

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

3

4      Katherine M. Bonefas, Ilakkiya Venkatachalam, and Shigeki Iwase.

5      **Abstract**

6      Mutations in numerous chromatin-modifying enzymes cause neurodevelopmental disorders (NDDs).  
7      Loss of repressive chromatin regulators can lead to the aberrant transcription of tissue-specific genes  
8      outside of their intended context, however the mechanisms and consequences of their dysregulation are  
9      largely unknown. Here, we explored this cellular identity crisis in NDDs by investigating lysine demethylase  
10     5c (KDM5C), an eraser of histone 3 lysine 4 di and tri-methylation (H3K4me2/3), in tissue identity. We  
11     found male *Kdm5c* knockout (-KO) mice, which recapitulate key behavioral phenotypes of Claes-Jensen  
12     X-linked intellectual disability, aberrantly expresses many liver, muscle, ovary, and testis genes within the  
13     amygdala and hippocampus. Gonad-enriched genes expressed in the *Kdm5c*-KO brain were typically  
14     unique to germ cells, indicating an erosion of the soma-germline boundary. Germline genes are usually  
15     decommissioned in somatic lineages in the post-implantation epiblast, yet *Kdm5c*-KO epiblast-like cells  
16     (EpiLCs) aberrantly expressed key regulators of germline identity and meiosis, including *Dazl* and *Stra8*.  
17     Germline gene suppression is sexually dimorphic, as female EpiLCs required a higher dose of KDM5C to  
18     maintain germline gene suppression. Using a curated list of mouse germline-enriched genes, we found  
19     KDM5C is selectively recruited to a subset of germline gene promoters that contain CpG islands (CGIs)  
20     to facilitate DNA CpG methylation (CpGme) during ESC to EpiLC differentiation. However, late stage  
21     spermatogenesis genes devoid of promoter CGIs can also become activated in *Kdm5c*-KO cells via ectopic  
22     activation by RFX transcription factors. Thus, distinct suppressive mechanisms are recruited to different  
23     germline gene classes and ectopic germline transcriptional programs can mirror germ cell development  
24     within somatic tissues.

25      **Introduction**

26      A single genome holds the instructions to generate the myriad of cell types found within an organism.  
27      This is, in part, accomplished by chromatin regulators that can either promote or impede lineage-specific

28 gene expression through DNA and histone modifications<sup>1–5</sup>. Human genetic studies revealed mutations in  
29 chromatin regulators are a major cause of neurodevelopmental disorders (NDDs)<sup>6</sup> and many studies have  
30 identified their importance for regulating brain-specific transcriptional programs. Loss of some chromatin  
31 regulators can also result in the ectopic expression of tissue-specific genes outside of their target environment,  
32 such as the misexpression of liver-specific genes within adult neurons<sup>7</sup>. However, the mechanisms underlying  
33 ectopic gene expression and its impact upon neurodevelopment are poorly understood.

34 To elucidate the role of tissue identity in chromatin-linked NDDs, it is essential to first characterize the  
35 nature of the dysregulated genes and the molecular mechanisms governing their de-repression. Here we  
36 focus on lysine demethylase 5C (KDM5C, also known as SMCX or JARID1C), which erases histone 3 lysine  
37 4 di- and trimethylation (H3K4me2/3), a permissive chromatin modification enriched at gene promoters<sup>8</sup>.  
38 Pathogenic mutations in *KDM5C* cause Intellectual Developmental Disorder, X-linked, Syndromic, Claes-  
39 Jensen Type (MRXSCJ, OMIM: 300534). MRXSCJ is more common and severe in males and its neurological  
40 phenotypes include intellectual disability, seizures, aberrant aggression, and autistic behaviors<sup>9–11</sup>. Male  
41 *Kdm5c* knockout (-KO) mice recapitulate key MRXSCJ phenotypes, including hyperaggression, increased  
42 seizure propensity, and learning impairments<sup>12,13</sup>. RNA sequencing (RNA-seq) of the *Kdm5c*-KO hippocam-  
43 pus surprisingly revealed ectopic expression of some germline genes within the brain<sup>13</sup>. However, it is unclear  
44 if other tissue-specific genes are aberrantly transcribed with KDM5C loss, at what point in development  
45 germline gene misexpression begins, and what mechanisms underlie their dysregulation.

46 Distinguishing between germ cells and somatic cells is a key feature of multicellularity<sup>14</sup> that occurs  
47 during early embryogenesis in many metazoans<sup>15</sup>. In mammals, chromatin regulators are crucial for  
48 decommissioning germline genes during the transition from naïve to primed pluripotency. Initially, germline  
49 gene promoters gain repressive histone H2A lysine 119 monoubiquitination (H2AK119ub1)<sup>16</sup> and histone 3  
50 lysine 9 trimethylation (H3K9me3)<sup>16,17</sup> in embryonic stem cells (ESCs) and are then decorated with DNA  
51 CpG methylation (CpGme) in the post-implantation embryo<sup>17–19</sup>. The contribution of KDM5C to this process  
52 remains unclear. Furthermore, studies on germline gene repression have primarily been conducted in males  
53 and focused on marker genes important for germ cell development rather than germline genes as a whole,  
54 given the lack of a curated list for germline-enriched genes. Therefore, it is unknown if the mechanism  
55 of repression differs between sexes or for certain classes of germline genes, e.g. meiotic genes versus  
56 spermatid differentiation genes.

57 To illuminate KDM5C's role in tissue identity, here we characterized the aberrant expression of tissue-  
58 enriched genes within the *Kdm5c*-KO brain and epiblast-like stem cells (EpiLCs), an *in vitro* model of the  
59 post-implantation embryo. We curated list of mouse germline-enriched genes, which enabled genome-wide  
60 analysis of germline gene silencing mechanisms for the first time. Based on the data presented below, we  
61 propose KDM5C plays a fundamental, sexually dimorphic role in the development of tissue identity during  
62 early embryogenesis, including the establishment of the soma-germline boundary.

63 **Results**

64 **Tissue-enriched genes are aberrantly expressed in the *Kdm5c*-KO brain**

65 Previous RNA sequencing (RNA-seq) of the adult male *Kdm5c*-KO hippocampus revealed ectopic  
66 expression of some germline genes unique to the testis<sup>13</sup>. It is currently unknown if the testis is the only  
67 tissue type misexpressed in the *Kdm5c*-KO brain. We thus characterized the role of KDM5C in brain tissue  
68 identity by systematically assessing the dysregulation of genes enriched in 17 mouse tissues<sup>20</sup>. We analyzed  
69 tissue-specific differentially expressed genes (DEGs) in our published mRNA-seq datasets<sup>21</sup> of the adult  
70 amygdala and hippocampus from wild-type and constitutive *Kdm5c*-KO male mice (DESeq2<sup>22</sup>, log2 fold  
71 change > 0.5, q < 0.1).

72 We found a large proportion of significantly upregulated genes within the *Kdm5c*-KO brain are typically  
73 enriched within non-brain tissues (Amygdala: 35%, Hippocampus: 24%) (Figure 1A-B). For both the  
74 amygdala and hippocampus, the majority of tissue-enriched (DEGs) were testis genes (Figure 1A-C). Even  
75 though the testis has the largest total number of tissue-biased genes (2,496 genes) compared to any other  
76 tissue, testis-biased DEGs were significantly enriched for both brain regions (Amygdala p = 1.83e-05, Odds  
77 Ratio = 5.13; Hippocampus p = 4.26e-11, Odds Ratio = 4.45, Fisher's Exact Test). One example of a  
78 testis-enriched gene misexpressed in the *Kdm5c*-KO brain is *FK506 binding protein 6* (*Fkbp6*), a known  
79 regulator of PIWI-interacting RNAs (piRNAs) and meiosis<sup>23,24</sup> (Figure 1C).

80 Interestingly, we also observed significant enrichment of ovary-biased DEGs in both the amygdala and  
81 hippocampus (Amygdala p = 0.00574, Odds Ratio = 18.7; Hippocampus p = 0.048, Odds Ratio = 5.88,  
82 Fisher's Exact) (Figure 1A-B). Ovary-enriched DEGs included *Zygotic arrest 1* (*Zar1*), which sequesters  
83 mRNAs in oocytes for meiotic maturation<sup>25</sup> (Figure 1D). Given that the *Kdm5c*-KO mice we analyzed are  
84 male, these data demonstrate that the ectopic expression of gonad-enriched genes is independent of  
85 organismal sex.

86 Although not consistent across brain regions, we also found significant enrichment of DEGs biased  
87 towards two non-gonadal tissues - the liver (Amygdala p = 0.04, Odds Ratio = 6.58, Fisher's Exact Test) and  
88 the muscle (Hippocampus p = 0.01, Odds Ratio = 6.95, Fisher's Exact Test) (Figure 1A-B). *Apolipoprotein*  
89 *C-1* (*Apoc1*) a lipoprotein metabolism and transport gene, is among the liver-biased DEG derepressed in both  
90 the hippocampus and amygdala<sup>26</sup> and its brain overexpression has been implicated in Alzheimer's disease<sup>27</sup>  
91 (Figure 1E).

92 For all *Kdm5c*-KO tissue-enriched DEGs, aberrantly expressed mRNAs are polyadenylated and spliced  
93 into mature transcripts (Figure 1C-E). Of note, we observed little to no dysregulation of brain-enriched genes  
94 (Amygdala p = 1, Odds Ratio = 1.22; Hippocampus p = 0.74, Odds Ratio = 1.22, Fisher's Exact), despite the  
95 fact these are brain samples and the brain has the second highest total number of tissue-enriched genes  
96 (708 genes). Altogether, these results suggest the aberrant expression of tissue-enriched genes within the  
97 brain is a major effect of KDM5C loss.

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

99     *Kdm5c*-KO brain expresses testicular germline genes<sup>13</sup>, however the testis also contains somatic cells that  
100 support hormone production and germline functions. To determine if *Kdm5c*-KO results in ectopic expression  
101 of somatic testicular genes, we first evaluated the known functions of testicular DEGs through gene ontology.  
102 We found *Kdm5c*-KO testis-enriched DEGs had high enrichment of germline-relevant ontologies, including  
103 spermatid development (GO: 0007286, p.adjust = 6.2e-12) and sperm axoneme assembly (GO: 0007288,  
104 p.adjust = 2.45e-14) (Figure 2A).

105     We then evaluated testicular DEG expression in wild-type testes versus testes with germ cell depletion<sup>28</sup>,  
106 which was accomplished by heterozygous *W* and *Wv* mutations in the enzymatic domain of *c-Kit* (*Kit*<sup>W/Wv</sup>)<sup>29</sup>.  
107 Almost all *Kdm5c*-KO testis-enriched DEGs lost expression with germ cell depletion (Figure 2B). We then  
108 assessed testis-enriched DEG expression in a published single cell RNA-seq dataset that identified cell  
109 type-specific markers within the testis<sup>30</sup>. Some *Kdm5c*-KO testis-enriched DEGs were classified as specific  
110 markers for different germ cell developmental stages (e.g. spermatogonia, spermatocytes, round spermatids,  
111 and elongating spermatids), yet none marked somatic cells (Figure 2C). Together, these data demonstrate  
112 that the *Kdm5c*-KO brain aberrantly expresses germline genes, but not somatic testicular genes, reflecting  
113 an erosion of the soma-germline boundary.

114     As of yet, research on germline gene silencing mechanisms has focused on a handful of key genes  
115 rather than assessing germline gene suppression genome-wide due to the lack of a comprehensive gene list.  
116 We therefore generated a list of mouse germline-enriched genes using RNA-seq datasets of *Kit*<sup>W/Wv</sup> mice  
117 that included males and females at embryonic day 12, 14, and 16<sup>31</sup> and adult male testes<sup>28</sup>. We defined  
118 genes as germline-enriched if their expression met the following criteria: 1) their expression is greater than  
119 1 FPKM in wild-type gonads 2) their expression in any non-gonadal tissue of adult wild type mice<sup>20</sup> does  
120 not exceed 20% of their maximum expression in the wild-type germline, and 3) their expression in the germ  
121 cell-depleted gonads, for any sex or time point, does not exceed 20% of their maximum expression in the  
122 wild-type germline. These criteria yielded 1,288 germline-enriched genes (Figure 2D), which was hereafter  
123 used as a resource to globally characterize germline gene misexpression with *Kdm5c* loss (Supplementary  
124 table 1).

