

Erosion of somatic tissue identity with loss of the X-linked intellectual disability factor KDM5C

3

4 Abstract

It is currently unclear why mutations in numerous chromatin-modifying enzymes cause neurodevelopmental disorders (NDDs). While some chromatin regulators are important for neuron-specific gene expression, loss of repressive chromatin regulators can also lead to the aberrant transcription of tissue-specific genes outside of their intended context. Because very few studies have characterized these aberrant, tissue-specific transcripts, the mechanisms behind their dysregulation and their functional consequences are largely unknown. Here, we explored this crisis in cellular identity in NDDs by investigating the ectopic expression of germline (e.g. sperm and egg) genes in mice lacking lysine demethylase 5c (KDM5C, also known as JARID1C or SMCX), an eraser of histone 3 lysine 4 di and tri-methylation (H3K4me2/3). We found male *Kdm5c* knockout (-KO) mice, which recapitulate key behavioral phenotypes of Claes-Jensen X-linked intellectual disability (MRXSCJ, OMIM: 300534), aberrantly express many liver, muscle, ovary, and testis genes within the amygdala and hippocampus. Many genes expressed in the *Kdm5c*-KO brain were typically unique to germ cells, indicating an erosion of the soma-germline boundary. Using germ cell-depleted RNA sequencing datasets, we curated a list of germline-enriched genes for unbiased characterization of KDM5C in germline gene repression. Germline genes are typically decommissioned in somatic lineages in the post-implantation epiblast, yet *Kdm5c*-KO epiblast-like cells (EpiLCs) aberrantly express key regulators of germline identity and meiosis, including *Dazl* and *Stra8*. We found that while KDM5C promotes the initial placement of DNA methylation at a subset of germline gene promoters in EpiLCs, germline genes can also become activated independent of direct KDM5C regulation. Intriguingly, many of the germline transcripts expressed in the *Kdm5c*-KO brain were late-stage spermatogenesis genes that are not directly bound by KDM5C. This suggests germline developmental programs can progress ectopically in the background of typical *Kdm5c*-KO development. Ultimately, this work provides novel insight into the demarcation of somatic and germline lineages in mammals while also characterizing the cellular identity crisis within a chromatin-linked neurodevelopmental disorder.

28 Introduction

29 A single genome holds the instructions to generate the myriad of cell types found within the adult organism.
30 This is, in part, accomplished by chromatin regulators that can either promote or impede lineage-specific
31 gene expression through DNA and histone modifications^{1,2}. While many chromatin regulators were initially
32 identified for their roles in shaping cellular and tissue identity^{3–5}, recent advancements in next generation
33 sequencing unexpectedly revealed many neurodevelopmental disorders (NDDs) are caused by mutations in
34 chromatin regulators⁶. Several studies have suggested this connection between chromatin regulators and
35 neurodevelopment is due to their regulation of brain-specific genes, such as orchestrating transcriptional
36 programs for synaptic maturation⁷ and transitioning between neuronal and glial fates during neural precursor
37 differentiation⁸. However, loss of some chromatin regulators can also result in the ectopic transcription of
38 tissue-specific genes outside of their target environment, such as the misexpression of liver-specific genes
39 within adult neurons⁹. Very few studies have investigated this cellular identity crisis in chromatin-linked
40 NDDs^{9,10} and it is currently unknown if partial loss of brain identity contributes to neurodevelopmental
41 impairments.

42 To elucidate the role of tissue identity in chromatin-linked neurodevelopmental disorders, it is essential
43 to first characterize the types of genes dysregulated and the molecular mechanisms governing their de-
44 repression. In this study, we characterized the aberrant expression of tissue-enriched genes with loss of
45 lysine demethylase 5C (KDM5C). KDM5C, , also known as SMCX or JARID1C, is a chromatin regulator that
46 can repress gene expression through erasure of histone 3 lysine 4 di- and trimethylation (H3K4me2/3)¹¹,
47 marks enriched at active gene promoters. Pathogenic mutations in *KDM5C* cause Intellectual Developmental
48 Disorder, X-linked, Syndromic, Claes-Jensen Type (MRXSCJ, OMIM: 300534), whose predominant features
49 include intellectual disability, seizures, aberrant aggression, and autistic behaviors^{12–14}. *Kdm5c* knockout
50 (-KO) mice recapitulate key MRXSCJ patient phenotypes, including hyperaggression, increased seizure
51 propensity, and learning impairments^{10,15}. Unexpectedly, RNA sequencing (RNA-seq) of the *Kdm5c*-KO
52 hippocampus revealed ectopic expression of testis genes within the brain¹⁰. It is currently unknown what types
53 of testis genes are dysregulated, when in *Kdm5c*-KO development this begins, and if other tissue-specific
54 genes are also aberrantly expressed with KDM5C loss.

55 The testis contains both germ cells (cells that pass on their genetic material, e.g. sperm) and somatic
56 cells (cells that perform all other bodily functions). Distinguishing between the germline and the soma is a
57 key feature of multicellularity and that typically occurs during early embryogenesis. In mammals, chromatin
58 regulators play a key role in decommissioning germline genes in somatic cells during the transition from naïve
59 to primed pluripotency. Initially, repressive histone H2A lysine 119 monoubiquitination (H2AK119ub1)¹⁶,
60 histone 3 lysine 9 trimethylation (H3K9me3)^{16,17}, are placed at germline gene promoters in embryonic stem
61 cells and are then decorated with DNA CpG methylation^{17–19} in the post-implantation embryo. How KDM5C
62 interacts with known germline gene silencers during this time point is currently unknown. Additionally, most

63 studies have focused on select genes important for early germ cell development rather than germline-enriched genes as a whole given the lack of a curated list of germline-enriched genes. Thus, it is unclear if the mechanism of repression differs for certain classes of germline genes, e.g. meiotic genes versus spermatid differentiation genes. Systematically characterizing KDM5C's role in germline gene repression during early embryogenesis will unveil key mechanisms underlying the demarcation between soma and germline identity and while also providing molecular footholds to test the impact of ectopic germline genes on neurodevelopment.

70 To illuminate KDM5C's role in tissue identity, here we characterized the aberrant transcription of tissue-enriched genes within the *Kdm5c*-KO brain and epiblast-like stem cells (EpiLCs), an *in vitro* model of the post-implantation embryo. We observed general dysregulation of tissue-enriched genes in both the *Kdm5c*-KO brain and EpiLCs, including misexpression of genes typically highly enriched within the testis, liver, muscle, and ovary. Both the *Kdm5c*-KO amygdala and hippocampus had significant enrichment of testis genes. We demonstrated testis and ovary-enriched genes are germline genes by generating a list of germline enriched-genes using germ cell-depleted RNA sequencing datasets. Using this curated list, we found *Kdm5c*-KO EpiLCs aberrantly expressed key drivers of germline identity and meiosis, including *Dazl* and *Stra8*, while the *Kdm5c*-KO brain primarily expressed germline genes important for late spermatogenesis. Additionally, KDM5C was highly enriched at germline gene promoters in EpiLCs but only bound to a subset of germline genes expressed in *Kdm5c*-KO cells, indicating germline-enriched mRNAs can be aberrantly transcribed through indirect mechanisms. Finally, we found KDM5C promotes the long-term silencing of germline genes in somatic cells by aiding the placement of DNA CpG methylation in EpiLCs. Thus, we propose KDM5C plays a fundamental role in the development of tissue identity during early embryogenesis, including the establishment of the soma-germline boundary.

85 Results

86 **Tissue-enriched genes, including testis genes, are aberrantly expressed in the 87 *Kdm5c*-KO brain**

88 Previous RNA sequencing (RNA-seq) in the adult hippocampus revealed ectopic expression of some testis genes within the *Kdm5c* knockout (-KO) brain¹⁰. Since tissue-enriched genes have not been systematically characterized in the *Kdm5c*-KO brain, if the testis is the only tissue type misexpressed and if loss of tissue identity a major consequence of KDM5C loss. Therefore, we globally assessed the expression of genes enriched in 17 mouse tissues²⁰ in our published mRNA-seq datasets of the adult amygdala and hippocampus in mice with constitutive knockout of *Kdm5c*²¹.

94 We found a large proportion of genes that are significantly upregulated within the *Kdm5c*-KO brain (DESeq2²², log2 fold change > 0.5, q < 0.1) are enriched in non-brain tissues in wild-type mice (Amygdala:

96 35%, Hippocampus: 24%) (Figure 1A-B). The majority of tissue-enriched differentially expressed genes
97 (tissue-enriched DEGs) were testis genes (Figure 1A-C). Even though the testis has the largest total number
98 of tissue-biased genes compared to any other tissue (2,496 genes), testis-biased DEGs were significantly
99 enriched for both brain regions (Amygdala $p = 1.83e-05$, Odds Ratio = 5.13; Hippocampus $p = 4.26e-11$,
100 Odds Ratio = 4.45, Fisher's Exact Test). One example of a testis-enriched DEG is *FK506 binding protein 6*
101 (*Fkbp6*), a known regulator of piRNA expression and meiosis in germ cells^{23,24} (Figure 1C).

102 In addition to the high enrichment of testis genes, we also identified aberrant expression of other tissue-
103 enriched genes within the *Kdm5c*-KO brain. Although the *Kdm5c*-KO mice we examined were male, we
104 observed significant enrichment of ovary-biased DEGs in both the amygdala and hippocampus (Amygdala
105 $p = 0.00574$, Odds Ratio = 18.7; Hippocampus $p = 0.048$, Odds Ratio = 5.88, Fisher's Exact) (Figure 1D).
106 Ovary-enriched DEGs included *Zygotic arrest 1* (*Zar1*), sequesters mRNAs in oocytes for meiotic maturation
107 and early zygote development²⁵ (Figure 1D). Although not consistent across brain regions, we also found
108 significant enrichment of DEGS biased towards two non-gonadal tissues - the liver (Amygdala $p = 0.0398$,
109 Odds Ratio = 6.58, Fisher's Exact Test) and the muscle (Hippocampus $p = 0.0104$, Odds Ratio = 6.95,
110 Fisher's Exact Test). An example of a liver-biased DEG dysregulated in both the hippocampus and amygdala
111 is *Apolipoprotein C-I* (*Apoc1*), a lipoprotein metabolism and transport gene²⁶ (Figure 1E). Testis, ovary,
112 and liver-enriched DEGs showed little to no expression in the developing and adult wild-type brain, yet our
113 mRNA-seq data indicate they are polyadenylated and spliced into mature transcripts (Figure 1C-E). Of
114 note, we did not observe enrichment of brain-enriched genes (Amygdala $p = 1$; Hippocampus $p = 0.74$,
115 Fisher's Exact), despite the fact these are brain samples and the brain has the second highest total number
116 of tissue-enriched genes (708 genes). Together, these results suggest misexpression of tissue-enriched
117 genes within the brain is a major effect of KDM5C loss.

118 **Germline genes are misexpressed in the *Kdm5c*-KO brain**

119 • Intriguingly, some of the ectopic testis transcripts identified within the *Kdm5c*-KO hippocampus have
120 known functions unique to germ cells¹⁰, suggesting KDM5C may play a role in demarcating somatic
121 versus germline identity.

122 The testis contains both germ cells (meiotic cells, e.g. spermatogonia) and somatic cells (non-meiotic,
123 e.g. Leydig cells) that support hormone production and germline functions. To determine if *Kdm5c*-KO
124 testis-enriched DEGs are somatic or germline genes, we first assessed their known functions through
125 gene ontology analysis. We found *Kdm5c*-KO testis-enriched DEGs high enrichment of germline-relevant
126 ontologies, including spermatid development (GO: 0007286, $p.adjust = 6.2e-12$) and sperm axoneme
127 assembly (GO: 0007288, $p.adjust = 2.45e-14$) (Figure 2A).

128 To further validate if these testis DEGs are truly germline genes, we then evaluated their expression
129 in somatic versus germ cells within the testis. We first compared their expression in the testis with germ

130 cell depletion²⁷, which was accomplished by heterozygous *W* and *Wv* mutations in the enzymatic domain
131 of *c-Kit* (*Kit*^{W/Wv}) that prevent the maturation of germ cells²⁸. Almost all *Kdm5c*-KO testis-enriched DEGs
132 lost expression with germ cell depletion (Figure 2B). We then assessed testis-enriched DEG expression
133 in a published single cell RNA-seq dataset that identified cell type-specific markers within the testis²⁹. We
134 found that while some testis-enriched DEGs were classified as specific markers for different germ cell types
135 (e.g. spermatogonia, spermatocytes, round spermatids, and elongating spermatids) none marked somatic
136 cells (Figure 2C). Together, these data demonstrate that the *Kdm5c*-KO brain aberrantly expresses germline
137 genes, reflecting an erosion between somatic versus germline identity.

