

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. Intriguingly, many *Kdm5c*-
124 KO testis and ovary enriched-DEGs have germline-specific functions, suggesting *Kdm5c*-KO cells fail to
125 distinguish between the soma and germline. To test if this holds true for all *Kdm5c*-KO testis-biased DEGs, we
126 first assessed their known functions through gene ontology analysis. We found *Kdm5c*-KO testis-enriched
127 DEGs high enrichment of germline-relevant ontologies, including spermatid development (GO: 0007286,
128 p.adjust = 6.2e-12) and sperm axoneme assembly (GO: 0007288, p.adjust = 2.45e-14) (Figure 2A).

129 To further validate if these testis DEGs are truly germline genes, we then evaluated their expression in

130 somatic versus germ cells within the testis. We first compared their expression in the testis with germ cell
131 depletion²⁷, which was accomplished by heterozygous *W* and *Wv* mutations in the enzymatic domain of
132 *c-Kit* (*Kit*^{WWv}) that prevent the maturation of germ cells²⁸. Almost all *Kdm5c*-KO testis-enriched DEGs lost
133 expression with germ cell depletion (Figure 2B). The only testis-enriched DEG that did not show considerable
134 downregulation with germline depletion was *FK506 binding protein 6* (*Fkbp6*), the aforementioned testis
135 gene that regulates piRNA expression and meiosis in germ cells^{23,24}. We then assessed testis-enriched
136 DEG expression in a published single cell RNA-seq dataset that identified cell type-specific markers within
137 the testis²⁹. We found that while some testis-enriched DEGs were classified as specific markers for different
138 germ cell types (e.g. spermatogonia, spermatocytes, round spermatids, and elongating spermatids) none
139 marked somatic cells (Figure 2C). Together, these data suggest the somatic *Kdm5c*-KO brain aberrantly
140 expresses germline genes.

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

152 ***Kdm5c*-KO epiblast-like cells aberrantly express master regulators of germline
153 identity**

154 Misexpression of germline genes in the adult brain suggests *Kdm5c*-KO cells fail to demarcate between
155 germline and somatic identity. Germ cells are typically distinguished from somatic cells soon after the embryo
156 implants into the uterine wall^{31,32} when a subset of epiblast stem cells become the primordial germ cells
157 (PGCs) while the remainder differentiate into the ectoderm, mesoderm, and endoderm to form the somatic
158 tissues³³. This developmental time point can be modeled *in vitro* through differentiation of embryonic stem
159 cells (ESCs) into post-implantation epiblast-like stem cells (EpiLCs) (Figure 3A, top). Previous studies have
160 demonstrated that while some germline-enriched genes are also expressed in embryonic stem cells (ESCs)
161 and in the 2-cell stage^{34–36}, they are silenced as they differentiate into EpiLCs¹⁷. Therefore, we assessed if
162 KDM5C was necessary for embryonic silencing of germline genes by evaluating the impact of *Kdm5c* loss in
163 EpiLCs.

164 We first identified *Kdm5c*-KO EpiLC DEGs in of our previously published RNA-seq dataset³⁷ (DESeq2,
165 log2 fold change > 0.5, q < 0.1) and assessed tissue-enriched gene expression. Similar to the *Kdm5c*-KO
166 brain, we observed general dysregulation of tissue-enriched genes, with the largest number of genes
167 belonging to the brain and testis, although they were not significantly enriched (Figure 3B). Using the curated
168 list of germline genes generated above, we found *Kdm5c*-KO EpiLCs aberrantly expressed 54 germline-
169 enriched genes, including the previously characterized hippocampal DEG¹⁰ *Cytochrome C, testis-specific*
170 (*Cyct*) (Figure 3C). Although we observed aberrant expression of many tissue-enriched genes, we did not
171 observe any significant difference in primed pluripotency genes (Figure 3D) or gross changes in *Kdm5c*-KO
172 cell morpholgy during differentiation (Figure 3E), indicating KDM5C loss does not impair EpiLC formation.

173 We then compared EpiLC germline DEGs to those expressed in the *Kdm5c*-KO brain to determine
174 if all germline DEGs, like *Cyct*, are constitutively dysregulated or if they can change over the course of
175 development. We found this was primarily not the case, as the majority of germline DEGs expressed
176 in EpiLCs and the brain were unique, with only *Cyct* shared across all sequencing datasets (Figure 3F).
177 We then compared the known functions of EpiLC and brain germline DEGs and found particularly high
178 enrichment of meiosis-related gene ontologies in EpiLCs (Figure 3G), such as meiotic cell cycle process (GO:
179 1903046, p.adjust = 1.59e-08) and meiotic nuclear division (GO: 0140013, p.adjust = 9.76e-09). While there
180 was modest enrichment of meiotic gene ontologies in both brain regions, the *Kdm5c*-KO hippocampus
181 primarily expressed late-stage spermatogenesis genes, such those involved in the sperm axoneme structure.

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

189 We were particularly interested in the aberrant transcription of *Dazl*, since it is essential for germ cell
190 development and promotes the translation of germline mRNAs⁴². A significant portion of germline transcripts
191 misexpressed in *Kdm5c*-KO EpiLCs are known binding targets of DAZL, including *Stra8*⁴³ (p = 1.698e-07,
192 Fisher's Exact Test). This suggests expression of DAZL protein could promote the translation of other
193 aberrant germline transcripts, influencing their ability to impact *Kdm5c*-KO cellular function. We thus tested
194 DAZL protein expression in *Kdm5c*-KO EpiLCs through immunocytochemistry (Figure 3I). We observed
195 about 25% of *Kdm5c*-KO EpiLCs expressed DAZL protein and it was localized to the cytoplasm (p = 0.0015,
196 Welch's t-test), consistent with the pattern of DAZL expression in spermatogonia⁴³. Altogether these results
197 suggest tissue-specific genes are misexpressed during *Kdm5c*-KO embryogenesis, including key drivers of
198 germline identity that can be translated into protein.

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

200 • KDM5C may also be involved in this early decommissioning of germline genes, as re-expression of
201 KDM5C in knockout neurons fails to suppress their dysregulation¹⁰.

202 Previous work suggests KDM5C represses germline genes during early development, as re-expression
203 of KDM5C in knockout neuronal cultures fails to suppress hippocampal germline DEGs and KDM5C is not
204 bound to their promoters in neurons¹⁰. There is some evidence KDM5C binds to select germline gene
205 promoters in ESCs¹⁰, including *Dazl*^{41,44}. As KDM5C's binding at germline gene promoters has not been
206 systematically characterized, it is currently unclear if KDM5C is enriched at germline gene promoters, what
207 types of germline genes KDM5C regulates, and if its binding is maintained at any germline genes in neurons.

208 To further characterize the mechanism behind *Kdm5c*-KO germline gene misexpression, we analyzed
209 KDM5C chromatin immunoprecipitation followed by DNA sequencing (ChIP-seq) datasets in EpiLCs³⁷ and
210 primary neuron cultures (PNCs) from the cortex and hippocampus¹⁵. EpiLCs had a higher total number of
211 KDM5C peaks than PNCs (EpiLCs: 5,808, PNCs: 1,276, MACS2 q < 0.1 and fold enrichment > 1, removal
212 of *Kdm5c*-KO false positive peaks) and KDM5C was primarily localized to gene promoters in both cell types
213 (EpiLCs: 4,190, PNCs: 745 +/- 500bp from the TSS), although PNCs showed increased localization to
214 non-promoter regions (Figure 4A).

