DEFINING A FUNCTIONAL ROLE FOR 5-HYDROXYMETHYLCYTOSINE IN DEVELOPMENTAL BRAIN DISORDERS

By

ANDY MADRID

A dissertation submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy
(Neuroscience)

at the
UNIVERSITY OF WISCONSIN – MADISON
2025

Date of oral examination: May 24, 2021

The dissertation is approved by the following members of the Final Committee:

Reid S. Alisch, Assistant Professor, Neuroscience Training Program
Anthony P. Auger, Professor, Neuroscience Training Program
Chiara Cirelli, Professor, Neuroscience Training Program
Sunduz Keles, Professor, Neuroscience Training Program
Ahna R. Skop, Professor, Genetics Training Program
Rupa Sridharan, Associate Professor, Genetics Training Program

To my mother, whose sacrifices and unconditional love

have forged me into the person that I am today.

ACKNOWLEDGMENTS

The works in this dissertation would not have been possible without support and aid from numerous people. First and foremost, I would like to thank Drs. Reid Alisch and Ligia Papale, two key members of the lab. It has been through their countless hours of effort in training me that any of this work was made possible. Before joining the lab, I had little-to-no wet lab experience. However, they both decided to take a chance on me, though – in my eyes – it was a very big risk. Though there were several highs and several lows and plenty of uncertainties both of them were always there to guide me and help mold me into the independent researcher that I am today. They both have given me numerous opportunities to prosper and flourish in the lab. I would like to believe that – in their eyes – all of this was worthwhile.

Although the lab was a critical source of my strength, my family was, likewise, an important factor in my completion of this thesis work. In particular, my mother's undying love and sacrifices have led me to become the person that I am today. From the time I was a child and just learning about the world, she was always there for me, being a guiding voice and always having my best interests at heart – even if it meant sacrifice and hardship on her part. It is safe to say that without her, none of this would have been possible. Were it not for my drive to make her proud and to show her that all of her years and years of hard work were not in vain, I would not have pushed myself further than I have ever gone before, such as by earning this degree. All of this, everything that I have ever accomplished in my life is my way of saying 'thank you' to her.

While my mother has been a driving force for me attempting to make something of myself, others have also left a lasting mark on my persona. My uncle Edward, who is like a father to me, has always been a fountain of wisdom. Whenever I needed help in making a difficult life decision, he has always been the person I turned to for advice. My sisters, Raylene and Linda, have always been there to make me laugh when I was jaded or distressed during these last few years, and they have always been my confidants. My grandmother, Herlinda, has always been like a second mother to me. From the time I was a child, whenever I would get home from school, she would always be there waiting for me with a snack and would help me with my homework. And, while I myself am not religious, seeing her dedication and love for her religion has always had a lasting impact on me and a source of inspiration for me.

Though I could fill the remaining pages of this thesis simply giving thanks to the endless number of individuals that have left an imprint in my life and my thinking, I will only list a few here. I would like to thank my best friend, Rey, who has always been there for me through thick and thin. From the moment we randomly met each other at a bus stop in San Francisco, to nearly nine years later and living in different states, he has been the person I know I can trust with anything. And, whenever I needed to vent or decompress, he has always been there to help me through my trying times. Martin and Kat and Odin have always been here for me in Madison. Whether it was going to get a beer to forget about the day, going over to their house for a meal, making our yearly Christmas tamales, or just taking care of Odin (he’s a dog) they have helped to keep me sane during these years of my PhD training. Eamon was always a voice of reason for me during my time

here in Madison. Though, sometimes, I did not listen to that voice of reason, if I needed some quick advice in life or with my science, he was always around to be a listening ear.

Lastly, while it may seem odd, I would also like to acknowledge my dog, Olive. Though she eats too much, takes up too much space on the bed, and is exceptionally needy at times, she has been there with me during my times of greatest need. She has seen me at my best and at my worst. Perhaps it is because she does not know any better and does not see any of my flaws, but her unconditional love for me and seeing the excitement in her eyes as she literally jumps up to greet me every day after lab have made even the hardest days during these last six years worth it.

Thank you, to all. None of this is solely my work, solely my accomplishment. Rather, my life and my achievements are a mosaic, a mosaic from the lessons, the laughs, and the moments that I have shared with so many people that have helped me to find myself. My sincerest thanks. I am forever in your debt.

TABLE OF CONTENTS

ABBREVIATIONS

5hmC	5-hydrxoymethylcytosine
5mC	5-methylcytosine
ASD	autism spectrum disorder
ChIP	chromatin immunoprecipitation
CLOCK	circadian locomotor output cycles kaput
CNS	central nervous system
DBD	developmental brain disorder
DhMR	differentially hydroxymethylated region
DMR	differentially methylated region
DNA	deoxyribonucleic acid
DNAm	DNA methylation
ELS	early-life stress
EMSA	electrophoretic mobility shift assay
HET	heterozygous
HPA-axis	hypothalamic-pituitary adrenal axis
HOMO	homozygous
RNA	ribonucleic acid
RNAseq	ribonucleic acid sequencing
TF	transcription factor
WB	western blot
WT	wildtype

ABSTRACT

Developmental brain disorders (DBD), such as autism spectrum disorders, usually present years after birth; however, the molecular pathogenesis is thought to arise earlier, either during pregnancy or just after birth. Genetic studies have revealed hundreds of rare risk genes that cumulatively account for less than 25% of developmental brain disorder cases. Thus, for the majority of cases, their etiology remains unknown. Emerging evidence indicates that environmental influences (*e.g.,* early-life stress) contribute to the development of brain-related disorders. 5-hydroxymethylcytosine (5hmC) is an environmentally sensitive epigenetic mark that is highly enriched in neuronal cells and is associated with the regulation of neuronal activity, suggesting that 5hmC plays an important role in coordinating transcriptional activity in the brain. Disruptions in 5hmC levels have been reported in a myriad of diseases and disorders, including cancers, developmental brain disorders (*e.g.,* autism), and neurodegenerative diseases (*e.g,* Alzheimer's disease). However, despite a wealth of genome-wide association studies characterizing 5hmC profiles in these diseases, a clear molecular function of 5hmC and its mechanistic role have been slow to emerge. Here, I begin by comprehensively discussing the emerging role of 5hmC in mental health and disease, and discuss the evidence for its putative function(s). Next, I follow up on previous research from our lab, by identifying sex-specific disruptions of striatal 5hmC in a mouse model of autism (*i.e., Cntnap2*$^{-/-}$), finding both common and unique sets of genes showing 5hmC perturbations in female and male mutants, and revealing parallels with genes and pathways with known involvement in autism. Considering that *Cntnap2*$^{-/-}$ mutants exhibited disruptions of 5hmC abundance and that 5hmC is an environmentally-sensitive epigenetic modification, I

tested the hypothesis that a gene by environment interaction can contribute to the development of DBD-like phenotypes by subjecting wildtype and *Cntnap2*$^{+/-}$ mutants, that display no phenotypic aberrations, to prenatal stress from embryonic day (E) 12-18, the timing of which coincides with the onset of *Cntnap2* expression (E14). At three months of age, the prenatally-stressed *Cntnap2*$^{+/-}$ female mice exhibited reduced sociability and repetitive behaviors, which parallel the phenotypes of the *Cntnap2*$^{-/-}$ mutants. In addition, genomic profiling of 5hmC levels from independent cohorts and brain tissues (hippocampus and striatum: regions associated with social and stress-related behaviors) revealed a significant overlap of differentially hydroxymethylated genes between both tissues (*N*=478 genes). Genomic regions displaying modulations in 5hmC abundance were found to be significantly enriched for sequence motifs that putatively bind transcription factors that are pivotal for the maintenance of neurodevelopment, such as CLOCK, suggesting that 5hmC may mediate transcription factor binding affinity. To test this claim, immunoblot, chromatin immunoprecipitation sequencing, and electrophoretic mobility shift assays were coupled to directly interrogate the role of 5hmC in CLOCK binding. These data revealed that 5hmC represses CLOCK binding to three genes involved in autism (*e.g., Gigyf1, Palld, Fry*) that exhibit alterations in 5hmC, gene expression, and CLOCK binding, as a consequence of the gene by environment interaction. Together, these data support a direct role of 5hmC in regulating the binding of transcription machinery, linking this mechanistic component of 5hmC in the molecular etiology of developmental brain disorders.

CHAPTER 1

The emerging role of 5-hydroxymethylcytosine in mental health and disease

This chapter is adapted from:

Madrid, A., Papale, L.A., and Alisch, R.S. (2016). New hope: the emerging role of 5-hydroxymethylcytosine in mental health and disease. *Epigenomics*

Abstract

Historically biomedical research has examined genetic influences on mental health but these approaches have had limited success, likely due to the broad heritability of brain-related disorders (*e.g.,* 30-90%). Epigenetic modifications, such as DNA methylation, are environmentally sensitive mechanisms that may play a role in the origins and progression of mental illness. Recently, genome-wide disruptions of 5-hydroxymethylcytosine (5hmC) were associated with the development of early- and late-onset mental illnesses such as autism and Alzheimer's disease, bringing new hope to the field of psychiatry. Here, I review the recent links of 5hmC to mental illness and discuss several putative functions of 5hmC in the context of its promising clinical relevance.

I. Introduction

A fundamental problem in psychiatry is that diagnoses are based on subjective measures that are primarily acquired through behavioral assessments. Modern genetics has held out hope that one could identify new genetic markers of mental illness early in childhood that could aid the subjective measures of behavior used today. New lab tests for these markers would then indicate to psychiatrists the most promising drug for each person or direct pharmaceutical companies to more specific personalized drug development. However, even though complete mapping of the human genome was announced in 2003, the genetic approach has had little impact on improving the lives of individuals struggling with mental illness. This outcome is partially caused by the broad range of heritability in mental illness (*i.e.,* 30-90%) but also is due to a failure in molecular psychiatry that has overlooked a crucial molecular step that causes variations in the expression of key genes in the brain. Past approaches have largely ignored how environmental stimuli can affect the unique amount of gene expression in the brain of each person. The use of this personalized information may provide new hope that one day will make it possible to develop a precise molecular diagnosis that will change the way mental health is treated today.

Epigenetic mechanisms are environmentally sensitive molecular modifications that are heritable but do not change the underlying DNA sequence. Much effort has been put forth toward understanding two specific epigenetic mechanisms: histone modifications, which alter gene expression through the restructuring of chromatin proteins, and DNA methylation, which modulate gene expression through the addition of a methyl group to

specific sites within the DNA sequence. The most studied epigenetic mechanism in the mammalian genome is DNA methylation, which is the addition of a methyl group on the fifth carbon of cytosine (*e.g.,* 5-methylcytosine (5mC)). This DNA mark functions in genomic imprinting, X-chromosome inactivation, chromatin structure, and gene silencing [1-5]. For a more definitive and technical perspective of 5mC, please refer to Box 1. Human studies support a role for 5mC in the development of mental pathologies including bipolar disorder, schizophrenia, and major depressive disorder, often resulting in concomitant changes in gene expression [6-10]. It was recently shown that 5mC can be oxidized to 5-hydroxymethylcytosine (5hmC) and that this modification is environmentally sensitive [11], accumulates throughout neurodevelopment[12] and highly enriched in the brain [13-17]. While the molecular role(s) of 5hmC remains elusive, its unique enrichment in the brain compared to peripheral tissue (unlike 5mC)[18] leads many to believe that 5hmC functions are independent from those of 5mC, and that 5hmC may be a critical molecular building block in mental health that is disrupted in disease. For a more definitive and technical perspective of 5hmC, please refer to Box 1. It is noteworthy that traditional DNA methylation detection methods utilizing sodium bisulfite treatment cannot distinguish between the methylated and hydroxymethylated forms of cytosine, meaning that past studies that utilized such methods report a composite of 5mC and 5hmC but have attributed any findings solely to 5mC. The following chapter will provide evidence to support the independent importance of 5hmC in mental health and the development of mental illness and present several putative molecular functions of 5hmC that may shed light on its promising clinical relevance. With recent findings implicating a unique role for

5hmC in developmental brain disorders, 5hmC brings a new frontier to the field of Psychiatry.

II. 5hmC: from the embryo to neurodegeneration

Embryogenesis

5hmC is essential for proper mammalian development, beginning with an early role in the distinction between embryonic stem cells and neural progenitor cells [19]. Here, the transition from embryonic stem to neural progenitor cells requires a genome-wide reduction of 5hmC, which is facilitated by a dramatic change in the expression of the modulatory proteins required for oxidizing 5mC to 5hmC (*Tet1-3*; Box 1) [20]. These findings link 5hmC and *Tet* expression with embryonic stem cell differentiation and suggest that aberrant 5hmC levels may impact neurogenesis. Consistent with this link, embryos simultaneously lacking TET1 and TET3 experience lower 5hmC and higher 5mC levels, resulting in altered expression of important neurodevelopmental genes, including those in the cholesterol biosynthesis pathway that lead to the development of holoprosencephaly, a cephalic disorder characterized by the failure of the two brain hemispheres to properly form [21]. However, mice individually lacking TET1, TET2, or TET3 survived beyond birth, suggesting redundancy among the TET proteins during early embryogenesis. Another study identified that a non-catalytic, non-DNA-oxidizing, action of TET3 is required in neural stem cells to prevent premature differentiation into non-neurogenic astrocytes, by binding to DNA encoding the gene *Snrpn,* repressing its expression[22]. Following up on this line of research concerning TET proteins and neuronal differentiation, it was found that neuronal differentiation requires an up-regulation of *Tet3*

and a down-regulation of *Tet1*[23]. Knockdown of *Tet3* in neural progenitor cells led to reactivation of pluripotency-associated genes, including *Oct4, Nanog,* and *Tcl1*, and to an increased rate of OCT4-positive cells, suggesting de-differentiation of cell populations. Together, these reports support that proper maintenance of 5hmC levels through the TET family of enzymes is crucial during early stages of neurogenesis, and that perturbations result in improper cell differentiation.

Neurodevelopment and Aging

While 5mC has been found to play a critical role in regulating and maintaining proper neurogenesis, in both developing and adult brains[24,25], less is known about the function and extent of 5hmC as a contributor in neurodevelopment. However, considering that a number of studies have demonstrated a role for 5hmC in the maintenance of embryogenesis, and the high abundance of 5hmC in brain tissue, via alterations in TET3 expression, these data suggest that 5hmC is essential for proper neurodevelopment, and that aberrations in 5hmC levels could lead to abnormal neurodevelopmental trajectories, contributing to neurological diseases. A recent report identified that distribution of 5hmC within chromatin changed throughout neurodevelopment, which paralleled changes in neuronal development and differentiation, indicating that 5hmC profiles are linked with neurodevelopment[26]. Indeed, 5hmC continues to be abundant postnatally and undergoes age-dependent accumulation throughout early life development and into adulthood, suggesting that 5hmC marks accumulate and are stable across neurodevelopment[12,27,28]. Moreover, when comparing human fetal and adult cerebellar tissue, specific loci were found to be fetus-specific, while others were found to be adult-specific, suggesting that

5hmC marks different regions of the genome at different stages of neurodevelopment[29]. Surprisingly, in the mouse hippocampus, age-associated increases of 5hmC were not consequential with fluctuations of *Tet1-3* mRNA expression, nor 8-hydroxy-2-deoxy-guanosine content, a marker of DNA oxidation, suggesting age-accumulated 5hmC is not related to oxidative stress and is not entirely regulated by TET proteins in aged mice[27]. Together, these findings suggest that 5hmC abundance that accumulates throughout neurodevelopment is important for neuron differentiation, development, and function in the developing and the adult brain.

Neurodegeneration

The emerging link between 5hmC and brain development, and its accumulation with age, have led researchers to examine this epigenetic mark in neurodegenerative diseases [30]. Indeed, age-associated genes that acquire 5hmC are associated with pathways related to neurodegenerative diseases [31]. For example, depletion of 5hmC was found in the hippocampus, cerebellum, and entorhinal cortex of patients suffering from Alzheimer's disease (AD) [32,33], a type of dementia that usually develops slowly and gets worse over time. On the other hand, enrichment of 5hmC in the frontal and mid-temporal gyrus was positively correlated with hallmarks of AD, including neurofibrillary tangles, amyloid beta levels, and ubiquitin load[34]. Additionally, in another cohort, while no global differences in 5hmC levels were identified in the mid-temporal gyrus of late-onset Alzheimer's disease patients, locus-specific differences of 5hmC in genes, such as *CHRNB1*, controlling acetylcholine signaling were found[35]. Together, these findings suggest that 5hmC may be driving the primary molecular components of AD progression within distinct regions of the

brain. Notably, AD-associated levels of 5hmC can be detected at preclinical stages as well as at later stages of AD [36], indicating that 5hmC may act as a viable biomarker of AD onset and progression. However, studies employing nuclear labeling or enzyme-linked immunosorbent assay (ELISA) were unable to find significant alterations in 5hmC associated with AD [37], suggesting that higher resolution methods are required to identify AD-associated changes in 5hmC. Connections between 5hmC and other neurodegenerative disorders including Huntington's, Ataxia-telangiectasia, and Fragile X-associated tremor/ataxia syndrome have recently surfaced[38-41]. These neurodegenerative disease-associated changes in 5hmC levels often arise within distinct cell-types and brain regions, supporting cell and tissue-specific development of these diseases. Since these studies have been largely descriptive, it will be imperative to determine the functional mechanism(s) played by 5hmC if we are to modify it toward healthy outcomes.

III. The sensitivity of 5hmC to environmental stimuli

While environmental stressors have been shown to alter gene expression [42], a definitive molecular culprit has been slow to emerge; perhaps 5hmC is a promising candidate. An initial investigation supporting this hypothesis showed that aged mice (18-months old) exposed to an enriched environment exhibited reduced 5hmC abundance in the hippocampus primarily in the gene bodies of genes involved in axon guidance. These changes in 5hmC levels also were associated with increased learning and memory, suggesting that environmental enrichment might positively modulate 5hmC in the hippocampus, underscoring the latent importance of 5hmC as a molecular regulator of

mental health and capacity [43]. Others have found that environmental influences, such as caloric restriction, can alter age-associated accumulation of 5hmC. Here, restricting caloric intake of mice prevented the age-related increase of 5hmC in hippocampal and cerebellar tissue [44]. This study was one of the first to find direct links between diet, aging, and 5hmC profiles, connecting the aging process with epigenetic changes and environmental stimuli. Further expanding on the link between environmental influences, such as diet and disruptions in 5hmC profiles, a recent study using a mouse model of perinatal protein malnutrition found associated changes in anxiety-like behaviors that were rescued by exposure to an enriched environment, suggesting environmental stimuli, without genetic mutations, influence behavioral outcomes, potentially through epigenetic mechanisms[45]. Indeed, when interrogating the ventral hippocampus of protein malnourished female mice compared to naïve female mice, genome-wide disruptions of 5hmC were identified in genes governing processes concerning dendrite outgrowth, such as *Fltr3*, and epigenetic machinery, such as *Dicer1*. These findings establish a role for 5hmC in early-life malnourishment that modulates phenotypic outcomes. Moreover, malnourished mice subjected to an enriched environment displayed a restoration of 5hmC levels in a number genes disrupted by protein malnutrition, further underscoring a role of 5hmC as either influencing behavioral outcomes, or its value as a marker of anxiety-like behaviors and in revealing underlying genes. Together, these studies highlight the sensitivity of 5hmC to environmental stimuli and suggest that it is a putative regulator of behavioral outcomes.

While diet and enriched environments can have long-lasting effects on 5hmC abundance, 5hmC also displays an immediate response on the genome following acute stress. For example, an increase in 5hmC levels in the 3'-UTR of a gene best known for its role in the stress response (the glucocorticoid receptor gene, *Nr3c1*) was identified in the hippocampus of mice stressed for 30 minutes and allowed to recover for 1 hour[46]. Moreover, genome-wide 5hmC analysis of these same mice revealed genome-wide disruptions of 5hmC and confirmed an overall increase in 5hmC levels following stress[47]. These altered 5hmC levels were found within the binding sites of several transcription factors (TFs) and near genes that were differentially expressed and have known roles in neurogenesis and neurological activities. These findings suggest that in response to stress 5hmC may influence TF binding to modulate gene expression levels. The fact that 5hmC changes were found within one hour following a short stress highlights the potential for rapid changes of 5hmC levels within the brain, and that 5hmC can be a bioindicator of acute stress. It will be interesting to examine the long-term effects of short stress (*i.e.,* more than one hour after exposure) or how chronic stress alters 5hmC levels to reveal novel regions that are susceptible to alterations brought about by environmental stress, potentially identifying stress-related 5hmC-assocaited biomarkers.

While these studies focused on hippocampal involvement in stress-response, others have shown that environmental stimuli can affect 5hmC levels in other brain regions. One example showed that mice exposed to repeated administrations of cocaine have increased 5hmC in the nucleus accumbens, primarily in coding regions and enhancer sequences of genes involved in drug addiction [48]. Notably, these 5hmC changes only

persisted in a small subset of loci one month after cocaine exposure, suggesting that these epigenetic changes are largely reversible. Another study showed that fear extinction, a form of reverse learning, results in dramatic 5hmC changes in the prefrontal cortex of mice, supporting that 5hmC aids in the regulation of rapid behavioral adaptation [49]. Interestingly, in another study, *Tet3*, but not *Tet1*, mediated the modulation of 5hmC levels and concomitant increased gene expression that was associated with fear extinction. Similarly, another group found that mice lacking the expression of *Tet1,* while not showing significant differences in learning or memory acquisition, exhibited impaired memory extinction, coupled with long-term synaptic depression and down-regulation of neuronal activity-related genes [50]. Thus, while *Tet3* may solely facilitate the accumulation of 5hmC in the prefrontal cortex of mice during rapid behavior adaptation in response to fear, *Tet1* governs alterations in 5hmC on synaptic plasticity genes during behavioral adaptation in response to stressful environmental exposures.

Taken together, these studies open up the possibility that 5hmC may function in the development of environmentally-sensitive neuronal adaptation and/or dysfunction. It will be important to investigate the role of each *Tet* enzyme coupled with the rapid and stable dynamics of 5hmC at different developmental time points to understand its role in synaptic plasticity, neuronal development, the maintenance of mental health, and the onset of mental illness.

IV. Sex-specific differences of 5hmC across development and following environmental exposures

Defining how the sexes differ at the molecular level, and how these basal differences contribute to sex-specific behavioral responses and outcomes, has become of much interest to the field of epigenetics research. These studies have focused efforts to identify sex-specific disruptions in DNA methylation and hydroxymethylation. For example, arsenic exposure resulted in sex-specific associations with leukocyte 5hmC levels, but not 5mC levels[51]. More specifically, male patients' 5hmC levels showed a positive correlation with increased arsenic exposure, while female patients' 5hmC levels showed a negative correlation. Both sex-dependent and sex-independent lead-associated changes in 5mC and 5hmC levels were identified across the entire genomes, of embryonic stem cell and umbilical cord blood models [52]. It was also identified that 5mC was a superior indicator of sex-dependent lead exposure, as compared to 5hmC.

Identifying sex-specific differences in 5hmC levels as a result of environmental insults has prompted investigations toward understanding basal differences of 5hmC between females and males. In early development, it was identified that 5hmC subnuclear localization patterns were maintained in female germ cells, while such patterns were lost in male germ cells during mitotic proliferation[53]. It is thought that the maintenance of 5hmC localization patterns in female germ cells aids in silencing major satellite repeats, which is exceptionally pertinent to female germ cell survival. Therefore, 5hmC shows importance in sex-dependent cellular maintenance even prior to embryonic development.

Identifying sex-dependent 5hmC patterns in germ cells prompted other investigations of differences in 5hmC levels between females and males across various developmental

windows. Indeed, sex-specific variations in 5mC and 5hmC abundances during human fetal brain development have been reported using tissues from samples spanning 23-183 days post-conception[54]. In particular, dynamic sex-specific changes of 5hmC were found across the genome, with several regions showing marked decreases in 5hmC abundance with age. While females showed an increase in 5mC levels across the X chromosome, likely due to X chromosome inactivation, males showed an increase in 5hmC abundance on the X chromosome, suggesting that 5mC and 5hmC levels differ in dynamic fashions across fetal brain development in females and males .

Others have reported distinguishing profiles of 5hmC between females and males in adulthood. Using hippocampal tissue derived from adult naïve mice, sex-specific 5hmC levels were identified between females and males[55]. Genes displaying 5hmC abundances in males were predominantly associated to neuronal developmental and differentiation pathways, while genes showing 5hmC abundances in females were primarily linked to cellular organization and synaptic plasticity. Together, these reports highlight the dynamic sex-specific nature of 5hmC, ranging from germ cells, to fetal development, to adulthood.

V. Links between 5hmC, early life adversity, and mental health

Several studies indicate that early life experiences have a profound impact on brain development and subsequent adult behavior [56-59]. Recent evidence indicates that the epigenome is a potential molecular mechanism governing the long-lasting effects of early life stress on brain and behavior. One such example involves rhesus macaques that were deprived of early life maternal interactions. Although altered behaviors were not

investigated, as adults, these monkeys have altered 5hmC in the prefrontal cortex on promoters of genes related to neurological functions and psychological disorders (*e.g.,* D_3 dopamine receptor (*DRD3*), serotonergic transporter (*5-HTT*), and GABAergic receptor (*GABRA2*)) [60]. Since these 5hmC disruptions were detected during adulthood, early life changes in 5hmC also can be stable throughout development and may represent the molecular origins of developmental brain disorders such as schizophrenia, bipolar disorder, and autism.

Schizophrenia, bipolar disorder, and major depressive disorder

Schizophrenia (SCZ) and bipolar disorder (BD) are psychiatric disorders with shared and distinct clinical and genetic features. In both disorders, stressful events increase the risk for onset and relapse mainly through the dysregulation of the hypothalamus-pituitary-adrenal (HPA) axis. Although many genes are linked to HPA dysfunction, the majority of SCZ and BD cases cannot be explained by genetics alone and the epigenome likely has a role in the molecular etiology of these disorders. Consistent with this hypothesis, rodent models exposed to prenatal stress exhibit long-lasting neurological, endocrinological, and behavioral changes that are thought to mirror the development of SCZ [61]. These models show increased DNA methylation (5mC + 5hmC) in GABAergic interneurons primarily in CpG-rich promoter regions of GABAergic genes, suggesting that these epigenetic signatures may regulate the expression of genes involved in GABAergic signaling[62,63]. Together, these studies implicate a role for DNA methylation in the molecular origins of SCZ and mood disorders[64]. In the case of SCZ, when interrogating peripheral blood from SCZ or control patients, global DNA hydroxymethylation differences were linked to SCZ,

and showed age-dependent correlations, through their role in SCZ pathogenesis, potentially though regulating gene expression, still remains elusive[65]. In humans, a role for 5hmC in the inferior parietal lobule (IPL) of SCZ and BD patients was characterized by increased levels of 5hmC and *TET1* expression, but not *TET2* or *TET3*, [66]. Remarkably, *TET1* expression levels were not altered in the cerebellum of these patients, suggesting that 5hmC may be involved in the development of psychosis through the inferior parietal lobule, but not the cerebellum, perhaps shedding light on the tissue-specific development of SCZ and BD. The global increases of 5hmC levels in these patients were associated with reduced expression of biologically relevant genes, including glutamic acid decarboxylase (GAD) 67 and APOBEC3A, an enzyme with critical roles in the active DNA demethylation pathway, suggesting that deregulation of the DNA methylation pathway and machinery may be a contributor in SCZ and BD development and pathogenesis. In contrast, patients suffering from severe major depressive disorder have significantly decreased global levels of 5hmC levels[67], suggesting that 5hmC has dynamic functions among closely related brain disorders. Together, these findings suggest a common etiology in psychosis, one that includes early life adversity and genome-wide changes in 5hmC.

Autism

The autism spectrum disorders (ASD) encompass a broad range of behaviorally-related and neurodevelopmental disorders with a high prevalence in children. Notably, only ~20% of ASD cases show a clear genetic etiology [68,69]. Prenatal factors shown to increase the risk of ASD in offspring [70] include environmental influences such as multiple births, *in vitro*

fertilization, and parental exposure to common drug treatments (*e.g.,* antiepileptic drugs (*e.g.,* valporate) or folic acid) [71]. Together, these findings effectively open the door for contributions from epigenetic modifications such as 5hmC to have an underlying role in the ASD etiology. Consistent with this hypothesis, during development 5hmC levels in the cerebellum are highly enriched in known autism genes [29]. In postmortem cerebellar tissue, ASD-associated changes in 5hmC profiles were observed genome-wide. Pathways and disease association analyses revealed that genes related to cell-cell communication and neurological/psychiatric disorders being most susceptible to 5hmC alterations[72]. Moreover, we profiled 5hmC in an established mouse model of autism (the *Cntnap2*$^{-/-}$ homozygous KO) that has behavioral and molecular alterations such as deficits in social interactions, reduced communication, repetitive behaviors, and aberrations in GABAergic signaling, consistent with abnormalities observed in humans presenting with ASD [73,74]. This study revealed that this mouse model harbored differential 5hmC on a remarkable number of established human autism genes, suggesting that 5hmC may be influencing the observed autistic-like phenotype in these mice [75]. Since these findings were observed post-symptomatically in adult mice, it is unclear if the altered 5hmC represents a cause or a consequence of having autistic-like behaviors. Thus, these findings warrant a deeper investigation of this mouse model at earlier developmental time points.

VI. 5hmC and gene by environment interactions

While a number of neurodevelopmental and neurodegenerative diseases can be attributed to genetic mutations involving a handful of genes, the cause of a large portion

of such disease remains unknown. This fact supports the hypothesis that these debilitating diseases/disorders are caused by mechanisms other than genetic mutations alone. One model that has gained traction in the field of neurodevelopmental/degenerative diseases is studies involving gene by environment (GxE) interactions[76,77]. A GxE interaction model involves acknowledging that two differing genotypes respond differentially to environmental variation. In the context of diseases and disorders, this would suggest the possibility that the sensitivity to environmental risk factors, rather than the disease itself, can be inherited, resulting in the wide variability in disease susceptibility and phenotypes. The inheritance of sensitivity, rather than the disease, makes it a difficult endeavor for researchers to pinpoint genes and pathways responsible for disease onset, progression, and treatment.

Epigenetic signatures affect transcriptional regulation and are sensitive to genotypic and environmental variations, suggesting that epigenetic mechanisms may govern GxE interactions[78]. As epigenetics bridges both genetic and environmental influences in phenotypic outcomes, studies investigating modulations in the epigenome, such as DNA methylation, have become of major focus in GxE interaction models. Humans contain several single nucleotide polymorphisms and DNA mutations between them. Researchers have found alterations in DNA methylation profiles as a consequence of GxE interactions in the cord blood of newborns[79]. While this study did not provide longitudinal analysis on how these changes in DNA methylation may have influenced the development of diseases and disorders later in life, it provides a proof in principle that genetic variability increases sensitivities to environmental stimuli.

GxE interaction models have been proposed in diseases and disorders ranging from cancer to Alzheimer's disease[80,81]. Research involving autism spectrum disorders (ASDs) has seen a spike in studies interrogating GxE interactions as a latent cause of the disorders. In particular, using a mouse model, one study investigated the sex-specific effects of maternal stress on the development of ASD-like phenotypes[82]. Using a *Cntnap2*$^{-/-}$ homozygous knockout model, in connection with a maternal immune activation paradigm, researchers identified male-specific social response deficits ranging from three days of postnatal development, to adulthood. It was found that the "three-hits" (*i.e.,* genotype, maternal immune activation, and sex) synergistically affected social recognition in this GxE interaction model. Moreover, a significant three-way interaction on the hippocampal expression of the gene *Crhr1* and the abundance in histone H3 N-terminal lysine 4 trimethylation (H3K4me3) in the promoter of *Crhr1* was identified. This study was one of the first to identify sex-specific ASD-like phenotypes as a consequence of a GxE interaction, revealing the involvement of the epigenome in regulating the transcriptional expression of a gene with well-established functions in stress response. Indeed, this opens the door for investigations of other epigenetic modifications involved in phenotypic outcomes associated with GxE interactions and the onset of developmental disorders, such as ASD. Of note, one limitation of the above study is the fact that the mouse model that was used already displayed ASD-like phenotypes. That is to say, the *Cntnap2*$^{-/-}$ homozygous knockout mice are an established model of ASD and inherently display autism-like behaviors that were exacerbated by the GxE interaction. However, considering the missing heritability in the etiology of ASDs and the low rate of genetic

mutations being associated with ASD development[83-85], it is of high interest to identify how generalizable and robust the effects of GxE interactions are in the development of ASD. For example, while *Cntnap2*$^{-/-}$ homozygous knockout mice display core ASD-like behaviors, *Cntnap2*$^{+/-}$ heterozygous mice do not display any abnormal behaviors or neurodevelopment. Therefore, subjecting *Cntnap2*$^{+/-}$ heterozygous mice to an environmental insult (GxE interaction) and finding that they have similar behavioral deficits as the homozygous knockout mice would prove the extent to which GxE interactions can contribute to the onset of developmental brain disorders, such as ASD. Moreover, identifying molecular alterations (*e.g.,* 5hmC) caused by the GxE interaction and their connections with any abnormal behaviors would build a strong foundation for a novel molecular etiology of ASD.

To date, there are no published works specifically investigating 5hmC using GxE interactions models, particularly in connection with developmental brain disorders. However, identifying that DNA methylation[79] and histone modifications[82] show links with GxE interactions, provides evidence that 5hmC levels may also be affected in these types of models, and may have a, currently, unknown role in the development of brain disorders.

VII. Putative functions of 5hmC

Many studies have shown that disruptions in 5hmC are linked with mental illness; however, the precise molecular function(s) of 5hmC remains unknown. Since 5hmC is enriched in post-mitotic neurons, it most likely has an impact on proper neurodevelopment

by regulating expression of genes responsible for neuronal propagation, development, and maintenance.

Several studies have revealed that altered 5hmC in differentially expressed genes is proximal to transcription factor (TF) binding motifs[40,47,86-89]. These studies suggest that 5hmC may regulate TF binding to DNA, resulting in either up- or down-regulation of gene expression (Figure 1). Indeed, when investigating the maintenance of cellular identity and molecular changes affiliated with reprogramming, a recent report found that absence of the Kdm3b gene product, a H3K9me2 demethylase, resulted in genomic loci that are required to be demethylated for reprogramming to occur remained in a hydroxymethylated state, unable to be fully demethylated[90]. Moreover, these loci that were perpetually trapped in a 5hmC state inhibited binding of POU5F1, a master pluripotency factor, supporting a role of 5hmC in mediated transcription factor binding. Despite the immense connections between 5hmC and its putative role in regulating TF binding, there is no consistency between which TF binding motifs are specifically targeted by 5hmC, likely due to 5hmC having a dynamic role that is tissue-, cell-, and age-specific.

Another potential role of 5hmC may be to regulate the expression of isoform production. For example, repeated cocaine administration in rodents resulted in decreased in TET1 expression was identified in the nucleus accumbens, a key reward brain structure. This decrease in TET1 expression correlated with an increase in 5hmC levels found in splice sites associated with upregulated spliced isoforms and a decrease of 5hmC in splice sites associated with downregulated spliced isoforms; hence, coupling 5hmC changes in the

brain with alternative splicing and the regulation of isoform production[48] (Figure 1). Interestingly, chronic variable stress on male mice alters microRNA content in sperm, and offspring from these sperm have a reduction in HPA axis stress responsivity [91]. By extension, environmentally sensitive epigenetic marks, such as 5hmC, may work together to reduce miRNA expression in response to early life stressors and contribute to the origins of mental illness. Together, these studies implicate 5hmC in the regulation of transcriptional abundance and diversity.

In reality, our understanding of 5hmC functions is obscure and in its infancy. These molecular roles are likely to contribute to a broad spectrum of cellular functions from lifelong neurogenesis to cell death. Environmentally sensitive molecular mechanisms, such as 5hmC, in the brain have become a significant focus of neuroscience research because of growing evidence that they are critical to the development of psychiatric disorders. Thus, in the coming years it will be of great interest to unravel these molecular functions contributing to developmental brain disorders.

VIII. Clinical utilities of 5hmC

The instability of 5hmC levels following prenatal and/or acute stress underscores the potential for 5hmC to be a novel biomarker in the diagnosis of mental health. For example, currently, a number of cancers are diagnosed based on aberrant levels of 5mC in genes such as *MGMT*, *GSTP1*, and *MLH1*[92]. Similarly, 5hmC could potentially be used as a biomarker, particularly in brain-related illnesses, especially considering that 5hmC has a high prevelance in the central nervous system. Indeed, others have already postulated

the utility of 5hmC as a biomarker in the use of clinical panels, though these primarily relate 5hmC to cancer diagnosis[93-95].

In addition to the potential of 5hmC being used as a biomarker for the diagnosis of cancer and brain-related diseases, the presence/absence of this epigenetic modification is reversible; thus, it may become relevant in therapeutic interventions [96], especially if methods to selectively modulate 5hmC *in vivo* are developed at the nucleotide level. For example, the highly popular CRISPR/Cas9 system has been developed to genetically modify the genome, by coupling a guide RNA to specific DNA targets and cleaving the DNA through use of the endonuclease enzyme Cas9[97]. Hypothetically, instead of tethering a Cas9 enzyme to cleave the DNA, one could tether a TET/DNMT enzyme and, using a guide RNA, could specifically target a region of the genome to methylate/hydroxymethylate. Therefore, a CRISPR/TET system could be developed to modulate 5hmC abundance throughout the genome at specific regions that increase the risk of developing diseases and disorders, making it a viable therapeutic option.

To this date, 5hmC alterations are still being identified and validated in the context of many diseases and disorders, though there is little consistency among similar studies. While we are still far away from using these findings in a clinical setting to diagnose such diseases, 5hmC does show the capacity to similarly be used as a biomarker for cancer, such as 5mC is utilized. Moreover, developing technologies, such as the CRISPR/Cas9 system, could be tweaked to specifically modify 5hmC profiles, which could prove to be

a therapeutic avenue in the future. Together, these highlight the potential utilities that 5hmC could have to the clinical community.

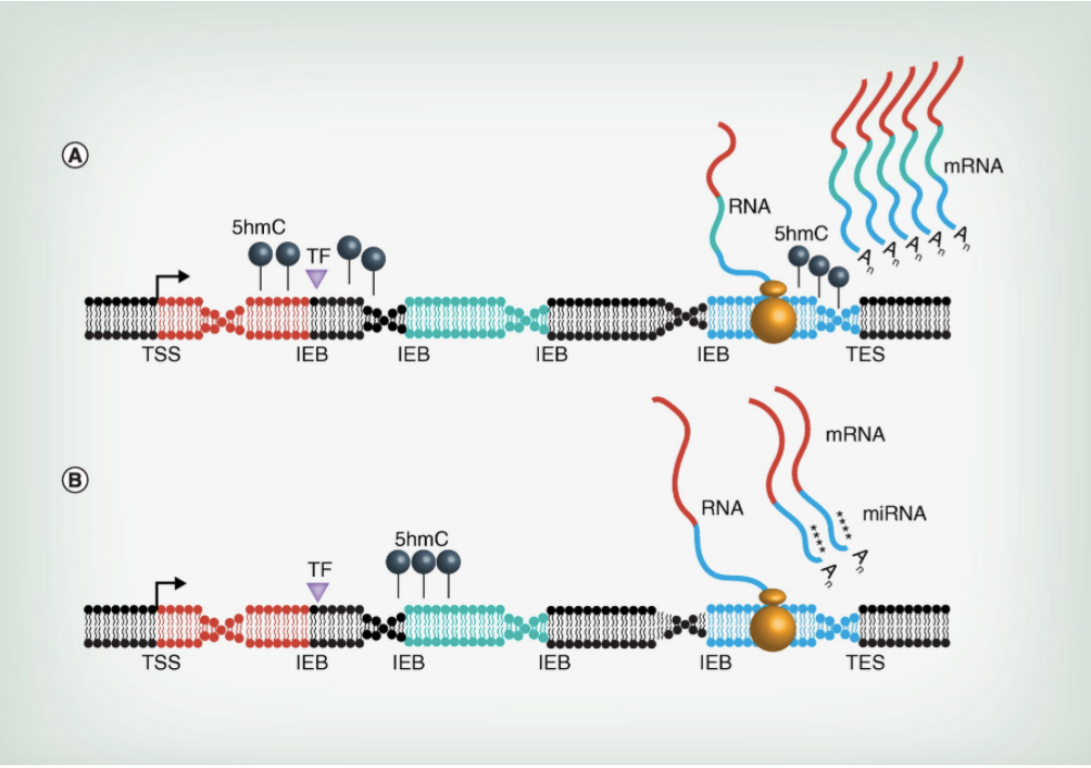
IX. Conclusions

Many of the studies discussed here suggest that mental illness is associated with disruptions in 5hmC throughout the genome. Since 5hmC is present embryonically and accumulates throughout life, primarily on genes associated with development, these studies support a role for 5hmC in developmental brain disorders. In addition, finding that 5hmC levels also can rapidly change at certain loci to reflect environmental stimuli and that these changes also can become stable, opens new perspectives in the study of epigenetic mechanisms underlying mental health and the molecular origins of mental illness. The growing number of links between 5hmC and psychiatry disorders leads many to suspect a functional role for 5hmC in mental health. While it is still unknown whether these mental illness-related disruptions in 5hmC are causative or correlative to the outcome, defining the molecular function(s) of 5hmC will likely reveal the clinical relevance and should lead to the development of new, more specific drug therapies with more personalized therapeutic options for those with mental illness. Indeed, identifying the molecular function(s) of 5hmC could corroborate it as a regulator of transcription, aiding to identify modifiable substrates that can be used as bioindicators of neurodevelopmental and neurodegenerative diseases. As many diseases and disorders show a low rate of genetic etiologies, identifying elusive molecular components that contribute to disease development and progression is of the utmost importance to identifying genes, pathways, and other molecular targets to treat and eradicate these

debilitating disorders. Together, these studies highlight the emerging role of 5hmC in mental health and disease and bring new hope to the field of psychiatry.

Figure Legends

Figure 1: Schematic of putative functions of 5hmC. When located near transcription factor (TF) binding motifs (purple triangle), 5hmC (black lollipops) may regulate gene expression levels, resulting in either up- (A) or down- (B) regulation. When located at intron/exon boundaries (IEB), 5hmC (black lollipops) may regulate alternative splicing of transcripts, resulting in a shift from full-length transcripts (A) to splice isoforms (B). Exons are colored red, green, and blue. Intergenic and intragenic regions are gray. When located near sequences complementary to microRNAs (miRs) seed regions, 5hmC (black lollipops) may have a latent function in repelling (A) or attracting (B) miRs as the DNA is being transcribed into RNA. Group of four asterisks represent miRs. Together, these studies implicate 5hmC in the regulation of transcriptional and translational abundance and diversity.



