Distribution Agreement

In presenting this thesis as a partial fulfillment of the requirements for a degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis in whole or in part in all forms of media, now or hereafter now, including display on the World Wide Web. I understand that I may select some access restrictions as part of the online submission of this thesis. I retain all ownership rights to the copyright of the thesis. I also retain the right to use in future works (such as articles or books) all or part of this thesis.

Sasicha Manupipatpong April 4, 2017

5-hydroxymethylcytosine in autism spectrum disorder: an investigation of 5-hydroxymethylcytosine profiles in Mecp2 and Fmr1 knockout and overexpression mouse models

By

Sasicha Manupipatpong

Dr. Peng Jin
Adviser

Department of Biology

Dr. Peng Jin

Adviser

Dr. Meleah Hickman

Committee Member

Dr. Zhexing Wen

Committee Member

2017

5-hydroxymethylcytosine in autism spectrum disorder: an investigation of 5-hydroxymethylcytosine profiles in Mecp2 and Fmr1 knockout and overexpression mouse models

By

Sasicha Manupipatpong

Dr. Peng Jin

Adviser

An abstract of
a thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Sciences with Honors

Department of Biology

2017

Abstract

5-hydroxymethylcytosine in autism spectrum disorder: an investigation of 5-hydroxymethylcytosine profiles in Mecp2 and Fmr1 knockout and overexpression mouse models
By Sasicha Manupipatpong

Autism spectrum disorders (ASDs) comprise a group of disorders characterized by impaired language, social, and communication skills, in addition to restrictive behaviors or stereotypies. Symptoms begin to manifest in early development and persist throughout the affected individual's life. With a prevalence of 1.5% in developed countries and relatively high comorbidity rates, but still no clear underlying mechanism that unifies the heterogeneous phenotypes, it is a field of research that warrants further exploration. 5-hydroxymethylcytosine (5-hmC) is an epigenetic modification whose importance in neurodevelopmental and neurodegenerative disorders has been demonstrated. By investigating the 5-hmC profiles of mouse models for Rett Syndrome and Fragile X Syndrome, two autism spectrum disorder-linked monogenic disorders, in addition to overexpression models for their associated genes (MECP2 and FMR1, respectively), we were able to identify possible pathways in which 5-hydroxymethylation could play a role. Additionally, we characterized the nature of 5-hydroxymethylation in these mouse models and proposed hypoxia-inducible factor 1-alpha (HIF1A) and glucocorticoid modulatory element-binding protein 1 (GMEB1) as a possible players based on transcription factor motif analysis of the data.

5-hydroxymethylcytosine in autism spectrum disorder: an investigation of 5-hydroxymethylcytosine profiles in Mecp2 and Fmr1 knockout and overexpression mouse models

By

Sasicha Manupipatpong

Dr. Peng Jin

Adviser

A thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Sciences with Honors

Department of Biology

2017

Acknowledgements

I would like to thank Bing Yao and Li Chen for teaching me and helping me with the bioinformatics analyses in this project. I also would like to thank Ying Cheng for his kind advice regarding HIF1A chromatin immunoprecipitation, Feiran Zhang for teaching me qPCR and chromatin immunoprecipitation, Zhiqin Wang and Ha Eun Kong for breeding and sacrificing the mice I used in my HIF1A chromatin immunoprecipitation, and Baoli Cheng and Li Lin for teaching me about 5-hmC capture. Most importantly, I would like to thank Dr. Peng Jin for allowing me the opportunity to work and learn in his lab.

Table of Contents

List of Abbreviations

5-hmC, 5-hydroxymethylcytosine

ASD, Autism Spectrum Disorder

ChIP, Chromatin immunoprecipitation

FMR1, Fragile X mental retardation 1

MECP2, Methyl-CpG binding protein 2

qPCR, Real-time PCR

Introduction

Autism spectrum disorders (ASDs) encompass a broad range of neurodevelopmental disorders whose symptoms manifest from early development as clinically significant functional deficits in linguistic and social aptitude, impaired communication skills, and restrictive or stereotypic behaviors[1]. A highly heterogeneous disorder, ASD symptoms can occur at various severities and as different specific combinations of behaviors (Table 1). Additionally, comorbidity rates for ASD are notable, with approximately 30% of ASD cases also manifesting impaired intellectual functioning. Psychiatric comorbidities include social anxiety disorder, attention-deficit hyperactivity disorder, and oppositional defiant disorder[2]. ASD has also been associated with abnormal immune system functioning, as well as gastrointestinal and sleep-related symptoms[3].

Reciprocal Social Interaction	Nonverbal Communication	Relationship Development and Maintenance
Abnormal approach to social interaction	Diminished integration of verbal and non-verbal communication	Inability to appropriately adapt to social environments
Inability to sustain mutual conversation	Abnormal usage of eye-contact, body language, and gestures	Trouble making friends and joining in imaginative play
Limited interests, emotions, and affect	Complete absence of facial affect and nonverbal communication	Lack of interest in others
Impaired social interaction initiation and response		

Restricted or Repetitive Behavior (at least two)			
Motor, object-use, or verbal stereotypies	Fixation on routine patterns and behaviors	High intensity focus on few interests	Abnormal reactivity to sensory stimuli

Table 1. DSM-V Diagnostic Criteria for ASD[1]. Individuals with ASD can manifest a range and combination of different symptoms that the DSM-V has outlined for the purposes of diagnosis.

Present in the population at a prevalence of 1.5% in developed countries, with males accounting for four times more cases than females, ASD not only affects the individuals with the disorder, but also those around them[4,5]. Children with ASD face a 12% increase in risk of injury, difficulties transitioning into adulthood, and discrimination when seeking employment as

adults[6,7,8]. At school, they often face bullying, leading to self-isolation that exacerbates their situation[5]. Parents of children with ASD face increased stress, anxiety, and depression, while their siblings face unique challenges growing up with siblings with ASD, both from caring for their siblings and dealing with their possibly frightening or unrestrained behaviors[5,9]. As it is a persistent disorder that affects many, continued research into ASD etiology and management is essential.

ASD development is predominantly attributed to underlying genetic factors, as shown from twin and family studies of the disorder[10]. A large number of risk factors have been identified, with 10% of ASD cases attributed to ASD related syndromes, 5% of cases due to rare chromosomal abnormalities, 5% of cases linked to rare copy number variations, and 5% of cases associated with risk genes identified for ASD. Rett Syndrome, Fragile X Syndrome, and Angelman Syndrome are examples of ASD related syndromes. Despite the progress that has been made in identifying genetic risk factors, the heterogeneity of the disorder, both in presentation and underlying cause, has proven a barrier to identifying the specific mechanisms by which ASD arises, thus demanding the consideration of other factors, such as environmental contributors.

Environmental contributors to ASD risk include both prenatal and perinatal factors. Maternal conditions found to influence ASD development include chemical exposure, infection, inflammation, emotional health, diet, and disease, as well as the age of the biological parents[11]. Specifically, endocrine-disrupting factors, air pollution, and heavy metals are chemicals that were found to increase ASD risk upon prenatal exposure, while maternal dietary concerns are focused on folic acid, iron, vitamin D, and poly-unsaturated fatty acid intake[4]. Perinatal factors include conditions shortly before and during childbirth, including preeclampsia, birth asphyxia, and breech birth[12]. Additionally, abnormal cytokine profiles in those with ASD point to immune

system dysfunction[13]. In order to understand the link between these environmental factors and the genetic factors influencing ASD development, however, epigenetics needs to be considered.

Epigenetics, the dynamic alteration of gene expression due to factors other than a change in genetic sequence, is an interface between genetic and environmental factors. There are many ways in which epigenetic regulation can occur, including via transcription factors, chromatin packaging, covalent histone and DNA modification, and non-coding RNA[14]. Of recent interest has been 5-hydroxymethylation of cytosine, a covalent DNA modification that is generated by the oxidation of 5-methylcytosine facilitated by the ten-eleven translocation enzyme (Figure 1). 5-hydroxymethylcytosine (5-hmC) is proposed to be both a transition state during 5-methylcytosine demethylation and a stable epigenetic marker that regulates transcription factor binding and chromatin structure. As it is more prevalent in the central nervous system than in other tissues, it is of particular interest in the context of neurodevelopmental disorders like ASD[15,16]. In fact, recent studies on chromatin remodeling complex components and genes in the context of ASD related syndromes have demonstrated the involvement of chromatin remodeling and DNA methylation in ASD[17]. In particular, it was shown that expression of *ARID1B*, *CHD8*, and *ANKRD1*, all of which are related to nucleosome and chromatin remodeling, are compromised in ASD, and monozygotic twin pairs in which only one twin has ASD exhibit different methylation profiles[17]. Additionally, 5-hmC was found to be specifically associated with both Fragile X Syndrome and Rett Syndrome, through differential 5-hydroxymethylation of cytosine in the *FMR1* gene promoter in Fragile X Syndrome patients and *MECP2* dosage-sensitive 5-hydroxymethylation in Rett Syndrome[18,19].

Figure Removed due to Copyright Restrictions

For more information, please see:

https://www.ncbi.nlm.nih.gov/core/lw/2.0/html/tileshop_pmc/tileshop_pmc_inline.html?title=Click%20on%20image%20to%20zoom&p=PMC3&id=4046508_nihms544443f2.jpg [20]

https://www.ncbi.nlm.nih.gov/core/lw/2.0/html/tileshop_pmc/tileshop_pmc_inline.html?title=Click%20on%20image%20to%20zoom&p=PMC3&id=4256999_fnins-08-00397-g0001.jpg [21]

Figure 1. Cytosine Modification Pathways[20, 21]. Cytosine modifications take place through multi-step pathways, with DNA methyltransferase (DNMT) responsible for the methylation of unmodified cytosine, leading to gene repression[20]. This can be followed by multiple oxidations facilitated by ten-eleven translocation (TET) enzymes, which converts the methyl group to a hydroxyl, formyl, and then carboxyl group. Thymine DNA glucosylase (TDG) then converts either 5-formylcytosine or 5-carboxylcytosine into an abasic site consisting of a hydroxyl group. Base excision repair pathways replace the abasic site with an unmodified cytosine, reactivating the gene. 5-hydroxymethylcytosine is also converted to unmodified cytosine by passive dilution during successive replications.

Fragile X Syndrome (OMIM #300624) and Rett Syndrome (OMIM #312750) are two ASD related monogenic disorders whose etiologies have been shown to include aberrant DNA methylation[17]. Fragile X Syndrome is caused by an abnormal expansion of CGG repeats on the X chromosome at the q27.3 region (Figure 2). This region is the 5'-non-coding region of the fragile X mental retardation 1 (*FMR1*) gene[22]. The *FMR1* gene encodes the fragile X mental retardation protein, which binds to mRNA and represses translation. Without normal levels of fragile X mental retardation protein, synapse transmission in the hippocampus undergoes long-term depression and synaptic plasticity is impaired[23]. Individuals who have Fragile X Syndrome

display a host of symptoms, including cognitive impairment, anxiety, hyperarousal, and hyperactivity, as well as enlarged testes for male individuals. In unaffected individuals, CGG repeats occur within the range of 6-53 repeats, and 200 or more repeats is considered a full mutation. The full mutation entails *FMR1* promoter hypermethylation[24]. As methylation is a repressive epigenetic modification, levels of the fragile X mental retardation protein are lowered[22].

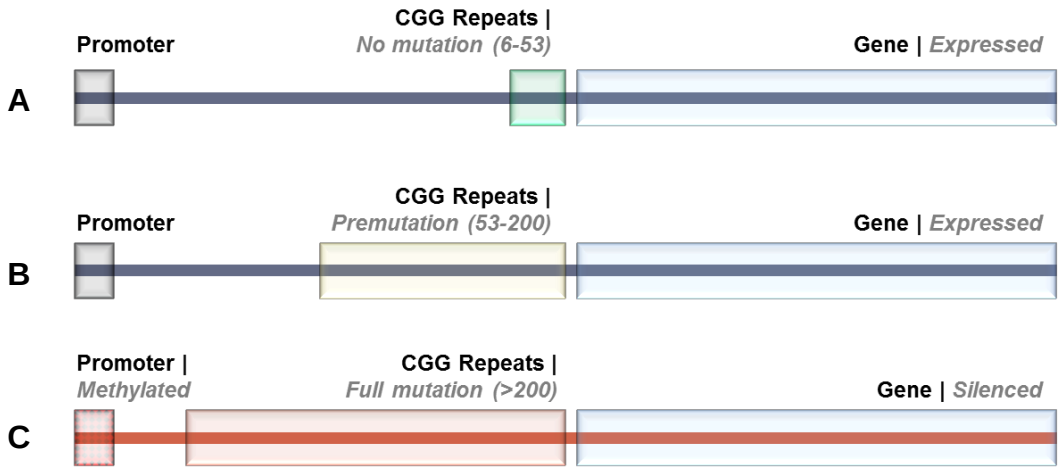


Figure 2. *FMR1* gene CGG repeat expansions in Fragile X Syndrome. (A) Normal *FMR1* gene composition, with 6-53 CGG repeats[24]. (B) Premutation expansion of 53-200 CGG repeats without promoter methylation. (C) Full mutation expansion of 200 or more CGG repeats, leading to promoter methylation, gene repression, and the Fragile X Syndrome phenotype.

