encyclopedia of DNA Elements (ENCODE) project is a useful database of functional elements, including chromatin regulation and histone marks with multiple layers of information from more than

8,000 assays (5). Epigenetic research is gradually becoming an important approach to resolve the missing heritability from large-scale genome-wide association studies (GWAS) (6).

DNA methylation

One of the best-understood epigenetic layers in vertebrates is DNA methylation, or the addition of a methyl group on the fifth carbon of cytosine, usually in CpG dinucleotides (7). CpG sites are unevenly distributed in the genome, concentrating in repetitive regions such as tandem repeats, CpG islands, and gene regulatory regions (8,9). CpG islands are characterized as regions with more than 200 bp in length with greater than 50% CpG content and are often localized at transcription start sites (10). Most epigenetic marks are removed and reestablished for each generation. Early in development, close to the time of embryo implantation, DNMT3A and DNMT3B start to methylate the majority of CpG sites (11). As the process continues, DNA methylation is modified to ensure specific cell fate in each cell differentiation stage (11). In most cancers, DNA methylation is highly dysregulated, with global hypomethylation and specific hypermethylation over gene regulatory regions compared to non-transformed cells of the same lineage (8,12).

DNA methyltransferases (DNMTs) are enzymes that transfer methyl groups to DNA, either *de novo* to previously unmethylated cytosines, or by maintenance on the daughter stand during DNA replication. DNMT3A and DNMT3B assisted by DNMT3L add *de novo* methylation marks with important functions in cell and tissue specification (13,14). DNMT1 is responsible for DNA methylation through the recognition of hemi-methylated DNA (15). DNA methylation can be

erased by the ten-eleven translocation (TET) enzymes which hydroxylate 5mC to produce 5-hydoxymethl cytosine (5hmC), followed by DNA repair back to unmethylated cytosine (16). Readers of DNA methylation contain methyl-CpG-binding domain (MBD) that bind DNA containing one or more symmetrically methylated CpGs. An example of a MBD-containing reader is methyl CpG binding 2 (MeCP2), which is deficient in neurodevelopmental disorder Rett syndrome (17,18).

Repetitive elements occupy half of the genome and generally have a high methylation level (10). Tandem repeats, long interspersed nuclear elements (LINEs), and short interspersed nuclear elements (SINEs) including Alu and SVA elements are found to be hypomethylated in cancer with large percentage of 5mC loss (12). Hypomethylation of repeats can result in chromatin decondensation, translocation, and rearrangements due to genome instabilities (19). Alu, SVA, and LINE elements are retrotransposons that make up around 30% of the genome and are able to amplify themselves with RNA intermediates, although the majority are nonfunctional (20). These CpG-rich elements have the capacity to act as promoters or enhancers thereby affecting gene transcription (21,22). Hypomethylation of LINE and Alu elements were found to be strongly linked with genome instability of multiple cancers (23) .

Histone modification

DNA wrapped around histone octamers to form nucleosomes is the major component of chromatin. Each nucleosome contains four core histone proteins pairs (H2A, H2B, H3, H4).

DNA methylation together with histone modifications correlate with gene expression differences in various ways (24). Chromatin conformation is controlled by chemical modification of the N-terminus tails of the histone proteins with covalent modification. Covalent modification of histone tails have following types: acetylation, phosphorylation, ubiquitination, and methylation (25). Histone PTMs involve dynamic process, with enzymes performing covalent modifications, removal, and recognition of previous markers (26). Those processes can alter nucleosome compaction directly or indirectly alter function in the binding sites for reader proteins with effects on transcription (27).

Histone methylation has been widely shown to define chromatin states that regulate transcription. Methylation at different histone tail residues can result in either activation or repression of transcription, depending on the residue. Histone methylation on the tail of histone proteins H3 and H4 occurs at both arginine and lysine. K-methyltransferases transfer methyl groups from the cofactor S-adenosyl methionine (SAM) to histone lysine residues. The mixed lineage leukemia (MLL), a K-methyltransferase, catalyzes H3K4 methylation. MLL activates genes involved in development and differentiation, which is opposite to polycomb repressive complex proteins (28). EZH2, a K-methyltransferase that catalyzes H3K27 trimethylation is important for stem cell maintenance and differentiation (29). EZH2 has also been shown to interact with DNMTs in the parallel control of DNA methylation (30). The H3K27me3 mark is normally involved with gene silencing with relation in stem cell differentiation and development. Histone methylation markers are known to interact with each other and with DNA methylation to regulate transcription (31).

Unlike histone methylation, which is associated with both transcription activation or repression depending on specific residues, histone acetylation is strongly associated with transcriptional activation. Histone acetylation occurs on lysine residues, resulting in neutralization of the positively charged histone that enhances transcription, since it decrease the interaction of the histone core proteins with the negatively charged DNA phosphate backbone (32). The histone acetylation marks are controlled by reciprocal actions of histone acetyltransferases (HATs), and histone deacetylases (HDACs). HATs function as a cofactor for open or permissive chromatin state by adding acetyl group to histone lysine (33). Wnt signaling pathway has been shown to be related with HAT dysregulation in cancer (34). The EP300/CBP family is an example of HATs that have the ability to acetylate core histones, as well as other proteins in cell proliferation, like p53 (35). EP300 are critical to sustain cell proliferation and regulate tumor suppressors (36). HDACs are involved in transcriptional repression by removing acetyl groups from histones (33). HDACs also can regulate other proteins important for cancer, such as p53, and STAT3 (37). As an example, HDAC3 regulates an important intracellular Wnt signaling pathway with sensitivity to growth regulation by Vitamin D (38).

Chromatin structure, such as folding and looping, also plays a vital role in transcriptional regulation in three-dimensional space by making gene promoters accessible to regulatory elements, such as enhancers, repressors, and insulators (39). Enhancers can regulate promoters by affecting regulatory elements, loop structures, and noncoding RNA (39). Chromosome are organized into functional compartments by intra-genomic interactions called topologically associating domains (TADs). Regions in TADs interact more frequently than regions located

outside with genes coregulated, expressed, and correlation in epigenetics marks (40). Genes and regulatory elements are insulated by TAD boundaries for epigenetics and functional insulation (41). The zinc-finger CCCTC-binding factor (CTCF) plays an important role in defining TAD boundaries and mediate long-range chromatin interactions (40). Noncoding RNAs also can play a regulatory role in long-range chromatin, including short non-coding RNAs (siRNAs, miRNAs, and piRNAs) and long noncoding RNAs (lncRNAs) that have significant roles in regulation of gene expression (42).

One-Carbon Metabolism

The one-carbon metabolism pathway links nutrition to epigenetic programming by regulating the supply of methyl donors for DNA and histone methylation. One-carbon metabolism includes the folate cycle, methionine remethylation, and transsulfuration pathways, with functions in molecular biosynthesis, genome maintenance, and epigenetic regulation (43). Dietary nutrients that function as substrates or cofactors in the one-carbon pathway include folate (vitamin B9), B vitamins (B2 (riboflavin), B6 (pyridoxine), and B12 (cobalamine), methionine, choline, and betaine (44). An example of how dietary nutrients can influence neurodevelopment is the reduction of neural-tube birth defects by 19%-32% since the implementation of folic acid fortification to grains since 1998 (45).

Folates are methyl donors that can activate the one-carbon metabolism pathway. Folate needs to come entirely from dietary sources because animals cannot synthesize folate endogenously (46). Folic acid is the synthetic form of folate that is fully oxidized and without active coenzyme

function (47). Folic acid reduces to dihydrofolate (DHF), and then to the biological active form tetrahydrofolate (THF) before entering the folate cycle (43). THF converts to 5,10- methylene-tetrahydrofolate, and is irreversibly reduced to 5-mTHF by the B2-dependent enzyme methylenetetrahydrofolate reductase (MTHFR). Other B12, and B6 dependent enzymes work together to transfer methyl group to betaine with producing dimethylglycine and methionine (48).

In the methionine cycle, methionine is adenylated to S-adenosylmethionine (SAM), which is the universal methyl donor that is involved in downstream cellular processes (49). During this process, DNMTs transfer methyl groups to DNA. During this methyl transfer step, SAM is converted to S-adenosylhomocysteine (SAH), which is hydrolyzed back to homocysteine (HCY) and adenosine by reversible reactions in methionine cycle completion (50). High SAH concentration can lead to inhibition of DNMTs, with some using SAM/SAH ratio as an index of the potential of methylation at the cellular level (51). Increased SAH is associated with decreased SAM/SAH ratio, which is associated with DNA hypomethylation (52). The inclusion of high doses of SAM from the 8-cell stage using bovine embryo culture led to global hypermethylation (53). In the transsulfuration pathway, high levels of SAM can induce HCY degradation when methionine and folate are sufficient, but requires B6 as cofactor (48). These studies show that dietary factors are crucial for the methionine cycle supply chain. Folate and B vitamins are known as environmental modifiers that can change DNA methylation such as in honeybees where queen bees are determined by royal jelly consumption (54,55).

Ethnic and interindividual variability in gene regulation can result in single nucleotide polymorphisms (SNPs) in genes involved in one-carbon metabolite pathways that are associated with changes in DNA methylation and other epigenetic marks (48). Furthermore, genetic differences resulting in differences to one-carbon metabolites, such as the *MTHFR* C677T polymorphism, can impair fertility (56). Deletions or modifications of the epigenetic regulators, *DNMT1* and euchromatin histone methyltransferase 1 (*EHMT1*), are associated with adverse pregnancy outcomes (57,58). Together, these findings show that one-carbon metabolism function is a key biochemical pathway that connects maternal environment, genetics, and epigenetics in early fetal development.

Autism Spectrum Disorder

Autism spectrum disorder (ASD) is a heterogeneous neurodevelopmental disorder with a range of symptoms and severities that is a growing concern worldwide, with 1 in 54 children diagnosed in the US (59). Children diagnosed as ASD show persistent difficulties in social communication and interaction, restricted and repetitive behaviors and interests, and language deficits. The etiology of ASD is complex and remains incompletely characterized by existing individual genetic and environmental factors (60,61). Genetic heritability is expected to be a major component of ASD risk based on studies of twins, siblings, and other family members (62,63). The genetic architecture of ASD contains both rare variants with strong effects and multiple common variants with weak individual effects (64). Genes mutated in rare cases of ASD are enriched for neuronal, embryonic development and chromatin regulation pathways. Large genome-wide association studies (GWAS) show the role of common genetics variants

with remaining challenges in ASD complexity and heterogeneity, as no single genetic variant can explain more than 1% of disease liability (64–66). ASD related genes also overlap with cancer pathways on cell differentiation, chromatin, and DNA repair functions (67). Environmental factors are also known to contribute to ASD risk, especially during the prenatal stage (68). During the gestation stage, environmental factors, such as environmental toxicants, and nutritional factors can affect cellular proliferation and differentiation processes (69–71). ASD etiology is thought to begin in early development and be influenced by both genetics and environmental factors influencing neuronal trajectories (72).

Genomics of ASD

Heritability is used to summarizes the degree variation in a trait that is due to genetic factors variation between individuals in the population which can be estimated from the correlation of offspring and parental, full or half siblings, or monozygotic and dizygotic twins (73). ASD is considered as highly heritable which was historic estimated above 90% based on monozygotic twin studies (74,75), but the ratio are lower with recent increased studies (76,77). Recent study on 37,570 twin pairs, estimated the heritability as 83% (77), lower than initial early 90% twin studies (78) and higher than 38% later California twin study (61). Infants born to families with at least one affected elder sibling have 24% chance of development ASD (79). Overall, the concordance rates among monozygotic twins, dizygotic twins, and siblings are 30-90%, 0-65%, and 3-30% with an estimated overall heritability of 0.7-0.8 (61,62,75,80). These studies estimate heritability often limited by heritability estimated based on twins. Recent population-

based studies in show a lower estimated heritability of 0.5-0.6 (81). All these studies and analysis shown ASD has strong genetic relatedness.

The genetic component of ASD includes both common variants with weak individual effects and rare variants with strong effects (64). Rare inherited and *de novo* variants are the major genetics contribution to individuals risk for ASD (65,66,82). Inherited truncated mutations are found to be more prevalence in ASD, containing genes function in synaptic formation, transcription regulation, and chromatin modifications (82,83). Partially of ASD compose of rare disorders with over 100 rare Mendelian syndromes have been associated with ASD (84,85). In contrast to the strong male bias in idiopathic ASD, syndromic ASD cases have different male-to-female ratio based on the types of causations (86). Classic neurodevelopmental syndromes such as fragile X syndrome (FXS) (mutations in *FMR1*), Rett syndrome (mutations in *MECP2*), tuberous sclerosis (mutations in *TSC1* and *TSC2*), and dup15q syndrome show partial comorbidity with ASD. In consanguineous families, rare recessive mutation on *CNTNAP2*, *SLC9A9*, and *BCKDK* were associated with ASD and epilepsy by linkage analysis and homozygosity mapping (87–89). Neuroligin family *NLGN3* and *NLGN4* are rare X-linked mutation that identified in male children with ASD together with mental retardation (90). *ANDP* mutation has been found to associated with ASD with function in facial dysmorphism (91). Three members of scaffolding proteins, *SHANK1*, *SHANK2*, and *SHANK3*, together with two members of the neurexin, *NRXN1* and *NRXN3* are rare inherited variants that have vital roles in ASD (92–96). With increasing researches in ASD, there is a substantial expand knowledge of rare inherited alleles in ASD.

Fragile X syndrome is the most common inherited from of intellectual disability as X-linked genetic disorder with higher frequency in males (97). FXS caused by expansions of CGG triplet repeat in the 5' untranslated region of the *FMR1* gene that located on the X chromosome (98). The expansion of CGG repeats leads to transcriptional silencing of *FMR1* that lead to lack of encoding fragile X mental retardation protein (FMRP) which is a synaptic RNA-binding protein with functions in modulating synaptic plasticity (99). Unaffected individuals have 6-44 CGG repeats, individuals with 50-200 CGG repeats are at risk for further expansion, while greater than 200 hypomethylated CGG trinucleotide repeats are observed in patients with FXS (100–102). *MECP2* mutation is one of the recurrent mutations in approximately 4% of ASD (103). MeCP2 is a reader of epigenetic information and a modulator of chromatin structure, which regulate mRNA splicing and epigenetic change in histone markers with effects in synaptic functions (104,105). SH3 and multiple ankyrin repeats domains (*SHANK3*) mutation were estimated to contribute about 1% of ASD cases and associated with zinc deficiency (106). *SHANK3* has crucial function in regulating synapses function and multiple postsynaptic signaling complex (107).

The largest exome sequencing study on ASD to date (ASD n = 11,986, TD n = 23,598) identified 102 risk genes at an FDR level less than 0.1, which included 49 shown higher frequencies of de novo variants in individuals with severe neurodevelopmental delay (108). The largest meta-analysis genome-wide association study (GWAS) study to date (ASD n = 18,381, TD n = 27,969) from the Psychiatric Genomics Consortium ASD workgroup (PGC-AUT) (64) focused on SNP variants from standard genotyping arrays. Other studies examined structural variants over

protein coding genes by whole exome sequencing (65). SNP-based GWAS has multiple limitations, including reliance on pre-exiting genetics variant reference panels, imputation errors due to imperfect linkage disequilibrium, and complex structural variance (109).

Structural variants (SV) are also polymorphic within humans and have been associated with many human diseases, especially involving immune responses, and cognitive disorders, such as schizophrenia (110–112). SVs exhibit a nonrandom distribution and form clusters into hotspots located within gene-poor regions in the primate genome (113,114). Genic sequences harbored with SV hotspots were enriched for oxygen transport, sensory perception, synapse assembly, and antigen-binding (114). Recent studies suggested that a large SV burden is associated with lower cognitive ability (115–117). SVs shown to be in association with autism have mapped to noncoding putative regulatory regions of the genome (118).

Environment factors in ASD risk

Environmental exposures can influence brain development in various ways including neural tube development, cell differentiation and migration. Immune system responses or oxidative stress from chemicals in the environment and their metabolites can affect digestive systems, which can then dysregulate nutrient or essential fatty acid uptake and lead to pathogenic changes. Environmental factors can also lead to epigenetic changes that regulate gene expression with short- or long-term effects. Furthermore, both nutrients and environmental chemicals can alter cellular biochemistry to change enzyme functions in multiple pathways.

Large population-based epidemiology studies have shown that high folate prenatal vitamin consumption near conception is associated with reduced ASD risk of the child (70,119,120). In the Childhood Autism Risks From Genetics and the Environment (CHARGE) case-control study, prenatal vitamin supplement use at an early pregnancy stage was associated with a 40% reduction in ASD risk (70,120). For mothers with a common risk allele in *MTHFR*, encoding the methylenetetrahydrofolate reductase enzyme in the one-carbon metabolic pathway, the child had a higher risk for ASD if they did not take a prenatal vitamin during the first month of pregnancy (119). Meta-analysis with eight case-control studies also shown that *MTHFR 677 C>T* polymorphism contributes to increase ASD risk, and periconceptional folic acid may reduce ASD risk (121). The finding was replicated with a Norwegian cohort study of 85,176 children with 40% lower risk of ASD in children whose mothers took folic acid supplements periconceptionally (119). In high-risk families, the association between maternal prenatal vitamin intake during the first month of pregnancy is also associated with reduction in ASD risk of the child (122). These studies suggest periconceptional folic acid can improve neurodevelopment and behavioral outcomes as well as reduce risk for neural tube defects (123). During the implantation stage, embryos undergo numerous DNA methylation changes with erasure and reestablishment (124), likely explaining the periconceptional stage as a critical period of protection.

Besides protective environment factors, several environmental chemicals have been shown to have adverse effects on reproduction and neurodevelopment, with stronger effects *in utero* and in children than in adults (125). As a group, both persistent and non-persistent chemicals

can alter mechanisms in nervous system, immune regulation, lipid metabolism, and mitochondrial functions that may influence ASD risk (126–128). For air pollution, a large population-based CHARGE study showed that living within 309 meters of a freeway nearly doubled the odds of having a child with ASD (129). The association between air pollution and ASD was strengthened when exposure was in the third trimester (129). Exposure to polychlorinated biphenyls (PCBs), a class of endocrine-disrupting chemicals within the category of persistent organic pollutants (POPs), has also been shown to associate with ASD risk (127,130). PCBs can affect levels of hormones during early gestation that regulate neuronal differentiation and growth in early gestation and also result in DNA methylation changes from an *in vitro* study (131,132). Pesticides are frequently designed to damage the nervous systems of insect pests, but can also have effects on mammalian neurotransmission. A prior study showed that self-reported pesticide exposure was significantly associated with alterations in placental DNA methylation in the ASD high risk MARBLES cohort (133). Residence in households located near agricultural pesticide applications during the first trimester was associated with an elevated level of ASD risk (128). Pesticides might influence neurodevelopment by various mechanisms, such as serotonergic systems, endocrine disruption, and calcium signaling (134).

Overall, numerous studies have demonstrated an environmental contribution to ASD. Considerable evidence supports potential roles for preconception or prenatal maternal nutrients, and exposures to some environmental chemicals in ASD risk. From those studies, we can further conclude there is no single or universal cause of ASD, and multiple interacting factors are likely involved. However, there are limitations for existence environmental factors

for ASD. Some studies have the limitation of inconsistent and/or insufficient replication among different populations or windows of exposure. Environmental factors are mostly association studies with few *in vivo* or *in vitro* experimental models of specific chemicals and pathways.

Epigenetic findings in ASD

Epigenetic mechanisms represent an important intersection of genetic and environmental influences to ASD, with previous research performed in human tissues such as brain, placenta, and blood (132,135–137). Previous studies of epigenome-wide association studies (EWAS) have demonstrated differential methylation on specific regions and epigenetic signatures associated with ASD diagnosis. The largest population EWAS in ASD have been restricted by study design limitations, including case-control studies where temporal relationships are not clear, and microarray based platforms that were limited to probe locations, covered less than 3% CpG sites, and were heavily biased towards CpG islands and existing gene promoters (138,139). In contrast, WGBS-based approaches enable full interrogation of the mappable human genome and cover the most epigenetically polymorphic regions missed by array-based approaches. Given the importance of brain development in ASD etiology, epigenetic studies performed in postmortem brain have shown differentially methylated regions for ASD at genes with functions in synaptic transmission, and histone acetylation (140,141). Studies have shown the DNA methylation changes during neuronal death with overexpression of *DNMT3A*, suggesting that DNMT3A inhibition protect neurons (142).

Placenta Interface and Methylome

Evolution of placenta

Placenta is the unique tissue that sits at the interface between the maternal and fetal vascular beds, functioning to provide nutrients and waste mediation to the developing conceptus. Placental cellular lineage is not unique to mammals, as the trophectoderm layer occurs within a wide range of both vertebrate and invertebrate species, but mammals have a tremendous variety in placenta anatomy that all serve the unique functional requirements of fetal provisions until live birth (143,144). Comparing with egg-laying species, offspring benefit from gaining internal development including environmental protection in maintenance on temperature, oxygen, osmolarity, and protection from predation (145). Mammals are amniotes distinct from reptiles and birds, possessing two other membranes, chorion and allantois, that form the placenta and umbilical cord (143). The fertilized embryo undergoes a series of cell division prior to implantation, which results in blastocyst formation consisting of two distinct cell populations, trophectoderm which differentiates into the placenta, and inner cell mass (ICM) which gives rise to the embryo (146). Placenta is the first organ to form in mammals to establish the maternal-fetal vascular interface with the capability to supply the bioenergetic needs of conceptus development (147). It is a complex organ with diverse functions in hormone production, immune protection, and nutrient transportation (147).

Placental metabolism

Placenta is not just simply a transporter organ, metabolically it needs to consume 40-60% of the body's oxygen to get glucose delivered to the uterus (148). Placenta metabolism can

significantly impact oxygen and nutrient delivery to the fetus by cell fate regulatory pathways. Aerobic lactate production enables the placenta to produce and transfer large amounts of lactate to the fetus, and placenta has the unique ability to metabolize lactate specifically during midgestation but not at term (149). Placenta alters its metabolic strategies under stressful states such as hypoxia, or in pathological states such as intrauterine growth restriction (IUGR). For instance, placenta can decrease its consumption of oxygen and increase glycolysis to maintain the basic bioenergetic needs under hypoxia at high altitude (150). Isolated hypoxia can result in increased hypoxia-inducible factor (HIF) levels and target gene expression in the human placenta (151,152). HIF is also critical for hypoxia metabolic adaptation, and regulates trophoblast cell fate decisions (152,153). Oxygen tension can also modulate extravillous trophoblast proliferation, differentiation, and invasion (154).

Placenta has long been recognized as steroid hormone-producing organ, but it can also produce neurotransmitters and regulate fetal brain development through those factors, such as serotonin, dopamine, and epinephrine/norepinephrine (155). Placenta is actually the sole source for some neurotransmitters during initial brain formation. Dopamine is a neurotransmitter that is produced by the placenta that can induce direct effects on brain development through different dopamine receptors, including D1 and D2 (156–158). Dopamine and norepinephrine are produced by placenta cytotrophoblasts, primarily expressed by syncytiotrophoblast cells during mid and late stages of pregnancy (159).

Placenta methylome

Term placenta is an accessible fetal tissue that maintains a distinctive embryonic bimodal methylation pattern, with higher methylation over expressed genes and relatively lower methylation over non-expressed genomic regions, a property which opens a unique window to study fetal development from placenta DNA methylation patterns (160,161). The placenta methylome shows developmental plasticity, which can reflect on both maternal and fetal signals. The key feature of epigenetic reprogramming is the erasure of DNA methylation in the zygotes and preimplantation embryo (162). Around implantation, ICM starts to gain do novo DNA methylation rapidly, while trophectoderm remains hypomethylated (163). Chorionic villi show 10%-25% less DNA methylation than fetal tissues (164). Placenta has global hypomethylation with reduced *DNMT1* levels and less maintenance of methylation during rapid growth phase in early development (165). X-chromosome inactivation (XCI) can be less stable in placenta on more regions or genes escape XCI with through mechanisms that remain unclear.

The landscape of DNA methylation is especially dynamic in the earliest stages from fertilization to implantation, with preimplantation embryos, and placenta both characterized by the presence of partially methylated domains (PMDs) (160,161,166). PMDs are large regions (>100 kb) with reduced DNA methylation, covering nearly 40% of the genome with enrichment in gene-poor regions (161). The presence of PMDs is limited to the methylomes of placenta and some cultured cell lines, including fetal lung fibroblasts, foreskin fibroblasts, adipocytes, and tumors (161,167,168). Genes within PMDs tend to be tissue-specific or inducible, with hypomethylation of gene bodies but hypermethylation of CpG islands compared with non-PMD

regions (161). The reasons for the presence of PMDs is still unclear, but they are hypothesized to arise as a consequence of rapid cell division for early placental development or programmed developmental mechanism. Regardless, because placental PMDs provide a historical footprint of gene activity in utero, the placental methylome could be useful for identifying epigenetic biomarkers relevant for human disease (168).

There are an increasing number of studies on placenta-brain axis using multi-omics datasets showing the key role of placenta in prenatal development. A recent study on schizophrenia polygenic risk scores showed significant association of the genetic risk with perinatal environment and at genes with expression in placenta (169). Placenta multi-omics kernel aggregation strongly predicted intellectual and social function via cross-validation which was significantly associated with ASD status (170). A prior study on placenta methylome and ASD identified a differentially methylated fetal brain enhancer near *DLL1* (171). Another study on self-reported exposures showed a significant association with pesticide exposure on placenta DNA methylation (133).

Due to previous findings on genomics, environment, and epigenetics, we investigated ASD risk through perinatal epigenomic signatures in placenta and brain, followed with a stud on maternal metabolites, studies incorporating genomics, and functional studies. Together these approaches bring novel and multi-dimensional insights into ASD etiology and shed light on perinatal signatures of ASD.

Chapter 2 – Placental DNA methylation levels at *CYP2E1* and *IRS2* are associated with child outcome in a prospective autism study

Yihui Zhu, Charles E. Mordaunt, Dag H. Yasui, Ria Marathe, Rochelle L. Coulson, Keith W. Dunaway, Julia M. Jianu, Cheryl K. Walker, Sally Ozonoff, Irva Hertz-Picciotto, Rebecca J. Schmidt*, Janine M. LaSalle*

* Joint last authors

Abstract

DNA methylation acts at the interface of genetic and environmental factors relevant for autism spectrum disorder (ASD). Placenta, normally discarded at birth, is a potentially rich source of DNA methylation patterns predictive of ASD in the child. Here, we performed whole methylome analyses of placentas from a prospective study MARBLES (Markers of Autism Risk in Babies-Learning Early Signs) of high-risk pregnancies. 400 differentially methylated regions (DMRs) discriminated placentas stored from children later diagnosed with ASD compared to typically developing controls. These ASD DMRs were significantly enriched at promoters, mapped to 596 genes functionally enriched in neuronal development, and overlapped genetic ASD risk. ASD DMRs at *CYP2E1* and *IRS2* reached genome-wide significance, replicated by pyrosequencing, and correlated with expression differences in brain. Methylation at *CYP2E1* associated with both ASD diagnosis and genotype within the DMR. In contrast, methylation at *IRS2* was unaffected by within DMR genotype, but modified by preconceptional maternal prenatal

vitamin use. This study therefore identified two potentially useful early epigenetic markers for ASD in placenta.

Introduction

Autism spectrum disorder (ASD) is a heterogeneous neurodevelopmental disorder diagnosed by a combination of behavioral features including restricted interests, repetitive behaviors, language deficits, and impairments in social communication (59). 1 in 54 children in the United States are diagnosed with ASD, at a mean age of 4.2 years (59). ASD is currently diagnosed by clinicians trained on the Autism Diagnostic Observation Schedule (ADOS) and the Autism Diagnostic Interview - Revised (ADI-R) according to the Statistical Manual of Mental Disorders (DSM-5) which is most accurate at or after 36 months (59). However, an early assessment of ASD risk could identify infants and toddlers who would benefit from behavioral interventions that improve cognitive, social, and language skills.

Monozygotic versus dizygotic twin and sibling studies suggest a strong genetic basis for ASD (172–174). However, mutations in any individual gene account for less than 1% of ASD cases (60,172). Genetic sequencing analyses can only identify a potentially causative genetic abnormality in ~25% of clinical ASD diagnoses (60,118,172). While genome-wide association studies (GWAS) also support common genetic variants in ASD, the complexity and heterogeneity of ASD has been a major challenge (65,175,176). Evidence for environmental risk factors in ASD point to *in utero* maternal exposures such as air pollution, fever, or asthma and nutrients, specifically the absence of periconceptional prenatal vitamin intake (70,120,177,178).

Maternal prenatal vitamins, which contain high levels of folic acid and other additional B vitamins, protect offspring by up to 70% for neural tube defects (179–184), and correlate with an overall 40% reduction in ASD risk if taken during the first month of pregnancy (P1) (70,119,120). This finding was replicated with a large prospective study in Norway including over 80,000 pregnancies (119).

DNA methylation shows dynamic changes during fetal development (7,67,166) and contains the molecular memory of *in utero* experiences such as maternal nutrition (180,185). Term placenta is an accessible fetal tissue that maintains the distinctive embryonic bimodal DNA methylation pattern, in which expressed genes are marked by higher methylation levels (161,186). Placenta therefore offers a unique window to study DNA methylation patterns that may reflect altered fetal development relevant to ASD genetic risk (166,186–188). Specifically, a recent study of polygenic risk scores for schizophrenia demonstrated a significant interaction of genetic risk with maternal perinatal environmental factors that affected placental gene expression (169). Previous analyses of DNA methylation patterns in placenta samples from a high-risk ASD cohort also identified an association between self-reported use of lawn and garden pesticides and large-scale changes in DNA methylation patterns, and identified a putative enhancer of the *DLL1* gene as differentially methylated in ASD (133,171).

Here, we continue the epigenetic investigation of ASD risk through the novel approach of identifying differentially methylated regions (DMRs) in whole methylomes from placenta samples from male children later diagnosed with ASD compared to children with typical

development (TD) controls. Two genome-wide significant ASD-associated DMRs at *CYP2E1* and *IRS2* were further validated and investigated for effects of genotype, RNA expression, and protein levels as well as interactions with preconception prenatal vitamin use. Understanding the epigenetic patterns of ASD associated with maternal prenatal vitamin use in placenta could lead to the development of preventative and therapeutic early interventions for high-risk children with ASD.

Results

Placenta ASD DMRs discriminate ASD from TD samples.

To identify novel differentially methylated gene loci between ASD and TD, a differentially methylated region (DMR) bioinformatic analysis was performed on placenta whole genome bisulfite sequencing (WGBS) data (133,171). 400 DMRs were identified with a threshold of > 10% methylation difference between ASD and TD groups, and these were associated with 596 genes using the Genomics Regions Enrichment of Annotations Tool (GREAT) on the default association settings (**Fig. 2.1A**, **Supplementary Table 2.1**). There was no bias for gene length in the ASD DMR associated genes compared to all human genes (**Supplementary Fig. 2.1**). 296 DMRs were hypermethylated, while 104 DMRs were hypomethylated in ASD compared to TD placenta (**Fig. 2.1A**). Principal component analysis (PCA) using methylation levels for each sample over the 400 DMRs demonstrated a clear separation of placental samples by child outcome of ASD versus TD (**Fig. 2.1B**). In addition, most ASD DMRs showed a highly significant association with Mullen Scales of Early Learning (MSEL) scores and autism severity score from

Autism Diagnostic Observation Schedule (ADOS), but not with potential confounding variables (**Supplementary Table 2.2**, **Supplementary Fig. 2.2**). Demographic and clinical variables of children and their mothers in the MARBLES study were not significantly associated with child diagnosis (**Supplementary Table 2.3**). Furthermore, there was no significant enrichment between placental cell type specific genes from single-cell RNA-seq data (189) and genes identified associated with ASD DMRs in placenta, suggesting that the differences were not due to major shifts in cell type composition (**Supplementary Fig. 2.3, Supplementary Table 2.4**).

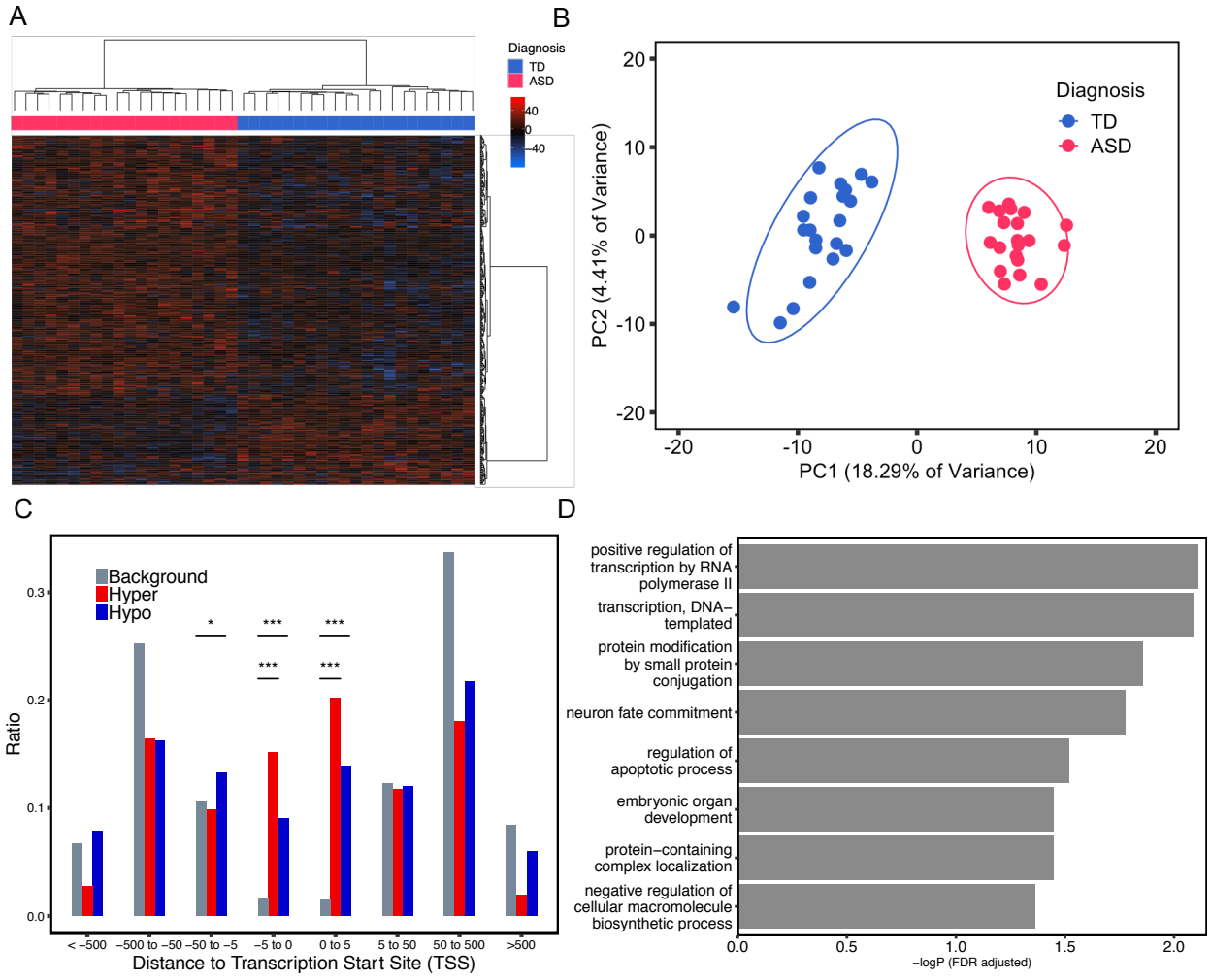


Fig. 2.1. Differentially methylated regions (DMRs) in placenta distinguished ASD from typical development (TD) child outcomes.

A. Heatmap and hierarchical clustering of 20 ASD versus 21 TD placental samples using methylation levels in the 400 identified ASD DMRs. Percent methylation for each sample relative to the mean methylation at each ASD DMR was plotted as a heatmap, with black representing no difference, hyper-methylation (red) and hypo-methylation (blue). Columns were clustered by child outcome, ASD (red) or TD (blue), while rows were clustered by methylation direction. **B**. Principal component analysis (PCA) of TD vs ASD placental samples on the basis of methylation at 400 ASD DMRs. Ellipses represent the 95% confidence interval for each group. The non-overlapping ellipses showed a significant difference between ASD and TD for these DMRs' methylation level ($p < 0.05$). **C**. Location relative to genes for hypermethylated (red) or hypomethylated (blue) ASD DMRs compared to background (grey). Distributions of locations relative to transcription start sites (TSS) are shown on the x-axis. The ratio plotted on the y-axis is calculated by the number of genes at each binned location divided by the total number of genes (**Supplementary Table 2.5**). *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$ by Fisher's exact test. **D**. Bar graph represents the significant results from gene ontology and pathway enrichment analysis of ASD DMRs associated genes compared to background by Fisher's exact test (FDR adjusted -log *p*-value, x-axis).

<u>Placenta ASD DMRs were enriched for transcription start sites and genes that function in transcriptional regulation and neuronal fate.</u>

To further study the location and function of ASD DMRs in placenta, we calculated the location of each ASD DMR relative to the assigned gene's transcription start site (TSS) (**Supplementary Table 2.5**). Both hyper- and hypomethylated ASD DMRs were enriched within 5 kb on either

side of TSS compared with background regions (**Fig. 2.1C**). Gene ontology (GO) analysis of ASD DMRs genes revealed significant enrichment for functions in transcription, protein modification, embryonic organ development, and neuron fate commitment by Fisher's exact test after false discovery rate (FDR) multiple test correction (**Fig. 2.1D, Supplementary Table 2.6**).

Placenta DMR genes were enriched in ASD but not ID risk genes

To test a hypothesized overlap between epigenetic and genetic ASD risk loci observed previously in ASD and neurodevelopmental disorder brain tissues (190), we investigated the possible overlap of placenta ASD DMR genes with identified genetic risk factors for ASD and other types of intellectual disability (ID). First, the curated Simons Foundation Autism Research Initiative (SFARI) gene list was separated into six categories based on SFARI ASD gene scores (191). The entire list of SFARI genes as well as the high confidence gene list both showed significant overlap with placenta ASD DMR genes (**Fig. 2.2A**, **Supplementary Table 2.7**). The 39 genes in common between the SFARI gene list and placenta ASD DMRs were significantly enriched for functions in positive regulation of histone H3K4 methylation, multicellular organ development, and system development. Second, high risk ASD genes from Sanders *et al*. (176) and likely gene-disrupting (LGD) recurrent mutations and missense mutation on *de novo* mutations to ASD gene lists from whole genome exome sequencing (65) were also significantly enriched for placental ASD DMRs. In contrast, no significant enrichment was observed for placental ASD DMRs with ID, Alzheimer's disease, or lung cancer genetic risk (192) or a random set of 400 genomic regions mapped to 600 genes (**Fig. 2.2A**, **Supplementary Table 2.7**). When placental ASD DMRs were separated by direction of change, hypomethylated ASD DMRs

exhibited more categories of significant enrichment with ASD genetic risk compared with hypermethylated ASD DMRs (**Supplementary Fig. 2.4**). Differential expressed genes in ASD postmortem brain were extracted from multiple studies (72,193–195). Placenta ASD DMR genes significantly overlapped with ASD cortex differentially expressed genes in the largest of these studies (72) (**Supplementary Fig. 2.5**, **Supplementary Table 2.8**).

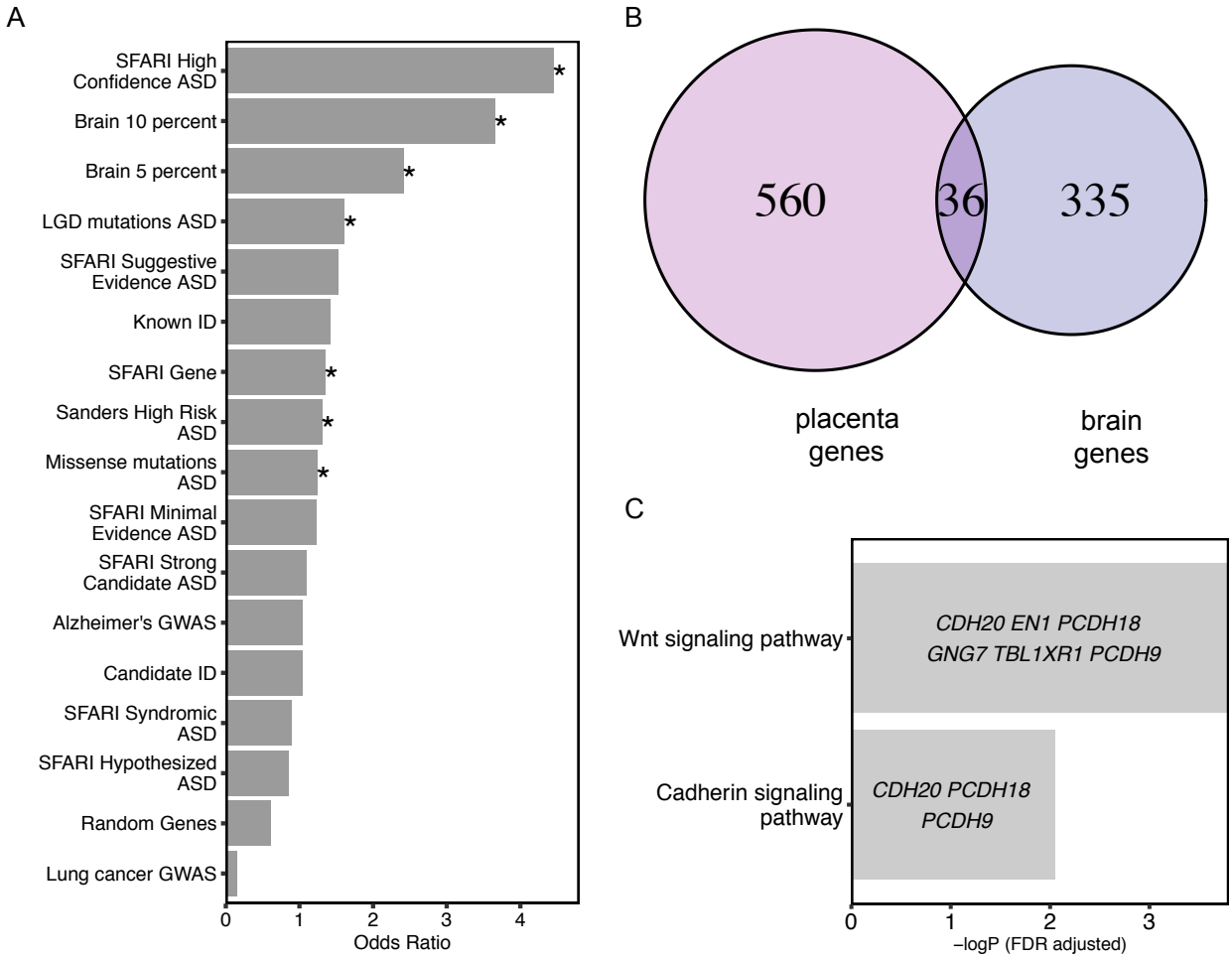


Figure 2.2. Placenta ASD DMR genes overlapped with ASD DMR associated genes from postmortem brain and known genetic risk for ASD, but not for other disorders.

A. Placenta ASD DMR associated genes were compared for significant overlap with ASD DMR genes identified from ASD postmortem brain (140) (based on 10% or 5% methylation difference

cutoffs), as well as multiple curated gene lists of ASD, intellectual disability, or unrelated disorder genetic risks, or a randomly generated gene list (*$p < 0.05$ FDR corrected two-tailed Fisher's exact test, ranked by odds ratio). SFARI: Simons Foundation Autism Research Initiative (191), LGD: likely gene disrupting mutation, ASD: autism spectrum disorder, Alzheimer: Alzheimer's Disease, ID: intellectual disability. **B**. Venn diagram represents the significant overlap of 36 genes associated with placenta ASD DMRs and brain ASD DMRs based on 10% methylation differences in brain between ASD versus TD by Fisher's exact test ($p < 0.001$***) (**Supplementary Table 2.7**). Methylation data from human postmortem brain was obtained from previous published datasets, GSE8154 (ASD and TD) (190). **C**. Gene ontology and pathway analysis on the 36 genes in common between placenta ASD DMRs and brain ASD DMRs associated genes. Enrichment tests were done on Fisher's exact test with FDR 0.05 correction. Genes in each gene ontology term are shown within each bar. **D**. Venn diagram represents significant overlap of 439 genes between placenta ASD DMRs and DMRs from brain with 5% methylation difference between ASD and TD (Fisher's exact test, $p < 0.001$***) (**Supplementary Table 2.7**). **E.** Multiple developmental pathways were significantly enriched on the overlapped 439 genes with Fisher's exact test after FDR 0.05 correction.

<u>Placenta ASD DMR genes significantly overlapped with brain ASD DMRs that were enriched at Wnt and cadherin signaling pathways and developmental functions.</u>

From a prior methylation analysis in brain frontal cortex, 210 ASD discriminating DMRs were identified (10% methylation difference between ASD and TD), which mapped to 371 genes (190). A significant overlap of ASD DMR genes was observed between placenta and brain

(Fisher's exact test, *p*-value < 0.001), with 36 genes in common at a 10% difference in ASD versus TD brain (**Fig. 2.2B**, **Supplementary Table 2.7**). Those 36 genes were significantly enriched for functions in the Wnt signaling and cadherin signaling pathways (**Fig. 2.2C**). Of these shared genes four overlapped with SFARI genetic risk: *GADD45B*, *MC4R*, *PCDH9,* and *TBL1XR1*. With a more inclusive 5% methylation difference in brain that has been described previously (132,190) , 74% of the placental ASD DMR genes significantly overlapped with those identified in brain (Fisher's exact test, *p*-value < 0.001) (**Fig. 2.2D**). Pathways significantly associated with this more extended list included neuronal fate commitment, neurogenesis, and neuronal differentiation, as well as embryonic developmental terms such as tube development (**Fig. 2.2E**). Among shared 439 genes, 32 genes were also identified as SFARI ASD genetics risk genes, including *AUTS2*, *KMT2A*, *SETBP1*, and *TBR1* (**Supplementary Table 2.7**).

Placenta ASD DMRs were enriched for placental and brain active promoter H3K4me3 peaks, promoter flanking regions, and CpG shores.

To functionally annotate the ASD DMRs identified by placenta WGBS, multiple histone modification ChIP-seq peaks and chromatin state predictions from multiple tissue types in the Roadmap Epigenomics Projects were compared to ASD DMR chromosomal locations for enrichment compared to background regions (196). Placental ASD DMRs were significantly enriched at epigenomic regulatory regions, specifically H3K4me3 and H3K4me1 marks across multiple tissues, although placental H3K4me3 marks showed the strongest (odds ratio = 17.08, FDR q < 1.8E-42) and brain H3K4me3 marks showed the second strongest enrichment (odds ratio = 13.75, FDR q < 3.55E-31) (**Fig. 2.3A**). Importantly, we also found a significant overlap

between genes we identified by placenta ASD DMR and those identified as differential in ASD prefrontal neuron H3K4me3 peaks (Fisher's exact test, *p*-value < 0.05) (**Supplementary Fig. 2.6**) (197). Next, we overlapped ASD DMRs with published chromatin state predictions that use histone modification ChIP-seq data to annotate the genome into 15 functional states (chromHMM) (198). Placental ASD DMRs showed significant enrichment in regions flanking transcription start site (TssAFlnk) and transcription start site (TssA) compared to background over multiple tissues (**Fig. 2.3B**). When separated by directional change in ASD, both hyper- and hypomethylated ASD DMRs were significantly enriched for H3K4me3 peaks, transcription start sites and their flanking regions, as well as enhancers (**Supplementary Fig. 2.7**).

We also separated the genome into four parts relative to CpG island location (199–201). CpG shores were defined as the region within 2 kb on both sides of CpG islands, while another 2 kb extension from the shores were defined as CpG shelves. The remaining regions were defined as "open sea". Placental ASD DMRs showed significant enrichment at CpG shores, and hypermethylated ASD DMRs more significantly overlapped CpG islands compared with hypomethylated DMRs (**Fig. 2.3C**).

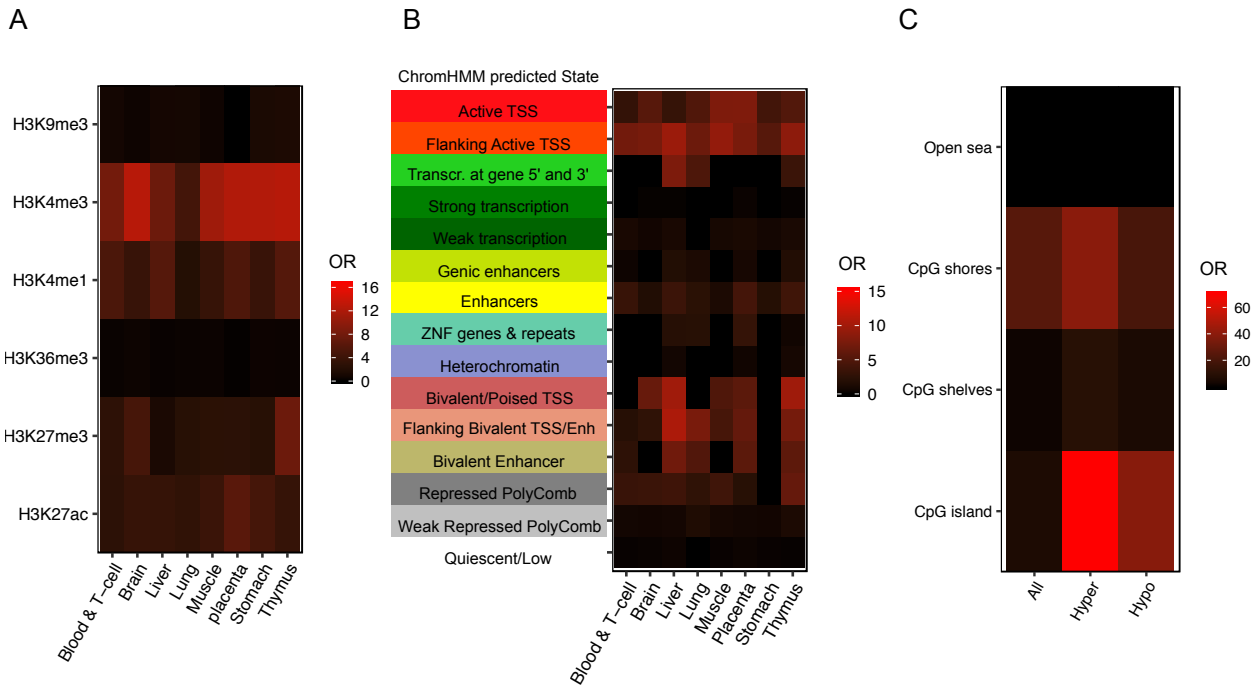


Figure 2.3. Placenta ASD DMRs were enriched at H3K4me3 regions, flanking promoter regions, and CpG shores.

A. Placenta ASD DMRs were examined for enrichment with histone modification ChIP-seq peaks from the Epigenome Roadmap using the LOLA package. Enrichments are plotted as the odds ratio) in a heat map for each of 8 different tissue types and 6 types of modified histone marks (202) . **B**. Enrichment tests on chromatin states from chromHMM categories in the Epigenome Roadmap and placental ASD DMRs from this study were performed using LOLA, with each row representing a different ChromHMM predicted state and each column a single tissue type. **C**. Placenta ASD DMRs (categorized as all, hypermethylated, or hypomethylated in ASD) were tested for enrichment based on CpG island location. The human genome was separated into CpG islands, CpG shores, CpG shelves and open sea.

Two genome-wide significant placental ASD DMRs at *CYP2E1* and *IRS2* validate by pyrosequencing and correlated with gene expression.

Two of the 400 ASD DMRs identified in ASD placenta reached genome-wide significance by family-wide error rate (FWER), including chr10: 133527713-133529507, located inside *CYP2E1* (cytochrome P450 2E1), and chr13: 109781111-109782389 located inside *IRS2* (insulin receptor substrate 2) (**Fig. 2.4**). The *CYP2E1* DMR was located after the first exon, included the first intron and part of the second exon, and was hypomethylated in ASD versus TD (**Fig. 2.4A**). The *IRS2* DMR, spanning the end of the first exon and the beginning of first intron and was hypermethylated in ASD versus TD (**Fig. 2.4B**). Both *CYP2E1* and *IRS2* were also present in the gene lists overlapping with brain ASD DMR related genes and high risk ASD genes (**Fig. 2.2A, Supplementary Table 2.7**) (176).

Pyrosequencing was performed as an independent method to verify methylation differences between ASD and TD placental samples at *CYP2E1* and *IRS2* DMRs (**Supplementary Table 2.9**). For the *CYP2E1* DMR, there was a significant difference in average percent methylation detected by pyrosequencing between ASD and TD samples (**Fig. 2.4C**). 13 CpG sites were included in the *CYP2E1* DMR pyrosequencing test, and all but 2 also showed individually significant differences between ASD and TD after false discovery rate (FDR) correction (**Supplementary Table 2.9**, **Supplementary Fig. 2.8A**). Pyrosequencing also confirmed a significant difference between ASD and TD average percent methylation at the *IRS2* DMR (**Fig. 2.4D**) and all of the 11 CpG sites individually assayed at *IRS2* (**Supplementary Table 2.9**, **Supplementary Fig. 2.8B**). There were no significance differences in the percent methylation

variances of ASD and TD samples on both *CYP2E1* DMR (two-tailed F-test, *p*-value = 0.8) and *IRS2* DMR (two-tailed F-test, *p*-value = 0.7). To test the relevance of these loci to females, 10 female placental samples were also verified by pyrosequencing and showed a significance difference between ASD and TD in Percent methylation at *CYP2E1* and *IRS2* DMRs in the same direction as that observed in males (**Supplementary Fig. 2.9**).

While MARBLES placenta samples were not collected in a manner conducive to RNA stability for gene expression analyses, we were able to examine expression level of both *CYP2E1* and *IRS2* in MARBLES umbilical cord blood during the same time period from an Affymetrix Human Gene 2.0 array analysis in a related study (203). No significant differences by ASD diagnosis were observed in cord blood *CYP2E1* or *IRS2* expression or placental IRS2 protein levels, although trends were consistent with the direction of methylation differences (**Supplementary Fig 2.10-2.11**). Furthermore, both *CYP2E1* and *IRS2* were previously identified as differentially expressed genes in ASD brain from three previous studies (193–195) as well as in a meta-analysis of ASD compared with TD in human male cortex (dbMDEGA database) (204). (**Supplementary Fig. 2.5**, **Supplementary Table 2.8**). Together, these results suggest that methylation differences identified in ASD placenta are also relevant to differential gene expression patterns identified in brain.

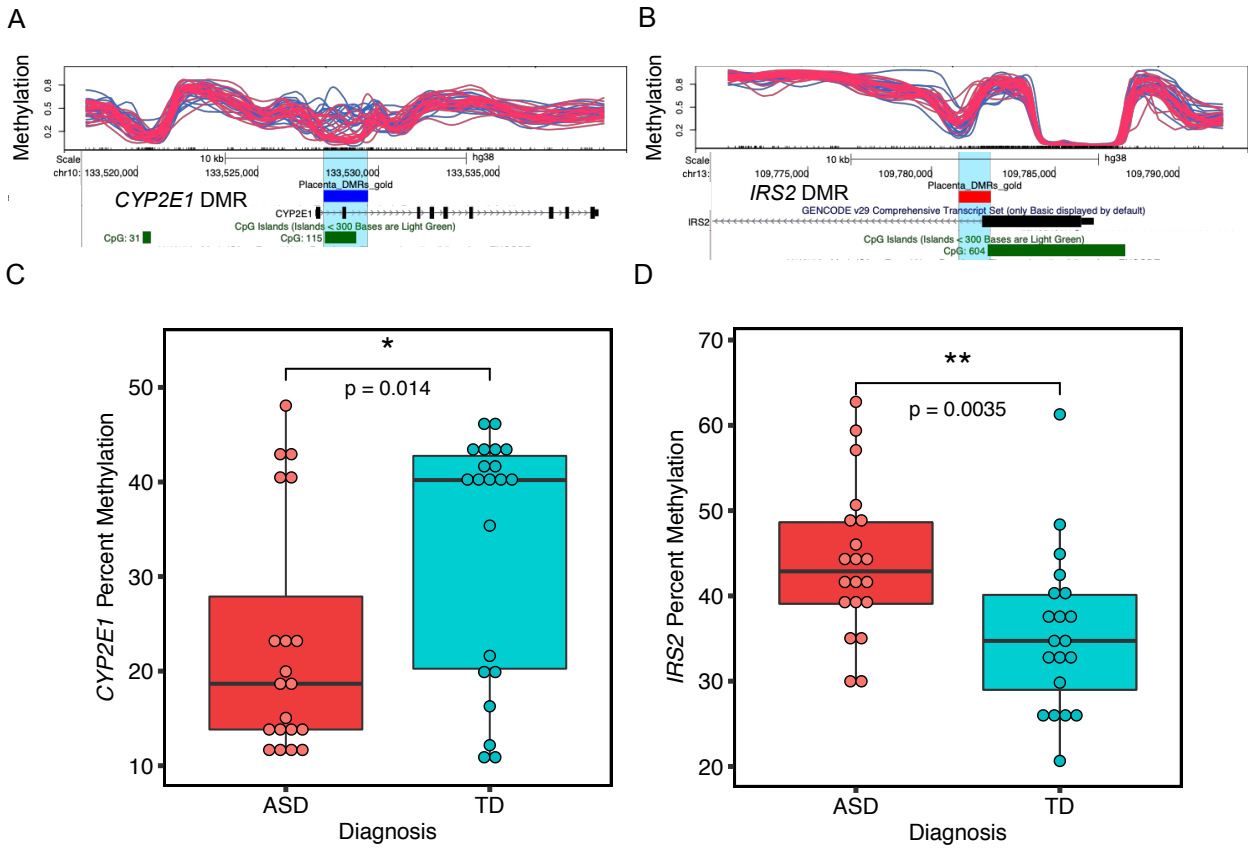


Figure 2.4. Two genome-wide significant placental DMRs located at *CYP2E1* and *IRS2* were validated by pyrosequencing.

A and **B** show the location relative to genes, and CpG islands of the two genome-wide significant DMRs (highlighted in pink and blue) in the UCSC Genome Browser. In the upper tracks, each line represents percent methylation (y-axis) of a single individual by WGBS analysis. Blue lines represent TD and red lines represent ASD samples. **A**. Hypomethylated DMR at *CYP2E1* with 10 kb upstream and 10 kb downstream. **B**. Hypermethylated DMR at *IRS2* with 10 kb upstream and 10 kb downstream. **C**. The *CYP2E1* DMR percent methylation was significantly associated with child outcome and verified by pyrosequencing (two-tailed t-test, *p*-value = 0.014). The y-axis represents the average percent DNA methylation across the DMR regions from pyrosequencing. Each dot represented one sample. **D**. Pyrosequencing validation on *IRS2*

DMR's methylation with child outcome (two-tailed t-test, *p*-value = 0.0035). *$p < 0.05$, **$p < 0.01$.

CYP2E1 but not *IRS2* DMR methylation levels were associated with common variants within each DMR.

We performed Sanger sequencing within the *CYP2E1* and *IRS2* ASD DMRs to identify single nucleotide polymorphisms (SNPs) located inside each DMR that could explain some of the methylation differences. Two SNPs (rs943975, rs1536828) were identified within the boundaries of the *CYP2E1* DMR in the 41 placenta samples (**Supplementary Table 2.10**). A significant association between rs1536828 (but not rs943975) genotype and *CYP2E1* DMR percent methylation was observed, with samples homozygous for the minor allele (G/G) showing the lowest methylation (**Fig. 2.5A).** This significant genotype-methylation association remained when the minor (GG) and heterozygous (CG) *CYP2E1* genotypes were combined and compared to the major (CC) samples (**Supplementary Fig. 2.12**). Further confirming the effect of *CYP2E1* genotype on methylation, 5 placental samples of each rs1536828 genotype showed a similar significant association of genotype with *CYP2E1* DMR percent methylation by pyrosequencing (**Supplementary Fig. 2.13**). A single informative SNP (rs9301411) was also identified within the *IRS2* DMR (**Supplementary Table 2.10**) but was not significantly associated with methylation level (**Fig. 2.5B**). However, genotypes at rs943975 (minor allele frequency: 0.4), rs1536828 (minor allele frequency: 0.32) and rs9301411 (minor allele frequency: 0.08) were not significantly associated with diagnosis. In addition, these SNPs were not located inside

any known transcription factor motifs identified by MEME(205) and there were no differences on the MEME motif structure on the samples with and without SNPs (**Supplementary Fig. 2.14**).

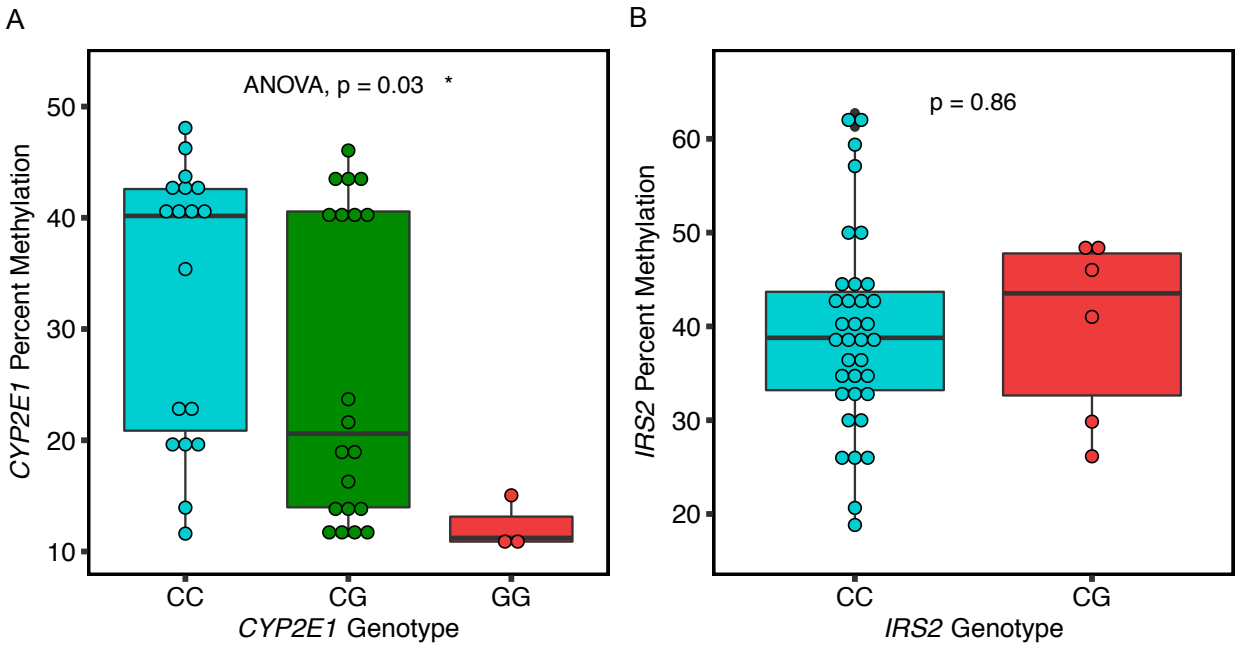


Figure 2.5. *Cis* genotype was significantly associated with *CYP2E1* but not *IRS2* DMR methylation levels.

A. *CYP2E1* genotype at rs1536828 within the ASD DMR was significantly associated with *CYP2E1* DMR average percent methylation tested by ANOVA (*p*-value = 0.03). **B**. *IRS2* genotype at rs9301411 within the ASD DMR was not significantly associated with *IRS2* DMR methylation by two-tailed t-test (*p*-value = 0.86).

Preconception prenatal vitamin use corresponded to protective placental DNA methylation patterns at *CYP2E1*, *IRS2*, and genome-wide.

Placental samples from mothers who took prenatal vitamins during the first month of pregnancy showed a trend for higher *CYP2E1* DMR methylation that was not significant, but in

the same direction expected for protection from ASD (**Fig. 2.6A**). At the *IRS2* DMR, however, there was a significant association with maternal prenatal vitamin use and lower methylation, also in ASD-protective direction (**Fig. 2.6B**).

To further explore the relationship between placental methylation patterns influenced by prenatal vitamin use in the first month of pregnancy, placental methylomes were analyzed for DMRs by prenatal vitamin use in the first month of pregnancy (PreVitM1) with more than 10% methylation difference, and 376 DMRs were identified in over 587 genes (**Supplementary Table 2.11**). 60 genes overlapped between PreVitM1 DMRs and ASD DMRs in placenta (**Supplementary Table 2.11**, **Fig. 2.6C**). Gene ontology analysis showed that genes common to PreVitM1 and ASD DMRs were significantly enriched for functions in neuron fate commitment, transcription regulation, central nervous system development, and regulation of phosphatidylinositol 3-kinase activity (**Fig. 2.6D**).

To further investigate the potential inter-relatedness of diagnosis, prenatal vitamin use, and *cis* genotype on methylation at *CYP2E1* and *IRS2* DMRs, we calculated associations between each factor and methylation separately by two-tailed t-test or ANOVA, as well as two-way diagnosis and PreVitM1; diagnosis and genotype; genotype and PreVitM1 by Pearson's chi-squared test. No significant association were found among each of the three factors and each DMR methylation level by ANOVA (**Supplementary Table 2.12**).

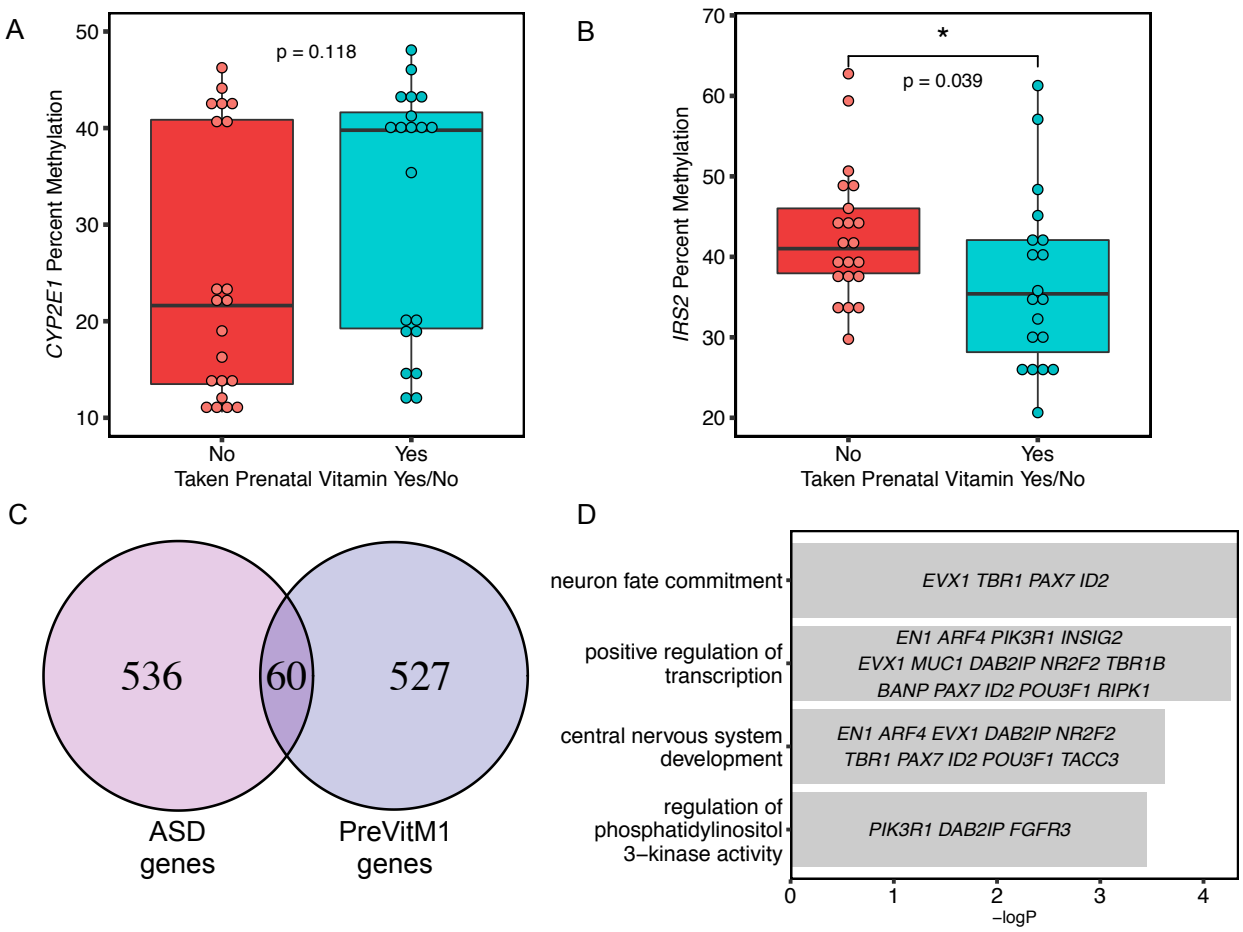


Figure 2.6. Preconception prenatal vitamin use was a significant modifier of *IRS2* methylation and associated DMRs overlapped ASD DMRs in placenta.

For **A** and **B**, the x-axis represents maternal prenatal vitamins use during the first month of pregnancy, while the y-axis shows the percent methylation. **A**. *CYP2E1* DMR methylation was not significantly altered by P1 prenatal vitamin use. **B**. Higher percent methylation at *IRS2* DMR was significantly associated with not taking prenatal vitamins at P1 (two-tailed t-test, *p*-value = 0.039), which is in the same direction as ASD risk. **C**. DMRs identified based on P1 prenatal vitamins use were associated with 587 genes, which showed a significant overlap with ASD DMR associated genes (Fisher's exact test, *p*-value < 2.528e-16). **D**. Gene ontology and pathway analysis was performed on the overlapped gene list (60 genes) (**Supplementary Table 2.11**)

between placenta ASD DMR and P1 prenatal vitamin DMR associated genes for enrichment by Fisher's exact test with -log (p-value) represented on the x-axis. Genes in each gene ontology (GO) term are shown within each bar.

Discussion

This study utilized whole methylome analysis of prospectively stored placenta samples in a high risk ASD cohort to bioinformatically identify novel gene loci that may be useful for understanding and predicting ASD risk. This unbiased analysis of ASD differentially methylated regions in placenta tissue resulted in several novel findings. First, the 596 genes identified from 400 placental ASD DMRs significantly overlapped with genetic risk for ASD from curated databases and gene functions in neurons. Second, two genome-wide significant placental ASD DMRs at *CYP2E1* and *IRS2* were discovered that were validated by pyrosequencing and also overlapped with ASD- associated genetic variation and gene expression changes. Lastly, we investigated genotype and nutrient factors correlating with methylation at *CYP2E1* and *IRS2*, demonstrating specific effects for *cis* genotype and diagnosis at *CYP2E1* and prenatal vitamin use at *IRS2*. These results therefore suggest that DNA methylation patterns in placenta provide a link between genetics, environment, and fetal epigenetic programming, which could reflect early development relevant to the complex etiology of ASD. The epigenomic signature of ASD in placenta also provides important insights into gene functions, pathways, gene-environment interactions, and potential biomarkers that may be useful in improving early detection of ASD.

This study is the first to identify 400 potential ASD DMRs that distinguish between ASD and TD placenta samples and highlights specific locations and gene functions of differential methylation in placental samples from children with ASD. First, these placental ASD DMRs were highly enriched around transcription start sites and H3K4me3 marks that are clear marks of gene regulatory functions (206,207). Furthermore, gene ontology analysis of the 596 genes mapped to placental ASD DMRs pointed to enriched gene functions in transcription, neuron fate, and embryonic development, which were expected based on previous studies (171,208,209).

Genes with ASD DMRs in both placenta and brain were enriched for Wnt and cadherin signaling pathways. Wnt signaling is important in embryogenesis, tissue regeneration, and neurodevelopment (210), while cadherin signaling plays a vital role in connecting major intracellular signaling pathways with adhesion protein complexes (211,212). Our results therefore complement previous studies that have shown the importance of Wnt and cadherin pathways in the etiology of ASD (213–215). We also replicated our previous finding of differential methylation at *DLL1* in ASD placenta (171) (**Supplementary Tables 2.1 and 2.7**). *DLL1* encodes a ligand of Notch, activated by Wnt signaling (216).