Box 1	DNA Methylation
5mC	5-methylcytosine (5mC) is the modification of cytosine through the addition of a methyl group to the fifth carbon of cytosine and is nearly exclusive at cytosine-phosphate-guanine (CpG) dinucleotide sites. This modification is catalyzed by DNA Methyltransferases (DNTMs) which utilize S-adenosyl-L-methionine as the methyl donor[98]. Several DNMTs have been found and have distinct functions. DNMT3a and DNMT3b have been shown to have an affinity towards previously unmethylated CpG sites, while DNMT1 is regarded as preserving methylation as it has been found to show preference towards hemimethylated CpG sites[99]. CpG rich-regions, known as CpG islands, and promoter regions of genes have been shown to show a significant reduction in 5mC levels, while the X chromosome has been shown to have an overabundance of 5mC[100,101]. 5mC has been implicated in processes such as gene silencing, X chromosome inactivation, genomic imprinting, and chromatin structure[102]. Bisulfite genomic sequencing is regarded as the gold-standard in the detection of 5mC[103]. Through this method, unmethylated cytosine is converted to uracil while methylated cytosine sites remain protected. Following this, PCR is used to amplify regions of interest with all uracils and thymines being amplified as thymine and only methylated cytosines being amplified as cytosines.
5hmC	5-hydroxymethylcytosine (5hmC) is the modified form of cytosine through the addition of a hydroxy group to the methyl group of 5mC. This modification is catalyzed by ten-eleven translocation (TET) enzymes[31]. An approximate ten-fold abundance of 5hmC is seen in brain tissues, other tissues of the central nervous system, and embryonic stem cells[104]. 5hmC levels show a reduction in CpG islands and intronic sequences, but show an enrichment in regions such as 5'-UTR, promoter regions, distal regulatory regions, and exonic regions[104]. Interestingly, alternatively spliced transcripts have been found to display a reduction of 5hmC on excised exons as compared to constitutive exons, suggesting that 5hmC may be playing a role in the production of isoform transcripts[105]. An increase in transcription levels have been correlated with 5hmC enrichment[104]. Neuronal- and synapse-related genes have been shown to have a significant increase in 5hmC levels[105]. Markedly, humans show a disproportionate reduction on 5hmC on chromosome 18 and the X chromosome[106]. The function of 5hmC has yet to be discovered, yet several putative functions have been put forth. Differentially hydroxymethylated regions (DhMRs) in mice show transcription factor

	(TF) binding motifs, suggesting that 5hmC may promote or disallow the binding of TFs to regulate gene expression levels[47]. Gender-specific profiles have been found[107], suggesting that 5hmC may function in regulating genes specific for organ development and mental health distinctly between sexes. Differential expression of isoforms have been found to be more variable than whole gene expression when correlated with 5hmC distributions, indicating that 5hmC may influence transcript diversity through interactions with the spliceosome[48]. As traditional bisulfite sequencing cannot distinguish between 5hmC and 5mC, Tet1-assisted bisulfite sequencing has become the established method for 5hmC detection[108]. Briefly, beta-glucosyltransferase is utilized to introduce glucose onto 5hmC, protecting 5hmC during bisulfite sequencing while converting all other forms of cytosine into uracil derivatives. Subsequent sequencing yields 5hmC as cytosine.

References

1 Bird, A. DNA methylation patterns and epigenetic memory. *Genes & development* **16**, 6-21, doi:10.1101/gad.947102 (2002).

2 Sharma, R. P., Gavin, D. P. & Grayson, D. R. CpG methylation in neurons: message, memory, or mask? *Neuropsychopharmacology* **35**, 2009-2020, doi:10.1038/npp.2010.85 (2010).

3 Han, J. A., An, J. & Ko, M. Functions of TET Proteins in Hematopoietic Transformation. *Molecules and cells* **38**, 925-935, doi:10.14348/molcells.2015.0294 (2015).

4 Suzuki, M. M. & Bird, A. DNA methylation landscapes: provocative insights from epigenomics. *Nat Rev Genet* **9**, 465-476, doi:10.1038/nrg2341 (2008).

5 Robertson, K. D. DNA methylation and human disease. *Nat Rev Genet* **6**, 597-610, doi:10.1038/nrg1655 (2005).

6 Abdolmaleky, H. M. *et al.* Hypomethylation of MB-COMT promoter is a major risk factor for schizophrenia and bipolar disorder. *Hum Mol Genet* **15**, 3132-3145, doi:10.1093/hmg/ddl253 (2006).

7 Poulter, M. O. *et al.* GABAA receptor promoter hypermethylation in suicide brain: implications for the involvement of epigenetic processes. *Biological psychiatry* **64**, 645-652, doi:10.1016/j.biopsych.2008.05.028 (2008).

8 Kuratomi, G. *et al.* Aberrant DNA methylation associated with bipolar disorder identified from discordant monozygotic twins. *Mol Psychiatry* **13**, 429-441, doi:10.1038/sj.mp.4002001 (2008).

9 Kappeler, L. & Meaney, M. J. Epigenetics and parental effects. *BioEssays : news and reviews in molecular, cellular and developmental biology* **32**, 818-827, doi:10.1002/bies.201000015 (2010).

10 Weaver, I. C. *et al.* Epigenetic programming by maternal behavior. *Nat Neurosci* **7**, 847-854, doi:10.1038/nn1276 (2004).

11 Wu, H. & Zhang, Y. Mechanisms and functions of Tet protein-mediated 5-methylcytosine oxidation. *Genes & development* **25**, 2436-2452, doi:10.1101/gad.179184.111 (2011).

12 Szulwach, K. E. *et al.* 5-hmC-mediated epigenetic dynamics during postnatal neurodevelopment and aging. *Nature neuroscience* **14**, 1607-1616, doi:10.1038/nn.2959 (2011).

13 Kriaucionis, S. & Heintz, N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science (New York, N.Y.)* **324**, 929-930, doi:10.1126/science.1169786 (2009).

14 Sun, W., Zang, L., Shu, Q. & Li, X. From development to diseases: the role of 5hmC in brain. *Genomics* **104**, 347-351, doi:10.1016/j.ygeno.2014.08.021 (2014).

15 Wang, J., Tang, J., Lai, M. & Zhang, H. 5-Hydroxymethylcytosine and disease. *Mutation research. Reviews in mutation research* **762**, 167-175, doi:10.1016/j.mrrev.2014.09.003 (2014).

16 He, Y. F. *et al.* Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **333**, 1303-1307, doi:10.1126/science.1210944 (2011).

17 Ito, S. *et al.* Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal

and inner cell mass specification. *Nature* **466**, 1129-1133, doi:10.1038/nature09303 (2010).

18 Kraus, T. F. *et al.* Profiling of methylation and demethylation pathways during brain development and ageing. *Journal of neural transmission (Vienna, Austria : 1996)*, doi:10.1007/s00702-015-1469-2 (2015).

19 Hahn, M. A. *et al.* Dynamics of 5-hydroxymethylcytosine and chromatin marks in Mammalian neurogenesis. *Cell reports* **3**, 291-300, doi:10.1016/j.celrep.2013.01.011 (2013).

20 Tan, L. *et al.* Genome-wide comparison of DNA hydroxymethylation in mouse embryonic stem cells and neural progenitor cells by a new comparative hMeDIP-seq method. *Nucleic acids research* **41**, e84, doi:10.1093/nar/gkt091 (2013).

21 Kang, J. *et al.* Simultaneous deletion of the methylcytosine oxidases Tet1 and Tet3 increases transcriptome variability in early embryogenesis. *Proceedings of the National Academy of Sciences of the United States of America* **112**, E4236-4245, doi:10.1073/pnas.1510510112 (2015).

22 Montalbán-Loro, R. *et al.* TET3 prevents terminal differentiation of adult NSCs by a non-catalytic action at Snrpn. *Nature communications* **10**, 1726, doi:10.1038/s41467-019-09665-1 (2019).

23 Santiago, M. *et al.* Tet3 regulates cellular identity and DNA methylation in neural progenitor cells. *Cellular and molecular life sciences : CMLS* **77**, 2871-2883, doi:10.1007/s00018-019-03335-7 (2020).

24 Hirabayashi, Y. & Gotoh, Y. Epigenetic control of neural precursor cell fate during development. *Nature reviews. Neuroscience* **11**, 377-388, doi:10.1038/nrn2810 (2010).

25 Ma, D. K. *et al.* Epigenetic choreographers of neurogenesis in the adult mammalian brain. *Nature neuroscience* **13**, 1338-1344, doi:10.1038/nn.2672 (2010).

26 Chen, Y., Damayanti, N. P., Irudayaraj, J., Dunn, K. & Zhou, F. C. Diversity of two forms of DNA methylation in the brain. *Frontiers in genetics* **5**, 46, doi:10.3389/fgene.2014.00046 (2014).

27 Chen, H., Dzitoyeva, S. & Manev, H. Effect of aging on 5-hydroxymethylcytosine in the mouse hippocampus. *Restorative neurology and neuroscience* **30**, 237-245, doi:10.3233/rnn-2012-110223 (2012).

28 Zampieri, M. *et al.* Reconfiguration of DNA methylation in aging. *Mechanisms of ageing and development* **151**, 60-70, doi:10.1016/j.mad.2015.02.002 (2015).

29 Wang, T. *et al.* Genome-wide DNA hydroxymethylation changes are associated with neurodevelopmental genes in the developing human cerebellum. *Human molecular genetics* **21**, 5500-5510, doi:10.1093/hmg/dds394 (2012).

30 Al-Mahdawi, S., Virmouni, S. A. & Pook, M. A. The emerging role of 5-hydroxymethylcytosine in neurodegenerative diseases. *Frontiers in neuroscience* **8**, 397, doi:10.3389/fnins.2014.00397 (2014).

31 Song, C. X. *et al.* Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. *Nature biotechnology* **29**, 68-72, doi:10.1038/nbt.1732 (2011).

32 Chouliaras, L. *et al.* Consistent decrease in global DNA methylation and hydroxymethylation in the hippocampus of Alzheimer's disease patients.

Neurobiology of aging **34**, 2091-2099, doi:10.1016/j.neurobiolaging.2013.02.021 (2013).

33 Condliffe, D. *et al.* Cross-region reduction in 5-hydroxymethylcytosine in Alzheimer's disease brain. *Neurobiology of aging* **35**, 1850-1854, doi:10.1016/j.neurobiolaging.2014.02.002 (2014).

34 Coppieters, N. *et al.* Global changes in DNA methylation and hydroxymethylation in Alzheimer's disease human brain. *Neurobiology of aging* **35**, 1334-1344, doi:10.1016/j.neurobiolaging.2013.11.031 (2014).

35 Lardenoije, R. *et al.* Alzheimer's disease-associated (hydroxy)methylomic changes in the brain and blood. *Clinical epigenetics* **11**, 164, doi:10.1186/s13148-019-0755-5 (2019).

36 Bradley-Whitman, M. A. & Lovell, M. A. Epigenetic changes in the progression of Alzheimer's disease. *Mechanisms of ageing and development* **134**, 486-495, doi:10.1016/j.mad.2013.08.005 (2013).

37 Lashley, T. *et al.* Alterations in global DNA methylation and hydroxymethylation are not detected in Alzheimer's disease. *Neuropathology and applied neurobiology* **41**, 497-506, doi:10.1111/nan.12183 (2015).

38 Villar-Menéndez, I. *et al.* Increased 5-methylcytosine and decreased 5-hydroxymethylcytosine levels are associated with reduced striatal A2AR levels in Huntington's disease. *Neuromolecular medicine* **15**, 295-309, doi:10.1007/s12017-013-8219-0 (2013).

39 Wang, F. *et al.* Genome-wide loss of 5-hmC is a novel epigenetic feature of Huntington's disease. *Human molecular genetics* **22**, 3641-3653, doi:10.1093/hmg/ddt214 (2013).

40 Yao, B. *et al.* Genome-wide alteration of 5-hydroxymethylcytosine in a mouse model of fragile X-associated tremor/ataxia syndrome. *Hum Mol Genet* **23**, 1095-1107, doi:10.1093/hmg/ddt504 (2014).

41 Jiang, D. *et al.* Alteration in 5-hydroxymethylcytosine-mediated epigenetic regulation leads to Purkinje cell vulnerability in ATM deficiency. *Brain : a journal of neurology* **138**, 3520-3536, doi:10.1093/brain/awv284 (2015).

42 Cirelli, C., Faraguna, U. & Tononi, G. Changes in brain gene expression after long-term sleep deprivation. *Journal of neurochemistry* **98**, 1632-1645, doi:10.1111/j.1471-4159.2006.04058.x (2006).

43 Irier, H. *et al.* Environmental enrichment modulates 5-hydroxymethylcytosine dynamics in hippocampus. *Genomics* **104**, 376-382, doi:10.1016/j.ygeno.2014.08.019 (2014).

44 Chouliaras, L. *et al.* Age-related increase in levels of 5-hydroxymethylcytosine in mouse hippocampus is prevented by caloric restriction. *Current Alzheimer research* **9**, 536-544 (2012).

45 Alberca, C. D. *et al.* Perinatal protein malnutrition results in genome-wide disruptions of 5-hydroxymethylcytosine at regions that can be restored to control levels by an enriched environment. *Epigenetics*, 1-17, doi:10.1080/15592294.2020.1841871 (2020).

46 Li, S. *et al.* Hippocampal increase of 5-hmC in the glucocorticoid receptor gene following acute stress. *Behav Brain Res* **286**, 236-240, doi:10.1016/j.bbr.2015.03.002 (2015).

47 Li, S. *et al.* Genome-wide alterations in hippocampal 5-hydroxymethylcytosine links plasticity genes to acute stress. *Neurobiology of disease* **86**, 99-108, doi:10.1016/j.nbd.2015.11.010 (2016).

48 Feng, J. *et al.* Role of Tet1 and 5-hydroxymethylcytosine in cocaine action. *Nature neuroscience* **18**, 536-544, doi:10.1038/nn.3976 (2015).

49 Li, X. *et al.* Neocortical Tet3-mediated accumulation of 5-hydroxymethylcytosine promotes rapid behavioral adaptation. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 7120-7125, doi:10.1073/pnas.1318906111 (2014).

50 Rudenko, A. *et al.* Tet1 is critical for neuronal activity-regulated gene expression and memory extinction. *Neuron* **79**, 1109-1122, doi:10.1016/j.neuron.2013.08.003 (2013).

51 Niedzwiecki, M. M. *et al.* Sex-specific associations of arsenic exposure with global DNA methylation and hydroxymethylation in leukocytes: results from two studies in Bangladesh. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* **24**, 1748-1757, doi:10.1158/1055-9965.epi-15-0432 (2015).

52 Sen, A. *et al.* Lead exposure induces changes in 5-hydroxymethylcytosine clusters in CpG islands in human embryonic stem cells and umbilical cord blood. *Epigenetics* **10**, 607-621, doi:10.1080/15592294.2015.1050172 (2015).

53 Yamaguchi, S. *et al.* Dynamics of 5-methylcytosine and 5-hydroxymethylcytosine during germ cell reprogramming. *Cell research* **23**, 329-339, doi:10.1038/cr.2013.22 (2013).

54 Spiers, H., Hannon, E., Schalkwyk, L. C., Bray, N. J. & Mill, J. 5-hydroxymethylcytosine is highly dynamic across human fetal brain development. *BMC genomics* **18**, 738, doi:10.1186/s12864-017-4091-x (2017).

55 Papale, L. A. *et al.* Sex-specific hippocampal 5-hydroxymethylcytosine is disrupted in response to acute stress. *Neurobiology of disease* **96**, 54-66, doi:10.1016/j.nbd.2016.08.014 (2016).

56 Roth, T. L. & Sweatt, J. D. Epigenetic marking of the BDNF gene by early-life adverse experiences. *Hormones and behavior* **59**, 315-320, doi:10.1016/j.yhbeh.2010.05.005 (2011).

57 Labonte, B. *et al.* Genome-wide epigenetic regulation by early-life trauma. *Archives of general psychiatry* **69**, 722-731, doi:10.1001/archgenpsychiatry.2011.2287 (2012).

58 McGowan, P. O. *et al.* Epigenetic regulation of the glucocorticoid receptor in human brain associates with childhood abuse. *Nature neuroscience* **12**, 342-348, doi:10.1038/nn.2270 (2009).

59 Oberlander, T. F. *et al.* Prenatal exposure to maternal depression, neonatal methylation of human glucocorticoid receptor gene (NR3C1) and infant cortisol stress responses. *Epigenetics* **3**, 97-106 (2008).

60 Massart, R. *et al.* Hydroxymethylation and DNA methylation profiles in the prefrontal cortex of the non-human primate rhesus macaque and the impact of maternal deprivation on hydroxymethylation. *Neuroscience* **268**, 139-148, doi:10.1016/j.neuroscience.2014.03.021 (2014).

61 Koenig, J. I. *et al.* Prenatal exposure to a repeated variable stress paradigm elicits behavioral and neuroendocrinological changes in the adult offspring: potential relevance to schizophrenia. *Behav Brain Res* **156**, 251-261, doi:10.1016/j.bbr.2004.05.030 (2005).
62 Guidotti, A., Dong, E., Tueting, P. & Grayson, D. R. Modeling the molecular epigenetic profile of psychosis in prenatally stressed mice. *Progress in molecular biology and translational science* **128**, 89-101, doi:10.1016/b978-0-12-800977-2.00004-8 (2014).
63 Matrisciano, F. *et al.* Epigenetic modifications of GABAergic interneurons are associated with the schizophrenia-like phenotype induced by prenatal stress in mice. *Neuropharmacology* **68**, 184-194, doi:10.1016/j.neuropharm.2012.04.013 (2013).
64 Brown, A. S. The environment and susceptibility to schizophrenia. *Progress in neurobiology* **93**, 23-58, doi:10.1016/j.pneurobio.2010.09.003 (2011).
65 Jiang, T. *et al.* Variation in global DNA hydroxymethylation with age associated with schizophrenia. *Psychiatry research* **257**, 497-500, doi:10.1016/j.psychres.2017.08.022 (2017).
66 Dong, E., Gavin, D. P., Chen, Y. & Davis, J. Upregulation of TET1 and downregulation of APOBEC3A and APOBEC3C in the parietal cortex of psychotic patients. *Translational psychiatry* **2**, e159, doi:10.1038/tp.2012.86 (2012).
67 Tseng, P. T. *et al.* Age-associated decrease in global DNA methylation in patients with major depression. *Neuropsychiatric disease and treatment* **10**, 2105-2114, doi:10.2147/ndt.s71997 (2014).
68 Gaugler, T. *et al.* Most genetic risk for autism resides with common variation. *Nat Genet* **46**, 881-885, doi:10.1038/ng.3039 (2014).
69 Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat Genet* **47**, 1236-1241, doi:10.1038/ng.3406 (2015).
70 Gardener, H., Spiegelman, D. & Buka, S. L. Prenatal risk factors for autism: comprehensive meta-analysis. *Br J Psychiatry* **195**, 7-14, doi:10.1192/bjp.bp.108.051672 (2009).
71 Rogers, E. J. Has enhanced folate status during pregnancy altered natural selection and possibly Autism prevalence? A closer look at a possible link. *Med Hypotheses* **71**, 406-410, doi:10.1016/j.mehy.2008.04.013 (2008).
72 Cheng, Y. *et al.* 5-Hydroxymethylcytosine alterations in the human postmortem brains of autism spectrum disorder. *Human molecular genetics* **27**, 2955-2964, doi:10.1093/hmg/ddy193 (2018).
73 Penagarikano, O. *et al.* Exogenous and evoked oxytocin restores social behavior in the Cntnap2 mouse model of autism. *Science translational medicine* **7**, 271ra278, doi:10.1126/scitranslmed.3010257 (2015).
74 Penagarikano, O. *et al.* Absence of CNTNAP2 leads to epilepsy, neuronal migration abnormalities, and core autism-related deficits. *Cell* **147**, 235-246, doi:10.1016/j.cell.2011.08.040 (2011).
75 Papale, L. A. *et al.* Genome-wide disruption of 5-hydroxymethylcytosine in a mouse model of autism. *Hum Mol Genet* **24**, 7121-7131, doi:10.1093/hmg/ddv411 (2015).
76 Hunter, D. J. Gene-environment interactions in human diseases. *Nature reviews.*

Genetics **6**, 287-298, doi:10.1038/nrg1578 (2005).
77 Esposito, G., Azhari, A. & Borelli, J. L. Gene × Environment Interaction in Developmental Disorders: Where Do We Stand and What's Next? *Frontiers in psychology* **9**, 2036, doi:10.3389/fpsyg.2018.02036 (2018).
78 Kubota, T., Miyake, K. & Hirasawa, T. Epigenetic understanding of gene-environment interactions in psychiatric disorders: a new concept of clinical genetics. *Clinical epigenetics* **4**, 1, doi:10.1186/1868-7083-4-1 (2012).
79 Czamara, D. *et al.* Integrated analysis of environmental and genetic influences on cord blood DNA methylation in new-borns. *Nature communications* **10**, 2548, doi:10.1038/s41467-019-10461-0 (2019).
80 Han, J., Hankinson, S. E., Colditz, G. A. & Hunter, D. J. Genetic variation in XRCC1, sun exposure, and risk of skin cancer. *British journal of cancer* **91**, 1604-1609, doi:10.1038/sj.bjc.6602174 (2004).
81 Eid, A., Mhatre, I. & Richardson, J. R. Gene-environment interactions in Alzheimer's disease: A potential path to precision medicine. *Pharmacology & therapeutics* **199**, 173-187, doi:10.1016/j.pharmthera.2019.03.005 (2019).
82 Schaafsma, S. M. *et al.* Sex-specific gene-environment interactions underlying ASD-like behaviors. *Proceedings of the National Academy of Sciences of the United States of America* **114**, 1383-1388, doi:10.1073/pnas.1619312114 (2017).
83 Sebat, J. *et al.* Strong association of de novo copy number mutations with autism. *Science (New York, N.Y.)* **316**, 445-449, doi:10.1126/science.1138659 (2007).
84 Glessner, J. T. *et al.* Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. *Nature* **459**, 569-573, doi:10.1038/nature07953 (2009).
85 Weiss, L. A., Arking, D. E., Daly, M. J. & Chakravarti, A. A genome-wide linkage and association scan reveals novel loci for autism. *Nature* **461**, 802-808, doi:10.1038/nature08490 (2009).
86 Tekpli, X. *et al.* Changes of 5-hydroxymethylcytosine distribution during myeloid and lymphoid differentiation of CD34+ cells. *Epigenetics & chromatin* **9**, 21, doi:10.1186/s13072-016-0070-8 (2016).
87 Wu, H. *et al.* Genome-wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells. *Genes & development* **25**, 679-684, doi:10.1101/gad.2036011 (2011).
88 Uribe-Lewis, S. *et al.* 5-hydroxymethylcytosine and gene activity in mouse intestinal differentiation. *Scientific reports* **10**, 546, doi:10.1038/s41598-019-57214-z (2020).
89 Green, B. B. *et al.* Hydroxymethylation is uniquely distributed within term placenta, and is associated with gene expression. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **30**, 2874-2884, doi:10.1096/fj.201600310R (2016).
90 Tran, K. A., Dillingham, C. M. & Sridharan, R. Coordinated removal of repressive epigenetic modifications during induced reversal of cell identity. *The EMBO journal* **38**, e101681, doi:10.15252/embj.2019101681 (2019).
91 Rodgers, A. B., Morgan, C. P., Bronson, S. L., Revello, S. & Bale, T. L. Paternal stress exposure alters sperm microRNA content and reprograms offspring HPA stress axis regulation. *J Neurosci* **33**, 9003-9012, doi:10.1523/jneurosci.0914-13.2013 (2013).

92 Locke, W. J. *et al.* DNA Methylation Cancer Biomarkers: Translation to the Clinic. *Frontiers in genetics* **10**, 1150, doi:10.3389/fgene.2019.01150 (2019).
93 Xu, T. & Gao, H. Hydroxymethylation and tumors: can 5-hydroxymethylation be used as a marker for tumor diagnosis and treatment? *Human genomics* **14**, 15, doi:10.1186/s40246-020-00265-5 (2020).
94 Zhang, Z. *et al.* Values of 5mC, 5hmC, and TET2 for identifying the presence and progression of breast precancerous lesion. *Journal of clinical laboratory analysis* **34**, e23162, doi:10.1002/jcla.23162 (2020).
95 Vasanthakumar, A. & Godley, L. A. 5-hydroxymethylcytosine in cancer: significance in diagnosis and therapy. *Cancer genetics* **208**, 167-177, doi:10.1016/j.cancergen.2015.02.009 (2015).
96 Szyf, M. Epigenetics, a key for unlocking complex CNS disorders? Therapeutic implications. *European neuropsychopharmacology : the journal of the European College of Neuropsychopharmacology* **25**, 682-702, doi:10.1016/j.euroneuro.2014.01.009 (2015).
97 Ma, Y., Zhang, L. & Huang, X. Genome modification by CRISPR/Cas9. *The FEBS journal* **281**, 5186-5193, doi:10.1111/febs.13110 (2014).
98 Cheng, X. Structure and function of DNA methyltransferases. *Annual review of biophysics and biomolecular structure* **24**, 293-318, doi:10.1146/annurev.bb.24.060195.001453 (1995).
99 Okano, M., Bell, D. W., Haber, D. A. & Li, E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99**, 247-257 (1999).
100 Sharp, A. J. *et al.* DNA methylation profiles of human active and inactive X chromosomes. *Genome research* **21**, 1592-1600, doi:10.1101/gr.112680.110 (2011).
101 Ioshikhes, I. P. & Zhang, M. Q. Large-scale human promoter mapping using CpG islands. *Nat Genet* **26**, 61-63, doi:10.1038/79189 (2000).
102 Irier, H. A. & Jin, P. Dynamics of DNA methylation in aging and Alzheimer's disease. *DNA Cell Biol* **31 Suppl 1**, S42-48, doi:10.1089/dna.2011.1565 (2012).
103 Frommer, M. *et al.* A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc Natl Acad Sci U S A* **89**, 1827-1831 (1992).
104 Branco, M. R., Ficz, G. & Reik, W. Uncovering the role of 5-hydroxymethylcytosine in the epigenome. *Nat Rev Genet* **13**, 7-13, doi:10.1038/nrg3080 (2012).
105 Khare, T. *et al.* 5-hmC in the brain is abundant in synaptic genes and shows differences at the exon-intron boundary. *Nature structural & molecular biology* **19**, 1037-1043, doi:10.1038/nsmb.2372 (2012).
106 Chopra, P. *et al.* Array-based assay detects genome-wide 5-mC and 5-hmC in the brains of humans, non-human primates, and mice. *BMC genomics* **15**, 131, doi:10.1186/1471-2164-15-131 (2014).
107 Gross, J. A. *et al.* Characterizing 5-hydroxymethylcytosine in human prefrontal cortex at single base resolution. *BMC genomics* **16**, 672, doi:10.1186/s12864-015-1875-8 (2015).
108 Yu, M. *et al.* Tet-assisted bisulfite sequencing of 5-hydroxymethylcytosine. *Nature protocols* **7**, 2159-2170, doi:10.1038/nprot.2012.137 (2012).

CHAPTER 2

Sex-specific 5-hydroxymethylcytosine perturbations in a mouse model of autism

Abstract

Recently we reported that male *Cntnap2* knockout mice exhibit a genome-wide disruption of 5hmC in a significant number of orthologs and pathways common to human neurodevelopmental disorders. Despite the fact that female mice lacking *Cntnap2* expression share the same behavioral and neuropathological abnormalities of the male knockout mice, we sought to examine the genome-wide 5hmC profiles in females to determine if there are sex-specific disruptions of 5hmC in this mouse model of autism. Comparison of genome-wide profiles of striatal 5hmC in *Cntnap2*$^{-/-}$ and wildtype female mice revealed that a significant number of the same genes harbor 5hmC disruptions in both female and male mutant mice. Further analyses found similar associations with neuronal-developmental and an enrichment of transcription factors binding sites within the differentially hydroxymethylated regions. Sex-specific alterations in 5hmC also were observed, revealing a broader and more diverse set of neurodevelopmentally important genes disrupted in the male mutant. Together, these data implicate a role for sex-specific 5hmC-mediated modulation in the pathogenesis of autism, shedding light on pathways and processes that are more susceptible in males, perhaps explaining the greater prevalence of autism in males.

Introduction

Autism spectrum disorders (ASD) encompass a heterogeneous continuum of neurodevelopmental disorders affecting >1% of the world's population, and are characterized by core deficits in sociability, communication, and repetitive/stereotypical behaviors and interests[1]. A number of other symptoms are frequently associated with ASD, including hyperactivity, epilepsy, aberrations in the sleep cycle, and abnormal gastrointestinal function[2]. Despite the prevalence of ASD, genetic studies have predicted rare variants in hundreds of genes that, cumulatively, account for <25% of cases of ASD[3-6]. Moreover, the rate of heritability of ASD shows a wide scale of variability, finding the rate of concordance ranging from 60-90% in monozygotic twins, and 0-20% in dizygotic twins, suggesting that genetic contributions, alone, are insufficient to explain entire etiology of ASD[7-9]. As such, ASD are now considered multifactorial hereditary disorders resulting from polygenic contributions and environmental influences. Surprisingly, ASD show a sex-specific preference, as males are three-times as likely to develop autism compared to their female counterparts, suggesting that inherent differences between females and males alter their susceptibility to ASD, though these risk factors remain unknown[10].

Several lines of research implicate both rare and common variants of contactin-associated protein-like 2 (*CNTNAP2*) with ASD[11-13]. *Cntnap2* encodes a neuronal member of the neurexin superfamily transmembrane protein that is involved in neuron-glia interactions and the clustering of potassium channels within myelinated axons[14]. Both male and female mice fostering a homozygous knockout of *Cntnap2* (*Cntnap2*$^{-/-}$) exhibit

striking parallels to the major neuropathological features of ASD, including cortical dysplasia and focal epilepsy, coupled with reduced social interactions, communication, and repetitive behaviors that are mirrored by abnormal neuronal migration of cortical projections and a decrease in the number of GABAergic interneurons in the striatum[15]. As described in chapter 1, our lab profiled striatal 5hmC levels in adult male *Cntnap2*$^{-/-}$ mutant mice and found genome-wide disruptions of 5hmC in genes with strong affiliations with ASD and neurological processes and pathways[16]. However, identifying sex-specific epigenetic modifications associated with ASD is pertinent towards understanding the male-specific bias of ASD. Here, we follow up on this line of research by profiling striatal 5hmC abundance in adult female *Cntnap2*$^{-/-}$ mutant mice and contrast these difference with male mice, providing sex-specific context in this mouse model of ASD. These genome-wide maps reveal known and novel genes contributing to the autistic-like phenotypes, demonstrate the sex-specific neuromolecular response to the absence of *Cntnap2*, and further establish a role for 5hmC in ASD.

Results

Disruption of 5hmC in the striatum of female Cntnap2$^{-/-}$ *mutant mice*

To determine the genome-wide 5hmC distribution in female *Cntnap2*$^{-/-}$ and wildtype (WT) littermates, we utilized an established chemical labeling and affinity purification method in conjunction with high-throughput sequencing technology[17,18]. Three female *Cntnap2*$^{-/-}$ (post-natal day 90) and three age-matched female WT littermates were sacrificed as independent biological replicates. DNA fragments containing 5hmC were enriched from striatum total DNA and sequenced, resulting in a range of ~30-46 million uniquely mapped

high-throughput sequence reads from each biological replicate. These data found no difference among chromosomes, except for a marked depletion on the X chromosome (**Fig. 1A**), which is consistent with previous observations. Together, this suggests that there are no gross differences in 5hmC abundance between mutant or WT mice[18].

To determine distinct 5hmC patterns between *Cntnap2*$^{-/-}$ and WT mice, we identified differentially hydroxymethylated regions (DhMRs) with respect to the *Cntnap2* mutant genome. In total, 1,699 mutant-specific increases in 5hmC levels (hyper-DhMRs) and 1,346 mutant-specific decreases in hydroxymethylation (hypo-DhMRs) were found across the entire genome (**Fig. 1A**; **Dataset 1**). As specific regions of the genome are differentially methylated based on the biological functions of the genes retained within the region, we next annotated the DhMRs to standard intragenic genomic structures, or to intergenic regions if >3kb away from any gene (**Fig. 1B**). Approximately 65% of DhMRs were annotated to gene structures, with the largest portion of these mapping to intronic regions, supporting the hypothesis the intronic regions are more dynamic to 5hmC alterations[19,20]. These data, taken together, suggest that changes in 5hmC levels, caused by the homozygous loss of CNTNAP2, are not randomly dispersed throughout the genome.

Annotation of DhMRs to genes identified 991 and 837 genes contained hyper-DhMRs and hypo-DhRMs, respectively. Notably, 123 genes contained both hyper- and hypo-DhMRs, though these DhMRs were located at distinct genomic loci (average >50kb away from each other), suggesting that when annotated to the same gene, hyper- and hypo-

DhMRs have unique function(s). An initial surveillance of DhMR-associated genes found several loci related to *Cntnap2* biology and ASD, including other contactin family members (*e.g., Cntnap1*), neurexin members (*e.g., Nrxn1-3*), along with several glutamate and GABA receptors. To determine the biological significance of the DhMR-associated genes, ontological and pathways analyses were separately performed on genes containing hyper- and hypo-DhMRs. Genes associated with hyper-DhMRs showed ties to neuronal development, synapse assembly, and axonogenesis, all of which are linked to ASD and the observed phenotypes in the *Cntnap2*$^{-/-}$ mutant mouse[21-23] (**Fig. 1C** [left panel]; **Dataset 2**). Similarly, genes containing hypo-DhMRs also were found to associate with neuronal- and synaptic-related processes, including terms such as axonogenesis, cell morphogenesis involved in neuron differentiation, and synapse organization (**Fig. 1C** [right panel]). Pathways analysis using the hyper-DhMR-associated genes identified a significant portion of genes related to steroid biosynthesis, oxytocin signaling, long-term potentiation, and cAMP signaling, all with published connections to ASD[24-27] (**Fig. 1D**). In contrast, hypo-DhMR-associated genes did not find any significantly enriched pathways. Together, these data suggest that hyper-DhMR associated genes may have larger contributions to the observed phenotype in the female mutant mouse, while hypo-DhMR associated genes may have novel relations to ASD.

Identifying links to neuronal- and synaptic-related pathways and processes among the DhMR-associated genes by ontological analyses prompted us to investigate the overlap of these genes with a validated list of genes linked to autism[28]. This overlay identified a significant portion of known developmental brain disorder genes (N = 60/233) foster

DhMRs (**Fig. 1E**). Taken together, these data suggest that 5hmC has an autism-related role in the absence of CNTNAP2.

As syndromic forms of autism have involved disruptions in transcription factor function[29,30], we tested the hypothesis that the DhMRs have an enrichment of transcription factor binding motifs by subjecting the DhMR genomic sequences to motif enrichment analysis. Hyper- and hypo-DhMRs identified distinct lists of significant enrichments of transcription factor binding motifs (**Fig. 1F Dataset 3**). Notably, top candidates included transcription factors previously found to associate with ASD phenotypes, such as HIF-1b and CLOCK (see Discussion). These findings suggest that in the absence of *Cntnap2,* 5hmC may regulate the expression of neuronal-related genes by altering the binding affinity of function of transcription factors.

Sex-specific Cntnap2$^{-/-}$-specific alterations in striatal 5hmC

Previously, we profiled 5hmC abundance in male *Cntnap2$^{-/-}$* mutant mice, compared to age-matched WT male mice[16]. To identify sex-specific changes in 5hmC associated with the loss of CNTNAP2, we compared female-*Cntnap2$^{-/-}$*-specific DhMR-associated genes and male-*Cntnap2$^{-/-}$*-specific DhMR-associated genes and found a significant overlap (**Fig. 2**). Nonetheless, each sex contained a unique list of DhMR-associated genes, suggesting that while female and male *Cntnap2$^{-/-}$* mice do not exhibit any significant differences in behavior or aberrant neuroanatomy, *Cntnap2$^{-/-}$* mice have sex-specific epigenetic profiles. Pathways analysis of the female-specific, male-specific, and common DhMR-associated genes identified distinct pathways (**Fig. 2; Dataset 2**). Female-specific

DhMR-associated genes were enriched in one pathway, steroid biosynthesis, while male-specific DhMR-associated genes were enriched in a number of ASD-related pathways, such as MAPK signaling, Rap1 signaling, cAMP signaling, and oxytocin signaling. Similarly, DhMR-associated genes common between the sexes were enriched for ASD-related pathways. Together, these data suggest that male-specific 5hmC changes more robustly correlate to the ASD-like phenotype.

The relationship of these sex-specific data to autism can be summarized through known receptors, signaling pathways, and synaptic proteins affiliated with ASD (**Fig. 3**). These pathways include several known autism genes that harbor DhMRs, such as the neurexin gene family, contactin proteins, and reelin, as well as a number of genes with common synaptic- and receptor-related functions, including genes encoding calcium channels, potassium channels, and post-synaptic glutamatergic and GABAergic transmembrane receptor proteins. Notably, a larger set of genes/pathways associated with ASD were found to contain DhMRs in males, compared to females, suggesting that males molecularly are more susceptible to alterations in 5hmC abundance, perhaps hinting at mechanisms contributing to the higher prevalence of autism in males. Together, these data indicate that the loss of CNTNAP2 results in epigenetic alterations in genes with common functions in ASD pathways. Further studies are warranted to investigate the role of these sex-specific epigenetic changes and how they may correlate with human patients to form the basis for earlier interventions and unraveling the backdrop of sex-specific neurodevelopmental disorders.

Discussion

Here, we profiled the genome-wide distribution of striatal 5hmC in female mice lacking the gene product encoded by *Cntnap2* and identified sex-specific hydroxymethylation between female and male mutant mice. The DhMRs revealed a significant number of known autism-related genes contained differential 5hmC levels, underscoring a role for the modulation of 5hmC in ASD, while also potentially identifying novel autism-related genes and pathways. Moreover, these data seemingly indicate that a wider range of ASD-related genes contain epigenetic alterations in male mice carrying the *Cntnap2* mutation, mirroring the fact that autism is diagnosed at nearly a three-times greater rate in males than in females. Together, these data provide a foundation that will facilitate future studies interrogating the complex interactions of genes and pathways underlying autism and sex-specific neurological outcomes. Improved understanding of the molecular mechanisms that regulate the transcription rates of these differentially hydroxymethylated genes may provide potentially modifiable substrates (*e.g.,* 5hmC levels) that could ultimately be targeted for earlier diagnosis and intervention of autism, attenuating its progression.

Finding sets of common genes harboring differential hydroxymethylation in both female and male *Cntnap2*$^{-/-}$ mutant mice revealed that numerous autism-related pathways were affected by the absence of CNTNAP2. The genes for these pathways included *Auts2*, *Foxp1*, and *Nrxn3*, along with genes related to GABAergic signaling (*e.g., Gabrg3* and *Gabarapl1*)[31-35]. These findings support a role of 5hmC in GABAergic interneuron development, which is consistent with previous reports showing that mice lacking CNTNAP2 exhibit a reduction in striatal GABAergic interneuron signaling[15]. Moreover, we

identified 7 potassium channel subunits exhibited differential hydroxymethylation in both sexes of the *Cntnap2*$^{-/-}$ mutant, which matches the known role of CNTNAP2 to cluster potassium channels within myelinated axons. Additionally, CNTNAP2 maintains and facilitates neuronal-glial cell interactions and we found, in the absence of *Cntnap2*, a number of cell-adhesion proteins, such as *Cadm1*, *Dscaml1*, *Nrxn1*, *Sdk2*, and *Tenm3*[36-40], fostered DhMRs in both sexes, suggesting that the loss of *Cntnap2* results in epigenetic modulations in a large set of adhesion proteins that contribute to the aberrant neuronal migration and development in the mutant mice. The loss of *Cntnap2* also contributed to common alterations in 5hmC on genes with known functions in epigenetic pathways, including DNA methyltransferases (*Dnmt3a*), histone deacetylases (*Hdac4* and *Hdac9*), proteins involved in chromatin architecture (*Smchd1*), and several microRNAs. These findings suggest extensive regulatory changes in epigenetic pathways underlie the mutant phenotypes, similar to findings previously identified in human autism and other related neurological disorders[41-44]. Together, these data underscore the molecular similarities of female and male mutant mice and highlight that an extraordinary number of known autism-related, neuronal- and synaptic- function, and epigenetic genes and pathways harbor 5hmC disruptions in the absence of *CNTNAP2*, implicating 5hmC as a regulator of these genes and their functions, particularly in their relation with the ASD-like behaviors in mutant mice. Importantly, these data also point to novel genes and pathways that may be associated with the autism-like phenotype and previously overlooked.

While a remarkable number of genes containing DhMRs were common between the sexes, sex-specific alterations in hydroxymethylation were also found. For example, only female mutants contained differing 5hmC profiles in the well-established developmental brain disorders genes (*e.g., Nfia*, *Sdk1*, and *Nrxn1*)[33,45-48]. In contrast, only male mutants contained DhMRs in distinct known autism genes (*Dlg2*, *Grip2b*, and *Mef2c*)[49-51]. Moreover, a significant number of genes affiliated with pathways dysregulated in autism, such as MAPK signaling, Rap1 signaling, and mTOR signaling contained DhMRs in male mutants, but not female mutant mice[52-54]. While these pathways show sex-specific regulation[55-57], whether or not they have sex-specific effects in autism has not been extensively explored. Nonetheless, data presented here suggest that male mutants exhibited a greater number of molecular alterations than their female mutant counterparts, suggesting that there are more extensive ASD-related changes in males, which parallels the greater rate of diagnosis of autism in males than females[10].

Several transcription factors that recognize sequence motifs in DhMRs have known roles in autism and related neurological activities and functions. For example, circadian locomotor output cycles kaput (CLOCK) is one of the main transcription factors governing circadian rhythm[58]. Notably, dysregulation of the sleep cycle is a common comorbidity associated with autism[59], suggesting that deficits in the circadian-controlling machinery, such as CLOCK, may be disrupted in patients with autism. Aryl hydrocarbon receptor nuclear translocator (*ARNT*; also known as *HIF-1b*) encodes a protein that promotes the expression of genes involved in xenobiotic metabolism and mTOR activity, and is also dysregulated in autism[54,60]. Notably, a number of transcription factors identified to

putatively bind to motifs enriched within DhMRs show preferential binding to E-box motifs, which bind to the canonical sequence of CACGTG[61]. E-box binding proteins show extensive roles in circadian rhythm, cell proliferation and apoptosis, myogenesis, synapse formation, and cell differentiation[62-66]. The fact that many of the transcription factors binding motifs identified here have a consensus E-box sequences suggest that E-box motifs may be especially susceptible to 5hmC modifications, which may result in altered DNA affinity and dysregulated expression of genes linked to behavioral deficits associated with autism.

Identifying sex-specific alterations of 5hmC in the *Cntnap2*$^{-/-}$ knockout mouse model of autism sheds light on the underlying molecular mechanisms that may contribute to variations in the origins of altered behaviors, leading to neurodevelopmental disorders, such as autism. The DhMR-associated genes presented here represent susceptible targets within the striatal network, a brain region strongly linked with autism[67], that respond to the loss of CNTNAP2 in a sex-specific fashion. These data may serve as a benchmark for future comparisons of sex-specific alterations in developmental brain disorders, revealing genes and pathways with sex-specific susceptibilities. These future studies of behavior-related 5hmC levels will broaden our understanding of the molecular role of 5hmC in sex-specific risk, progression, pathology, and severity of neurodevelopmental disorders.

Methods

Mice

Heterozygous male *Cntnap2*$^{+/-}$ mice were purchased from the Jackson laboratories (Bar Harbor, ME) and maintained on C57BL/6J background. The mice were housed under uniform conditions in a pathogen-free mouse facility with a 12-hour light/dark cycle. Food and water were available *ad libitum*. All experiments were approved by the University of Wisconsin–Madison Institutional Animal Care and Use Committee (M02529).

Genotyping

Cntnap2 mutants and WT littermates were genotyped using the following primers: Mutant Rev: CGCTTCCTCGTGCTTTACGGTAT, Common: CTGCCAGCCCAGAACTGG, WT Rev 1: GCCTGCTCTCAGAGACATCA. PCR amplification was performed with one cycle of 95°C for 5 min and 31 cycles of 95°C for 30 s, 56°C for 30 s, 68°C for 30 s, 56°C for 30 s and 68°C for 10 min. The mutant allele was obtained with a 350-bp and WT allele with a 197-bp PCR products.

DNA extraction

Seven-week old male *Cntnap2*$^{-/-}$ mice and their WT littermates (N = 3 per group) were sacrificed (2 h after lights on), and whole brains were extracted and immediately flash-frozen in 2-methylbutane and dry ice. Striatum tissue was excised by micropunch (1.53 to −0.95 mm posterior to bregma), and ~30 milligrams of tissue was homogenized with glass beads (Sigma) and DNA was extracted using AllPrep DNA/RNA mini kit (Qiagen).

5hmC Enrichment of genomic DNA

Chemical labeling-based 5hmC enrichment was described previously[17]. Briefly, a total of 10 μg of striatum DNA was sonicated to 300 bp and incubated for 1 h at 37°C in the

following labeling reaction: 1.5 µl of N3-UDPG (2 mm); 1.5 µlβ of -GT (60 µm) and 3 µl of 10× β-GT buffer, in a total of 30 µl. Biotin was added and the reaction was incubated at 37°C for 2 h prior to capture on streptavidin-coupled dynabeads (Invitrogen, 65001). Enriched DNA was released from the beads during a 2-h incubation at room temperature with 100 mm DTT (Invitrogen, 15508013), which was removed using a Bio-Rad column (Bio-Rad, 732-6227). Capture efficiency was ~5–7% for each sample.