125 ***Kdm5c*-KO epiblast-like cells aberrantly express key regulators of germline identity**

126     Germ cells are typically distinguished from somatic cells soon after the embryo implants into the uterine  
127 wall<sup>32,33</sup>, when they are silenced in epiblast stem cells that will differentiate into the ectoderm, mesoderm,  
128 and endoderm to form the somatic tissues<sup>34</sup>. This developmental time point can be modeled *in vitro* through  
129 differentiation of naïve embryonic stem cells (nESCs) into epiblast-like stem cells (EpiLCs) (Figure 3A)<sup>35,36</sup>.  
130 While some germline-enriched genes are also expressed in nESCs and in the 2-cell stage<sup>37-39</sup>, they are  
131 silenced as they differentiate into EpiLCs<sup>17,40</sup>. Therefore, we tested if KDM5C was necessary for the initial

132 silencing germline genes in somatic lineages by evaluating the impact of *Kdm5c* loss in male EpiLCs.  
133 *Kdm5c*-KO cell morpholgy during ESC to EpiLC differentiation appeared normal (Figure 3B) and EpiLCs  
134 properly expressed markers of primed pluripotency, such as *Dnmt3b*, *Fgf5*, *Pou3f1*, and *Otx2* (Figure 3C).  
135 We then identified tissue-enriched DEGs in our previously published RNA-seq dataset of wild-type and  
136 *Kdm5c*-KO EpiLCs<sup>41</sup> (DESeq2, log2 fold change > 0.5, q < 0.1). Similar to the *Kdm5c*-KO brain, we observed  
137 general dysregulation of tissue-enriched genes, with the largest number of genes belonging to the brain and  
138 testis, although they were not significantly enriched (Figure 3D). Using our list of mouse germline-enriched  
139 genes assembled above, we found 54 germline genes were misexpressed in male *Kdm5c*-KO EpiLCs.

140 We then compared EpiLC germline DEGs to those expressed in the *Kdm5c*-KO brain to determine if  
141 germline genes are constitutively dysregulated or change over the course of development. The majority  
142 of germline DEGs were unique to either EpiLCs or the brain, with only *Cyct* shared across all tissue/cell  
143 types (Figure 3E-F). EpiLCs had particularly high enrichment of meiosis-related gene ontologies (Figure  
144 3G), such as meiotic cell cycle process (GO:1903046, p.adjust = 1.59e-08) and meiotic nuclear division  
145 (GO:0140013, p.adjust = 9.76e-09). While there was modest enrichment of meiotic gene ontologies in both  
146 brain regions, the *Kdm5c*-KO hippocampus primarily expressed late-stage spermatogenesis genes involved  
147 in sperm axoneme assembly (GO:0007288, p.adjust = 0.00621) and sperm motility (GO:0097722, p.adjust =  
148 0.00612).

149 Notably, DEGs unique to *Kdm5c*-KO EpiLCs included key drivers of germline identity, such as *Stimulated*  
150 *by retinoic acid 8* (*Stra8*: log2 fold change = 3.7, q = 2.05e-26) and *Deleted in azoospermia like* (*Dazl*: log2  
151 fold change = 3.16, q = 4.08e-06) (Figure 3H). These genes are typically expressed when primordial germ  
152 cells (PGCs) are committed to the germline fate, but are also expressed later in life to trigger meiotic gene  
153 expression programs<sup>42-44</sup>. Of note, some germline genes, including *Dazl*, are also expressed in the two-cell  
154 embryo<sup>38,45</sup>. However, we did not see derepression of two-cell stage-specific genes, like *Duxf3* (*Dux*) (log2  
155 fold change = -0.282, q = 0.337) and *Zscan4d* (log2 fold change = 0.25, q = 0.381) (Figure 3H), indicating  
156 *Kdm5c*-KO EpiLCs do not revert back to a 2-cell-like state. Altogether, *Kdm5c*-KO EpiLCs expressing key  
157 drivers of germline identity and meiosis while the brain primarily expresses spermiogenesis genes indicate  
158 germline gene misexpression parallels germline development during proper *Kdm5c*-KO differentiation.

159 **Female epiblast-like cells have increased sensitivity to germline gene misexpression  
160 with *Kdm5c* loss**

161 It is currently unknown if the misexpression of germline genes is influenced by sex, as previous studies  
162 on germline gene repressors have focused on males<sup>16-18,46,47</sup>. Sex is particularly pertinent in the case  
163 of KDM5C because it partially escapes X chromosome inactivation (XCI), resulting in a higher dosage in  
164 females<sup>48-51</sup>. We therefore explored the impact of chromosomal sex upon germline gene suppression by  
165 comparing their dysregulation in male *Kdm5c* hemizygous knockout (XY *Kdm5c*-KO), female homozygous

166 knockout (XX *Kdm5c*-KO), and female heterozygous knockout (XX *Kdm5c*-HET) EpiLCs.<sup>41</sup>.  
167 Homozygous and heterozygous *Kdm5c* knockout females expressed over double the number of germline-  
168 enriched genes than hemizygous males (Figure 4A). While the majority of germline DEGs in *Kdm5c*-KO  
169 males were also dysregulated in females (74%), many were male-specific and female-specific, such as  
170 *Tktl2* and *Esx1* (Figure 4B). We then compared the known functions of germline genes dysregulated only in  
171 females (XX only - dysregulated in XX *Kdm5c*-KO, XX *Kdm5c*-HET, or both), only in males (XY only), or in  
172 all samples (shared) (Figure 4C). Female-specific germline DEGs were enriched for meiotic (GO:0051321  
173 meiotic cell cycle) and flagellar (GO:0003341 cilium movement) functions, while male-specific DEGs had  
174 roles in mitochondrial and cell signaling (GO:0070585 protein localization to mitochondrion). Germline  
175 transcripts expressed in both sexes were enriched for meiotic (GO:0140013 meiotic nuclear division) and  
176 egg-specific functions (GO:0007292 female gamete generation).

177 The majority of germline genes expressed in both sexes had a greater log2 fold change in females  
178 compared to males (Figure 4D-F). This increased degree of dysregulation in females, along with the  
179 increased total number of germline genes, indicates females are more sensitive to losing KDM5C-mediated  
180 germline gene suppression. Female sensitivity could be due to impaired XCI in *Kdm5c* mutants<sup>41</sup>, as many  
181 spermatogenesis genes lie on the X chromosome<sup>52,53</sup>. However, female germline DEGs were not biased  
182 towards the X chromosome and had a similar overall proportion of X chromosome DEGs compared to  
183 males (XY *Kdm5c*-KO - 10.29%, XX *Kdm5c*-HET - 7.43%, XX *Kdm5c*-KO - 10.59%) (Figure 4G). The  
184 majority of germline DEGs instead lie on autosomes for both male and female *Kdm5c* mutants (Figure 4G).  
185 Thus, while female EpiLCs are more prone to germline gene misexpression with KDM5C loss, it is likely  
186 independent of XCI defects.

#### 187 **Germline gene misexpression in *Kdm5c* mutants is independent of germ cell sex**

188 Although many germline genes have shared functions in the male and female germline, some have  
189 unique or sex-biased expression. Therefore, we wondered if *Kdm5c* mutant males would primarily express  
190 sperm genes while mutant females primarily expressed egg genes. To comprehensively assess whether  
191 germline gene sex corresponds with *Kdm5c* mutant sex, we first filtered our list of germline-enriched genes  
192 for egg and sperm-biased genes (Figure 4H). We defined germ cell sex-biased genes as those whose  
193 expression in the opposite sex, at any time point, is no greater than 20% of the gene's maximum expression  
194 in a given sex. This yielded 67 egg-biased, 1,024 sperm-biased, and 197 unbiased germline-enriched genes.  
195 We found egg, sperm, and unbiased germline genes were dysregulated in all *Kdm5c* mutants at similar  
196 proportions (Figure 4I-J). Furthermore, germline genes dysregulated exclusively in either male or female  
197 mutants were also not biased towards their corresponding germ cell sex (Figure 4I). Altogether, these results  
198 demonstrate sex differences in germline gene dysregulation is not due to sex-specific activation of sperm or  
199 egg transcriptional programs.

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

201 KDM5C binds to the promoters of several germline genes in embryonic stem cells (ESCs) but its binding  
202 is absent in neurons<sup>13</sup>. However, the lack of a comprehensive list of germline-enriched genes prohibited  
203 genome-wide characterization of KDM5C binding at germline gene promoters. Thus, it is unclear if KDM5C  
204 is enriched at germline gene promoters, what types of germline genes KDM5C regulates, and if its binding is  
205 maintained at any germline genes in neurons.

206 To address these questions, we analyzed KDM5C chromatin immunoprecipitation followed by DNA  
207 sequencing (ChIP-seq) datasets in EpiLCs<sup>41</sup> and primary forebrain neuron cultures (PNCs)<sup>12</sup>. EpiLCs had a  
208 higher total number of high-confidence KDM5C peaks than PNCs (EpiLCs: 5,808, PNCs: 1,276, MACS2 q <  
209 0.1 and fold enrichment > 1). KDM5C was primarily localized to gene promoters in both cell types (EpiLCs:  
210 4,190, PNCs: 745 ± 500bp from the TSS), although PNCs showed increased localization to non-promoter  
211 regions (Figure 5A).

212 The majority of promoters bound by KDM5C in PNCs were also bound in EpiLCs (513 shared promoters),  
213 however a large portion of gene promoters were only bound by KDM5C in EpiLCs (3,677 EpiLC only  
214 promoters) (Figure 5B). Genes bound by KDM5C in both PNCs and EpiLCs were enriched for functions  
215 involving nucleic acid turnover, such as deoxyribonucleotide metabolic process (GO:0009262, p.adjust =  
216 8.28e-05) (Figure 5C). Germline-specific ontologies were only enriched in KDM5C-bound promoters unique  
217 to EpiLCs, such as meiotic nuclear division (GO: 0007127 p.adjust = 6.77e-16) and meiotic cell cycle process  
218 (GO:1903046, p.adjust = 5.05e-16) (Figure 5C). There were no ontologies significantly enriched for genes  
219 bound by KDM5C only in PNCs. Using our mouse germline gene list, we observed evident KDM5C signal  
220 around the TSS of many germline genes in EpiLCs, but not in PNCs (Figure 5D). Based on our ChIP-seq  
221 peak cut-off criteria, KDM5C was highly enriched at 211 germline gene promoters in EpiLCs (0.164% of all  
222 germline genes) (Figure 5E). Of note, KDM5C was only bound to about one third of *Kdm5c*-KO RNA-seq  
223 DEG promoters (EpiLC only DEGs: 36%, Brain only DEGs: 33.3%) (Supplementary figure 1A-C). However,  
224 KDM5C was bound to the promoter at 3 out of the 4 genes dysregulated in both the brain and EpiLCs.  
225 Representative examples of KDM5C-bound and unbound germline DEGs are *Dazl* and *Stra8*, respectively  
226 (Figure 5F). Together, these results demonstrate KDM5C is recruited to a subset of germline genes in EpiLCs,  
227 including meiotic genes, but does not directly regulate germline genes in neurons. Furthermore, the majority  
228 of germline mRNAs expressed in *Kdm5c*-KO cells are dysregulated independent of direct KDM5C binding to  
229 their promoters.