Rett Syndrome, like Fragile X Syndrome, is also a monogenic disorder, and primarily occurs in females[25]. Rett Syndrome is characterized by specific stages. In the first six to eight months of their lives, Rett Syndrome patients undergo normal development, but then reach stagnation until ages one to four years. This stagnation continues on to become active regression in motor ability and language skills, in addition to development of social withdrawal characteristics of ASD in some cases. After regression is the plateau stage, when the loss of aforementioned skills ends and may begin to improve with work over time, coinciding with the development of hand

stereotypies. The last stage is late motor decline, in which patients no longer are able to walk. Rett Syndrome is typically (in 70-80% of cases) caused by a mutation in one copy of the methyl-CpG binding protein 2 (*MECP2*) gene located on the X chromosome, which is responsible for facilitating the downregulation of methylated genes[26]. The MECP2 protein is comprised of three main regions: the methyl-binding domain, transcription repression domain, which contains the nuclear localization signal, and C-terminal domain[27]. The methyl-binding domain recognizes and preferentially binds 5-methylcytosine, working with SIN3A, while the transcription repression domain binds to corepressors C-SKI and N-COR, as well as transcription factor IIB to obstruct transcription. Lastly, the C-terminal domain maintains protein stability. MECP2 also cooperates with histone deacetylase complexes and histone methylase to alter chromatin structure[28]. Over 100 missense and protein truncating mutations within and between these domains have been found to cause Rett Syndrome by preventing the protein from efficiently silencing genes[27].

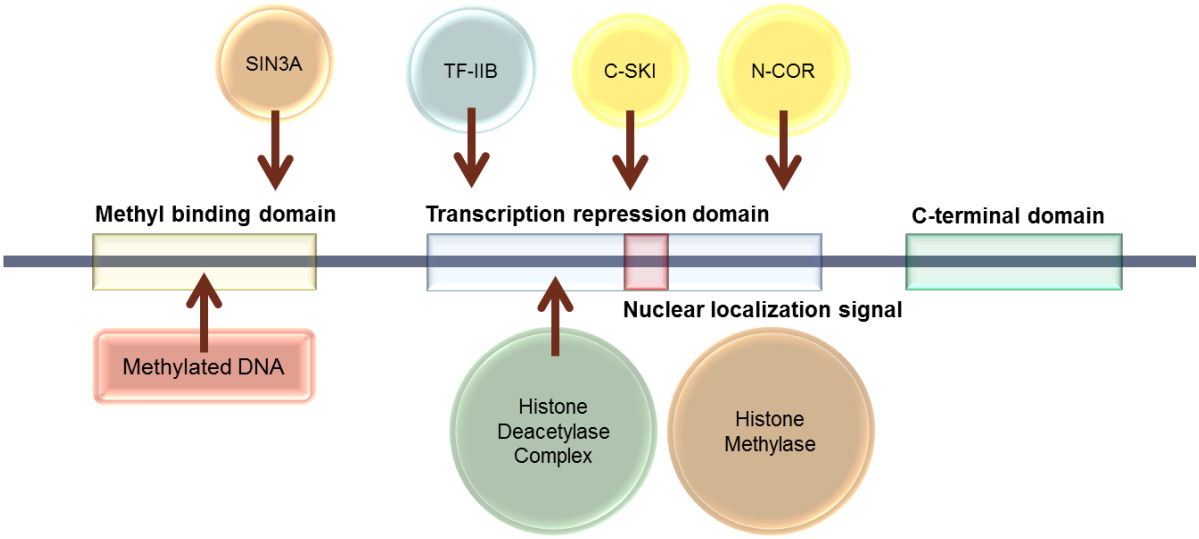


Figure 3. MECP2 protein schematic detailing main functional domains and interactions. MECP2 has three main important regions: the methyl binding domain, transcription repression domain, and C-terminal domain[27]. The methyl binding domain preferentially binds to methylated DNA and localizes the protein to its target, interacting with corepressor SIN3A. The transcription repression domain contains the nuclear localization signal and binds to histone deacetylase complex to cause chromatin remodeling. This is then followed by MECP2-bound histone methylase activity for further remodeling[28]. Mutations that

cause an early stop in translation prior to the nuclear localization signal prevent the protein from entering the nucleus and binding to DNA. The C-terminal domain is important to protein stability.

Gene knockout mouse models for both ASD-linked monogenic disorders exist, facilitating research into the two disorders. *Fmr1* knockout mice display abnormal dendritic spine morphology and density in very early development and adulthood, increased long-term depression in the hippocampus, decreased long-term potentiation, hypersensitivity, hyperactivity, repetitive behaviors, intellectual impairment, and enlarged testes for male *Fmr1* knockout mice[22,24,29]. Social impairment in the form of decreased vocalization is also exhibited[30]. At six weeks, *Mecp2* knockout mice begin to display behavioral abnormalities, such as decreased motor control and range, repetitive clasping of hind-limbs, decreased sociability, and internal testes for male mice[26,31]. Although due to a slightly different mechanism compared to *Fmr1* knockout mice, *Mecp2* knockout mice also show decreased long-term potentiation[32,33]. It is interesting to note that overexpression of *Mecp2* in *Mecp2*-duplication mice show similar phenotypes to *Fmr1* knockout mice and Fragile X Syndrome patients, including increased anxiety, decreased long-term potentiation, and cognitive dysfunction[34].

In the present study, we conduct an exploration of the 5-hmC profiles of *Fmr1* and *Mecp2* knockout and overexpression mouse models for Fragile X Syndrome and Rett Syndrome, respectively. We aim to identify possible common pathways to inform our understanding of the mechanisms underlying ASD pathology. Cerebella, where Purkinje cell number and size are pathologically decreased in ASD, were extracted from six week-old mice, followed by 5-hmC capture[35]. 5-hmC capture was achieved through selective chemical labelling of 5-hmC via T4 bacteriophage β-glucosyltransferase-facilitated transfer of a glucose-derived entity capable of binding biotin[36]. DNA obtained was subsequently sequenced to produce genome-wide maps of

5-hmC and analyzed for gene ontology, gene annotation, and centrally enriched transcription factor motifs.

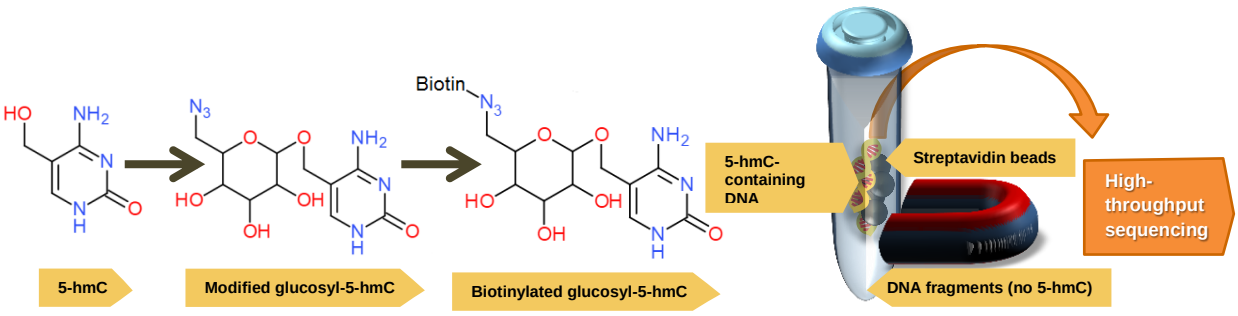


Figure 4. T4 bacteriophage β-glucosyltransferase-facilitated 5-hmC capture. 5-hmC is converted to a molecule that contains N_3 by the addition of a modified glucosyl group[36]. Biotin binds to the N_3 on the glucosyl group, allowing for pull-down of 5-hmC containing DNA fragments by magnetic Streptavidin beads that bind biotin. These DNA fragments are then processed for high-throughput sequencing.

Results

5-hmC loss dominates *Mecp2* and *Fmr1* knockout and overexpression mouse cerebellum 5-hmC profiles, with differentially 5-hydroxymethylated regions enriched in intron and exon regions

HOMER-identified peak annotation results for each mouse model showed a majority of peaks falling in intron regions (63.14% to 74.42% of peaks amongst all conditions), followed by intergenic regions (13.02-19.25%), and then exon regions (8.50-18.65%) (Figures 6A-H). HOMER peak annotation results for bins of 450 base pairs across the mouse genome were used as expected distribution for comparison, with results showing intergenic regions as the majority of regions (62.30%), followed by intron (33.73%), and then exon regions (1.21%) (Figure 5A). Therefore, differentially 5-hydroxymethylated regions are enriched in intron and exon regions, and depleted in intergenic regions. Across all mouse models, there is an overall global loss of 5-hmC. There is an overall gain in 5-hmC in exon regions for *Mecp2* knockout and overexpression mice, and an overall loss of 5-hmC in exon regions for *Fmr1* knockout and overexpression mice.

Overall loss of 5-hmC is seen in intron, transcription termination site, and 3' untranslated regions across all four mouse models tested. Less notable trends include an overall increase in 5-hmC in intergenic regions only in the *Mecp2* knockout mice, in 5' untranslated regions only in the *Mecp2* overexpression mice, and in promoter regions only in the *Fmr1* knockout mice.

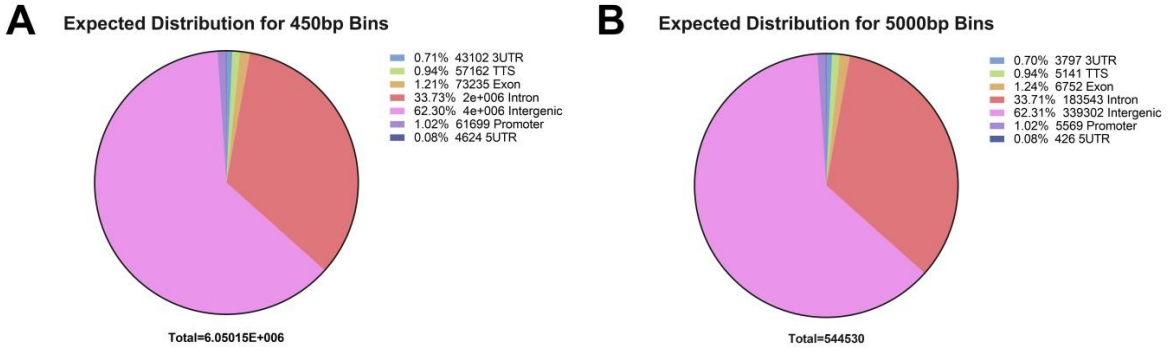
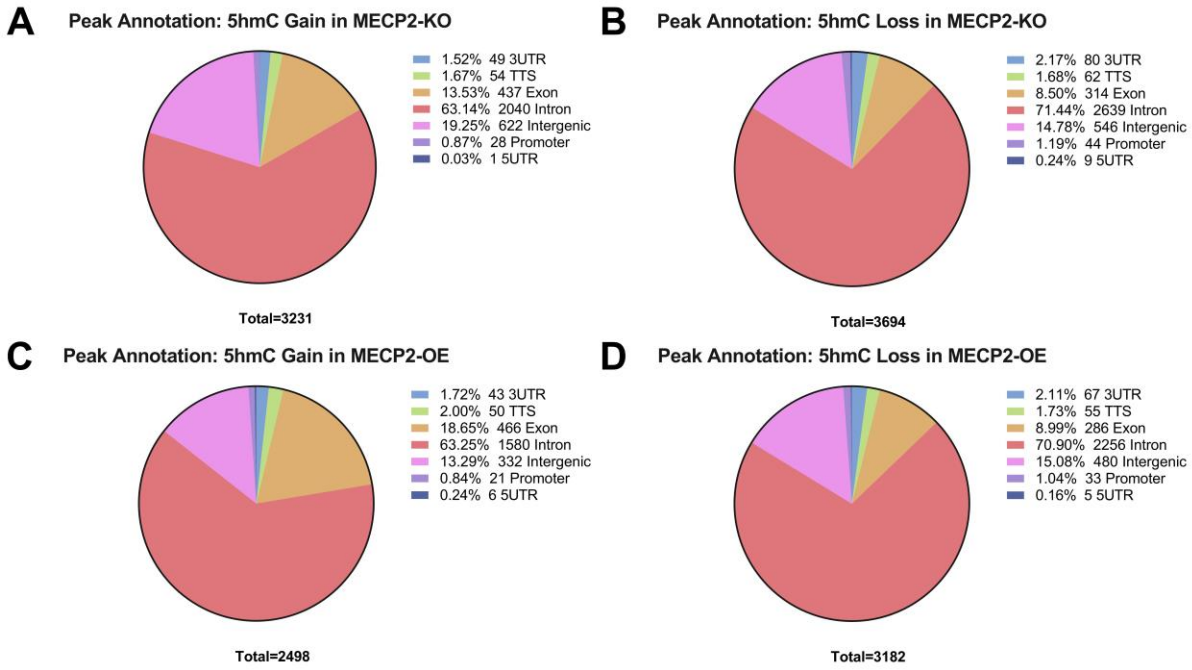


Figure 5. Expected gene annotation for mouse genome. (A) HOMER peak annotation statistics for mouse genome determined from 450 base pair bin regions (for comparison with HOMER-identified peak analyses, as 450 base pairs is the average peak length for those regions). (B) HOMER peak annotation statistics for mouse genome determined from 5000 base pair bin regions (for comparison with 5000 base pair bin analyses).