138 We then wanted to more deeply characterize germline gene misexpression with *Kdm5c* loss, but lacked
139 a comprehensive list of mouse germline-enriched genes. To facilitate downstream analyses, we generated
140 a list of germline-enriched genes using currently available RNA-seq datasets in *Kit*^{W/Wv} mice. Wild-type
141 and *Kit*^{W/Wv} datasets included males and females at embryonic day 12, 14, and 16³⁰, as well as adult
142 male testes²⁷. We defined genes as germline-enriched if their expression met the following criteria: 1)
143 their expression is greater than 1 FPKM in wild-type gonads 2) their expression in any adult wild-type,
144 non-gonadal tissue²⁰ does not exceed 20% of their maximum expression in the wild-type germline, and
145 3) their expression in the germ cell-depleted germline, for any sex or time point, does not exceed 20% of
146 their maximum expression in the wild-type germline. These criteria yielded 1,288 germline-enriched genes
147 (Figure 2D), which was hereafter used as a resource to globally assess germline gene misexpression with
148 *Kdm5c* loss (Supplementary table 1).

149 ***Kdm5c*-KO epiblast-like cells aberrantly express master regulators of germline
150 identity**

151 Germ cells are typically distinguished from somatic cells soon after the embryo implants into the uter-
152 ine wall^{31,32} when a subset of epiblast stem cells become the primordial germ cells (PGCs) while the
153 remainder differentiate into the ectoderm, mesoderm, and endoderm to form the somatic tissues³³. This
154 developmental time point can be modeled *in vitro* through differentiation of embryonic stem cells (ESCs) into
155 post-implantation epiblast-like stem cells (EpiLCs) (Figure 3A, top). While some germline-enriched genes
156 are also expressed in embryonic stem cells (ESCs) and in the 2-cell stage^{34–36}, they are silenced as they
157 differentiate into EpiLCs¹⁷. Therefore, we tested if KDM5C was necessary for silencing germline genes in
158 the post-implantation embryo by evaluating the impact of *Kdm5c* loss in EpiLCs.

159 We first identified *Kdm5c*-KO EpiLC DEGs in our previously published RNA-seq dataset³⁷ (DESeq2, log2
160 fold change > 0.5, q < 0.1) and assessed tissue-enriched gene expression. Similar to the *Kdm5c*-KO brain,
161 we observed general dysregulation of tissue-enriched genes, with the largest number of genes belonging
162 to the brain and testis, although they were not significantly enriched (Figure 3B). Although we observed
163 aberrant expression of many tissue-enriched genes, including 54 germline-enriched genes, we did not find

164 any significant difference in primed pluripotency genes, such as *Dnmt3b*, *Fgf5*, *Pou3f1*, and *Otx2* (Figure 3C).
165 We also did not observe any gross changes in *Kdm5c*-KO cell morphology during ESC to EpiLC differentiation
166 (Figure 3D), altogether indicating KDM5C loss does not impair EpiLC formation.

167 To determine if germline DEGs are constitutively dysregulated or if they can change over the course of
168 development, we compared EpiLC germline DEGs to those expressed in the *Kdm5c*-KO brain.
169 We found the majority of germline DEGs expressed in EpiLCs and the brain were unique, with only *Cyct*
170 shared across all sequencing datasets (Figure 3E-F). We then evaluated the known functions of EpiLC and
171 brain germline DEGs and found particularly high enrichment of meiosis-related gene ontologies in EpiLCs
172 (Figure 3G), such as meiotic cell cycle process (GO: 1903046, p.adjust = 1.59e-08) and meiotic nuclear
173 division (GO: 0140013, p.adjust = 9.76e-09). While there was modest enrichment of meiotic gene ontologies
174 in both brain regions, the *Kdm5c*-KO hippocampus primarily expressed late-stage spermatogenesis genes,
175 such those involved in the sperm axoneme structure.

176 Interestingly, DEGs unique to *Kdm5c*-KO EpiLCs included key drivers of germline identity, such as
177 *Stimulated by retinoic acid 8 (Stra8)* and *Deleted in azoospermia like (Dazl)* (Figure 3H). These genes are
178 typically expressed during embryonic germ cell development to commit PGCs to the germline fate, but are
179 also expressed later in life to trigger meiotic gene expression programs^{38–40}. Of note, some germline genes,
180 including *Dazl*, are also expressed in the two-cell embryo^{35,41}. However, we did not see misexpression of
181 two-cell embryo-specific genes, like *Duxf3 (Dux)* ($q = 0.337$) and *Zscan4d* ($q = 0.381$), indicating *Kdm5c*-KO
182 in EpiLCs do not revert back to a 2-cell state (Figure 3H).

183 **KDM5C binds to a subset of germline gene promoters during early embryogenesis**

184 Previous work suggests KDM5C represses germline genes during early development, as re-expression
185 of KDM5C in knockout neuronal cultures fails to suppress two hippocampal germline DEGs¹⁰. There is some
186 evidence KDM5C binds to select germline gene promoters in embryonic stem cells (ESCs) but is lost in
187 neurons¹⁰. However, the lack of a comprehensive list of germline-enriched genes prohibited systematic
188 characterization of KDM5C binding at germline gene promoters. Thus, it is currently unclear if KDM5C is
189 enriched at germline gene promoters, what types of germline genes KDM5C regulates, and if its binding is
190 maintained at any germline genes in neurons.

191 To further characterize KDM5C's role in germline gene silencing, we analyzed KDM5C chromatin
192 immunoprecipitation followed by DNA sequencing (ChIP-seq) datasets in EpiLCs³⁷ and primary neuron
193 cultures (PNCs) from the cortex and hippocampus¹⁵. EpiLCs had a higher total number of KDM5C peaks
194 than PNCs (EpiLCs: 5,808, PNCs: 1,276, MACS2 $q < 0.1$ and fold enrichment > 1 , removal of *Kdm5c*-KO
195 false positive peaks) and KDM5C was primarily localized to gene promoters in both cell types (EpiLCs: 4,190,
196 PNCs: 745 +/- 500bp from the TSS), although PNCs showed increased localization to non-promoter regions
197 (Figure 4A).

198 The majority of promoters bound by KDM5C in PNCs were also bound in EpiLCs (513 shared promoters),

however a large portion of gene promoters were only bound by KDM5C in EpiLCs (3677 EpiLC only promoters) (Figure 4B). We then performed gene ontology analysis to compare the known functions of genes with KDM5C bound to their promoter in EpiLCs and PNCs. While there were no ontologies significantly enriched for genes only bound by KDM5C in PNCs, gene ontologies for genes shared between PNCs and EpiLCs were enriched for functions involving nucleic acid turnover, such as deoxyribonucleotide metabolic process (GO:0009262, p.adjust = 8.28e-05) (Figure 4C). Germline-specific ontologies were only enriched in promoters unique to EpiLCs, such as meiotic nuclear division (GO: 0007127 p.adjust = 6.77e-16) and meiotic cell cycle process (GO:1903046, p.adjust = 5.05e-16) (Figure 3C). We then evaluated KDM5C binding around the transcription start site (TSS) of all germline-enriched genes. While KDM5C was bound to a subset of germline gene promoters in EpiLCs, it was not bound to any in PNCs (Figure 4D-E). Together, these results demonstrate KDM5C is significantly enriched at a subset of germline gene promoters in EpiLCs, including meiotic genes, but does not regulate germline genes in neurons.

We then compared KDM5C binding at RNA-seq DEG promoters to determine if the germline mRNAs expressed in the *Kdm5c*-KO brain and EpiLCs are direct targets of KDM5C (Figure 4F). About one third of EpiLC-specific and brain-specific (hippocampus, amygdala, or both) germline DEGs were bound by KDM5C in EpiLCs (EpiLC only: 36%, Brain only: 33.3%). Again, we did not observe any KDM5C binding at germline gene promoters in PNCs, even for brain-specific DEGs (Figure 4H). Some notable differences in KDM5C binding for EpiLC-specific DEGs included *Dazl* and *Stra8*, two key drivers of germline identity discussed above. Although both mRNAs are expressed in *Kdm5c*-KO EpiLCs, KDM5C is only bound to *Dazl*'s promoter and not *Stra8*'s in EpiLCs (Figure 4F). In contrast to the unique DEGs, 3 out of the 4 genes dysregulated in both the brain and EpiLCs are bound by KDM5C around the TSS (Figure 4G). Altogether, this suggests KDM5C decommissions germline genes in EpiLCs but is no longer recruited to germline gene promoters in neurons. Furthermore, majority of germline mRNAs expressed in *Kdm5c*-KO cells are dysregulated independent of direct KDM5C recruitment to their promoters in EpiLCs.

KDM5C promotes *de novo* DNA methylation at germline genes

In the early embryo, germline gene promoters are initially decorated with repressive histone modifications before long-term silencing by *de novo* DNA CpG methylation (CpGme)^{16,17,42}. Our results above suggest KDM5C also acts during this time period, however how KDM5C interacts with other germline gene silencing mechanism is currently unclear. KDM5C is generally thought to suppress transcription through erasure of histone 3 lysine 4 di- and trimethylation (H3K4me2/3), marks enriched at active gene promoters¹¹. However, KDM5C's catalytic activity was recently shown to not be required for suppressing *Dazl* in ESCs⁴¹. Since H3K4me3 impedes *de novo* CpGme placement^{43,44}, KDM5C's catalytic activity may instead be required for long-term germline gene silencing. In support of this, CpGme is markedly reduced at two germline gene promoters in the adult hippocampus¹⁰. However, because KDM5C's role in germline gene repression has only been characterized at select germline genes in ESCs and in the mature brain, it is currently unclear to

234 what extent KDM5C is involved during initial CpGme placement.

235 To elucidate KDM5C's role in germline gene silencing, we first characterized KDM5C's substrates, histone
236 3 lysine 4 di- and trimethylation (H3K4me2/3) in our previously published ChIP-seq datasets in wild type and
237 *Kdm5c*-KO amygdala²¹ and EpiLCs³⁷. In congruence with previous work in the *Kdm5c*-KO hippocampus¹⁰,
238 we observed aberrant accumulation of H3K4me3 around the transcription start site (TSS) of germline genes
239 in the *Kdm5c*-KO amygdala (Figure 5A). We additionally found a marked increase in H3K4me2 around the
240 TSS of germline genes in *Kdm5c*-KO EpiLCs (Figure 5B).

241 We then evaluated KDM5C's role in germline gene silencing during ESC to EpiLC differentiation. We
242 first characterized KDM5C's mRNA and protein expression at 0, 24, and 48 hours of differentiation (Figure
243 5C). While *Kdm5c* mRNA steadily decreased from 0 to 48 hours of differentiation (Figure 5D), KDM5C
244 protein initially increased from 0 to 24 hours but then decreased to near knockout levels by 48 hours (Figure
245 5E). To determine KDM5C's role in the initial placement of DNA methylation at germline gene promoters,
246 we performed whole genome bisulfite sequencing (WGBS) in wild-type and *Kdm5c*-KO ESCs and 96 hour
247 extended EpiLCs (exEpiLCs). Although wild-type cells accumulated high levels of DNA methylation at
248 germline gene promoters over the course of ESCs to exEpiLC differentiation, DNA methylation was markedly
249 reduced in *Kdm5c*-KO exEpiLCs (Figure 5F).

250 • Altogether, this suggests KDM5C is dynamically regulated during ESC to EpiLC differentiation and
251 promotes germline gene silencing via removal of H3K4me2/3 during early embryogenesis.