230 Many germline-specific genes are suppressed by the transcription factor heterodimers E2F6/DP1 and  
231 MGA/MAX, which respectively bind E2F and E-box motifs<sup>18,46,47,54,55</sup>. To elucidate if KDM5C is recruited to  
232 germline genes by a similar mechanism, we used HOMER<sup>56</sup> to identify transcription factor motifs enriched at  
233 KDM5C-bound or unbound germline gene promoters (TSS ± 500 bp, q-value < 0.1). MAX and E2F6 binding  
234 sites were significantly enriched at germline genes bound by KDM5C in EpiLCs, but not at germline genes  
235 unbound by KDM5C (MAX q-value: 0.0068, E2F6 q-value: 0.0673, E2F q-value: 0.0917) (Figure 5G). One

236 third of KDM5C-bound promoters contained the consensus sequence for either E2F6 (E2F, 5'-TCCCGC-3'),  
237 MGA (E-box, 5'-CACGTG-3'), or both, but only 17% of KDM5C-unbound genes contained these motifs (Figure  
238 5H). KDM5C-unbound germline genes were instead enriched for multiple RFX transcription factor binding  
239 sites (RFX q-value < 0.0001, RFX2 q-value < 0.0001, RFX5 q-value < 0.0001) (Figure 5I, Supplementary  
240 figure 1D). RFX transcription factors bind X-box motifs<sup>57</sup> to promote ciliogenesis<sup>58,59</sup> and among them is  
241 RFX2, a central regulator of post-meiotic spermatogenesis<sup>60,61</sup>. Interestingly, RFX2 mRNA is derepressed  
242 in *Kdm5c*-KO EpiLCs (Figure 5J), however it is also not a direct target of KDM5C (Supplementary figure  
243 1E). Thus, RFX2 is a candidate transcription factor for driving the ectopic expression of KDM5C-unbound  
244 germline genes in *Kdm5c*-KO cells.

245 **KDM5C is recruited to germline gene promoters harboring CpG islands to facilitate  
246 *de novo* DNA methylation**

247 In the early embryo, germline gene promoters are initially decorated with repressive histone modifications  
248 and are then silenced long-term via DNA CpG methylation (CpGme)<sup>16,17,40,62</sup>. Our results above indicate  
249 KDM5C also acts at germline gene promoters during this time period. However, how KDM5C interacts with  
250 other germline gene silencing mechanisms is currently unclear. KDM5C is generally thought to suppress  
251 transcription through erasure of histone 3 lysine 4 di- and trimethylation (H3K4me2/3)<sup>8</sup>, yet KDM5C's  
252 catalytic activity was recently shown to be dispensable for suppressing *Dazl* in undifferentiated ESCs<sup>45</sup>. Since  
253 H3K4me3 impedes *de novo* CpGme placement<sup>63,64</sup>, KDM5C's catalytic activity may instead be required  
254 in the post-implantation embryo for long-term silencing of germline genes. In support of this, CpGme is  
255 markedly reduced at two germline gene promoters in the *Kdm5c*-KO adult hippocampus<sup>13</sup>.

256 Based on the above observations, we hypothesized KDM5C erases H3K4me3 to promote the initial  
257 placement of CpGme at germline gene promoters in EpiLCs. To test this hypothesis, we first characterized  
258 KDM5C's substrates, histone 3 lysine 4 di- and trimethylation (H3K4me2/3), at germline gene promoters in  
259 our previously published ChIP-seq datasets in male wild type and *Kdm5c*-KO amygdala<sup>21</sup> and EpiLCs<sup>41</sup>. In  
260 congruence with previous work in the *Kdm5c*-KO hippocampus<sup>13</sup>, we observed aberrant accumulation of  
261 H3K4me3 around the transcription start site (TSS) of germline genes in the *Kdm5c*-KO amygdala (Figure  
262 6A). We additionally found a marked increase in H3K4me2 around the TSS of germline genes in *Kdm5c*-KO  
263 EpiLCs (Figure 6B). To elucidate KDM5C's embryonic role, we then characterized KDM5C's mRNA and protein  
264 expression during male ESC to EpiLC differentiation (Figure 6C). While *Kdm5c* mRNA steadily decreased  
265 from 0 to 48 hours of differentiation (Figure 6D), KDM5C protein initially increased from 0 to 24 hours but  
266 then decreased to near knockout levels by 48 hours (Figure 6E). Together, these data suggest KDM5C acts  
267 during the transition between ESCs and EpiLCs to remove H3K4me at germline gene promoters.

268 In wild-type cells, germline genes accumulate CpG methylation (CpGme) at CpG islands (CGIs) during  
269 the transition from naïve to primed pluripotency<sup>19,40,65</sup>, reaching peak methylation levels when differentiated

270 into EpiLCs for 96 hours (extended EpiLCs, exEpiLCs)<sup>17</sup>. We first identified how many germline genes  
271 contained CGIs using the UCSC genome browser<sup>66</sup> and found out of 1,288 germline-enriched genes, only  
272 356 (27.64%) contained CGIs within their promoters (TSS ± 500 bp) (Figure 6F). CGI-containing germline  
273 genes were enriched for meiotic gene ontologies, including meiotic nuclear division (GO:XXXX, p.adj) and  
274 meiosis I (GO:XXXX, p.adj) when compared to CGI-free genes (Figure 6G). Although a minor portion of  
275 germline gene promoters contained CGIs, CGIs strongly determined KDM5C's recruitment to germline gene  
276 promoters (FISHER'S XXXX), with 79% of KDM5C-bound germline genes containing CGIs (Figure 6G).

277 We then performed whole genome bisulfite sequencing (WGBS) in male wild-type and *Kdm5c*-KO ESCs  
278 and 96 hour extended EpiLCs (exEpiLCs) to assess how KDM5C loss impacted initial CpGme placement  
279 at germline gene promoters (Figure 6H). We first identified which germline gene promoters significantly  
280 gained CpGme in wild-type cells during ESC to exEpiLCs differentiation (methylKit<sup>67</sup>, q < 0.01, |methylation  
281 difference| >= 25%, TSS ± 500 bp). In wild-type cells, the majority of germline genes gained substantial  
282 CpGme at their promoter during differentiation (60.08%), regardless if their promoter contained a CGI (Figure  
283 6I).

284 We then identified germline gene promoters differentially methylated in wild-type versus *Kdm5c*-KO  
285 exEpiLCs (methylKit, q < 0.01, |methylation difference| >= 25%, TSS ± 500 bp) and found 28 germline  
286 promoters were significantly hypomethylated with *Kdm5c* loss (Figure 6J). Approximately half of germline  
287 promoters hypomethylated in *Kdm5c*-KO exEpiLCs are direct targets of KDM5C in EpiLCs (13 out of 28  
288 hypomethylated DMRs). We then evaluated promoter CpGme at germline genes ectopically transcribed in  
289 either *Kdm5c*-KO EpiLCs or within the brain and found promoter CpGme was substantially reduced in about  
290 half of germline DEGs (Figure 6K). Significantly hypomethylated promoters included genes consistently  
291 dysregulated across multiple *Kdm5c*-KO RNA-seq datasets, such as *D1Pas1* (methylation difference =  
292 -60.03%, q-value = 3.26e-153) (Figure 6L). Surprisingly, we only observed a modest reduction in CpGme  
293 at *Dazl*'s promoter (methylation difference = -6.525%, q-value = 0.0159) (Figure 6M). Altogether, these  
294 results demonstrate KDM5C is recruited to germline gene CGIs to promote CpGme at select germline gene  
295 promoters during early embryogenesis, however other germline gene silencing mechanism can sufficiently  
296 compensate for KDM5C's loss at select germline gene promoters.

## 297 Discussion

298 In the above study, we demonstrate KDM5C's pivotal role in the development of tissue identity. We  
299 first characterized the misexpression of tissue-enriched genes within the *Kdm5c*-KO brain and identified  
300 substantial dysregulation of testis, liver, muscle, and ovary-enriched genes. Testis genes significantly  
301 enriched within the *Kdm5c*-KO amygdala and hippocampus are specific to germ cells and not expressed  
302 within testicular somatic cells. *Kdm5c*-KO epiblast-like cells (EpiLCs) aberrantly express key drivers of  
303 germline identity and meiosis, including *Dazl* and *Stra8*, while the *Kdm5c*-KO brain primarily expresses

304 genes important for late spermatogenesis. We demonstrated that while *Kdm5c* mutant sex did not influence  
305 whether sperm or egg-specific genes were misexpressed, female EpiLCs are more sensitive to germline  
306 gene de-repression. Germline genes can become aberrantly expressed in *Kdm5c*-KO cells via an indirect  
307 mechanism, as KDM5C is only bound to a subset of germline-enriched DEGs. Finally, we found KDM5C is  
308 dynamically regulated during ESC to EpiLC differentiation and promotes long-term germline gene silencing  
309 through DNA methylation at CpG islands. Therefore, we propose KDM5C plays a fundamental role in  
310 the development of tissue identity during early embryogenesis, including the establishment of the soma-  
311 germline boundary. By systematically characterizing KDM5C's role in germline gene repression, including  
312 its interaction with known silencing mechanisms, we unveiled unique repressive mechanisms governing  
313 distinct classes of germline gene in somatic lineages. Furthermore, these data provide molecular footholds  
314 that can then be exploited to test the ultimate contribution of ectopic germline gene expression upon  
315 neurodevelopment.

316 It is important to note that while we highlighted KDM5C's regulation of germline genes, some germline-  
317 enriched genes are also expressed at the 2-cell stage and in naïve ESCs for their role in pluripotency  
318 and self-renewal. For example, *Dazl* is primarily known for committing PGCs to the germline fate and  
319 regulating the translation of germline-specific mRNAs, but is also expressed at the 2-cell stage<sup>68</sup>, in naïve  
320 ESCs<sup>37</sup>, and in the inner cell mass<sup>37</sup>. Based on the de-repression of *Dazl* and *Zscan4c* in *Kdm5c*-KO  
321 ESCs, KDM5C was thought to promote the 2-cell-to-ESC transition<sup>[45]</sup>. Although expressed in naïve ESCs,  
322 *Dazl* and other "self-renewal" germline genes are silenced during ESC differentiation into epiblast stem  
323 cells/EpiLCs<sup>17,40</sup>. We found that while *Kdm5c*-KO EpiLCs also expressed *Dazl*, they did not express 2-  
324 cell specific genes. Misexpression of many germline genes in *Kdm5c*-KO EpiLCs may indicate they are  
325 differentiating into primordial germ cell-like cells (PGCLCs)<sup>32,33,35</sup>. Yet, *Kdm5c*-KO EpiLCs had normal  
326 cellular morphology and properly expressed markers for primed pluripotency, including *Otx2* which is known  
327 to repress EpiLC differentiation into PGCs/PGCLCs<sup>69</sup>. Altogether, these data suggest *Kdm5c*-KO germline  
328 gene misexpression occurs ectopically in conjunction with typical developmental programs and the 2-cell-like  
329 state observed in *Kdm5c*-KO ESCs likely reflect KDM5C's primary role in germline gene repression.

330 Although eggs and sperm employ the same transcriptional programs for shared functions, e.g. PGC  
331 formation, meiosis, and genome defense, some germline genes are sex specific. We found *Kdm5c* mutant  
332 males and females expressed both sperm and egg-biased genes, indicating the mechanism of derepression  
333 is independent of a given germline gene's sex. However, organismal sex did greatly influence the degree of  
334 germline gene dysregulation, as female *Kdm5c*-KO EpiLCs had over double the number of germline-enriched  
335 DEGs compared to males. The lack of X-linked gene enrichment in females suggests that this greater  
336 sensitivity to germline gene misexpress is not due to XCI defects previously reported in *Kdm5c*-KO females<sup>41</sup>.  
337 Sex differences in germline gene suppression may be a consequence of females having a higher dose of  
338 KDM5C than males, due to its escape from XCI<sup>48-51</sup>. Intriguingly, females with heterozygous loss of *Kdm5c*  
339 also had over double the number of germline DEGs than hemizygous knockout males, even though their

340 level of KDM5C should be roughly equivalent to that of wild-type males. Altogether, these results suggests  
341 germline gene silencing mechanims differ between males and females, which warrants further study to  
342 identify the biological implications and underlying mechanisms.

343 Emerging work indicates many histone-modifying enzymes have non-catalytic functions that influnce  
344 gene expression, sometimes even more potently than their catalytic roles<sup>70,71</sup>. KDM5C's catalytic activity  
345 may not be required for germline gene silencing, as it was recently found to be dispensible for repressing  
346 *Dazl* in ESCs<sup>45</sup>.

347 • Our results indicate KDM5C's removal of H3K4me2/3 may be necessary at a subset of CGI-containing  
348 germline genes, given they fail to accumulate CpGme during EpiLC differentiation. These CpGme-  
349 sensitive germline genes, such as *D1Pas1* and *Naa11*, are also de-repressed in the mature *Kdm5c*-KO  
350 brain and consistently dysregulated in multiple *Kdm5c*-KO datasets<sup>13</sup>.

