

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

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

28 Intriguingly, some of the ectopic testis transcripts identified within the *Kdm5c*-KO hippocampus have known functions  
29 unique to germ cells<sup>10</sup>, suggesting KDM5C may play a role in demarcating somatic versus germline identity. Distinguishing  
30 between the germline (cells that pass on their genetic material, e.g. sperm and eggs) and the soma (cells that perform all other  
31 bodily functions) is a key feature of multicellularity and occurs during early embryogenesis. In mammals, chromatin regulators  
32 play a key role in decommissioning germline genes in somatic cells as the embryo implants into the uterine wall and transitions  
33 from naive to primed pluripotency by placing repressive histone H2A lysine 119 monoubiquitination (H2AK119ub1)<sup>16</sup>, histone

34 3 lysine 9 trimethylation (H3K9me3)<sup>16,17</sup>, and DNA CpG methylation<sup>17-19</sup> at germline gene promoters. KDM5C may also be  
35 involved in this early decommissioning of germline genes, as re-expression of KDM5C in knockout neurons fails to suppress  
36 their dysregulation<sup>10</sup>. In support of this, KDM5C was very recently shown to repress *Deleted in azoospermia like (Dazl)*, a key  
37 regulator of germline development, in mouse embryonic stem cells (ESCs)<sup>20,21</sup>. However, KDM5C binding and *Kdm5c*-KO  
38 germline gene misexpression has yet to be globally characterized during early embryogenesis. Given that *Dazl* and other  
39 germline-enriched genes can also be expressed in ESCs and at the 2-cell stage, it is unclear if KDM5C has a direct role in the  
40 long-term germline gene silencing that occurs in the post-implantation epiblast. Systematically characterizing KDM5C's role  
41 in germline gene repression during early embryogenesis will unveil key mechanisms underlying the demarcation between  
42 soma and germline identity and while also providing molecular footholds to test the impact of ectopic germline genes on  
43 neurodevelopment.

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

## 56 Results

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

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

59 Previous RNA sequencing (RNA-seq) in the adult hippocampus revealed ectopic expression of some testis genes within  
60 the *Kdm5c* knockout (-KO) brain<sup>10</sup>. Since tissue-enriched genes have not been systematically characterized in the *Kdm5c*-KO  
61 brain, it is currently unclear if this erosion of brain tissue identity is a major consequence of *Kdm5c* loss and if it is unique to  
62 testis-enriched genes. Therefore, we first globally assessed the expression of genes enriched in 17 mouse tissues<sup>22</sup> in our  
63 published mRNA-seq datasets of the adult amygdala and hippocampus in mice with constitutive knockout of *Kdm5c*<sup>23</sup>.

64 We found a large proportion of genes that are significantly upregulated within the *Kdm5c*-KO brain (DESeq2<sup>24</sup>, log2  
65 fold change > 0.5, q < 0.1) are typically enriched in non-brain tissues (Amygdala: 35%, Hippocampus: 24%) (Figure 1A-B).  
66 The majority of tissue-enriched differentially expressed genes (tissue-enriched DEGs) were testis genes (Figure 1A-C).  
67 Even though the testis has the largest total number of tissue-biased genes compared to any other tissue (2,496 genes),  
68 testis-biased DEGs were significantly enriched for both brain regions (Amygdala p = 1.83e-05; Hippocampus p = 4.26e-11,  
69 Fisher's Exact Test). One example of a testis-enriched DEG is *FK506 binding protein 6 (Fkbp6)*, a known regulator of piRNA  
70 expression and meiosis in germ cells<sup>25,26</sup> (Figure 1C).

71 In addition to the high enrichment of testis genes, we also identified aberrant expression of other tissue-enriched genes  
72 within the *Kdm5c*-KO brain. Although the *Kdm5c*-KO mice were male, we observed significant enrichment of ovary-biased  
73 DEGs in both the amygdala and hippocampus (Amygdala p = 0.00574; Hippocampus p = 0.048, Fisher's Exact) (Figure  
74 1D). Ovary-enriched DEGs included *Zygotic arrest 1 (Zar1)*, which was recently shown to sequester mRNAs in oocytes  
75 for meiotic maturation and early zygote development<sup>27</sup> (Figure 1D). Although not consistent across brain regions, we also  
76 found significant enrichment of DEGS biased towards two non-gonadal tissues - the liver (Amygdala p = 0.0398, Fisher's  
77 Exact Test) and the muscle (Hippocampus p = 0.0104, Fisher's Exact Test). An example of a liver-biased DEG dysregulated  
78 in both the hippocampus and amygdala is *Apolipoprotein C-I (Apoc1)*, which is involved in lipoprotein metabolism (Figure  
79 1E). Testis, ovary, and liver-enriched DEGs showed little to no expression in the developing and adult wild-type brain, yet  
80 our mRNA-seq data indicate they are polyadenylated and spliced into mature transcripts (Figure 1C-E). Of note, we did not  
81 observe enrichment of brain-enriched genes (Amygdala p = 1; Hippocampus p = 0.74, Fisher's Exact), despite the fact these  
82 are brain samples and the brain has the second highest total number of tissue-enriched genes (708 genes). Together, these  
83 results suggest misexpression of tissue-enriched genes within the brain is a major effect of KDM5C loss.

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

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

91 To further validate if these testis DEGs are truly germline genes, we then evaluated their expression in somatic versus germ  
92 cells within the testis. We first compared their expression in the testis with germ cell depletion<sup>28</sup>, which was accomplished by  
93 heterozygous *W* and *Wv* mutations in the enzymatic domain of *c-Kit* (*Kit*<sup>W/Wv</sup>) that prevent the maturation of germ cells<sup>29</sup>.  
94 Almost all *Kdm5c*-KO testis-enriched DEGs lost expression with germ cell depletion (Figure 2B). The only testis-enriched  
95 DEG that did not show considerable downregulation with germline depletion was *FK506 binding protein 6 (Fkbp6)*, the  
96 aforementioned testis gene that regulates piRNA expression and meiosis in germ cells<sup>25,26</sup>. We then assessed testis-  
97 enriched DEG expression in a published single cell RNA-seq dataset that identified cell type-specific markers within the  
98 testis<sup>30</sup>. We found that while some testis-enriched DEGs were classified as specific markers for different germ cell types  
99 (e.g. spermatogonia, spermatocytes, round spermatids, and elongating spermatids) none marked somatic cells (Figure 2C).  
100 Together, these data suggest the somatic *Kdm5c*-KO brain aberrantly expresses germline genes.