When overlapped with datasets of genetic risk for neurodevelopmental disorders including ASD and intellectual disability (65,176,191), placenta ASD DMRs were significantly enriched for ASD but not for intellectual disability genetic risk, illustrating the specificity of the ASD DMRs

identified in our study. The highest overlap of ASD DMR associated genes was with the SFARI high confidence genes, including *KMT2A*, *MYT1L*, and *TBR1* (191).

Our study identified novel methylation differences at *CYP2E1* and *IRS2*, which exhibited genome-wide significant differences between ASD and TD. Both *CYP2E1* and *IRS2* are identified as ASD genetic risk genes in multiple databases related to ASD genetic risk across different tissues and populations (176,190). Both *CYP2E1* and *IRS2* DMRs are located close to the TSS site at CpG shore intragenic regions, which is also consistent with the enrichment for TSS flanking regions and H3K4me3 promoter marks in the 400 ASD DMRs. In our prior studies of differential methylation associated with *UBE3A* duplication in Dup15q syndrome, a genetic cause of ASD, both *CYP2E1* and *IRS2* were identified with differential methylation and H3K4me3 peaks in a Dup15q cell line model, and differential methylation in Dup15q brain (132,217). Furthermore, *CYP2E1* and *IRS2* both showed differential H3K4me3 marks at their promoters in ASD vs control in the prefrontal cortex (197). Structural variants and SNPs in cis-regulatory elements also showed significant contribution to ASD (141,218). In contrast, neither *CYP2E1* nor *IRS2* were included as significant differential splicing events for ASD (72,219).

CYP2E1 encodes a member of the cytochrome P450 superfamily that is involved in the metabolism of drugs (220). Previous studies showed those proteins essential for embryonic development in human, rat and zebrafish (221–225). We observed a significant association between methylation and *cis* genotype at the *CYP2E1* DMR, a finding which is consistent with the identification of this locus in a screen for human metastable epialleles variability between

individuals (226). In addition, in immune models of ASD, maternal interleukin-6 (IL6) crosses the placenta, disrupting development of hippocampal spatial learning (227–229). Previous studies showed that IL6 inhibits *CYP1A1*, *CYP1A2* and *CYP2E1* expression (230–233), consistent with the lower methylation and expression levels in ASD versus TD observed in our study. In addition, *CYP2E1* expression is transcriptionally regulated by the JAK2/STAT3 pathway, providing a potential convergent pathway with *IRS2* (**Fig. 2.7**) (230).

IRS2 encodes for insulin receptor substrate 2, a cytoplasmic signaling molecule that mediates the effects of insulin and insulin-like growth factor 1 (*IGF1*) (234) and cytokine receptors (235). A GWAS noise reduction (GWAS-NR) method to correct for false-positive association with ASD identified cadherin and signaling transduction pathways that included *IRS2* as high confidence ASD genes (236). IRS2 has a phosphotyrosine-binding domain which contributes to the intracellular affinity to cell membrane receptors (**Fig. 2.7**) (235). PI3K/AKT/mTOR and MAPK signaling pathways are linked with IRS2 in the regulation of protein synthesis and cell proliferation (237). Furthermore, a previous study using a reverse pathway genetic approach identified the MAPK signaling pathway associated with ASD (238). When activated by cytokine and hormone receptors, IRS2 stimulates JAK2, leading to STAT and MAPK signaling activation (237,239). The insulin-like growth factors (IGF1) pathway, which includes IRS2, also mediates *de novo* DNA methylation by DNA methyltransferase (DNMT) through AKT (240). This pathway may explain why methylation at the *IRS2* DMR was sensitive to maternal prenatal vitamin intake, since *IRS2* stimulates the mTOR (mechanistic target of rapamycin) pathway, which responds to nutrients and growth factors signaling to regulate protein synthesis (239). The link

between epigenetic alterations in *IRS2* and risk for ASD is particularly intriguing given a growing body of epidemiologic evidence demonstrating higher ASD risk in offspring born to mothers who experienced diabetes during pregnancy, including some very large and methodologically sound studies with clinical diagnoses of both maternal diabetes and child ASD (241).

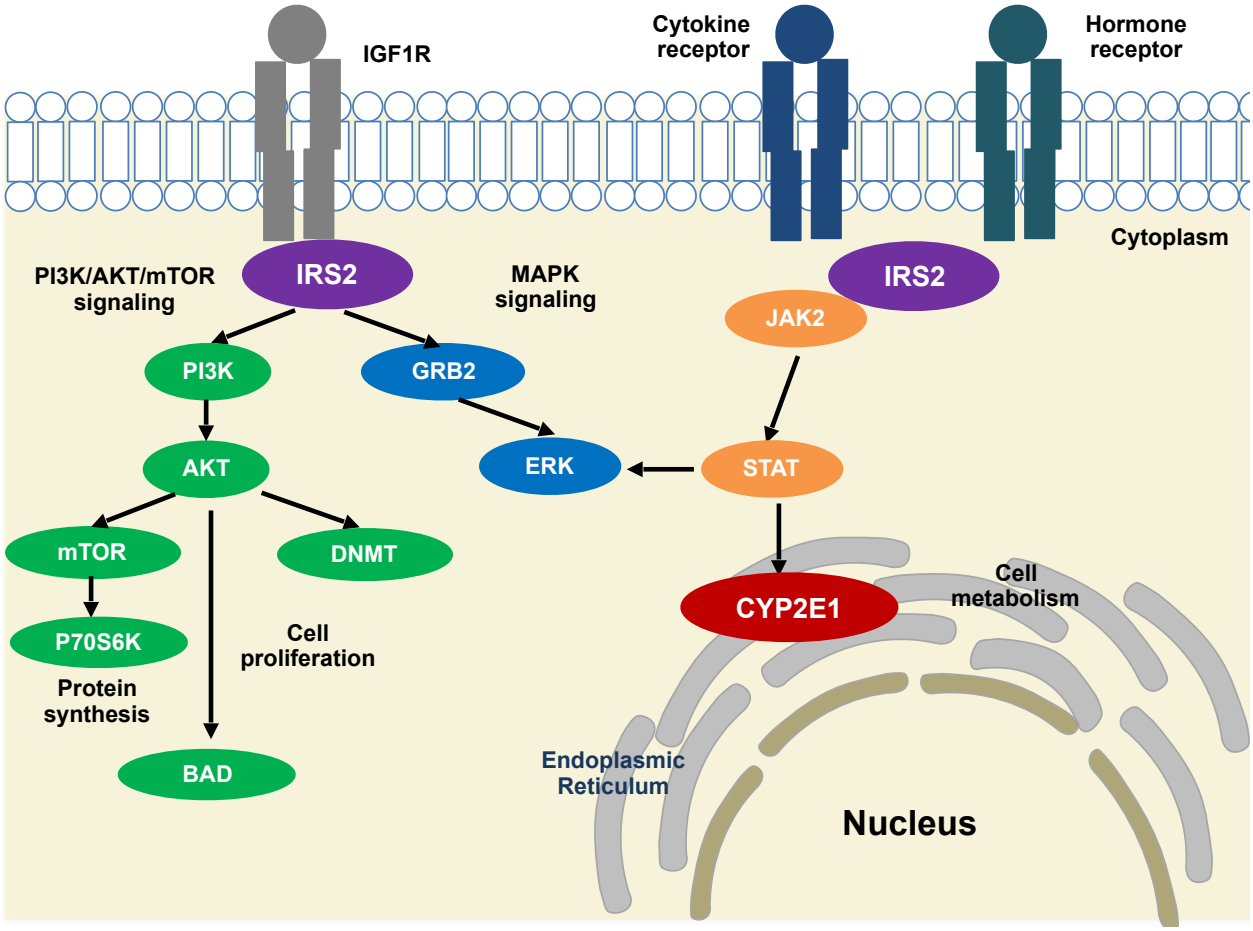


Figure 2.7. Potential pathway convergence of proteins encoded by both ASD DMRs. IRS2 interacts with transmembrane protein insulin-like growth factor receptor (IGF1R) at the intracellular membrane, resulting in activation of the PI3K/AKT/mTOR and MAPK signaling pathways involved in protein synthesis, cell proliferation and gene expression (237) . An AKT-mediated ubiquitin pathway leads to *de novo* DNA methylation changes by DNMT (240). IRS2

also interacts with cytokine and hormone receptors and induces JAK2/STAT3 signaling (237,239). STAT activation leads to *CYP2E1* localization at the endoplasmic reticulum, changing cellular metabolism (230).

We did not observe any significant associations between other potential cofounders such as maternal age, pregnancy BMI, gestational age, or pre-eclampsia with ASD diagnosis in the MARBLES study (133). Cell type heterogeneity in the placenta may complicate the interpretation of our results, however, placental ASD DMRs were not significantly associated with cell-type markers and our previous study did not detect differences in methylation levels by placental region at specific gene loci (171). This study serves as a proof-of-principle that placenta methylation patterns detected by WGBS may be informative in ASD. Replication with additional samples, on female samples, and other similar prospective cohorts, and improved sequencing and bioinformatic strategies will be important in future studies. Approaches such as methylation quantitative trait loci (meQTLs) analysis will be important in future studies with larger sample sizes for identifying possible common genetic variants associated with placental ASD DMRs. In addition, functional analyses on JAK, mTOR and MAPK signaling pathways in ASD placenta and brain would be important to provide mechanical insight.

In conclusion, we identified two high confidence genes differentially methylated in ASD from an unbiased analysis of DNA methylation in placenta from high-risk pregnancies and investigated possible genetic and environmental modifiers of methylation at both loci. Methylation levels at the *CYP2E1* DMR were associated with genotype, while the methylation levels at the *IRS2* DMR

were associated with prenatal vitamin use. Our results are consistent with a previous study using the Illumina 450K array, which showed that both genetic and environmental effects influence DNA methylation levels (174). Placenta reflects the essential interface between the fetus and mother, mediating the impacts of endocrine and growth factors in the maternal environment on fetal development (242,243). Both *CYP2E1* and *IRS2* are related to protein synthesis, cell proliferation, and cell metabolism, consistent with previous studies of convergent gene pathways in ASD (176,190,204,219,244). These results therefore provide evidence that placental methylation levels reflect the intersection of genetic and environmental risk and protective factors that are expected to be useful for early intervention and prevention of ASD.

Methods

MARBLES study design, sample selection, and DNA isolation

The Markers of Autism Risk in Babies: Learning Early Signs (MARBLES) study design was described in a previous publication (79). In MARBLES, mothers of at least one child with confirmed ASD who were pregnant or planning a pregnancy were recruited in the Northern California area. Inclusion criteria for the study were: 1) mother or father has one or more biological child(ren) with ASD; 2) mother is 18 years or older; 3) mother is pregnant; 4) mother speaks, reads, and understands English sufficiently to complete the protocol and the younger sibling will be taught to speak English; 5) mother lives within 2.5 hours of the Davis/Sacramento, California region at time of enrollment. With shared genetics, the next child has a 15-fold higher risk for developing ASD compared to the general population (79). Demographic, diet, lifestyle, environmental, and medical information were prospectively

collected through telephone-assisted interviews and mailed questionnaires throughout pregnancy and the postnatal period. Infants received standardized neurodevelopmental assessments beginning at 6 months and concluding at 3 years of age (79). Diagnostic assessments at 3 years included the gold standard Autism Diagnostic Observation Schedule (ADOS) (245), the Autism Diagnostic Interview-Revised (ADI-R) (246), and the Mullen Scales of Early Learning (MSEL) (247). Participants were classified into outcome groups including ASD and Typical Development (TD), based on a previously published algorithm that uses ADOS and MSEL scores (248,249). Children with ASD outcomes have scores over the ADOS cutoff and meet DSM-5 criteria for ASD. Children with TD outcomes have all MSEL scores within 2.0 SD and no more than one MSEL subscale that is 1.5 SD below the normative mean and scores on the ADOS at least three more points below the ASD cutoff. This study utilized 41 male MARBLES placenta samples, including 20 samples from children later diagnosed with ASD and 21 children determined to have TD, matched for enrollment time frame and date of birth. DNA was isolated from 50-100 mg frozen placental tissues (20 ASD and 21 TD) using the Gentra Puregene tissue kit (Qiagen).

Whole Genome Bisulfite Sequencing (WGBS)

Raw sequencing data (fastq files) were published previously (171). Briefly, WGBS libraries were made with the sonicated genomic DNA (around 300 bp) and ligated with methylated Illumina adapters using NEB's NEBNext DNA library prep kit (161,171). The library was bisulfite converted using EZ DNA Methylation lighting kit (Zymo), amplified for 12 cycles using PfuTurbo Cx Hotstart DNA Polymerase (Agilent) and purified with Agencourt AMPure XP Beads (Beckman

Coulter). The quality and quantity of libraries were measured on a Bioanalyzer (Agilent) and sequenced on Illumina HiSeq 2000 with each sample per single lane. Reads after trimming were uniquely mapped to human reference genome (hg38) as described previously using BS-Seeker2 on average 1.6X genome converge with 99.3% bisulfite conversion efficiency (measured through the percentage of non-CpG cytosines that were unconverted) (132,161,250).

ASD Differentially Methylated Regions (DMRs) and genome-wide significant DMRs

DMRs were called as described in previous publications (132,251) using the default settings. In this case, each ASD DMR contained greater than 10% methylation difference between ASD and TD samples at least three CpGs within 300 base pairs (bp) and a *p*-value < 0.05. Background regions were defined using the same conditions as DMRs but without any percent methylation filters to identify all possible DMR locations based on CpG density and sample sequencing coverage. Hypermethylated ASD DMRs were defined as higher percent methylation in ASD versus TD, while hypomethylated ASD DMRs was were defined as lower percent methylation in ASD versus TD samples. Genome-wide, significant DMRs were identified based on a family-wide error rate (FWER) < 0.05, determined by permuting the samples 1000 times by chromosome, and counting the number of null permutations with equal or better DMRs ranked by number of CpGs and areaStat (252).

Hierarchical clustering and principal component analysis (PCA)

Methylation was extracted at each ASD DMR for every sample. Percent methylation of each sample was normalized to the mean methylation of each ASD DMR. ASD DMRs were grouped

by Ward's Method of hierarchical clustering (253). Principal component analysis was performed on methylation at all ASD DMRs across all samples using the prcomp function and ggbiplot package in R. The ellipses for each group were illustrated as the 95% confidence interval. The lack of overlapping ellipses for ASD and TD samples indicated significant methylation difference in ASD DMRs between groups ($p < 0.05$).

Placental Cell Type Specific Genes Identification

Cell type specific genes were extracted from Vermo-Tormo *et al*. (189) on 38 different types of cells with the top 30 cell-type specific signature genes. Genes for each cell types were filtered as unique genes that were only expressed in specific cell types. A test for significance of overlapping genes was done using Fisher's exact test.

Assignment of DMRs to genes and relative location to TSS

Genes were assigned to DMRs using the Genomics Regions Enrichment of Annotations Tool (GREAT) on the default association settings (5.0 kilo-base (kb) upstream and 1.0 kb downstream, up to 1000.0 kb max extension) (254). The distance (kb) was calculated from the ASD DMRs, hypermethylated ASD DMRs, hypomethylated ASD DMRs and background regions to transcription start site (TSS) of the GREAT assigned gene. The gene length was calculated for both placental ASD DMR genes and all genes in human genome and tested for potential distribution differences by Pearson's chi-squared test.

Gene Ontology Term and Pathway Enrichment Analysis

Gene ontology (GO) analysis was done using PANTHER (Protein Analysis Through Evolutionary Relationships) overrepresentation test, with the GO Ontology database (255,256) and Fisher's exact test with false discovery rate (FDR) multiple test correction. GO term enrichments were presented by the hierarchical terms rather than specific subclass functional classes, as described previously (257,258).

Tests for ASD DMR Enrichments

All tests of enrichment for ASD DMRs were compared to a set of all possible background regions that are calculated in the DMR analysis pipeline. Enrichment tests for placenta ASD DMRs associated genes and published gene lists were done using the GeneOverlap R package which implements Fisher's exact test and adjusted for FDR correction (259). *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$ by Fisher's exact test with FDR corrected. Brain cortex (BA9) ASD DMRs were defined as either a 5% or 10% methylation difference between ASD and TD and were described previously using the same method as placenta ASD DMRs (190). The SFARI (Simons Foundation Autism Research Initiative) database was used for the five categories of ASD risk genes (191). High effect ASD risk gene lists were also identified from Sanders *et al.* (176). Likely gene-disrupting (LGD) recurrent ASD mutations and missense mutation on *de novo* mutations were obtained from Iossifov *et al*. (65). Gene lists on intellectual disability (ID) were obtained from Gilissen *et al*. (192). Alzheimer's disease GWAS gene lists were extracted from SNPs showing association with Alzheimer's disease ($P \leq 1x10^{-3}$) (260). Lung cancer GWAS gene lists were acquired from Landi *et al*. (261). The random genes category contains the same number of

regions as the placenta ASD DMRs to serve as a specificity control. ASD DMRs were examined for enrichment with known chromatin marks compared to the background using LOLA R package with two-tailed Fisher's exact test after FDR correction (202). Placenta histone marks H3K4me1, H3K4me3, H4K9me3, H3K36me3, H3K27me3 and H3K27ac were extracted from ENCODE (Encyclopedia of DNA Elements) placenta ChIP-seq dataset (262,263). ASD DMRs were also analyzed for overlap with chromatin states predicted by chromHMM, which use histone modification ChIP-seq data to separate the genome into 15 functional states in the Roadmap Epigenomics Project using a Hidden Markov Model (196,264). For promoters, chromHMM separates active transcription start site (TssA), TSS flank (TssAFlnk), bivalent TSS (TssBiv), and bivalent TSS flank (BivFlnk) states. For enhancers, genic enhancer (EnhG), enhancer (Enh), and bivalent enhancer (EnhBiv) are the different states. Human CpG island locations were extracted from UCSC genome browser (265). CpG island shores were defined as 2 kb flanking regions on both sides of CpG island. CpG island shelf was characterized as 2 kb flanking regions on both sides of CpG island shore, not including CpG island or CpG island shore. CpG island "open sea" includes all genomic regions except CpG island, CpG island shore and CpG island shelf. A custom R script was used to generate the locations of CpG islands (https://github.com/Yihui-Zhu/AutismPlacentaEpigenome).

Pyrosequencing

Genomic DNA (500 ng) was bisulfite converted using the EZ DNA Methylation kit (Zymo). Amplification and sequencing primers were designed using the PyroMark Assay Design Software 2.0 (Qiagen). DMRs were amplified using the PyroMark PCR kit (Qiagen).

Pyrosequencing of 13 CpG sites at *CYP2E1* gene, and 11 CpG sites in human *IRS2* gene was performed in triplicate. Pyrosequencing was performed on a Pyromark Q24 Pyrosequencer (Qiagen) with the manufacturers recommended protocol. Enzyme, substrate, and dNTPs were from the Pyromark Gold Q24 Reagents (Qiagen) and the methylation levels were analyzed using Pyromark Q24 software.

CYP2E1 related DMR pyrosequencing primers:

Forward: GGTGTTTTGTTTTGGGGTTGA

Reverse: ACCCATTCAATATTCACAACAATC (5' Biotin)

Sequencing: GGTTGATGATGGGGA

Amplification region: chr10: 133527817 – 133527938 (hg38)

IRS2 related DMR pyrosequencing primers:

Forward: TTAGGAATATAGGGAAAGGTGAAAGT

Reverse: CCACCCATTCACCCATTCTA (5' Biotin)

Sequencing: GGGAAAGGTGAAAGTT

Amplification region: chr13: 109781623 – 109781794 (hg38)

Gene Expression in Umbilical Cord Blood

Data for gene expression assessed by Affymetrix Human Gene 2.0 array were extracted a previous publication on umbilical cord blood from subjects in the MARBLES study (GEO ID: GSE123302) (203). Placenta and cord blood were collected at the same time period in the same

study. Raw intensity values from cord blood samples were normalized by RMA and data from 70 male samples were extracted, including 30 ASD and 40 TD samples. Normalized expression was examined at the only probe annotated to *CYP2E1* (16711001) and the only probe annotated to *IRS2* (16780917). Analysis was done on those two probes with 70 samples on the normalized matrix data.

Western Blot

In Western blot experiments, placental proteins were isolated with RIPA buffer containing 10mM Tris-Cl (pH 8.0), 1mM EDTA, 1% Triton X-100, 0.1% sodium deozydholate, 0.1% SDS, 140mM NaCl, 1mM PMSF and complete protease inhibitors (ThermoFisher), incubated at 37°C for 30 minutes, sonicated and heated at 95°C for 5 min. BCA (Bicinchoninic Acid) protein assay (ThermoFisher) was used to determine protein concentration. Protein samples (20-30 ug) were resolved on 4-20% tris-gylcine polyacrylamide gels (Biorad). Proteins were separated and transferred to nitrocellulose membranes for 60 minutes at a constant voltage of 100. The membranes were blocked in Odyssey Blocking Buffer (PBS) (Licor, 927-40000) for 40 min. Anti-IRS2 (1:5,000, Cell Signaling, 3089S) and anti-GAPDH (1:10,000, Advanced Immunochemical, Inc., 2-RGM2) were incubated with the membrane with Odyssey Blocking Buffer containing 0.2% Tween overnight at 4°C. Membranes were washed with 1 X PBS (Phosphate-buffered saline) containing 0.2% Tween and then incubated with secondary antibodies, IRDye 800CW Donkey anti-Mouse IgG (1:50,000, Licor, 926-32212) and IRDye 680RD Donkey anti-Rabbit IgG (1:50,000, Licor, 926-68073) for 1 hour. Membranes were scanned using a Licor Odyssey infrared imaging system based on the manufacturer's guidance (with resolution: 84; quality:

medium, 600-channel: 6; 800-channel: 5). Relative protein quantification was done using the ImageJ software program (266) in densitometry mode. IRS2 signals were normalized to GAPDH (Glyceraldehyde 3-phosphate dehydrogenase) for each sample.

Sanger Sequencing

PCR amplification was performed on each sample using PCR 10x buffer, 25 mM $MgCl_2$, 5 M betaine, 10 mM dNTPs, DMSO, and HotStart Taq (Qiagen). Each PCR program was unique to the region being amplified with specific primers. The PCR product was then resolved by gel electrophoresis using a 1% Agarose gel in 1 X TE to later be extracted using the gel extraction kit (Qiagen) based on the default protocol. After DNA quantitation by NanoDrop, the samples were sent to the UC Davis Sequencing Facility for sequencing on the 3730 Genetic Analyzer (Applied Biosystems Prism) with DNA sequencing Analysis software v.5.2 (Applied Biosystems Prism). The sequencing results were assembled and analyzed using CodonCode Aligner version 7.0 (CodonCode).

CYP2E1 related SNP (rs943975, rs1536828) primers:

Forward: CTACAAGGCGGTGAAGGAAG

Reverse: CCCATCCCCATAAACTCTCC

IRS2 related SNP (rs943975) primers:

Forward: TTAGGAATATAGGGAAAGGTGAAAGT

Reverse: CCACCCATTCACCCATTCTA

Code availability

Custom scripts for WGBS analysis are available at https://github.com/kwdunaway/WGBS_Tools with the instructions. Custom Scripts for DMR finder are available at https://github.com/cemordaunt/DMRfinder with the instructions. The rest of code and scripts for each figure and tables are available at https://github.com/Yihui-Zhu/AutismPlacentaEpigenome.

Data availability

WGBS data were previously published, Gene Expression Omnibus (GEO) accession number GSE67615 (171). The rest of the relevant data and information are included in supplementary tables.

Chapter 3 – Expression changes in epigenetic gene pathways associated with one-carbon nutritional metabolites in maternal blood from pregnancies resulting in autism and non-typical neurodevelopment

Yihui Zhu, Charles E. Mordaunt, Blythe P Durbin-Johnson, Marie A Caudill, Olga V. Malysheva, Joshua W. Miller, Ralph Green, S. Jill James, Stepan B. Melnyk, M. Daniele Fallin, Irva Hertz-Picciotto, Rebecca J. Schmidt, Janine M. LaSalle

Abstract

The prenatal period is a critical window for the development of autism spectrum disorder (ASD). The relationship between prenatal nutrients and gestational gene expression in mothers of children later diagnosed with ASD or non-typical development (Non-TD) is poorly understood. Maternal blood collected prospectively during pregnancy provides insights into the effects of nutrition, particularly one-carbon metabolites, on gene pathways and neurodevelopment. Genome-wide transcriptomes were measured with microarrays in 300 maternal blood samples in Markers of Autism Risk in Babies -Learning Early Signs (MARBLES). Sixteen different one-carbon metabolites, including folic acid, betaine, 5'-methyltretrahydrofolate (5-MeTHF), and dimethylglycine (DMG) were measured. Differential expression analysis and weighted gene correlation network analysis (WGCNA) were used to compare gene expression between children later diagnosed as typical development (TD), Non-TD and ASD, and to one-carbon metabolites. Using differential gene expression analysis, six transcripts (*TGR-AS1*, *SQSTM1*, *HLA-C* and *RFESD*) were associated with child outcomes (ASD,

Non-TD and TD) with genome-wide significance (FDR q <0.05). Genes nominally differentially expressed between ASD and TD significantly overlapped with seven high confidence ASD genes. WGCNA identified co-expressed gene modules significantly correlated with 5-MeTHF, folic acid, DMG, and betaine. A module enriched in DNA methylation functions showed a suggestive protective association with folic acid/5-MeTHF concentrations and ASD risk. Independent of child outcome, maternal plasma betaine and DMG concentrations were associated with a block of co-expressed genes enriched for adaptive immune, histone modification, and RNA processing functions. These results support the premise that the prenatal maternal blood transcriptome is a sensitive indicator of gestational nutrition and children's later neurodevelopmental outcomes.

Lay Summary

Pregnancy is a time when maternal nutrition could interact with genetic risk for autism spectrum disorder. Blood samples collected during pregnancy from mothers who had a prior child with autism were examined for gene expression and nutrient metabolites, then compared to the diagnosis of the child at age three. Expression differences in gene pathways related to the immune system and gene regulation were observed for pregnancies of children with autism and non-typical neurodevelopment and were associated with maternal nutrients.

Introduction

Autism spectrum disorder (ASD) is a group of neurodevelopmental disorders characterized by persistent impairment in social interactions, communication, restricted interests or repetitive

behaviors, and language deficits (59). Current data show that one in every 54 children in the United States has ASD (59). One major component of ASD risk is genetic heritability, based on studies of twins, siblings, and other family members (63,81,173). Common genetic variants each having small effects dominate most ASD risk compared with rare gene variants with large effects (267). Large genome-wide association studies (GWAS) support the role of common genetic variants in ASD with remaining challenges in ASD complexity and heterogeneity (64,65,176). Mutations in single genes can explain less than 1% of ASD cases (172,268).

Accumulating lines of evidence suggest that ASD risk arises from both genetic and environmental risk factors. *In utero* maternal exposures can contribute as ASD risk factors, including air pollution, fever, asthma, and nutrition, especially nutrients involved in the one-carbon metabolic pathway (70,120,178,269). Other studies suggest that one-carbon metabolism is implicated in gene-environment interactions in ASD (270,271). Maternal prenatal nutritional supplements containing folic acid and additional B vitamins that play a role in one-carbon metabolism are associated with ASD risk reduction (119,120,122). A common genetic polymorphism affecting folic acid metabolism, *MTHFR* C677T, interacts with maternal nutrition, as association between folic acid and reduced ASD risk was strongest for mothers and children with *MTHFR* 677 C>T variant genotypes (70,272). These findings suggest that additional gene-environment interactions relevant to ASD may be identified from investigations into maternal factors, since maternal and fetal metabolisms are shaped by both shared genetics and nutritional environment during pregnancy that can coordinately impact neurodevelopment.

Gene expression levels are also influenced by both genetic and environmental factors, especially by *in utero* maternal nutrition (273,274). Maternal peripheral blood therefore offers a unique window to study transcriptome alterations during pregnancy that may reflect altered fetal development associated with nutrition (275,276). Numerous environmental factors during pregnancy can alter gene expression levels (178,277). Other neurodevelopment disorders, such as schizophrenia, have also demonstrated a significant interaction of genetic risk with maternal perinatal environmental factors that affected the transcriptome (169,278). Postmortem brain gene expression studies revealed gene co-expression modules enriched for immune response and neuronal development functions in ASD (194,219). Other studies using child peripheral blood and cord blood showed that differential gene expression in ASD was enriched for immune and inflammatory processes (279–281).

While numerous studies have investigated specific genes or pathways in children with ASD, none have focused on the maternal transcriptome during pregnancy. Further, most previous ASD transcriptome studies used data from specimens collected postmortem or after childbirth, as opposed to prospective studies to help understand potential etiologic changes that occur before behavioral symptoms. Other large epidemiology studies examined environmental effects in ASD, but how the environment influences alterations at the molecular level remains to be understood. The goal of this study was to examine maternal prenatal gene expression profiles associated with both maternal serum one-carbon metabolites and the child outcome (ASD, Non-TD, TD) to shed light on molecular changes during pregnancy.

Results

Study sample characteristics and nutrient measurements

High quality RNA was isolated from 300 maternal peripheral blood samples collected during pregnancy from the MARBLES high risk ASD cohort (**Supplementary Table 3.1**). Children from MARBLES pregnancies were diagnosed at 3 years old as ASD (67, including 47 male and 20 female), Non-TD (79, including 46 male and 33 female), and TD (154, including 79 male and 75 female) (**Supplementary Table 3.2**).

Nutrients in the one-carbon metabolism pathway, including methionine, SAM, SAH, adenosine, homocysteine, 5-MeTHF, folic acid, vitamin (Vit) B6, Vit B12, choline, DMG, betaine, cystine, cysteine, GSH, and GSSG were directly measured from maternal blood in 14% - 62% of all samples (**Supplementary Table 3.3**). None of these metabolites in maternal blood were significantly associated with clinical outcomes of children (**Supplementary Table 3.3**). Measurements for one-carbon metabolites and transcriptomes were conducted on samples collected throughout pregnancy (**Supplementary Fig. 3.1**).

Differential gene expression analyses by child outcome

Expression was measured using the Human Gene 2.0 Affymetrix microarray and adjusted for all surrogate variables, followed by differential gene expression analysis for child diagnosis (ASD, Non-TD, TD) on 36,459 transcripts. There were 28 surrogate variables (SVs) identified, including 5 SVs significantly associated with batch effect and 2 SVs significantly associated with gestational age of maternal blood draw (**Supplementary Fig. 3.2**). Six transcripts located at four

genes (*TGR-AS1*, *SQSTM1*, *HLA-C*, and *RFESD*) were associated with child outcomes (ASD, Non-TD, TD) with genome-wide significance by F-test (FDR adjusted *p*-value <0.05) (**Supplementary Table 3.4**). Three out of these six transcripts mapped to *HLA-C* (Major Histocompatibility Complex, Class I, C) (FDR adjusted *p*-value <0.05).

Comparing the maternal blood transcriptome between ASD and TD outcomes revealed 2,012 differentially expressed transcripts at a nominal confidence level (unadjusted *p*-value <0.05) that mapped to 1,912 genes, including 980 up-regulated and 1,032 down-regulated transcripts, with none significant after FDR adjustment (**Fig. 3.1A**, **Supplementary Table 3.5**). There was a significant overlap between these 1,912 differentially expressed genes and a list of strong ASD candidate genes from the Simons Foundation Autism Research Initiative (SFARI Gene, including *TRIO*, *GRIA1*, *SMARCC2*, *SPAST*, *DIP2C*, *FOXP1*, and *CNTN4*, Fisher's exact test, *p*-value <0.05) (191).

Comparing the maternal blood transcriptome between Non-TD and TD outcomes revealed 1,987 differentially expressed transcripts at a nominal confidence level (unadjusted *p*-value <0.05) that mapped to 1,919 genes, including 1,044 up-regulated and 943 down-regulated transcripts (**Fig. 3.1B**, **Supplementary Table 3.6**). Two of these transcripts, *RFESD* and *TRG-AS1*, also passed genome-wide significance (FDR adjusted *p*-value <0.05). Unlike the ASD vs TD comparison, however, no significant overlap was observed between Non-TD vs TD differentially expressed genes and SFARI gene lists. An alternative approach of adjustment for known confounders revealed a significant overlap with SVA adjusted differentially expressed

transcripts for both ASD vs TD and Non-TD vs TD comparisons (Fisher's exact test, *p*-value <2.2E-16) (**Supplementary Fig. 3.3**, **Supplementary Table 3.4**).

Differential gene expression analysis using SVA resulted in a significant overlap of 218 transcripts between ASD vs TD differentially expressed transcripts and Non-TD vs TD differentially expressed transcripts (Fisher's exact test, *p*-value <2.2E-16) (**Fig. 3.1C**). Gene ontology (GO) analysis of these 218 transcripts revealed significant enrichment for the interferon-gamma mediated signaling pathway, apoptosis in muscle, response to interferon gamma, and metal ion transport (**Fig. 3.1D**, **Supplementary Fig. 3.4**). CaMK (calmodulin-dependent protein kinase) families (*CAMK2A*, *CAMK2B*, *CAMK2D* and *CAMK2G*) and HLA (human leukocyte antigen) systems (*HLA-B*, *HLA-C* and *HLA-E*) were included in those significant signaling pathways (**Fig. 3.1D**). In contrast, neither list of ASD- or Non-TD-specific differentially expressed transcripts were significantly enriched for any GO terms. To control for potential bias towards expressed genes in the GO terms, we also used an alternative approach of filtering the probes based on intensity that removed the lowest 5% of expressed transcripts (**Supplementary Fig. 3.5**), which resulted in an identical list of significant GO terms (**Supplementary Table 3.7**).

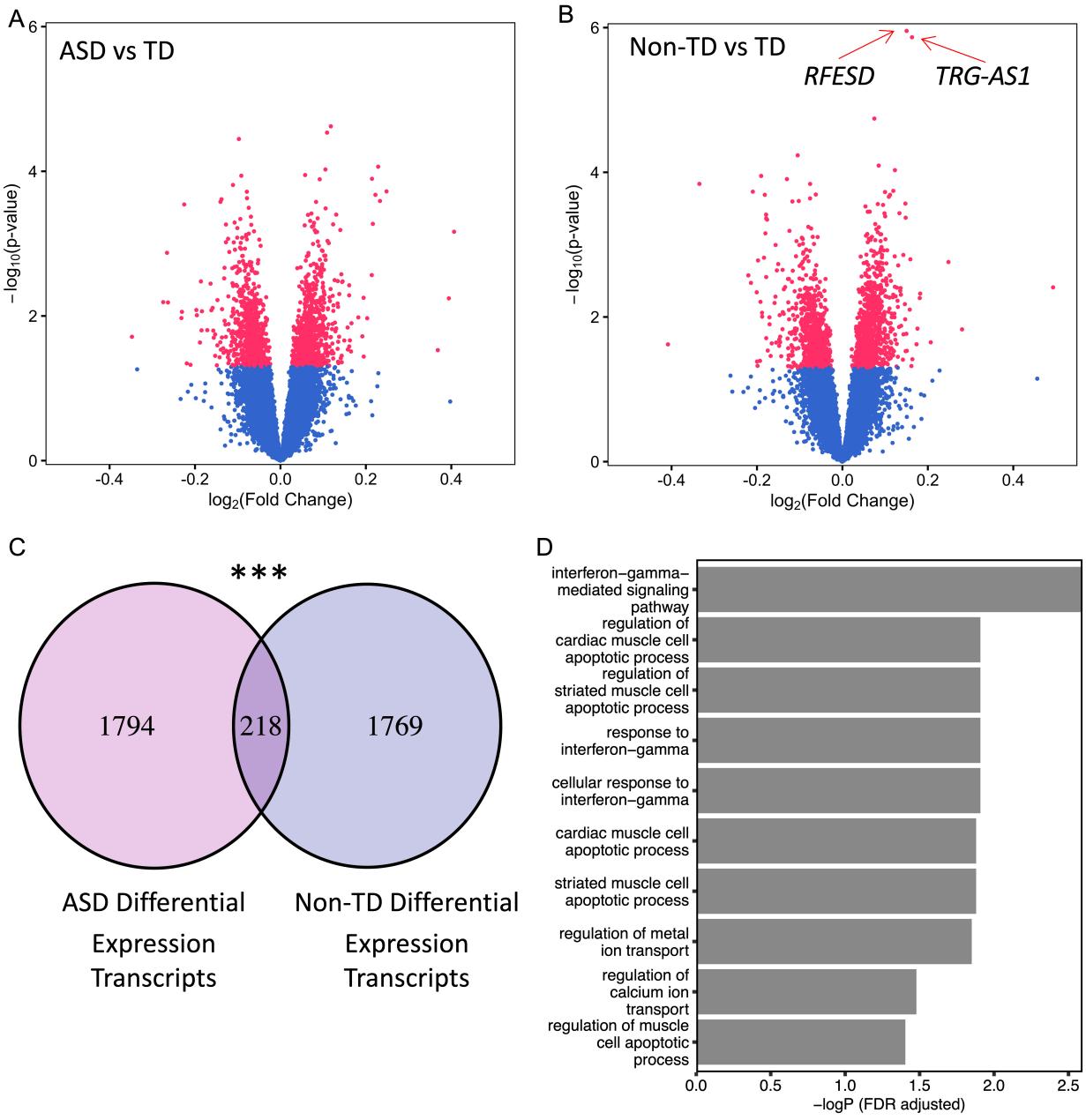


Figure 3.1. Identification and function of ASD associated and Non-TD associated differentially expressed genes in maternal peripheral blood.

Differential expression analysis was performed in maternal peripheral blood transcriptomes (n = 300) after adjustment for surrogate variables.

A. Identification of 1,912 differentially expressed genes (2,012 transcripts, *p*-value <0.05) compared between children diagnosed as ASD (n = 67) and TD (n = 154).

B. Identification of 1,919 differential expressed genes (1,987 transcripts, *p*-value <0.05) compared between children diagnosed as Non-TD (n = 79) and TD (n =154). Two transcripts located at *RFESD* and *TRG-AS1* were genome-wide significant in the Non-TD to TD comparison (Supplementary Table 2).

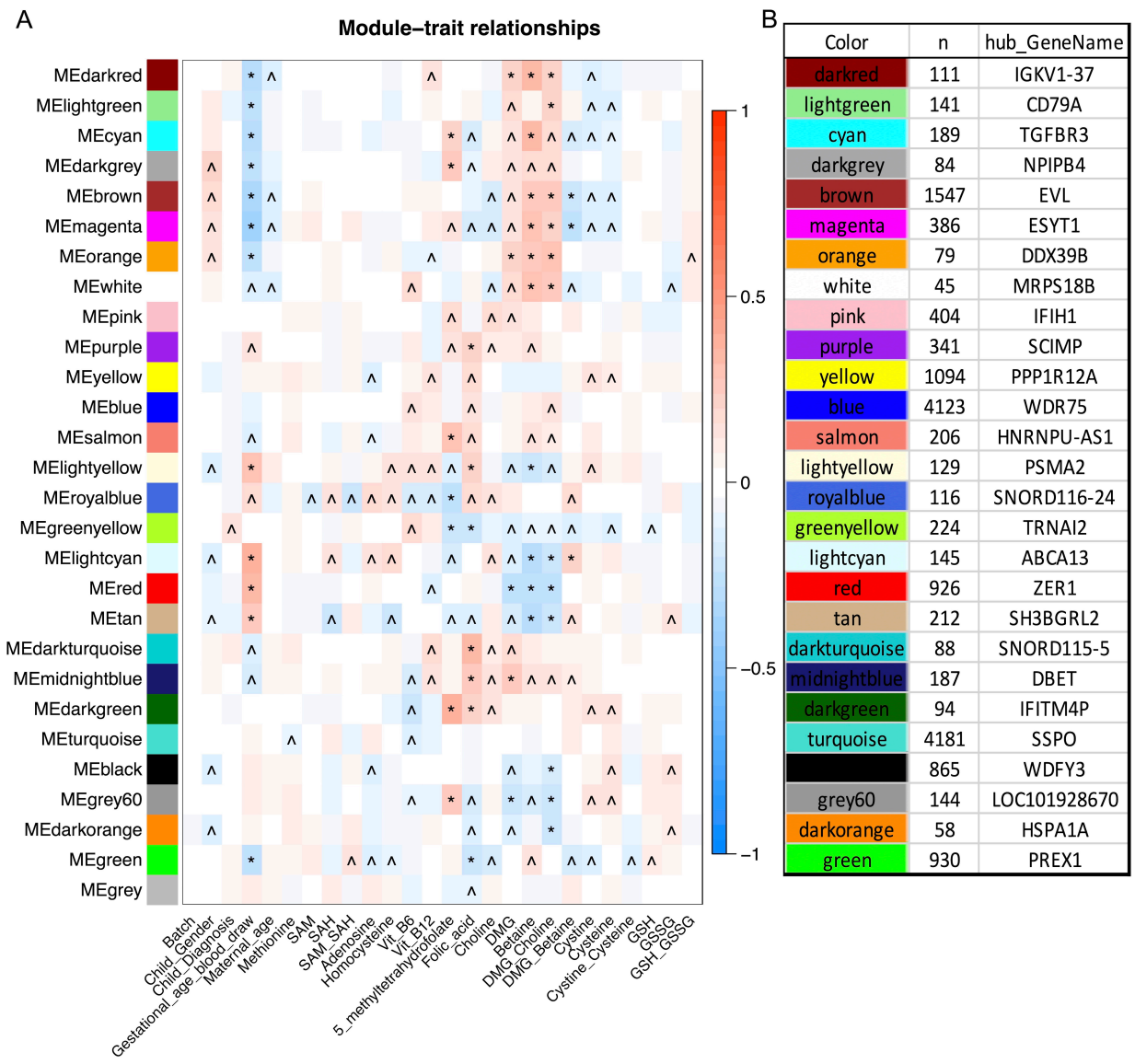
C. Venn diagram represents the overlap in differentially expressed transcripts (unadjusted p<0.05) identified in ASD to TD versus Non-TD to TD comparisons, which was greater than expected by random using a Fisher's exact test (*p*-value <0.001***).

D) Gene ontology (GO) and pathway analysis was performed on the 218 transcripts differentially expressed in both ASD-TD and non-TD-TD comparisons, with significant enrichments (Fisher's exact test, FDR *p*-value <0.05). In contrast, the differentially expressed transcripts uniquely associating with either ASD or non-TD were not significantly enriched for any GO terms.

Weighted gene co-expression network analysis (WGCNA) identified gene modules correlating with specific maternal nutrient levels

WGCNA was performed as a complementary bioinformatic approach that incorporates the independent and inter-related associations of transcript levels with measured concentrations of maternal nutrients. First, expression values were adjusted for batch effects, then correlation patterns among all transcripts were analyzed across all 300 samples. WGCNA identified 27 co-expressed gene modules in our dataset, representing 17,049 transcripts, distinguished from

19,410 transcripts without evidence of co-expression were grouped into the “grey” module (**Fig. 3.2A, Fig. 3.2B**, **Supplementary Fig. 3.6**, **Supplementary Table 3.8**). For each module, the number of transcripts, as well as the hub gene, defined as the gene with the highest correlation with the module eigengene, were determined (**Fig. 3.2B**, **Supplementary Table 3.9**). Out of those 27 co-expression modules, 23 modules showed associations between eigengene expression level and at least one variable related to demographics, diagnosis, or maternal nutrients, after FDR correction (FDR adjusted *p*-value <0.05) (**Fig. 3.2A**, **Supplementary Fig. 3.6**). All 27 modules were significantly associated with one or more traits, including child clinical outcome, demographic factors, and maternal blood metabolite concentrations at unadjusted *p*-value <0.05 (**Fig. 3.2A**, **Supplementary Fig. 3.7**).



Color	n	hub_GeneName
darkred	111	IGKV1-37
lightgreen	141	CD79A
cyan	189	TGFBR3
darkgrey	84	NPIPB4
brown	1547	EVL
magenta	386	ESYT1
orange	79	DDX39B
white	45	MRPS18B
pink	404	IFIH1
purple	341	SCIMP
yellow	1094	PPP1R12A
blue	4123	WDR75
salmon	206	HNRNPU-AS1
lightyellow	129	PSMA2
royalblue	116	SNORD116-24
greenyellow	224	TRNAI2
lightcyan	145	ABCA13
red	926	ZER1
tan	212	SH3BGRL2
darkturquoise	88	SNORD115-5
midnightblue	187	DBET
darkgreen	94	IFITM4P
turquoise	4181	SSPO
	865	WDFY3
grey60	144	LOC101928670
darkorange	58	HSPA1A
green	930	PREX1

Figure 3.2. Co-expression network modules with demographic factors and maternal peripheral blood one-carbon metabolites.

A. Heatmap of Z-scores of modules eigengenes with sample covariates with 27 co-expression network modules on all 300 maternal blood samples. Each row represents a different module eigengene and each column is the associated trait, which include child clinical outcome, demographic factors, and maternal blood metabolite concentrations. The matrix was calculated by Pearson correlation and *p*-values adjusted for the total number of comparisons. Color

represents the direction (red, positive correlation; blue, negative correlation) and intensity reflects the significance. (^ unadjusted *p*-value <0.05 and FDR adjusted *p*-value > 0.05; * FDR adjusted *p*-value <0.05).

B. Number of transcripts and hub genes from all 27 co-expressed modules are listed.

Multiple co-expression modules were significantly correlated (FDR adjusted *p*-value <0.05) with gestational age at blood draw and four maternal metabolites, including 5-MeTHF, folic acid, DMG, and betaine (**Fig. 3.2A**). None of the additional measured variables was significantly associated with any co-expression gene modules, including clinical outcome. However, the module "greenyellow" showed a nominally significant positive correlation with outcome (unadjusted *p*-value = 0.02, FDR adjusted *p*-value = 0.14) and a negative significant correlation with both 5-MeTHF (FDR adjusted *p*-value = 0.02) and folic acid levels (FDR adjusted *p*-value = 0.02) (**Fig. 3.2A, Fig. 3.3A, Fig. 3.3B, Supplementary Fig. 3.6, Supplementary Fig. 3.7**, **Supplementary Table 3.10**). Interestingly, the greenyellow module eigengene was correlated in opposite directions with ASD and 5-MeTHF, consistent with a putative 5-MeTHF protective effect in ASD (**Fig. 3.3A**, **Fig. 3.3B**).

This "greenyellow" module contained 224 transcripts with *TRNAI2* as the hub gene (**Supplementary Table 3.10**). These 224 transcripts showed a significant enrichment for gene ontology functions in methylation-CpG binding, methyl-dependent chromatin silencing, and keratinocyte differentiation (Fisher's exact test, FDR adjusted *p*-value <0.05) (**Fig. 3.3C**, **Supplementary Table 3.11**). The three known genes with methyl-binding functions included

MBD3L3, *MBD3L4* and *MBD3L5*, represented by 16857547, 16867905, and 16867910 transcripts (**Fig. 3.3C**). Normalized expression of those three transcripts was also significantly associated with the "greenyellow" module eigengene, supporting their membership in the module (**Fig. 3.3D**).

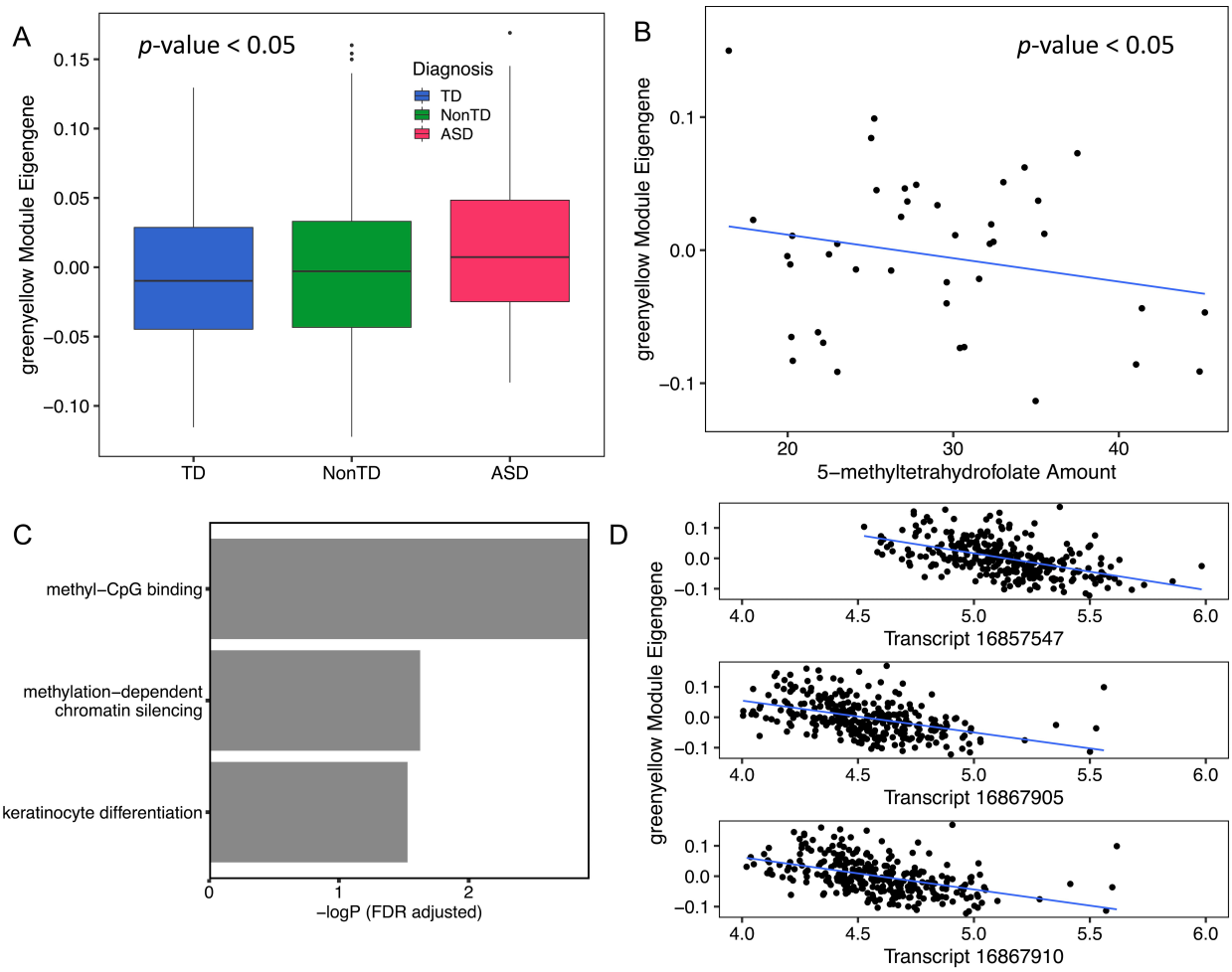


Figure 3.3. "Greenyellow" module was positively associated with diagnosis and negatively associated with folic acid and 5-MeTHF.

A. "Greenyellow" module eigengene was significantly associated with child diagnosis (one-way ANOVA, unadjusted *p*-value <0.05). Greenyellow eigengene values were higher in maternal blood from ASD pregnancies than TD or non-TD pregnancies.

B. "Greenyellow" module eigengene level was significantly negatively associated with 5-MeTHF concentrations in maternal blood (ANOVA, *p*-value <0.05).

C. Bar graph shows gene ontology (GO) and pathway significant enrichments from the 224 transcripts in "greenyellow" module (Supplementary Table 8).

D. Transcripts (16857547, 16867905 and 16867910) from MBD3L3-5 genes encoding proteins involved in methylation-CpG binding functions were significantly negatively associated with "greenyellow" module eigengene.

There were 12 modules transcriptome significantly associated with gestational age, including 8 positively correlated modules (darkred, lightgreen, cyan, darkgrey, brown, magenta, orange and green) and 4 negatively correlated modules (lightyellow, lightcyan, red, tan) (**Fig. 3.2A**, **Supplementary Table 3.12**). Genes within modules with a positive correlation with gestational age were significant enriched for functions in RNA binding, chromatin biding, and ATP binding, among others (**Supplementary Table 3.12**). In contract, genes within modules negatively correlated with gestational age were significantly enriched in functions related to granulocyte activation, neutrophil mediated immunity, coagulation, and other blood functions (**Supplementary Table 3.12**).

Eight co-expression modules strongly clustered with betaine and DMG

Among the 27 identified co-expression modules, eight modules (darkred, lightgreen, cyan, darkgrey, brown, magenta, orange and white) were highly correlated with each other and clustered based on unsupervised hierarchical clustering, representing a total of 2,582 transcripts (**Fig. 3.4, Supplementary Fig. 3.8, Supplementary Table 3.13**). Betaine and DMG were significantly associated and clustered together with this distinct block of co-expression modules.

Transcripts inside these eight clustered co-expression modules associated with betaine and DMG showed significant enrichment for 18 gene pathways involved in adaptive immune response, RNA processing, histone modification, inflammatory response, and Rett syndrome (Fisher's exact test, FDR adjusted *p*-value <0.05) (**Supplementary Fig. 3.9**). Network analysis using GeneMANIA (282) identified a network with *EVL* in the center, linked with other hub genes (**Supplementary Fig. 3.10**).

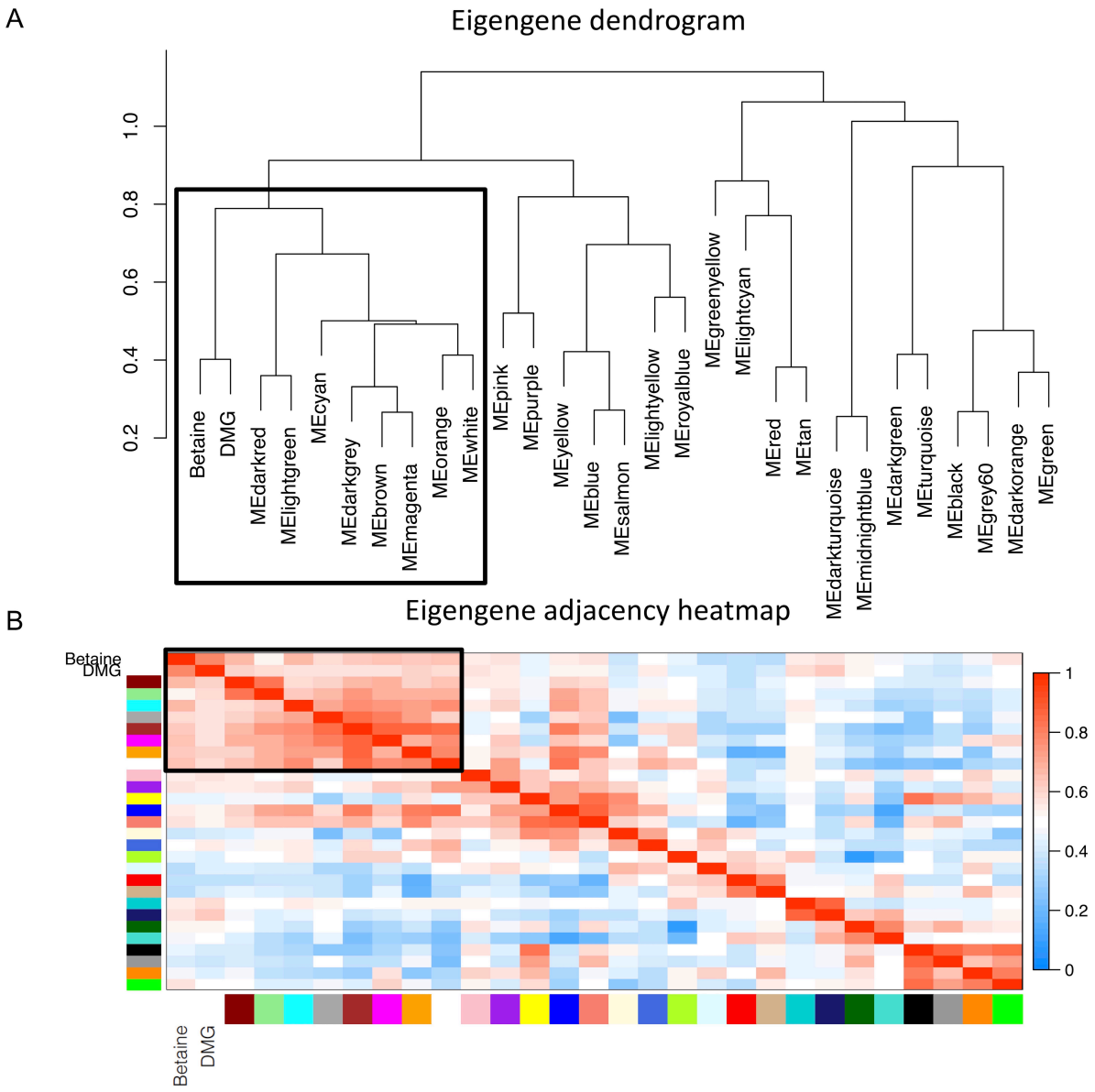


Figure 3.4. Eight weighted gene co-expression modules associated with maternal betaine and DMG concentrations were strongly clustered.

A. Unsupervised hierarchical clustering dendrogram was performed with module eigengenes, betaine and DMG. The height of each node represents the intergroup dissimilarity. Similar nodes clustered together under one branch.

B. Unsupervised hierarchical clustering adjacency heatmap, with color and intensity representing the degree of correlation (dark, high; light, low correlation).
Black box indicates the block of eight weighted gene co-expression modules associated with betaine and DMG concentrations.

Cell type composition in maternal peripheral blood was associated with maternal metabolites but not child clinical outcomes

In order to determine the effects of cell composition differences on the findings associated with maternal transcriptomes, cell type specific information from 22 immune cell types was deconvoluted using the CIBERSORT web tool. Maternal peripheral blood samples reflected a mixture of cell types, with neutrophils as the largest and most variable population ranging from 17% to 48% (**Fig. 3.5A, Supplementary Table 3.14**). The eigengenes for 21 out of 27 modules were significantly correlated with at least one cell type (FDR adjusted *p*-value <0.05) (**Supplementary Fig. 3.11**). No significant difference was observed in cell type composition between child diagnosis outcomes or gender (**Fig. 3.5A. 3.5B**, **Supplementary Fig. 3.12**). Furthermore, neither the "greenyellow" module, nor the betaine and DMG variables were significantly associated with cell type proportions, suggesting that the associations identified with these modules were largely cell type independent (**Fig. 3.5B**, **Supplementary Fig. 3.11**, **Supplementary Fig. 3.12**).

In contrast, some cell type proportions were significantly correlated with some maternal metabolites. Vit B6, 5-MeTHF, choline, cysteine, the ratio of DMG/betaine, and the ratio of

cystine/cysteine were separately associated with six cell types (FDR adjusted *p*-value <0.05) (**Fig. 3.5B, Supplementary Fig. 3.12**). Vit B12, folic acid, the ratio of DMG/betaine, and the ratio of SAM/SAH were associated with more than one cell type (FDR adjusted *p*-value <0.05) (**Fig. 3.5B, Supplementary Fig. 3.12**). The most significant association was between vit B12 and memory B cells (FDR adjusted *p*-value = 0.0001) (**Fig. 3.5B, Supplementary Fig. 3.12**).

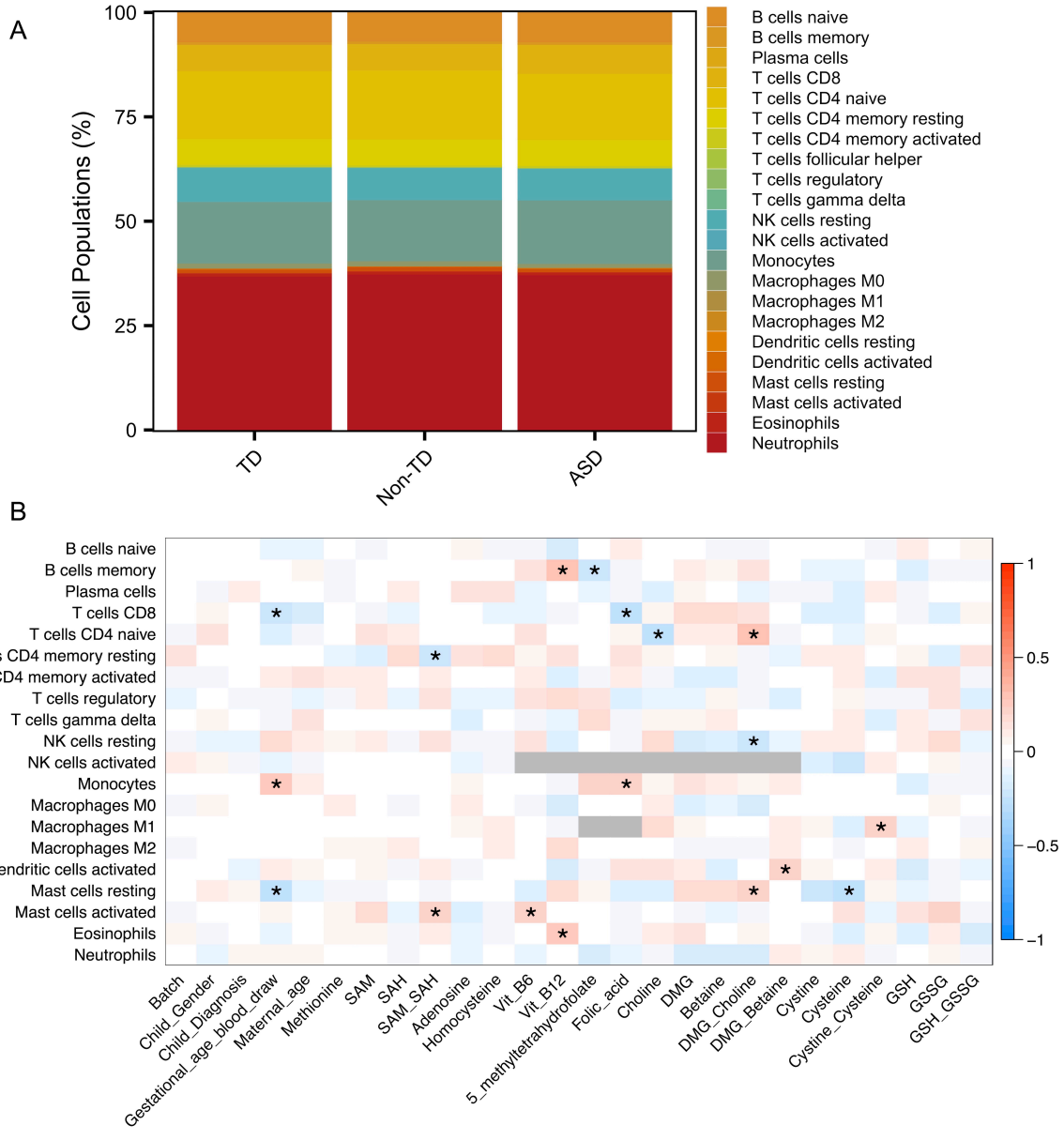


Figure 3.5. Imputed cell type proportions in material peripheral blood associated with demographic factors and maternal nutrients.

A. Barplot of each cell type mean estimated proportion separated by children diagnosis outcomes using peripheral blood reference panel in CIBERSORT.

B. Heatmap of correlation between sample demographic factors and maternal nutrients with cell type proportions. Each row represents a cell type proportion and columns represent traits, including child diagnostic outcome, demographic factors, and maternal blood nutrient concentrations. *p*-values adjusted for the total number of comparisons. Color represents the direction (red, positive correlation; blue, negative correlation) and intensity reflects the significance, * *p*-value <0.05 after FDR correction.

Discussion

Maternal blood collected during pregnancy can provide molecular insights into the *in utero* environment relevant to the etiology of ASD. This was the first study to our knowledge to examine gene expression differences together with one-carbon metabolites in peripheral blood during pregnancy from mothers of children that went on to develop ASD, Non-TD, or TD at 36 months. Using complementary bioinformatics approaches, we identified several genes and gene pathways consistent with proinflammatory and oxidative stress responses in mothers of children with adverse neurodevelopmental outcomes. We also identified eight novel co-regulated gene modules associated with maternal blood betaine and DMG concentration.

Genes and gene patterns common to mothers of children with ASD and non-typical neurodevelopment

Using differential gene expression analysis of individual genes, we describe four genes (*SQSTM1, HLA-C, TRG-AS1,* and *RFESD)* that were differentially expressed in mothers by child diagnosis outcome of either TD, ASD, or Non-TD. *SQSTM1* encodes the p62 sequestosome that acts as a receptor for ubiquitinated cargo in the selective autophagy response induced by oxidative stress (283), and also links the mTOR and GABA signaling pathways in brain (284). *RFESD*, encoding an iron-sulfur cluster binding protein with oxidoreductase activity, is located on 5q15, a hotspot for copy number variants in intellectual and developmental disabilities (285,286). *TRG-AS1*, T-cell receptor gamma locus antisense RNA 1, is located on 7p14.1, another locus previously associated with developmental delay, intellectual disability, and ASD (285,287,288). *HLA-C* belongs to the HLA (human leukocyte antigen) polymorphic loci encoding major histocompatibility class I (MHC I) proteins involved in antigen presentation to CD8+ T cells and NK cells. HLA-C is important for both tolerance to fetal allo-antigens and viral immunity during pregnancy (289). Proinflammatory cytokines such as interferon gamma (IFNγ) induce HLA-C expression in both lymphocytes and placental trophoblasts. In the WGCNA modules, *HLA-C*, *SQSTM1* and *TGR-AS1* were all members of the same module, as evidence of co-expression. A number of previous studies have shown that the HLA locus is associated with ASD (290–292) or HLA locus activation in ASD children and their mothers (293–295), which is consistent with our findings at *HLA-C*. Furthermore, two additional class I loci, *HLA-B* and *HLA-E,* were also differentially expressed in mothers of children classified as ASD and Non-TD compared to TD children in this study, providing further evidence of an MHC I response in

pregnancies of non-typical neurodevelopment. These findings are consistent with the known maternal effect of viral infection during pregnancy of increasing susceptibility to ASD and schizophrenia during pregnancy (178,296). They also suggest how common genetic polymorphisms at these HLA loci may interact with common environmental factors such as viral infection during pregnancy to impact diverse neurodevelopmental outcomes.

Furthermore, gene pathway analysis of differentially expressed genes common between ASD and Non-TD revealed enrichment for the interferon-gamma mediated signaling pathway, which has been previously found to be elevated in mothers of children with ASD and other neurodevelopmental disorders (297,298). In one such study, elevated interferon-gamma levels in maternal midgestation peripheral blood was associated with a 50% increased risk of offspring ASD risk (297). A second enriched pathway included CaMK family members which play an important role in neuronal connectivity and synaptic plasticity (268,299,300) as well as immune response and inflammation (301). Prior ASD studies have implicated the CaMK pathway in dendritic growth and local connectivity alterations related to gene-environment interactions in ASD (268,299,300).

Although genome-wide significance of individual differentially expressed genes was not observed between samples from mothers whose children developed ASD compared to TD after adjusting for multiple comparisons, seven nominally-significant genes were also on the SFARI list of strong ASD candidate genes. *TRIO*, Trio Rho guanine nucleotide exchange factor, promotes exchange of GDP for GTP and provides necessary support for cell migration and cell

growth related to Alzheimer disease and other neurological conditions (82,302). *GRIA1*, encoding a receptor for glutamate, the predominant excitatory neurotransmitter in brain, is activated by normal neurophysiologic processes (82,303). *SMARCC2* encodes a chromatin remodeling protein with helicase and ATPase activities which has been implicated in altering chromatin structure in ASD (304). *CNTN4* functions in neuronal network formation and plasticity, and is associated with nervous system development at the transcriptome level (305,306). Mutations in *FOXP1*, a developmental transcription factor, are observed in rare cases of intellectual disability with ASD (307,308). Dysregulated expression of these ASD risk genes in maternal blood could reflect an underlying shared genetic risk for ASD in these high-risk families.

Methylation and methyl-binding functions in a gene module oppositely associated with folic acid and ASD risk

The complementary co-expression network analysis further revealed a module of 224 co-expressed genes in maternal blood showing an association with folic acid and 5-MeTHF levels in the opposite direction from ASD risk that could not be explained by cell type differences. Interestingly, these ASD and nutrient associated genes were functionally enriched for DNA methylation binding and methylation-dependent chromatin silencing, consistent with prior DNA methylation changes observed in ASD (136,140,251,309) as well as ASD-like syndromes associated with methyl binding proteins (310,311). *MBD3L*, which has methyl-binding function, is predicted to assist with demethylation reactions and functions as a transcriptional repressor (312–314). Folic acid, the synthetic form of folate that contributes the substrate for one-carbon

metabolism, and 5-MeTHF, one of the active biological forms of folate that plays a critical role in one-carbon metabolism, have also been shown to be inversely associated with developmental delay (70,270,315,316).

One-carbon metabolites associated with changes in gene expression in this study have also been associated with the prevention of neurodevelopmental conditions (136,317–319). The co-regulated block of betaine and DMG co-expression modules contained genes enriched in the adaptive immune system and chromatin modification functions, as well as Rett syndrome, a known syndromic form of ASD (18,43,320,321). Choline is metabolized to betaine, which converts homocysteine to form methionine, generating DMG in the one-carbon pathway (322,323). A previous study of maternal peripheral blood collected at term showed that changes in betaine and DMG were in the opposite direction from choline when compared with nonpregnant women (324). *EVL* (Enah/Vasp-like) is involved in actin cytoskeleton remodeling and is crucial for central nervous system processes and immune system functions (325–327). One study also showed *EVL* as a differentially expressed gene in schizophrenia in peripheral blood (326).

Previous studies in ASD have been focused on post mortem brain tissue (219,328), as a tissue relevant to the condition, but collected after diagnoses were made, raising concerns about reverse causation in determining etiologically-relevant expression changes. Few studies have focused on prospective transcriptomic profiles collected prior to the presentation of the condition (281,329), and none have examined maternal gene expression profiles during

pregnancies at high risk for developing ASD. In addition, few studies have integrated maternal transcriptome and one-carbon metabolite data within biospecimens. Furthermore, most studies of ASD expression biomarkers have not considered the roles of nutritional factors during pregnancy that could be relevant to fetal development.

A limitation of using maternal peripheral blood to examine expression is that it contains multiple cell types, and proportions can differ across samples. However, estimated cell type composition of maternal blood was not significantly associated with the child's clinical outcomes, the "greenyellow" module, betaine, or DMG, which suggests that our main findings were not driven by differential cell type proportions. After correcting for multiple comparisons, this study did not identify any individual differentially maternally expressed genes specifically associated with ASD, although 6 transcripts from 4 genes reached genome-wide significance with diagnosis of either ASD or Non-TD. Another potential limitation is the relatively low number of samples with certain metabolite measurements available. Future studies on studies on maternal nutritional factors and child outcomes would be beneficial in confirming our data. Furthermore, lack of genome-wide evidence of individual differentially expressed genes specific to a pairwise comparison of ASD vs TD is likely due to the relatively small sample size that is inherent to a prospective ASD study, that is underpowered to detect small differences in transcript levels. However, this does not eliminate the importance of identifying and understanding the biologically significant gene set enrichments and co-expression network modules using differential gene expression and WGCNA analysis. Additionally, other factors, including genetics, epigenetics, gestational age, and other environmental factors can influence

the transcriptome and ASD risk. Approaches incorporating those factors will be important in future studies.

In summary, genome-wide gene expression analysis of maternal peripheral blood samples revealed transcriptome changes associated with maternal one-carbon metabolites and child neurodevelopmental outcomes implicating maternal immune, apoptotic, and epigenetic mechanisms in the development of ASD in offspring. In addition, folic acid and 5-MeTHF were associated with expression of genes involved in methylated-CpG binding in an opposite direction to that of ASD, consistent with prior evidence of protection. Finally, maternal betaine and DMG levels clustered with co-expressed genes related to immune, chromatin modification, and development functions. These results therefore provide important biological insights into maternal gene pathways associated with adverse neurodevelopment in the child, as well as the suggested protective association with one carbon metabolites in the complex etiology of ASD.

