

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

3

4      **Abstract**

5      **Introduction**

6      A single genome holds the instructions to generate the myriad of cell types found within the adult organism.  
7      This is, in part, accomplished by chromatin regulators that can either promote or impede lineage-specific  
8      gene expression through DNA and histone modifications<sup>1,2</sup>. Although many chromatin regulators were initially  
9      identified for their roles in shaping cellular and tissue identity<sup>3,4</sup>, recent advancements in next generation  
10     sequencing unexpectedly revealed many neurodevelopmental disorders (NDDs) are caused by mutations in  
11     chromatin regulators<sup>5</sup>. Several studies have suggested this connection between chromatin regulators and  
12     neurodevelopment is due to their regulation of brain-specific genes, such as orchestrating transcriptional  
13     programs for synaptic maturation<sup>6</sup> and transitioning between neuronal and glial fates during neural precursor  
14     differentiation<sup>7</sup>. However, loss of some chromatin regulators can also result in the ectopic transcription of  
15     tissue-specific genes outside of their target environment<sup>3,4,8</sup>, such as the misexpression of liver-specific  
16     genes within adult neurons<sup>9</sup>. Very few studies have investigated the misexpression of tissue-specific genes  
17     in chromatin-linked NDDs<sup>9,10</sup> and it is currently unknown if this partial loss of brain identity contributes to  
18     neurodevelopmental impairments.

19      To elucidate the role of tissue identity in chromatin-linked neurodevelopmental disorders, it is essential  
20     to first characterize the types of genes dysregulated and the molecular mechanisms governing their de-  
21     repression. In this study, we characterized the aberrant expression of tissue-enriched genes with loss of  
22     lysine demethylase 5C (KDM5C). KDM5C, , also known as SMCX or JARID1C, is a chromatin regulator that  
23     can repress gene expression through erasure of histone 3 lysine 4 di- and trimethylation (H3K4me2/3)<sup>11</sup>,  
24     marks enriched at active gene promoters. Pathogenic mutations in *KDM5C* cause Intellectual Developmental  
25     Disorder, X-linked, Syndromic, Claes-Jensen Type (MRXSCJ, OMIM: 300534), whose predominant features  
26     include intellectual disability, seizures, aberrant aggression, and autistic behaviors<sup>12–14</sup>. *Kdm5c* knockout  
27     (-KO) mice recapitulate key MRXSCJ patient phenotypes, including hyperaggression, increased seizure  
28     propensity, and learning impairments<sup>10,15</sup>. Unexpectedly, RNA sequencing (RNA-seq) of the *Kdm5c*-KO  
29     hippocampus revealed ectopic expression of testis genes within the brain<sup>10</sup>. It is currently unknown if this

30 dysregulation of brain tissue identity further impairs *Kdm5c*-KO neurodevelopment and if ectopic gene  
31 expression within the *Kdm5c*-KO brain is unique to testis genes.

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

33 Distinguishing between the germline (cells that pass on their genetic material, e.g. sperm and eggs) and  
34 the soma (cells that perform all other bodily functions) is a key feature of multicellularity and occurs during  
35 early embryogenesis. In mammals, chromatin regulators play a key role in decommissioning germline genes  
36 in somatic cells during the transition from naïve to primed pluripotency by placing repressive histone H2A  
37 lysine 119 monoubiquitination (H2AK119ub1)<sup>16</sup>, histone 3 lysine 9 trimethylation (H3K9me3)<sup>16,17</sup>, and DNA  
38 CpG methylation<sup>17–19</sup> at germline gene promoters. Systematically characterizing KDM5C's role in germline  
39 gene repression during early embryogenesis will unveil key mechanisms underlying the demarcation between  
40 soma and germline identity and while also providing molecular footholds to test the impact of ectopic germline  
41 genes on neurodevelopment.

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

## 56 **Results**

### 57 **Tissue-enriched genes, including testis genes, are aberrantly expressed in the** 58 ***Kdm5c*-KO brain**

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

60 Previous RNA sequencing (RNA-seq) in the adult hippocampus revealed ectopic expression of some testis  
61 genes within the *Kdm5c* knockout (-KO) brain<sup>10</sup>. Since tissue-enriched genes have not been systematically

62 characterized in the *Kdm5c*-KO brain, it is currently unclear if this erosion of brain tissue identity is a major  
63 consequence of *Kdm5c* loss and if it is unique to testis-enriched genes. Therefore, we first globally assessed  
64 the expression of genes enriched in 17 mouse tissues<sup>20</sup> in our published mRNA-seq datasets of the adult  
65 amygdala and hippocampus in mice with constitutive knockout of *Kdm5c*<sup>21</sup>.

66 We found a large proportion of genes that are significantly upregulated within the *Kdm5c*-KO brain  
67 (DESeq2<sup>22</sup>, log2 fold change > 0.5, q < 0.1) are typically enriched in non-brain tissues (Amygdala: 35%,  
68 Hippocampus: 24%) (Figure 1A-B). The majority of tissue-enriched differentially expressed genes (tissue-  
69 enriched DEGs) were testis genes (Figure 1A-C). Even though the testis has the largest total number of  
70 tissue-biased genes compared to any other tissue (2,496 genes), testis-biased DEGs were significantly  
71 enriched for both brain regions (Amygdala p = 1.83e-05, Odds Ratio = 5.13; Hippocampus p = 4.26e-11,  
72 Odds Ratio = 4.45, Fisher's Exact Test). One example of a testis-enriched DEG is *FK506 binding protein 6*  
73 (*Fkbp6*), a known regulator of piRNA expression and meiosis in germ cells<sup>23,24</sup> (Figure 1C).

74 In addition to the high enrichment of testis genes, we also identified aberrant expression of other  
75 tissue-enriched genes within the *Kdm5c*-KO brain. Although the *Kdm5c*-KO mice were male, we observed  
76 significant enrichment of ovary-biased DEGs in both the amygdala and hippocampus (Amygdala p = 0.00574,  
77 Odds Ratio = 18.7; Hippocampus p = 0.048, Odds Ratio = 5.88, Fisher's Exact) (Figure 1D). Ovary-enriched  
78 DEGs included *Zygotic arrest 1 (Zar1)*, which was recently shown to sequester mRNAs in oocytes for meiotic  
79 maturation and early zygote development<sup>25</sup> (Figure 1D). Although not consistent across brain regions, we  
80 also found significant enrichment of DEGS biased towards two non-gonadal tissues - the liver (Amygdala p =  
81 0.0398, Odds Ratio = 6.58, Fisher's Exact Test) and the muscle (Hippocampus p = 0.0104, Odds Ratio =  
82 6.95, Fisher's Exact Test). An example of a liver-biased DEG dysregulated in both the hippocampus and  
83 amygdala is *Apolipoprotein C-I (Apoc1)*, which is involved in lipoprotein metabolism (Figure 1E). Testis,  
84 ovary, and liver-enriched DEGs showed little to no expression in the developing and adult wild-type brain,  
85 yet our mRNA-seq data indicate they are polyadenylated and spliced into mature transcripts (Figure 1C-E).  
86 Of note, we did not observe enrichment of brain-enriched genes (Amygdala p = 1; Hippocampus p = 0.74,  
87 Fisher's Exact), despite the fact these are brain samples and the brain has the second highest total number  
88 of tissue-enriched genes (708 genes). Together, these results suggest misexpression of tissue-enriched  
89 genes within the brain is a major effect of KDM5C loss.

## 90 **Germline genes are misexpressed in the *Kdm5c*-KO brain**

91 • Intriguingly, some of the ectopic testis transcripts identified within the *Kdm5c*-KO hippocampus have  
92 known functions unique to germ cells<sup>10</sup>, suggesting KDM5C may play a role in demarcating somatic  
93 versus germline identity.