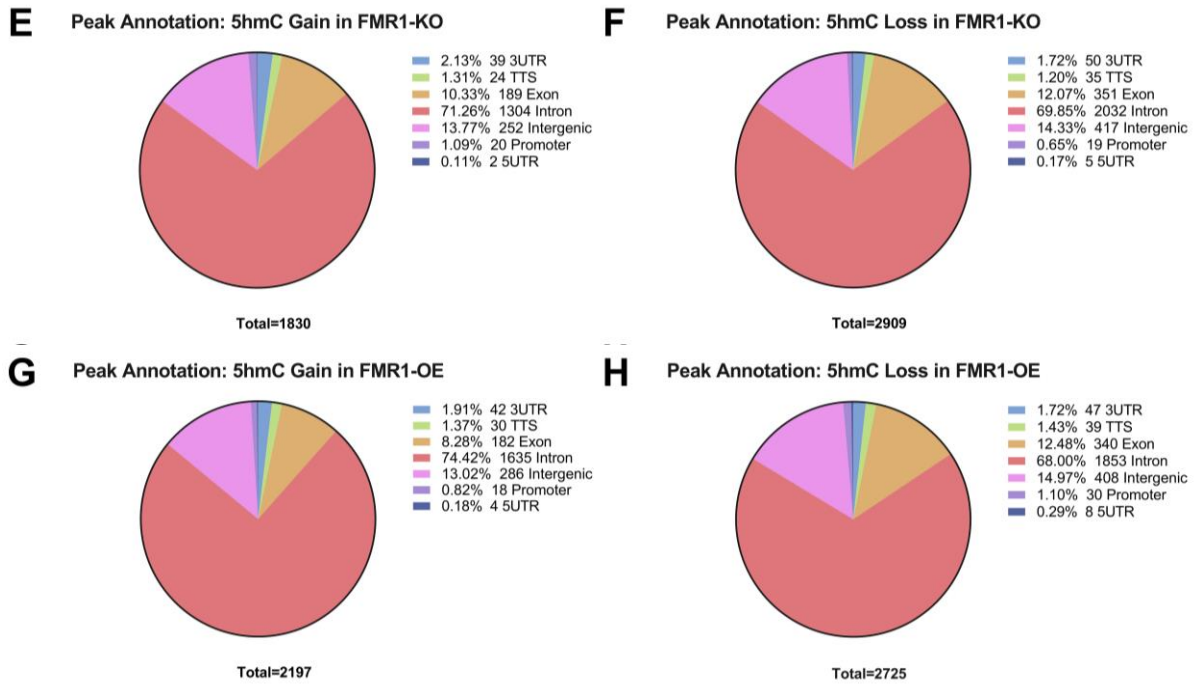


Figure 6. Gene annotation for differentially 5-hydroxymethylated regions in each mouse model. (A) HOMER peak annotation statistics for *Mecp2* knockout mice (n=2) for regions with increased 5-hmC compared to wildtype littermate (n=1). (B) HOMER peak annotation statistics for *Mecp2* knockout mice for regions with decreased 5-hmC compared to wildtype littermate. (C) HOMER peak annotation statistics for *Mecp2* overexpression mice (n=2) for regions with increased 5-hmC compared to wildtype littermate (n=1). (D) HOMER peak annotation statistics for *Mecp2* overexpression mice for regions with decreased 5-hmC compared to wildtype littermate. (E) HOMER peak annotation statistics for *Fmr1* knockout mice (n=2) for regions with increased 5-hmC compared to wildtype (n=2). (F) HOMER peak annotation statistics for *Fmr1* knockout mice for regions with decreased 5-hmC compared to wildtype. (G) HOMER peak annotation statistics for *Fmr1* overexpression mice (n=2) for regions with increased 5-hmC compared to wildtype (n=2). (H) HOMER peak annotation statistics for *Fmr1* overexpression mice for regions with decreased 5-hmC compared to wildtype.

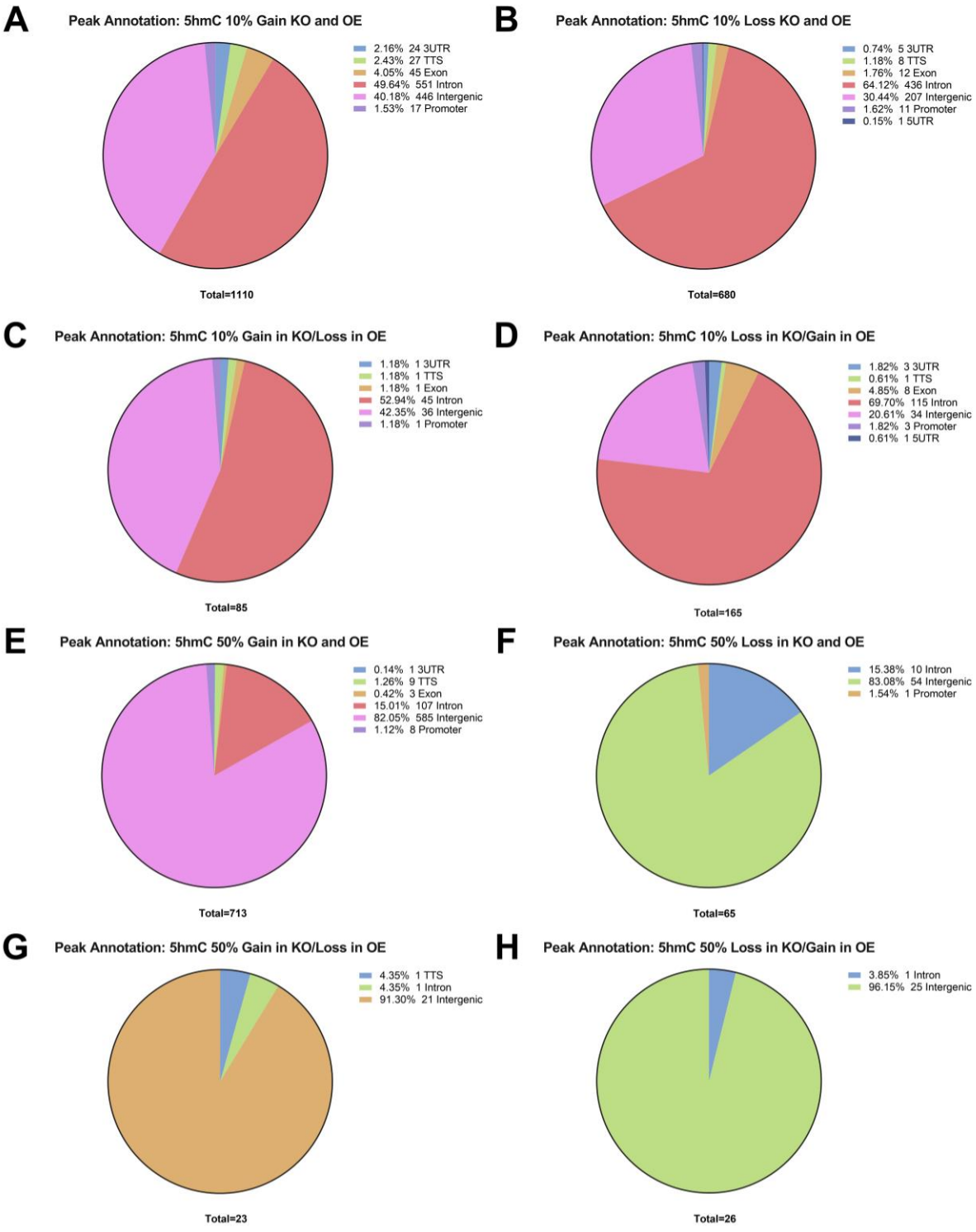
5-hmC loss in *Fmr1* and *Mecp2* knockout mice with opposite change in overexpression mice is more common than gain with opposite change

Broad 5000 base pair bins and HOMER-identified peaks were analyzed for 5-hmC enrichment and compared between models at different percent changes compared to wildtype. Three comparisons were conducted in this way. Firstly, 5000 base pair bins with at least a 10% percent difference in 5-hmC reads between wildtype and knockout or overexpression models were considered. Regions were compared to input reads to ensure the enrichment was significant with a Benjamini and Hochberg adjusted p-value of 0.05 (Figures 7A-D). Secondly, 5000 base pair

bins with at least a 50% difference between wildtype and knockout or overexpression models were considered (Figures 7E-H). These gave a broader view of 5-hydroxymethylation change. Lastly, HOMER-identified peaks with at least a 25% 5-hmC level difference between wildtype and knockout or overexpression models were considered in order to look at changes in regions centrally enriched in 5-hmC (Figures 7I-L). For all three conditions, all of the mouse models have more 5-hmC increase in common than 5-hmC decrease. Additionally, there are more areas of 5-hmC loss in knockout mice with the opposite change (gain) in overexpression mice as there are areas of gain in knockout mice with an opposite change (loss) in overexpression mice (Figures 7C, D, G, H, K, L). Conversely, this entails that there are more areas of 5-hydroxymethyaltion gain in overexpression mice with an opposite change in knockout mice than there are areas of 5-hmC loss in overexpression mice with an opposite change in knockout mice. At a 10% change threshold, there was more loss of 5-hydroxymethylation in knockout mice with an opposite change in overexpression mice in the intron, 3' untranslated, exon, promoter, and 5' untranslated regions (Figures 7C-D). Only in intergenic regions was there overall gain in 5-hydroxymethylation in knockout mice with an opposite change in overexpression mice. Transcription termination sites were tied for peaks with knockout mouse 5-hmC gain and loss with an opposite change in overexpression mice. Relatively similar trends were seen for the 25% threshold comparison, except for transcription termination sites showing more loss and promoter regions being tied for peaks with gain and loss. For regions with at least a 50% change in 5-hydroxymethylation compared to wildtype, a large majority (82.05-96.15%) of investigated changes in 5-hydroxymethylation occurred in intergenic regions (Figures 7E-H), whereas for regions with at least 10% and 25% changes compared to wildtype (Figures 7A-D, I-L), the

majority of 5-hydroxymethylation gain and loss fell in intron regions (49.64-69.70% in the 10% threshold comparison and 51.32-59.19% in the 25% threshold comparison).

For comparison, HOMER peak annotation was conducted on 450 base pair and 5000 base pair bin regions across the mouse genome to generate an expected distribution. Both distributions showed intergenic regions as the largest majority (62.30% for 450 base pair regions, 62.31% for 5000 base pair regions), followed by intron (33.73% for 450 base pair regions, 33.71% for 5000 base pair regions), exons (1.21% for 450 base pair regions, 1.24% for 5000 base pair regions), promoter (1.02% for both 450 base pair regions and 5000 base pair regions), and then transcription termination site (0.94% for both 450 base pair regions and 5000 base pair regions) (Figure 5). This indicates that the tested differentially 5-hydroxymethylated regions in these various groups for 10% change in 5-hmC level and 25% change in 5-hmC level were enriched for intron regions and depleted for intergenic regions (Figures 7A-D, I-L). However, the sets of regions demonstrating 50% change in 5-hmC level followed the same trend in terms of distribution ranking for components, with intergenic regions representing the majority of regions (Figures 7E-H).


