

1 Erosion of somatic tissue identity with loss of the X-linked

2 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). Male *Kdm5c* knockout (-KO) mice recapitulate key phenotypes of Claes-Jensen X-linked intellectual disability (MRXSCJ, OMIM: 300534), including heightened aggression and learning impairments. We found male *Kdm5c* knockout (-KO) mice that recapitulate key phenotypes of Claes-Jensen X-linked intellectual disability (MRXSCJ, OMIM: 300534) aberrantly express liver, muscle, ovary, and testis genes within the amygdala and hippocampus. The majority of tissue-enriched genes were typically unique to germ cells, indicating the erosion of the soma-germline boundary is a major effect of KDM5C loss. Using RNA sequencing datasets of germ cell-depleted mice, we curated a list of germline-enriched genes and found KDM5C regulates many key drivers of meiosis and germline identity in an *in vitro* model of the post-implantation epiblast. While KDM5C promotes the initial placement of DNA methylation at a subset of germline gene promoters during early embryogenesis, germline genes can also be aberrantly expressed in *Kdm5c*-KO cells indirectly of KDM5C binding. Intriguingly, many of the germline genes expressed in the *Kdm5c*-KO brain are late-stage spermatogenesis genes that are not directly bound by KDM5C. This suggests germline developmental programs can occur ectopically in the background of typical *Kdm5c*-KO development. Ultimately, this work provides insight into the demarcation of somatic and germline identity in mammals while also providing the foundational mechanisms by which to test the impact of germline gene dysregulation in neurodevelopmental disorders.

29 Introduction

30 A single genome holds the instructions to generate the myriad of cell types found within the adult organism.
31 This is, in part, accomplished by chromatin regulators that can either promote or impede lineage-specific
32 gene expression through DNA and histone modifications^{1,2}. Although many chromatin regulators were initially
33 identified for their roles in shaping cellular and tissue identity^{3,4}, recent advancements in next generation
34 sequencing unexpectedly revealed many neurodevelopmental disorders (NDDs) are caused by mutations in
35 chromatin regulators⁵. Several studies have suggested this connection between chromatin regulators and
36 neurodevelopment is due to their regulation of brain-specific genes, such as orchestrating transcriptional
37 programs for synaptic maturation⁶ and transitioning between neuronal and glial fates during neural precursor
38 differentiation⁷. However, loss of some chromatin regulators can also result in the ectopic transcription of
39 tissue-specific genes outside of their target environment^{3,4,8}, such as the misexpression of liver-specific
40 genes within adult neurons⁹. Very few studies have investigated the misexpression of tissue-specific genes
41 in chromatin-linked NDDs^{9,10} and it is currently unknown if this partial loss of brain identity contributes to
42 neurodevelopmental impairments.

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

- 56 • one key process chrom reg are involved in are is distinguishing bw germ + soma cell identity

57 Distinguishing between the germline (cells that pass on their genetic material, e.g. sperm and eggs) and
58 the soma (cells that perform all other bodily functions) is a key feature of multicellularity and occurs during
59 early embryogenesis. In mammals, chromatin regulators play a key role in decommissioning germline genes
60 in somatic cells during the transition from naïve to primed pluripotency by placing repressive histone H2A
61 lysine 119 monoubiquitination (H2AK119ub1)¹⁶, histone 3 lysine 9 trimethylation (H3K9me3)^{16,17}, and DNA
62 CpG methylation^{17–19} at germline gene promoters. Systematically characterizing KDM5C's role in germline
63 gene repression during early embryogenesis will unveil key mechanisms underlying the demarcation between

64 soma and germline identity and while also providing molecular footholds to test the impact of ectopic germline
65 genes on neurodevelopment.

66 To illuminate KDM5C's role in tissue identity, here we characterized the aberrant transcription of tissue-
67 enriched genes within the *Kdm5c*-KO brain and epiblast-like stem cells (EpiLCs), an *in vitro* model of
68 the post-implantation embryo. We observed general dysregulation of tissue-enriched genes in both the
69 *Kdm5c*-KO brain and EpiLCs, including misexpression of genes typically highly enriched within the testis,
70 liver, muscle, and ovary. Both the *Kdm5c*-KO amygdala and hippocampus had significant enrichment of testis
71 genes that are typically uniquely expressed in germ cells. While the *Kdm5c*-KO brain primarily expressed
72 germline genes important for late spermatogenesis, *Kdm5c*-KO EpiLCs aberrantly expressed key drivers
73 of germline identity and meiosis, including *Dazl* and *Stra8*. KDM5C was highly enriched at germline gene
74 promoters in EpiLCs but only bound to a subset of germline genes expressed in *Kdm5c*-KO cells, indicating
75 germline-enriched mRNAs can be aberrantly transcribed through indirect mechanisms. Finally, we found
76 KDM5C promotes the long-term silencing of germline genes in somatic cells by aiding the placement of
77 DNA methylation in EpiLCs through H3K4me2/3 removal. Thus, we propose KDM5C plays a fundamental
78 role in the development of tissue identity during early embryogenesis, including the establishment of the
79 soma-germline boundary.

80 Results

81 **Tissue-enriched genes, including testis genes, are aberrantly expressed in the** 82 ***Kdm5c*-KO brain**

- 83 • note: should I compare amygdala and hippocampus? scandaglia only looked at hippocampus

84 Previous RNA sequencing (RNA-seq) in the adult hippocampus revealed ectopic expression of some testis
85 genes within the *Kdm5c* knockout (-KO) brain¹⁰. Since tissue-enriched genes have not been systematically
86 characterized in the *Kdm5c*-KO brain, it is currently unclear if this erosion of brain tissue identity is a major
87 consequence of *Kdm5c* loss and if it is unique to testis-enriched genes. Therefore, we first globally assessed
88 the expression of genes enriched in 17 mouse tissues²⁰ in our published mRNA-seq datasets of the adult
89 amygdala and hippocampus in mice with constitutive knockout of *Kdm5c*²¹.

90 We found a large proportion of genes that are significantly upregulated within the *Kdm5c*-KO brain
91 (DESeq2²², log2 fold change > 0.5, q < 0.1) are typically enriched in non-brain tissues (Amygdala: 35%,
92 Hippocampus: 24%) (Figure 1A-B). The majority of tissue-enriched differentially expressed genes (tissue-
93 enriched DEGs) were testis genes (Figure 1A-C). Even though the testis has the largest total number of
94 tissue-biased genes compared to any other tissue (2,496 genes), testis-biased DEGs were significantly
95 enriched for both brain regions (Amygdala p = 1.83e-05, Odds Ratio = 5.13; Hippocampus p = 4.26e-11,
96 Odds Ratio = 4.45, Fisher's Exact Test). One example of a testis-enriched DEG is *FK506 binding protein 6*

97 (*Fkbp6*), a known regulator of piRNA expression and meiosis in germ cells^{23,24} (Figure 1C).

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

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

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

118 The testis contains both germ cells (meiotic cells, e.g. spermatogonia) and somatic cells (non-meiotic,
119 e.g. Leydig cells) that support hormone production and germline functions. Intriguingly, many *Kdm5c*-
120 KO testis and ovary enriched-DEGs have germline-specific functions, suggesting *Kdm5c*-KO cells fail to
121 distinguish between the soma and germline. To test if this holds true for all *Kdm5c*-KO testis-biased DEGs, we
122 first assessed their known functions through gene ontology analysis. We found *Kdm5c*-KO testis-enriched
123 DEGs high enrichment of germline-relevant ontologies, including spermatid development (GO: 0007286,
124 p.adjust = 6.2e-12) and sperm axoneme assembly (GO: 0007288, p.adjust = 2.45e-14) (Figure 2A).