94 The testis contains both germ cells (meiotic cells, e.g. spermatogonia) and somatic cells (non-meiotic,  
95 e.g. Leydig cells) that support hormone production and germline functions. Intriguingly, many *Kdm5c*-

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

101 To further validate if these testis DEGs are truly germline genes, we then evaluated their expression in  
102 somatic versus germ cells within the testis. We first compared their expression in the testis with germ cell  
103 depletion<sup>26</sup>, which was accomplished by heterozygous *W* and *Wv* mutations in the enzymatic domain of  
104 *c-Kit* (*Kit*<sup>W/Wv</sup>) that prevent the maturation of germ cells<sup>27</sup>. Almost all *Kdm5c*-KO testis-enriched DEGs lost  
105 expression with germ cell depletion (Figure 2B). The only testis-enriched DEG that did not show considerable  
106 downregulation with germline depletion was *FK506 binding protein 6* (*Fkbp6*), the aforementioned testis  
107 gene that regulates piRNA expression and meiosis in germ cells<sup>23,24</sup>. We then assessed testis-enriched  
108 DEG expression in a published single cell RNA-seq dataset that identified cell type-specific markers within  
109 the testis<sup>28</sup>. We found that while some testis-enriched DEGs were classified as specific markers for different  
110 germ cell types (e.g. spermatogonia, spermatocytes, round spermatids, and elongating spermatids) none  
111 marked somatic cells (Figure 2C). Together, these data suggest the somatic *Kdm5c*-KO brain aberrantly  
112 expresses germline genes.

113 We then wanted to more deeply characterize germline gene misexpression with *Kdm5c* loss, but lacked  
114 a comprehensive list of mouse germline-enriched genes. To facilitate downstream analyses, we generated a  
115 curated list of germline-enriched genes using currently available RNA-seq datasets in *Kit*<sup>W/Wv</sup> mice. Wild-type  
116 and *Kit*<sup>W/Wv</sup> datasets included males and females at embryonic day 12, 14, and 16<sup>29</sup>, as well as adult  
117 male testes<sup>26</sup>. We defined genes as germline-enriched if their expression met the following criteria: 1)  
118 their expression is greater than 1 FPKM in wild-type gonads 2) their expression in any adult wild-type,  
119 non-gonadal tissue<sup>20</sup> does not exceed 20% of their maximum expression in the wild-type germline, and  
120 3) their expression in the germ cell-depleted germline, for any sex or time point, does not exceed 20% of  
121 their maximum expression in the wild-type germline. These criteria yielded 1,288 germline-enriched genes  
122 (Figure 2D), which was hereafter used as a resource for assessing germline gene misexpression with *Kdm5c*  
123 loss (Supplementary table 1).

## 124 ***Kdm5c*-KO epiblast-like cells aberrantly express master regulators of germline 125 identity**

126 Misexpression of germline genes in the adult brain suggests *Kdm5c*-KO cells fail to demarcate between  
127 germline and somatic identity. Germ cells are typically distinguished from somatic cells soon after the embryo  
128 implants into the uterine wall<sup>30,31</sup> when a subset of epiblast stem cells become the primordial germ cells  
129 (PGCs) while the remainder differentiate into the ectoderm, mesoderm, and endoderm to form the somatic

130 tissues<sup>32</sup>. This developmental time point can be modeled *in vitro* through differentiation of embryonic stem  
131 cells (ESCs) into post-implantation epiblast-like stem cells (EpiLCs) (Figure 3A, top). Previous studies have  
132 demonstrated that while some germline-enriched genes are also expressed in embryonic stem cells (ESCs)  
133 and in the 2-cell stage<sup>33–35</sup>, they are silenced as they differentiate into EpiLCs<sup>17</sup>. Therefore, we assessed if  
134 KDM5C was necessary for embryonic silencing of germline genes by evaluating the impact of *Kdm5c* loss in  
135 EpiLCs.

136 We first identified *Kdm5c*-KO EpiLC DEGs in of our previously published RNA-seq dataset<sup>36</sup> (DESeq2,  
137 log2 fold change > 0.5, q < 0.1) and assessed tissue-enriched gene expression. Similar to the *Kdm5c*-KO  
138 brain, we observed general dysregulation of tissue-enriched genes, with the largest number of genes  
139 belonging to the brain and testis, although they were not significantly enriched (Figure 3B). Using the curated  
140 list of germline genes generated above, we found *Kdm5c*-KO EpiLCs aberrantly expressed 54 germline-  
141 enriched genes, including the previously characterized hippocampal DEG<sup>10</sup> *Cytochrome C, testis-specific*  
142 (*Cyct*) (Figure 3C). Although we observed aberrant expression of many tissue-enriched genes, we did not  
143 observe any significant difference in primed pluripotency genes (Figure 3D) or gross changes in *Kdm5c*-KO  
144 cell morpholgy during differentiation (Figure 3E), indicating KDM5C loss does not impair EpiLC formation.

145 We then compared EpiLC germline DEGs to those expressed in the *Kdm5c*-KO brain to determine  
146 if all germline DEGs, like *Cyct*, are constitutively dysregulated or if they can change over the course of  
147 development. We found this was primarily not the case, as the majority of germline DEGs expressed  
148 in EpiLCs and the brain were unique, with only *Cyct* shared across all sequencing datasets (Figure 3F).  
149 We then compared the known functions of EpiLC and brain germline DEGs and found particularly high  
150 enrichment of meiosis-related gene ontologies in EpiLCs (Figure 3G), such as meiotic cell cycle process (GO:  
151 1903046, p.adjust = 1.59e-08) and meiotic nuclear division (GO: 0140013, p.adjust = 9.76e-09). While there  
152 was modest enrichment of meiotic gene ontologies in both brain regions, the *Kdm5c*-KO hippocampus  
153 primarily expressed late-stage spermatogenesis genes, such those involved in the sperm axoneme structure.

154 Interestingly, DEGs unique to *Kdm5c*-KO EpiLCs included key drivers of germline identity, such as  
155 *Stimulated by retinoic acid 8 (Stra8)* and *Deleted in azoospermia like (Dazl)* (Figure 3H). These genes are  
156 typically expressed during embryonic germ cell development to commit PGCs to the germline fate, but are  
157 also expressed later in life to trigger meiotic gene expression programs<sup>37–39</sup>. Of note, some germline genes,  
158 including *Dazl*, are also expressed in the two-cell embryo<sup>34,40</sup>. However, we did not see misexpression of  
159 two-cell embryo-specific genes, like *Duxf3 (Dux)* (q = 0.337) and *Zscan4d* (q = 0.381), indicating *Kdm5c*-KO  
160 in EpiLCs do not revert back to a 2-cell state (Figure 3H).

161 We were particularly interested in the aberrant transcription of *Dazl*, since it is essential for germ cell  
162 development and promotes the translation of germline mRNAs<sup>41</sup>. A significant portion of germline transcripts  
163 misexpressed in *Kdm5c*-KO EpiLCs are known binding targets of DAZL, including *Stra8*<sup>42</sup> (p = 1.698e-07,  
164 Fisher's Exact Test). This suggests expression of DAZL protein could promote the translation of other  
165 aberrant germline transcripts, influencing their ability to impact *Kdm5c*-KO cellular function. We thus tested

166 DAZL protein expression in *Kdm5c*-KO EpiLCs through immunocytochemistry (Figure 3I). We observed  
167 about 25% of *Kdm5c*-KO EpiLCs expressed DAZL protein and it was localized to the cytoplasm ( $p = 0.0015$ ,  
168 Welch's t-test), consistent with the pattern of DAZL expression in spermatogonia<sup>42</sup>. Altogether these results  
169 suggest tissue-specific genes are misexpressed during *Kdm5c*-KO embryogenesis, including key drivers of  
170 germline identity that can be translated into protein.

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

172 • KDM5C may also be involved in this early decommissioning of germline genes, as re-expression of  
173 KDM5C in knockout neurons fails to suppress their dysregulation<sup>10</sup>.

174 Previous work suggests KDM5C represses germline genes during early development, as re-expression  
175 of KDM5C in knockout neuronal cultures fails to suppress hippocampal germline DEGs and KDM5C is not  
176 bound to their promoters in neurons<sup>10</sup>. There is some evidence KDM5C binds to select germline gene  
177 promoters in ESCs<sup>10</sup>, including *Dazl*<sup>40,43</sup>. As KDM5C's binding at germline gene promoters has not been  
178 systematically characterized, it is currently unclear if KDM5C is enriched at germline gene promoters, what  
179 types of germline genes KDM5C regulates, and if its binding is maintained at any germline genes in neurons.