351 • H3K4me3 and CpGme typically do not colocalize

- 352 – <https://pubmed.ncbi.nlm.nih.gov/17334365/>  
353 – <https://www.nature.com/articles/s41594-017-0013-5>  
354 – <https://pubmed.ncbi.nlm.nih.gov/23664763/>  
355 – <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2718496/>  
356 – <https://pubmed.ncbi.nlm.nih.gov/19581485/>

357 • Contrastingly, many germline genes had only a modest reduction in CpGme, including *Dazl* even  
358 though its TSS accumulated substantial H3K4me2.

359 • Multiple regulators repress *dazl*, *dazl* is not expressed in the mature *Kdm5c*-KO brain E2F6/MGA/MAX  
360 - Loss of many chromatin regulators affects CpGme (Multiple, non-redundant silencing mechanisms) -  
361 seems to be the crux point of germline gene silencing - Because CGIs are typically resistant to CpGme  
362 (accurate?), germline CGIs may require a highly repressive histone landscape to recruit sufficient  
363 DNMTs to these loci

364 • Altogether, our results indicates a given chromatin modifier's catalytic and non-catlytic regulation of  
365 gene expression can change depending upon the locus and developmental stage.

366 Although the classic model for germline gene silencing is accumulation of CpGme at CGIs within their  
367 promoter. We found less than 30% of germline-enriched genes have CGIs, yet the majority of CGI-free  
368 germline genes still gained significant CpGme around their TSS.

369 • CGIs highly determined KDM5C recruitment to germline gene promoters.

370 • KDM5C loss also impacted CpGme at some KDM5C-unbound, CGI-free germline genes. Unclear  
371 what the repressive mechanism is for these genes.

372 In somatic cells, germline genes are highly methylated at promoter CGIs, which are typically unmethylated  
373 for other classes of genes.

374 • Germline gene suppression has focused on the accumulation of CpGme at CGIs

375 • DNAm and CpG islands

376 • Combine germline gene list, CGI, RFX information together.

377 – Benefit of the list is finding CGI difference and kdm5c indirect mechanisms of de-repression

378 – List of genes enables us to propose this model

379 – CGI positive meiosis/germline formation regulators turned on → these turn on CGI-negative late  
380 stage regulators → These promote downstream dysregulation long term

381 • CGIs highly determine KDM5C recruitment to germline gene promoters. - KDM5C is known to be  
382 enriched at CGIs (in neurons? are these methylated? or is its germline CGI function different from its  
383 somatic CGI function?). - Other studies on germline gene repressors have shown they are important  
384 for CGIme, unclear if they participate in non-CGI TSS CpGme

385 – Altogether demonstrates different classes of germline genes use distinct silencing mechanisms

386 It is yet unknown how KDM5C is recruited to germline gene promoters, given that KDM5C itself does  
387 not contain domains for sequence-specific binding<sup>8</sup>. In HeLa cells and ESCs<sup>45,72</sup>, KDM5C associates with  
388 members of the polycomb repressive complex 1.6 (PRC1.6), which is recruited to germline gene promoters  
389 through E2F6/DP1 and MGA/MAX heterodimers<sup>16,55</sup>. While MAX and E2F6 motifs were enriched at KDM5C-  
390 bound germline genes in EpiLCs, only about one-third of promoters contained their consensus sequence.  
391 Additionally, some germline genes can become active in *Kdm5c*-KO cells independent of KDM5C binding to  
392 their promoters. Germline genes unbound by KDM5C were primarily late-stage spermatogenesis genes and  
393 significantly enriched for RFX transcription factors, including RFX2, a central regulator of spermiogenesis<sup>60,61</sup>.  
394 Another notable KDM5C-unbound DEG is *Stra8*, a meiotic transcription factor that is turned on in germ cells  
395 with retionic acid signaling and DAZL expression<sup>73,74</sup>. Misexpression of *Dazl* and *Rfx2* and their downstream  
396 targets in *Kdm5c*-KO cells suggests that once activated, ectopic germline genes can then turn on other  
397 aberrant germline programs to loosely mimic germ cell development.

398 The above work provides the mechanistic foundation for KDM5C-mediated repression of tissue and  
399 germline-specific genes. However, the contribution of these ectopic, tissue-specific genes towards *Kdm5c*-  
400 KO neurological impairments is still unknown. In addition to germline genes, we also identified significant  
401 enrichment of muscle, liver, and even ovary-biased transcripts within the male *Kdm5c*-KO brain. Intriguingly,  
402 select liver and muscle-biased DEGs do have known roles within the brain, such as the liver-enriched lipid  
403 metabolism gene *Apolipoprotein C-I (Apoc1)*<sup>26</sup>. *APOC1* dysregulation is implicated in Alzheimer's disease in  
404 humans<sup>27</sup> and overexpression of *Apoc1* in the mouse brain can impair learning and memory<sup>75</sup>. KDM5C may

405 therefore be crucial for neurodevelopment by fine-tuning the expression of tissue-enriched, dosage-sensitive  
406 genes like *Apoc1*. Given germline genes have no known functions within the brain, their impact upon  
407 neurodevelopment is currently unknown. Ectopic testicular germline transcripts have been observed in a va-  
408 riety of cancers<sup>76,77</sup>, including brain tumors in *Drosophila* and mammals<sup>78,ghafouri-fardExpressionCancerTestis2012?</sup>,  
409 indicating their dysregulation may promote genome instability and cellular de-differentiation. Intriguingly,  
410 some models for other chromatin-linked neurodevelopmental disorders also display impaired soma-germline  
411 demarcation<sup>7,79-82</sup>. Like KDM5C, the chromatin regulators underlying these conditions - DNA methyltrans-  
412 ferase 3b (DNMT3B), H3K9me1/2 methyltransferases G9A/GLP, methyl-CpG -binding protein 2 (MECP2) -  
413 primarily silence gene expression. Thus, KDM5C is among a growing cohort of chromatin-linked neurodevel-  
414 opmental disorders with similar erosion of the germline versus soma boundary. Further research is required  
415 to determine the impact of these germline genes and the extent to which this phenomenon occurs in humans.

## 416 Materials and Methods

### 417 Classifying tissue-enriched and germline-enriched genes

418 Tissue-enriched differentially expresssd genes (DEGs) were determined by their classification in a previ-  
419 ously published dataset from 17 male and female mouse tissues<sup>20</sup>. This study defined tissue expression as  
420 greater than 1 Fragments Per Kilobase of transcript per Million mapped read (FPKM) and tissue enrichment  
421 as at least 4-fold higher expression than any other tissue.

422 We curated a list of germline-enriched genes using an RNA-seq dataset from wild-type and germline-  
423 depleted (Kit<sup>W/W<sup>v</sup></sup>) male and female mouse embryos from embryonic day 12, 14, and 16<sup>31</sup>, as well as adult  
424 male testes<sup>28</sup>. Germline-enriched genes met the following criteria: 1) their expression is greater than 1  
425 FPKM in wild-type germline 2) their expression in any wild-type somatic tissues<sup>20</sup> does not exceed 20%  
426 of maximum expression in wild-type germline, and 3) their expression in the germ cell-depleted (Kit<sup>W/W<sup>v</sup></sup>)  
427 germline, for any sex or time point, does not exceed 20% of maximum expression in wild-type germline.

### 428 Cell culture

429 We utilized our previously established cultures of male wild-type and *Kdm5c* knockout (-KO) embryonic  
430 stem cells<sup>41</sup>. Sex was confirmed by genotyping *Uba1/Uba1y* on the X and Y chromosomes with the following  
431 primers: 5'-TGGATGGTGTGCCAATG-3', 5'-CACCTGCACGTTGCCCT-3'. Deletion of *Kdm5c* was  
432 confirmed through the primers 5'-ATGCCCATATTAAGAGTCCCTG-3', 5'-TCTGCCTTGATGGGACTGTT-3',  
433 and 5'-GGTTCTCAACACTCACATAGTG-3'.

434 Embryonic stem cells (ESCs) and epiblast-like cells were cultured using previously established  
435 methods<sup>36</sup>. Briefly, ESCs were initially cultured in primed ESC (pESC) media consisting of KnockOut  
436 DMEM (Gibco#10829–018), fetal bovine serum (Gibco#A5209501), KnockOut serum replacement

437 (Invitrogen#10828–028), Glutamax (Gibco#35050-061), Anti-Anti (Gibco#15240-062), MEM Non-essential  
438 amino acids (Gibco#11140-050), and beta-mercaptoethanol (Sigma#M7522). They were then transitioned  
439 into ground-state, “naïve” ESCs (nESCs) by culturing for four passages in N2B27 media containing  
440 DMEM/F12 (Gibco#11330–032), Neurobasal media (Gibco#21103–049), Gluamax, Anti-Anti, N2 sup-  
441 plement (Invitrogen#17502048), and B27 supplement without vitamin A (Invitrogen#12587-010), and  
442 beta-mercaptoethanol. Both pESC and nESC media were supplemented with 3  $\mu$ M GSK3 inhibitor  
443 CHIR99021 (Sigma #SML1046-5MG), 1  $\mu$ M MEK inhibitor PD0325901 (Sigma #PZ0162-5MG), and 1,000  
444 units/mL leukemia inhibitory factor (LIF, Millipore#ESG1107).

445 nESCs were differentiated into epiblast-like cells (EpiLCs, 48 hours) and extendend EpiLCs (exEpiLCs,  
446 96 hours) by culturing in N2B27 media containing DMEM/F12, Neurobasal media, Gluamax, Anti-Anti,  
447 N2 supplement, B27 supplement (Invitrogen#17504044), beta-mercaptoethanol, fibroblast growth factor 2  
448 (FGF2, R&D Biotechne 233-FB), and activin A (R&D Biotechne 338AC050CF), as previously described<sup>36</sup>.

#### 449 **Immunocytochemistry (ICC)**

450 ICC of DAZL in EpiLCs was performed by first growing cells on fibronectin-coated coverslips. Cells were  
451 then washed thrice with phosphobuffered saline (PBS), fixed in 4% paraformaldehyde, washed thrice in PBS,  
452 and blacked in PBS containing 0.3% Triton X-100, and 5% fetal bovine serum for 1 hour. They were then  
453 washed thrice with PBS and incubating in primary antibody (Rabbit anti DAZL, abcam ab34139, 1:200) in  
454 the blocking solution overnight at 4°C with gentle rocking. The next day, cells were rinsed thrice with PBS,  
455 and incubated in secondary antibody (Alexaflouor 488 Invitrogen #710369, 1:1,000) with DAPI (1:1,000) in  
456 blocking buffer for 1 hour at room temperature. Coverslips were then rinsed thrice in PBS, and mounted onto  
457 slides using Prolong Gold (Invitrogen #P36930). Images were taken blinded for genotype, chosen based on  
458 similar levels of DAPI signal, and quantified via ImageJ before unblinding.

#### 459 **RNA sequencing (RNA-seq)**

460 After ensuring read quality via FastQC (v0.11.8), reads were then mapped to the mm10 *Mus musculus*  
461 genome (Gencode) using STAR (v2.5.3a), during which we removed duplicates and kept only uniquely  
462 mapped reads. Count files were generated by FeatureCounts (Subread v1.5.0), and BAM files were  
463 converted to bigwigs using deeptools (v3.1.3) and visualized by the UCSC genome browser. RStudio (v3.6.0)  
464 was then used to analyze counts files by DESeq2 (v1.26.0)<sup>22</sup> to identify differentially expressed genes  
465 (DEGs) with a q-value (p-adjusted via FDR/Benjamini–Hochberg correction) less than 0.1 and a log2 fold  
466 change greater than 0.5. For all DESeq2 analyses, log2 fold changes were calculated with IfcShrink using  
467 the ashR package<sup>83</sup>. MA-plots were generated by ggpubr (v0.6.0), and Eulerr diagrams were generated by  
468 eulerr (v6.1.1). Boxplots and scatterplots were generated by ggpubr (v0.6.0) and ggplot2 (v3.3.2). The Upset  
469 plot was generated via the package UpSetR (v1.4.0)<sup>84</sup>. Gene ontology (GO) analyses were performed by

470 the R package enrichPlot (v1.16.2) using the biological processes setting and compareCluster.