Library preparation and high-throughput sequencing

5hmC-enriched libraries were generated using the NEBNext ChIP-Seq Library Prep Reagent Set for Illumina sequencing, according to the manufacturer's protocol. Briefly, the 5hmC-enriched DNA fragments were purified after the adapter ligation step using AMPure XP beads (Agencourt A63880). An Agilent 2100 BioAnalyzer was used to quantify the amplified library DNA and 20-pM of diluted libraries were used for sequencing. 50-cycle single-end sequencing was performed by Beckman Coulter Genomics. Image processing and sequence extraction were done using the standard Illumina Pipeline.

Analysis of 5hmC data: sequence alignment, fragment length estimation and peak identification

We mapped the reads to mouse NCBI37v1/ mm9 reference genome using Bowtie2 2.3.4[68], only keeping the uniquely mapped reads. The Model-based Analysis of ChIP-Seq 2 (MACS2) algorithm v2.1.2[69] was used to estimate fragment size, call peaks, and identify peak summits from aligned single-end reads using the following parameters: single-end

format, effective genome size of 1.87e9, band width of 300bp, an FDR cutoff of 0.01, auto pair model process enabled, local bias computed in a surrounding 1kb window, and a maximum of one duplicate fragment to avoid PCR bias. Summits were extracted for each peak for each sample and extended +/-500bp for downstream analysis. We defined the peaks for each group as follows using the peaks from all three subjects: We first merged the peaks from all subjects in this group, and call one such region a peak for this group if it overlapped with the peaks from at least two subjects in this group.

Identification of differentially hydroxymethylated regions (DhMRs) and annotation

For each genotype, the stress and control groups were pooled and merged to form the candidate regions in the comparison within each genotype, and then Bioconductor package edgeR was used to test whether a difference of read counts exists between the two groups in each candidate region (p-value<0.05)[70]. The types of DhMRs (specific to each genotype/sex) were determined by the average log fold change in a normalized read count (logFC) between the *Cntnap2*$^{-/-}$ and WT mice. To annotate DhMRs to genomic structures and gene symbols, R package *ChIPseeker* was used, using packages *TxDb.Mmusculus.UCSC.mm9.knownGene* and *org.Mm.eg.db* for the annotation of genomic feature and gene symbol, respectively, with promoter regions deemed +/-3000bp surrounding transcription start sites[71].

Enrichment tests of genes and GO analysis

To test for the enrichment of autism genes among the DhMR-associated genes, we used a chi-square test to compare the DhMR-associated genes to a list of orthologs of well-

known autism genes (*N* = 232; C. L. Martin, personal communication). Bioconductor package *clusterProfiler* was used to test GO Biological Process (BP) term enrichment or KEGG pathways[72]. For the GO and KEGG enrichment analyses of DhMR associated genes, the gene universe consisted of all the genes associated with 5hmC peaks in tested mice (*i.e.,* genotype and sex). An FDR threshold of 0.3 was used to identify significant terms.

Sequence motif analysis

For motif discovery analysis, the Hypergeometric Optimization of Motif EnRichment (HOMER) suite of tools was utilized[73]. DNA sequences corresponding to DhMR coordinates were obtained from the mm9 genome and compared against candidate region background sequences, using the given size of the regions. Enriched known motifs of vertebrate transcription factors (*N* = 428) were determined using binomial testing and a q-value cutoff of 0.01.

Supplemental Data Availability

Supporting documents and information can be found at https://app.box.com/s/mvuc4zh2yf2asvpx5qrz9gnahn4y4254

Figure Legends

Figure 1: Characterization of DhMRs identified between homozygous mutant and wildtype mice. (A) A Manhattan plot depicts, genome-wide, the regions interrogated for differential 5hmC abundance between mutant and wildtype mice. Chromosomes (x-axis)

and the –log10(*P*-value) (y-axis) are shown. 5hmC levels were investigated across all chromosomes of the mouse genome, with a marked reduction of regions found on the X chromosome. (B) A pie chart displays the distribution of identified DhMRs in relation to standard genomic structures. (C) Enrichment map plots show the connectivity of the top identified ontological terms for hyper-DhMR-associated genes (left panel) and hypo-DhMR-associated genes (right panel). Color of circles are based on FDR *P*-value, while the size of the circles represents the relative number of DhMR-associated genes identified in each ontological term. (D) A bar plot shows the top KEGG pathway terms identified for hyper-DhMR-associated genes. The ontological term (y-axis) and the number of DhMR-associated genes associated with each term (x-axis) are displayed. The colors are based on the FDR *P*-value of each term. (E) A Venn diagram shows the number DhMR-associated genes (blue circle) overlapping with a list of well-established developmental brain disorder genes (yellow circle). Asterisks (***) represent a *P*-value <0.001, as determined by a hypergeometric enrichment test. (F) The top enriched motifs and its associated transcription factor identified from hyper-DhMR (top panel) and hypo-DhMR (bottom panel) sequences.

Figure 2: Sex-specific 5hmC levels in mutant mice. A Venn diagram (top panel) depicts the overlap of DhMR-associated genes identified between female mutant mice versus control female mice (green circle) and male mutant mice versus control male mice (grey circle). Asterisks (***) represent a *P*-value <0.001, as determined by a hypergeometric enrichment test. Dot plots of the top terms from KEGG pathway analysis are shown for the female-specific DhMR-associated genes (bottom left panel), the genes common

between the female and male comparisons (bottom middle panel), and the male-specific DhMR-associated genes (bottom right panel). The ontological term (y-axis) and proportion of DhMR-associated genes (GeneRatio; x-axis) linked to each term are shown. Color is based on FDR *P*-value of each term, and size represents the relative number of DhMR-associated genes relating to each term.

Figure 3: Summary of the DhMR associated to genes encoding proteins linked to ASD. These include synaptic proteins, receptors, and signaling pathways. Proteins/pathways highlighted in red depict DhMRs in related genes specific from the female comparison. Proteins/pathways highlighted in blue depict DhMRs in related genes specific from the male comparison. Proteins/pathways highlighted in purple depict DhMRs in related genes identified in both the female and male comparisons.

References

1 Battle, D. E. Diagnostic and Statistical Manual of Mental Disorders (DSM). *CoDAS* **25**, 191-192, doi:10.1590/s2317-17822013000200017 (2013).

2 Lai, M. C., Lombardo, M. V. & Baron-Cohen, S. Autism. *Lancet (London, England)* **383**, 896-910, doi:10.1016/s0140-6736(13)61539-1 (2014).

3 Ramaswami, G. & Geschwind, D. H. Genetics of autism spectrum disorder. *Handbook of clinical neurology* **147**, 321-329, doi:10.1016/b978-0-444-63233-3.00021-x (2018).

4 Sebat, J. *et al.* Strong association of de novo copy number mutations with autism. *Science (New York, N.Y.)* **316**, 445-449, doi:10.1126/science.1138659 (2007).

5 Shailesh, H., Gupta, I., Sif, S. & Ouhtit, A. Towards understanding the genetics of Autism. *Frontiers in bioscience (Elite edition)* **8**, 412-426, doi:10.2741/e776 (2016).

6 Weiss, L. A., Arking, D. E., Daly, M. J. & Chakravarti, A. A genome-wide linkage and association scan reveals novel loci for autism. *Nature* **461**, 802-808, doi:10.1038/nature08490 (2009).

7 Tordjman, S. *et al.* Gene x Environment interactions in autism spectrum disorders: role of epigenetic mechanisms. *Frontiers in psychiatry* **5**, 53, doi:10.3389/fpsyt.2014.00053 (2014).

8 Siu, M. T. & Weksberg, R. Epigenetics of Autism Spectrum Disorder. *Advances in*

experimental medicine and biology **978**, 63-90, doi:10.1007/978-3-319-53889-1_4 (2017).

9 Tick, B., Bolton, P., Happé, F., Rutter, M. & Rijsdijk, F. Heritability of autism spectrum disorders: a meta-analysis of twin studies. *Journal of child psychology and psychiatry, and allied disciplines* **57**, 585-595, doi:10.1111/jcpp.12499 (2016).

10 Loomes, R., Hull, L. & Mandy, W. P. L. What Is the Male-to-Female Ratio in Autism Spectrum Disorder? A Systematic Review and Meta-Analysis. *Journal of the American Academy of Child and Adolescent Psychiatry* **56**, 466-474, doi:10.1016/j.jaac.2017.03.013 (2017).

11 Strauss, K. A. *et al.* Recessive symptomatic focal epilepsy and mutant contactin-associated protein-like 2. *The New England journal of medicine* **354**, 1370-1377, doi:10.1056/NEJMoa052773 (2006).

12 Alarcón, M. *et al.* Linkage, association, and gene-expression analyses identify CNTNAP2 as an autism-susceptibility gene. *American journal of human genetics* **82**, 150-159, doi:10.1016/j.ajhg.2007.09.005 (2008).

13 Arking, D. E. *et al.* A common genetic variant in the neurexin superfamily member CNTNAP2 increases familial risk of autism. *American journal of human genetics* **82**, 160-164, doi:10.1016/j.ajhg.2007.09.015 (2008).

14 Poliak, S. *et al.* Caspr2, a new member of the neurexin superfamily, is localized at the juxtaparanodes of myelinated axons and associates with K+ channels. *Neuron* **24**, 1037-1047, doi:10.1016/s0896-6273(00)81049-1 (1999).

15 Penagarikano, O. *et al.* Absence of CNTNAP2 leads to epilepsy, neuronal migration abnormalities, and core autism-related deficits. *Cell* **147**, 235-246, doi:10.1016/j.cell.2011.08.040 (2011).

16 Papale, L. A. *et al.* Genome-wide disruption of 5-hydroxymethylcytosine in a mouse model of autism. *Human molecular genetics* **24**, 7121-7131, doi:10.1093/hmg/ddv411 (2015).

17 Song, C. X. *et al.* Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. *Nature biotechnology* **29**, 68-72, doi:10.1038/nbt.1732 (2011).

18 Szulwach, K. E. *et al.* 5-hmC-mediated epigenetic dynamics during postnatal neurodevelopment and aging. *Nature neuroscience* **14**, 1607-1616, doi:10.1038/nn.2959 (2011).

19 Khare, T. *et al.* 5-hmC in the brain is abundant in synaptic genes and shows differences at the exon-intron boundary. *Nature structural & molecular biology* **19**, 1037-1043, doi:10.1038/nsmb.2372 (2012).

20 Tabish, A. M. *et al.* Association of intronic DNA methylation and hydroxymethylation alterations in the epigenetic etiology of dilated cardiomyopathy. *American journal of physiology. Heart and circulatory physiology* **317**, H168-h180, doi:10.1152/ajpheart.00758.2018 (2019).

21 Guang, S. *et al.* Synaptopathology Involved in Autism Spectrum Disorder. *Frontiers in cellular neuroscience* **12**, 470, doi:10.3389/fncel.2018.00470 (2018).

22 Doll, C. A. & Broadie, K. Impaired activity-dependent neural circuit assembly and refinement in autism spectrum disorder genetic models. *Frontiers in cellular neuroscience* **8**, 30, doi:10.3389/fncel.2014.00030 (2014).

23 Garcia-Forn, M., Boitnott, A., Akpinar, Z. & De Rubeis, S. Linking Autism Risk

Genes to Disruption of Cortical Development. *Cells* **9**, doi:10.3390/cells9112500 (2020).

24 Janšáková, K. *et al.* Alteration of the steroidogenesis in boys with autism spectrum disorders. *Translational psychiatry* **10**, 340, doi:10.1038/s41398-020-01017-8 (2020).

25 Gregory, S. G. *et al.* Genomic and epigenetic evidence for oxytocin receptor deficiency in autism. *BMC medicine* **7**, 62, doi:10.1186/1741-7015-7-62 (2009).

26 Jung, N. H. *et al.* Impaired induction of long-term potentiation-like plasticity in patients with high-functioning autism and Asperger syndrome. *Developmental medicine and child neurology* **55**, 83-89, doi:10.1111/dmcn.12012 (2013).

27 Zamarbide, M. *et al.* Male-Specific cAMP Signaling in the Hippocampus Controls Spatial Memory Deficits in a Mouse Model of Autism and Intellectual Disability. *Biological psychiatry* **85**, 760-768, doi:10.1016/j.biopsych.2018.12.013 (2019).

28 Gonzalez-Mantilla, A. J., Moreno-De-Luca, A., Ledbetter, D. H. & Martin, C. L. A Cross-Disorder Method to Identify Novel Candidate Genes for Developmental Brain Disorders. *JAMA psychiatry* **73**, 275-283, doi:10.1001/jamapsychiatry.2015.2692 (2016).

29 Bienvenu, T. & Chelly, J. Molecular genetics of Rett syndrome: when DNA methylation goes unrecognized. *Nature reviews. Genetics* **7**, 415-426, doi:10.1038/nrg1878 (2006).

30 Gharani, N., Benayed, R., Mancuso, V., Brzustowicz, L. M. & Millonig, J. H. Association of the homeobox transcription factor, ENGRAILED 2, 3, with autism spectrum disorder. *Molecular psychiatry* **9**, 474-484, doi:10.1038/sj.mp.4001498 (2004).

31 Beunders, G. *et al.* Exonic deletions in AUTS2 cause a syndromic form of intellectual disability and suggest a critical role for the C terminus. *American journal of human genetics* **92**, 210-220, doi:10.1016/j.ajhg.2012.12.011 (2013).

32 Hamdan, F. F. *et al.* De novo mutations in FOXP1 in cases with intellectual disability, autism, and language impairment. *American journal of human genetics* **87**, 671-678, doi:10.1016/j.ajhg.2010.09.017 (2010).

33 Vaags, A. K. *et al.* Rare deletions at the neurexin 3 locus in autism spectrum disorder. *American journal of human genetics* **90**, 133-141, doi:10.1016/j.ajhg.2011.11.025 (2012).

34 Chen, C. H. *et al.* Genetic analysis of GABRB3 as a candidate gene of autism spectrum disorders. *Molecular autism* **5**, 36, doi:10.1186/2040-2392-5-36 (2014).

35 Somekh, J. *et al.* A model-driven methodology for exploring complex disease comorbidities applied to autism spectrum disorder and inflammatory bowel disease. *Journal of biomedical informatics* **63**, 366-378, doi:10.1016/j.jbi.2016.08.008 (2016).

36 Jin, J. *et al.* The Implicated Roles of Cell Adhesion Molecule 1 (CADM1) Gene and Altered Prefrontal Neuronal Activity in Attention-Deficit/Hyperactivity Disorder: A "Gene-Brain-Behavior Relationship"? *Frontiers in genetics* **10**, 882, doi:10.3389/fgene.2019.00882 (2019).

37 Agarwala, K. L. *et al.* Cloning and functional characterization of DSCAML1, a novel DSCAM-like cell adhesion molecule that mediates homophilic intercellular adhesion. *Biochemical and biophysical research communications* **285**, 760-772,

doi:10.1006/bbrc.2001.5214 (2001).
38 Südhof, T. C. Synaptic Neurexin Complexes: A Molecular Code for the Logic of Neural Circuits. *Cell* **171**, 745-769, doi:10.1016/j.cell.2017.10.024 (2017).
39 Goodman, K. M. *et al.* Molecular basis of sidekick-mediated cell-cell adhesion and specificity. *eLife* **5**, doi:10.7554/eLife.19058 (2016).
40 Antinucci, P., Nikolaou, N., Meyer, M. P. & Hindges, R. Teneurin-3 specifies morphological and functional connectivity of retinal ganglion cells in the vertebrate visual system. *Cell reports* **5**, 582-592, doi:10.1016/j.celrep.2013.09.045 (2013).
41 Wu, Y. *et al.* Aberrant Expression of Histone Deacetylases 4 in Cognitive Disorders: Molecular Mechanisms and a Potential Target. *Frontiers in molecular neuroscience* **9**, 114, doi:10.3389/fnmol.2016.00114 (2016).
42 Lang, B. *et al.* HDAC9 is implicated in schizophrenia and expressed specifically in post-mitotic neurons but not in adult neural stem cells. *American journal of stem cells* **1**, 31-41 (2012).
43 Christian, D. L. *et al.* DNMT3A Haploinsufficiency Results in Behavioral Deficits and Global Epigenomic Dysregulation Shared across Neurodevelopmental Disorders. *Cell reports* **33**, 108416, doi:10.1016/j.celrep.2020.108416 (2020).
44 Kim, K. H. *et al.* Transcriptomic Analysis of Induced Pluripotent Stem Cells Derived from Patients with Bipolar Disorder from an Old Order Amish Pedigree. *PloS one* **10**, e0142693, doi:10.1371/journal.pone.0142693 (2015).
45 Yang, R. *et al.* ANK2 autism mutation targeting giant ankyrin-B promotes axon branching and ectopic connectivity. *Proceedings of the National Academy of Sciences of the United States of America* **116**, 15262-15271, doi:10.1073/pnas.1904348116 (2019).
46 Wang, T. *et al.* Large-scale targeted sequencing identifies risk genes for neurodevelopmental disorders. *Nature communications* **11**, 4932, doi:10.1038/s41467-020-18723-y (2020).
47 Corley, M. J. *et al.* Epigenetic Delay in the Neurodevelopmental Trajectory of DNA Methylation States in Autism Spectrum Disorders. *Frontiers in genetics* **10**, 907, doi:10.3389/fgene.2019.00907 (2019).
48 Onay, H. *et al.* Mutation analysis of the NRXN1 gene in autism spectrum disorders. *Balkan journal of medical genetics : BJMG* **19**, 17-22, doi:10.1515/bjmg-2016-0031 (2016).
49 Yoo, T. *et al.* A DLG2 deficiency in mice leads to reduced sociability and increased repetitive behavior accompanied by aberrant synaptic transmission in the dorsal striatum. *Molecular autism* **11**, 19, doi:10.1186/s13229-020-00324-7 (2020).
50 Mejias, R. *et al.* Gain-of-function glutamate receptor interacting protein 1 variants alter GluA2 recycling and surface distribution in patients with autism. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 4920-4925, doi:10.1073/pnas.1102233108 (2011).
51 Harrington, A. J. *et al.* MEF2C Hypofunction in Neuronal and Neuroimmune Populations Produces MEF2C Haploinsufficiency Syndrome-like Behaviors in Mice. *Biological psychiatry* **88**, 488-499, doi:10.1016/j.biopsych.2020.03.011 (2020).
52 Vithayathil, J., Pucilowska, J. & Landreth, G. E. ERK/MAPK signaling and autism spectrum disorders. *Progress in brain research* **241**, 63-112,

doi:10.1016/bs.pbr.2018.09.008 (2018).
53 Huang, M. *et al.* Two Autism/Dyslexia Linked Variations of DOCK4 Disrupt the Gene Function on Rac1/Rap1 Activation, Neurite Outgrowth, and Synapse Development. *Frontiers in cellular neuroscience* **13**, 577, doi:10.3389/fncel.2019.00577 (2019).
54 Ganesan, H. *et al.* mTOR signalling pathway - A root cause for idiopathic autism? *BMB reports* **52**, 424-433, doi:10.5483/BMBRep.2019.52.7.137 (2019).
55 Krementsov, D. N. *et al.* Sex-specific control of central nervous system autoimmunity by p38 mitogen-activated protein kinase signaling in myeloid cells. *Annals of neurology* **75**, 50-66, doi:10.1002/ana.24020 (2014).
56 Martínez, P. *et al.* RAP1 protects from obesity through its extratelomeric role regulating gene expression. *Cell reports* **3**, 2059-2074, doi:10.1016/j.celrep.2013.05.030 (2013).
57 Baar, E. L., Carbajal, K. A., Ong, I. M. & Lamming, D. W. Sex- and tissue-specific changes in mTOR signaling with age in C57BL/6J mice. *Aging cell* **15**, 155-166, doi:10.1111/acel.12425 (2016).
58 Menet, J. S., Pescatore, S. & Rosbash, M. CLOCK:BMAL1 is a pioneer-like transcription factor. *Genes & development* **28**, 8-13, doi:10.1101/gad.228536.113 (2014).
59 Bourgeron, T. The possible interplay of synaptic and clock genes in autism spectrum disorders. *Cold Spring Harbor symposia on quantitative biology* **72**, 645-654, doi:10.1101/sqb.2007.72.020 (2007).
60 Serajee, F. J., Nabi, R., Zhong, H. & Huq, M. Polymorphisms in xenobiotic metabolism genes and autism. *Journal of child neurology* **19**, 413-417, doi:10.1177/088307380401900603 (2004).
61 Chaudhary, J. & Skinner, M. K. Basic helix-loop-helix proteins can act at the E-box within the serum response element of the c-fos promoter to influence hormone-induced promoter activation in Sertoli cells. *Molecular endocrinology (Baltimore, Md.)* **13**, 774-786, doi:10.1210/mend.13.5.0271 (1999).
62 Yoshitane, H. *et al.* CLOCK-controlled polyphonic regulation of circadian rhythms through canonical and noncanonical E-boxes. *Molecular and cellular biology* **34**, 1776-1787, doi:10.1128/mcb.01465-13 (2014).
63 Desbarats, L., Gaubatz, S. & Eilers, M. Discrimination between different E-box-binding proteins at an endogenous target gene of c-myc. *Genes & development* **10**, 447-460, doi:10.1101/gad.10.4.447 (1996).
64 Malik, S., Huang, C. F. & Schmidt, J. The role of the CANNTG promoter element (E box) and the myocyte-enhancer-binding-factor-2 (MEF-2) site in the transcriptional regulation of the chick myogenin gene. *European journal of biochemistry* **230**, 88-96, doi:10.1111/j.1432-1033.1995.tb20537.x (1995).
65 Carrasco-Serrano, C., Campos-Caro, A., Viniegra, S., Ballesta, J. J. & Criado, M. GC- and E-box motifs as regulatory elements in the proximal promoter region of the neuronal nicotinic receptor alpha7 subunit gene. *The Journal of biological chemistry* **273**, 20021-20028, doi:10.1074/jbc.273.32.20021 (1998).
66 Uribesalgo, I. *et al.* E-box-independent regulation of transcription and differentiation by MYC. *Nature cell biology* **13**, 1443-1449, doi:10.1038/ncb2355 (2011).

67 Fuccillo, M. V. Striatal Circuits as a Common Node for Autism Pathophysiology. *Frontiers in neuroscience* **10**, 27, doi:10.3389/fnins.2016.00027 (2016).
68 Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology* **10**, R25, doi:10.1186/gb-2009-10-3-r25 (2009).
69 Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome biology* **9**, R137, doi:10.1186/gb-2008-9-9-r137 (2008).
70 Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics (Oxford, England)* **26**, 139-140, doi:10.1093/bioinformatics/btp616 (2010).
71 Yu, G., Wang, L. G. & He, Q. Y. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics (Oxford, England)* **31**, 2382-2383, doi:10.1093/bioinformatics/btv145 (2015).
72 Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics : a journal of integrative biology* **16**, 284-287, doi:10.1089/omi.2011.0118 (2012).
73 Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular cell* **38**, 576-589, doi:10.1016/j.molcel.2010.05.004 (2010).

Figure 1

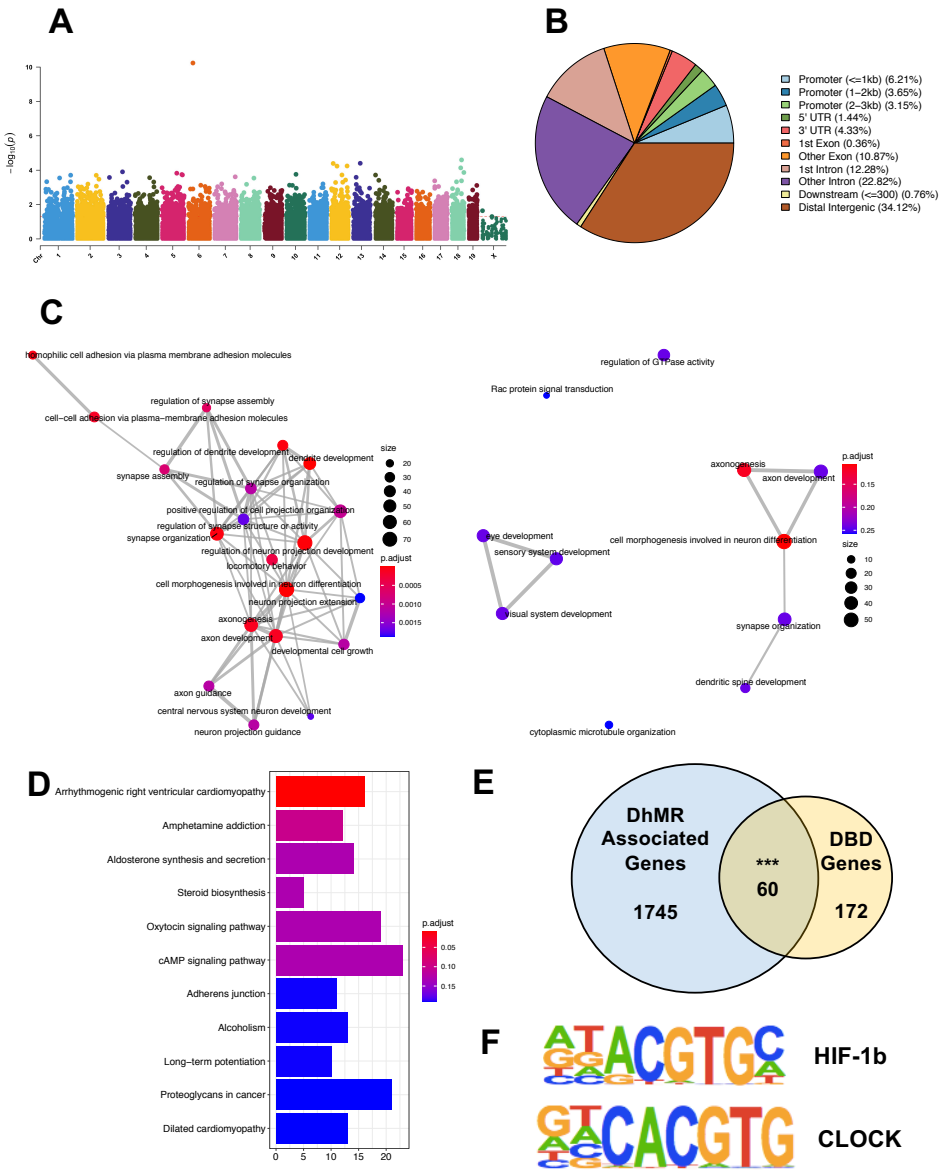

A
B
Promoter (<=1kb) (6.21%)
Promoter (1–2kb) (3.65%)
Promoter (2–3kb) (3.15%)
5' UTR (1.44%)
3' UTR (4.33%)
1st Exon (0.36%)
Other Exon (10.87%)
1st Intron (12.28%)
Other Intron (22.82%)
Downstream (<=300) (0.76%)
Distal Intergenic (34.12%)
C
homophilic cell adhesion via plasma membrane adhesion molecules
regulation of synapse assembly
cell–cell adhesion via plasma–membrane adhesion molecules
regulation of dendrite development
dendrite development
synapse assembly
regulation of synapse organization
positive regulation of cell projection organization
regulation of synapse structure or activity
synapse organization
regulation of neuron projection development
locomotory behavior
cell morphogenesis involved in neuron differentiation
neuron projection extension
axonogenesis
axon development
developmental cell growth
axon guidance
central nervous system neuron development
neuron projection guidance
size
p.adjust
regulation of GTPase activity
Rac protein signal transduction
axonogenesis
axon development
cell morphogenesis involved in neuron differentiation
sensory system development
visual system development
synapse organization
dendritic spine development
cytoplasmic microtubule organization
p.adjust
size
D
Arrhythmogenic right ventricular cardiomyopathy
Amphetamine addiction
Aldosterone synthesis and secretion
Steroid biosynthesis
Oxytocin signaling pathway
cAMP signaling pathway
Adherens junction
Alcoholism
Long–term potentiation
Proteoglycans in cancer
Dilated cardiomyopathy
p.adjust
E
DhMR Associated Genes
1745
***
60
DBD Genes
172
F
HIF-1b
CLOCK

Figure 2

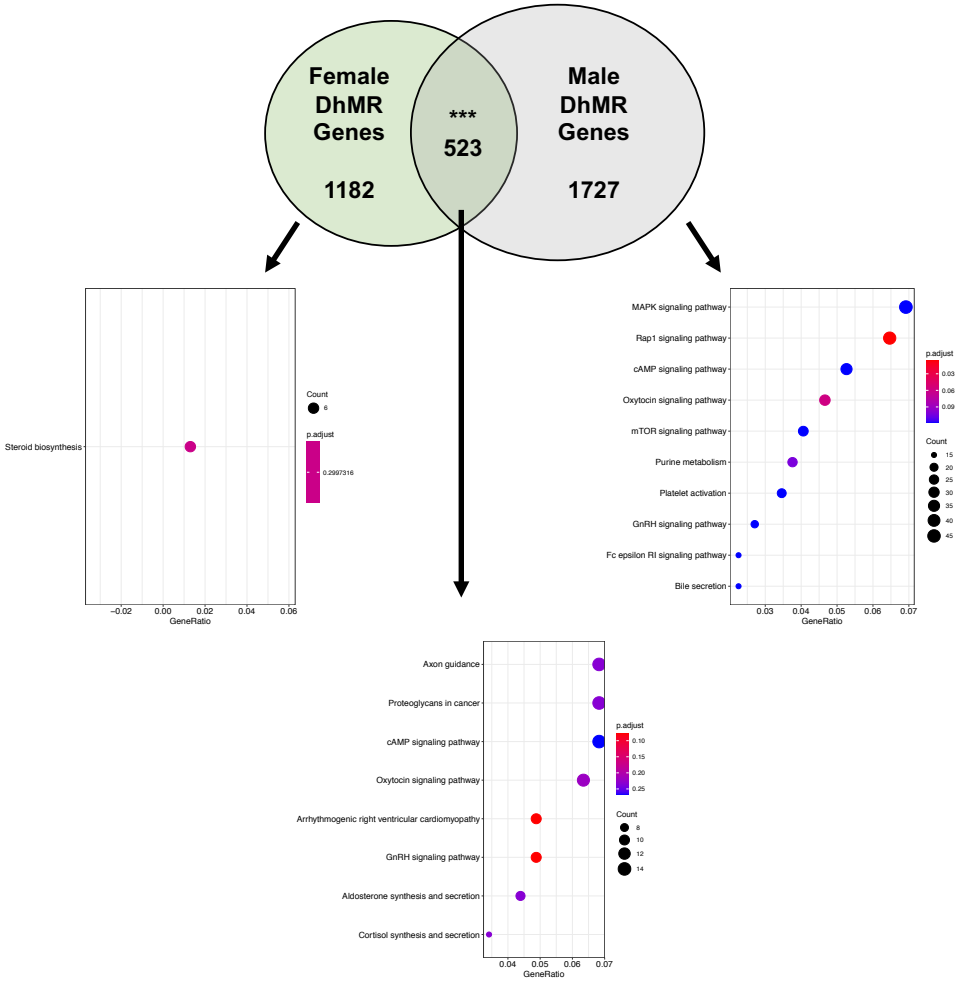
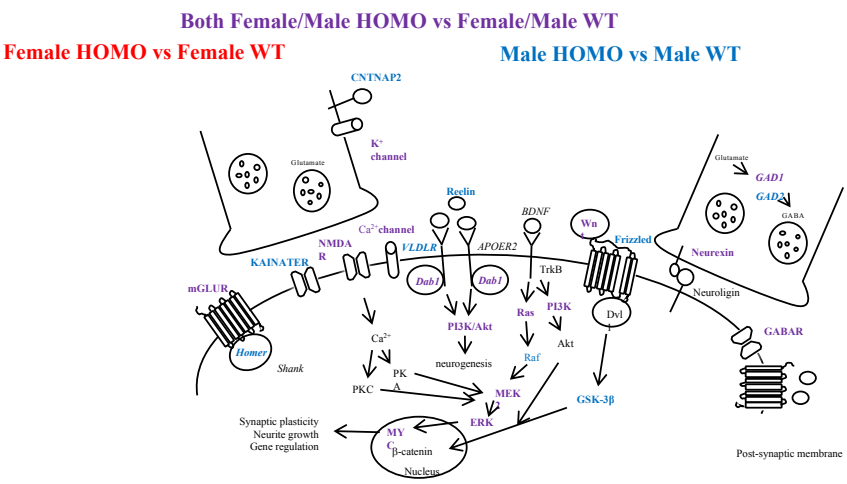

Female DhMR Genes
1182
***
523
Male DhMR Genes
1727
Steroid biosynthesis
Count
6
p.adjust
0.2997316
−0.02
0.00
0.02
0.04
0.06
GeneRatio
MAPK signaling pathway
Rap1 signaling pathway
cAMP signaling pathway
Oxytocin signaling pathway
mTOR signaling pathway
Purine metabolism
Platelet activation
GnRH signaling pathway
Fc epsilon RI signaling pathway
Bile secretion
p.adjust
0.03
0.06
0.09
Count
15
20
25
30
35
40
45
0.03
0.04
0.05
0.06
0.07
GeneRatio
Axon guidance
Proteoglycans in cancer
cAMP signaling pathway
Oxytocin signaling pathway
Arrhythmogenic right ventricular cardiomyopathy
GnRH signaling pathway
Aldosterone synthesis and secretion
Cortisol synthesis and secretion
p.adjust
0.10
0.15
0.20
0.25
Count
8
10
12
14
0.04
0.05
0.06
0.07
GeneRatio

Figure 3


Subset of DhMR-associated genes
Both Female/Male HOMO vs Female/Male WT
Female HOMO vs Female WT
Male HOMO vs Male WT
CNTNAP2
K+ channel
Glutamate
Reelin
BDNF
Wn
Frizzled
Glutamate
GAD1
GAD2
GABA
Ca2+ channel
NMDA R
VLDLR
APOER2
TrkB
Neurexin
KAINATER
Dab1
Dab1
PI3K
Ras
Dvl
Neuroligin
mGLUR
PI3K/Akt
Ca2+
Akt
GABAR
Homer
Shank
neurogenesis
Raf
PK A
PKC
MEK
GSK-3β
ERK
Synaptic plasticity
Neurite growth
Gene regulation
MY C
β-catenin
Nucleus
Post-synaptic membrane

CHAPTER 3

Early-life stress of *Cntnap2* heterozygous mice demonstrates a gene by environment interaction model for developmental brain disorders.

This chapter is adapted from:

Papale, L.A.*, Madrid, A.*, Zhang, Q.*, Chen, K., Sak, L., Keles, S., Alisch, R.S. (2021). Gene by environment interaction mouse model reveals a functional role for 5-hydroxymethylcytosine in neurodevelopmental disorders. (*under review*)

Abstract

Humans harboring a loss of contactin-associated protein-like 2 (*CNTNAP2*) function present with severe developmental delays and intellectual disabilities. Mouse knockouts of *Cntnap2* (*Cntnap2*$^{-/-}$) exhibit altered neurodevelopmental behavior and deficits in striatal GABAergic signaling and have a genome-wide disruption of an environmentally sensitive DNA methylation modification (5-hydroxymethylcytosine, 5hmC) in the orthologs of a significant number of genes implicated in human neurodevelopmental disorders. Since 5hmC levels are sensitive to environmental stress, we examined whether adult *Cntnap2* heterozygous mice (*Cntnap2*$^{+/-}$, lacking behavioral or neuropathological abnormalities) subjected to a prenatal stress would have disruptions in brain 5hmC levels and exhibit altered behaviors similar to the knockout mice. Adult early-life stressed *Cntnap2*$^{+/-}$ female mice showed repetitive behaviors and altered sociability, similar to the homozygote phenotype. Genomic profiling revealed disruptions in hippocampal and striatal 5hmC levels that were correlated to altered transcript levels of genes linked to these phenotypes (*e.g., Reln*, *Dst*, *Gigyf1, Palld* and *Fry*). Sequence motif analysis of the differentially hydroxymethylated regions found significant enrichments of transcription factor sequence motifs, including transcription factors pivotal in the maintenance of neurodevelopment, such as CLOCK. Together, these data support gene by environment hypotheses for the origins of mental illness and provide a means to identify the elusive factors contributing to complex human diseases.

Keywords: Developmental brain disorders; Gene by Environment; 5hmC

Introduction

Developmental brain disorders usually present years after birth; however, the molecular pathogenesis is thought to arise earlier, either during pregnancy or just after birth. Genetic studies have revealed hundreds of rare risk genes that cumulatively account for less than 25% of neurodevelopmental disorder cases[1-3]. The search for environmental factors contributing to the origins of neurodevelopmental disorders has been motivated by their multifactorial genetic heredity and incomplete concordance in monozygotic twins[4]. There is clear evidence that individual differences in the developmental-timing and severity of early-life stressors can predict emotional dispositions later in life and influence the risk for the development of psychopathology[5,6]. Among the earliest adverse experiences is early-life exposure to maternal depression and anxiety, which confers a lifelong risk for behavioral disturbances in childhood and beyond[7]. This process of “fetal programming” may be mediated, in part, by the impact of the early-life experience on the developing hypothalamic-pituitary-adrenal (HPA) axis[8,9]. The HPA axis is a dynamic metabolic system that regulates homeostatic mechanisms, such as the ability to respond to stressors in early embryonic development, and is highly sensitive to early life adversities[10]. The HPA axis also governs the activity of sex-specific endocrine mechanisms and responds to stress by altering the neuronal epigenome[11].

5-hydroxymethylcytosine (5hmC) is an environmentally sensitive epigenetic mark that is highly enriched in neuronal cells[12-15] and is associated with the regulation of neuronal activity, suggesting that 5hmC plays an important role in coordinating transcriptional activity in the brain[16]. These findings have prompted investigations into the potential role(s) of 5hmC in mental illness, where it has been linked to developmental

brain disorders (*e.g.,* fragile X syndrome, Rett syndrome, and autism) and neurodegenerative diseases (*e.g.,* Huntington's and Alzheimer's)[17-23]. Recently we reported that the *Cntnap2* knockout mouse model of autism exhibits a genome-wide disruption of 5hmC in a significant number of orthologous genes and pathways common to human neurodevelopmental disorders (N = 57/232)[24], supporting an interaction between *Cntnap2* and 5hmC in developmental brain disorders.

The *Cntnap2* (or *Caspr2*) gene encodes a neuronal transmembrane protein member of the neurexin superfamily involved in neuron–glia interactions and clustering of potassium channels in myelinated axons[25,26]. Humans harboring a homozygous loss of *CNTNAP2* function exhibit severe developmental delays and intellectual disabilities, contributing to a variety of neurodevelopmental disorders such as Gilles de la Tourette syndrome, obsessive-compulsive disorder, cortical dysplasia-focal epilepsy syndrome, autism, schizophrenia, Pitt-Hopkins syndrome, and attention deficit hyperactivity disorder[27]. Mice harboring a homozygous loss of *Cntnap2* function exhibit parallels to the major neuropathological features in autism spectrum disorders, including abnormal social behavior, deficits in communication, and stereotypic repetitive behaviors, cortical dysplasia and focal epilepsy[28]. In addition, these mice exhibit defects in neuronal migration of cortical projection neurons and a reduction in the number of striatal GABAergic interneurons. In contrast, heterozygous *Cntnap2* mutant mice lack any of the behavioral or neuropathological abnormalities found in the homozygous mutant[28].

Recent evidence in humans does not support a relationship between heterozygous deletions of *CNTNAP2* and neurodevelopmental disorders[29]; however, it remains possible that the combination of heterozygous *CNTNAP2* deletions in a genomic

background of increased risk may lead to mental illness[29]. With the onset of *Cntnap2* expression in mice beginning on embryonic day 14, its expression may be sensitive to prenatal (early-life) stressors that alter the neuronal epigenome. Thus, we hypothesized that *Cntnap2* heterozygous mutants exposed to early-life stress would have genome-wide disruptions in 5hmC levels and behavioral deficits, similar to the homozygous mutant. Using a stress paradigm that included seven-days of mild and variable prenatal stress from embryonic day (E) 12 to E18, we provide evidence for a gene by environment model of sex-specific social-related deficits and identify a mechanistic role for 5hmC that contributes to this behavioral outcome.

Results

Early-life stressed Cntnap2 heterozygous mice show altered social & repetitive behaviors

To test the long-lasting effects of a prenatal (early-life) environmental stress on the development of altered behaviors, we subjected WT pregnant mice (carrying WT and *Cntnap2*$^{+/-}$ pups) to seven-days of variable prenatal stress from embryonic day (E) 12 to E18, which overlaps the onset of *Cntnap2* expression (E14; Methods)[28,30]. These variable stressors included: 36 hours of constant light, 15 min of fox odor beginning 2 hours after lights on, overnight exposure to novel objects, 5 minutes of restraint stress beginning 2 hours after lights on, overnight novel noise, 12 cage changes during light period, and overnight saturated bedding[31,32]. These mild stressors were previously published and were selected because they do not include pain or directly influence maternal food intake or weight gain. Importantly, both WT and HET mice were taken from the same litter, providing an ideal internal control for our findings. Offspring were monitored for consistent

maternal/offspring interactions and left undisturbed until weaning day (postnatal day 18). At weaning, same sex offspring of both genotypes were randomly distributed into groups for behavioral or molecular testing and left undisturbed until three months of age. A maximum of three mice from each litter were assigned for behavioral and a maximum of two mice from each litter were assigned for molecular testing (Methods). At three months of age, early-life stressed (ELS) heterozygous *Cntnap2* (ELS-HET) and ELS wild-type (ELS-WT), as well as non-ELS HET (control-HET) and non-ELS WT (control-WT) offspring, were subjected to behavioral testing or sacrificed for molecular analysis. All behavioral groups (ELS and controls for both genotypes and sex, N = 9-12 per genotype and sex) were assessed using a variety of behavioral tests for general locomotor activity, anxiety and depression-like levels and social aptitude, including open field, elevated plus maze, light and dark box, forced swim tests and social interaction tests. Only female ELS-HET mice showed altered social and repetitive behaviors in two independent social behavioral tests, the 3-chamber social test and a 10-minute reciprocal social interaction test, respectively. As expected for the highly social strain C57/B6J, female controls (control-WT, ELS-WT, and control-HET) showed a significant preference to be in the chamber and interacting with the cup containing an unfamiliar mouse in the 3-chamber social test. In contrast, female ELS-HET mice mirror the homozygous *Cntnap2* mutants by not showing a significant preference (differential in mean time spent in each chamber was 65.3 seconds (s) for control-WT, 80.3s for ELS-WT, 102.4s for control-HET, and 27.8s for ELS-HET; Two-way Anova, P-value < 0.05 for controls only; **Figure 1A-B; Supplemental Fig. 1A&D**)[28]. Similarly, the time spent interacting with the unfamiliar mouse instead of the empty cup during the 3-chamber test was significant for the controls

but not for the ELS-HET female mice (differential in mean time spent interacting in each chamber was 50.2s for control-WT, 53.6s for ELS-WT, 65.4s for control-HET, 16.1s for ELS-HET; Two-way Anova, *P*-value < 0.05 for controls only; **Figure 1C-D**). The mice were then subjected to an additional social interaction test that confirmed the female ELS-HET mice spent significantly more time grooming/digging than interacting with an unfamiliar mouse compared to all controls (total time spent grooming/digging among pairs of mice matched for treatment, genotype, and sex are 96.4s for control-WT, 74.4s for ELS-WT, 101.2s for control-HET, and 160.8s for ELS-HET; One-way Anova, *P*-value < 0.05, **Figure 1E**). This repetitive grooming/digging behavior phenotype was only present in a social setting, as it was not observed in mice allowed to explore a cage alone for 10 min (total time spent grooming/digging during a 10 minute cage exploratory test are 64.5s for control-WT, 52.2s for ELS-WT, 59.3s for control-HET, and 44.2s for ELS-HET; Two-way Anova, *P*-value > 0.05, **Table 1**). Deficits in these 3 social-related measures were reproducible in an independent cohort of mice (*N* = 8-10 per treatment, sex, and genotype, **Supplemental Fig. 1)**. Notably, male mice (control-WT, control-HET, ELS-WT, and ELS-HET) did not show any altered social or repetitive behavior phenotypes (**Supplemental Fig. 2**). In addition, all behavioral groups showed no significant differences in general locomotor activity and anxiety and depression-like levels (*P*-value > 0.05; **Table 1 and Supplemental Table 1**). Together, these data support altered social and repetitive behaviors in female ELS-HET mice, revealing a sex-specific gene by environment (GxE) interaction effect on behavior. Since homozygous mutants show similar altered social and repetitive-like behaviors[28] and genome-wide disruptions in 5hmC[24], we next sought to examine the 5hmC profile in this GxE interaction model.