215 The majority of promoters bound by KDM5C in PNCs were also bound in EpiLCs (513 shared promoters),
216 however a large portion of gene promoters were only bound by KDM5C in EpiLCs (3677 EpiLC only
217 promoters) (Figure 4B). We then performed gene ontology analysis to compare the known functions of genes
218 with KDM5C bound to their promoter in EpiLCs and PNCs. While there were no ontologies significantly
219 enriched for genes only bound by KDM5C in PNCs, gene ontologies for genes shared between PNCs and
220 EpiLCs were enriched for functions involving nucleic acid turnover, such as deoxyribonucleotide metabolic
221 process (GO:0009262, p.adjust = 8.28e-05) (Figure 4C). Germline-specific ontologies were only enriched
222 in promoters unique to EpiLCs, such as meiotic nuclear division (GO: 0007127 p.adjust = 6.77e-16) and
223 meiotic cell cycle process (GO:1903046, p.adjust = 5.05e-16) (Figure 3C). We then evaluated KDM5C
224 binding around the transcription start site (TSS) of all germline-enriched genes. While KDM5C was bound
225 to a subset of germline gene promoters in EpiLCs, it was not bound to any in PNCs (Figure 4D). Together,
226 this suggests KDM5C is significantly enriched at a subset of germline gene promoters in EpiLCs, including
227 meiotic genes, but does not regulate germline genes in neurons.

228 We then compared KDM5C binding at RNA-seq DEG promoters to determine if the germline mRNAs
229 expressed in the *Kdm5c*-KO brain and EpiLCs are direct targets of KDM5C (Figure 4E). About one third
230 of EpiLC-specific and brain-specific (hippocampus, amygdala, or both) germline DEGs were bound by
231 KDM5C in EpiLCs (EpiLC only: 36%, Brain only: 33.3%). Some notable differences in KDM5C binding
232 for EpiLC-specific DEGs included *Dazl* and *Stra8*, two key drivers of germline identity discussed above.
233 Although both mRNAs are expressed in *Kdm5c*-KO EpiLCs, KDM5C is only bound to *Dazl*'s promoter and

234 not *Stra8*'s in EpiLCs (Figure 4F). In contrast to the unique DEGs, 3 out of the 4 genes dysregulated in both
235 the brain and EpiLCs are bound by KDM5C around the TSS (Figure 4G). Again, we did not observe any
236 KDM5C binding at germline gene promoters in PNCs, even for brain-specific DEGs (Figure 4H). Altogether,
237 this suggests the majority of germline mRNAs expressed in *Kdm5c*-KO cells are dysregulated independent
238 of direct KDM5C recruitment to their promoters during embryogenesis.

239 **notes:** - do Direct vs indirect DEGs motif analysis

240 **KDM5C erases H3K4me3 to promote long-term repression of germline genes via**
241 **DNA methylation**

242 Although KDM5C is generally thought to suppress transcription through erasure of histone 3 lysine 4 di-
243 and trimethylation (H3K4me2/3)¹¹, recent studies in ESCs have suggested KDM5C's repression *Dazl* is
244 independent of its catalytic activity⁴¹. Somatic repression of germline genes is typically established during the
245 transition between naïve and primed pluripotency, which modeled *in vitro* as ESCs to EpiLC differentiation.
246 In ESCs, chromatin regulators place repressive histone modifications at germline gene promoters, including
247 histone 2A 119 monoubiquitination (H2AK119ub1) and histone 3 lysine 9 trimethylation (H3K9me3)^{16,17,45}.
248 Germline genes are then silenced long-term in EpiLCs by *de novo* placement of DNA CpG methylation
249 (DNAme)¹⁷. Although KDM5C's catalytic activity is not required to suppress *Dazl* in ESCs, KDM5C may
250 promote long-term germline gene silencing via H3K4me3 removal since H3K4me3 can impede DNAme
251 placement^{46,47} and DNAme is lost at select germline gene promoters in the hippocampus¹⁰. Because
252 KDM5C's role in germline gene repression has only been characterized in ESCs and in the mature brain, it is
253 currently unclear to what extent KDM5C is involved during transition between ESCs and EpiLCs and if its
254 catalytic activity is required for long-term silencing.

255 To elucidate KDM5C's role in germline gene silencing, we first characterized KDM5C's substrates, histone
256 3 lysine 4 di- and trimethylation (H3K4me2/3) in our previously published ChIP-seq datasets in wild type and
257 *Kdm5c*-KO amygdala²¹ and EpiLCs³⁷. In congruence with previous work in the *Kdm5c*-KO hippocampus¹⁰,
258 we observed aberrant accumulation of H3K4me3 around the transcription start site (TSS) of germline genes
259 in the *Kdm5c*-KO amygdala (Figure 5A). We additionally found a marked increase in H3K4me2 around the
260 TSS of germline genes in *Kdm5c*-KO EpiLCs (Figure 5B).

261 We then evaluated KDM5C's embryonic role in germline gene silencing during ESC to EpiLC differentiation.
262 We first characterized KDM5C's mRNA and protein expression at 0, 24, and 48 hours of differentiation
263 (Figure 5C). While *Kdm5c* mRNA steadily decreased from 0 to 48 hours of differentiation (Figure 5D), KDM5C
264 protein initially increased from 0 to 24 hours but then decreased to near knockout levels by 48 hours (Figure
265 5E). We then performed whole genome bisulfite sequencing in wild-type and *Kdm5c*-KO ESCs and 96 hour
266 extended EpiLCs (exEpiLCs) to determine KDM5C's role in the initial placement of DNA methylation at
267 germline gene promoters. While wild-type cells accumulated high levels of DNA methylation at germline

268 gene promoters over the course of ESCs to exEpiLC differentiation, DNA methylation was markedly reduced
269 in *Kdm5c*-KO exEpiLCs (Figure 5F).

- 270 • Catalytic activity
271 • Altogether, this suggests KDM5C is dynamically regulated during ESC to EpiLC differentiation and
272 promotes germline gene silencing via removal of H3K4me2/3 during early embryogenesis.
273 • **note:** should I test if increase in H3K4me2 and H3K4me3 was highest in *Kdm5c*-KO cells at genes
274 bound by KDM5C at their promoter in EpiLCs? Don't think it's that impactful

275 **Discussion**

276 In the above study, we demonstrate KDM5C's pivotal role in the development of tissue identity, resulting in
277 substantial misexpression of tissue-enriched genes within the adult *Kdm5c*-KO hippocampus and amygdala.
278 We defined tissue-enriched genes based on the criteria used in a previously published dataset of 17 C57/Bl6J
279 mouse tissues²⁰, which defined genes as tissue-enriched if they had more than 4-fold higher expression
280 when compared to any other tissue. In addition to testis genes identified previously¹⁰, we found significant
281 enrichment of muscle, liver, and even ovary-biased genes aberrantly expressed within the male *Kdm5c*-KO
282 brain. It is currently unclear what the impact of these tissue-enriched genes are upon *Kdm5c*-KO brain
283 function and if they ultimately contribute to MRXSCJ-related phenotypes. While the majority of tissue-
284 enriched DEGs were testis and ovary genes with no known brain functions, select liver and muscle-biased
285 DEGs do have known roles in the wild-type brain. For example, *Apolipoprotein C-I (Apoc1)* is a lipid transport
286 gene highly enriched within the liver, but is also lowly expressed in the wild-type brain. Intriguingly, *Apoc1*
287 overexpression in the mouse brain can impair learning and memory⁴⁸ and is implicated in Alzheimer's
288 disease in humans⁴⁹. Our results suggest KDM5C fine-tunes the expression of tissue-enriched genes like
289 *Apoc1* to match the level required for proper brain function. However, further studies are required to determine
290 if these genes contribute to *Kdm5c*-KO neuronal impairments and MRXSCJ pathology.

291 A substantial portion of genes upregulated in *Kdm5c*-KO brain are testis genes that have no known
292 function within the brain. Through the use of publically available RNA-seq datasets, we demonstrated these
293 testis-enriched DEGs are typically unique to germ cells. Misexpression of germline genes in the brain sug-
294 gests *Kdm5c*-KOs fail to demarcate between somatic and germline lineages, a key feature of multicellularity
295 and sexual reproduction. Previous studies have demonstrated germline genes are decommissioned in
296 somatic cells by multiple chromatin-modifying enzymes, but have primarily focused on the repression of
297 key marker genes. To globally characterize KDM5C's role in germline gene repression, we curated a list
298 of mouse germline-enriched genes using publically available germ cell-depleted RNA-seq datasets. This
299 resource enabled us to identify 1) the extent of germline gene dysregulation in *Kdm5c*-KO cells, 2) the types

300 of germline genes misexpressed at different developmental time points, and 3) which types of germline
301 genes are directly or indirectly regulated by KDM5C.