180 To further characterize the mechanism behind *Kdm5c*-KO germline gene misexpression, we analyzed  
181 KDM5C chromatin immunoprecipitation followed by DNA sequencing (ChIP-seq) datasets in EpiLCs<sup>36</sup> and  
182 primary neuron cultures (PNCs) from the cortex and hippocampus<sup>15</sup>. EpiLCs had a higher total number of  
183 KDM5C peaks than PNCs (EpiLCs: 5,808, PNCs: 1,276, MACS2 q < 0.1 and fold enrichment > 1, removal  
184 of *Kdm5c*-KO false positive peaks) and KDM5C was primarily localized to gene promoters in both cell types  
185 (EpiLCs: 4,190, PNCs: 745 +/- 500bp from the TSS), although PNCs showed increased localization to  
186 non-promoter regions (Figure 4A).

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

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

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

## 212 **KDM5C erases H3K4me3 to promote long-term repression of germline genes via 213 DNA methylation**

214 Although KDM5C is generally thought to suppress transcription through erasure of histone 3 lysine 4 di-  
215 and trimethylation (H3K4me2/3)<sup>11</sup>, recent studies in ESCs have suggested KDM5C's repression *Dazl* is  
216 independent of its catalytic activity<sup>40</sup>. Somatic repression of germline genes is typically established during the  
217 transition between naïve and primed pluripotency, which modeled *in vitro* as ESCs to EpiLC differentiation.  
218 In ESCs, chromatin regulators place repressive histone modifications at germline gene promoters, including  
219 histone 2A 119 monoubiquitination (H2AK119ub1) and histone 3 lysine 9 trimethylation (H3K9me3)<sup>16,17,44</sup>.  
220 Germline genes are then silenced long-term in EpiLCs by *de novo* placement of DNA CpG methylation  
221 (DNAm)<sup>17</sup>. Although KDM5C's catalytic activity is not required to suppress *Dazl* in ESCs, KDM5C may  
222 promote long-term germline gene silencing via H3K4me3 removal since H3K4me3 can impede DNAm  
223 placement<sup>45,46</sup> and DNAm is lost at select germline gene promoters in the hippocampus<sup>10</sup>. Because  
224 KDM5C's role in germline gene repression has only been characterized in ESCs and in the mature brain, it is  
225 currently unclear to what extent KDM5C is involved during transition between ESCs and EpiLCs and if its  
226 catalytic activity is required for long-term silencing.

227 To elucidate KDM5C's role in germline gene silencing, we first characterized KDM5C's substrates, histone  
228 3 lysine 4 di- and trimethylation (H3K4me2/3) in our previously published ChIP-seq datasets in wild type and  
229 *Kdm5c*-KO amygdala<sup>21</sup> and EpiLCs<sup>36</sup>. In congruence with previous work in the *Kdm5c*-KO hippocampus<sup>10</sup>,  
230 we observed aberrant accumulation of H3K4me3 around the transcription start site (TSS) of germline genes  
231 in the *Kdm5c*-KO amygdala (Figure 5A). We additionally found a marked increase in H3K4me2 around the  
232 TSS of germline genes in *Kdm5c*-KO EpiLCs (Figure 5B).

233 We then evaluated KDM5C's embryonic role in germline gene silencing during ESC to EpiLC differentiation.

234 We first characterized KDM5C's mRNA and protein expression at 0, 24, and 48 hours of differentiation  
235 (Figure 5C). While *Kdm5c* mRNA steadily decreased from 0 to 48 hours of differentiation (Figure 5D), KDM5C  
236 protein initially increased from 0 to 24 hours but then decreased to near knockout levels by 48 hours (Figure  
237 5E). We then performed whole genome bisulfite sequencing in wild-type and *Kdm5c*-KO ESCs and 96 hour  
238 extended EpiLCs (exEpiLCs) to determine KDM5C's role in the initial placement of DNA methylation at  
239 germline gene promoters. While wild-type cells accumulated high levels of DNA methylation at germline  
240 gene promoters over the course of ESCs to exEpiLC differentiation, DNA methylation was markedly reduced  
241 in *Kdm5c*-KO exEpiLCs (Figure 5F).

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

## 247 Experimental Procedures

## 248 Discussion

249 In the above study, we demonstrate KDM5C's pivotal role in the development of tissue identity, resulting in  
250 substantial misexpression of tissue-enriched genes within the adult *Kdm5c*-KO hippocampus and amygdala.  
251 We defined tissue-enriched genes based on the criteria used in a previously published dataset of 17 C57Bl6J  
252 mouse tissues<sup>20</sup>, which defined genes as tissue-enriched if they had more than 4-fold higher expression  
253 when compared to any other tissue. In addition to testis genes identified previously<sup>10</sup>, we found significant  
254 enrichment of muscle, liver, and even ovary-biased genes aberrantly expressed within the male *Kdm5c*-KO  
255 brain. It is currently unclear what the impact of these tissue-enriched genes are upon *Kdm5c*-KO brain  
256 function and if they ultimately contribute to MRXSCJ-related phenotypes. While the majority of tissue-  
257 enriched DEGs were testis and ovary genes with no known brain functions, select liver and muscle-biased  
258 DEGs do have known roles in the wild-type brain. For example, *Apolipoprotein C-I* (*Apoc1*) is a lipid transport  
259 gene highly enriched within the liver[XXX], but is lowly expressed in the wild-type brain. Intriguingly, *Apoc1*  
260 overexpression in the mouse brain can impair learning and memory<sup>47</sup> and is implicated in Alzheimer's  
261 disease in humans<sup>48</sup>. Our results suggest KDM5C fine-tunes the expression of tissue-enriched genes like  
262 *Apoc1* to match the level required for proper brain function. However, further studies are required to determine  
263 if these genes contribute to *Kdm5c*-KO neuronal impairments and MRXSCJ pathology.

264 A substantial portion of genes upregulated in *Kdm5c*-KO brain are testis genes that have no known  
265 function within the brain. Through the use of publicly available RNA-seq datasets, we demonstrated these

266 testis-enriched DEGs are typically unique to germ cells. Misexpression of germline genes in the brain sug-  
267 gests *Kdm5c*-KO fail to demarcate between somatic and germline lineages, a key feature of multicellularity  
268 and sexual reproduction. Previous studies have demonstrated germline genes are decommissioned in  
269 somatic cells by multiple chromatin-modifying enzymes, but have primarily focused on the repression of  
270 key marker genes. To globally characterize KDM5C's role in germline gene repression, we curated a list  
271 of mouse germline-enriched genes using publically available germ cell-depleted RNA-seq datasets. This  
272 resource enabled us to identify 1) the extent of germline gene dysregulation in *Kdm5c*-KO cells, 2) the types  
273 of germline genes misexpressed at different developmental time points, and 3) which types of germline  
274 genes are directly or indirectly regulated by KDM5C.

275 We found this dysregulation in tissue identity likely begins during early emrbyogenesis, as germline  
276 genes and other tissue-enriched genes are aberrantly expressed in *Kdm5c*-KO epiblast-like cells (EpiLCs),  
277 an *in vitro* model of the post-implantation embryo. *In vivo*, germline genes are typically decommissioned  
278 in epiblast stem cells and remain silenced as the epiblast differentiates into the body's somatic tissues[].  
279 However, a small subset of epiblast stem cells will receive signals to reactivate germline gene expression  
280 to become the primordial germ cells (PGCs) that will ultimately form the mature germline[]. This process  
281 can be mimicked *in vitro* by differentiating EpiLCs into primordial germ cell-like cells (PGCLCs)[XXX]. Thus,  
282 misexpression of germline genes in *Kdm5c*-KO EpiLCs might suggest they are progressing beyond EpiLC  
283 differentiation and becoming PGCLCs. Yet, *Kdm5c*-KO EpiLCs had proper expression of EpiLC marker  
284 genes for primed pluripotency and we observed no difference in cellular morphology during ESC to EpiLC  
285 differentiation. Furthermore, we saw no significant change in *Otx2* expression, an epiblast stem cell marker  
286 that is known to repress epiblast and EpiLC differentiation into PGC and PGCLCs<sup>zhangOTX2RestrictsEntry2018?</sup>.  
287 This, in conjunction with *Kdm5c*-KO mice being viable, suggests germline gene expression is occurring  
288 ectopically in conjunction with typical developmental programs, rather than a complete shift to germline  
289 identity.