A
Peak Annotation: 5hmC 10% Gain KO and OE
2.16% 24 3UTR
2.43% 27 TTS
4.05% 45 Exon
49.64% 551 Intron
40.18% 446 Intergenic
1.53% 17 Promoter
Total=1110
B
Peak Annotation: 5hmC 10% Loss KO and OE
0.74% 5 3UTR
1.18% 8 TTS
1.76% 12 Exon
64.12% 436 Intron
30.44% 207 Intergenic
1.62% 11 Promoter
0.15% 1 5UTR
Total=680
C
Peak Annotation: 5hmC 10% Gain in KO/Loss in OE
1.18% 1 3UTR
1.18% 1 TTS
1.18% 1 Exon
52.94% 45 Intron
42.35% 36 Intergenic
1.18% 1 Promoter
Total=85
D
Peak Annotation: 5hmC 10% Loss in KO/Gain in OE
1.82% 3 3UTR
0.61% 1 TTS
4.85% 8 Exon
69.70% 115 Intron
20.61% 34 Intergenic
1.82% 3 Promoter
0.61% 1 5UTR
Total=165
E
Peak Annotation: 5hmC 50% Gain in KO and OE
0.14% 1 3UTR
1.26% 9 TTS
0.42% 3 Exon
15.01% 107 Intron
82.05% 585 Intergenic
1.12% 8 Promoter
Total=713
F
Peak Annotation: 5hmC 50% Loss in KO and OE
15.38% 10 Intron
83.08% 54 Intergenic
1.54% 1 Promoter
Total=65
G
Peak Annotation: 5hmC 50% Gain in KO/Loss in OE
4.35% 1 TTS
4.35% 1 Intron
91.30% 21 Intergenic
Total=23
H
Peak Annotation: 5hmC 50% Loss in KO/Gain in OE
3.85% 1 Intron
96.15% 25 Intergenic
Total=26

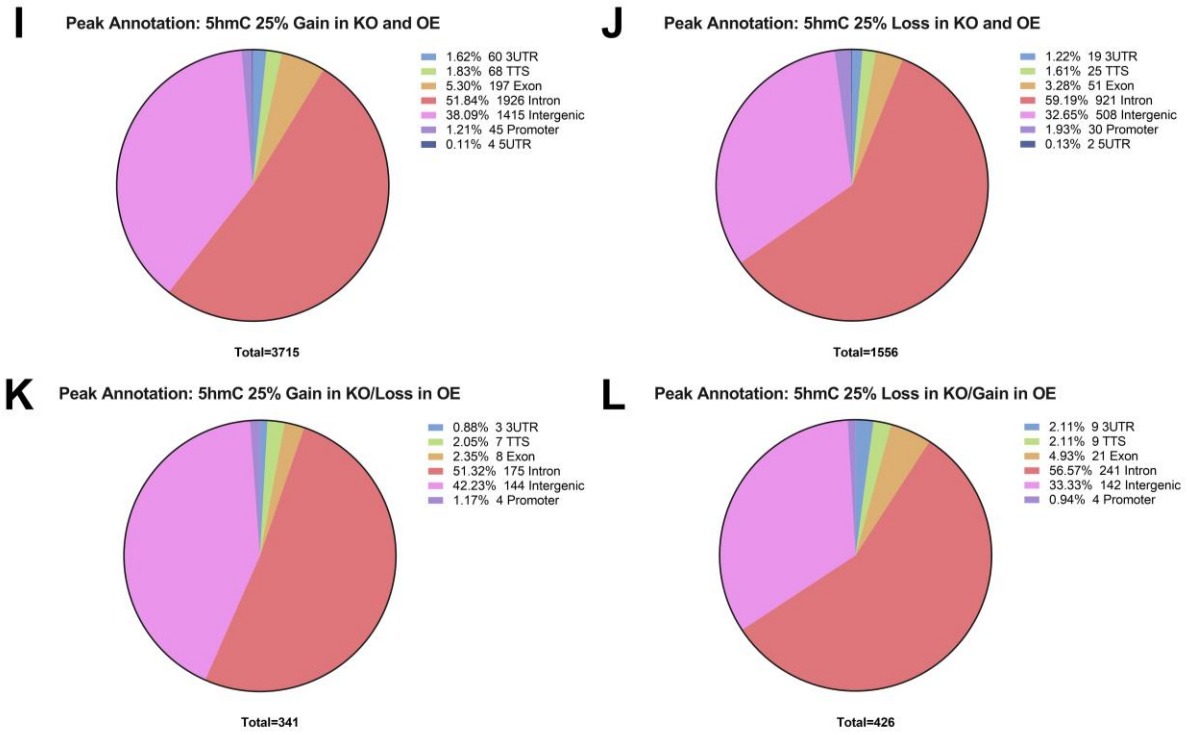


Figure 7. Gene annotation for differentially 5-hydroxymethlated regions by percent difference of 5-hmC in knockout and overexpression groups. (A) HOMER annotation on 5000 base pair regions with 10% increase of 5-hmC compared to wildtype in knockout and overexpression models (adjusted p-value < 0.05). (B) HOMER annotation on 5000 base pair regions with 10% decrease of 5-hmC compared to wildtype in knockout and overexpression models (adjusted p-value < 0.05). (C) HOMER annotation on 5000 base pair regions with 10% increase of 5-hmC compared to wildtype in knockout models, and 10% decrease compared to wildtype in overexpression models (adjusted p-value < 0.05). (D) HOMER annotation on 5000 base pair regions with 10% decrease of 5-hmC compared to wildtype in knockout models, and 10% increase compared to wildtype in overexpression models (adjusted p-value < 0.05). (E) HOMER annotation on 5000 base pair regions with 50% increase of 5-hmC compared to wildtype in knockout and overexpression models. (F) HOMER annotation on 5000 base pair regions with 50% decrease of 5-hmC compared to wildtype in knockout and overexpression models. (G) HOMER annotation on 5000 base pair regions with 50% increase of 5-hmC compared to wildtype in knockout models, and 50% decrease compared to wildtype in overexpression models. (H) HOMER annotation on 5000 base pair regions with 50% decrease of 5-hmC compared to wildtype in knockout models, and 50% increase compared to wildtype in overexpression models. (I) HOMER annotation on HOMER-identified peak regions with 25% increase of 5-hmC compared to wildtype in knockout and overexpression models. (J) HOMER annotation on HOMER-identified peak regions with 25% decrease of 5-hmC compared to wildtype in knockout and overexpression models. (K) HOMER annotation on HOMER-identified peak regions with 25% increase of 5-hmC compared to wildtype in knockout models, and 25% decrease compared to wildtype in overexpression models. (L) HOMER annotation on HOMER-identified peak regions with 25% decrease of 5-hmC compared to wildtype in knockout models, and 25% increase compared to wildtype in overexpression models.

5-hmC enrichment plots show differences in 5-hmC levels compared to wildtype between *Mecp2* mouse models and *Fmr1* mouse models as well as between knockout and overexpression models

5-hmC enrichment plots for promoter, enhancer, exon, gene body, transcription start site, transcription end site, and CpG island regions were also generated to show read distribution profiles throughout regions of interest. While we were able compare different groups with our data, it is important to note the discrepancies between the different wildtype groups (general wildtype, wildtype littermate for *Mecp2* knockout, and wildtype littermate for *Mecp2* overexpression). As such, wildtype comparisons were only made here for *Mecp2* knockout and overexpression to maintain validity. For both *Fmr1* and *Mecp2* mouse models, knockout 5-hmC levels in enhancer regions were higher than in their corresponding overexpression mouse models (Figure 8B). Additionally, within exon regions, *Mecp2* knockout and overexpression mice show higher levels of 5-hmC than their wildtype littermates (Figure 8C). It also appears as though there are lower levels of 5-hmC within exon regions for *Fmr1* knockout and overexpression mice than the wildtype mice, however, the comparison should be repeated with wildtype littermates. These same trends were observed in gene body regions, with *Mecp2* knockout and overexpression mice showing higher 5-hmC levels than their wildtype littermates near the end of the gene body regions, and *Fmr1* knockout and overexpression showing lower 5-hmC levels than wildtype near the front of the gene body regions (Figure 8D). Lastly, *Mecp2* knockout mice showed higher levels of 5-hydroxymethylation than their wildtype littermates slightly downstream of transcription start sites (Figure 8E).

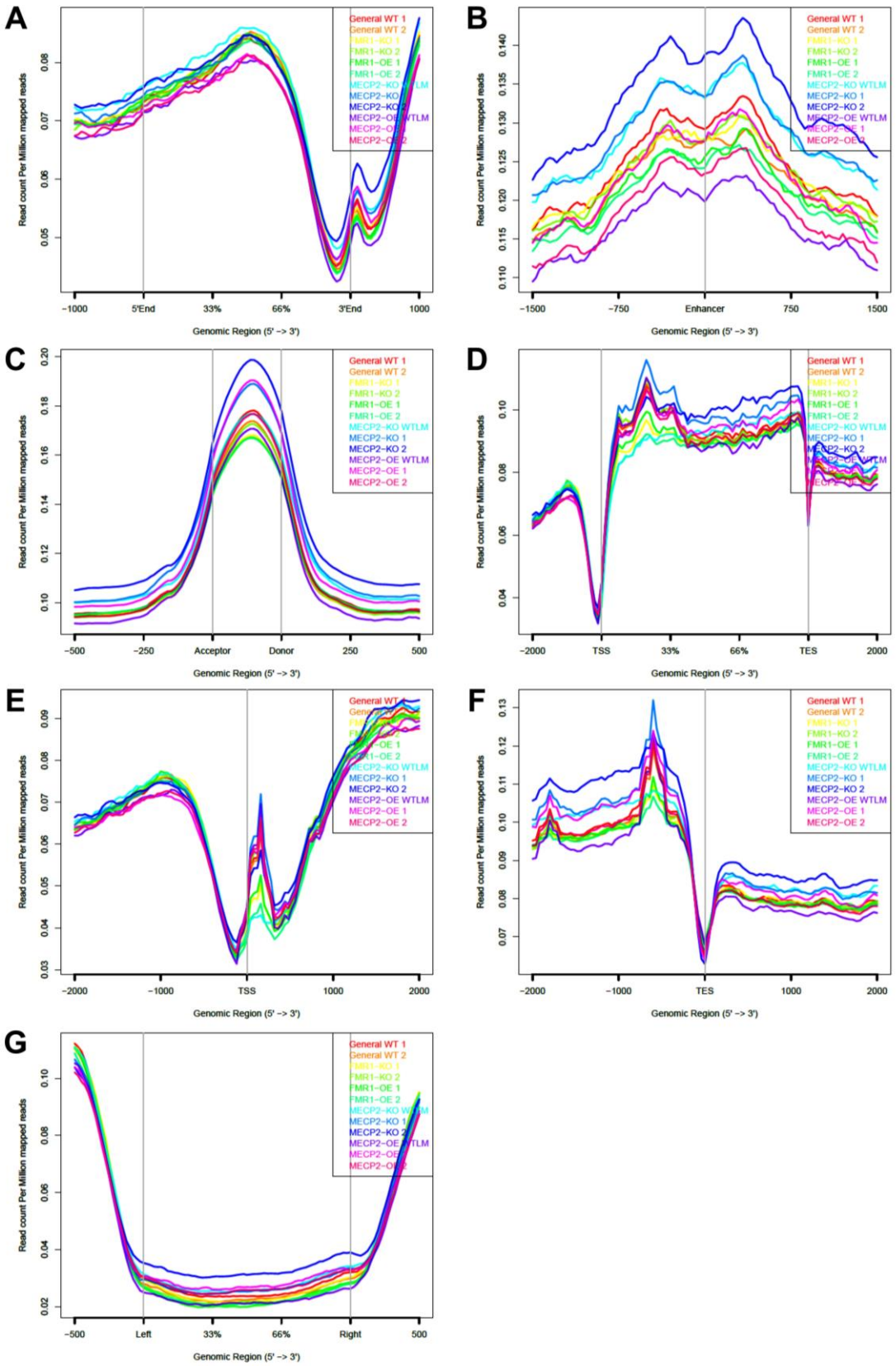
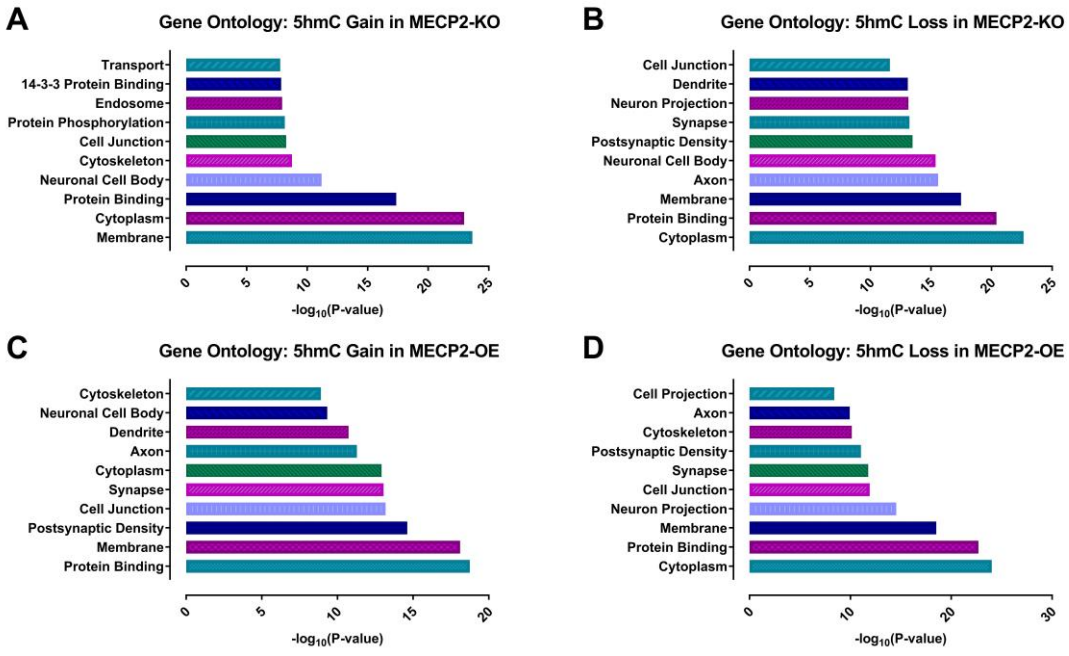


Figure 8. 5-hmC capture enrichment plots for various genomic regions. ngs.plot read distribution over (A)

promoter (2000 base pairs upstream of known UCSC RefSeq genes), (B) enhancer, (C) exon, (D) gene body (E) transcription start site, (F) transcription end site, and (G) CpG island regions.