Long-lasting disruption of 5hmC in the female GxE interaction model

To determine the effect of this GxE interaction on the genome-wide distribution of 5hmC, we employed an established chemical labeling and affinity purification method, coupled with high-throughput sequencing technology[12,24,33,34]. Age-matched ELS-HET and control HET, as well as ELS-WT and control-WT female offspring (*N* = 3/group) were sacrificed at 3 months of age as independent biological replicates. 5hmC-containing DNA sequences were enriched from pools (*N* = 2) of hippocampal tissues, a brain region involved in the pathology of social behavior with high expression of the *Cntnap2* gene from embryonic day 14 and previously shown to be correlated with altered sociability in the *Cntnap2*$^{-/-}$ mice[35]. High-throughput sequencing resulted in a range of ~25-43 million uniquely mapped reads from each pool of biological replicates (**Supplemental Table 2**, methods). These data showed no visible differences among the chromosomes, except depletion on chromosomes X, which is consistent with previous observations[12,24,34], and an overall equal distribution of 5hmC levels in both genotypes and in both conditions on defined genomic structures and repetitive elements. (**Supplemental Fig. 3**).

Identification and characterization of differentially hydroxymethylated regions (DhMRs) in the female GxE interaction model

To identify distinct 5hmC distribution patterns following early-life stress throughout the genome of both *Cntnap2* heterozygous and WT female mice, we characterized differentially hydroxymethylated regions (DhMRs) between: 1) ELS-HET and control-HET groups, to investigate the long lasting effects of the GxE interaction on 5hmC levels (HET-

DhMRs); as well as 2) ELS-WT and control-WT groups, to determine the long-lasting effects of early-life stress on 5hmC levels (WT-DhMRs). A total of 1,381 ELS-HET increases in hydroxymethylation (ELS-HET-specific 5hmC) and 1,231 ELS-HET decreases in hydroxymethylation (control-HET-specific 5hmC) were found and these loci were distributed across all chromosomes (**Figure 2A**; **Dataset 1**; methods). In contrast, only a total of 735 significant ELS-WT increases in hydroxymethylation (ELS-WT-specific 5hmC) and 710 significant ELS-WT decreases in hydroxymethylation (control-WT-specific 5hmC) were found and these loci were distributed across all chromosomes (**Supplemental Fig. 4A**; **Dataset 1;** methods). These data indicate that the GxE interaction results in nearly twice the number of changes in 5hmC levels, but also suggest that early-life stress alone is sufficient to cause stable changes in 5hmC levels. Since specific regions of the genome are differentially methylated based on the biological functions of the genes contained within each region, we next determined if GxE and ELS-induced DhMRs are enriched or depleted on certain chromosomes using a binomial test of all detected 5hmC peaks as the background (Methods). This analysis revealed that DhMRs were not significantly enriched or depleted on any chromosomes (binomial *P*-value < 0.05; **Figure 2A; Supplemental Fig. 4A**).

We proceeded to define the genomic features associated with the GxE interaction by annotating HET-DhMRs to overlapping gene structures or to intergenic regions if more than 10 kilobases (kb) away from any gene structure. The overall distribution of these data indicated that the largest fractions of HET-DhMRs were intronic (~34%) or in distal intergenic regions (~30%) (**Figure 2B**). Binomial testing revealed that the HET-DhMRs were significantly depleted in exons and promoter regions (**Supplemental Fig. 5A**;

binomial *P*-value < 0.05). While the distribution of WT-DhMRs also revealed structures with significant fluctuations, the genomic location of the WT-DhMR fluctuations were unique to those found for the HET-DhMRs (**Supplemental Figs. 4B and 5B**). Together, these data indicate that the GxE interaction- and stress-alone-induced DhMRs are unique and not randomly distributed throughout the genome.

Finally, DNA hypermethylation on repetitive elements is believed to play a critical role in maintaining genomic stability[36]. To investigate the genome-wide disruptions of 5hmC on repetitive elements, we aligned the total 5hmC reads and DhMRs to RepeatMasker and the segmental duplication tracks of NCBI37v1/mm9 and found that the largest fraction of HET-DhMRs (~70%) was in the SINE and simple repeat regions of the genome (**Supplemental Fig. 5C**). The WT-DhMRs were enriched in LTR repeats and were specifically enriched in DNA repeats and SINEs (ELS-WT- and Control-WT-specific, respectively; **Supplemental Fig. 5D**; binomial *P*-value < 0.05). These data suggest a long-term stability of 5hmC at the majority of repeat sequences in female mice following early-life stress, which could either imply stabilization of 5mC in repetitive elements and appear as hypermethylation, a response previously observed in response to stress[37], or hypomethylation at these sites (removing the ability to generate 5hmC), a response previously observed in PTSD patients (but not controls) following repeated stress[38].

Annotation of GxE interaction DhMRs to genes reveals known and potentially novel stress and neurodevelopmental-related genes

To determine the genes and pathways affected by the long-lasting molecular effects of a GxE interaction on female mice, we annotated the HET-DhMRs to genes and

found 895 and 770 genes that contained ELS-HET- and control-HET-specific 5hmC levels, respectively. A significant number of HET-DhMR-associated genes were found in well-known stress genes such as *Homer1*[39] *and Ncam1*[40] (524/1600 χ^2 test *P*-value < 0.0001). Moreover, several of the HET-DhMR-associated genes were previously linked with developmental brain disorders including genes that encode synaptic proteins/receptors (*e.g.,* glutamate, GABAergic, calcium channels, potassium channels, sodium channels, as well as neurexin and neuroligin transmembrane proteins). To examine whether the DhMR-associated genes were enriched for neurodevelopmental-related genes, we compared them to a list of developmental brain disorder genes (*N* = 232 genes; methods)[41]. This comparison revealed that a significant number of orthologs (*N* = 56 of 232; hypergeometric test *P*-value < 0.01) harbor HET-DhMRs, suggesting that 5hmC has a molecular role in the disruption of neurodevelopment following a GxE interaction. Indeed, this finding is similar to the DhMR-associated genes found in the *Cntnap2* homozygous mouse[24]. The HET-DhMR-associated genes also included several genes known to function in epigenetic pathways, including *Dnmt3a*, *Tet2*, and *Hdac4.* To further examine the biological significance of the HET DhMRs-associated genes, we performed separate gene ontological (GO) analyses of the ELS- and control-specific HET DhMR-associated genes (*N* = 895 and 770, respectively) and found a number of neuronal-related ontological terms only among the ELS-specific HET DhMR-associated genes, which included regulation of synaptic organization, axon development, and neuron projection morphogenesis (*P*-value < 0.05 & FC > 1.5 for GO enrichment tests; **Figure 2C**; **Dataset 2**). In contrast, the control-specific HET DhMR-associated genes were related to organ-related developmental and morphological processes (**Figure 2D**;

Dataset 2). Together, increased 5hmC in the GxE model results in disruptions of genes that encode synaptic proteins/receptors and signaling pathways, supporting a link to brain-related disorders.

Genomic annotation of the WT-DhMRs to genes revealed that 510 and 450 genes contained ELS-WT- and control-WT-specific 5hmC, respectively. Notably, nearly 80% of WT-DhMR-associated genes were unique from the HET-DhMR-associated genes (**Figure 2E**), indicating that the majority of hydroxymethylation changes were specific to genotype. A significant number of WT DhMR-associated genes were previously linked to stress, including *Dnmt1*, *Pten*, and *Fkbp5*[42-44] _ENREF_41 (312/931; hypergeometric test *P*-value < 0.01). Moreover, a significant number of orthologs, best known for their role in human neurodevelopmental disorders, have disruptions in 5hmC levels (35/232; hypergeometric *P*-value < 0.01; *e.g., Aut2*, *Nckap5*, *Setbp1*[24,41]), further implicating stress-alone in the pathogenesis of neurodevelopmental disorders. Finally, WT DhMR-associated genes included *Dnmt1* from the epigenetic pathway. Annotation of the ELS- and control-specific WT DhMR-associated genes (*N* = 510 and 450 genes, respectively) to GO terms found a significant overrepresentation of terms connected with metabolism/catabolism and limb development/morphogenesis (**Supplemental Fig. 4C and D**; **Dataset 2**). Together, these data indicate that early-life stress alone disrupts 5hmC on biologically relevant genes and pathways.

Confirmation of long-lasting disruptions of 5hmC in the GxE interaction model

As an initial confirmation of genome-wide disruption of 5hmC following early-life stress, we profiled 5hmC levels in an independent cohort of mice and another brain region

associated with social and stress-related behaviors (striatum) that was recently shown to contribute to pathways of risk in female autism spectrum disorders[45]. Age-matched ELS-HET and control-HET, as well as ELS-WT and control-WT female mice (*N* = 3/group) were sacrificed at 3 months of age as independent biological replicates. 5hmC-containing DNA sequences were enriched from striatal total DNA and high-throughput sequencing resulted in a range of ~20-70 million uniquely mapped reads from each biological replicate (**Supplemental Table 3**, methods). Similar to the hippocampal 5hmC data, sequence read density mapping showed no visible differences among the chromosomes, except depletion on chromosome X, and an overall equal distribution of 5hmC levels in both genotypes and in both conditions on defined genomic structures and repetitive elements (**Supplemental Fig. 6**). To compare the long-lasting molecular effects of a gene by environment interaction on female mice in the striatum and hippocampus, we annotated the striatal HET-DhMRs to genes and found a significant overlap of differentially hydroxymethylated genes between both tissues (*N* = 389 genes; *P*-value < 0.05), including genes well known to play a role in neurodevelopmental disorders (*e.g., Nrxn1, Nlgn1,* and *Grip1/Ncoa2*; **Supplemental Fig. 7A**). Again, similar to the hippocampal findings, the gene ontologies of the striatal DhMR-associated gene contained a significant neuronal enrichment of terms (*e.g.,* synapse organization/transmission (glutamatergic) and negative regulation of axonogenesis; (*P*-value < 0.05 & FC > 1.5 for GO enrichment tests; Dataset 2). The DNA hydroxymethylation levels of several regions also were validated using an alternative method (**Supplemental Fig. 7B**; Methods). Together, these striatal 5hmC data provide a validation that 5hmC levels are stably disrupted following a GxE interaction in brain regions linked to stress response and social behavior.

Recently we reported that male *Cntnap2* knockout mice exhibit a genome-wide disruption of 5hmC in a significant number of orthologous genes and pathways common to human neurodevelopmental disorders[24]. Since the early-life stressed *Cntnap2*$^{+/-}$ female mice showed repetitive behaviors and altered sociability, similar to the homozygote phenotype, we compared the DhMR-associated genes between these two models (*e.g.,* ELS-HET and female homozygotes) and found a significant overlap of genes in the striatum (N = 443; P-value < 0.05; **Supplemental Fig. 7C**). In addition, the DhMR-associated genes from the homozygote striatum also had a significant overlap of gene when compared to the ELS-HET genes from the hippocampus (N = 356; P-value < 0.05; **Supplemental Fig. 7C**). Together, these data further validate the finding that 5hmC disruptions are linked to neurodevelopment-related disorders and support gene by environment hypotheses for complex disorders.

Candidate functional DhMRs in the female GxE interaction model

To gain insight into the potential molecular mechanism(s) for long-lasting GxE interaction-induced DhMRs, we used RNA sequencing (RNAseq) to profile gene expression in the hippocampus of the same mice surveyed for 5hmC (see methods). Comparison of transcript levels in ELS-HET and control-HET mice revealed 524 differentially expressed genes, 162 upregulated and 363 downregulated (FDR P-value < 0.1; **Figure 3A**; **Dataset 3**). Separate gene ontologies of these 162 upregulated and 363 downregulated genes found terms related to metabolic-related processes (upregulated genes; **Figure 3B**; **Dataset 4**) and axonal development and histone modifications (downregulated genes; **Figure 3C; Dataset 4**; methods). DhMRs associated with

differentially expressed genes represent candidate functional DhMRs that may have a direct role in gene regulation. An overlay of the HET-DhMR data with these HET-RNAseq data revealed that 72 differentially expressed genes harbor a HET-DhMR, which included neurodevelopmental disorder genes (*e.g., Reln*, *Dst,* and *Trio*)[41], known stress-related genes (*e.g., Ncoa2*)[46] and epigenetic genes (*Dnmt3a* and *Hdac4*; **Dataset 3**). While general relationships between 5hmC abundance and gene expression levels were not observed in these genes, there was a significant enrichment of DhMRs located in the introns of differentially expressed genes (*P*-value < 0.05). These results indicate that the GxE interaction induces long-lasting effects in the regulation of genes functioning in neuronal development and synaptic plasticity and supports an association between 5hmC and the altered social behavior observed in these adult mice.

Comparison of transcript levels in ELS-WT and control-WT mice revealed only 5 genes that were altered by early-life stress, all downregulated (FDR *P*-value < 0.1; **Supplemental Fig. 8**; **Dataset 3**). An overlay of the WT-DhMR data with the WT-RNAseq data did not reveal any potentially functional DhMRs. These results suggest that WT-DhMRs represent stable molecular alterations that may have affected gene expression at earlier developmental time-points, suggesting that the ELS-WT mice may have earlier phenotypes related to the early-life stress.

The GxE interaction may alter transcription factor binding in DhMRs

Syndromic forms of neurodevelopmental disorders can involve the disruption of transcription factor function[47,48]; thus, we investigated the potential for the HET-DhMRs to alter DNA-binding of transcription factors by testing for enrichments of known

transcription factor sequence motifs among the DhMRs (Methods). Several transcription factor sequence motifs were significantly enriched in the hippocampal HET-DhMRs. Many of the transcription factors that bind these motifs have links to neurodevelopmental behaviors and disorders, such as *Clock*, *Npas2, Pax5, Hif-1a/b, Elk1, and Usf1* (**Table 2**)[49-56]. Together, these findings suggest that 5hmC influences transcription factor binding, which may explain the observed correlated disruptions in gene expression. An initial test of this hypothesis examined transcription factor sequence motif enrichments among only the potentially functional DhMRs (*i.e.,* genes with correlated disruptions in 5hmC and expression; *N* = 72; Methods). Five transcription factor sequence motifs were significantly enriched among the potentially functional HET-DhMRs (bHLHE40, c-Myc, CLOCK, HIF-1b, and USF1). Together, these data suggest that differential hydroxymethylation levels may mediate transcription factor binding, suggesting a mechanistic role for 5hmC in the regulation of gene expression of genes associated with the observed phenotype (**Fig. 4**).

Discussion

Here we provide evidence of a role for 5hmC in a gene (partial loss of *Cntnap2*) by environment (early-life stress; GxE) interaction model that results in sex-specific alterations of neurodevelopmental behaviors. These 5hmC disruptions were found in the orthologs of a significant number of genes implicated in human developmental brain disorders, significantly overlapping with those found in both male and female *Cntnap2* homozygous mouse mutants[57]. Integration of epigenetic and gene expression data from the same mice revealed relations with genes (*e.g., Nrxn1*, *Reln*, *Grip1*, *Dlg2* and *Sdk1*) known to contribute to the behavioral deficits of neurodevelopmental disorders. Finally,

we identified that disruptions in 5hmC overlapped with motifs of transcription factors pivotal in the maintenance of neurodevelopment, supporting the hypothesis that 5hmC may mediate protein:DNA binding affinity. Together these data provide evidence of a putative molecular mechanism for 5hmC in the regulation of developmentally important genes that when altered can result in social-related disorders. A GxE model and 5hmC disruptions may represent a common etiology for developmental brain disorders in humans.

Having data from both striatal and hippocampal tissues provide a unique opportunity to examine the combined and differential contributions of these tissues to neurodevelopment-related behaviors, which both supported that the epigenome plays a role in the long lasting effects of early-life environmental exposures on brain and behavior. For example, both tissues carried GxE DhMRs in orthologous genes common to human neurodevelopmental disorders, such as nuclear receptor genes, glucocorticoid receptor-interaction protein 1 (*Grip1* also called *Ncoa2*)[46], as well as several genes associated with neurotransmission and synaptic plasticity (*Nrxn1*, *Nlgn1*, *Dlg2*, and *Negr1*)[41,58] _ENREF_28. The hippocampus-specific GxE DhMRs also included genes related to human neurodevelopmental disorders, including *Homer1*, which forms high-order complexes with *Shank1* that are necessary for the structural and functional integrity of dendritic spines[59]. In addition, the hippocampus-specific GxE DhMRs contained links to circadian function (*e.g., Clock*, *Homer1*, and *Shank2*); the prenatal stress administered here is sleep disruptive, supporting a hypothesis that circadian rhythms contribute to the pathogenesis of neurodevelopmental disorders[60]. Finally, investigation of the hippocampal-specific GxE DhMRs with correlated disruptions in gene expression

revealed links to the observed adult behavioral deficits, including specific links to the limbic system and synaptic plasticity (*e.g., Dst, Reln, Lpp, Trio*, *Notch1 Col4a1*, and *Utrn*)[61-64]. Together, these data suggest insights into the molecular pathogenesis of GxE interactions that result in neurobehavioral alterations. Importantly, while early-life-stress-alone disruptions in 5hmC (WT-DhMRs) were largely unique from the GxE DhMRs (HET-DhMRs), WT-DhMR-associated genes also included previous links to neurodevelopmental disorders (*e.g., Auts2*, *Cacna1c* and *Setd5*)[41]. Although 5hmC disruption on these genes was not correlated with altered transcript levels and were not sufficient to cause a behavioral phenotype, many of them have been implicated in disorders with social deficits[41], suggesting that early-life stressed WT mice have uncharacterized deficits and may be susceptible to additional environmental stressors that trigger later life brain-related disorders.

Mechanistically, it is intriguing that the binding motifs of several transcription factors with known roles in neuronal development were found in the GxE DhMRs. These findings included binding motifs for transcription factors best known for their roles in circadian rhythm, immune response, learning and memory, and hypoxia. For example, *Clock*, *Bmal1*, *Bhlhe40*, and *Bhlhe41* all regulate the circadian clock; in addition, they are also linked to neurodevelopmental disorders, such as major depressive disorders, bipolar disorders, and autism spectrum disorders[51,53-55]; again, consistent with the prenatal stress paradigm used here being sleep disruptive. CLOCK functions as a pioneer-like transcription factor[65], while also displaying histone acetyltransferase activity[66], corroborating CLOCK as a regulator of chromatin accessibility and contributor to the epigenetic framework in cells. While CLOCK generally acts as a transcriptional activator,

it also exhibits repressive activity that is dependent on the assembly of larger protein complexes[67-69]. Paired box protein 5 (*Pax5*) and aryl hydrocarbon receptor (*Ahr*) have defined roles in immune response and also are connected to autism spectrum disorders and major depressive disorders[50,51,70]. The stress paradigm used in this study increases placental inflammation[71]; indeed, immune activation history in the mother is associated to increased symptom severity in children with mental illness[72]. It is notable that the GxE interaction resulted in distinct differences between ELS-HET and ELS-WT enriched transcription factor sequence motifs, perhaps suggesting an interaction between *CNTNAP2* and 5hmC in transcription factor binding and function.

Exposure to stress during different gestational periods has unique effects on epigenetic programming of the developing embryo. During early gestation, early-life stress-induced epigenetic changes in neuronal precursor cells are predominantly mediated by hormonal mechanisms, which in part can be transmitted via placental pathways[73]. Stress during this time period predominately affects behavioral development in males[30,74]. However, in late gestational stages, the mechanism underlying early-life stress-induced epigenetic changes become more complex involving the activation of the excitatory and inhibitory endocrine modulatory systems, which may interfere with stress-induced synaptic reorganization[73]. Stress during this time period predominately affects behavioral development in females[30,74]; thus, there is a sex-specific effect on behavior depending on the gestational timing of the stress exposure. Consistent with these reports, late gestational stress in this present study results in long-lasting behavioral and molecular effects in females on genes known to be associated with synaptic plasticity and mental disorders. The repetitive behaviors and social deficits render the ELS-HET mice

an attractive model for other psychiatric disorders in humans. Since both insults (*i.e.,* genetic and environment) are required for the expression of an altered behavioral phenotype, this model (*i.e.,* a two-hit model) directs studies to consider environmentally sensitive molecular mechanisms contributing to stress-induced behavioral outcomes. Moreover, it supports the study of other "multi-hit" hypotheses in the origins of mental illness to identify elusive factors that could contribute to the complexity of psychiatric disorders in humans. While there is precedence for multi-hit models disrupting brain development[75-78], it will be important to examine the generalizability of these findings using other developmentally important genes in the brain that also may contribute to the vulnerability/resilience toward mental illness.

Importantly, *CNTNAP2* has been implicated in numerous neurodevelopmental disorders and heterozygous disruptions have been found in humans with intellectual disability (ID), seizures, and signs of autism spectrum disorders (*e.g.,* repetitive behaviors, and social deficits). However, heterozygous losses of *CNTNAP2* also have been inherited from healthy parents, suggesting that heterozygous disruptions of *CNTNAP2* by themselves are not sufficient to elicit cellular or organismal phenotypes. Aside from striatum and hippocampus, high levels of *CNTNAP2* expression have been found in the olfactory bulb, ventricular zone, and thalamus[28,79], warranting future examinations of 5hmC in other brain-related behaviors that are linked to these additional brain regions. *CNTNAP2* also has an organizing function to assemble neurons into neural circuits[80], suggesting future studies should examine the electrophysiology of the excitatory and inhibitory synaptic transmissions in the ELS-HET mice. Finally, early-life stress may have revealed a molecular connection between 5hmC and *Cntnap2* involving

the binding of a transcription-enhancing transcription factor *TCF4,* which can transactivate the *CNTNAP2* and *NRXN1* promoters. Heterozygous deletions of *TCF4* causes Pitt-Hopkins syndrome: a syndrome that also is linked to the loss of either *CNTNAP2* or *NRXN1* function[81-83]. Thus, *TCF4* may modulate the expression of *CNTNAP2* and *NRXN1* in the regulatory network involved in Pitt-Hopkins syndrome[84]. Interestingly, we find several DhMRs associated with *Tcf4* and *Nrxn1* following GxE interaction, suggesting that the regulatory network involved in Pitt-Hopkins syndrome may be similarly disrupted in the ELS-HET mice. Together, these findings potentially provide novel insight into the molecular connection between 5hmC, *Cntnap2*, *Tcf4*, and *Nrxn1* in neurodevelopmental disorders, such as Pitt-Hopkins syndrome.

Molecular factors influencing vulnerability and resilience to environmental stress clearly involve environmentally sensitive epigenetic mechanisms such as DNA methylation. Thus, perturbations of gene by environment interactions may underlie common pathways for many mood-related disorders in humans. The identification and characterization of a potentially modifiable substrate (*e.g.,* 5hmC) contributing to gene by environment interaction-induced sex-specific neurologic behaviors is significant and comes at a time when there is great interest to harness the diagnostic and therapeutic power of these substrates toward healthy outcomes.

Methods

Mice and genotyping

All experiments were approved by the University of Wisconsin – Madison Institutional Animal Care and Use Committee (M02529). Heterozygous male *Cntnap2*$^{+/-}$

mice were purchased from the Jackson laboratories (Bar Harbor, ME) and maintained on C57BL/6J background, as previously reported[24]. *Cntnap2*$^{+/-}$ mutants were genotyped using the following primers: Mutant Rev: CGCTTCCTCGTGCTTTACGGTAT, Common: CTGCCAGCCCAGAACTGG, WT Rev 1: GCCTGCTCTCAGAGACATCA. PCR amplification was performed with one cycle of 95°C for 5 min and 31 cycles of 95°C for 30 sec, 56°C for 30 s, 68°C for 30 s, followed by 68°C for 10 min. The mutant allele was obtained with a 350-bp and wild-type (WT) allele with a 197-bp PCR products.

Breeding scheme and prenatal stress paradigm

To minimize the stress of animal handling, all of the following were conducted by a single researcher: animal colony maintenance; breeding, prenatal stress; and behavioral tests. For breeding, a three month old *Cntnap2*$^{+/-}$ male mouse was placed together with a virgin 3 month old WT C57/Bl6J female mouse at 6 pm (one hour prior to lights off); every morning before 9 am (two hours after lights on) female mice were checked for vaginal copulation plug and separated from male. Presence of a copulation plug denoted day 1 of gestation and the pregnant female was individually housed and given a cotton nestlet. At day 12 of gestation (E12), pregnant females were randomly assigned to either a variable stress or non-stressed control group. Pregnant mice assigned to the variable stress group experienced a different daily stressor on each of the seven days during late pregnancy (E12 to E18; the timing of which was chosen because it overlaps with the onset of *Cntnap2* expression (E14)). These variable stressors included: 36 hours of constant light, 15 min of fox odor (Cat# W332518) beginning 2 hours after lights on, overnight exposure to novel objects (8 marbles), 5 minutes of restraint

stress (beginning 2 hours after lights on), overnight novel noise (white noise, nature sound-sleep machine®, Brookstone), 12 cage changes during light period, and overnight saturated bedding (700 mL, 23°C water)[31,32]. These mild stressors were previously published and were selected because they do not include pain or directly influence maternal food intake or weight gain. Importantly, both WT and HET mice were taken from the same litter, providing an ideal internal control for our findings and negates the need for cross-fostering of the prenatally stressed pups to non-stressed moms to find effects of early-life exposure to stress. Litter sizes of less than 5 and more than 8 pups were removed from the experiment. Offspring were ear-tagged at postnatal day (P) 12 and left undisturbed until weaning day (P18), at which time the mice were group housed with same sex. Female offspring were left intact, and were not cycled[32]. Finally, to minimize the effect of parent-to-offspring interaction per litter, a maximum of 3 pups/litter were randomly selected for the behavioral or molecular experiments (total of *N* = 16 litters for both experimental and validation cohorts). Importantly, the mice used for behavioral and molecular analysis were left undisturbed until behavioral testing or sacrifice at 3 months of age.

Behavioral Tests

Early-life stressed and non-early-life stressed *Cntnap2*$^{+/-}$ and WT mice (both sexes, 12-17 weeks of age) were submitted to tests of anxiety, depression, and sociability. All testing was performed during a period of 4 hours in dark period (beginning 2 hours after lights off). With one-week between tests, the same group of mice were subjected to the following sequence of behavioral tests: open field, light/dark box,

elevated plus maze, forced swim test, 3-chamber social test. For the validation of the repetitive behavior and social deficits the same group of mice were subjected to the following sequence of behavioral tests, with one-week between tests: ten-minute observational test, 3-chamber social test, and Ten-minute reciprocal social interaction test. An experimenter who was blind to the animal's group and genotype scored video recordings of each test.

Light/Dark Box Test - The light/dark box test was performed in a rectangular box divided in two compartments (light and dark). The walls of the light and dark compartments were constructed of Plexiglass (same areas 319.3 cm^2). A removable dark Plexiglass partition was used to divide the box into light and dark sides. Each animal was placed into the light side of the box, facing away from the dark side and allowed to explore both chambers of the apparatus for 10 min. The time spent in the light side was scored for each mouse.

Open Field Test - The open field apparatus consisted of a circular arena (104 cm diameter). A marker was used to inscribe a smaller circle 25 cm from the walls. Each mouse was placed in the inner circle of the apparatus and allowed to explore for 10 min. The time spent in the center and numbers of entries to inner cycle were scored for each mouse.

Elevated Plus Maze - The elevated plus maze consisted of two open and two closed arms, elevated 52 cm above the floor, with each arm projecting 50 cm from the center (a 10 X 10 cm area). Each mouse was individually placed in the center area, facing the open arm, and allowed to explore for 7 min. The measurements scored for each

mouse were as follows: time spent in closed arms, time spent in open arms, and time spent in center.

Forced Swim Test -The mice were individually placed into a glass cylinder (14-cm internal diameter, 38 cm high) filled with water (28-cm deep, 25–26 °C) for 6 min. During the last 4 min of the test, the time spent floating was scored for each mouse. Floating was defined as immobility or minimal movements necessary to maintain the head above the water.

3-Chamber Social Test - The social interaction test was performed as previously described[85]. Briefly, an experimental mouse was placed in the center third of a Plexiglass box (77.5 x 44.4 cm, divided into three equal chambers) for 10 minutes. Next the experimental mouse was allowed to explore all three interconnected chambers for 10 minutes. Finally, an empty wire cup was placed in one of the end chambers and an identical wire cup containing an unfamiliar mouse of the same sex was placed in the chamber at the opposite end; the experimental mouse was allowed to explore all three interconnected chambers for 10 minutes. Time spent in each chamber and spent sniffing each cup as well as number of entries to each chamber was measured for each mouse.

Ten-minute reciprocal social interaction test: Mice were placed in a cage (previously habituated to it) with an unfamiliar mouse matched for treatment, genotype, and sex for 10 minutes. The time mice were engaged in repetitive behaviors (grooming and digging) was measured.

Ten-minute observational test: Mice were individually placed in a cage and the time engaged in repetitive behaviors (grooming and digging) was measured.

Statistics for behavioral tests

Data is reported as mean ± S.E.M. Homogeneity of variance was assessed using the Brown-Forsythe and the normal distribution of the data was tested by the Shapiro-Wilk test. Two-way ANOVA also was used to detect differences between genotype (WT x HET) and treatment (Control x ELS) for time spent in the inner circle and number of entries into that circle (Open field test); time spent in closed arm, time spent in open arm, and time spent in center (Elevated plus maze); time spent in light side (Light/Dark box test) and time spent floating (Forced swim test). Two-way ANOVA was used to detect differences between treatment (Control x ELS) for time spent in the chamber (3-chamber social test) and interacting with the novel mouse and time spent in the chamber with the empty cup. One-way ANOVA was used to detect differences between groups for time spent engaged in repetitive behaviors (grooming and digging). Post hoc analysis was performed using a Bonferroni correction. Statistics were performed using SigmaPlot version 14.

DNA and RNA extractions

DNA methylation and gene expression was obtained from early-life stressed and non-early-life stressed three-month old female *Cntnap2*$^{+/-}$ and WT mice (N = 3 per group). Mice were sacrificed (2 hours after lights on) and whole brains were extracted without perfusion and immediately flash frozen in 2-methylbutane and dry ice. Striatum tissue was excised by micropunch (1.53 to -0.95 mm posterior to bregma), while whole hippocampal tissue was excised by morphology. Approximately 30 milligrams of tissue was homogenized with glass beads (Sigma) and DNA and RNA were extracted using

AllPrep DNA/RNA mini kit (Qiagen).

5hmC Enrichment of Genomic DNA

Chemical labeling-based 5hmC enrichment was described previously[15,24]. Briefly, a total of 10ug of striatum DNA was sonicated to 300 bp and incubated for 1 hour at 37°C in the following labeling reaction: 1.5 ul of N3-UDPG (2mM); 1.5ulβ of -GT (60uM); and 3ul of 10X β-GT buffer, in a total of 30ul. Biotin was added and the reaction was incubated at 37°C for 2 hours prior to capture on streptavidin-coupled dynabeads (Invitrogen, 65001). Enriched DNA was released from the beads during a 2 hour incubation at room temperature with 100mM DTT (Invitrogen, 15508013), which was removed using a Bio-Rad column (Bio-Rad, 732-6227). Capture efficiency was approximately 5-7% for each sample.

Library Preparation and high-throughput sequencing

5hmC-enriched libraries were generated using the NEBNext ChIP-Seq Library Prep Reagent Set for Illumina sequencing, according to the manufacturer's protocol. Briefly, the 5hmC-enriched DNA fragments were purified after the adapter ligation step using AMPure XP beads (Agencourt A63880). An Agilent 2100 BioAnalyzer was used to quantify the amplified library DNA and 20-pM of diluted libraries were used for sequencing. 50-cycle single-end (striatum) or paired-end (hippocampus) sequencing was performed by Beckman Coulter Genomics or the University of Wisconsin Biotechnology Center, respectively. Paired-end sequencing for hippocampal tissue was chosen for reasons unrelated to this project. Image processing and sequence extraction were done

using the standard Illumina Pipeline.

Analysis of 5hmC data: sequence alignment, fragment length estimation and peak identification

We mapped the reads to mouse NCBI37v1/ mm9 reference genome using Bowtie 0.12.7[86], allowing for no more than two mismatches throughout the entire read and only keeping the uniquely mapped reads. For striatal data: the Model-based Analysis of ChIP-Seq 2 (MACS2) algorithm v2.1.2 [87] was used to estimate fragment size, call peaks, and identify peak summits from aligned single-end reads using the following parameters: single-end format, effective genome size of 1.87e9, band width of 300bp, an FDR cutoff of 0.01, auto pair model process enabled, local bias computed in a surrounding 1kb window, and a maximum of one duplicate fragment to avoid PCR bias. For hippocampal data: MACS2 also was used, employing the same parameters except in the paired-end mode, and using an FDR cutoff of 0.05, which could be relaxed (compared to striatal data) due to having paired-end sequence data, which increases the signal in the data. Summits from both striatal and hippocampal data were extracted for each peak for each sample and extended +/-500bp for downstream analysis. We defined the peaks for each group as follows using the peaks from all three subjects: peaks were merged from all subjects in a group and group peaks were identified if individual peaks overlapped between a minimum of two subjects in the group.

Identification of differentially hydroxymethylated regions (DhMRs)

For each genotype, the stress and control groups were pooled and merged to form

the candidate differential regions within each genotype. The Bioconductor package edgeR was used to test whether a difference of read counts exists between the two groups in each candidate region (*P*-value<0.05)[88]. The types of DhMRs (specific to each genotype/condition) were determined by the average log fold change in a normalized read count (logFC) between the early-life stressed mice and controls.

Annotation of sequence reads, peaks, and DhMRs

All the reads were extended to their fragment lengths estimated by MACS2. We extracted genomic features and their associated gene symbols from GENCODE M1 and R Bioconductor package org.Mm.eg.db[89], and downloaded the repetitive elements from UCSC genome browser (https://genome.ucsc.edu/)[90]. We used Binomial tests to determine the significance in read density and DhMR distribution differences over the chromosomes, genomic features and repetitive elements. When conducting the binomial test for DhMRs, the background proportions were calculated from all the peaks in the two groups of animals in each comparison. We used ngsplot to draw the profile plots of the genotype/condition 5hmC enrichments and other peaks[91].

Simultaneous targeted methylation sequencing for molecular validation

Molecular validation of DhMR data was performed as previously described[92]. PCR-amplicons were sequenced on an Illumina miSeq following standard protocols. Illumina adapter sequences were removed from paired-end FASTQ files using Trim Galore v0.4.4. Alignment of trimmed reads was performed using Bismark v0.19.0, coupled with Bowtie2 v2.2.0[86,93], using a seed mismatch parameter of one base pair, a maximum insertion

length of 1000bp, with all other parameters set to default. Mapping efficiency ranged from ~78-85%. Read coverage and methylation calling was extracted using bismark_methylation_extractor with the following parameters: paired-end files used, no overlapping reads reported, with methylation output reported as a sorted bedGraph file[92]. Coverage files were imported to R environment and *methylKit* was used to determine differentially hydroxymethylated loci (DhMLs)[94]. A *P*-value threshold of 0.05 was used to determine significance.

RNA sequencing

Approximately 100 ng of total RNA was used for sequence library construction following instructions of NuGen mRNA sample prep kit (cat# 0348). In brief, total RNA was copied into first strand cDNA using reverse transcriptase and random primers. This was followed by second strand cDNA synthesis using DNA Polymerase I and RNaseH. These cDNA fragments went through an end repair process, the addition of a single 'A' base, and ligation of adapters. These products were gel purified and enriched with PCR to create the final cDNA libraries. The library constructs were run on the bioanalyzer to verify the size and concentration before sequencing on the Illumina HiSeq2500 machine where 100-cycle single-end sequencing was performed by the University of Wisconsin Biotechnology Center. In total, three libraries (the same mice and combinations that were used to generate the 5hmC data) were sequenced for each experimental condition.

Analysis of RNA-sequence data

Read alignment and calculation of transcript expression levels: we used the mm9

assembly as our reference genome and the GENCODE M1 (NCBIM37) as our gene annotation library (the Y chromosome was deleted for the alignment of female samples). We ran RSEM to calculate the expression at both gene-level and isoform-level, where RSEM automates the alignment of reads to reference transcripts using the Bowtie aligner to assign fractional counts for multi-mapping reads[95]. Bowtie was set to report all valid hits with up to 2 mismatches allowed, and to suppress all the alignments if a read has more than 200 hits. Differentially expressed (DE) genes/transcripts detection: EBSeq was used to detect the DE genes and transcripts in each of the pairwise comparisons of ELS-WT vs control WT, and ELS-HET vs control HET. We tested on both the gene level and the isoform level with a maximum number of differentially expressed groups of isoforms to be three for each gene. Genes or transcripts with more than 75% values < 10 (*i.e.,* 4 out of 6 samples in each contrast) are filtered to ensure better model fitting, and the normalization factors are the median count for each gene/transcript. We applied a FDR control to the testing and picked out genes/transcripts with FDR < 0.1.

Enrichment tests of genes and GO analysis

DhMRs were annotated to all genes within 10k. To test for the enrichment of known neurodevelopmental genes among the DhMR-associated genes, we used a chi-square test to compare the DhMR-associated genes to a list of orthologs of well-documented human developmental brain disorder genes (N = 232)[41]. Bioconductor package clusterProfiler was used to test GO Biological Process (BP) term enrichment with P-value cutoff of 0.05 and an enrichment fold-change > 1.5[96]. For the GO enrichment analysis of DhMR associated genes, the gene universe consisted of all the genes associated with

5hmC peaks in both the early-life stressed and the control mice. For the GO analysis of DE genes, the gene universe is all genes that survived the filtering of EBSeq. To test for an enrichment of neuronal related GO BP terms among the DhMR-associated GO BP terms, we used a chi-square test and a previously published list of neuronal related GO BP terms (N = 3,046)[97].

Sequence motif analysis

For motif discovery analysis, the Hypergeometric Optimization of Motif EnRichment (HOMER) suite of tools was utilized[98]. DNA sequences corresponding to DhMR coordinates were obtained from the mm9 genome and compared against background sequences, using the given size of the region. Enriched known motifs of vertebrate transcription factors (N = 428) were determined using binomial testing and an FDR cutoff of 0.05.

References

1 Sebat, J. *et al.* Strong association of de novo copy number mutations with autism. *Science* **316**, 445-449, doi:10.1126/science.1138659 (2007).

2 Glessner, J. T. *et al.* Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. *Nature* **459**, 569-573, doi:10.1038/nature07953 (2009).

3 Weiss, L. A., Arking, D. E., Daly, M. J. & Chakravarti, A. A genome-wide linkage and association scan reveals novel loci for autism. *Nature* **461**, 802-808, doi:10.1038/nature08490 (2009).

4 Tordjman, S. *et al.* Gene x Environment interactions in autism spectrum disorders: role of epigenetic mechanisms. *Frontiers in psychiatry* **5**, 53, doi:10.3389/fpsyt.2014.00053 (2014).

5 Heim, C., Newport, D. J., Mletzko, T., Miller, A. H. & Nemeroff, C. B. The link between childhood trauma and depression: insights from HPA axis studies in humans. *Psychoneuroendocrinology* **33**, 693-710, doi:10.1016/j.psyneuen.2008.03.008 (2008).

6 Ceschi, G., Billieux, J., Hearn, M., Furst, G. & Van der Linden, M. Trauma exposure interacts with impulsivity in predicting emotion regulation and depressive mood. *European journal of psychotraumatology* **5**, doi:10.3402/ejpt.v5.24104 (2014).

7 O'Connor, T. G., Heron, J., Golding, J. & Glover, V. Maternal antenatal anxiety and behavioural/emotional problems in children: a test of a programming hypothesis. *J. Child Psychol. Psychiatry* **44**, 1025-1036 (2003).

8 Van den Hove, D. L. *et al.* Prenatal stress and neonatal rat brain development. *Neuroscience* **137**, 145-155, doi:10.1016/j.neuroscience.2005.08.060 (2006).

9 Weinstock, M. Does prenatal stress impair coping and regulation of hypothalamic-pituitary-adrenal axis? *Neurosci. Biobehav. Rev.* **21**, 1-10 (1997).

10 Meaney, M. J. Maternal care, gene expression, and the transmission of individual differences in stress reactivity across generations. *Annu. Rev. Neurosci.* **24**, 1161-1192, doi:10.1146/annurev.neuro.24.1.1161 (2001).

11 McCarthy, M. M. *et al.* The epigenetics of sex differences in the brain. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **29**, 12815-12823, doi:10.1523/JNEUROSCI.3331-09.2009 (2009).

12 Szulwach, K. E. *et al.* 5-hmC-mediated epigenetic dynamics during postnatal neurodevelopment and aging. *Nature neuroscience* **14**, 1607-1616, doi:10.1038/nn.2959 (2011).

13 Szulwach, K. E. *et al.* Integrating 5-hydroxymethylcytosine into the epigenomic landscape of human embryonic stem cells. *PLoS Genet* **7**, e1002154, doi:10.1371/journal.pgen.1002154 (2011).

14 Kriaucionis, S. & Heintz, N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science (New York, N.Y.)* **324**, 929-930, doi:10.1126/science.1169786 (2009).

15 Song, C. X. *et al.* Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. *Nature biotechnology* **29**, 68-72, doi:10.1038/nbt.1732 (2011).