302 Dysregulation of *Kdm5c*-KO tissue identity begins during early embryogenesis, as germline and other
303 tissue-enriched genes are aberrantly expressed in epiblast-like cells (EpiLCs). *In vivo*, germline genes
304 are typically decommissioned in epiblast stem cells and remain silenced as the epiblast differentiates
305 into the body's somatic tissues³³. However, a small subset of epiblast stem cells will receive signals to
306 reactivate germline gene expression to become the primordial germ cells (PGCs) that will ultimately form
307 the mature germline^{31,32}. This process can be mimicked *in vitro* by differentiating EpiLCs into primordial
308 germ cell-like cells (PGCLCs)⁵⁰. Thus, misexpression of germline genes in *Kdm5c*-KO EpiLCs might
309 suggest they are progressing beyond EpiLC differentiation and becoming PGCLCs. Yet, *Kdm5c*-KO EpiLCs
310 had proper expression of EpiLC marker genes for primed pluripotency and we observed no difference in
311 cellular morphology during ESC to EpiLC differentiation. Furthermore, we saw no significant change in *Otx2*
312 expression, an epiblast stem cell marker that is known to repress epiblast and EpiLC differentiation into PGC
313 and PGCLCs⁵¹. This, in conjunction with *Kdm5c*-KO mice being viable, suggests germline gene expression
314 is occurring ectopically in conjunction with typical developmental programs, rather than a complete shift to
315 germline identity.

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

318 We then globally characterized KDM5C binding at germline-enriched gene promoters through analysis
319 of KDM5C ChIP-seq in EpiLCs and primary hippocampus and cortex neuron cultures (PNCs). While we
320 observed high enrichment of KDM5C peaks at germline genes in EpiLCs, KDM5C was not bound to germline
321 gene promoters in PNCs. This suggests dysregulation of germline genes in the mature *Kdm5c*-KO brain is
322 due to loss of repression during embryogenesis, which is consistent with previous work that found introducing
323 human KDM5C into *Kdm5c*-KO PNCs does not repress germline transcripts¹⁰. Although KDM5C is enriched
324 at germline gene promoters in EpiLCs, KDM5C was only bound to about a third of EpiLC germline-enriched
325 DEGs. One notable DEG not bound by KDM5C is *Stra8*, a transcription factor activated by retinoic acid
326 signaling in germ cells that promotes meiotic initiation^{52,53}. Retinoic acid can only activate *Stra8* expression
327 when DAZL is present, which is a direct target of KDM5C and is aberrantly transcribed and translated in
328 *Kdm5c*-KO EpiLCs. This indicates germline genes can be aberrantly expressed in *Kdm5c*-KO cells indirectly
329 of KDM5C regulation through activation by other ectopic germline programs. Consistent with this idea, many
330 *Kdm5c*-KO EpiLC germline DEGs are important for early germ cell development and meiosis while those
331 expressed in the mature *Kdm5c*-KO brain are involved in late sperm development. Altogether, this indicates
332 ectopic germline programs are, to some extent, progressing through germ cell developmental stages over
333 the course of *Kdm5c*-KO development.

334 • Talk about different motifs if we do see differences there and if it explains direct vs indirect dysregulation

- 335 – Could also mention PRC1.6/KDM5C binds pcgf6 in ESCs
- 336 • Maybe talk about other regulators/if anything is known about them long term? Or save for 2nd paper.
- 337 – another germline repressor: <https://pubmed.ncbi.nlm.nih.gov/35345626/>
- 338 • 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>
- 340 It is important to note that while we focused on KDM5C's regulation of germline genes, some germline-enriched genes are also expressed in naïve ESCs and at the 2-cell stage due to their roles in pluripotency and self-renewal. For example, although primarily known for committing PGCs to the germline fate and regulating the translation of germline-specific RNAs, *Dazl* is also expressed in naïve ESCs *in vitro*³⁴, the inner cell mass *in vivo*³⁴, and at the 2-cell stage⁵⁴, but is silenced when cells transition from naïve to primed pluripotency during ESC to EpiLC differentiation¹⁷. Very recently, two screens of *Dazl*-repressors in ESCs identified KDM5C as a direct regulator of *Dazl* expression^{41,44}. Interestingly, one screen found *Kdm5c*-KO ESCs also aberrantly expressed 2-cell-specific genes, indicating KDM5C negatively regulates totipotency in ESCs⁴¹. We found that while KDM5C also represses *Dazl* expression in EpiLCs, *Kdm5c*-KO EpiLCs do not express 2-cell specific genes like *Dux* and *Zscan4c*. Out of the four 2-cell regulators characterized in ESCs, KDM5C was the only factor whose repression of *Dazl* was independent of *Dux* expression⁴¹. Together, this suggests the 2-cell-like state in *Kdm5c*-KO ESCs is an indirect consequence of KDM5C's primary role in germline gene repression, including germline genes that are also involved in pluripotency and self-renewal.
- 353 It is hypothesized distinct repressive chromatin marks are placed at germline gene promoters as the embryo transitions from naïve to primed pluripotency. Initially, germline genes are repressed by placement of histone 2A lysine 119 monoubiquitination (H2AK119ub1) and histone 3 lysine 9 trimethylation (H3K9me3) in ESCs and then gain *de novo* DNA CpG methylation (CpGme) in EpiLCs^{17–19,45,55,56}. KDM5C may be instead required to remove an active mark, histone 3 lysine 4 trimethylation (H3K4me3), since H3K4me3 can impede *de novo* CpGme placement^{46,47}. This is supported by previous work in the *Kdm5c*-KO adult hippocampus, which found an increase in H3K4me3 around the transcription start site (TSS) of germline DEGs and loss of CpGme at at least two germline promoters¹⁰. However, KDM5C's role in embryonic germline gene silencing is still unclear, as KDM5C's catalytic activity was recently found to be dispensable for silencing DAZL in ESCs⁴¹. In this study, we observed a global increase in H3K4me3 around the TSS of germline-enriched gene in the *Kdm5c*-KO amygdala and an increase in H3K4me2 in *Kdm5c*-KO EpiLCs. We found KDM5C's expression is dynamically regulated during ESC to EpiLC differentiation and loss of KDM5C leads to impaired placement of CpGme in extended EpiLCs. Altogether, this suggests KDM5C is necessary during the transition from naïve to primed pluripotency to promote the initial placement of CpGme at CpG islands for the long-term silencing of germline genes.
- 368 • Talk about impact - Other germline gene chromatin regulators, NDDs, cancer. Maybe talk about myt1l

369 • include cancer in there somewhere (Somatic misexpression of germline genes has been implicated in
370 many cancers.)

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

383 Materials and Methods

384 Classifying tissue-enriched and germline-enriched genes

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

389 We curated a list of germline-enriched genes using an RNA-seq dataset from wild-type and germline-
390 depleted (Kit^{W/Wv}) male and female mouse embryos from embryonic day 12, 14, and 16³⁰, as well as adult
391 male testes²⁷. Germline-enriched genes met the following criteria: 1) their expression is greater than 1
392 FPKM in wild-type germline 2) their expression in any wild-type somatic tissues²⁰ does not exceed 20%
393 of maximum expression in wild-type germline, and 3) their expression in the germ cell-depleted (Kit^{W/Wv})
394 germline, for any sex or time point, does not exceed 20% of maximum expression in wild-type germline.

395 Cell culture

396 We utilized our previously established cultures of male wild-type and *Kdm5c* knockout (-KO) embryonic
397 stem cells³⁷. Sex was confirmed by genotyping *Uba1* on the X and Y chromosomes with the following
398 primers: 5'-TGGATGGTGTGCCAATG-3', 5'-CACCTGCACGTTGCCCTT-3'. Deletion of *Kdm5c* was
399 confirmed through the primers 5'-ATGCCCATATTAAGAGTCCCTG-3', 5'-TCTGCCCTGATGGGACTGTT-3',
400 and 5'-GGTTCTAACACTCACATAGTG-3'.