125 To further validate if these testis DEGs are truly germline genes, we then evaluated their expression in
126 somatic versus germ cells within the testis. We first compared their expression in the testis with germ cell
127 depletion²⁶, which was accomplished by heterozygous *W* and *Wv* mutations in the enzymatic domain of
128 *c-Kit* (*Kit*^{W/Wv}) that prevent the maturation of germ cells²⁷. Almost all *Kdm5c*-KO testis-enriched DEGs lost
129 expression with germ cell depletion (Figure 2B). The only testis-enriched DEG that did not show considerable
130 downregulation with germline depletion was *FK506 binding protein 6 (Fkbp6)*, the aforementioned testis

131 gene that regulates piRNA expression and meiosis in germ cells^{23,24}. We then assessed testis-enriched
132 DEG expression in a published single cell RNA-seq dataset that identified cell type-specific markers within
133 the testis²⁸. We found that while some testis-enriched DEGs were classified as specific markers for different
134 germ cell types (e.g. spermatogonia, spermatocytes, round spermatids, and elongating spermatids) none
135 marked somatic cells (Figure 2C). Together, these data suggest the somatic *Kdm5c*-KO brain aberrantly
136 expresses germline genes.

137 We then wanted to more deeply characterize germline gene misexpression with *Kdm5c* loss, but lacked
138 a comprehensive list of mouse germline-enriched genes. To facilitate downstream analyses, we generated a
139 curated list of germline-enriched genes using currently available RNA-seq datasets in *Kit*^{W/Wv} mice. Wild-type
140 and *Kit*^{W/Wv} datasets included males and females at embryonic day 12, 14, and 16²⁹, as well as adult
141 male testes²⁶. We defined genes as germline-enriched if their expression met the following criteria: 1)
142 their expression is greater than 1 FPKM in wild-type gonads 2) their expression in any adult wild-type,
143 non-gonadal tissue²⁰ does not exceed 20% of their maximum expression in the wild-type germline, and
144 3) their expression in the germ cell-depleted germline, for any sex or time point, does not exceed 20% of
145 their maximum expression in the wild-type germline. These criteria yielded 1,288 germline-enriched genes
146 (Figure 2D), which was hereafter used as a resource for assessing germline gene misexpression with *Kdm5c*
147 loss (Supplementary table 1).

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

150 Misexpression of germline genes in the adult brain suggests *Kdm5c*-KO cells fail to demarcate between
151 germline and somatic identity. Germ cells are typically distinguished from somatic cells soon after the embryo
152 implants into the uterine wall^{30,31} when a subset of epiblast stem cells become the primordial germ cells
153 (PGCs) while the remainder differentiate into the ectoderm, mesoderm, and endoderm to form the somatic
154 tissues³². This developmental time point can be modeled *in vitro* through differentiation of embryonic stem
155 cells (ESCs) into post-implantation epiblast-like stem cells (EpiLCs) (Figure 3A, top). Previous studies have
156 demonstrated that while some germline-enriched genes are also expressed in embryonic stem cells (ESCs)
157 and in the 2-cell stage^{33–35}, they are silenced as they differentiate into EpiLCs¹⁷. Therefore, we assessed if
158 KDM5C was necessary for embryonic silencing of germline genes by evaluating the impact of *Kdm5c* loss in
159 EpiLCs.

160 We first identified *Kdm5c*-KO EpiLC DEGs in of our previously published RNA-seq dataset³⁶ (DESeq2,
161 log2 fold change > 0.5, q < 0.1) and assessed tissue-enriched gene expression. Similar to the *Kdm5c*-KO
162 brain, we observed general dysregulation of tissue-enriched genes, with the largest number of genes
163 belonging to the brain and testis, although they were not significantly enriched (Figure 3B). Using the curated
164 list of germline genes generated above, we found *Kdm5c*-KO EpiLCs aberrantly expressed 54 germline-

165 enriched genes, including the previously characterized hippocampal DEG¹⁰ *Cytochrome C, testis-specific*
166 (*Cyct*) (Figure 3C). Although we observed aberrant expression of many tissue-enriched genes, we did not
167 observe any significant difference in primed pluripotency genes (Figure 3D) or gross changes in *Kdm5c*-KO
168 cell morphology during differentiation (Figure 3E), indicating KDM5C loss does not impair EpiLC formation.

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

178 Interestingly, DEGs unique to *Kdm5c*-KO EpiLCs included key drivers of germline identity, such as
179 *Stimulated by retinoic acid 8 (Stra8)* and *Deleted in azoospermia like (Dazl)* (Figure 3H). These genes are
180 typically expressed during embryonic germ cell development to commit PGCs to the germline fate, but are
181 also expressed later in life to trigger meiotic gene expression programs³⁷⁻³⁹. Of note, some germline genes,
182 including *Dazl*, are also expressed in the two-cell embryo^{34,40}. However, we did not see misexpression of
183 two-cell embryo-specific genes, like *Duxf3 (Dux)* (q = 0.337) and *Zscan4d* (q = 0.381), indicating *Kdm5c*-KO
184 in EpiLCs do not revert back to a 2-cell state (Figure 3H).

185 We were particularly interested in the aberrant transcription of *Dazl*, since it is essential for germ cell
186 development and promotes the translation of germline mRNAs⁴¹. A significant portion of germline transcripts
187 misexpressed in *Kdm5c*-KO EpiLCs are known binding targets of DAZL, including *Stra8*⁴² (p = 1.698e-07,
188 Fisher's Exact Test). This suggests expression of DAZL protein could promote the translation of other
189 aberrant germline transcripts, influencing their ability to impact *Kdm5c*-KO cellular function. We thus tested
190 DAZL protein expression in *Kdm5c*-KO EpiLCs through immunocytochemistry (Figure 3I). We observed
191 about 25% of *Kdm5c*-KO EpiLCs expressed DAZL protein and it was localized to the cytoplasm (p = 0.0015,
192 Welch's t-test), consistent with the pattern of DAZL expression in spermatogonia⁴². Altogether these results
193 suggest tissue-specific genes are misexpressed during *Kdm5c*-KO embryogenesis, including key drivers of
194 germline identity that can be translated into protein.

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

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

198 Previous work suggests KDM5C represses germline genes during early development, as re-expression

199 of KDM5C in knockout neuronal cultures fails to suppress hippocampal germline DEGs and KDM5C is not
200 bound to their promoters in neurons¹⁰. There is some evidence KDM5C binds to select germline gene
201 promoters in ESCs¹⁰, including *Dazl*^{40,43}. As KDM5C's binding at germline gene promoters has not been
202 systematically characterized, it is currently unclear if KDM5C is enriched at germline gene promoters, what
203 types of germline genes KDM5C regulates, and if its binding is maintained at any germline genes in neurons.

204 To further characterize the mechanism behind *Kdm5c*-KO germline gene misexpression, we analyzed
205 KDM5C chromatin immunoprecipitation followed by DNA sequencing (ChIP-seq) datasets in EpiLCs³⁶ and
206 primary neuron cultures (PNCs) from the cortex and hippocampus¹⁵. EpiLCs had a higher total number of
207 KDM5C peaks than PNCs (EpiLCs: 5,808, PNCs: 1,276, MACS2 q < 0.1 and fold enrichment > 1, removal
208 of *Kdm5c*-KO false positive peaks) and KDM5C was primarily localized to gene promoters in both cell types
209 (EpiLCs: 4,190, PNCs: 745 +/- 500bp from the TSS), although PNCs showed increased localization to
210 non-promoter regions (Figure 4A).

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

224 We then compared KDM5C binding at RNA-seq DEG promoters to determine if the germline mRNAs
225 expressed in the *Kdm5c*-KO brain and EpiLCs are direct targets of KDM5C (Figure 4E). About one third
226 of EpiLC-specific and brain-specific (hippocampus, amygdala, or both) germline DEGs were bound by
227 KDM5C in EpiLCs (EpiLC only: 36%, Brain only: 33.3%). Some notable differences in KDM5C binding
228 for EpiLC-specific DEGs included *Dazl* and *Stra8*, two key drivers of germline identity discussed above.
229 Although both mRNAs are expressed in *Kdm5c*-KO EpiLCs, KDM5C is only bound to *Dazl*'s promoter and
230 not *Stra8*'s in EpiLCs (Figure 4F). In contrast to the unique DEGs, 3 out of the 4 genes dysregulated in both
231 the brain and EpiLCs are bound by KDM5C around the TSS (Figure 4G). Again, we did not observe any
232 KDM5C binding at germline gene promoters in PNCs, even for brain-specific DEGs (Figure 4H). Altogether,
233 this suggests the majority of germline mRNAs expressed in *Kdm5c*-KO cells are dysregulated independent
234 of direct KDM5C recruitment to their promoters during embryogenesis.