Methods

MARBLES study design

The MARBLES study recruited mothers in Northern California with at least one child with ASD who were pregnant or planning another pregnancy. Due to a shared genetic background, 24% of the next children met the criteria for ASD. A previous publication detailed the study design of MARBLES (79,281). In order to enroll into MARBLES, all five of the following criteria needed to be met: (1) the prospective child has one or more first or second degree relatives with ASD; (2) mother is 18 years or older; (3) mother is pregnant or able to become pregnant; (4) mother is

able to speak, read, and understand English and plans to raise the child with English spoken at home; (5) mother lives within a 2.5-hour drive from Davis/Sacramento, California. Demographic, diet, environmental, and medical information were collected by telephone interviews or questionnaires throughout the pregnancy. Infants received standardized neurodevelopmental assessments from 6 months until 3 years old (79). At 3 years old, the child was assessed clinically using the gold standard Autism Diagnostic Observation Schedule (ADOS) (245), the Autism Diagnostic Interview – Revised (ADI-R) (330), and the Mullen Scales of Early Learning (MSEL) (247). Based on a previously published algorithm using ADOS and MSEL scores (122,281), participants were classified into three outcome groups including ASD, TD, and Non-TD (248,249). Children with ASD had scores over the ADOS cutoff and fit ASD DSM-5 criteria. Children with Non-TD outcomes were defined as children with low MSEL scores (two or more MSEL subscales with more than 1.5 standard deviations (SD) below averages or at least one MSEL subscale more than 2 SD below average) and elevated ADOS scores. Children with TD outcome had all MSEL scores within 2.0 SD and no more than one MSEL subscale that is 1.5 SD below the normative mean and scores on the ADOS at least three points lower than the ASD cutoff.

RNA isolation and expression microarray

Maternal peripheral blood was collected at study visits during all three trimesters of pregnancy in PAXgene Blood RNA tubes with the RNA stabilization reagent (BD Biosciences) and stored frozen at -80°C. The first timepoint sample was used for mothers who had multiple blood draws (n = 12) during pregnancy. RNA was isolated using the PAXgene Blood RNA Kit (Qiagen)

according to the default protocol. Total RNA was converted to cDNA and biotin labeled. Expression was measured using Human Gene 2.0 Affymetrix microarray chips with three batches by the John Hopkins Sequencing and Microarray core following washing, staining, and scanning procedures based on manufacturer's protocol.

Data preprocessing and normalization

Robust Multi-Chip Average (RMA) (331–333) from the oligo R package was used for normalization of Affymetrix CEL files. For quality control, we used the oligo and ArrayQualityMetrics R packages (334,335). No samples were identified as outliers by principal component analysis, the Kolmogorov-Smirnov test, or Euclidean distance to other arrays. Probes were mapped at the transcript level using the pd.hugene.2.0.st R package, and those annotated to genes (36,459) were used in subsequent analyses.

One-carbon nutrient metabolite measurements

Serum and plasma samples from the same blood draw as specimens used for RNA expression analysis were used to measure one-carbon and nutrient metabolites. Metabolites were selected for their role in one-carbon metabolism and may be relevant to gene expression. S-adenosylmethionine (SAM) and S-adenosylhomocysteine (SAH), adenosine, homocysteine, cystine, cysteine, glutathione (GSH) and glutathione disulfide (GSSG) were measured in the James' laboratory at the Arkansas Children's Research Institute using HPLC with electrochemical detection as previously described (336,337). Serum pyridoxal phosphate (PLP), the biologically active form of vitamin B6 (Vit B6), was measured by HPLC using fluorescence detection in the

Green-Miller laboratory at the UC Davis Medical Center (inter-assay coefficient of variation (CV) = 4.8%) (338). Total serum vitamin B12 (Vit B12) was measured using automated chemiluminescence in the CLIA-approved Medicine Clinical Laboratories at UC Davis Medical Center (inter-assay CV = 6.2%). Plasma choline, betaine and dimethylglycine (DMG) were measured using LC-MS/MS stable isotope dilution methods in the Caudill laboratory (339,340) with modifications to include measurements of trimethylamine N-oxide (TMAO) and methionine (341,342). Intra-assay and inter-assay CVs of the in-house controls were 3.0% and 3.6% for choline; 1.5% and 1.7% for betaine, 2.5% and 2.4% for DMG; 2.6% and 2.6% for methionine; and 3.1% and 3.4% for TMAO. Serum 5-methyltetrahydrolate (5-MeTHF) and folic acid were quantified in the Caudill laboratory using LC-MS/MS stable-isotope dilution methods (343) with modifications based on the instrumentation (344). Intra-assay and inter-assay CVs of in -house controls were 1.8% and 1.9% for 5-methyltetrahydrofolate; and 4.9 and 8.5% for folic acid.

Differential gene expression

After normalization, surrogate variable analysis (SVA) was used to estimate and adjust for hidden confounding variables on gene expression (345). Differential gene expression was identified using the limma R package by a linear model that included the children's diagnosis outcome (ASD, Non-TD and TD) and all surrogate variables (346). Differential gene expression analysis with children diagnosed as ASD, Non-TD and TD were included in the same model with three levels of diagnosis using the F-test (346). Pairwise fold change, standard error and p-value between ASD vs TD, and Non-TD vs TD were extracted from the same model using the limma R

package. Differentially expressed transcripts were identified as those with an unadjusted *p*-value <0.05. Genome-wide significant differentially expressed transcripts were classified as those with a false discovery rate (FDR) adjusted *p*-value (*q*-value) <0.05. To complement adjustment by SVA, we used an alternative approach that adjusted for known confounders, including batches, trimesters, child gender and cell types using limFit function in limma R package. Differentially expressed transcripts between ASD vs TD, and Non-TD vs TD were extracted using the same approach as SVA adjustment and compared.

Gene overlap analysis

Gene overlap analysis was performed by Fisher's exact test using the GeneOverlap R package (347). In each comparison, the null distribution was generated from 1000 random samples of all genes annotated to transcripts on the array. Gene symbols annotated to differentially expressed transcripts were compared to 943 genes in the Simons Foundation Autism Research Initiative (SFARI) Gene database (191). As an alternative approach to using the whole probe set on the Affymetrix arrays (36,459) as background, an analysis of filtered probe values was performed based on removing probe intensity on the lowest 5% intensity probes (34,636) to remove genes with the lowest expression levels.

Gene ontology (GO) term and pathway enrichment analysis

Transcripts with significant expression levels or selected gene lists were exported to DAVID bioinformatics software with default settings for GO analysis (255,256). The analysis was done using the GO ontology database and Fisher's exact test with multiple test correction by the FDR

method (256). GO term enrichments were presented with hierarchical terms. GO terms with an FDR *q*-value <0.05 were considered statistically significant. GO terms enrichment analyses was performed separately for both all genes and filtered probes representing expressed genes.

Weighted Gene Co-Expression Network Analysis (WGCNA)

A weighted gene co-expression network was built using the WGCNA R package (348,349) with normalized expression levels after adjustment for batch effects (3 batches) using the ComBat function from the sva R package (350). The correlation matrix included all probes and all samples. To construct a signed adjacency matrix, estimated soft thresholding power (power = 6) was used to achieve approximately scale-free topology (R^2 fit >0.8). Adjacency values were transformed into a signed topological overlap matrix (TOM). Co-expression modules were identified from the dissimilarity matrix (1-TOM) with a minimum module size of 30 probes using Pearson's coefficient. Module eigengenes were clustered based on correlation. Similar modules were merged based on a cut height of 0.25 to generate co-expression modules. Each module's expression profile was summarized into a module eigengene (ME) using the matched module's first principal component. The correlation between each gene in the module with the ME was represented as intramodule connectivity (kME). Module hub probes were defined as the probe in each module with the highest module membership. Hierarchical clustering was done using the standard R function hclust with the default setting using ward's agglomeration method (351). Pearson's correlation coefficient was used to measure the correlation between traits and modules. Highly correlated modules were defined as those with an FDR-adjusted *p*-value <0.05.

Cell type proportion deconvolution

CIBERSORT was used to estimate the proportions of each cell type using the default settings and the LM22 adult peripheral blood signature gene expression profiles (352). Normalized expression levels adjusted for batch effects were used to estimate cell type proportions. Both relative and absolute modes were performed together with 100 permutation tests. *P*-values were calculated using FDR multiple test adjustment. Significant associations were defined based on FDR *q*-value <0.05.

Availability of data and material

Data are shared in the Gene Expression Omnibus (GEO) accession number (GSE148450) based on participates consent. Code and scripts for this study are available on GitHub (https://github.com/Yihui-Zhu/AutismMaternalBloodExpression). Other related data and information are included in supplementary tables.

Chapter 4 – Placental methylome reveals a 22q13.33 brain regulatory gene locus associated with autism risk

Yihui Zhu, J. Antonio Gomez, Benjamin I. Laufer, Dag H. Yasui, Charles E. Mordaunt, Daniela C. Soto, Megan Y. Dennis, Kelly S. Benke, Kelly M. Bakulski, John Dou, Julia M. Jianu, Ria Marathe, Cheryl K. Walker, Sally Ozonoff, Jason I. Feinberg, M. Daniele Fallin, Irva Hertz-Picciotto, Rebecca J. Schmidt, Janine M. LaSalle

Abstract

Most autism spectrum disorder (ASD) cases involve complex genetics interacting with perinatal environment, complicating the discovery of common genetic risk. The epigenetic layer of DNA methylation shows dynamic developmental changes and molecular memory of *in utero* experiences, particularly in placenta, a fetal tissue discarded at birth. However, current array-based methods to identify novel ASD risk genes lack coverage of the most structurally and epigenetically variable regions of the human genome. Here we used whole genome bisulfite sequencing (WGBS) in placenta samples from prospective ASD studies to discover a previously uncharacterized ASD risk gene in a comethylated block at 22q13.33. Differentially methylated region (DMR) analysis identified 139 DMRs discriminating ASD in placental samples, including a high-confidence 118 kb hypomethylated block at 22q13.33 that replicated in two additional cohorts. A polymorphic H3K4me3 peak in placenta corresponded to novel transcript *LOC105373085* (renamed *NHIP*) with high expression in brain, increased expression following neuronal differentiation or hypoxia, but decreased expression in ASD placenta and brain. Transient *NHIP* overexpression increased cellular proliferation and altered expression of genes

regulating synapses and neurogenesis in response to hypoxia, significantly overlapping with *NHIP*-associated transcript levels in ASD brain and known ASD risk genes. A common structural variant near a fetal brain enhancer was associated with *NHIP* placental and brain expression levels and ASD risk, demonstrating a common genetic influence on DNA methylation levels. Together, these results demonstrate a novel environmentally-responsive ASD risk gene relevant to brain development in a relatively uncharacterized region of the human genome.

Introduction

Autism spectrum disorders (ASD) are of growing prevalence, with 1 in 54 children diagnosed in the US (59). Diagnosis of ASD is based on a child's behavioral difficulties in social communication and interactions, restricted and repetitive behaviors and interests, and language deficits. The etiology of ASD is complex and heterogeneous, and expected to involve multiple genetic and environmental factors, as well as poorly understood gene-environment interactions (60,61). Twin and sibling studies have shown a strong heritability of ASD risk within families, and most genetic risk for ASD is expected to come from common variants (353). Exome sequencing of ASD trios has identified genes mutated in rare genetic ASD cases, which are enriched for neuronal, embryonic development and chromatin regulation functions, but no single gene cause can explain more than 1% of disease liability (65,176). A large genome-wide association study (GWAS) calculated that an individual's ASD risk depends on the level of polygenic burden of thousands of common variants in a dose-dependent manner (64). Furthermore, ASD genetic susceptibility could be improved by adding SNP weights using polygenic risk scores (PRS) from ASD-correlated traits, including schizophrenia, depression, and

educational attainment (64,108,354) . In comparison, GWAS for schizophrenia risk has identified 108 loci, but PRS for schizophrenia is more than five times greater in the presence of early-life maternal complications, and corresponded to differences in placental gene expression (355). Term placenta is an accessible tissue normally discarded at birth, but the convergence between placental biology and genetic risk for ASD is understudied.

Placenta maintains a distinct landscape of DNA methylation characterized by partially methylated domains (PMDs), which is more similar to oocytes and preimplantation embryos than fetal or adult tissues (160,161,166). However, prior investigations have demonstrated that placenta is a promising tissue for identifying methylation alterations for genes relevant to fetal brain and gene by environment interactions in ASD (136,170,171,356). Most epigenome-wide association studies (EWAS) for ASD have used array-based reduced-representation probes for DNA methylation differences which lack coverage over the most epigenetically and genetically polymorphic regions of the human genome (357). Correlated regions of systemic interindividual variation (CoRSIVs) are sites of individual variation in DNA methylation that are sensitive to periconceptional environment, observed across diverse tissues, and associated with human disease genes (357,358). Structural variant (SV) have also been associated with many human phenotypes, especially immune response, and cognitive disorders, such as schizophrenia(110–112). SVs exhibit a nonrandom distribution in hotspots within relatively gene-poor regions in primate genomes, but are enriched for gene functions in oxygen transport, sensory perception, synapse assembly, and antigen-binding (113,114). Recent studies suggested that a large SV burden was associated with lower cognitive ability (115–117) and ASD (118), but GWAS and

EWAS studies are blind to most SVs and CoRSIVs in the genome, leaving many unexplored genomic regions to be discovered through sequencing-based epigenome-wide investigations in placenta.

Here, we investigated the association of ASD risk with placental DNA methylation in two high-familial ASD risk cohorts using WGBS on 203 individuals. We identified a block of differential methylation in ASD at 22q13.33, a region previously described as CoRSIV and SV hotspot but not previously associated with ASD. A novel gene *LOC105373085* (renamed *NHIP* for neuronal hypoxia inducible peptide) within 22q13.33 was demonstrated to be expressed in brain, responsive to oxidative stress, and a transcriptional regulator of additional known ASD risk genes. A common SV insertion within 22q13.33 was significantly associated with increased ASD risk, reduced expression of *NHIP,* and reduced methylation, but first month prenatal vitamin use counteracted this effect. Together, these results demonstrate a novel ASD risk gene regulatory locus at the interface of common genetics and perinatal environmental resilience.

Results

Differential methylation analysis using WGBS identifies a hypomethylated block at 22q13.33 in ASD placenta

To identify novel regions of epigenetic differences in placenta discriminating later ASD diagnosis, we performed WGBS analysis of genome-wide DNA methylation on 204 subjects from the two prospective high-risk ASD cohorts (MARBLES and EARLI) with diagnosis outcome at 36 months (**Fig. 4.1a**, **Supplementary Table 4.1**, **Supplementary Table 4.2**). ADOS and MSEL

scores that were used in the diagnostic algorithm for ASD were significantly associated with ASD outcome, but no other demographic or technical variables were significantly associated (**Supplementary Table 4.1**). Global methylation levels over 20 kb windows were also not significantly different by diagnostic group (**Supplementary Fig. 4.1**). Discovery, external replication and internal replication groups were analyzed separately, since sequencing platform differences impacted global methylation levels (**Supplementary Fig. 4.2**). Differentially methylated regions (DMRs) distinguishing ASD from TD placental samples were identified with a permutation-based statistical approach adjusted for sex and placental cell types to identify broad epigenomic signatures of multiple gene regulatory regions at a genome-wide level in the discovery group (ASD n = 46, TD n = 46) (**Supplementary Table 4.2**). 134 DMRs, including 77 hyper- and 57 hypo-methylated in ASD, mapped to 183 genes (**Fig. 4.1b**, **Supplementary Table 4.3**). A cluster of 12 ASD DMRs mapped to 22q13.33, including one that also passed genome-wide significance (FDR adjusted *p*-value < 0.05). Methylation levels within ASD DMRs were specifically associated with ADOS and MSEL scores, but not other demographic and technical variables (**Supplementary Fig. 4.3**). Further evidence that DMRs identified in placenta reflect epigenetic differences relevant to brain came from the significant enrichment of ASD DMRs in enhancers of fetal brain, as well as bivalent enhancer and repressed polycomb regions of placenta compared to background regions (**Fig. 4.1c**). Demonstrating their functional relevance, both hyper- and hypo methylated ASD DMRs were enriched 5 kb and 5 - 50 kb downstream of TSS, at CpG islands and shores (**Supplementary Fig. 4.4-4.5**), and at known transcription factor binding sites (**Supplementary Table 4.4**). Genes mapped to placental ASD DMRs significantly overlapped with ASD risk genes from the Simons Foundation Autism Research Initiative (SFARI)

dataset (191) (**Supplementary Fig. 4.6**, **Supplementary Table 4.5**). The overrepresentation of 12 DMRs at 22q13.33 hypomethylated in ASD drove the additional enrichment at >500 kb of TSS as well as gene ontology (GO) enrichment for functions in histone acetyltransferase (HATs) and chromatin modification, due to the assignment of the 22q13.33 hypomethylated DMRs to the nearest downstream gene *BRD1*, a histone acetyltransferase (**Supplementary Fig. 4.7**, **Supplementary Table 4.6**). Based on these results, we decided to focus downstream analyses on further understanding the impact of the 22q13.33 hypomethylated locus on ASD risk.

The 22q13.33 DMRs hypomethylated in ASD were highly positively correlated with each other and formed a 118 kb hypomethylation cluster that was also detected as a hypomethylated block (chr22: 49044669 - 49162642, hg38) (**Fig. 4.1d, Supplemental Fig. 4.8**), which was also previously described as a CoRSIV (357). We therefore examined smoothed methylation levels over the 118 kb 22q13.33 block for replication in a different ASD enriched risk cohort (EARLI, external replication group). Similar to the discovery group, 22q13.33 block methylation levels in ASD were significantly lower in TD (**Fig. 4.1e**, **Supplementary Table 4.7**). Furthermore, an independent "internal replication group" of MARBLES subjects using a different sequencing platform also showed significantly lower DNA methylation levels in ASD compared to either TD or the additional diagnostic Non-TD samples, defined as atypical scores but not reaching ASD diagnosis (**Fig. 4.1e**, **Supplementary Table 4.7**). These results demonstrate that hypomethylation over the 118 kb 22q13.33 comethylated block is a reproducible finding across different cohorts and platforms, specifically distinguishing placental samples from newborns later diagnosed with ASD.

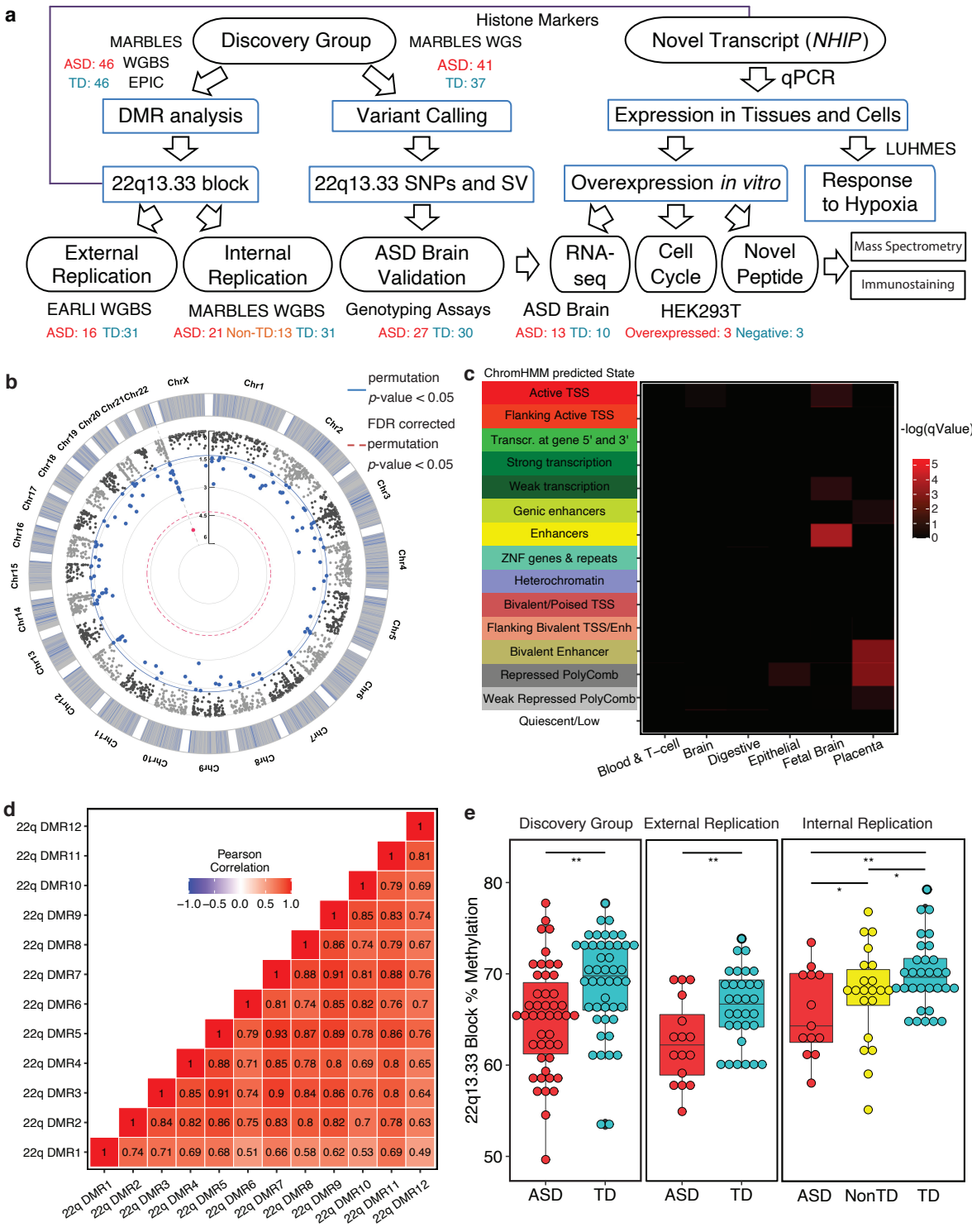

a
MARBLES
ASD: 46 WGBS
TD: 46 EPIC
Discovery Group
Histone Markers
MARBLES WGS
ASD: 41
TD: 37
Novel Transcript (NHIP)
qPCR
DMR analysis
Variant Calling
Expression in Tissues and Cells
LUHMES
22q13.33 block
22q13.33 SNPs and SV
Overexpression in vitro
Response to Hypoxia
External Replication
Internal Replication
ASD Brain Validation
RNA-seq
Cell Cycle
Novel Peptide
Mass Spectrometry
Immunostaining
EARLI WGBS
ASD: 16 TD:31
MARBLES WGBS
ASD: 21 Non-TD:13 TD: 31
Genotyping Assays
ASD: 27 TD: 30
ASD Brain
ASD: 13 TD: 10
HEK293T
Overexpressed: 3 Negative: 3
b
permutation
p-value < 0.05
FDR corrected
permutation
p-value < 0.05
c
ChromHMM predicted State
Active TSS
Flanking Active TSS
Transcr. at gene 5' and 3'
Strong transcription
Weak transcription
Genic enhancers
Enhancers
ZNF genes & repeats
Heterochromatin
Bivalent/Poised TSS
Flanking Bivalent TSS/Enh
Bivalent Enhancer
Repressed PolyComb
Weak Repressed PolyComb
Quiescent/Low
-log(qValue)
Blood & T-cell
Brain
Digestive
Epithelial
Fetal Brain
Placenta
d
Pearson Correlation
e
Discovery Group
External Replication
Internal Replication
22q13.33 Block % Methylation
ASD
TD
NonTD

Figure 4.1. ASD associated DMRs are enriched at fetal brains enhancers and a comethylated block at 22q13.33 replicated across studies and platforms.

(**a**) Schematic of the experimental design for discovery of ASD DMRs, replication of the comethylated 22q13.33 locus, genetic associations, and functional follow-up of novel transcript (*NHIP*).

(**b**) Circular Manhattan plot of the epigenome-wide association of DNA methylation in placenta with ASD diagnosis at 36 months. Results are represented as DMR association test results ($-\log_{10}(p)$) ordered by genomics position. Significant thresholds are blue for permutation *p*-value < 0.05, red for FDR adjusted permutation *q*-value < 0.05, and grey for nonsignificant.

(**c**) 134 ASD DMRs (permutation *p*-value < 0.05) tested for enrichment within chromatin states defined by Epigenome Roadmap ChromHMM states. Each row represents a different ChromHMM predicted state and each column a single tissue type, with the heatmap plotting the $-\log_{10}(q\text{-value})$ significance of ASD DMR enrichment.

(**d**) The triangle correlation matrix of methylation levels using the Pearson correction r score for the 12 DMRs located in the 22q13.33 hypomethylated block.

(**e**) The smooth methylation values were averaged over the 22q13.33 hypomethylated block (y-axis) and compared across diagnosis groups (x-axis). In discovery group, ASD samples had significant lower methylation than TD samples (MARBLES, HiSeq X, ASD n = 46, TD n = 46) (*p*-value = 0.003). The same result and direction were observed in the external replication group (EARLI, HiSeq 2500, ASD n = 16, TD n = 31) (*p*-value = 0.006). For the internal replication group (MARBLES, NovaSeq, ASD n = 21, Non-TD n = 13, TD n = 31), ASD methylation levels were also significantly lower than TD samples (*p*-value = 0.003). Non-TD had lower methylation than TD

(*p*-value = 0.048) and higher methylation than ASD on 22q13 block (*p*-value = 0.049). Comparisons used two-tailed t-test. Each dot represented one sample. Different sequencing platforms used in the three cohorts led to the absolute differences in global DNA methylation levels (**Supplementary Fig. 4.1**).

NHIP is a primate-specific gene dynamically expressed during neuronal differentiation but exhibiting reduced expression in ASD

The 22q13.33 comethylated block was within an apparent gene desert, located more the 500 kb away from the closest annotated protein coding genes, *FAM19A5 (TAFA5)* and *BRD1*. Epigenetic evidence for promoter and enhancer activity within 22q13.33 was obtained from placenta, ovary, and brain ENCODE datasets (**Supplementary Fig. 4.9**) (359). Within 22q13.33, an active promoter peak identified by H3K4me3 histone markers was observed in 1/2 of ovary, 2/3 of placenta, and 2/3 of brain samples, suggesting polymorphic promoter marks between individuals (**Supplementary Fig. 4.10**). This H3K4me3 peak overlapped a CpG island and the TSS of the uncharacterized transcript, *LOC105373085* (also named *AK057312*) identified from a human testis cDNA library (**Supplementary Fig. 4.10**) (360). We renamed *LOC105373085* as *NHIP*, for *Neuronal Hypoxia Inducible Peptide*. *NHIP* was also variably expressed among brain regions from the Genotype-Tissue Expression (GTEx) database (**Supplementary Fig. 4.11**) (361). *NHIP* was highly conserved in primates, but not in other vertebrates including mouse (Sup**plementary Fig. 4.12**, **Supplementary Table 4.8**). When quantified by RT-PCR in human tissues, *NHIP* was expressed in placenta, testis, adult and fetal brain, with relatively lower expression in placenta (**Fig. 4.2a**). ASD placental samples showed significantly lower *NHIP*

transcript levels than TD samples, in the same direction as methylation changes in the 22q13 block (**Fig. 4.2b**). Since gene body methylation in placenta predicts active gene expression and the 22q13.33 comethylated block mapped to a previously-defined partially methylated domain in placenta (161) (**Supplementary Fig. 4.13**), these results indicate that hypomethylation of the 22q13.33 block in ASD was reflective of lower past or current expression of *NHIP* expression *in utero* for ASD compared to TD.

NHIP was expressed in multiple human cell lines (IMR90, LUHMES, SH-SY5Y) but neuronal differentiation in LUHMES cells resulted in a significant increase in *NHIP* transcript levels compared to undifferentiated (**Fig. 4.2c**). Differentiated LUHMES neurons were more sensitive to treatment with a hypoxia mimetic ($CoCl_2$) than undifferentiated, with significant decrease in cell viability and increase in reactive oxygen species (ROS) levels (**Fig. 4.2d**, **Supplementary Fig. 4.14**). *NHIP* transcript levels increased after exposure $CoCl_2$ specifically in differentiated, but not undifferentiated LUHMES cells (**Fig. 4.2e**, **Supplementary Fig. 4.15**). Among the tested human cell lines, embryonic origin HEK293T had the lowest endogenous *NHIP* transcript levels (**Fig. 4.2c**). Since response to hypoxia is a developmental signal regulating cell proliferation decision in embryos (362), we experimentally tested this phenotype by transiently transfecting HEK293T cells with either a plasmid encoding *NHIP* with a dual GFP-Puromycin selection cassette or a matched vector control lacking *NHIP* (**Supplementary Fig. 4.16**, **Supplementary Table 4.9**). A significantly shortened doubling time was observed in response to *NHIP* overexpression compared to control cells (20.23 h vs.24.91 h) (**Fig. 4.2f**, **Supplementary Fig. 4.17**).

To examine whether *NHIP* encoded a protein, we identified a 20 amino acid putative peptide containing a Kozak sequence and tested the existence of the peptide by designing the pb-NHIP-peptide-eGFP vector so that the peptide sequence would be in frame with ATG-less GFP transiently transfected in HEK293T cells (**Fig. 4.2g**, **Supplementary Table 4.9**). The presence of both transfection control (red, mCherry) and reporter (green, eGFP) confirmed the existence of the 20 aa peptide from *NHIP* (**Fig. 4.2g**). The NHIP encoded peptide sequence was confirmed using mass spectrometry after pull-down with anti-GFP antibody (**Supplementary Fig. 4.18**, **Supplementary Table 4.10**). A blastp search of human databases demonstrated that the *NHIP* peptide partially overlapped protein sequences within BRCA2 and CHD4 (**Supplementary Fig. 4.19**, **Supplementary Table 4.10**). Lastly, using a custom antibody against the *NHIP* encoded peptide, immunostaining was performed on sections of human postmortem prefrontal cortex, demonstrating nuclear staining in a subset of neuronal nuclei (**Fig. 4.2h**). Together, these results demonstrate the existence of a nuclear peptide encoded by *NHIP*.

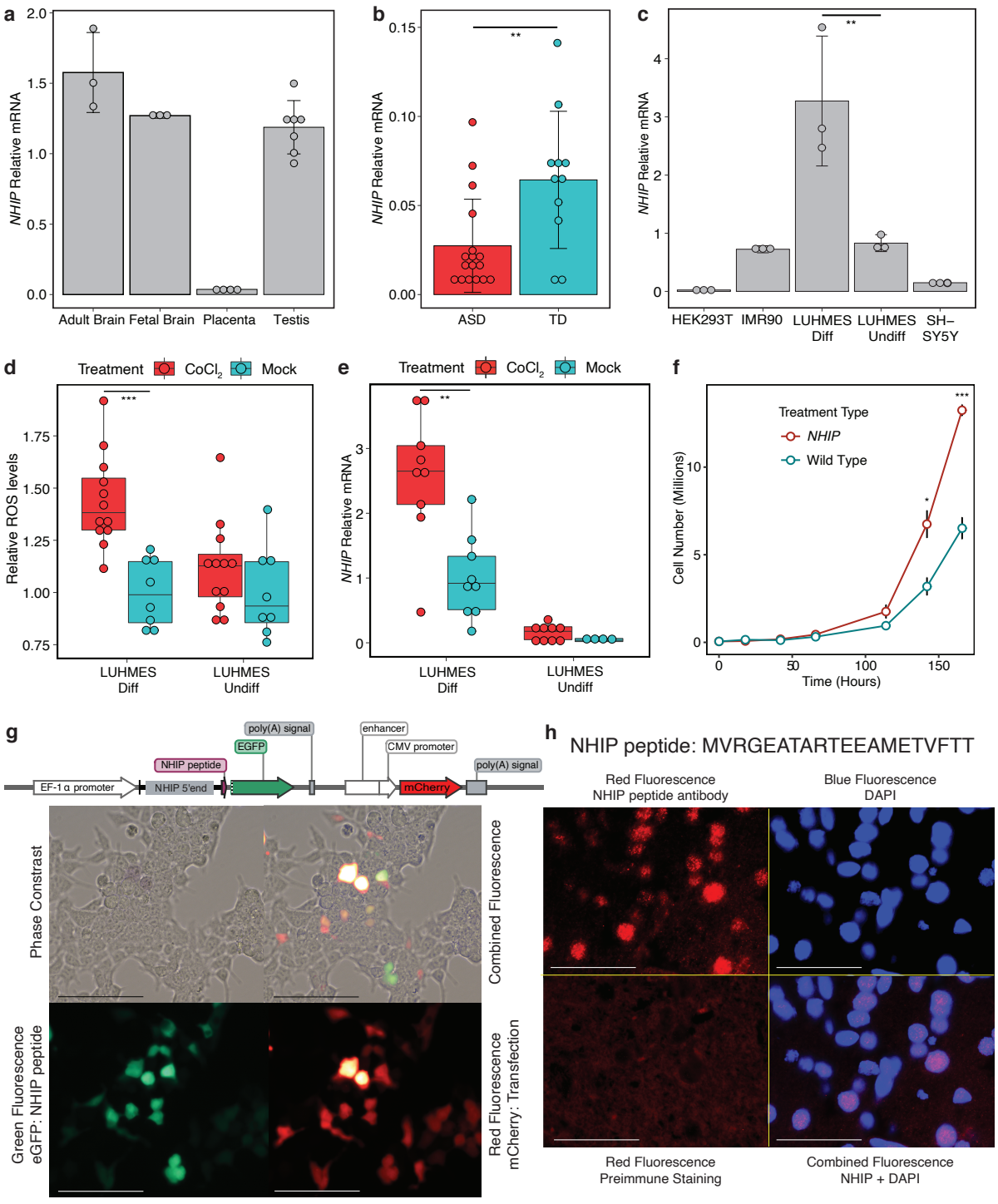

a
NHIP Relative mRNA
Adult Brain
Fetal Brain
Placenta
Testis
b
NHIP Relative mRNA
ASD
TD
c
NHIP Relative mRNA
HEK293T
IMR90
LUHMES Diff
LUHMES Undiff
SH–SY5Y
d
Treatment
CoCl2
Mock
Relative ROS levels
LUHMES Diff
LUHMES Undiff
e
Treatment
CoCl2
Mock
NHIP Relative mRNA
LUHMES Diff
LUHMES Undiff
f
Treatment Type
NHIP
Wild Type
Cell Number (Millions)
Time (Hours)
g
poly(A) signal
EGFP
enhancer
CMV promoter
NHIP peptide
poly(A) signal
EF-1 α promoter
NHIP 5'end
mCherry
Phase Constrast
Combined Fluorescence
Green Fluorescence
eGFP: NHIP peptide
Red Fluorescence
mCherry: Transfection
h
NHIP peptide: MVRGEATARTEEAMETVFTT
Red Fluorescence
NHIP peptide antibody
Blue Fluorescence
DAPI
Red Fluorescence
Preimmune Staining
Combined Fluorescence
NHIP + DAPI

Figure 4.2. *NHIP* transcript levels in tissues and cells and in response to hypoxia and evidence for *NHIP* encoded nuclear peptide

In RT-qPCR assays, *NHIP* levels were normalized to *GAPDH* with at least three independent experiments per condition.

(**a**) *NHIP* levels in human tissues, including adult brain, fetal brain, placenta, and testis.

(**b**) *NHIP* levels in placenta samples from discovery group (ASD n = 17, TD n = 11). ASD samples show significantly lower *NHIP* levels than TD samples (two-tailed t-test, *p*-value = 0.005).

(**c**) *NHIP* levels in human cell lines, HEK293T, IMR90, LUHMES, and SH-SY5Y. In LUHMES cells, *NHIP* levels were significantly higher in differentiated neurons compared to undifferentiated cells (two-tailed t-test, *p*-value = 0.020).

(**d**) Differentiated LUHMES cells are more sensitive to hypoxia than undifferentiated LUHMES cells. Formation of reactive oxygen species (ROS) was measured in differentiated and undifferentiated LUHMES cells treated with 100nM $CoCl_2$, a hyopoxia mimetic, or vehicle (mock) (two-tailed t-test, *p*-value = 0.001).

(**e**) *NHIP* levels increase in response to hypoxia, specifically in differentiated neurons. Differentiated or undifferentiated LUHMES cells were treated with 100nM $CoCl_2$. In differentiated LUHMES cells, $CoCl_2$ treatment significantly increased *NHIP* levels (two-tailed t-test, *p*-value = 0.016).

(**f**) *NHIP* transient overexpression in HEK293T cells resulted in a faster doubling time than vector control cells, indicating increased cell proliferation.

(**g**) Vector design of pb-NHIP-peptide-eGFP (dotted line represents excised ATG of EGFP) and combined phase and fluorescent microscopy. Green, eGFP linked to NHIP peptide; red, mCherry, transfection positive control. Scale bars, 100 µm.

(**h**) Immunofluorescent staining of human frontal cortex, showing nuclear localization with anti-NHIP, but not pre-immune control. Blue, DAPI nuclear counterstain; red, anti-NHIP staining. Scale bars, 100 µm.

A common genetic structural variant at 22q13.33 is associated with reduced placental DNA methylation, reduced *NHIP* expression, and increased ASD risk

To examine genetic factors associated with 22q13.33 methylation levels and polymorphic expression of *NHIP* in ASD, we tested the association between 22q13.33 block methylation level and common variants from individual-matched WGS, including SNPs, InDels, CNVs, and SVs. Methylation levels in five out of 12 ASD DMRs within 22q13.33 were significantly associated with common SNPs located inside the DMRs (**Supplementary Fig. 4.20**). An upstream 1,674 bp SV insertion (chr22: 49029657, hg38) located 15,013 bp from the start site of the 22q13.33 comethylated block (**Fig. 4.3a**, **Supplementary Table 4.11**) was identified in which methylation levels of all 12 22q13.33 ASD DMRs were significant associated (**Fig. 4.3b**). In the MARBLES cohort, this SV insertion was identified in significantly more ASD compared to TD samples (**Supplementary Fig. 4.21**). Placenta samples with the 22q13.33 insertion and ASD, but not TD, showed significantly lower methylation levels (**Fig. 4.3c**). While not present in the reference genome, the 22q13.33 insertion was also identified as structural variant in a long-read PacBio sequence CHM1_chr22-49029645-INS-1673 contig (363), which also can be found in the NCBI

GenBank ID QPKN01007947.1 (364) (**Supplementary Table 4.11, Supplementary Fig. 4.22**). We also confirmed the WGS evidence of SV using two PCR genotyping primer sets that discriminated two alleles, with homozygosity of the insertion significantly associated with ASD (**Supplementary Table 4.11**, **Supplementary Fig. 4.23**). The insertion sequence showed high similarity with retrotransposon elements, including SVA and Alu (**Supplementary Fig. 4.24**). This 22q13.33 SV also corresponded to INS_22_115103 in Genome Aggregation Database (gnomAD) (**Supplementary Fig. 4.25**) (365,366).

Since the 22q13.33 block only showed lower methylation in ASD but not TD placental samples, we tested the hypothesis that prenatal vitamin use during the first month of pregnancy, previously shown to be associated with decreased ASD risk (122), may mediate a protective effect on the ASD SV risk associated with the insertion. At the 22q13.33 block, there was a significant positive association with prenatal vitamins use in the first month and methylation level, in the protective direction (**Fig. 4.3d**). When samples were separated by 22q13.33 insertion genotype, prenatal vitamin use during the first month of pregnancy showed a significant protective effect specifically in individuals with the insertion (**Supplementary Fig. 4.26**). Unlike the 22q13.33 insertion, the GWAS-based PRS (64) calculated for the MARBLES cohort was not significantly predictive of either ASD diagnosis or 22q13.33 block methylation by ANOVA in the MARBLES discovery cohort (**Supplementary Table 4.12**). Together, these results are consistent with the hypothesis that ASD risk associated with the 22q13.33 SV and comethylated block is distinguishable from polygenic ASD risk and tempered by a common nutrient intervention with demonstrated ASD protection.

Since SVs have been previously implicated in altering chromatin loops regulating promoter-enhancer interactions (367), we hypothesized that this 1.7 kb insertion may be located within an enhancer-promoter loop relevant to fetal brain. Using the recent EpiMap database of chromatin states across multiple humans and tissue types (368), we identified two CTCF sites flanking the SV insertion (**Fig. 4.3e**). ChromHMM maps (264) demonstrate a fetal brain enhancer that aligns with the distal CTCF binding site. The proximal CTCF site is adjacent to the *NHIP* TSS, which ChromHMM predicts as an active promoter in brain, ovary, and placenta. Neither CTCF binding site corresponded to the boundaries of a topologically associated domain (TAD), as this locus was inside a large ~2Mb TAD spanning from the 48.5 Mb position to the telomere off 22q (**Supplementary Fig. 4.27**) (369). Together, these results suggest a model whereby the SV insertion allele disrupts the fetal brain enhancer-promoter interaction within a large telomeric TAD, thereby reducing the responsiveness of *NHIP* expression to neuronal differentiation and/or oxidative stress. Early pregnancy prenatal vitamin use is expected to counteract the effects of oxidative stress through provision of dietary methyl groups, thereby increasing DNA methylation at the *NHIP* locus in individuals homozygous for the 22q13.33 insertion.

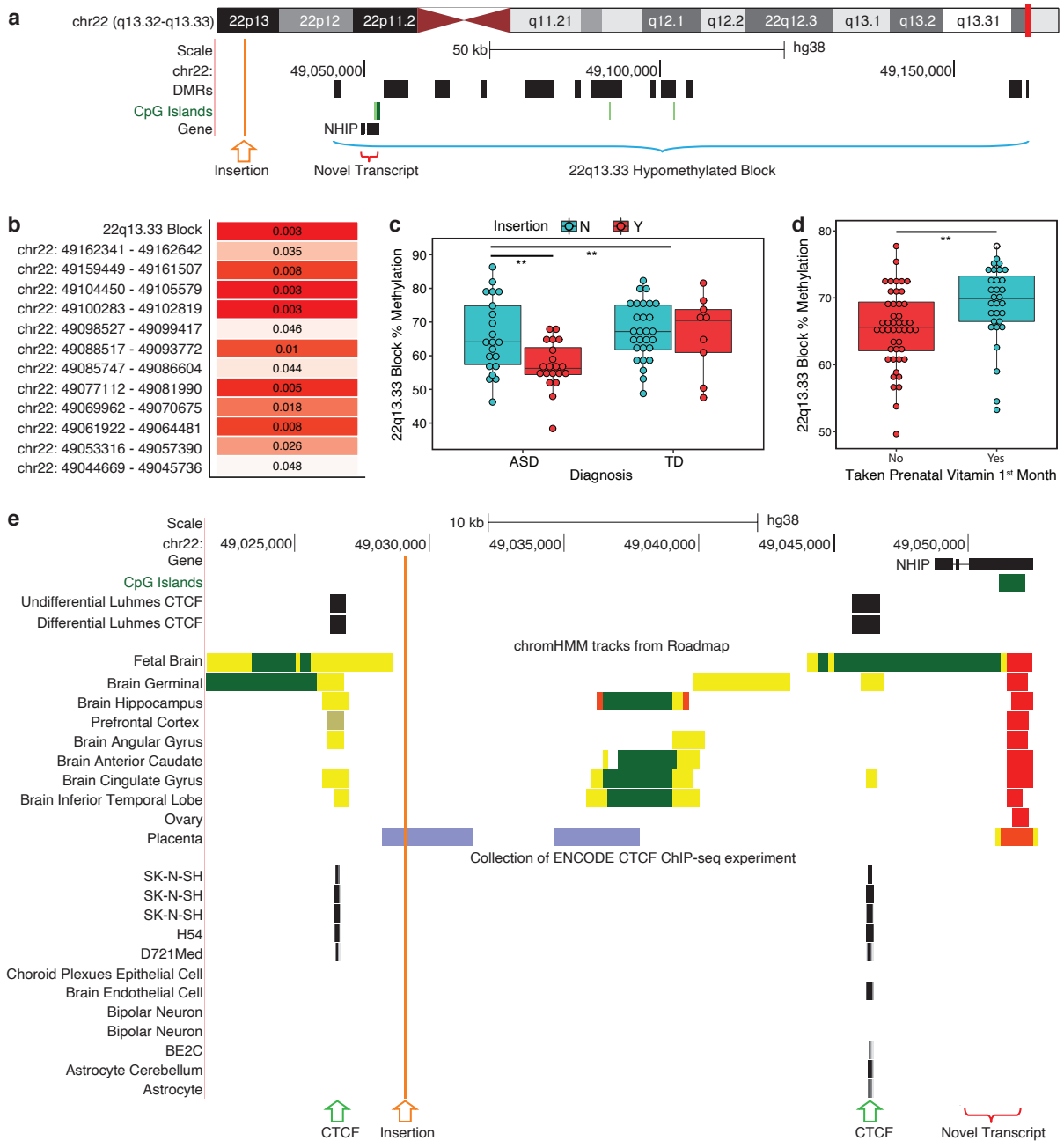


Figure 4.3. A common genetic structural variant is significantly associated with 22q13.33 DNA methylation and ASD.

(**a**) Insertion location (orange) relative to the 22q13.33 hypomethylated block (blue), and the novel transcript, *NHIP* (red) in the UCSC genome browser. The 22q13.33 comethylated block was 117,974 bp in length (blue). *NHIP* TSS was located 7,881 bp downstream from the start site

of the 22q13.33 hypomethylated block. The insertion (not in the reference genome) is 15,013 bp upstream from the start site of the 22q13.33 hypomethylated block.

(**b**) The association matrix shows ANOVA *p*-values for the comparison of the insertion genotype (homozygous for insertion versus not) with smoothed methylation levels within each of 12 DMRs located in 22q13.33 hypomethylated block from discovery group (ASD n = 41, TD n = 37).

(**c**) Association was tested between insertion genotype (Y, homozygous for insertion; N, not) and 22q13.33 comethylated block methylation levels (discovery group, ASD n = 41, TD n = 37). ASD showed significantly lower DNA methylation levels compared to TD placenta samples within the entire 22q13.33 comethylated block (*p*-value = 0.006). Samples homozygous for the insertion had a significant lower methylation than those not having insertion on both alleles (*p*-value = 0.008). When broken down by diagnosis, samples with insertion had significantly lower methylation specifically in ASD samples (*p*-value = 0.003), not TD samples (*p*-value = 0.63).

(**d**) Periconception prenatal vitamin use was a significant modifier of 22q13.33 block methylation in placenta (discovery group, ASD n = 41, TD n =37). Lower percent methylation at the 22q13.33 comethylated block was significantly associated with not taking prenatal vitamins during the first month of pregnancy (p-value = 0.007), which was in the same direction as ASD risk.

(**e**) UCSC genome browser map shows the insertion location (orange vertical line) relative to two adjacent CTCF sites (green arrows), and *NHIP*. Both undifferentiated and differentiated LUHMES cells have both CTCF sites, consistent with them being homozygous for the reference sequence. Additional brain tracks show the variability of the upstream CTCF site between human samples. ChromHMM tracks were derived from fetal brain, multiple brain regions,

ovary, and placenta. Red, active promoter; yellow, active enhancer; green; active transcriptional elongation; purple, bivalent poised chromatin.

NHIP expression is reduced in ASD brain and associated with the regulation of genes enriched for synaptic functions and ASD risk

We then tested the hypothesis that the 22q13.33 insertion was associated with *NHIP* expression in ASD versus TD postmortem brain (**Supplementary Table 4.13**). Similar to the MARBLES cohort of placenta samples, the 22q13.33 insertion showed a significantly higher frequency in ASD compared with TD in 58 cortical samples (**Supplementary Fig. 4.28**). RNA-seq was performed on a subset of 20 cortical samples representing all three SV insertion genotypes, matched for age and sex between ASD and TD. Brain samples homozygous for the 22q13.33 insertion (Y) exhibited lower *NHIP* levels compared to those with one or no insertion alleles (N) specifically in ASD samples, but not in TD samples (**Fig. 4.4a**).

We then performed a genome-wide analysis of transcript levels associated with variable *NHIP* levels in brain samples as a continuous trait. 851 *NHIP*-associated genes passed FDR significance, including 195 positively and 656 negatively associated (**Fig. 4.4b**, **Supplementary Table 4.14**). Downregulated genes included ASD candidate genes such as *CHD8* (370), and a gene previously implicated in ASD from placenta, *IRS2* (136) (**Supplementary Table 4.14**). Gene ontology (GO) enrichment analysis revealed 277 terms passing permutation significant gene ontology terms, including 202 terms from biological processes, 18 terms from molecular functions, and 57 terms from cellular components (**Fig. 4.4c**, **Supplementary Table 4.15**).

Regulation of nervous system, glial cell differentiation, synaptic membrane, neurogenesis, and response to oxidative stress were identified as GO terms enriched for transcripts negatively associated with *NHIP* levels (**Fig. 4.4c**, **Supplementary Table 4.15**). GO term functions related to the dendritic spine, synaptic plasticity, and regulation of synaptic transmission formed a functional module of genes negatively associated with *NHIP* levels (**Supplementary Fig. 4.29**). In contrast, transcripts positively associated with *NHIP* levels were enriched for distinct functions in epidermal development, G-protein coupled receptors, and negative regulation of secretion. To further examine the relevance of *NHIP* expression to ASD etiology, we overlapped brain *NHIP*-associated transcripts with SFARI ASD risk genes and observed a significant overlap of 85 genes (**Supplementary Fig. 4.30**, **Supplementary Table 4.16**). The 85 genes in common were significantly enriched for 49 GO terms, including nervous system development, synapse, and dendrite (**Supplementary Fig. 4.31**, **Supplementary Table 4.17**), demonstrating associations of *NHIP* levels with functionally relevant gene pathways in brain and ASD.

Overexpression of *NHIP* in HEK293T cells results in large-scale transcriptional changes to genes relevant to brain and ASD risk

In order to experimentally model the transcriptional impact of *NHIP* induction, RNA-seq and differential expression analyses were performed on HEK293T cells transiently transfected with *NHIP* or vector control. We identified 4,756 differentially expressed genes with genome-wide significance (FDR adjusted p-value < 0.05). *NHIP* transient overexpression increased 1,490 genes and decreased 3,266 genes (**Supplementary Fig. 4.32**, **Supplementary Table 4.18**). Genes decreased with transient *NHIP* expression included the downstream flanking gene *BRD1*, as well as *IRS2*, *CHD8*, and *DLL1* (**Supplementary Fig. 4.33**, **Supplementary Table 4.18**). *NHIP* overexpression and reduced *BRD1* in

overexpression cell lines were confirmed with RT-PCR (**Supplementary Fig. 4.34**). Genes differentially expressed with transient *NHIP* overexpression were enriched for GO terms associated with non-coding RNA processing, histone modification, placental development, cell cycle, and p53 binding (**Supplementary Fig. 4.35**, **Supplementary Table 4.19**), consistent with the proliferation phenotype (**Fig. 4.2f**). KEGG gene set enrichment analysis (371) showed enrichment for brain disorders, including Parkinson, Alzheimer, and Huntington diseases and metabolism, such as fatty acid metabolism and drug metabolism (**Supplementary Fig. 4.36**, **Supplementary Table 4.20**), further demonstrating the relevance of *NHIP* regulated genes to brain functions. Immune system process, oxidative phosphorylation, and respiratory chain complex were enriched GO terms in both transient (FDR adjusted p-value < 0.05) and stable (raw p-value < 0.05) *NHIP* overexpressing cells (**Supplementary Table 4.21**). When compared with KEGG gene set for enrichment, all significant GO terms in stable overexpressing cells were also included in those of transient expression. Fewer genome-wide significant differentially expressed genes were observed with stable (23) compared to transient (4,756) *NHIP* expression, however, consistent with the responsive inducible effects of *NHIP* (**Fig. 4.2c-4.2f**).

In a comparison of *in vivo* and *in vitro* RNA-seq analyses, a significant overlap of 284 genes was observed between those differentially expressed in response to experimental *NHIP* transient overexpression and those associated with *NHIP* transcript levels in human brain. Genes negatively associated with *NHIP* were enriched for functions in synapse, dendrite, cell-cell signaling, regulation of nervous system development, and upregulation of cell cycle process (**Supplementary Fig. 4.37-4.38**, **Supplementary Table 4.22-4.23**). Furthermore, genes differentially expressed with *NHIP* overexpression also showed a significant overlap of 263 genes with ASD risk genes from the SFARI database, enriched for functions in central nervous

system development, synaptic signaling, and response to oxygen levels (**Supplementary Fig. 4.39-4.40**, **Supplementary Table 4.24-4.25**). There were 45 genes in common among ASD risk, *NHIP* association in brain, and *NHIP* overexpression, including *BRD4*, *SETD5*, *CHD2*, *EP300*, and *FOXG1* (**Fig. 4.4c**, **Table 4.1**, **Supplementary Table 4.26**). Genes common to ASD risk, *NHIP* association in brain, and *NHIP* overexpression were enriched for chromatin organization, regulation of transcription by RNA polymerase II, regulation of cell differentiation, neurogenesis, rhythmic processes, and response to decreased oxygen (**Table 4.1**, **Supplementary Fig. 41**, **Supplementary Table 4.27**). Together, these results demonstrate that *NHIP* is a novel regulatory gene with functions relevant to known ASD risk factors.

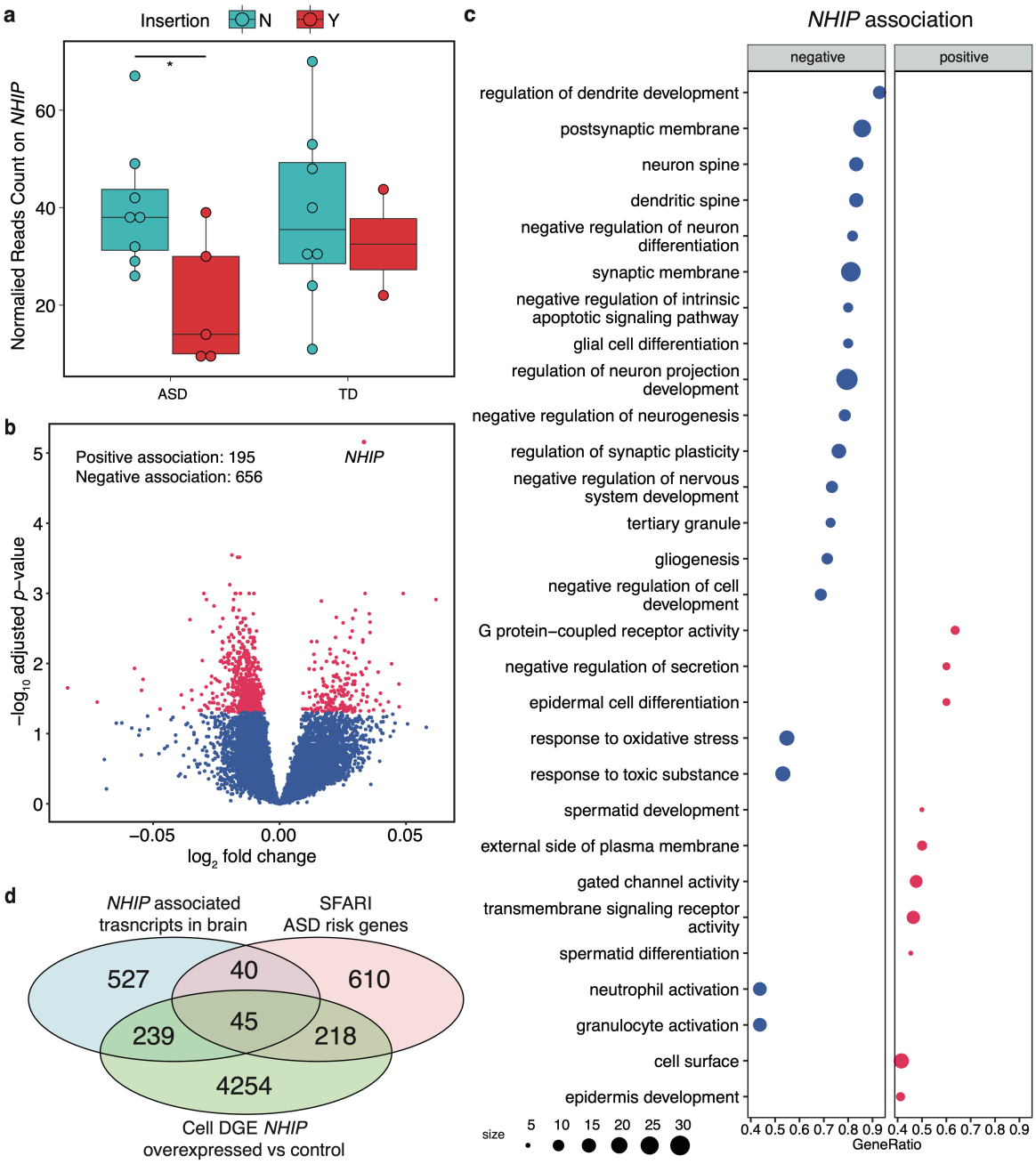


Figure 4.4. *NHIP* levels in brain are reduced in ASD and associated with expression of genes enriched for synaptic functions, response to oxidative stress, and ASD risk.

(**a**) Brain samples homozygous for the 22q13.33 insertion had significantly lower *NHIP* levels compared to those who were not (*p*-value = 0.035). The association *NHIP* levels and insertion

was observed specifically in ASD (*p*-value = 0.024), not in TD (*p*-value = 0.692) (two-tailed t-test, brain, ASD n = 13, TD n = 10).

(**b**) *NHIP*-associated differential expression analysis was performed from brain RNA-seq, identifying 851 genome-wide significant genes (FDR adjusted *q*-value < 0.05).

(**c**) Gene ontology (GO) enrichment analysis of the 851 *NHIP*-associated genes in brain with significant enrichment (FDR adjusted q-value < 0.05). Positively associated GO terms are shown in red and negatively associated GO terms are colored in blue.

(**d**) Venn diagram representing the overlapped 45 genes between *NHIP* association in brain, differential gene expression (DGE) in NHIP transiently expressing cell line, and SFARI ASD risk gene list.

Table 4.1. 45 genes in common to *NHIP* differential expression in brain and cells and ASD risk.

	GO Terms						
Genes	Chromatin organization	Regulation of transcription by RNA polymerase II	regulation of cell differentiation	Neurogenesis	Rhythmic process	Response to decreased oxygen levels	Count
EP300	X	X	X	X	X	X	6
CREBBP	X	X	X		X	X	5
SMAD4	X	X	X	X		X	5
RORA		X	X	X	X	X	5
HNRNPU	X	X	X		X		4
KMT2A	X	X	X		X		4
NIPBL	X	X	X	X			4
ADNP		X	X	X	X		4
NR1D1		X	X	X	X		4
ARID1B	X	X		X			3
PRKCA	X		X	X			3
RERE	X	X		X			3
CUX1		X	X	X			3
GRIN1		X	X	X			3
NR2F1		X	X	X			3
SPEN		X	X	X			3
ASH1L	X	X					2
BRD4	X	X					2
CHD2	X	X					2
HCFC1	X	X					2
HUWE1	X				X		2
KAT6A	X	X					2
KMT2E	X		X				2
CDK13		X	X				2
PCM1			X	X			2
JMJD1C	X						1
KANSL1	X						1
SETD5	X						1
CIC		X					1
FOXG1		X					1
MED13		X					1
POGZ		X					1
ZNF292		X					1
TNRC6B			X				1
CPEB4						X	1

Discussion

This study has taken the innovative approach of utilizing placental tissue from a high-risk prospective cohort to discover a novel ASD risk gene locus that integrates responsiveness to oxidative stress with inheritance of a common structural variant. The combination of WGBS and the distinctive DNA methylation landscape of the placenta characterized by partially methylated domains and higher gene body methylation over expressed genes (160,161,166) enabled the discovery of a novel gene associated with ASD that had been missed by more standard genetic and epigenetic array-based approaches. The 22q13.33 comethylated block identified in this study was previously identified by WGBS as a region of increased methylation variance (CoRSIV) (357,358) as well as a region of increased SV (114) in the human genome. We confirmed the hypothesis that CoRSIV and SV locations overlap more frequently than expected at random (**Supplementary Fig. 4.42**). Although this 22q13.33 region has not been previously associated with ASD risk, the neighboring distal long arm of 22q13.3 harbors multiple genes implicated in neuropsychiatric disorders, including ASD, intellectual disability, schizophrenia, and bipolar disease (92,372,373). *SHANK3* is 1.5 Mb telomeric from the 22q13.33 hypomethylated block identified in this study. In addition to rare *SHANK3* mutations in ASD, large structural variations including *SHANK3* are observed in rare ASD cases. *SHANK3* encodes a postsynaptic protein required for maturation of glutamatergic synapses (107). In addition, 22q13.33, 22q13.32 and 22q13.31 are disease-associated hotspot regions in ASD (117). While these highly polymorphic regions of the genome have the potential to contain regulatory genes such as *NHIP*, as well as primate-specific sequences relevant to brain development (374), they are often excluded from the design of array-based platforms because of their complexities. The

NHIP locus is sparsely covered by probes in the most current genetic and epigenetic array designs (**Supplementary Figure 4.13**), a likely explanation for why it was not identified by prior ASD studies. In contrast, sequencing-based approaches, such as the integrated WGS and WGBS approach employed here, are a promising alternative for disease association testing.

Placenta is an often misunderstood and overlooked tissue, despite its importance in regulating and thereby reflecting events critical to brain development *in utero*. Placenta regulates metabolism and provides steroid hormones as well as neurotransmitters critical for the developing brain (148,155). Placenta regulates oxygen supply, as it consumes 40-60% of the body's oxygen and hypoxia metabolic adaptation regulates trophoblast cell fate decisions (152,153). Oxygen tension can also modulate extravillous trophoblast proliferation, differentiation, and invasion (154), all important for successful implantation and placentation.

We have demonstrated that *NHIP* is a primate-specific, variably expressed gene responsive to hypoxia in human placenta and brain tissues. The variability in *NHIP* expression was influenced by both non-genetic and genetic factors. *NHIP* was induced with neuronal differentiation, but also with a hypoxia mimetic. Interestingly, the responsiveness of *NHIP* expression as well as oxidative stress was specific to differentiated neurons but not seen in undifferentiated LUHMES. Oxidative stress is a common convergent mechanism that occurs in normal neurodevelopment but can be excessive in cases of many environmental exposures associated with in ASD, including air pollution and pesticides. Prenatal vitamin use in the first month of pregnancy provides essential methyl donors to the one-carbon metabolism pathway that may

counteract excessive oxidative stress, a prediction consistent with the higher methylation over the 22q13.33 comethylated block in placentas from pregnancies with first month prenatal vitamin use. Common genetic variants were also associated with 22q13.33 methylation levels. While we identified 12 SNPs within the 22q13.33 comethylated block that were significantly associated with methylation, the strongest genetic factor was a 1.7 kb insertion with a high allele frequency in all ethnicities. Homozygosity for this 22q13.33 SV was a better predictor of ASD risk than GWAS-base PRS. 22q13.33 SV homozygosity was also strongly associated with hypomethylation of this locus and reduced expression of *NHIP* in ASD compared to TD placenta and brain samples.

Large insertions such as the 22q13.33 SV that occur outside of coding regions can still modify gene expression through alterations in promoter-enhancer loop size. The *NHIP* promoter shows differences in active chromatin marks between individuals and is associated with two CTCF binding sites that apparently anchor an intra-TAD loop between the promoter and a distal fetal brain enhancer. These results suggest a model by which the presence of at least one copy of the reference allele without the insertion would allow *NHIP* to be induced during neurodevelopment and hypoxia, thereby protect the developing brain through its regulation of downstream regulatory gene pathways. Homozygosity for the 22q13.33 SV allele is associated with lower *NHIP* expression and less protection, however, likely because the enhancer-promoter loop forms less efficiently because of the >15% increased size of the loop. For the minority of TD cases who were also homozygous for the 22q13.33 SV, the use of prenatal vitamins that reduce the consequences of oxidative stress might have been one source of

protection from risk, although other genetic and environmental factors not investigated may also be involved.

Methods

Sample Population and Diagnostic Classification

The Markers of Autism Risk in Babies – Learning Early Signs (MARBLES) study recruited mothers with at least one child with ASD who were pregnant or planning another pregnancy in Northern California, primarily through lists provided by the California Department of Development Services (71,79,136,375). The following criteria were required for MARBLES study's enrollment: the prospective child has at least one first or second degree relative diagnosed with ASD; the mother is at least 18 years old; the mother is pregnant or planning for a pregnancy; the mother speaks, reads and understands English proficiently enough in order to complete the protocol; the mother lives within 2.5 h drive distance of Davis/Sacramento region. Demographic, diet and medical information were collected by telephone interviews or questionnaires throughout the pregnancy.

The Early Autism Risk Longitudinal Investigation (EARLI) study recruited pregnant mothers who have another child diagnosed as ASD and has been described in detail previously (376). EARLI families were recruited from four sites (Drexel/Children's Hospital of Philadelphia, Johns Hopkins/Kennedy Krieger Institute, Kaiser Permanente Northern California, and University of California, Davis) across three US regions (Southeast Pennsylvania, Northeast Maryland, and Northern California). Enrollment criteria for EARLI are: having a biological child diagnosed as

ASD; communicating fluently in English or Spanish; being 18 years or older; living within 2 h drive distance from the study site; being less than 29 weeks of pregnancy.

In both MARBLES and EARLI studies, child diagnosis was clinically assessed by trained, professional examiners at 36 months using standardized instruments including the Autism Diagnostic Observation Schedule (ADOS) (245), Autism Diagnostic Interview – Revised (ADI-R) (330), and Mullen Scales of Early Learning (MSEL) (247). Based on a previously published algorithm, children were classified into three outcome groups: ASD, TD and Non-TD (122,248,249). Children with ASD outcome had scores over the ADOS cutoff and fit ASD DSM-5 criteria. Children with TD outcome had all MSEL scores less than 2 standard deviations (SD) and no more than one MSEL subscale 1.5 SD below the normative mean together, with scores on the ADOS at least three points lower than the ASD cutoff. Children with Non-TD outcome did not meet ASD or TD criteria but had elevated ADOS scores and low MSEL scores, defined as two or more MSEL subscales with more than 1.5 SD below averages or at least one MSEL subscale more than 2 SD below average.

Whole Genome Bisulfite Sequencing (WGBS) Library Preparation

The placental samples were frozen within 4 hours after birth. DNA was extracted from placenta tissue with the Gentra Puregene kit (Qiagen, Hilden, Germany) and quantified with the Qubit DNA Assay Kit (Thermo Fisher Scientific, Waltham, MA, USA). The discovery group included 92 samples (ASD n = 46, TD n = 46) from the MARBLES study. DNA was bisulfite converted with the EZ DNA Methylation Lightning kit (Zymo, Irvine, CA, USA). WGBS libraries were prepared from

bisulfite-converted DNA using the TruSeq DNA Methylation kit (Illumina, San Diego, CA, USA) with indexed PCR primers and a 14 cycle PCR programs. Libraries were sequenced at 2 per lane with 150 bp paired-end reads in Illumina HiSeq X (San Diego, CA, USA) by Novogene (Sacramento, CA, USA). The external replication group included 47 samples (ASD n = 16, TD n = 31) from the EARLI study, with details described previously (377). The internal replication group included 65 samples (ASD n = 21, Non-TD n = 13, TD n = 31) from the MARBLES study. DNA were sonicated to ~ 350 bp using Covaris E220 (Woburn, MA, USA). Sonicated and size selected DNA was bisulfite converted using the EZ DNA Methylation Lightning kit (Zymo, Irvine, CA, USA). WGBS libraries were prepared using Accel-NGS Methyl-Seq DNA library kit (Swift Biosciences, Ann Arbor, MI, USA) with indexed PCR primers and a 12 cycle PCR programs. Libraries were pooled and sequenced on 2 lanes with 150 bp paired-end reads of Illumina NovaSeq 6000 S4 (San Diego, CA, USA) by DNA Tech Core at University of California, Davis (Davis, CA, USA).

WGBS Alignment and Quality Control

Raw sequencing files were preprocessed, aligned to the human reference genome and converted to CpG methylation count matrices with the default parameters in CpG_Me (251,378,379). Reads were trimmed to remove adapters and methylation bias on both 5' and 3' end. After trimming, reads were aligned to human reference genome hg38, and filtered for PCR duplicates. Cytosine methylation reports were generated using all covered sites CpG methylation. Quality control was examined for each sample. Libraries with CHH methylation greater than 2% were excluded as incomplete bisulfite conversion. The CpG_Me workflow incorporates Trim Galore, Bismark, Bowtie2, SAMtools, and MultiQC (379–383).

Window methylation and principal component analysis (PCA)

DNA methylation at 20 kb windows sliding across the genome was extracted using getMeth function in bsseq (319,384). Percent methylation for each sample at each window was calculated using the average methylation value from the window. Correlations between DMRs were done using Pearson's correlation coefficient (r score). PCA analysis was performed using prcomp function in stats and visualized using ggbiplot (385). The ellipses for each group were illustrated as 95% confidence.

Methylation Array Analysis and Cell Type Estimation

The same 92 placenta DNA samples aliquots in discovery group (ASD n = 46, TD n = 46) were used for methylation array analysis. DNA was treated and cleaned with the EZ DNA methylation gold kit (Zymo, Irvine, CA, USA). Samples were assayed on the Infinium MethylationEPIC array (Illumina, San Diego, CA, USA) at John Hopkins University CIDR (Baltimore, MD, USA). Raw image files were analyzed using minfi package (201). Data were corrected for background and dye bias with the normal-exponential by out-of-band probe (noob) method (386). Cell type composition of placenta (trophoblast cells, stromal cells, Hofbauer cells, endothelial cells, and nucleated red blood cells) were estimated from methylation using a sorted placenta cell reference using PlaNET (387).

Detection of DMRs

DMRs were identified between ASD and TD in the discovery group with 100 permutation test, and adjusted for sex and cell types using DMRichR (378,388). DMRichR utilized dmrseq and

bsseq algorithms to extract methylation levels from CpG count matrix to identify DMRs (384,389). The DMR analysis approach used a smoothing and weighting algorithm that weights CpGs more heavily on high coverage and low variation within each group. CpGs in physical proximity with similar methylation values were grouped into candidate background regions to estimate region statistics. Permutation testing was done on the pooled null distribution to calculate empirical p-values to identify significant DMRs and then further corrected for genome-wide significance at an FDR of 0.05. Individual smoothed methylation levels and chr22q block methylation levels were obtained using bsseq (384). Genes were assigned to DMRs using the Genomic Regions Enrichment of Annotation Tool (GREAT) tool with the default association settings (5 kb upstream, 1 kb downstream and 1000 kb max extension) (390). The distances (kb) were calculated from DMRs to the transcription start sit (TSS) of the GREAT assigned genes. Gene Ontology (GO) enrichment analysis for DMRs, hypermethylation DMRs, and hypomethylation DMRs relative to background regions was done using GREAT (390). Significant terms were called with FDR corrected p-values less than 0.05.

DMR Enrichment Analysis

DMRs were examined for enrichment with chromatin marks compared to the background regions using LOLA R package with Fisher's exact test after FDR correction (202). Chromatin states were predicted by chromHMM using the Hidden Markov Model to separate human genome into 15 functional states in the Roadmap Epigenomics Project (196,264). Promoter related states included active TSS (TssA) (red), TSS flank (TssAFlnk) (orange red), bivalent TSS (TssBiv) (Indian Red), and bivalent TSS flank (BivFlnk) (Dark Salmon) states. Enhancer related

states included genic enhancer (EnhG) (Green Yellow), enhancer (Enh) (Yellow), and bivalent enhancer (EnhBiv) (Dark Khaki). CpG island, shore, shelf and open sea coordinates were obtained from the annotatr R package (391). Encyclopedia of DNA Elements (ENCODE) datasets were used to extract histone post-translational modifications (PTMs), including H3K4me1, H3K4me3, H4K9me3, H3K36me3, H3K27me3 and H3K27ac datasets (263,359). Enrichment for known transcription factor binding site motif sequences in DMRs was obtained using Hypergeometric Optimization of Motif EnRichment (HOMER) (392).

Whole Genome Sequencing (WGS) and Variant Calling

WGS was performed using cord blood genomics data on subset of the same individuals from in the discovery group (ASD n = 41, TD n = 37). Sequencing libraries were generated using NEBNest DNA library prep kit (NEB, Ipswich, MA, USA) with 150 bp paired-end reads in Illumina HiSeq X (San Diego, CA, USA) by Novogene (Sacramento, CA, USA) with at least 30x coverage per sample. Raw read files were mapped to human reference genome hg38 using Burrows-Wheeler Aligner (BWA) with the default setting (393). SAMtools was utilized to sort the bam files and Picard was used to merge bam files from the same sample identify duplicate reads (382,394). Single nucleotide polymorphisms (SNPs), small insertion, and deletions (InDels) were called using GATK and annotated variant using ANNOVAR (395,396). Copy number variations (CNVs), longer than 50 bp, were identified using control-FREEC and CREST (397,398). Structural variants (SVs) detection and genotyping, larger than 50 bp were performed using DELLY with the default settings (399).

PGS generation

A subset of individuals from in the discovery group were also genotyped using Illumina Multi-Ethnic genotyping array (ASD n = 31, TD n = 35). Stringent QC criteria was used on the raw genotypes in order to remove low quality SNPs and samples (400). Our criteria included removal of samples with call rates < 98%, sex discrepancy, and relatedness (pi-hat < 0.18) to non-familial samples with filtering for minor allele frequency (MAF) < 5% using PLINK software (401). After data cleaning, the imputation pipeline was performed using University of Michigan Imputation Server (402) using minimac4 software (403) to the 1000G Phase v5 reference panel (hg19) (404,405). Phasing was performed using Eagle software with measured and imputed SNPs (406).