16 Yao, B. & Jin, P. Cytosine modifications in neurodevelopment and diseases. *Cellular and molecular life sciences : CMLS* **71**, 405-418, doi:10.1007/s00018-013-

1433-y (2014).
17 Al-Mahdawi, S., Virmouni, S. A. & Pook, M. A. The emerging role of 5-hydroxymethylcytosine in neurodegenerative diseases. *Frontiers in neuroscience* **8**, 397, doi:10.3389/fnins.2014.00397 (2014).
18 Mellen, M., Ayata, P., Dewell, S., Kriaucionis, S. & Heintz, N. MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. *Cell* **151**, 1417-1430, doi:10.1016/j.cell.2012.11.022 (2012).
19 Zhubi, A. *et al.* Increased binding of MeCP2 to the GAD1 and RELN promoters may be mediated by an enrichment of 5-hmC in autism spectrum disorder (ASD) cerebellum. *Transl. Psychiatry* **4**, e349, doi:10.1038/tp.2013.123 (2014).
20 Villar-Menéndez, I. *et al.* Increased 5-methylcytosine and decreased 5-hydroxymethylcytosine levels are associated with reduced striatal A2AR levels in Huntington's disease. *Neuromolecular medicine* **15**, 295-309, doi:10.1007/s12017-013-8219-0 (2013).
21 Wang, F. *et al.* Genome-wide loss of 5-hmC is a novel epigenetic feature of Huntington's disease. *Human molecular genetics* **22**, 3641-3653, doi:10.1093/hmg/ddt214 (2013).
22 Chouliaras, L. *et al.* Consistent decrease in global DNA methylation and hydroxymethylation in the hippocampus of Alzheimer's disease patients. *Neurobiology of aging* **34**, 2091-2099, doi:10.1016/j.neurobiolaging.2013.02.021 (2013).
23 Wang, T. *et al.* Genome-wide DNA hydroxymethylation changes are associated with neurodevelopmental genes in the developing human cerebellum. *Human molecular genetics* **21**, 5500-5510, doi:10.1093/hmg/dds394 (2012).
24 Papale, L. A. *et al.* Genome-wide disruption of 5-hydroxymethylcytosine in a mouse model of autism. *Hum Mol Genet* **24**, 7121-7131, doi:10.1093/hmg/ddv411 (2015).
25 Poliak, S. *et al.* Caspr2, a new member of the neurexin superfamily, is localized at the juxtaparanodes of myelinated axons and associates with K+ channels. *Neuron* **24**, 1037-1047, doi:10.1016/s0896-6273(00)81049-1 (1999).
26 Poliak, S. *et al.* Juxtaparanodal clustering of Shaker-like K+ channels in myelinated axons depends on Caspr2 and TAG-1. *The Journal of cell biology* **162**, 1149-1160, doi:10.1083/jcb.200305018 (2003).
27 Poot, M. Connecting the CNTNAP2 Networks with Neurodevelopmental Disorders. *Molecular syndromology* **6**, 7-22, doi:10.1159/000371594 (2015).
28 Penagarikano, O. *et al.* Absence of CNTNAP2 leads to epilepsy, neuronal migration abnormalities, and core autism-related deficits. *Cell* **147**, 235-246, doi:10.1016/j.cell.2011.08.040 (2011).
29 Toma, C. *et al.* Comprehensive cross-disorder analyses of CNTNAP2 suggest it is unlikely to be a primary risk gene for psychiatric disorders. *PLoS Genet* **14**, e1007535, doi:10.1371/journal.pgen.1007535 (2018).
30 Mueller, B. R. & Bale, T. L. Sex-specific programming of offspring emotionality after stress early in pregnancy. *J. Neurosci.* **28**, 9055-9065, doi:10.1523/jneurosci.1424-08.2008 (2008).
31 Mueller, B. R. & Bale, T. L. Impact of prenatal stress on long term body weight is dependent on timing and maternal sensitivity. *Physiology & behavior* **88**, 605-614,

doi:10.1016/j.physbeh.2006.05.019 (2006).
32 Mueller, B. R. & Bale, T. L. Early prenatal stress impact on coping strategies and learning performance is sex dependent. *Physiol. Behav.* **91**, 55-65, doi:10.1016/j.physbeh.2007.01.017 (2007).
33 Li, S. *et al.* Genome-wide alterations in hippocampal 5-hydroxymethylcytosine links plasticity genes to acute stress. *Neurobiology of disease* **86**, 99-108, doi:10.1016/j.nbd.2015.11.010 (2016).
34 Papale, L. A. *et al.* Sex-specific hippocampal 5-hydroxymethylcytosine is disrupted in response to acute stress. *Neurobiol. Dis.* **96**, 54-66, doi:10.1016/j.nbd.2016.08.014 (2016).
35 Penagarikano, O. *et al.* Exogenous and evoked oxytocin restores social behavior in the Cntnap2 mouse model of autism. *Science translational medicine* **7**, 271ra278, doi:10.1126/scitranslmed.3010257 (2015).
36 Coufal, N. G. *et al.* L1 retrotransposition in human neural progenitor cells. *Nature* **460**, 1127-1131, doi:10.1038/nature08248 (2009).
37 Secco, D. *et al.* Stress induced gene expression drives transient DNA methylation changes at adjacent repetitive elements. *eLife* **4**, doi:10.7554/eLife.09343 (2015).
38 Rusiecki, J. A. *et al.* DNA methylation in repetitive elements and post-traumatic stress disorder: a case-control study of US military service members. *Epigenomics* **4**, 29-40, doi:10.2217/epi.11.116 (2012).
39 Wagner, K. V. *et al.* Homer1/mGluR5 activity moderates vulnerability to chronic social stress. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology* **40**, 1222-1233, doi:10.1038/npp.2014.308 (2015).
40 Bisaz, R., Conboy, L. & Sandi, C. Learning under stress: a role for the neural cell adhesion molecule NCAM. *Neurobiology of learning and memory* **91**, 333-342, doi:10.1016/j.nlm.2008.11.003 (2009).
41 Gonzalez-Mantilla, A. J., Moreno-De-Luca, A., Ledbetter, D. H. & Martin, C. L. A Cross-Disorder Method to Identify Novel Candidate Genes for Developmental Brain Disorders. *JAMA psychiatry* **73**, 275-283, doi:10.1001/jamapsychiatry.2015.2692 (2016).
42 Bassi, C. *et al.* Nuclear PTEN controls DNA repair and sensitivity to genotoxic stress. *Science* **341**, 395-399, doi:10.1126/science.1236188 (2013).
43 Unterberger, A., Andrews, S. D., Weaver, I. C. & Szyf, M. DNA methyltransferase 1 knockdown activates a replication stress checkpoint. *Mol Cell Biol* **26**, 7575-7586, doi:10.1128/MCB.01887-05 (2006).
44 Zannas, A. S., Wiechmann, T., Gassen, N. C. & Binder, E. B. Gene-Stress-Epigenetic Regulation of FKBP5: Clinical and Translational Implications. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology* **41**, 261-274, doi:10.1038/npp.2015.235 (2016).
45 Jack, A. *et al.* A neurogenetic analysis of female autism. *Brain : a journal of neurology*, doi:10.1093/brain/awab064 (2021).
46 Barko, K., Paden, W., Cahill, K. M., Seney, M. L. & Logan, R. W. Sex-Specific Effects of Stress on Mood-Related Gene Expression. *Mol Neuropsychiatry* **5**, 162-175, doi:10.1159/000499105 (2019).
47 Bienvenu, T. & Chelly, J. Molecular genetics of Rett syndrome: when DNA

methylation goes unrecognized. *Nat Rev Genet* **7**, 415-426, doi:10.1038/nrg1878 (2006).

48 Gharani, N., Benayed, R., Mancuso, V., Brzustowicz, L. M. & Millonig, J. H. Association of the homeobox transcription factor, ENGRAILED 2, 3, with autism spectrum disorder. *Mol Psychiatry* **9**, 474-484, doi:10.1038/sj.mp.4001498 (2004).

49 Schmidt-Kastner, R., van Os, J., H, W. M. S. & Schmitz, C. Gene regulation by hypoxia and the neurodevelopmental origin of schizophrenia. *Schizophr Res* **84**, 253-271, doi:10.1016/j.schres.2006.02.022 (2006).

50 Wong, M. L., Dong, C., Maestre-Mesa, J. & Licinio, J. Polymorphisms in inflammation-related genes are associated with susceptibility to major depression and antidepressant response. *Mol Psychiatry* **13**, 800-812, doi:10.1038/mp.2008.59 (2008).

51 O'Roak, B. J. *et al.* Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature* **485**, 246-250, doi:10.1038/nature10989 (2012).

52 Apazoglou, K. *et al.* Antidepressive effects of targeting ELK-1 signal transduction. *Nature medicine* **24**, 591-597, doi:10.1038/s41591-018-0011-0 (2018).

53 Nicholas, B. *et al.* Association of Per1 and Npas2 with autistic disorder: support for the clock genes/social timing hypothesis. *Mol Psychiatry* **12**, 581-592, doi:10.1038/sj.mp.4001953 (2007).

54 Hamilton, K. A. *et al.* Mice lacking the transcriptional regulator Bhlhe40 have enhanced neuronal excitability and impaired synaptic plasticity in the hippocampus. *PLoS One* **13**, e0196223, doi:10.1371/journal.pone.0196223 (2018).

55 Shi, S. Q. *et al.* Molecular analyses of circadian gene variants reveal sex-dependent links between depression and clocks. *Translational psychiatry* **6**, e748, doi:10.1038/tp.2016.9 (2016).

56 Aguiniga, L. M. *et al.* Acyloxyacyl hydrolase modulates depressive-like behaviors through aryl hydrocarbon receptor. *Am J Physiol Regul Integr Comp Physiol* **317**, R289-R300, doi:10.1152/ajpregu.00029.2019 (2019).

57 Papale, L. A. *et al.* Genome-wide disruption of 5-hydroxymethylcytosine in a mouse model of autism. *Hum Mol Genet*, doi:10.1093/hmg/ddv411 (2015).

58 Szczurkowska, J. *et al.* NEGR1 and FGFR2 cooperatively regulate cortical development and core behaviours related to autism disorders in mice. *Brain : a journal of neurology* **141**, 2772-2794, doi:10.1093/brain/awy190 (2018).

59 Orlowski, D., Elfving, B., Muller, H. K., Wegener, G. & Bjarkam, C. R. Wistar rats subjected to chronic restraint stress display increased hippocampal spine density paralleled by increased expression levels of synaptic scaffolding proteins. *Stress* **15**, 514-523, doi:10.3109/10253890.2011.643516 (2012).

60 Geoffray, M. M., Nicolas, A., Speranza, M. & Georgieff, N. Are circadian rhythms new pathways to understand Autism Spectrum Disorder? *J Physiol Paris* **110**, 434-438, doi:10.1016/j.jphysparis.2017.06.002 (2016).

61 Bhandare, R. *et al.* Glucocorticoid receptor interacting protein-1 restores glucocorticoid responsiveness in steroid-resistant airway structural cells. *Am J Respir Cell Mol Biol* **42**, 9-15, doi:10.1165/rcmb.2009-0239RC (2010).

62 Sibbe, M., Forster, E., Basak, O., Taylor, V. & Frotscher, M. Reelin and Notch1

cooperate in the development of the dentate gyrus. *J Neurosci* **29**, 8578-8585, doi:10.1523/JNEUROSCI.0958-09.2009 (2009).

63 Wong, T. H. *et al.* PRKAR1B mutation associated with a new neurodegenerative disorder with unique pathology. *Brain : a journal of neurology* **137**, 1361-1373, doi:10.1093/brain/awu067 (2014).

64 Wang, H. Y. *et al.* RBFOX3/NeuN is Required for Hippocampal Circuit Balance and Function. *Scientific reports* **5**, 17383, doi:10.1038/srep17383 (2015).

65 Menet, J. S., Pescatore, S. & Rosbash, M. CLOCK:BMAL1 is a pioneer-like transcription factor. *Genes & development* **28**, 8-13, doi:10.1101/gad.228536.113 (2014).

66 Doi, M., Hirayama, J. & Sassone-Corsi, P. Circadian regulator CLOCK is a histone acetyltransferase. *Cell* **125**, 497-508, doi:10.1016/j.cell.2006.03.033 (2006).

67 Yang, Y., Li, N., Qiu, J., Ge, H. & Qin, X. Identification of the Repressive Domain of the Negative Circadian Clock Component CHRONO. *International journal of molecular sciences* **21**, doi:10.3390/ijms21072469 (2020).

68 Michael, A. K. *et al.* Formation of a repressive complex in the mammalian circadian clock is mediated by the secondary pocket of CRY1. *Proceedings of the National Academy of Sciences of the United States of America* **114**, 1560-1565, doi:10.1073/pnas.1615310114 (2017).

69 Cao, X., Yang, Y., Selby, C. P., Liu, Z. & Sancar, A. Molecular mechanism of the repressive phase of the mammalian circadian clock. *Proceedings of the National Academy of Sciences of the United States of America* **118**, doi:10.1073/pnas.2021174118 (2021).

70 Neavin, D. R., Liu, D., Ray, B. & Weinshilboum, R. M. The Role of the Aryl Hydrocarbon Receptor (AHR) in Immune and Inflammatory Diseases. *Int J Mol Sci* **19**, doi:10.3390/ijms19123851 (2018).

71 Bronson, S. L. & Bale, T. L. Prenatal stress-induced increases in placental inflammation and offspring hyperactivity are male-specific and ameliorated by maternal antiinflammatory treatment. *Endocrinology* **155**, 2635-2646, doi:10.1210/en.2014-1040 (2014).

72 Patel, S. *et al.* Social impairments in autism spectrum disorder are related to maternal immune history profile. *Mol Psychiatry* **23**, 1794-1797, doi:10.1038/mp.2017.201 (2018).

73 Bock, J., Wainstock, T., Braun, K. & Segal, M. Stress In Utero: Prenatal Programming of Brain Plasticity and Cognition. *Biol. Psychiatry* **78**, 315-326, doi:10.1016/j.biopsych.2015.02.036 (2015).

74 Li, H. *et al.* NF-kappaB regulates prenatal stress-induced cognitive impairment in offspring rats. *Behav. Neurosci.* **122**, 331-339, doi:10.1037/0735-7044.122.2.331 (2008).

75 Bell, M. R., Hart, B. G. & Gore, A. C. Two-hit exposure to polychlorinated biphenyls at gestational and juvenile life stages: 2. Sex-specific neuromolecular effects in the brain. *Mol. Cell. Endocrinol.* **420**, 125-137, doi:10.1016/j.mce.2015.11.024 (2016).

76 Schraut, K. G. *et al.* Prenatal stress-induced programming of genome-wide promoter DNA methylation in 5-HTT-deficient mice. *Translational psychiatry* **4**, e473, doi:10.1038/tp.2014.107 (2014).

77 van den Hove, D. L. *et al.* Differential effects of prenatal stress in 5-Htt deficient

mice: towards molecular mechanisms of gene x environment interactions. *PLoS One* **6**, e22715, doi:10.1371/journal.pone.0022715 (2011).

78 Schaafsma, S. M. *et al.* Sex-specific gene-environment interactions underlying ASD-like behaviors. *Proc Natl Acad Sci U S A* **114**, 1383-1388, doi:10.1073/pnas.1619312114 (2017).

79 Abrahams, B. S. *et al.* Genome-wide analyses of human perisylvian cerebral cortical patterning. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 17849-17854, doi:10.1073/pnas.0706128104 (2007).

80 Anderson, G. R. *et al.* Candidate autism gene screen identifies critical role for cell-adhesion molecule CASPR2 in dendritic arborization and spine development. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 18120-18125, doi:10.1073/pnas.1216398109 (2012).

81 Amiel, J. *et al.* Mutations in TCF4, encoding a class I basic helix-loop-helix transcription factor, are responsible for Pitt-Hopkins syndrome, a severe epileptic encephalopathy associated with autonomic dysfunction. *Am. J. Hum. Genet.* **80**, 988-993, doi:10.1086/515582 (2007).

82 Brockschmidt, A. *et al.* Severe mental retardation with breathing abnormalities (Pitt-Hopkins syndrome) is caused by haploinsufficiency of the neuronal bHLH transcription factor TCF4. *Hum. Mol. Genet.* **16**, 1488-1494, doi:10.1093/hmg/ddm099 (2007).

83 Zweier, C. *et al.* Haploinsufficiency of TCF4 causes syndromal mental retardation with intermittent hyperventilation (Pitt-Hopkins syndrome). *Am. J. Hum. Genet.* **80**, 994-1001, doi:10.1086/515583 (2007).

84 Forrest, M. *et al.* Functional analysis of TCF4 missense mutations that cause Pitt-Hopkins syndrome. *Hum. Mutat.* **33**, 1676-1686, doi:10.1002/humu.22160 (2012).

85 Silverman, J. L., Yang, M., Lord, C. & Crawley, J. N. Behavioural phenotyping assays for mouse models of autism. *Nature reviews. Neuroscience* **11**, 490-502, doi:10.1038/nrn2851 (2010).

86 Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology* **10**, R25, doi:10.1186/gb-2009-10-3-r25 (2009).

87 Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome biology* **9**, R137, doi:10.1186/gb-2008-9-9-r137 (2008).

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

89 org.Mm.eg.db: Genome wide annotation for Mouse (2015).

90 Bailey, J. A., Yavor, A. M., Massa, H. F., Trask, B. J. & Eichler, E. E. Segmental duplications: organization and impact within the current human genome project assembly. *Genome Res* **11**, 1005-1017, doi:10.1101/gr.187101 (2001).

91 Shen, L. *et al.* diffReps: detecting differential chromatin modification sites from ChIP-seq data with biological replicates. *PLoS One* **8**, e65598, doi:10.1371/journal.pone.0065598 (2013).

92 Chen, G. G. *et al.* Medium throughput bisulfite sequencing for accurate detection of 5-methylcytosine and 5-hydroxymethylcytosine. *BMC genomics* **18**, 96, doi:10.1186/s12864-017-3489-9 (2017).

93 Krueger, F. & Andrews, S. R. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**, 1571-1572, doi:10.1093/bioinformatics/btr167 (2011).
94 Akalin, A. *et al.* methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome biology* **13**, R87, doi:10.1186/gb-2012-13-10-r87 (2012).
95 Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323, doi:10.1186/1471-2105-12-323 (2011).
96 Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics : a journal of integrative biology* **16**, 284-287, doi:10.1089/omi.2011.0118 (2012).
97 Geifman, N., Monsonego, A. & Rubin, E. The Neural/Immune Gene Ontology: clipping the Gene Ontology for neurological and immunological systems. *BMC bioinformatics* **11**, 458, doi:10.1186/1471-2105-11-458 (2010).
98 Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular cell* **38**, 576-589, doi:10.1016/j.molcel.2010.05.004 (2010).

Figure Legends

Figure 1: Results from social behavioral tests in female offspring. (A-D) 3-chamber social interaction test: (A-B) time spent in the chamber with either an unfamiliar mouse inside a cup (mouse) or with an empty cup (empty). Data analysis details: (A) two way ANOVA (ANOVA) for control-WT versus ELS-WT groups: effect of time spent in either chamber: $F_{(1,42)} = 28.5$, P-value < 0.001. (B) ANOVA for control-HET versus ELS-HET groups: effect of time spent in either chamber: $F_{(1,32)} = 16.2$, P-value < 0.001 and interaction between time spent in either chamber and treatment: $F_{(1,32)} = 5.3$, P-value = 0.002. (C-D) Time spent interacting with either an unfamiliar mouse inside a cup (mouse) or with an empty cup (empty). Data analysis details: (C) ANOVA for control-WT versus ELS-WT groups: effect of time spent interacting: $F_{(1,42)} = 29$, P-value < 0.001. (D) ANOVA for control-HET versus ELS-HET groups: effect of time spent interacting in either chamber: $F_{(1,32)} = 9.9$, P-value = 0.003). All asterisks denote P-value ≤ 0.05, two-way ANOVA with Bonferroni multiple comparison test determined post hoc significance. (E) Ten-minute reciprocal social interaction test in female offspring. One-way ANOVA detected effect of groups on time spent grooming/digging during a 10-minute reciprocal social interaction test ($F_{(3,16)} = 10.9$, P-value < 0.001). Hashtag (#) denotes significance for time spent grooming/digging between ELS-HET in comparison to Control-WT, Control-HET, and ELS-WT (P-value < 0.05). N = 9-10 per treatment and genotype.

Figure 2: Distribution of hippocampal HET DhMRs. (A) A Manhattan plot shows the distribution of 5hmC sequenced data across the genome (x-axis; alternating black/grey for alternating chromosomes) and the level of significance (-log10(P-value)) along the y-

axis. Dots above the top blue line represent ELS-HET-specific (hyper)-DhMRs, while dots below the bottom blue line represent Control-HET-specific (hypo)-DhMRs (*P*-value < 0.05). (B) A pie chart displays the proportion of DhMRs across standard genomic structures. (C-D) Bar plot showing the top ten significantly over-represented ontological terms (y-axis) based on gene ontological analysis of ELS-HET-specific DhMR-associated genes (C) and Control-HET-specific DhMR-associated genes (D). The number of DhMR-associated genes linked to the ontological terms are displayed on the x-axis. Bar color is based on *P*-value. (E) Venn diagram showing the overlap of HET-DhMR-associated genes and WT-DhMR-associated genes.

Figure 3: Characterization of differentially expressed genes (DEGs) between ELS-HET and Control-HET. (A) A modified volcano plot depicts the log2(posterior fold change; x-axis) versus the posterior probability of differential expression (y-axis). Open circles (red/black) represent each gene examined and differentially expressed genes (red) are shown above the significance line (blue, FDR *P*-value < 0.1). (B-C) A bar plot showing the top significantly over-represented ontological terms based on gene ontological analysis of ELS-HET-specific (B) differentially expressed genes and Control-HET-specific (C) differentially expressed genes. The number of DEGs linked to the ontological terms are displayed on the x-axis. Bar color is based on *P*-value. (D) A Venn diagram depicts the overlap of DhMR-associated genes (green circle) and differentially expressed genes (orange circle) from the ELS-HET versus Control-HET comparison. The asterisk (*) denotes a significant overlap (hypergeometric enrichment *P*-value < 0.05).

Figure 4: A working model of a putative functional role for 5hmC in developmental brain disorder-like behaviors. WT mice (Cntnap2$^{+/+}$ (left panel) and ELS-HET (*Cntnap2*$^{+/-}$ (right panel)) 5hmC levels (orange lollipops) are shown. Following the gene by environment interaction, a disruption of 5hmC results. The baseline levels of 5hmC in WT mice allow for transcription factor (blue semicircle) binding to occur uninterrupted, allowing for proper rates of transcription (black curved lines) to occur, resulting in phenotypically unaffected behaviors. The disruptions of 5hmC in the ELS-HET mice disrupt this transcription factor binding, resulting in dysregulated gene expression, contributing to the etiology of developmental brain disorder-like behaviors.

Supplemental Figure Legends

Supplemental Figure 1: Results from 3-chamber social interaction test in females from an independent cohort. (A-C) test time spent in the chamber with either an unfamiliar mouse inside a cup or with an empty cup. (A) *Cntnap2* homozygote: Data analysis details: Paired T-test did not detect any significant differences. (B) two way ANOVA for control-WT versus ELS-WT groups: effect of time spent in chamber: $F_{(1,16)} = 58$, *P*-value < 0.001. (C) two way ANOVA for control-HET versus ELS-HET groups: interaction between treatment and time spent in chamber: $F_{(1,24)} = 25$, *P*-value < 0.001. There was a significant three-way interaction between effect of treatment, time spent in chamber, and genotype $F_{(1,40)} = 12.5$, *P*-value = 0.001. (D-F) Time spent interacting with either an unfamiliar mouse inside a cup or with an inanimate object. (D) *Cntnap2* homozygote: Data analysis details: Paired T-test did not detect any significant differences. (E) two way ANOVA for control-WT versus ELS-WT groups: effect of time spent interacting: $F_{(1,16)} = 12.3$, *P*-value = 0.003. (F) two way ANOVA for control-HET versus ELS-HET groups: effect of time spent interacting $F_{(1,24)} = 4.7$, *P*-value = 0.004 and interaction between treatment and time spent interacting $F_{(1,24)} = 6.5$, *P*-value = 0.01. The asterisk (*) denotes significance within group (*P*-value < 0.05). Hashtag (#) denotes significance for time spent in empty chamber between ELS-HET in comparison to Control-WT, Control-HET, and ELS-WT (*P*-value < 0.05).

Supplemental Figure 2: Results from social behavioral tests in male offspring. (A-B) 3-chamber social interaction test, time spent in the chamber with either an unfamiliar mouse inside a cup or with an empty cup. (A) two way ANOVA for control-WT versus ELS-WT

groups: effect of time spent in chamber: $F_{(1,28)}$ = 22.4, *P*-value < 0.001. (B) two way ANOVA for control-HET versus ELS-HET groups: effect of time spent in chamber: $F_{(1,24)}$ = 13.1, *P*-value = 0.001. (C-D) Time spent interacting with either an unfamiliar mouse inside a cup or with an empty cup. (C) two way ANOVA for control-WT versus ELS-WT groups: effect of time spent in chamber: $F_{(1,28)}$ = 18.5, *P*-value < 0.001. (D) two way ANOVA for control-HET versus ELS-HET groups: effect of time spent in chamber: $F_{(1,24)}$ = 10.7, *P*-value = 0.003. The asterisk (*) denotes significance of *P*-value < 0.05.

Supplemental Figure 3: Distribution and characterization of aligned reads derived from hippocampal tissue from the four experimental groups. (A) A bar plot depicts the percent of mapped reads (y-axis) relative to chromosomes (x-axis) for Control-WT (black bars), ELS-WT (light grey bars), Control-HET (white bars), and ELS-HET (dark grey bars) mice. (B) A bar plot shows the proportion of mapped reads (y-axis) as they relate to standard genomic structures (x-axis) for Control-WT (black bars), ELS-WT (light grey bars), Control-HET (white bars), and ELS-HET (dark grey bars) mice. (C) A bar plot shows the proportion of mapped reads (y-axis) as they relate to repetitive elements across the genome (x-axis) for Control-WT (black bars), ELS-WT (light grey bars), Control-HET (white bars), and ELS-HET (dark grey bars) mice.

Supplemental Figure 4: Distribution of hippocampal WT DhMRs. (A) A Manhattan plot shows the distribution of 5hmC sequenced data across the genome (x-axis; alternating black/grey for alternating chromosomes) and the level of significance (y-axis; -log10(*P*-value)). Dots above the top blue line represent ELS-WT-specific-(hyper) DhMRs, while

dots below the bottom blue line represent Control-WT-specific-(hypo) DhMRs (*P*-value < 0.05). (B) A pie chart displays the proportion of WT-DhMRs across standard genomic structures. (C-D) Bar plots showing the top ten significantly over-represented ontological terms (y-axis) based on gene ontological analysis of ELS-WT-specific DhMR-associated genes (C) and Control-WT-specific DhMR-associated genes (D). The number of DhMR-associated genes linked to the ontological terms are displayed on the x-axis. Bar color is based on *P*-value.

Supplemental Figure 5: Distribution of hippocampal DhMRs. (A and B) Bar plots showing the proportion of DhMRs (y-axis) from the ELS-HET vs Control-HET (A) and ELS-WT vs Control-WT (B) comparisons as they relate to standard genomic structures (x-axis). Depicted are all regions across the genome tested for differential 5hmC abundance (*i.e.,* background regions; black bars), all DhMRs (light grey bars), ELS-specific (hyper) DhMRs (white bars), and Control-specific (hypo) DhMRs (dark grey bars). An asterisk (*) represents a significant increase or decrease of 5hmC abundance relative to the background regions, as determined by binomial testing (*P*-value < 0.05). (C and D) Bar plots showing the proportion of DhMRs (y-axis) from the ELS-HET vs Control-HET (C) and ELS-WT vs Control-WT (D) comparisons as they relate to repetitive elements (x-axis). Depicted are all regions across the genome tested for differential 5hmC abundance (*i.e.,* background regions; black bars), all DhMRs (light grey bars), ELS-specific (hyper) DhMRs (white bars), and Control-specific (hypo) DhMRs (dark grey bars). An asterisk (*) represents a significant increase or decrease of 5hmC disruptions relative to the background regions, as determined by binomial testing (*P*-value < 0.05).

Supplemental Figure 6: Distribution and characterization of aligned reads derived from striatal tissue from the four experimental groups. (A) A bar plot depicts the percent of mapped reads (y-axis) relative to chromosomes (x-axis) for ELS-HET (black bars), Control-HET (white bars), ELS-WT (striped bars), and Control-WT (grey bars) mice. (B) A bar plot shows the percent of mapped reads (y-axis) as they relate to standard genomic structures (x-axis) for ELS-HET (black bars), Control-HET (white bars), ELS-WT (striped bars), and Control-WT (grey bars) mice. (C) A bar plot shows the proportion of mapped reads (y-axis) as they relate to repetitive elements across the genome (x-axis) for for ELS-HET (black bars), Control-HET (white bars), ELS-WT (striped bars), and Control-WT (dark grey bars) mice. An asterisk (*) represents a significant increase or decrease of 5hmC disruptions relative to the background regions, as determined by binomial testing (*P*-value < 0.05).

Supplemental Figure 7: Molecular validation of hippocampal DhMR data. (A) A Venn diagram displays the overlap of DhMR-associated genes found in hippocampal and striatal tissues. An asterisk (*) represents a significant enrichment of DhMR-associated genes between the two tissues, as determined by a hypergeometric enrichment test (*P*-value < 0.05). (B) Molecular validation of a DhMR (*Epha5*) and a non-DhMR (*Pak7*). The genomic position (x-axis) and percent hydroxymethylation (y-axis) as determined by simultaneous targeted methylation sequencing of hippocampal tissue from ELS-HET (blue) and Control-HET (red) mice. An asterisk (*) represents a significant increase or decrease of 5hmC abundance of ELS-HET mice compared to Control-HET mice, as

determined by Fisher's exact testing (*P*-value < 0.05). (C) A Venn diagram shows the overlap of DhMR-associated genes found in hippocampal and striatal tissues from ELS *Cntnap2*$^{+/-}$ (ELS-HET) and *Cntnap2*$^{-/-}$ (HOMO) mice. An asterisk (*) represents a significant enrichment of DhMR-associated genes between the two tissues, as determined by a hypergeometric enrichment test (*P*-value < 0.05).

Supplemental Figure 8: Characterization of differentially expressed genes (DEGs) between ELS-WT and Control-WT. (A) A modified volcano plot depicts the log2(posterior fold change; x-axis) versus the posterior probability of differential expression (y-axis). Open circles (red/black) represent each gene examined and differentially expressed genes (red) are shown above the significance line (blue, FDR *P*-value < 0.1).

Figure 1

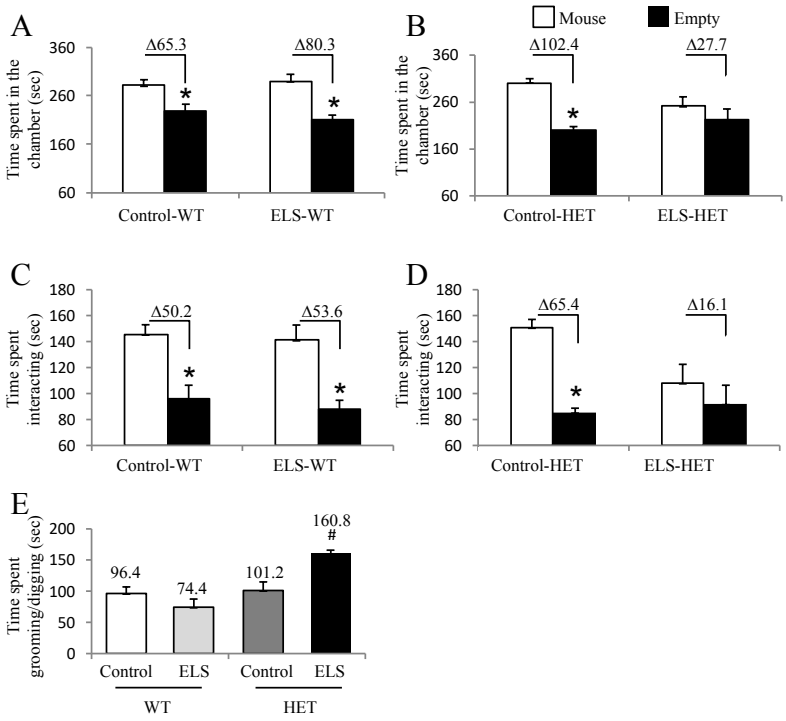

A
Δ65.3
Δ80.3
Time spent in the chamber (sec)
360
260
160
60
Control-WT
ELS-WT
B
Mouse
Empty
Δ102.4
Δ27.7
Time spent in the chamber (sec)
Control-HET
ELS-HET
C
Δ50.2
Δ53.6
Time spent interacting (sec)
180
160
140
120
100
80
60
Control-WT
ELS-WT
D
Δ65.4
Δ16.1
Time spent interacting (sec)
Control-HET
ELS-HET
E
Time spent grooming/digging (sec)
200
150
100
50
0
96.4
74.4
101.2
160.8
#
Control
ELS
Control
ELS
WT
HET

Figure 2

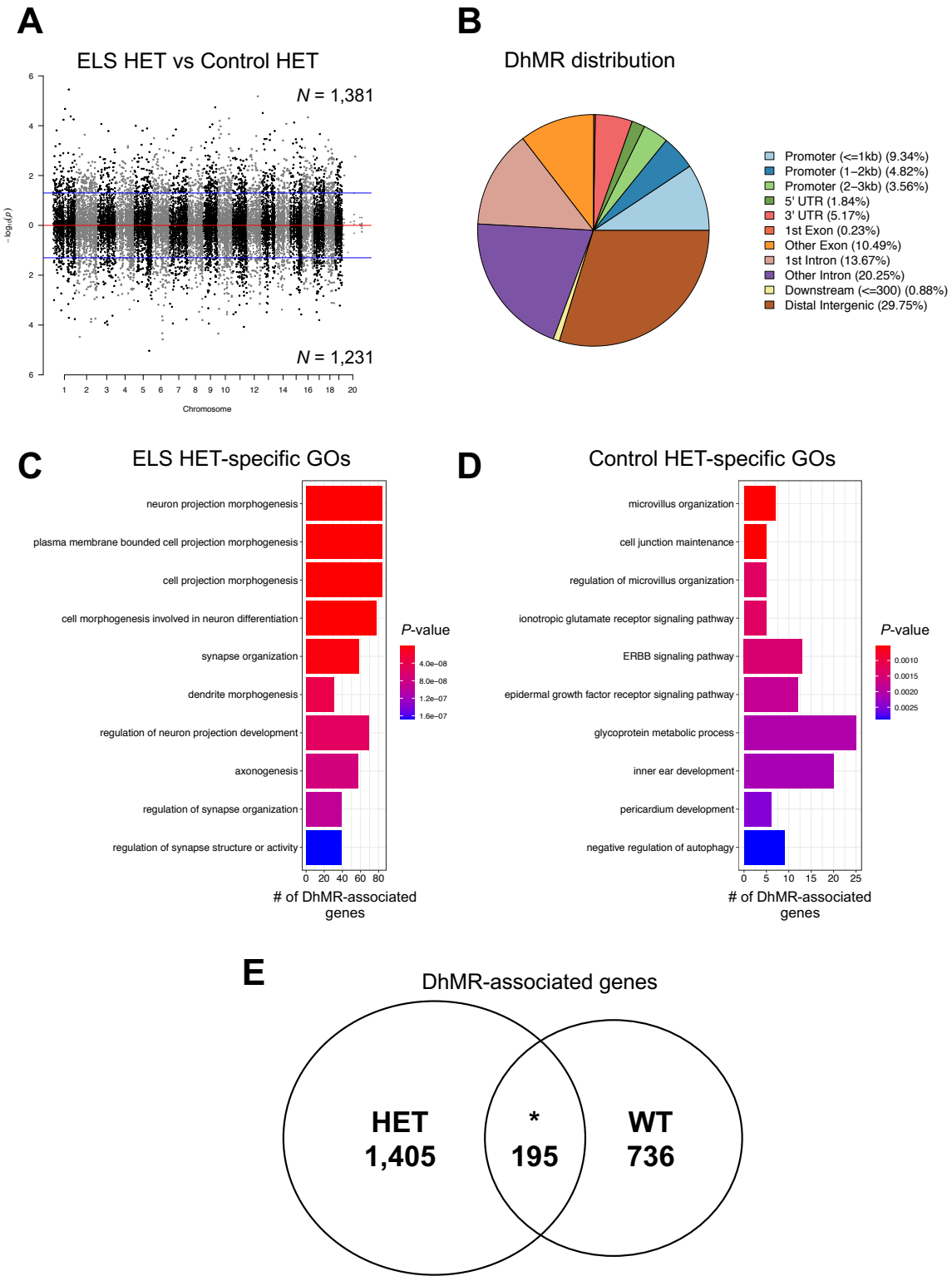

A
ELS HET vs Control HET
N = 1,381
N = 1,231
Chromosome
B
DhMR distribution
Promoter (<=1kb) (9.34%)
Promoter (1–2kb) (4.82%)
Promoter (2–3kb) (3.56%)
5' UTR (1.84%)
3' UTR (5.17%)
1st Exon (0.23%)
Other Exon (10.49%)
1st Intron (13.67%)
Other Intron (20.25%)
Downstream (<=300) (0.88%)
Distal Intergenic (29.75%)
C
ELS HET-specific GOs
neuron projection morphogenesis
plasma membrane bounded cell projection morphogenesis
cell projection morphogenesis
cell morphogenesis involved in neuron differentiation
synapse organization
dendrite morphogenesis
regulation of neuron projection development
axonogenesis
regulation of synapse organization
regulation of synapse structure or activity
P-value
# of DhMR-associated genes
D
Control HET-specific GOs
microvillus organization
cell junction maintenance
regulation of microvillus organization
ionotropic glutamate receptor signaling pathway
ERBB signaling pathway
epidermal growth factor receptor signaling pathway
glycoprotein metabolic process
inner ear development
pericardium development
negative regulation of autophagy
P-value
# of DhMR-associated genes
E
DhMR-associated genes
HET
1,405
*
195
WT
736

Figure 3

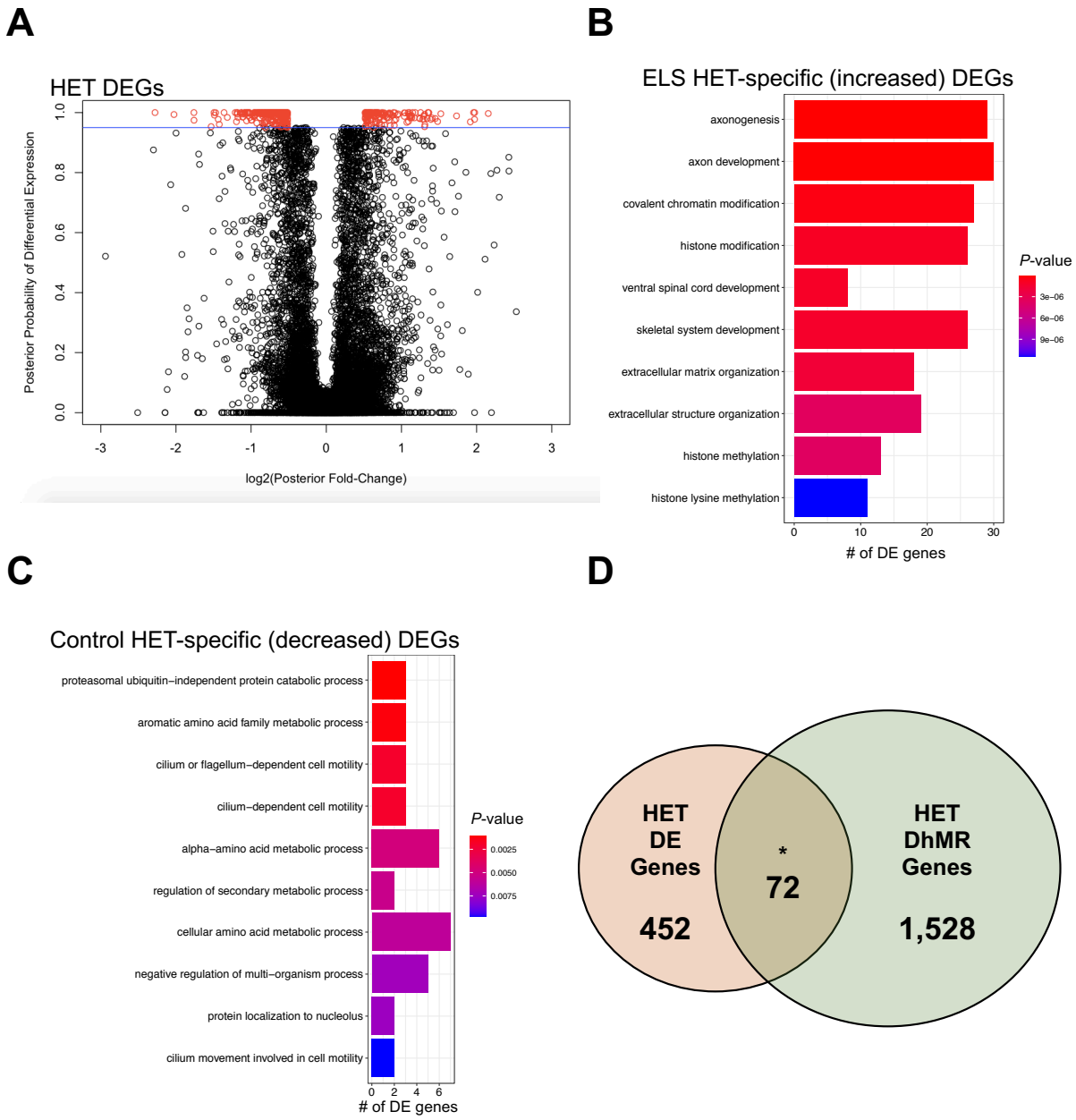

A
HET DEGs
Posterior Probability of Differential Expression
log2(Posterior Fold-Change)
B
ELS HET-specific (increased) DEGs
axonogenesis
axon development
covalent chromatin modification
histone modification
ventral spinal cord development
skeletal system development
extracellular matrix organization
extracellular structure organization
histone methylation
histone lysine methylation
P-value
# of DE genes
C
Control HET-specific (decreased) DEGs
proteasomal ubiquitin−independent protein catabolic process
aromatic amino acid family metabolic process
cilium or flagellum−dependent cell motility
cilium−dependent cell motility
alpha−amino acid metabolic process
regulation of secondary metabolic process
cellular amino acid metabolic process
negative regulation of multi−organism process
protein localization to nucleolus
cilium movement involved in cell motility
P-value
# of DE genes
D
HET DE Genes
452
*
72
HET DhMR Genes
1,528

Figure 4

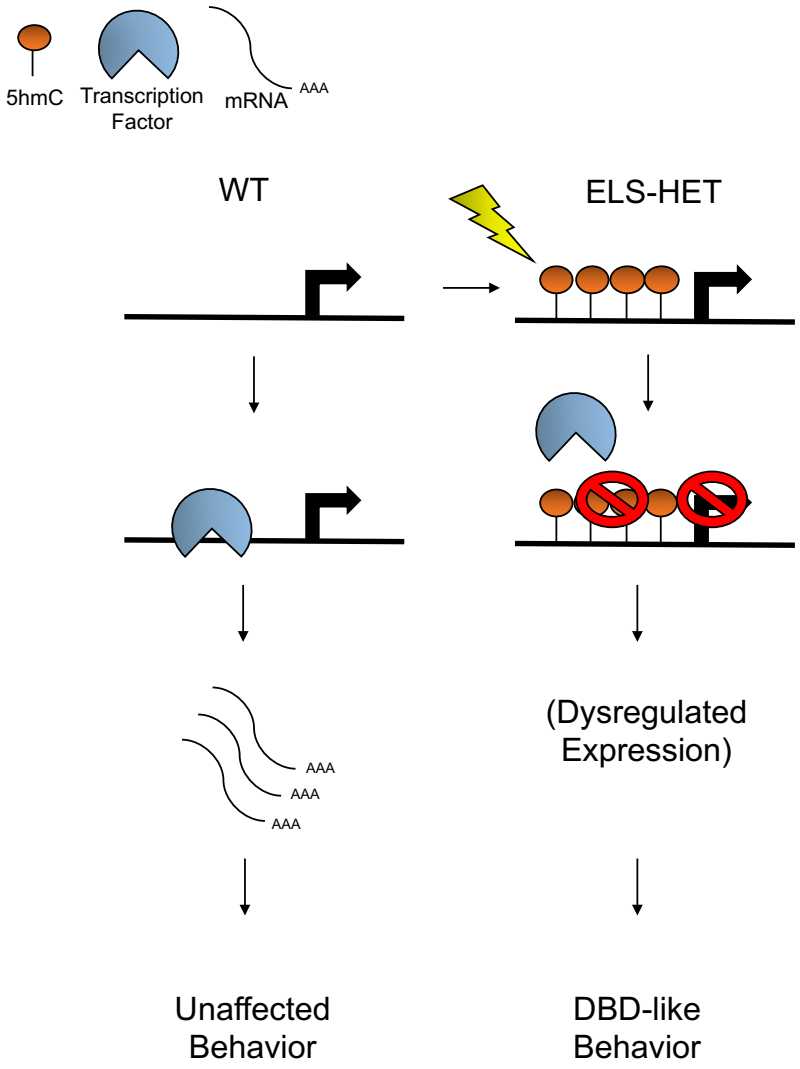

5hmC
Transcription Factor
mRNA
AAA
WT
ELS-HET
AAA
AAA
AAA
(Dysregulated Expression)
Unaffected Behavior
DBD-like Behavior

Table 1: Summary of behavioral tests in female offspring

	Control		*Early-Life Stressed*	
	WT	*Cntnap2*$^{+/-}$	WT	*Cntnap2*$^{+/-}$
Open Field				
Time spent in the center (s)	44 ± 10	35 ± 3	42 ± 4	45 ± 3
Number of entries to the center	19 ± 3	17 ± 1	19 ± 1	19 ± 0
Elevated Plus Maze				
Time spent in closed arm (s)	222 ± 9	244 ± 8	240 ± 12	233 ± 9
Time spent in open arm (s)	204 ± 11	166 ± 8	182 ± 13	183 ± 9
Light/Dark Box				
Time spent in light side (s)	352 ± 18	357 ± 19	392 ± 11	376 ± 24
Forced swim test				
Time spent floating	194 ± 7	152 ± 13	172 ± 16	192 ± 11
Ten-minute grooming test				
Time spent grooming	64.5±8.8	52.2±4.8	59.3±7.8	44.2±5

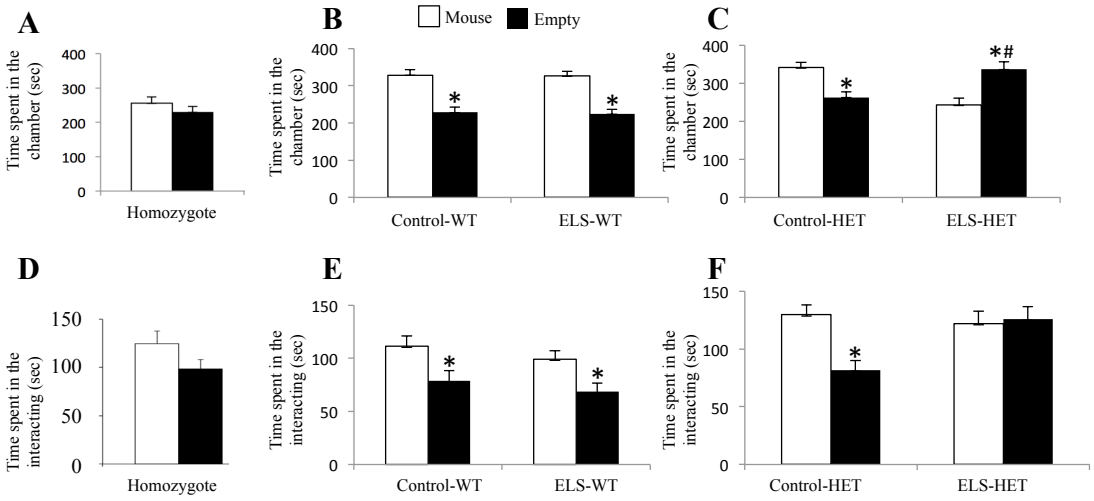
Comparison of different behavioral tests between early-life stressed (ELS) and non-stressed (Control) *Cntnap2*$^{+/-}$ and WT mice. The value are presented as mean ± S.E.M. *p≤ 0.05. N = 6-8. Data analysis: 1) Open Field (time spent in the center: effect of ELS $F_{(1,31)} = 0.59$, $P = 0.4$; effect of genotype $F_{(1,31)} = 0.9$, $P = 0.3$; number of entries to the center: effect of ELS $F_{(1,31)} = 0.27$, $P = 0.6$; effect of genotype $F_{(1,31)} = 0.02$, $P = 0.8$), 2) Elevated Plus Maze (time spent in closed arm: effect of ELS $F_{(1,34)} = 0.08$, $P = 0.7$; effect of genotype $F_{(1,34)} = 0.4$, $P = 0.5$; time spent in open arm: effect of ELS $F_{(1,34)} = 0.04$, $P = 0.8$; effect of genotype $F_{(1,34)} = 2.8$, $P = 0.1$), 3) Light/Dark Box (time spent in light side: effect of ELS $F_{(1,19)} = 2.4$, $P = 0.13$; effect of genotype $F_{(1,19)} = 0.1$, $P = 0.7$) and 4) Forced Swim Test (time spent floating: effect of ELS $F_{(1,26)} = 0.31$, $P = 0.5$; effect of genotype $F_{(1,26)} = 0.49$, $P = 0.4$). 5) Ten-minutes grooming test: effect of ELS $F_{(1.34)} = 0.8$, $P = 0.3$; effect of genotype $F_{(1.34)} = 3.6$, $P = 0.06$.