101 We then wanted to more deeply characterize germline gene misexpression with *Kdm5c* loss, but lacked a comprehensive  
102 list of mouse germline-enriched genes. To facilitate downstream analyses, we generated a curated list of germline-enriched  
103 genes using currently available RNA-seq datasets in *Kit*<sup>W/Wv</sup> mice. Wild-type and *Kit*<sup>W/Wv</sup> datasets included males and females  
104 at embryonic day 12, 14, and 16<sup>31</sup>, as well as adult male testes<sup>28</sup>. We defined genes as germline-enriched if their expression  
105 met the following criteria: 1) their expression is greater than 1 FPKM in wild-type gonads 2) their expression in any adult  
106 wild-type, non-gonadal tissue<sup>22</sup> does not exceed 20% of their maximum expression in the wild-type germline, and 3) their  
107 expression in the germ cell-depleted germline, for any sex or time point, does not exceed 20% of their maximum expression  
108 in the wild-type germline. These criteria yielded 1,288 germline-enriched genes (Figure 2D), which was hereafter used as a  
109 resource for assessing germline gene misexpression with *Kdm5c* loss (Supplementary table 1).

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

111 Misexpression of germline genes in the adult brain suggests *Kdm5c*-KO cells fail to demarcate between germline and  
112 somatic identity. Germ cells are typically distinguished from somatic cells soon after the embryo implants into the uterine  
113 wall<sup>32,33</sup> when a subset of epiblast stem cells become the primordial germ cells (PGCs) while the remainder differentiate into  
114 the ectoderm, mesoderm, and endoderm to form the somatic tissues<sup>34</sup>. This developmental time point can be modeled *in*  
115 *vitro* through differentiation of embryonic stem cells (ESCs) into post-implantation epiblast-like stem cells (EpiLCs) (Figure 3A,  
116 top). Previous studies have demonstrated that while some germline-enriched genes are also expressed in embryonic stem  
117 cells (ESCs) and in the 2-cell stage<sup>35-37</sup>, they are silenced as they differentiate into EpiLCs<sup>17</sup>. Therefore, we assessed if  
118 KDM5C was necessary for embryonic silencing of germline genes by evaluating the impact of *Kdm5c* loss in EpiLCs.

119 We first identified *Kdm5c*-KO EpiLC DEGs in of our previously published RNA-seq dataset<sup>38</sup> (DESeq2, log2 fold change  
120 > 0.5, q < 0.1) and assessed tissue-enriched gene expression. Similar to the *Kdm5c*-KO brain, we observed general  
121 dysregulation of tissue-enriched genes, with the largest number of genes belonging to the brain and testis, although they were  
122 not significantly enriched (Figure 3B). Using the curated list of germline genes generated above, we found *Kdm5c*-KO EpiLCs  
123 aberrantly expressed 54 germline-enriched genes, including the previously characterized hippocampal DEG<sup>10</sup> *Cytochrome*  
124 *C, testis-specific (Cyct)* (Figure 3C). Although we observed aberrant expression of many tissue-enriched genes, we did not  
125 observe any significant difference in primed pluripotency genes (Figure 3D) or gross changes in *Kdm5c*-KO cell morpholgy  
126 during differentiation (Figure 3E), indicating KDM5C loss does not impair EpiLC formation.

127 We then compared EpiLC germline DEGs to those expressed in the *Kdm5c*-KO brain to determine if all germline DEGs,  
128 like *Cyct*, are constitutively dysregulated or if they can change over the course of development. We found this was primarily  
129 not the case, as the majority of germline DEGs expressed in EpiLCs and the brain were unique, with only *Cyct* shared across  
130 all sequencing datasets (Figure 3F). We then compared the known functions of EpiLC and brain germline DEGs and found  
131 particularly high enrichment of meiosis-related gene ontologies in EpiLCs (Figure 3G), such as meiotic cell cycle process (GO:  
132 1903046, p.adjust = 1.59e-08) and meiotic nuclear division (GO: 0140013, p.adjust = 9.76e-09). While there was modest  
133 enrichment of meiotic gene ontologies in both brain regions, the *Kdm5c*-KO hippocampus primarily expressed late-stage  
134 spermatogenesis genes, such those involved in the sperm axoneme structure.

135 Interestingly, DEGs unique to *Kdm5c*-KO EpiLCs included key drivers of germline identity, such as *Stimulated by retinoic*  
136 *acid 8 (Stra8)* and *Deleted in azoospermia like (Dazl)* (Figure 3H). These genes are typically expressed during embryonic  
137 germ cell development to commit PGCs to the germline fate, but are also expressed later in life to trigger meiotic gene  
138 expression programs<sup>39-41</sup>. Of note, some germline genes, including *Dazl*, are also expressed in the two-cell embryo<sup>20,36</sup>.  
139 However, we did not see misexpression of two-cell embryo-specific genes, like *Duxf3 (Dux)* (q = 0.337) and *Zscan4d* (q =  
140 0.381), indicating *Kdm5c*-KO in EpiLCs do not revert back to a 2-cell state (Figure 3H).

141 We were particularly interested in the aberrant transcription of *Dazl*, since it is essential for germ cell development and  
142 promotes the translation of germline mRNAs<sup>42</sup>. A significant portion of germline transcripts misexpressed in *Kdm5c*-KO  
143 EpiLCs are known binding targets of DAZL, including *Stra8*<sup>43</sup> (p = 1.698e-07, Fisher's Exact Test). This suggests expression  
144 of DAZL protein could promote the translation of other aberrant germline transcripts, influencing their ability to impact  
145 *Kdm5c*-KO cellular function. We thus tested DAZL protein expression in *Kdm5c*-KO EpiLCs through immunocytochemistry  
146 (Figure 3I). We observed about 25% of *Kdm5c*-KO EpiLCs expressed DAZL protein and it was localized to the cytoplasm (p =  
147 0.0015, Welch's t-test), consistent with the pattern of DAZL expression in spermatogonia<sup>43</sup>. Altogether these results suggest  
148 tissue-specific genes are misexpressed during *Kdm5c*-KO embryogenesis, including key drivers of germline identity that can  
149 be translated into protein.

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

- 151 • note: do Direct vs indirect DEGs motif analysis  
152 • However, it is currently unclear if KDM5C binds to all germline DEGs and if its binding is maintained at any germline  
153 genes in neurons.

154 Previous work suggests KDM5C represses germline genes during early development, as re-expression of KDM5C in  
155 knockout neuronal cultures fails to suppress hippocampal germline DEGs and KDM5C is not bound to their promoters  
156 in neurons<sup>10</sup>. There is some evidence KDM5C binds to select germline gene promoters in ESCs<sup>10</sup>, including two recent  
157 independent screens that found KDM5C binds to Dazl's promoter<sup>20,21</sup>. As KDM5C's binding at germline gene promoters has  
158 not been systematically characterized, it is currently unclear what types of germline genes KDM5C regulates and if its binding  
159 is maintained at any germline genes in neurons.