252 Discussion

253 In the above study, we demonstrate KDM5C's pivotal role in the development of tissue identity, resulting in
254 substantial misexpression of tissue-enriched genes within the adult *Kdm5c*-KO hippocampus and amygdala.
255 We defined tissue-enriched genes based on the criteria used in a previously published dataset of 17 C57/Bl6J
256 mouse tissues²⁰, which defined genes as tissue-enriched if they had more than 4-fold higher expression
257 when compared to any other tissue. In addition to testis genes identified previously¹⁰, we found significant
258 enrichment of muscle, liver, and even ovary-biased genes aberrantly expressed within the male *Kdm5c*-KO
259 brain. It is currently unclear what the impact of these tissue-enriched genes are upon *Kdm5c*-KO brain
260 function and if they ultimately contribute to MRXSCJ-related phenotypes. While the majority of tissue-
261 enriched DEGs were testis and ovary genes with no known brain functions, select liver and muscle-biased
262 DEGs do have known roles in the wild-type brain. For example, *Apolipoprotein C-I* (*Apoc1*) is a lipid transport
263 gene highly enriched within the liver, but is also lowly expressed in the wild-type brain. Intriguingly, *Apoc1*
264 overexpression in the mouse brain can impair learning and memory⁴⁵ and is implicated in Alzheimer's
265 disease in humans⁴⁶. Our results suggest KDM5C fine-tunes the expression of tissue-enriched genes like
266 *Apoc1* to match the level required for proper brain function. However, further studies are required to determine
267 if these genes contribute to *Kdm5c*-KO neuronal impairments and MRXSCJ pathology.

268 A substantial portion of genes upregulated in *Kdm5c*-KO brain are testis genes that have no known
269 function within the brain. Through the use of publically available RNA-seq datasets, we demonstrated these
270 testis-enriched DEGs are typically unique to germ cells. Misexpression of germline genes in the brain sug-
271 gests *Kdm5c*-KOs fail to demarcate between somatic and germline lineages, a key feature of multicellularity
272 and sexual reproduction. Previous studies have demonstrated germline genes are decommissioned in
273 somatic cells by multiple chromatin-modifying enzymes, but have primarily focused on the repression of
274 key marker genes. To globally characterize KDM5C's role in germline gene repression, we curated a list
275 of mouse germline-enriched genes using publically available germ cell-depleted RNA-seq datasets. This
276 resource enabled us to identify 1) the extent of germline gene dysregulation in *Kdm5c*-KO cells, 2) the types
277 of germline genes misexpressed at different developmental time points, and 3) which types of germline
278 genes are directly or indirectly regulated by KDM5C.

279 Dysregulation of *Kdm5c*-KO tissue identity begins during early emrbyogenesis, as germline and other
280 tissue-enriched genes are aberrantly expressed in epiblast-like cells (EpiLCs). *In vivo*, germline genes
281 are typically decommissioned in epiblast stem cells and remain silenced as the epiblast differentiates
282 into the body's somatic tissues³³. However, a small subset of epiblast stem cells will receive signals to
283 reactivate germline gene expression to become the primordial germ cells (PGCs) that will ultimately form
284 the mature germline^{31,32}. This process can be mimicked *in vitro* by differentiating EpiLCs into primordial
285 germ cell-like cells (PGCLCs)⁴⁷. Thus, misexpression of germline genes in *Kdm5c*-KO EpiLCs might
286 suggest they are progressing beyond EpiLC differentiation and becoming PGCLCs. Yet, *Kdm5c*-KO EpiLCs
287 had proper expression of EpiLC marker genes for primed pluripotency and we observed no difference in
288 cellular morphology during ESC to EpiLC differentiation. Furthermore, we saw no significant change in *Otx2*
289 expression, an epiblast stem cell marker that is known to repress epiblast and EpiLC differentiation into PGC
290 and PGCLCs⁴⁸. This, in conjunction with *Kdm5c*-KO mice being viable, suggests germline gene expression
291 is occuring ectopically in conjunction with typical developmental programs, rather than a complete shift to
292 germline identity.

293 • change first sentence: Other chromatin regulators involved in germline gene repression act in ESCs
294 and EpiLCs. We observed this is true for KDM5C

295 We then globally characterized KDM5C binding at germline-enriched gene promoters through analysis
296 of KDM5C ChIP-seq in EpiLCs and primary hippocampus and cortex neuron cultures (PNCs). While we
297 observed high enrichment of KDM5C peaks at germline genes in EpiLCs, KDM5C was not bound to germline
298 gene promoters in PNCs. This suggests dysregulation of germline genes in the mature *Kdm5c*-KO brain is
299 due to loss of repression during embryogenesis, which is consistent with previous work that found introducing
300 human KDM5C into *Kdm5c*-KO PNCs does not repress germline transcripts¹⁰. Although KDM5C is enriched
301 at germline gene promoters in EpiLCs, KDM5C was only bound to about a third of EpiLC germline-enriched
302 DEGs. One notable DEG not bound by KDM5C is *Stra8*, a transcription factor activated by retinoic acid

303 signaling in germ cells that promotes meiotic initiation^{49,50}. Retinoic acid can only activate *Stra8* expression
304 when DAZL is present, which is a direct target of KDM5C and is aberrantly transcribed and translated in
305 *Kdm5c*-KO EpiLCs. This indicates germline genes can be aberrantly expressed in *Kdm5c*-KO cells indirectly
306 of KDM5C regulation through activation by other ectopic germline programs. Consistent with this idea, many
307 *Kdm5c*-KO EpiLC germline DEGs are important for early germ cell development and meiosis while those
308 expressed in the mature *Kdm5c*-KO brain are involved in late sperm development. Altogether, this indicates
309 ectopic germline programs are, to some extent, progressing through germ cell developmental stages over
310 the course of *Kdm5c*-KO development.

- 311 • Talk about different motifs if we do see differences there and if it explains direct vs indirect dysregulation
312 – Could also mention PRC1.6/KDM5C binds pcgf6 in ESCs
313 • Maybe talk about other regulators/if anything is known about them long term? Or save for 2nd paper.
314 – another germline repressor: <https://pubmed.ncbi.nlm.nih.gov/35345626/>
315 • Might be good to look at for retinoic acid paper (WT germ expression): <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0005654>

317 It is important to note that while we focused on KDM5C's regulation of germline genes, some germline-
318 enriched genes are also expressed in naïve ESCs and at the 2-cell stage due to their roles in pluripotency
319 and self-renewal. For example, although primarily known for committing PGCs to the germline fate and
320 regulating the translation of germline-specific RNAs, *Dazl* is also expressed in naïve ESCs *in vitro*³⁴, the
321 inner cell mass *in vivo*³⁴, and at the 2-cell stage⁵¹, but is silenced when cells transition from naïve to primed
322 pluripotency during ESC to EpiLC differentiation¹⁷. Very recently, two screens of *Dazl*-repressors in ESCs
323 identified KDM5C as a direct regulator of *Dazl* expression^{41,52}. Interestingly, one screen found *Kdm5c*-KO
324 ESCs also aberrantly expressed 2-cell-specific genes, indicating KDM5C negatively regulates totipotency in
325 ESCs⁴¹. We found that while KDM5C also represses *Dazl* expression in EpiLCs, *Kdm5c*-KO EpiLCs do not
326 express 2-cell specific genes like *Dux* and *Zscan4c*. Out of the four 2-cell regulators characterized in ESCs,
327 KDM5C was the only factor whose repression of *Dazl* was independent of *Dux* expression⁴¹. Together, this
328 suggests the 2-cell-like state in *Kdm5c*-KO ESCs is an indirect consequence of KDM5C's primary role in
329 germline gene repression, including germline genes that are also involved in pluripotency and self-renewal.

330 It is hypothesized distinct repressive chromatin marks are placed at germline gene promoters as the
331 embryo transitions from naïve to primed pluripotency. Initially, germline genes are repressed by placement of
332 histone 2A lysine 119 monoubiquitination (H2AK119ub1) and histone 3 lysine 9 trimethylation (H3K9me3)
333 in ESCs and then gain *de novo* DNA CpG methylation (CpGme) in EpiLCs^{17–19,42,53,54}. KDM5C may be
334 instead required to remove an active mark, histone 3 lysine 4 trimethylation (H3K4me3), since H3K4me3
335 can impede *de novo* CpGme placement^{43,44}. This is supported by previous work in the *Kdm5C*-KO adult
336 hippocampus, which found an increase in H3K4me3 around the transcription start site (TSS) of germline

337 DEGs and loss of CpGme at at least two germline promoters¹⁰. However, KDM5C's role in embryonic
338 germline gene silencing is still unclear, ask KDM5C's catalytic activity was recently found to be dispensible
339 for silencing DAZL in ESCs⁴¹. In this study, we observed a global increase in H3K4me3 around the TSS of
340 germline-enriched gene in the *Kdm5c*-KO amygdala and an increase in H3K4me2 in *Kdm5c*-KO EpiLCs. We
341 found KDM5C's expression is dynamically regulated during ESC to EpiLC differentiation and loss of KDM5C
342 leads to impaired placement of CpGme in extended EpiLCs. Altogether, this suggests KDM5C is necessary
343 during the transition from naïve to primed pluripotency to promote the initial placement of CpGme at CpG
344 islands for the long-term silencing of germline genes.

- 345 • Talk about impact - Other germline gene chromatin regulators, NDDs, cancer. Maybe talk about myt1l
346 • include cancer in there somewhere (Somatic misexpression of germline genes has been implicated in
347 many cancers.)

348 In summary, we demonstrate loss of the X-linked intellectual disability gene KDM5C leads to widespread
349 dysregulation of brain tissue identity, including misexpression of germline genes in the somatic brain. In
350 EpiLCs, KDM5C directly represses key drivers of germline identity like *Dazl*, by promoting intial CpG methy-
351 lation placement in the post-implantation embryo. However, germline genes can also become ectopically
352 expresssesed in *Kdm5c*-KO cells independent of direct KDM5C regulation, including the meiotic transcription
353 factor *Stra8*. These ectopic germline developmental programs can, to some extent, mimic typical germ
354 cell development, resulting in aberrant transcription early developmental and meiotic genes in *Kdm5c*-KO
355 EpiLCs and late-stage spermatogenesis genes in the *Kdm5c*-KO brain. Altogether, these results define
356 KDM5C's role in the demarcation between soma and gemrline identity and offers novel insight into how this
357 dysregulation of tissue identity changes over the course of development. Additionally, this study provides
358 the mechanistic foundation required to ultimately investigate the impact of aberrant germline identity upon
359 neurodevelopment.

360 Materials and Methods

361 Classifying tissue-enriched and germline-enriched genes

362 Tissue-enriched differentially expresssd genes were determined by their classification in a previously
363 published list of genes enriched in 17 male and female mouse tissues²⁰. This study defined expression as
364 greather than 1 Fragments Per Kilobase of transcript per Million mapped read (FPKM) and tissue enrichment
365 as at least 4-fold higher expression than any other tissue.

366 We curated a list of germline-enriched genes using an RNA-seq dataset from wild-type and germline-
367 depleted (*Kit^{W/W^v}*) male and female mouse embryos from embryonic day 12, 14, and 16³⁰, as well as adult
368 male testes²⁷. Germline-enriched genes met the following criteria: 1) their expression is greater than 1
369 FPKM in wild-type germline 2) their expression in any wild-type somatic tissues²⁰ does not exceed 20%

370 of maximum expression in wild-type germline, and 3) their expression in the germ cell-depleted ($Kit^{W/Wv}$)
371 germline, for any sex or time point, does not exceed 20% of maximum expression in wild-type germline.

372 **Cell culture**

373 We utilized our previously established cultures of male wild-type and *Kdm5c* knockout (-KO) embryonic
374 stem cells³⁷. Sex was confirmed by genotyping *Uba1* on the X and Y chromosomes with the following
375 primers: 5'-TGGATGGTGTGCCAATG-3', 5'-CACCTGCACGTTGCCCT-3'. Deletion of *Kdm5c* was
376 confirmed through the primers 5'-ATGCCCATATTAAAGAGTCCTG-3', 5'-TCTGCCTTGATGGGACTGTT-3',
377 and 5'-GGTTCTAACACTCACATAGTG-3'.

378 Embryonic stem cells (ESCs) were initially cultured in primed ESC (pESC) media consisting of Knock-
379 Out DMEM (Gibco#10829–018), fetal bovine serum (Gibco#A5209501), KnockOut serum replacement
380 (Invitrogen#10828–028), Glutamax (Gibco#35050-061), Anti-Anti (Gibco#15240-062), MEM Non-essential
381 amino acids (Gibco#11140-050), and beta-mercaptoethanol (Sigma#M7522). They were then transitioned
382 into ground-state “naive” ESCs (nESCs) by culturing for four passages in N2B27 media containing DMEM/F12
383 (Gibco#11330–032), Neurobasal media (Gibco#21103–049), Gluamax, Anti-Anti, N2 supplement (Invitro-
384 gen#17502048), and B27 supplement without vitamin A (Invitrogen#12587-010), and beta-mercaptoethanol.
385 Both pESC and nESC media were supplemented with the GSK3 inhibitor CHIR99021 (Sigma #SML1046-
386 5MG), the MEK inhibitor PD0325901 (Sigma #PZ0162-5MG), and leukemia inhibitory factor (LIF, Milli-
387 pore#ESG1107).

