

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 polycomb repressive complex 1.6 (PRC1.6), which  
231 contains transcription factor heterodimers E2F6/DP1 and MGA/MAX that respectively bind E2F and E-box  
232 motifs<sup>56</sup>. PRC1.6 members have been proposed to recruit KDM5C to germline gene promoters, given  
233 their association in HeLa cells and ESCs<sup>45,57</sup> and that KDM5C itself does not contain sequence-specific  
234 binding domains<sup>8</sup>. To elucidate how KDM5C is recruited to germline genes, we used HOMER<sup>58</sup> to identify  
235 transcription factor motifs enriched at KDM5C-bound or unbound germline gene promoters (TSS ± 500

236 bp, q-value < 0.1). MAX and E2F6 binding sites were significantly enriched at germline genes bound by  
237 KDM5C in EpiLCs, but not at germline genes unbound by KDM5C (MAX q-value: 0.0068, E2F6 q-value:  
238 0.0673, E2F q-value: 0.0917) (Figure 5G). One third of KDM5C-bound promoters contained the consensus  
239 sequence for either E2F6 (E2F, 5'-TCCCGC-3'), MGA (E-box, 5'-CACGTG-3'), or both, but only 17% of  
240 KDM5C-unbound genes contained these motifs (Figure 5H). KDM5C-unbound germline genes were instead  
241 enriched for multiple RFX transcription factor binding sites (RFX q-value < 0.0001, RFX2 q-value < 0.0001,  
242 RFX5 q-value < 0.0001) (Figure 5I, Supplementary figure 1D). RFX transcription factors bind X-box motifs<sup>59</sup>  
243 to promote ciliogenesis<sup>60,61</sup> and among them is RFX2, a central regulator of post-meiotic spermatogenesis<sup>62,63</sup>.  
244 Interestingly, RFX2 mRNA is derepressed in *Kdm5c*-KO EpiLCs (Figure 5J), however it is also not a direct  
245 target of KDM5C (Supplementary figure 1E). Thus, RFX2 is a candidate transcription factor for driving the  
246 ectopic expression of KDM5C-unbound germline genes in *Kdm5c*-KO cells.

247 **KDM5C is recruited to CpG islands within germline gene promoters to facilitate de  
248 novo DNA methylation**

249 In the early embryo, germline gene promoters are initially decorated with repressive histone modifications  
250 and are then silenced long-term via DNA CpG methylation (CpGme)<sup>16,17,40,64</sup>. Our results above indicate  
251 KDM5C also acts at germline gene promoters during this time period. However, how KDM5C interacts with  
252 other germline gene silencing mechanisms is currently unclear. KDM5C is generally thought to suppress  
253 transcription through erasure of histone 3 lysine 4 di- and trimethylation (H3K4me2/3)<sup>8</sup>, yet KDM5C's  
254 catalytic activity was recently shown to be dispensable for suppressing *Dazl* in undifferentiated ESCs<sup>45</sup>. Since  
255 H3K4me3 impedes *de novo* CpGme placement<sup>65,66</sup>, KDM5C's catalytic activity may instead be required  
256 in the post-implantation embryo for long-term silencing of germline genes. In support of this, CpGme is  
257 markedly reduced at two germline gene promoters in the *Kdm5c*-KO adult hippocampus<sup>13</sup>.

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

270 Germline genes accumulate CpG methylation (CpGme) at CpG islands (CGIs) during the transition from  
271 naïve to primed pluripotency<sup>19,40,67</sup>, reaching peak methylation levels when differentiated into EpiLCs for  
272 96 hours (extended EpiLCs, exEpiLCs)<sup>17</sup>. We first identified how many germline genes contained CGIs  
273 using the UCSC genome browser<sup>68</sup> and found out of 1,288 germline-enriched genes, only 356 (27.64%)  
274 had promoter CGIs (TSS ± 500 bp) (Figure 6F). CGI-containing germline genes were enriched for meiotic  
275 gene ontologies, including meiotic nuclear division (GO:XXXX, p.adj) and meiosis I (GO:XXXX, p.adj) when  
276 compared to CGI-free genes (Figure 6G). Although a minor portion of germline gene promoters contained  
277 CGIs, CGIs strongly determined KDM5C's recruitment to germline genes (FISHER'S XXXX), with 79.15% of  
278 KDM5C-bound germline genes containing CGIs (Figure 6G).