160 To further characterize the mechanism behind *Kdm5c*-KO germline gene misexpression, we analyzed KDM5C chromatin  
161 immunoprecipitation followed by DNA sequencing (ChIP-seq) datasets in EpiLCs<sup>38</sup> and primary neuron cultures (PNCs) from  
162 the cortex and hippocampus<sup>15</sup>. EpiLCs had a higher total number of KDM5C peaks than PNCs (EpiLCs: 5,808, PNCs: 1,276,  
163 MACS2 q < 0.1 and fold enrichment > 1) and KDM5C was primarily localized to gene promoters in both cell types (EpiLCs:  
164 4,190, PNCs: 745 +/- 500kb from TSS), although PNCs showed increased localization to non-promoter regions (Figure 4A).

165 The majority of promoters bound by KDM5C in PNCs were also bound in EpiLCs (513 shared promoters), however a  
166 large portion of gene promoters were only bound by KDM5C in EpiLCs (3677 EpiLC only promoters) (Figure 4B). We then  
167 performed gene ontology analysis to compare the functions of genes bound by KDM5C at the promoter in different cell  
168 types. While there were no significant ontologies for genes with KDM5C only bound in PNCs, gene ontologies for peaks  
169 shared between PNCs and EpiLCs were enriched for functions involving nucleic acid turnover, such as deoxyribonucleotide  
170 metabolic process (GO:0009262, p.adjust = 8.28e-05) (Figure 4C). Germline-specific ontologies were only enriched in  
171 peaks unique to EpiLCs, such as meiotic nuclear division (GO: 0007127 p.adjust = 6.77e-16) and meiotic cell cycle process  
172 (GO:1903046, p.adjust = 5.05e-16) (Figure 3C). When comparing KDM5C binding at all germline gene promoters, KDM5C  
173 was only bound to a subset of germline gene promoters in EpiLCs and was not bound to any in PNCs (Figure 4D). Together,  
174 this suggests KDM5C is recruited to a subset of germline gene promoters in EpiLCs, including meiotic genes, but does not  
175 regulate germline genes in neurons.

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

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

186 Although KDM5C is generally thought to suppress transcription through erasure of histone 3 lysine 4 di- and trimethylation  
187 (H3K4me2/3)<sup>11</sup>, recent studies in ESCs have suggested KDM5C's repression *Dazl* is independent of its catalytic activity<sup>20</sup>.

188 Somatic repression of germline genes is typically established during the transition between naive and primed pluripotency,  
189 which modeled in vitro as ESCs to EpiLC differentiation. In ESCs, chromatin regulators place repressive histone modifications  
190 at germline gene promoters, including histone 2A 119 monoubiquitination (H2AK119ub1) and histone 3 lysine 9 trimethylation  
191 (H3K9me3)<sup>16,17,44</sup>. Germline genes are then silenced long-term in EpiLCs by *de novo* placement of DNA CpG methylation<sup>17</sup>.  
192 It has been proposed KDM5C may promote germline gene silencing via H3K4me3 removal since DNA methylation is lost at  
193 select germline gene promoters in the hippocampus<sup>10</sup> and H3K4me3 can impair DNA methylation placement<sup>45,46</sup>. However,  
194 KDM5C was recently shown to repress *Dazl* in ESCs independent of its catalytic activity<sup>20</sup>. Because KDM5C's role in germline  
195 gene repression has only been characterized in ESCs and in the mature brain, it is currently unclear to what extent KDM5C is  
196 involved during transition between ESCs and EpiLCs and if its catalytic activity is required for long-term silencing.

197 To elucidate KDM5C's role in germline gene silencing, we first characterized KDM5C's substrates, histone 3 lysine 4  
198 di- and trimethylation (H3K4me2/3) in our previously published ChIP-seq datasets in wild type and *Kdm5c*-KO amygdala<sup>23</sup>  
199 and EpiLCs<sup>38</sup>. In congruence with previous work in the *Kdm5c*-KO hippocampus<sup>10</sup>, we observed aberrant accumulation of  
200 H3K4me3 around the transcription start site (TSS) of germline genes in the *Kdm5c*-KO amygdala (Figure 5A). We additionally  
201 found a marked increase in H3K4me2 germline gene TSSs in *Kdm5c*-KO EpiLCs (Figure 5B). Increase in H3K4me2 and  
202 H3K4me3 was highest in *Kdm5c*-KO cells that were highest genes that are bound by KDM5C at their promoter in EpiLCs  
203 (**note: do analysis to check if true**).

204 To assess KDM5C's embryonic role in germline gene silencing, we first characterized KDM5C's expression in ESCs and  
205 EpiLCs by harvesting RNA and protein at 0 hours (ESCs), 24 hours, and 48 hours (EpiLCs). While *Kdm5c* mRNA steadily  
206 decreased from 0 to 48 hours of differentiation, KDM5C protein initially increased from 0 to 24 hours but then decreased to  
207 near knockout levels by 48 hours (Figure 5C).

208 We then determined the role of KDM5C in the initial placement of DNA methylation at germline gene promoters by  
209 performing whole genome bisulfite sequencing in wild-type and *Kdm5c*-KO ESCs and 96 hour extended EpiLCs (exEpiLCs)  
210 (**note: check**). - While wild-type cells accumulate high levels of DNA methylation over the course of ESCs to exEpiLC  
211 differentiation, DNA methylation is markedly reduced in *Kdm5c*-KO

## 212 Discussion

213 Random thoughts - The demarcation of the germ vs soma is a key feature of multicellularity - Anything known about tissue-  
214 biased gene expression in other H3K4me regulators? - Our data suggests the germline developmental program is occurring  
215 ectopically as *Kdm5c*-KOs progresses through somatic tissue development - tissue-biased gene expression: - However unlike  
216 the gonadal-biased DEGs, many liver and muscle-biased DEGs have a known involvement in brain function. For example, the  
217 liver-biased DEG is *Apolipoprotein C-I* (*Apoc1*), is important for lipoprotein metabolism but has also been shown to influence  
218 learning and memory. - *Otx2* is expressed in EpiLCs and is known to repress PGC identity. - It's properly expressed in *Kdm5c*-  
219 KO EpiLCs, further supporting they aren't just becoming PGCs <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6485399/>

220 • Altogether, these data indicate that while some germline genes are misexpressed due to direct loss of KDM5C binding  
221 during emryogenesis, secondary downstream mechanisms can also promote their aberrant transcription.