Gene ontology analyses for each mouse model show most significantly enriched terms to be related to membrane, cytoplasm, and protein binding

Using DAVID gene ontology to determine the top ten most significant enriched gene ontology terms within HOMER-identified peaks, terms that stood out with the lowest p-values were membrane, cytoplasm, and protein binding throughout all four models for regions of 5-hmC gain and loss (Figures 9A-H). Other terms that appeared in the top ten most significant enriched terms throughout all conditions were axon, postsynaptic density, neuronal cell body, and cell junction. Interestingly, the gene ontology term dendrite appeared in 5-hmC loss regions for *Mecp2* knockout, *Fmr1* knockout, and *Fmr1* overexpression, and 5-hmC gain regions for *Mecp2* overexpression. Additionally, the gene ontology term cytoskeleton appeared within the top ten most significant terms only for *Mecp2* knockout and overexpression mice (gain of 5-hmC, and both gain and loss, respectively) (Figures 9A, C, D).



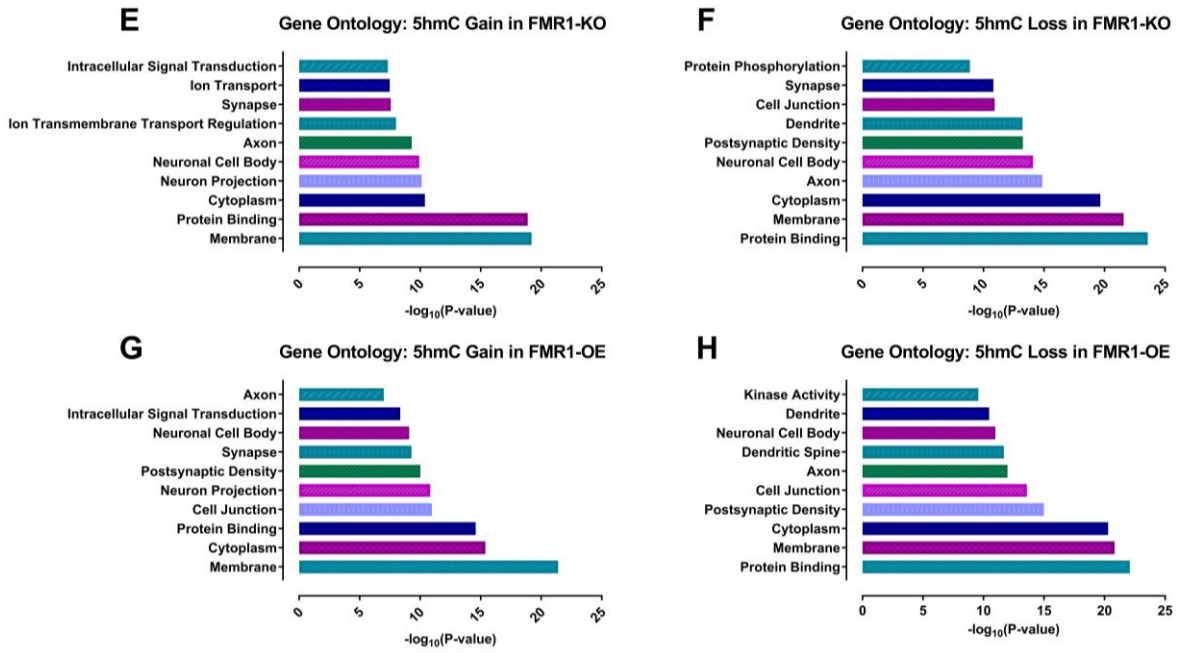
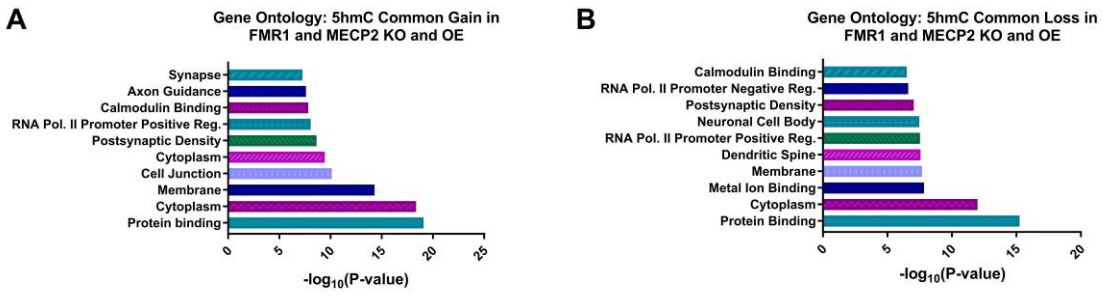


Figure 9. DAVID gene ontology for HOMER-identified 5-hmC peaks for each mouse model. Biological process, cellular component, and metabolic function categories were considered. (A) Gene ontology for HOMER-identified peaks demonstrating 5-hydroxymethylation gain compared to wildtype littermate in *Mecp2* knockout mice. (B) Gene ontology for HOMER-identified peaks demonstrating 5-hydroxymethylation loss compared to wildtype littermate in *Mecp2* knockout mice. (C) Gene ontology for HOMER-identified peaks demonstrating 5-hydroxymethylation gain compared to wildtype littermate in *Mecp2* overexpression mice. (D) Gene ontology for HOMER-identified peaks demonstrating 5-hydroxymethylation loss compared to wildtype littermate in *Mecp2* overexpression mice. (E) Gene ontology for HOMER-identified peaks demonstrating 5-hydroxymethylation gain compared to wildtype in *Fmr1* knockout mice. (F) Gene ontology for HOMER-identified peaks demonstrating 5-hydroxymethylation loss compared to wildtype in *Fmr1* knockout mice. (G) Gene ontology for HOMER-identified peaks demonstrating 5-hydroxymethylation gain compared to wildtype in *Fmr1* overexpression mice. (H) Gene ontology for HOMER-identified peaks demonstrating 5-hydroxymethylation loss compared to wildtype in *Fmr1* overexpression mice.

Enriched gene ontology terms for regions with 5-hmC gain in knockout mice and loss in overexpression mice are widely different from those for regions of common gain or loss between all mouse models

Membrane and membrane-related gene ontology terms were present in all comparisons made for common 5-hmC gain and loss across all groups, as well as gain and loss in knockout or overexpression groups with opposite change in overexpression or knockout groups, respectively

(Figures 10A-H). These terms include membrane, postsynaptic membrane, plasma membrane integral component, and organelle membrane. The top ten most significant enriched gene ontology terms for HOMER-identified regions with at least a 25% change in 5-hydroxymethlcytosine level from wildtype (Figures 10A-D) for the most part maintained similar composition as in the individual gene ontology analyses (Figures 9A-H). However, gene ontology terms for regions with gain in 5-hmC in knockout mice (Figure 10D) with an opposite change in overexpression mice were quite different from the terms found in the individual mouse model gene ontology analyses. New terms that arose were kainite selective glutamate receptor (Figure 10D), B-cell proliferation in immune response and lipid binding (Figure 10C), and calmodulin binding and RNA polymerase II promoter regulation (Figures 10A-B). Another quite different set of enriched gene ontology terms arose from the 5000 base pair regions with at least a 10% change in 5-hmC level compared to wildtype and adjusted p-values of 0.05 or less (Figures 10E-H). These terms were more specific, especially in the regions showing 5-hmC gain in knockout and loss in overexpression, which generated terms that included aromatase activity, arachidonic acid epoxygenase activity, steroid hydroxylase activity, monooxygenase activity, iron ion binding, homophilic cell adhesion, and epoxygenase P450 pathway (Figure 10H). Gene ontology terms for regions demonstrating 5-hmC loss in knockout and gain in overexpression introduced new terms like perikaryon, aspartate metabolic process, terminal bouton, and positive regulation of apoptotic process (Figure 10G).



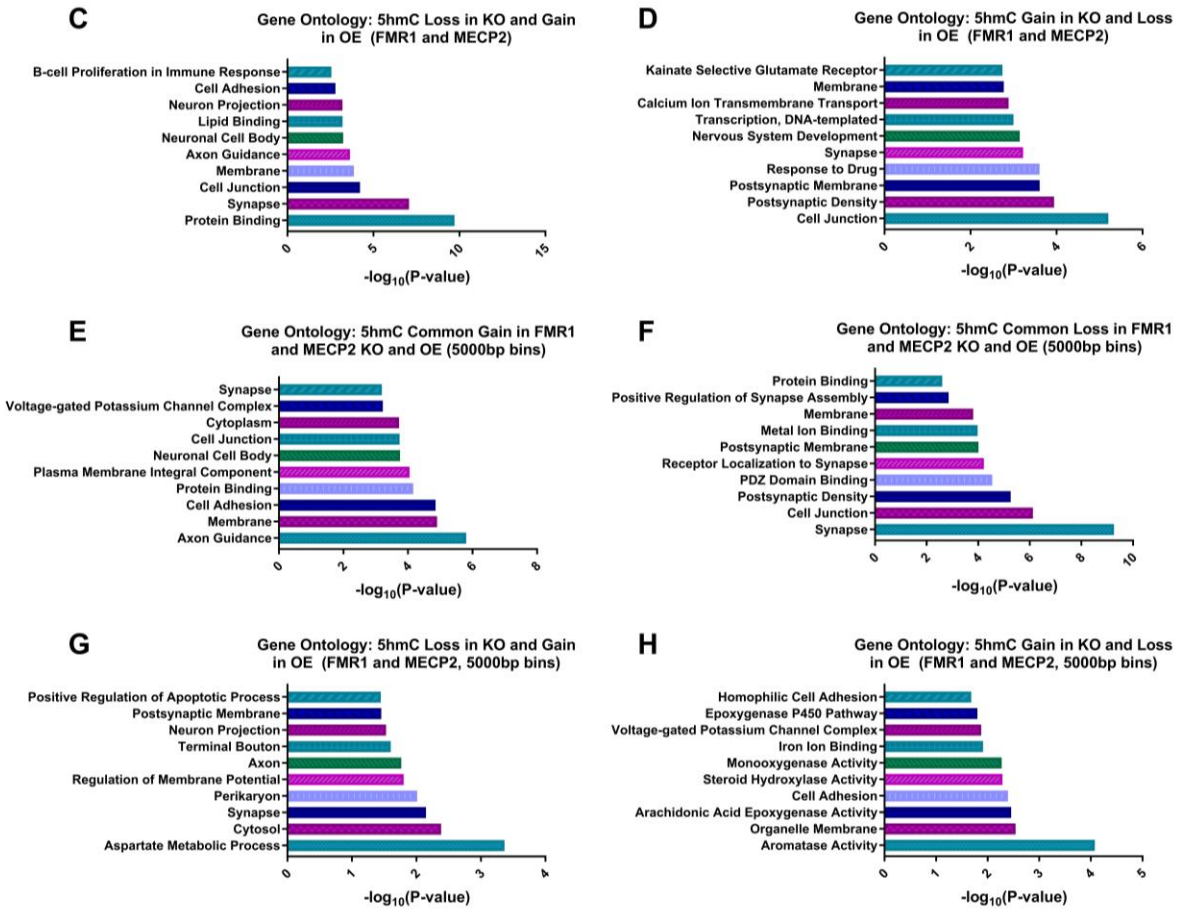


Figure 10. DAVID gene ontology for differentially 5-hydroxymethlated regions by percent difference of 5-hmC in knockout and overexpression groups (HOMER-identified peaks and 5000 base pair bins). (A) Gene ontology for HOMER-identified peak regions with 25% increase of 5-hmC compared to wildtype in knockout and overexpression models. (B) Gene annotation for HOMER-identified peak regions with 25% decrease of 5-hmC compared to wildtype in knockout and overexpression models. (C) Gene annotation for HOMER-identified peak regions with 25% decrease of 5-hmC compared to wildtype in knockout models, and 25% increase compared to wildtype in overexpression models. (D) Gene annotation for HOMER-identified peak regions with 25% increase of 5-hmC compared to wildtype in knockout models, and 25% decrease compared to wildtype in overexpression models. (E) Gene ontology for 5000 base pair regions with 10% increase of 5-hmC compared to wildtype in knockout and overexpression models (adjusted p-value < 0.05). (F) Gene ontology for 5000 base pair regions with 10% decrease of 5-hmC compared to wildtype in knockout and overexpression models (adjusted p-value < 0.05). (G) Gene ontology for 5000 base pair regions with 10% decrease of 5-hmC compared to wildtype in knockout models, and 10% increase compared to wildtype in overexpression models (adjusted p-value < 0.05). (H) Gene ontology for 5000 base pair regions with 10% increase of 5-hmC compared to wildtype in knockout models, and 10% decrease compared to wildtype in overexpression models (adjusted p-value < 0.05).