PRS calculation was imputed from genetic data generated at the a range of $p_{discovery}$ thresholds based on GWAS results from the combined PGC-iPSYCH genome-wide meta-analysis (64). Using PLINK software (401), we removed correlated SNPs and applied from 2 to > 20,000 effect sizes and weights ($p_{discovery}$ threshold range from $1*10^{-8}$ to 1.0) to achieve a weighted summation of alleles on PRS scores generations that reflected ASD risk in the discovery group. Based on PRS model fit R^2, $p_{discover}$ used at 0.05 for the best-fit PRS. The association between 22q13.33 comethylated block % methylation or diagnosis with PRS was measured by analysis of variance (ANOVA).

Insertion Characterization and Sanger Sequencing

To validate the 22q13.33 insertion from Illumina WGS data, the expected genomic location of the insertion was queried in a published PacBio long read sequencing dataset (363). The insertion was identified located at CHM1_chr22-49029645-INS-1673 contig (363). The contig was in a fasta file with accession number GCA_003709635.1 with the correspondence table, it also named with GenBank ID QPKN01007947.1 in NCBI database (364). SAMtools was utilized to isolate the fasta sequence from the contig with 85,271 bp in length and extracted the insertion sequence with 1,673 bp in length (**Supplementary Table 4.11**). The QPKN01007947.1 contig mapped to chr22: 49,381,532 – 49,466,902 (reference genome: hg19) using blast search (407) and visualized the insertion using Miropeats (408).

In addition to characterizing the insertion using PacBio long read sequencing, primer sets were designed to span the insertion location for PCR-based genotyping (**Supplementary Table 4.11**). A 25 ul PCR reaction mixture contained 100 ng genomics DNA, 5 µl 5X LongAmp *Taq* reaction buffer (NEB, Ipswich, MA, USA), 1 µl LongAmp *Taq* DNA polymerase (NEB, Ipswich, MA, USA), 1 µl 10 mM dNTPs and 2 µl of 10 µM forward and reverse primer. The PCR amplifications were performed using following conditions: initial denaturation at 94 °C for 30 s; 30 cycles of denaturing at 94 °C for 30 s, 52 °C for 30 s and 65 °C for 2 min with a final extension at 65 °C for 10 min. PCR products were subjected to Topoisomerase (TOPO) PCR Cloning Kit (Thermo Fisher Scientific, Waltham, MA, USA) followed by a 1.5% agarose gel electrophoresis with purification and Sanger sequencing by University of California, Davis, DNA Sequencing Facility (Davis, CA, USA) and chromatograms were analyzed using SnapGene (Genewiz, South Plainfield, NJ, USA).

PCR products genotype and size were characterized using Bioanalyzer 2100 (Agilent, Santa Clara, CA, USA). The sequence of the insertion was analyzed for repetitive elements using CENSOR and RepeatMasker (409,410).

Cell Culture, Cell-based Assays and Transfection

LUHMES cells (ATCC, Manassas, VA, USA, CRL-2927) were seeded on fibronectin coated plates (Thermo Fisher Scientific, Waltham, MA, USA, CWP001, 354402). Undifferentiated cells were maintained in proliferation medium: Advanced DMEM/F12 (Invitrogen, Carlsbad, CA, USA), supplemented with N2 supplement (Invitrogen, Carlsbad, CA, USA), Penicillin-streptomycin-glutamine (Thermo Fisher Scientific, Waltham, MA, USA), and 40 ng/ml recombinant bFGF (Invitrogen, Carlsbad, CA, USA). To generate differentiated LUHMES, cells were switched to differentiation media for five days. Differentiation media: Advanced DMEM/F12, supplemented with N2 supplement, Penicillin-streptomycin-glutamine, 1 mM dbcAMP (MilliporeSigma, Burlington, MA, USA), 1 µg/ml tetracycline (Neta Scientific, Hainesport, NJ, USA), and 2 ng/ml recombinant human GDNF (Thermo Fisher Scientific, Waltham, MA, USA). For cell viability and hydrogen peroxide production experiments, differentiated cells were growth in 96-well plates for six days prior to treatment with CellTiter Blue or ROS-Glo visualization reagent (Promega, Madison, WI, USA). Undifferentiated cells, were plated in 96-well plates at same densities as differentiated neurons and treated identically for cell viability and hydrogen peroxide measurements. For RNA quantification, cells were maintained in 6-well plates. Challenges with hydrogen peroxide (MilliporeSigma, Burlington, MA, USA), cobalt chloride (Thermo Fisher

Scientific, Waltham, MA, USA) or mock treatment were carried out after five days of differentiation and cells were treated for 24 hours before analysis.

A PiggyBac-compatible expression plasmid, pb-NHIP-eGFP was synthesized by VectorBuilder (Chicago, IL, USA) with EF-1α as promoter for *NHIP* and CMV as promoter for eGFP fused with puromycin resistant gene (**Supplementary Table 4.9**, **Supplementary Fig. 4.16**). A control plasmid was cut using XbaI and AbaI restriction endonucleases based on the pb-NHIP-eGFP, named pb-NEG-eGFP to remove *NHIP* and maintained the rest of plasmid structure (**Supplementary Table 4.9**, **Supplementary Fig. 4.16**). The backbone piggyBac plasmid was obtained from Sanger Institute Archives and described previous as hyperactive *PB* transposase (pCMV-hyPBase) (**Supplementary Table 4.9**) (411). Plasmid for NHIP plasmid, pb-NHIP-peptide-eGFP was synthesized by VectorBuilder with EF-1α as promoter for the NHIP peptide, removed the stop codon and fused the end of the NHIP peptide with eGFP, together with CMV as promoter for mCherry fused with puromycin resistant gene (**Supplementary Table 4.9**, **Fig. 4.2g**). All constructs were sequenced by Sanger sequencing by University of California, Davis, DNA Sequencing Facility (Davis, CA, USA) and analyzed using SnapGene (Genewiz, South Plainfield, NJ, USA) to confirm the expected sequence.

HEK293T cells (ATCC, Manassas, VA, USA, CRL-11268) were grown in DMEM/F12, GlutaMAX medium (Thermo Fisher Scientific, Waltham, MA, USA) supplemented with MEM non-essential amino acids (Thermo Fisher Scientific, Waltham, MA, USA) and 10% fetal bovine serum (Invitrogen, Carlsbad, CA, USA). Low passage HEK293T cells were transfected with plasmids

using Lipofectamine 3000 and Opti-MEM (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's instructions. Transfections were performed with or without piggyBac backbone plasmids for both transient and stable cell lines for each condition. Transfection medium was replaced 24 h post-transfection with complete growth media with puromycin at 3 µg/ml for 7 days.

All cells were maintained at 37 °C containing 95% O_2 and 5% CO_2. Images were taken using EVOS under magnification labeled in the images. Cell numbers were measured using disposable countess chamber slide on Countess II FL automated cell counter (Thermo Fisher Scientific, Waltham, MA, USA) under the default steps with mixing 10 µl of samples with 10 µl of trypan blue. CellTiter Blue reagent was used for measured cell viability using luminescence based on manufacturer instruction (Promega, Madison, WI, USA). H_2O_2 production represented relative reactive oxygen species (ROS) level was measured with the ROS-Glo H_2O_2 assay system using 50nM with the default setting with level measured by luminometer (Promega, Madison, WI, USA).

HEK293T whole cell lysates were prepared by resuspension in 1x RIPA buffer and sonication using a Diagenode Bioruptor 300 (Diagenode, Denville, NJ, USA) followed by centrifugation at 21,130 x g at 4°C to remove insoluble material and then resolved on a 4-15% SDS-PAGE gel (Biorad, Hercules, CA, USA). The SDS-PAGE gel was rinsed in three changes of water to remove SDS and stained with Imperial protein stain (Thermo Fisher Scientific, Waltham, MA, USA) to visualize proteins. Stained bands between 25 kd and 37 kd were carefully excised from the gel,

washed in three changes of 50 mM ammonium bicarbonate followed by three washes with acetonitrile then swollen in 10 mM DTT in acetonitrile and incubated at 56 °C for 30 minutes do reduce disulfide bonds. The gel pieces were next shrunk by incubation in acetonitrile then incubated in 55 mM iodoacetamide (IAA) in 50 mM ammonium bicarbonate prior to washing with 50mM ammonium bicarbonate, shrunk with acetonitrile and dried in a speed vac. Gel pieces were suspended in 50mM ammonium bicarbonate with 0.01 % Protease Max (Promega, Madison, WI, USA) and treated with trypsin (Promega) for four hours at 50°C. The NHIP/GFP fusion protein was detected from the resulting peptides by (LC/MS-MS). MS was performed at University of California, Davis Proteomics Core Facility. Significant level was called based on FDR adjusted p-value < 0.05. The protein probabilities were calculated use the protein prophet algorithm. The peptide probabilities were calculated use naïve Bayesian classifier.

NHIP peptide immunofluorescence staining utilized a custom polyclonal antibody that was produced in Rabbit by GenScript Inc (Piscataway, NJ, USA) to a truncated NHIP peptide MVRGEATARTEEAMC and affinity purified. Flash frozen human cortical tissues were fixed in 4% formaldehyde in 1x PBS for 72 hours then dehydrated by immersion in 70% ethanol for seven days and embedded in paraffin. 5µm sections were cut from embedded brain tissue and mounted on glass slides then baked for 4 hours at 56 °C. Tissues on slides were washed four changes of xylene to remove paraffin. Next, slides were washed in two changes of 100% ethanol which was removed by heating to 50 °C on a heat block. The slides were then treated with 1x DAKO antigen retrieval solution (Agilent, Santa Clara, CA, USA) at 95 °C for one hour in a water bath. Slides were washed five times in 1x PBS with agitation. To reduce endogenous

autofluorescence slides were immersed in 1x PBS and exposed to LED light for 24 hours. Slides were next incubated with 1x PBS/0.5% Tween 20/3% BSA 1 hour at 37°C to block background signals then washed three times in 1x PBS/0.5% Tween 20. Anti-NHIP peptide and control pre-immune antibodies were diluted 1/200 in 1x PBS/0.5% Tween 20/3% BSA and incubated on slides at 37°C overnight in a humid chamber before three washes in 1x PBS/ 0.5% Tween. Goat anti-Rabbit Alexa 594 (Thermo Fisher Scientific, Waltham, MA, USA, Catalog #A32740) was diluted in 1x PBS/ 0.5% Tween20/ 3% BSA with 5 µg/ml DAPI and added to slides for two hours at 37°C in a humid chamber. Slides were washed five times in 1x PBS/ 0.5% Tween 20 with shaking before mounting in 5 µg/ml DAPI in 50% glycerol and application of glass coverslips.

RNA Extraction, cDNA Synthesis and RT-PCR

Total RNA was isolated from HEK293T cells transiently or stably transfected with pb-NHIP-eGFP or negative control pb-NEG-eGFP using AllPrep DNA/RNA/Protein mini kit (Qiagen, Hilden, Germany). Human tissue total RNA samples were obtained commercially, including placenta (Life Technology, Carlsbad, CA, USA), testes (TaKaRa Bio, Kusatsu, Shiga, Japan), and fetal brain (Cell Applications, San Diego, CA, USA). RNA was extracted from frozen placenta samples in the Discovery group samples using TRIzol Reagent (Invitrogen, Carlsbad, CA, USA). cDNA was synthesized using the High-Capacity cDNA Reverse Transcription Kit (Thermo Fisher Scientific, Waltham, MA, USA) based on the manufacturer's protocol. TaqMan Gene Expression Assays for *LOC105373085* (renamed as *NHIP*) (assay ID: Hs01034248_s1), *BRD1* (Hs00205849_m1), *FAM19A5* (Hs00395354_m1) and *GAPDH* (assay ID: Hs02786624_g1) were used (Thermo Fisher Scientific, Waltham, MA, USA). The expression of 3 genes of interest and 1 reference genes

were examined by real-time TaqMan PCR assay (Thermo Fisher Scientific, Waltham, MA, USA). Expression levels were determined by the probes with optimized primer and probe concentrations. Quantification was accomplished with RT-PCR machine using TaqMan Fast Advanced Master Mix with the default parameters by the manufacturer (Thermo Fisher Scientific, Waltham, MA, USA). Reactions were performed with three biological replicates. Fold changes of transcript levels were measured using the Fluidigm Real-Time PCR Analysis software calculated fold change of gene expression as the delta delta CT normalized to *GAPDH* (Fluidigm, San Fracisco, CA, USA).

Brain Sample Acquisition

Human brain samples were obtained from the NICHD Brain and Tissue Bank for Developmental Disorders at the University of Maryland (Baltimore, MD, USA) (**Supplementary Table 4.13**). RNA from the frozen human brain was purified using AllPrep DNA/RNA/Protein mini kit (Qiagen, Hilden, Germany).

RNA-seq Library Preparation and Sequencing

RNA from cells and brain was prepared for RNA-seq library using Kapa RNA HyperPrep kits (Roche, Basel, Switzerland) together with the QIAseq FastSelect Human ribodepletion kit (Qiagen, Hilden, Germany). Libraries were assessed for quality and quantify on Agilent Bioanalyzer 2100 and pooled for multiplex sequencing with at least 25 million reads with 150 bp paired-end on Illumina NovaSeq 6000 S4 (San Diego, CA, USA) by DNA Tech Core at University of California, Davis (Davis, CA, USA).

RNA-seq Data Processing and Differential Gene Expression (DGE)

Raw fastq files were processed and aligned using STAR (412). After quality control steps by FASTQC, the count matrixes were generated by featureCounts (413,414). Count matrixes were filtered for at least one count in any sample. Size factors estimation and normalization were performed by DESeq2 (415). DGE for transient cell was generated compared between overexpressed *NHIP* and negative control cells using DESeq2 (FDR corrected *p*-value < 0.05) (415). DGE for stable cell compared overexpressed *NHIP* and negative control cells using DESeq2 (raw p-value < 0.05) (415). DGE for brain was analyzed by using normalized read count for *NHIP* levels as continuous trait using DESeq2 (FDR corrected *p*-value < 0.05) (415). Gene overlaps between different experiments were tested for significance using Fisher's exact test in the GeneOverlap R package (347).

Gene Ontology terms for DGE were identified using clusterProfiler on Gene Set Enrichment Analysis using gseGO function with 1,000 permutation tests (416). Normalized enrichment scores (NES) were calculated for enrichment after correcting for FDR multiple testing. The dotplots illustrate significant GO terms based on GeneRatio, calculated from the number of overlapped genes divided by the total number of genes in the gene set (416). GO terms to be included in the plots were selected based of GeneRatio ranking. The enrichment map was plotted using emapplot function on clustering mutually overlapping gene sets to form functional modules (416). The ridgeplot was plotted using ridgeplot R function to visualize

expression distributions of core enriched genes (416). The cnetplot depicted the linkages of genes and biological concepts as networks (416).

Data and Code Availability

Datasets supporting the conclusions are available in the Gene Expression Omnibus repository (GEO). Code and scripts for this study are available on GitHub (417).

Chapter 5 – Discussion

Together, the research in this dissertation integrates genome-wide sequencing with targeted molecular approaches, resulting in multiple novel insights into the relationship among maternal and fetal epigenetics, genetics, and environment during pregnancies at high risk for the development of ASD (**Fig. 5**). First, we demonstrated the utility of WGBS approaches for identifying ASD associated DMRs in human placenta that reflect the developing brain and interactions with prenatal vitamin use (**Chapter 2**). Next, we demonstrated transcriptomic patterns associated with maternal one carbon nutrients (**Chapter 3**). Lastly, we performed a large, multi-cohort and multi-tiered analysis of placenta and brain to further understand genetic and environmental risk for ASD (**Chapter 4**). This multi-omics analysis utilized the largest WGBS data set in ASD with a unique sample type, placenta, in two prospective longitudinal studies. These studies have led to the discovery of a novel ASD risk regulatory gene locus at 22q13.33 with methylation level affected by both genetics and environment. We have also discovered a novel transcript at 22q13.33 and further characterized its functional relevance using transcriptome analysis in both human brain and *in vitro* cell models.

A particularly novel aspect of our findings was that placenta ASD DMRs were enriched for fetal brain enhancers and genes with neuronal functions. Previous studies had shown that placenta and brain shared similar Wnt and cadherin signaling pathways (136,210). Prior work from our laboratory identified a putative fetal brain enhancer near *DLL1* (171). *DLL1* encodes a ligand of Notch activated by Wnt signaling, and was replicated in our list of genes from the SFARI Gene list and also changed with *NHIP* overexpression in cell culture (171,191,216). Genes identified in

SFARI ASD database were also overrepresented among placenta ASD DMR genes, demonstrating an overlap between genetic and epigenetic risk for ASD. Of these, *PRKCA* is particularly interesting as it has also been identified as differentially expressed in both brain and *NHIP* overexpressing cells. *PRKCA* has been previously associated with ASD in post-mortem brain and umbilical cord blood by studies of DNA methylation (140,375,418), in brain a as differentially expressed gene (419), in large GWAS studies as a high risk ASD gene with missense or synonymous mutations (65,176), as well as in prefrontal neuron with differential H3K4me3 peaks for ASD (420). *PRKCA* encodes protein kinase C alpha, part of the MAP kinase signaling that regulates cell proliferation previously implicated in ASD (421). Furthermore, placental ASD DMRs were enriched for bivalent enhancer and repressed polycomb regions, which are known regions of hypomethylation in all tissues, but with relatively higher methylation in placenta (161). Together, these results suggest that the ASD associated differential methylation identified in placenta is more likely to be a remaining signal of past dysregulation in early neuronal development rather than gene expression differences at birth. In further support of this, genes vital for early prenatal development were identified as regulatory domains in placenta ASD DMRs.

Placenta ASD DMR associated genes were enriched for functions in a distinct but related epigenetic mark of histone acetylation. Histone acetylation dysregulation is implicated in a variety of neurodegenerative diseases, including Alzheimer's disease, ASD, Huntington's disease, and amyotrophic lateral sclerosis from studies of cultured neural progenitors and *in vivo* models (422–424). A histone acetylome-wide association study discovered thousands of

differentially acetylated genes with functions in synaptic transmission, DNA replication, and histone deacetylase activity in ASD cortex (141). Histone deacetylase (*HDAC*) inhibition can rescue *SHANK3* deficient mouse models (425). The closest annotated protein coding gene to the 22q13.33 comethylated block is *BRD1*, which was also regulated by the novel transcript in the overexpressed cell model. *BRD1* encodes a bromodomain protein that can stimulate acetylation of histones H3and H4 as an epigenetic reader (426,427). *BRD1* is predicted to have functions in transcriptional regulation and brain development, and genetic susceptibility to schizophrenia from multiple human GWA studies with high replicability (355,428–430). *HDAC4*, histone deacetylase 4, identified as differentially expressed with *NHIP* overexpression is also an ASD risk gene in the SFARI dataset. *HDAC4* is known to regulate the transcription of synaptic genes, and rare genetic aberrations in *HDAC4* may cause epigenetic dysregulation in autistic brain (431–433).

The 22q13.33 comethylated block identified as hypomethylated in ASD with genome-wide significance in our study was replicated across studies and different sequencing platforms. In contrast, due to lack of annotated protein coding gene in this 118 kb block, very few probes were designed at this region in Infinium methylation EPIC array (434). This comethylated block was also identified as CoRSIV with interindividual epigenetic variation modulates risk of disease (357). Although this 22q13.33 region has not been previously associated with ASD risk, the neighboring distal long arm of 22q13.3 harbors multiple genes implicated in neuropsychiatric disorders, including ASD, intellectual disability, schizophrenia, and bipolar disease (92,372,373) . *SHANK3*, a leading ASD candidate gene with 1-2% of ASD individuals reported to have *SHANK3*

mutations and encoding a protein essential for postsynaptic function in glutamate signaling (92), is 1.5 Mb downstream from our novel 22q13.33 hypomethylated block. In addition to rare *SHANK3* mutations in ASD, large structural variations over this locus are observed in ASD. *SHANK3* encodes a postsynaptic protein required for maturation of glutamatergic synapses (107). In addition to 22q13.33, 22q13.31 and 22q13.32 are both identified as disease-associated hotspot regions in ASD (117).

The *NHIP* histone H3K4me3 peaks in ovary, placenta, and brain appeared to be polymorphic, suggesting variation in expression that was consistent with RNA-seq data from brain (359). *LOC105373085* (renamed as *NHIP*) was originally described by Ota et al. analysis of 21,243 full-length human cDNA sequences as noncoding, due the bioinformatic evidence of splicing but no coding exons predicted (360). Therefore, ours is the first study to functionally characterize this novel transcript. Studies on noncoding RNAs differentially expressed in ASD suggest functions in cell migration, and neuronal connectivity in brain and peripheral blood (435,436). Based on query location in GTEx (361), *NHIP* was dynamically expressed in various brain regions. This novel transcript was specifically conserved in primates but not other mammals, including mouse. Previous studies on noncoding RNAs have shown they lack of primary sequence conservation, showing less conservation than protein coding genes, with a subset of them exhibiting tissue-specific and stress-responsive expression patterns (437,438). The majority of mouse noncoding RNAs are expressed in neuronal tissue and distinct cortical regions with epigenetic regulatory functions that can direct chromatin states, thereby impacting gene expression (438–440). *NHIP* had higher expression in brain compared with placenta, consistent

with our hypothesis that DMRs in placenta reflect remnants of past dysregulation related to neuronal development, a finding also seen previously in the analysis of DNA methylation differences in ASD umbilical cord blood (375). Hypomethylated gene body methylation in placenta predicts reduced gene expression (187) which suggests that reduced *NHIP* levels in ASD pregnancies are insufficient to protect against ASD *in utero*. Recently, an increasing number of experiments have indicated that noncoding genes can be translated into peptides (441–443). Peptides derived from noncoding genes facilitate embryonic development, response to stress, and modulate tumor development (444–446). Future experiments are being designed to test the hypothesis that part of the function of *NHIP* may be related to a peptide.

In the analysis of potential genetic effects on differential methylation and expression at the *NHIP* gene locus, a novel 1.7 kb insertion not present in the reference genome was significantly associated with 22q13.33 methylation levels in placenta and *NHIP* expression levels in brain. While some SNPs near *NHIP* were also associated with 22q13.33 methylation, this region was poorly covered with probes on the Infinium Multi-Ethnic Genotyping Array (447). The 22q13.33 insertion location was previously annotated as a SV hotspot and INS_22_115103 in gnomAD, with an allele frequency of 0.7, despite being absent from the reference genome (114,365,366). We also determined that CoRSIV regions of high polymorphic DNA methylation significantly overlapped with hotspots of polymorphic structural variants (114,357). Genome-wide assessments of ASD have mapped the relevance to the larger 22q13 region, particularly at the *SHANK3* gene encoding a postsynaptic density protein (107). Other studies have shown associations with common SVs and ASD risk located in putative noncoding regulatory regions

(118). Our study demonstrating the discovery of a novel common genetic locus for ASD associated with expression of a novel gene as well as DNA methylation reveals the importance of performing alternative approaches to understanding common genetics of ASD beyond GWAS.

Previous studies have shown common genetic variants, each having small effects, dominate most ASD risk compared with rare gene variants (267). Researchers have shown that ASD is a multi-dimensional disorder with both genetic and environmental risk and protective factors (71). In our study, prenatal vitamin use during the first month pregnancy showed significant protective effects in individuals with the genetic risk factor of homozygosity for the 22q13.33 insertion. This finding supports our hypothesis that DNA methylation in placenta is an intermediate interface between both an inherited structural variant and an environmental protective factor. Maternal prenatal vitamins, containing high concentrations of folate and other additional B-vitamins important in one-carbon metabolism and methylation pathways, have been shown to protect offspring by up to 70% from neural tube defects (181,182,184). Periconceptionally, prenatal vitamin intake has been associated with a 40% reduction in ASD risk in multiple studies, including the MARBLES high risk cohort (119,120,122,448). A study on schizophrenia risk demonstrated significant interactions between genetic risk using polygenic risk score and maternal perinatal environmental factors reflecting changes on placental gene expression (169). Previous analyses of DNA methylation patterns in MARBLES identified associations between self-reported use of lawn and garden pesticides and large-scale changes in DNA methylation patterns (133). In addition, genes differentially methylated with prenatal

vitamin use also significantly overlapped with those differentially methylated in ASD placenta in MARBLES, with enrichment for functions in neuron fate commitment and central nervous system development, especially demonstrated at *IRS2* (136). *IRS2*, a gene that encodes for insulin receptor substrate 2 that mediates the effects of insulin and insulin-like growth factor 1 (*IGF1*), has also been discovered as gene associated with *NHIP* expression in both brain and an overexpression cell line model together with other studies relating its functions with ASD (136,234,236,241).

Overall, the genes associated with differential expression of *NHIP* in postmortem brain were enriched for synaptic functions, nervous system, and response to oxidative stress, and significantly overlapping with ASD risk genes in the SFARI database. There is a deep literature examining neuronal dendritic spines, synapses and their correlation with mTOR pathways in ASD (449,450). *SHANK3* mutations in 22q13 have also been shown to lead to dendritic spine morphology changes (451). Environmental and genetic factors increase vulnerability to oxidative stress together with the interconnected transmethylation cycle and transsulfuration pathway in ASD that may contribute to the development of the disease (452–454). Oxidative stress can lead to impaired methylation, which can inhibit the folate and vitamin B12 - dependent methionine synthase in one-carbon metabolism pathway, which is consistent with DNA hypomethylation in peripheral blood in ASD (453,455,456). Neuronal responses of apoptosis and regeneration lead to accumulated oxidative insult and attempts at cell cycle re-entry and could be protected by nutrients such as vitamin D and folate in Alzheimer and

Parkinson disease (457). Studies have also shown that increased vulnerability to oxidative stress and decreased capacity for DNA methylation contributes to ASD risk (180,458).

Interestingly, genes differentially expressed in response to *NHIP* overexpression in HEK293T cells were enriched for functions in histone modification, developmental gene regulation, and cell cycle. The link back to epigenetic pathways was encouraging, as many ASD-risk genes are involved in chromatin remodeling and synaptic plasticity (60). Histone methylation as an epigenetic mark is also at the interface of genetic and environmental factors, including folate, the major dietary methyl donor for DNA and histone methylation reactions (120,459). During the methionine cycle, methionine converts to SAM which serves as the universal methyl donor for downstream methylation processes by DNMT (49). In our genome-wide transcriptome analysis on maternal peripheral blood, betaine and DMG concentrations in maternal plasma were associated with genes in histone modification, adaptive immune, and RNA processing functions. These results support that pregnancy is a sensitive time when nutrients play a vital role on genetics and epigenetics changes for ASD which reflect the immune system and gene regulation changes.

Previous studies have shown altered proliferation in neural cells derived from idiopathic individuals with autism (460). Cell cycle networks have shown links with gene expression dysregulation, mutation, and abnormal brain development in ASD (419). Cell cycle gene networks detected in peripheral blood were highly preserved in brain during prenatal stages of development with the ASD risk gene *CHD8* as central regulator of downstream genes with

functions in neurogenesis and cell adhesion processes (419). p53 binding was also identified as a *NHIP* regulated gene ontology function. p53 is one of the most important links between proliferation and cell death, by inhibiting cell cycle and inducing apoptosis in response to environmental signals (461). Interestingly, DNA hypomethylation, such as that observed in the 22q13.33 comethylated block in placenta has long been implicated in the genome instability observed in tumors (12). Furthermore, ASD and cancer share similar dysregulated pathways including mTOR, chromatin remodeling, histone modification, cell cycle, and synaptic pathways (67).

CHD8, another high-confidence ASD risk gene we identified as responsive to *NHIP* overexpression, functions as chromodomain helicase DNA-binding protein 8. *CHD8* is one of the major risk genes for ASD and is identified in 0.2% patients. CHD8 contributes to early brain development by acting as a negative regulator of the WNT signaling pathway (370). Studies have shown that *CHD8* can regulate cell cycle of neuronal progenitors in ASD brain (419,462). RNA-mediated knockdown of *CHD8* in human neural progenitor cells showed that *CHD8* regulates noncoding RNAs that function as central hub regulators of neuronal development and ASD risk (463).

The lysine methyltransferases *KMT2E* is also one of the 5 ASD GWAS hits, and mutated in rare ASD cases (64). In our study, *KMT2E* was identified as one of the 45 genes common to *NHIP* associated expression in brain and the experimental model as well as ASD genetic risk. *KMT2E* encodes a chromatin regulator that is recruited to methylated histone at promoters of actively

expressed genes, specifically H3K4me3. It was also identified as a tumor-suppressor gene, highly pleiotropic, with roles in response to DNA damage (464). *KMT2E* has been identified a key risk factor for ASD in chromatin regulation at fetal development with high expression in brain especially during fetal development (465–467). *KMT2E* tightly regulates cell cycle progression and reportedly if overexpressed in HEK293T cells, *KMT2E* arrests cell proliferation, which replicated in both tumor and normal diploid cells (468,469). *KMT2E* is predicted to play a role in H3K4me3 peak alterations affecting neuronal chromatin of some individuals with autism (470,471).

In a complementary approach, the 45 genes in common among ASD risk and associated with *NHIP* levels in brain and the experimental cell model were enriched for histone modification, neuronal function, cell cycle, and oxygen regulation functions. Specifically, *CREBBP*, *EP300*, *BRD4*, *SMAD4*, and *RORA* existence in multiple pathways and negatively are associated with the novel transcript in both brain and the cell line model, suggesting they may be involved in multiple pathways dysregulated in ASD. *CREBBP* and *EP300* are HATs that share conserved HAT domain and bromodomain, play critical roles in embryonic development, growth control, and chromatin remodeling with histone acetyltransferase activity involved in maintenance of super-enhancers and *BRD4* recruitment (36,472) and both identified as high confidence ASD risk genes in SFARI dataset (191). Aberrant histone acetylation such as *CBP* and *EP300* histone acetyltransferase can lead to Rubinstein-Taybi syndrome (473). *BRD4*, bromodomain-containing protein recruit acetylated sites of the genome and mediate transcriptional initiation and elongation (474). Together, *CREBBP*, *EP300*, *BRD4* have functions in histone acetylation and

interfere with oncogene-driven transcriptional programs in cancer and also regulate neuronal differentiation in neurodegenerative disorders (36,475–477). *SMAD4* acts as a tumor suppressor and inhibits cell proliferation in response to growth factors (478) and has been a missense mutation hotspot for autism (65). *RORA*, regulated gene in circadian rhythm in brain development and hormones, is identified as a ASD candidate gene through global methylation profile in idiopathic autism (479–481) . Previous studies have also shown that circadian clock is dysregulated in neurodevelopment disorders (251,482). Together, these results demonstrate that multiple genes and terms related with the novel transcript in brain and HEK293T cell are related to histone modification and cell cycle in neurodevelopment and cancer, suggesting the novel transcript identified by 22q13.33 block methylation changes in ASD can have wide-spread genomic effects on histone modifications.

One potential explanation for polymorphic 22q13.33 hypomethylated patterns in human placenta may be related to the “multiple hit” hypothesized etiology of most ASD cases. The very high allele frequency of the 22q13.33 insertion suggests that more than just homozygosity for this locus is required for the etiology of ASD. Our finding that the novel *NHIP* gene is responsive to oxidative stress and hypoxia suggests that the second hit could be one or more prenatal insults that result in oxidative stress. Prenatal vitamins that function as a protective factor during preimplantation embryos may therefore protect individuals with genetic risk and exposure to excessive oxidative stress. The novel transcript, *NHIP*, regulates genes involved in oxygen transfer, histone modification, and cell cycle pathways that protect brain and placenta

from gene pathway dysregulated leading to ASD. The higher expression of *NHIP* in TD compared to ASD brain and placenta samples also suggests the protective nature of this novel gene locus.

In summary, this dissertation focused on perinatal epigenetic signatures of ASD using multi-omics genome-wide approach on epigenetics, genetics, and environment. In **Chapter 2,** we identified two genes differential methylated in ASD in placenta and studied the genetic and environmental modifiers of genotype and prenatal vitamin use. These studies led to the conclusion that placenta reflects the developing brain in its DNA methylation signatures, mediating the impacts of maternal environment on fetal development. In **Chapter 3** we investigated maternal peripheral blood transcriptome profile with maternal one-carbon metabolites and discovered that maternal immune, apoptotic, and epigenetic mechanisms are important for children's neurodevelopmental outcomes. These results provided new insights on maternal gene pathways associated with child neurodevelopment and suggested protective association with one carbon metabolites in ASD. In **Chapter 4**, we developed the largest placenta methylome in two prospective cohorts in ASD, identified 22q13.33 hypomethylated brain regulatory locus, and described initial functions of a previously uncharacterized transcript *NHIP* using genome-wide multiple-omics studies. *NHIP* increased expression after neuronal differentiation, but decreased expression in ASD placenta and brain. Overexpressed *NHIP* increased cell proliferation in HEK293T cells and regulated genes with functions in histone modification, synapses, and oxidative stress, pathways also identified in ASD postmortem human brain. Both genetic factors and prenatal vitamin intake were associated with 22q13.33 methylation. Our findings underscore the potential utility of placenta DNA methylation profiles

on 22q13.33 block as basis for epigenetic dysregulation in non-invasive, sensitive and accurate early ASD detection. In addition, further characterization of the novel transcript and peptide could lead to a molecular therapeutic protective for ASD in high risk pregnancies.

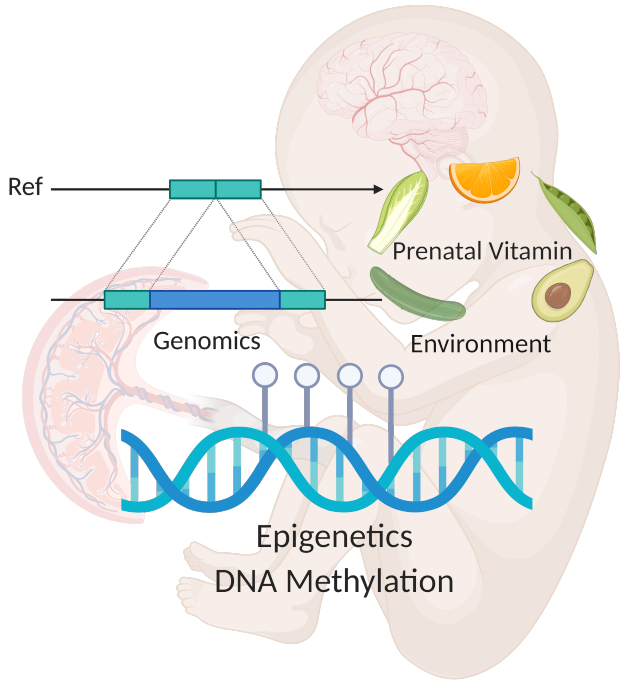
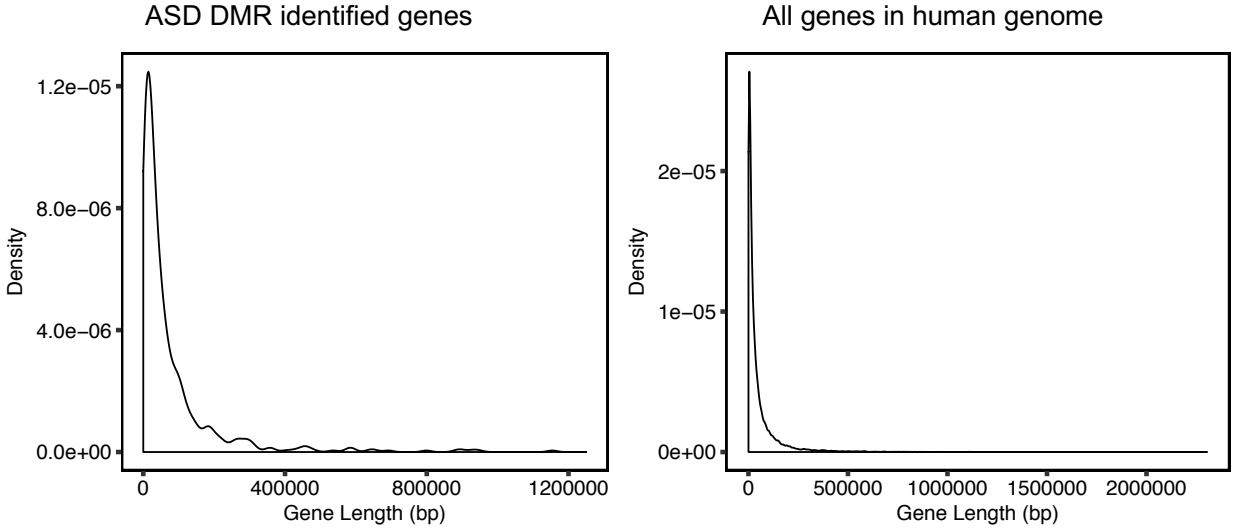


Figure 5. Genetics and perinatal environment interact on epigenetic signatures of ASD. Environment factors and their effects on neurodevelopment through maternal to fetal interaction by placenta-brain axis. Polymorphic genomic backgrounds determine environmental dynamic responses, leaving distinct epigenomic signatures. Integration of multi-disciplinary fields of genetics, environment, and epigenetics are vital to solve the complex etiologies of ASD.

Supplementary Figures

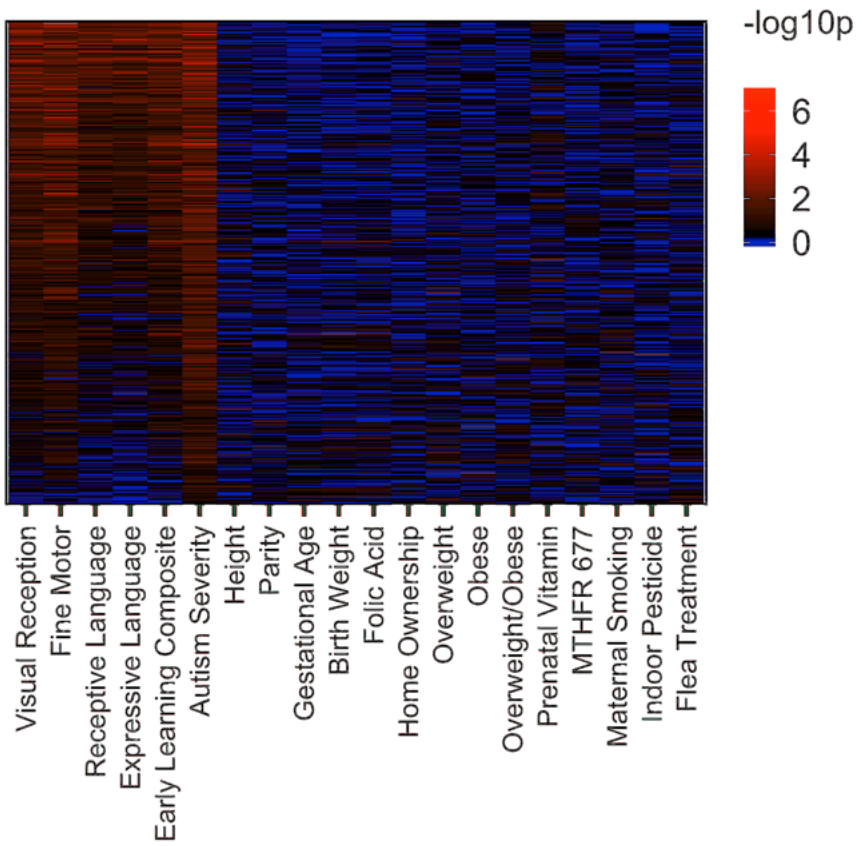
Chapter 2: Supplementary Figures



Supplementary Figure 2.1

The distribution of gene length for ASD DMR genes was similar to all genes in the human genome.

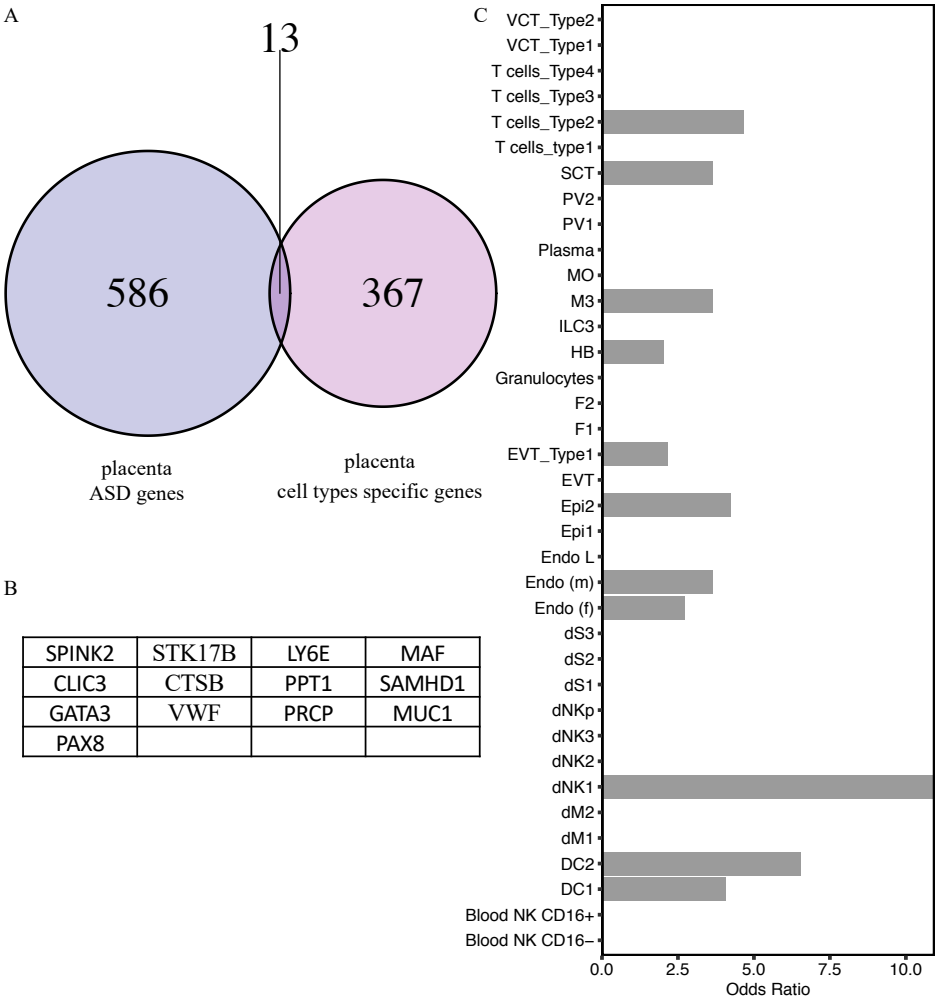
The density plot showed the distribution of gene length of ASD DMR identified genes (left) and all genes (right). The x-axis illustrated the gene length in base pairs. The y-axis was the density for gene length. Person's chi-squared test showed no significant difference between two distributions (*p*-value = 0.9994).



Supplementary Figure 2.2

ASD DMRs heatmap by child outcome continuous measurements of cognition and autism severity versus potential cofounding variables.

The plot shows a heatmap of ASD DMRs (y-axis) and the association of % methylation at each DMR with other measured variables. Each row in the heatmap represents one DMR and each column showed one measured variable. The first 5 child outcome variables on the x-axis includes 4 sub-categories of Mullen scores as well as composite score and autism severity score from the ADOS. Significant associations are red ($p < 0.05$). While ASD DMRs were highly associated with autism severity and to a lesser degree with early learning Mullen's scores, other potential confounding variables from MARBLES exhibited only rare associations with individual ASD DMRs.



SPINK2	STK17B	LY6E	MAF
CLIC3	CTSB	PPT1	SAMHD1
GATA3	VWF	PRCP	MUC1
PAX8			

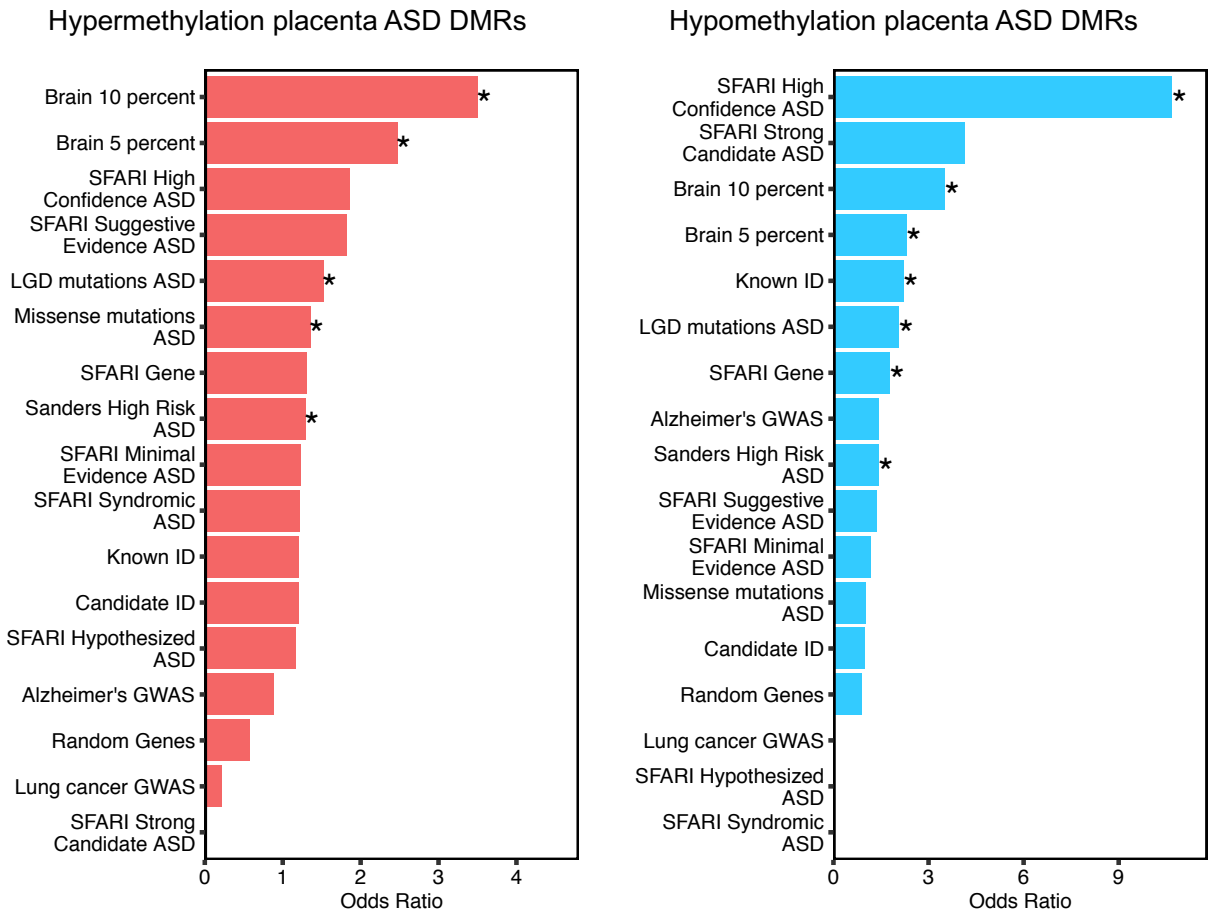
Supplementary Figure 2.3

Placental ASD DMRs genes did not significantly overlap with cell type specific genes in placenta.

A. Venn diagram shows a non-significant overlap of 13 genes associated with placenta ASD DMRs and placental cell type specific genes by two-tailed Fisher's exact test (*p*-value = 0.84).

B. Table of those 13 placental ASD DMR genes that were placental cell type markers.

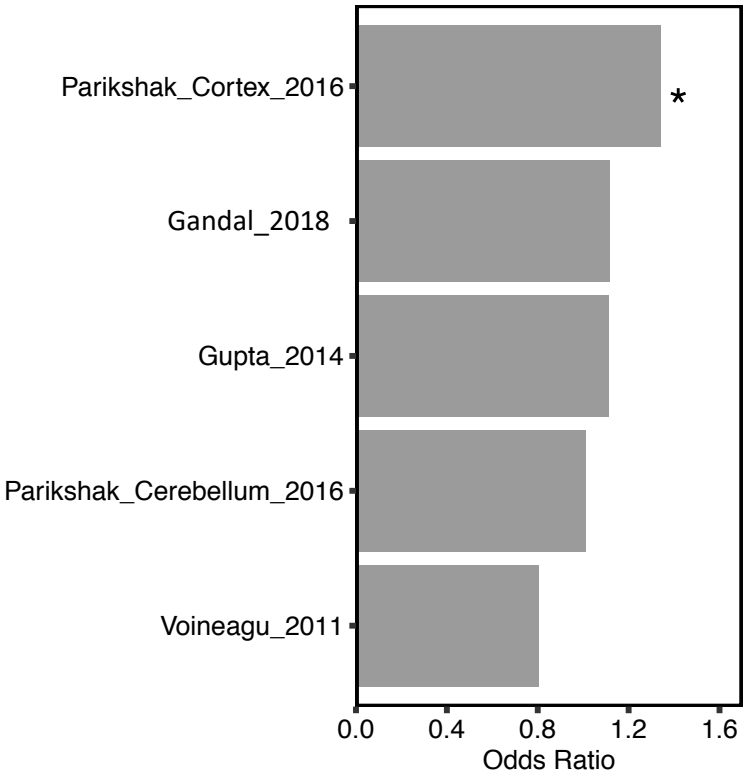
C. Placenta ASD DMR associated genes were compared for overlap by two-tailed Fisher's exact test with each of 38 different identified placental cell types. None of cell type specific genes were significant using two-tailed Fisher's exact test.



Supplementary Figure 2.4

Enrichment test on hyper- versus hypomethylated placenta ASD DMR genes for overlap with ASD risk genes and other databases.

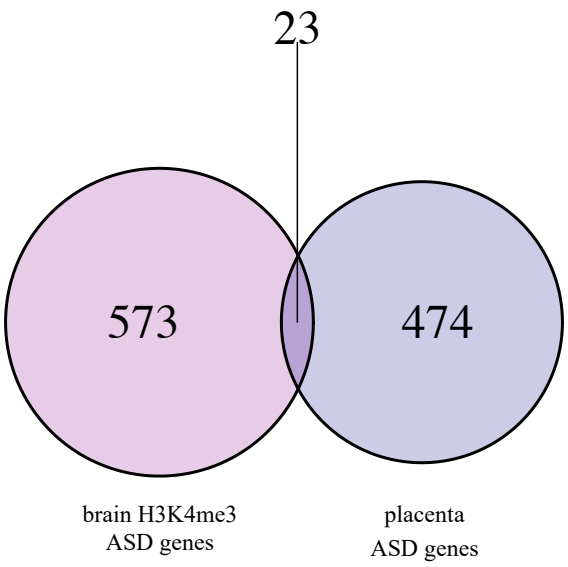
Hypermethylated (left) and Hypomethylated (right) placenta ASD DMRs associated genes overlapped with ASD genetics risk database, other intellectual disability and a random gene list ranked by odds ratio. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$ by two tailed Fisher's exact test after the FDR correction. SFARI: Simons Foundation Autism Research Initiative, LGD: likely gene disrupting mutation, ASD: autism spectrum disorder, Alzheimer: Alzheimer's Disease, ID: intellectual disability.



Supplementary Figure 2.5

Enrichment test on placenta ASD DMR genes and differential expressed genes in ASD postmortem brain.

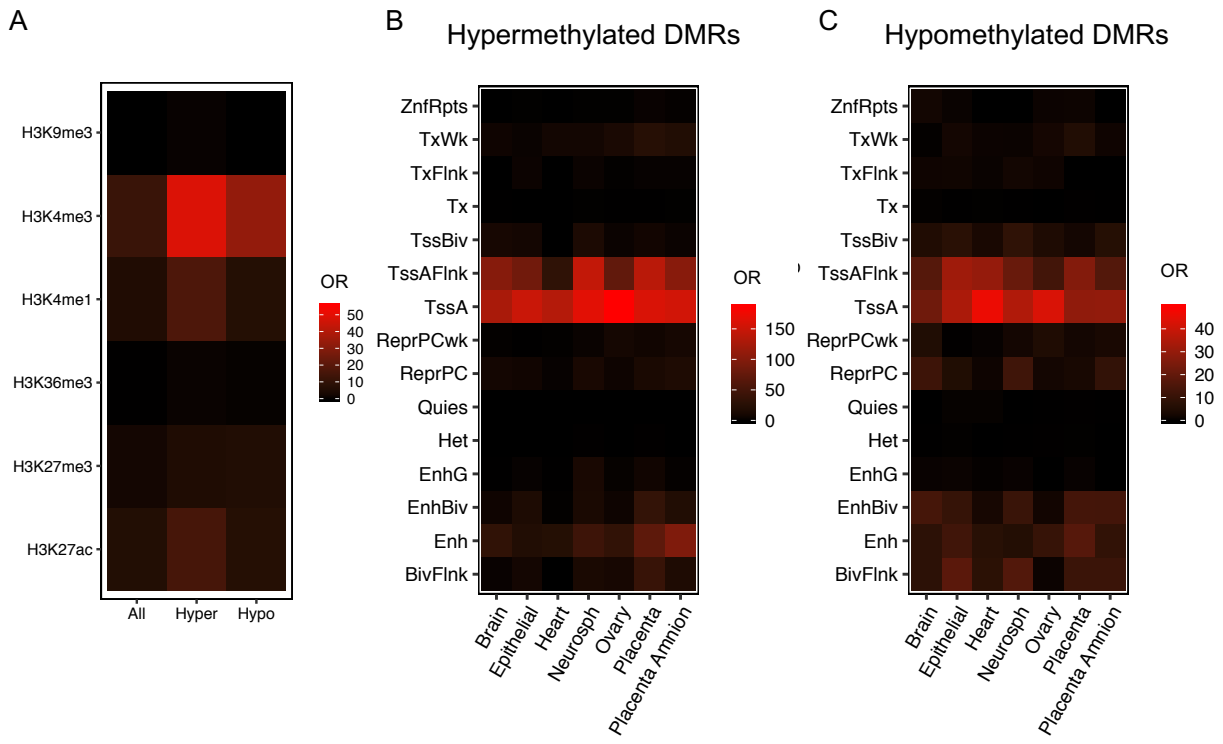
Placenta ASD DMRs associated genes overlapped with four ASD brain differential expressed genes databases ranked by odds ratio. *$p < 0.05$ by two tailed Fisher's exact test after the FDR correction.



Supplementary Figure 2.6

Placental ASD DMR genes significantly overlapped with genes associated woth H3K4me3 changes in ASD brain study.

Venn diagram shows the significant overlap between placenta ASD DMRs genes and genes associated with H3K4me3 changes in brain (Shulha et al, 2012) by Fisher's exact test (*p*-value < 0.05).



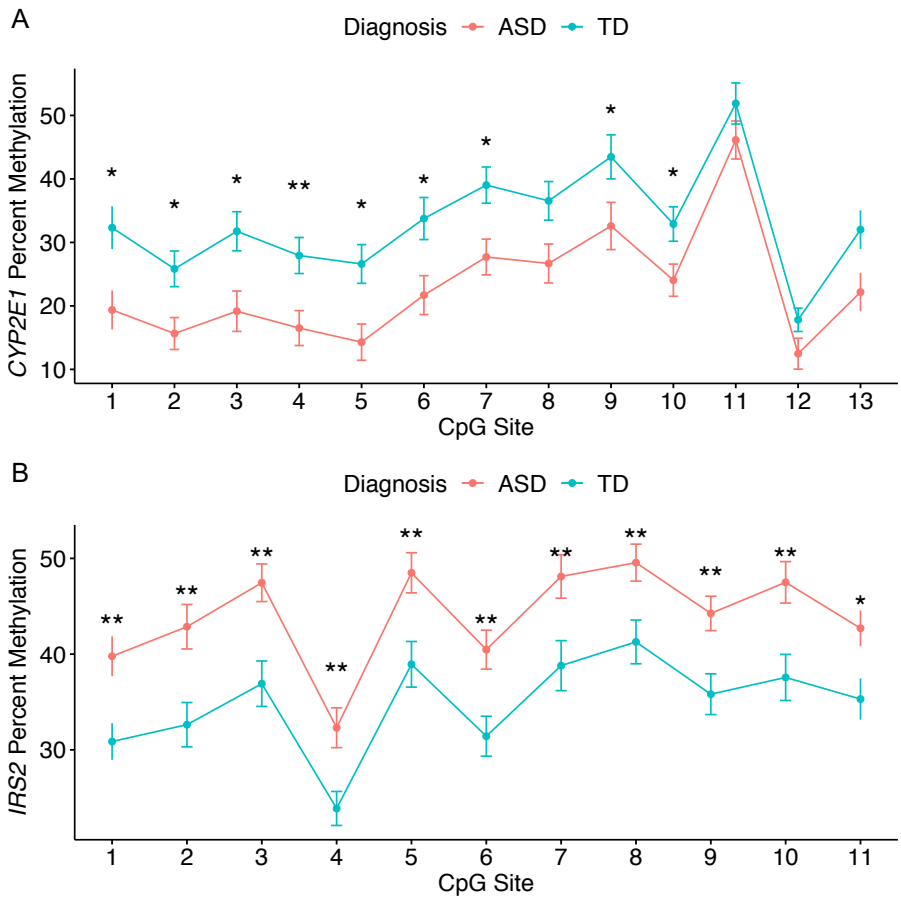
Supplementary Figure 2.7

Placenta ASD hyper- and hypomethylated DMRs were both enriched at H3K4me3 regions, active promoters and their flanking regions.

A. Hyper- and hypomethylated placenta ASD DMRs were overlapped with histone modification human placenta ChIP-seq peaks from the Epigenome Roadmap with odds ratio plotted.

B. Hypermethylated placenta ASD DMRs were tested on chromatin states from the Epigenome Roadmap. The x-axis was different tissue type, the y-axis was represented chromatin states.

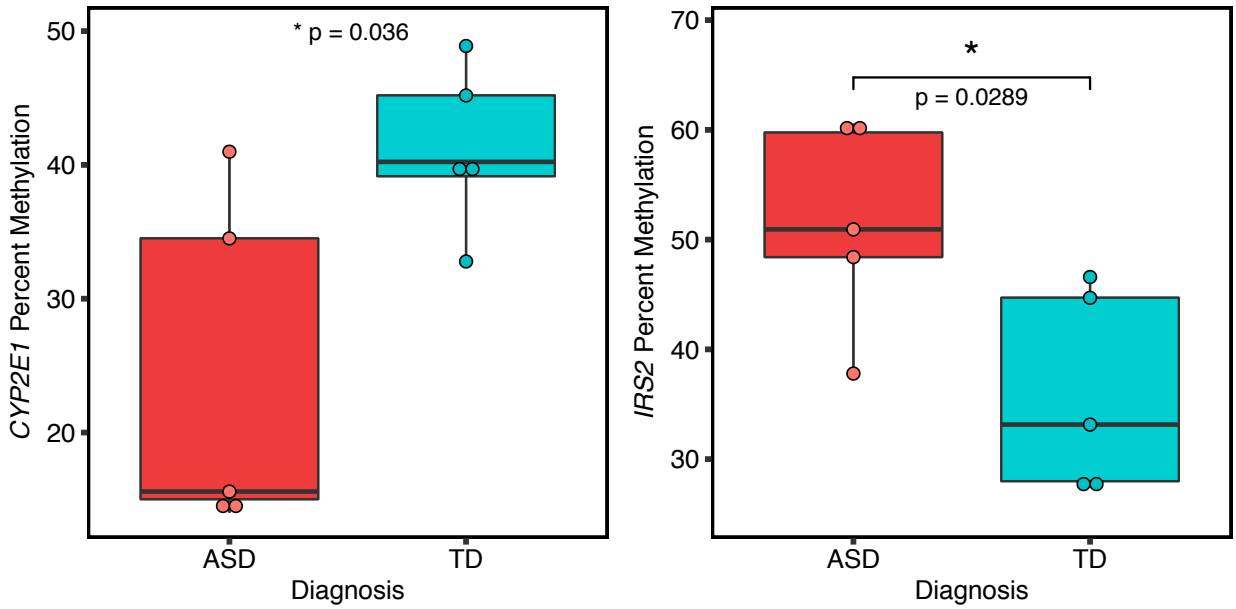
C. Hypomethylated placenta ASD DMRs were tested on chromatin states from the Epigenome Roadmap.



Supplementary Figure 2.8

Pyrosequencing results from *CYP2E1* and *IRS2* DMRs for each CpG site.

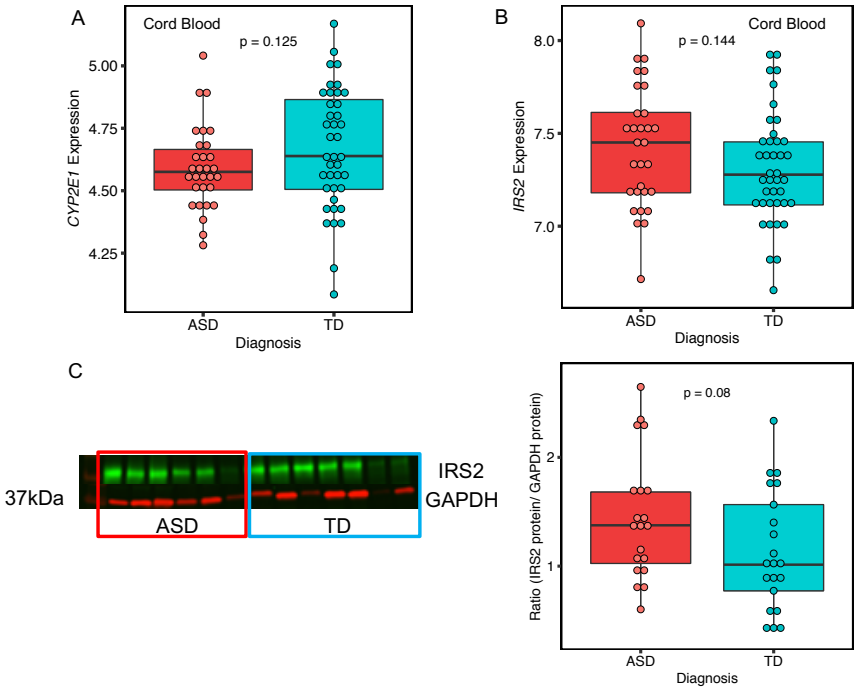
The x-axis represents each CpG sites included in pyrosequencing DMR regions. The y-axis plots the percent methylation at each CpG site. The red line showed the average methylation of ASD samples and the blue line represented the average methylations of TD samples and error bars represent the standard error of the mean. Each CpG site was tested on the significance level with FDR corrected for the numbers of CpGs. *$p < 0.05$, **$p < 0.01$. **A**. 13 CpG sites tested at the *CYP2E1* DMRs with 10 of them showed significant association with diagnosis after FDR correction. **B**. All 11 CpG sites at the *IRS2* DMRs showed significant association with diagnosis after FDR correction.



Supplementary Figure 2.9

Pyrosequencing results from female ASD versus TD placental samples at *CYP2E1* and *IRS2* DMRs.

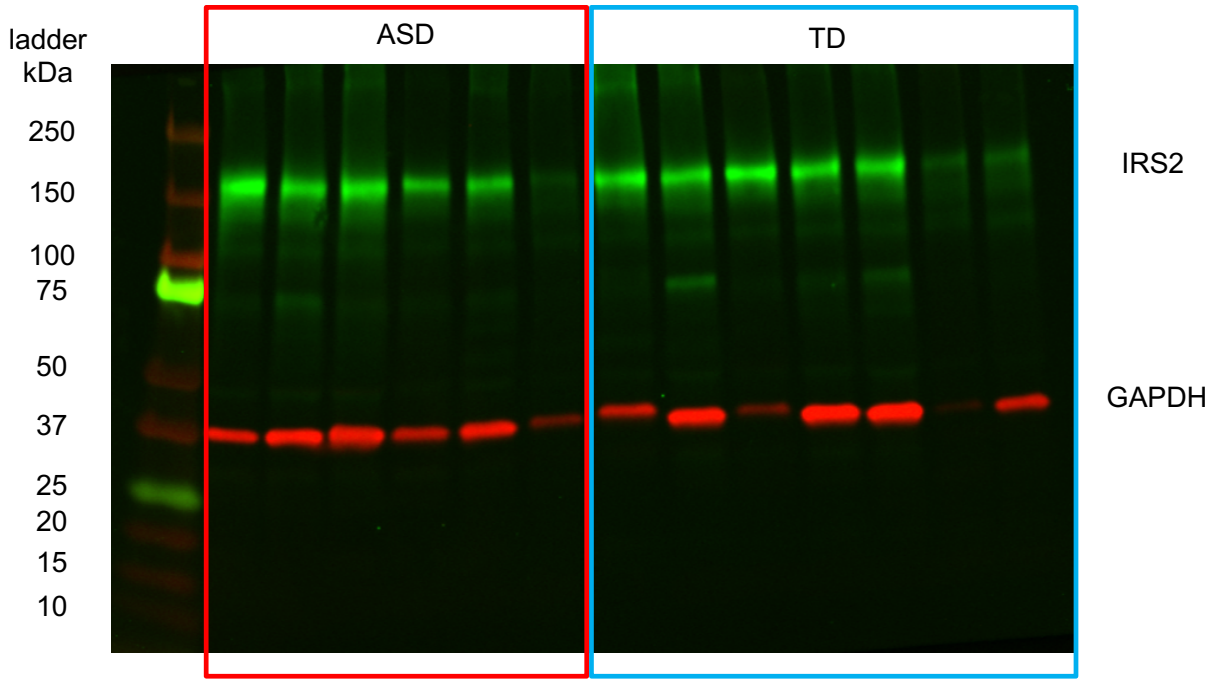
Five ASD female placental samples and five TD female placental samples were used in the analysis. The left panel shows *CYP2E1* DMR percent methylation was significantly lower in ASD compared with TD (two-tailed t-test, *p*-value = 0.036) in the same direction as that observed in ASD males (**Fig. 2.4C**). The right panel shows *IRS2* DMR percent methylation is significantly higher in ASD (two-tailed t-test, *p*-value = 0.029), in the same direction as ASD males (**Fig. 2.4D**).



Supplementary Figure 2.10

For both placental ASD DMRs at *CYP2E1* and *IRS2*, expression trended towards positive correlation with methylation.

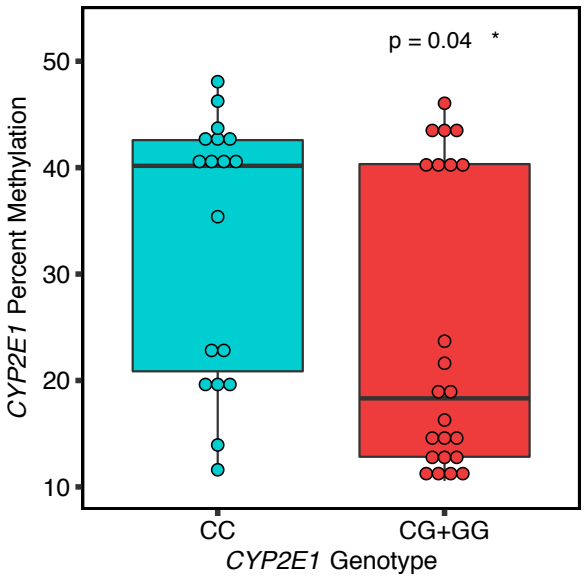
A. 30 ASD and 40 TD umbilical cord blood sample in MARBLES were included in this analysis. Affymetric array matrix data on the probe 16711001 was used to represent the expression of *CYP2E1* on the y-axis. Each dot was used to represent one individual (two-tailed t-test, *p*-value = 0.125). **B**. The same umbilical cord blood samples were used for measuring *IRS2* expression at the probe 16780917 (two-tailed t-test, *p*-value = 0.144). **C**. Representative Westerns blots are shown for the ratio of IRS2 to GAPDH (normalization control) in all 41 placenta samples of ASD and TD comparison with each dot representing one sample (two-tailed t-test, *p*-value = 0.08). A Western blot with 6 samples in ASD and 7 samples TD were showed at the left panel. IRS2 protein was labeled with green fluorescence at 185 kDa and GAPDH was marked with red fluorescence at 37 kDa.



Supplementary Figure 2.11

Representative IRS2 Western blot image.

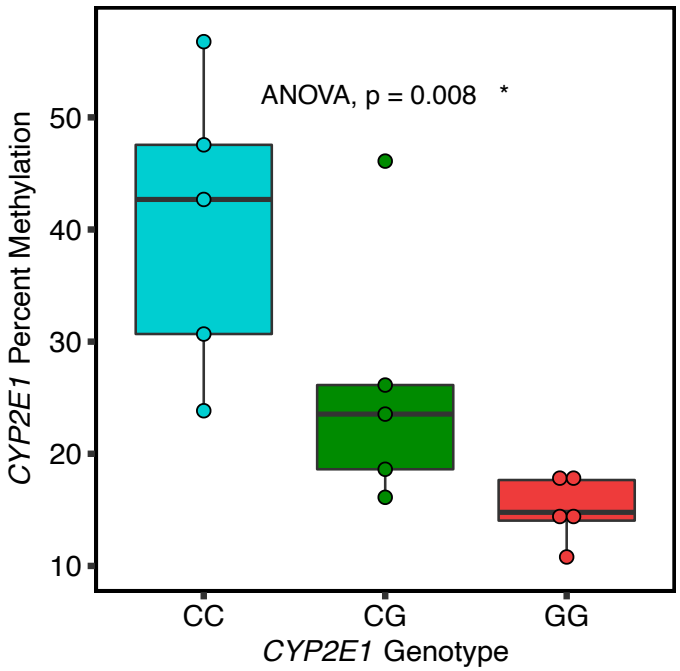
A representative Western blot image of placental protein extracts detected by anti-IRS2 (green) or anti-GAPDH (red) antibodies. The left column is the protein size ladder.



Supplementary Figure 2.12

CYP2E1 genotype (rs1536828) was significantly associated with *CYP2E1* DMR methylation levels.

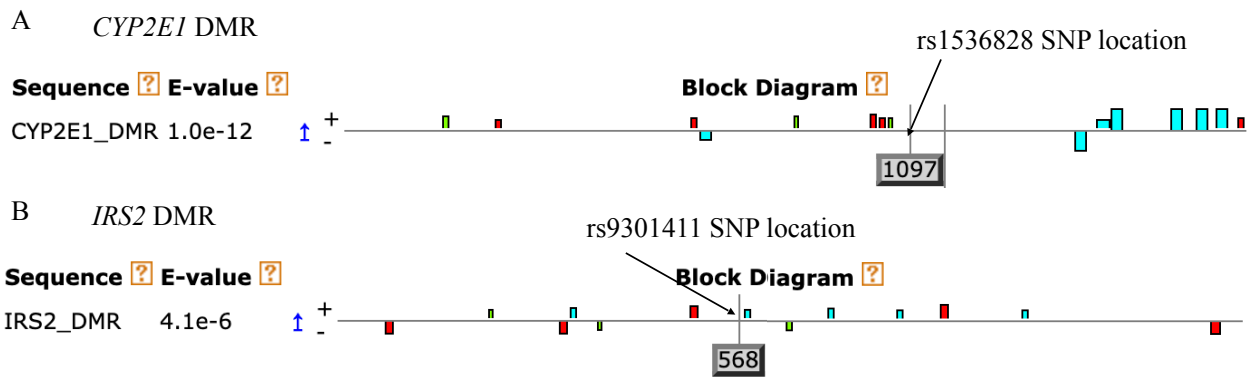
CYP2E1 genotype at rs1536828 combining the minor homozygous genotype (GG) and heterozygous genotype together (CG) together within the ASD DMR was significantly associated with *CYP2E1* DMR average percent methylation tested by two-tailed t-test (*p*-value = 0.04).



Supplementary Figure 2.13

CYP2E1 genotype (rs1536828) was significantly associated with *CYP2E1* DMR methylation levels on an additional 15 MARBLES placental samples.

Sanger sequencing and pyrosequencing were performed on an additional 15 MARBLES placental samples, with 5 samples in each genotype groups (CC, CG, and GG). *CYP2E1* genotype at rs1536828 was significantly associated with *CYP2E1* DMR percent methylation using AVOVA (*p*-value = 0.008).