- 290 • Only a small number of germline genes were shared between *Kdm5c*-KO EpiLCs and in the brain.
- 291 dysregulation of tissue identity
- 292 • In line with previous work on other chromatin regulators, we found dysregulation of *Kdm5c*-KO tissue  
293 identity began during early embryogenesis. While *Kdm5c*-KO epiblast-like cells (EpiLCs)
- 294 • KDM5C represses germline genes during early embryogenesis
- 295 – Found this dysregulation in tissue identity begins during *Kdm5c*-KO early embryogenesis
- 296 – EpiLC express early regualtors, Brain late sperm - program seems to be continuing across  
297 development
- 298 – Importantly, even though germline genes are misexpressed, markers of EpiLC differentiation and  
299 EpiLC morphology is unimpaired.

- 300                  \* Epiblast can become primordial germ cells
- 301                  \* Otx2 is expressed in EpiLCs and is known to repress PGC identity. It's properly expressed in
- 302                  *Kdm5c*-KO EpiLCs, further supporting they aren't just becoming PGCs <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6485399/>
- 303                  \* Germline Program happening in the background of typical development.
- 304
- 305                  • One of the genes misexpressed in *Kdm5c*-KO EpiLCs is Dazl, *Deleted in azoospermia like (Dazl)*, a key
- 306                  regulator of germline development that promotes the translation of germline-specific transcripts[XXX].
- 307                  – 2 other studies of dazl regulators in ESCs also found KDM5C represses dazl.
- 308                  – While crucial for germ cell development, Dazl is also expressed in ESCs and at the 2-cell stage.
- 309                  – Some genes important for germ cell, ESC, and 2-cell development given they are important for
- 310                  self-renewal.
- 311
- 312                  – We did not observe dysregulation of 2-cell-specific regulators like Dux.
- 313                  – We found Dazl was expressed in *Kdm5c*-KO EpiLCs, when *Dazl* it typically decommissioned. This
- 314                  indicates KDM5C loss results in long-term ectopic expression.
- 315                  – We additionally found DAZL protein is ectopically expressed and localized to the cytoplasm. Given
- 316                  that a significant number of KDM5C-KO germline DEGs are DAZL targets, it may promote the
- 317                  ectopic germline program.
- 318                  – Very recently two other studies identified KDM5C represses Dazl in ESCs.
- 319                  – \* KDM5C was very recently shown to repress *Deleted in azoospermia like (Dazl)*, a key regulator
- 320                  of germline development, in mouse embryonic stem cells (ESCs)<sup>40,43</sup>. However, KDM5C
- 321                  binding and *Kdm5c*-KO germline gene misexpression has yet to be globally characterized
- 322                  during early embryogenesis. Given that *Dazl* and other germline-enriched genes can also
- 323                  be expressed in ESCs and at the 2-cell stage, it is unclear if KDM5C has a direct role in the
- 324                  long-term germline gene silencing that occurs in the post-implantation epiblast.
- 325                  • We globally characterized KDM5C's repression of germline genes during early embryogenesis and in
- 326                  mature neurons.
- 327                  – In line with previous work, we found KDM5C did not regulate germline genes in neurons.
- 328                  – KDM5C-bound genes in EpiLCs were enriched for germline ontologies, suggesting a major role of
- 329                  KDM5C during embryogenesis
- 330                  – While KDM5C directly binds some germline genes, including *Dazl*, many of the genes dysregulated
- 331                  in *Kdm5c*-KO were not direct targets.
- 332                  \* Stra8 notable exception

- 333            \* Brain genes, including late-stage spermatogenesis genes  
334            \* Suggests germline genes can be dysregulated direct and indirect of KDM5C regulation  
335            \* Further supports germline programs can be ectopically activated  
336            \* Talk about different motifs if we do see differences there and if it explains direct vs indirect  
337            dysregulation  
338                 · Could also mention PRC1.6/KDM5C binds pcgf6 in ESCs
- 339         • We demonstrated KDM5C is important for the transition between histone-mediated to DNAme-mediated  
340            silencing of germline genes during the transition from naïve to primed pluripotency.
- 341                 – Loss of DNAme can last throughout life at at least two germline gene promoters (hippocampus  
342                 barco)
- 343                 – KDM5C could be important for DNAme is that KDM5C erases H3K4me3 which can impede  
344                 CpGme.
- 345                 \* In support of this, KDM5C is highly enriched at the CpG islands near germline TSS that are  
346                 methylated in EpiLCs
- 347                 – However, Recently KDM5C's catalytic activity was found to be unnecessary for dazl suppression in  
348                 ESCs.
- 349
- 350                 – Since DNAme is not placed until EpiLC stage, KDM5C's catalytic activity may be required for  
351                 long-term silencing of germline genes.
- 352                 – This would be the first (?) example in which removal of an active mark is required for germline  
353                 gene repression.
- 354         • In summary, we demonstrate loss of the X-linked intellectual disability gene KDM5C leads to widespread  
355                 dysregulation of brain tissue identity, including misexpression of germline genes in the somatic brain.
- 356                 – In EpiLCs, KDM5C directly represses key drivers of germline identity like Dazl, likely through  
357                 PRC1.6 recruitment and promoting CpG methylation. However, *Kdm5c*-KO also ectopically  
358                 expresses germline genes activated indirectly, including *Stra8*.
- 359
- 360                 – The germline developmental program to some extent continues ectopically during *Kdm5c*-KO  
361                 development, resulting in aberrant transcription of late stage spermatogenesis genes later in life.
- 362                 – These results define KDM5C's role in the demarcation between soma and germline identity and  
363                 offer a window into potential targets to assess the deleterious effects these ectopic genes on  
364                 neurodevelopment.

365 **Discussion notes**

- 366     • Our results indicate KDM5C is a pivotal repressor of germline programs in somatic cells.
  - 367     • Studies on other chromatin regulators have focused on key marker genes to identify germline gene  
368       misexpression, such as *Dazl*.
  - 369     • Using a germline-depleted mouse model, we curated a list of mouse germline-enriched genes to  
370       globally assess germline gene dysregulation.
  - 371     • Many of the genes dysregulated in the mature *Kdm5c*-KO brain were important for late stages of  
372       spermatogenesis, such as those important for sperm flagellar structure. Contrastingly, *Kdm5c*-KO  
373       EpiLCs aberrantly expressed key drivers of germ cell identity, including *Dazl* and *Stra8*.
  - 374     • While a significant portion of KDM5C-bound promoters are germline genes, many germline genes  
375       expressed during *Kdm5c*-KO embryogenesis are not directly bound by *kdm5c*.
  - 376     • One notable example is *Stra8*, a transcription factor important for germ cell specific transcription and  
377       meiotic initiation
  - 378     •
- 379     While Directly binds to a subset of germline genes during early embryogenesis - KDM5C binds to a  
380     subset of germline genes during early embryogenesis
- 381     • The including the demarcation between soma and germline fates.
  - 382       the *Kdm5c*-KO brain also Previous work identified testis-enriched genes in *Kdm5c* genes
  - 383       –
  - 384       – However unlike the gonadal-biased DEGs,
- 385     • The demarcation of the germline versus soma is a key feature of multicellularity and sexually dimorphic  
386       reproduction
- 387     • Anything known about tissue-biased gene expression in other H3K4me regulators?
- 388     • Our data suggests the germline developmental program is occurring ectopically as *Kdm5c*-KOs pro-  
389       gresses through somatic tissue development
- 390     • tissue-biased gene expression:
- 391     • Altogether, these data indicate that while some germline genes are misexpressed due to direct loss of  
392       KDM5C binding during embryogenesis, secondary downstream mechanisms can also promote their  
393       aberrant transcription.