222 – This shift from meiotic genes to later spermatogenesis genes in the hippocampus suggests the germline devel-  
223 opmental program could occur ectopically as *Kdm5c*-KO cells progresses through somatic tissue development.  
224 **note: this is strengthened by the ChIP-seq data since KDM5C is not directly bound to many brain/flagellar**

225      **DEGs. This point might be stronger in the ChIPseq figure**

- 226      • KDM5C is dynamically regulated during the window of embryonic germline gene silencing, our EpiLC ChIP-seq is likely  
227      catching the tail end of KDM5C's main involvement.
- 228      • Papers to read/reference:  
229        – Reconstitution of the Mouse Germ Cell Specification Pathway in Culture by Pluripotent Stem Cells: [https://www.cell.com/fulltext/S0092-8674\(11\)00771-9](https://www.cell.com/fulltext/S0092-8674(11)00771-9)  
230        – 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/>
- 231
- 232

233      **References**

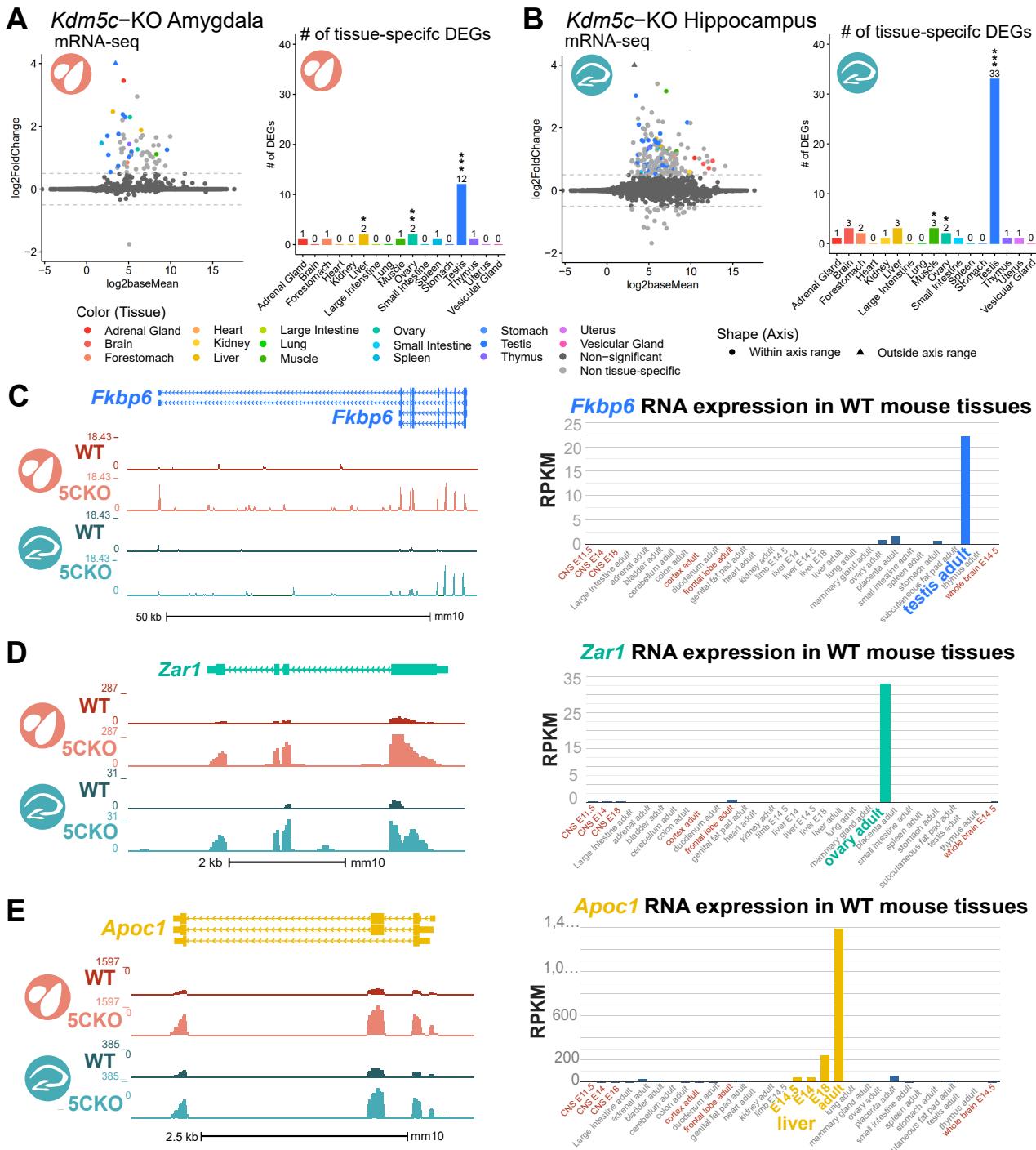
- 234      1. Strahl, B.D., and Allis, C.D. (2000). The language of covalent histone modifications. *Nature* **403**, 41–45. <https://doi.org/10.1038/47412>.
- 235
- 236      2. Jenuwein, T., and Allis, C.D. (2001). Translating the Histone Code. *Science* **293**, 1074–1080. <https://doi.org/10.1126/science.1063127>.
- 237
- 238      3. Lewis, E.B. (1978). A gene complex controlling segmentation in *Drosophila*. *Nature* **276**, 565–570. <https://doi.org/10.1038/276565a0>.
- 239
- 240      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>.
- 241
- 242      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>.
- 243
- 244      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>.
- 245
- 246      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 Astrocytic Fate Transition. *Neuron* **63**, 600–613. <https://doi.org/10.1016/j.neuron.2009.08.021>.
- 247
- 248      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>.
- 249
- 250      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>.
- 251
- 252      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>.
- 253

- 254 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.,  
255 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>.
- 256 12. Claes, S., Devriendt, K., Van Goethem, G., Roelen, L., Meireleire, J., Raeymaekers, P., Cassiman, J.J., and Fryns, J.P.  
257 (2000). Novel syndromic form of X-linked complicated spastic paraplegia. *Am J Med Genet* *94*, 1–4.
- 258 13. Jensen, L.R., Amende, M., Gurok, U., Moser, B., Gimmel, V., Tzschach, A., Janecke, A.R., Tariverdian, G., Chelly, J.,  
259 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>.
- 260 14. Carmignac, V., Nambot, S., Lehalle, D., Callier, P., Moortgat, S., Benoit, V., Ghoumid, J., Delobel, B., Smol, T., Thuillier,  
261 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>.
- 262 15. Iwase, S., Brookes, E., Agarwal, S., Badeaux, A.I., Ito, H., Vallianatos, C.N., Tomassy, G.S., Kasza, T., Lin, G.,  
263 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>.
- 264 16. Endoh, M., Endo, T.A., Shinga, J., Hayashi, K., Farcas, A., Ma, K.W., Ito, S., Sharif, J., Endoh, T., Onaga, N., et  
265 al. (2017). PCGF6-PRC1 suppresses premature differentiation of mouse embryonic stem cells by regulating germ  
cell-related genes. *eLife* *6*. <https://doi.org/10.7554/eLife.21064>.
- 266 17. Mochizuki, K., Sharif, J., Shirane, K., Uranishi, K., Bogutz, A.B., Janssen, S.M., Suzuki, A., Okuda, A., Koseki, H.,  
267 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>.
- 268 18. Velasco, G., Hubé, F., Rollin, J., Neuillet, D., Philippe, C., Bouzinba-Segard, H., Galvani, A., Viegas-Péquignot, E., and  
269 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>.
- 270 19. Hackett, J.A., Reddington, J.P., Nestor, C.E., Dunican, D.S., Branco, M.R., Reichmann, J., Reik, W., Surani, M.A.,  
271 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>.
- 272 20. Gupta, N., Yakhou, L., Albert, J.R., Azogui, A., Ferry, L., Kirsh, O., Miura, F., Battault, S., Yamaguchi, K., Laisné,  
273 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>.
- 274 21. Al Adhami, H., Vallet, J., Schaal, C., Schumacher, P., Bardet, A.F., Dumas, M., Chicher, J., Hammann, P., Daujat,  
275 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>.
- 276 22. Li, B., Qing, T., Zhu, J., Wen, Z., Yu, Y., Fukumura, R., Zheng, Y., Gondo, Y., and Shi, L. (2017). A Comprehensive  
277 Mouse Transcriptomic BodyMap across 17 Tissues by RNA-seq. *Sci Rep* *7*, 4200. <https://doi.org/10.1038/s41598-017-04520-z>.
- 278 23. 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.,  
279 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>.