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

472 ChIP-seq reads were aligned to mm10 using Bowtie1 (v1.1.2) allowing up to two mismatches. Only  
473 uniquely mapped reads were used for analysis. Peaks were called using MACS2 software (v2.2.9.1) using  
474 input BAM files for normalization, with filters for a q-value < 0.1 and a fold enrichment > 1. We removed  
475 “black-listed” genomic regions that often give aberrant signals. Common peak sets were obtained in R via  
476 DiffBind (v3.6.5). In the case of KDM5C ChIP-seq, *Kdm5c*-KO peaks were then subtracted from wild-type  
477 samples using bedtools (v2.25.0). Peak proximity to genome annotations was determined by ChIPSeeker  
478 (v1.32.1). Gene ontology (GO) analyses were performed by the R package enrichPlot (v1.16.2) using the  
479 biological processes setting and compareCluster. Enriched motifs were identified using HOMER<sup>56</sup>. Average  
480 binding across the genome was visualized using deeptools (v3.1.3). Bigwigs were visualized using the  
481 UCSC genome browser.

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

483 Genomic DNA (gDNA) from naïve ESCs and extended EpiLCs was extracted using the Wizard Genomic  
484 DNA Purification Kit (Promega A1120), following the instructions for Tissue Culture Cells. gDNA from  
485 two wild-types and two *Kdm5c*-KOs of each cell type was sent to Novogene for WGBS using the Illumina  
486 NovaSeq X Plus platform and sequenced for 150bp paired-end reads (PE150). - bismark - Methylkit

487 **Data availability**

488 **Published datasets**

489 All published datasets are available at the Gene Expression Omnibus (GEO) <https://www.ncbi.nlm.nih>  
490 .gov/geo. Published RNA-seq datasets analyzed in this study included the male wild-type and *Kdm5c*-KO  
491 adult amygdala and hippocampus<sup>21</sup> (available at GEO: GSE127722) and male wild-type and *Kdm5c*-KO  
492 EpiLCs<sup>41</sup> (available at GEO: GSE96797).

493 Previously published ChIP-seq experiments included KDM5C in wild-type and *Kdm5c*-KO EpiLCs<sup>41</sup> (avail-  
494 able at GEO: GSE96797) and mouse primary neuron cultures (PNCs) from the cortex and hippocampus<sup>12</sup>  
495 (available at GEO: GSE61036). ChIP-seq of histone 3 lysine 4 dimethylation in male wild-type and *Kdm5c*-KO  
496 EpiLCs<sup>41</sup> is also available at GEO: GSE96797. ChIP-seq of histone 3 lysine 4 trimethylation in wild-type and  
497 *Kdm5c*-KO male amygdala<sup>21</sup> are available at GEO: GSE127817.

498 **Data analysis**

499 Scripts used to generate the results, tables, and figures of this study are available via a GitHub repository:  
500 XXX

501 **Acknowledgements**

- 502 • Jacob Mueller for providing insight in germline gene regulation.  
503 • Sundeep Kalantry for providing reagents and expertise in culturing mouse embryonic stem cells and  
504 epiblast-like cells  
505 • Ilakkija  
506 • Funding sources

507 **References**

- 508 1. Strahl, B.D., and Allis, C.D. (2000). The language of covalent histone modifications. *Nature* **403**,  
509 41–45. <https://doi.org/10.1038/47412>.
- 510 2. Jenuwein, T., and Allis, C.D. (2001). Translating the Histone Code. *Science* **293**, 1074–1080.  
511 <https://doi.org/10.1126/science.1063127>.
- 512 3. Lewis, E.B. (1978). A gene complex controlling segmentation in Drosophila. *Nature* **276**, 565–570.  
513 <https://doi.org/10.1038/276565a0>.
- 514 4. Kennison, J.A., and Tamkun, J.W. (1988). Dosage-dependent modifiers of polycomb and antennapedia  
515 mutations in Drosophila. *Proc. Natl. Acad. Sci. U.S.A.* **85**, 8136–8140. <https://doi.org/10.1073/pnas.85.21.8136>.
- 516 5. Kassis, J.A., Kennison, J.A., and Tamkun, J.W. (2017). Polycomb and Trithorax Group Genes in  
517 *Drosophila*. *Genetics* **206**, 1699–1725. <https://doi.org/10.1534/genetics.115.185116>.
- 518 6. Gabriele, M., Lopez Tobon, A., D'Agostino, G., and Testa, G. (2018). The chromatin basis of  
519 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>.
- 520 7. Schaefer, A., Sampath, S.C., Intrator, A., Min, A., Gertler, T.S., Surmeier, D.J., Tarakhovsky, A.,  
521 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>.
- 522 8. Iwase, S., Lan, F., Bayliss, P., De La Torre-Ubieta, L., Huarte, M., Qi, H.H., Whetstone, 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>.

- 523
- 524 9. 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 Med Genet* **94**, 1–4.
- 525
- 526 10. Jensen, L.R., Amende, M., Gurok, U., Moser, B., Gimmel, V., Tschach, A., Janecke, A.R., Tariverdian, G., Chelly, J., Fryns, J.P., et al. (2005). Mutations in the JARID1C gene, which is involved in transcriptional regulation and chromatin remodeling, cause X-linked mental retardation. *Am J Hum Genet* **76**, 227–236. <https://doi.org/10.1086/427563>.
- 527
- 528 11. 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. <https://doi.org/10.1111/cge.13755>.
- 529
- 530 12. 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>.
- 531
- 532 13. 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>.
- 533
- 534 14. Devlin, D.K., Ganley, A.R.D., and Takeuchi, N. (2023). A pan-metazoan view of germline-soma distinction challenges our understanding of how the metazoan germline evolves. *Current Opinion in Systems Biology* **36**, 100486. <https://doi.org/10.1016/j.coisb.2023.100486>.
- 535
- 536 15. Lehmann, R. (2012). Germline Stem Cells: Origin and Destiny. *Cell Stem Cell* **10**, 729–739. <https://doi.org/10.1016/j.stem.2012.05.016>.
- 537
- 538 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 stem cells by regulating germ cell-related genes. *eLife* **6**. <https://doi.org/10.7554/eLife.21064>.
- 539
- 540 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>.
- 541

- 542 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>.
- 543
- 544 19. Hackett, J.A., Reddington, J.P., Nestor, C.E., Dunican, D.S., Branco, M.R., Reichmann, J., Reik, W., Surani, M.A., Adams, I.R., and Meehan, R.R. (2012). Promoter DNA methylation couples genome-defence mechanisms to epigenetic reprogramming in the mouse germline. *Development* **139**, 3623–3632. <https://doi.org/10.1242/dev.081661>.
- 545
- 546 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>.
- 547
- 548 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>.
- 549
- 550 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>.
- 551
- 552 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>.
- 553
- 554 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>.
- 555
- 556 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>.
- 557
- 558 26. Rouland, A., Masson, D., Lagrost, L., Vergès, B., Gautier, T., and Bouillet, B. (2022). Role of apolipoprotein C1 in lipoprotein metabolism, atherosclerosis and diabetes: A systematic review. *Cardiovasc Diabetol* **21**, 272. <https://doi.org/10.1186/s12933-022-01703-5>.
- 559
- 560 27. Leduc, V., Jasmin-Bélanger, S., and Poirier, J. (2010). APOE and cholesterol homeostasis in Alzheimer's disease. *Trends in Molecular Medicine* **16**, 469–477. <https://doi.org/10.1016/j.molmed.2010.07.008>.

- 562 28. Mueller, J.L., Skaletsky, H., Brown, L.G., Zaghlul, S., Rock, S., Graves, T., Auger, K., Warren,  
563 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>.
- 564 29. Handel, M.A., and Eppig, J.J. (1979). Sertoli Cell Differentiation in the Testes of Mice Genetically  
565 Deficient in Germ Cells. *Biology of Reproduction* *20*, 1031–1038. <https://doi.org/10.1095/biolreprod20>  
.5.1031.
- 566 30. 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.
- 567 31. 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>.
- 568 32. Magnúsdóttir, E., and Surani, M.A. (2014). How to make a primordial germ cell. *Development* *141*,  
571 245–252. <https://doi.org/10.1242/dev.098269>.
- 572 33. 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*.  
573 <https://doi.org/10.1098/rstb.2013.0543>.
- 574 34. 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>.
- 576 35. Hayashi, K., Ohta, H., Kurimoto, K., Aramaki, S., and Saitou, M. (2011). Reconstitution of the  
mouse germ cell specification pathway in culture by pluripotent stem cells. *Cell* *146*, 519–532.  
577 <https://doi.org/10.1016/j.cell.2011.06.052>.
- 578 36. Samanta, M., and Kalantry, S. (2020). Generating primed pluripotent epiblast stem cells: A methodol-  
579 ogy chapter. *Curr Top Dev Biol* *138*, 139–174. <https://doi.org/10.1016/bs.ctdb.2020.01.005>.
- 580 37. 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.  
581 *EMBO Reports* *16*, 791–802. <https://doi.org/10.15252/embr.201540538>.
- 582 38. 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  
583 activity. *Nature* *487*, 57–63. <https://doi.org/10.1038/nature11244>.

- 584 39. 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>.
- 585
- 586 40. Borgel, J., Guibert, S., Li, Y., Chiba, H., Schübeler, D., Sasaki, H., Forné, T., and Weber, M. (2010). Targets and dynamics of promoter DNA methylation during early mouse development. *Nat Genet* 42, 1093–1100. <https://doi.org/10.1038/ng.708>.
- 587
- 588 41. 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>.
- 589
- 590 42. 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>.
- 591
- 592 43. 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>.
- 593
- 594 44. Endo, T., Mikedis, M.M., Nicholls, P.K., Page, D.C., and De Rooij, D.G. (2019). Retinoic Acid and Germ Cell Development in the Ovary and Testis. *Biomolecules* 9, 775. <https://doi.org/10.3390/biom9120775>.
- 595
- 596 45. Gupta, N., Yakhou, L., Albert, J.R., Azogui, A., Ferry, L., Kirsh, O., Miura, F., Battault, S., Yamaguchi, K., Laisné, M., et al. (2023). A genome-wide screen reveals new regulators of the 2-cell-like cell state. *Nat Struct Mol Biol*. <https://doi.org/10.1038/s41594-023-01038-z>.
- 597
- 598 46. Pohlers, M., Truss, M., Frede, U., Scholz, A., Strehle, M., Kuban, R.-J., Hoffmann, B., Morkel, M., Birchmeier, C., and Hagemeier, C. (2005). A Role for E2F6 in the Restriction of Male-Germ-Cell-Specific Gene Expression. *Current Biology* 15, 1051–1057. <https://doi.org/10.1016/j.cub.2005.04.060>.
- 599
- 600 47. Dahlet, T., Truss, M., Frede, U., Al Adhami, H., Bardet, A.F., Dumas, M., Vallet, J., Chicher, J., Hammann, P., Kottnik, S., et al. (2021). E2F6 initiates stable epigenetic silencing of germline genes during embryonic development. *Nat Commun* 12, 3582. <https://doi.org/10.1038/s41467-021-23596-w>.
- 601
- 602 48. Agulnik, A.I., Mitchell, M.J., Mattei, M.G., Borsani, G., Avner, P.A., Lerner, J.L., and Bishop, C.E. (1994). A novel X gene with a widely transcribed Y-linked homologue escapes X-inactivation in mouse and human. *Hum Mol Genet* 3, 879–884. <https://doi.org/10.1093/hmg/3.6.879>.
- 603
- 604 49. Carrel, L., Hunt, P.A., and Willard, H.F. (1996). Tissue and lineage-specific variation in inactive X chromosome expression of the murine Smcx gene. *Hum Mol Genet* 5, 1361–1366. <https://doi.org/10.1093/hmg/5.9.1361>.
- 605