Overrepresented enriched centrally enriched transcription factor motifs identified in HOMER-identified peak regions include hypoxia-inducible factor 1-alpha (HIF1A) and glucocorticoid modulatory element-binding protein 1 (GMEB1)

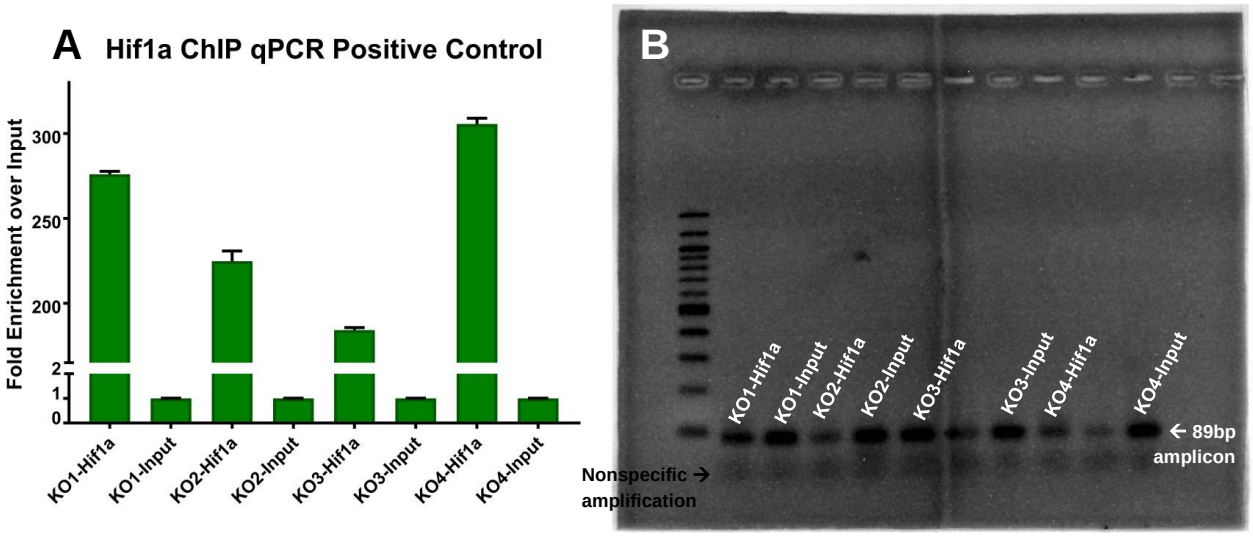
For HOMER-identified peak regions demonstrating at least a 25% change in 5-hmC level compared to wildtype, regions showing common gain and loss across all four mouse models were most significantly centrally enriched for HIF1A, ARNT, and GMEB1 transcription factor motifs. For peak regions showing loss in 5-hmC in knockout mice and gain in 5-hmC in overexpression mice, HIF1A, ZFP161, and GMEB1 were identified. Lastly, for peak regions showing gain in 5-hmC in knockout mice and loss in 5-hmC in overexpression mice, HIF1A and GMEB1 were identified. Throughout the groups, HIF1A and GMEB1 were the most common motifs out of the most significant motifs identified (Figure 11).

Top Motifs	Transcription Factor	CentriMo p-value			
		(A)	(B)	(C)	(D)
	Hypoxia-inducible factor 1-alpha (HIF1A)	7.2e-46	6.3e-26	1.2e-6	8.6e-4
	Glucocorticoid modulatory element-binding protein 1 (GMEB1)	5.1e-22	-	1.2e-3	2.2e-3

Figure 11. CentriMo centrally enriched transcription factor motifs in differentially 5-hydroxymethylated regions. Homer-motifs for most significant motifs identified, along with p-value for the four sets of regions (A-D). (A) Transcription factor motifs centrally enriched in HOMER-identified peak regions with 25% increase of 5-hmC compared to wildtype in knockout and overexpression models. HIF1A: 947 region matches; ARNT: 464 region matches; GMEB1 primary: 942 region matches. (B) Transcription factor motifs centrally enriched in HOMER-identified peak regions with 25% decrease of 5-hmC compared to wildtype in knockout and overexpression models. HIF1A: 471 region matches; ARNT: 239 region matches; GMEB1 secondary: 392 region matches. (C) Transcription factor motifs centrally enriched in HOMER-identified peak regions with 25% decrease of 5-hmC compared to wildtype in knockout, and 25% increase compared to wildtype in overexpression models. HIF1A: 88 region matches; ZFP161: 71 region matches; GMEB1 primary: 113 region matches. (D) Transcription factor motifs centrally enriched in HOMER-identified peak regions with 25% increase of 5-hmC compared to wildtype in knockout, and 25% decrease compared to wildtype in overexpression models. HIF1A: 78 region matches; GMEB1 primary: 56 region matches; GMEB1 secondary: 128 region matches.

HIF1A chromatin immunoprecipitation (ChIP) followed by real-time PCR (qPCR) illustrates possible enrichment of HIF1A-bound DNA in samples

HIF1A ChIP was employed to obtain HIF1A-bound DNA fragments for further exploration of the identified overrepresented transcription binding motifs in the cerebella of *Fmr1* knockout mice. qPCR with positive primers was used to determine whether or not the ChIP was successful, with all ChIP samples showing between 184-305 fold enrichment over input, indicating that HIF1A-bound regions were possibly enriched by the ChIP (Figure 12A). However, as melt curves and gel electrophoresis of qPCR products showed some non-specific amplification in addition to the expected 89 base pair amplicon, it is best to also find other primers to confirm ChIP efficiency (Figures 12B-C). As input samples had very high C_T values, indicating low starting amount and thus more likelihood of generating nonspecific amplification, Figure 12D was added to illlustrate slightly better melting curves for HIF1A ChIP samples. Nonetheless, these results need to be confirmed via qPCR with decreased nonspecific amplification.



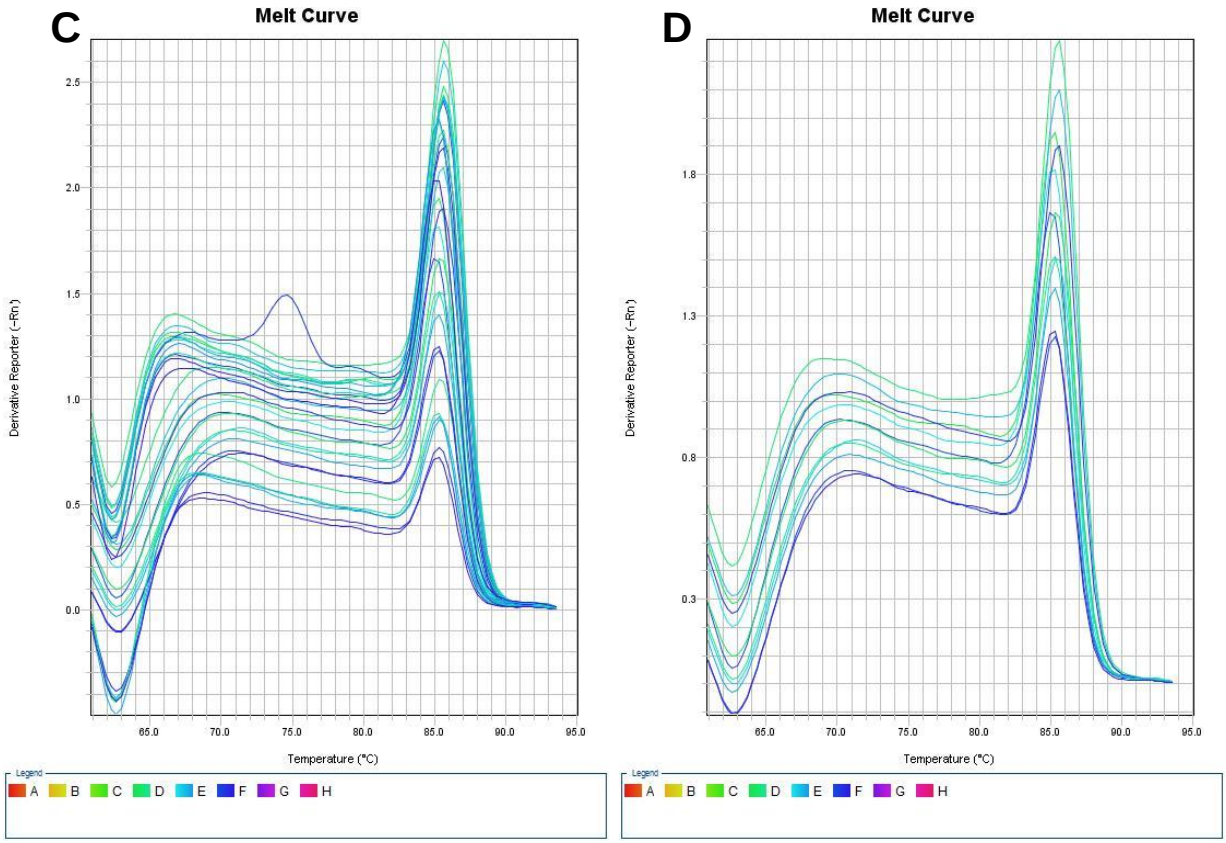


Figure 12. HIF1A ChIP real-time PCR results. (A) HIF1A ChIP fold enrichment over input. qPCR results obtained using a positive primer for HIF1A were analyzed to show fold enrichment of ChIP samples for HIF1A-binding regions over input. Mean fold enrichment over input for triplicate qPCR wells are graphed in each column, with propagated standard deviation error bars. (B) Gel electrophoresis results for qPCR products, two lanes excluded due to insufficient sample to run the condition for all four samples (IgG control). 89 base pair desired amplicon band was observed, but nonspecific amplication was also observed as smears for all samples. (C) Melt curves for all samples. (D) Melt curves for only HIF1A ChIP samples.

Comparison of differentially 5-hydroxymethylated regions and ASD candidate genes (SFARI) shows little overlap between genes associated with both increased 5-hmC in knockout and decreased 5-hmC in overexpression conditions and both increased 5-hmC in overexpression and decreased 5-hmC in knockout conditions

In comparing genes associated with HOMER-identified peaks with differential 5-hydroxymethylation (at least 25% change from wildtype) with SFARI ASD candidate genes, we

were able to identify genes of interest[37]. Genes that appeared in more than one group were sometimes cited for multiple regions, for example, intron and intergenic. There was very little overlap between ASD candidate genes that appeared in the two groups where opposite changes in 5-hmC level were observed in knockout and overexpression mice (Figure 13). For the regions demonstrating gain of 5-hmC in knockout mice and loss in overexpression mice, ASD candidate genes *ANKRD11, APP, ATP2B2, DAB1, DSCAM, GRIK4, GRIN2B, NLGN1, RBFOX1, SLC1A2,* and *STAT1* were identified as associated. For the regions demonstrating loss of 5-hmC in knockout mice and gain in overexpression mice, ASD candidate genes *ATP2B2, AUTS2, CNTN4, CNTNAP4, FOXP1, HOMER1*, and *NTNG1*were identified as associated, of which only *ATP2B2* overlaps with the former group. ASD genes were associated with 2.3% of peaks identified for regions showing gain of 5-hmC across all groups, 3.1% of peaks identified for regions showing loss across all groups, 3.8% in regions showing loss in knockout and gain in overexpression mice, and 1.6% in regions showing gain in knockout and loss in overexpression mice.