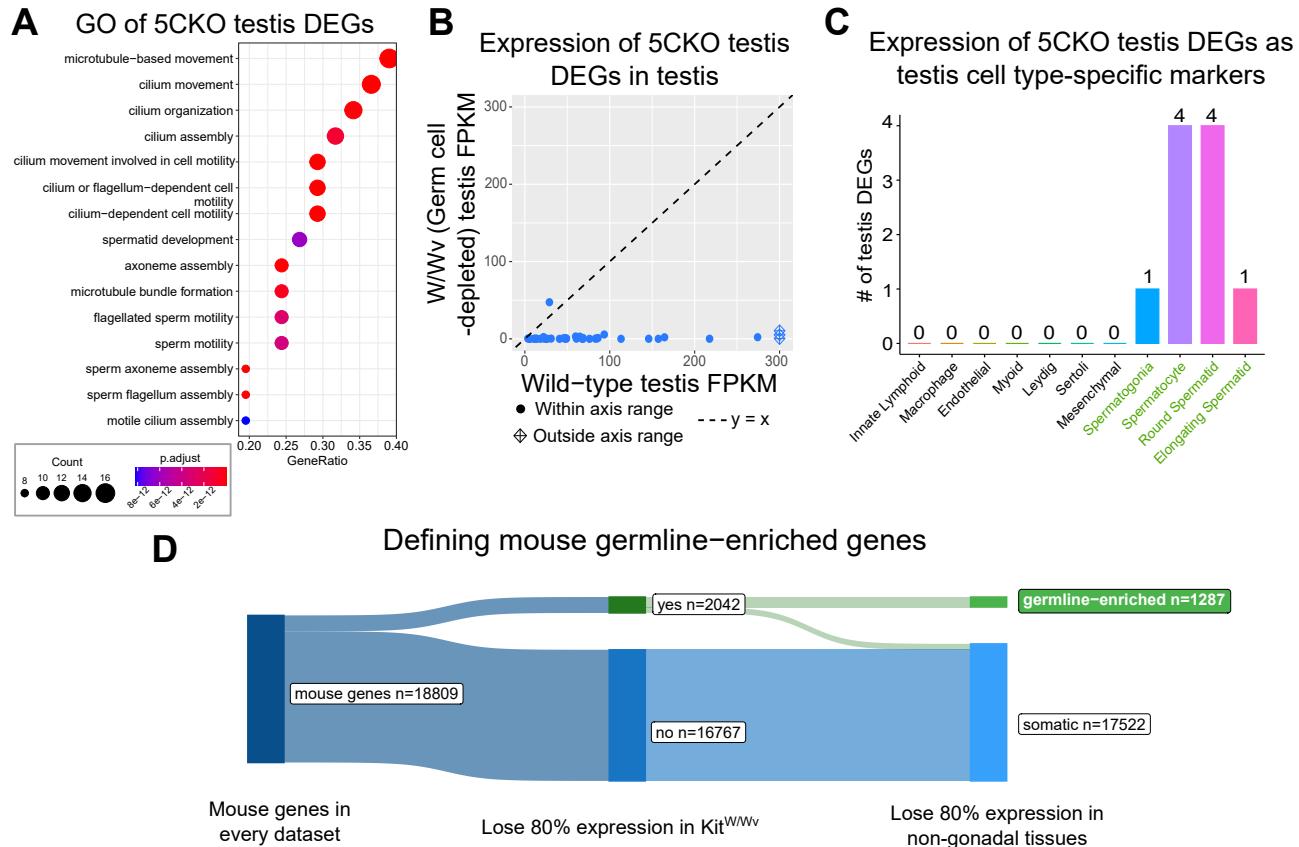
- 280 24. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data  
281 with DESeq2. *Genome Biol* **15**, 550. <https://doi.org/10.1186/s13059-014-0550-8>.
- 282 25. Crackower, M.A., Kolas, N.K., Noguchi, J., Sarao, R., Kikuchi, K., Kaneko, H., Kobayashi, E., Kawai, Y., Kozieradzki, I.,  
283 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>.
- 284 26. Xiol, J., Cora, E., Koglgruber, R., Chuma, S., Subramanian, S., Hosokawa, M., Reuter, M., Yang, Z., Berninger, P.,  
285 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>.
- 286 27. Cheng, S., Altmeppen, G., So, C., Welp, L.M., Penir, S., Ruhwedel, T., Menelaou, K., Harasimov, K., Stützer, A.,  
287 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>.
- 288 28. Mueller, J.L., Skaletsky, H., Brown, L.G., Zaghoul, S., Rock, S., Graves, T., Auger, K., Warren, W.C., Wilson, R.K., and  
289 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>.
- 290 29. Handel, M.A., and Eppig, J.J. (1979). Sertoli Cell Differentiation in the Testes of Mice Genetically Deficient in Germ  
291 Cells. *Biology of Reproduction* **20**, 1031–1038. <https://doi.org/10.1095/biolreprod20.5.1031>.
- 292 30. Green, C.D., Ma, Q., Manske, G.L., Shami, A.N., Zheng, X., Marini, S., Moritz, L., Sultan, C., Gurczynski, S.J., Moore,  
293 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>.
- 294 31. Soh, Y.Q., Junker, J.P., Gill, M.E., Mueller, J.L., van Oudenaarden, A., and Page, D.C. (2015). A Gene Regulatory  
295 Program for Meiotic Prophase in the Fetal Ovary. *PLoS Genet* **11**, e1005531. <https://doi.org/10.1371/journal.pgen.1005531>.
- 296 32. Magnúsdóttir, E., and Surani, M.A. (2014). How to make a primordial germ cell. *Development* **141**, 245–252.  
297 <https://doi.org/10.1242/dev.098269>.
- 298 33. Günesdogan, U., Magnúsdóttir, E., and Surani, M.A. (2014). Primordial germ cell specification: A context-dependent  
299 cellular differentiation event [corrected]. *Philos Trans R Soc Lond B Biol Sci* **369**. <https://doi.org/10.1098/rstb.2013.0543>.
- 300 34. Bardot, E.S., and Hadjantonakis, A.-K. (2020). Mouse gastrulation: Coordination of tissue patterning, specification and  
301 diversification of cell fate. *Mechanisms of Development* **163**, 103617. <https://doi.org/10.1016/j.mod.2020.103617>.
- 302 35. Welling, M., Chen, H., Muñoz, J., Musheev, M.U., Kester, L., Junker, J.P., Mischerikow, N., Arbab, M., Kuijk, E.,  
303 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>.
- 304 36. Macfarlan, T.S., Gifford, W.D., Driscoll, S., Lettieri, K., Rowe, H.M., Bonanomi, D., Firth, A., Singer, O., Trono, D., and  
305 Pfaff, S.L. (2012). Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* **487**, 57–63.  
<https://doi.org/10.1038/nature11244>.
- 306 37. Suzuki, A., Hirasaki, M., Hishida, T., Wu, J., Okamura, D., Ueda, A., Nishimoto, M., Nakachi, Y., Mizuno, Y., Okazaki,  
307 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>.