388 nESCs were differentiated into epiblast-like cells (EpiLCs, 48 hours) and extendend EpiLCs (exEpiLCs,
389 96 hours) by culturing in N2B27 media containing DMEM/F12, Neurobasal media, Gluamax, Anti-Anti,
390 N2 supplement, B27 supplement (Invitrogen#17504044), beta-mercaptoethanol, fibroblast growth factor 2
391 (FGF2, R&D Biotechne 233-FB), and activin A (R&D Biotechne 338AC050CF).

392 **Immunocytochemistry (ICC)**

393 ICC of DAZL in EpiLCs was performed by washing cells thrice with phosphobuffered saline (PBS), fixing
394 cells in 4% paraformaldehyde, washing thrice in PBS, blocking in PBS containing 0.3% Triton X-100, and 5%
395 fetal bovine serum for 1 hour, washing thrice with PBS, and incubating in primary antibody (Rabbit anti DAZL,
396 abcam ab34139, 1:200) in the blocking solution overnight at 4C. The next day cells were rinsed thrice with
397 PBS, incubated in secondary antibody (Alexafluor 488 Invitrogen #710369, 1:1000) in blocking buffer, rinsed
398 thrice in PBS, and then imaged. Images were taken blinded for genotype, chosen based on similar levels of
399 DAPI signal, and then quantified via ImageJ.

400 **RNA sequencing**

401 **Published datasets**

402 All published datasets are available at the Gene Expression Omnibus (GEO) [https://www.ncbi.nlm.nih.g](https://www.ncbi.nlm.nih.gov/geo)
403 ov/geo. Previously published RNA sequencing datasets analyzed in this study included the male wild-type
404 and *Kdm5c*-KO adult amygdala and hippocampus²¹ (available at GEO: GSE127722) and male wild-type and
405 *Kdm5c*-KO EpiLCs³⁷ (available at GSE: GSE96797).

406 **Alignment and analysis**

407 After ensuring read quality via FastQC (v0.11.8), reads were then mapped to the mm10 *Mus musculus*
408 genome (Gencode) using STAR (v2.5.3a), during which we removed duplicates and kept only uniquely
409 mapped reads. Count files were generated by FeatureCounts (Subread v1.5.0), and BAM files were
410 converted to bigwigs using deeptools (v3.1.3) and visualized by the UCSC genome browser. RStudio (v3.6.0)
411 was then used to analyze counts files by DESeq2 (v1.26.0)²² to identify differentially expressed genes
412 (DEGs) with a q-value (p-adjusted via FDR/Benjamini–Hochberg correction) less than 0.1 and a log2 fold
413 change greater than 0.5. For all DESeq2 analyses, log2 fold changes were calculated with IfcShrink using
414 the ashR package⁵⁵. MA-plots were generated by ggpubr (v0.4.0), and Eulerr diagrams were generated by
415 eulerr (v6.1.1). Boxplots and scatterplots were generated by ggpubr (v0.6.0) and ggplot2 (v3.3.2). Heatmaps
416 of gene expression were generated using the base R functions scale and hclust and visualized using the R
417 package ComplexHeatmap (v2.12.1). The Upset plot was generated via the package UpSetR (v1.4.0)⁵⁶.
418 Gene ontology (GO) analyses were performed by the R package enrichPlot (v1.16.2) using the biological
419 processes setting.

420 **Chromatin immunoprecipitation followed by DNA sequencing (ChIP-seq)**

421 We analyzed our previously published KDM5C ChIP-seq experiments in primary neuron cultures from
422 the cortex and hippocampus¹⁵ (available at GEO: GSE61036) and EpiLCs³⁷ (available at GEO: GSE96797). ChIP-seq
423 of histone 3 lysine 4 dimethylation in male EpiLCs³⁷ is also available at GEO: GSE96797. ChIP-seq
424 of histone 3 lysine 4 trimethylation in the male amygdala²¹ are available at GEO: GSE127817.

425 ChIP-seq reads were aligned to mm10 using Bowtie1 (v1.1.2) allowing up to two mismatches. Only
426 uniquely mapped reads were used for analysis. Peaks were called using MACS2 software (v2.1.0.20140616)
427 using input BAM files for normalization, with filters for a q-value < 0.1 and a fold enrichment > 1. Common
428 peak sets were obtained in R via DiffBind (v3.6.5), and count tables for the common peaks were generated
429 with the Bedtools multicov command. We removed “black-listed” genomic regions that often give aberrant
430 signals. Peak proximity to genome annotations was determined by ChIPSeeker (v1.32.1). Enriched motifs
431 were identified using HOMER⁵⁷. Gene ontology (GO) analyses were performed by the R package enrichPlot

432 (v1.16.2) using the biological processes setting. Average binding across the genome was visualized using
433 deeptools (v3.1.3). Bigwigs were visualized using the UCSC genome browser.

434 **Whole genome bisulfite sequencing (WGBS)**

435 **Data availability**

436 **Acknowledgements**

437 **References**

- 438 1. Strahl, B.D., and Allis, C.D. (2000). The language of covalent histone modifications. *Nature* **403**,
439 41–45. <https://doi.org/10.1038/47412>.
- 440 2. Jenuwein, T., and Allis, C.D. (2001). Translating the Histone Code. *Science* **293**, 1074–1080.
441 <https://doi.org/10.1126/science.1063127>.
- 442 3. Lewis, E.B. (1978). A gene complex controlling segmentation in *Drosophila*. *Nature* **276**, 565–570.
443 <https://doi.org/10.1038/276565a0>.
- 444 4. Kennison, J.A., and Tamkun, J.W. (1988). Dosage-dependent modifiers of polycomb and antennapedia
mutations in *Drosophila*. *Proc. Natl. Acad. Sci. U.S.A.* **85**, 8136–8140. <https://doi.org/10.1073/pnas.8>
445 5.21.8136.
- 446 5. Kassis, J.A., Kennison, J.A., and Tamkun, J.W. (2017). Polycomb and Trithorax Group Genes in
447 *Drosophila*. *Genetics* **206**, 1699–1725. <https://doi.org/10.1534/genetics.115.185116>.
- 448 6. Gabriele, M., Lopez Tobon, A., D'Agostino, G., and Testa, G. (2018). The chromatin basis of
neurodevelopmental disorders: Rethinking dysfunction along the molecular and temporal axes. *Prog
449 Neuropsychopharmacol Biol Psychiatry* **84**, 306–327. <https://doi.org/10.1016/j.pnpbp.2017.12.013>.
- 450 7. Zhou, Z., Hong, E.J., Cohen, S., Zhao, W.-N., Ho, H.-Y.H., Schmidt, L., Chen, W.G., Lin, Y., Savner,
E., Griffith, E.C., et al. (2006). Brain-specific phosphorylation of MeCP2 regulates activity-dependent
Bdnf transcription, dendritic growth, and spine maturation. *Neuron* **52**, 255–269. [https://doi.org/10.1
016/j.neuron.2006.09.037](https://doi.org/10.1).
- 452 8. Hirabayashi, Y., Suzuki, N., Tsuboi, M., Endo, T.A., Toyoda, T., Shinga, J., Koseki, H., Vidal, M., and
Gotoh, Y. (2009). Polycomb Limits the Neurogenic Competence of Neural Precursor Cells to Promote
453 Astrogenic Fate Transition. *Neuron* **63**, 600–613. <https://doi.org/10.1016/j.neuron.2009.08.021>.
- 454 9. Schaefer, A., Sampath, S.C., Intrator, A., Min, A., Gertler, T.S., Surmeier, D.J., Tarakhovsky, A.,
and Greengard, P. (2009). Control of cognition and adaptive behavior by the GLP/G9a epigenetic
455 suppressor complex. *Neuron* **64**, 678–691. <https://doi.org/10.1016/j.neuron.2009.11.019>.

- 456 10. Scandaglia, M., Lopez-Atalaya, J.P., Medrano-Fernandez, A., Lopez-Cascales, M.T., Del Blanco, B.,
Lipinski, M., Benito, E., Olivares, R., Iwase, S., Shi, Y., et al. (2017). Loss of Kdm5c Causes Spurious
Transcription and Prevents the Fine-Tuning of Activity-Regulated Enhancers in Neurons. *Cell Rep* **21**,
457 47–59. <https://doi.org/10.1016/j.celrep.2017.09.014>.
- 458 11. Iwase, S., Lan, F., Bayliss, P., De La Torre-Ubieta, L., Huarte, M., Qi, H.H., Whetstine, J.R., Bonni, A.,
Roberts, T.M., and Shi, Y. (2007). The X-Linked Mental Retardation Gene SMCX/JARID1C Defines a
Family of Histone H3 Lysine 4 Demethylases. *Cell* **128**, 1077–1088. <https://doi.org/10.1016/j.cell.2007.02.017>.
- 459 12. Claes, S., Devriendt, K., Van Goethem, G., Roelen, L., Meireleire, J., Raeymaekers, P., Cassiman,
J.J., and Fryns, J.P. (2000). Novel syndromic form of X-linked complicated spastic paraplegia. *Am J
460 Med Genet* **94**, 1–4.
- 461 13. Jensen, L.R., Amende, M., Gurok, U., Moser, B., Gimmel, V., Tschach, A., Janecke, A.R., Tariverdian,
G., Chelly, J., Fryns, J.P., et al. (2005). Mutations in the JARID1C gene, which is involved in
transcriptional regulation and chromatin remodeling, cause X-linked mental retardation. *Am J Hum
462 Genet* **76**, 227–236. <https://doi.org/10.1086/427563>.
- 463 14. Carmignac, V., Nambot, S., Lehalle, D., Callier, P., Moortgat, S., Benoit, V., Ghoumid, J., Delobel,
B., Smol, T., Thuillier, C., et al. (2020). Further delineation of the female phenotype with KDM5C
disease causing variants: 19 new individuals and review of the literature. *Clin Genet* **98**, 43–55.
464 <https://doi.org/10.1111/cge.13755>.
- 465 15. Iwase, S., Brookes, E., Agarwal, S., Badeaux, A.I., Ito, H., Vallianatos, C.N., Tomassy, G.S., Kasza, T.,
Lin, G., Thompson, A., et al. (2016). A Mouse Model of X-linked Intellectual Disability Associated with
466 Impaired Removal of Histone Methylation. *Cell Reports* **14**, 1000–1009. <https://doi.org/10.1016/j.celrep.2015.12.091>.
- 467 16. Endoh, M., Endo, T.A., Shinga, J., Hayashi, K., Farcas, A., Ma, K.W., Ito, S., Sharif, J., Endoh, T.,
Onaga, N., et al. (2017). PCGF6-PRC1 suppresses premature differentiation of mouse embryonic
468 stem cells by regulating germ cell-related genes. *eLife* **6**. <https://doi.org/10.7554/eLife.21064>.
- 469 17. Mochizuki, K., Sharif, J., Shirane, K., Uranishi, K., Bogutz, A.B., Janssen, S.M., Suzuki, A., Okuda,
A., Koseki, H., and Lorincz, M.C. (2021). Repression of germline genes by PRC1.6 and SETDB1
470 in the early embryo precedes DNA methylation-mediated silencing. *Nat Commun* **12**, 7020. <https://doi.org/10.1038/s41467-021-27345-x>.
- 471 18. Velasco, G., Hubé, F., Rollin, J., Neuillet, D., Philippe, C., Bouzinba-Segard, H., Galvani, A., Viegas-
Péquignot, E., and Francastel, C. (2010). Dnmt3b recruitment through E2F6 transcriptional repressor
472 mediates germ-line gene silencing in murine somatic tissues. *Proc Natl Acad Sci U S A* **107**, 9281–
9286. <https://doi.org/10.1073/pnas.1000473107>.
- 473