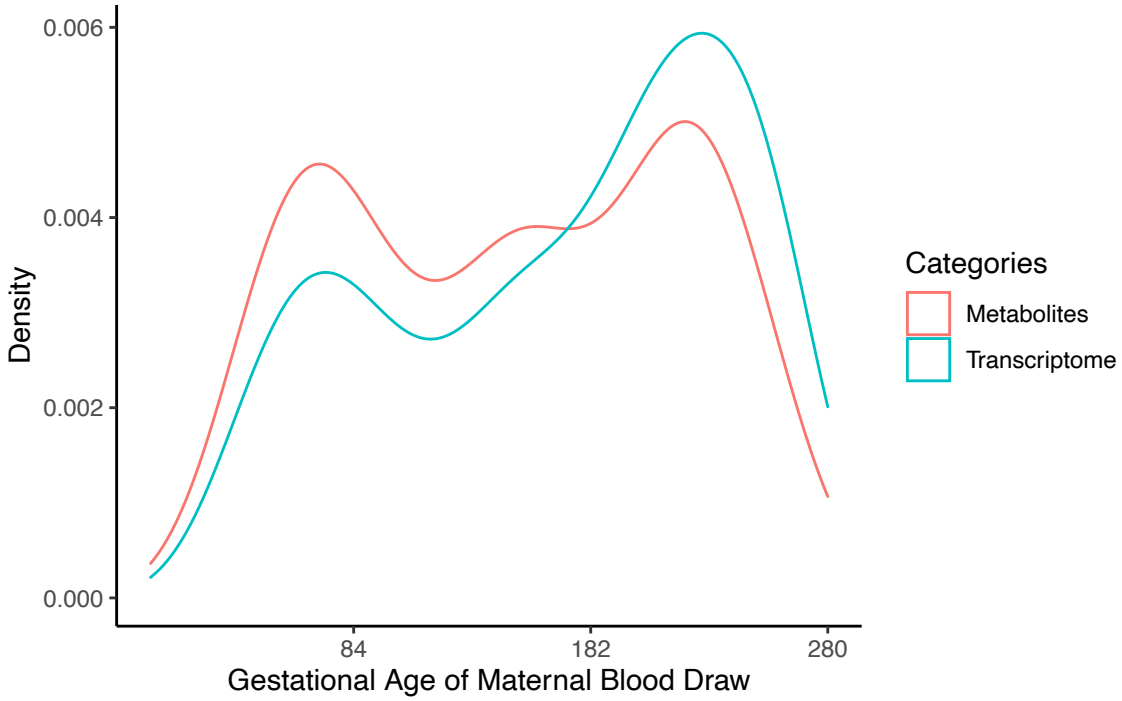


Supplementary Figure 2.14

CYP2E1 genotype (rs1536828) and *IRS2* genotype (rs9301411) do not change transcription factor motifs.

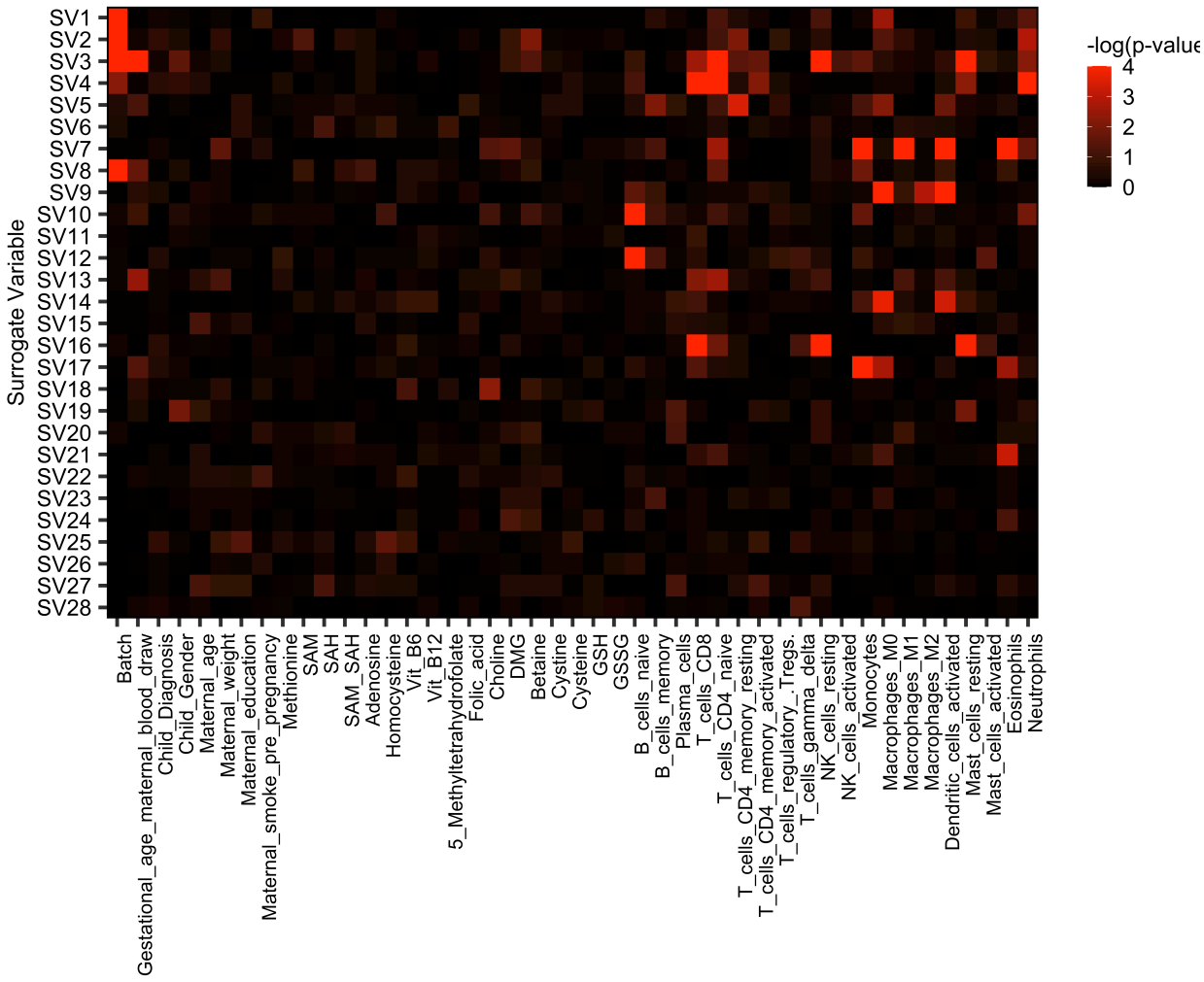
Motif structure identified by MEME at *CYP2E1* DMR (**A**) and *IRS2* (**B**) DMR. Horizontal line represents the DMR DNA sequence. Each block shows the location of a transcription factor motif. Arrows point to the SNP location on the sequence. Numbers within grey boxes represent the relative SNP location to the DMR DNA sequence.

Chapter 3: Supplementary Figures



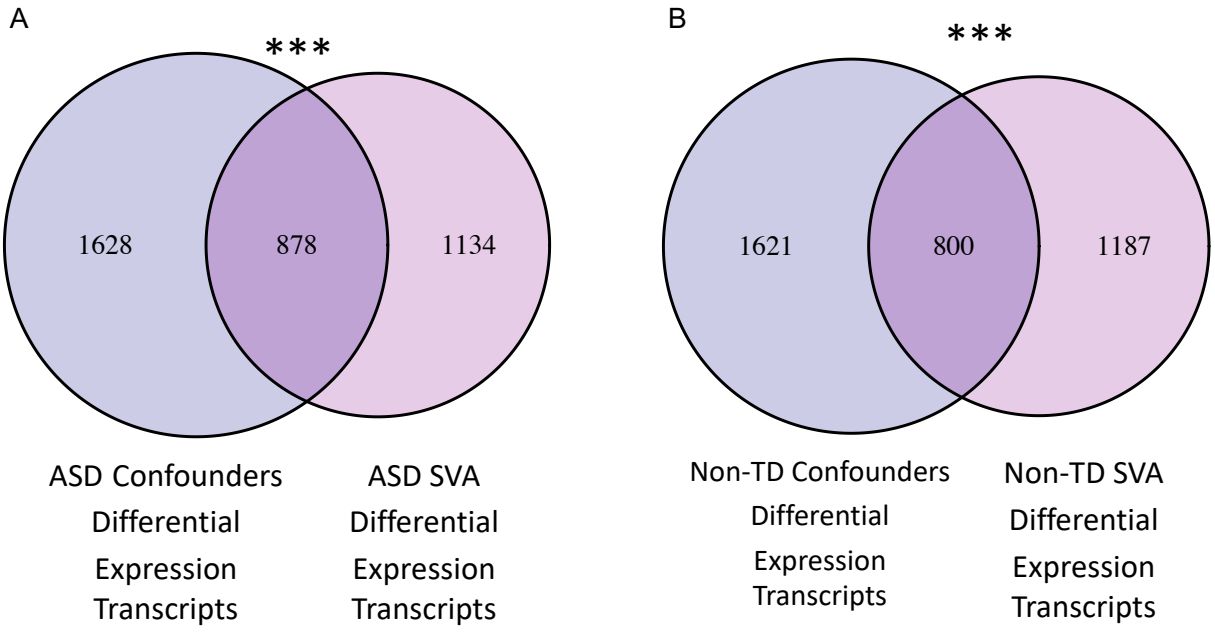
Supplementary Figure 3.1

The number of samples assayed for metabolites and transcriptome for each gestational age of maternal blood draw in this study. Since not all metabolites were able to be measured in each sample, the metabolites curve represents the number of samples with >30% of metabolite measurements detectable.



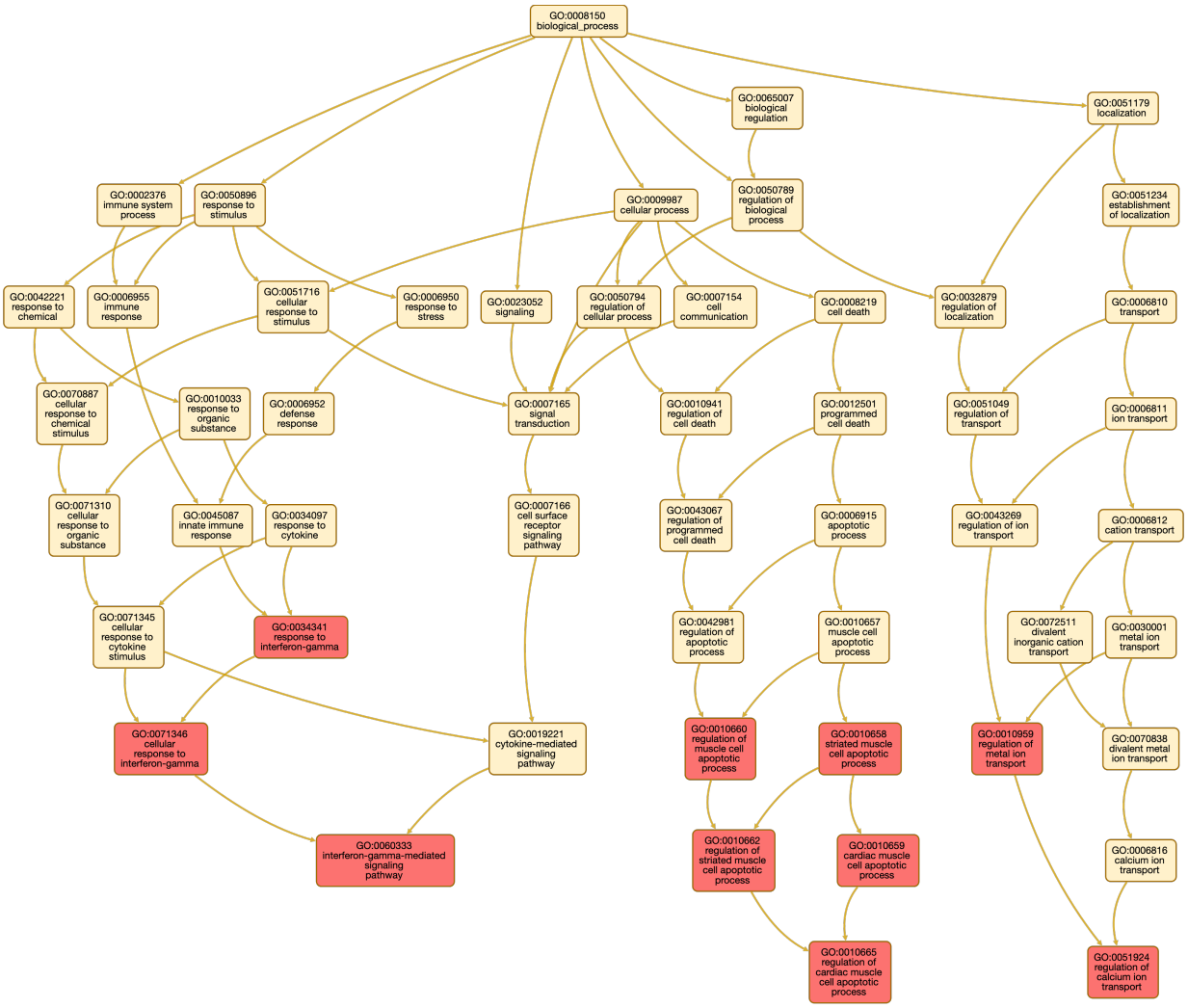
Supplementary Figure 3.2

Surrogate variable analysis in MARBLES subjects.



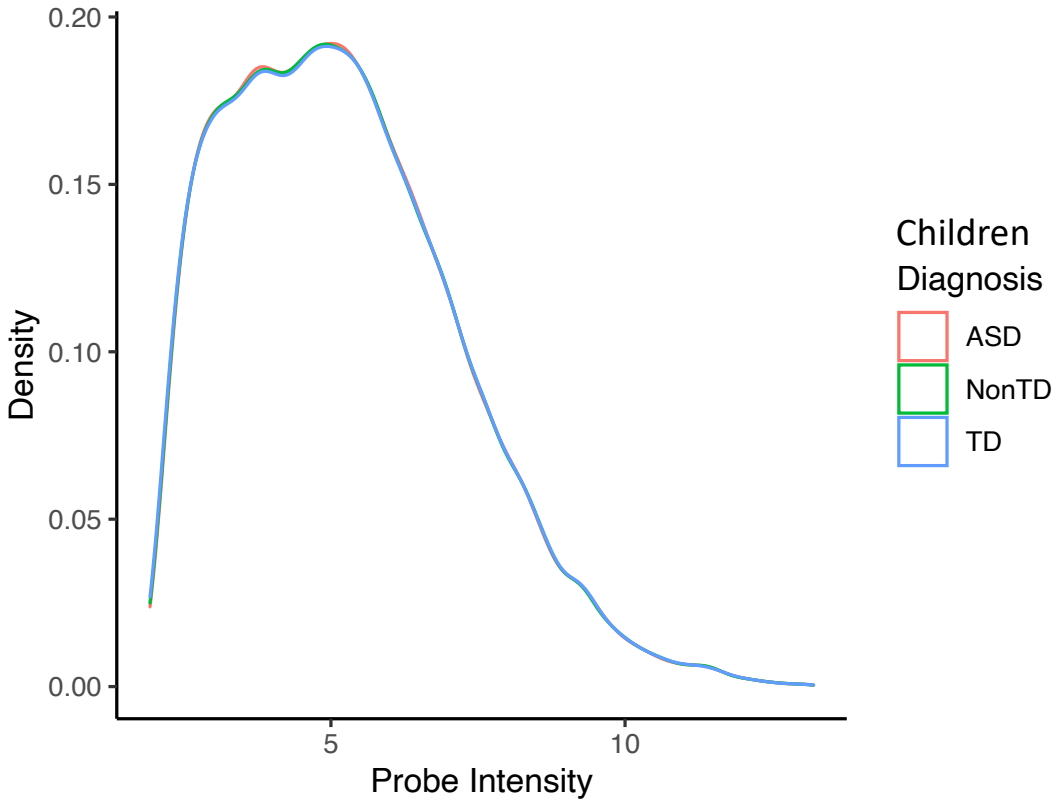
Supplementary Figure 3.3

Differential expressed transcripts on using known confounders compared to SVA adjustments.



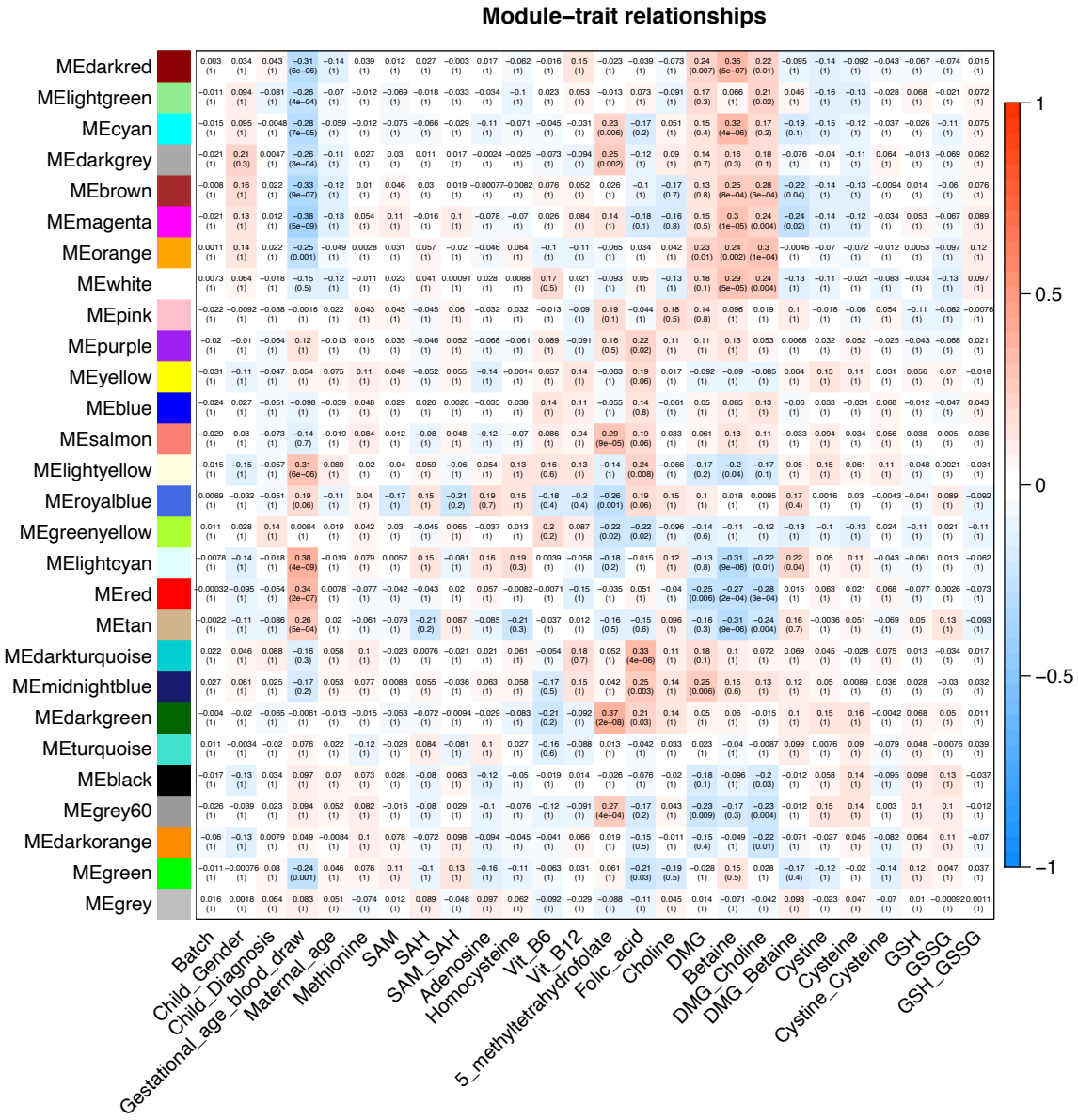
Supplementary Figure 3.4

Gene ontology and pathway analysis using directed acyclic graph (DAG) on the 218 transcripts common to ASD vs. TD and Non-TD vs. TD differentially expressed gene lists.



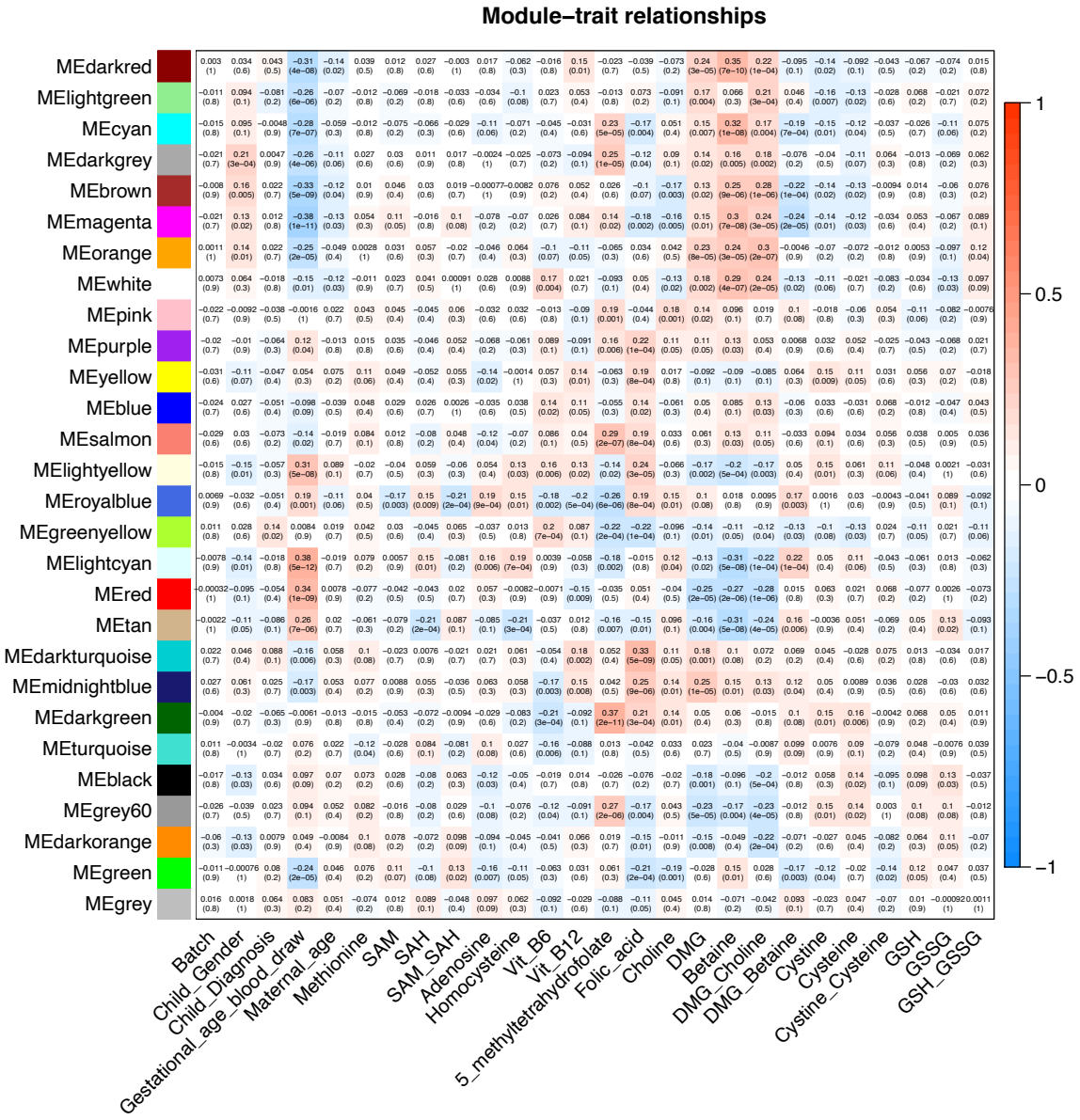
Supplementary Figure 3.5

Probe intensity density plot based all 300 maternal blood Affymetrix data.



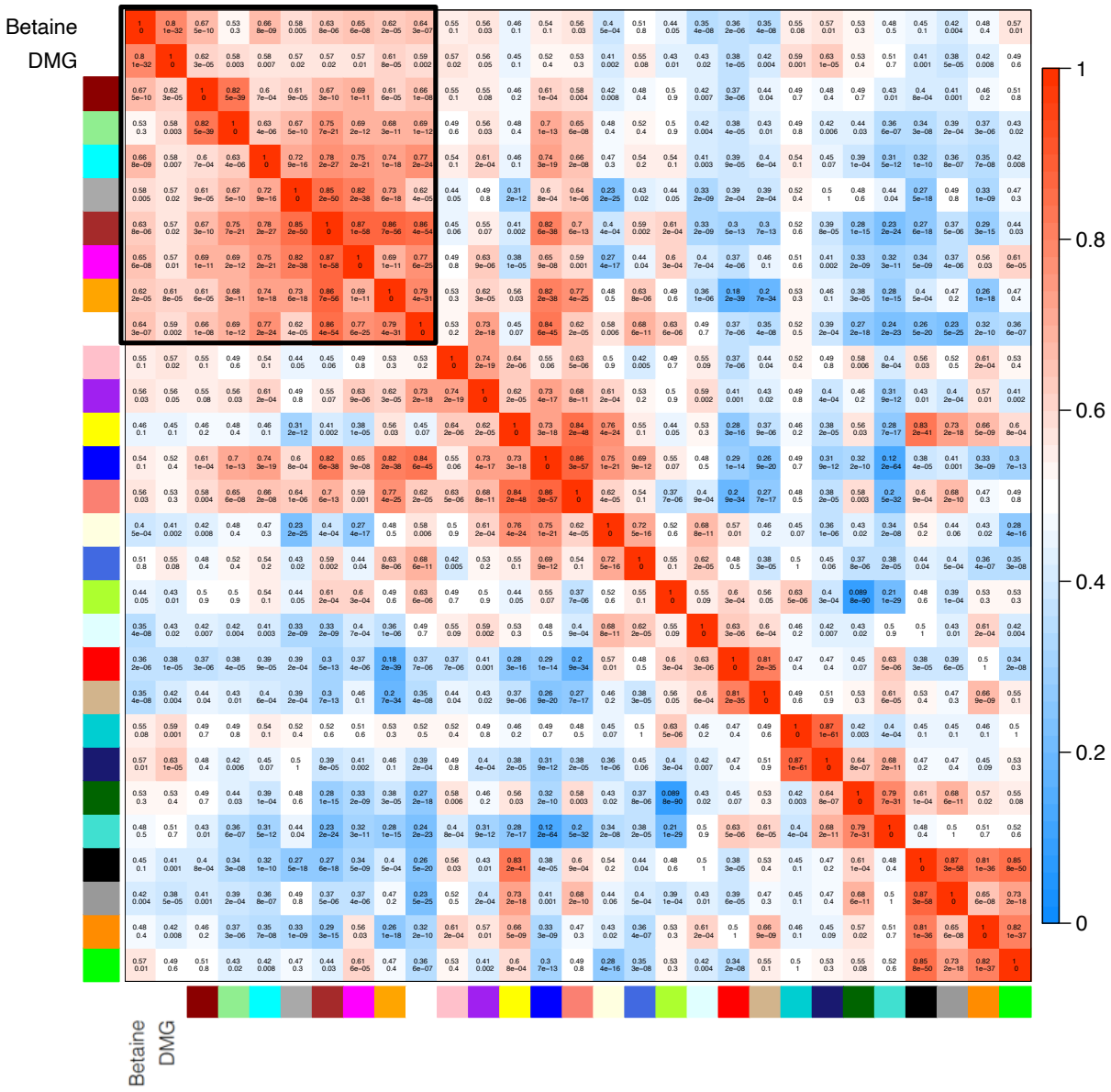
Supplementary Figure 3.6

Co-expression network modules with diagnosis, demographic factors, and maternal blood nutrient concentrations. The values in the cells represents Pearson r (adjusted *p*-value). *p*-value were adjusted for all comparisons.



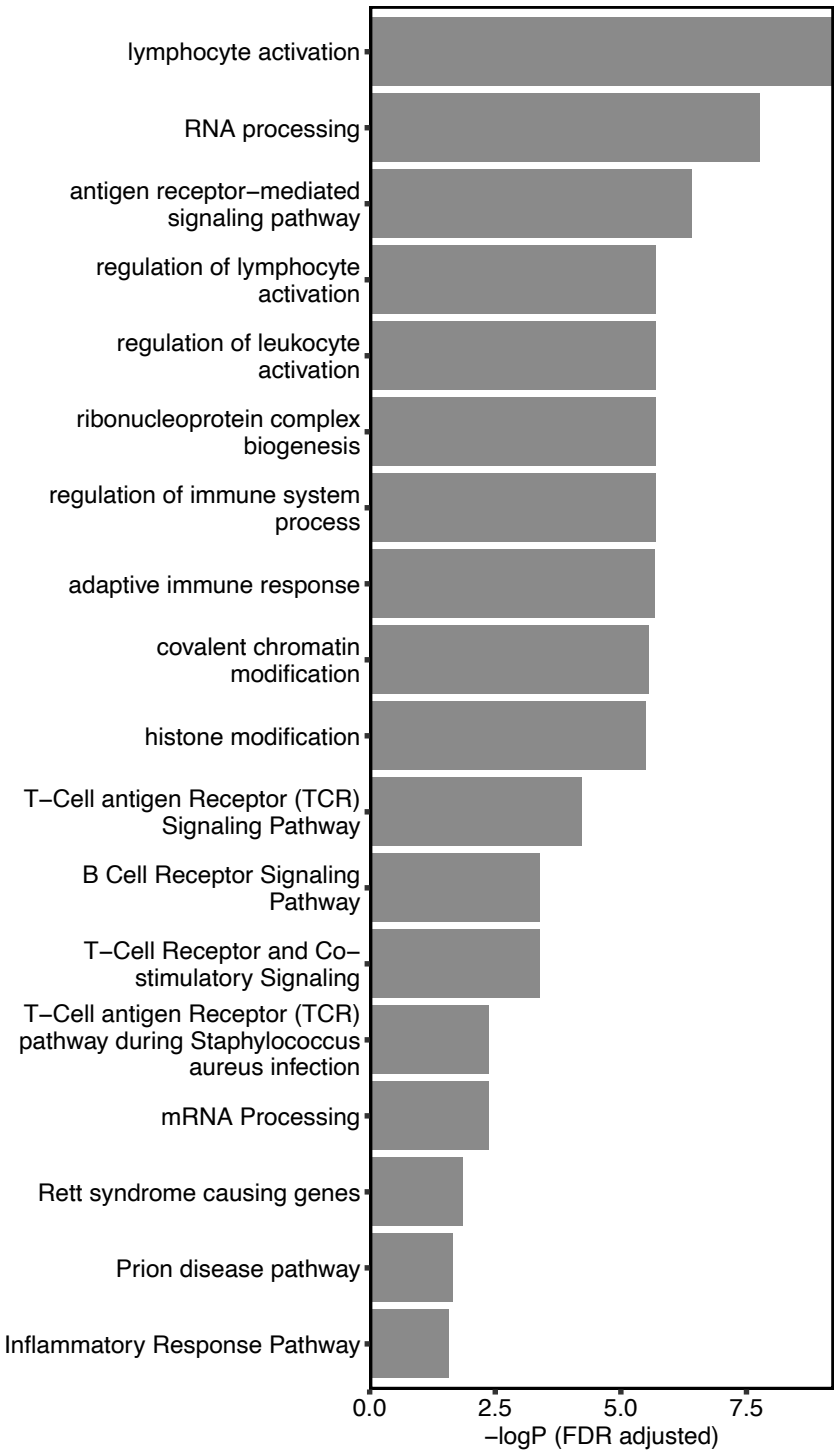
Supplementary Figure 3.7

Co-expression network modules with diagnosis, demographic factors, and maternal blood nutrient concentrations. The values in the cells represents Pearson r (*p*-value). *p*-value shown here were raw *p*-value without adjustment.



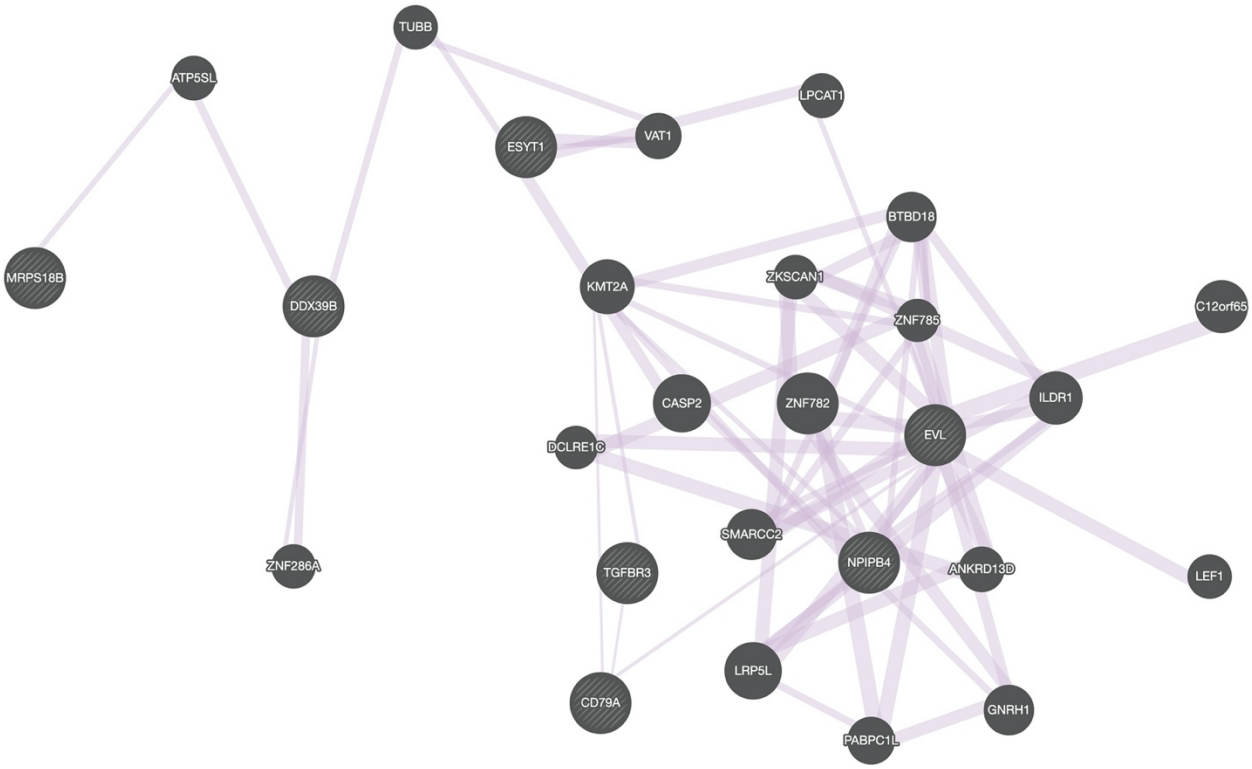
Supplementary Figure 3.8

Unsupervised hierarchical clustering adjacency heatmap correlation and *p*-value.



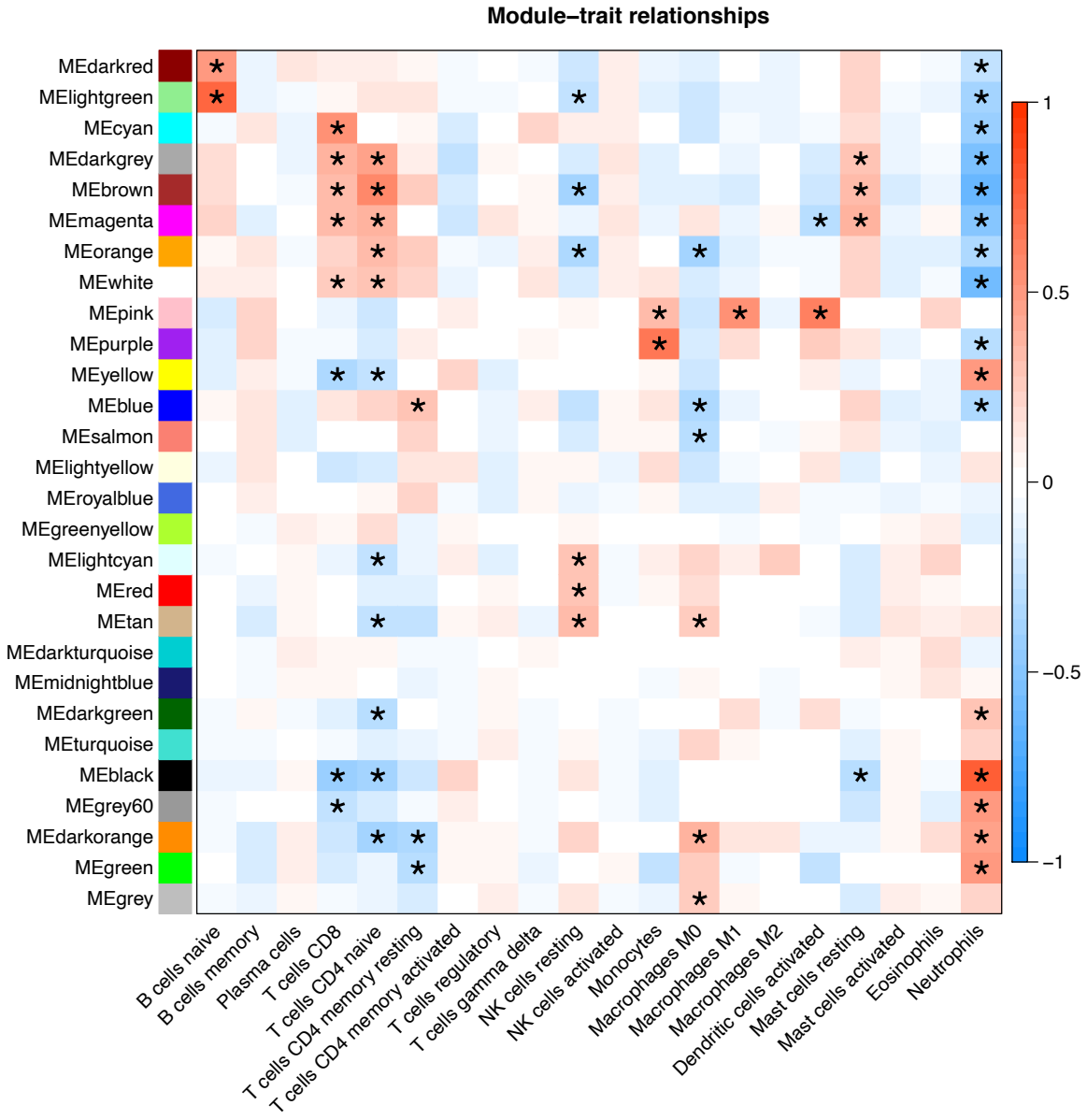
Supplementary Figure 3.9

Gene ontology and pathway analysis for the block of eight weighted gene co-expression modules associated with betaine and DMG.



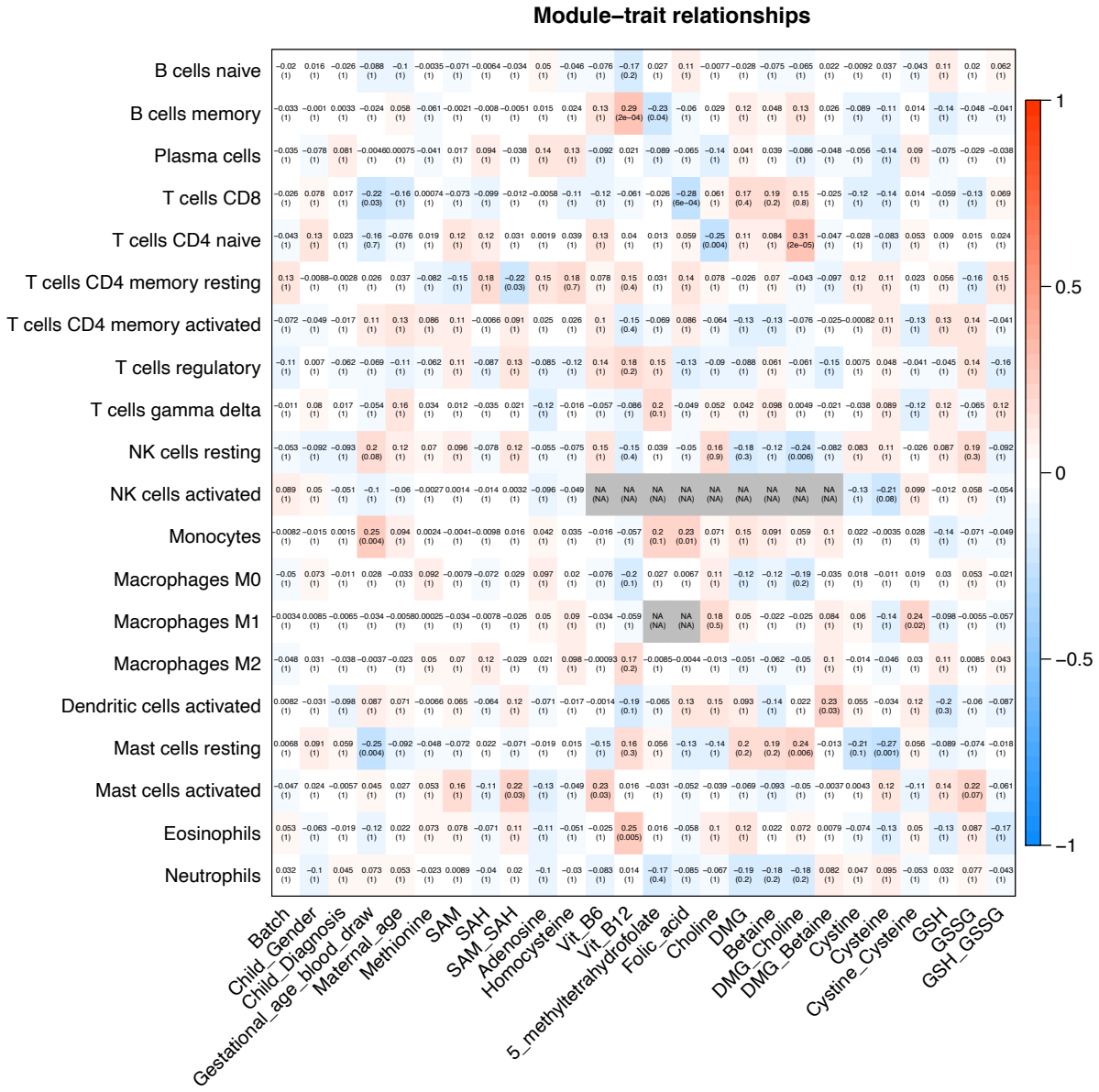
Supplementary Figure 3.10

Association network on hub genes from the block of eight weighted gene co-expression modules associated with betaine and DMG.



Supplementary Figure 3.11

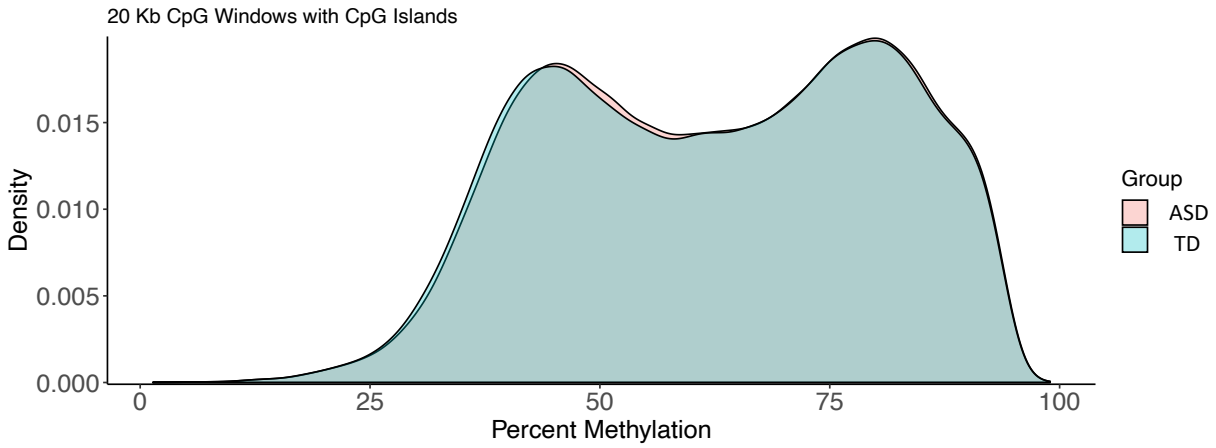
Heatmap of correlation between module eigengenes and cell type proportions with FDR adjusted *p*-value.



Supplementary Figure 3.12

Heatmap of correlation between sample demographic factors and nutrients and cell type proportions with FDR adjusted p-value.

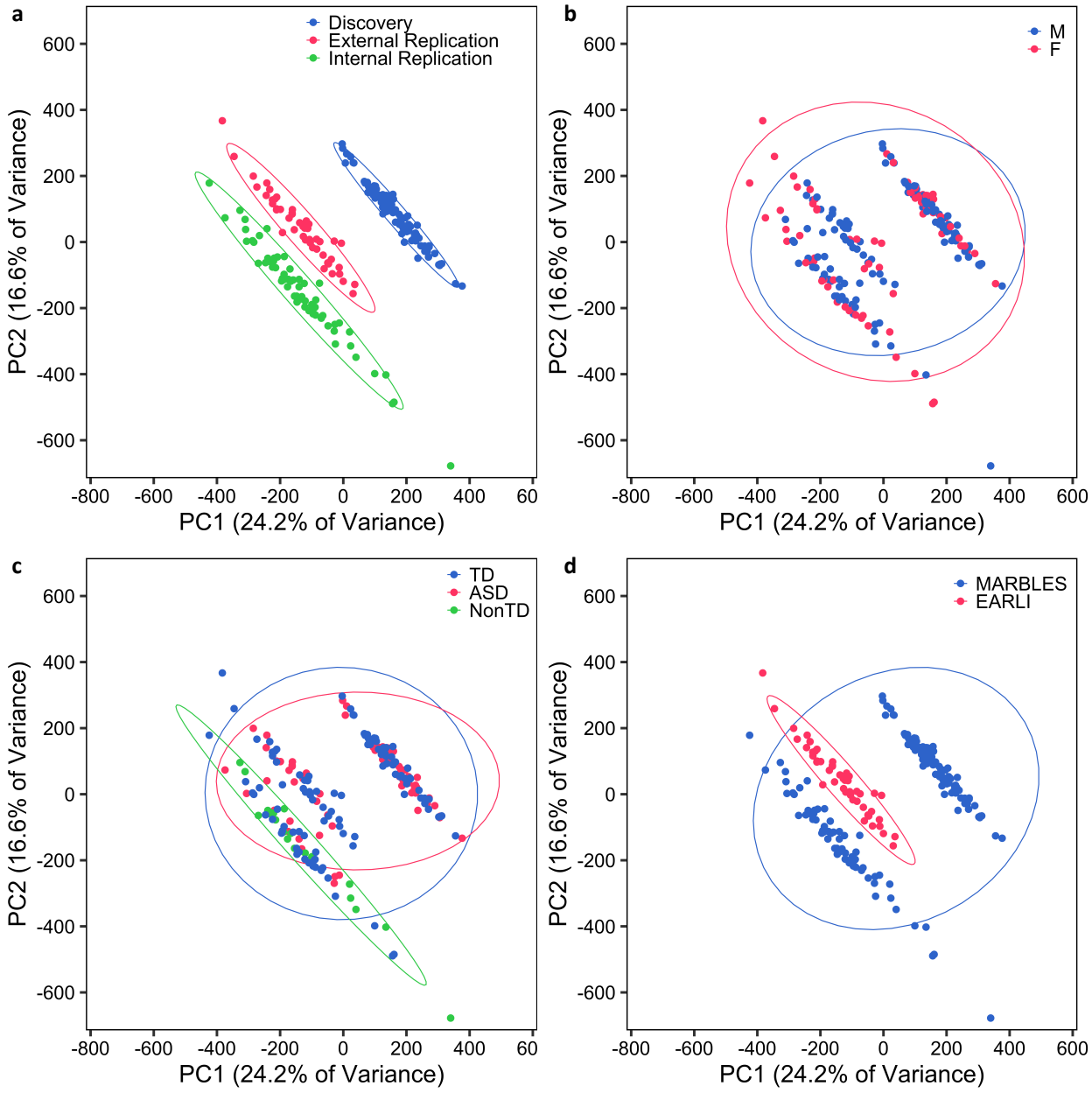
Chapter 4: Supplementary Figures



Supplementary Figure 4.1

Sequencing platform has a larger effect on 20 kb window methylation than ASD diagnosis, sex, or study.

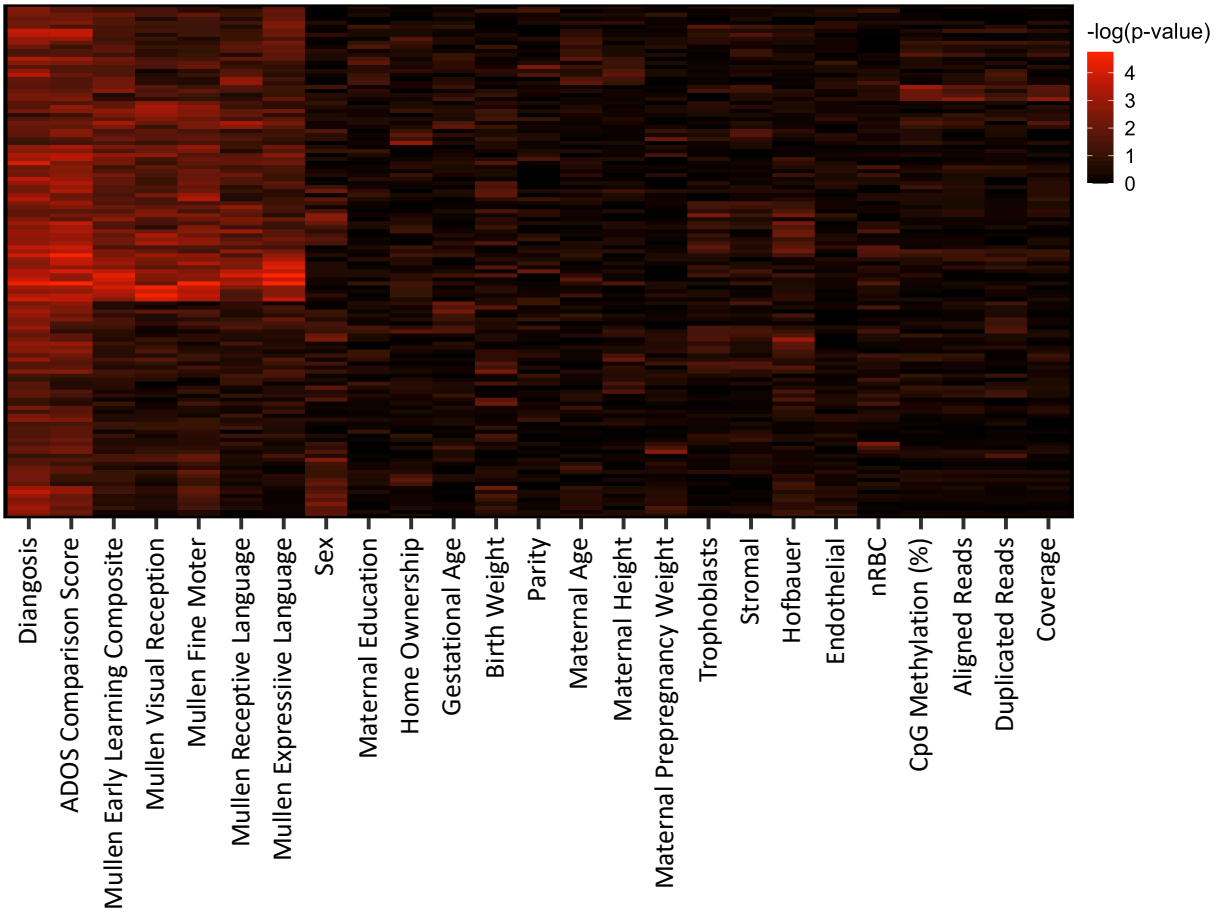
PCA plots using percent methylation at 20 kb windows across the genome for each sample. Points were colored by group, sex, diagnosis, or study. Ellipses indicate 95% confidence limits.



Supplementary Figure 4.2

No global methylation difference was observed between ASD and TD using 20 kb window methylation.

Density plot of average percent smoothed methylation for CpGs covered in the discovery group.

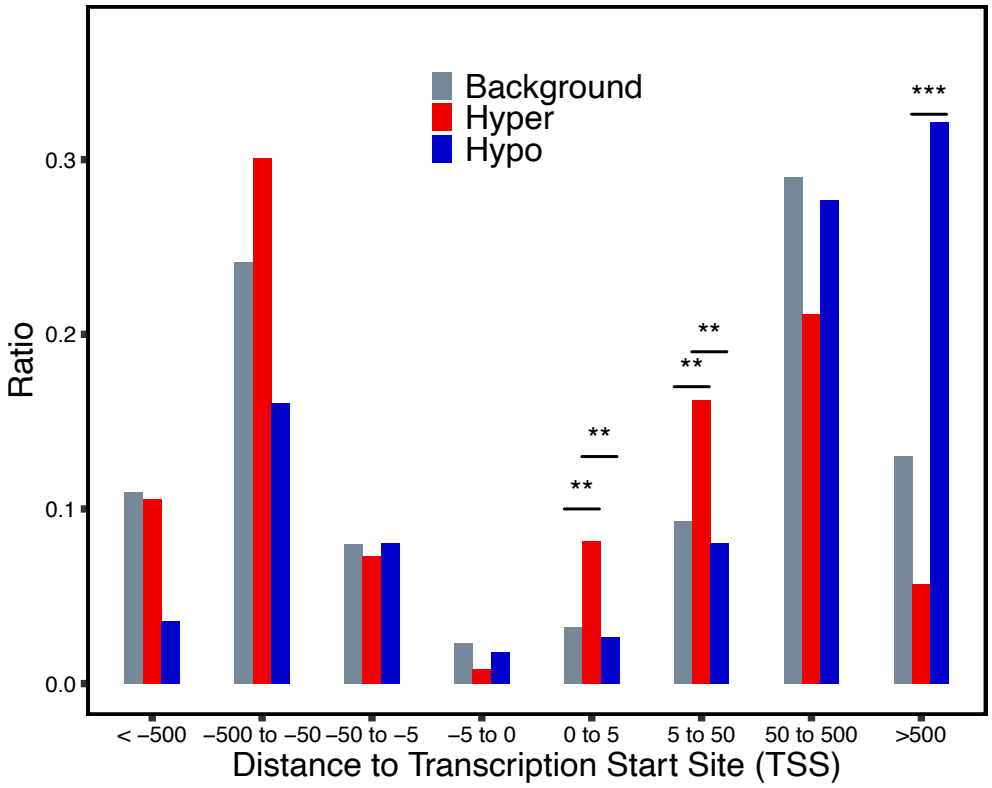


Supplementary Figure 4.3

Methylation level at DMRs is specifically associated with behavioral outcome.

Smooth methylation at DMRs was compared with demographic and technical variables.

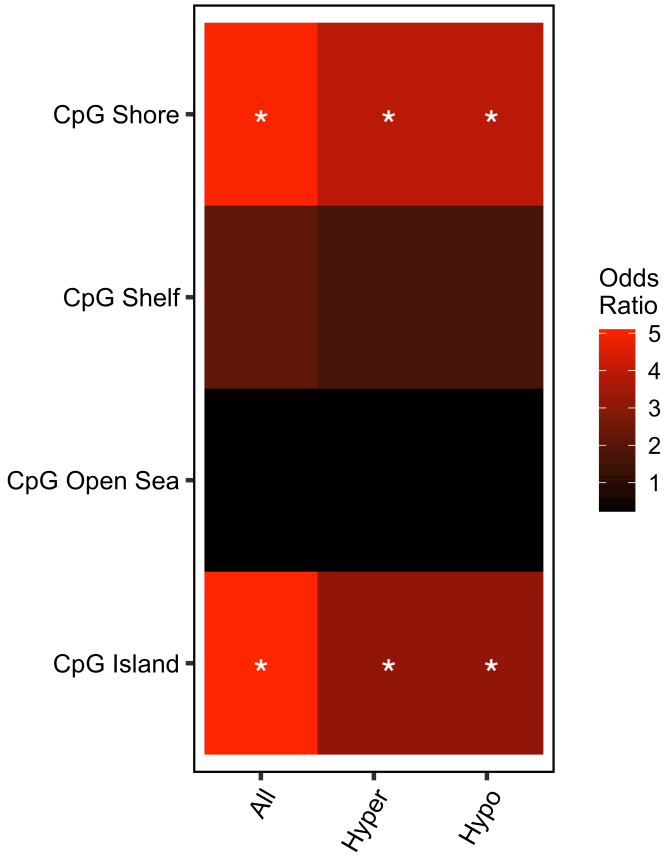
Significant test was performed using linear regression.



Supplementary Figure 4.4

DMRs were enriched at 5 kb and 5 – 50 kb downstream off TSS compared to the background regions.

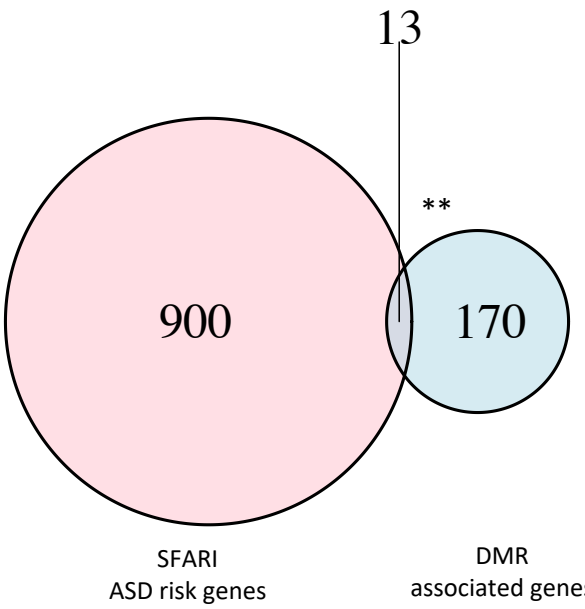
Locations relative to genes for hypermethylated (red) and hypomethylated (blue) DMRs compared to background regions (grey). Distributions of locations relative to transcription start sites (TSS). The ratio was plotted based on the number of genes at each binned location divided by the total number of genes. Significant levels were calculated using Fisher's exact test. (* *p*-value < 0.05, ** *p*-value < 0.01, *** *p*-value < 0.001).



Supplementary Figure 4.5

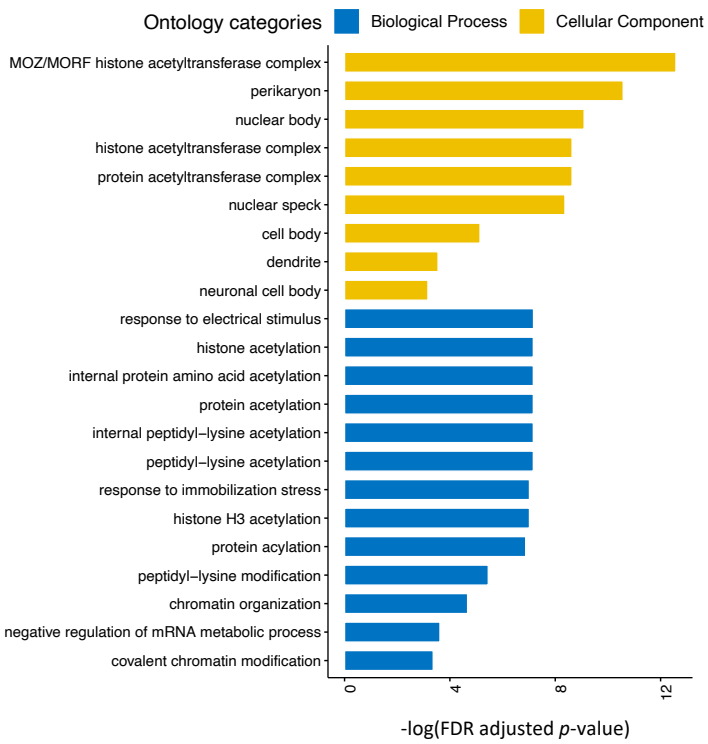
ASD DMRs were enriched at CpG islands and CpG shores.

ASD DMRs (categorized as all, hypermethylated or hypomethylated in ASD) were tested for enrichment based on CpG island location.



Supplementary Figure 4.6

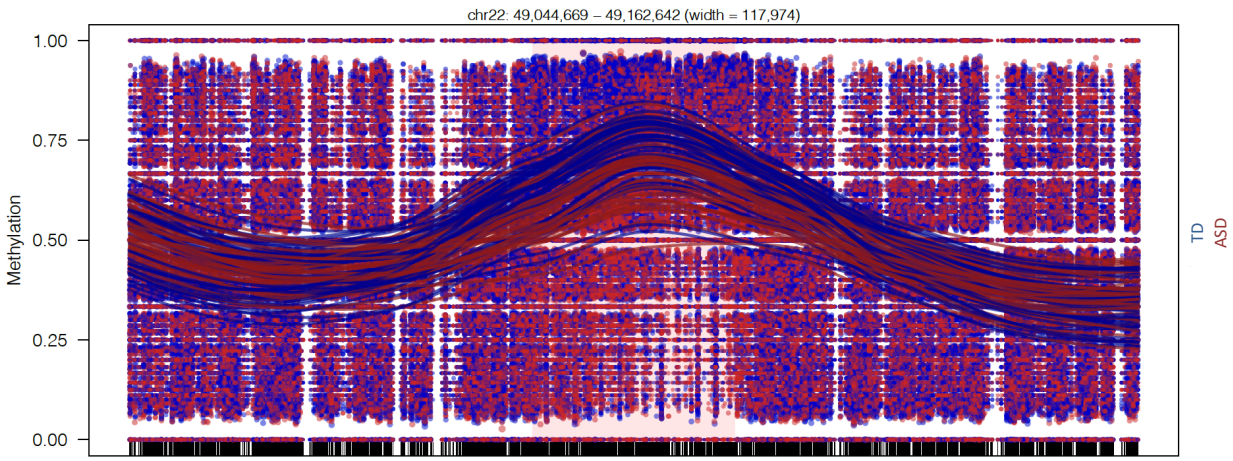
DMRs associated gene has significantly overlapped with SFARI ASD genes (Fisher's exact test, *p*-value < 0.01).



Supplementary Figure 4.7

ASD DMRs were enriched for histone acetyltransferase and chromatin modification.

Bar graph represented the significant results from GREAT gene ontology and pathway enrichment analysis of ASD DMRs associated gene with background regions using Fisher's exact test (FDR adjusted *p*-value).

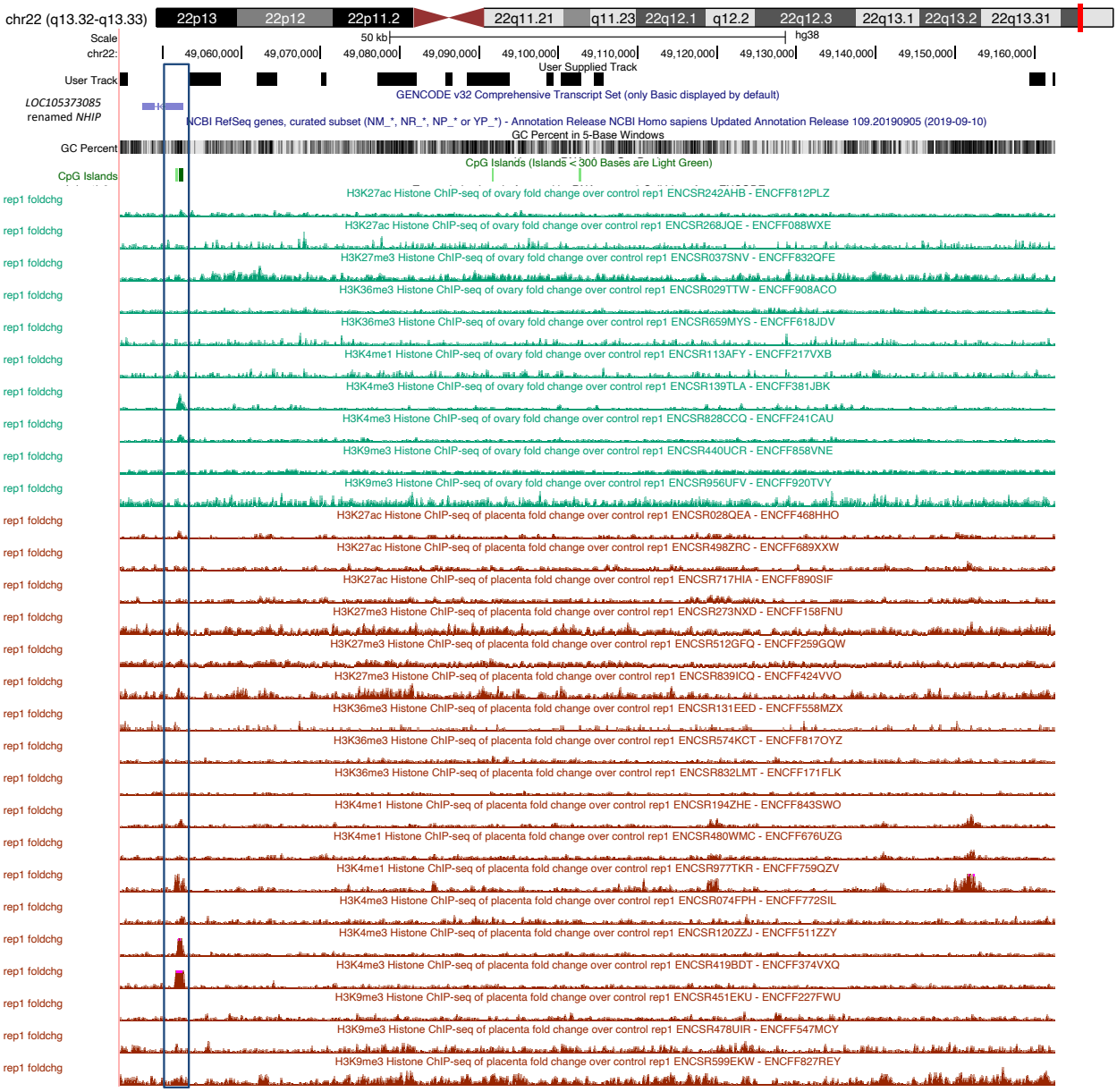


Supplementary Figure 4.8

Significant hypomethylated on comethylated 22q13.33 block.

The lines represent individual smoothed methylation level estimated for ASD (red) or TD (blue).

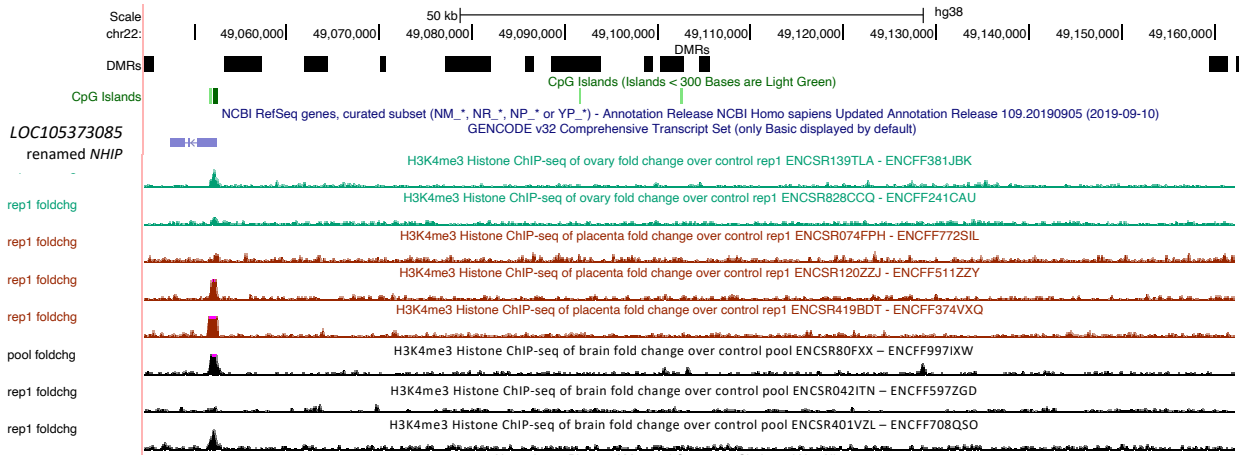
The dots represent the methylation level estimated of individual CpG.



Supplementary Figure 4.9

H3K4me3 peaks were observed at 22q13.33 comethylated block.

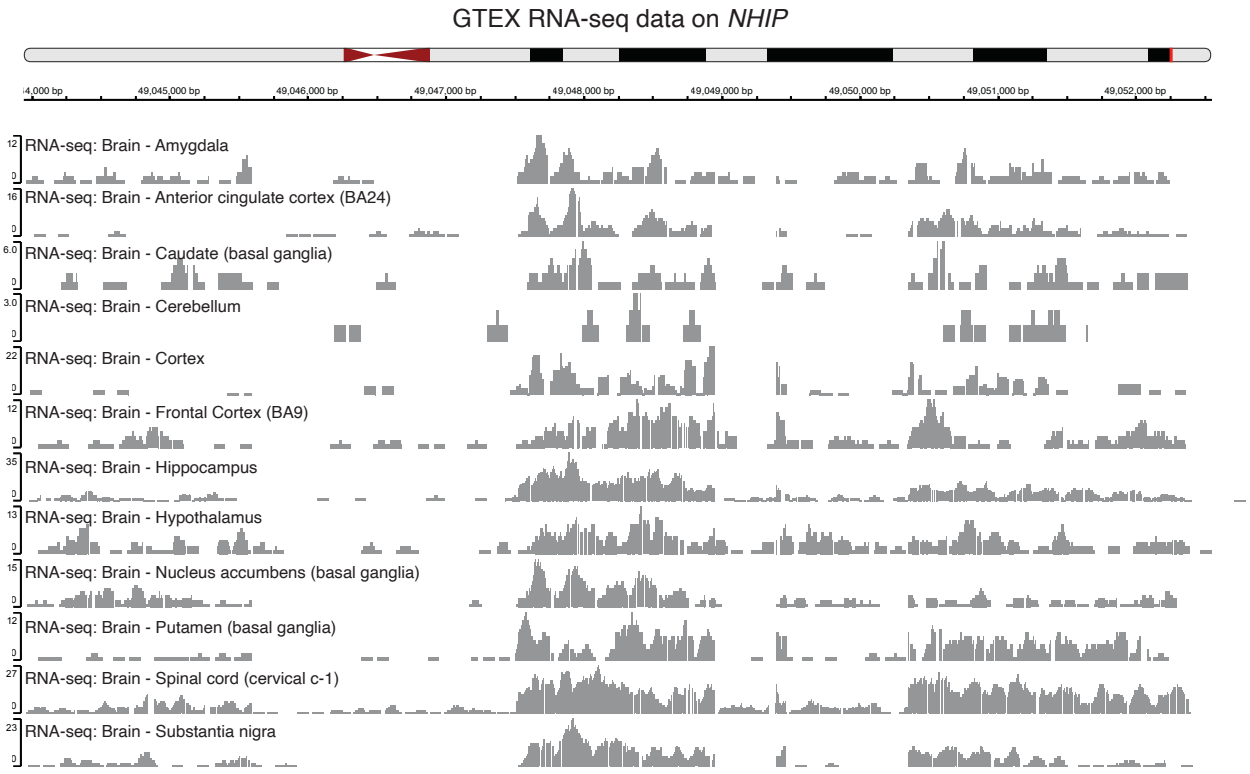
Histone post-translational modifications (PTMs) markers for ovary and placenta were extracted from ENCODE data (359,483) at the 22q13.33 comethylated block region to UCSC genome browser. Black box illustrated the H3K4me3 peak observed on CpG island at the TSS site of *NHIP*.



Supplementary Figure 4.10

The H3K4me3 peak at the TSS of *NHIP* shown polymorphic among individuals in ovary, placenta, and brain.

H3K4me3 peaks from multiple individuals on ovary, placenta, and brain were extracted from ENCODE data (359,483) at the 22q13.33 comethylated block regions to UCSC genome browser. There was a variability on the existence of the H3K4me3 peak among individuals.



Supplementary Figure 4.11

NHIP was variable among brain regions.