- 308 38. Samanta, M.K., Gayen, S., Harris, C., Maclary, E., Murata-Nakamura, Y., Malcore, R.M., Porter, R.S., Garay, P.M.,  
309 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>.
- 310 39. Koubova, J., Menke, D.B., Zhou, Q., Capel, B., Griswold, M.D., and Page, D.C. (2006). Retinoic acid regulates  
311 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>.
- 312 40. Lin, Y., Gill, M.E., Koubova, J., and Page, D.C. (2008). Germ Cell-Intrinsic and -Extrinsic Factors Govern Meiotic  
313 Initiation in Mouse Embryos. *Science* *322*, 1685–1687. <https://doi.org/10.1126/science.1166340>.
- 314 41. Endo, T., Mikedis, M.M., Nicholls, P.K., Page, D.C., and De Rooij, D.G. (2019). Retinoic Acid and Germ Cell  
315 Development in the Ovary and Testis. *Biomolecules* *9*, 775. <https://doi.org/10.3390/biom9120775>.
- 316 42. Li, H., Liang, Z., Yang, J., Wang, D., Wang, H., Zhu, M., Geng, B., and Xu, E.Y. (2019). DAZL is a master translational  
317 regulator of murine spermatogenesis. *Natl Sci Rev* *6*, 455–468. <https://doi.org/10.1093/nsr/nwy163>.
- 318 43. 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.  
319 *eLife* *9*, e56523. <https://doi.org/10.7554/eLife.56523>.
- 320 44. Liu, M., Zhu, Y., Xing, F., Liu, S., Xia, Y., Jiang, Q., and Qin, J. (2020). The polycomb group protein PCGF6  
321 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>.
- 322 45. Otani, J., Nankumo, T., Arita, K., Inamoto, S., Ariyoshi, M., and Shirakawa, M. (2009). Structural basis for recognition  
323 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>.
- 324 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  
325 autoinhibition and histone H3-induced activation of DNMT3A. *Nature* *517*, 640–644. <https://doi.org/10.1038/nature13899>.

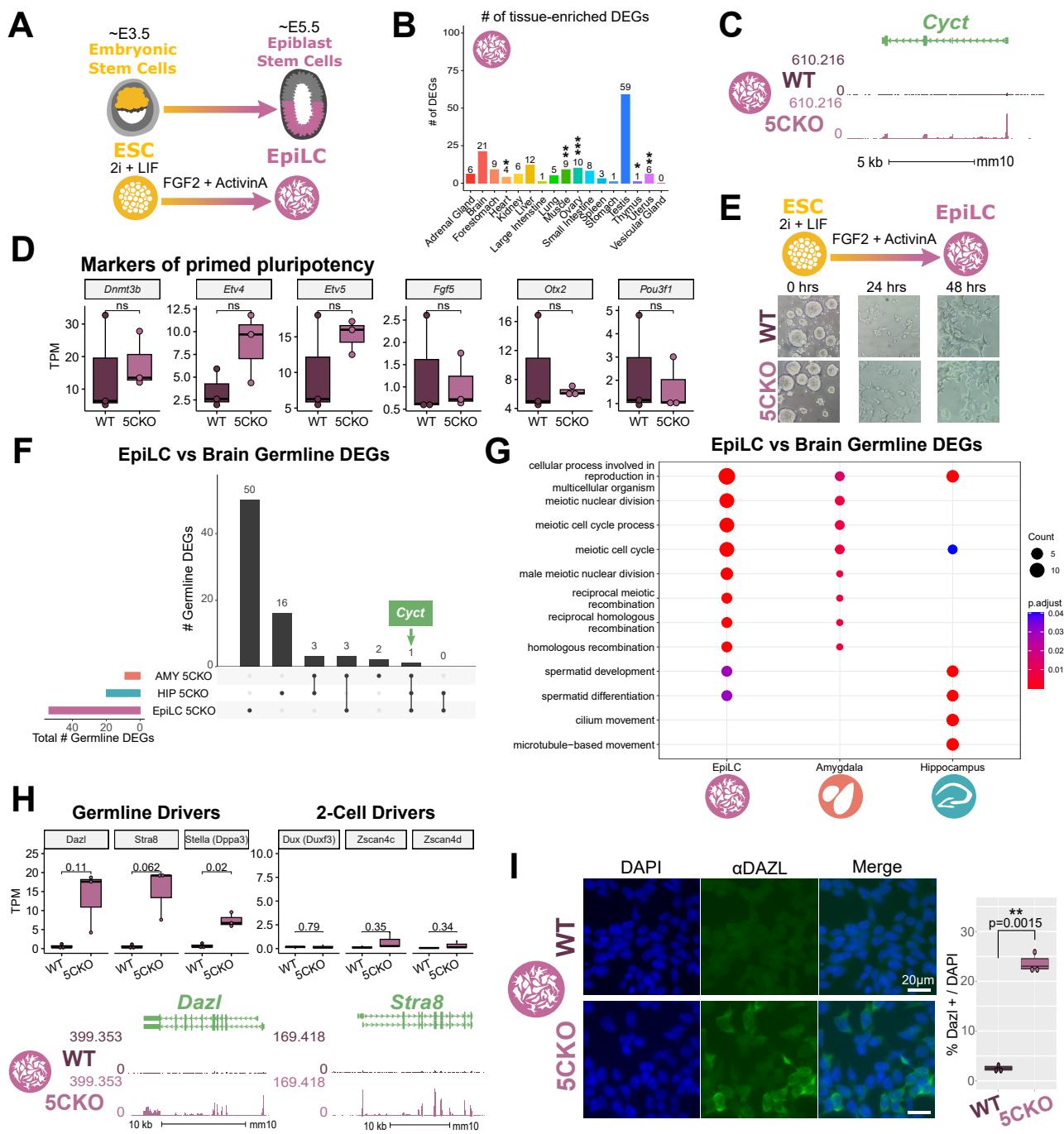
326 **Figures and Tables**



**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-seq. 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 (~E5.5). 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).