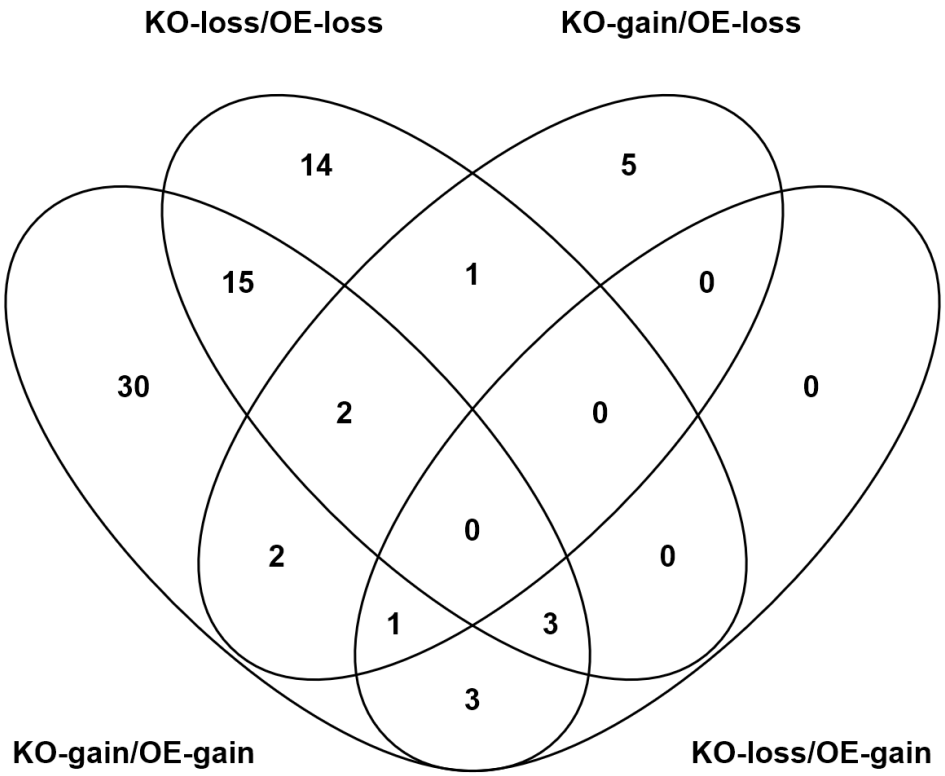


Figure 13. Venn diagram of autism spectrum disorder candidate genes (SFARI) within differentially 5-hydroxymethylated regions. HOMER-identified peaks showing at least 25% change in 5-hydroxymethylation level from wildtype by reads were annotated using HOMER and associated genes were compared with SFARI autism spectrum disorder candidate genes. These gene subsets were then used to generate a Venn diagram.

Discussion

The aim of the present study was to identify similarities and differences underlying the pathogenesis of Rett Syndrome and Fragile X Syndrome using mouse models for *Mecp2* and *Fmr1* knockout and overexpression. Most of the analysis centered on identifying aspects of 5-hydroxymethylation—its distribution, gene ontology, and associated transcription factor motifs—in order to identify how 5-hydroxymethylation could fit into the pathogenic pathways of ASD.

Differentially 5-hydroxymethylated regions were found to be most prevalent in intron regions, followed by intergenic and exon regions. A similar distribution of differentially 5-hydroxymethylated regions was found in a mouse model of Alzheimer's disease[38]. Interestingly, a study on normal human brain 5-hmC profiles found it to be enriched in enhancer, exon, and intron regions, but not promoter or intergenic regions[39]. Additionally, there was an overall loss of 5-hydroxymethylation throughout all four mouse models in the cerebellum. Overall loss of 5-hydroxymethylation was also observed in an Alzheimer's disease mouse model, however, this was specific to the hippocampus[38]. The difference in location for 5-hydroxymethylation loss might be due to the different natures of the disorders.

We compared common change across all models to identify abnormal 5-hydroxymetylation patterns, as all models produced abnormal phenotypes in the mice, in addition to opposite trends in knockout and overexpression models in order to determine how gene dosage may affect 5-hydroxymethylation. Introns were also found to be most enriched in differentially 5-hydroxymethylated regions with smaller changes in 5-hmC compared to the wildtype (10% and 25% differences). For regions with a 50% change in 5-hmC compared to wildtype, the majority are associated with intergenic regions. These regions are also much fewer in number. It would be interesting to explore whether these few regions with large change in 5-hmC might play a role in pathogenesis, especially since normal human brains are not enriched for 5-hmC in intergenic regions[39].

5-hmC enrichment plots showed trends in 5-hmC distribution in various genomic regions. Of note were distributions in enhancer, exon, gene body, and transcription start site regions. While the enhancer regions showed the difference between knockout and overexpression models, with higher 5-hmC levels in knockout mice than overexpression mice, the other regions showed

opposite effects in *Mecp2* models and *Fmr1* models compared to wildtype. Namely, *Mecp2* knockout and overexpression mice showed higher 5-hmC levels than wildtype, and *Fmr1* knockout and overexpression mice showed lower 5-hmC levels than wildtype in exon, gene body, and transcription start site regions. These relationships also varied throughout different portions of the gene body, suggesting that there might be different elements in the gene body that are getting 5-hydroxymethylated in models involving *Mecp2* and those involving *Fmr1*. Despite these observations, however, more experiments should be conducted with wildtype littermates for the *Fmr1* models, as a comparison to wildtype mice that are not littermates limits the conclusions we can draw from the data.

Perhaps most interesting are the gene ontology results, which show protein binding, membrane, and cytoplasm as the most significant enriched terms throughout regions of gain and loss in all mouse models. The enriched gene ontology term membrane, in addition to the presence of other membrane-related terms, could possibly point to the involvement of 5-hmC in key parts of Fragile X Syndrome and Rett Syndrome pathology: abnormal dendritic spine morphology leading to deficient synaptic plasticity[40,41]. In Fragile X Syndrome, the protein product of the *FMR1* gene is involved in regulating diacylglycerol and phosphatidic acid signaling molecules, which in turn are responsible for membrane structure and dendritic spine morphology[41]. *FMR1* is also expressed in neuronal cell bodies, dendrites, and postsynaptic spines, which could explain the presence of these gene ontology terms as well[22]. It also might be of interest to note the gene ontology term B-cell proliferation in immune response, as immune dysfunction has also been implicated in ASD due to the finding that children with ASD show elevated levels of B cells in addition to abnormal natural killer cell and cytokine levels[13].

Abnormal long-term depression and its involvement in both Rett Syndrome and Fragile X Syndrome are also linked to some of the gene ontology terms that arose, especially in regions of opposite change in knockout and overexpression conditions. Kainate selective glutamate receptor, aspartate metabolic process, arachidonic acid epoxygenase activity, and calmodulin binding are all related to long-term depression[42,43,44,45,46,47]. Kainate receptors are involved in both excitatory and inhibitory signals—exciting neurons but also regulating GABA and glutamate[42]. Aspartate is also important in that N-methyl-d-aspartate receptors are key in both Fragile X and Rett Syndrome—in Fragile X Syndrome, long-term depression and potentiation that rely on input from these receptors is impaired, and in Rett Syndrome, abnormal extrasynaptic N-methyl-d-aspartate receptor-mediated excitation contributes to excitation/inhibition imbalance[43,44]. Arachidonic acid epoxygenase activity through the cytochrome 450 pathway generates epoxyeicosatrienoic acids with involvement from metabolic glutamate receptor in the brain, protecting the brain from ischemia[45]. Metabolic glutamate receptor-mediated long-term depression has shown to be abnormal in ASD mouse models[46]. Lastly, calcium/calmodulin-dependent serine protein kinase is responsible for regulating associative memory extinction, and is associated with ASD[47].

The centrally enriched transcription factor motifs provide an avenue for continued research, with chromatin immunoprecipitation and sequencing to characterize the possible relationship between these transcription factors and pathogenesis in these mouse models. HIF1A is responsible for the cellular response to low levels of oxygen, and is predicted to bind to 9458 sites in 3980 genetic regions[48]. HIF1A had 78 (regions with 5-hmC gain in knockout, loss in overexpression) to 942 (regions with gain of 5-hydroxymethylcyosine gain throughout all 4 mouse models) region matches across the four region subsets tested. HIF1A chromatin immunoprecipitation was

carried out, followed by real-time PCR to determine success of chromatin immunoprecipitation in order to obtain HIF1A-bound DNA fragments for future sequencing. However, this was unable to be completed within the allotted time. GMEB1 had fewer region matches, but was the second most common significant centrally enriched transcription factor motif, and as such is also of interest.

In conclusion, we illustrated the distribution of differentially 5-hydroxymethylated regions, identified possible pathways in which 5-hydroxymethylation could play a role that were common to both ASD-linked monogenic disorders, and suggest a possible area for further research with regards to transcription factor involvement with 5-hyroxymethylcytosine in an ASD pathogenic pathway.

Materials and Methods

Mice

Mecp2 knockout. Mouse strain 129P2(C)-Mecp2 $^{tm1.1Bird}$/J purchased from The Jackson Laboratory was used for this experiment. Knockout mice were reared with their wildtype littermate and sacrificed at 6 weeks.

Mecp2 overexpression. Mouse strain Cg-Mapt$^{tm1(Mecp2)Jae}$/LimmJ from The Jackson Laboratory was used for this experiment. Overexpression mice were reared with their wildtype littermate and sacrificed at 6 weeks.

Fmr1 knockout. Mouse strain 129P2-Fmr1^{tm1Cgr}/J purchased from the Jackson Laboratory with a C57BL/6J genetic background was used in this experiment. Mice were sacrificed at 6 weeks.

Fmr1 overexpression. Yeast artificial chromosome transgenic mice overexpressing *Fmr1* were purchased from Baylor University for this experiment[49]. Mice were sacrificed at 6 weeks.

5-hydroxymethylcytosine capture

Cerebellum tissue was processed to obtain genomic DNA, which was then purified and sonicated. A mixture of the genomic DNA, β-glucosyltransferase reaction buffer, UDP-6-N_3-Glu, and β-glucosyltransferase was then incubated at 37˚C for one hour, followed by purification by Qiagen QIAquick Nucleotide Removal Kit. DNA was then eluted according to instructions provided by Bio-Rad. Disulfide-biotin linker was then added, followed by a 2 hour incubation at 37˚C, and another purification by Qiagen QIAquick Nucleotide Removal Kit. Lastly, Invitrogen Dynabeads MyOne Streptavidin C1 beads were added to the sample and manufacturer's instructions were followed for 5-hmC pulldown. Upon completion of the reaction, DTT added to release the DNA from the beads was removed and DNA was purified using Qiagen MinElute PCR Purification Kit.

Bioinformatics analysis

High-throughput sequencing results from 5-hmC capture was first aligned using bowtie2, followed by peak identification and annotation using HOMER suites[50,51]. From there, BEDTools was used to identify overlapping and intersecting peaks between different samples[52]. edgeR was used to compare different samples based on fold change in reads such that we could identify dosage-dependent differences in 5-hmC in addition to simple peak presence or absence[53]. Enriched gene ontology terms were identified based off gene lists obtained from HOMER annotation of peaks using DAVID Bioinformatics Resources 6.8 Functional Annotation Tool[54,55]. The top ten most significant results were reported. In order to illustrate read distribution in a 5-

hmC enrichment plot, ngs.plot was utilized[56]. Lastly, to identify centrally enriched transcription factor motifs, we used MEME suite tool CentriMo with default parameters[57,58].

Chromatin immunoprecipitation and qPCR

Chromatin immunoprecipitation was carried out on four *Fmr1* knockout mice at six weeks of age using HIF1A antibody N100-134 from Novus Biologicals. Each sample consisted of cerebellum tissue from one mouse which was cross-linked in 1% formaldehyde shortly after it was extracted. Seven cycles of sonication at 6 to 9 watts (30 seconds sonication, followed by 60 seconds break) were used to obtain an average fragment length of 300 base pairs. After sonication, samples were diluted 4.5 times, 10% was saved as input, and the rest was incubated at 4˚C overnight with HIF1A antibody. Low salt, high salt, LiCl salt, and TE buffer washes using Dynabeads Protein G were followed by elution and overnight reversal of cross-linking. Lastly, QIAquick PCR Purification Kit buffers and QIAquick Spin Columns (purple) were used to elute the DNA. The final elution was accomplished using nuclease free water. Sample concentrations were quantified using a Qubit 2.0 Fluorometer, and then 0.5ng was used for each qPCR reaction. The following positive primers were used for the qPCR: forward: CTACGTGAGGCTTCGCTG, reverse: CGTAAGTTCTGCTCTACTGCG. SYBR green reagents were used for the qPCR, and results were analyzed based on C_T values.

Figures

Figures were created using GraphPad Prism version 7.00 for Windows (GraphPad Software, La Jolla California USA, www.graphpad.com). The Venn diagram in Figure 8 was made using

Venny (Oliveros, J.C. (2007-2015) Venny. An interactive tool for comparing lists with Venn's diagrams. http://bioinfogp.cnb.csic.es/tools/venny/index.html).