401 Embryonic stem cells (ESCs) were initially cultured in primed ESC (pESC) media consisting of Knock-
402 Out DMEM (Gibco#10829-018), fetal bovine serum (Gibco#A5209501), KnockOut serum replacement
403 (Invitrogen#10828-028), Glutamax (Gibco#35050-061), Anti-Anti (Gibco#15240-062), MEM Non-essential
404 amino acids (Gibco#11140-050), and beta-mercaptoethanol (Sigma#M7522). They were then transitioned
405 into ground-state “naive” ESCs (nESCs) by culturing for four passages in N2B27 media containing DMEM/F12
406 (Gibco#11330-032), Neurobasal media (Gibco#21103-049), Gluamax, Anti-Anti, N2 supplement (Invitro-
407 gen#17502048), and B27 supplement without vitamin A (Invitrogen#12587-010), and beta-mercaptoethanol.
408 Both pESC and nESC media were supplemented with the GSK3 inhibitor CHIR99021 (Sigma #SML1046-
409 5MG), the MEK inhibitor PD0325901 (Sigma #PZ0162-5MG), and leukemia inhibitory factor (LIF, Milli-
410 pore#ESG1107).

411 nESCs were differentiated into epiblast-like cells (EpiLCs, 48 hours) and extendend EpiLCs (exEpiLCs,
412 96 hours) by culturing in N2B27 media containing DMEM/F12, Neurobasal media, Gluamax, Anti-Anti,
413 N2 supplement, B27 supplement (Invitrogen#17504044), beta-mercaptoethanol, fibroblast growth factor 2
414 (FGF2, R&D Biotechne 233-FB), and activin A (R&D Biotechne 338AC050CF).

415 **Immunocytochemistry (ICC)**

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

423 **RNA sequencing**

424 **Published datasets**

425 All published datasets are available at the Gene Expression Omnibus (GEO) <https://www.ncbi.nlm.nih.gov/geo/>. Previously published RNA sequencing datasets analyzed in this study included the male wild-type
426 and *Kdm5c*-KO adult amygdala and hippocampus²¹ (available at GEO: GSE127722) and male wild-type and
427 *Kdm5c*-KO EpiLCs³⁷ (available at GSE: GSE96797).

429 **Alignment and analysis**

430 After ensuring read quality via FastQC (v0.11.8), reads were then mapped to the mm10 *Mus musculus*
431 genome (Gencode) using STAR (v2.5.3a), during which we removed duplicates and kept only uniquely

432 mapped reads. Count files were generated by FeatureCounts (Subread v1.5.0), and BAM files were
433 converted to bigwigs using deeptools (v3.1.3) and visualized by the UCSC genome browser. RStudio (v3.6.0)
434 was then used to analyze counts files by DESeq2 (v1.26.0)²² to identify differentially expressed genes
435 (DEGs) with a q-value (p-adjusted via FDR/Benjamini–Hochberg correction) less than 0.1 and a log2 fold
436 change greater than 0.5. For all DESeq2 analyses, log2 fold changes were calculated with IfcShrink using
437 the ashr package⁵⁷. MA-plots were generated by ggpahr (v0.4.0), and Eulerr diagrams were generated by
438 eulerr (v6.1.1). Boxplots and scatterplots were generated by ggpahr (v0.6.0) and ggplot2 (v3.3.2). Heatmaps
439 of gene expression were generated using the base R functions scale and hclust and visualized using the R
440 package ComplexHeatmap (v2.12.1). The Upset plot was generated via the package UpSetR (v1.4.0)⁵⁸.
441 Gene ontology (GO) analyses were performed by the R package enrichPlot (v1.16.2) using the biological
442 processes setting.

443 Chromatin immunoprecipitation followed by DNA sequencing (ChIP-seq)

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

448 ChIP-seq reads were aligned to mm10 using Bowtie1 (v1.1.2) allowing up to two mismatches. Only
449 uniquely mapped reads were used for analysis. Peaks were called using MACS2 software (v2.1.0.20140616)
450 using input BAM files for normalization, with filters for a q-value < 0.1 and a fold enrichment > 1. Common
451 peak sets were obtained in R via DiffBind (v3.6.5), and count tables for the common peaks were generated
452 with the Bedtools multicov command. We removed “black-listed” genomic regions that often give aberrant
453 signals. Peak proximity to genome annotations was determined by ChIPSeeker (v1.32.1). Enriched motifs
454 were identified using HOMER⁵⁹. Gene ontology (GO) analyses were performed by the R package enrichPlot
455 (v1.16.2) using the biological processes setting. Average binding across the genome was visualized using
456 deeptools (v3.1.3). Bigwigs were visualized using the UCSC genome browser.

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

458 **Data availability**

459 **Acknowledgements**

460 **References**

- 461 1. Strahl, B.D., and Allis, C.D. (2000). The language of covalent histone modifications. *Nature* *403*,
462 41–45. <https://doi.org/10.1038/47412>.
- 463 2. Jenuwein, T., and Allis, C.D. (2001). Translating the Histone Code. *Science* *293*, 1074–1080.
464 <https://doi.org/10.1126/science.1063127>.
- 465 3. Lewis, E.B. (1978). A gene complex controlling segmentation in *Drosophila*. *Nature* *276*, 565–570.
466 <https://doi.org/10.1038/276565a0>.
- 467 4. Kennison, J.A., and Tamkun, J.W. (1988). Dosage-dependent modifiers of polycomb and antennapedia
468 mutations in *Drosophila*. *Proc. Natl. Acad. Sci. U.S.A.* *85*, 8136–8140. <https://doi.org/10.1073/pnas.8>
5.21.8136.
- 469 5. Kassis, J.A., Kennison, J.A., and Tamkun, J.W. (2017). Polycomb and Trithorax Group Genes in
470 *Drosophila*. *Genetics* *206*, 1699–1725. <https://doi.org/10.1534/genetics.115.185116>.
- 471 6. Gabriele, M., Lopez Tobon, A., D'Agostino, G., and Testa, G. (2018). The chromatin basis of
472 neurodevelopmental disorders: Rethinking dysfunction along the molecular and temporal axes. *Prog
Neuropsychopharmacol Biol Psychiatry* *84*, 306–327. <https://doi.org/10.1016/j.pnpbp.2017.12.013>.
- 473 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
474 Bdnf transcription, dendritic growth, and spine maturation. *Neuron* *52*, 255–269. <https://doi.org/10.1016/j.neuron.2006.09.037>.
- 475 8. Hirabayashi, Y., Suzki, N., Tsuboi, M., Endo, T.A., Toyoda, T., Shinga, J., Koseki, H., Vidal, M., and
476 Gotoh, Y. (2009). Polycomb Limits the Neurogenic Competence of Neural Precursor Cells to Promote
Astrogenic Fate Transition. *Neuron* *63*, 600–613. <https://doi.org/10.1016/j.neuron.2009.08.021>.
- 477 9. Schaefer, A., Sampath, S.C., Intrator, A., Min, A., Gertler, T.S., Surmeier, D.J., Tarakhovsky, A.,
478 and Greengard, P. (2009). Control of cognition and adaptive behavior by the GLP/G9a epigenetic
suppressor complex. *Neuron* *64*, 678–691. <https://doi.org/10.1016/j.neuron.2009.11.019>.

- 479 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**,
480 47–59. <https://doi.org/10.1016/j.celrep.2017.09.014>.
- 481 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>.
- 482 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
483 Med Genet* **94**, 1–4.
- 484 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
485 Genet* **76**, 227–236. <https://doi.org/10.1086/427563>.
- 486 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.
487 <https://doi.org/10.1111/cge.13755>.
- 488 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
489 Impaired Removal of Histone Methylation. *Cell Reports* **14**, 1000–1009. <https://doi.org/10.1016/j.celrep.2015.12.091>.
- 490 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
491 stem cells by regulating germ cell-related genes. *eLife* **6**. <https://doi.org/10.7554/eLife.21064>.
- 492 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
in the early embryo precedes DNA methylation-mediated silencing. *Nat Commun* **12**, 7020. <https://doi.org/10.1038/s41467-021-27345-x>.
- 493 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
494 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>.