- 394 – This shift from meiotic genes to later spermatogenesis genes in the hippocampus suggests the  
395 germline developmental program could occur ectopically as *Kdm5c*-KO cells progresses through  
396 somatic tissue development. **note: this is strengthened by the ChIP-seq data since KDM5C**  
397 **is not directly bound to many brain/flagellar DEGs. This point might be stronger in the**  
398 **ChIPseq figure**
- 399 • KDM5C is dynamically regulated during the window of embryonic germline gene silencing, our EpiLC  
400 ChIP-seq is likely catching the tail end of KDM5C's main involvement.
- 401 • Papers to read/reference:
- 402 – Reconstitution of the Mouse Germ Cell Specification Pathway in Culture by Pluripotent Stem Cells:  
403 [https://www.cell.com/fulltext/S0092-8674\(11\)00771-9](https://www.cell.com/fulltext/S0092-8674(11)00771-9)
- 404 – 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/>

## 406 References

- 407 1. Strahl, B.D., and Allis, C.D. (2000). The language of covalent histone modifications. *Nature* **403**,  
408 41–45. <https://doi.org/10.1038/47412>.
- 409 2. Jenuwein, T., and Allis, C.D. (2001). Translating the Histone Code. *Science* **293**, 1074–1080.  
410 <https://doi.org/10.1126/science.1063127>.
- 411 3. Lewis, E.B. (1978). A gene complex controlling segmentation in Drosophila. *Nature* **276**, 565–570.  
412 <https://doi.org/10.1038/276565a0>.
- 413 4. Kennison, J.A., and Tamkun, J.W. (1988). Dosage-dependent modifiers of polycomb and antennapedia  
414 mutations in Drosophila. *Proc. Natl. Acad. Sci. U.S.A.* **85**, 8136–8140. <https://doi.org/10.1073/pnas.85.21.8136>.
- 415 5. Gabriele, M., Lopez Tobon, A., D'Agostino, G., and Testa, G. (2018). The chromatin basis of  
416 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>.
- 417 6. Zhou, Z., Hong, E.J., Cohen, S., Zhao, W.-N., Ho, H.-Y.H., Schmidt, L., Chen, W.G., Lin, Y., Savner,  
418 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>.
- 419 7. Hirabayashi, Y., Suzuki, N., Tsuboi, M., Endo, T.A., Toyoda, T., Shinga, J., Koseki, H., Vidal, M., and  
420 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>.

- 421 8. Kassis, J.A., Kennison, J.A., and Tamkun, J.W. (2017). Polycomb and Trithorax Group Genes in  
422 *Drosophila*. *Genetics* **206**, 1699–1725. <https://doi.org/10.1534/genetics.115.185116>.
- 423 9. Schaefer, A., Sampath, S.C., Intrator, A., Min, A., Gertler, T.S., Surmeier, D.J., Tarakhovsky, A.,  
424 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>.
- 425 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**,  
426 47–59. <https://doi.org/10.1016/j.celrep.2017.09.014>.
- 427 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>.
- 428 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 paraparesis. *Am J  
430 Med Genet* **94**, 1–4.
- 429 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  
432 Genet* **76**, 227–236. <https://doi.org/10.1086/427563>.
- 433 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.  
434 <https://doi.org/10.1111/cge.13755>.
- 435 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  
Impaired Removal of Histone Methylation. *Cell Reports* **14**, 1000–1009. <https://doi.org/10.1016/j.celrep.2015.12.091>.
- 436 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  
438 stem cells by regulating germ cell-related genes. *eLife* **6**. <https://doi.org/10.7554/eLife.21064>.
- 439 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>.

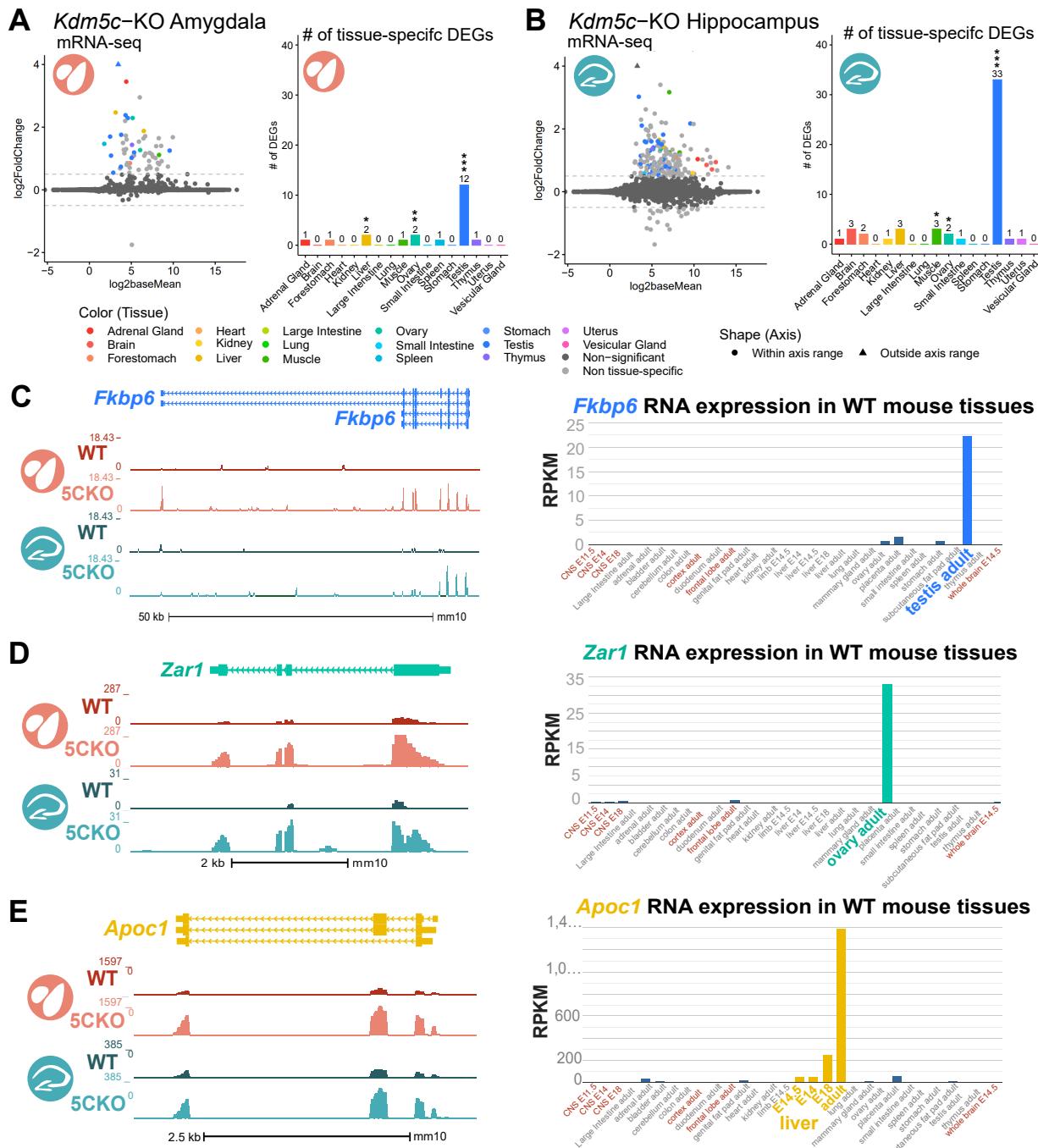
- 440
- 441 18. Velasco, G., Hubé, F., Rollin, J., Neuillet, D., Philippe, C., Bouzinba-Segard, H., Galvani, A., Viegas-Péquignot, E., and Francastel, C. (2010). Dnmt3b recruitment through E2F6 transcriptional repressor mediates germ-line gene silencing in murine somatic tissues. *Proc Natl Acad Sci U S A* **107**, 9281–9286. <https://doi.org/10.1073/pnas.1000473107>.
- 442
- 443 19. Hackett, J.A., Reddington, J.P., Nestor, C.E., Duncan, 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>.
- 444
- 445 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>.
- 446
- 447 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>.
- 448
- 449 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>.
- 450
- 451 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>.
- 452
- 453 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>.
- 454
- 455 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>.
- 456
- 457 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>.
- 458
- 459 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>.
- 460

- 461 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>.
- 462 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>.
- 463 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>.
- 464 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>.
- 465 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>.
- 466 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>.
- 467 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>.
- 468 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>.
- 469 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>.
- 470 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>.
- 471 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>.