- 606 50. Sheardown, S., Norris, D., Fisher, A., and Brockdorff, N. (1996). The mouse Smcx gene exhibits  
developmental and tissue specific variation in degree of escape from X inactivation. *Hum Mol Genet*  
607 5, 1355–1360. <https://doi.org/10.1093/hmg/5.9.1355>.
- 608 51. Xu, J., Deng, X., and Disteche, C.M. (2008). Sex-Specific Expression of the X-Linked Histone  
Demethylase Gene Jarid1c in Brain. *PLoS ONE* 3, e2553. <https://doi.org/10.1371/journal.pone.00025>  
609 53.
- 610 52. Wang, P.J., McCarrey, J.R., Yang, F., and Page, D.C. (2001). An abundance of X-linked genes  
expressed in spermatogonia. *Nat Genet* 27, 422–426. <https://doi.org/10.1038/86927>.
- 612 53. Khil, P.P., Smirnova, N.A., Romanienko, P.J., and Camerini-Otero, R.D. (2004). The mouse X  
chromosome is enriched for sex-biased genes not subject to selection by meiotic sex chromosome  
613 inactivation. *Nat Genet* 36, 642–646. <https://doi.org/10.1038/ng1368>.
- 614 54. Hurlin, P.J. (1999). Mga, a dual-specificity transcription factor that interacts with Max and contains a  
T-domain DNA-binding motif. *The EMBO Journal* 18, 7019–7028. <https://doi.org/10.1093/emboj/18.2>  
615 4.7019.
- 616 55. Stielow, B., Finkernagel, F., Stiewe, T., Nist, A., and Suske, G. (2018). MGA, L3MBTL2 and E2F6  
determine genomic binding of the non-canonical Polycomb repressive complex PRC1.6. *PLoS Genet*  
617 14, e1007193. <https://doi.org/10.1371/journal.pgen.1007193>.
- 618 56. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H.,  
and Glass, C.K. (2010). Simple Combinations of Lineage-Determining Transcription Factors Prime  
cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Molecular Cell* 38, 576–589.  
619 <https://doi.org/10.1016/j.molcel.2010.05.004>.
- 620 57. Gajiwala, K.S., Chen, H., Cornille, F., Roques, B.P., Reith, W., Mach, B., and Burley, S.K. (2000).  
Structure of the winged-helix protein hRFX1 reveals a new mode of DNA binding. *Nature* 403,  
621 916–921. <https://doi.org/10.1038/35002634>.
- 622 58. Swoboda, P., Adler, H.T., and Thomas, J.H. (2000). The RFX-Type Transcription Factor DAF-19  
Regulates Sensory Neuron Cilium Formation in *C. elegans*. *Molecular Cell* 5, 411–421. [https://doi.org/10.1016/S1097-2765\(00\)80436-0](https://doi.org/10.1016/S1097-2765(00)80436-0).
- 624 59. Ashique, A.M., Choe, Y., Karlen, M., May, S.R., Phamluong, K., Solloway, M.J., Ericson, J., and  
Peterson, A.S. (2009). The Rfx4 Transcription Factor Modulates Shh Signaling by Regional Control of  
625 Ciliogenesis. *Sci. Signal.* 2. <https://doi.org/10.1126/scisignal.2000602>.
- 626 60. Kistler, W.S., Baas, D., Lemeille, S., Paschaki, M., Seguin-Estevez, Q., Barras, E., Ma, W., Duteyrat, J.-  
L., Morlé, L., Durand, B., et al. (2015). RFX2 Is a Major Transcriptional Regulator of Spermiogenesis.  
627 *PLoS Genet* 11, e1005368. <https://doi.org/10.1371/journal.pgen.1005368>.

- 628 61. Wu, Y., Hu, X., Li, Z., Wang, M., Li, S., Wang, X., Lin, X., Liao, S., Zhang, Z., Feng, X., et al.  
629 (2016). Transcription Factor RFX2 Is a Key Regulator of Mouse Spermiogenesis. *Sci Rep* 6, 20435.  
<https://doi.org/10.1038/srep20435>.
- 630 62. Liu, M., Zhu, Y., Xing, F., Liu, S., Xia, Y., Jiang, Q., and Qin, J. (2020). The polycomb group protein  
631 PCGF6 mediates germline gene silencing by recruiting histone-modifying proteins to target gene  
632 promoters. *J Biol Chem* 295, 9712–9724. <https://doi.org/10.1074/jbc.RA119.012121>.
- 633 63. Otani, J., Nankumo, T., Arita, K., Inamoto, S., Ariyoshi, M., and Shirakawa, M. (2009). Structural basis  
634 for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX–DNMT3–DNMT3L  
635 domain. *EMBO Reports* 10, 1235–1241. <https://doi.org/10.1038/embor.2009.218>.
- 636 64. Guo, X., Wang, L., Li, J., Ding, Z., Xiao, J., Yin, X., He, S., Shi, P., Dong, L., Li, G., et al. (2015).  
637 Structural insight into autoinhibition and histone H3-induced activation of DNMT3A. *Nature* 517,  
638 640–644. <https://doi.org/10.1038/nature13899>.
- 639 65. Meissner, A., Mikkelsen, T.S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein,  
640 B.E., Nusbaum, C., Jaffe, D.B., et al. (2008). Genome-scale DNA methylation maps of pluripotent and  
641 differentiated cells. *Nature* 454, 766–770. <https://doi.org/10.1038/nature07107>.
- 642 66. Nassar, L.R., Barber, G.P., Benet-Pagès, A., Casper, J., Clawson, H., Diekhans, M., Fischer, C.,  
643 Gonzalez, J.N., Hinrichs, A.S., Lee, B.T., et al. (2023). The UCSC Genome Browser database: 2023  
644 update. *Nucleic Acids Research* 51, D1188–D1195. <https://doi.org/10.1093/nar/gkac1072>.
- 645 67. Akalin, A., Kormaksson, M., Li, S., Garrett-Bakelman, F.E., Figueiroa, M.E., Melnick, A., and Mason,  
646 C.E. (2012). methylKit: A comprehensive R package for the analysis of genome-wide DNA methylation  
647 profiles. *Genome Biol* 13, R87. <https://doi.org/10.1186/gb-2012-13-10-r87>.
- 648 68. Yang, M., Yu, H., Yu, X., Liang, S., Hu, Y., Luo, Y., Izsvák, Z., Sun, C., and Wang, J. (2022). Chemical-  
649 induced chromatin remodeling reprograms mouse ESCs to totipotent-like stem cells. *Cell Stem Cell*  
29, 400–418.e13. <https://doi.org/10.1016/j.stem.2022.01.010>.
- 650 69. Zhang, J., Zhang, M., Acampora, D., Vojtek, M., Yuan, D., Simeone, A., and Chambers, I. (2018).  
OTX2 restricts entry to the mouse germline. *Nature* 562, 595–599. [https://doi.org/10.1038/s41586-018-0581-5](https://doi.org/10.1038/s41586-<br/>018-0581-5).
- 651 70. Aubert, Y., Egolf, S., and Capell, B.C. (2019). The Unexpected Noncatalytic Roles of Histone Modifiers  
652 in Development and Disease. *Trends in Genetics* 35, 645–657. <https://doi.org/10.1016/j.tig.2019.06.004>.
- 653 71. Morgan, M.A.J., and Shilatifard, A. (2020). Reevaluating the roles of histone-modifying enzymes  
654 and their associated chromatin modifications in transcriptional regulation. *Nat Genet* 52, 1271–1281.  
655 <https://doi.org/10.1038/s41588-020-00736-4>.

- 650 72. Tahiliani, M., Mei, P., Fang, R., Leonor, T., Rutenberg, M., Shimizu, F., Li, J., Rao, A., and Shi, Y. (2007). The histone H3K4 demethylase SMCX links REST target genes to X-linked mental retardation.
- 651 Nature 447, 601–605. <https://doi.org/10.1038/nature05823>.
- 652 73. Endo, T., Romer, K.A., Anderson, E.L., Baltus, A.E., De Rooij, D.G., and Page, D.C. (2015). Periodic retinoic acid–STRA8 signaling intersects with periodic germ-cell competencies to regulate spermatogenesis. Proc. Natl. Acad. Sci. U.S.A. 112. <https://doi.org/10.1073/pnas.1505683112>.
- 653 74. Bowles, J., Knight, D., Smith, C., Wilhelm, D., Richman, J., Mamiya, S., Yashiro, K., Chawengsaksophak, K., Wilson, M.J., Rossant, J., et al. (2006). Retinoid Signaling Determines Germ Cell Fate in Mice. Science 312, 596–600. <https://doi.org/10.1126/science.1125691>.
- 654 75. Abildayeva, K., Berbée, J.F.P., Blokland, A., Jansen, P.J., Hoek, F.J., Meijer, O., Lütjohann, D., Gautier, T., Pillot, T., De Vente, J., et al. (2008). Human apolipoprotein C-I expression in mice impairs learning and memory functions. Journal of Lipid Research 49, 856–869. <https://doi.org/10.1194/jlr.M700518JLR200>.
- 655 76. Nielsen, A.Y., and Gjerstorff, M.F. (2016). Ectopic Expression of Testis Germ Cell Proteins in Cancer and Its Potential Role in Genomic Instability. Int J Mol Sci 17. <https://doi.org/10.3390/ijms17060890>.
- 656 77. Adebayo Babatunde, K., Najafi, A., Salehipour, P., Modarressi, M.H., and Mobasher, M.B. (2017). Cancer/Testis genes in relation to sperm biology and function. Iranian Journal of Basic Medical Sciences 20. <https://doi.org/10.22038/ijbms.2017.9259>.
- 657 78. Janic, A., Mendizabal, L., Llamazares, S., Rossell, D., and Gonzalez, C. (2010). Ectopic expression of germline genes drives malignant brain tumor growth in Drosophila. Science 330, 1824–1827. <https://doi.org/10.1126/science.1195481>.
- 658 79. Bonefas, K.M., and Iwase, S. (2021). Soma-to-germline transformation in chromatin-linked neurodevelopmental disorders? FEBS J. <https://doi.org/10.1111/febs.16196>.
- 659 80. Velasco, G., Walton, E.L., Sterlin, D., Hédouin, S., Nitta, H., Ito, Y., Fouyssac, F., Mégarbané, A., Sasaki, H., Picard, C., et al. (2014). Germline genes hypomethylation and expression define a molecular signature in peripheral blood of ICF patients: Implications for diagnosis and etiology. Orphanet J Rare Dis 9, 56. <https://doi.org/10.1186/1750-1172-9-56>.
- 660 81. Walton, E.L., Francastel, C., and Velasco, G. (2014). Dnmt3b Prefers Germ Line Genes and Centromeric Regions: Lessons from the ICF Syndrome and Cancer and Implications for Diseases. Biology (Basel) 3, 578–605. <https://doi.org/10.3390/biology3030578>.
- 661 82. Samaco, R.C., Mandel-Brehm, C., McGraw, C.M., Shaw, C.A., McGill, B.E., and Zoghbi, H.Y. (2012). Crh and Oprm1 mediate anxiety-related behavior and social approach in a mouse model of MECP2 duplication syndrome. Nat Genet 44, 206–211. <https://doi.org/10.1038/ng.1066>.