- 497 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>.
- 498
- 499 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>.
- 500
- 501 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>.
- 502
- 503 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>.
- 504
- 505 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>.
- 506
- 507 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>.
- 508
- 509 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>.
- 510
- 511 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>.
- 512
- 513 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>.
- 514
- 515 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>.
- 516

- 517 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>.
- 518 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>.
- 520 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>.
- 522 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>.
- 524 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>.
- 526 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>.
- 528 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>.
- 530 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>.
- 532 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>.
- 534 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>.
- 536 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>.

- 539 40. Endo, T., Mikedis, M.M., Nicholls, P.K., Page, D.C., and De Rooij, D.G. (2019). Retinoic Acid and Germ
540 Cell Development in the Ovary and Testis. *Biomolecules* *9*, 775. <https://doi.org/10.3390/biom9120775>.
- 541 41. Gupta, N., Yakhou, L., Albert, J.R., Azogui, A., Ferry, L., Kirsh, O., Miura, F., Battault, S., Yamaguchi,
542 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>.
- 543 42. Li, H., Liang, Z., Yang, J., Wang, D., Wang, H., Zhu, M., Geng, B., and Xu, E.Y. (2019). DAZL
544 is a master translational regulator of murine spermatogenesis. *Natl Sci Rev* *6*, 455–468. <https://doi.org/10.1093/nsr/nwy163>.
- 545 43. Mikedis, M.M., Fan, Y., Nicholls, P.K., Endo, T., Jackson, E.K., Cobb, S.A., De Rooij, D.G., and Page,
546 D.C. (2020). DAZL mediates a broad translational program regulating expansion and differentiation of
547 spermatogonial progenitors. *eLife* *9*, e56523. <https://doi.org/10.7554/eLife.56523>.
- 548 44. 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
549 of germline genes in mouse embryonic stem cells. *Nucleic Acids Research* *51*, 3130–3149. <https://doi.org/10.1093/nar/gkad071>.
- 550 45. Liu, M., Zhu, Y., Xing, F., Liu, S., Xia, Y., Jiang, Q., and Qin, J. (2020). The polycomb group protein
PCGF6 mediates germline gene silencing by recruiting histone-modifying proteins to target gene
551 promoters. *J Biol Chem* *295*, 9712–9724. <https://doi.org/10.1074/jbc.RA119.012121>.
- 552 46. Otani, J., Nankumo, T., Arita, K., Inamoto, S., Ariyoshi, M., and Shirakawa, M. (2009). Structural basis
for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX–DNMT3–DNMT3L
553 domain. *EMBO Reports* *10*, 1235–1241. <https://doi.org/10.1038/embor.2009.218>.
- 554 47. Guo, X., Wang, L., Li, J., Ding, Z., Xiao, J., Yin, X., He, S., Shi, P., Dong, L., Li, G., et al. (2015).
Structural insight into autoinhibition and histone H3-induced activation of DNMT3A. *Nature* *517*,
640–644. <https://doi.org/10.1038/nature13899>.
- 555 48. Abildayeva, K., Berbée, J.F.P., Blokland, A., Jansen, P.J., Hoek, F.J., Meijer, O., Lütjohann, D., Gautier,
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>.
- 556 49. Leduc, V., Jasmin-Bélanger, S., and Poirier, J. (2010). APOE and cholesterol homeostasis in
Alzheimer's disease. *Trends in Molecular Medicine* *16*, 469–477. <https://doi.org/10.1016/j.molmed.2010.07.008>.
- 557 50. Hayashi, K., Ohta, H., Kurimoto, K., Aramaki, S., and Saitou, M. (2011). Reconstitution of the
558 mouse germ cell specification pathway in culture by pluripotent stem cells. *Cell* *146*, 519–532.
559 <https://doi.org/10.1016/j.cell.2011.06.052>.

- 561 51. Zhang, J., Zhang, M., Acampora, D., Vojtek, M., Yuan, D., Simeone, A., and Chambers, I. (2018).
OTX2 restricts entry to the mouse germline. *Nature* **562**, 595–599. <https://doi.org/10.1038/s41586-018-0581-5>.
- 562
- 563 52. Endo, T., Romer, K.A., Anderson, E.L., Baltus, A.E., De Rooij, D.G., and Page, D.C. (2015). Periodic 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>.
- 564
- 565 53. Bowles, J., Knight, D., Smith, C., Wilhelm, D., Richman, J., Mamiya, S., Yashiro, K., Chawengsaksophak, 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>.
- 566
- 567 54. 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>.
- 568
- 569 55. 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>.
- 570
- 571 56. 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>.
- 572
- 573 57. Stephens, M. (2016). False discovery rates: A new deal. *Biostat*, kxw041. <https://doi.org/10.1093/biostatistics/kxw041>.
- 574
- 575 58. 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>.
- 576
- 577 59. 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>.
- 578

579 **Figures and Tables**

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

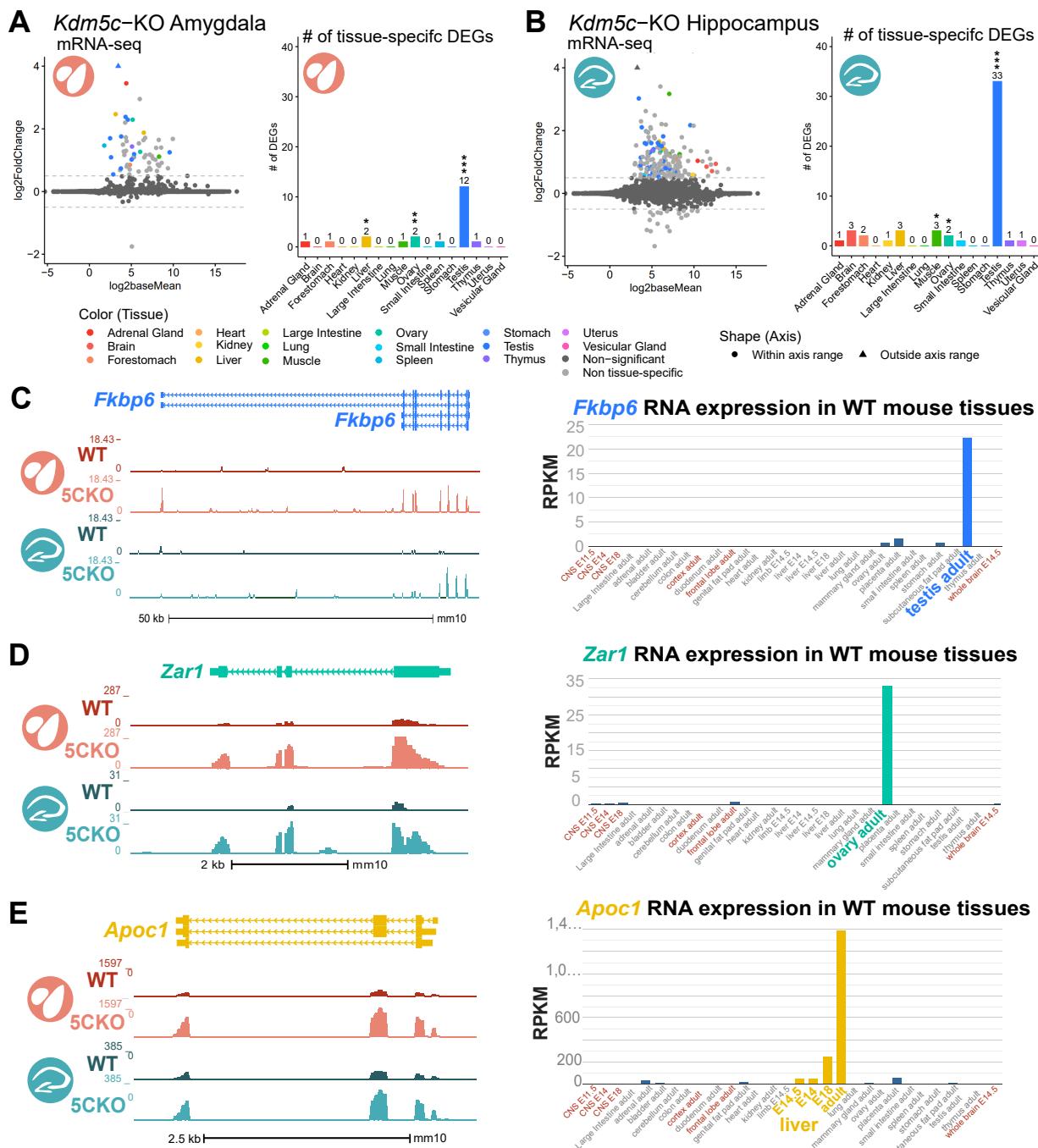


Figure 1: Tissue-enriched genes are misexpressed in the *Kdm5c*-KO brain. **A-B.** Expression of tissue-enriched genes (Li et al 2017) in the male *Kdm5c*-KO amygdala (A) and hippocampus (B). Left - MA plot of mRNA-sequencing. Right - Number of tissue-enriched differentially expressed genes (DEGs). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, 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.