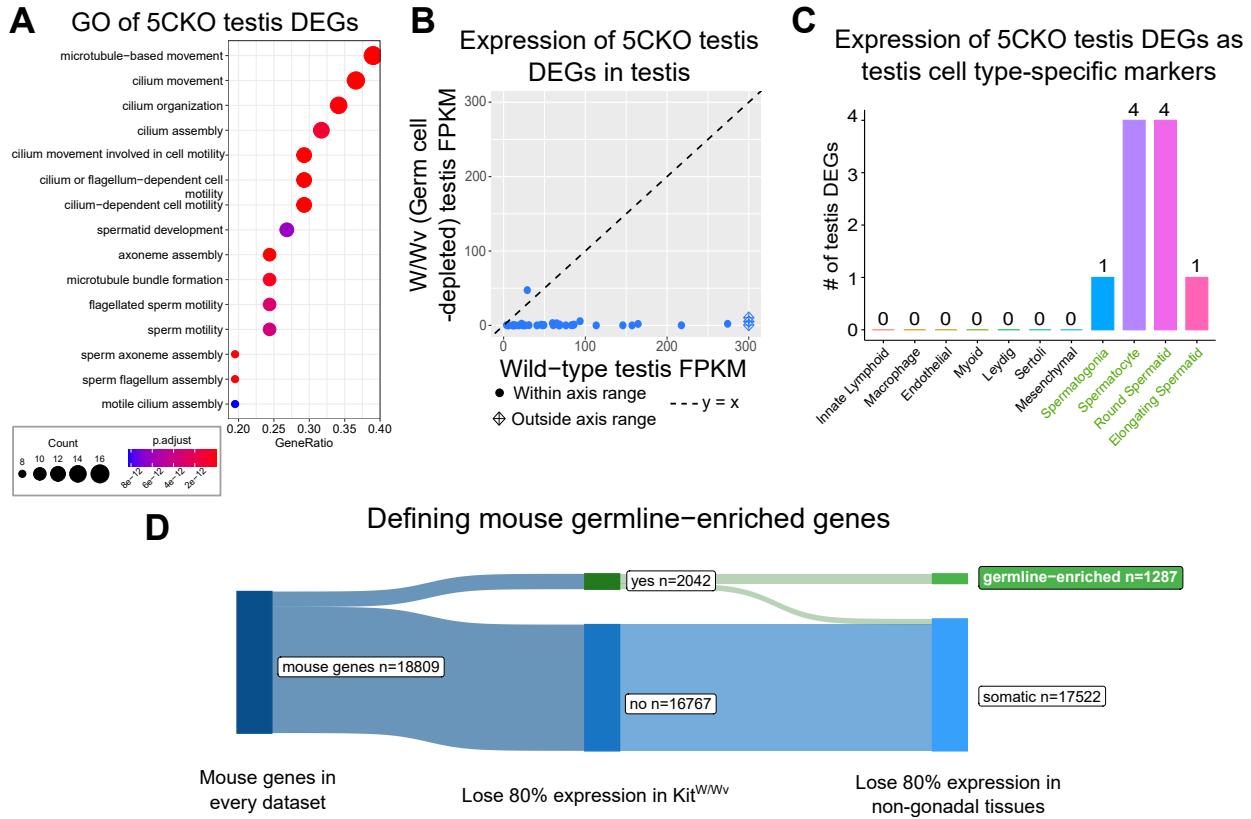
- 483 39. Endo, T., Mikedis, M.M., Nicholls, P.K., Page, D.C., and De Rooij, D.G. (2019). Retinoic Acid and Germ  
484 Cell Development in the Ovary and Testis. *Biomolecules* *9*, 775. <https://doi.org/10.3390/biom9120775>.
- 485 40. Gupta, N., Yakhou, L., Albert, J.R., Azogui, A., Ferry, L., Kirsh, O., Miura, F., Battault, S., Yamaguchi,  
486 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>.
- 487 41. Li, H., Liang, Z., Yang, J., Wang, D., Wang, H., Zhu, M., Geng, B., and Xu, E.Y. (2019). DAZL  
488 is a master translational regulator of murine spermatogenesis. *Natl Sci Rev* *6*, 455–468. <https://doi.org/10.1093/nsr/nwy163>.
- 489 42. Mikedis, M.M., Fan, Y., Nicholls, P.K., Endo, T., Jackson, E.K., Cobb, S.A., De Rooij, D.G., and Page,  
490 D.C. (2020). DAZL mediates a broad translational program regulating expansion and differentiation of  
491 spermatogonial progenitors. *eLife* *9*, e56523. <https://doi.org/10.7554/eLife.56523>.
- 492 43. Al Adhami, H., Vallet, J., Schaal, C., Schumacher, P., Bardet, A.F., Dumas, M., Chicher, J., Hammann,  
493 P., Daujat, S., and Weber, M. (2023). Systematic identification of factors involved in the silencing  
494 of germline genes in mouse embryonic stem cells. *Nucleic Acids Research* *51*, 3130–3149. <https://doi.org/10.1093/nar/gkad071>.
- 495 44. Liu, M., Zhu, Y., Xing, F., Liu, S., Xia, Y., Jiang, Q., and Qin, J. (2020). The polycomb group protein  
496 PCGF6 mediates germline gene silencing by recruiting histone-modifying proteins to target gene  
497 promoters. *J Biol Chem* *295*, 9712–9724. <https://doi.org/10.1074/jbc.RA119.012121>.
- 498 45. Otani, J., Nankumo, T., Arita, K., Inamoto, S., Ariyoshi, M., and Shirakawa, M. (2009). Structural basis  
499 for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX–DNMT3–DNMT3L  
500 domain. *EMBO Reports* *10*, 1235–1241. <https://doi.org/10.1038/embor.2009.218>.
- 501 46. Guo, X., Wang, L., Li, J., Ding, Z., Xiao, J., Yin, X., He, S., Shi, P., Dong, L., Li, G., et al. (2015).  
502 Structural insight into autoinhibition and histone H3-induced activation of DNMT3A. *Nature* *517*,  
640–644. <https://doi.org/10.1038/nature13899>.
- 503 47. Abildayeva, K., Berbée, J.F.P., Blokland, A., Jansen, P.J., Hoek, F.J., Meijer, O., Lütjohann, D., Gautier,  
504 T., Pillot, T., De Vente, J., et al. (2008). Human apolipoprotein C-I expression in mice impairs learning  
505 and memory functions. *Journal of Lipid Research* *49*, 856–869. <https://doi.org/10.1194/jlr.M700518JLR200>.
- 506 48. Leduc, V., Jasmin-Bélanger, S., and Poirier, J. (2010). APOE and cholesterol homeostasis in  
507 Alzheimer's disease. *Trends in Molecular Medicine* *16*, 469–477. <https://doi.org/10.1016/j.molmed.2010.07.008>.