279 To assess how KDM5C loss impacts initial CpGme placement at germline gene promoters, we performed  
280 whole genome bisulfite sequencing (WGBS) in male wild-type and *Kdm5c*-KO ESCs and 96 hour extended  
281 EpiLCs (exEpiLCs) (Figure 6H). We first identified which germline gene promoters significantly gained  
282 CpGme in wild-type cells during nESC to exEpiLCs differentiation (methylKit<sup>69</sup>, q < 0.01, |methylation  
283 difference| >= 25%, TSS ± 500 bp). In wild-type cells, the majority of germline genes gained substantial  
284 CpGme at their promoter during differentiation (60.08%), regardless if their promoter contained a CGI (Figure  
285 6I).

286 We then identified germline gene promoters differentially methylated in wild-type versus *Kdm5c*-KO  
287 exEpiLCs (methylKit, q < 0.01, |methylation difference| > 25%, TSS ± 500 bp) and found 28 germline  
288 promoters were significantly hypomethylated with *Kdm5c* loss (Figure 6J). Approximately half of germline  
289 promoters hypomethylated in *Kdm5c*-KO exEpiLCs are direct targets of KDM5C in EpiLCs (13 out of 28  
290 hypomethylated DMRs). Promoters that showed the most robust loss of CpGme (low q-values) harbored  
291 CGIs (Figure 6J).

292 We then compared average percent CpGme at germline gene promoters with or without CGIs. While  
293 the degree of CpGme was largely unchanged for non-CGI promoters, CGI promoters on average had a  
294 significant reduction in CpGme with KDM5C loss (Figure 6K) (Non-CGI promoters p = 0.0846, CGI promoters  
295 p = 0.0081, Mann-Whitney U test). Significantly hypomethylated promoters included genes consistently  
296 dysregulated across multiple *Kdm5c*-KO RNA-seq datasets<sup>13</sup>, such as *Naa11* and *D1PAs1* (methylation  
297 difference = -60.03%, q-value = 3.26e-153) (Figure 6L). Surprisingly, we found only a modest reduction in  
298 CpGme at *Dazl*'s promoter (methylation difference = -6.525%, q-value = 0.0159) (Figure 6M). Altogether,  
299 these results demonstrate KDM5C is recruited to germline gene CGIs to promote CpGme at germline gene  
300 promoters during early embryogenesis, however some loci can compensate for KDM5C loss through other  
301 silencing mechanisms, even when retaining H3K4me.

302 **Discussion**

303 In the above study, we demonstrate KDM5C's pivotal role in the development of tissue identity. We  
304 first characterized tissue-enriched genes expressed within the *Kdm5c*-KO brain and identified substantial  
305 dysregulation of testis, liver, muscle, and ovary-enriched genes. Testis genes significantly enriched within  
306 the *Kdm5c*-KO amygdala and hippocampus are specific to the germline and not expressed within somatic  
307 cells. *Kdm5c*-KO epiblast-like cells (EpiLCs) aberrantly express key drivers of germline identity and meiosis,  
308 including *Dazl* and *Stra8*, while the adult brain primarily expresses genes important for late spermatogenesis.  
309 We demonstrated that although *Kdm5c* mutant sex did not influence whether sperm or egg-specific genes  
310 were misexpressed, female EpiLCs are more sensitive to germline gene de-repression. Germline genes  
311 can become aberrantly expressed in *Kdm5c*-KO cells via an indirect mechanism, such as activation via  
312 ectopic RFX transcription factors. Finally, we found KDM5C is dynamically regulated during ESC to EpiLC  
313 differentiation to promote long-term germline gene silencing through DNA methylation at CpG islands.  
314 Therefore, we propose KDM5C plays a fundamental role in the development of tissue identity during  
315 early embryogenesis, including the establishment of the soma-germline boundary. By systematically  
316 characterizing KDM5C's role in germline gene repression, including its interaction with known silencing  
317 mechanisms, we unveiled repressive mechanisms governing distinct classes of germline gene in somatic  
318 lineages. Furthermore, these data provide molecular footholds which can be exploited to test the overarching  
319 contribution of ectopic germline gene expression upon neurodevelopment.