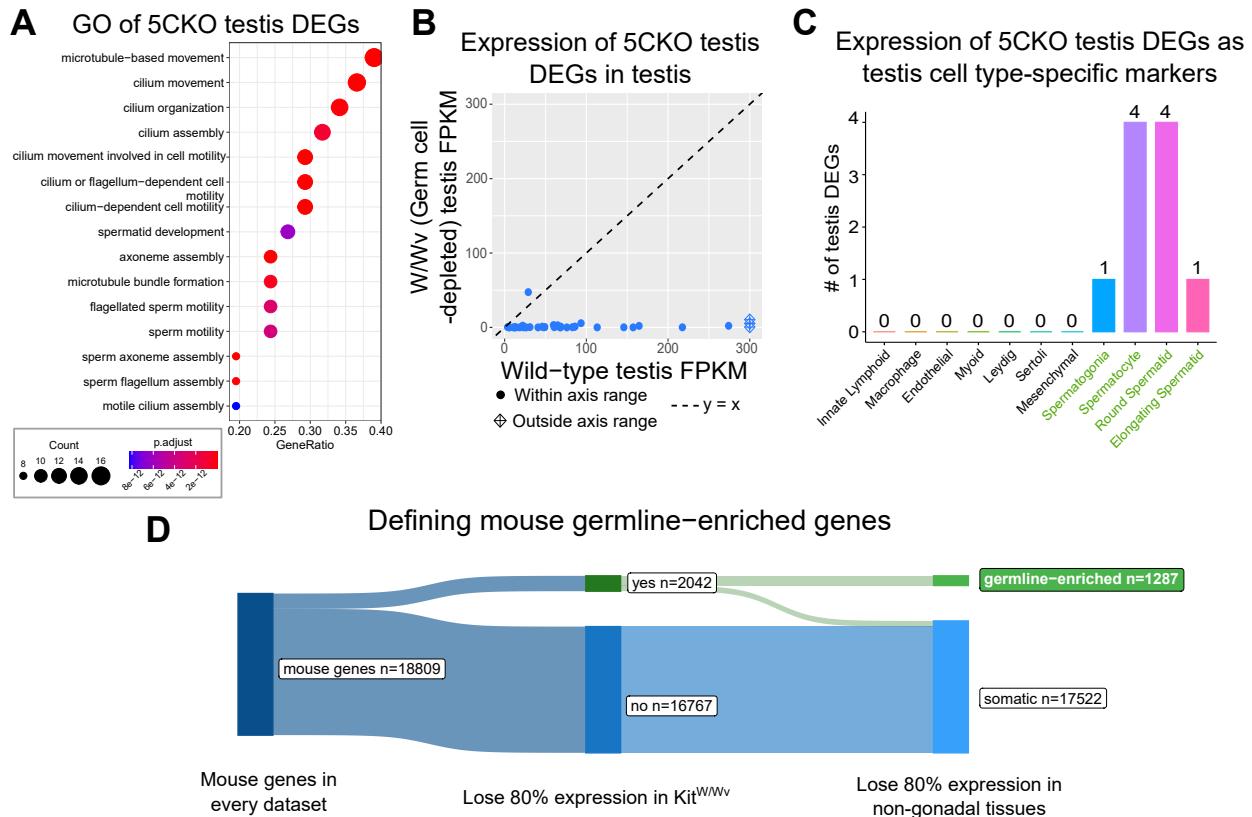


Figure 2: Aberrant transcription of germline genes in the *Kdm5c*-KO in the brain. **A.** enrichPlot gene ontology (GO) of *Kdm5c*-KO amygdala and hippocampus testis-enriched DEGs **B.** Expression of testis DEGs in wild-type (WT) testis versus germ cell-depleted (W/W_v) testis (Mueller et al 2013). 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).

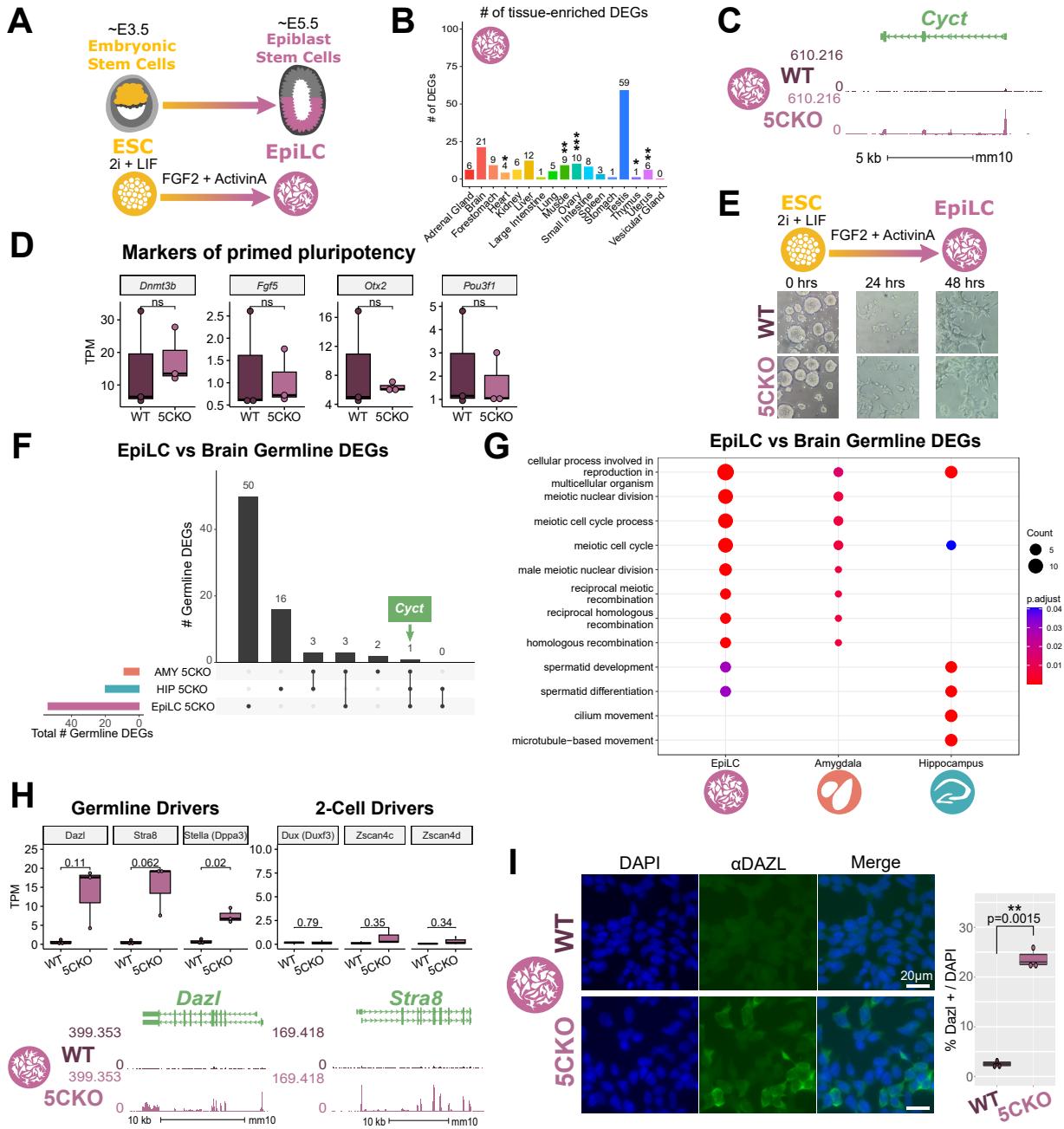


Figure 3: Kdm5c-KO epiblast-like cells express key drivers of germline identity **A.** Top - Diagram of *in vivo* differentiation of embryonic stem cells (ESCs) of the inner cell mass into epiblast stem cells. Bottom - *in vitro* differentiation of ESCs into epiblast-like cells (EpiLCs). **B.** Number 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. **I.** 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.