503 **Figures and Tables**

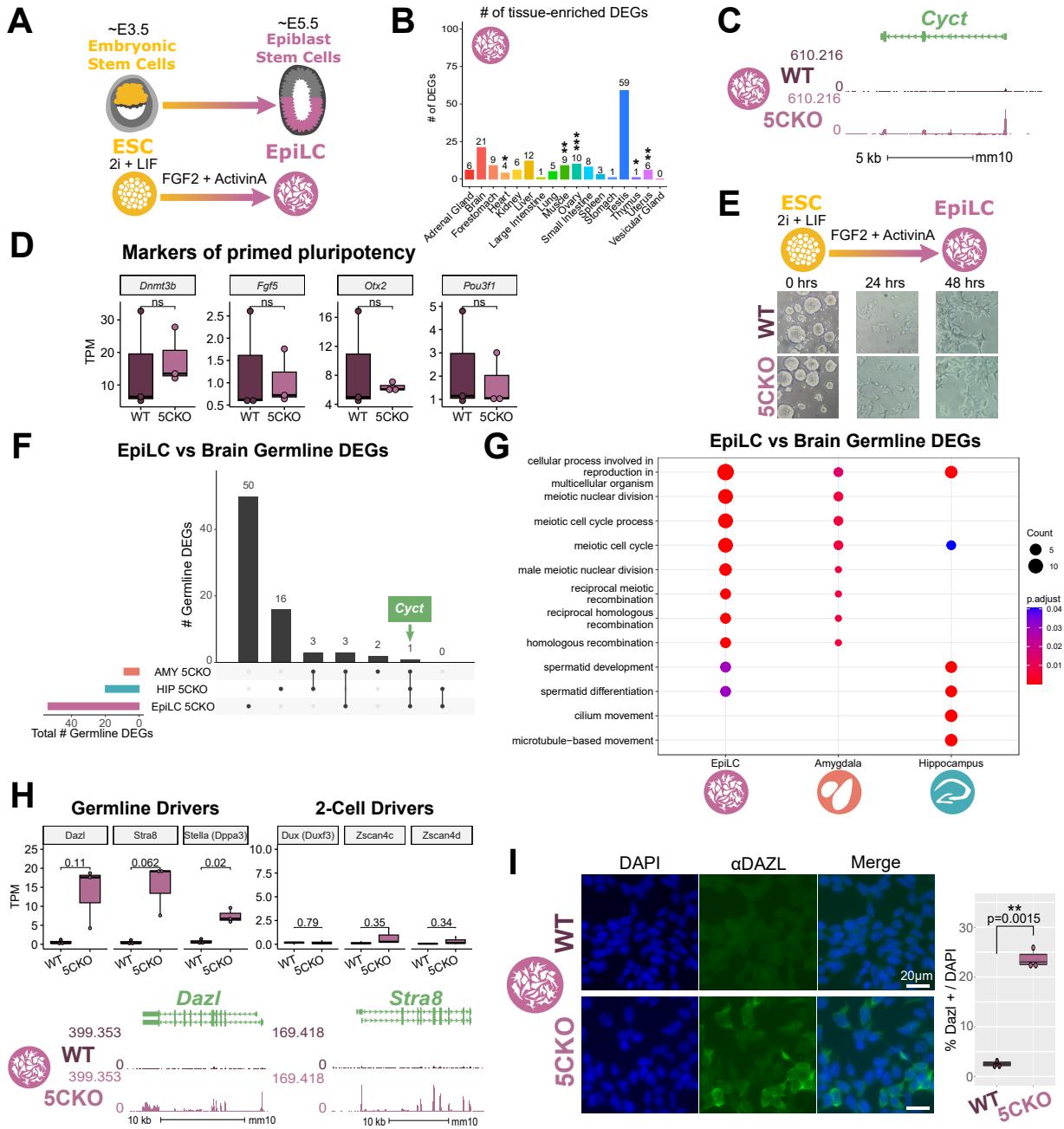
- 504     • Supplementary table 1: list of all germline genes.
- 505       – Columns to include:
- 506           \* KDM5C bound vs not
- 507           \* DEG in EpiLC, brain, both, neither (separate columns?)



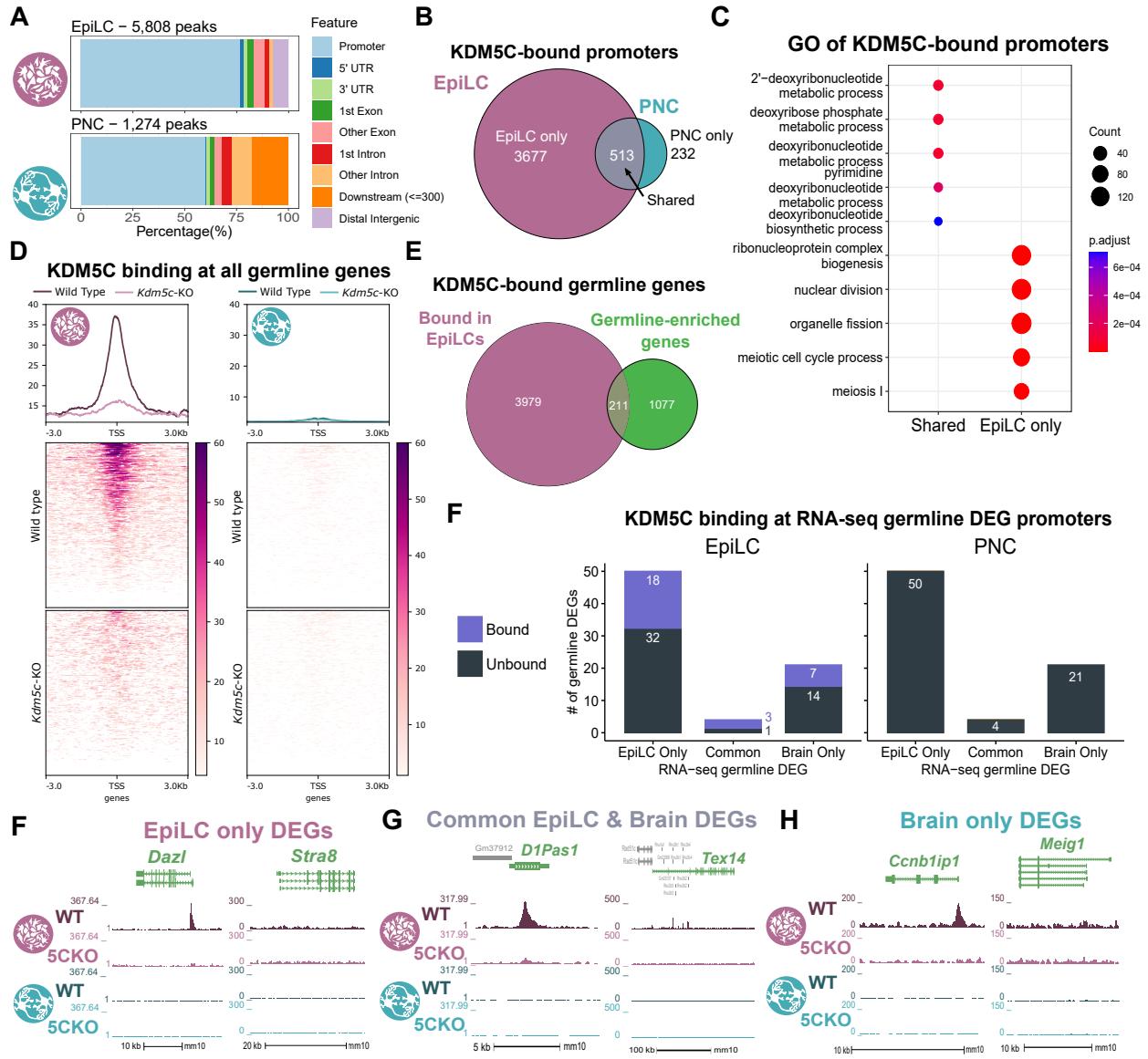
**Figure 1: Tissue-enriched genes are misexpressed in the *Kdm5c*-KO brain.** **A-B.** Expression of tissue-enriched genes (Li et al 2017) in the male *Kdm5c*-KO amygdala (A) and hippocampus (B). Left - MA plot of mRNA-sequencing. Right - Number of tissue-enriched differentially expressed genes (DEGs).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