- 474 19. Hackett, J.A., Reddington, J.P., Nestor, C.E., Dunican, D.S., Branco, M.R., Reichmann, J., Reik, W., Surani, M.A., Adams, I.R., and Meehan, R.R. (2012). Promoter DNA methylation couples genome-defence mechanisms to epigenetic reprogramming in the mouse germline. *Development* *139*, 3623–3632. <https://doi.org/10.1242/dev.081661>.
- 475
- 476 20. Li, B., Qing, T., Zhu, J., Wen, Z., Yu, Y., Fukumura, R., Zheng, Y., Gondo, Y., and Shi, L. (2017). A Comprehensive Mouse Transcriptomic BodyMap across 17 Tissues by RNA-seq. *Sci Rep* *7*, 4200. <https://doi.org/10.1038/s41598-017-04520-z>.
- 477
- 478 21. Vallianatos, C.N., Raines, B., Porter, R.S., Bonefas, K.M., Wu, M.C., Garay, P.M., Collette, K.M., Seo, Y.A., Dou, Y., Keegan, C.E., et al. (2020). Mutually suppressive roles of KMT2A and KDM5C in behaviour, neuronal structure, and histone H3K4 methylation. *Commun Biol* *3*, 278. <https://doi.org/10.1038/s42003-020-1001-6>.
- 479
- 480 22. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* *15*, 550. <https://doi.org/10.1186/s13059-014-0550-8>.
- 481
- 482 23. Crackower, M.A., Kolas, N.K., Noguchi, J., Sarao, R., Kikuchi, K., Kaneko, H., Kobayashi, E., Kawai, Y., Kozieradzki, I., Landers, R., et al. (2003). Essential Role of Fkbp6 in Male Fertility and Homologous Chromosome Pairing in Meiosis. *Science* *300*, 1291–1295. <https://doi.org/10.1126/science.1083022>.
- 483
- 484 24. Xiol, J., Cora, E., Koglgruber, R., Chuma, S., Subramanian, S., Hosokawa, M., Reuter, M., Yang, Z., Berninger, P., Palencia, A., et al. (2012). A Role for Fkbp6 and the Chaperone Machinery in piRNA Amplification and Transposon Silencing. *Molecular Cell* *47*, 970–979. <https://doi.org/10.1016/j.molcel.2012.07.019>.
- 485
- 486 25. Cheng, S., Altmeppen, G., So, C., Welp, L.M., Penir, S., Ruhwedel, T., Menelaou, K., Harasimov, K., Stützer, A., Blayney, M., et al. (2022). Mammalian oocytes store mRNAs in a mitochondria-associated membraneless compartment. *Science* *378*, eabq4835. <https://doi.org/10.1126/science.abq4835>.
- 487
- 488 26. Rouland, A., Masson, D., Lagrost, L., Vergès, B., Gautier, T., and Bouillet, B. (2022). Role of apolipoprotein C1 in lipoprotein metabolism, atherosclerosis and diabetes: A systematic review. *Cardiovasc Diabetol* *21*, 272. <https://doi.org/10.1186/s12933-022-01703-5>.
- 489
- 490 27. Mueller, J.L., Skaletsky, H., Brown, L.G., Zaghlul, S., Rock, S., Graves, T., Auger, K., Warren, W.C., Wilson, R.K., and Page, D.C. (2013). Independent specialization of the human and mouse X chromosomes for the male germ line. *Nat Genet* *45*, 1083–1087. <https://doi.org/10.1038/ng.2705>.
- 491
- 492 28. Handel, M.A., and Eppig, J.J. (1979). Sertoli Cell Differentiation in the Testes of Mice Genetically Deficient in Germ Cells. *Biology of Reproduction* *20*, 1031–1038. <https://doi.org/10.1095/biolreprod20.5.1031>.
- 493

- 494 29. Green, C.D., Ma, Q., Manske, G.L., Shami, A.N., Zheng, X., Marini, S., Moritz, L., Sultan, C.,
Gurczynski, S.J., Moore, B.B., et al. (2018). A Comprehensive Roadmap of Murine Spermatogenesis
Defined by Single-Cell RNA-Seq. *Dev Cell* *46*, 651–667.e10. <https://doi.org/10.1016/j.devcel.2018.07.025>.
- 495 30. Soh, Y.Q., Junker, J.P., Gill, M.E., Mueller, J.L., van Oudenaarden, A., and Page, D.C. (2015). A
Gene Regulatory Program for Meiotic Prophase in the Fetal Ovary. *PLoS Genet* *11*, e1005531.
<https://doi.org/10.1371/journal.pgen.1005531>.
- 496 31. Magnúsdóttir, E., and Surani, M.A. (2014). How to make a primordial germ cell. *Development* *141*,
245–252. <https://doi.org/10.1242/dev.098269>.
- 497 32. Günesdogan, U., Magnúsdóttir, E., and Surani, M.A. (2014). Primordial germ cell specification: A
context-dependent cellular differentiation event [corrected]. *Philos Trans R Soc Lond B Biol Sci* *369*.
<https://doi.org/10.1098/rstb.2013.0543>.
- 501 33. Bardot, E.S., and Hadjantonakis, A.-K. (2020). Mouse gastrulation: Coordination of tissue patterning,
specification and diversification of cell fate. *Mechanisms of Development* *163*, 103617. <https://doi.org/10.1016/j.mod.2020.103617>.
- 503 34. Welling, M., Chen, H., Muñoz, J., Musheev, M.U., Kester, L., Junker, J.P., Mischerikow, N., Arbab, M.,
Kuijk, E., Silberstein, L., et al. (2015). DAZL regulates Tet1 translation in murine embryonic stem cells.
EMBO Reports *16*, 791–802. <https://doi.org/10.15252/embr.201540538>.
- 505 35. Macfarlan, T.S., Gifford, W.D., Driscoll, S., Lettieri, K., Rowe, H.M., Bonanomi, D., Firth, A., Singer, O.,
Trono, D., and Pfaff, S.L. (2012). Embryonic stem cell potency fluctuates with endogenous retrovirus
activity. *Nature* *487*, 57–63. <https://doi.org/10.1038/nature11244>.
- 507 36. Suzuki, A., Hirasaki, M., Hishida, T., Wu, J., Okamura, D., Ueda, A., Nishimoto, M., Nakachi, Y.,
Mizuno, Y., Okazaki, Y., et al. (2016). Loss of MAX results in meiotic entry in mouse embryonic and
germline stem cells. *Nat Commun* *7*, 11056. <https://doi.org/10.1038/ncomms11056>.
- 510 37. Samanta, M.K., Gayen, S., Harris, C., Maclary, E., Murata-Nakamura, Y., Malcore, R.M., Porter, R.S.,
Garay, P.M., Vallianatos, C.N., Samollow, P.B., et al. (2022). Activation of Xist by an evolutionarily
conserved function of KDM5C demethylase. *Nat Commun* *13*, 2602. <https://doi.org/10.1038/s41467-022-30352-1>.
- 512 38. Koubova, J., Menke, D.B., Zhou, Q., Capel, B., Griswold, M.D., and Page, D.C. (2006). Retinoic
acid regulates sex-specific timing of meiotic initiation in mice. *Proc. Natl. Acad. Sci. U.S.A.* *103*,
2474–2479. <https://doi.org/10.1073/pnas.0510813103>.
- 513 39. Lin, Y., Gill, M.E., Koubova, J., and Page, D.C. (2008). Germ Cell-Intrinsic and -Extrinsic Factors
Govern Meiotic Initiation in Mouse Embryos. *Science* *322*, 1685–1687. <https://doi.org/10.1126/science.1166340>.
- 515

- 516 40. Endo, T., Mikedis, M.M., Nicholls, P.K., Page, D.C., and De Rooij, D.G. (2019). Retinoic Acid and Germ
517 Cell Development in the Ovary and Testis. *Biomolecules* *9*, 775. <https://doi.org/10.3390/biom9120775>.
- 518 41. Gupta, N., Yakhou, L., Albert, J.R., Azogui, A., Ferry, L., Kirsh, O., Miura, F., Battault, S., Yamaguchi,
519 K., Laisné, M., et al. (2023). A genome-wide screen reveals new regulators of the 2-cell-like cell state.
Nat Struct Mol Biol. <https://doi.org/10.1038/s41594-023-01038-z>.
- 520 42. Liu, M., Zhu, Y., Xing, F., Liu, S., Xia, Y., Jiang, Q., and Qin, J. (2020). The polycomb group protein
521 PCGF6 mediates germline gene silencing by recruiting histone-modifying proteins to target gene
promoters. *J Biol Chem* *295*, 9712–9724. <https://doi.org/10.1074/jbc.RA119.012121>.
- 522 43. Otani, J., Nankumo, T., Arita, K., Inamoto, S., Ariyoshi, M., and Shirakawa, M. (2009). Structural basis
523 for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX–DNMT3–DNMT3L
domain. *EMBO Reports* *10*, 1235–1241. <https://doi.org/10.1038/embor.2009.218>.
- 524 44. Guo, X., Wang, L., Li, J., Ding, Z., Xiao, J., Yin, X., He, S., Shi, P., Dong, L., Li, G., et al. (2015).
525 Structural insight into autoinhibition and histone H3-induced activation of DNMT3A. *Nature* *517*,
640–644. <https://doi.org/10.1038/nature13899>.
- 526 45. Abildayeva, K., Berbée, J.F.P., Blokland, A., Jansen, P.J., Hoek, F.J., Meijer, O., Lütjohann, D., Gautier,
527 T., Pillot, T., De Vente, J., et al. (2008). Human apolipoprotein C-I expression in mice impairs learning
and memory functions. *Journal of Lipid Research* *49*, 856–869. <https://doi.org/10.1194/jlr.M700518JLR200>.
- 528 46. Leduc, V., Jasmin-Bélanger, S., and Poirier, J. (2010). APOE and cholesterol homeostasis in
529 Alzheimer's disease. *Trends in Molecular Medicine* *16*, 469–477. <https://doi.org/10.1016/j.molmed.2010.07.008>.
- 530 47. Hayashi, K., Ohta, H., Kurimoto, K., Aramaki, S., and Saitou, M. (2011). Reconstitution of the
531 mouse germ cell specification pathway in culture by pluripotent stem cells. *Cell* *146*, 519–532.
<https://doi.org/10.1016/j.cell.2011.06.052>.
- 532 48. Zhang, J., Zhang, M., Acampora, D., Vojtek, M., Yuan, D., Simeone, A., and Chambers, I. (2018).
533 OTX2 restricts entry to the mouse germline. *Nature* *562*, 595–599. <https://doi.org/10.1038/s41586-018-0581-5>.
- 534 49. Endo, T., Romer, K.A., Anderson, E.L., Baltus, A.E., De Rooij, D.G., and Page, D.C. (2015). Pe-
535 riodic retinoic acid–STRA8 signaling intersects with periodic germ-cell competencies to regulate
spermatogenesis. *Proc. Natl. Acad. Sci. U.S.A.* *112*. <https://doi.org/10.1073/pnas.1505683112>.
- 536 50. Bowles, J., Knight, D., Smith, C., Wilhelm, D., Richman, J., Mamiya, S., Yashiro, K., Chawengsak-
537 sophak, K., Wilson, M.J., Rossant, J., et al. (2006). Retinoid Signaling Determines Germ Cell Fate in
Mice. *Science* *312*, 596–600. <https://doi.org/10.1126/science.1125691>.

- 538 51. Yang, M., Yu, H., Yu, X., Liang, S., Hu, Y., Luo, Y., Izsvák, Z., Sun, C., and Wang, J. (2022). Chemical-induced chromatin remodeling reprograms mouse ESCs to totipotent-like stem cells. *Cell Stem Cell* 29, 400–418.e13. <https://doi.org/10.1016/j.stem.2022.01.010>.
- 539
- 540 52. Al Adhami, H., Vallet, J., Schaal, C., Schumacher, P., Bardet, A.F., Dumas, M., Chicher, J., Hammann, P., Daujat, S., and Weber, M. (2023). Systematic identification of factors involved in the silencing of germline genes in mouse embryonic stem cells. *Nucleic Acids Research* 51, 3130–3149. <https://doi.org/10.1093/nar/gkad071>.
- 541
- 542 53. Karimi, M.M., Goyal, P., Maksakova, I.A., Bilenky, M., Leung, D., Tang, J.X., Shinkai, Y., Mager, D.L., Jones, S., Hirst, M., et al. (2011). DNA Methylation and SETDB1/H3K9me3 Regulate Predominantly Distinct Sets of Genes, Retroelements, and Chimeric Transcripts in mESCs. *Cell Stem Cell* 8, 676–687. <https://doi.org/10.1016/j.stem.2011.04.004>.
- 543
- 544 54. Tatsumi, D., Hayashi, Y., Endo, M., Kobayashi, H., Yoshioka, T., Kiso, K., Kanno, S., Nakai, Y., Maeda, I., Mochizuki, K., et al. (2018). DNMTs and SETDB1 function as co-repressors in MAX-mediated repression of germ cell-related genes in mouse embryonic stem cells. *PLoS ONE* 13, e0205969. <https://doi.org/10.1371/journal.pone.0205969>.
- 545
- 546 55. Stephens, M. (2016). False discovery rates: A new deal. *Biostat*, kxw041. <https://doi.org/10.1093/biostatistics/kxw041>.
- 547
- 548 56. Conway, J.R., Lex, A., and Gehlenborg, N. (2017). UpSetR: An R package for the visualization of intersecting sets and their properties. *Bioinformatics* 33, 2938–2940. <https://doi.org/10.1093/bioinformatics/btx364>.
- 549
- 550 57. 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. *Molecular Cell* 38, 576–589. <https://doi.org/10.1016/j.molcel.2010.05.004>.
- 551
- 552 58. Li, H., Liang, Z., Yang, J., Wang, D., Wang, H., Zhu, M., Geng, B., and Xu, E.Y. (2019). DAZL is a master translational regulator of murine spermatogenesis. *Natl Sci Rev* 6, 455–468. <https://doi.org/10.1093/nsr/nwy163>.
- 553
- 554 59. Mikedis, M.M., Fan, Y., Nicholls, P.K., Endo, T., Jackson, E.K., Cobb, S.A., De Rooij, D.G., and Page, D.C. (2020). DAZL mediates a broad translational program regulating expansion and differentiation of spermatogonial progenitors. *eLife* 9, e56523. <https://doi.org/10.7554/eLife.56523>.
- 555

556 **Figures and Tables**

- 557 • Supplementary table 1: list of all germline genes.
- 558 – Columns to include:
- 559 * KDM5C bound vs not
- 560 * DEG in EpiLC, brain, both, neither (separate columns?)