Table 2: A subset of the transcription factors with binding motifs associated with DhMRs in hippocampus ELS-HET x Control-HET

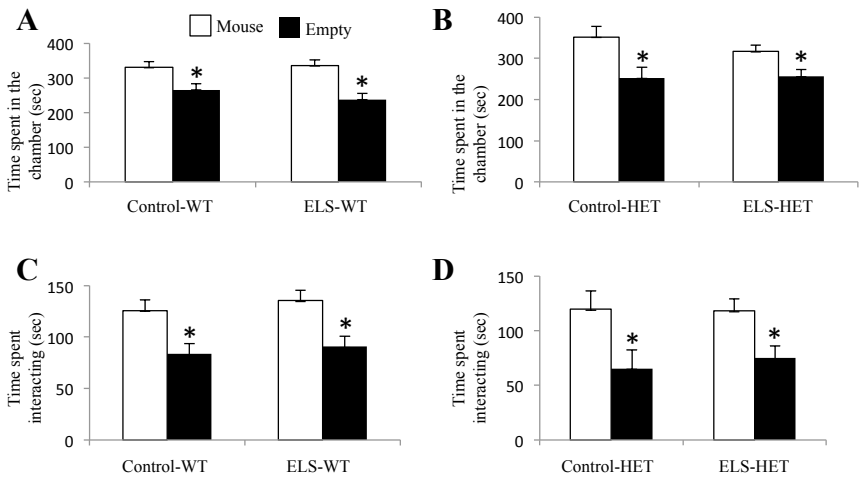
Gene	Name	Role	FDR *P*-value*	DhMR type
bHLHE40/41	Basic helix-loop-helix family	Regulation of circadian rhythms	0.00001	ELS-HET, Ctrl-HET
BMAL1	Brain and Muscle ARNT-Like 1	Regulates biological functions under circadian control	0.00001	Ctrl-HET
c-Myc	Cellular Myelocytomatosis	Regulates cell division and proto-oncogenes	0.00001	ELS-HET, Ctrl-HET
HIF1a/1b/2a	Hypoxia-inducible factors	Cellular oxygen sensing	0.00001	ELS-HET, Ctrl-HET
Max	Myc-associated factor X	Cell proliferation and apoptosis	0.00001	Ctrl-HET
MNT	Max-binding protein MNT	Gene-specific transcriptional activation or repression	0.00001	Ctrl-HET
NF1	Nuclear factor I	Neuronal regulation	0.00001	ELS-Het, Ctrl-HET
NPAS/2	Neuronal PAS domain protein	Paralogous to Clock, maintenance of circadian rhythms	0.00001	Ctrl-HET
USF1/2	Upstream stimulator factor	Cell metabolism	0.00001	ELS-HET, Ctrl-HET
Clock	Circadian locomotor output cycles kaput	Maintenance of circadian rhythms	0.00001	ELS-HET, Ctrl-HET

Underlined transcription factors are also found in hippocampal ELS-WT x Control WT
*FDR *P*-value for the enrichment of each transcription factor binding motif in hippocampal tissue

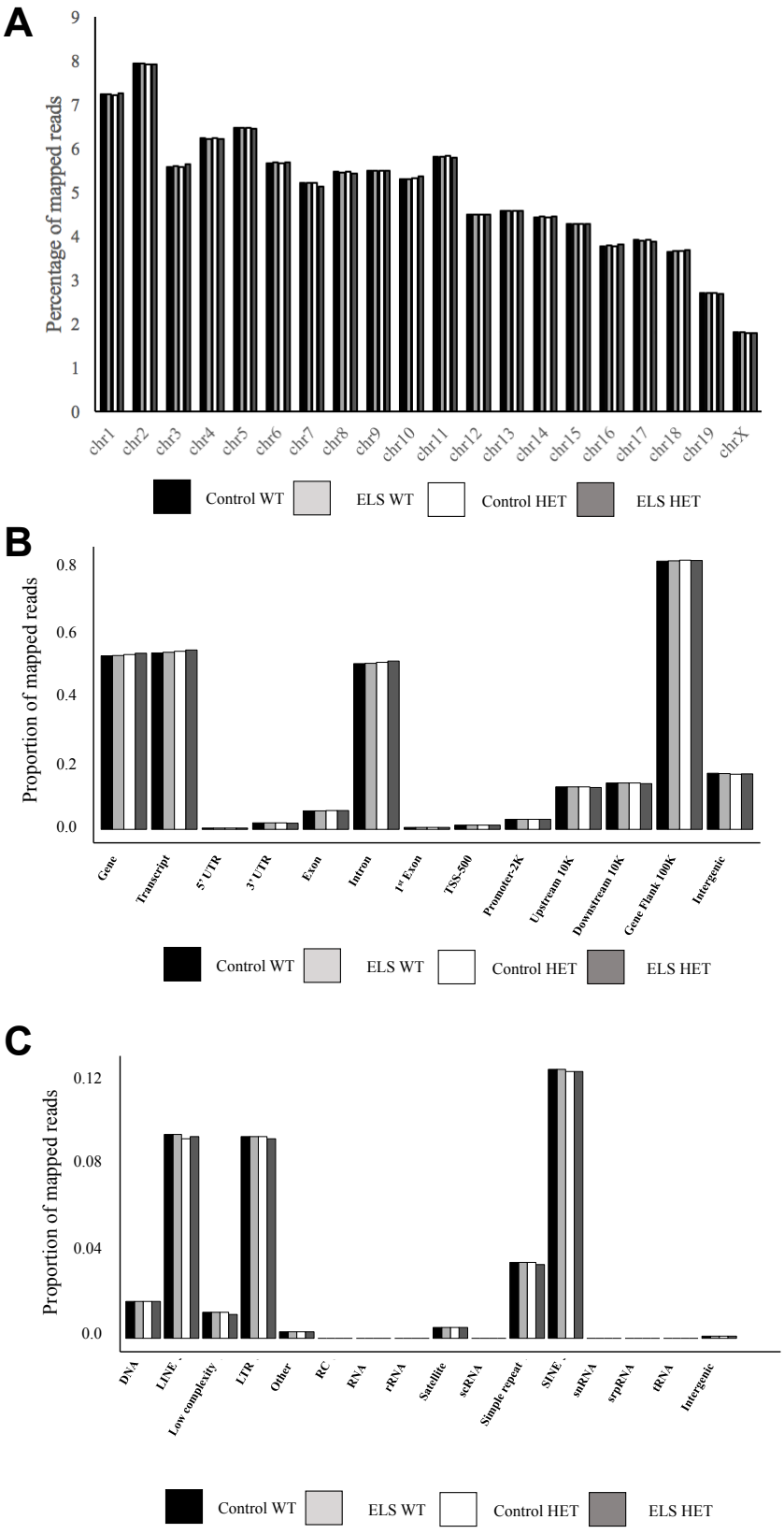
Supplementary Figure 1



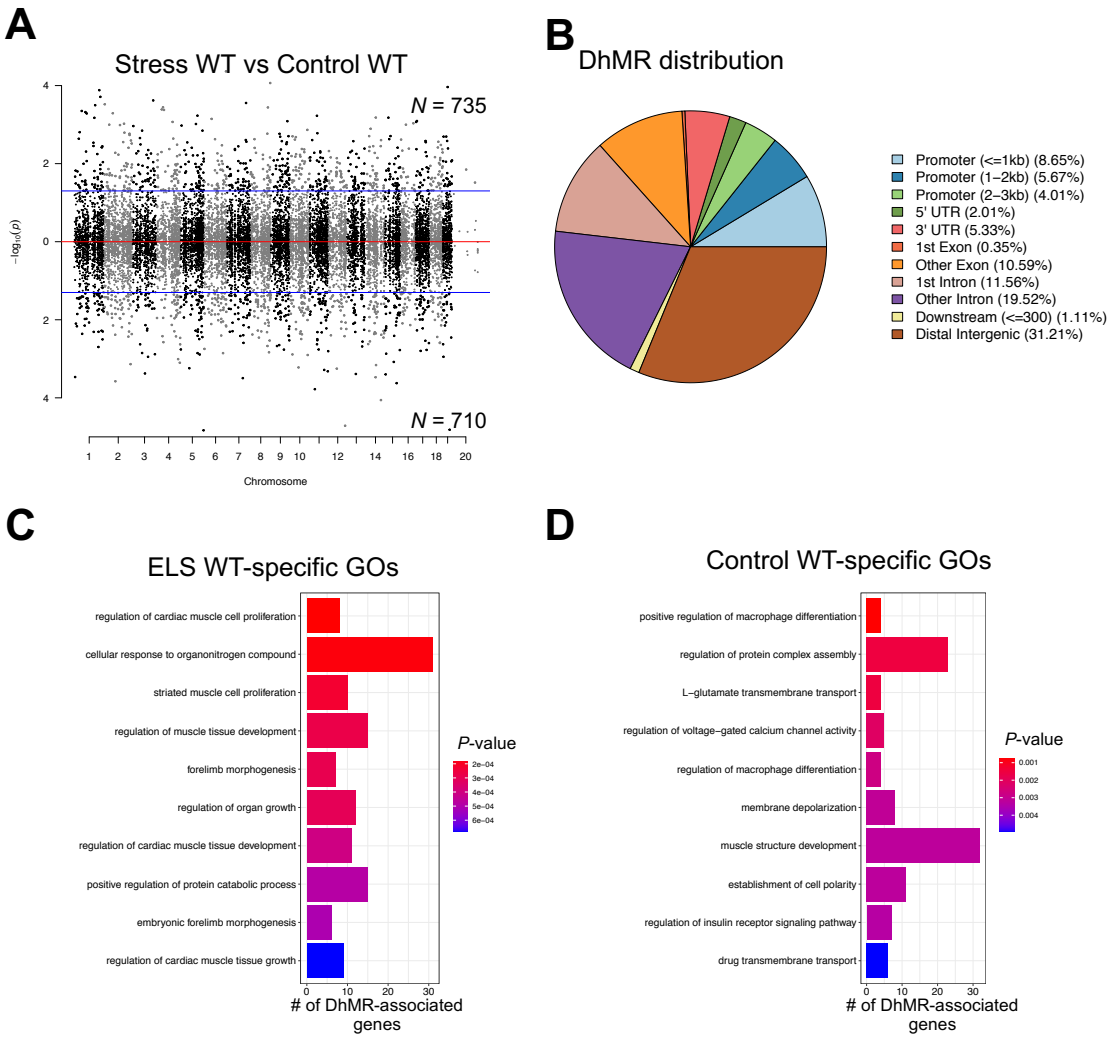
Supplementary Figure 2



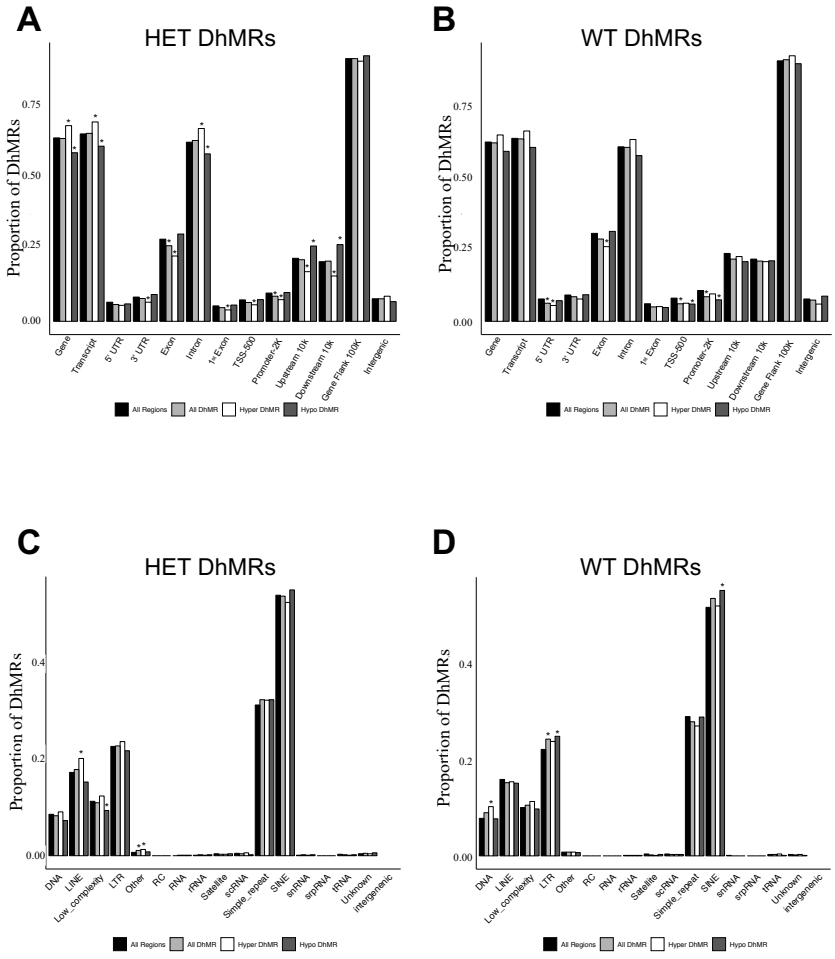
Supplementary Figure 3



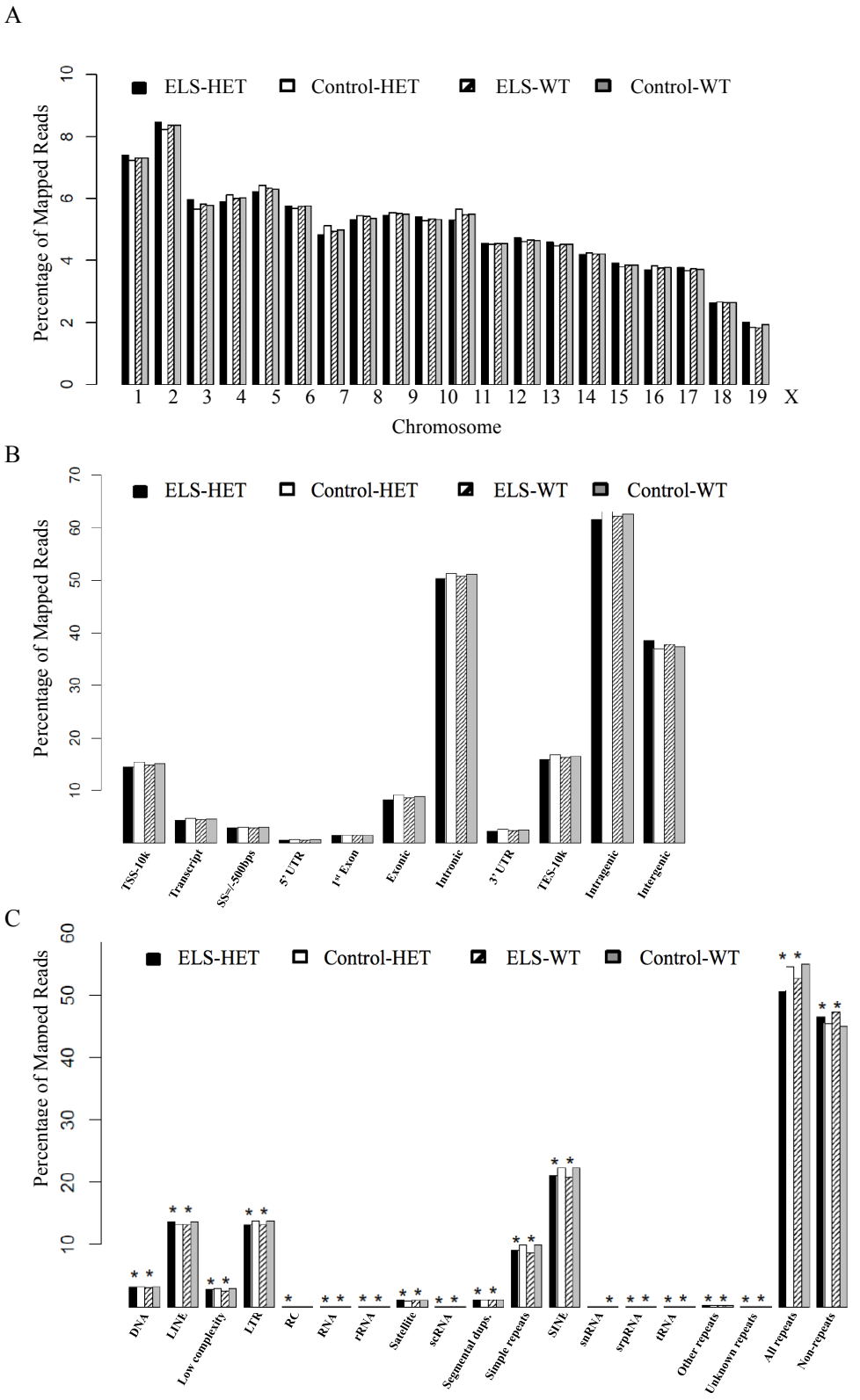
Supplementary Figure 4



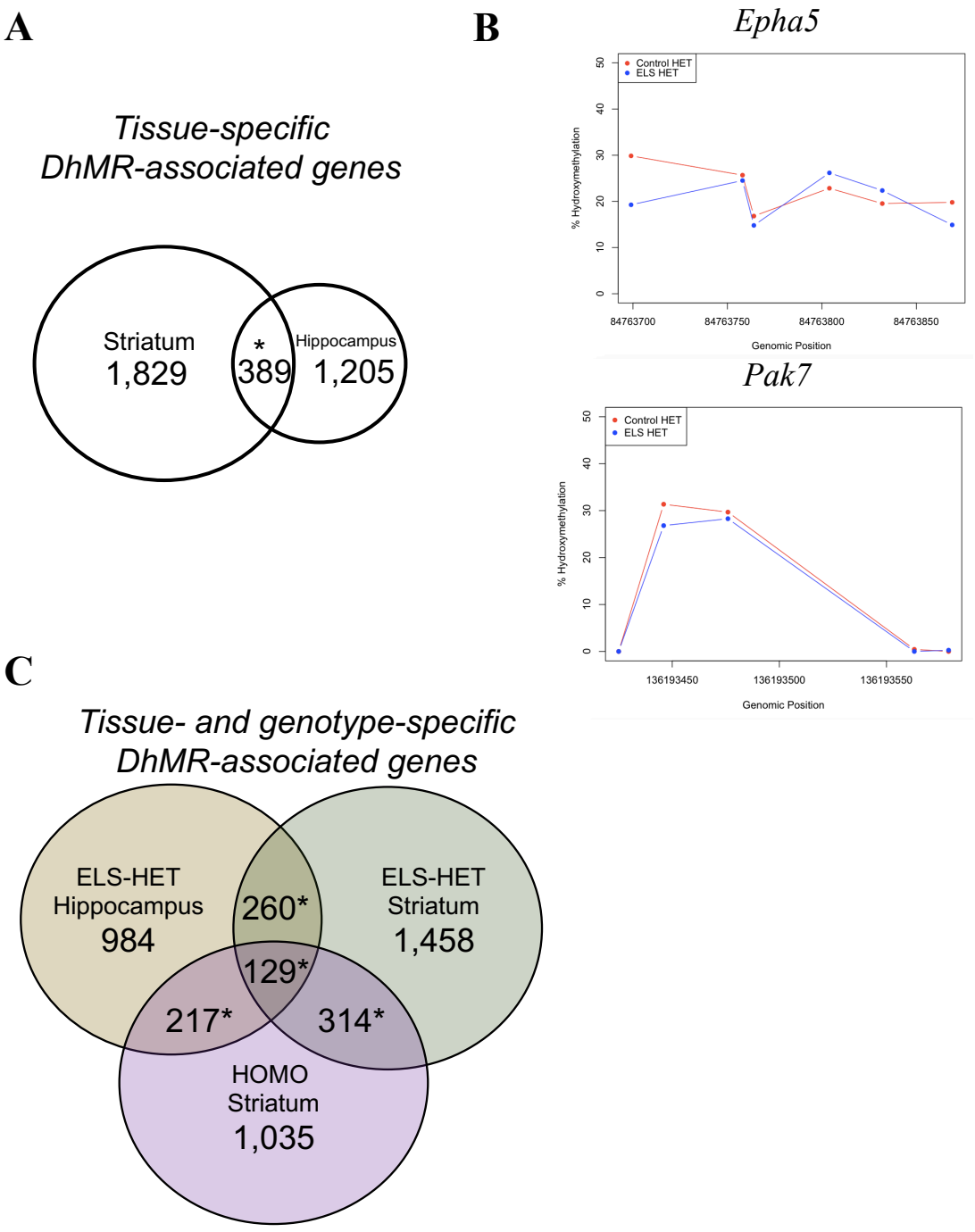
Supplementary Figure 5



Supplementary Figure 6

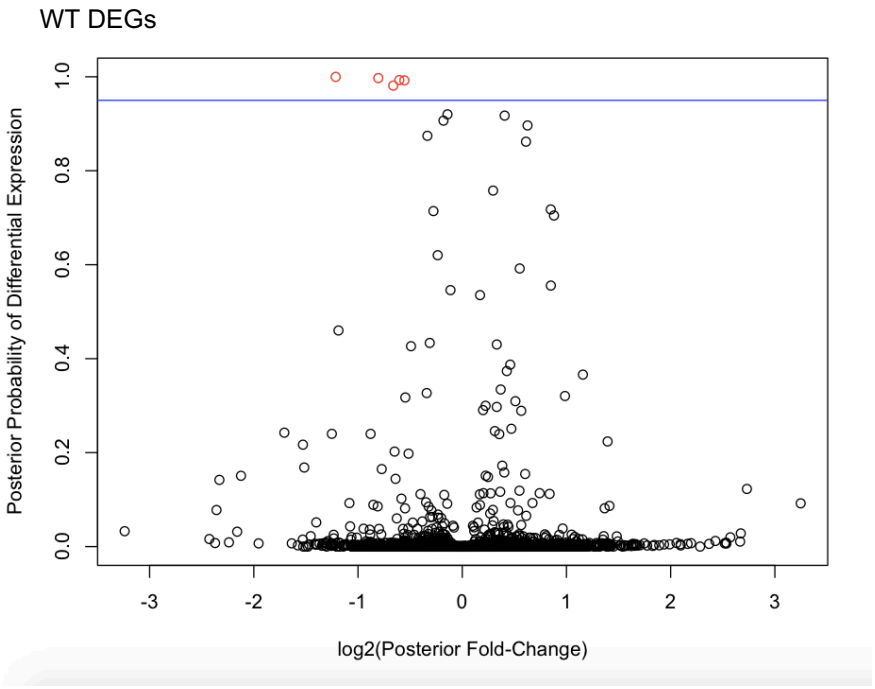


Supplementary Figure 7



Supplementary Figure 8

A



Supplemental Table 1: Summary of behavioral tests in male offspring

	Control		***Early-Life Stress***	
	WT	***Cntnap2*$^{+/-}$**	**WT**	***Cntnap2*$^{+/-}$**
Open Field				
Time spent in the center (s)	47 ± 6	35 ± 7	42 ± 4	54 ± 6
Number of entries to center	19 ± 3	14 ± 1	19 ± 2	20.9 ± 2
Elevated Plus Maze				
Time spent in closed arm (s)	209 ± 13	213 ± 13	213 ± 12	214 ± 8
Time spent in open arm (s)	204 ± 16	208 ± 10	202 ± 10	294 ± 8
Light/Dark Box				
Time spent in light side (s)	386 ± 14	378 ± 25	381 ± 12	376 ± 17
Forced swim test				
Time spent floating	157 ± 8	156 ± 11	145 ± 9	155 ± 8

Comparison of different behavioral tests between early-life stress (ELS) and non-stressed (Control) *Cntnap2*$^{+/-}$ and WT mice. The value are presented as mean ± S.E.M. *p≤ 0.05. N = 8-10. Data analysis: 1) Open Field (time spent in the center: effect of ELS $F_{(1,32)} = 1.06$, $P = 0.3$; effect of genotype $F_{(1,32)} = 0.0009$, $P = 0.97$; number of entries to the center: effect of ELS $F_{(1,32)} = 1.28$, $P = 0.26$; effect of genotype $F_{(1,32)} = 0.29$, $P = 0.59$), 2) Elevated Plus Maze (time spent in closed arm: effect of ELS $F_{(1,35)} = 0.04$, $P = 0.82$; effect of genotype $F_{(1,35)} = 0.04$, $P = 0.82$; time spent in open arm: effect of ELS $F_{(1,35)} = 0.5$, $P = 0.48$; effect of genotype $F_{(1,35)} = 0.02$, $P = 0.87$), 3) Light/Dark Box (time spent in light side: effect of ELS $F_{(1,30)} = 0.03$, $P = 0.86$; effect of genotype $F_{(1,30)} = 0.14$, $P = 0.7$) and 4) Forced Swim Test (time spent floating: effect of ELS $F_{(1,35)} = 0.42$, $P = 0.5$; effect of genotype $F_{(1,35)} = 0.23$, $P = 0.63$).

Supplementary Table 2: Hippocampal DNA sequence read information for each biological replicate pool

Replicates	Total Reads	Monoclonal Reads
H1_Control_WT	31009917	28354228
H2_Control_WT	34141544	31281271
H3_Control_WT	27708730	25349461
H4_Stess_HET	31616944	29011105
H5_Stress_HET	30770924	28164679
H6_Stress_HET	32608364	29846521
H7_Stress_WT	35939256	32905840
H8_Stress_WT	35131002	31950304
H9_Stress_WT	25443415	23313623
H10_Control_HET	47224464	43340378
H11_Control_HET	29163771	26757595
H12_Control_HET	28371476	25917067

Supplemental Table 3: Striatal DNA sequence read information for each biological replicate

Replicates	Total Reads	Monoclonal Reads
38_ELS-Het_ATTCAGAA	43,094,864	38,500,647
40_ELS-Het_ATTCAGAA	22,573,229	19,664,376
42_ELS-Het_CGCTCATT	28,328,088	24,400,876
47_Control-Het_GAGATTCC	33,182,079	29,348,069
52_Control-Het_ATTCAG	62,699,786	56,850,400
58_Control-Het_GAATTCGT	42,042,894	37,823,405
39_ELS-WT_GAATTCGT	25,976,627	22,394,417
41_ELS-WT_ATTCAGAA	36,328,292	32,336,451
43_ELS-WT_GAATTCGT	45,858,981	41,135,682
45_Control-WT_GAGATTCC	30,715,252	26,590,295
46_Control-WT_ATTCAG	75,558,071	68,062,741
54_Control-WT_GAGATTCC	48,954,247	43,406,149

CHAPTER 4

5-hydroxymethylcytosine disrupts CLOCK binding in a gene by environment interaction mouse model of developmental brain disorders

This chapter is adapted from:

Papale, L.A.*, Madrid, A.*, Zhang, Q.*, Chen, K., Sak, L., Keles, S., Alisch, R.S. (2021). Gene by environment interaction mouse model reveals a functional role for 5-hydroxymethylcytosine in neurodevelopmental disorders. (*under review*)

ABSTRACT

Despite a growing body of evidence that 5hmC is a stable modification and that it may affect transcription factor binding affinity, studies have remained largely correlative, with few works exploring the molecular function of 5hmC. In chapter 3 we found that the CLOCK transcription factor sequence motif (CACGTG) was significantly enriched (577/2,612) in the differentially hydroxymethylated regions (DhMRs) found in hippocampal tissue from a gene by environment (GxE) interaction mouse model of developmental brain disorders. Here we tested the hypothesis that variations in 5hmC levels affect the binding of CLOCK in these DhMRs. After establishing that CLOCK expression levels were not affected at the RNA and protein levels, we conducted chromatin immunoprecipitation coupled with high-throughput sequencing on hippocampal tissue from GxE and WT mice and identified 19/577 sites with significant perturbations in CLOCK binding. Three of these 19 sites were located in genes (*i.e., Fry, Gigyf1, Palld*) that exhibited correlated differential expression in the GxE model, suggesting that altered 5hmC levels disrupted the CLOCK binding and gene expression levels of these genes. To formally test if 5hmC disrupts CLOCK binding, we utilized hippocampal nuclear lysates in mobility shift assays and found that 5hmC prevents the binding of CLOCK near the promoters of these genes, providing a mechanistic role for 5hmC (disruption of transcription factor binding) in gene regulation of developmentally important genes.

INTRODUCTION

Environmental adversity is linked with an increased risk to develop developmental brain disorders (DBDs), including anxiety, bipolar, major depression, and personality disorders[6-10]. Indeed, recent work from our lab demonstrated that a model utilizing a gene by environment interaction resulted in phenotypic behaviors in mice that are associated with DBDs, bolstering evidence for gene by environment interactions as an elusive contributor in the complex etiology of DBDs (https://biorxiv.org/cgi/content/short/2021.05.04.441625v1). In addition, we found genome-wide disruptions of 5hmC in this mouse model and a significant enrichment of transcription factor binding motifs in these differential hydroxymethylated regions. While these findings and those of others suggests 5hmC may play a role in regulating transcription factor binding affinity, these studies are correlative, and few studies have formally tested this putative mechanistic function for 5hmC[21-24].

One of the enriched transcription factor binding motifs found in the DhMRs was the binding motif for a key transcription factor in the tight regulation of endogenous circadian rhythms, circadian locomotor output cycles protein kaput (CLOCK)[26]. The primary function of CLOCK is transcriptional activation of downstream genes, while simultaneously promoting rhythmic chromatin opening and DNA accessibility for other transcription factors[27,28]. Likewise, CLOCK displays acetyltransferase activity of histones[29], linking it as an epigenetic modifier. CLOCK binding to DNA shows preferential affinity to E-box motif (CACGTG) enhancer elements[30]. Anomalies in CLOCK function and expression have been found in a wide range of human diseases and illnesses, including tumor progression in breast cancer[31], focal epilepsy[32], obesity[33].

Neurodevelopmental disorders, such as autism, are often associated with genetic and transcript alterations in *CLOCK,* as well as with comorbidities in circadian disturbances[34], highlighting a role for CLOCK in the etiology and progression of DBDs. Despite knowledge of CLOCK as a pioneer transcription factor, few studies have interrogated differential binding affinity of CLOCK caused by DNA methylation, particularly in conjunction with DBD etiologies.

Here, using hippocampal tissue and nuclear protein lysates derived from the GxE (chapter 3) and WT mice, we explored the role of 5hmC on CLOCK binding as a functional epigenetic regulator in the complex etiology of DBDs. Coupling RNA sequencing, western blot, chromatin immunoprecipitation sequencing, and electrophoretic mobility shift assays, we found no evidence of differences in CLOCK RNA or protein expression levels in any genotype or experimental condition, discovered genome-wide sites with altered CLOCK binding as a consequence of the gene by environment interaction, and identified that 5hmC significantly hindered CLOCK binding *in vitro*. Together, these findings indicate a functional role of 5hmC as a regulator of transcription factor binding and provide evidence for 5hmC-mediated transcriptional alterations in the complex molecular etiology of DBDs.

RESULTS

CLOCK expression is not affected in a gene by environment interaction model of DBDs

To identify the effects of 5hmC on DNA binding of CLOCK, we first investigated if CLOCK was differentially expressed in the GxE and WT mice. Four experimental groups were generated (as described in chapter 3, Methods) and tested in downstream analyses: wildtype controls (WT-controls), early-life stress wildtype (ELS-WT), heterozygous controls (HET-controls), and early-life stress heterozygous (ELS-HET). As previous findings identified behavioral and molecular abnormalities using this model at postnatal day 90 (P90), hippocampi were extracted at P90 and nuclear protein lysates were extracted (Methods). Lysates were subjected to western blot analysis using antibodies against CLOCK and these protein immunoblot assays found no evidence of differential CLOCK expression among all experimental groups (**Fig 1A**), nor any differences in loading controls (**Fig 1B and C**). These findings of CLOCK expression at the protein level were corroborated at the RNA level by RNAseq analysis (**Fig 1D**). As such, differences in DNA binding of CLOCK identified in downstream analyses could not be attributed to variable protein abundance between experimental groups, suggesting it would be the result of influence from other factors (*e.g.,* 5hmC).

ChIP-sequencing identifies DhMRs differentially bound by CLOCK

To determine whether or not CLOCK is differentially bound to differentially hydroxymethylated regions (DhMRs) associated with the gene by environment interaction, hippocampal DNA was subjected to chromatin immunoprecipitation sequencing (ChIP-seq), to identify genome-wide regions where CLOCK was differentially bound to DNA in GxE and WT mice (Methods). These methods generated ~10million uniquely mapped, non-duplicated, sequence reads to the mouse genome from each

biological replicate (Methods). DhMRs containing a CLOCK binding site (CACGTG) were assessed for differential CLOCK binding by binning mapped reads to these regions of interest (*N* = 577). Differential analysis of these regions of interest identified that 19 DhMRs exhibited differential CLOCK binding (**Fig. 2A**). Three of these regions were associated with genes also identified to be differentially expressed following the gene by environment interaction. These genes included *Gigyf1*, *Palld*, and *Fry*, all of which have previously been found to be molecularly modified in autism[35-38] (**Fig. 2B, C, D**). Thus, these data suggest that differential hydroxymethylation levels as a consequence of the gene by environment interaction may be influencing CLOCK binding, potentially altering the expression of genes linked with autism.

Hydroxymethylated E-box motifs alter CLOCK binding

To directly test the ability of 5hmC to influence CLOCK binding to DNA, we employed electrophoretic mobility shift assays (EMSAs) using hippocampal nuclear lysates to examine the electrophoretic migration of short DNA fragments that contained sequences found to be differentially bound by CLOCK through ChIP-seq analysis. Double-stranded oligonucleotide probes corresponding to the differentially bound CLOCK sites located in the DhMRs of *Gigyf1*, *Palld*, and *Fry* were generated containing either a cytosine, 5-methylcytosine (5mC), or 5-hydroxymethylcytosine (5hmC) at the cytosine position in the CpG dinucleotide of the E-box motif (CACGTG) of the CLOCK bound regions (Methods). Shift assays identified that all three oligonucleotide probe sets (*i.e., Gigyf1, Palld, Fry*) only exhibited slower electrophoretic migration with the addition of hippocampal nuclear lysates (**Fig. 3A, C, E**). These probe migration shifts were abrogated in 1:1000 cold

competition assays using identically corresponding non-biotinylated probes (Methods), suggesting proteins in the lysate are specifically binding to the oligonucleotide probe sequences to slow their migration. To determine if CLOCK itself was binding to the oligonucleotide probes, resulting in the observed shifts, antibodies specific to CLOCK were pre-incubated with the probes prior to lysates. In these lanes a slower migration of the labeled probe was observed, indicating that CLOCK and its specific antibody were binding to the probe sets (Methods). Notably, antibodies specific to other E-box-motif binding transcription factors (*e.g.,* MAX, USF1), other non-E-box-motif binding transcription factors (*e.g.,* SMAD3), or IgG did not alter the observed initial migration shifts caused by the hippocampal nuclear lysates (data not shown), suggesting that the migration shifts were specific to CLOCK. Subsequently, the presence of 5mC in the oligonucleotide probe resulted in a quantitative decrease in the migration shift, compared to cytosine oligonucleotide probe sets, which was near significance for two of the three tested genes (*i.e., Palld* and *Fry*; **Fig. 3B, D, F**). In contrast, the presence of 5hmC in the oligonucleotide probe resulted in significant decreases in CLOCK binding in all three genes (P-value < 0.05); **Fig. 3B, D, F**). Together, these data suggest that the GxE-induced disruptions of 5hmC facilitate CLOCK binding to the promoters of these genes and prevent their expression (**Fig. 4**). Furthermore, these data corroborate 5hmC as a mediator of CLOCK binding affinity, revealing a mechanistic role for 5hmC as a regulator of transcription of developmentally important genes.

Discussion

Genome-wide disruptions of molecular modifications, such as 5hmC, have been linked to the complex etiology of DBDs[16,17,19], yet the function of these perturbations and the role they play in phenotypic outcomes remain elusive. 5hmC has been hypothesized to hold a CNS-specific function due to its high abundance throughout the CNS and connection with neuronal activity[14]. However, few studies have found direct and definitive molecular mechanism(s) of 5hmC, specifically regarding its effects on the binding properties of transcription factors to DNA. Using a combination of western blot, ChIP-seq, and electrophoretic mobility shift assays, we determined that hydroxymethylation in an E-box motif decreases the binding of a core circadian and neurodevelopment-related transcription factor, CLOCK. These findings identify a functional role of 5hmC, particularly in the context of a GxE interaction model that exhibits DBD-related behavioral deficits, providing evidence for epigenetic regulation of transcription factors. While CLOCK is best known as a transcriptional activator, the 5hmC repressive properties found here parallel previous findings of pluripotency transcription factors[55]. It should be noted that the binding affinity of transcription factors can depend on the specific residue(s) that are molecularly modified[56,57]; the specific residue that is methylated can either increase or decrease transcription factor binding[58,59]. Together, and with these previous findings, our data also suggests that the type of modification (*e.g.,* 5mC and/or 5hmC) determines the affinity of transcription factor binding. Finding that 5mC and 5hmC exhibit unique effects on transcription factor binding affinity may indicate a first step towards unravelling the means by which neuronal cells can dynamically control gene expression in response to a stimulus, and how deviations from these mechanisms can result in aberrations of neuronal function, potentially DBDs.

DBDs often present with circadian disturbances as comorbidities[34], suggesting a link between abnormal circadian rhythm in neuropsychiatric disorders, as either a cause or consequence. Here, the chronic variable stress paradigm used on experimental mice is a circadian disruptor to the pregnant dams, further supporting the hypothesis that disruptions in circadian rhythms contribute to the pathogenesis of these behavior-related disorders. Moreover, modulations in the circadian-controlled melatonin synthesis pathway are thought to be the primary cause of the abnormal circadian rhythms linked with autism spectrum disorders[60,61]. Melatonin synthesis is controlled by core clock genes and transcription factors, including CLOCK[62]. Thus, alterations in CLOCK function and binding can cause abnormal melatonin synthesis, resulting in abnormal circadian rhythms, which may contribute to DBD etiologies. Identifying that 5mC and 5hmC both can influence CLOCK binding affinity provides diverse molecular mechanisms modulating CLOCK function in critical pathways, such as melatonin synthesis. The findings described here highlight a possible clinical application of epigenetic signatures, such as 5hmC, to guide therapeutic options that restore CLOCK binding/function to ameliorate circadian abnormalities that result in DBDs.

The current working model, based on our findings, depicts 5hmC as a transcriptional activating mark, by decreasing repressive transcription factor (*i.e.,* CLOCK) binding. As a consequence of ELS, there is a loss of 5hmC abundance in the promoters and intronic regions of genes such as *Gigyf1, Palld,* and *Fry*, which have links to neurodevelopmental disorders[35,36,38]. The loss of 5hmC results in an increased capacity for CLOCK to bind to

these differentially hydroxymethylated regions, based on ChIP-seq results, which were validated through EMSAs, which directly identified that incorporation of 5hmC decreased CLOCK binding. Notably, all three of these genes (*i.e., Gigyf1, Palld, Fry*) displayed repressed expression, suggesting that the increased binding of CLOCK may be act as a transcriptional repressor, which previously has been reported[63-65]. Moreover, the differential expression of these three genes may be contributing to the DBD-like phenotypes observed following the GxE interaction. GRB10 interacting GYF protein 1 (*Gigyf1*) encodes a gyf family adaptor protein and has recently been identified as the most autism-specific candidate risk gene based on a high frequency of *de novo* variants in patients with autism and other neurodevelopmental disorders[35]. Palladin cytoskeleton associated protein (*Palld*) encodes a cytoskeletal protein required for organization of the actin cytoskeleton and has been found to be differentially expressed in the temporal cortex of autistic patients[36]. FRY microtubule binding protein (*Fry*) encodes a protein that supports the structural integrity of mitotic spindles and the spindle poles during early mitosis and was recently found to be involved in developmental delay, following a homozygous mutation[38]. Together, these findings support that 5hmC acts as a repressive mark for CLOCK binding, which may contribute to the differential expression of genes linked to DBDs and result in the observed abnormal behavioral phenotypes of the GxE interaction model. These findings illustrate the importance of 5hmC as a molecular regulator of transcription throughout the course of CNS development, highlighting its potential as a therapeutic target in the diagnosis and amelioration of DBDs.