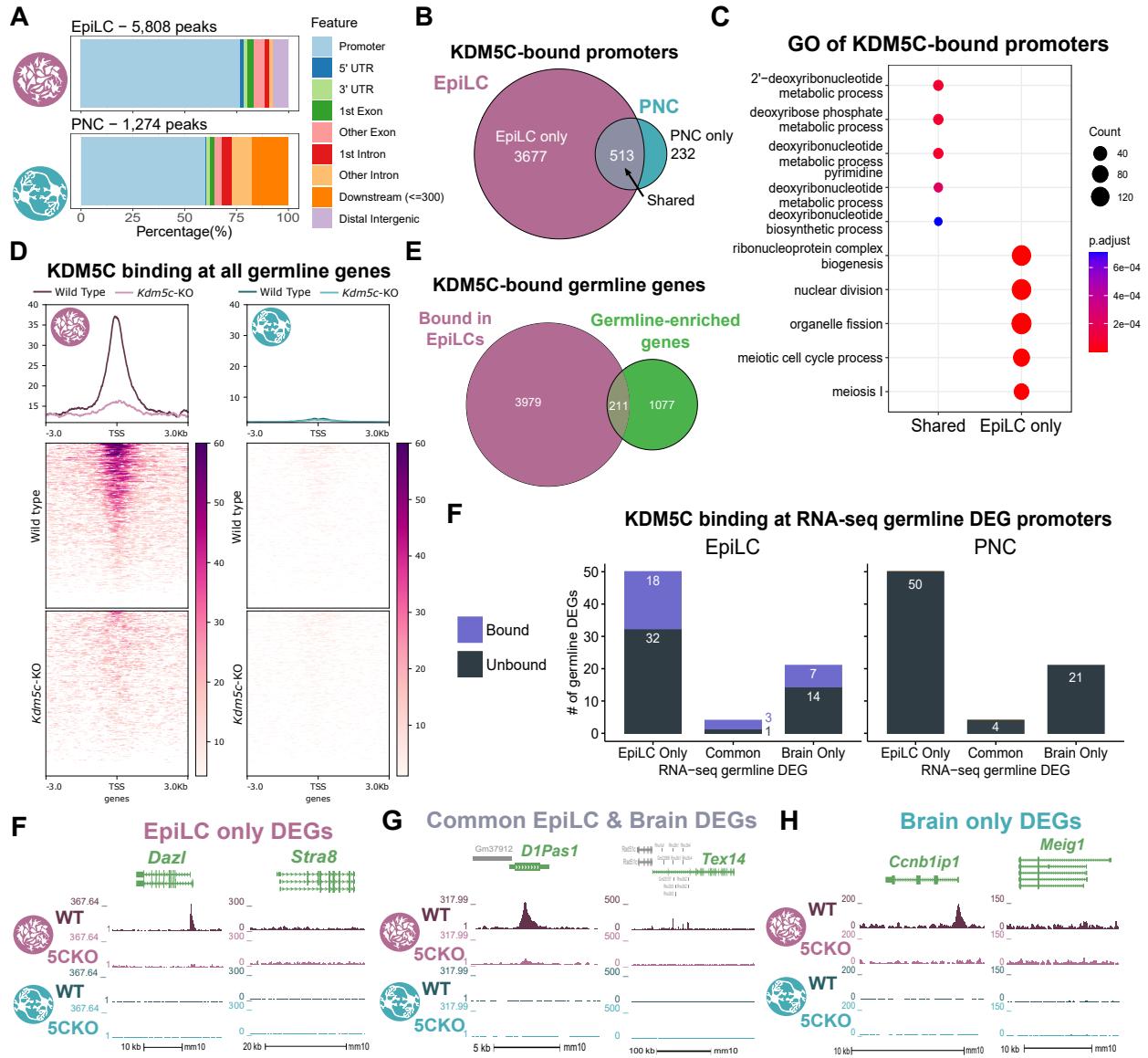
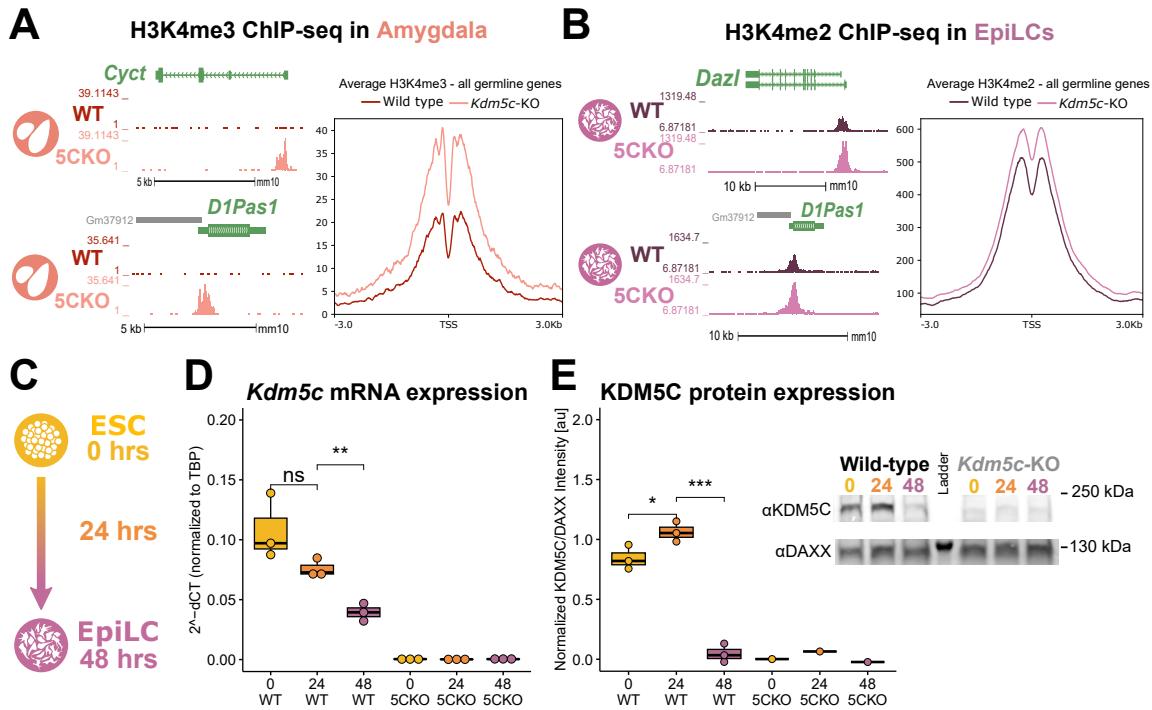


Figure 4: KDM5C binds to a subset of germline gene promoters during early embryogenesis **A.** ChIPseeker localization of KDM5C peaks at different genomic regions in EpiLCs (top) and hippocampal and cortex primary neuron cultures (PNCs, bottom). **B.** Overlap of genes with KDM5C bound to their promoters 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