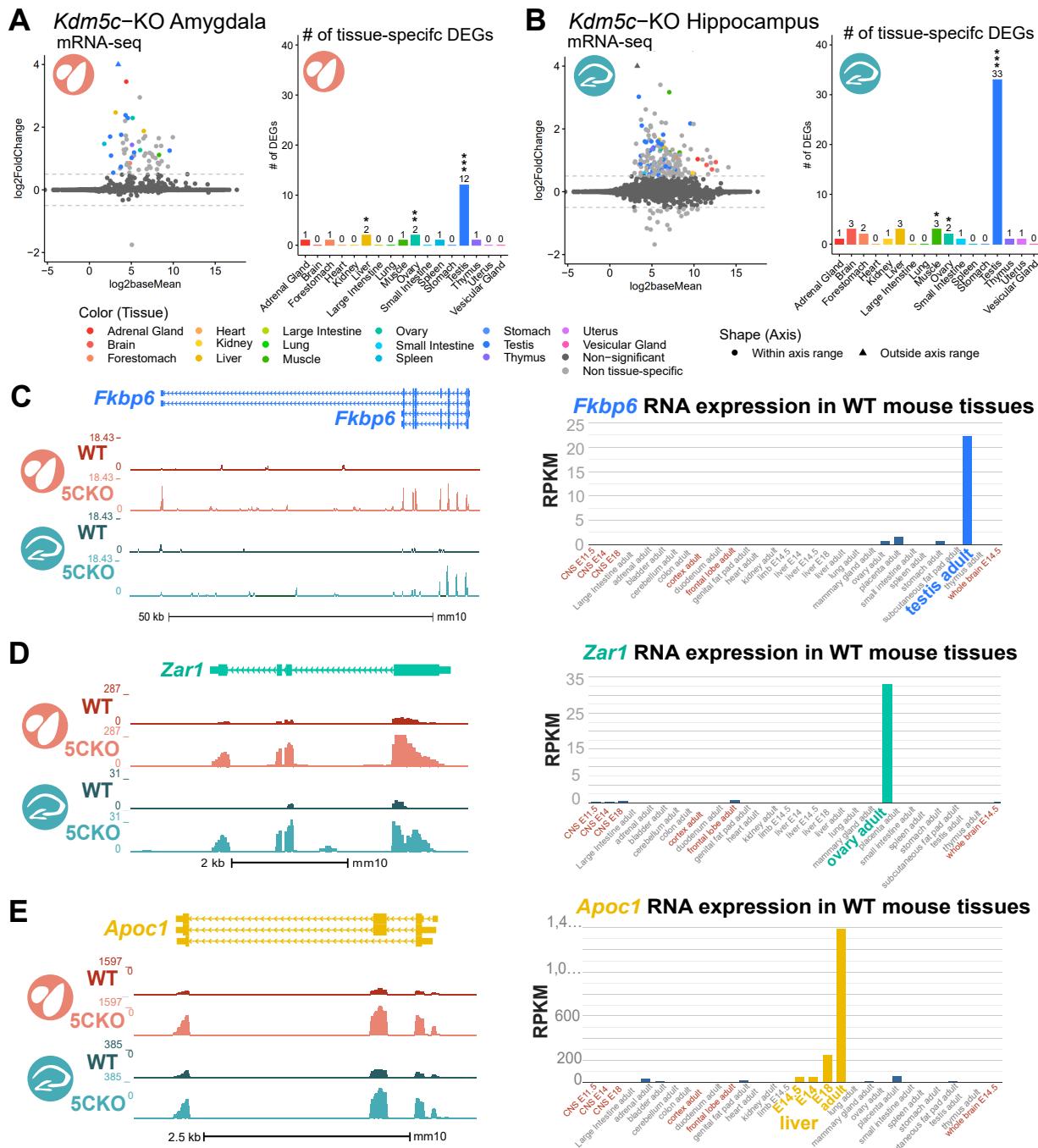


Figure 1: Tissue-enriched genes are misexpressed in the *Kdm5c*-KO brain. **A-B.** Expression of tissue-enriched genes (Li et al 2017) in the male *Kdm5c*-KO amygdala (A) and hippocampus (B). Left - MA plot of mRNA-sequencing. Right - Number of tissue-enriched differentially expressed genes (DEGs). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, Fisher's exact test. **C.** Left - Average bigwigs of an example aberrantly expressed testis-enriched DEG, *FK506 binding protein 6* (*Fkbp6*) in the wild-type (WT) and *Kdm5c*-KO (5CKO) amygdala (red) and hippocampus (teal). Right - Expression of *Cyct* in wild-type tissues from NCBI Gene, with testis highlighted in blue and brain tissues highlighted in red. **D.** Left - Average bigwigs of an example ovary-enriched DEG, *Zygotic arrest 1* (*Zar1*). Right - Expression of *Zar1* in wild-type tissues from NCBI Gene, with ovary highlighted in teal and brain tissues highlighted in red. **E.** Left - Average bigwigs of an example liver-enriched DEG, *Apolipoprotein C-I* (*Apoc1*). Right - Expression of *Apoc1* in wild-type tissues from NCBI Gene, with liver highlighted in orange and brain tissues highlighted in red.

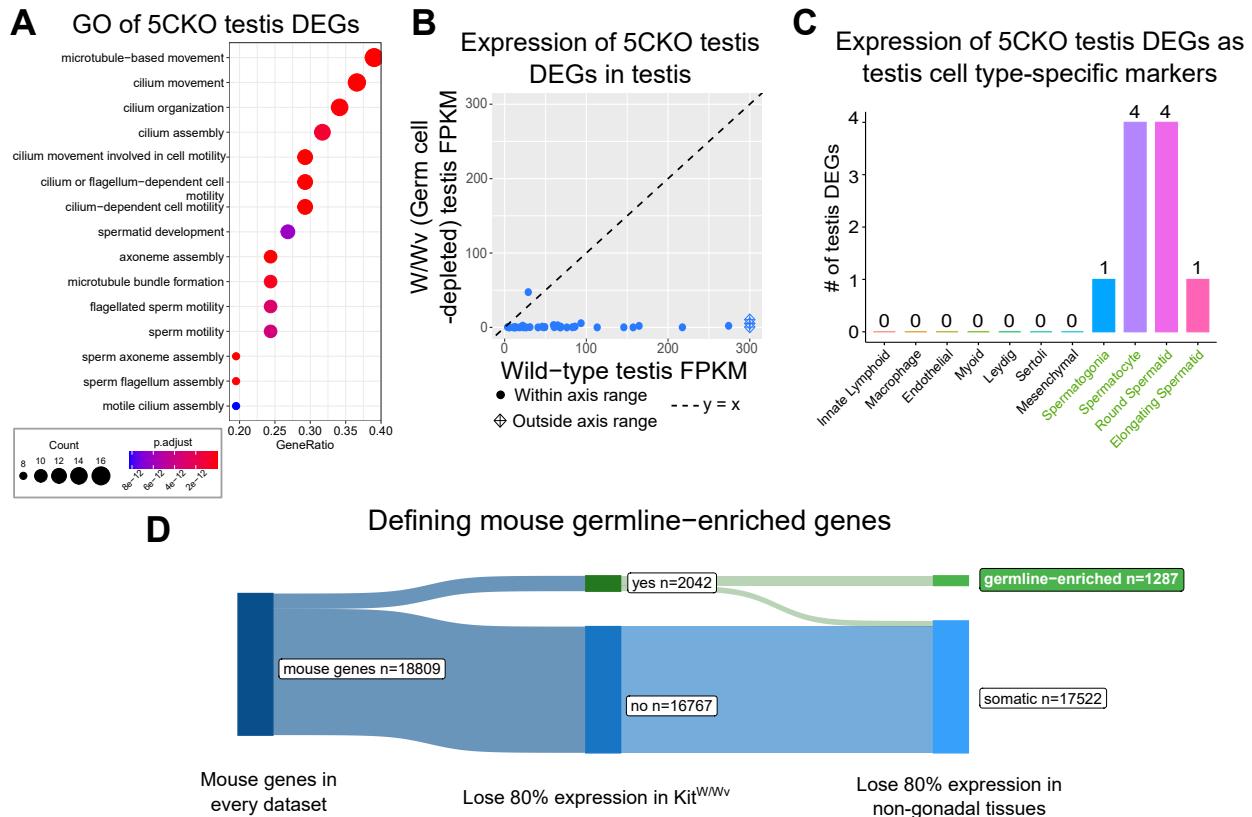


Figure 2: Aberrant transcription of germline genes in the *Kdm5c*-KO in the brain. **A.** enrichPlot gene ontology (GO) of *Kdm5c*-KO amygdala and hippocampus testis-enriched DEGs **B.** Expression of testis DEGs in wild-type (WT) testis versus germ cell-depleted (W/Wv) testis (Mueller et al 2013). Expression is in Fragments Per Kilobase of transcript per Million mapped reads (FPKM). **C.** Number of testis DEGs that were classified as cell-type specific markers in a single cell RNA-seq dataset of the testis (Green et al 2018). Germline cell types are highlighted in green, somatic cell types in black. **D.** Sankey diagram of mouse genes filtered for germline enrichment based on their expression in wild-type and germline-depleted mice and in adult mouse non-gonadal tissues (Li et al 2017).