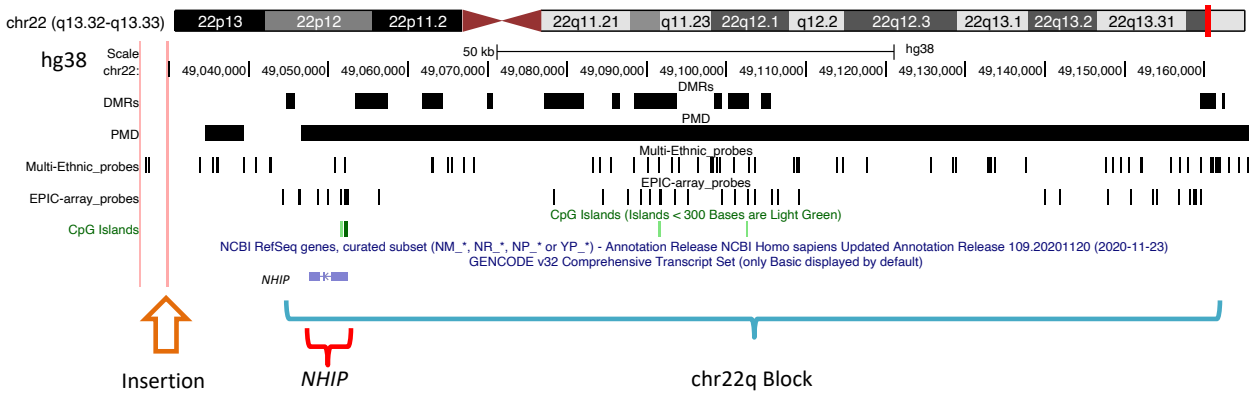
NHIP location was queried regarding of various brain region using the Genotype-Tissue Expression (GTEx) database (357,361). Various brain regions shown polymorphic expression level on *NHIP* with high expression in cortex and low expression in cerebellum.

Species	Query start	Query end	Length	Score	E-val	%ID	Species
Human (Homo sapiens)	1	8609	8609	17016	0	100	Primates
Gorilla (Gorilla gorilla gorilla)	1	8609	8672	13460	0	97.16	Primates
Bonobo (Pan paniscus)	1	8609	8669	12011	0	97.87	Primates
Chimpanzee (Pan troglodytes)	1	8609	8677	11964	0	97.78	Primates
Orangutan (Pongo abelii)	1	8609	8638	7979	0	94.55	Primates
Gibbon (Nomascus leucogenys)	40	8609	7663	7232	0	94.39	Primates
Crab-eating macaque (Macaca fascicularis)	46	8609	8570	6327	0	90.15	Primates
Gelada (Theropithecus gelada)	46	8609	8609	6327	0	89.98	Primates
Olive baboon (Papio anubis)	46	8609	8563	6275	0	90.02	Primates
Pig-tailed macaque (Macaca nemestrina)	46	8609	8609	6260	0	89.98	Primates
Macaque (Macaca mulatta)	46	8077	8567	6204	0	89.79	Primates
Drill (Mandrillus leucophaeus)	46	8077	7979	6147	0	89.74	Primates
Black snub-nosed monkey (Rhinopithecus bieti)	40	7499	9243	6078	0	89.36	Primates
Angola colobus (Colobus angolensis palliatus)	40	8601	8631	6060	0	89.33	Primates
Sooty mangabey (Cercocebus atys)	46	8609	8580	6032	0	89.3	Primates
Golden snub-nosed monkey (Rhinopithecus roxellana)	40	8609	9014	6024	0	89.23	Primates
Vervet-AGM (Chlorocebus sabaeus)	40	8609	8601	5340	0	90.66	Primates
Ugandan red Colobus (Piliocolobus tephrosceles)	980	8609	7628	5178	0	90.19	Primates
Ma's night monkey (Aotus nancymaae)	977	7176	7351	2170	0	84.83	Primates
Bolivian squirrel monkey (Saimiri boliviensis boliviensis)	3697	7499	3576	1391	0	86.14	Primates
Capuchin (Cebus capucinus imitator)	46	7176	6416	1077	0	84.13	Primates
Marmoset (Callithrix jacchus)	980	1808	852	557	3.00E-155	83.45	Primates
Tarsier (Carlito syrichta)	1832	6760	1596	557	3.00E-155	83.9	Primates
American bison (Bison bison bison)	7250	7508	259	251	5.00E-63	87.26	Ungulates
Greater bamboo lemur (Prolemur simus)	7237	7476	241	229	1.00E-56	87.14	Primates
American black bear (Ursus americanus)	1894	2072	179	204	8.00E-49	89.39	Carnivores
Polar bear (Ursus maritimus)	1894	2072	179	204	7.00E-49	89.39	Carnivores
Alpaca (Vicugna pacos)	7252	7464	214	192	3.00E-45	86.45	Ungulates
Panda (Ailuropoda melanoleuca)	1894	2072	179	188	4.00E-44	88.27	Carnivores
Agassiz's desert tortoise (Gopherus agassizii)	7250	7491	243	186	2.00E-43	84.77	Birds and Re
Coquerel's sifaka (Propithecus coquereli)	7247	7463	217	184	8.00E-43	85.71	Primates
Cat (Felis catus)	7252	7508	265	178	4.00E-41	84.53	Carnivores
Donkey (Equus asinus asinus)	1862	2058	197	166	1.00E-37	85.79	Ungulates
Horse (Equus caballus)	1862	2058	198	164	6.00E-37	85.86	Ungulates
Common canary (Serinus canaria)	7251	7414	164	164	3.00E-37	87.8	Birds and Re
Goat (Capra hircus)	1906	2034	129	160	1.00E-35	90.7	Ungulates
Rat (Rattus norvegicus)	1899	2038	140	158	4.00E-35	89.29	Rats & Mice

Supplementary Figure 4.12

NHIP was highly conserved in primates by blast search.

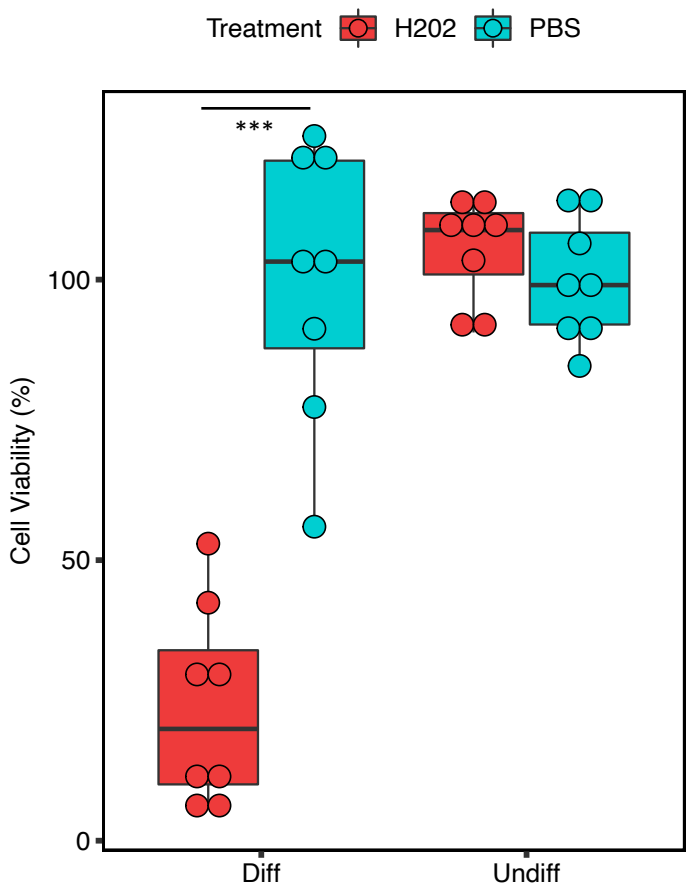
NHIP DNA sequence was extracted to blast search against vertebrate databases (407).



Supplementary Figure 4.13

22q13.33 hypomethylated block location relative to PMD, multi-ethnic SNP probes, and EPIC-array probe location.

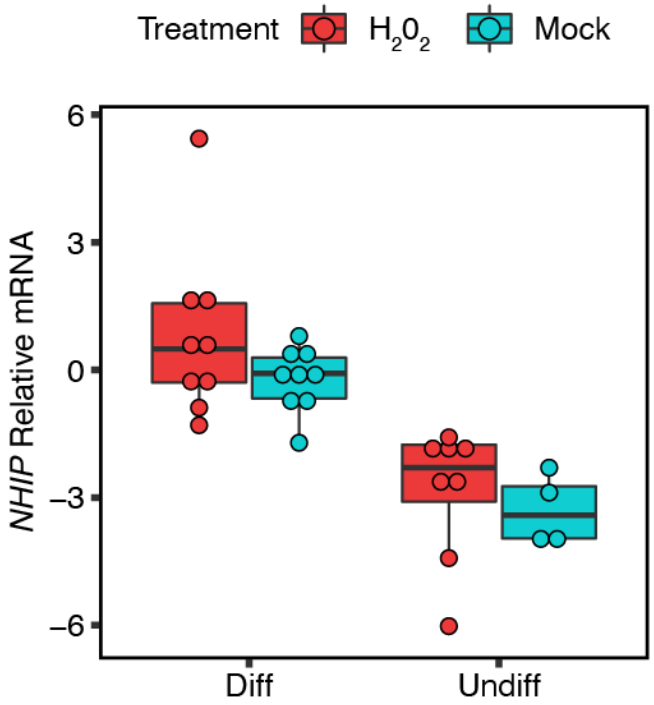
UCSC genome browser (genome assembly: hg38) with the22q13.33 hypomethylated block (blue), the insertion location (orange), and *NHIP* (red). 12 DMRs located inside the 22q13.33 hypomethylated block were shown in the DMRs track. PMD location were illustrated in the PMD track (161). SNP-array probe location was extracted from Illumina Infinium multi-ethnic genotyping array (447). EPIC-array probe location was extracted from Illumina Infinium MethylationEPIC (434).



Supplementary Figure 4.14

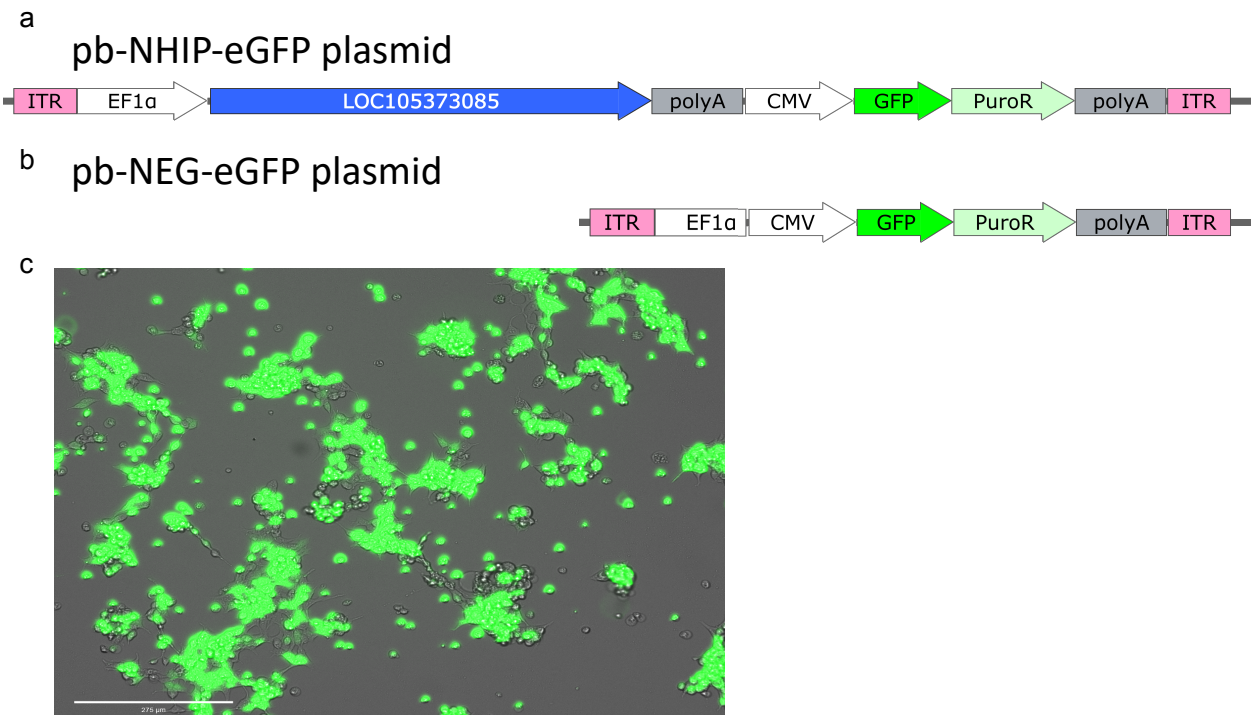
Differentiated LUHMES cells are more sensitive to oxidative stress than undifferentiated LUHMES cells.

Cell viabilities were detected by fluorescent intensity of CellTiter luminescent assay. The relative levels in treated cells were generated by comparing with the average of PBS-treated cells for both differentiated and undifferentiated experiments.



Supplementary Figure 4.15

NHIP levels in differentiated or undifferentiated LUHMES cells before or after treatment with 10nM H_2O_2. In differentiated LUHMES cells, H_2O_2-treated cells increased *NHIP* levels.



Supplementary Figure 4.16

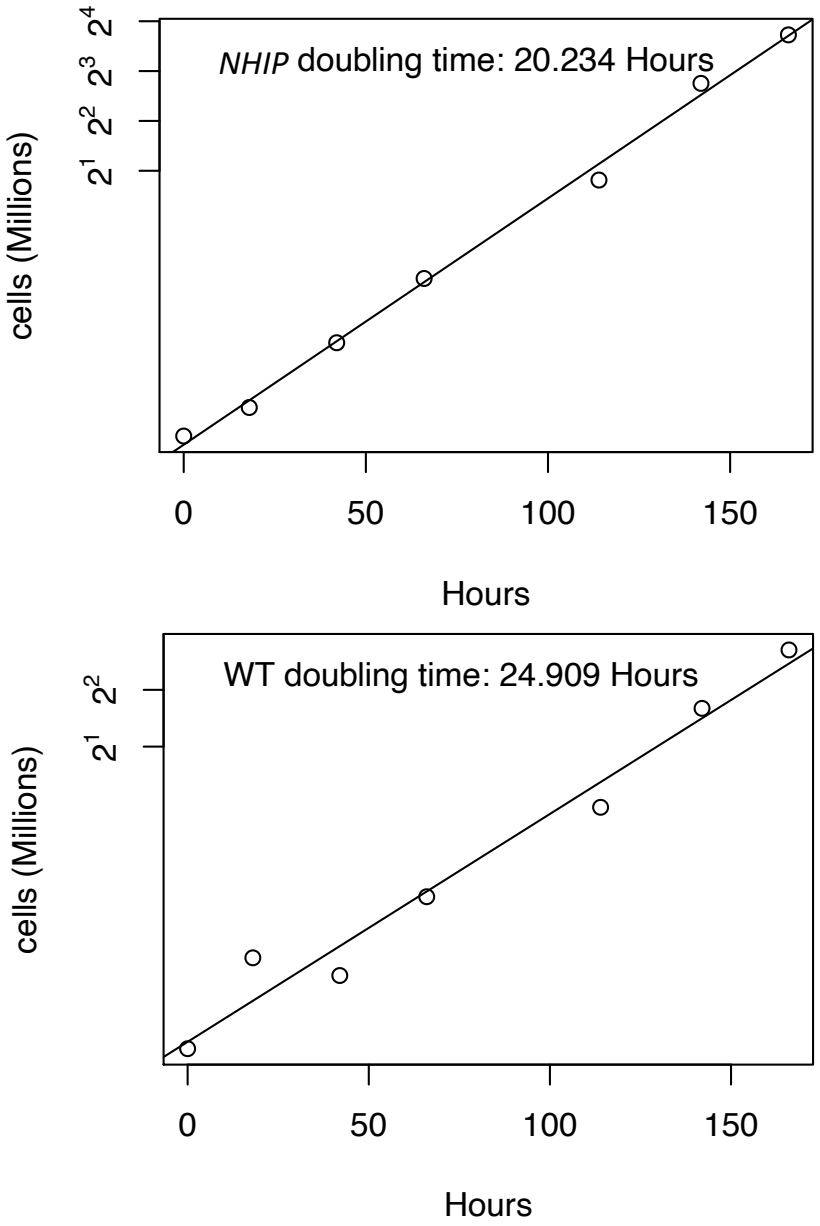
Cells were successfully transfected with overexpressed *NHIP*.

Schematic diagram of vector used for HEK293T cell transfection.

(**a**) Plasmid with *NHIP*, pb-NHIP-eGFP. EF-1α as promoter for *NHIP* and CMB as promoter for eGFP fused with puromycin resistant gene.

(**b**) Negative control plasmid without *NHIP*, pb-NEG-eGFP. Remove *NHIP* from pb-NHIP-eGFP and remove the rest of plasmid structure.

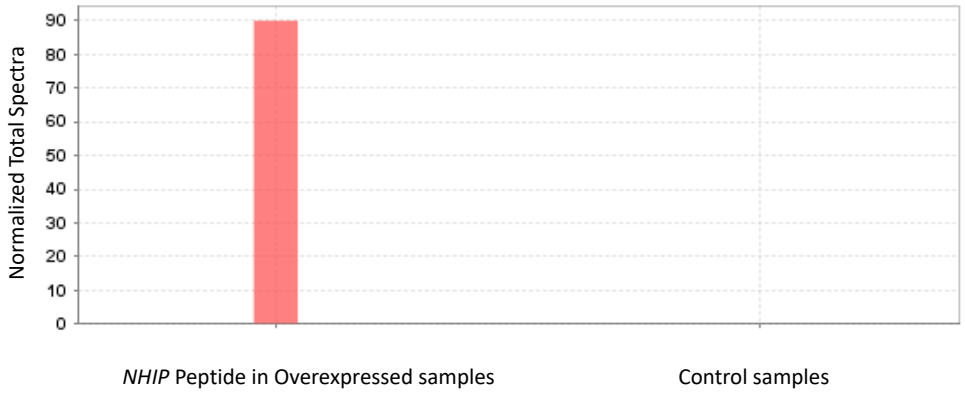
(**c**) Successful transfection image on HEK293T cell using EVOS microscopy.



Supplementary Figure 4.17

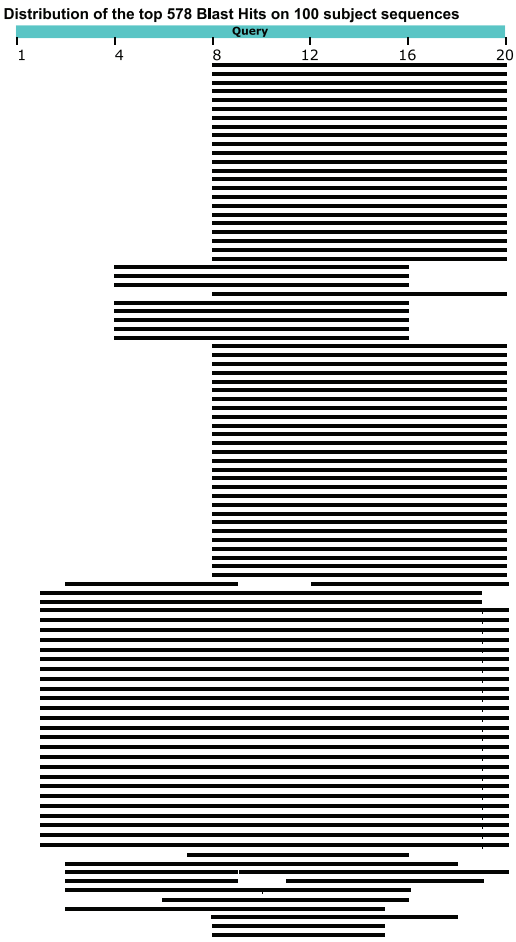
Cells with overexpressed *NHIP* significantly changed the cell cycle.

For overexpressed cell phenotype, cells were significantly shortened the doubling time after overexpressed *NHIP* in HEK293T cell (overexpression cells doubling time = 20.23 h, wide type cells doubling time = 24.91 h).



Supplementary Figure 4.18

Mass spectrometry of NHIP peptide and negative control.



BLAST® » **blastp suite** » results for RID-DBJE4WD201R

Your search parameters were adjusted to search for a short input sequence.
Your search is limited to records that include: Home sapiens (taxid:9606)

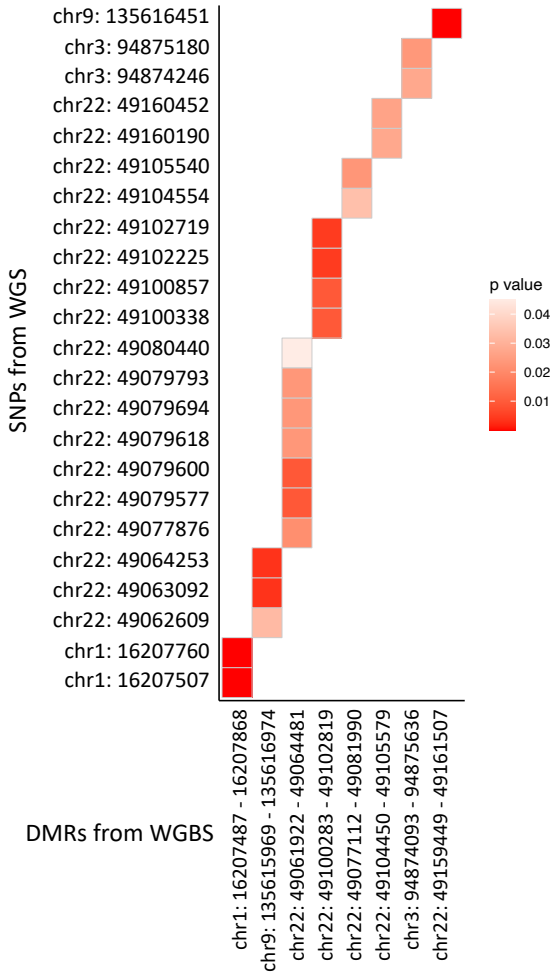
Job Title	Protein Sequence ...
RID	DBJE4WD201R Search expires on 06-03 11:38 am
Program	Quick BLASTP
Database	nr
Query ID	lcl|Query_38615
Description	None ...
Molecule type	amino acid
Query Length	20

Descriptions

Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
truncated breast and ovarian cancer susceptibility protein 2 [Homo sapiens]	26.9	26.9	60%	3.7	66.67%	AYD59793.1
truncated breast and ovarian cancer susceptibility protein 2 [Homo sapiens]	26.9	26.9	60%	3.8	66.67%	AYD59803.1
DNA repair-associated BRCA2 [Homo sapiens]	26.9	26.9	60%	3.8	66.67%	QCL11117.1
truncated breast and ovarian cancer susceptibility protein 2 [Homo sapiens]	26.9	26.9	60%	3.8	66.67%	AYD59769.1
truncated breast and ovarian cancer susceptibility protein 2 [Homo sapiens]	26.9	26.9	60%	3.8	66.67%	AYD59770.1
truncated breast and ovarian cancer susceptibility protein 2 [Homo sapiens]	26.9	26.9	60%	3.8	66.67%	AYD59771.1
truncated breast and ovarian cancer susceptibility protein 2 [Homo sapiens]	26.9	26.9	60%	3.8	66.67%	AYD59772.1
truncated breast and ovarian cancer susceptibility protein 2 [Homo sapiens]	26.9	26.9	60%	3.8	66.67%	AYD59773.1
truncated breast and ovarian cancer susceptibility protein 2 [Homo sapiens]	26.9	26.9	60%	3.8	66.67%	QGJ03648.1
truncated breast and ovarian cancer susceptibility protein 2 [Homo sapiens]	26.9	26.9	60%	3.8	66.67%	AYD59774.1
truncated breast and ovarian cancer susceptibility protein 2 [Homo sapiens]	26.9	26.9	60%	3.8	66.67%	AYD59775.1
truncated breast and ovarian cancer susceptibility protein 2 [Homo sapiens]	26.9	26.9	60%	3.8	66.67%	AYD59776.1
truncated breast and ovarian cancer susceptibility protein 2 [Homo sapiens]	26.9	26.9	60%	3.8	66.67%	QGJ03647.1
truncated breast and ovarian cancer susceptibility protein 2 [Homo sapiens]	26.9	26.9	60%	3.8	66.67%	AYD59777.1
truncated breast and ovarian cancer susceptibility protein 2 [Homo sapiens]	26.9	26.9	60%	3.8	66.67%	AYD59778.1
truncated BRCA2 DNA repair-associated protein [Homo sapiens]	26.9	26.9	60%	3.8	66.67%	QIM56776.1
truncated breast and ovarian cancer susceptibility protein 2 [Homo sapiens]	26.9	26.9	60%	3.8	66.67%	AYD59779.1
truncated breast and ovarian cancer susceptibility protein 2 [Homo sapiens]	26.9	26.9	60%	3.8	66.67%	AYD59780.1
truncated breast and ovarian cancer susceptibility protein 2 [Homo sapiens]	26.9	26.9	60%	3.8	66.67%	AYD59781.1
truncated breast and ovarian cancer susceptibility protein 2 [Homo sapiens]	26.9	26.9	60%	3.8	66.67%	AYD59782.1
DNA repair-associated BRCA2 [Homo sapiens]	26.9	26.9	60%	3.8	66.67%	QCL11116.1
truncated breast and ovarian cancer susceptibility protein 2 [Homo sapiens]	26.9	26.9	60%	3.8	66.67%	AYD59784.1
truncated breast and ovarian cancer susceptibility protein 2 [Homo sapiens]	26.9	26.9	60%	3.8	66.67%	QGJ03646.1
unnamed protein product [Homo sapiens]	26.9	26.9	60%	3.8	66.67%	BAG54725.1
chromodomain-helicase-DNA-binding protein 4 isoform 3 [Homo sapiens]	26.9	41.1	60%	3.8	66.67%	NP_001350535.1
chromodomain-helicase-DNA-binding protein 4 isoform 2 [Homo sapiens]	26.9	41.1	60%	3.8	66.67%	NP_001284482.1
truncated breast and ovarian cancer susceptibility protein 2 [Homo sapiens]	26.9	26.9	60%	3.8	66.67%	AYD59785.1
chromodomain helicase DNA binding protein 4, isoform CRA_c [Homo sapiens]	26.9	41.1	60%	3.8	66.67%	EAW88779.1
chromodomain helicase DNA binding protein 4, isoform CRA_a [Homo sapiens]	26.9	41.1	60%	3.8	66.67%	EAW88777.1
chromodomain helicase DNA binding protein 4, isoform CRA_b [Homo sapiens]	26.9	41.1	60%	3.8	66.67%	EAW88778.1
chromodomain-helicase-DNA-binding protein 4 isoform 1 [Homo sapiens]	26.9	41.1	60%	3.8	66.67%	NP_001264.2
CHD4 protein [Homo sapiens]	26.9	41.1	60%	3.8	66.67%	AAH38596.1

Supplementary Figure 4.19

NHIP peptide in blastp search with hit on BRCA2 AND CHD4.

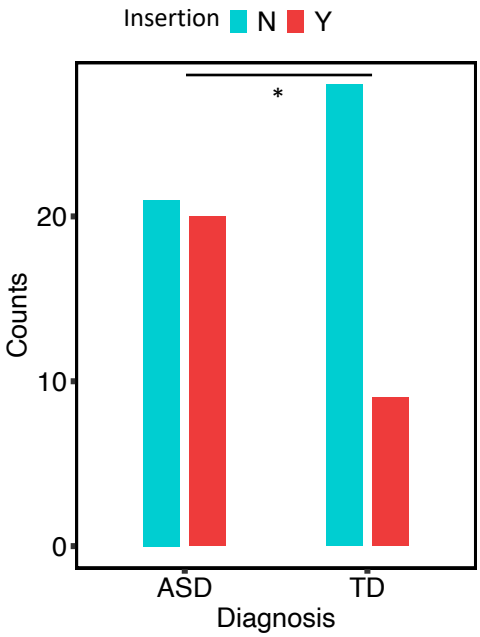


Supplementary Figure 4.20

Eight DMRs percent methylation were significant associated with *in cis* SNP.

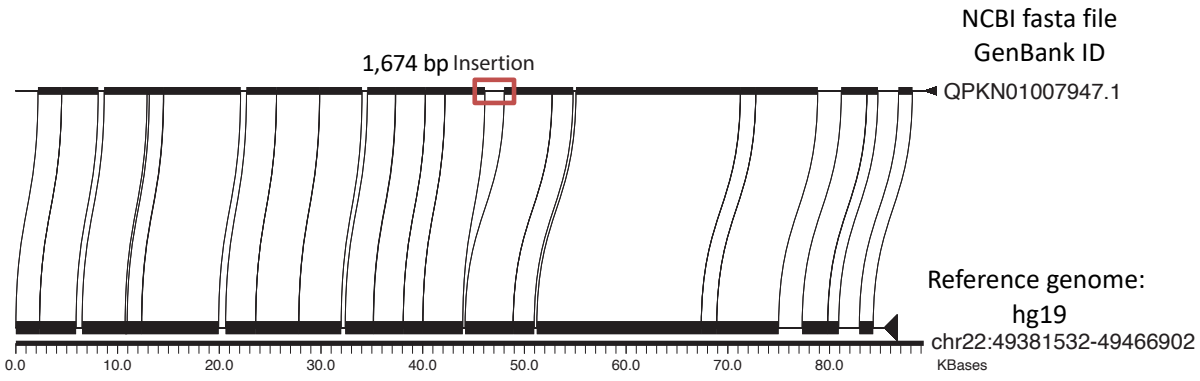
Percent methylation for each DMR was extracted from WGBS on average smooth methylation of the DMR region. WGS data was used to extract the *in cis* SNPs that located inside the DMRs. Significant association tests were done between the *in cis* SNPs and DMRs using linear regression on *p*-value < 0.05 (discovery group, ASD n = 41, TD n = 37).

Diagnosis \ Insertion	Yes	No
ASD	20	21
TD	9	27



Supplementary Figure 4.21

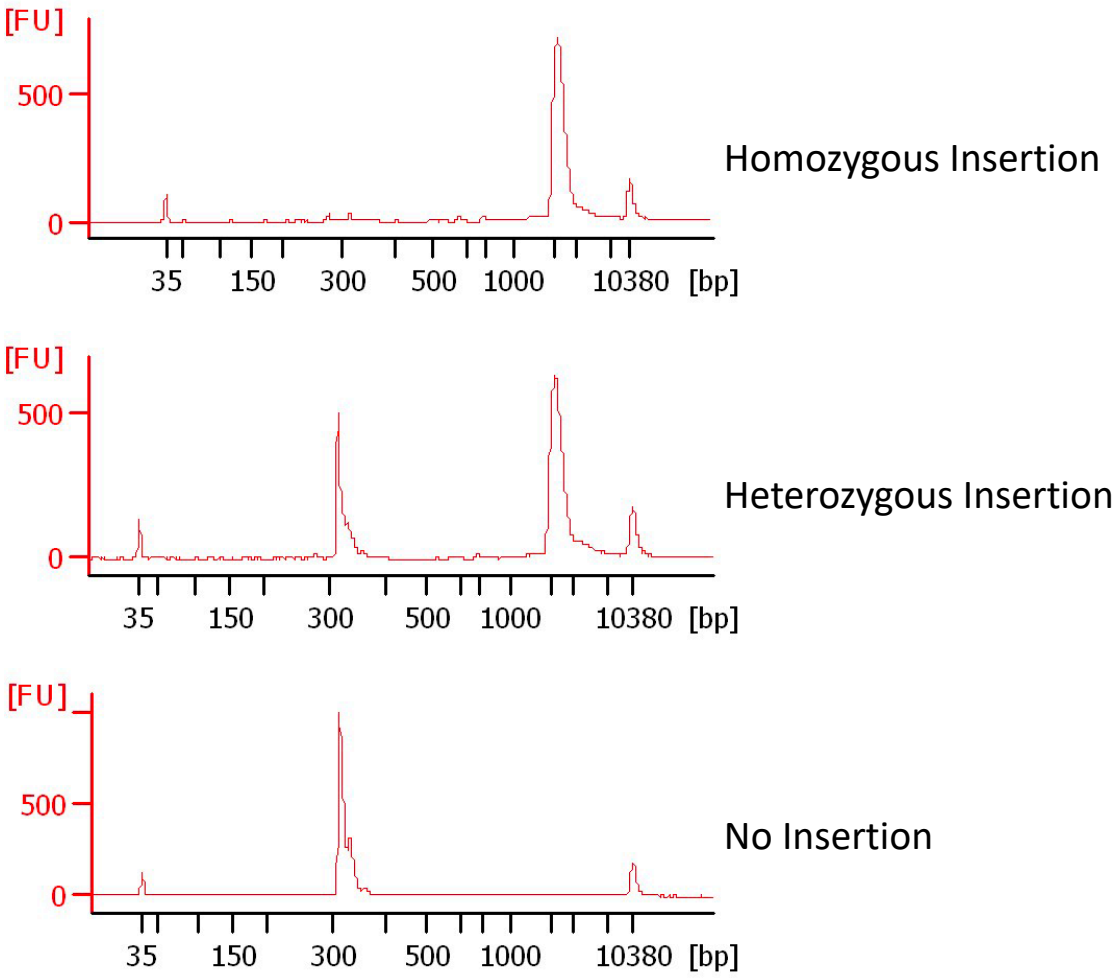
Count matrix was based on the insertion with ASD and TD samples shown more individuals with the insertion in ASD compared to TD (discovery group, ASD n = 41, TD n = 37) (chi-square test, *p*-value = 0.045).



Supplementary Figure 4.22

Insertion was characterized with PacBio long-read sequencing.

QPKN01007947.1 contig (363) mapped to reference genome chr22: 49,381,532 – 49,466,902 with Miropeat (408) used for visualization. The orange box shown the clear insertion compared the QPKN01007947.1 contig with the reference genome on 1,674 bp in length.



Supplementary Figure 4.23

Insertion profiles were validated using PCR genotyping on the same sample.

Genotyping primer designed on the two side of the insertion to capture the difference based on the PCR size. Agarose gel electrophoresis and bioanalyzer shown the same result (discovery group, ASD n = 41, TD n = 37).

QPKN01007947.1:46099-47772 (SVG Plot; Alignments; Masked)

Name	From	To	Name	From	To	Class	Dir	Sim	Pos/Mm:Ts	Score
QPKN01007947.1:46099-47772	59	1673	SVA_A	1	1640	NonLTR/SINE	d	0.9514	1.4043	10850

Masked Sequence

```
>QPKN01007947.1:46099-47772
TAAGAAAACTCCTGCTCTCCCTCCTCTCCCTCTCCCTCCTCTCCCTCTCCCTCTCCCTXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXA
```

Local Alignments*

Name	From	To	Name	From	To	Dir	Sim	Pos/Mm:Ts	Score
QPKN01007947.1:46099-47772	59	1673	SVA_A	1	1640	d	0.9514	1.4043	10850

```
 59 CTCCCTCTCCCACGGTCTCCCTCTCCCCACGGTCTCCCTCTCCCTCTCTTTCCACGGTCTCCCCCTGAT 128
    |||||||||||-----|| |||||||||||:|||||||||||||||||||||||||||||||||:|||||
  1 CTCCCTCTCCC-----TCACCCTCTCCCCATGGTCTCCCTCTCCCTCTCTTTCCACGGTCTCCCTCTGAT 65

129 GCTGAGCCGAAGCTGGACTGTACTGCTGCCATCTCGGCTCACTGCAACCTCCCTGCCTGATTCTCCTGCC 198
    ||:||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
 66 GCCGAGCCGAAGCTGGACGGTACTGCTGCCATCTCGGCTCACTGCAACCTCCCTGCCTGATTCTCCTGCC 135

199 TCAGCCTGCCGAGTGCCTGCGATTGCAGGCGCGTGCCGCCACGCCTGACTGGTTTTCGTATTTTTTTGGT 268
    |||||:|||||||||||||||||||||||||||:|||||||||||||||||||||||||||||| ||:||
136 TCAGCTTGCCGAGTGCCTGCGATTGCAGGCGCGCGCCGCCACGCCTGACTGGTTTTCGTATTTTGTTAGT 205

269 GGAGACGGGGTTTCGCTGTGTTGGCCGGGCTGGTCTCCAGCTCCTAATCACGAGTGATCCGCCAGCCTTG 338
    |||||||||||||||||||||||||||||||||||||||||||||||:|:||||||||||:|||||||:|
206 GGAGACGGGGTTTCGCTGTGTTGGCCGGGCTGGTCTCCAGCTCCTAACCGCGAGTGATCCACCAGCCTCG 275

339 GCCTCCCGAGGTGCCGGGATTGCAGACGGAGTCTCGTTCACTCAGTGCTCAATGGTGCCCAGGCTGGAGT 408
    ||||||||||||||:|||||||||||||||||||||||||||||||||||||||:|||||||||||||||
276 GCCTCCCGAGGTGCTGGGATTGCAGACGGAGTCTCGTTCACTCAGTGCTCAATGATGCCCAGGCTGGAGT 345

409 GCAGTGGCGTGATCTCGGCTTGCTACAACCTCCACCTCCCAGCCGCCTGCCTTGGCCTCCCAAAGTGCCG 478
    ||||||||||||||||||||:|||||||||||||||||||||| ||||||||||||||||||||||||||
346 GCAGTGGCGTGATCTCGGCTCGCTACAACCTCCACCTCCCAGCAGCCTGCCTTGGCCTCCCAAAGTGCCG 415

479 AGATTGCAGCCTCTGCCCGGCCGCCACCCCGTCTGGGAAGTGAGGAGCGTCTCTGCCTGGCCGCCCATCG 548
    |||||||||||||||||||||||||||||||||||||||||||||||:|||||:||||||||:|||||||
416 AGATTGCAGCCTCTGCCCGGCCGCCACCCCGTCTGGGAAGTGAGGAGTGTCTCCGCCTGGCCACCCATCG 485

549 TCTGGGACGTGAGGAGCCCCTCTGCCTGGCTGCCCAGTCTGGAAAGTGAGGAGCGTCTCTGCCCGGCCGC 618
    |||||||:|||||||||-------------------------------------|||||||||| |||||
486 TCTGGGATGTGAGGAGC-------------------------------------GTCTCTGCCCTGCCGC 518

619 CATCCCATCTAGGAAGTGAGGAGCGCCTCTGCCAGGCCGCCCATCGTCTGAGATGTGGGGAGCGCCTCTG 688
    |----------------------------------------||||||||||||||||||||||:||||||
519 C----------------------------------------CATCGTCTGAGATGTGGGGAGCACCTCTG 548

689 CCCTGCCACCCCGTCTGGGATGTGAGGAGCGTCTCTGCCCGGCCGCCCCGTCTGAGAAGTGAGGAGACCC 758
    ||| |||:|||||||:||||||||||||||||| ||||||||||||||||||||||||||||||||||||
549 CCCGGCCGCCCCGTCCGGGATGTGAGGAGCGTCGCTGCCCGGCCGCCCCGTCTGAGAAGTGAGGAGACCC 618

759 TCTGCCTGGCAACCGCCCCGTCTGAGAAGTGAGGAGCCCCTCCGCCCGGCAGCCACACCCTCTGAGAAGT 828
    ||||||||||||||||:||:||||||||||||||||||||||||||||||||||:| |: ||||||||||
619 TCTGCCTGGCAACCGCTCCATCTGAGAAGTGAGGAGCCCCTCCGCCCGGCAGCCGCCCTGTCTGAGAAGT 688

829 GAGGAGCGTCTCTGCCTGGCAGCCACCCCGTCTGGGAGGGAGGTGGGGG--TCAGCCCCCTGCCCCGCCA 896

 965 GGGGCGCCTCTGCCCGGCCGCCCCTACTGGGAAGTGAGGAGCC--------------------------- 1007
     |||||||||||||||||||||||||||||||||||||||||||---------------------------
 829 GGGGCGCCTCTGCCCGGCCGCCCCTACTGGGAAGTGAGGAGCCACTTTGCCCGGCCAGCCACTCTGTCCG 898

1008 ----------------------CCTCTGCCTGGCCAGCCGCCCCGTCGGGGAGGGAGGTGGGGGA---- 1051
     ----------------------||:|:|||:||||||||||||||| |||||||||||||||||||----
 899 GGAGGGAGGTGGGGGGGTCAGCCCCCCGCCCGGCCAGCCGCCCCGTCCGGGAGGGAGGTGGGGGGATCAG 968

1052 ----------------------------------------------CAGCCCCCCGCCCGGCCAGCCGCC 1075
     ----------------------------------------------||||||||||||||||||||||||
 969 CCCCCCGCCCAGCCAGCCGCCCCGTCCGGGAGGGAGGTGGGGGGTCAGCCCCCCGCCCGGCCAGCCGCC 1038

1076 CCGTCCGGGAGGTGAGGGGCGCCTCTGCCCGGCCGCCCCTACTGGGAAGTGAGGAGCCCCTCTGCCCGGC 1145
     |:|||||||||||||||||||||||||||||||||| ||||||||:||||||||||||||||||||||||
1039 CTGTCCGGGAGGTGAGGGGCGCCTCTGCCCGGCCGCGCCTACTGGAAAGTGAGGAGCCCCTCTGCCCGGC 1108

1146 CACCACCCCGTCTGGGAGGTGTACTCAACAGCTCATTGAGAACGGGCCATGATGACAATGGCGGTTTTGT 1215
     ||||||||||||||||||||||:|:||||||||||||||||| |||||||||||||||||||||||||||
1109 CACCACCCCGTCTGGGAGGTGTGCCCAACAGCTCATTGAGAAGGGGCCATGATGACAATGGCGGTTTTGT 1178

1216 GGAATAGAAAGGGGGGAAAGGTGGGGAAAAGATTGAGAAATCGGATGGTTGCCGTGTCTGTGTAGAAAGA 1285
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1179 GGAATAGAAAGGGGGGAAAGGTGGGGAAAAGATTGAGAAATCGGATGGTTGCCGTGTCTGTGTAGAAAGA 1248

1286 GGTAGACATGGGAGACTTTTCATTTTGTTCTGTACTAAGAAAAATTCTTCTGCCTTGGGATCCTGTTGAT 1355
     ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1249 GGTAGACCTGGGAGACTTTTCATTTTGTTCTGTACTAAGAAAAATTCTTCTGCCTTGGGATCCTGTTGAT 1318

1356 CTGTGACCTTACCCCCAACCCTGTGCTCTCTGAAACATGTGCTGTGTCCACTCAGGGTTGAATGGATTAA 1425
     | |||||||||||||||||||||||||||||||||||||||||||||:||||||||||||||||||||||||
1319 CGGTGACCTTACCCCCAACCCTGTGCTCTCTGAAACATGTGCTGTATCCACTCAGGGTTGAATGGATTAA 1388

1426 GGGTGGTGCAAGATGTGCTTTGTTAAACAGATGCTTGAAGGCAGCATGCTCGTTAAGAGTCATCACCACT 1495
     |:|:||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
1389 GAGCGGTGCAAGATGTGCTTTGTTAAACAGATGCTTGAAGGCAGCATGCTCCTTAAGAGTCATCACCACT 1458

1496 CCCTAATCTCAAGTACCCAGGGACACAAACACTGCGGAAGGCCGCAGGGTCCTCTGCCTAGGAAAACCAG 1565
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1459 CCCTAATCTCAAGTACCCAGGGACACAAACACTGCGGAAGGCCGCAGGGTCCTCTGCCTAGGAAAACCAG 1528

1566 AGACCTTTGTTCACTT----ATCTGCTGACCTTCCCTCCACTATTGTCCTGTGACCCTGCCAAATCCCCC 1631
     ||||||||||||||||----||||||||||||||||||||||||||||||||||||||||||||||||||
1529 AGACCTTTGTTCACTTGTTTATCTGCTGACCTTCCCTCCACTATTGTCCTGTGACCCTGCCAAATCCCCC 1598

1632 TCTGCGAGAAACACCCAAGAATGATCAATAAAAAAAAAAAAAA 1673
     ||||:|||||||||||||||||||||||||||||||||||||
1599 TCTGTGAGAAACACCCAAGAATGATCAATAAAAAAAAAAAAAA 1640
```

<u>Supplementary Figure 4.24</u>

Insertion contained retrotransposon elements.

Insertion sequence was used to query against the collection of repeat databases with default settings.

Structural variant: INS_22_115103

Dataset gnomAD SVs v2.1

Filter Pass
Allele Count 14350
Allele Number 20614
Allele Frequency 0.6961
Quality score 999
Position 22:49425456
Size 1,314 bp
Class insertion (ME:SVA)
Evidence Split reads
Algorithms MELT

References

- UCSC

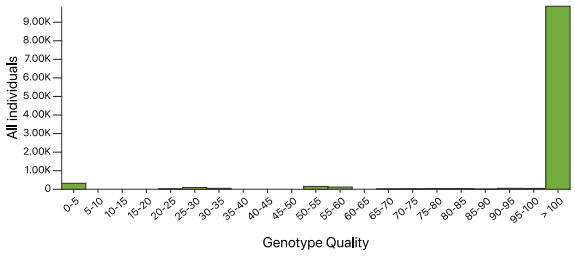
Report

- Report this variant
- Request additional information

Consequences

This variant has consequences in 0 genes.

Genotype Quality Metrics

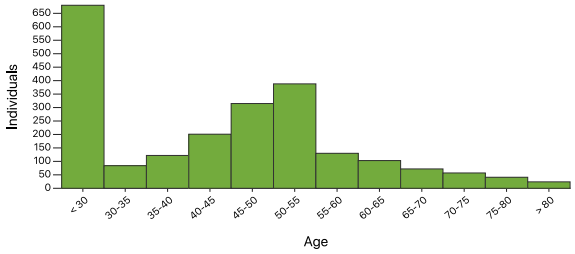


All | Variant Carriers

Population Frequencies

Population	Allele Count	Allele Number	Number of Homozygotes	Allele Frequency
Latino	1178	1470	463	0.8014
European	5437	7452	1908	0.7296
African/African-American	6121	9086	1988	0.6737
Other	125	190	40	0.6579
East Asian	1489	2416	404	0.6163
Female	7131	10132	2422	0.7038
Male	7189	10444	2369	0.6883
Total	**14350**	**20614**	**4803**	**0.6961**

Age Distribution

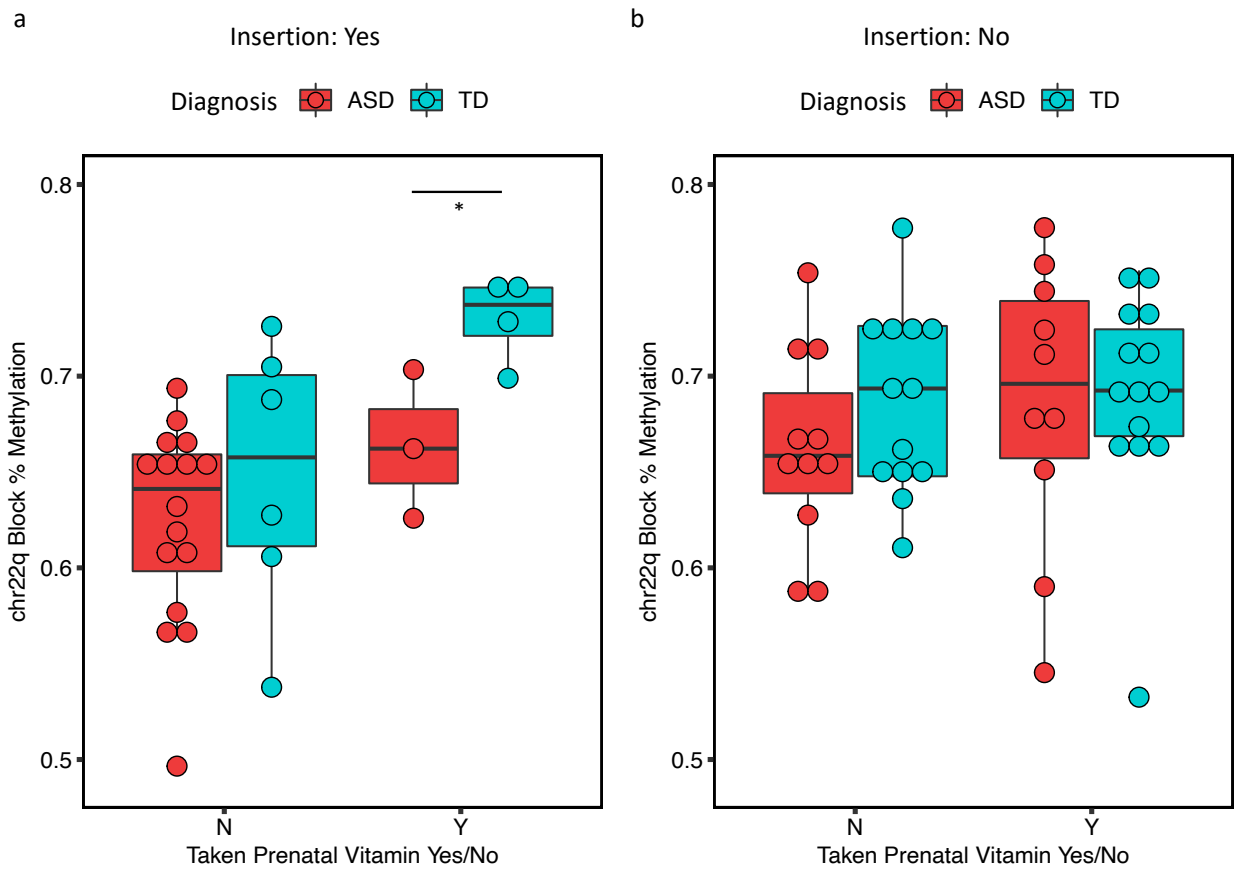


Heterozygous Variant Carriers

Supplementary Figure 4.25

Insertion has also identified in the gnomAD database.

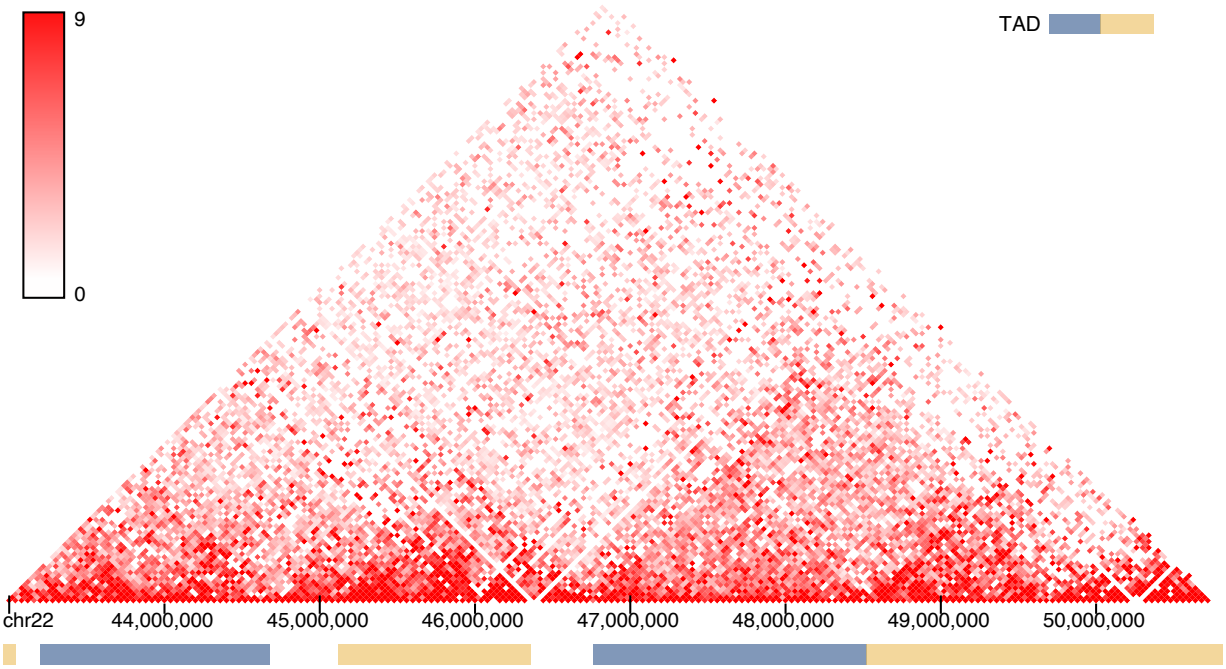
gnomAD dataset (365,366) snapshot of the insertion named as INS_22_115103, classified as insertion structural variant with SVA element.



Supplementary Figure 4.26

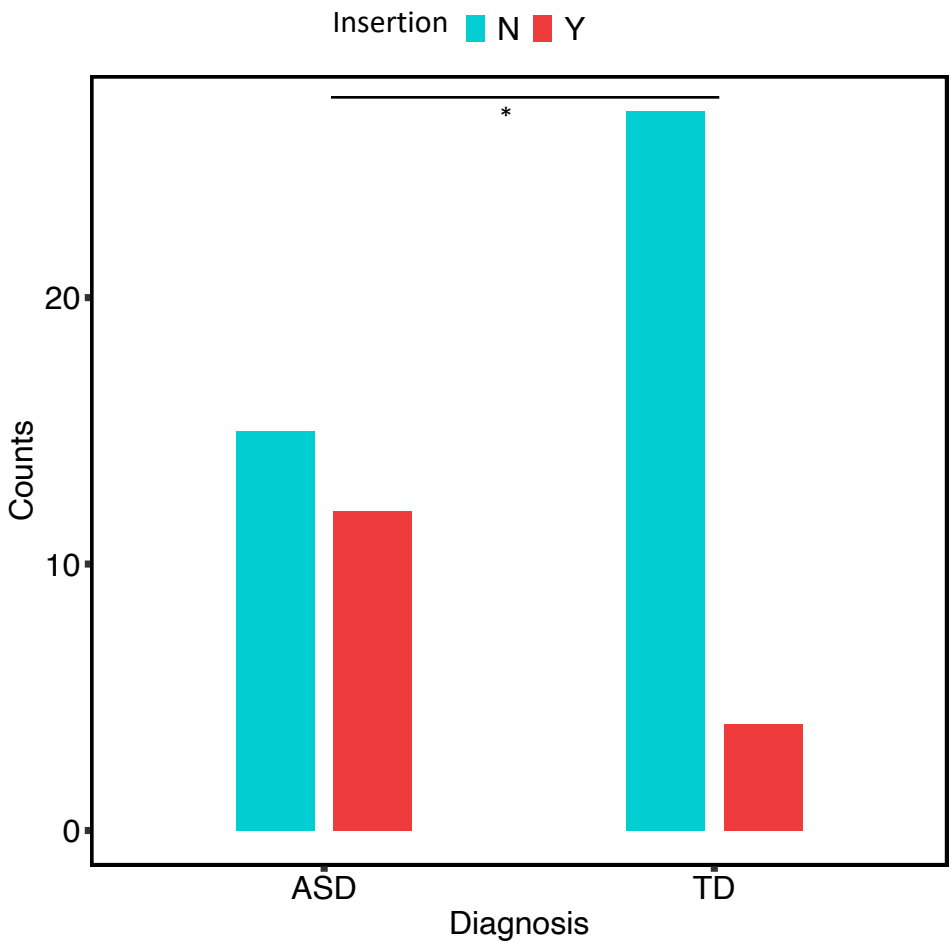
When separate into whether carry the insertion risk, individual methylation at 22q13.33 altered by P1 prenatal vitamin use when exposed to the genetic risk.

In individuals with the genetics risk, taking prenatal vitamins at P1 significantly altered 22q13.33 block methylation, which is in the same direction as TD (*p*-value = 0.034), which was not significant in not taking prenatal vitamins but still care the insertion (*p*-value = 0.400).



Supplementary Figure 4.27

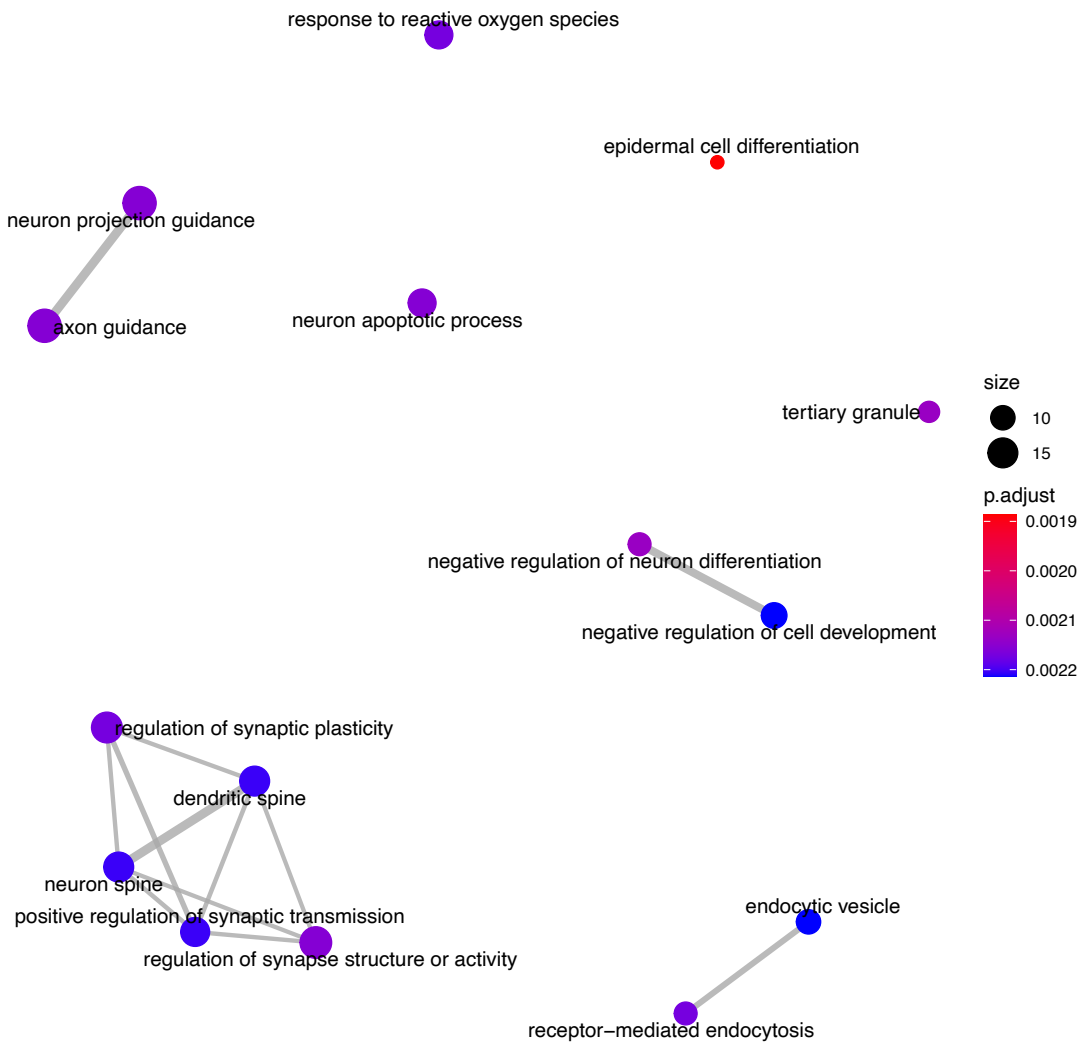
Topologically associated domain (TAD) from hippocampus on 22q13.31 – 22q13.33 regions (till telomeres), assembly in hg38 with resolution at 40 kb (369).



Supplementary Figure 4.28

The insertion had significant higher frequency in ASD compared with TD samples in brain.

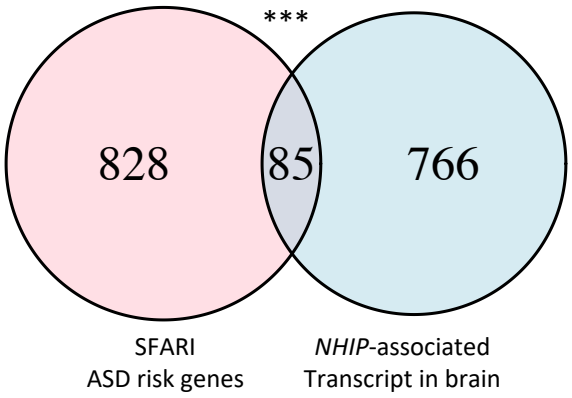
Count matrix was based on the insertion in brain and shown more individuals with insertion in ASD compared to TD (ASD n = 27, TD n =30) (chi-square test, *p*-value = 0.023).



Supplementary Figure 4.29

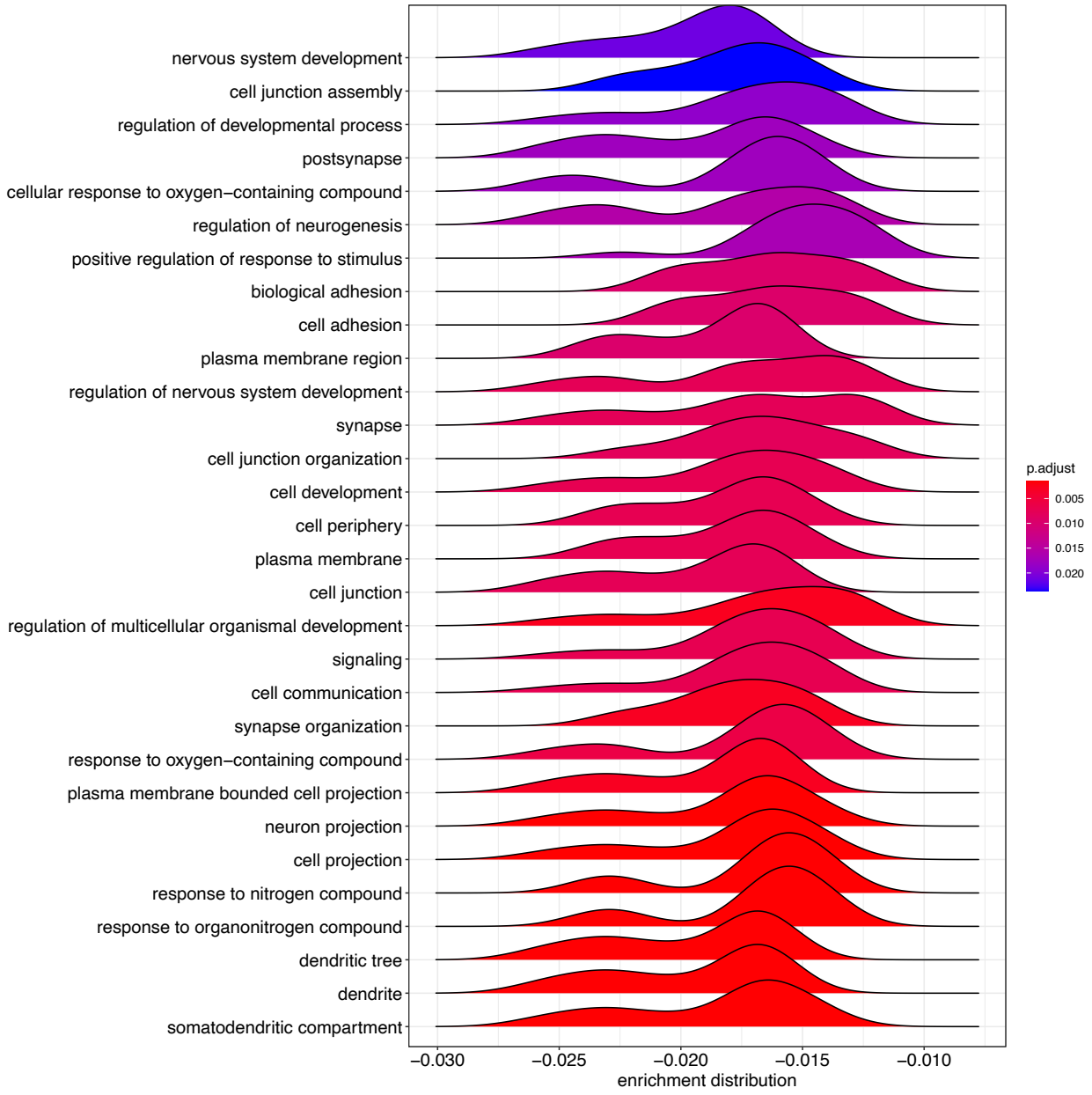
Genes significant associated with *NHIP* were enriched for neuronal functions.

GO terms analysis was based on the significant DGE associated with *NHIP* in brain. The enrichment map organized significant enriched terms into a network with edges connecting overlapping gene sets. Clustered gene sets were identified as functional modules with mutually overlapping genes together.



Supplementary Figure 4.30

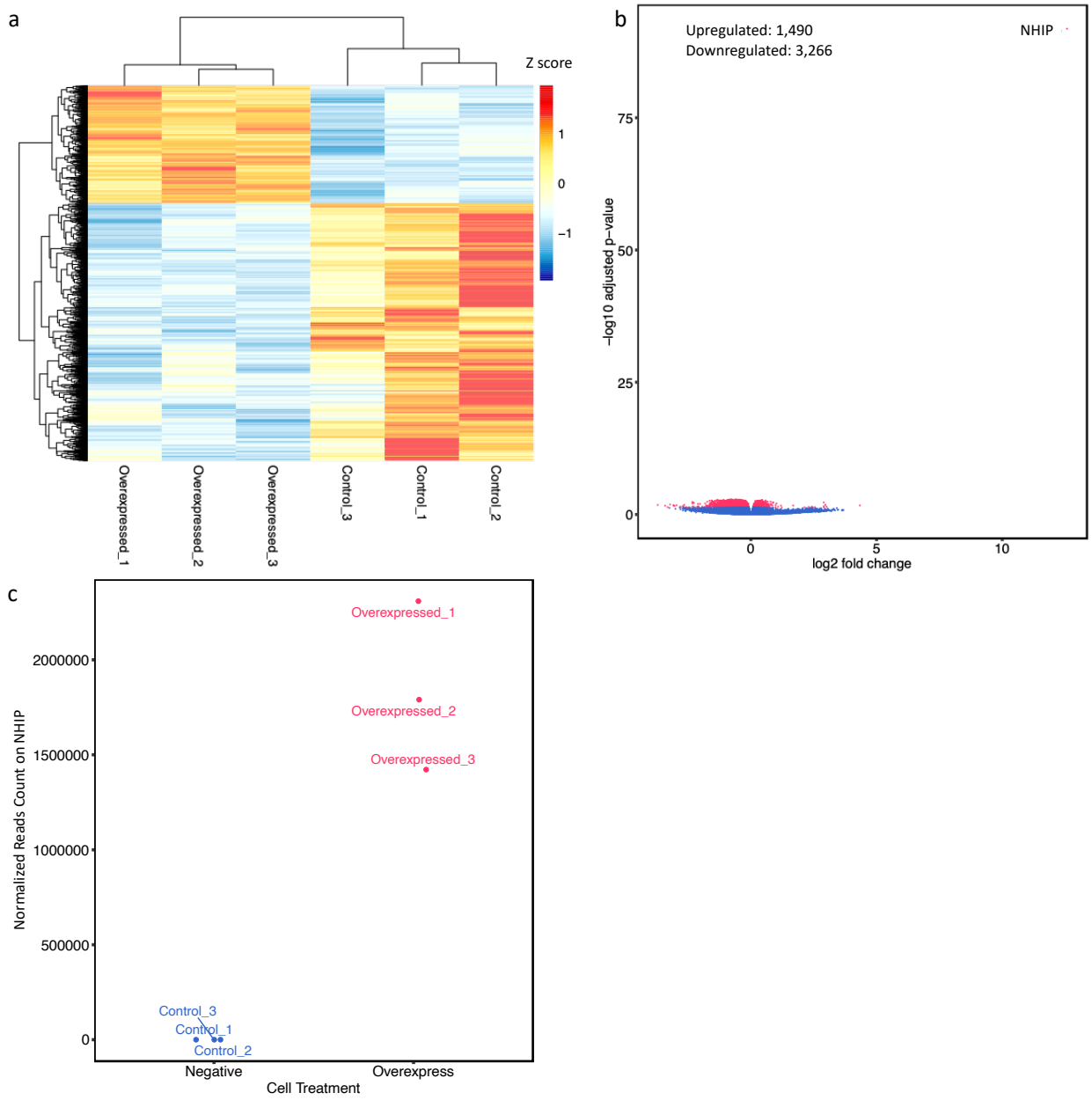
NHIP-associated genes in brain show a significant overlap with SFARI ASD risk genes (Fisher's exact test, *p*-value < 0.001).



Supplementary Figure 4.31

Genes in common between brain DGE and SFARI ASD genes shown enrichment at synapse, dendrite, and nervous system development.

GO term analysis was based on the common 85 genes in common between DGE associated with *NHIP* in brain and SFARI ASD genes. The ridgeline plot was generated using the frequency of fold change values per gene within each enrichment GO term set.



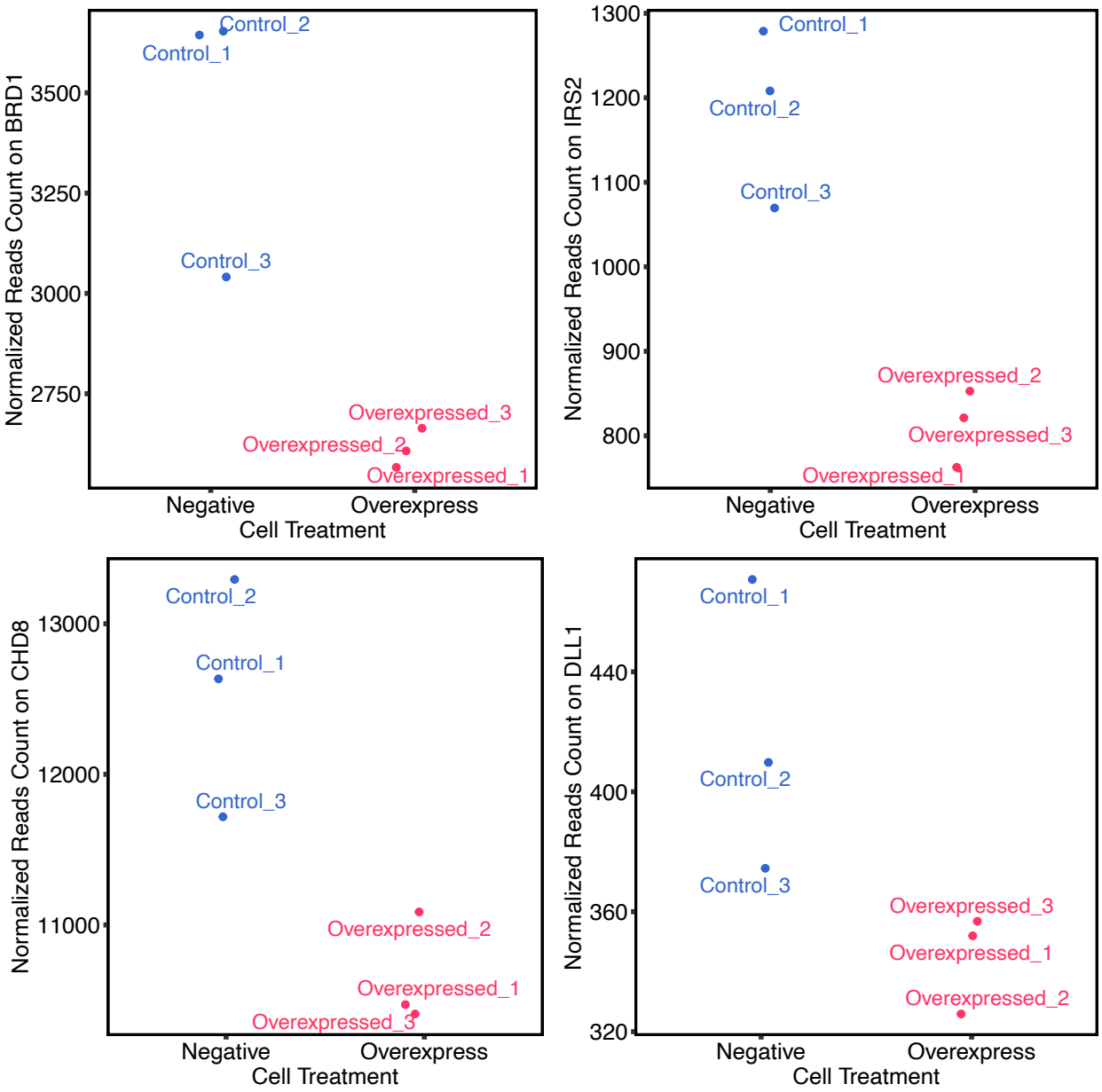
Supplementary Figure 4.32

RNA-seq from overexpressed *NHIP* and negative control cells, identifying 4,756 differential expressed genes passed at FDR significant with clear separation from treatment.