Methods

Mice and genotyping

All experiments were approved by the University of Wisconsin – Madison Institutional Animal Care and Use Committee (M02529). Heterozygous male *Cntnap2* mice were purchased from the Jackson laboratories (Bar Harbor, ME) and maintained on C57BL/6J background. *Cntnap2*$^{+/-}$ mutants were genotyped using the following primers: Mutant Rev: CGCTTCCTCGTGCTTTACGGTAT, Common: CTGCCAGCCCAGAACTGG, WT Rev: GCCTGCTCTCAGAGACATA. PCR amplification was performed with one cycle of 95C for 5min and 31 cycles of 95C for 30sec, 56C for 30sec, 68C for 30sec, followed by 68C for 10min. The mutant allele was obtained with a 350bp and WT allele with a 197bp PCR products.

Breeding scheme and prenatal stress paradigm

To minimize for the stress of animal handling, all of the following were conducted by a single researcher: animal colony maintenance, breeding, prenatal stress, and behavioral tests. For breeding, a three month old *Cntnap2+/-* male mouse was placed together with a virgin 3 month old WT C57/BL6J female mouse at 6pm (one hour prior to lights off); every morning before 9am (two hours after lights on) female mice were checked for vaginal copulation plug and separated from the male. Presence of a copulation plug denoted day 1 of gestation and the pregnant female was individually housed and given a cotton nestlet. At day 12 of gestation (E12), pregnant females were randomly assigned to either variable stress or to a non-stressed control group. Pregnant mice assigned to

the variable stress group experienced a daily stressor on each of the seven days during late pregnancy (E12 to E18; the time of which was chosen because it overlaps with the onset of *Cntnap2* expression (E14)). These variable stressors included: 36 hours of constant light, 15 min of fox odor (Cat# W332518) beginning 2 hours after lights on, novel objects (8 marbles) exposure overnight, 5 minutes of restraint stress (beginning 2 hours after lights on), novel noise (white noise, nature sound-sleep machine®, Brookstone) overnight, 12 cage changes during light period, and saturated bedding (700 mL, 23°C water) overnight[66]. These mild stressors were previously published and were selected because they do not include pain or directly influence maternal food intake or weight gain. Importantly, both WT and HET mice were taken from the same litter, providing an ideal internal control for our findings and negates the need for cross-fostering of the prenatally stress pups to non-stressed moms to find effects of early-life exposure to stress. Litter sizes of less than 5 and more than 8 pups were removed from the experiment. Offspring were ear-tagged at (postnatal day) (P) 12 and left undisturbed until weaning day (P18), at which time the mice were group housed with same sex. Females were left intact, and were not cycled. Finally, to minimize the effect of parent-to-offspring interaction per litter, 3 pups/litter were randomly selected for molecular experiments. Importantly, the mice used for behavioral and molecular analysis were left undisturbed until sacrifice by perfusion at 3 months of age, after which whole brains were extracted, and hippocampal tissue was excised on ice, then frozen at -80^{C} until use. To control for circadian effects, such as changes in CLOCK expression throughout the 24h period, all mice were sacrificed between the hours of 9am-11am.

Western blot

Mouse hippocampal nuclear extracts were isolated from adult female mice (P90) using a nuclear and cytoplasmic extraction kit per the manufacturer's instructions (Thermo Scientific 78833). For western blot analysis, 40ug of total protein from hippocampal nuclear lysates was boiled for 5min to dissociate complexes and run in a 4-20% gradient gel (Bio-Rad 4561096) for 1hr at 150V in 1x Tris/Glycine/SDS (Bio-Rad 1610732). Following separation of proteins, gels were transferred to nylon membrane in 1x Tris/Glycine for 1hr on ice at 100V. Following transfer, membranes were washed in 1X TBST and blocked in 5% milk/TBST for 1hr at room temperature. After blocking, membranes were incubated with primary antibodies for 12hr at 4^{C}. Primary antibodies included: 1:200 dilution anti-CLOCK (ab3517) and 1:4000 dilution anti-Actin (ab8226). Excess primary antibody was removed and membranes were washed in 1x TBST for 1hr at room temperature. Secondary antibody in 5% milk/TBST was incubated at room temperature with membranes. Secondary antibody included: (ab97051). Excess secondary antibody was removed and membranes were washed at room temperature 1hr with 1x TBST. Chemilumiescence was achieved by employing the Pico Chemiluminescence Kit (Thermo Fisher 34579) following manufacturer's instructions. Visualization was achieved utilizing an Odyssey® Fc imaging system (Li-Cor) using a 30sec exposure.

Chromatin immunoprecipitation

Hippocampi were extracted as described above, and were kept at -80C until further processing. Antibodies were bound and pre-blocked to magnetic beads. Fresh block

solution was made (225mg bovine serum albumin, 45ml ice cold 1x PBS). Dynabead protein A (50ul) (Thermo Fisher 10001D) and protein G (50ul) (Thermo Fisher 10003D) were added to 1ml of block solution. Magnetic beads were collected on a magnetic stand for 5min. Block solution was suctioned off and this wash cycle was repeated a total of three times. After the final wash, 5ug of anti-CLOCK antibody (ab3517) or IgG (ab171870) were added to magnetic beads, and block solution was added to a final volume of 250ul. This solution was constantly rotated and incubated at 4C for the duration of chromatin preparation.

Chromatin preparation and shearing was performed using a Covaris truCHIP chromatin shearing tissue kit (Covaris 520237). Following antibody pre-blocking, quenching buffer and fixing buffer were made following manufacturer's instructions (Covaris 520237). Hippocampi (25-40mg) were removed from -80C and kept on dry ice. Each hippocampus was individually cut to $1mm^3$ segments in a petri dish, kept frozen using liquid nitrogen, then placed in tubes on dry ice. Once all hippocampi were prepared for further processing, tubes were placed at RT and 400ul of ice cold 1x PBS was added. Tubes were centrifuged at 4C for 5min at 3300rpm. While tubes were centrifuged, 1ml of 16% formaldehyde was added to the fixing buffer, to create a final 11.1% formaldehyde fixing buffer. Supernatant was removed from tubes and tubes were placed on ice. Tissue was resuspended in 400ul of fixing buffer and rocked at RT for 8min. Tissue fixation was quenched using 24ul of quenching buffer and rocked at RT for 5min. Tissue was centrifuged for at 4C at 3300rpm for 5min. Supernatant was removed and tissue was washed in 400ul of ice cold 1x PBS, centrifuged at 4C at 3300rpm for 5min. This 1x PBS wash was repeated twice. After the

final wash, supernatant was removed and tubes were placed in dry ice to flash freeze the tissue, then placed on dry ice.

Tissue pulverization was performed using a Covaris CP02 cryoPREP automated dry pulverizer (Covaris 500001). Freeze dried tissue was placed in Covaris tissue bags (Covaris 520001) and pulverized using a setting of 5 on the cryoPREP, twice. Following pulverization, tissue was placed in liquid nitrogen then placed on dry ice. Pulverization was repeated for all tissue samples. Following pulverization, lysis buffer, protease inhibitor cocktail, wash buffer, and shearing buffer were made following manufacturer's instructions (Covaris 520237). Tissue was thawed on ice. 400ul of lysis buffer was added to each tissue bag, resuspended by pipetting, and contents were transferred to new tubes on ice. Tubes were rotated at 4C for 20min, then centrifuged at 4C for 5min at 2000xg. Supernatant was removed and 400ul of wash buffer was added to tubes, then samples were rotated at 4C for 10min. Samples were centrifuged at 4C for 5min at 2000xg. This wash was repeated a total of two times. Following the final wash, 125ul of shearing buffer was added to tubes and incubated on ice for 10min. Following incubation, 130ul of the prepared nuclei were transferred to microTUBES and placed on ice. Shearing of chromatin was performed using a Covaris S220 focused-ultrasonicator (500217), using the following parameters: PIP 105, 2% duty factor, CPB 200, treatment time of 8min, setpoint temperature of 6C, minimum temperature of 3C, maximum temperature of 9C, and continuous degassing. Following sonication, samples were placed on ice. ChIP dilution buffer was made following manufacturer's instructions (Covaris 520237). 160ul of ChIP dilution buffer was added to new tubes, followed by 130ul of sheared chromatin

samples. Samples were centrifuged for 10min at 4C at max speed (20,000xg). Supernatant was transferred to new tubes set on ice and protein quantification was performed using a Qubit protein assay kit (Thermo Fisher Q33212) and a Qubit 4 fluorometer (Q33238).

Pre-blocked magnetic bead-bound antibodies were washed in fresh block solution as described above, for an additional three washes. Following the washes, magnetic beads were resuspended in 100ul of block solution. 600ug of protein from sheared samples were placed in corresponding tubes containing anti-CLOCK antibody or IgG control. ChIP dilution buffer was added to a final total volume of 300ul. Immunoprecipitation and IgG samples were rotated overnight at 4C. Following overnight incubation, magnetic beads were collected using a magnetic stand for 5min and the supernatant was removed. Tubes were removed from the magnetic stand and resuspended in 1ml of RIPA buffer (50 mM Hepes–KOH, pH 7.5; 500 mM LiCl; 1 mM EDTA; 1% NP-40 or Igepal CA-630; 0.7% Na– Deoxycholate). Magnetic beads were collected on a magnetic stand for 5min and RIPA buffer was removed. This wash cycle was repeated for a total of five washes. Following the RIPA buffer washes, 1ml of TBS (20 mM Tris–HCl, pH 7.6; 150 mM NaCl) was added to each tube, beads were resuspended, and then collected on a magnetic stand for 5min. TBS was removed and tubes were centrifuged at 1000rpm for 3min at 4C. Tubes were placed on a magnetic stand and any residual TBS was suctioned off. 200ul of elution buffer (20 mM Tris–HCl, pH 7.6; 150 mM NaCl) was added to each tube. Samples were placed in a water bath set at 65C overnight to reverse-crosslink proteins from DNA.

Following reverse-crosslinking, samples were placed on a magnetic stand and magnetic beads were collected for 5min. Supernatant from immunoprecipitated and IgG samples were transferred to new tubes, and 200ul of TE (10 mM Tris-HCl; pH 8.0; 1mM EDTA) was added, along with 10ug of RNaseA (EN0531) and incubated in a 37C water bath for 30min. After the incubation, 4ul of proteinase K (Invitrogen, 25530-049) was added to each sample, mixed, and incubated at 55C for one hour, after which 400ul of phenol-chloroform-isoamyl alcohol, mixed, and each sample was transferred to a 2ml phase lock gel light tube (FPR5101) and centrifuged for 5min at 16,000xg. The aqueous layer was removed and transferred to a new tube containing 16ul of 5M NaCl and 1ul of GlycoBlue (AM9516). 800ul of 100% ethanol was added, mixed, and tubes were incubated for 30min at -80C. After the 30min incubation, samples were centrifuged at 4C for 10min at 20,000xg to pellet the DNA. Pellets were washed with 500ul of 80% ethanol and centrifuged for 5min at 4C at 20,000xg. Supernatant was removed and DNA pellets were allowed to air dry for 30min at room temperature. DNA pellets were eluted in 50ul of ultra-pure water 10 mM Tris–HCl, pH 8.0, placed in a 55C water bath for 10min, and vortexed to suspend the DNA.

ChIP-sequencing

Libraries were generated from eluted DNA from immunoprecipitated and IgG samples using the NEBNext ChIP-seq library prep reagent set for Illumina sequencing, according to the manufacturer's instructions. Briefly, eluted DNA fragments were purified after the adapter ligation step using using AMPure XP beads (Agencourt A63880). An Agilent 2100 Bioanalyzer was used to quantify the amplified library DNA and 20pM of diluted

libraries were used for sequencing. A Sage Science Pippin HT was used to size select DNA from libraries, for an average size of ~400bp per sample. 50-cycle paired-end sequencing was performed by the University of Wisconsin – Madison Biotechnology Center. Image processing and sequence extraction were done using the standard Illumina Pipeline.

ChIP-sequencing analysis

Raw paired-end sequencing files were assessed for quality using FASTQC (https://www.bioinformatics.babraham.ac.uk/). Adapters were removed and reads were quality trimmed, using a quality score cutoff of 30, using trim_galore (https://www.bioinformatics.babraham.ac.uk/). Trimmed reads were aligned to the *mm9* genome using bowtie2 v2.3.4[67], using local alignment. Following alignment, only uniquely mapped reads were used for downstream analysis, using samtools to filter multi-mapping reads[68], followed by the removal of PCR duplicates using samtools, as well. As we sought to determine whether 5hmC alters CLOCK binding, uniquely-mapped, filtered reads which showed enrichment compared to sequenced reads of IgG control samples that mapped to differentially hydroxymethylated regions (DhMRs) found to contain a putative CLOCK binding site (CACGTG) by motif enrichment analysis (N = 577), were used for differential binding analyses. R package *edgeR*[69] was used to perform differential analysis, normalizing read counts to background IgG reads. A P-value threshold of 0.05 was used to determine significant differential binding of CLOCK between groups.

Electrophoretic mobility shift assays

Complementary 5'-biotinylated (0.5uM) and unlabeled oligonucleotides (Integrated DNA Technologies) were annealed by heating equal concentrations of sense and anti-sense oligonucleotides to 95C and lowering the temperature by 2C in 3min intervals until reaching 23C. Mouse hippocampal nuclear extracts were isolated from adult female mice (P90) using a nuclear and cytoplasmic extraction kit per the manufacturer's instructions (Thermo Scientific 78833). Shift assays were performed using the LightShift® Chemilumiescent EMSA Kit (Thermo Scientific 20148). DNA binding reactions were performed in a 20ul system containing biotinylated oligonucleotides and nuclear extracts (12ug). For cold competition assays, a 1:1000 concentration of unlabeled oligonucleotides was incubated on ice for 1hr with nuclear extracts prior to the addition on biotinylated oligonucleotides. For supershift assays, 5ug of anti-CLOCK antibody (ab3517) or anti-BMAL1 antibody (ab93806) were incubated on ice with nuclear extract for 1hr prior to the addition of biotinylated oligonucleotides. In such assays, nuclear extracts without the addition of unlabeled oligonucleotides or antibody were also incubated on ice for 1hr prior to the addition of biotinylated oligonucleotides, to ensure the absence of protein degradation during the incubation period. Following incubation on ice, biotinylated oligonucleotides (0.5uM) were added to binding reactions and incubated at room temperature for 20min. Reaction products were separated by electrophoresis at 100V for 60min. Following separation, protein::DNA complexes were transferred onto a positively-charged nylon membrane (Thermo Scientific 77016) for 60min at 380A on ice. Protein::DNA complexes were detected using the Nucleic Acid Detection Module Kit (Thermo Scientific 89880) per the manufacturer's instructions, and imaged using an Odyssey® Fc imaging system (Li-Cor) using a 60min exposure.

ImageJ band quantification

Band intensity percentages were quantified using ImageJ software. First, images were set to grey-scale. Bands of interest were selected across all lanes and their inverted means were obtained. Background noise intensity was subtracted out by subtracting the intensity of staining above the bands of interest. Band intensities were normalized to loading controls lanes (*i.e.,* beta-actin (western blot)). For EMSA, a two-sided T-test (R environment) was used to determine significant alterations in binding affinity to methylated- and hydroxymethylated-oligonucleotides, using three independent EMSA replicates. For western blot, a two-sided T-test (R environment) was used to determine significant changes in CLOCK expression between groups/genotypes.

Oligonucleotide sequences

Unmodified Probes:
Fry FW: 5'-Biotin-TATGTTCATCCACGTGATGC-3'
Fry RV: 5'-Biotin-GCATCACGTGGATGAACATA-3'

Gigyf1 FW: 5'-Biotin-GGGTACACGTGCGCCATGGC-3'
Gigyf1 RV: 5'-Biotin-GCCATGGCGCACGTGTACCC-3'

Palld FW: 5'-Biotin-GAATGTCCACACGTGTATGC-3'
Palld RV: 5'-Biotin-GCATACACGTGTGGACATTC-3'

5hmC Probes
Fry FW: 5'-Biotin-TATGTTCATCCA^{hme}CGTGATGC-3'
Fry RV: 5'-Biotin-GCATCA^{hme}CGTGGATGAACATA-3'

Gigyf1 FW: 5'-Biotin-GGGTACA^{hme}CGTGCGCCATGGC-3'
Gigyf1 RV: 5'-Biotin-GCCATGGCGCA^{hme}CGTGTACCC-3'

Palld FW: 5'-Biotin-GAATGTCCACA^{hme}CGTGTATGC-3'
Palld RV: 5'-Biotin-GCATACA^{hme}CGTGTGGACATTC-3'

5mC Probes
Fry FW: 5'-Biotin-TATGTTCATCCA^{me}CGTGATGC-3'
Fry RV: 5'-Biotin-GCATCA^{me}CGTGGATGAACATA-3'

Gigyf1 FW: 5'-Biotin-GGGTACA^{me}CGTGCGCCATGGC-3'
Gigyf1 RV: 5'-Biotin-GCCATGGCGCA^{me}CGTGTACCC-3'

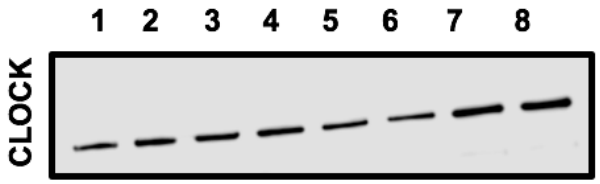
Palld FW: 5'-Biotin-GAATGTCCACA^{me}CGTGTATGC-3'
Palld RV: 5'-Biotin-GCATACA^{me}CGTGTGGACATTC-3'

Unlabeled Probes:
Fry FW: 5'-TATGTTCATCCACGTGATGC-3'
Fry RV: 5'-GCATCACGTGGATGAACATA-3'

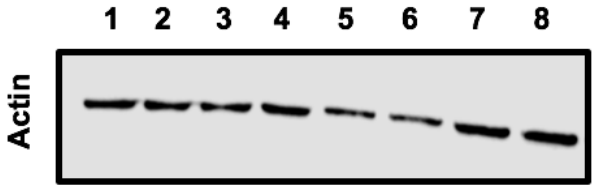
Gigyf1 FW: 5'-GGGTACACGTGCGCCATGGC-3'
Gigyf1 RV: 5'-GCCATGGCGCACGTGTACCC-3'

Palld FW: 5'-GAATGTCCACACGTGTATGC-3'
Palld RV: 5'-GCATACACGTGTGGACATTC-3'

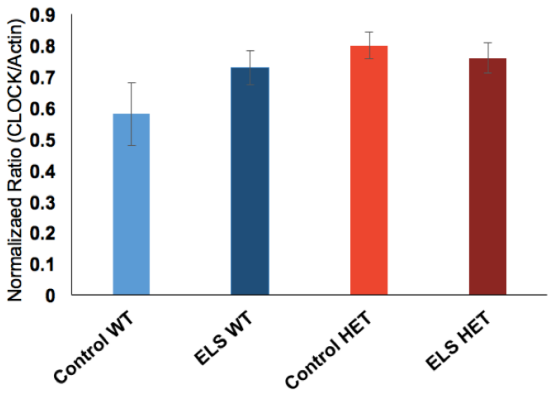
A


1 2 3 4 5 6 7 8
CLOCK

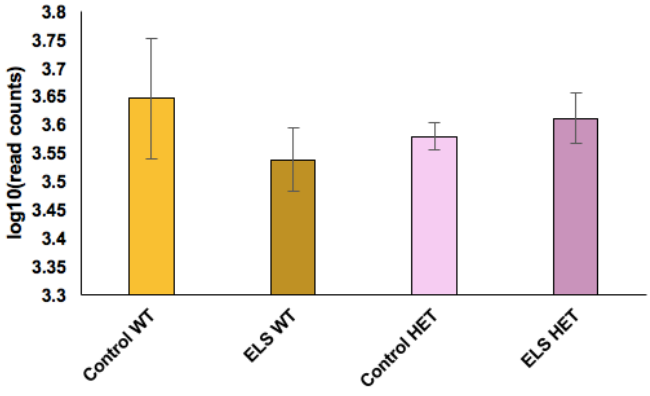
B

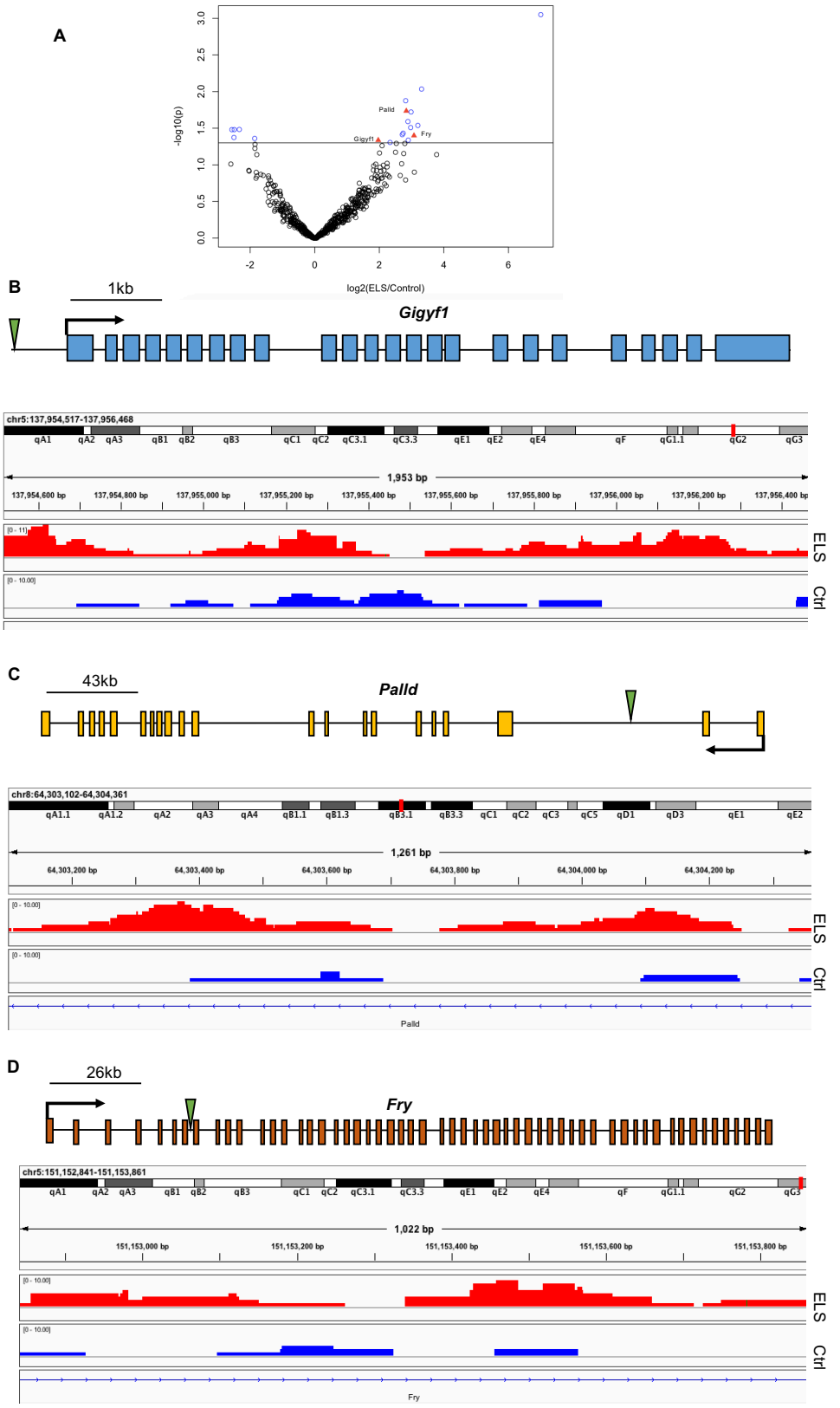

1 2 3 4 5 6 7 8
Actin

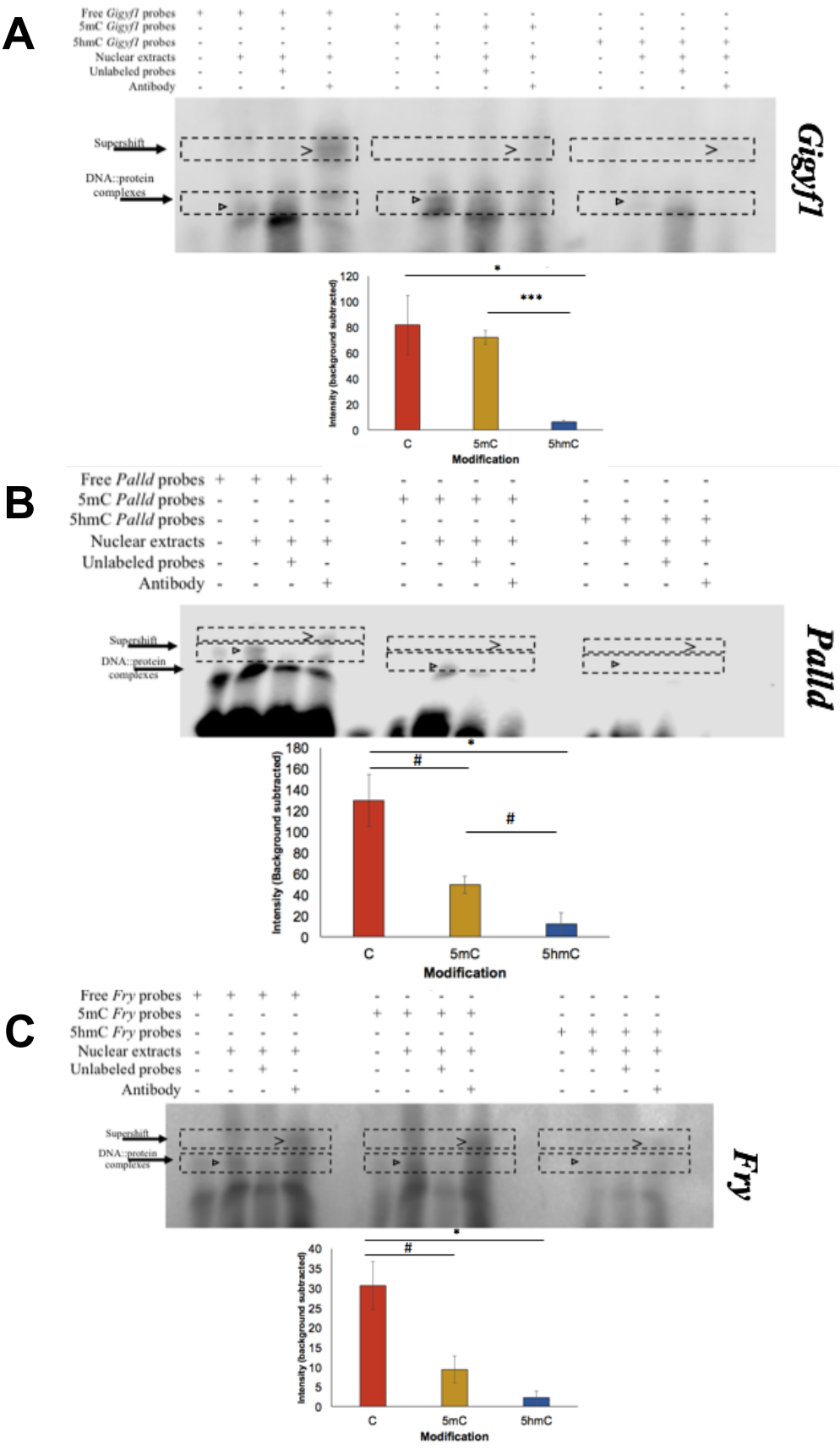
C


Normalizaed Ratio (CLOCK/Actin)
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Control WT
ELS WT
Control HET
ELS HET

D


log10(read counts)
3.8
3.75
3.7
3.65
3.6
3.55
3.5
3.45
3.4
3.35
3.3
Control WT
ELS WT
Control HET
ELS HET


A
-log10(p)
log2(ELS/Control)
Palld
Gigyf1
Fry
B
1kb
Gigyf1
chr5:137,954,517-137,956,468
1,953 bp
ELS
Ctrl
C
43kb
Palld
chr8:64,303,102-64,304,361
1,261 bp
ELS
Ctrl
Palld
D
26kb
Fry
chr5:151,152,841-151,153,861
1,022 bp
ELS
Ctrl
Fry


A
Free Gigyf1 probes
5mC Gigyf1 probes
5hmC Gigyf1 probes
Nuclear extracts
Unlabeled probes
Antibody
Supershift
DNA::protein complexes
Gigyf1
Intensity (background subtracted)
Modification
C
5mC
5hmC
B
Free Palld probes
5mC Palld probes
5hmC Palld probes
Nuclear extracts
Unlabeled probes
Antibody
Palld
C
Fry
Free Fry probes
5mC Fry probes
5hmC Fry probes

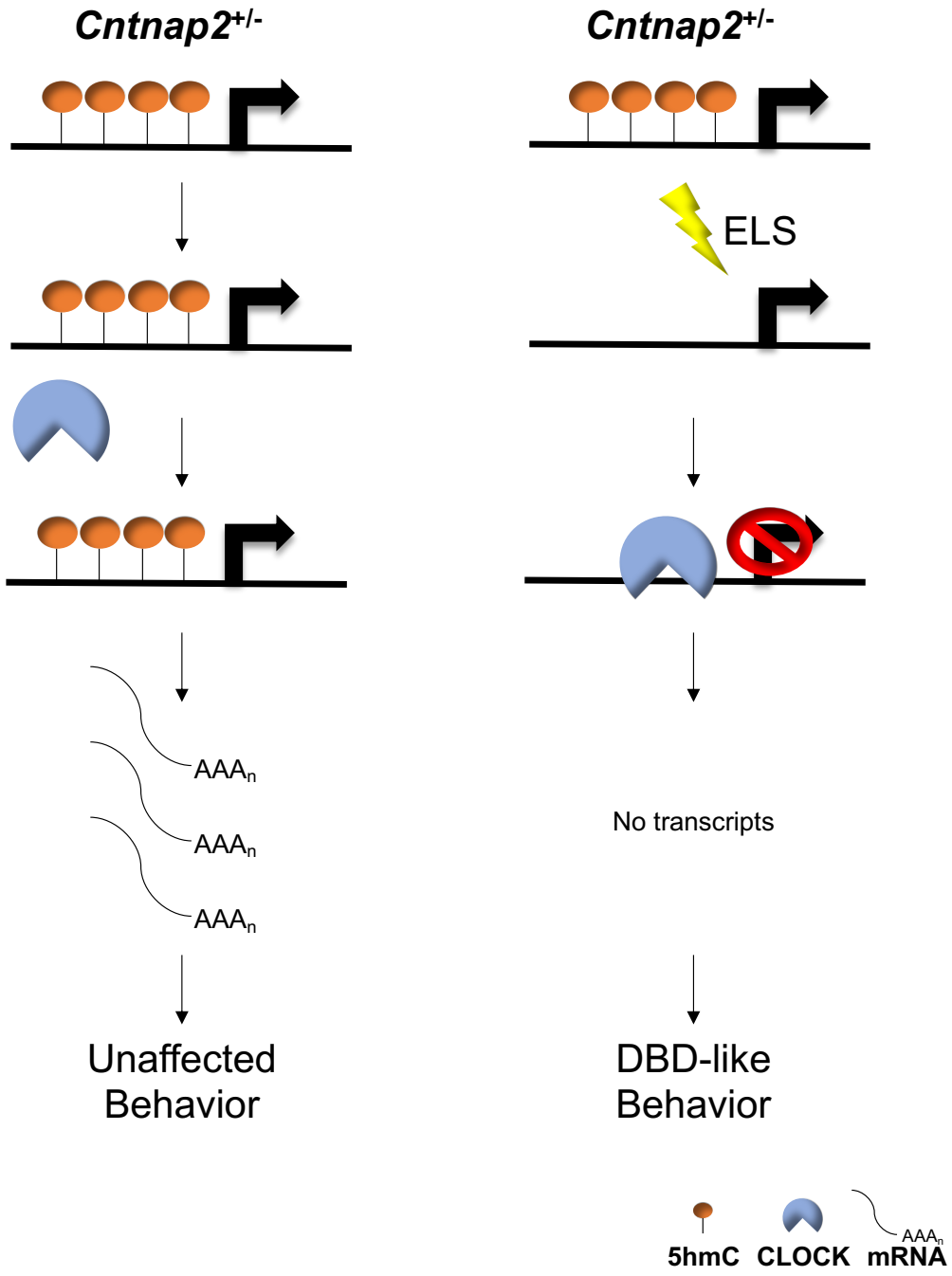

Cntnap2+/-
Cntnap2+/-
ELS
AAAn
AAAn
AAAn
No transcripts
Unaffected Behavior
DBD-like Behavior
5hmC
CLOCK
AAAn
mRNA

Figure Legends

Figure 1: CLOCK protein expression is not affected by genotype or ELS. (A) Western blot identified no gross differences between Control WT (lanes 1 and 2), ELS WT (lanes 3 and 4), Control HET (lanes 5 and 6), or ELS HET (lanes 7 and 8) in CLOCK protein expression (A) or in beta-actin (B), used as a loading control. (C) A bar plot depicts the quantifications used to determine any significant differences in CLOCK protein expression between Control WT (light blue), ELS WT (dark blue), Control HET (light red), ELS HET (dark red). Values are depicted as CLOCK band intensities normalized to beta-actin bands from each respective lane. Error bars show the standard error of the mean. (D) A bar plot depicts the log10(read counts) (y-axis) of Control WT (light yellow), ELS WT (dark yellow), Control HET (light pink), and ELS HET (dark pink), as determined by RNA-seq analysis, identifying no differential gene expression among any of the four groups.

Figure 2: ChIP-seq identified regions displaying differentially bound CLOCK. (A) A volcano plot displays the –log2(fold-change) of ELS HET compared to Control HET (x-axis) and the significance (y-axis) from differential analysis of ChIP-seq data. Blue dots above the black line represent differentially hydroxymethylated regions (DhMRs) differentially bound by CLOCK (*P*-value <0.05), while red triangles show the three DhMRs that were differentially bound by CLOCK, while also exhibiting differential gene expression which included the genes *Gigyf1*, *Palld*, and *Fry*. (B) Top panel: a gene schematic of *Gigyf1* shows the location of the DhMR exhibiting differential CLOCK binding (green triangle) as a result of the ELS relative to exons (blue boxes) of *Gigyf1*. Bottom panel: zoomed-in tracks of the DhMR associated to *Gigyf1* exhibiting differentially bound CLOCK display the abundance of ChIP-seq reads binned to this region for ELS HET animals (red tracks) and Control HET animals (blue tracks). (C) Top panel: a gene schematic of *Palld* shows the location of the DhMR exhibiting differential CLOCK binding (green triangle) as a result of the ELS relative to exons (yellow boxes) of *Palld*. Bottom panel: zoomed-in tracks of the DhMR associated to *Palld* exhibiting differentially bound CLOCK display the abundance of ChIP-seq reads binned to this region for ELS HET animals (red tracks) and Control HET animals (blue tracks). (D) Top panel: a gene schematic of *Fry* shows the location of the DhMR exhibiting differential CLOCK binding (green triangle) as a result of the ELS relative to exons (orange boxes) of *Fry*. Bottom panel: zoomed-in tracks of the DhMR associated to *Fry* exhibiting differentially bound CLOCK display the abundance of ChIP-seq reads binned to this region for ELS HET animals (red tracks) and Control HET animals (blue tracks).

Figure 3: EMSA identifies that 5hmC inhibits CLOCK binding. (A) Top panel: shift assays (lanes 2 6, and 10), cold competition assays (lanes 3, 7, and 11), and supershift assays (lanes 4, 8, and 12) of an E-box motif in the DhMR of *Gigyf1* that exhibited differential CLOCK binding through ChIP-seq analysis were performed using hippocampal nuclear extracts. Biotinylated oligonucleotides contained either an unmodified E-box motif (lanes 1-4), a methylated E-box motif (lanes 5-8), or a hydroxymethylated E-box motif (lanes 9-12). Black arrows indicate free probes, bands of interest (DNA::protein complexes) that are abrogated with cold competition, and supershifts. Bottom panel: band intensity

quantifications of the area encompassing the bands of interest using ImageJ software. The E-box modifications (x-axis) and mean of band intensity (y-axis) are depicted +/- SEM. The methylated E-box motif (yellow) did not affect CLOCK binding compared to the unmodified (red) motif. The hydroxymethylated E-box motif (blue) significantly decreased binding compared to the unmodified E-box (* *P*-value <0.05) and the methylated E-box motif (*** *P*-value <0.001). (B) Top panel: shift assays (lanes 2 6, and 10), cold competition assays (lanes 3, 7, and 11), and supershift assays (lanes 4, 8, and 12) of an E-box motif in the DhMR of *Palld* that exhibited differential CLOCK binding through ChIP-seq analysis were performed using hippocampal nuclear extracts. Biotinylated oligonucleotides contained either an unmodified E-box motif (lanes 1-4), a methylated E-box motif (lanes 5-8), or a hydroxymethylated E-box motif (lanes 9-12). Black arrows indicate free probes, bands of interest (DNA::protein complexes) that are abrogated with cold competition, and supershifts. Bottom panel: band intensity quantifications of the area encompassing the bands of interest using ImageJ software. The E-box modifications (x-axis) and mean of band intensity (y-axis) are depicted +/- SEM. The methylated E-box motif (yellow) showed a trend towards significantly decreasing CLOCK binding compared to the unmodified (red) motif (# *P*-value = 0.054). The hydroxymethylated E-box motif (blue) significantly decreased binding compared to the unmodified E-box (* *P*-value <0.05) and trended towards significance compared to the methylated E-box motif (# *P*-value = 0.073). (C) Top panel: shift assays (lanes 2 6, and 10), cold competition assays (lanes 3, 7, and 11), and supershift assays (lanes 4, 8, and 12) of an E-box motif in the DhMR of *Fry* that exhibited differential CLOCK binding through ChIP-seq analysis were performed using hippocampal nuclear extracts. Biotinylated oligonucleotides contained either an unmodified E-box motif (lanes 1-4), a methylated E-box motif (lanes 5-8), or a hydroxymethylated E-box motif (lanes 9-12). Black arrows indicate free probes, bands of interest (DNA::protein complexes) that are abrogated with cold competition, and supershifts. Bottom panel: band intensity quantifications of the area encompassing the bands of interest using ImageJ software. The E-box modifications (x-axis) and mean of band intensity (y-axis) are depicted +/- SEM. The methylated E-box motif (yellow) showed a trend towards significantly decreasing CLOCK binding compared to the unmodified (red) motif (# *P*-value = 0.058). The hydroxymethylated E-box motif (blue) significantly decreased binding compared to the unmodified E-box (* *P*-value <0.05).

Figure 4: A working model of a functional role for 5hmC in developmental brain disorder-like behaviors. Control *Cntnap2*$^{+/-}$ mice (left panel) 5hmC levels (orange lollipops) remain abundant and prevent CLOCK (blue semicircle) from binding to genes implicated in developmental brain disorders (*e.g., Gigyf1*), which results in adequate gene expression levels (black curved lines) and unaffected behaviors. *Cntnap2*$^{+/-}$ mice exposed to early life stress (ELS; lightning bolt; right panel) exhibit a loss of 5hmC abundance in genes implicated in developmental brain disorders (*e.g., Gigyf1*). This reduction in 5hmC levels allows CLOCK to bind to DNA and repress transcription, which results in developmental brain disorder-like phenotypes.

References

1 Parenti, I., Rabaneda, L. G., Schoen, H. & Novarino, G. Neurodevelopmental Disorders: From Genetics to Functional Pathways. *Trends in neurosciences* **43**, 608-621, doi:10.1016/j.tins.2020.05.004 (2020).

2 Sebat, J. *et al.* Strong association of de novo copy number mutations with autism. *Science (New York, N.Y.)* **316**, 445-449, doi:10.1126/science.1138659 (2007).

3 Glessner, J. T. *et al.* Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. *Nature* **459**, 569-573, doi:10.1038/nature07953 (2009).

4 Weiss, L. A., Arking, D. E., Daly, M. J. & Chakravarti, A. A genome-wide linkage and association scan reveals novel loci for autism. *Nature* **461**, 802-808, doi:10.1038/nature08490 (2009).

5 Tordjman, S. *et al.* Gene x Environment interactions in autism spectrum disorders: role of epigenetic mechanisms. *Frontiers in psychiatry* **5**, 53, doi:10.3389/fpsyt.2014.00053 (2014).

6 Conway, C. C., Rutter, L. A. & Brown, T. A. Chronic environmental stress and the temporal course of depression and panic disorder: A trait-state-occasion modeling approach. *Journal of abnormal psychology* **125**, 53-63, doi:10.1037/abn0000122 (2016).

7 Comes, A. L. *et al.* The role of environmental stress and DNA methylation in the longitudinal course of bipolar disorder. *International journal of bipolar disorders* **8**, 9, doi:10.1186/s40345-019-0176-6 (2020).

8 van den Bosch, M. & Meyer-Lindenberg, A. Environmental Exposures and Depression: Biological Mechanisms and Epidemiological Evidence. *Annual review of public health* **40**, 239-259, doi:10.1146/annurev-publhealth-040218-044106 (2019).

9 Bourvis, N., Aouidad, A., Cabelguen, C., Cohen, D. & Xavier, J. How Do Stress Exposure and Stress Regulation Relate to Borderline Personality Disorder? *Frontiers in psychology* **8**, 2054, doi:10.3389/fpsyg.2017.02054 (2017).

10 Heim, C. & Binder, E. B. Current research trends in early life stress and depression: review of human studies on sensitive periods, gene-environment interactions, and epigenetics. *Experimental neurology* **233**, 102-111, doi:10.1016/j.expneurol.2011.10.032 (2012).

11 Szulwach, K. E. *et al.* 5-hmC-mediated epigenetic dynamics during postnatal neurodevelopment and aging. *Nature neuroscience* **14**, 1607-1616, doi:10.1038/nn.2959 (2011).

12 Kriaucionis, S. & Heintz, N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science (New York, N.Y.)* **324**, 929-930, doi:10.1126/science.1169786 (2009).

13 Song, C. X. *et al.* Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. *Nature biotechnology* **29**, 68-72, doi:10.1038/nbt.1732 (2011).

14 Yao, B. & Jin, P. Cytosine modifications in neurodevelopment and diseases. *Cellular and molecular life sciences : CMLS* **71**, 405-418, doi:10.1007/s00018-013-1433-y (2014).