- 672 83. Stephens, M. (2016). False discovery rates: A new deal. Biostat, kxw041. <https://doi.org/10.1093/biostatistics/kxw041>.
- 673
- 674 84. Conway, J.R., Lex, A., and Gehlenborg, N. (2017). UpSetR: An R package for the visualization of intersecting sets and their properties. Bioinformatics 33, 2938–2940. <https://doi.org/10.1093/bioinformatics/btx364>.
- 675
- 676 85. Li, H., Liang, Z., Yang, J., Wang, D., Wang, H., Zhu, M., Geng, B., and Xu, E.Y. (2019). DAZL is a master translational regulator of murine spermatogenesis. Natl Sci Rev 6, 455–468. <https://doi.org/10.1093/nsr/nwy163>.
- 677
- 678 86. Mikedis, M.M., Fan, Y., Nicholls, P.K., Endo, T., Jackson, E.K., Cobb, S.A., De Rooij, D.G., and Page, D.C. (2020). DAZL mediates a broad translational program regulating expansion and differentiation of spermatogonial progenitors. eLife 9, e56523. <https://doi.org/10.7554/eLife.56523>.
- 679
- 680 87. Al Adhami, H., Vallet, J., Schaal, C., Schumacher, P., Bardet, A.F., Dumas, M., Chicher, J., Hammann, P., Daujat, S., and Weber, M. (2023). Systematic identification of factors involved in the silencing of germline genes in mouse embryonic stem cells. Nucleic Acids Research 51, 3130–3149. <https://doi.org/10.1093/nar/gkad071>.
- 681

682 **Figures and Tables**

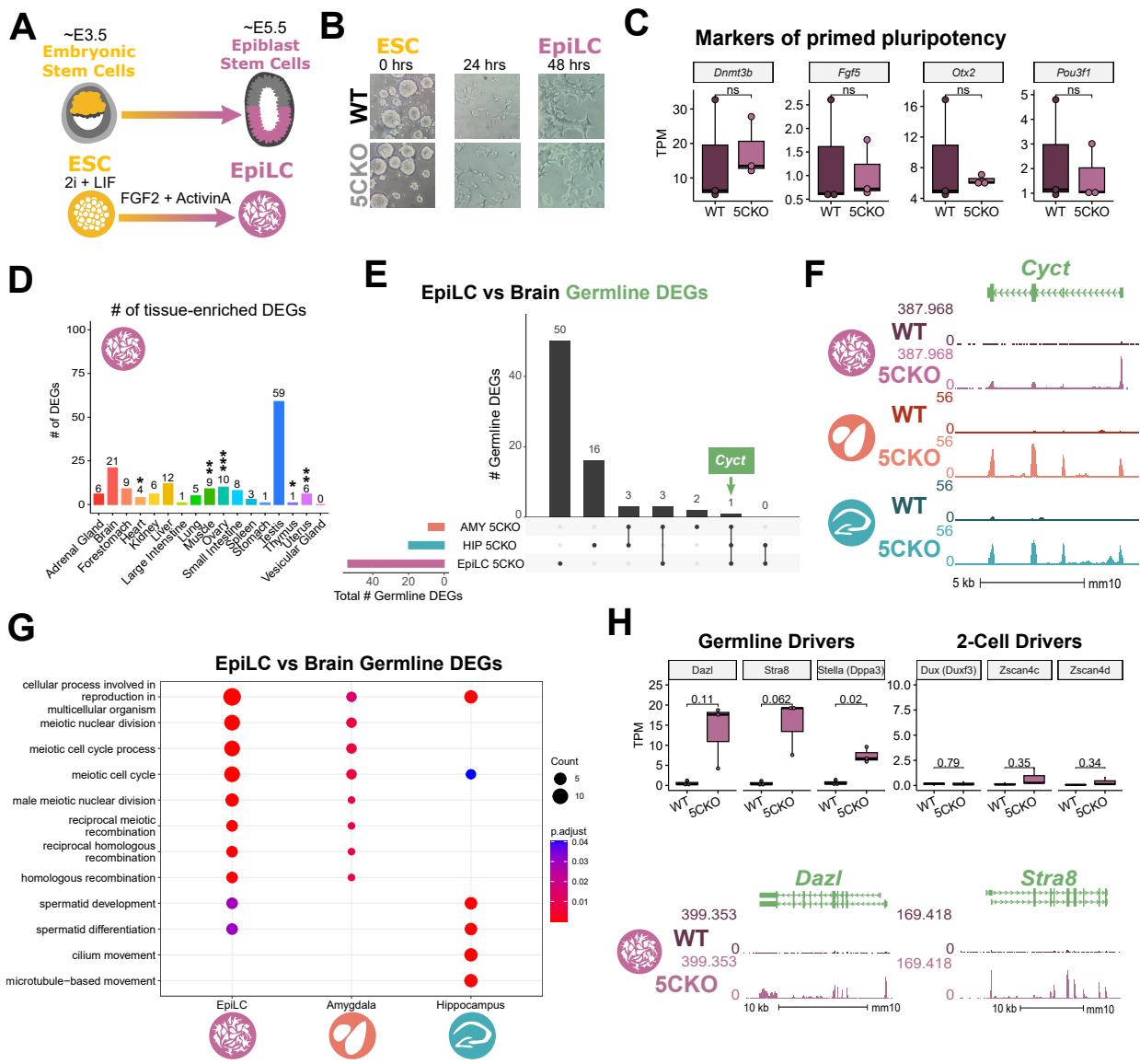
- 683     • Supplementary table 1: list of all germline genes.
- 684       – Columns to include:
- 685           \* KDM5C bound vs not
- 686           \* DEG in EpiLC, brain, both, neither (separate columns?)



**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, *Cyclin T*, that is dysregulated *Kdm5c*-KO EpiLCs. Welch's t-test, expression in transcripts per million (TPM).

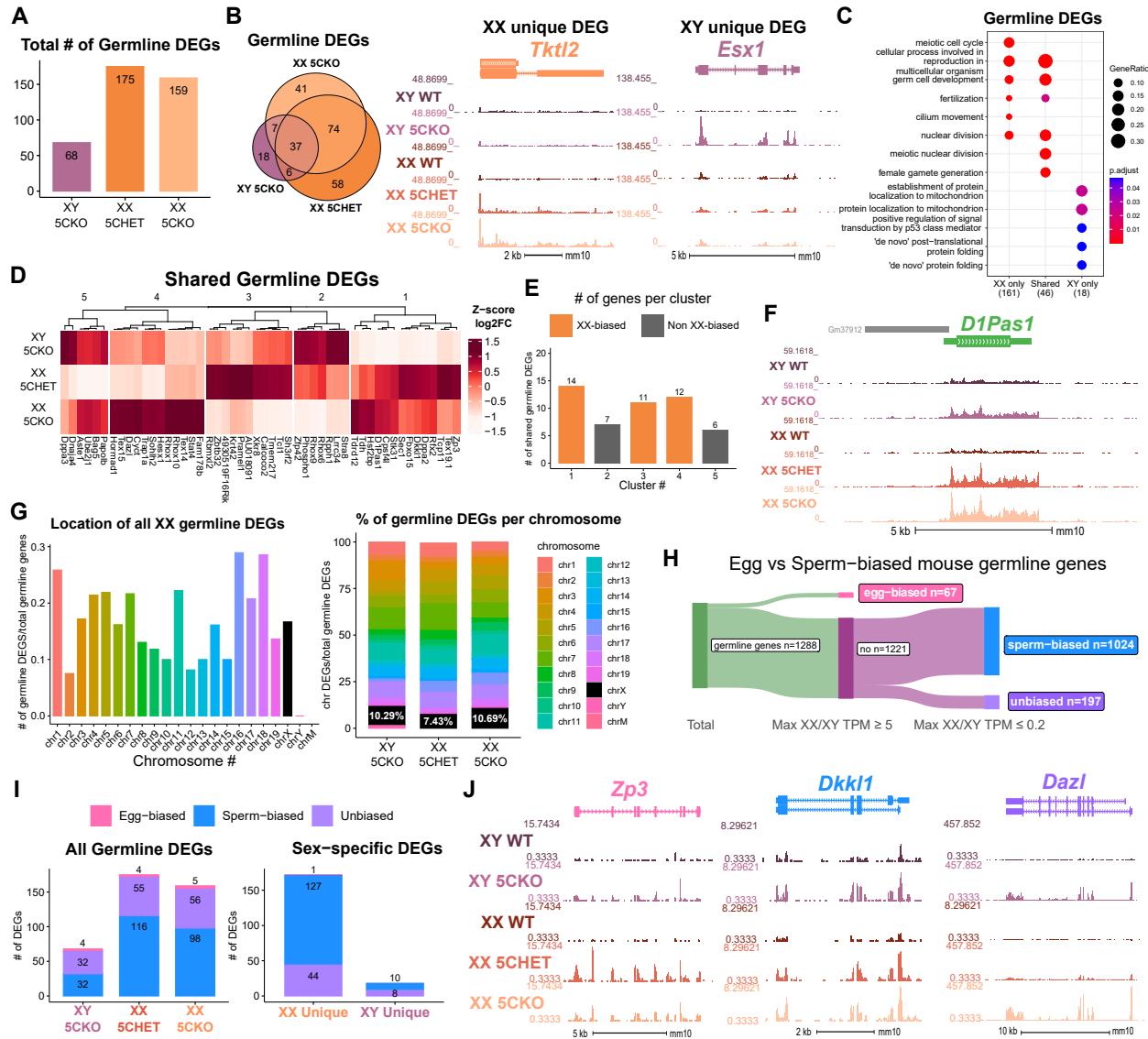
**D.** No significant difference in primed pluripotency marker expression in wild-type versus *Kdm5c*-KO EpiLCs. Welch's t-test, expression in transcripts per million (TPM).

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

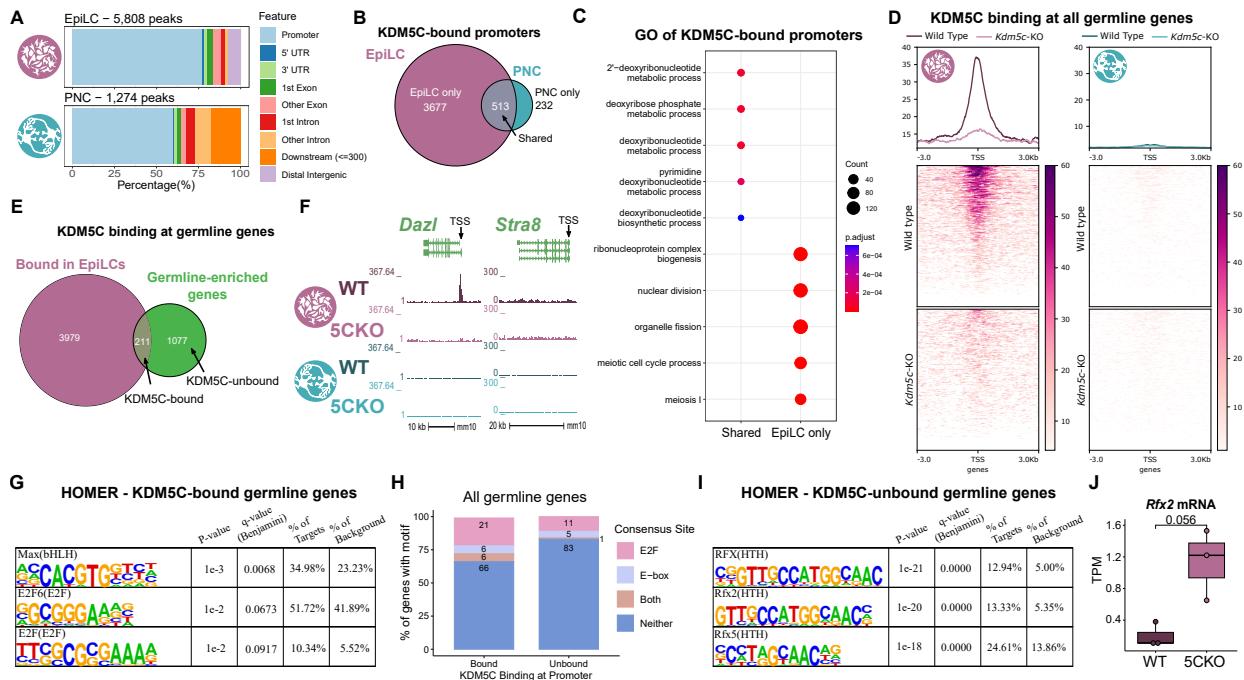
**F.** Upset plot displaying the overlap of germline DEGs expressed in *Kdm5c*-KO EpiLCs, amygdala (AMY), and hippocampus (HIP) RNA-seq datasets.