(**a**) Heatmap and hierarchical clustering of 3 overexpressed *NHIP* and 3 negative control with clear separation between treatment.

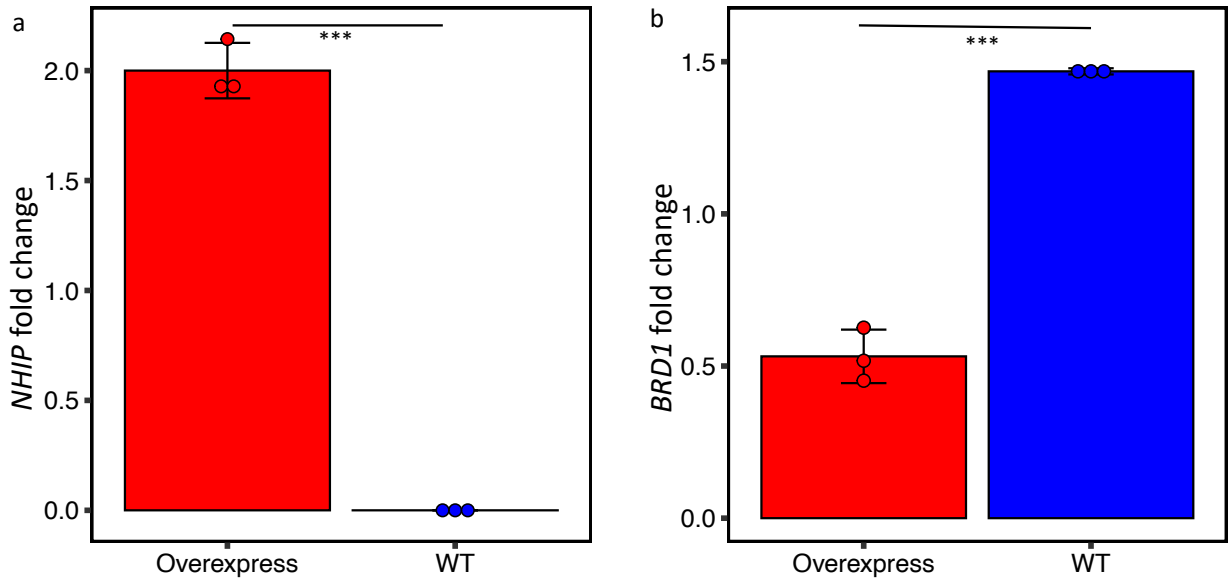
(**b**) Volcano plot shown the differential expressed gene in cell regarding of the treatment with 4,756 DGEs including 1,490 upregulated and 3,266 downregulated gene.

(**c**) Dot plot was plotted to illustrate the difference on *NHIP* on overexpressed and negative control cells transcript level.



Supplementary Figure 4.33

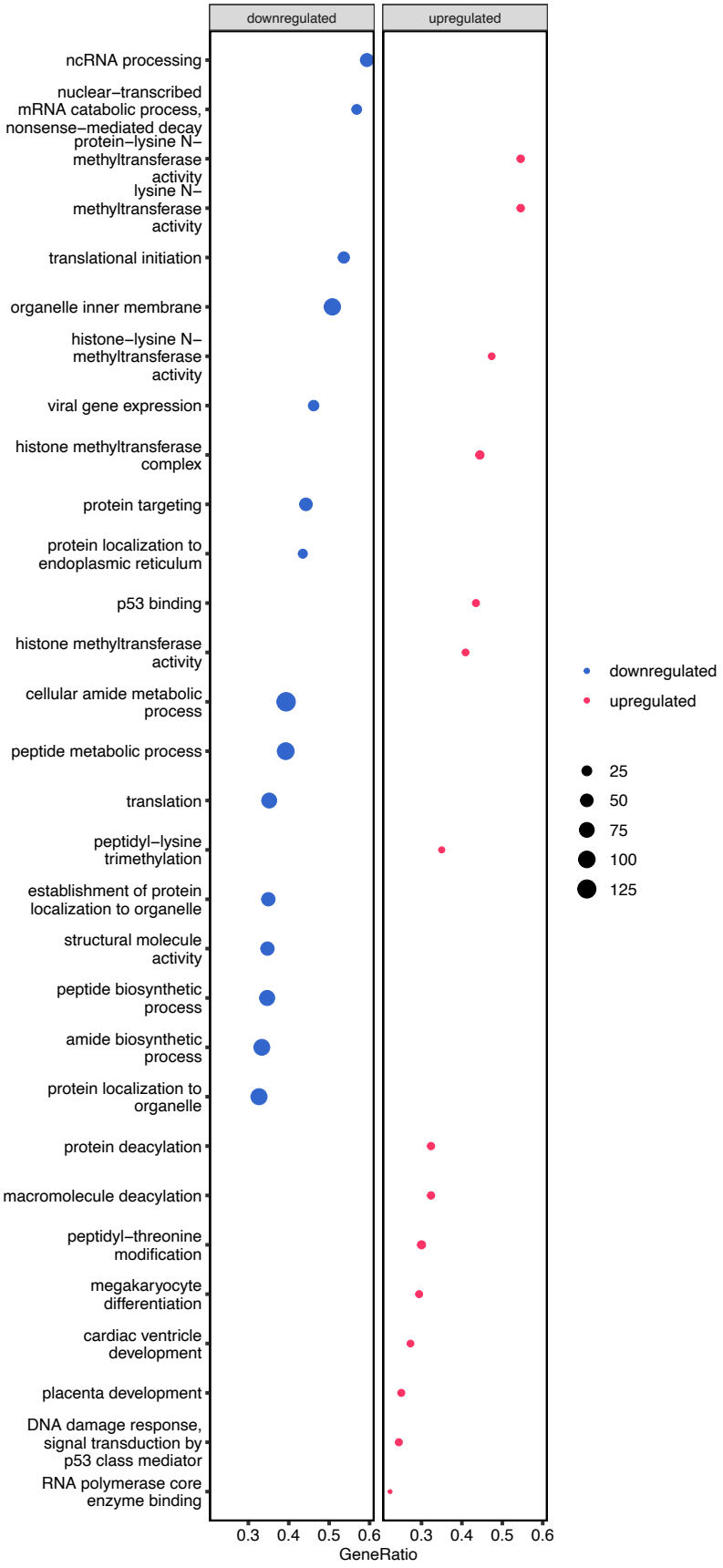
Several genes were downregulated with the overexpressed *NHIP* in HEK283T cells. Normalize reads count was plotted on the y-axis and the x-axis was separated by cell treatment on whether expressed *NHIP* or not. *BRD1* (adjusted *p*-value: 0.004), *IRS2* (adjusted *p*-value: 0.003), *CHD8* (adjusted *p*-value: 0.003), and *DLL1* (adjusted *p*-value: 0.048) were significant downregulated when overexpressed *NHIP*.



Supplementary Figure 4.34

Results on *NHIP* and *BRD1* were validated used RT-qPCR.

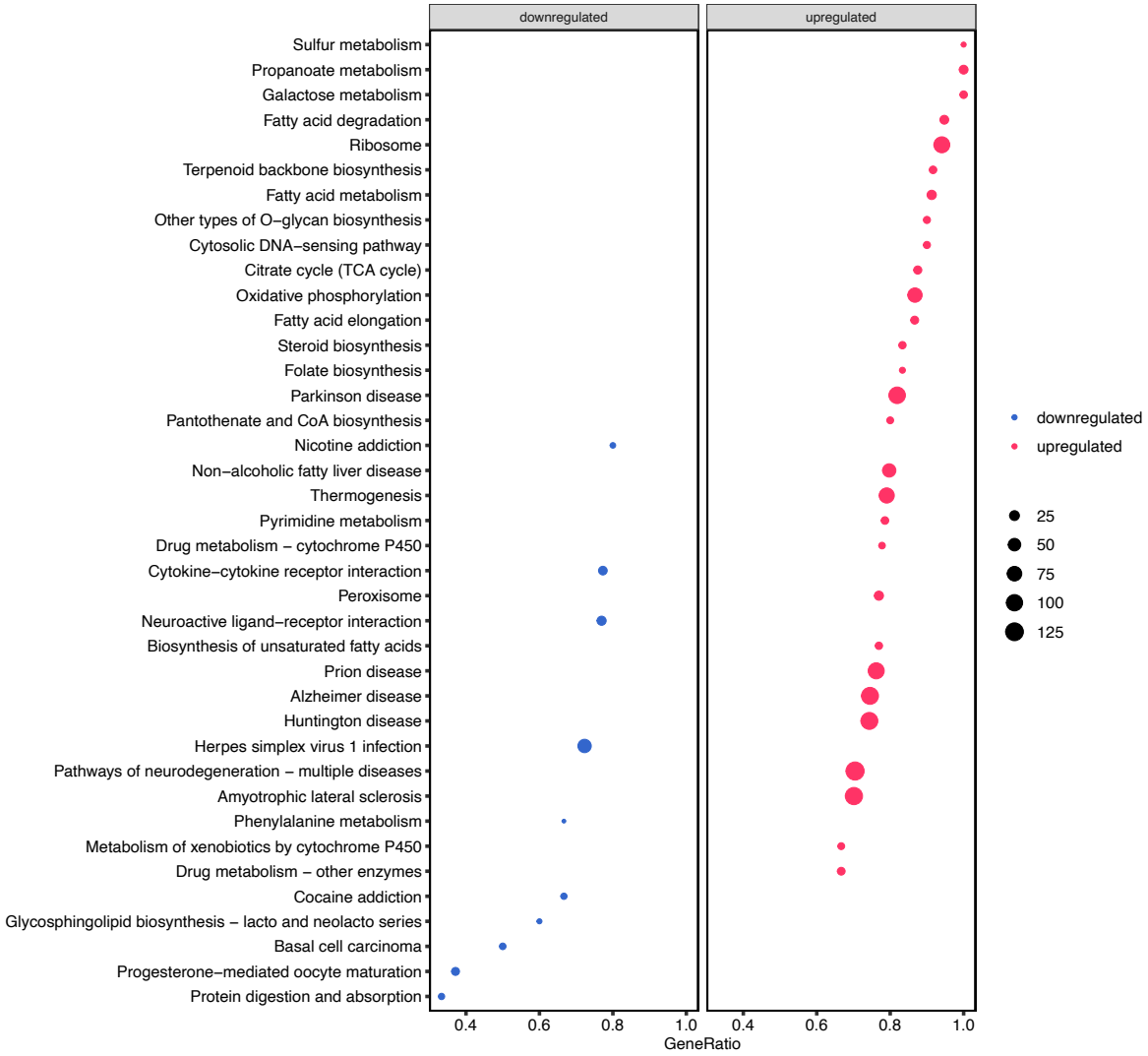
To validate the results from RNA-seq. RT-qPCR was performed on *NHIP* and *BRD1*. *NHIP* was significant upregulated in the overexpressed cell line (two-tailed t-test, *p*-value = 1.05E-05). *BRD1* was significant downregulated in the overexpressed cell line (two-tailed t-test, *p*-value = 5.21E-05).


downregulated
upregulated
ncRNA processing
nuclear−transcribed mRNA catabolic process, nonsense−mediated decay
protein−lysine N−methyltransferase activity
lysine N−methyltransferase activity
translational initiation
organelle inner membrane
histone−lysine N−methyltransferase activity
viral gene expression
histone methyltransferase complex
protein targeting
protein localization to endoplasmic reticulum
p53 binding
histone methyltransferase activity
cellular amide metabolic process
peptide metabolic process
translation
peptidyl−lysine trimethylation
establishment of protein localization to organelle
structural molecule activity
peptide biosynthetic process
amide biosynthetic process
protein localization to organelle
protein deacylation
macromolecule deacylation
peptidyl−threonine modification
megakaryocyte differentiation
cardiac ventricle development
placenta development
DNA damage response, signal transduction by p53 class mediator
RNA polymerase core enzyme binding
0.3
0.4
0.5
0.6
0.3
0.4
0.5
0.6
GeneRatio
downregulated
upregulated
25
50
75
100
125

Supplementary Figure 4.35

Overexpression of *NHIP* in HEK293T cells led to transcriptional changes in genes with functions in histone methylation and cell cycle.

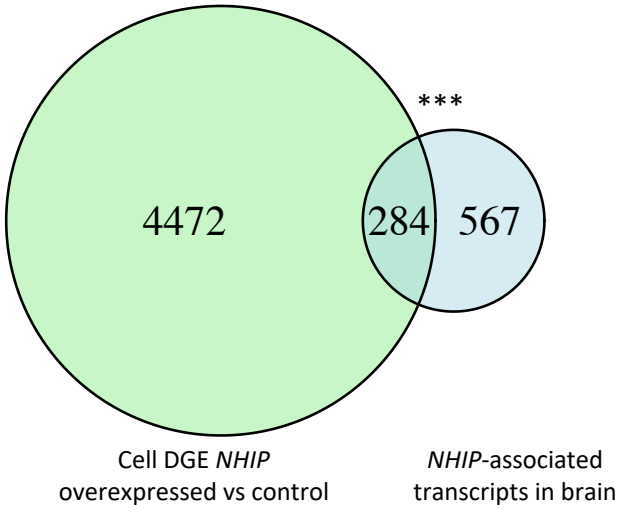
GO analysis was based on the 4,756 DGE compared *NHIP* overexpression and negative control cells with significant enrichment colored in red and blue to represent upregulated and downregulated GO terms (FDR adjusted *p*-value < 0.05).



Supplementary Figure 4.36

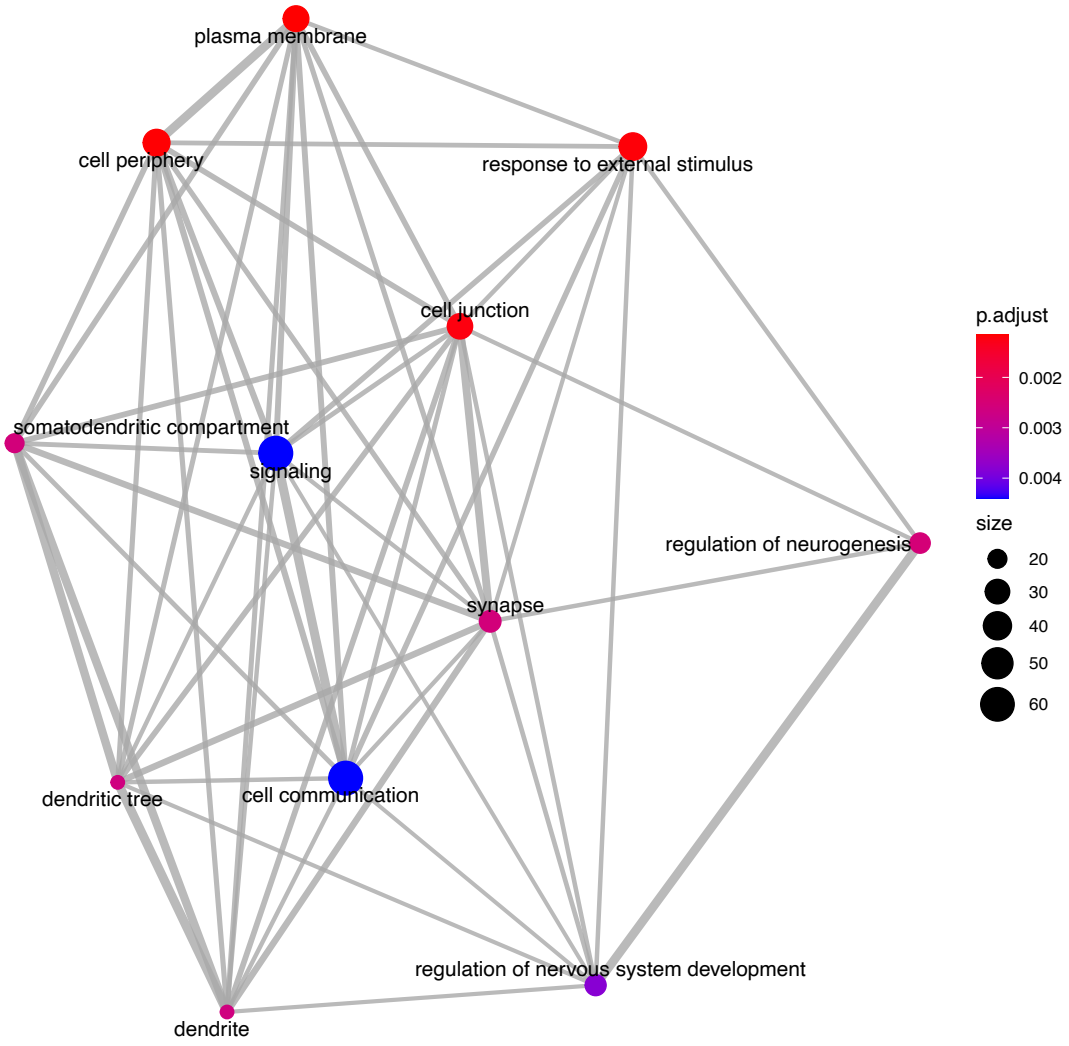
Genes significant associated with overexpression *NHIP* in HEK293T cells were enriched at neuronal diseases.

Enrichment analyses of DGE in cell were also compared to Kyoto Encyclopaedia of Genes and Genomes (KEGG) (371,484) gene set using clusterProfiler package (416). Several terms related with neuronal diseases, Parkinson disease, Alzheimer disease, and Huntington diseases were upregulated with overexpressing *NHIP* in HEK293T cells. Metabolism terms, including fatty acid and drug metabolism, were also upregulated with *NHIP* overexpression.



Supplementary Figure 4.37

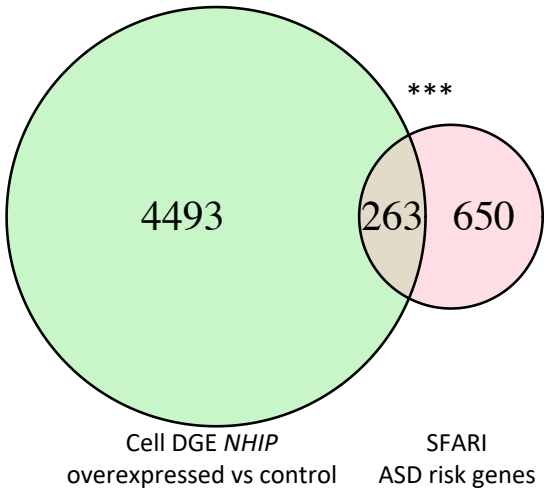
DGE in cell has significantly overlapped with DGE in brain (Fisher's exact test, *p*-value < 0.001).



Supplementary Figure 4.38

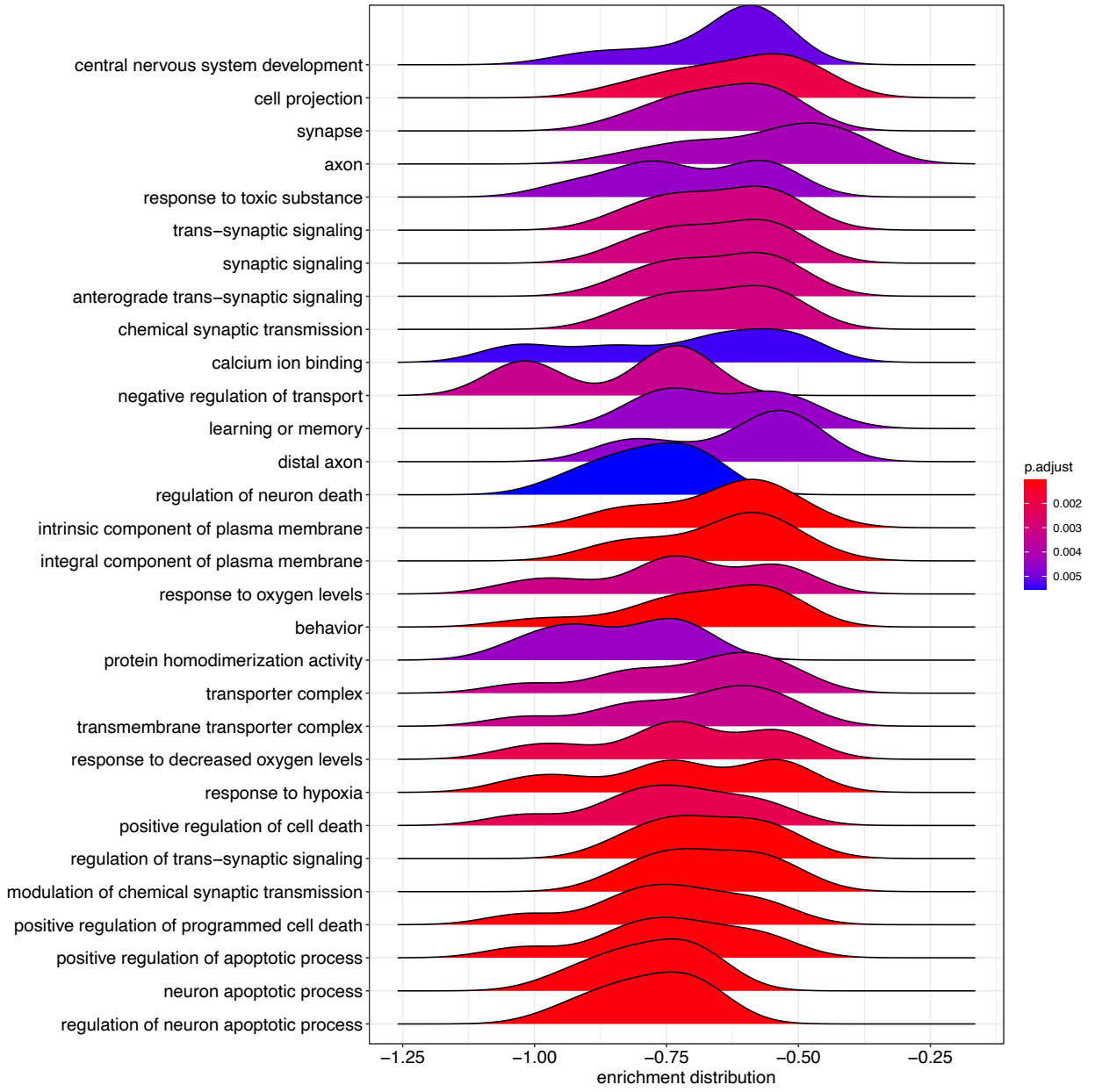
Genes in common between DGE in cell and DGE in brain were significantly enriched for neuronal functions.

The overlapped 284 genes were used for GO term analysis were enriched for neuronal functions, including dendrite, nervous system development, and regulation of neurogenesis. The enrichment map was organized into network on the significant enriched term. Functional modules were clustered together.



Supplementary Figure 4.39

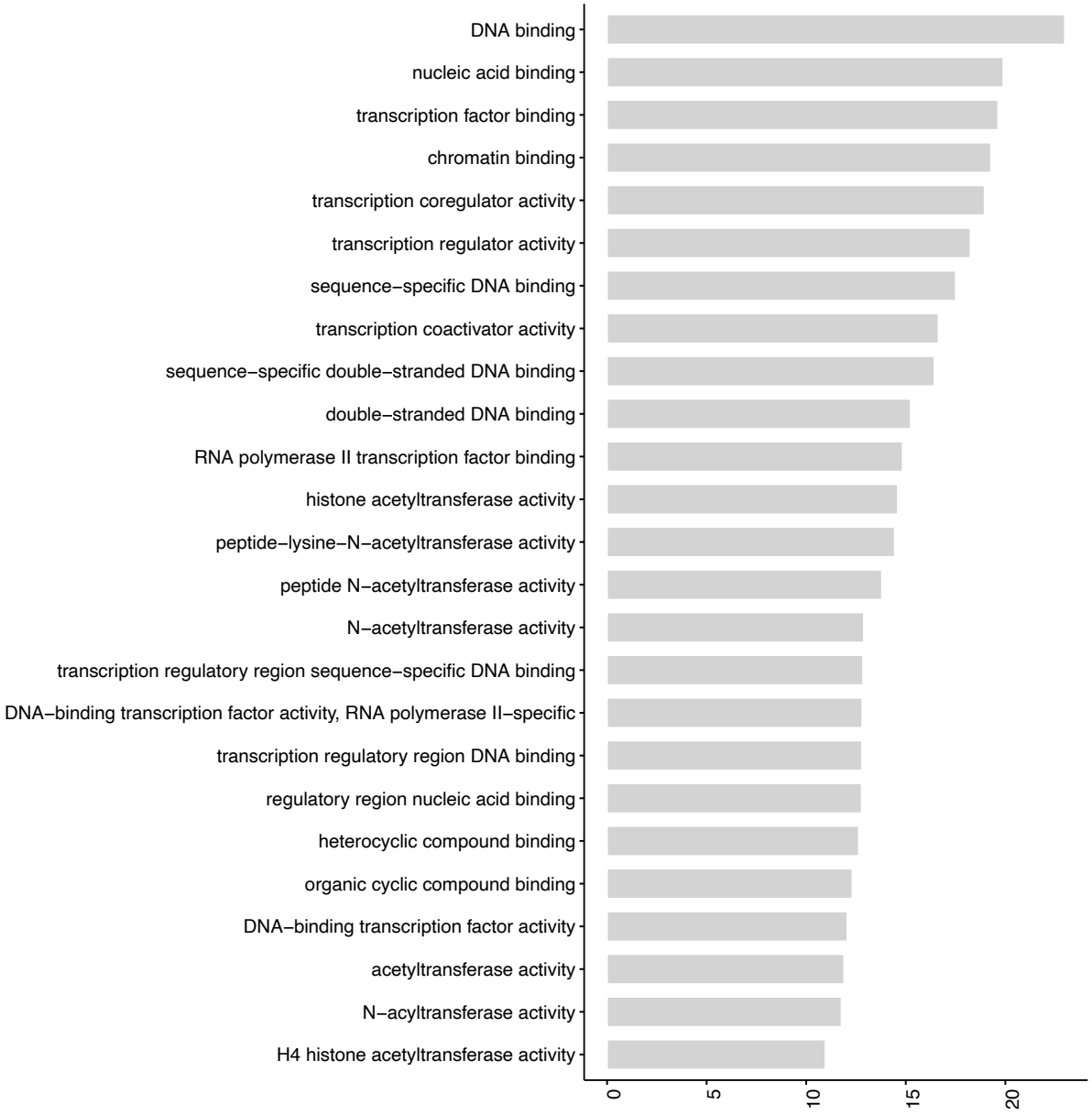
DGE in cell has significantly overlapped with SFARI ASD genes (Fisher's exact test, *p*-value < 0.001).



Supplementary Figure 4.40

Genes in common between cell DGE and SFARI ASD genes shown enrichment as neuronal terms and response to oxygen levels.

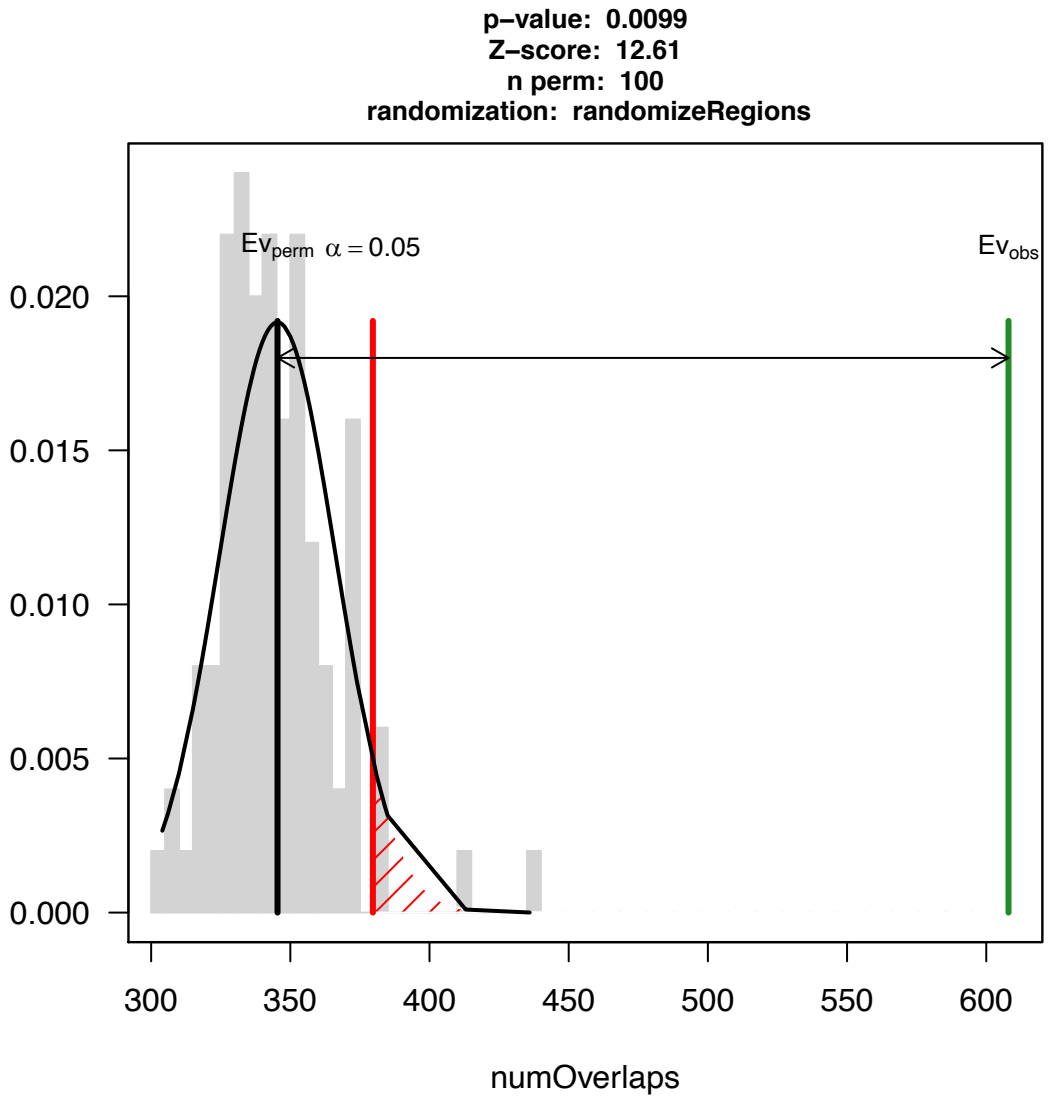
GO term analysis was based on the 263 genes in common between DGE in the overexpressed *NHIP* at HEK293T and SFARI ASD genes. The ridgeline plot was generated using the frequency of the fold change values per gene within each enriched GO terms.



Supplementary Figure 4.41

Genes in common among brain DGE, cell DGE, and SFARI ASD genes enriched for histone acetyltransferase.

GO term analysis was based on the common 45 genes among brain DGE, cell DGE, and SFARI ASD genes. The bar plot was generated by the top 25 GO terms in the molecular function category and shown enrichment with histone acetyltransferase.



Supplementary Figure 4.42

CoRSIV regions significant overlapped with SV hotspots.

Analysis was performed using regioneR R package with 100 permutation tests to show the significant overlapping between CoRSIV region and SV hotspots.

Supplementary Tables

Chapter 2: Supplementary Tables

Supplementary Table 2.1

400 differentially methylated regions (DMRs) in placenta that distinguish ASD and TD samples and their association with 597 genes.

Supplementary Table 2.2

Neurodevelopmental outcomes and additional variables for each placenta sample.

Supplementary Table 2.3

Demographic and clinical variables of children and their mothers in the MARBLES study, stratified by child diagnosis.

Supplementary Table 2.4

Overlapping genes between placenta ASD DMR associated genes and cell type specific genes.

Supplementary Table 2.5

597 genes associated with ASD DMRs and the distance (bp) to the transcription start site (TSS).

Supplementary Table 2.6

Gene ontology (GO) analysis of ASD DMR genes by Fisher's exact test after FDR (false discovery rate) correction.

Supplementary Table 2.7

Overlapping genes between placenta ASD DMR associated genes and other databases, including brain ASD DMR associated genes, ASD genetic risk factors, intellectual disability, Alzheimer's GWAS, and lung cancer GWAS.

Supplementary Table 2.8

Overlapping genes between placenta ASD DMR associated genes and differential expressions genes from ASD postmortem brain studies.

Supplementary Table 2.9

Methylation data from both *CYP2E1* and *IRS2* DMRs for each sample and CpG site (13 CpG sites for *CYP2E1* and 12 CpG sites for *IRS2*). The average represents all CpG sites for each sample. Methods and primers were described in methods sections.

Supplementary Table 2.10

Sanger sequencing results from *CYP2E1* DMR and *IRS2* DMR on rs943975, rs1536828 and rs9301411.

Supplementary Table 2.11

376 differentially methylated regions (DMRs) in placenta separated by whether prenatal vitamins were taken or not during the first month of pregnancy. 587 genes were associated with PreVitM1 DMRs. The interaction set of 60 genes with ASD DMRs is also shown.

Supplementary Table 2.12

Methylation at the *CYP2E1* DMR and IRS2 DMR was tested for interaction with diagnosis, genotype, and PreVitM1.

Chapter 3: Supplementary Tables

Supplementary Table 3.1

Sample variables and RNA quality in MARBLES subjects.

Supplementary Table 3.2

Demographic characteristics of mother participants and their children in MARBLES, stratified by child diagnosis outcomes.

Supplementary Table 3.3

Descriptive statistics of maternal peripheral blood nutrients level in MARBLES, stratified by children diagnosis.

Supplementary Table 3.4

Differential expression analysis on maternal gene expression and diagnosis on adjustment using either SVA or known confounders.

Supplementary Table 3.5

ASD related significant genes from differential expression analysis.

Supplementary Table 3.6

Non-TD related significant genes from differential expression analysis.

Supplementary Table 3.7

Gene Ontology terms on all transcripts and filtered transcripts as background.

Supplementary Table 3.8

Weighted gene co-expression network module memberships for each transcript on the array (MM: Pearson correlation coefficient for module membership; p.MM: P value for the preceding relationship).

Supplementary Table 3.9

Weighted gene co-expression network module features, including number of transcripts and hub genes characters.

Supplementary Table 3.10

“Greenyellow” gene co-expression network module memberships on 224 transcripts (MM: Pearson correlation coefficient for module membership; p.MM: P value for the preceding relationship).

Supplementary Table 3.11

“Greenyellow” gene co-expression network module 224 transcripts gene ontology terms and gene lists.

Supplementary Table 3.12

Modules significant associated with trimester during pregnancy and their enriched gene ontology terms.

Supplementary Table 3.13

Eight weighted gene co-expression modules block memberships on 2,582 transcripts.

Supplementary Table 3.14

Cell type proportions in all 300 maternal peripheral blood samples estimated with CIBERSORT.

Chapter 4: Supplementary Tables

Supplementary Table 4.1

Subject characteristics in relation to outcomes at 36 months.

Supplementary Table 4.2

Subject characteristics by subject.

Supplementary Table 4.3

ASD DMRs identified in discovery group with annotations.

Supplementary Table 4.4

Transcription factor motif enrichment for DMRs.

Supplementary Table 4.5

Genes in common between DMRs associated genes and SFARI ASD genes.

Supplementary Table 4.6

Gene Ontology terms on ASD DMRs.

Supplementary Table 4.7

22q13.33 comethylated block replicate with independent studies, and sequencing platforms.

Supplementary Table 4.8

Blast results of *NHIP* in vertebrates.

Supplementary Table 4.9

Plasmid structure for HEK293T cells and *NHIP* peptide sequence.

Supplementary Table 4.10

NHIP peptide blastp research results.

Supplementary Table 4.11

Insertion characteristics in the discovery group and primers.

Supplementary Table 4.12

PRS was tested for association on 22q13.33 comethylated block % methylation or diagnosis.

Supplementary Table 4.13

Human postmortem brain sample subjects.

Supplementary Table 4.14

Differential gene expression related with *NHIP* in brain.

Supplementary Table 4.15

Gene ontology analysis on differential expressed genes (FDR corrected *p*-value < 0.05) related with *NHIP* in brain.

Supplementary Table 4.16

Genes in common between DGE in brain and SFARI ASD genes.

Supplementary Table 4.17

Gene ontology analysis on overlapped genes between DGE in brain and SFARI ASD genes.

Supplementary Table 4.18

Differential gene expression related with overexpressed *NHIP* in transient cell.

Supplementary Table 4.19

Gene ontology analysis on differential expressed genes (FDR corrected *p*-value < 0.05) related with overexpressed *NHIP* treatment in transient cell.

Supplementary Table 4.20

KEGG pathway enrichment on differential expressed genes (FDR corrected *p*-value < 0.05) with overexpressed *NHIP* treatment in transient cell.

Supplementary Table 4.21

Stable cell differential gene expression, gene ontology on stable differential expressed genes (raw *p*-value < 0.05), gene ontology in common between transient and stable cells, KEGG pathway enrichment on DGE on stable DGE, KEGG pathway in common between transient and stable cells

Supplementary Table 4.22

Genes in common between DGE in brain and DGE in cell.

Supplementary Table 4.23

Gene ontology analysis on the overlapped genes between DGE in brain and DGE in cell.

Supplementary Table 4.24

Genes in common between DGE in brain and SFARI ASD genes.

Supplementary Table 4.25

Gene ontology analysis on the overlapped genes between DGE in cell and SFARI ASD genes.

Supplementary Table 4.26

Genes in common among DGE in brain, DGE in cell, and SFARI ASD genes.

Supplementary Table 4.27

Gene ontology analysis on the overlapped genes among DGE in brain, DGE in cell, and SFARI ASD genes.

References

1. Srivastava D, DeWitt N. In Vivo Cellular Reprogramming: The Next Generation. Cell. 2016.
2. Dor Y, Cedar H. Principles of DNA methylation and their implications for biology and medicine. The Lancet. 2018.
3. Smith ZD, Meissner A. DNA methylation: Roles in mammalian development. Nature Reviews Genetics. 2013.
4. LaSalle JM, Powell WT, Yasui DH. Epigenetic layers and players underlying neurodevelopment. Trends in Neurosciences. 2013.
5. Abascal F, Acosta R, Addleman NJ, Adrian J, Afzal V, Aken B, et al. Perspectives on ENCODE [Internet]. Vol. 583, Nature. Nature Research; 2020 [cited 2021 Mar 25]. p. 693–8. Available from: https://doi.org/10.1038/s41586-020-2449-8
6. Tost J. 10 years of Epigenomics: A journey with the epigenetic community through exciting times. Epigenomics. 2020.
7. Vogel Ciernia A, LaSalle J. The landscape of DNA methylation amid a perfect storm of autism aetiologies. Nat Rev Neurosci [Internet]. 2016;17(7):411–23. Available from: http://www.nature.com/doifinder/10.1038/nrn.2016.41
8. Bird A. DNA methylation patterns and epigenetic memory. Genes and Development. 2002.
9. Virani S, Virani S, Colacino JA, Kim JH, Rozek LS. Cancer epigenetics: a brief review. ILAR journal / National Research Council, Institute of Laboratory Animal Resources. 2012.
10. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. Nature. 2001;
11. Cedar H, Bergman Y. Programming of DNA methylation patterns. Annu Rev Biochem. 2012;
12. Ehrlich M. DNA hypomethylation in cancer cells. Epigenomics. 2009.
13. Lei H, Oh SP, Okano M, Jüttermann R, Goss KA, Jaenisch R, et al. De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells. Development. 1996;
14. Okano M, Bell DW, Haber DA, Li E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell. 1999;
15. Bestor TH. The DNA methyltransferases of mammals. Human Molecular Genetics. 2000.
16. Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science (80-). 2009;
17. Nan X, Meehan RR, Bird A. Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2. Nucleic Acids Res. 1993;
18. Amir RE, Van Den Veyver IB, Wan M, Tran CQ, Francke U, Zoghbi HY. Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl- CpG-binding protein 2. Nat Genet. 1999 Oct;23(2):185–8.
19. Eden A, Gaudet F, Waghmare A, Jaenisch R. Chromosomal instability and tumors promoted by DNA hypomethylation. Science. 2003.
20. Chen RZ, Pettersson U, Beard C, Jackson-Grusby L, Jaenisch R. DNA hypomethylation leads to elevated mutation rates. Nature. 1998;
21. Cordaux R, Batzer MA. The impact of retrotransposons on human genome evolution.

Nature Reviews Genetics. 2009.
22. Kazazian HH, Goodier JL. LINE drive: Retrotransposition and genome instability. Cell. 2002.
23. Daskalos A, Nikolaidis G, Xinarianos G, Savvari P, Cassidy A, Zakopoulou R, et al. Hypomethylation of retrotransposable elements correlates with genomic instability in non-small cell lung cancer. Int J Cancer. 2009;
24. Rose NR, Klose RJ. Understanding the relationship between DNA methylation and histone lysine methylation. Biochimica et Biophysica Acta - Gene Regulatory Mechanisms. 2014.
25. Shilatifard A. Chromatin modifications by methylation and ubiquitination: Implications in the regulation of gene expression. Annual Review of Biochemistry. 2006.
26. Wang GG, Allis CD, Chi P. Chromatin remodeling and cancer, part I: covalent histone modifications. Trends in Molecular Medicine. 2007.
27. Hyun K, Jeon J, Park K, Kim J. Writing, erasing and reading histone lysine methylations. Experimental and Molecular Medicine. 2017.
28. Milne TA, Briggs SD, Brock HW, Martin ME, Gibbs D, Allis CD, et al. MLL targets SET domain methyltransferase activity to Hox gene promoters. Mol Cell. 2002;
29. Cao R, Wang L, Wang H, Xia L, Erdjument-Bromage H, Tempst P, et al. Role of histone H3 lysine 27 methylation in polycomb-group silencing. Science (80-). 2002;
30. Viré E, Brenner C, Deplus R, Blanchon L, Fraga M, Didelot C, et al. The Polycomb group protein EZH2 directly controls DNA methylation. Nature. 2006;
31. Jeltsch A, Broche J, Bashtrykov P. Molecular processes connecting DNA methylation patterns with DNA methyltransferases and histone modifications in mammalian genomes. Genes. 2018.
32. Bannister AJ, Kouzarides T. Regulation of chromatin by histone modifications. Cell Research. 2011.
33. Roth SY, Denu JM, Allis CD. Histone acetyltransferases. Annual Review of Biochemistry. 2001.
34. Chen J, Luo Q, Yuan Y, Huang X, Cai W, Li C, et al. Pygo2 Associates with MLL2 Histone Methyltransferase and GCN5 Histone Acetyltransferase Complexes To Augment Wnt Target Gene Expression and Breast Cancer Stem-Like Cell Expansion. Mol Cell Biol. 2010;
35. Iyer NG, Özdag H, Caldas C. p300/CBP and cancer [Internet]. Vol. 23, Oncogene. Nature Publishing Group; 2004 [cited 2021 Mar 28]. p. 4225–31. Available from: www.nature.com/onc
36. Garcia-Carpizo V, Ruiz-Llorente S, Sarmentero J, Graña-Castro O, Pisano DG, Barrero MJ. CREBBP/EP300 bromodomains are critical to sustain the GATA1/MYC regulatory axis in proliferation. Epigenetics and Chromatin. 2018;
37. Glozak MA, Sengupta N, Zhang X, Seto E. Acetylation and deacetylation of non-histone proteins. Gene. 2005.
38. Godman CA, Joshi R, Tierney BR, Greenspan E, Rasmussen TP, Wang HW, et al. HDAC3 impacts multiple oncogenic pathways in colon cancer cells with effects on Wnt and vitamin D signaling. Cancer Biol Ther. 2008;
39. Matharu N, Ahituv N. Minor Loops in Major Folds: Enhancer–Promoter Looping, Chromatin Restructuring, and Their Association with Transcriptional Regulation and Disease. PLoS Genetics. 2015.

40. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature. 2012;
41. Lazar NH, Nevonen KA, O'Connell B, McCann C, O'Neill RJ, Green RE, et al. Epigenetic maintenance of topological domains in the highly rearranged gibbon genome. Genome Res. 2018;
42. Wei JW, Huang K, Yang C, Kang CS. Non-coding RNAs as regulators in epigenetics (Review). Oncol Rep. 2017;
43. Ducker GS, Rabinowitz JD. One-Carbon Metabolism in Health and Disease. Cell Metabolism. 2017.
44. Mason JB. Biomarkers of nutrient exposure and status in one-carbon (methyl) metabolism. J Nutr. 2003;
45. Crider KS, Bailey LB, Berry RJ. Folic acid food fortification-its history, effect, concerns, and future directions. Nutrients. 2011.
46. Garratt LC, Ortori CA, Tucker GA, Sablitzky F, Bennett MJ, Barrett DA. Comprehensive metabolic profiling of mono- and polyglutamated folates and their precursors in plant and animal tissue using liquid chromatography/negative ion electrospray ionisation tandem mass spectrometry. Rapid Commun Mass Spectrom. 2005;
47. Shane B. Folate and vitamin B12, metabolism: Overview and interaction with riboflavin, vitamin B6, and polymorphisms. In: Food and Nutrition Bulletin. 2008.
48. Clare CE, Brassington AH, Kwong WY, Sinclair KD. One-Carbon Metabolism: Linking Nutritional Biochemistry to Epigenetic Programming of Long-Term Development. Annual Review of Animal Biosciences. 2019.
49. Roje S. S-Adenosyl-l-methionine: Beyond the universal methyl group donor. Phytochemistry. 2006.
50. Ulrey CL, Liu L, Andrews LG, Tollefsbol TO. The impact of metabolism on DNA methylation. Human Molecular Genetics. 2005.
51. Mato JM, Corrales FJ, Lu SC, Avila MA. S-Adenosylmethionine: a control switch that regulates liver function. FASEB J. 2002;
52. Yi P, Melnyk S, Pogribna M, Pogribny IP, Hine RJ, James SJ. Increase in plasma homocysteine associated with parallel increases in plasma S-adenosylhomocysteine and lymphocyte DNA hypomethylation. J Biol Chem. 2000;
53. Shojaei Saadi HA, Gagné D, Fournier É, Baldoceda Baldeon LM, Sirard MA, Robert C. Responses of bovine early embryos to S-adenosyl methionine supplementation in culture. Epigenomics. 2016;
54. Kucharski R, Maleszka J, Foret S, Maleszka R. Nutritional control of reproductive status in honeybees via DNA methylation. Science (80-). 2008;
55. Sun C, Huang J, Wang Y, Zhao X, Su L, Thomas GWC, et al. Genus-Wide Characterization of Bumblebee Genomes Provides Insights into Their Evolution and Variation in Ecological and Behavioral Traits. Mol Biol Evol. 2021;
56. Gong M, Dong W, He T, Shi Z, Huang G, Ren R, et al. MTHFR 677C>T polymorphism increases the male infertility risk: A meta-analysis involving 26 studies. PLoS One. 2015;
57. Lambrot R, Xu C, Saint-Phar S, Chountalos G, Cohen T, Paquet M, et al. Low paternal dietary folate alters the mouse sperm epigenome and is associated with negative pregnancy outcomes. Nat Commun. 2013;

58. Blackburn PR, Tischer A, Zimmermann MT, Kemppainen JL, Sastry S, Knight Johnson AE, et al. A novel Kleefstra syndrome-associated variant that affects the conserved TPLX motif within the ankyrin repeat of EHMT1 leads to abnormal protein folding. J Biol Chem. 2017;
59. Maenner MJ, Shaw KA, Baio J, Washington A, Patrick M, DiRienzo M, et al. Prevalence of autism spectrum disorder among children aged 8 Years-Autism and developmental disabilities monitoring network, 11 Sites, United States, 2016. MMWR Surveill Summ [Internet]. 2020 [cited 2020 Jul 22];69(4):1–12. Available from: https://www.cdc.gov/mmwr/volumes/69/ss/ss6904a1.htm
60. Bourgeron T. From the genetic architecture to synaptic plasticity in autism spectrum disorder. Nat Rev Neurosci [Internet]. 2015;16(9):551–63. Available from: http://dx.doi.org/10.1038/nrn3992
61. Hallmayer J, Cleveland S, Torres A, Phillips J, Cohen B, Torigoe T, et al. Genetic heritability and shared environmental factors among twin pairs with autism. Arch Gen Psychiatry. 2011;
62. Colvert E, Tick B, McEwen F, Stewart C, Curran SR, Woodhouse E, et al. Heritability of autism spectrum disorder in a UK population-based twin sample. JAMA Psychiatry. 2015;
63. Ozonoff S, Young GS, Carter A, Messinger D, Yirmiya N, Zwaigenbaum L, et al. Recurrence risk for autism spectrum disorders: a Baby Siblings Research Consortium study. Pediatrics [Internet]. 2011 Sep [cited 2018 Aug 15];128(3):e488-95. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21844053
64. Grove J, Ripke S, Als TD, Mattheisen M, Walters RK, Won H, et al. Identification of common genetic risk variants for autism spectrum disorder. Nat Genet. 2019;
65. Iossifov I, O'Roak BJ, Sanders SJ, Ronemus M, Krumm N, Levy D, et al. The contribution of de novo coding mutations to autism spectrum disorder. Nature [Internet]. 2014 Nov 29 [cited 2018 Sep 10];515(7526):216–21. Available from: http://www.nature.com/articles/nature13908
66. Sanders SJ, He X, Willsey AJ, Ercan-Sencicek AG, Samocha KE, Cicek AE, et al. Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. Neuron. 2015;
67. Crawley JN, Heyer WD, LaSalle JM. Autism and Cancer Share Risk Genes, Pathways, and Drug Targets. Trends Genet [Internet]. 2016;32(3):139–46. Available from: http://dx.doi.org/10.1016/j.tig.2016.01.001
68. Lyall K, Croen L, Daniels J, Fallin MD, Ladd-Acosta C, Lee BK, et al. The Changing Epidemiology of Autism Spectrum Disorders. Annu Rev Public Health. 2017;
69. Hertz-Picciotto I, Schmidt RJ, Krakowiak P. Understanding environmental contributions to autism: Causal concepts and the state of science. Autism Research. 2018.
70. Schmidt RJ, Tancredi DJ, Ozonoff S, Hansen RL, Hartiala J, Allayee H, et al. Maternal periconceptional folic acid intake and risk of autism spectrum disorders and developmental delay in the CHARGE (CHildhood Autism Risks from Genetics and Environment) case-control study. Am J Clin Nutr. 2012;96(1):80–9.
71. Zhu Y, Mordaunt CE, Durbin-Johnson BP, Caudill MA, Malysheva O V., Miller JW, et al. Expression Changes in Epigenetic Gene Pathways Associated With One-Carbon Nutritional Metabolites in Maternal Blood From Pregnancies Resulting in Autism and

Non-Typical Neurodevelopment. Autism Res [Internet]. 2020 [cited 2021 Jan 3]; Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/aur.2428
72. Parikshak NN, Swarup V, Belgard TG, Irimia M, Ramaswami G, Gandal MJ, et al. Genome-wide changes in lncRNA, splicing and regional gene expression patterns in autism. Nature [Internet]. 2016 Dec 5 [cited 2019 Mar 4];540(7633):423–7. Available from: http://www.nature.com/articles/nature20612
73. Wray NR, Visscher PM. Estimating trait heritability. Nat Educ. 2008;
74. Steffenburg S, Gillberg C, Hellgren L, Andersson L, Gillberg IC, Jakobsson G, et al. A Twin Study of Autism in Denmark, Finland, Iceland, Norway and Sweden. J Child Psychol Psychiatry. 1989;
75. Bailey A, Le Couteur A, Gottesman I, Bolton P, Simonoff E, Yuzda E, et al. Autism as a strongly genetic disorder: Evidence from a british twin study. In: The Science of Mental Health: Volume 2: Autism. 2013.
76. Ronald A, Hoekstra RA. Autism spectrum disorders and autistic traits: A decade of new twin studies. American Journal of Medical Genetics, Part B: Neuropsychiatric Genetics. 2011.
77. Sandin S, Lichtenstein P, Kuja-Halkola R, Hultman C, Larsson H, Reichenberg A. The heritability of autism spectrum disorder. JAMA - J Am Med Assoc. 2017;
78. Tick B, Bolton P, Happé F, Rutter M, Rijsdijk F. Heritability of autism spectrum disorders: A meta-analysis of twin studies. J Child Psychol Psychiatry Allied Discip. 2016;
79. Hertz-Picciotto I, Schmidt RJ, Walker CK, Bennett DH, Oliver M, Shedd-Wise KM, et al. A Prospective Study of Environmental Exposures and Early Biomarkers in Autism Spectrum Disorder: Design, Protocols, and Preliminary Data from the MARBLES Study. Environ Health Perspect [Internet]. 2018 Nov [cited 2018 Nov 26];126(11):117004. Available from: https://ehp.niehs.nih.gov/doi/10.1289/EHP535
80. Rosenberg RE, Law JK, Yenokyan G, McGready J, Kaufmann WE, Law PA. Characteristics and concordance of autism spectrum disorders among 277 twin pairs. Arch Pediatr Adolesc Med. 2009;
81. Sandin S, Lichtenstein P, Kuja-Halkola R, Larsson H, Hultman CM, Reichenberg A. The familial risk of autism. JAMA - J Am Med Assoc. 2014;311(17):1770–7.
82. De Rubeis S, He X, Goldberg AP, Poultney CS, Samocha K, Cicek AE, et al. Synaptic, transcriptional and chromatin genes disrupted in autism. Nature. 2014;
83. Krumm N, Turner TN, Baker C, Vives L, Mohajeri K, Witherspoon K, et al. Excess of rare, inherited truncating mutations in autism. Nat Genet. 2015;
84. Abrahams BS, Geschwind DH. Advances in autism genetics: On the threshold of a new neurobiology. Nature Reviews Genetics. 2008.
85. Betancur C. Etiological heterogeneity in autism spectrum disorders: More than 100 genetic and genomic disorders and still counting. Brain Research. 2011.
86. Yoo H. Genetics of Autism Spectrum Disorder: Current Status and Possible Clinical Applications. Exp Neurobiol. 2015;
87. Strauss KA, Puffenberger EG, Huentelman MJ, Gottlieb S, Dobrin SE, Parod JM, et al. Recessive Symptomatic Focal Epilepsy and Mutant Contactin-Associated Protein-like 2. N Engl J Med. 2006;
88. Morrow EM, Yoo SY, Flavell SW, Kim TK, Lin Y, Hill RS, et al. Identifying autism loci and

genes by tracing recent shared ancestry. Science (80-). 2008;
89. Novarino G, El-Fishawy P, Kayserili H, Meguid NA, Scott EM, Schroth J, et al. Mutations in BCKD-kinase lead to a potentially treatable form of autism with epilepsy. Science (80-). 2012;
90. Jamain S, Quach H, Betancur C, Råstam M, Colineaux C, Gillberg C, et al. Mutations of the X-linked genes encoding neuroligins NLGN3 and NLGN4 are associated with autism. Nat Genet. 2003;
91. Helsmoortel C, Vulto-Van Silfhout AT, Coe BP, Vandeweyer G, Rooms L, Van Den Ende J, et al. A SWI/SNF-related autism syndrome caused by de novo mutations in ADNP. Nat Genet. 2014;
92. Moessner R, Marshall CR, Sutcliffe JS, Skaug J, Pinto D, Vincent J, et al. Contribution of SHANK3 mutations to autism spectrum disorder. Am J Hum Genet. 2007;
93. Kim HG, Kishikawa S, Higgins AW, Seong IS, Donovan DJ, Shen Y, et al. Disruption of Neurexin 1 Associated with Autism Spectrum Disorder. Am J Hum Genet. 2008;
94. Berkel S, Marshall CR, Weiss B, Howe J, Roeth R, Moog U, et al. Mutations in the SHANK2 synaptic scaffolding gene in autism spectrum disorder and mental retardation. Nat Genet. 2010;
95. Sato D, Lionel AC, Leblond CS, Prasad A, Pinto D, Walker S, et al. SHANK1 deletions in males with autism spectrum disorder. Am J Hum Genet. 2012;
96. Vaags AK, Lionel AC, Sato D, Goodenberger M, Stein QP, Curran S, et al. Rare deletions at the neurexin 3 locus in autism spectrum disorder. Am J Hum Genet. 2012;
97. Pieretti M, Zhang F, Fu YH, Warren ST, Oostra BA, Caskey CT, et al. Absence of expression of the FMR-1 gene in fragile X syndrome. Cell. 1991;
98. Coffee B, Keith K, Albizua I, Malone T, Mowrey J, Sherman SL, et al. Incidence of Fragile X Syndrome by Newborn Screening for Methylated FMR1 DNA. Am J Hum Genet. 2009;
99. Myrick LK, Hashimoto H, Cheng X, Warren ST. Human FMRP contains an integral tandem Agenet (Tudor) and KH motif in the amino terminal domain. Hum Mol Genet. 2015;
100. Hagerman PJ, Hagerman R. Fragile X syndrome. Curr Biol. 2021;
101. Hagerman R, Hagerman P. Advances in clinical and molecular understanding of the FMR1 premutation and fragile X-associated tremor/ataxia syndrome. The Lancet Neurology. 2013.
102. Sitzmann AF, Hagelstrom RT, Tassone F, Hagerman RJ, Butler MG. Rare FMR1 gene mutations causing fragile X syndrome: A review. Am J Med Genet Part A. 2018;
103. Wang T, Guo H, Xiong B, Stessman HAF, Wu H, Coe BP, et al. De novo genic mutations among a Chinese autism spectrum disorder cohort. Nat Commun. 2016;
104. Ragione F Della, Vacca M, Fioriniello S, Pepe G, D'Esposito M. MECP2, a multi-talented modulator of chromatin architecture. Brief Funct Genomics. 2016;
105. Cheng TL, Chen J, Wan H, Tang B, Tian W, Liao L, et al. Regulation of mRNA splicing by MeCP2 via epigenetic modifications in the brain. Sci Rep. 2017;
106. Pfaender S, Sauer AK, Hagmeyer S, Mangus K, Linta L, Liebau S, et al. Zinc deficiency and low enterocyte zinc transporter expression in human patients with autism related mutations in SHANK3. Sci Rep. 2017;
107. Marshall CR, Noor A, Vincent JB, Lionel AC, Feuk L, Skaug J, et al. Structural Variation of Chromosomes in Autism Spectrum Disorder. Am J Hum Genet. 2008;

108. Satterstrom FK, Kosmicki JA, Wang J, Breen MS, De Rubeis S, An JY, et al. Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. Cell [Internet]. 2020 Feb 6 [cited 2021 Mar 19];180(3):568-584.e23. Available from: https://doi.org/10.1016/j.cell.2019.12.036
109. Tam V, Patel N, Turcotte M, Bossé Y, Paré G, Meyre D. Benefits and limitations of genome-wide association studies. Nature Reviews Genetics. 2019.
110. Hollox EJ, Huffmeier U, Zeeuwen PLJM, Palla R, Lascorz J, Rodijk-Olthuis D, et al. Psoriasis is associated with increased β-defensin genomic copy number. Nat Genet. 2008;
111. Stefansson H, Rujescu D, Cichon S, Pietiläinen OPH, Ingason A, Steinberg S, et al. Large recurrent microdeletions associated with schizophrenia. Nature. 2008;
112. Sekar A, Bialas AR, De Rivera H, Davis A, Hammond TR, Kamitaki N, et al. Schizophrenia risk from complex variation of complement component 4. Nature. 2016;
113. Gokcumen O, Babb PL, Iskow RC, Zhu Q, Shi X, Mills RE, et al. Refinement of primate copy number variation hotspots identifies candidate genomic regions evolving under positive selection. Genome Biol. 2011;
114. Lin YL, Gokcumen O. Fine-scale characterization of genomic structural variation in the human genome reveals adaptive and biomedically relevant hotspots. Genome Biol Evol. 2019;
115. Girirajan S, Brkanac Z, Coe BP, Baker C, Vives L, Vu TH, et al. Relative burden of large CNVs on a range of neurodevelopmental phenotypes. PLoS Genet. 2011;
116. Pinto D, Pagnamenta AT, Klei L, Anney R, Merico D, Regan R, et al. Functional impact of global rare copy number variation in autism spectrum disorders. Nature. 2010;
117. Girirajan S, Dennis MY, Baker C, Malig M, Coe BP, Campbell CD, et al. Refinement and discovery of new hotspots of copy-number variation associated with autism spectrum disorder. Am J Hum Genet. 2013;
118. Turner TN, Hormozdiari F, Duyzend MH, McClymont SA, Hook PW, Iossifov I, et al. Genome Sequencing of Autism-Affected Families Reveals Disruption of Putative Noncoding Regulatory DNA. Am J Hum Genet [Internet]. 2016;98(1):58–74. Available from: http://dx.doi.org/10.1016/j.ajhg.2015.11.023
119. Suren P, Roth C, Bresnahan M, Haugen M, Hornig M, Hirtz D, et al. Association between maternal use of folic acid supplements and risk of autism spectrum disorders in children. J Am Med Assoc. 2013;309(6):570–7.
120. Schmidt RJ, Hansen RL, Hartiala J, Allayee H, Schmidt LC, Tancredi DJ, et al. Prenatal vitamins, one-carbon metabolism gene variants, and risk for autism. Epidemiology [Internet]. 2011 Jul [cited 2018 Oct 23];22(4):476–85. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21610500
121. Pu D, Shen Y, Wu J. Association between mthfr gene polymorphisms and the risk of autism spectrum disorders: A meta-analysis. Autism Res. 2013;
122. Schmidt RJ, Iosif A-M, Guerrero Angel E, Ozonoff S. Association of Maternal Prenatal Vitamin Use With Risk for Autism Spectrum Disorder Recurrence in Young Siblings. JAMA Psychiatry [Internet]. 2019 Apr 1 [cited 2019 Apr 29];76(4):391. Available from: http://archpsyc.jamanetwork.com/article.aspx?doi=10.1001/jamapsychiatry.2018.3901
123. Lyall K, Schmidt RJ, Hertz-Picciotto I. Maternal lifestyle and environmental risk factors for autism spectrum disorders. Int J Epidemiol. 2014;

124. Reik W, Walter J. Genomic imprinting: Parental influence on the genome. Nature Reviews Genetics. 2001.
125. Landrigan PJ, Garg A. Chronic effects of toxic environmental exposures on children's health. In: Journal of Toxicology - Clinical Toxicology. 2002.
126. Windham GC, Zhang L, Gunier R, Croen LA, Grether JK. Autism spectrum disorders in relation to distribution of hazardous air pollutants in the San Francisco Bay area. Environ Health Perspect. 2006;
127. Cheslack-Postava K, Rantakokko P V., Hinkka-Yli-Salomäki S, Surcel HM, McKeague IW, Kiviranta HA, et al. Maternal serum persistent organic pollutants in the Finnish Prenatal Study of Autism: A pilot study. Neurotoxicol Teratol. 2013;
128. Roberts EM, English PB, Grether JK, Windham GC, Somberg L, Wolff C. Maternal residence near agricultural pesticide applications and autism spectrum disorders among children in the California Central Valley. Environ Health Perspect. 2007;
129. Volk HE, Hertz-Picciotto I, Delwiche L, Lurmann F, McConnell R. Residential proximity to freeways and autism in the CHARGE study. Environ Health Perspect. 2011;
130. Granillo L, Sethi S, Keil KP, Lin Y, Ozonoff S, Iosif AM, et al. Polychlorinated biphenyls influence on autism spectrum disorder risk in the MARBLES cohort. Environ Res. 2019;
131. Boas M, Feldt-Rasmussen U, Main KM. Thyroid effects of endocrine disrupting chemicals. Molecular and Cellular Endocrinology. 2012.
132. Dunaway KW, Islam MS, Coulson RL, Lopez SJ, Vogel Ciernia A, Chu RG, et al. Cumulative Impact of Polychlorinated Biphenyl and Large Chromosomal Duplications on DNA Methylation, Chromatin, and Expression of Autism Candidate Genes. Cell Rep. 2016;17(11):3035–48.
133. Schmidt RJ, Schroeder DI, Crary-Dooley FK, Barkoski JM, Tancredi DJ, Walker CK, et al. Self-reported pregnancy exposures and placental DNA methylation in the MARBLES prospective autism sibling study. Environ Epigenetics [Internet]. 2016;2(4):dvw024–dvw024. Available from: http://dx.doi.org/10.1093/eep/dvw024
134. Lawrence LJ, Casida JE. Stereospecific action of pyrethroid insecticides on the γ-aminobutyric acid receptor-ionophore complex. Science (80-) [Internet]. 1983 Sep 30 [cited 2021 Apr 1];221(4618):1399–401. Available from: https://science.sciencemag.org/content/221/4618/1399
135. Ladd-Acosta C, Hansen KD, Briem E, Fallin MD, Kaufmann WE, Feinberg AP. Common DNA methylation alterations in multiple brain regions in autism. Mol Psychiatry. 2014;
136. Zhu Y, Mordaunt CE, Yasui DH, Marathe R, Coulson RL, Dunaway KW, et al. Placental DNA methylation levels at CYP2E1 and IRS2 are associated with child outcome in a prospective autism study. Hum Mol Genet [Internet]. 2019 Aug 15 [cited 2019 Sep 10];28(16):2659–74. Available from: https://academic.oup.com/hmg/article/28/16/2659/5476403
137. Wong CCY, Meaburn EL, Ronald A, Price TS, Jeffries AR, Schalkwyk LC, et al. Methylomic analysis of monozygotic twins discordant for autism spectrum disorder and related behavioural traits. Mol Psychiatry. 2014;
138. Lappalainen T, Greally JM. Associating cellular epigenetic models with human phenotypes. Nature Reviews Genetics. 2017.
139. Illumina. Illumina Sequencing Methods [Internet]. 2016 [cited 2021 Jan 30]. Available from: www.illumina.com

140. Vogel Ciernia A, Laufer BI, Hwang H, Dunaway KW, Mordaunt CE, Coulson RL, et al. Epigenomic Convergence of Neural-Immune Risk Factors in Neurodevelopmental Disorder Cortex. Cereb Cortex. 2019;
141. Sun W, Poschmann J, Cruz-Herrera del Rosario R, Parikshak NN, Hajan HS, Kumar V, et al. Histone Acetylome-wide Association Study of Autism Spectrum Disorder. Cell [Internet]. 2016 Nov 17 [cited 2018 Oct 2];167(5):1385-1397.e11. Available from: https://www.sciencedirect.com/science/article/pii/S0092867416314519#!
142. Chestnut BA, Chang Q, Price A, Lesuisse C, Wong M, Martin LJ. Epigenetic regulation of motor neuron cell death through DNA methylation. J Neurosci. 2011;
143. Power ML, Schulkin J. The evolution of the human placenta. The Evolution of the Human Placenta. 2012.
144. Bergsjø P. Pathology of the Human Placenta. Acta Obstet Gynecol Scand. 2001;
145. Clutton-Brock TH. The Evolution of Parental Care [Internet]. Princeton University Press. 1991. p. Vol. 64. Available from: https://books.google.com/books?hl=en&lr=&id=RV65DwAAQBAJ&oi=fnd&pg=PR10&ots=CupeiTRVcD&sig=fB7LEGDFu1w7RQY4-bzCAo_gI9E#v=onepage&q&f=false
146. Rossant J, Tam PPL. Blastocyst lineage formation, early embryonic asymmetries and axis patterning in the mouse. Development. 2009.
147. Maltepe E, Bakardjiev AI, Fisher SJ. The placenta: Transcriptional, epigenetic, and physiological integration during development. Journal of Clinical Investigation. 2010.
148. Carter AM. Placental oxygen consumption. Part I: In vivo studies - A review. Placenta. 2000;
149. Carter BS, Moores RR, Battaglia FC, Meschia G. Ovine fetal placental lactate exchange and decarboxylation at midgestation. Am J Physiol - Endocrinol Metab. 1993;
150. Postigo L, Heredia G, Illsley NP, Torricos T, Dolan C, Echalar L, et al. Where the O2 goes to: Preservation of human fetal oxygen delivery and consumption at high altitude. J Physiol. 2009;
151. Ietta F, Wu Y, Romagnoli R, Soleymanlou N, Orsini B, Zamudio S, et al. Oxygen regulation of macrophage migration inhibitory factor in human placenta. Am J Physiol - Endocrinol Metab. 2007;
152. Zamudio S, Wu Y, Ietta F, Rolfo A, Cross A, Wheeler T, et al. Human placental hypoxia-inducible factor-1α expression correlates with clinical outcomes in chronic hypoxia in vivo. Am J Pathol. 2007;
153. Semenza GL. Regulation of oxygen homeostasis by hypoxia-Inducible factor 1. Physiology. 2009.
154. Genbacev O, Zhou Y, Ludlow JW, Fisher SJ. Regulation of human placental development by oxygen tension. Science (80-). 1997;
155. Rosenfeld CS. The placenta-brain-axis. Journal of Neuroscience Research. 2021.
156. Ben-Jonathan N, Munsick RA. Dopamine and prolactin in human pregnancy. J Clin Endocrinol Metab. 1980;
157. Kim MO, Kim JH, Choi WS, Lee BH, Cho GJ, Roh SM, et al. Colocalization of Dopamine D1 and D2 Receptor mRNAs in Rat Placenta. Mol Cells. 1997;
158. Mao J, Jain A, Denslow ND, Nouri MZ, Chen S, Wang T, et al. Bisphenol A and bisphenol S disruptions of the mouse placenta and potential effects on the placenta-brain axis. Proc

Natl Acad Sci U S A. 2020;
159. Zhu Y, Zhang W, Chen M, Liu N, Guo J. Study on expression of norepinephrine and dopamine placental tissues of normal pregnancy and pregnancy induced hypertension syndrome. Zhonghua Fu Chan Ke Za Zhi. 2002;
160. Schroeder DI, Lott P, Korf I, LaSalle JM. Large-scale methylation domains mark a functional subset of neuronally expressed genes. Genome Res [Internet]. 2011 Oct 1 [cited 2018 Oct 30];21(10):1583–91. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21784875
161. Schroeder DI, Blair JD, Lott P, Yu HOK, Hong D, Crary F, et al. The human placenta methylome. Proc Natl Acad Sci U S A [Internet]. 2013;110(15):6037–42. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3625261&tool=pmcentrez&rendertype=abstract
162. Smith ZD, Chan MM, Mikkelsen TS, Gu H, Gnirke A, Regev A, et al. A unique regulatory phase of DNA methylation in the early mammalian embryo. Nature. 2012;
163. Santos F, Hyslop L, Stojkovic P, Leary C, Murdoch A, Reik W, et al. Evaluation of epigenetic marks in human embryos derived from IVF and ICSI. Hum Reprod. 2010;
164. Ehrlich M, Gama-Sosa MA, Huang LH, Midgett RM, Kuo KC, Mccune RA, et al. Amount and distribution of 5-methylcytosine in human DNA from different types of tissues or cells. Nucleic Acids Res. 1982;
165. Novakovic B, Wong NC, Sibson M, Ng HK, Morley R, Manuelpillai U, et al. DNA methylation-mediated down-regulation of DNA methyltransferase-1 (DNMT1) is coincident with, but not essential for, global hypomethylation in human placenta. J Biol Chem. 2010;
166. Smallwood SA, Kelsey G. De novo DNA methylation: A germ cell perspective. Trends Genet [Internet]. 2012;28(1):33–42. Available from: http://dx.doi.org/10.1016/j.tig.2011.09.004
167. Lister R, Pelizzola M, Kida YS, Hawkins RD, Nery JR, Hon G, et al. Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. Nature. 2011;
168. Schroeder DI, LaSalle JM. How has the study of the human placenta aided our understanding of partially methylated genes? Epigenomics [Internet]. 2013 Dec [cited 2018 Jul 30];5(6):645–54. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24283879
169. Ursini G, Punzi G, Chen Q, Marenco S, Robinson JF, Porcelli A, et al. Convergence of placenta biology and genetic risk for schizophrenia. Nat Med [Internet]. 2018 Jun 28 [cited 2018 Jun 8];24(6):792–801. Available from: http://www.nature.com/articles/s41591-018-0021-y
170. Santos HP, Bhattacharya A, Joseph RM, Smeester L, Kuban KCK, Marsit CJ, et al. Evidence for the placenta-brain axis: multi-omic kernel aggregation predicts intellectual and social impairment in children born extremely preterm. Mol Autism [Internet]. 2020 Dec 1 [cited 2021 Mar 19];11(1):97. Available from: https://doi.org/10.1186/s13229-020-00402-w
171. Schroeder DI, Schmidt RJ, Crary-Dooley FK, Walker CK, Ozonoff S, Tancredi DJ, et al. Placental methylome analysis from a prospective autism study. Mol Autism [Internet]. 2016;7(1):51. Available from: http://molecularautism.biomedcentral.com/articles/10.1186/s13229-016-0114-