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

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

238 Although KDM5C is generally thought to suppress transcription through erasure of histone 3 lysine 4 di-
239 and trimethylation (H3K4me2/3)¹¹, recent studies in ESCs have suggested KDM5C's repression *Dazl* is
240 independent of its catalytic activity⁴⁰. Somatic repression of germline genes is typically established during the
241 transition between naïve and primed pluripotency, which modeled *in vitro* as ESCs to EpiLC differentiation.
242 In ESCs, chromatin regulators place repressive histone modifications at germline gene promoters, including
243 histone 2A 119 monoubiquitination (H2AK119ub1) and histone 3 lysine 9 trimethylation (H3K9me3)^{16,17,44}.
244 Germline genes are then silenced long-term in EpiLCs by *de novo* placement of DNA CpG methylation
245 (DNAme)¹⁷. Although KDM5C's catalytic activity is not required to suppress *Dazl* in ESCs, KDM5C may
246 promote long-term germline gene silencing via H3K4me3 removal since H3K4me3 can impede DNAme
247 placement^{45,46} and DNAme is lost at select germline gene promoters in the hippocampus¹⁰. Because
248 KDM5C's role in germline gene repression has only been characterized in ESCs and in the mature brain, it is
249 currently unclear to what extent KDM5C is involved during transition between ESCs and EpiLCs and if its
250 catalytic activity is required for long-term silencing.

251 To elucidate KDM5C's role in germline gene silencing, we first characterized KDM5C's substrates, histone
252 3 lysine 4 di- and trimethylation (H3K4me2/3) in our previously published ChIP-seq datasets in wild type and
253 *Kdm5c*-KO amygdala²¹ and EpiLCs³⁶. In congruence with previous work in the *Kdm5c*-KO hippocampus¹⁰,
254 we observed aberrant accumulation of H3K4me3 around the transcription start site (TSS) of germline genes
255 in the *Kdm5c*-KO amygdala (Figure 5A). We additionally found a marked increase in H3K4me2 around the
256 TSS of germline genes in *Kdm5c*-KO EpiLCs (Figure 5B).

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

- 266 • Catalytic activity
267 • Altogether, this suggests KDM5C is dynamically regulated during ESC to EpiLC differentiation and
268 promotes germline gene silencing via removal of H3K4me2/3 during early embryogenesis.

- 269 • **note:** should I test if increase in H3K4me2 and H3K4me3 was highest in *Kdm5c*-KO cells at genes
270 bound by KDM5C at their promoter in EpiLCs? Don't think it's that impactful

271 **Discussion**

272 In the above study, we demonstrate KDM5C's pivotal role in the development of tissue identity, resulting in
273 substantial misexpression of tissue-enriched genes within the adult *Kdm5c*-KO hippocampus and amygdala.
274 We defined tissue-enriched genes based on the criteria used in a previously published dataset of 17 C57/Bl6J
275 mouse tissues²⁰, which defined genes as tissue-enriched if they had more than 4-fold higher expression
276 when compared to any other tissue. In addition to testis genes identified previously¹⁰, we found significant
277 enrichment of muscle, liver, and even ovary-biased genes aberrantly expressed within the male *Kdm5c*-KO
278 brain. It is currently unclear what the impact of these tissue-enriched genes are upon *Kdm5c*-KO brain
279 function and if they ultimately contribute to MRXSCJ-related phenotypes. While the majority of tissue-
280 enriched DEGs were testis and ovary genes with no known brain functions, select liver and muscle-biased
281 DEGs do have known roles in the wild-type brain. For example, *Apolipoprotein C-I (Apoc1)* is a lipid transport
282 gene highly enriched within the liver, but is also lowly expressed in the wild-type brain. Intriguingly, *Apoc1*
283 overexpression in the mouse brain can impair learning and memory⁴⁷ and is implicated in Alzheimer's
284 disease in humans⁴⁸. Our results suggest KDM5C fine-tunes the expression of tissue-enriched genes like
285 *Apoc1* to match the level required for proper brain function. However, further studies are required to determine
286 if these genes contribute to *Kdm5c*-KO neuronal impairments and MRXSCJ pathology.

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

298 Dysregulation of *Kdm5c*-KO tissue identity begins during early embryogenesis, as germline and other
299 tissue-enriched genes are aberrantly expressed in epiblast-like cells (EpiLCs). *In vivo*, germline genes
300 are typically decommissioned in epiblast stem cells and remain silenced as the epiblast differentiates
301 into the body's somatic tissues³². However, a small subset of epiblast stem cells will receive signals to
302 reactivate germline gene expression to become the primordial germ cells (PGCs) that will ultimately form

303 the mature germline^{30,31}. This process can be mimicked *in vitro* by differentiating EpiLCs into primordial
304 germ cell-like cells (PGCLCs)⁴⁹. Thus, misexpression of germline genes in *Kdm5c*-KO EpiLCs might
305 suggest they are progressing beyond EpiLC differentiation and becoming PGCLCs. Yet, *Kdm5c*-KO EpiLCs
306 had proper expression of EpiLC marker genes for primed pluripotency and we observed no difference in
307 cellular morphology during ESC to EpiLC differentiation. Furthermore, we saw no significant change in *Otx2*
308 expression, an epiblast stem cell marker that is known to repress epiblast and EpiLC differentiation into PGC
309 and PGCLCs⁵⁰. This, in conjunction with *Kdm5c*-KO mice being viable, suggests germline gene expression
310 is occurring ectopically in conjunction with typical developmental programs, rather than a complete shift to
311 germline identity.

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

314 We then globally characterized KDM5C binding at germline-enriched gene promoters through analysis
315 of KDM5C ChIP-seq in EpiLCs and primary hippocampus and cortex neuron cultures (PNCs). While we
316 observed high enrichment of KDM5C peaks at germline genes in EpiLCs, KDM5C was not bound to germline
317 gene promoters in PNCs. This suggests dysregulation of germline genes in the mature *Kdm5c*-KO brain is
318 due to loss of repression during embryogenesis, which is consistent with previous work that found introducing
319 human KDM5C into *Kdm5c*-KO PNCs does not repress germline transcripts¹⁰. Although KDM5C is enriched
320 at germline gene promoters in EpiLCs, KDM5C was only bound to about a third of EpiLC germline-enriched
321 DEGs. One notable DEG not bound by KDM5C is *Stra8*, a transcription factor activated by retinoic acid
322 signaling in germ cells that promotes meiotic initiation^{51,52}. Retinoic acid can only activate *Stra8* expression
323 when DAZL is present, which is a direct target of KDM5C and is aberrantly transcribed and translated in
324 *Kdm5c*-KO EpiLCs. This indicates germline genes can be aberrantly expressed in *Kdm5c*-KO cells indirectly
325 of KDM5C regulation through activation by other ectopic germline programs. Consistent with this idea, many
326 *Kdm5c*-KO EpiLC germline DEGs are important for early germ cell development and meiosis while those
327 expressed in the mature *Kdm5c*-KO brain are involved in late sperm development. Altogether, this indicates
328 ectopic germline programs are, to some extent, progressing through germ cell developmental stages over
329 the course of *Kdm5c*-KO development.