584 **Notes**

585 **Discussion notes**

- 586 • We additionally found *Kdm5c*-KO EpiLCs ectopically express DAZL protein that is localized to the
587 cytoplasm, similar to its morphology in spermatogonia⁴³. **note: maybe just put in results.** Could
588 move around depending upon if I get pheno working.
- 589 • end with something like: this indicates kdm5c not only modulates pluripotency and self-renewal in
590 ESCs, but also has a role in long-term silencing of germline genes
 - 591 – then transition into the long term silencing mechanism paragraph
- 592 • KDM5C is dynamically regulated during the window of embryonic germline gene silencing, our EpiLC
593 ChIP-seq is likely catching the tail end of KDM5C's main involvement.
- 594 • Our results indicate KDM5C is a pivotal repressor of germline programs in somatic cells.
- 595 • Studies on other chromatin regulators have focused on key marker genes to identify germline gene
596 misexpression, such as *Dazl*.
- 597 • Using a germline-depleted mouse model, we curated a list of mouse germline-enriched genes to
598 globally assess germline gene dysregulation.
- 599 • Many of the genes dysregulated in the mature *Kdm5c*-KO brain were important for late stages of
600 spermatogenesis, such as those important for sperm flagellar structure. Contrastingly, *Kdm5c*-KO
601 EpiLCs aberrantly expressed key drivers of germ cell identity, including *Dazl* and *Stra8*.
- 602 • While a significant portion of KDM5C-bound promoters are germline genes, many germline genes
603 expressed during *Kdm5c*-KO embryogenesis are not directly bound by kdm5c.
- 604 • One notable example is *Stra8*, a transcription factor important for germ cell specific transcription and
605 meiotic initiation
- 606 • The including the demarcation between soma and germline fates.
607 the *Kdm5c*-KO brain also Previous work identified testis-enriched genes in *Kdm5c* genes
- 608 –
609 – However unlike the gonadal-biased DEGs,
- 610 • The demarcation of the germline versus soma is a key feature of multicellularity and sexually dimorphic
611 reproduction
- 612 • Anything known about tissue-biased gene expression in other H3K4me regulators?

- 613 • Our data suggests the germline developmental program is occurring ectopically as *Kdm5c*-KOs pro-
614 gresses through somatic tissue development
- 615 • tissue-biased gene expression:
- 616 • Altogether, these data indicate that while some germline genes are misexpressed due to direct loss of
617 KDM5C binding during embryogenesis, secondary downstream mechanisms can also promote their
618 aberrant transcription.
- 619 • Papers to read/reference:
- 620 – Reconstitution of the Mouse Germ Cell Specification Pathway in Culture by Pluripotent Stem Cells:
621 [https://www.cell.com/fulltext/S0092-8674\(11\)00771-9](https://www.cell.com/fulltext/S0092-8674(11)00771-9)
- 622 – 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/>

624 **Figure outline:**

625 **Figure 1: Misexpression of tissue-specific genes in the *Kdm5c*-KO brain** * MA-plot and bar graphs of
626 tissue-enriched genes * Example testis-specific genes (NCBI and bigwigs) * An example ovary tissue-specific
627 gene * An example muscle/liver tissue-specific gene (NCBI and bigwigs)

628 **Figure 2: The male *Kdm5c*-KO brain expresses male and female germline-enriched genes** * Gene
629 ontology of testis DEGs in the amygdala and hippocampus - ontologies are germline ontologies * Expression
630 of testis DEGs in germline-depleted testis (this is adult testis data) * scRNASeq of testis - # of testis DEGs that
631 are germline-specific markers * Although far fewer, 5CKO brain also expresses ovary-enriched genes(NCBI
632 and bigwigs of Zar1) * These ovary enriched genes are also germline specific (NCBI/Li tissues are in adult
633 ovary, so it would be best to show they're oocyte-specific in adult ovary. But I don't think there's a published
634 adult female W/Wv dataset. Could try looking at scRNASeq or just do TPM in embryonic W/Wv data since
635 oocytes are developed at this point? Or both?) * Defining what is/isn't a germline gene, and which are
636 male/female biased using embryonic W/Wv data

637 **Figure 3: *Kdm5c*-KO epiblast-like cells express key drivers regulators of germline identity** * A) ESC
638 to EpiLC differentiation Left - Morphology is unchanged, * B) 5CKO EpiLCs express EpiLC differentiation
639 genes similar to WT lvs * C) Male EpiLCs express germline genes (example Cyct again) * Overlap between
640 brain and EpiLC germline genes - show they're mostly unique * GO of Brain and EpiLC germline genes
641 (meiotic enriched) * Bigwigs or TPM of master regulators * Show that while some are also 2-cell genes
642 (Dazl), 2-cell specific genes aren't dysregulated (Zscan4). Important point because published KDM5C dazl
643 paper is saying KDM5C is a 2-cell regulator, but as far as I can tell only genes shared between germline and
644 2-cell are dysregulated.

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

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

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

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

664 Introduction: * Chromatin regulators are important for cellular identity * H3K4me1-3 linked to active
665 gene promoters and enhancers * Surprisingly, mutations in Chromatin regulators lead to many NDDs
666 (including many H3K4 regulators) * Recent studies have shown some chromatin regulators are important
667 for regulating neuron-specific gene expression/chromatin stat_compare_means * However, loss of some
668 chromatin regulators can also lead to ectopic expression of tissue-enriched genes * Very few studies have
669 looked at these genes and it's unclear if these genes contribute to NDD impairments. * Necessary to
670 first characterize the mechanism behind their derepression to identify molecular footholds into testing their
671 contribution to neuronal impairments and potential for therapeutic intervention

- 672 • Loss of KDM5C can result in the misexpression of genes typically only found in the testis
- 673 – Misexpression of tissue-enriched genes hasn't been systematically characterized - Unclear if
674 these genes are exceptions or if other tissue-specific genes are dysregulated
- 675 – Interestingly, these genes (Cyct, D1pas1) typically function in the germline
- 676 – Germ cells (meiotic cells) are typically distinguished from somatic cells very early on in embryogenesis
677 and is a key feature of multicellularity
- 678 – Chromatin regulators are very important for decommissioning germline genes and act successively
679 the embryo implants into the uterine wall
- 680 * Most studies have focused on ESCs, which have a similar transcriptome to germ cells / 2-cells

- 681 * recently, KDM5C was shown to repress DAZL in ESCs, independent of its catalytic activity
682 * However, DNA methylation is lost in the mature 5CKO brain, DNA methylation is placed later
683 and it's Unclear if it's required for long-term repression (maybe too specific, just trying to go
684 into the fact that the mechanism is partially understood but unclear)
685 – Systematic characterization of ectopic germline genes hasn't been done
686 * unknown if other germline-enriched genes are dysregulated, including oocyte-specific genes
687 * Crucially, it's unknown if misexpression of the germline program leads to functional conse-
688 quences in 5CKO cells.

689 **Germline gene repression background:**

690 Interestingly, some of the ectopic testis transcripts identified in the *Kdm5c*-KO brain are typically ex-
691 pressed in germ cells¹⁰. Unlike somatic cells, germ cells (e.g. sperm and eggs) undergo meiosis and pass
692 on their genetic material to the next generation. The germline and the soma are typically distinguished during
693 early embryogenesis, when germline genes are silenced in epiblast stem cells soon after implantation and
694 only reactivated in a subset to form the germline. Chromatin regulators play a key role in decommissioning
695 germline genes as the embryo transitions from naïve to primed pluripotency by placing repressive histone
696 H2A lysine 119 monoubiquitination (H2AK119ub1)¹⁶, histone 3 lysine 9 trimethylation (H3K9me3)^{16,17}, and
697 DNA CpG methylation^{17–19} at germline gene promoters. KDM5C may also be involved in this early decom-
698 missioning of germline genes, as re-expression of KDM5C in neurons fails to suppress their dysregulation¹⁰.
699 In support of this, KDM5C was very recently shown to repress *Deleted in azoospermia like* (*Dazl*), a key
700 regulator of germline development, in mouse embryonic stem cells (ESCs)^{41,44}. In support of this, two
701 independent screens in mouse embryonic stem cells (ESCs) recently identified KDM5C as a repressor of
702 *Deleted in azoospermia like* (*Dazl*), a key regulator of germline development. However, KDM5C's role in
703 embryonic germline gene repression is currently unclear, given that *Dazl* is also expressed in ESCs and in
704 the 2-cell stage and germline gene misexpression has yet to be globally characterized during *Kdm5c*-KO
705 embryogenesis.