8%5Cnhttp://www.ncbi.nlm.nih.gov/pubmed/28018572
172. Tsai PC, Bell JT. Power and sample size estimation for epigenome-wide association scans to detect differential DNA methylation. Int J Epidemiol. 2015;44(4):1429–41.
173. Wessels WH, Pompe van Meerdervoort M. Monozygotic twins with early infantile autism. A case report. S Afr Med J [Internet]. 1979 Jun 2 [cited 2017 Aug 10];55(23):955–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/572995
174. Hannon E, Knox O, Sugden K, Burrage J, Wong CCY, Belsky DW, et al. Characterizing genetic and environmental influences on variable DNA methylation using monozygotic and dizygotic twins. Greally JM, editor. PLOS Genet [Internet]. 2018 Aug 9 [cited 2018 Aug 16];14(8):e1007544. Available from: http://dx.plos.org/10.1371/journal.pgen.1007544
175. Grove J, Ripke S, Als TD, Mattheisen M, Walters R, Won H, et al. Common risk variants identified in autism spectrum disorder. bioRxiv [Internet]. 2017 Nov 27 [cited 2018 Jun 8];224774. Available from: https://www.biorxiv.org/content/early/2017/11/27/224774
176. Sanders SJ, He X, Willsey AJ, Ercan-Sencicek AG, Samocha KE, Cicek AE, et al. Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. Neuron [Internet]. 2015 Sep 23 [cited 2018 Aug 5];87(6):1215–33. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26402605
177. Raz R, Roberts AL, Lyall K, Hart JE, Just AC, Laden F, et al. Autism Spectrum Disorder and Particulate Matter Air Pollution before, during, and after Pregnancy: A Nested Case–Control Analysis within the Nurses' Health Study II Cohort. Environ Health Perspect [Internet]. 2015 Mar [cited 2018 Oct 23];123(3):264–70. Available from: https://ehp.niehs.nih.gov/doi/10.1289/ehp.1408133
178. Zerbo O, Iosif AM, Walker C, Ozonoff S, Hansen RL, Hertz-Picciotto I. Is Maternal Influenza or Fever during Pregnancy Associated with Autism or Developmental Delays? Results from the CHARGE (childhood Autism Risks from Genetics and Environment) Study. J Autism Dev Disord. 2013;
179. Relton CL, Wilding CS, Laffling AJ, Jonas PA, Burgess T, Binks K, et al. Low erythrocyte folate status and polymorphic variation in folate-related genes are associated with risk of neural tube defect pregnancy. Mol Genet Metab. 2004;81(4):273–81.
180. Howsmon DP, Kruger U, Melnyk S, James SJ, Hahn J. Classification and adaptive behavior prediction of children with autism spectrum disorder based upon multivariate data analysis of markers of oxidative stress and DNA methylation. PLoS Comput Biol. 2017;13(3):1–15.
181. Rush EC, Katre P, Yajnik CS. Vitamin B12: one carbon metabolism, fetal growth and programming for chronic disease. Eur J Clin Nutr [Internet]. 2014;68(1):2–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24219896
182. Kalkbrenner AE, Schmidt RJ, Penlesky AC. Environmental chemical exposures and autism spectrum disorders: a review of the epidemiological evidence. Curr Probl Pediatr Adolesc Health Care [Internet]. 2014;44(10):277–318. Available from: http://dx.doi.org/10.1016/j.cppeds.2014.06.001
183. Caramaschi D, Sharp GC, Nohr EA, Berryman K, Lewis SJ, Davey Smith G, et al. Exploring a causal role of DNA methylation in the relationship between maternal vitamin B12 during pregnancy and child's IQ at age 8, cognitive performance and educational attainment: A

two-step Mendelian randomization study. Hum Mol Genet. 2017;26(15):3001–13.
184. Zeisel SH. Importance of methyl donors during reproduction. Am J Clin Nutr [Internet]. 2009 Feb 1 [cited 2018 Sep 13];89(2):673S-677S. Available from: https://academic.oup.com/ajcn/article/89/2/673S/4596765
185. Teh AL, Pan H, Chen L, Ong ML, Dogra S, Wong J, et al. The effect of genotype and in utero environment on interindividual variation in neonate DNA methylomes. Genome Res. 2014;24(7):1064–74.
186. Smallwood SA, Kelsey G. De novo DNA methylation: a germ cell perspective. Trends Genet [Internet]. 2012 Jan [cited 2017 Aug 1];28(1):33–42. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0168952511001582
187. Schroeder DI, Jayashankar K, Douglas KC, Thirkill TL, York D, Dickinson PJ, et al. Early Developmental and Evolutionary Origins of Gene Body DNA Methylation Patterns in Mammalian Placentas. PLoS Genet. 2015;11(8):1–20.
188. Watson ED, Cross JC. Development of Structures and Transport Functions in the Mouse Placenta. Physiology [Internet]. 2005 Jun [cited 2018 Sep 25];20(3):180–93. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15888575
189. Vento-Tormo R, Efremova M, Botting RA, Turco MY, Vento-Tormo M, Meyer KB, et al. Single-cell reconstruction of the early maternal–fetal interface in humans. Nature [Internet]. 2018 Nov 14 [cited 2019 Feb 28];563(7731):347–53. Available from: http://www.nature.com/articles/s41586-018-0698-6
190. Vogel Ciernia A, Laufer BI, Dunaway KW, Mordaunt CE, Coulson RL, Yasui DH, et al. Epigenomic convergence of genetic and immune risk factors in autism brain. bioRxiv [Internet]. 2018 Feb 23 [cited 2018 Aug 5];270827. Available from: https://www.biorxiv.org/content/early/2018/02/23/270827
191. Abrahams BS, Arking DE, Campbell DB, Mefford HC, Morrow EM, Weiss LA, et al. SFARI Gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs). Mol Autism [Internet]. 2013 Oct 3 [cited 2018 Aug 15];4(1):36. Available from: http://molecularautism.biomedcentral.com/articles/10.1186/2040-2392-4-36
192. Gilissen C, Hehir-Kwa JY, Thung DT, van de Vorst M, van Bon BWM, Willemsen MH, et al. Genome sequencing identifies major causes of severe intellectual disability. Nature [Internet]. 2014 Jul 4 [cited 2018 Sep 10];511(7509):344–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24896178
193. Voineagu I, Wang X, Johnston P, Lowe JK, Tian Y, Horvath S, et al. Transcriptomic analysis of autistic brain reveals convergent molecular pathology. Nature [Internet]. 2011 Jun 25 [cited 2018 Sep 28];474(7351):380–4. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21614001
194. Gupta S, Ellis SE, Ashar FN, Moes A, Bader JS, Zhan J, et al. Transcriptome analysis reveals dysregulation of innate immune response genes and neuronal activity-dependent genes in autism. Nat Commun [Internet]. 2014 Dec 10 [cited 2019 Mar 4];5(1):5748. Available from: http://www.nature.com/articles/ncomms6748
195. Gandal MJ, Haney JR, Parikshak NN, Leppa V, Ramaswami G, Hartl C, et al. Shared molecular neuropathology across major psychiatric disorders parallels polygenic overlap. Science [Internet]. 2018 Feb 9 [cited 2019 Mar 4];359(6376):693–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/29439242

196. Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, et al. Integrative analysis of 111 reference human epigenomes. Nature [Internet]. 2015 Feb 19 [cited 2018 Jul 2];518(7539):317–30. Available from: http://www.nature.com/articles/nature14248
197. Shulha HP, Cheung I, Whittle C, Wang J, Virgil D, Lin CL, et al. Epigenetic Signatures of Autism. Arch Gen Psychiatry [Internet]. 2012 Mar 1 [cited 2018 Oct 30];69(3):314. Available from: http://archpsyc.jamanetwork.com/article.aspx?doi=10.1001/archgenpsychiatry.2011.151
198. Ernst J, Kellis M. ChromHMM: automating chromatin-state discovery and characterization. Nat Methods [Internet]. 2012 Mar 1 [cited 2018 Aug 15];9(3):215–6. Available from: http://www.nature.com/articles/nmeth.1906
199. Timp W, Bravo HC, McDonald OG, Goggins M, Umbricht C, Zeiger M, et al. Large hypomethylated blocks as a universal defining epigenetic alteration in human solid tumors. Genome Med [Internet]. 2014 Aug 26 [cited 2018 Oct 30];6(8):61. Available from: http://genomemedicine.biomedcentral.com/articles/10.1186/s13073-014-0061-y
200. Sandoval J, Heyn H, Moran S, Serra-Musach J, Pujana MA, Bibikova M, et al. Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. Epigenetics [Internet]. 2011 Jun 27 [cited 2018 Oct 30];6(6):692–702. Available from: http://www.tandfonline.com/doi/abs/10.4161/epi.6.6.16196
201. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics [Internet]. 2014 May 15 [cited 2018 Oct 30];30(10):1363–9. Available from: https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btu049
202. Sheffield NC, Bock C. LOLA: enrichment analysis for genomic region sets and regulatory elements in R and Bioconductor. Bioinformatics [Internet]. 2016 Feb 15 [cited 2018 Aug 15];32(4):587–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26508757
203. Mordaunt CE, Park BY, Bakulski KM, Feinberg JI, Croen LA, Ladd-Acosta C, et al. A meta-analysis of two high-risk prospective cohort studies reveals autism-specific transcriptional changes to chromatin, autoimmune, and environmental response genes in umbilical cord blood. bioRxiv [Internet]. 2018 Dec 4 [cited 2018 Dec 5];486498. Available from: https://www.biorxiv.org/content/early/2018/12/04/486498
204. Zhang S, Deng L, Jia Q, Huang S, Gu J, Zhou F, et al. dbMDEGA: A database for meta-analysis of differentially expressed genes in autism spectrum disorder. BMC Bioinformatics [Internet]. 2017 [cited 2018 Aug 8];18(1). Available from: http://www.ncbi.nlm.nih.gov/gds
205. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, et al. MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res [Internet]. 2009 Jul 1 [cited 2019 Mar 21];37(Web Server):W202–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19458158
206. Koudritsky M, Domany E. Positional distribution of human transcription factor binding sites. Nucleic Acids Res [Internet]. 2008 Dec [cited 2018 Sep 14];36(21):6795–805. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18953043
207. Carninci P, Sandelin A, Lenhard B, Katayama S, Shimokawa K, Ponjavic J, et al. Genome-

wide analysis of mammalian promoter architecture and evolution. Nat Genet [Internet]. 2006 Jun 28 [cited 2018 Sep 15];38(6):626–35. Available from: http://www.nature.com/articles/ng1789
208. Geschwind DH, Levitt P. Autism spectrum disorders: developmental disconnection syndromes. Curr Opin Neurobiol [Internet]. 2007 Feb 1 [cited 2018 Oct 1];17(1):103–11. Available from: https://www.sciencedirect.com/science/article/pii/S0959438807000116
209. Dapretto M, Davies MS, Pfeifer JH, Scott AA, Sigman M, Bookheimer SY, et al. Understanding emotions in others: mirror neuron dysfunction in children with autism spectrum disorders. Nat Neurosci [Internet]. 2006 Jan 4 [cited 2018 Oct 1];9(1):28–30. Available from: http://www.nature.com/articles/nn1611
210. Katoh M, Katoh M. Notch ligand, JAG1, is evolutionarily conserved target of canonical WNT signaling pathway in progenitor cells. Int J Mol Med [Internet]. 2006 Apr 1 [cited 2018 Sep 15];17(4):681–5. Available from: http://www.spandidos-publications.com/10.3892/ijmm.17.4.681
211. Klezovitch O, Vasioukhin V. Cadherin signaling: keeping cells in touch. F1000Research [Internet]. 2015 [cited 2018 Oct 1];4(F1000 Faculty Rev):550. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26339481
212. Yap AS, Kovacs EM. Direct cadherin-activated cell signaling: a view from the plasma membrane. J Cell Biol [Internet]. 2003 Jan 6 [cited 2018 Oct 1];160(1):11–6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12507993
213. Kalkman H. A review of the evidence for the canonical Wnt pathway in autism spectrum disorders. Mol Autism [Internet]. 2012 Oct 19 [cited 2018 Oct 1];3(1):10. Available from: http://molecularautism.biomedcentral.com/articles/10.1186/2040-2392-3-10
214. Krey JF, Dolmetsch RE. Molecular mechanisms of autism: a possible role for Ca2+ signaling. Curr Opin Neurobiol [Internet]. 2007 Feb 1 [cited 2018 Oct 1];17(1):112–9. Available from: https://www.sciencedirect.com/science/article/pii/S0959438807000128
215. Betancur C, Sakurai T, Buxbaum JD. The emerging role of synaptic cell-adhesion pathways in the pathogenesis of autism spectrum disorders. Trends Neurosci [Internet]. 2009 Jul 1 [cited 2018 Oct 1];32(7):402–12. Available from: https://www.sciencedirect.com/science/article/pii/S0166223609000927
216. Hofmann M, Schuster-Gossler K, Watabe-Rudolph M, Aulehla A, Herrmann BG, Gossler A. WNT signaling, in synergy with T/TBX6, controls Notch signaling by regulating Dll1 expression in the presomitic mesoderm of mouse embryos. Genes Dev [Internet]. 2004 Nov 15 [cited 2018 Nov 15];18(22):2712–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15545628
217. Lopez SJ, Dunaway K, Islam MS, Mordaunt C, Vogel Ciernia A, Meguro-Horike M, et al. UBE3A-mediated regulation of imprinted genes and epigenome-wide marks in human neurons. Epigenetics [Internet]. 2017 Nov 2 [cited 2019 Feb 28];12(11):982–90. Available from: https://www.tandfonline.com/doi/full/10.1080/15592294.2017.1376151
218. Brandler WM, Antaki D, Gujral M, Kleiber ML, Whitney J, Maile MS, et al. Paternally inherited cis-regulatory structural variants are associated with autism. Science [Internet]. 2018 Apr 20 [cited 2018 Oct 2];360(6386):327–31. Available from: http://www.ncbi.nlm.nih.gov/pubmed/29674594
219. Voineagu I, Wang X, Johnston P, Lowe JK, Tian Y, Horvath S, et al. Transcriptomic analysis

of autistic brain reveals convergent molecular pathology. Nature [Internet]. 2011 Jun 25 [cited 2018 Oct 2];474(7351):380–4. Available from: http://www.nature.com/articles/nature10110
220. Gonzalez FJ. The molecular biology of cytochrome P450s. Pharmacol Rev. 1988;40(4).
221. Jones SM, Boobis AR, Moore GE, Stanier PM. Expression of CYP2E1 during human fetal development: methylation of the CYP2E1 gene in human fetal and adult liver samples. Biochem Pharmacol [Internet]. 1992 Apr 15 [cited 2018 Dec 5];43(8):1876–9. Available from: https://www.sciencedirect.com/science/article/pii/000629529290726Y?via%3Dihub
222. Kishida M, Callard G V. Distinct Cytochrome P450 Aromatase Isoforms in Zebrafish (*Danio rerio*) Brain and Ovary Are Differentially Programmed and Estrogen Regulated during Early Development [1]. Endocrinology [Internet]. 2001 Feb 1 [cited 2018 Oct 2];142(2):740–50. Available from: https://academic.oup.com/endo/article-lookup/doi/10.1210/endo.142.2.7928
223. Ko Y, Choi I, Green ML, Simmen FA, Simmen RCM. Transient expression of the cytochrome P450 aromatase gene in elongating porcine blastocysts is correlated with uterine insulin-like growth factor levels during peri-implantation development. Mol Reprod Dev [Internet]. 1994 Jan 1 [cited 2018 Oct 2];37(1):1–11. Available from: http://doi.wiley.com/10.1002/mrd.1080370102
224. Majdic G, Sharpe RM, O'Shaughnessy PJ, Saunders PT. Expression of cytochrome P450 17alpha-hydroxylase/C17-20 lyase in the fetal rat testis is reduced by maternal exposure to exogenous estrogens. Endocrinology [Internet]. 1996 Mar 1 [cited 2018 Oct 2];137(3):1063–70. Available from: https://academic.oup.com/endo/article-lookup/doi/10.1210/endo.137.3.8603575
225. Hakkola J, Raunio H, Purkunen R, Pelkonen O, Saarikoski S, Cresteil T, et al. Detection of cytochrome P450 gene expression in human placenta in first trimester of pregnancy. Biochem Pharmacol [Internet]. 1996 Jul 26 [cited 2018 Oct 2];52(2):379–83. Available from: https://www.sciencedirect.com/science/article/pii/000629529600216X
226. Silver MJ, Kessler NJ, Hennig BJ, Dominguez-Salas P, Laritsky E, Baker MS, et al. Independent genomewide screens identify the tumor suppressor VTRNA2-1 as a human epiallele responsive to periconceptional environment. Genome Biol [Internet]. 2015 Jun 11 [cited 2018 Nov 15];16(1):118. Available from: http://genomebiology.com/2015/16/1/118
227. Jonakait GM. The effects of maternal inflammation on neuronal development: possible mechanisms. Int J Dev Neurosci [Internet]. 2007 Nov 1 [cited 2018 Sep 17];25(7):415–25. Available from: https://www.sciencedirect.com/science/article/pii/S073657480700113X?via%3Dihub
228. Boksa P. Effects of prenatal infection on brain development and behavior: A review of findings from animal models. Brain Behav Immun [Internet]. 2010 Aug 1 [cited 2018 Sep 17];24(6):881–97. Available from: https://www.sciencedirect.com/science/article/pii/S0889159110000589?via%3Dihub
229. Krakowiak P, Walker CK, Bremer AA, Baker AS, Ozonoff S, Hansen RL, et al. Maternal Metabolic Conditions and Risk for Autism and Other Neurodevelopmental Disorders. Pediatrics [Internet]. 2012;129(5):e1121–8. Available from:

http://pediatrics.aappublications.org/cgi/doi/10.1542/peds.2011-2583
230. Patel SAA, Bhambra U, Charalambous MP, David RM, Edwards RJ, Lightfoot T, et al. Interleukin-6 mediated upregulation of CYP1B1 and CYP2E1 in colorectal cancer involves DNA methylation, miR27b and STAT3. Br J Cancer [Internet]. 2014 Dec 21 [cited 2018 Sep 17];111(12):2287–96. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25333344
231. Jover R, Bort R, Gómez-Lechón MJ, Castell J V. Down-regulation of human CYP3A4 by the inflammatory signal interleukin-6: molecular mechanism and transcription factors involved. FASEB J [Internet]. 2002 Nov [cited 2018 Sep 17];16(13):1799–801. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12354697
232. Abdel-Razzak Z, Loyer P, Fautrel A, Gautier JC, Corcos L, Turlin B, et al. Cytokines down-regulate expression of major cytochrome P-450 enzymes in adult human hepatocytes in primary culture. Mol Pharmacol [Internet]. 1993 Oct [cited 2018 Sep 17];44(4):707–15. Available from: http://www.ncbi.nlm.nih.gov/pubmed/8232220
233. Hakkola J, Hu Y, Ingelman-Sundberg M. Mechanisms of Down-Regulation of CYP2E1 Expression by Inflammatory Cytokines in Rat Hepatoma Cells. J Pharmacol Exp Ther [Internet]. 2002 Nov 25 [cited 2018 Sep 17];304(3):1048–54. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12604681
234. Park HJ, Kim SK, Kang WS, Park JK, Kim YJ, Nam M, et al. Association between IRS1 Gene Polymorphism and Autism Spectrum Disorder: A Pilot Case-Control Study in Korean Males. Int J Mol Sci. 2016;17(8):1–8.
235. Sun XJ, Wang L-M, Zhang Y, Yenush L, Myers Jr MG, Glasheen E, et al. Role of IRS-2 in insulin and cytokine signalling. Nature [Internet]. 1995 Sep 14 [cited 2018 Oct 30];377(6545):173–7. Available from: http://www.nature.com/doifinder/10.1038/377173a0
236. Hussman JP, Chung R-H, Griswold AJ, Jaworski JM, Salyakina D, Ma D, et al. A noise-reduction GWAS analysis implicates altered regulation of neurite outgrowth and guidance in autism. Mol Autism [Internet]. 2011 Jan 19 [cited 2018 Aug 8];2(1):1. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21247446
237. Machado-Neto JA, de Melo Campos P, Traina F. IRS2 (insulin receptor substrate 2). Atlas Genet Cytogenet Oncol Haematol [Internet]. 2018 Jul [cited 2018 Oct 2];(2). Available from: http://hdl.handle.net/2042/68156
238. Mitra I, Lavillaureix A, Yeh E, Traglia M, Tsang K, Bearden CE, et al. Reverse Pathway Genetic Approach Identifies Epistasis in Autism Spectrum Disorders. Flint J, editor. PLOS Genet [Internet]. 2017 Jan 11 [cited 2019 Mar 5];13(1):e1006516. Available from: https://dx.plos.org/10.1371/journal.pgen.1006516
239. Carvalheira JBC, Ribeiro EB, Folli F, Velloso LA, Saad MJA. Interaction between Leptin and Insulin Signaling Pathways Differentially Affects JAK-STAT and PI 3-Kinase-Mediated Signaling in Rat Liver. Biol Chem [Internet]. 2003 Jan 27 [cited 2018 Oct 2];384(1):151–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12674509
240. Fang QL, Yin YR, Xie CR, Zhang S, Zhao WX, Pan C, et al. Mechanistic and biological significance of DNA methyltransferase 1 upregulated by growth factors in human hepatocellular carcinoma. Int J Oncol [Internet]. 2015 Feb 1 [cited 2018 Sep 24];46(2):782–90. Available from: https://www.spandidos-publications.com/10.3892/ijo.2014.2776

241. Xu G, Jing J, Bowers K, Liu B, Bao W. Maternal Diabetes and the Risk of Autism Spectrum Disorders in the Offspring: A Systematic Review and Meta-Analysis. J Autism Dev Disord [Internet]. 2014 Apr 22 [cited 2018 Dec 18];44(4):766–75. Available from: http://link.springer.com/10.1007/s10803-013-1928-2
242. Koukoura O, Sifakis S, Spandidos DA. DNA methylation in the human placenta and fetal growth (review). Mol Med Rep [Internet]. 2012 Apr [cited 2018 Jul 31];5(4):883–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22294146
243. Zeltser LM, Leibel RL. Roles of the placenta in fetal brain development. Proc Natl Acad Sci U S A [Internet]. 2011 Sep 20 [cited 2018 Oct 2];108(38):15667–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21890794
244. Xu L-M, Li J-R, Huang Y, Zhao M, Tang X, Wei L. AutismKB: an evidence-based knowledgebase of autism genetics. Nucleic Acids Res [Internet]. 2012 Jan [cited 2018 Oct 2];40(Database issue):D1016-1022. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22139918
245. Lord C, Risi S, Lambrecht L, Cook EH, Leventhal BL, DiLavore PC, et al. Autism Diagnostic Observation Schedule (ADOS). J Autism Dev Disord. 2000;30(3):205–23.
246. Lord C, Rutter M, Le Couteur A. Autism Diagnostic Interview-Revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. J Autism Dev Disord. 1994;24(5):659–85.
247. Mullen E. Mullen scales of early learning [Internet]. 1995 [cited 2018 Oct 23]. Available from: http://www.v-psyche.com/doc/special-cases/Mullen Scales of Early Learning.docx
248. Chawarska K, Shic F, Macari S, Campbell DJ, Brian J, Landa R, et al. 18-month predictors of later outcomes in younger siblings of children with autism spectrum disorder: a baby siblings research consortium study. J Am Acad Child Adolesc Psychiatry [Internet]. 2014 Dec [cited 2018 Oct 23];53(12):1317-1327.e1. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25457930
249. Ozonoff S, Young GS, Belding A, Hill M, Hill A, Hutman T, et al. The broader autism phenotype in infancy: When does it emerge? J Am Acad Child Adolesc Psychiatry. 2014;53(4).
250. Guo W, Fiziev P, Yan W, Cokus S, Sun X, Zhang MQ, et al. BS-Seeker2: a versatile aligning pipeline for bisulfite sequencing data. BMC Genomics [Internet]. 2013 Nov 10 [cited 2018 Jun 9];14(1):774. Available from: http://bmcgenomics.biomedcentral.com/articles/10.1186/1471-2164-14-774
251. Coulson RL, Yasui DH, Dunaway KW, Laufer BI, Vogel Ciernia A, Zhu Y, et al. Snord116-dependent diurnal rhythm of DNA methylation in mouse cortex. Nat Commun [Internet]. 2018 Dec 24;9(1):1616. Available from: http://dx.doi.org/10.1038/s41467-018-03676-0
252. Box JF. R. A. Fisher and the Design of Experiments, 1922-1926. Am Stat [Internet]. 1980 Feb [cited 2018 Jul 31];34(1):1. Available from: https://www.jstor.org/stable/2682986?origin=crossref
253. Wilks DS. Cluster Analysis. Int Geophys [Internet]. 2011 Jan 1 [cited 2018 Dec 4];100:603–16. Available from: https://www.sciencedirect.com/science/article/pii/B9780123850225000154
254. McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, et al. GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol [Internet]. 2010 May 2

[cited 2018 Jun 15];28(5):495–501. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20436461
255. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene Ontology: tool for the unification of biology. Nat Genet [Internet]. 2000 May [cited 2018 Aug 14];25(1):25–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10802651
256. The Gene Ontology Consortium. Expansion of the Gene Ontology knowledgebase and resources. Nucleic Acids Res [Internet]. 2017 Jan 4 [cited 2018 Aug 14];45(D1):D331–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27899567
257. Thomas PD, Campbell MJ, Kejariwal A, Mi H, Karlak B, Daverman R, et al. PANTHER: a library of protein families and subfamilies indexed by function. Genome Res [Internet]. 2003 Sep 1 [cited 2018 Aug 15];13(9):2129–41. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12952881
258. Mi H, Muruganujan A, Thomas PD. PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. Nucleic Acids Res [Internet]. 2012 Nov 26 [cited 2018 Aug 15];41(D1):D377–86. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23193289
259. Shen L, Shao N-Y, Liu X, Maze I, Feng J, Nestler EJ. diffReps: Detecting Differential Chromatin Modification Sites from ChIP-seq Data with Biological Replicates. Mantovani R, editor. PLoS One [Internet]. 2013 Jun 10 [cited 2018 Aug 15];8(6):e65598. Available from: http://dx.plos.org/10.1371/journal.pone.0065598
260. Harold D, Abraham R, Hollingworth P, Sims R, Gerrish A, Hamshere ML, et al. Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. Nat Genet [Internet]. 2009 Oct 6 [cited 2018 Oct 30];41(10):1088–93. Available from: http://www.nature.com/articles/ng.440
261. Landi MT, Chatterjee N, Yu K, Goldin LR, Goldstein AM, Rotunno M, et al. A Genome-wide Association Study of Lung Cancer Identifies a Region of Chromosome 5p15 Associated with Risk for Adenocarcinoma. Am J Hum Genet [Internet]. 2009 Nov 13 [cited 2018 Oct 29];85(5):679–91. Available from: https://www.sciencedirect.com/science/article/pii/S000292970900408X
262. Sloan CA, Chan ET, Davidson JM, Malladi VS, Strattan JS, Hitz BC, et al. ENCODE data at the ENCODE portal. Nucleic Acids Res [Internet]. 2016 Jan 4 [cited 2018 Aug 15];44(D1):D726–32. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26527727
263. ENCODE Project Consortium TEP. An integrated encyclopedia of DNA elements in the human genome. Nature [Internet]. 2012 Sep 6 [cited 2018 Aug 15];489(7414):57–74. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22955616
264. Ernst J, Kellis M. Chromatin-state discovery and genome annotation with ChromHMM. Nat Protoc [Internet]. 2017 Nov 9 [cited 2018 Jul 2];12(12):2478–92. Available from: http://www.nature.com/doifinder/10.1038/nprot.2017.124
265. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The human genome browser at UCSC. Genome Res [Internet]. 2002 Jun 1 [cited 2018 Aug 15];12(6):996–1006. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12045153
266. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. Nat Methods [Internet]. 2012 Jul 1 [cited 2018 Aug 15];9(7):671–5. Available from: http://www.nature.com/articles/nmeth.2089

267. Weiner DJ, Wigdor EM, Ripke S, Walters RK, Kosmicki JA, Grove J, et al. Polygenic transmission disequilibrium confirms that common and rare variation act additively to create risk for autism spectrum disorders. Nat Genet. 2017 Jul 1;49(7):978–85.
268. Bourgeron T. From the genetic architecture to synaptic plasticity in autism spectrum disorder. Nat Rev Neurosci [Internet]. 2015 Aug 20 [cited 2017 Aug 1];16(9):551–63. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26289574
269. Raz R, Roberts AL, Lyall K, Hart JE, Just AC, Laden F, et al. Autism spectrum disorder and particulate matter air pollution before, during, and after pregnancy: A nested case–control analysis within the nurses' health study II cohort. Environ Health Perspect. 2015;
270. Schaevitz LR, Berger-Sweeney JE. Gene-Environment Interactions and Epigenetic Pathways in Autism: The Importance of One-Carbon Metabolism. ILAR J. 2012;
271. Schaevitz L, Berger-Sweeney J, Ricceri L. One-carbon metabolism in neurodevelopmental disorders: Using broad-based nutraceutics to treat cognitive deficits in complex spectrum disorders. Neuroscience and Biobehavioral Reviews. 2014.
272. Li Y, Qiu S, Shi J, Guo Y, Li Z, Cheng Y, et al. Association between MTHFR C677T/A1298C and susceptibility to autism spectrum disorders: a meta-analysis. BMC Pediatr [Internet]. 2020 Sep 24 [cited 2020 Oct 3];20(1):449. Available from: /pmc/articles/PMC7517654/?report=abstract
273. Vucetic Z, Kimmel J, Totoki K, Hollenbeck E, Reyes TM. Maternal high-fat diet alters methylation and gene expression of dopamine and opioid-related genes. Endocrinology. 2010;
274. Yajnik CS, Deshmukh US. Fetal programming: Maternal nutrition and role of one-carbon metabolism. Rev Endocr Metab Disord [Internet]. 2012 Jun 14 [cited 2018 Sep 25];13(2):121–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22415298
275. Costello PM, Rowlerson A, Astaman NA, Anthony FEW, Sayer AA, Cooper C, et al. Peri-implantation and late gestation maternal undernutrition differentially affect fetal sheep skeletal muscle development. J Physiol. 2008;
276. Croen LA, Goines P, Braunschweig D, Yolken R, Yoshida CK, Grether JK, et al. Brain-derived neurotrophic factor and autism: Maternal and infant peripheral blood levels in the early markers for autism (EMA) study. Autism Res. 2008;
277. Haugen AC, Schug TT, Collman G, Heindel JJ. Evolution of DOHaD: The impact of environmental health sciences. Journal of Developmental Origins of Health and Disease. 2015.
278. Xu J, He G, Zhu J, Zhou X, Clair DS, Wang T, et al. Prenatal nutritional deficiency reprogrammed postnatal gene expression in mammal brains: Implications for schizophrenia. Int J Neuropsychopharmacol. 2014;
279. Tylee DS, Hess JL, Quinn TP, Barve R, Huang H, Zhang-James Y, et al. Blood transcriptomic comparison of individuals with and without autism spectrum disorder: A combined-samples mega-analysis. Am J Med Genet Part B Neuropsychiatr Genet. 2017;
280. Ansel A, Rosenzweig JP, Zisman PD, Melamed M, Gesundheit B. Variation in gene expression in autism spectrum disorders: An extensive review of transcriptomic studies. Frontiers in Neuroscience. 2017.
281. Mordaunt CE, Park BY, Bakulski KM, Feinberg JI, Croen LA, Ladd-Acosta C, et al. A meta-analysis of two high-risk prospective cohort studies reveals autism-specific

transcriptional changes to chromatin, autoimmune, and environmental response genes in umbilical cord blood. Mol Autism. 2019;
282. Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, et al. The GeneMANIA prediction server: Biological network integration for gene prioritization and predicting gene function. Nucleic Acids Res. 2010;
283. Sánchez-Martín P, Komatsu M. Physiological Stress Response by Selective Autophagy. J Mol Biol [Internet]. 2020 Jan 3 [cited 2020 Apr 2];432(1):53–62. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0022283619303778
284. Hui KK, Tanaka M. Autophagy links MTOR and GABA signaling in the brain. Autophagy. 2019.
285. Kaminsky EB, Kaul V, Paschall J, Church DM, Bunke B, Kunig D, et al. An evidence-based approach to establish the functional and clinical significance of copy number variants in intellectual and developmental disabilities. Genet Med. 2011 Sep;13(9):777–84.
286. Sajan SA, Fernandez L, Nieh SE, Rider E, Bukshpun P, Wakahiro M, et al. Both Rare and De Novo Copy Number Variants Are Prevalent in Agenesis of the Corpus Callosum but Not in Cerebellar Hypoplasia or Polymicrogyria. PLoS Genet. 2013 Oct;9(10).
287. Klamt J, Hofmann A, Böhmer AC, Hoebel AK, Gölz L, Becker J, et al. Further evidence for deletions in 7p14.1 contributing to nonsyndromic cleft lip with or without cleft palate. Birth Defects Res Part A - Clin Mol Teratol. 2016;
288. Wenger TL, Kao C, McDonald-Mcginn DM, Zackai EH, Bailey A, Schultz RT, et al. The Role of mGluR Copy Number Variation in Genetic and Environmental Forms of Syndromic Autism Spectrum Disorder. Sci Rep. 2016;
289. Papúchová H, Meissner TB, Li Q, Strominger JL, Tilburgs T. The Dual Role of HLA-C in Tolerance and Immunity at the Maternal-Fetal Interface. Vol. 10, Frontiers in Immunology. Frontiers Media S.A.; 2019. p. 2730.
290. Torres AR, Maciulis A, Stubbs EG, Cutler A, Odell D. The transmission disequilibrium test suggests that HLA-DR4 and DR13 are linked to autism spectrum disorder. Hum Immunol. 2002;
291. Torres AR, Sweeten TL, Cutler A, Bedke BJ, Fillmore M, Stubbs EG, et al. The Association and Linkage of the HLA-A2 Class I Allele with Autism. Human Immunology. 2006.
292. Saresella M, Marventano I, Guerini FR, Mancuso R, Ceresa L, Zanzottera M, et al. An Autistic Endophenotype Results in Complex Immune Dysfunction in Healthy Siblings of Autistic Children. Biol Psychiatry. 2009;
293. Guerini FR, Bolognesi E, Manca S, Sotgiu S, Zanzottera M, Agliardi C, et al. Family-based transmission analysis of HLA genetic markers in Sardinian children with autistic spectrum disorders. Hum Immunol. 2009;
294. Torres AR, Westover JB, Gibbons C, Johnson RC, Ward DC. Activating killer-cell immunoglobulin-like receptors (KIR) and their cognate HLA ligands are significantly increased in autism. Brain Behav Immun. 2012;
295. Guerini FR, Bolognesi E, Chiappedi M, Manca S, Ghezzo A, Agliardi C, et al. Activating KIR molecules and their cognate ligands prevail in children with a diagnosis of ASD and in their mothers. Brain Behav Immun. 2014;
296. al-Haddad BJS, Oler E, Armistead B, Elsayed NA, Weinberger DR, Bernier R, et al. The fetal origins of mental illness [Internet]. Vol. 221, American Journal of Obstetrics and

Gynecology. Mosby Inc.; 2019 [cited 2020 Oct 3]. p. 549–62. Available from: https://pubmed.ncbi.nlm.nih.gov/31207234/
297. Goines PE, Croen LA, Braunschweig D, Yoshida CK, Grether J, Hansen R, et al. Increased midgestational IFN-γ, IL-4 and IL-5 in women bearing a child with autism: A case-control study. Mol Autism. 2011;
298. Krakowiak P, Goines PE, Tancredi DJ, Ashwood P, Hansen RL, Hertz-Picciotto I, et al. Neonatal Cytokine Profiles Associated With Autism Spectrum Disorder. Biol Psychiatry. 2017;
299. Zimmerman AW, Pessah IN, Lein PJ. Evidence for Environmental Susceptibility in Autism. In: Autism. Humana Press; 2008. p. 409–28.
300. Stamou M, Streifel KM, Goines PE, Lein PJ. Neuronal connectivity as a convergent target of gene × environment interactions that confer risk for Autism Spectrum Disorders. Vol. 36, Neurotoxicology and Teratology. 2013. p. 3–16.
301. Racioppi L, Means AR. Calcium/calmodulin-dependent kinase IV in immune and inflammatory responses: novel routes for an ancient traveller. Vol. 29, Trends in Immunology. Elsevier; 2008. p. 600–7.
302. Katrancha SM, Wu Y, Zhu M, Eipper BA, Koleske AJ, Mains RE. Neurodevelopmental disease-associated de novo mutations and rare sequence variants affect TRIO GDP/GTP exchange factor activity. Hum Mol Genet [Internet]. 2017 Dec 1 [cited 2019 Dec 30];26(23):4728–40. Available from: https://academic.oup.com/hmg/article/26/23/4728/4158143
303. Geisheker MR, Heymann G, Wang T, Coe BP, Turner TN, Stessman HAF, et al. Hotspots of missense mutation identify neurodevelopmental disorder genes and functional domains. Nat Neurosci. 2017;
304. Devlin B, Boone BE, Levy SE, Lihm J, Buxbaum JD, Wu Y, et al. Patterns and rates of exonic de novo mutations in autism spectrum disorders. Nature. 2012.
305. Yoshihara Y, Kawasaki M, Tamada A, Nagata S, Kagamiyama H, Mori K. Overlapping and differential expression of BIG-2, BIG-1, TAG-1, and F3: Four members of an axon-associated cell adhesion molecule subgroup of the immunoglobulin superfamily. J Neurobiol. 1995;
306. Fernandez T, Morgan T, Davis N, Klin A, Morris A, Farhi A, et al. Disruption of contactin 4 (CNTN4) results in developmental delay and other features of 3p deletion syndrome. Am J Hum Genet. 2004;
307. Ferland RJ, Cherry TJ, Preware PO, Morrisey EE, Walsh CA. Characterization of Foxp2 and Foxp1 mRNA and protein in the developing and mature brain. J Comp Neurol. 2003;
308. Teramitsu I, Kudo LC, London SE, Geschwind DH, White SA. Parallel FoxP1 and FoxP2 Expression in Songbird and Human Brain Predicts Functional Interaction. J Neurosci. 2004;
309. Mordaunt CE, Jianu JM, Laufer B, Zhu Y, Dunaway KW, Bakulski KM, et al. Cord blood DNA methylome in newborns later diagnosed with autism spectrum disorder reflects early dysregulation of neurodevelopmental and X-linked genes. bioRxiv. 2019 Nov 21;850529.
310. Cukier HN, Lee JM, Ma D, Young JI, Mayo V, Butler BL, et al. The Expanding Role of MBD Genes in Autism: Identification of a MECP2 Duplication and Novel Alterations in MBD5,

MBD6, and SETDB1. Autism Res [Internet]. 2012 Dec [cited 2020 Apr 2];5(6):385–97. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23055267
311. Cukier HN, Rabionet R, Konidari I, Rayner-Evans MY, Baltos ML, Wright HH, et al. Novel variants identified in methyl-CpG-binding domain genes in autistic individuals. Neurogenetics. 2010 Nov 18;11(3):291–303.
312. Mungall AJ. Meeting review: Epigenetics in development and disease. In: Comparative and Functional Genomics. 2002. p. 277–81.
313. Fouse SD, Nagarajan RP, Costello JF. Genome-scale DNA methylation analysis. Vol. 2, Epigenomics. 2010. p. 105–17.
314. Zhou C, Wang Y, Zhang J, Su J, An Q, Liu X, et al. H3K27me3 is an epigenetic barrier while KDM6A overexpression improves nuclear reprogramming efficiency. FASEB J. 2019 Mar;33(3):4638–52.
315. Scaglione F, Panzavolta G. Folate, folic acid and 5-methyltetrahydrofolate are not the same thing [Internet]. Vol. 44, Xenobiotica. Informa Healthcare; 2014 [cited 2020 Apr 2]. p. 480–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24494987
316. Moretti P, Sahoo T, Hyland K, Bottiglieri T, Peters S, Del Gaudio D, et al. Cerebral folate deficiency with developmental delay, autism, and response to folinic acid. Neurology. 2005 Mar 22;64(6):1088–90.
317. Waterland RA, Jirtle RL. Early nutrition, epigenetic changes at transposons and imprinted genes, and enhanced susceptibility to adult chronic diseases. Nutrition. 2004;
318. Afman L, Müller M. Nutrigenomics: From molecular nutrition to prevention of disease. J Am Diet Assoc. 2006;
319. Mordaunt CE, Kieffer DA, Shibata NM, Członkowska A, Litwin T, Weiss K-H, et al. Epigenomic signatures in liver and blood of Wilson disease patients include hypermethylation of liver-specific enhancers. Epigenetics Chromatin [Internet]. 2019 Dec 1 [cited 2019 Feb 3];12(1):10. Available from: https://epigeneticsandchromatin.biomedcentral.com/articles/10.1186/s13072-019-0255-z
320. Craig SAS. Betaine in human nutrition. American Journal of Clinical Nutrition. 2004.
321. Paparo L, Di Costanzo M, Di Scala C, Cosenza L, Leone L, Nocerino R, et al. The influence of early life nutrition on epigenetic regulatory mechanisms of the immune system. Nutrients. 2014;
322. Zeisel SH, Blusztajn JK. Choline and Human Nutrition. Annu Rev Nutr. 1994;
323. Ueland PM, Holm PI, Hustad S. Betaine: A key modulator of one-carbon metabolism and homocysteine status. Clinical Chemistry and Laboratory Medicine. 2005.
324. Friesen RW, Novak EM, Hasman D, Innis SM. Relationship of Dimethylglycine, Choline, and Betaine with Oxoproline in Plasma of Pregnant Women and Their Newborn Infants. J Nutr. 2007;
325. Krause M, Dent EW, Bear JE, Loureiro JJ, Gertler FB. Ena/VASP Proteins: Regulators of the Actin Cytoskeleton and Cell Migration. Annu Rev Cell Dev Biol. 2003;
326. Gardiner EJ, Cairns MJ, Liu B, Beveridge NJ, Carr V, Kelly B, et al. Gene expression analysis reveals schizophrenia-associated dysregulation of immune pathways in peripheral blood mononuclear cells. J Psychiatr Res. 2013;
327. Tsunoda F, Lamon-Fava S, Asztalos BF, Iyer LK, Richardson K, Schaefer EJ. Effects of oral

eicosapentaenoic acid versus docosahexaenoic acid on human peripheral blood mononuclear cell gene expression. Atherosclerosis. 2015;
328. Ginsberg MR, Rubin RA, Falcone T, Ting AH, Natowicz MR. Brain Transcriptional and Epigenetic Associations with Autism. PLoS One. 2012;
329. Glatt SJ, Tsuang MT, Winn M, Chandler SD, Collins M, Lopez L, et al. Blood-based gene expression signatures of infants and toddlers with autism. J Am Acad Child Adolesc Psychiatry. 2012;
330. Rutter M, LeCouteur A, Lord C. Autism Diagnostic Interview - Revised (ADI-R). Statew Agric L Use Baseline 2015. 2015;1.
331. Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, Speed TP. Summaries of Affymetrix GeneChip probe level data. Nucleic Acids Res [Internet]. 2003 Feb 15 [cited 2019 Apr 24];31(4):e15. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12582260
332. Bolstad BM, Irizarry RA, Astrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics [Internet]. 2003 Jan 22 [cited 2019 Apr 24];19(2):185–93. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12538238
333. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics [Internet]. 2003 Apr 1 [cited 2019 Apr 24];4(2):249–64. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12925520
334. Carvalho BS, Irizarry RA. A framework for oligonucleotide microarray preprocessing. Bioinformatics [Internet]. 2010 Oct 1 [cited 2019 Apr 24];26(19):2363–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20688976
335. Kauffmann A, Gentleman R, Huber W. arrayQualityMetrics—a bioconductor package for quality assessment of microarray data. Bioinformatics [Internet]. 2009 Feb 1 [cited 2019 Apr 24];25(3):415–6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19106121
336. Hollowood K, Melnyk S, Pavliv O, Evans T, Sides A, Schmidt RJ, et al. Maternal metabolic profile predicts high or low risk of an autism pregnancy outcome. Res Autism Spectr Disord. 2018 Dec 1;56:72–82.
337. Melnyk S, Fuchs GJ, Schulz E, Lopez M, Kahler SG, Fussell JJ, et al. Metabolic imbalance associated with methylation dysregulation and oxidative damage in children with autism. J Autism Dev Disord [Internet]. 2012 Mar [cited 2020 Mar 9];42(3):367–77. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21519954
338. Talwar D, Quasim T, McMillan DC, Kinsella J, Williamson C, O'Reilly DSJ. Optimisation and validation of a sensitive high-performance liquid chromatography assay for routine measurement of pyridoxal 5-phosphate in human plasma and red cells using pre-column semicarbazide derivatisation. J Chromatogr B Anal Technol Biomed Life Sci. 2003;
339. Holm PI, Ueland PM, Kvalheim G, Lien EA. Determination of choline, betaine, and dimethylglycine in plasma by a high-throughput method based on normal-phase chromatography-tandem mass spectrometry. Clin Chem. 2003;
340. Yan J, Wang W, Gregory JF, Malysheva O, Brenna JT, Stabler SP, et al. MTHFR C677T genotype influences the isotopic enrichment of one-carbon metabolites in folate-compromised men consuming d9-choline. Am J Clin Nutr. 2011;
341. Wang Y, Wang T, Shi X, Wan D, Zhang P, He X, et al. Analysis of acetylcholine, choline and

butyrobetaine in human liver tissues by hydrophilic interaction liquid chromatography-tandem mass spectrometry. J Pharm Biomed Anal. 2008;
342. Yan J, Jiang X, West AA, Perry CA, Malysheva O V., Devapatla S, et al. Maternal choline intake modulates maternal and fetal biomarkers of choline metabolism in humans. Am J Clin Nutr. 2012;
343. Pfeiffer CM, Fazili Z, McCoy L, Zhang M, Gunter EW. Determination of Folate Vitamers in Human Serum by Stable-Isotope-Dilution Tandem Mass Spectrometry and Comparison with Radioassay and Microbiologic Assay. Clin Chem. 2004;
344. West AA, Yan J, Perry CA, Jiang X, Malysheva O V., Caudill MA. Folate-status response to a controlled folate intake in nonpregnant, pregnant, and lactating women. Am J Clin Nutr. 2012;
345. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The SVA package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics. 2012;
346. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;
347. Li S. GeneOverlap [Internet]. GeneOverlap: Test and visualize gene overlaps. 2019. Available from: https://www.bioconductor.org/packages/release/bioc/html/GeneOverlap.html
348. Zhang B, Horvath S. A General Framework for Weighted Gene Co-Expression Network Analysis. Stat Appl Genet Mol Biol [Internet]. 2005 Jan 12 [cited 2019 Apr 29];4(1). Available from: https://www.degruyter.com/view/j/sagmb.2005.4.issue-1/sagmb.2005.4.1.1128/sagmb.2005.4.1.1128.xml
349. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics [Internet]. 2008 Dec 29 [cited 2019 Apr 29];9(1):559. Available from: https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-9-559
350. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics. 2007;
351. Gentle JE, Kaufman L, Rousseuw PJ. Finding Groups in Data: An Introduction to Cluster Analysis. Biometrics. 1991;
352. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. Nat Methods. 2015 Apr 29;12(5):453–7.
353. Gaugler T, Klei L, Sanders SJ, Bodea CA, Goldberg AP, Lee AB, et al. Most genetic risk for autism resides with common variation. Nat Genet. 2014;
354. Clarke TK, Lupton MK, Fernandez-Pujals AM, Starr J, Davies G, Cox S, et al. Common polygenic risk for autism spectrum disorder (ASD) is associated with cognitive ability in the general population. Mol Psychiatry [Internet]. 2016 Mar 1 [cited 2021 Mar 19];21(3):419–25. Available from: https://pubmed.ncbi.nlm.nih.gov/25754080/
355. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. Nature [Internet]. 2014 Jul 22 [cited 2018 Jun 8];511(7510):421–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25056061
356. Corley MJ, Vargas-Maya N, Pang APS, Lum-Jones A, Li D, Khadka V, et al. Epigenetic Delay

in the Neurodevelopmental Trajectory of DNA Methylation States in Autism Spectrum Disorders. Front Genet [Internet]. 2019 Oct 1 [cited 2021 Mar 19];10:907. Available from: https://www.frontiersin.org/article/10.3389/fgene.2019.00907/full
357. Gunasekara CJ, Scott CA, Laritsky E, Baker MS, MacKay H, Duryea JD, et al. A genomic atlas of systemic interindividual epigenetic variation in humans. Genome Biol. 2019;
358. Kessler NJ, Waterland RA, Prentice AM, Silver MJ. Establishment of environmentally sensitive DNA methylation states in the very early human embryo. Sci Adv. 2018;
359. Sloan CA, Chan ET, Davidson JM, Malladi VS, Strattan JS, Hitz BC, et al. ENCODE data at the ENCODE portal. Nucleic Acids Res [Internet]. 2016 Jan 4 [cited 2018 Aug 15];44(D1):D726–32. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26527727
360. Ota T, Suzuki Y, Nishikawa T, Otsuki T, Sugiyama T, Irie R, et al. Complete sequencing and characterization of 21,243 full-length human cDNAs. Nat Genet. 2004;
361. Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, et al. The Genotype-Tissue Expression (GTEx) project. Nature Genetics. 2013.
362. Simon MC, Keith B. The role of oxygen availability in embryonic development and stem cell function. Nature Reviews Molecular Cell Biology. 2008.
363. Audano PA, Sulovari A, Graves-Lindsay TA, Cantsilieris S, Sorensen M, Welch AME, et al. Characterizing the Major Structural Variant Alleles of the Human Genome. Cell. 2019 Jan 24;176(3):663-675.e19.
364. Homo sapiens isolate CHM1 chromosome 22 22-49000000:0, whole genome sh - Nucleotide - NCBI [Internet]. [cited 2021 Mar 20]. Available from: https://www.ncbi.nlm.nih.gov/nuccore/QPKN01007947
365. Collins RL, Brand H, Karczewski KJ, Zhao X, Alföldi J, Francioli LC, et al. A structural variation reference for medical and population genetics. Nature. 2020;
366. Collins RL, Brand H, Karczewski KJ, Zhao X, Alföldi J, Francioli LC, et al. INS_22_115103 [Internet]. Available from: https://gnomad.broadinstitute.org/variant/INS_22_115103?dataset=gnomad_sv_r2_1
367. Kaiser VB, Semple CA. Chromatin loop anchors are associated with genome instability in cancer and recombination hotspots in the germline. Genome Biol [Internet]. 2018 Jul 30 [cited 2021 Mar 20];19(1). Available from: https://pubmed.ncbi.nlm.nih.gov/30060743/
368. Boix CA, James BT, Park YP, Meuleman W, Kellis M. Regulatory genomic circuitry of human disease loci by integrative epigenomics. Nature [Internet]. 2021 Feb 11 [cited 2021 Mar 20];590(7845):300–7. Available from: https://doi.org/10.1038/s41586-020-03145-z
369. Schmitt AD, Hu M, Jung I, Xu Z, Qiu Y, Tan CL, et al. A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome. Cell Rep. 2016;
370. Bernier R, Golzio C, Xiong B, Stessman HA, Coe BP, Penn O, et al. Disruptive CHD8 mutations define a subtype of autism early in development. Cell. 2014;
371. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Research. 2000.
372. Johannessen M, Haugen IB, Bakken TL, Braaten Ø. A 22q13.33 duplication harbouring the SHANK3 gene: Does it cause neuropsychiatric disorders? BMJ Case Rep. 2019;
373. Han K, Holder JL, Schaaf CP, Lu H, Chen H, Kang H, et al. SHANK3 overexpression causes manic-like behaviour with unique pharmacogenetic properties. Nature. 2013;

374. Dennis MY, Eichler EE. Human adaptation and evolution by segmental duplication. Current Opinion in Genetics and Development. 2016.
375. Mordaunt CE, Jianu JM, Laufer BI, Zhu Y, Hwang H, Dunaway KW, et al. Cord blood DNA methylome in newborns later diagnosed with autism spectrum disorder reflects early dysregulation of neurodevelopmental and X-linked genes. Genome Med [Internet]. 2020 Dec 14 [cited 2020 Oct 15];12(1):88. Available from: https://genomemedicine.biomedcentral.com/articles/10.1186/s13073-020-00785-8
376. Newschaffer CJ, Croen LA, Fallin MD, Hertz-Picciotto I, Nguyen D V, Lee NL, et al. Infant siblings and the investigation of autism risk factors. J Neurodev Disord [Internet]. 2012;4(1):7. Available from: http://www.jneurodevdisorders.com/content/4/1/7
377. Ladd-Acosta C, Andrews S V, Bakulski KM, Feinberg JI, Tryggvadottir R, Yao R, et al. Placenta DNA methylation at ZNF300 is associated with fetal sex and placental morphology. bioRxiv [Internet]. 2021 Mar 8 [cited 2021 Mar 19];2021.03.05.433992. Available from: https://doi.org/10.1101/2021.03.05.433992
378. Laufer BI, Hwang H, Jianu JM, Mordaunt CE, Korf IF, Hertz-Picciotto I, et al. Low-Pass Whole Genome Bisulfite Sequencing of Neonatal Dried Blood Spots Identifies a Role for RUNX1 in Down Syndrome DNA Methylation Profiles. Hum Mol Genet. 2020;
379. Krueger F, Andrews SR. Bismark: A flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics. 2011;
380. Krueger F. Trim Galore!: A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files. Babraham Inst. 2015;
381. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;
382. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics [Internet]. 2009 Aug [cited 2021 Jan 4];25(16):2078–9. Available from: https://pubmed.ncbi.nlm.nih.gov/19505943/
383. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: Summarize analysis results for multiple tools and samples in a single report. Bioinformatics. 2016;
384. Hansen KD, Langmead B, Irizarry RA. BSmooth: from whole genome bisulfite sequencing reads to differentially methylated regions. Genome Biol. 2012;
385. Vu VQ. ggbiplot: A ggplot2 based biplot. R package version 0.55. Vu, Vincent Q. 2011.
386. Triche TJ, Weisenberger DJ, Van Den Berg D, Laird PW, Siegmund KD. Low-level processing of Illumina Infinium DNA Methylation BeadArrays. Nucleic Acids Res. 2013;
387. Yuan V, Price EM, Del Gobbo G, Mostafavi S, Cox B, Binder AM, et al. Accurate ethnicity prediction from placental DNA methylation data. Epigenetics and Chromatin. 2019;
388. Laufer B. GitHub - ben-laufer/DMRichR: An executable and package for the statistical analysis and visualization of differentially methylated regions (DMRs) from CpG count matrices (Bismark cytosine reports) [Internet]. Available from: https://github.com/ben-laufer/DMRichR
389. Korthauer K, Chakraborty S, Benjamini Y, Irizarry RA. Detection and accurate false discovery rate control of differentially methylated regions from whole genome bisulfite sequencing. Biostatistics [Internet]. 2019 Jul 1 [cited 2019 Sep 11];20(3):367–83. Available from: https://academic.oup.com/biostatistics/article/20/3/367/4899074
390. McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, et al. GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol [Internet].

2010;28(5):495–501. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20436461
391. Cavalcante RG, Sartor MA. Annotatr: Genomic regions in context. Bioinformatics. 2017;
392. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. Mol Cell. 2010;
393. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;
394. Broad Institute. Picard Tools - By Broad Institute. Github. 2009.
395. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res [Internet]. 2010 Sep 1 [cited 2019 Sep 11];20(9):1297–303. Available from: http://genome.cshlp.org/cgi/doi/10.1101/gr.107524.110
396. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res [Internet]. 2010 Sep 1 [cited 2019 Sep 11];38(16):e164–e164. Available from: https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkq603
397. Boeva V, Popova T, Bleakley K, Chiche P, Cappo J, Schleiermacher G, et al. Control-FREEC: A tool for assessing copy number and allelic content using next-generation sequencing data. Bioinformatics. 2012;
398. Wang J, Mullighan CG, Easton J, Roberts S, Heatley SL, Ma J, et al. CREST maps somatic structural variation in cancer genomes with base-pair resolution. Nat Methods. 2011;
399. Rausch T, Zichner T, Schlattl A, Stütz AM, Benes V, Korbel JO. DELLY: Structural variant discovery by integrated paired-end and split-read analysis. Bioinformatics. 2012;
400. Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT. Data quality control in genetic case-control association studies. Nat Protoc [Internet]. 2010 Aug 26 [cited 2021 Mar 21];5(9):1564–73. Available from: https://pubmed.ncbi.nlm.nih.gov/21085122/
401. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet [Internet]. 2007 [cited 2021 Mar 21];81(3):559–75. Available from: /pmc/articles/PMC1950838/
402. Michigan Imputation Server [Internet]. [cited 2021 Mar 21]. Available from: https://imputationserver.sph.umich.edu/index.html#!pages/home
403. Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, et al. Next-generation genotype imputation service and methods. Nat Genet [Internet]. 2016 Oct 1 [cited 2021 Mar 21];48(10):1284–7. Available from: https://pubmed.ncbi.nlm.nih.gov/27571263/
404. Auton A, Abecasis GR, Altshuler DM, Durbin RM, Bentley DR, Chakravarti A, et al. A global reference for human genetic variation [Internet]. Vol. 526, Nature. Nature Publishing Group; 2015 [cited 2021 Mar 21]. p. 68–74. Available from: https://pubmed.ncbi.nlm.nih.gov/26432245/
405. Sudmant PH, Rausch T, Gardner EJ, Handsaker RE, Abyzov A, Huddleston J, et al. An integrated map of structural variation in 2,504 human genomes. Nature [Internet]. 2015 Sep 30 [cited 2021 Mar 21];526(7571):75–81. Available from: https://pubmed.ncbi.nlm.nih.gov/26432246/

406. Loh PR, Danecek P, Palamara PF, Fuchsberger C, Reshef YA, Finucane HK, et al. Reference-based phasing using the Haplotype Reference Consortium panel. Nat Genet [Internet]. 2016 Nov 1 [cited 2021 Mar 21];48(11):1443–8. Available from: https://pubmed.ncbi.nlm.nih.gov/27694958/
407. Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TL. NCBI BLAST: a better web interface. Nucleic Acids Res. 2008;
408. Parsons JD. Miropeats: graphical DNA sequence comparisons. Bioinformatics [Internet]. 1995 Dec 1 [cited 2021 Mar 21];11(6):615–9. Available from: https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/11.6.615
409. Jurka J. Repeats in genomic DNA: Mining and meaning. Curr Opin Struct Biol. 1998;
410. RepeatMasker Home Page [Internet]. [cited 2021 Jan 6]. Available from: http://www.repeatmasker.org/
411. Yusa K, Zhou L, Li MA, Bradley A, Craig NL. A hyperactive piggyBac transposase for mammalian applications. Proc Natl Acad Sci U S A. 2011;
412. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: Ultrafast universal RNA-seq aligner. Bioinformatics. 2013;
413. Andrews S, Krueger F, Seconds-Pichon A, Biggins F, Wingett S. FastQC. A quality control tool for high throughput sequence data. Babraham Bioinformatics. Babraham Institute. 2015.
414. Liao Y, Smyth GK, Shi W. FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;
415. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;
416. Yu G, Wang LG, Han Y, He QY. ClusterProfiler: An R package for comparing biological themes among gene clusters. Omi A J Integr Biol. 2012;
417. Zhu Y. GitHub - ASD Epigenetics and Genetics Biomarkers [Internet]. Available from: https://github.com/Yihui-Zhu/Epigenetics_Genetics_ASD_Biomarker
418. Andrews S V., Sheppard B, Windham GC, Schieve LA, Schendel DE, Croen LA, et al. Case-control meta-analysis of blood DNA methylation and autism spectrum disorder. Mol Autism. 2018;
419. Pramparo T, Lombardo M V, Campbell K, Barnes CC, Marinero S, Solso S, et al. Cell cycle networks link gene expression dysregulation, mutation, and brain maldevelopment in autistic toddlers. Mol Syst Biol. 2015;
420. Shulha HP, Cheung I, Whittle C, Wang J, Virgil D, Lin CL, et al. Epigenetic Signatures of Autism. Arch Gen Psychiatry [Internet]. 2012 Mar 1 [cited 2019 Mar 19];69(3):314. Available from: http://archpsyc.jamanetwork.com/article.aspx?doi=10.1001/archgenpsychiatry.2011.151
421. Di Nanni N, Bersanelli M, Cupaioli FA, Milanesi L, Mezzelani A, Mosca E. Network-based integrative analysis of genomics, epigenomics and transcriptomics in autism spectrum disorders. Int J Mol Sci. 2019;
422. Berson A, Nativio R, Berger SL, Bonini NM. Epigenetic Regulation in Neurodegenerative Diseases. Trends in Neurosciences. 2018.

423. Stilling RM, Fischer A. The role of histone acetylation in age-associated memory impairment and Alzheimer's disease. Neurobiol Learn Mem. 2011;
424. Contestabile A, Sintoni S. Histone Acetylation in Neurodevelopment. Curr Pharm Des. 2013;
425. Qin L, Ma K, Wang ZJ, Hu Z, Matas E, Wei J, et al. Social deficits in Shank3-deficient mouse models of autism are rescued by histone deacetylase (HDAC) inhibition. Nat Neurosci. 2018;
426. Hu Z, Zhou J, Jiang J, Yuan J, Zhang Y, Wei X, et al. Genomic characterization of genes encoding histone acetylation modulator proteins identifies therapeutic targets for cancer treatment. Nat Commun. 2019;
427. Mishima Y, Wang C, Miyagi S, Saraya A, Hosokawa H, Mochizuki-Kashio M, et al. Histone acetylation mediated by Brd1 is crucial for Cd8 gene activation during early thymocyte development. Nat Commun. 2014;
428. Severinsen JE, Bjarkam CR, Kiar-Larsen S, Olsen IM, Nielsen MM, Blechingberg J, et al. Evidence implicating BRD1 with brain development and susceptibility to both schizophrenia and bipolar affective disorder. Mol Psychiatry. 2006;
429. Fryland T, Christensen JH, Pallesen J, Mattheisen M, Palmfeldt J, Bak M, et al. Identification of the BRD1 interaction network and its impact on mental disorder risk. Genome Med. 2016;
430. Andreassen OA, Thompson WK, Dale AM. Boosting the power of schizophrenia genetics by leveraging new statistical tools. Schizophr Bull. 2014;
431. Sando R, Gounko N, Pieraut S, Liao L, Yates J, Maximov A. HDAC4 governs a transcriptional program essential for synaptic plasticity and memory. Cell. 2012;
432. Nardone S, Sharan Sams D, Reuveni E, Getselter D, Oron O, Karpuj M, et al. DNA methylation analysis of the autistic brain reveals multiple dysregulated biological pathways. Transl Psychiatry. 2014;
433. Pinto D, Delaby E, Merico D, Barbosa M, Merikangas A, Klei L, et al. Convergence of genes and cellular pathways dysregulated in autism spectrum disorders. Am J Hum Genet. 2014;
434. Illumina. Infinium MethylationEPIC Product Files [Internet]. Available from: https://support.illumina.com/downloads/infinium-methylationepic-v1-0-product-files.html
435. Ziats MN, Rennert OM. Aberrant Expression of Long Noncoding RNAs in Autistic Brain. J Mol Neurosci. 2013;
436. Wang Y, Zhao X, Ju W, Flory M, Zhong J, Jiang S, et al. Genome-wide differential expression of synaptic long noncoding RNAs in autism spectrum disorder. Transl Psychiatry. 2015;
437. Quinn JJ, Chang HY. Unique features of long non-coding RNA biogenesis and function. Nature Reviews Genetics. 2016.
438. Ransohoff JD, Wei Y, Khavari PA. The functions and unique features of long intergenic non-coding RNA. Nature Reviews Molecular Cell Biology. 2018.
439. Pollard KS, Salama SR, Lambert N, Lambot MA, Coppens S, Pedersen JS, et al. An RNA gene expressed during cortical development evolved rapidly in humans. Nature. 2006;
440. Mercer TR, Mattick JS. Structure and function of long noncoding RNAs in epigenetic

regulation. Nature Structural and Molecular Biology. 2013.
441. Galindo MI, Pueyo JI, Fouix S, Bishop SA, Couso JP. Peptides encoded by short ORFs control development and define a new eukaryotic gene family. PLoS Biol. 2007;
442. Pang Y, Mao C, Liu S. Encoding activities of non-coding RNAs. Theranostics. 2018.
443. Wang S, Mao C, Liu S. Peptides encoded by noncoding genes: Challenges and perspectives. Signal Transduction and Targeted Therapy. 2019.
444. Kondo T, Hashimoto Y, Kato K, Inagaki S, Hayashi S, Kageyama Y. Small peptide regulators of actin-based cell morphogenesis encoded by a polycistronic mRNA. Nat Cell Biol. 2007;
445. Detzer A, Engel C, Wünsche W, Sczakiel G. Cell stress is related to re-localization of Argonaute 2 and to decreased RNA interference in human cells. Nucleic Acids Res. 2011;
446. Huang JZ, Chen M, Chen D, Gao XC, Zhu S, Huang H, et al. A Peptide Encoded by a Putative lncRNA HOXB-AS3 Suppresses Colon Cancer Growth. Mol Cell. 2017;
447. Infinium Multi-Ethnic Genotyping Array [Internet]. Available from: https://support.illumina.com/array/array_kits/infinium-multi-ethnic-global-8-kit/downloads.html
448. Dunford AR, Sangster JM. Maternal and paternal periconceptional nutrition as an indicator of offspring metabolic syndrome risk in later life through epigenetic imprinting: A systematic review. Diabetes and Metabolic Syndrome: Clinical Research and Reviews. 2017.
449. Hutsler JJ, Zhang H. Increased dendritic spine densities on cortical projection neurons in autism spectrum disorders. Brain Res. 2010;
450. Phillips M, Pozzo-Miller L. Dendritic spine dysgenesis in autism related disorders. Neuroscience Letters. 2014.
451. Durand CM, Perroy J, Loll F, Perrais D, Fagni L, Bourgeron T, et al. SHANK3 mutations identified in autism lead to modification of dendritic spine morphology via an actin-dependent mechanism. Mol Psychiatry. 2012;
452. Chauhan A, Chauhan V. Oxidative stress in autism. Pathophysiology. 2006.
453. Frustaci A, Neri M, Cesario A, Adams JB, Domenici E, Dalla Bernardina B, et al. Oxidative stress-related biomarkers in autism: Systematic review and meta-analyses. Free Radical Biology and Medicine. 2012.
454. James SJ, Melnyk S, Jernigan S, Cleves MA, Halsted CH, Wong DH, et al. Metabolic endophenotype and related genotypes are associated with oxidative stress in children with autism. Am J Med Genet Part B Neuropsychiatr Genet. 2006;
455. Deth R, Muratore C, Benzecry J, Power-Charnitsky VA, Waly M. How environmental and genetic factors combine to cause autism: A redox/methylation hypothesis. NeuroToxicology. 2008.
456. Melnyk S, Jernigan S, Savenka A, James SJ. Elevation in S-adenosylhomocysteine and DNA hypomethylation in parents and children with autism. FASEB J. 2007;
457. Nunomura A, Moreira P, Lee H, Zhu X, Castellani R, Smith M, et al. Neuronal Death and Survival Under Oxidative Stress in Alzheimer and Parkinson Diseases. CNS Neurol Disord - Drug Targets. 2008;
458. James SJ, Cutler P, Melnyk S, Jernigan S, Janak L, Gaylor DW, et al. Metabolic biomarkers of increased oxidative stress and impaired methylation capacity in children with autism. Am J Clin Nutr. 2004;

459. Fagiolini M, Jensen CL, Champagne FA. Epigenetic influences on brain development and plasticity. Current Opinion in Neurobiology. 2009.
460. Marchetto MC, Belinson H, Tian Y, Freitas BC, Fu C, Vadodaria KC, et al. Altered proliferation and networks in neural cells derived from idiopathic autistic individuals. Mol Psychiatry. 2017;
461. Prives C. Signaling to p53: Breaking the MDM2-p53 circuit. Cell. 1998.
462. Chenn A, Walsh CA. Regulation of cerebral cortical size by control of cell cycle exit in neural precursors. Science (80-). 2002;
463. Wilkinson B, Grepo N, Thompson BL, Kim J, Wang K, Evgrafov O V., et al. The autism-associated gene chromodomain helicase DNA-binding protein 8 (CHD8) regulates noncoding RNAs and autism-related genes. Transl Psychiatry. 2015;
464. Ali M, Rincón-Arano H, Zhao W, Rothbart SB, Tong Q, Parkhurst SM, et al. Molecular basis for chromatin binding and regulation of MLL5. Proc Natl Acad Sci U S A. 2013;
465. O ' Roak BJ, Vives L, Girirajan S, Karakoc E, Krumm N, Coe BP, et al. Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. Nature. 2012;
466. Willsey AJ, Sanders SJ, Li M, Dong S, Tebbenkamp AT, Muhle RA, et al. Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. Cell. 2013 Nov 21;155(5):997.
467. Kang HJ, Kawasawa YI, Cheng F, Zhu Y, Xu X, Li M, et al. Spatio-temporal transcriptome of the human brain. Nature. 2011;
468. Deng LW, Chiu I, Strominger JL. MLL5 protein forms intranuclear foci, and overexpression inhibits cell cycle progression. Proc Natl Acad Sci U S A. 2004;
469. Cheng F, Liu J, Zhou SH, Wang XN, Chew JF, Deng LW. RNA interference against mixed lineage leukemia 5 resulted in cell cycle arrest. Int J Biochem Cell Biol. 2008;
470. Shen E, Shulha H, Weng Z, Akbarian S. Regulation of histone H3K4 methylation in brain development and disease. Vol. 369, Philosophical Transactions of the Royal Society B: Biological Sciences. 2014.
471. Shulha HP, Cheung I, Whittle C, Wang J, Virgil D, Lin CL, et al. Epigenetic signatures of autism: Trimethylated H3K4 landscapes in prefrontal neurons. Arch Gen Psychiatry. 2012;
472. Hnisz D, Abraham BJ, Lee TI, Lau A, Saint-André V, Sigova AA, et al. Super-enhancers in the control of cell identity and disease. Cell. 2013;
473. Roelfsema JH, White SJ, Ariyürek Y, Bartholdi D, Niedrist D, Papadia F, et al. Genetic heterogeneity in Rubinstein-Taybi syndrome: Mutations in both the CBP and EP300 genes cause disease. Am J Hum Genet. 2005;
474. Moon KJ, Mochizuki K, Zhou M, Jeong HS, Brady JN, Ozato K. The bromodomain protein Brd4 is a positive regulatory component of P-TEFb and stimulates RNA polymerase II-dependent transcription. Mol Cell. 2005;
475. Valor L, Viosca J, Lopez-Atalaya J, Barco A. Lysine Acetyltransferases CBP and p300 as Therapeutic Targets in Cognitive and Neurodegenerative Disorders. Curr Pharm Des. 2013;
476. Chekler ELP, Pellegrino JA, Lanz TA, Denny RA, Flick AC, Coe J, et al. Transcriptional Profiling of a Selective CREB Binding Protein Bromodomain Inhibitor Highlights

Therapeutic Opportunities. Chem Biol. 2015;
477. Iossifov I, Levy D, Allen J, Ye K, Ronemus M, Lee YH, et al. Low load for disruptive mutations in autism genes and their biased transmission. Proc Natl Acad Sci U S A. 2015;
478. Zawel L, Le Dai J, Buckhaults P, Zhou S, Kinzler KW, Vogelstein B, et al. Human Smad3 and Smad4 are sequence-specific transcription activators. Mol Cell. 1998;
479. Nguyen A, Rauch TA, Pfeifer GP, Hu VW. Global methylation profiling of lymphoblastoid cell lines reveals epigenetic contributions to autism spectrum disorders and a novel autism candidate gene, RORA , whose protein product is reduced in autistic brain. FASEB J. 2010;
480. Sayad A, Noroozi R, Omrani MD, Taheri M, Ghafouri-Fard S. Retinoic acid-related orphan receptor alpha (RORA) variants are associated with autism spectrum disorder. Metab Brain Dis. 2017;
481. Sato TK, Panda S, Miraglia LJ, Reyes TM, Rudic RD, McNamara P, et al. A functional genomics strategy reveals rora as a component of the mammalian circadian clock. Neuron. 2004;
482. Glickman G. Circadian rhythms and sleep in children with autism. Neuroscience and Biobehavioral Reviews. 2010.
483. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature [Internet]. 2012 Sep 6 [cited 2018 Aug 15];489(7414):57–74. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22955616
484. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res. 2016;