- 330 • Talk about different motifs if we do see differences there and if it explains direct vs indirect dysregulation
331 – Could also mention PRC1.6/KDM5C binds pcgf6 in ESCs
332 • Maybe talk about other regulators/if anything is known about them long term? Or save for 2nd paper.
333 – another germline repressor: <https://pubmed.ncbi.nlm.nih.gov/35345626/>
334 • 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>

336 It is important to note that while we focused on KDM5C's regulation of germline genes, some germline-
337 enriched genes are also expressed in naïve ESCs and at the 2-cell stage due to their roles in pluripotency
338 and self-renewal. For example, although primarily known for committing PGCs to the germline fate and
339 regulating the translation of germline-specific RNAs, *Dazl* is also expressed in naïve ESCs *in vitro*³³, the
340 inner cell mass *in vivo*³³, and at the 2-cell stage⁵³, but is silenced when cells transition from naïve to primed
341 pluripotency during ESC to EpiLC differentiation¹⁷. Very recently, two screens of *Dazl*-repressors in ESCs
342 identified KDM5C as a direct regulator of *Dazl* expression^{40,43}. Interestingly, one screen found *Kdm5c*-KO
343 ESCs also aberrantly expressed 2-cell-specific genes, indicating KDM5C negatively regulates totipotency in
344 ESCs⁴⁰. We found that while KDM5C also represses *Dazl* expression in EpiLCs, *Kdm5c*-KO EpiLCs do not
345 express 2-cell specific genes like *Dux* and *Zscan4c*. Out of the four 2-cell regulators characterized in ESCs,
346 KDM5C was the only factor whose repression of *Dazl* was independent of *Dux* expression⁴⁰. Together, this
347 suggests the 2-cell-like state in *Kdm5c*-KO ESCs is an indirect consequence of KDM5C's primary role in
348 germline gene repression, including germline genes that are also involved in pluripotency and self-renewal.

349 It is hypothesized distinct repressive chromatin marks are placed at germline gene promoters as the
350 embryo transitions from naïve to primed pluripotency. Initially, germline genes are repressed by placement of
351 histone 2A lysine 119 monoubiquitination (H2AK119ub1) and histone 3 lysine 9 trimethylation (H3K9me3)
352 in ESCs and then gain *de novo* DNA CpG methylation (CpGme) in EpiLCs^{17–19,44,54,55}. KDM5C may be
353 instead required to remove an active mark, histone 3 lysine 4 trimethylation (H3K4me3), since H3K4me3
354 can impede *de novo* CpGme placement^{45,46}. This is supported by previous work in the *Kdm5c*-KO adult
355 hippocampus, which found an increase in H3K4me3 around the transcription start site (TSS) of germline
356 DEGs and loss of CpGme at at least two germline promoters¹⁰. However, KDM5C's role in embryonic
357 germline gene silencing is still unclear, as KDM5C's catalytic activity was recently found to be dispensable
358 for silencing DAZL in ESCs⁴⁰. In this study, we observed a global increase in H3K4me3 around the TSS of
359 germline-enriched gene in the *Kdm5c*-KO amygdala and an increase in H3K4me2 in *Kdm5c*-KO EpiLCs. We
360 found KDM5C's expression is dynamically regulated during ESC to EpiLC differentiation and loss of KDM5C
361 leads to impaired placement of CpGme in extended EpiLCs. Altogether, this suggests KDM5C is necessary
362 during the transition from naïve to primed pluripotency to promote the initial placement of CpGme at CpG
363 islands for the long-term silencing of germline genes.

364 In summary, we demonstrate loss of the X-linked intellectual disability gene KDM5C leads to widespread
365 dysregulation of brain tissue identity, including misexpression of germline genes in the somatic brain. In
366 EpiLCs, KDM5C directly represses key drivers of germline identity like *Dazl*, by promoting initial CpG methy-
367 lation placement in the post-implantation embryo. However, germline genes can also become ectopically
368 expressed in *Kdm5c*-KO cells independent of direct KDM5C regulation, including the meiotic transcription
369 factor *Stra8*. These ectopic germline developmental programs can, to some extent, mimic typical germ
370 cell development, resulting in aberrant transcription early developmental and meiotic genes in *Kdm5c*-KO
371 EpiLCs and late-stage spermatogenesis genes in the *Kdm5c*-KO brain. Altogether, these results define

372 KDM5C's role in the demarcation between soma and germline identity and offers novel insight into how this
373 dysregulation of tissue identity changes over the course of development. Additionally, this study provides
374 the mechanistic foundation required to ultimately investigate the impact of aberrant germline identity upon
375 neurodevelopment.

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

378 Materials and Methods

379 Classifying tissue-enriched and germline-enriched genes

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

384 We curated a list of germline-enriched genes using an RNA-seq dataset from wild-type and germline-
385 depleted (Kit^{W/Wv}) male and female mouse embryos from embryonic day 12, 14, and 16²⁹, as well as adult
386 male testes²⁶. Germline-enriched genes met the following criteria: 1) their expression is greater than 1
387 FPKM in wild-type germline 2) their expression in any wild-type somatic tissues²⁰ does not exceed 20%
388 of maximum expression in wild-type germline, and 3) their expression in the germ cell-depleted (Kit^{W/Wv})
389 germline, for any sex or time point, does not exceed 20% of maximum expression in wild-type germline.

390 Cell culture

391 We utilized our previously established cultures of male wild-type and *Kdm5c* knockout (-KO) embryonic
392 stem cells³⁶. Sex was confirmed by genotyping *Uba1* on the X and Y chromosomes with the following
393 primers: 5'-TGGATGGTGTGCCAATG-3', 5'-CACCTGCACGTTGCCCT-3'. Deletion of *Kdm5c* was
394 confirmed through the primers 5'-ATGCCCATATTAAGAGTCCTG-3', 5'-TCTGCCCTGATGGACTGTT-3',
395 and 5'-GGTTCTCAACACTCACATAGTG-3'.

396 Embryonic stem cells (ESCs) were initially cultured in primed ESC (pESC) media consisting of Knock-
397 Out DMEM (Gibco#10829-018), fetal bovine serum (Gibco#A5209501), KnockOut serum replacement
398 (Invitrogen#10828-028), Glutamax (Gibco#35050-061), Anti-Anti (Gibco#15240-062), MEM Non-essential
399 amino acids (Gibco#11140-050), and beta-mercaptoethanol (Sigma#M7522). They were then transitioned
400 into ground-state "naive" ESCs (nESCs) by culturing for four passages in N2B27 media containing DMEM/F12
401 (Gibco#11330-032), Neurobasal media (Gibco#21103-049), Gluamax, Anti-Anti, N2 supplement (Invitro-
402 gen#17502048), and B27 supplement without vitamin A (Invitrogen#12587-010), and beta-mercaptoethanol.

403 Both pESC and nESC media were supplemented with the GSK3 inhibitor CHIR99021 (Sigma #SML1046-
404 5MG), the MEK inhibitor PD0325901 (Sigma #PZ0162-5MG), and leukemia inhibitory factor (LIF, Milli-
405 pore#ESG1107).

406 nESCs were differentiated into epiblast-like cells (EpiLCs, 48 hours) and extendend EpiLCs (exEpiLCs,
407 96 hours) by culturing in N2B27 media containing DMEM/F12, Neurobasal media, Gluamax, Anti-Anti,
408 N2 supplement, B27 supplement (Invitrogen#17504044), beta-mercaptoethanol, fibroblast growth factor 2
409 (FGF2, R&D Biotechne 233-FB), and activin A (R&D Biotechne 338AC050CF).

410 **Immunocytochemistry (ICC)**

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

418 **RNA sequencing**

419 **Published datasets**

420 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
421 and *Kdm5c*-KO adult amygdala and hippocampus²¹ (available at GEO: GSE127722) and male wild-type and
422 *Kdm5c*-KO EpiLCs³⁶ (available at GSE: GSE96797).