References

1. American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed). Arlington, VA: American Psychiatric Association.
2. Simonoff, E., Pickles, A., Charman, T., Chandler, S., Loucas, T. and Baird, G. (2008) Psychiatric disorders in children with autism spectrum disorders: prevalence, comorbidity, and associated factors in a population-derived sample. *J Am Acad Child Adolesc Psychiatry*, **47**, 921-929.
3. Mannion, A. and Leader, G. (2016) An investigation of comorbid psychological disorders, sleep problems, gastrointestinal symptoms and epilepsy in children and adolescents with autism spectrum disorder: A two year follow-up. *Research in Autism Spectrum Disorders*, **22**, 20-33.
4. Lyall, K., Croen, L., Daniels, J., Fallin, M.D., Ladd-Acosta, C., Lee, B.K., Park, B.Y., Snyder, N.W., Schendel, D., Volk, H.E. et al. (2016) The Changing Epidemiology of Autism Spectrum Disorders. *Annu Rev Public Health*, in press.
5. Bonis, S.A. and Sawin, K.J. (2016) Risks and Protective Factors for Stress Self-Management in Parents of Children With Autism Spectrum Disorder: An Integrated Review of the Literature. *J Pediatr Nurs*, **31**, 567-579.
6. Jain, A., Spencer, D., Yang, W., Kelly, J.P., Newschaffer, C.J., Johnson, J., Marshall, J., Azocar, F., Tabb, L.P. and Dennen, T. (2014) Injuries among children with autism spectrum disorder. *Acad Pediatr*, **14**, 390-397.

7. van Schalkwyk, G.I. and Volkmar, F.R. (2017) Autism Spectrum Disorders: Challenges and Opportunities for Transition to Adulthood. *Child Adolesc Psychiatr Clin N Am*, **26**, 329-339.
8. Ohl, A., Grice Sheff, M., Little, S., Nguyen, J., Paskor, K. and Zanjirian, A. (2017) Predictors of employment status among adults with Autism Spectrum Disorder. *Work*, **56**, 345-355.
9. Ward, B., Tanner, B.S., Mandleco, B., Dyches, T.T. and Freeborn, D. (2016) Sibling Experiences: Living with Young Persons with Autism Spectrum Disorders. *Pediatr Nurs*, **42**, 69-76.
10. Devlin, B. and Scherer, S.W. (2012) Genetic architecture in autism spectrum disorder. *Curr Opin Genet Dev*, **22**, 229-237.
11. Ornoy, A., Weinstein-Fudim, L. and Ergaz, Z. (2015) Prenatal factors associated with autism spectrum disorder (ASD). *Reprod Toxicol*, **56**, 155-169.
12. Getahun, D., Fassett, M.J., Peltier, M.R., Wing, D.A., Xiang, A.H., Chiu, V. and Jacobsen, S.J. (2017) Association of Perinatal Risk Factors with Autism Spectrum Disorder. *Am J Perinatol*, **34**, 295-304.
13. Masi, A., Glozier, N., Dale, R. and Guastella, A.J. (2017) The Immune System, Cytokines, and Biomarkers in Autism Spectrum Disorder. *Neurosci Bull*, in press.
14. Minarovits, J., Banati, F., Szenthe, K. and Niller, H.H. (2016) Epigenetic Regulation. *Adv Exp Med Biol*, **879**, 1-25.
15. Al-Mahdawi, S., Virmouni, S.A. and Pook, M.A. (2014) The emerging role of 5-hydroxymethylcytosine in neurodegenerative diseases. *Front Neurosci*, **8**, 397.

16. Kohli, R.M. and Zhang, Y. (2013) TET enzymes, TDG and the dynamics of DNA demethylation. *Nature*, **502**, 472-479.

17. Vijayakumar, N.T. and Judy, M.V. (2016) Autism spectrum disorders: Integration of the genome, transcriptome and the environment. *J Neurol Sci*, **364**, 167-176.

18. Szulwach, K.E., Li, X., Li, Y., Song, C.X., Wu, H., Dai, Q., Irier, H., Upadhyay, A.K., Gearing, M., Levey, A.I. et al. (2011) 5-hmC-mediated epigenetic dynamics during postnatal neurodevelopment and aging. *Nat Neurosci*, **14**, 1607-1616.

19. Esanov, R., Andrade, N.S., Bennison, S., Wahlestedt, C. and Zeier, Z. (2016) The FMR1 promoter is selectively hydroxymethylated in primary neurons of fragile X syndrome patients. *Hum Mol Genet*, **25**, 4870-4880.

20. Kohli, R.M. and Zhang, Y. (2013) TET enzymes, TDG and the dynamics of DNA demethylation. *Nature*, **502**, 472-479.

21. Al-Mahdawi, S., Virmouni, S.A. and Pook, M.A. (2014) The emerging role of 5-hydroxymethylcytosine in neurodegenerative diseases. *Front Neurosci*, **8**, 397.

22. Kazdoba, T.M., Leach, P.T., Silverman, J.L. and Crawley, J.N. (2014) Modeling fragile X syndrome in the Fmr1 knockout mouse. *Intractable Rare Dis Res*, **3**, 118-133.

23. Bear, M.F., Huber, K.M. and Warren, S.T. (2004) The mGluR theory of fragile X mental retardation. *Trends Neurosci*, **27**, 370-377.

24. Bakker, C.E., Verheij, C., Willemsen, R., van der Helm, R., Oerlemans, F., Vermey, M., Bygrave, A., Hoogeveen, T., Oostra, B.A., Reyniers, E., et al. (1994) Fmr1 knockout mice: a model to study fragile X mental retardation. The Dutch-Belgian Fragile X Consortium. *Cell*, **78**, 23-33.

25. Neul, J.L. (2012) The relationship of Rett syndrome and MECP2 disorders to autism. *Dialogues Clin Neurosci*, **14**, 253-262.

26. Guy, J., Hendrich, B., Holmes, M., Martin, J.E. and Bird, A. (2001) A mouse Mecp2-null mutation causes neurological symptoms that mimic Rett syndrome. *Nat Genet*, **27**, 322-326.

27. Renieri, A., Meloni, I., Longo, I., Ariani, F., Mari, F., Pescucci, C. and Cambi, F. (2003) Rett syndrome: the complex nature of a monogenic disease. *J Mol Med (Berl)*, **81**, 346-354.

28. Clouaire, T. and Stancheva, I. (2008) Methyl-CpG binding proteins: specialized transcriptional repressors or structural components of chromatin? *Cell Mol Life Sci*, **65**, 1509-1522.

29. Galvez, R. and Greenough, W.T. (2005) Sequence of abnormal dendritic spine development in primary somatosensory cortex of a mouse model of the fragile X mental retardation syndrome. *Am J Med Genet A*, **135**, 155-160.

30. Reynolds, C.D., Nolan, S.O., Jefferson, T. and Lugo, J.N. (2016) Sex-specific and genotype-specific differences in vocalization development in FMR1 knockout mice. *Neuroreport*, **27**, 1331-1335.

31. Park, M.J., Aja, S., Li, Q., Degano, A.L., Penati, J., Zhuo, J., Roe, C.R. and Ronnett, G.V. (2014) Anaplerotic triheptanoin diet enhances mitochondrial substrate use to remodel the metabolome and improve lifespan, motor function, and sociability in MeCP2-null mice. *PLoS One*, **9**, e109527.

32. Kang, J.Y., Chadchankar, J., Vien, T.N., Mighdoll, M.I., Hyde, T.M., Mather, R.J., Deeb, T.Z., Pangalos, M.N., Brandon, N.J., Dunlop, J. et al. (2017) Deficits in the activity of

presynaptic gamma-Aminobutyric acid type B receptors contribute to altered neuronal excitability in Fragile X Syndrome. *J Biol Chem*, in press.

33. Medrihan, L., Tantalaki, E., Aramuni, G., Sargsyan, V., Dudanova, I., Missler, M. and Zhang, W. (2008) Early defects of GABAergic synapses in the brain stem of a MeCP2 mouse model of Rett syndrome. *J Neurophysiol*, **99**, 112-121.
34. Na, E.S., Nelson, E.D., Adachi, M., Autry, A.E., Mahgoub, M.A., Kavalali, E.T. and Monteggia, L.M. (2012) A mouse model for MeCP2 duplication syndrome: MeCP2 overexpression impairs learning and memory and synaptic transmission. *J Neurosci*, **32**, 3109-3117.
35. Chen, J.A., Penagarikano, O., Belgard, T.G., Swarup, V. and Geschwind, D.H. (2015) The emerging picture of autism spectrum disorder: genetics and pathology. *Annu Rev Pathol*, **10**, 111-144.
36. Li, Y., Song, C.X., He, C. and Jin, P. (2012) Selective capture of 5-hydroxymethylcytosine from genomic DNA. *J Vis Exp*, in press.
37. Abrahams, B.S., Arking, D.E., Campbell, D.B., Mefford, H.C., Morrow, E.M., Weiss, L.A., Menashe, I., Wadkins, T., Banerjee-Basu, S. and Packer, A. (2013) SFARI Gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs). *Mol Autism*, **4**, 36.
38. Shu, L., Sun, W., Li, L., Xu, Z., Lin, L., Xie, P., Shen, H., Huang, L., Xu, Q., Jin, P. et al. (2016) Genome-wide alteration of 5-hydroxymenthylcytosine in a mouse model of Alzheimer's disease. *BMC Genomics*, **17**, 381.

39. Wen, L., Li, X., Yan, L., Tan, Y., Li, R., Zhao, Y., Wang, Y., Xie, J., Zhang, Y., Song, C. et al. (2014) Whole-genome analysis of 5-hydroxymethylcytosine and 5-methylcytosine at base resolution in the human brain. *Genome Biol*, **15**, R49.

40. Li, W., Xu, X. and Pozzo-Miller, L. (2016) Excitatory synapses are stronger in the hippocampus of Rett syndrome mice due to altered synaptic trafficking of AMPA-type glutamate receptors. *Proc Natl Acad Sci U S A*, **113**, E1575-1584.

41. Tabet, R., Vitale, N. and Moine, H. (2016) Fragile X syndrome: Are signaling lipids the missing culprits? *Biochimie*, **130**, 188-194.

42. Fritsch, B., Reis, J., Gasior, M., Kaminski, R.M. and Rogawski, M.A. (2014) Role of GluK1 kainate receptors in seizures, epileptic discharges, and epileptogenesis. *J Neurosci*, **34**, 5765-5775.

43. Yau, S.Y., Bostrom, C.A., Chiu, J., Fontaine, C.J., Sawchuk, S., Meconi, A., Wortman, R.C., Truesdell, E., Truesdell, A., Chiu, C. et al. (2016) Impaired bidirectional NMDA receptor dependent synaptic plasticity in the dentate gyrus of adult female Fmr1 heterozygous knockout mice. *Neurobiol Dis*, **96**, 261-270.

44. Lo, F.S., Blue, M.E. and Erzurumlu, R.S. (2016) Enhancement of postsynaptic GABAA and extrasynaptic NMDA receptor-mediated responses in the barrel cortex of Mecp2-null mice. *J Neurophysiol*, **115**, 1298-1306.

45. Spector, A.A. (2009) Arachidonic acid cytochrome P450 epoxygenase pathway. *J Lipid Res*, **50 Suppl**, S52-56.

46. Tao, J., Wu, H., Coronado, A.A., de Laittre, E., Osterweil, E.K., Zhang, Y. and Bear, M.F. (2016) Negative Allosteric Modulation of mGluR5 Partially Corrects Pathophysiology in a Mouse Model of Rett Syndrome. *J Neurosci*, **36**, 11946-11958.

47. Huang, T.N. and Hsueh, Y.P. (2017) Calcium/calmodulin-dependent serine protein kinase (CASK), a protein implicated in mental retardation and autism-spectrum disorders, interacts with T-Brain-1 (TBR1) to control extinction of associative memory in male mice. *J Psychiatry Neurosci*, **42**, 37-47.

48. Ortiz-Barahona, A., Villar, D., Pescador, N., Amigo, J. and del Peso, L. (2010) Genome-wide identification of hypoxia-inducible factor binding sites and target genes by a probabilistic model integrating transcription-profiling data and in silico binding site prediction. *Nucleic Acids Res*, **38**, 2332-2345.

49. Peier, A.M., McIlwain, K.L., Kenneson, A., Warren, S.T., Paylor, R. and Nelson, D.L. (2000) (Over)correction of FMR1 deficiency with YAC transgenics: behavioral and physical features. *Hum Mol Genet*, **9**, 1145-1159.

50. Langmead, B. and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods*, **9**, 357-359.

51. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H. and Glass, C.K. (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell*, **38**, 576-589.

52. Quinlan, A.R. and Hall, I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841-842.

53. Robinson, M.D., McCarthy, D.J. and Smyth, G.K. (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, **26**, 139-140.

54. Huang da, W., Sherman, B.T. and Lempicki, R.A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*, **4**, 44-57.

55. Huang da, W., Sherman, B.T. and Lempicki, R.A. (2009) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res*, **37**, 1-13.

56. Shen, L., Shao, N., Liu, X. and Nestler, E. (2014) ngs.plot: Quick mining and visualization of next-generation sequencing data by integrating genomic databases. *BMC Genomics*, **15**, 284.

57. Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W. and Noble, W.S. (2009) MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res*, **37**, W202-208.

58. Bailey, T.L. and Machanick, P. (2012) Inferring direct DNA binding from ChIP-seq. *Nucleic Acids Res*, **40**, e128.