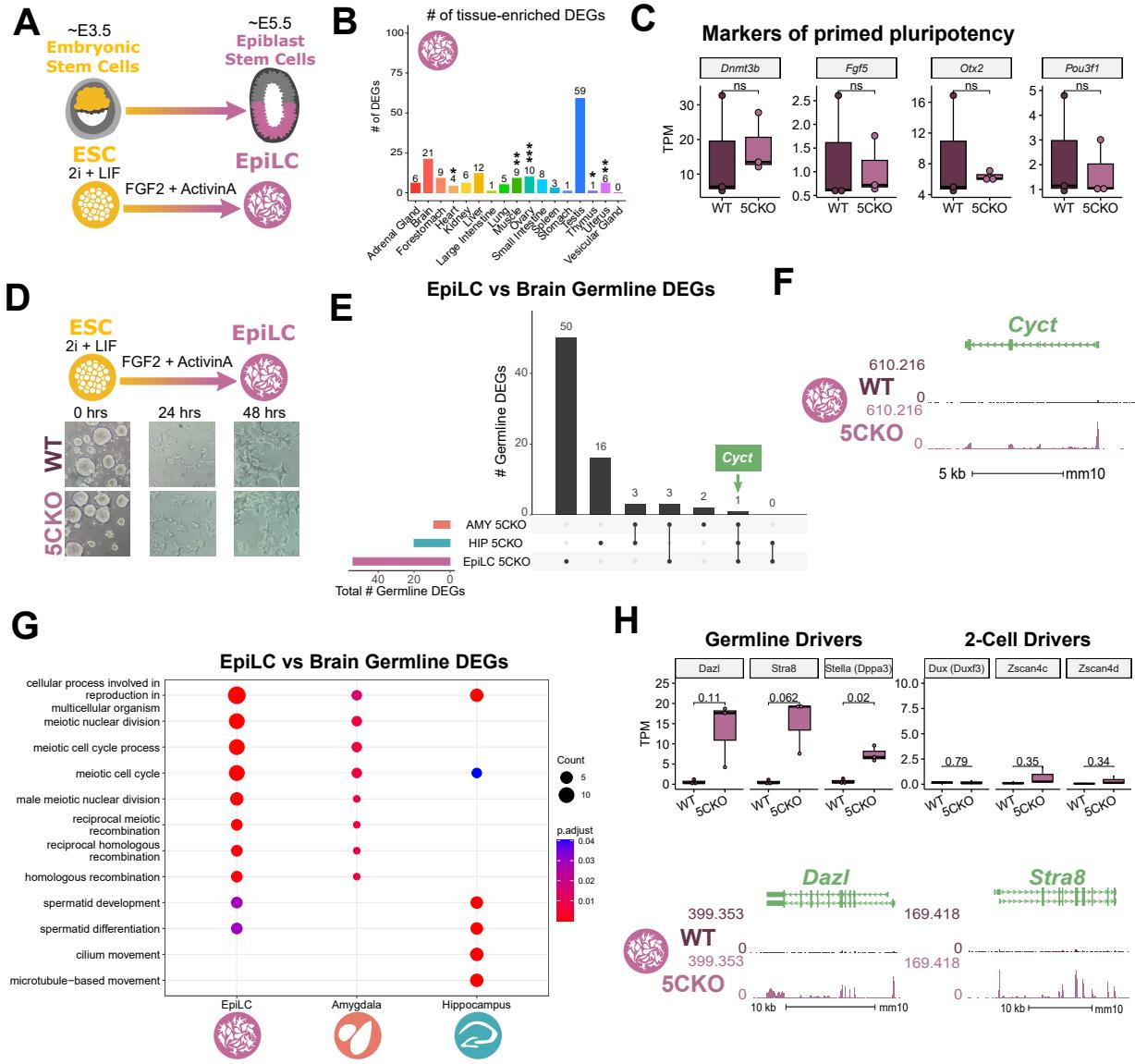


Figure 3: *Kdm5c*-KO epiblast-like cells express key drivers of germline identity

A. Top - Diagram of *in vivo* differentiation of embryonic stem cells (ESCs) of the inner cell mass into epiblast stem cells. Bottom - *in vitro* differentiation of ESCs into epiblast-like cells (EpiLCs).

B. Number of tissue-enriched differentially expressed genes (DEGs) in *Kdm5c*-KO EpiLCs. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, Fisher's exact test.

C. Average bigwigs of an example germline gene, *Cyct*, that is dysregulated *Kdm5c*-KO EpiLCs.

D. No significant difference in primed pluripotency marker expression in wild-type versus *Kdm5c*-KO EpiLCs. Welch's t-test, expression in transcripts per million (TPM).

E. Representative images of wild-type (WT) and *Kdm5c*-KO cells during ESC to EpiLC differentiation. Brightfield images taken at 20X.

F. Upset plot displaying the overlap of germline DEGs expressed in *Kdm5c*-KO EpiLCs, amygdala (AMY), and hippocampus (HIP) RNA-seq datasets.

G. enrichPlot comparing enriched biological process gene ontologies for *Kdm5c*-KO EpiLC, amygdala, and hippocampus germline DEGs.

H. Top left - Example germline identity DEGs unique to EpiLCs. Top right - Example 2-cell genes that are not dysregulated in *Kdm5c*-KO EpiLCs. p-values for Welch's t-test. Bottom - Average bigwigs of *Dazl* and *Stra8* expression in wild-type and *Kdm5c*-KO EpiLCs. Immunocytochemistry of DAZL in male wild-type (WT) and *Kdm5c*-KO (5CKO) EpiLCs. Percentage of DAZL-positive cells normalized to number of DAPI-positive cells, p-value for Welch's t-test.

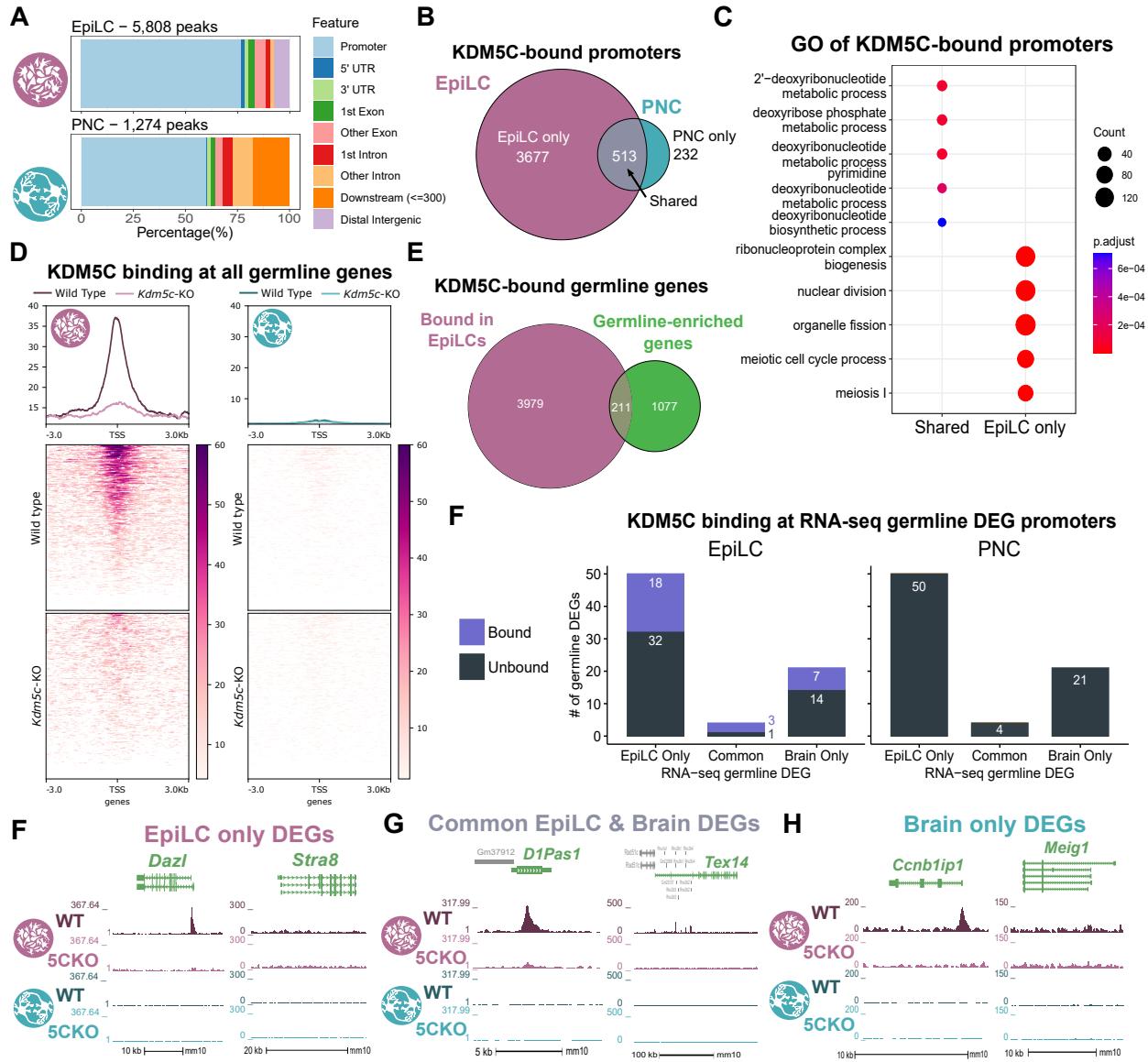
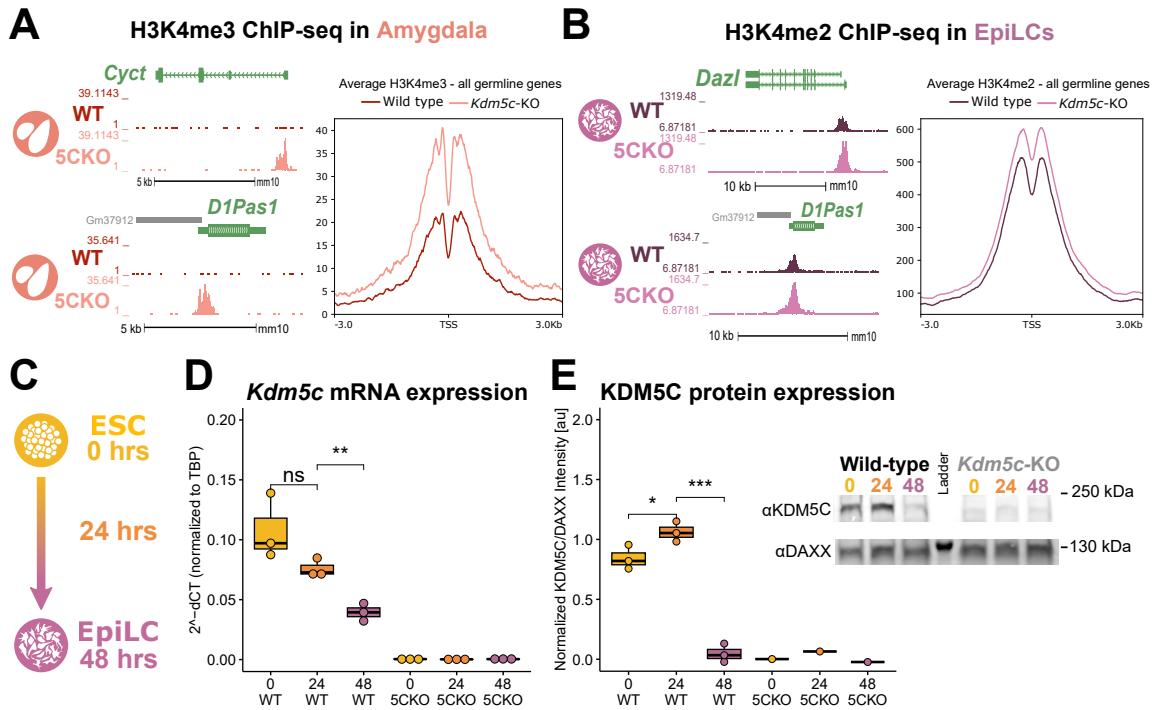


Figure 4: KDM5C binds to a subset of germline gene promoters during early embryogenesis **A.** ChIPseeker localization of KDM5C peaks at different genomic regions in EpiLCs (top) and hippocampal and cortex primary neuron cultures (PNCs, bottom). **B.** Overlap of genes with KDM5C bound to their promoters in EpiLCs (purple) and PNCs (blue). **C.** Gene ontology (GO) comparison of genes with KDM5C bound to their promoter in EpiLCs and PNCs. Genes were classified as either bound in EpiLCs only (EpiLC only), unique to PNCs (PNC only, no significant ontologies) or bound in both PNCs and EpiLCs (shared). **D.** Average KDM5C binding around the transcription start site (TSS) of all germline-enriched genes in EpiLCs (left) and PNCs (right). **E.** KDM5C binding at the promoters of RNA-seq germline DEGs. Genes were classified as either only dysregulated in EpiLCs (EpiLC only), genes dysregulated in the hippocampus or amygdala but not EpiLCs (brain only), or genes dysregulated in both EpiLCs and the brain (common). (Legend continued on next page.)