424 **Alignment and analysis**

425 After ensuring read quality via FastQC (v0.11.8), reads were then mapped to the mm10 *Mus musculus*
426 genome (Gencode) using STAR (v2.5.3a), during which we removed duplicates and kept only uniquely
427 mapped reads. Count files were generated by FeatureCounts (Subread v1.5.0), and BAM files were
428 converted to bigwigs using deeptools (v3.1.3) and visualized by the UCSC genome browser. RStudio (v3.6.0)
429 was then used to analyze counts files by DESeq2 (v1.26.0)²² to identify differentially expressed genes
430 (DEGs) with a q-value (p-adjusted via FDR/Benjamini–Hochberg correction) less than 0.1 and a log2 fold
431 change greater than 0.5. For all DESeq2 analyses, log2 fold changes were calculated with IfcShrink using
432 the ashR package⁵⁶. MA-plots were generated by ggpubr (v0.4.0), and Eulerr diagrams were generated by
433 eulerr (v6.1.1). Boxplots and scatterplots were generated by ggpubr (v0.6.0) and ggplot2 (v3.3.2). Heatmaps

434 of gene expression were generated using the base R functions scale and hclust and visualized using the R
435 package ComplexHeatmap (v2.12.1). The Upset plot was generated via the package UpSetR (v1.4.0)⁵⁷.
436 Gene ontology (GO) analyses were performed by the R package enrichPlot (v1.16.2) using the biological
437 processes setting.

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

439 We analyzed our previously published KDM5C ChIP-seq experiments in primary neuron cultures from
440 the cortex and hippocampus¹⁵ (available at GEO: GSE61036) and EpiLCs³⁶ (available at GEO: GSE96797).
441 ChIP-seq of histone 3 lysine 4 dimethylation in male EpiLCs³⁶ is also available at GEO: GSE96797. ChIP-seq
442 of histone 3 lysine 4 trimethylation in the male amygdala²¹ are available at GEO: GSE127817.

443 ChIP-seq reads were aligned to mm10 using Bowtie1 (v1.1.2) allowing up to two mismatches. Only
444 uniquely mapped reads were used for analysis. Peaks were called using MACS2 software (v2.1.0.20140616)
445 using input BAM files for normalization, with filters for a q-value < 0.1 and a fold enrichment > 1. Common
446 peak sets were obtained in R via DiffBind (v3.6.5), and count tables for the common peaks were generated
447 with the Bedtools multicov command. We removed “black-listed” genomic regions that often give aberrant
448 signals. Peak proximity to genome annotations was determined by ChIPSeeker (v1.32.1). Enriched motifs
449 were identified using HOMER⁵⁸. Gene ontology (GO) analyses were performed by the R package enrichPlot
450 (v1.16.2) using the biological processes setting. Average binding across the genome was visualized using
451 deeptools (v3.1.3). Bigwigs were visualized using the UCSC genome browser.

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

453 **Data availability**

454 **Acknowledgements**

455 **References**

- 456 1. Strahl, B.D., and Allis, C.D. (2000). The language of covalent histone modifications. *Nature* **403**,
457 41–45. <https://doi.org/10.1038/47412>.
- 458 2. Jenuwein, T., and Allis, C.D. (2001). Translating the Histone Code. *Science* **293**, 1074–1080.
459 <https://doi.org/10.1126/science.1063127>.
- 460 3. Lewis, E.B. (1978). A gene complex controlling segmentation in Drosophila. *Nature* **276**, 565–570.
461 <https://doi.org/10.1038/276565a0>.

- 462 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.85.21.8136>.
- 463 5. 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 Neuropsychopharmacol Biol Psychiatry 84, 306–327. <https://doi.org/10.1016/j.pnpbp.2017.12.013>.
- 464 6. 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/10.1016/j.neuron.2006.09.037>.
- 465 7. 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 Astrogenic Fate Transition. Neuron 63, 600–613. <https://doi.org/10.1016/j.neuron.2009.08.021>.
- 466 8. Kassis, J.A., Kennison, J.A., and Tamkun, J.W. (2017). Polycomb and Trithorax Group Genes in *Drosophila*. Genetics 206, 1699–1725. <https://doi.org/10.1534/genetics.115.185116>.
- 467 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 suppressor complex. Neuron 64, 678–691. <https://doi.org/10.1016/j.neuron.2009.11.019>.
- 468 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, 47–59. <https://doi.org/10.1016/j.celrep.2017.09.014>.
- 469 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>.
- 470 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 Med Genet 94, 1–4.
- 471 13. Jensen, L.R., Amende, M., Gurok, U., Moser, B., Gimmel, V., Tzschach, 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 Genet 76, 227–236. <https://doi.org/10.1086/427563>.

- 482 14. Carmignac, V., Nambot, S., Lehalle, D., Callier, P., Moortgat, S., Benoit, V., Ghoumid, J., Delobel,
483 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.
<https://doi.org/10.1111/cge.13755>.
- 484 15. Iwase, S., Brookes, E., Agarwal, S., Badeaux, A.I., Ito, H., Vallianatos, C.N., Tomassy, G.S., Kasza, T.,
485 Lin, G., Thompson, A., et al. (2016). A Mouse Model of X-linked Intellectual Disability Associated with
Impaired Removal of Histone Methylation. *Cell Reports* **14**, 1000–1009. <https://doi.org/10.1016/j.celrep.2015.12.091>.
- 486 16. Endoh, M., Endo, T.A., Shinga, J., Hayashi, K., Farcas, A., Ma, K.W., Ito, S., Sharif, J., Endoh, T.,
487 Onaga, N., et al. (2017). PCGF6-PRC1 suppresses premature differentiation of mouse embryonic
488 stem cells by regulating germ cell-related genes. *Elife* **6**. <https://doi.org/10.7554/elife.21064>.
- 489 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>.
- 490 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
mediates germ-line gene silencing in murine somatic tissues. *Proc Natl Acad Sci U S A* **107**, 9281–
491 9286. <https://doi.org/10.1073/pnas.1000473107>.
- 492 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
493 genome-defence mechanisms to epigenetic reprogramming in the mouse germline. *Development*
139, 3623–3632. <https://doi.org/10.1242/dev.081661>.
- 494 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.
495 <https://doi.org/10.1038/s41598-017-04520-z>.
- 496 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>.
- 498 22. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for
499 RNA-seq data with DESeq2. *Genome Biol* **15**, 550. <https://doi.org/10.1186/s13059-014-0550-8>.
- 500 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>.

- 501
- 502 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>.
- 503
- 504 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>.
- 505
- 506 26. Mueller, J.L., Skaletsky, H., Brown, L.G., Zaghloul, 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>.
- 507
- 508 27. 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>.
- 509
- 510 28. 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>.
- 511
- 512 29. 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>.
- 513
- 514 30. 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>.
- 515
- 516 31. 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>.
- 517
- 518 32. 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>.
- 519
- 520 33. 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>.
- 521
- 522 34. 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>.

- 523
- 524 35. 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>.
- 525
- 526 36. 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>.
- 527
- 528 37. 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>.
- 529
- 530 38. 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>.
- 531
- 532 39. Endo, T., Mikedis, M.M., Nicholls, P.K., Page, D.C., and De Rooij, D.G. (2019). Retinoic Acid and Germ Cell Development in the Ovary and Testis. *Biomolecules* 9, 775. <https://doi.org/10.3390/biom9120775>.
- 533
- 534 40. Gupta, N., Yakhou, L., Albert, J.R., Azogui, A., Ferry, L., Kirsh, O., Miura, F., Battault, S., Yamaguchi, 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>.
- 535
- 536 41. 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>.
- 537
- 538 42. 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>.
- 539
- 540 43. 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 44. 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 promoters. *J Biol Chem* 295, 9712–9724. <https://doi.org/10.1074/jbc.RA119.012121>.
- 543
- 544 45. 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 domain. *EMBO Reports* 10, 1235–1241. <https://doi.org/10.1038/embor.2009.218>.