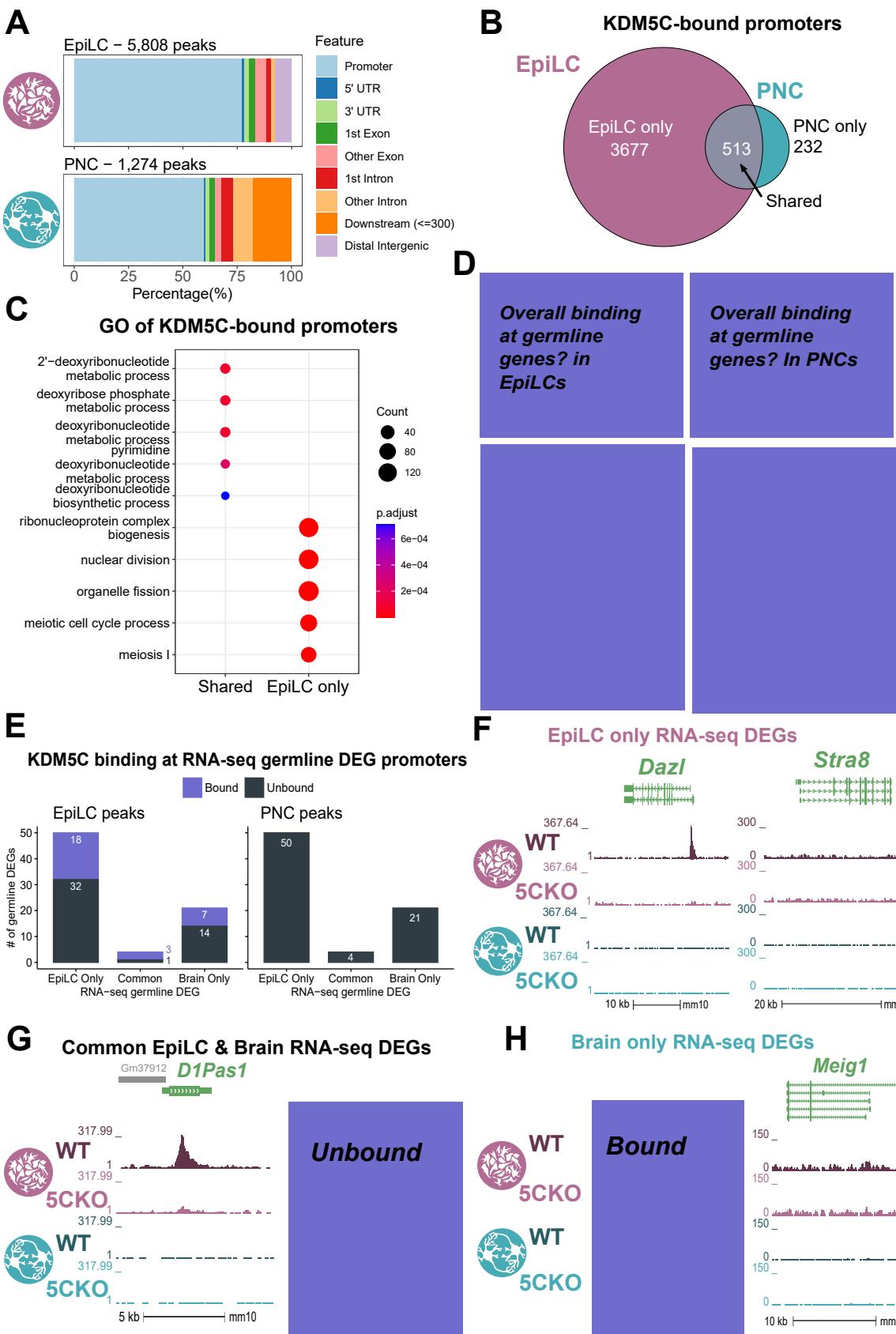
**E.** Representative images of wild-type (WT) and *Kdm5c*-KO cells during ESC to EpiLC differentiation. Brightfield images taken at 20X.

**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.

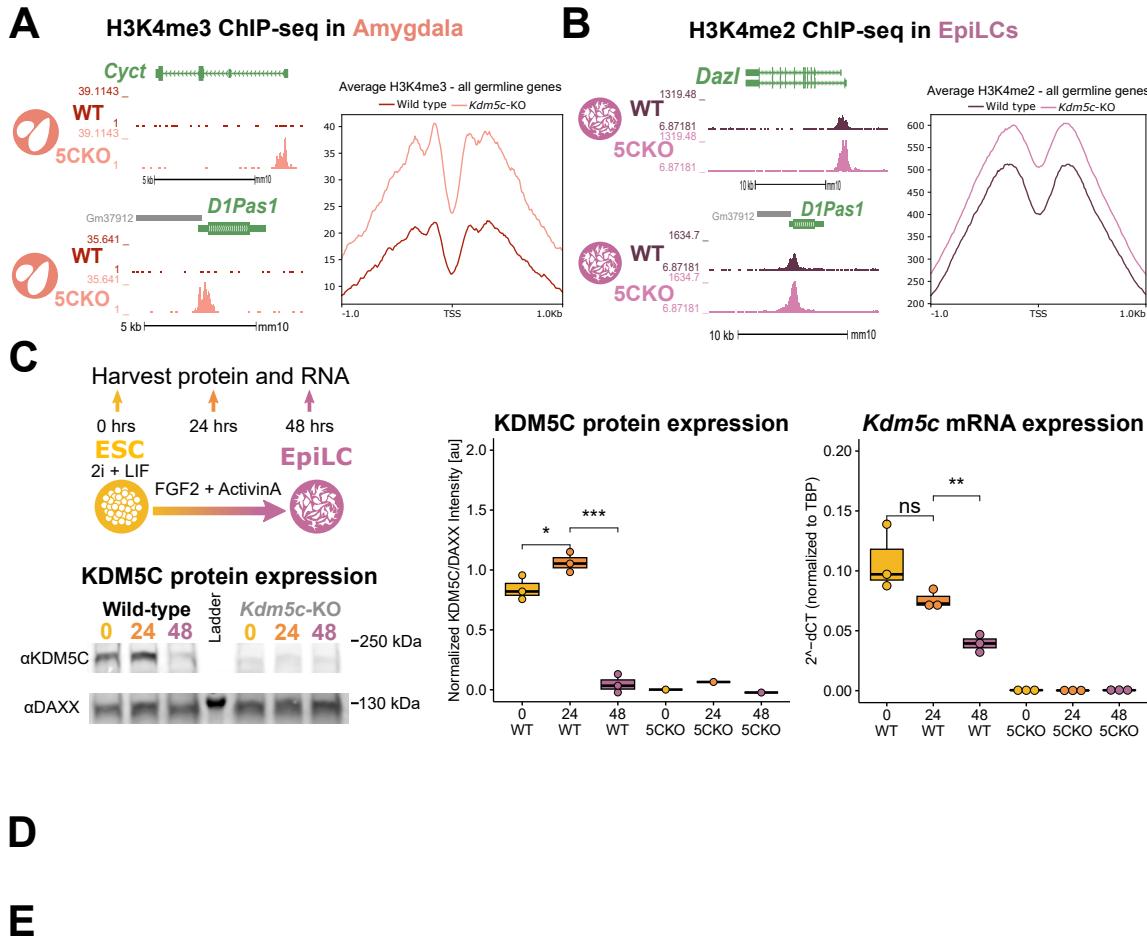
**H.** Top left - Example germline identity DEGs unique to EpiLCs. Top right - Example 2-cell genes that are not dysregulated in *Kdm5c*-KO EpiLCs. p-values for Welch's t-test. Bottom - Average bigwigs of *Dazl* and *Stra8* expression in wild-type and *Kdm5c*-KO EpiLCs.