15 Al-Mahdawi, S., Virmouni, S. A. & Pook, M. A. The emerging role of 5-hydroxymethylcytosine in neurodegenerative diseases. *Frontiers in neuroscience*

8, 397, doi:10.3389/fnins.2014.00397 (2014).
16 Mellen, M., Ayata, P., Dewell, S., Kriaucionis, S. & Heintz, N. MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. *Cell* **151**, 1417-1430, doi:10.1016/j.cell.2012.11.022 (2012).
17 Zhubi, A. *et al.* Increased binding of MeCP2 to the GAD1 and RELN promoters may be mediated by an enrichment of 5-hmC in autism spectrum disorder (ASD) cerebellum. *Translational psychiatry* **4**, e349, doi:10.1038/tp.2013.123 (2014).
18 Wang, F. *et al.* Genome-wide loss of 5-hmC is a novel epigenetic feature of Huntington's disease. *Human molecular genetics* **22**, 3641-3653, doi:10.1093/hmg/ddt214 (2013).
19 Villar-Menéndez, I. *et al.* Increased 5-methylcytosine and decreased 5-hydroxymethylcytosine levels are associated with reduced striatal A2AR levels in Huntington's disease. *Neuromolecular medicine* **15**, 295-309, doi:10.1007/s12017-013-8219-0 (2013).
20 Chouliaras, L. *et al.* Consistent decrease in global DNA methylation and hydroxymethylation in the hippocampus of Alzheimer's disease patients. *Neurobiology of aging* **34**, 2091-2099, doi:10.1016/j.neurobiolaging.2013.02.021 (2013).
21 Choi, I., Kim, R., Lim, H. W., Kaestner, K. H. & Won, K. J. 5-hydroxymethylcytosine represses the activity of enhancers in embryonic stem cells: a new epigenetic signature for gene regulation. *BMC genomics* **15**, 670, doi:10.1186/1471-2164-15-670 (2014).
22 Green, B. B. *et al.* Hydroxymethylation is uniquely distributed within term placenta, and is associated with gene expression. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **30**, 2874-2884, doi:10.1096/fj.201600310R (2016).
23 Liu, H. *et al.* Altered 5-Hydroxymethylcytosine Landscape in Primary Gastric Adenocarcinoma. *DNA and cell biology* **38**, 1460-1469, doi:10.1089/dna.2019.4965 (2019).
24 Qin, L. *et al.* Ethnicity-specific and overlapping alterations of brain hydroxymethylome in Alzheimer's disease. *Human molecular genetics* **29**, 149-158, doi:10.1093/hmg/ddz273 (2020).
25 Papale, L. A. *et al.* Genome-wide disruption of 5-hydroxymethylcytosine in a mouse model of autism. *Human molecular genetics* **24**, 7121-7131, doi:10.1093/hmg/ddv411 (2015).
26 Trott, A. J. & Menet, J. S. Regulation of circadian clock transcriptional output by CLOCK:BMAL1. *PLoS genetics* **14**, e1007156, doi:10.1371/journal.pgen.1007156 (2018).
27 Lande-Diner, L., Boyault, C., Kim, J. Y. & Weitz, C. J. A positive feedback loop links circadian clock factor CLOCK-BMAL1 to the basic transcriptional machinery. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 16021-16026, doi:10.1073/pnas.1305980110 (2013).
28 Menet, J. S., Pescatore, S. & Rosbash, M. CLOCK:BMAL1 is a pioneer-like transcription factor. *Genes & development* **28**, 8-13, doi:10.1101/gad.228536.113 (2014).
29 Doi, M., Hirayama, J. & Sassone-Corsi, P. Circadian regulator CLOCK is a histone

acetyltransferase. *Cell* **125**, 497-508, doi:10.1016/j.cell.2006.03.033 (2006).
30 Yoshitane, H. *et al.* CLOCK-controlled polyphonic regulation of circadian rhythms through canonical and noncanonical E-boxes. *Molecular and cellular biology* **34**, 1776-1787, doi:10.1128/mcb.01465-13 (2014).
31 Blakeman, V., Williams, J. L., Meng, Q. J. & Streuli, C. H. Circadian clocks and breast cancer. *Breast cancer research : BCR* **18**, 89, doi:10.1186/s13058-016-0743-z (2016).
32 Refinetti, R. Focal Epilepsy and the Clock Gene. *Pediatric neurology briefs* **32**, 6, doi:10.15844/pedneurbriefs-32-6 (2018).
33 Valladares, M., Obregón, A. M. & Chaput, J. P. Association between genetic variants of the clock gene and obesity and sleep duration. *Journal of physiology and biochemistry* **71**, 855-860, doi:10.1007/s13105-015-0447-3 (2015).
34 Schuch, J. B., Genro, J. P., Bastos, C. R., Ghisleni, G. & Tovo-Rodrigues, L. The role of CLOCK gene in psychiatric disorders: Evidence from human and animal research. *American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics* **177**, 181-198, doi:10.1002/ajmg.b.32599 (2018).
35 Satterstrom, F. K. *et al.* Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. *Cell* **180**, 568-584.e523, doi:10.1016/j.cell.2019.12.036 (2020).
36 Garbett, K. *et al.* Immune transcriptome alterations in the temporal cortex of subjects with autism. *Neurobiology of disease* **30**, 303-311, doi:10.1016/j.nbd.2008.01.012 (2008).
37 Voineagu, I. & Eapen, V. Converging Pathways in Autism Spectrum Disorders: Interplay between Synaptic Dysfunction and Immune Responses. *Frontiers in human neuroscience* **7**, 738, doi:10.3389/fnhum.2013.00738 (2013).
38 Paulraj, P. *et al.* A Novel Homozygous Deletion within the FRY Gene Associated with Nonsyndromic Developmental Delay. *Cytogenetic and genome research* **159**, 19-25, doi:10.1159/000502598 (2019).
39 Schaafsma, S. M. *et al.* Sex-specific gene-environment interactions underlying ASD-like behaviors. *Proceedings of the National Academy of Sciences of the United States of America* **114**, 1383-1388, doi:10.1073/pnas.1619312114 (2017).
40 Penagarikano, O. *et al.* Absence of CNTNAP2 leads to epilepsy, neuronal migration abnormalities, and core autism-related deficits. *Cell* **147**, 235-246, doi:10.1016/j.cell.2011.08.040 (2011).
41 Lauber, E., Filice, F. & Schwaller, B. Dysregulation of Parvalbumin Expression in the Cntnap2-/- Mouse Model of Autism Spectrum Disorder. *Frontiers in molecular neuroscience* **11**, 262, doi:10.3389/fnmol.2018.00262 (2018).
42 Vogt, D. *et al.* Mouse Cntnap2 and Human CNTNAP2 ASD Alleles Cell Autonomously Regulate PV+ Cortical Interneurons. *Cerebral cortex (New York, N.Y. : 1991)* **28**, 3868-3879, doi:10.1093/cercor/bhx248 (2018).
43 Báez-Mendoza, R. & Schultz, W. The role of the striatum in social behavior. *Frontiers in neuroscience* **7**, 233, doi:10.3389/fnins.2013.00233 (2013).
44 Sano, H., Nagai, Y., Miyakawa, T., Shigemoto, R. & Yokoi, M. Increased social interaction in mice deficient of the striatal medium spiny neuron-specific phosphodiesterase 10A2. *Journal of neurochemistry* **105**, 546-556,

doi:10.1111/j.1471-4159.2007.05152.x (2008).
45 Kohls, G., Yerys, B. E. & Schultz, R. T. Striatal development in autism: repetitive behaviors and the reward circuitry. *Biological psychiatry* **76**, 358-359, doi:10.1016/j.biopsych.2014.07.010 (2014).
46 Jacobson, L. & Sapolsky, R. The role of the hippocampus in feedback regulation of the hypothalamic-pituitary-adrenocortical axis. *Endocrine reviews* **12**, 118-134, doi:10.1210/edrv-12-2-118 (1991).
47 Jankord, R. & Herman, J. P. Limbic regulation of hypothalamo-pituitary-adrenocortical function during acute and chronic stress. *Annals of the New York Academy of Sciences* **1148**, 64-73, doi:10.1196/annals.1410.012 (2008).
48 Phillips, M. L., Robinson, H. A. & Pozzo-Miller, L. Ventral hippocampal projections to the medial prefrontal cortex regulate social memory. *eLife* **8**, doi:10.7554/eLife.44182 (2019).
49 Felix-Ortiz, A. C. & Tye, K. M. Amygdala inputs to the ventral hippocampus bidirectionally modulate social behavior. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **34**, 586-595, doi:10.1523/jneurosci.4257-13.2014 (2014).
50 Kalman, E. & Keay, K. A. Hippocampal volume, social interactions, and the expression of the normal repertoire of resident-intruder behavior. *Brain and behavior* **7**, e00775, doi:10.1002/brb3.775 (2017).
51 Han, S., Tai, C., Jones, C. J., Scheuer, T. & Catterall, W. A. Enhancement of inhibitory neurotransmission by GABAA receptors having α2,3-subunits ameliorates behavioral deficits in a mouse model of autism. *Neuron* **81**, 1282-1289, doi:10.1016/j.neuron.2014.01.016 (2014).
52 Folsom, T. D. & Fatemi, S. H. The involvement of Reelin in neurodevelopmental disorders. *Neuropharmacology* **68**, 122-135, doi:10.1016/j.neuropharm.2012.08.015 (2013).
53 Curran, S., Ahn, J. W., Grayton, H., Collier, D. A. & Ogilvie, C. M. NRXN1 deletions identified by array comparative genome hybridisation in a clinical case series - further understanding of the relevance of NRXN1 to neurodevelopmental disorders. *Journal of molecular psychiatry* **1**, 4, doi:10.1186/2049-9256-1-4 (2013).
54 Rademacher, S. & Eickholt, B. J. PTEN in Autism and Neurodevelopmental Disorders. *Cold Spring Harbor perspectives in medicine* **9**, doi:10.1101/cshperspect.a036780 (2019).
55 Tran, K. A., Dillingham, C. M. & Sridharan, R. Coordinated removal of repressive epigenetic modifications during induced reversal of cell identity. *The EMBO journal* **38**, e101681, doi:10.15252/embj.2019101681 (2019).
56 Héberlé, É. & Bardet, A. F. Sensitivity of transcription factors to DNA methylation. *Essays in biochemistry* **63**, 727-741, doi:10.1042/ebc20190033 (2019).
57 Hu, S. *et al.* DNA methylation presents distinct binding sites for human transcription factors. *eLife* **2**, e00726, doi:10.7554/eLife.00726 (2013).
58 Jin, J. *et al.* The effects of cytosine methylation on general transcription factors. *Scientific reports* **6**, 29119, doi:10.1038/srep29119 (2016).
59 Griswold, M. D. & Kim, J. S. Site-specific methylation of the promoter alters deoxyribonucleic acid-protein interactions and prevents follicle-stimulating hormone receptor gene transcription. *Biology of reproduction* **64**, 602-610,

doi:10.1095/biolreprod64.2.602 (2001).
60 Bourgeron, T. The possible interplay of synaptic and clock genes in autism spectrum disorders. *Cold Spring Harbor symposia on quantitative biology* **72**, 645-654, doi:10.1101/sqb.2007.72.020 (2007).
61 Wu, Z. Y. *et al.* Autism spectrum disorder (ASD): Disturbance of the melatonin system and its implications. *Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie* **130**, 110496, doi:10.1016/j.biopha.2020.110496 (2020).
62 Saha, S., Singh, K. M. & Gupta, B. B. P. Melatonin synthesis and clock gene regulation in the pineal organ of teleost fish compared to mammals: Similarities and differences. *General and comparative endocrinology* **279**, 27-34, doi:10.1016/j.ygcen.2018.07.010 (2019).
63 Yang, Y., Li, N., Qiu, J., Ge, H. & Qin, X. Identification of the Repressive Domain of the Negative Circadian Clock Component CHRONO. *International journal of molecular sciences* **21**, doi:10.3390/ijms21072469 (2020).
64 Michael, A. K. *et al.* Formation of a repressive complex in the mammalian circadian clock is mediated by the secondary pocket of CRY1. *Proceedings of the National Academy of Sciences of the United States of America* **114**, 1560-1565, doi:10.1073/pnas.1615310114 (2017).
65 Cao, X., Yang, Y., Selby, C. P., Liu, Z. & Sancar, A. Molecular mechanism of the repressive phase of the mammalian circadian clock. *Proceedings of the National Academy of Sciences of the United States of America* **118**, doi:10.1073/pnas.2021174118 (2021).
66 Mueller, B. R. & Bale, T. L. Impact of prenatal stress on long term body weight is dependent on timing and maternal sensitivity. *Physiology & behavior* **88**, 605-614, doi:10.1016/j.physbeh.2006.05.019 (2006).
67 Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology* **10**, R25, doi:10.1186/gb-2009-10-3-r25 (2009).
68 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).
69 Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics (Oxford, England)* **26**, 139-140, doi:10.1093/bioinformatics/btp616 (2010).

CHAPTER 5

Summary, Future Directions, and Concluding Thoughts

SUMMARY

Here, through my doctoral work, I interrogated genome-wide disruptions of a relatively novel epigenetic mark, 5-hydroxymethylcytosine (5hmC), beginning with characterizing its sex-specific distribution in a mouse model of autism. Next, I contextualized 5hmC as a molecular component in a gene by environment interaction mouse model of developmental brain disorders (DBDs), bolstering support for epigenetic mechanisms as an underlying cause of DBDs. Experiments presented in a subsequent chapter provided foundational and direct evidence that 5hmC abundance is sufficient to disrupt transcription factor (TF) binding, corroborating this DNA methylation mark as a molecular component of transcriptional regulation involved in the development of DBD-like phenotypes. Further defining molecular function(s) of 5hmC will likely reveal its clinical and therapeutic potential, leading to novel, more precise and personalized drug therapies. Together, this research has highlighted an emerging role of 5hmC in DBDs.

Chapter summaries:

5hmC sex-specific landscapes in a mouse model of autism

- 5hmC profiles derived from striatal tissue of female mice with a knockout of *Cntnap2* were contrasted against wildtype female mice
- Genomic loci throughout the entire genome were found to display significant differences in 5hmC levels between mutant and wildtype mice
- Genes found to harbor mutant-related changes in epigenomic landscapes were associated with neuron- and synaptic-related functions and pathways implicated in autism

- A significant number of genes exhibiting mutant-related differences in 5hmC were linked to DBDs and were shown to have similar profiles in both male and female mutant mice
- Mutant male mice were found to harbor a larger number of molecular alterations than female mice, with greater parallels with autism-related molecular deficits in humans
- Sequence analysis associated to regions displaying differential 5hmC landscapes were enriched for motifs associated to TFs that previously have been found to play crucial roles in neurodevelopment and autism
- Together, these investigations provide substantial sex-specific evidence that 5hmC is an epigenetic contributor in the etiology of neurodevelopmental disorders, potentially through altering TF binding/expression of genes contributing to neurodevelopment

Long-lasting aberrations of 5hmC in the hippocampus of a gene by environment interaction mouse model of DBDs

- Early-life stress in mice with a partial genetic mutation (*Cntnap2*$^{+/-}$) resulted in behavioral deficits associated with DBDs, but not in mice lacking genetic aberrations
- Genomic profiling of the hippocampus of these mice identified regional disruptions of 5hmC associated to stress alone, and as a result of the gene by environment interaction

- A significant number of orthologs associated with human DBDs exhibited long-lasting changes in 5hmC abundance
- Gene clusters fostering 5hmC alterations were associated with neurodevelopmentally-related biological processes
- Sequence motif analysis of genomic regions displaying differential 5hmC showed enrichments of TFs with known functions in neurodevelopment (*e.g.,* CLOCK)
- These interrogations support a role of 5hmC as a molecular contributor in long-standing behavioral deficits associated with DBDs, suggesting this GxE model as a possible etiology of DBDs and bolsters the importance of 5hmC as a molecular component in neurodevelopmental disorders

5hmC functions in modulating CLOCK binding affinity

- RNA-seq and western blot analyses indicated that RNA and protein expression levels were not significantly differ between genotypes (*i.e.,* WT and HET), nor between experimental conditions (*i.e.,* controls and exposure to early-life stress)
- Chromatin immunoprecipitation coupled with high-throughput sequencing (ChIP-seq) identified 19 regions among the GxE DhMRs that displayed altered CLOCK binding
- Three of these 19 regions with altered CLOCK binding and 5hmC abundances overlapped with genes that also were differentially expressed, including *Fry, Gigyf1*, and *Palld*
- In vitro studies showed that CLOCK binds to the E-box motifs located in regions of *Fry*, *Gigyf1*, and *Palld*, which also harbored DhMRs and were differentially bound by CLOCK in the GxE model

- *In vitro* studies showed that the presence of a 5hmC modification significantly reduced CLOCK binding, while the presence of a 5mC modification showed a near significant decrease in CLOCK binding.
- These studies provide crucial evidence of 5hmC involvement in TF binding affinity in genes associated with DBD phenotypes. Furthermore, these findings underscore 5hmC as a modulator to the transcription machinery, adding another component to the complex process of transcriptional regulation

Taken together, my doctoral research provides evidence of a functional role for 5hmC in the etiology of DBDs through influencing affinitive binding of TFs critical in neurodevelopment, such as CLOCK. My studies also show the importance of 5hmC in modulating the transcription machinery which may have phenotypic consequences across the span of neurodevelopment, supporting 5hmC as a therapeutic target in the diagnosis and amelioration of DBDs such as ASD.

Future Directions

I wish to continue these studies to further understand the mechanistic importance of 5hmC on gene expression dysregulation and diversity and long-range chromatin interactions associated with DBDs. My thesis research posits the following questions:

Do 5hmC modulations directly influence the expression of underlying genes?

While the ChIP-seq experiments performed in Chapter 4 revealed that a loss of 5hmC (*i.e.,* hypo-DhMRs) resulted in an increase binding of CLOCK to the DhMRs in *Fry, Gigyf1,* and *Palld*, and EMSAs showed that the presence of 5hmC hindered the binding of CLOCK

to a DNA sequence containing an E-box motif, this incident may be necessary but not be sufficient to result in altered transcription levels. Additional *in vitro* functional assays can directly test if 5hmC is sufficient to alter gene expression using a reporter construct, such as a luciferase assay. Other labs have found success at incorporating 5mC into luciferase reporter vectors by using synthetic double-stranded synthetic DNA, M.SssI CpG methyltransferase, and ethanol precipitation to methylate DNA[1]. Similar methods could be used to incorporate 5hmC into a reporter vector. For example, differentially hydroxymethylated regions from genes of interest (*e.g., Fry*, *Gigyf1*, and *Palld*) could be synthesized, treated with M.SssI CpG methyltransferase to methylate DNA, and followed by treatment with recombinant TET to oxidize methylated DNA to 5hmC in reporter vectors. Notably, previous studies investigating **5mC** alterations associated with DBDs in human patients have been limited in finding positive results linked with transcriptional variations[2,3]. While these findings suggests that 5mC may not influence the transcription of genes in close proximity, long-range effects of DNA methylation on the expression of distant developmentally important genes cannot be captured using common approaches to link epigenetic modifications to gene expression. Upon identifying sex-specific differences in 5hmC in a mouse model of autism (Chapter 2), genome-wide alterations of 5hmC in a gene by environment interaction mouse model of DBDs (Chapter 3), and differential binding of CLOCK caused by 5hmC (Chapter 4) the timing is ideal to determine the full function(s) and importance of 5hmC within the CNS related to gene regulation.

Does 5hmC influence long-range DNA::DNA interactions?

Studies from various labs interrogating 5hmC abundance and gene expression have found modest simultaneous overlaps between differential 5hmC and RNA levels, usually finding at most ~20% of differentially expressed genes contain DhMRs. It is hard to accept that ~80% of 5hmC changes would have no function. Yet, these relationships are only characterizing when 5hmC may be directly influencing expression of genes that reside proximally to a DhMR, which is only the *cis*-mode of action that 5hmC may have on gene expression levels. Indeed, DNA is a three-dimensional structure and stretches of DNA that are millions of basepairs apart, or on completely different chromosomes, may interact with each other. As such, these long-range *trans*-acting interactions of DNA that result in differential gene expression may, in part, be mediated by 5hmC that are not captured solely by limiting associations of 5hmC to *cis*-genes in proximal regions. With the development of chromatin conformation capture (*i.e.,* Hi-C) technologies, long-range interactions of DNA can now be sequenced and studied[4], and can be coupled with an antibody specific to 5hmC to specifically pull-down regions of the genome exhibiting 5hmC levels to study their long-range interactions. Understanding and exploring the long-range effects of 5hmC can shed light on how 5hmC dysregulates gene expression linked to DBDs, especially considering that the majority of DBD causes remain unknown.

Are transcriptional variants a result of 5hmC?

Transcriptional variations previously have been reported in DBDs[5-7], and understanding how these variants are produced, such as through mediation by 5hmC levels, can aid in understanding the complex and elusive etiology of these disorders. Previous work from our lab, and from other labs characterizing 5hmC changes, have identified an enrichment

of 5hmC alterations overlapping or near exon-intron boundaries[8,9]. Moreover, coupled with RNAseq data, DhMRs have been shown to overlap with dysregulated transcriptional variants (*i.e.,* isoforms), suggesting that 5hmC may be a mark for the inclusion or exclusion of exons/introns by the spliceosome. Further underscoring this possibility of 5hmC as a component in transcriptional diversity is that 5hmC has been identified on mRNA transcripts[10,11], meaning it is not solely confined to DNA. Other labs have reported Tet-mediated 5hmC modifications on RNA in the promotion of infection-induced myelopoiesis by mRNA oxidation[12], in the regulation of endogenous retroviruses[13], and on mRNAs in key pluripotency-related transcripts[10]. However, the function and effects of 5hmC incorporation on RNAs remains unknown. I wish to explore the relation of 5hmC and isoform production by using high-affinity capture/chemical label assays for 5hmC at both the DNA and RNA level, and identify if there are significant overlaps between 5hmC abundance on DNA that is mirrored at the RNA level of transcriptional variants. For example, a combination of hydroxymethylated DNA immunoprecipitation sequencing (hMeDIP-seq)[14], hydroxymethylated RNA immunoprecipitation sequencing (hMeRIP-seq)[15], and mRNA sequencing could be employed to test if 5hmC correlates with changes in isoform production. These assays would identify regions of the genome (*i.e.,* DNA) that display 5hmC abundances, identify transcripts (*i.e.,* RNA) that exhibit 5hmC markings, and identify any differential isoform expression, respectively. Together, these data could corroborate if 5hmC presence on DNA is also found on transcripts that undergo differential splicing, suggesting that 5hmC influences isoform production. To investigate a more mechanistic action of 5hmC in transcript diversity associated with DBDs, RNA-immunoprecipitation sequencing (RIP-seq) assays immunoprecipitating for components

of the spliceosome, derived from wild-type and *Cntnap2* homozygous knockout mice (*i.e.,* murine model of ASD), could be used to identify differential binding of the spliceosome that may overlap with DhMRs, suggesting involvement of 5hmC in alternative splicing.

Considerations

No experimental framework is without its inherent limitations and considerations. Indeed, shortcomings stem from the experiments performed here in this thesis work. For example, for the functional assays (Chapter 4) I focused solely on information derived from hippocampal tissue because there is a breadth of previous data supporting molecular alterations in the hippocampus in mouse models of ASD[16-18], deformations/morphological changes of the hippocampus in several DBDs[19-21], cellular/neuronal abnormalities in the hippocampus of ASD children[22], and its association with sociability and other behaviors[23-26]. Moreover, the hippocampus is strongly associated with the main stress-response pathway, the hypothalamic-pituitary-adrenal (HPA-axis), and previous work from our lab has shown molecular alterations in the hippocampus following acute and chronic stresses[27,28], further corroborating a putative role of the hippocampus as a modulator of DBD-like behaviors observed in the gene by environment interaction employed in the studies here. Together, these data underscore the neurodevelopmental impact of modulations on the hippocampus stemming from genetic mutations and environmental influences, bolstering this tissue as a strong candidate to investigate for my thesis experiments. While there is extensive research corroborating hippocampal involvement in the development of DBDs, it is only one candidate tissue contributing to ASD-related

outcomes. Others have reported cellular, molecular, and systemic perturbations associated with ASD in various other tissues of the central nervous system, such as the cerebellum, prefrontal cortex, motor cortex, striatum, and the amygdala[29-33]. Together, these data suggest that, in the case of ASD, there is no one “hub” or central tissue contributing to the development of ASD. Rather, there is an extensive involvement of the whole-brain. Experiments described in previous chapters provide an invaluable foundation for future research to investigate other tissues to study tissue-specific alterations and functions of 5hmC.

A vital component of my thesis work involved the selection of a transcription factor to study for functional assays (Chapter 4). Here, I selected the transcription factor CLOCK, which is best known to function in the regulation of the circadian rhythm. While CLOCK is highly expressed in hippocampal tissue, the hippocampus is not the primary tissue controlling cyclic and rhythmic circadian entrainment. In the central nervous system, the suprachiasmatic nucleus (SCN) is the master pacemaker of the biological clock[34]. While experiments performed in my thesis were not directly testing differences in circadian rhythms following gene by environment interactions, by studying CLOCK, there is an inherent assumption that these experiments involve circadian regulation. Therefore, the SCN would be an appropriate tissue to study, considering its strong association with circadian rhythm, and the dysregulation of circadian entrainment in several DBDs[34,35]. Moreover, while the chronic variable stress paradigm used in my experiments can be construed as a disruption of the circadian cycle in mice, I did not investigate any alterations in circadian-controlled behaviors in either the pregnant dams, nor in the early-

life stressed offspring, which would have supported the claim that this stress paradigm is a true disruptor of the endogenous molecular clock. Performing experiments such as electroencephalograph (EEG) during light cycles in mice could have identified any abnormal activity in response in the stress, which would suggest a dysregulation of the circadian cycle. Thus, while experiments performed here involved CLOCK, they did not investigate any circadian differences caused by the gene by environment interaction, which could have increased the reasoning of finding differentially bound CLOCK in this model.

Another consideration of the studies presented here is the timepoint at which molecular studies were performed. To study long-lasting effects from the gene by environment interaction, mice were experimentally studied in adulthood, at three months of age. Considering that the stress paradigm was administered prenatally, studying changes at three months of age seems counterintuitive, as most molecular reprogramming would be hypothesized to fluctuate immediately following the stress. However, the studies employed here aimed to study and identify stable modifications that persist to adulthood. To gain a deeper and more thorough understanding of the molecular modulations of 5hmC and gene expression as a consequence of the gene by environment interaction several timepoints should be interrogated, such as immediately after birth (P0), at weaning (P21), at one month of age (early-adulthood; P30), and at three months of age (P90). A timepoint at P0 can be used to determine the immediate effects of the early-life stress, as the week-long stress paradigm ends right before birth. A timepoint at P21 in mice correlates with ~2-3 years of age in humans, the age at which DBD symptoms begin

to clearly emerge[36]. At this age, increased rates of myelination, increased synaptic density, and increased activity and sociability are present in mice[37-39]. Moreover, epigenetic machinery, such as DNA methyltransferases have a marked decrease in expression in the hippocampus at this timepoint[40], indicating a shift in epigenetic regulation that may show abnormalities following the GxE interaction that can be pinpointed by investigating mice at P21. Sexual maturity and early-adulthood begins approximately at one month of age (P30) in mice, and correlates with ~12-16 years of age in humans. Adolescence, both in rodents and humans, is a period of ongoing refinement and maturation of neural circuitry, coupled with high rates of synaptic pruning[41]. Synaptic pruning during adolescence is partially dependent on GABAergic signaling[41], which was identified to be perturbed in *Cntnap2*$^{-/-}$ mutant mice[16], and may, likewise, be dysregulated following the GxE interaction. Therefore, investigating 5hmC profiles and functions at P30 overlaps with a developmental window crucial in neuronal maturation that is altered in DBDs. These timepoints would allow the opportunity to differentiate what alterations are long-lasting from those that are transient. Notably, the primary goal of using and establishing the gene by environment interaction in Chapter 3 was to determine if any behavioral alterations and molecular perturbations mirrored those of the *Cntnap2* homozygous knockout mice, which were studied at 3 months of age. Identifying long-lasting behavioral and molecular changes at three months of age is of much value, as it supports that our gene by environment interaction is significantly altering pathways and processes in the central nervous system that persist throughout neurodevelopment. Despite this late cross-sectional timepoint used in these studies, there is still much more that we can learn from studying other timepoints throughout

neurodevelopment to gain a broader understanding of the molecular timing of the reprogramming that occurs and shapes nervous system development, and how alterations in these processes can lead to DBDs.

It is worth noting that only female mice were used for molecular studies, as only female mice exhibited behavioral changes following the gene by environment interaction. Notably there also were molecular changes in female wild-type mice subjected to the early-life stress paradigm, though they lacked behavioral changes in adulthood. Thus, it is likely that there also would be molecular alterations in male GxE mice, though they also did not display abnormal behaviors in adulthood. Considering that human males have a nearly three-fold predisposition to develop autism, compared to females, it is counterintuitive that female mice subjected to this gene by environment interaction display autistic-like social deficits, while males do not. One reason for this could be the timing of the early-life stress. The early-life stress paradigm was administered during the third trimester of pregnancy in mice (E12-18). During this developmental window, the central nervous system undergoes extensive changes, such as neuronal migration, neuronal apoptosis, synaptogenesis and myelination[42-44]. All of these processes are essential for proper CNS development and, moreover, show sex-specific differences in their regulation[45-47]. Previous studies have identified an increased susceptibility to later phenotypic outcomes in females following exposure to adversities during the third trimester of pregnancy[48]. Therefore, the timepoint chosen in the gene by environment interaction described here, largely selected due to the overlapping onset of *Cntnap2* expression (E14), may be a significant contributing factor to the sex-specific outcomes observed here. Prenatal stress

using this paradigm during other prenatal timeframes, such as during the first trimester of pregnancy, may lead to varying outcomes in offspring due to inherent prenatal sex-specific differences in susceptibilities[49,50]. Thus, further investigations are warranted as these findings suggest gene by environment interactions are a cause of DBDs, and support that the timing of environmental insults lead to sex-specific pathogeneses in DBDs. However, as behavioral and molecular studies were only performed at three months of age, when male mice and wild-type mice did not display behavioral changes, this does not rule out the possibility that abnormal behaviors exist in these groups prior to three months of age. Together, the cross-sectional investigations presented in this thesis provide a foundational snapshot of molecular underpinnings that may drive differences in adult phenotypic outcomes, but may not provide the whole story.

Concluding Thoughts

While experiments performed in my doctoral thesis work provide important foundational evidence for 5hmC involvement in DBD etiology, these studies barely scratch the tip of the iceberg in fully unravelling the true nature and potential of 5hmC in the context of its role in transcription and as a molecular component with clinical implications in human health. A direct role in transcription by modulating TF binding is only one function that 5hmC may harbor. Other possible functions for which evidence is currently available for include: influencing transcript diversity, altering DNA::DNA interactions, and influencing the rate of translation of mRNA. Indeed, experiments in this thesis provide the basis and the framework for unveiling other functions for 5hmC as a component of DBD etiology.

The current dogma of DNA methylation is that methylated genes are repressed, while unmethylated/oxidized genes are expressed; that is to say, 5mC is a repressive mark and 5hmC has the opposite effect. However, recent studies have provided strong evidence that the function of 5mC, and to some extent 5hmC, is highly context-specific and is more complex than previously thought[51]. For example, in Chapter 4 we investigated the genes *Fry, Gigyf1*, and *Palld* and found evidence that 5mC nominally decreased CLOCK binding, while 5hmC significantly decreased CLOCK binding. Together, this finding would suggest that 5mC and 5hmC can act as repressive marks to CLOCK binding to DNA of genes (*e.g., Gigyf1*) that are linked to DBD. However, previous findings indicate that DNA methylation (*i.e.,* 5mC) has the capacity to increase binding of transcription factors[52], suggesting that the context of DNA methylation modifications can have varying and dynamic impacts on protein binding to DNA. Moreover, another study reported that site-specific methylation at various cytosines within the same short span on DNA of the promoter region of the *FSHR* gene either promoted or prevented protein complexes to bind the DNA, further highlighting the context-specific and locus-specific modulations that DNA methylation can have on TF binding[53]. Together, these data suggest that the type of modification (*e.g.,* 5mC and/or 5hmC) and the specific genetic locus being altered determine the effect of these epigenetic signatures. By showing that 5mC and 5hmC exhibit alterations in TF binding affinity through my thesis research is a key step towards unravelling the means by which neuronal cells can quickly and efficiently modulate gene expression in response to an insult, and to understand how deviations from these mechanisms can result in aberrations of neuronal function, potentially leading to DBDs.

Defining a function of 5hmC in the context of DBD etiology is a new and growing avenue of research. Current studies of 5hmC have primarily been confined to bulk tissue analysis; however, I expect that in the coming years disruptions of 5hmC commonly will be examined in a cell-specific manner, allowing researchers to study 5hmC at a higher resolution, as bulk tissue cannot account for confounding differences in cell population heterogeneity. Being able to identify specific cell populations that exhibit changes in 5hmC may be vital in teasing apart the complex molecular etiologies of DBDs. Future cell-specific examinations of genome-wide 5hmC in specific neural circuits that contribute to behavioral outcomes will provide an objective marker for a much quicker more reliable diagnosis that will lead to personalized treatments for those at high risk of developing DBD.

References

1 Zhu, Z. *et al.* DNA hypomethylation of a transcription factor binding site within the promoter of a gout risk gene NRBP1 upregulates its expression by inhibition of TFAP2A binding. *Clinical epigenetics* **9**, 99, doi:10.1186/s13148-017-0401-z (2017).

2 Wong, C. C. Y. *et al.* Genome-wide DNA methylation profiling identifies convergent molecular signatures associated with idiopathic and syndromic autism in post-mortem human brain tissue. *Human molecular genetics* **28**, 2201-2211, doi:10.1093/hmg/ddz052 (2019).

3 Stathopoulos, S. *et al.* DNA Methylation Associated with Mitochondrial Dysfunction in a South African Autism Spectrum Disorder Cohort. *Autism research : official journal of the International Society for Autism Research* **13**, 1079-1093, doi:10.1002/aur.2310 (2020).

4 Belton, J. M. *et al.* Hi-C: a comprehensive technique to capture the conformation of genomes. *Methods (San Diego, Calif.)* **58**, 268-276, doi:10.1016/j.ymeth.2012.05.001 (2012).

5 Porter, R. S., Jaamour, F. & Iwase, S. Neuron-specific alternative splicing of transcriptional machineries: Implications for neurodevelopmental disorders. *Molecular and cellular neurosciences* **87**, 35-45, doi:10.1016/j.mcn.2017.10.006 (2018).

6 Fliedner, A. *et al.* Variants in SCAF4 Cause a Neurodevelopmental Disorder and Are Associated with Impaired mRNA Processing. *American journal of human genetics* **107**, 544-554, doi:10.1016/j.ajhg.2020.06.019 (2020).

7 Doorenweerd, N. *et al.* Timing and localization of human dystrophin isoform expression provide insights into the cognitive phenotype of Duchenne muscular dystrophy. *Scientific reports* **7**, 12575, doi:10.1038/s41598-017-12981-5 (2017).

8 Feng, J. *et al.* Role of Tet1 and 5-hydroxymethylcytosine in cocaine action. *Nature neuroscience* **18**, 536-544, doi:10.1038/nn.3976 (2015).

9 Papale, L. A. *et al.* Sex-specific hippocampal 5-hydroxymethylcytosine is disrupted in response to acute stress. *Neurobiology of disease* **96**, 54-66, doi:10.1016/j.nbd.2016.08.014 (2016).

10 Lan, J. *et al.* Functional role of Tet-mediated RNA hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature communications* **11**, 4956, doi:10.1038/s41467-020-18729-6 (2020).

11 Xu, Q. *et al.* IDH1/2 Mutants Inhibit TET-Promoted Oxidation of RNA 5mC to 5hmC. *PloS one* **11**, e0161261, doi:10.1371/journal.pone.0161261 (2016).

12 Shen, Q. *et al.* Tet2 promotes pathogen infection-induced myelopoiesis through mRNA oxidation. *Nature* **554**, 123-127, doi:10.1038/nature25434 (2018).

13 Guallar, D. *et al.* RNA-dependent chromatin targeting of TET2 for endogenous retrovirus control in pluripotent stem cells. *Nature genetics* **50**, 443-451, doi:10.1038/s41588-018-0060-9 (2018).

14 Tan, L. *et al.* Genome-wide comparison of DNA hydroxymethylation in mouse embryonic stem cells and neural progenitor cells by a new comparative hMeDIP-seq method. *Nucleic acids research* **41**, e84, doi:10.1093/nar/gkt091 (2013).

15 Delatte, B. *et al.* RNA biochemistry. Transcriptome-wide distribution and function of RNA hydroxymethylcytosine. *Science (New York, N.Y.)* **351**, 282-285, doi:10.1126/science.aac5253 (2016).
16 Penagarikano, O. *et al.* Absence of CNTNAP2 leads to epilepsy, neuronal migration abnormalities, and core autism-related deficits. *Cell* **147**, 235-246, doi:10.1016/j.cell.2011.08.040 (2011).
17 Chung, L., Bey, A. L. & Jiang, Y. H. Synaptic plasticity in mouse models of autism spectrum disorders. *The Korean journal of physiology & pharmacology : official journal of the Korean Physiological Society and the Korean Society of Pharmacology* **16**, 369-378, doi:10.4196/kjpp.2012.16.6.369 (2012).
18 Provenzano, G. *et al.* Comparative Gene Expression Analysis of Two Mouse Models of Autism: Transcriptome Profiling of the BTBR and En2 (-/-) Hippocampus. *Frontiers in neuroscience* **10**, 396, doi:10.3389/fnins.2016.00396 (2016).
19 Dager, S. R. *et al.* Shape mapping of the hippocampus in young children with autism spectrum disorder. *AJNR. American journal of neuroradiology* **28**, 672-677 (2007).
20 Richards, R. *et al.* Increased hippocampal shape asymmetry and volumetric ventricular asymmetry in autism spectrum disorder. *NeuroImage. Clinical* **26**, 102207, doi:10.1016/j.nicl.2020.102207 (2020).
21 Chaddad, A., Desrosiers, C., Hassan, L. & Tanougast, C. Hippocampus and amygdala radiomic biomarkers for the study of autism spectrum disorder. *BMC neuroscience* **18**, 52, doi:10.1186/s12868-017-0373-0 (2017).
22 Belmonte, M. K. *et al.* Autism and abnormal development of brain connectivity. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **24**, 9228-9231, doi:10.1523/jneurosci.3340-04.2004 (2004).
23 Phillips, M. L., Robinson, H. A. & Pozzo-Miller, L. Ventral hippocampal projections to the medial prefrontal cortex regulate social memory. *eLife* **8**, doi:10.7554/eLife.44182 (2019).
24 Felix-Ortiz, A. C. & Tye, K. M. Amygdala inputs to the ventral hippocampus bidirectionally modulate social behavior. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **34**, 586-595, doi:10.1523/jneurosci.4257-13.2014 (2014).
25 Kalman, E. & Keay, K. A. Hippocampal volume, social interactions, and the expression of the normal repertoire of resident-intruder behavior. *Brain and behavior* **7**, e00775, doi:10.1002/brb3.775 (2017).
26 Kim, Y. S. *et al.* Altered cerebellar development in nuclear receptor TAK1/ TR4 null mice is associated with deficits in GLAST(+) glia, alterations in social behavior, motor learning, startle reactivity, and microglia. *Cerebellum (London, England)* **9**, 310-323, doi:10.1007/s12311-010-0163-z (2010).
27 Jacobson, L. & Sapolsky, R. The role of the hippocampus in feedback regulation of the hypothalamic-pituitary-adrenocortical axis. *Endocrine reviews* **12**, 118-134, doi:10.1210/edrv-12-2-118 (1991).
28 Jankord, R. & Herman, J. P. Limbic regulation of hypothalamo-pituitary-adrenocortical function during acute and chronic stress. *Annals of the New York Academy of Sciences* **1148**, 64-73, doi:10.1196/annals.1410.012 (2008).

29 Wang, S. S., Kloth, A. D. & Badura, A. The cerebellum, sensitive periods, and autism. *Neuron* **83**, 518-532, doi:10.1016/j.neuron.2014.07.016 (2014).
30 Courchesne, E. & Pierce, K. Why the frontal cortex in autism might be talking only to itself: local over-connectivity but long-distance disconnection. *Current opinion in neurobiology* **15**, 225-230, doi:10.1016/j.conb.2005.03.001 (2005).
31 Nebel, M. B. *et al.* Disruption of functional organization within the primary motor cortex in children with autism. *Human brain mapping* **35**, 567-580, doi:10.1002/hbm.22188 (2014).
32 Fuccillo, M. V. Striatal Circuits as a Common Node for Autism Pathophysiology. *Frontiers in neuroscience* **10**, 27, doi:10.3389/fnins.2016.00027 (2016).
33 Amaral, D. G. & Corbett, B. A. The amygdala, autism and anxiety. *Novartis Foundation symposium* **251**, 177-187; discussion 187-197, 281-197 (2003).
34 Mieda, M. The Network Mechanism of the Central Circadian Pacemaker of the SCN: Do AVP Neurons Play a More Critical Role Than Expected? *Frontiers in neuroscience* **13**, 139, doi:10.3389/fnins.2019.00139 (2019).
35 Pinato, L., Galina Spilla, C. S., Markus, R. P. & da Silveira Cruz-Machado, S. Dysregulation of Circadian Rhythms in Autism Spectrum Disorders. *Current pharmaceutical design* **25**, 4379-4393, doi:10.2174/1381612825666191102170450 (2019).
36 Lord, C. *et al.* Autism from 2 to 9 years of age. *Archives of general psychiatry* **63**, 694-701, doi:10.1001/archpsyc.63.6.694 (2006).
37 Semple, B. D., Blomgren, K., Gimlin, K., Ferriero, D. M. & Noble-Haeusslein, L. J. Brain development in rodents and humans: Identifying benchmarks of maturation and vulnerability to injury across species. *Progress in neurobiology* **106-107**, 1-16, doi:10.1016/j.pneurobio.2013.04.001 (2013).
38 Aranmolate, A., Tse, N. & Colognato, H. Myelination is delayed during postnatal brain development in the mdx mouse model of Duchenne muscular dystrophy. *BMC neuroscience* **18**, 63, doi:10.1186/s12868-017-0381-0 (2017).
39 Fairless, A. H. *et al.* Sociability and brain development in BALB/cJ and C57BL/6J mice. *Behavioural brain research* **228**, 299-310, doi:10.1016/j.bbr.2011.12.001 (2012).
40 Simmons, R. K., Stringfellow, S. A., Glover, M. E., Wagle, A. A. & Clinton, S. M. DNA methylation markers in the postnatal developing rat brain. *Brain research* **1533**, 26-36, doi:10.1016/j.brainres.2013.08.005 (2013).
41 Afroz, S., Parato, J., Shen, H. & Smith, S. S. Synaptic pruning in the female hippocampus is triggered at puberty by extrasynaptic GABAA receptors on dendritic spines. *eLife* **5**, doi:10.7554/eLife.15106 (2016).
42 Malik, S. *et al.* Neurogenesis continues in the third trimester of pregnancy and is suppressed by premature birth. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **33**, 411-423, doi:10.1523/jneurosci.4445-12.2013 (2013).
43 Raff, M. C. *et al.* Programmed cell death and the control of cell survival: lessons from the nervous system. *Science (New York, N.Y.)* **262**, 695-700, doi:10.1126/science.8235590 (1993).
44 Mazarakis, N. D., Edwards, A. D. & Mehmet, H. Apoptosis in neural development and disease. *Archives of disease in childhood. Fetal and neonatal edition* **77**,

F165-170, doi:10.1136/fn.77.3.f165 (1997).
45 Brandt, N., Vierk, R. & Rune, G. M. Sexual dimorphism in estrogen-induced synaptogenesis in the adult hippocampus. *The International journal of developmental biology* **57**, 351-356, doi:10.1387/ijdb.120217gr (2013).
46 Keil, K. P., Sethi, S., Wilson, M. D., Chen, H. & Lein, P. J. In vivo and in vitro sex differences in the dendritic morphology of developing murine hippocampal and cortical neurons. *Scientific reports* **7**, 8486, doi:10.1038/s41598-017-08459-z (2017).
47 Lang, J. T. & McCullough, L. D. Pathways to ischemic neuronal cell death: are sex differences relevant? *Journal of translational medicine* **6**, 33, doi:10.1186/1479-5876-6-33 (2008).
48 Walsh, K. *et al.* Maternal prenatal stress phenotypes associate with fetal neurodevelopment and birth outcomes. *Proceedings of the National Academy of Sciences of the United States of America* **116**, 23996-24005, doi:10.1073/pnas.1905890116 (2019).
49 Mueller, B. R. & Bale, T. L. Sex-specific programming of offspring emotionality after stress early in pregnancy. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **28**, 9055-9065, doi:10.1523/jneurosci.1424-08.2008 (2008).
50 Morgan, C. P. & Bale, T. L. Early prenatal stress epigenetically programs dysmasculinization in second-generation offspring via the paternal lineage. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **31**, 11748-11755, doi:10.1523/jneurosci.1887-11.2011 (2011).
51 Jin, J. *et al.* The effects of cytosine methylation on general transcription factors. *Scientific reports* **6**, 29119, doi:10.1038/srep29119 (2016).
52 Hu, S. *et al.* DNA methylation presents distinct binding sites for human transcription factors. *eLife* **2**, e00726, doi:10.7554/eLife.00726 (2013).
53 Griswold, M. D. & Kim, J. S. Site-specific methylation of the promoter alters deoxyribonucleic acid-protein interactions and prevents follicle-stimulating hormone receptor gene transcription. *Biology of reproduction* **64**, 602-610, doi:10.1095/biolreprod64.2.602 (2001).