**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. p-value for Welch's t-test.

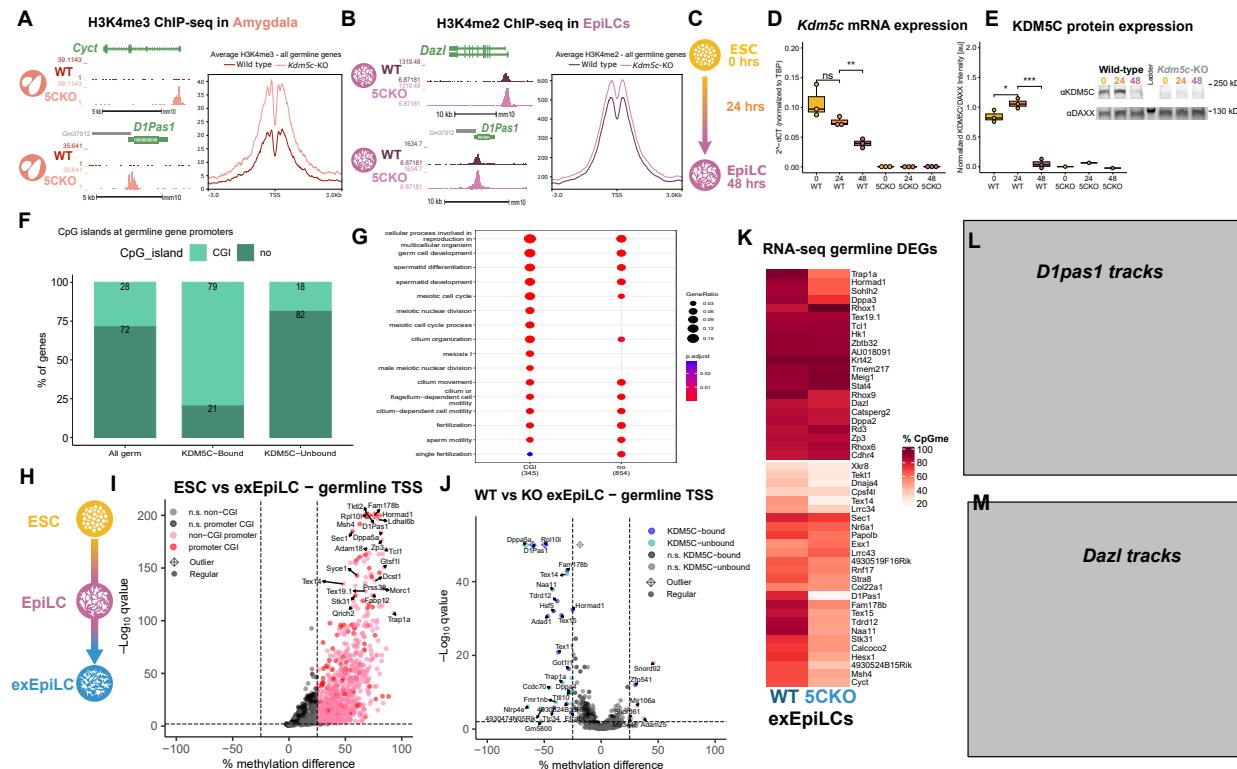


**Figure 4: Chromosomal sex influences *Kdm5c*-KO germline gene misexpression.** **A-B.** Expression of tissue-enriched genes (Li et al 2017) in the male *Kdm5c*-KO amygdala (A) and hippocampus (B). Left - MA plot of mRNA-seq. Right - Number of tissue-enriched differentially expressed genes (DEGs). \* 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 5: 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 5: 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



**Figure 6: 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 (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 comparison 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

## 687 Notes

### 688 Things to do

- 689     • Move dazl to new figure if other staining works
- 690     • Use Ddx4 staining (with dazl?). If not should add RNA to diagram because its a key germline gene.
- 691     • Motif analysis
  - 692         – Discussion - talk about motifs

### 693 Dazl

694     We were particularly interested in the aberrant transcription of *Dazl*, since it is essential for germ cell  
695 development and promotes the translation of germline mRNAs<sup>85</sup>. A significant portion of germline transcripts  
696 misexpressed in *Kdm5c*-KO EpiLCs are known binding targets of DAZL, including *Stra8*<sup>86</sup> ( $p = 1.698e-07$ ,  
697 Fisher's Exact Test). This suggests expression of DAZL protein could promote the translation of other  
698 aberrant germline transcripts, influencing their ability to impact *Kdm5c*-KO cellular function. We thus tested  
699 DAZL protein expression in *Kdm5c*-KO EpiLCs through immunocytochemistry (Figure 3I). We observed  
700 about 25% of *Kdm5c*-KO EpiLCs expressed DAZL protein and it was localized to the cytoplasm ( $p = 0.0015$ ,  
701 Welch's t-test), consistent with the pattern of DAZL expression in spermatogonia<sup>86</sup>. Altogether these results  
702 suggest tissue-specific genes are misexpressed during *Kdm5c*-KO embryogenesis, including key drivers of  
703 germline identity that can be translated into protein.

- 704     • We additionally found *Kdm5c*-KO EpiLCs ectopically express DAZL protein that is localized to the  
705         cytoplasm, similar to its morphology in spermatogonia<sup>86</sup>. **note: maybe just put in results.** Could  
706         move around depending upon if I get pheno working.

### 707 Discussion notes

- 708     • For other paper:
  - 709         – for methods: Heatmaps of gene expression were generated using the base R functions scale and  
710             hclust and visualized using the R package ComplexHeatmap (v2.12.1).
  - 711         –     \* Might be good to look at for retinoic acid paper (WT germ expression): <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0005654>
- 713     • Talk about different motifs if we do see differences there and if it explains direct vs indirect dysregulation
  - 714         – Could also mention PRC1.6/KDM5C binds pcgf6 in ESCs
- 715     • Maybe talk about other regulators/if anything is known about them long term? Or save for 2nd paper.

- 716            – another germline repressor: <https://pubmed.ncbi.nlm.nih.gov/35345626/>
- 717        • end with something like: this indicates kdm5c not only modulates pluripotency and self-renewal in  
718            ESCs, but also has a role in long-term silencing of germline genes
- 719            – then transition into the long term silencing mechanism paragraph
- 720        • KDM5C is dynamically regulated during the window of embryonic germline gene silencing, our EpiLC  
721            ChIP-seq is likely catching the tail end of KDM5C's main involvement.
- 722        • Our results indicate KDM5C is a pivotal repressor of germline programs in somatic cells.
- 723        • Studies on other chromatin regulators have focused on key marker genes to identify germline gene  
724            misexpression, such as *Dazl*.
- 725        • Using a germline-depleted mouse model, we curated a list of mouse germline-enriched genes to  
726            globally assess germline gene dysregulation.
- 727        • Many of the genes dysregulated in the mature *Kdm5c*-KO brain were important for late stages of  
728            spermatogenesis, such as those important for sperm flagellar structure. Contrastingly, *Kdm5c*-KO  
729            EpiLCs aberrantly expressed key drivers of germ cell identity, including *Dazl* and *Stra8*.
- 730        • While a significant portion of KDM5C-bound promoters are germline genes, many germline genes  
731            expressed during *Kdm5c*-KO embryogenesis are not directly bound by kdm5c.
- 732        • One notable example is *Stra8*, a transcription factor important for germ cell specific transcription and  
733            meiotic initiation
- 734        • The including the demarcation between soma and germline fates.
- 735            the *Kdm5c*-KO brain also Previous work identified testis-enriched genes in *Kdm5c* genes
- 736            –
- 737            – However unlike the gonadal-biased DEGs,
- 738        • The demarcation of the germline versus soma is a key feature of multicellularity and sexually dimorphic  
739            reproduction
- 740        • Anything known about tissue-biased gene expression in other H3K4me regulators?
- 741        • Our data suggests the germline developmental program is occurring ectopically as *Kdm5c*-KOs pro-  
742            gresses through somatic tissue development
- 743        • tissue-biased gene expression:

- 744 • Altogether, these data indicate that while some germline genes are misexpressed due to direct loss of  
745 KDM5C binding during emryogenesis, secondary downstream mechanisms can also promote their  
746 aberrant transcription.
- 747 • Papers to read/reference:  
748 – Reconstitution of the Mouse Germ Cell Specification Pathway in Culture by Pluripotent Stem Cells:  
749 [https://www.cell.com/fulltext/S0092-8674\(11\)00771-9](https://www.cell.com/fulltext/S0092-8674(11)00771-9)  
750 – 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/>

752 **Figure outline:**

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

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

765 **Figure 3: *Kdm5c*-KO epiblast-like cells express key drivers regulators of germline identity** \* A) ESC  
766 to EpiLC differentiation Left - Morphology is unchanged, \* B) 5CKO EpiLCs express EpiLC differentiation  
767 genes similar to WT lvs \* C) Male EpiLCs express germline genes (example Cyct again) \* Overlap between  
768 brain and EpiLC germline genes - show they're mostly unique \* GO of Brain and EpiLC germline genes  
769 (meiotic enriched) \* Bigwigs or TPM of master regulators \* Show that while some are also 2-cell genes  
770 (Dazl), 2-cell specific genes aren't dysregulated (Zscan4). Important point because published KDM5C dazl  
771 paper is saying KDM5C is a 2-cell regulator, but as far as I can tell only genes shared between germline and  
772 2-cell are dysregulated.

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

774 **Figure 4: Loss of KDM5C's catalytic activity impairs DNAme placement and long-term silencing of  
775 germline genes** \* Increase in H3K4me3 in *Kdm5c*-KO amygdala at germline genes \* Increase in H3K4me2  
776 in EpiLCs at germline genes \* *Kdm5c* binding in EpiLCs vs PNCs to show that germline repression is  
777 happening in early embryo \* Previous studies only looked at ESCs, unknown if catalytic activity is required

778 for long-term repression, especially since DNA methylation is placed later). KDM5C RNA and protein ESC →  
779 EpiLC (increasing then decreasing) \* RNA expression of germline genes with catalytic dead rescue (Ilakkia)  
780 \* DNA methylation in WT and 5CKO EpiLCs (Ilakkia)

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

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

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

- 800 • Loss of KDM5C can result in the misexpression of genes typically only found in the testis  
801 – Misexpression of tissue-enriched genes hasn't been systematically characterized - Unclear if  
802 these genes are exceptions or if other tissue-specific genes are dysregulated  
803 – Interestingly, these genes (Cyct, D1pas1) typically function in the germline  
804 – Germ cells (meiotic cells) are typically distinguished from somatic cells very early on in embryogen-  
805 esis and is a key feature of multicellularity  
806 – Chromatin regulators are very important for decommissioning germline genes and act successively  
807 the embryo implants into the uterine wall  
808 – Most studies have focused on ESCs, which have a similar transcriptome to germ cells / 2-cells  
809 – recently, KDM5C was shown to repress DAZL in ESCs, independent of its catalytic activity  
810 – However, DNA methylation is lost in the mature 5CKO brain, DNA methylation is placed later  
811 and it's Unclear if it's required for long-term repression (maybe too specific, just trying to go  
812 into the fact that the mechanism is partially understood but unclear)

813 – Systematic characterization of ectopic germline genes hasn't been done  
814 \* unknown if other germline-enriched genes are dysregulated, including oocyte-specific genes  
815 \* Crucially, it's unknown if misexpression of the germline program leads to functional conse-  
816 quences in 5CKO cells.

817 **Germline gene repression background:**

818 Interestingly, some of the ectopic testis transcripts identified in the *Kdm5c*-KO brain are typically ex-  
819 pressed in germ cells<sup>13</sup>. Unlike somatic cells, germ cells (e.g. sperm and eggs) undergo meiosis and pass  
820 on their genetic material to the next generation. The germline and the soma are typically distinguished during  
821 early embryogenesis, when germline genes are silenced in epiblast stem cells soon after implantation and  
822 only reactivated in a subset to form the germline. Chromatin regulators play a key role in decommissioning  
823 germline genes as the embryo transitions from naïve to primed pluripotency by placing repressive histone  
824 H2A lysine 119 monoubiquitination (H2AK119ub1)<sup>16</sup>, histone 3 lysine 9 trimethylation (H3K9me3)<sup>16,17</sup>, and  
825 DNA CpG methylation<sup>17–19</sup> at germline gene promoters. KDM5C may also be involved in this early decom-  
826 missioning of germline genes, as re-expression of KDM5C in neurons fails to suppress their dysregulation<sup>13</sup>.  
827 In support of this, KDM5C was very recently shown to repress *Deleted in azoospermia like (Dazl)*, a key  
828 regulator of germline development, in mouse embryonic stem cells (ESCs)<sup>45,87</sup>. In support of this, two  
829 independent screens in mouse embryonic stem cells (ESCs) recently identified KDM5C as a repressor of  
830 *Deleted in azoospermia like (Dazl)*, a key regulator of germline development. However, KDM5C's role in  
831 embryonic germline gene repression is currently unclear, given that Dazl is also expressed in ESCs and in  
832 the 2-cell stage and germline gene misexpression has yet to be globally characterized during *Kdm5c*-KO  
833 embryogenesis.