- 545
- 546 46. 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*, 547 640–644. <https://doi.org/10.1038/nature13899>.
- 548 47. 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>.
- 549 48. 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>.
- 550 49. Hayashi, K., Ohta, H., Kurimoto, K., Aramaki, S., and Saitou, M. (2011). Reconstitution of the mouse germ cell specification pathway in culture by pluripotent stem cells. *Cell* *146*, 519–532. 551 <https://doi.org/10.1016/j.cell.2011.06.052>.
- 552 50. 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>.
- 553 51. 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 554 spermatogenesis. *Proc. Natl. Acad. Sci. U.S.A.* *112*. <https://doi.org/10.1073/pnas.1505683112>.
- 555 52. Bowles, J., Knight, D., Smith, C., Wilhelm, D., Richman, J., Mamiya, S., Yashiro, K., Chawengsak- 556 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>.
- 557 53. 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*, 558 400–418.e13. <https://doi.org/10.1016/j.stem.2022.01.010>.
- 559 54. 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*, 560 676–687. <https://doi.org/10.1016/j.stem.2011.04.004>.
- 561 55. 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. 562 <https://doi.org/10.1371/journal.pone.0205969>.
- 563
- 564
- 565

- 566 56. Stephens, M. (2016). False discovery rates: A new deal. Biostat, kxw041. <https://doi.org/10.1093/biostatistics/kxw041>.
- 567
- 568 57. 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>.
- 569
- 570 58. 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.
- 571 <https://doi.org/10.1016/j.molcel.2010.05.004>.

572 **Figures and Tables**

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

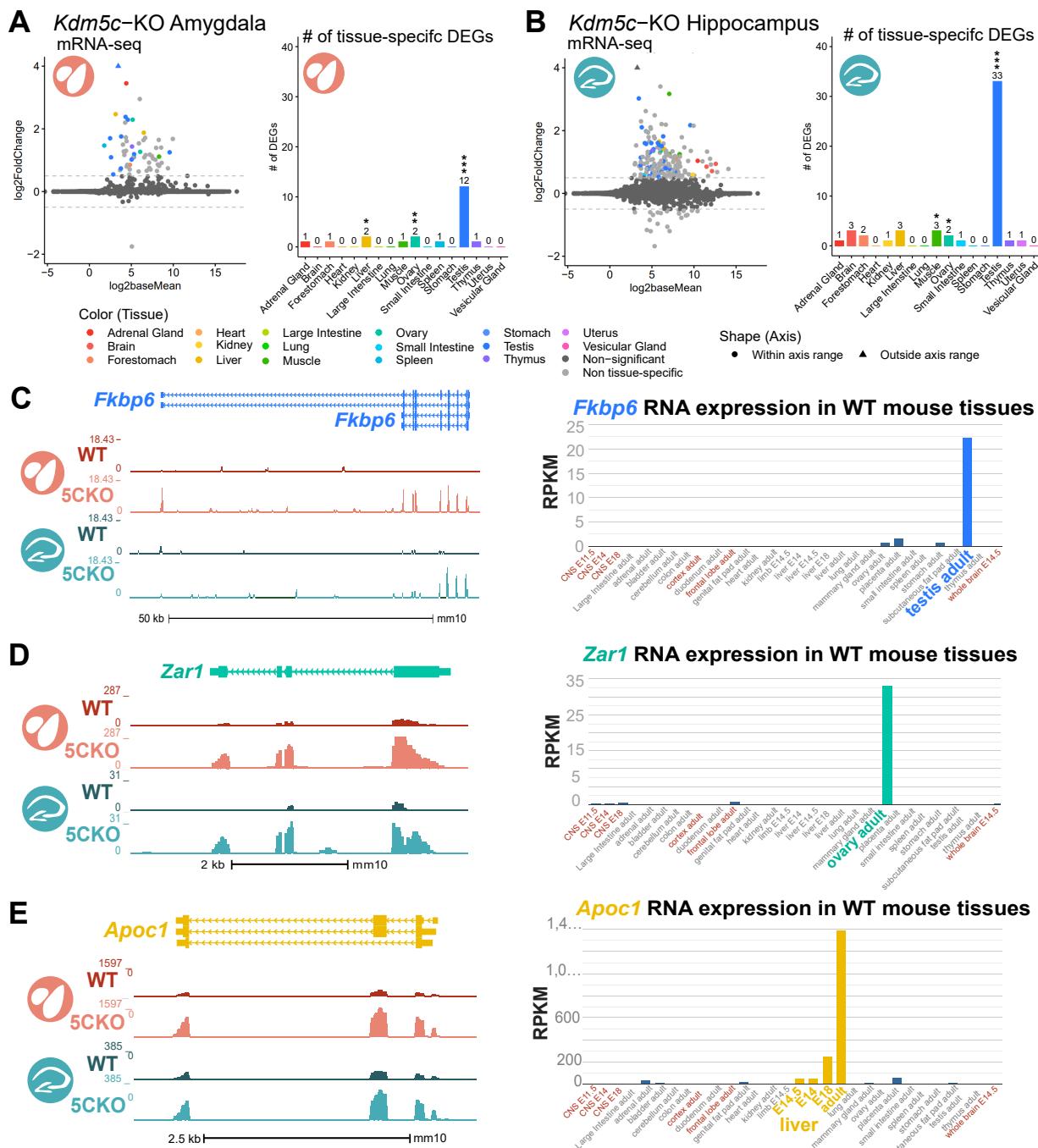


Figure 1: Tissue-enriched genes are misexpressed in the *Kdm5c*-KO brain. **A-B.** Expression of tissue-enriched genes (Li et al 2017) in the male *Kdm5c*-KO amygdala (A) and hippocampus (B). Left - MA plot of mRNA-sequencing. Right - Number of tissue-enriched differentially expressed genes (DEGs). *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

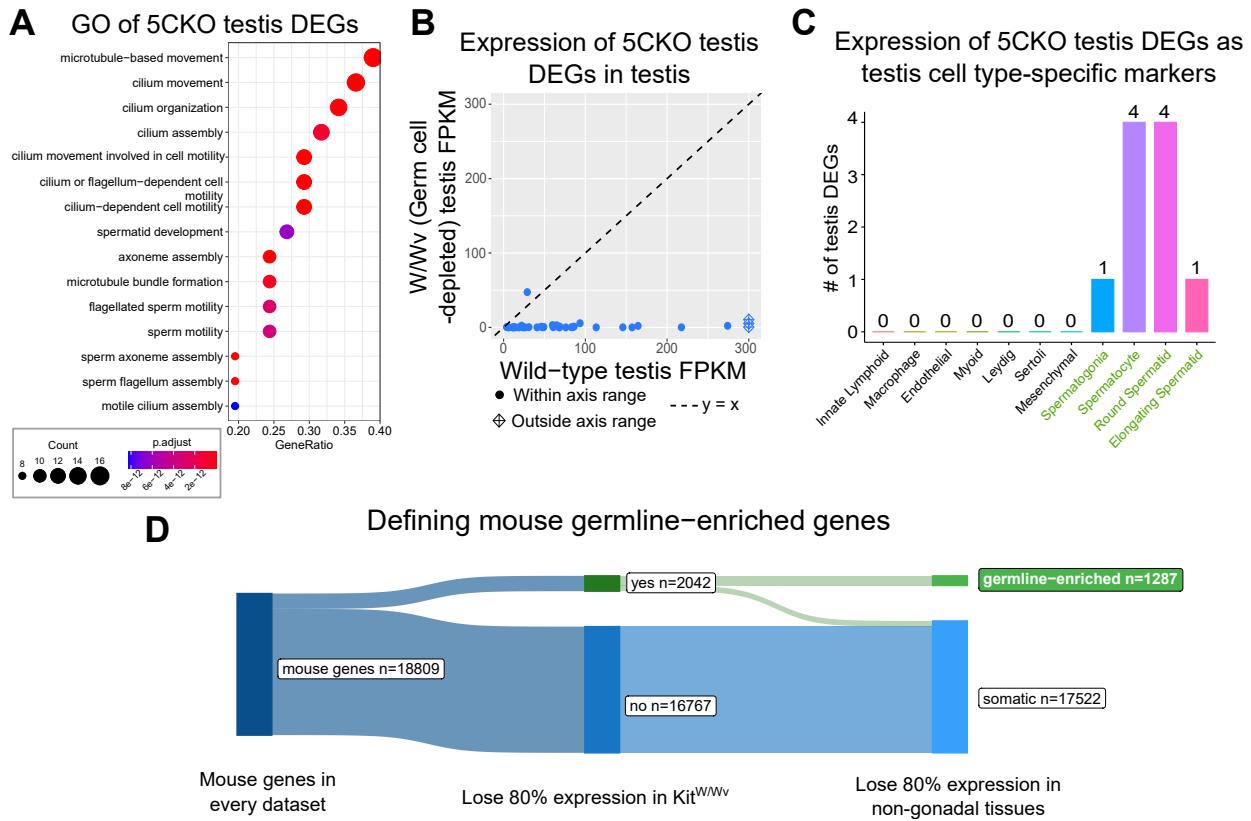


Figure 2: Aberrant transcription of germline genes in the *Kdm5c*-KO in the brain. **A.** enrichPlot gene ontology (GO) of *Kdm5c*-KO amygdala and hippocampus testis-enriched DEGs **B.** Expression of testis DEGs in wild-type (WT) testis versus germ cell-depleted (W/Wv) testis (Mueller et al 2013). Expression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