**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. 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 at 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 ChIP-seq bigwigs of DEGs unique to EpiLCs. Although both are expressed in EpiLCs, KDM5C is bound to the *Dazl* promoter but not the *Stra8* promoter in EpiLCs. **F.** Bigwigs of the upregulated imprinted gene *Dlk1* that shows rescue in the double mutant brain. **G.** Example ChIP-seq bigwigs of DEGs common between brain and EpiLCs. KDM5C is bound to the *D1Pas1* promoter but not the *XXX* promoter in EpiLCs. **H.** Example ChIP-seq bigwigs of DEGs unique to the brain. KDM5C is bound to the *XXX* promoter but not the *Meig1* promoter



**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 transcription start site (TSS) of all germline-enriched genes in wild-type (dark red) and *Kdm5c*-KO 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. **C.** RNA and protein expression of KDM5C across ESC to EpiLC differentiation. Top left - diagram of differentiation protocol and collection time points. Bottom left - representative lanes of Western blot for KDM5C and DAXX. Middle - KDM5C protein expression normalized to DAXX. Quantified intensity using ImageJ (artificial units - au). Right - RT-qPCR of *Kdm5c* RNA expression, calculated in comparison to TBP expression ( $2^{-\Delta\Delta CT}$ ). **D.** XXX. **E.** XXX.

327 **Notes**

328 **Figure outline:**

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

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

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

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

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

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

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

364    Introduction: \* Chromatin regulators are important for cellular identity \* H3K4me1-3 linked to active gene promoters and  
365    enhancers \* Surprisingly, mutations in Chromatin regulators lead to many NDDs (including many H3K4 regulators) \* Recent

366 studies have shown some chromatin regulators are important for regulating neuron-specific gene expression/chromatin  
367 stat\_compare\_means \* However, loss of some chromatin regulators can also lead to ectopic expression of tissue-enriched  
368 genes \* Very few studies have looked at these genes and it's unclear if these genes contribute to NDD impairments. \*  
369 Necessary to first characterize the mechanism behind their derepression to identify molecular footholds into testing their  
370 contribution to neuronal impairments and potential for therapeutic intervention

- 371 • Loss of KDM5C can result in the misexpression of genes typically only found in the testis
- 372     – Misexpression of tissue-enriched genes hasn't been systematically characterized - Unclear if these genes are  
373         exceptions or if other tissue-specific genes are dysregulated
- 374     – Interestingly, these genes (Cyct, D1pas1) typically function in the germline
- 375     – Germ cells (meiotic cells) are typically distinguished from somatic cells very early on in embryogenesis and is a  
376         key feature of multicellularity
- 377     – Chromatin regulators are very important for decommissioning germline genes and act successively the embryo  
378         implants into the uterine wall
- 379         \* Most studies have focused on ESCs, which have a similar transcriptome to germ cells / 2-cells
- 380         \* recently, KDM5C was shown to repress DAZL in ESCs, independent of its catalytic activity
- 381         \* However, DNA methylation is lost in the mature 5CKO brain, DNA methylation is placed later and it's unclear if  
382             it's required for long-term repression (maybe too specific, just trying to go into the fact that the mechanism is  
383             partially understood but unclear)
- 384     – Systematic characterization of ectopic germline genes hasn't been done
- 385         \* unknown if other germline-enriched genes are dysregulated, including oocyte-specific genes
- 386         \* Crucially, it's unknown if misexpression of the germline program leads to functional consequences in 5CKO  
387         cells.

388 **Germline gene repression background:**

389 Interestingly, some of the ectopic testis transcripts identified in the *Kdm5c*-KO brain are typically expressed in germ cells<sup>10</sup>.  
390 Unlike somatic cells, germ cells (e.g. sperm and eggs) undergo meiosis and pass on their genetic material to the next genera-  
391 tion. The germline and the soma are typically distinguished during early embryogenesis, when germline genes are silenced  
392 in epiblast stem cells soon after implantation and only reactivated in a subset to form the germline. Chromatin regulators  
393 play a key role in decommissioning germline genes as the embryo transitions from naive to primed pluripotency by placing  
394 repressive histone H2A lysine 119 monoubiquitination (H2AK119ub1)<sup>16</sup>, histone 3 lysine 9 trimethylation (H3K9me3)<sup>16,17</sup>,  
395 and DNA CpG methylation<sup>17-19</sup> at germline gene promoters. KDM5C may also be involved in this early decommissioning of  
396 germline genes, as re-expression of KDM5C in neurons fails to suppress their dysregulation<sup>10</sup>. In support of this, KDM5C  
397 was very recently shown to repress *Deleted in azoospermia like* (*Dazl*), a key regulator of germline development, in mouse  
398 embryonic stem cells (ESCs)<sup>20,21</sup>. In support of this, two independent screens in mouse embryonic stem cells (ESCs) recently  
399 identified KDM5C as a repressor of *Deleted in azoospermia like* (*Dazl*), a key regulator of germline development. However,  
400 KDM5C's role in embryonic germline gene repression is currently unclear, given that *Dazl* is also expressed in ESCs and in  
401 the 2-cell stage and germline gene misexpression has yet to be globally characterized during *Kdm5c*-KO embryogenesis.