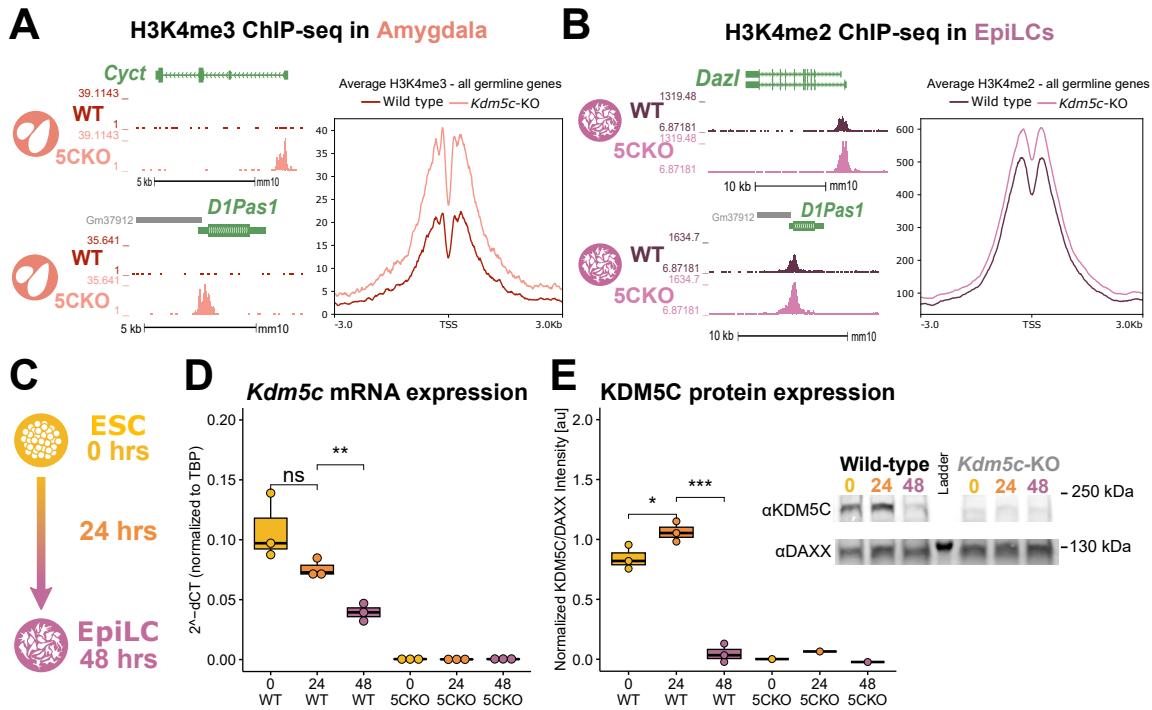


**Figure 3: Kdm5c-KO epiblast-like cells express key drivers of germline identity** **A.** Top - Diagram of *in vivo* differentiation of embryonic stem cells (ESCs) of the inner cell mass into epiblast stem cells. Bottom - *in vitro* differentiation of ESCs into epiblast-like cells (EpiLCs). **B.** Number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

508 **Notes**

509 **Figure outline:**

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

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

522     **Figure 3: *Kdm5c*-KO epiblast-like cells express key drivers regulators of germline identity** \* A) ESC  
523     to EpiLC differentiation Left - Morphology is unchanged, \* B) 5CKO EpiLCs express EpiLC differentiation  
524     genes similar to WT lvs \* C) Male EpiLCs express germline genes (example Cyct again) \* Overlap between  
525     brain and EpiLC germline genes - show they're mostly unique \* GO of Brain and EpiLC germline genes  
526     (meiotic enriched) \* Bigwigs or TPM of master regulators \* Show that while some are also 2-cell genes  
527     (Dazl), 2-cell specific genes aren't dysregulated (Zscan4). Important point because published KDM5C dazl  
528     paper is saying KDM5C is a 2-cell regulator, but as far as I can tell only genes shared between germline and  
529     2-cell are dysregulated.

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

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

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

540     Gaps in knowledge addressed: \* Are other tissue-enriched genes dysregulated, or only testis, germline  
541     genes? \* Curating a robust list of male and female germline genes \* Should talk about 2-cell genes  
542     vs germline genes - way to systematically categorize? \* Mechanism behind long-term germline gene

543 misexpression \* Recent evidence suggests loss of KDM5C in ESCs express some germline genes \* Unclear  
544 if catalytic activity is required for long-term silencing \* Unclear if their dysregulation lasts throughout life or  
545 the same between brain or not \* When in development does it begin? - Recent evidence suggests some  
546 germline genes expressed in 5CKO ESCs but unclear if their dysregulation lasts throughout differentiation  
547 and if the identity of germline genes are different compared to the brain \* Are there functional consequences  
548 to germline gene misexpression?

549      Introduction: \* Chromatin regulators are important for cellular identity \* H3K4me1-3 linked to active  
550 gene promoters and enhancers \* Surprisingly, mutations in Chromatin regulators lead to many NDDs  
551 (including many H3K4 regulators) \* Recent studies have shown some chromatin regulators are important  
552 for regulating neuron-specific gene expression/chromatin stat\_compare\_means \* However, loss of some  
553 chromatin regulators can also lead to ectopic expression of tissue-enriched genes \* Very few studies have  
554 looked at these genes and it's unclear if these genes contribute to NDD impairments. \* Necessary to  
555 first characterize the mechanism behind their derepression to identify molecular footholds into testing their  
556 contribution to neuronal impairments and potential for therapeutic intervention

- 557      • Loss of KDM5C can result in the misexpression of genes typically only found in the testis
  - 558           – Misexpression of tissue-enriched genes hasn't been systematically characterized - Unclear if  
559           these genes are exceptions or if other tissue-specific genes are dysregulated
  - 560           – Interestingly, these genes (Cyct, D1pas1) typically function in the germline
  - 561           – Germ cells (meiotic cells) are typically distinguished from somatic cells very early on in embryogenesis and is a key feature of multicellularity
  - 563           – Chromatin regulators are very important for decommissioning germline genes and act successively  
564           the embryo implants into the uterine wall
    - 565              \* Most studies have focused on ESCs, which have a similar transcriptome to germ cells / 2-cells
    - 566              \* recently, KDM5C was shown to repress DAZL in ESCs, independent of its catalytic activity
    - 567              \* However, DNA methylation is lost in the mature 5CKO brain, DNA methylation is placed later  
568              and it's Unclear if it's required for long-term repression (maybe too specific, just trying to go  
569              into the fact that the mechanism is partially understood but unclear)
  - 570           – Systematic characterization of ectopic germline genes hasn't been done
    - 571              \* unknown if other germline-enriched genes are dysregulated, including oocyte-specific genes
    - 572              \* Crucially, it's unknown if misexpression of the germline program leads to functional conse-  
573              quences in 5CKO cells.

574      **Germline gene repression background:**

575      Interestingly, some of the ectopic testis transcripts identified in the *Kdm5c*-KO brain are typically ex-  
576 pressed in germ cells<sup>10</sup>. Unlike somatic cells, germ cells (e.g. sperm and eggs) undergo meiosis and pass  
577 on their genetic material to the next generation. The germline and the soma are typically distinguished during

578 early embryogenesis, when germline genes are silenced in epiblast stem cells soon after implantation and  
579 only reactivated in a subset to form the germline. Chromatin regulators play a key role in decommissioning  
580 germline genes as the embryo transitions from naïve to primed pluripotency by placing repressive histone  
581 H2A lysine 119 monoubiquitination (H2AK119ub1)<sup>16</sup>, histone 3 lysine 9 trimethylation (H3K9me3)<sup>16,17</sup>, and  
582 DNA CpG methylation<sup>17–19</sup> at germline gene promoters. KDM5C may also be involved in this early decom-  
583 missioning of germline genes, as re-expression of KDM5C in neurons fails to suppress their dysregulation<sup>10</sup>.  
584 In support of this, KDM5C was very recently shown to repress *Deleted in azoospermia like (Dazl)*, a key  
585 regulator of germline development, in mouse embryonic stem cells (ESCs)<sup>40,43</sup>. In support of this, two  
586 independent screens in mouse embryonic stem cells (ESCs) recently identified KDM5C as a repressor of  
587 *Deleted in azoospermia like (Dazl)*, a key regulator of germline development. However, KDM5C's role in  
588 embryonic germline gene repression is currently unclear, given that Dazl is also expressed in ESCs and in  
589 the 2-cell stage and germline gene misexpression has yet to be globally characterized during *Kdm5c*-KO  
590 embryogenesis.