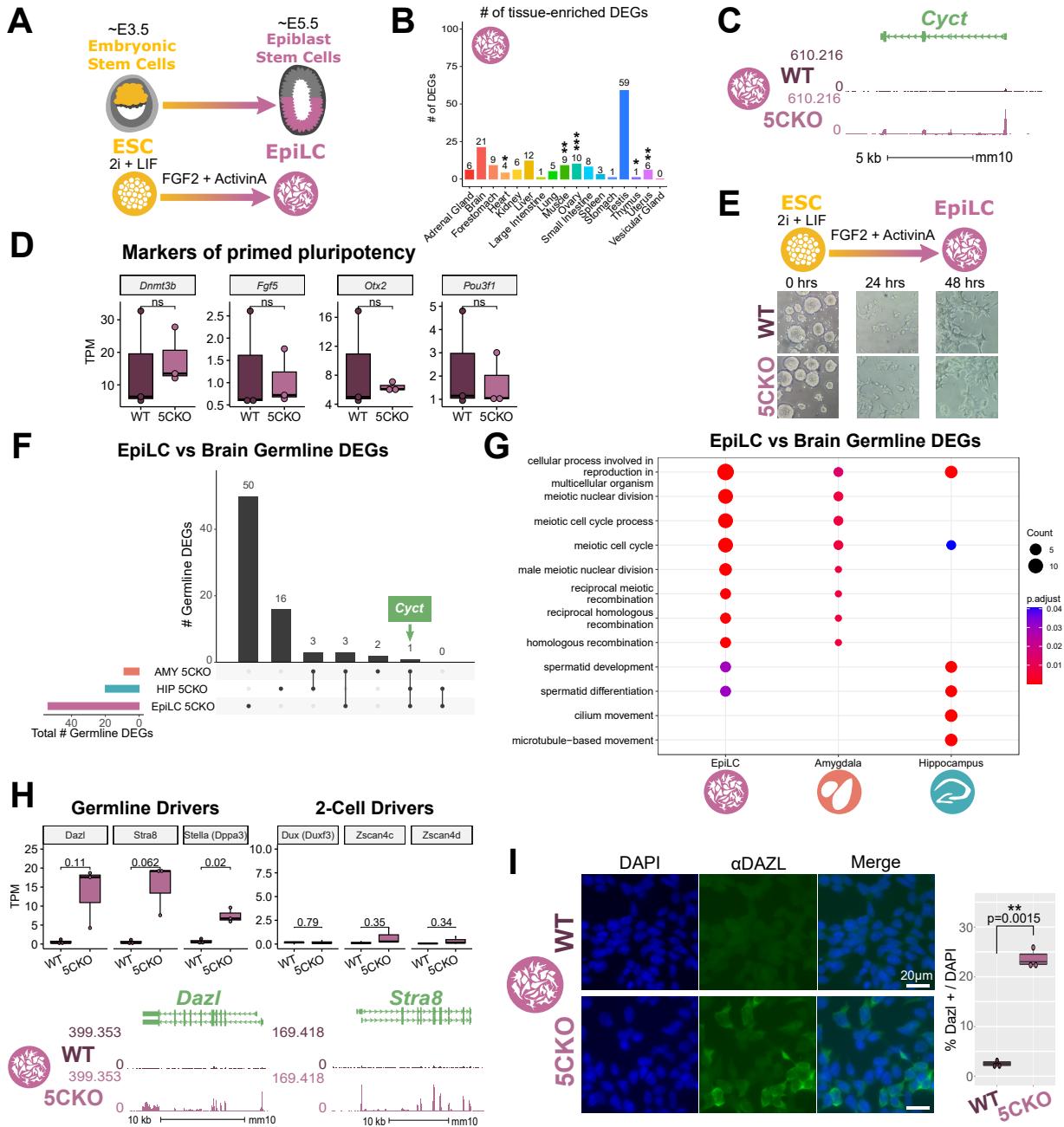


Figure 3: Kdm5c-KO epiblast-like cells express key drivers of germline identity **A.** Top - Diagram of *in vivo* differentiation of embryonic stem cells (ESCs) of the inner cell mass into epiblast stem cells. Bottom - *in vitro* differentiation of ESCs into epiblast-like cells (EpiLCs). **B.** Number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