320 Although eggs and sperm employ the same transcriptional programs for shared functions, e.g. PGC  
321 formation, meiosis, and genome defense, some germline genes are sex specific. We found *Kdm5c* mutant  
322 males and females expressed both sperm and egg-biased genes, indicating the mechanism of derepression  
323 is independent of a given germline gene's sex. However, organismal sex did greatly influence the degree of  
324 germline gene dysregulation, as female *Kdm5c*-KO EpiLCs had over double the number of germline-enriched  
325 DEGs compared to males. The lack of X-linked gene enrichment in females suggests that this greater  
326 sensitivity to germline gene misexpress is not due to XCI defects previously reported in *Kdm5c*-KO females<sup>41</sup>.  
327 Sex differences in germline gene suppression may be a consequence of females having a higher dose of  
328 KDM5C than males, due to its escape from XCI<sup>48–51</sup>. Intriguingly, females with heterozygous loss of *Kdm5c*  
329 also had over double the number of germline DEGs than hemizygous knockout males, even though their  
330 level of KDM5C should be roughly equivalent to that of wild-type males. Altogether, these results suggests  
331 germline gene silencing mechanims differ between males and females, which warrants further study to  
332 identify the biological implications and underlying mechanisms.

333 It is important to note that while we highlighted KDM5C's regulation of germline genes, some germline-  
334 enriched genes are also expressed at the 2-cell stage and in naïve ESCs for their role in pluripotency and  
335 self-renewal<sup>39,45,70,71</sup>. For example, while *Dazl* is primarily known for regulating the translation of germline-  
336 specific mRNAs, it is also expressed at the 2-cell stage<sup>71</sup>, in naïve ESCs<sup>37</sup>, and in the inner cell mass<sup>37</sup>.

337 KDM5C was thought to promote the 2-cell-to-ESC transition, given *Dazl* and *Zscan4c* are de-repressed  
338 in *Kdm5c*-KO ESCs<sup>45</sup>. Although expressed in naïve ESCs, “self-renewal” germline genes like *Dazl* are  
339 silenced during ESC differentiation into epiblast stem cells/EpiLCs<sup>17,40</sup>. We found that while *Kdm5c*-KO  
340 EpiLCs also express *Dazl*, they did not express 2-cell-specific genes. These data suggest the 2-cell-like  
341 state observed in *Kdm5c*-KO ESCs likely reflects KDM5C’s primary role in germline gene repression.  
342 Germline gene misexpression in *Kdm5c*-KO EpiLCs may indicate they are differentiating into primordial germ  
343 cell-like cells (PGCLCs), rather than de-differentiating into 2-cell-like cells<sup>32,33,35</sup>. Yet, *Kdm5c*-KO EpiLCs  
344 had normal cellular morphology and properly expressed markers for primed pluripotency, including *Otx2*  
345 which is known to repress EpiLC differentiation into PGCs/PGCLCs<sup>72</sup>. In addition to unimpaired EpiLC  
346 differentiation, *Kdm5c*-KO gross brain morphology is overall normal<sup>12</sup> and hardly any brain-specific genes  
347 were significantly dysregulated. Thus, ectopic germline gene expression occurs in parallel with proper  
348 *Kdm5c*-KO differentiation and somatic development.

349 While promoter CGIs are typically resistant to CpGme<sup>73</sup>, CGIs at germline promoters are highly methylated in somatic cells<sup>74</sup>. In EpiLCs, loss of KDM5C binding at a subset of germline gene promoters,  
350 e.g. *D1Pas1* and *Naa11*, strongly impaired CGI methylation, which resulted in their long-lasting de-repression  
351 into adulthood. Removal of H3K4me2/3 at CGIs is a plausible mechanism for KDM5C-mediated germline  
352 gene suppression<sup>13,75</sup>, given H3K4me2/3 primarily do not colocalize with CpGme<sup>74</sup> and can oppose DNMT3  
353 activity<sup>65,66</sup>. However, emerging work indicates many histone-modifying enzymes have non-catalytic functions that influence gene expression, sometimes even more potently than their catalytic roles<sup>76,77</sup>. Indeed,  
354 KDM5C’s catalytic activity was recently found to be dispensable for repressing *Dazl* in ESCs<sup>45</sup>. In our study,  
355 *Dazl*’s promoter still gained CpGme in *Kdm5c*-KO exEpiLCs, even with elevated H3K4me2. *Dazl* and a few  
356 other germline gene CGIs use multiple repressive mechanisms to facilitate CpGme<sup>16,17,46,47</sup>. Together, this  
357 suggests alternative silencing mechanisms are sufficient to recruit DNMT3s to some germline CGIs, while  
358 others may require KDM5C-mediated H3K4me removal to overcome promoter CGI escape from CpGme.  
359 Furthermore, these results indicate the requirement for catalytic activity for the same class of genes can  
360 change depending upon the locus and developmental stage.