Figure 4: KDM5C binds to a subset of germline gene promoters during early embryogenesis. (Legend continued.) **F.** Example KDM5C ChIP-seq bigwigs of DEGs unique to EpiLCs. Although both are expressed in *Kdm5c*-KO EpiLCs, KDM5C is only bound to the *Dazl* promoter and not the *Stra8* promoter in EpiLCs. **G.** Example KDM5C ChIP-seq bigwigs of RNA-seq DEGs common between the brain and EpiLCs. KDM5C is bound to the *D1Pas1* promoter but not the *XXX* promoter in EpiLCs. **H.** Example KDM5C ChIP-seq bigwigs of RNA-seq DEGs unique to the brain. KDM5C is bound to the *XXX* promoter but not the *Meig1* promoter



F

G

Figure 5: KDM5C's catalytic activity promotes long-term silencing of germline genes via DNA methylation. **A.** Left - Bigwigs of representative histone 3 lysine 4 trimethylation (H3K4me3) ChIP-seq peaks at two germline genes in the wild-type (WT) and *Kdm5c*-KO (5CKO) adult amygdala. Right - Average H3K4me3 at the transcripton start site (TSS) of all germline-enriched genes in wild-type (dark red) and *Kdm5c*-KO (light red) amygdala. **B.** Left - Bigwigs of representative histone 3 lysine 4 dimethylation (H3K4me2) ChIP-seq peaks at representative germline genes in wild-type and *Kdm5c*-KO EpiLCs. Right - Average H3K4me2 at the TSS of all germline-enriched genes in wild-type (dark purple) and *Kdm5c*-KO (light purple) EpiLCs. **C.** Diagram of embryonic stem cell (ESC) to epiblast-like cell (EpiLC) differentiation protocol and collection time points for RNA and protein. **D.** Real time quantitative PCR (RT-qPCR) of *Kdm5c* RNA expression, calculated in comparision to TBP expression ($2^{-\Delta\Delta CT}$). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, Welch's t-test. **E.** KDM5C protein expression normalized to DAXX. Quantified intensity using ImageJ (artificial units - au). Right - representative lanes of Western blot for KDM5C and DAXX. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, Welch's t-test **F.** XXX **G.** XXX

561 **Notes**

562 **Dazl**

563 We were particularly interested in the aberrant transcription of *Dazl*, since it is essential for germ cell
564 development and promotes the translation of germline mRNAs⁵⁸. A significant portion of germline transcripts
565 misexpressed in *Kdm5c*-KO EpiLCs are known binding targets of DAZL, including *Stra8*⁵⁹ ($p = 1.698e-07$,
566 Fisher's Exact Test). This suggests expression of DAZL protein could promote the translation of other
567 aberrant germline transcripts, influencing their ability to impact *Kdm5c*-KO cellular function. We thus tested
568 DAZL protein expression in *Kdm5c*-KO EpiLCs through immunocytochemistry (Figure 3I). We observed
569 about 25% of *Kdm5c*-KO EpiLCs expressed DAZL protein and it was localized to the cytoplasm ($p = 0.0015$,
570 Welch's t-test), consistent with the pattern of DAZL expression in spermatogonia⁵⁹. Altogether these results
571 suggest tissue-specific genes are misexpressed during *Kdm5c*-KO embryogenesis, including key drivers of
572 germline identity that can be translated into protein.

573 **Discussion notes**

- 574 • We additionally found *Kdm5c*-KO EpiLCs ectopically express DAZL protein that is localized to the
575 cytoplasm, similar to its morphology in spermatogonia⁵⁹. **note: maybe just put in results**. Could
576 move around depending upon if I get pheno working.
- 577 • end with something like: this indicates kdm5c not only modulates pluripotency and self-renewal in
578 ESCs, but also has a role in long-term silencing of germline genes
 - 579 – then transition into the long term silencing mechanism paragraph
- 580 • KDM5C is dynamically regulated during the window of embryonic germline gene silencing, our EpiLC
581 ChIP-seq is likely catching the tail end of KDM5C's main involvement.
- 582 • Our results indicate KDM5C is a pivotal repressor of germline programs in somatic cells.
- 583 • Studies on other chromatin regulators have focused on key marker genes to identify germline gene
584 misexpression, such as *Dazl*.
- 585 • Using a germline-depleted mouse model, we curated a list of mouse germline-enriched genes to
586 globally assess germline gene dysregulation.
- 587 • Many of the genes dysregulated in the mature *Kdm5c*-KO brain were important for late stages of
588 spermatogenesis, such as those important for sperm flagellar structure. Contrastingly, *Kdm5c*-KO
589 EpiLCs aberrantly expressed key drivers of germ cell identity, including *Dazl* and *Stra8*.

- 590 • While a significant portion of KDM5C-bound promoters are germline genes, many germline genes
591 expressed during *Kdm5c*-KO embryogenesis are not directly bound by *kdm5c*.
592 • One notable example is *Stra8*, a transcription factor important for germ cell specific transcription and
593 meiotic initiation
594 • The including the demarcation between soma and germline fates.
595 the *Kdm5c*-KO brain also Previous work identified testis-enriched genes in *Kdm5c* genes
596 –
597 – However unlike the gonadal-biased DEGs,
598 • The demarcation of the germline versus soma is a key feature of multicellularity and sexually dimorphic
599 reproduction
600 • Anything known about tissue-biased gene expression in other H3K4me regulators?
601 • Our data suggests the germline developmental program is occurring ectopically as *Kdm5c*-KOs pro-
602 gresses through somatic tissue development
603 • tissue-biased gene expression:
604 • Altogether, these data indicate that while some germline genes are misexpressed due to direct loss of
605 KDM5C binding during emryogenesis, secondary downstream mechanisms can also promote their
606 aberrant transcription.
607 • Papers to read/reference:
608 – Reconstitution of the Mouse Germ Cell Specification Pathway in Culture by Pluripotent Stem Cells:
609 [https://www.cell.com/fulltext/S0092-8674\(11\)00771-9](https://www.cell.com/fulltext/S0092-8674(11)00771-9)
610 – two cell gene list used by Suzuki et al Max paper is based on 2 cell sequencing: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3395470/>
611

612 **Figure outline:**

- 613 **Figure 1: Misexpression of tissue-specific genes in the *Kdm5c*-KO brain** * MA-plot and bar graphs of
614 tissue-enriched genes * Example testis-specific genes (NCBI and bigwigs) * An example ovary tissue-specific
615 gene * An example muscle/liver tissue-specific gene (NCBI and bigwigs)
616 **Figure 2: The male *Kdm5c*-KO brain expresses male and female germline-enriched genes** * Gene
617 ontology of testis DEGs in the amygdala and hippocampus - ontologies are germline ontologies * Expression
618 of testis DEGs in germline-depleted testis (this is adult testis data) * scRNASeq of testis - # of testis DEGs that
619 are germline-specific markers * Although far fewer, 5CKO brain also expresses ovary-enriched genes(NCBI

620 and bigwigs of Zar1) * These ovary enriched genes are also germline specific (NCBI/Li tissues are in adult
621 ovary, so it would be best to show they're oocyte-specific in adult ovary. But I don't think there's a published
622 adult female W/Wv dataset. Could try looking at scRNASeq or just do TPM in embryonic W/Wv data since
623 oocytes are developed at this point? Or both?) * Defining what is/isn't a germline gene, and which are
624 male/female biased using embryonic W/Wv data

625 **Figure 3: Kdm5c-KO epiblast-like cells express key drivers regulators of germline identity** * A) ESC
626 to EpiLC differentiation Left - Morphology is unchanged, * B) 5CKO EpiLCs express EpiLC differentiation
627 genes similar to WT lvs * C) Male EpiLCs express germline genes (example Cyct again) * Overlap between
628 brain and EpiLC germline genes - show they're mostly unique * GO of Brain and EpiLC germline genes
629 (meiotic enriched) * Bigwigs or TPM of master regulators * Show that while some are also 2-cell genes
630 (Dazl), 2-cell specific genes aren't dysregulated (Zscan4). Important point because published KDM5C dazl
631 paper is saying KDM5C is a 2-cell regulator, but as far as I can tell only genes shared between germline and
632 2-cell are dysregulated.

633 Staining of Dazl (+ Stra8 if I can get it to work)

634 **Figure 4: Loss of KDM5C's catalytic activity impairs DNAme placement and long-term silencing of
635 germline genes** * Increase in H3K4me3 in Kdm5c-KO amygdala at germline genes * Increase in H3K4me2
636 in EpiLCs at germline genes * Kdm5c binding in EpiLCs vs PNCs to show that germline repression is
637 happening in early embryo * Previous studies only looked at ESCs, unknown if catalytic activity is required
638 for long-term repression, especially since DNA methylation is placed later). KDM5C RNA and protein ESC →
639 EpiLC (increasing then decreasing) * RNA expression of germline genes with catalytic dead rescue (Ilakkia)
640 * DNA methylation in WT and 5CKO EpiLCs (Ilakkia)

641 **Figure 5: Ectopic, germline-like phenotypes in Kdm5c-KO ESCs/EpiLCs** * Sycp3 staining * DDX4
642 staining and repression of retrotransposons * Cilia??

643 Gaps in knowledge addressed: * Are other tissue-enriched genes dysregulated, or only testis, germline
644 genes? * Curating a robust list of male and female germline genes * Should talk about 2-cell genes
645 vs germline genes - way to systematically categorize? * Mechanism behind long-term germline gene
646 misexpression * Recent evidence suggests loss of KDM5C in ESCs express some germline genes * Unclear
647 if catalytic activity is required for long-term silencing * Unclear if their dysregulation lasts throughout life or
648 the same between brain or not * When in development does it begin? - Recent evidence suggests some
649 germline genes expressed in 5CKO ESCs but unclear if their dysregulation lasts throughout differentiation
650 and if the identity of germline genes are different compared to the brain * Are there functional consequences
651 to germline gene misexpression?

652 Introduction: * Chromatin regulators are important for cellular identity * H3K4me1-3 linked to active
653 gene promoters and enhancers * Surprisingly, mutations in Chromatin regulators lead to many NDDs
654 (including many H3K4 regulators) * Recent studies have shown some chromatin regulators are important
655 for regulating neuron-specific gene expression/chromatin stat_compare_means * However, loss of some

656 chromatin regulators can also lead to ectopic expression of tissue-enriched genes * Very few studies have
657 looked at these genes and it's unclear if these genes contribute to NDD impairments. * Necessary to
658 first characterize the mechanism behind their derepression to identify molecular footholds into testing their
659 contribution to neuronal impairments and potential for therapeutic intervention

- 660 • Loss of KDM5C can result in the misexpression of genes typically only found in the testis
- 661 – Misexpression of tissue-enriched genes hasn't been systematically characterized - Unclear if
662 these genes are exceptions or if other tissue-specific genes are dysregulated
- 663 – Interestingly, these genes (Cyct, D1pas1) typically function in the germline
- 664 – Germ cells (meiotic cells) are typically distinguished from somatic cells very early on in embryoge-
665 nesis and is a key feature of multicellularity
- 666 – Chromatin regulators are very important for decommissioning germline genes and act successively
667 the embryo implants into the uterine wall
- 668 * Most studies have focused on ESCs, which have a similar transcriptome to germ cells / 2-cells
- 669 * recently, KDM5C was shown to repress DAZL in ESCs, independent of its catalytic activity
- 670 * However, DNA methylation is lost in the mature 5CKO brain, DNA methylation is placed later
671 and it's Unclear if it's required for long-term repression (maybe too specific, just trying to go
672 into the fact that the mechanism is partially understood but unclear)
- 673 – Systematic characterization of ectopic germline genes hasn't been done
- 674 * unknown if other germline-enriched genes are dysregulated, including oocyte-specific genes
- 675 * Crucially, it's unknown if misexpression of the germline program leads to functional conse-
676 quences in 5CKO cells.

677 **Germline gene repression background:**

678 Interestingly, some of the ectopic testis transcripts identified in the *Kdm5c*-KO brain are typically ex-
679 pressed in germ cells¹⁰. Unlike somatic cells, germ cells (e.g. sperm and eggs) undergo meiosis and pass
680 on their genetic material to the next generation. The germline and the soma are typically distinguished during
681 early embryogenesis, when germline genes are silenced in epiblast stem cells soon after implantation and
682 only reactivated in a subset to form the germline. Chromatin regulators play a key role in decommissioning
683 germline genes as the embryo transitions from naïve to primed pluripotency by placing repressive histone
684 H2A lysine 119 monoubiquitination (H2AK119ub1)¹⁶, histone 3 lysine 9 trimethylation (H3K9me3)^{16,17}, and
685 DNA CpG methylation^{17–19} at germline gene promoters. KDM5C may also be involved in this early decom-
686 missioning of germline genes, as re-expression of KDM5C in neurons fails to suppress their dysregulation¹⁰.
687 In support of this, KDM5C was very recently shown to repress *Deleted in azoospermia like (Dazl)*, a key
688 regulator of germline development, in mouse embryonic stem cells (ESCs)^{41,52}. In support of this, two
689 independent screens in mouse embryonic stem cells (ESCs) recently identified KDM5C as a repressor of
690 *Deleted in azoospermia like (Dazl)*, a key regulator of germline development. However, KDM5C's role in

691 embryonic germline gene repression is currently unclear, given that Dazl is also expressed in ESCs and in
692 the 2-cell stage and germline gene misexpression has yet to be globally characterized during *Kdm5c*-KO
693 embryogenesis.