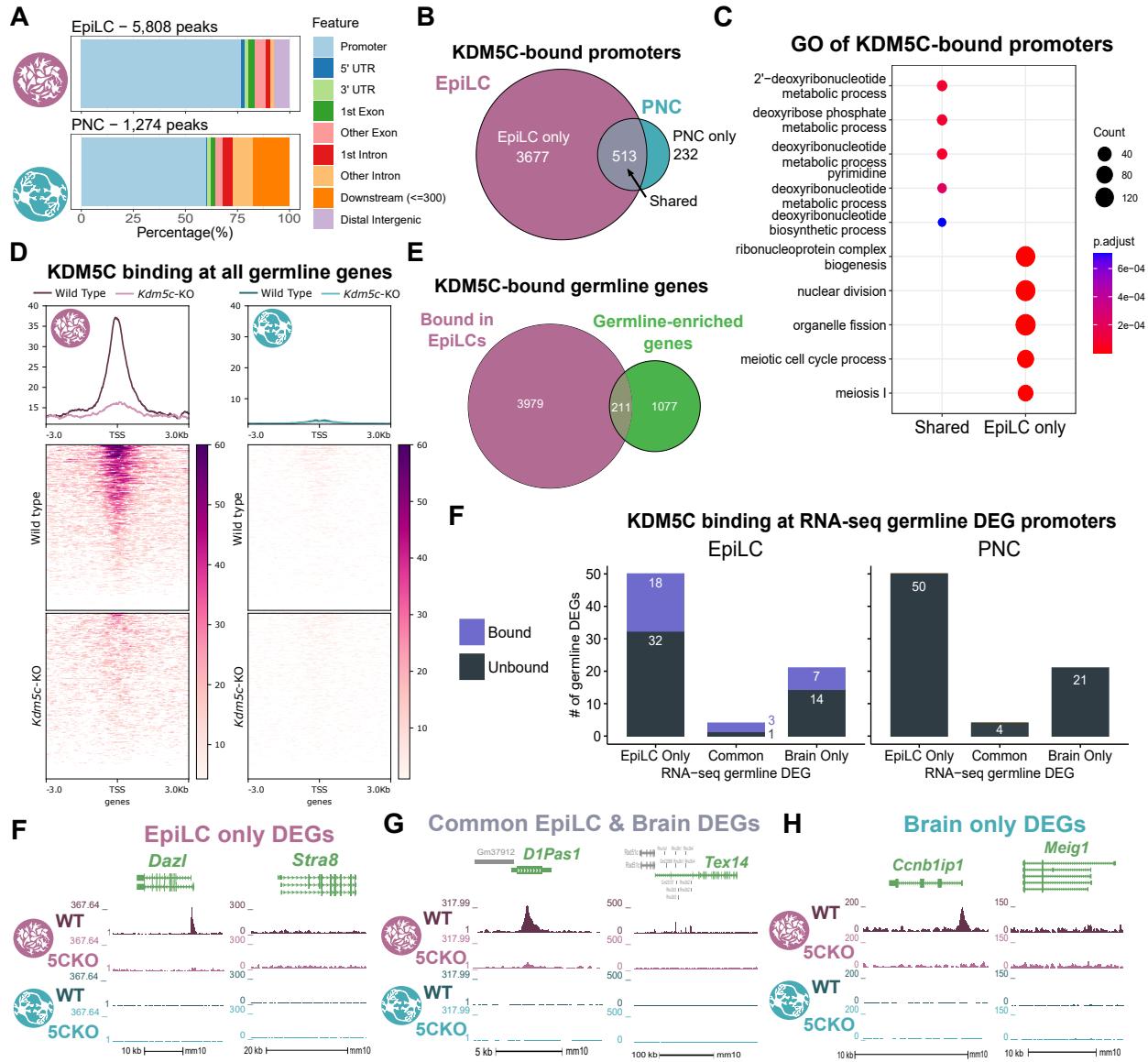
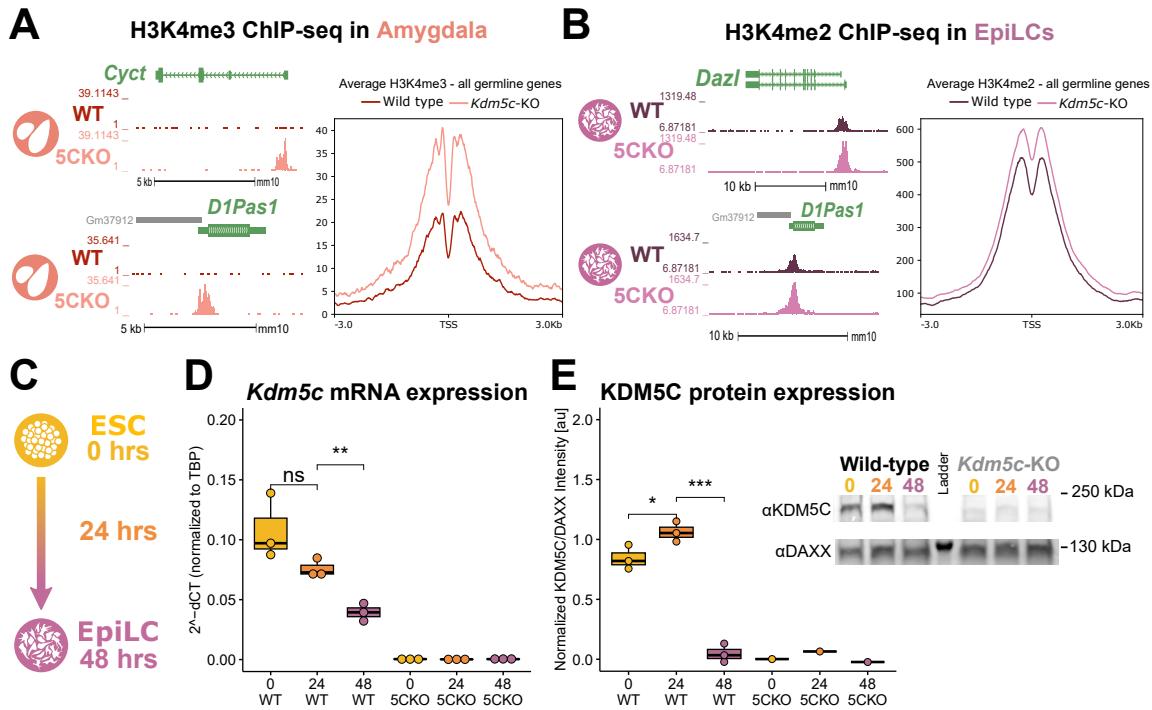


Figure 4: KDM5C binds to a subset of germline gene promoters during early embryogenesis **A.** ChIPseeker localization of KDM5C peaks at different genomic regions in EpiLCs (top) and hippocampal and cortex primary neuron cultures (PNCs, bottom). **B.** Overlap of genes with KDM5C bound to their promoters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

577 **Notes**

578 **Discussion notes**

- 579 • We additionally found *Kdm5c*-KO EpiLCs ectopically express DAZL protein that is localized to the
580 cytoplasm, similar to its morphology in spermatogonia⁴². **note: maybe just put in results.** Could
581 move around depending upon if I get pheno working.
 - 582 • end with something like: this indicates kdm5c not only modulates pluripotency and self-renewal in
583 ESCs, but also has a role in long-term silencing of germline genes
 - 584 – then transition into the long term silencing mechanism paragraph
 - 585 • KDM5C is dynamically regulated during the window of embryonic germline gene silencing, our EpiLC
586 ChIP-seq is likely catching the tail end of KDM5C's main involvement.
 - 587 • Our results indicate KDM5C is a pivotal repressor of germline programs in somatic cells.
 - 588 • Studies on other chromatin regulators have focused on key marker genes to identify germline gene
589 misexpression, such as *Dazl*.
 - 590 • Using a germline-depleted mouse model, we curated a list of mouse germline-enriched genes to
591 globally assess germline gene dysregulation.
 - 592 • Many of the genes dysregulated in the mature *Kdm5c*-KO brain were important for late stages of
593 spermatogenesis, such as those important for sperm flagellar structure. Contrastingly, *Kdm5c*-KO
594 EpiLCs aberrantly expressed key drivers of germ cell identity, including *Dazl* and *Stra8*.
 - 595 • While a significant portion of KDM5C-bound promoters are germline genes, many germline genes
596 expressed during *Kdm5c*-KO embryogenesis are not directly bound by kdm5c.
 - 597 • One notable example is *Stra8*, a transcription factor important for germ cell specific transcription and
598 meiotic initiation
 - 599 • The including the demarcation between soma and germline fates.
- 600 the *Kdm5c*-KO brain also Previous work identified testis-enriched genes in *Kdm5c* genes
- 601 –
- 602 – However unlike the gonadal-biased DEGs,
- 603 • The demarcation of the germline versus soma is a key feature of multicellularity and sexually dimorphic
604 reproduction
- 605 • Anything known about tissue-biased gene expression in other H3K4me regulators?

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

617 **Figure outline:**

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

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

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

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

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

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

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

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

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

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

682 **Germline gene repression background:**

683 Interestingly, some of the ectopic testis transcripts identified in the *Kdm5c*-KO brain are typically ex-
684 pressed in germ cells¹⁰. Unlike somatic cells, germ cells (e.g. sperm and eggs) undergo meiosis and pass
685 on their genetic material to the next generation. The germline and the soma are typically distinguished during
686 early embryogenesis, when germline genes are silenced in epiblast stem cells soon after implantation and
687 only reactivated in a subset to form the germline. Chromatin regulators play a key role in decommissioning
688 germline genes as the embryo transitions from naïve to primed pluripotency by placing repressive histone
689 H2A lysine 119 monoubiquitination (H2AK119ub1)¹⁶, histone 3 lysine 9 trimethylation (H3K9me3)^{16,17}, and
690 DNA CpG methylation^{17–19} at germline gene promoters. KDM5C may also be involved in this early decom-
691 missioning of germline genes, as re-expression of KDM5C in neurons fails to suppress their dysregulation¹⁰.
692 In support of this, KDM5C was very recently shown to repress *Deleted in azoospermia like* (*Dazl*), a key
693 regulator of germline development, in mouse embryonic stem cells (ESCs)^{40,43}. In support of this, two
694 independent screens in mouse embryonic stem cells (ESCs) recently identified KDM5C as a repressor of
695 *Deleted in azoospermia like* (*Dazl*), a key regulator of germline development. However, KDM5C's role in
696 embryonic germline gene repression is currently unclear, given that *Dazl* is also expressed in ESCs and in
697 the 2-cell stage and germline gene misexpression has yet to be globally characterized during *Kdm5c*-KO
698 embryogenesis.