363 By generating a comprehensive list of mouse germline-enriched genes, we were able to reveal distinct  
364 repressive mechanisms governing early versus late-stage germline developmental programs. In EpiLCs,  
365 KDM5C was highly enriched at germline promoters containing CGIs, many of which had roles in early  
366 germ cell formation and meiosis. However, over 70% of germline-enriched gene promoters lacked CGIs,  
367 including the many KDM5C-unbound germline genes that were de-repressed in *Kdm5c*-KO cells. CGI-free,  
368 KDM5C-unbound germline genes were primarily late-stage spermatogenesis genes and significantly  
369 enriched for RFX2 binding sites, a central regulator of spermiogenesis<sup>62,63</sup>. These data suggest that once  
370 activated during early embryogenesis, drivers of germline identity like *Rfx2* and *Dazl* turn on downstream  
371 germline programs that can loosely mimic germ cell development, ultimately culminating in the expression of  
372 spermiogenesis genes in the adult *Kdm5c*-KO brain. Therefore, we propose KDM5C is recruited to germline

373 genes that shape germ cell formation via their promoter CGIs to act as break against runaway activation of  
374 germline-specific programs. Therefore, we propose KDM5C is recruited via promoter CGIs to genes that  
375 shape germ cell formation and acts as break against runaway activation of germline-specific programs.

376 The above work provides the mechanistic foundation for KDM5C-mediated repression of tissue and  
377 germline-specific genes. However, the contribution of these ectopic, tissue-specific genes towards *Kdm5c*-  
378 KO neurological impairments is still unknown. In addition to germline genes, we also identified significant  
379 enrichment of muscle, liver, and even ovary-biased transcripts within the male *Kdm5c*-KO brain. Intriguingly,  
380 select liver and muscle-biased DEGs do have known roles within the brain, such as the liver-enriched lipid  
381 metabolism gene *Apolipoprotein C-I (Apoc1)*<sup>26</sup>. *APOC1* dysregulation is implicated in Alzheimer's disease in  
382 humans<sup>27</sup> and overexpression of *Apoc1* in the mouse brain can impair learning and memory<sup>78</sup>. KDM5C may  
383 therefore be crucial for neurodevelopment by fine-tuning the expression of tissue-enriched, dosage-sensitive  
384 genes like *Apoc1*. Given germline genes have no known functions within the brain, their impact upon  
385 neurodevelopment is currently unknown. Ectopic testicular germline transcripts have been observed in a  
386 variety of cancers<sup>79,80</sup>, including brain tumors in *Drosophila* and mammals<sup>81,82</sup>, indicating their dysregulation  
387 may promote genome instability and cellular de-differentiation. Intriguingly, some models for other chromatin-  
388 linked neurodevelopmental disorders also display impaired soma-germline demarcation<sup>7,83-86</sup>. Like KDM5C,  
389 the chromatin regulators underlying these conditions - DNA methyltransferase 3b (DNMT3B), H3K9me1/2  
390 methyltransferases G9A/GLP, methyl-CpG -binding protein 2 (MECP2) - primarily silence gene expression.  
391 Thus, KDM5C is among a growing cohort of chromatin-linked neurodevelopmental disorders with similar  
392 erosion of the germline versus soma boundary. Further research is required to determine the impact of these  
393 germline genes and the extent to which this phenomenon occurs in humans.

## 394 Materials and Methods

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

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

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

406 **Cell culture**

407 We utilized our previously established cultures of male wild-type and *Kdm5c* knockout (-KO) embryonic  
408 stem cells<sup>41</sup>. Sex was confirmed by genotyping *Uba1/Uba1y* on the X and Y chromosomes with the following  
409 primers: 5'-TGGATGGTGTGCCAATG-3', 5'-CACCTGCACGTTGCCCT-3'. Deletion of *Kdm5c* was  
410 confirmed through the primers 5'-ATGCCCATATTAAGAGTCCTG-3', 5'-TCTGCCTTGATGGGACTGTT-3',  
411 and 5'-GGTTCTAACACTCACATAGTG-3'.

412 Embryonic stem cells (ESCs) and epiblast-like cells were cultured using previously established  
413 methods<sup>36</sup>. Briefly, ESCs were initially cultured in primed ESC (pESC) media consisting of KnockOut  
414 DMEM (Gibco#10829-018), fetal bovine serum (Gibco#A5209501), KnockOut serum replacement  
415 (Invitrogen#10828-028), Glutamax (Gibco#35050-061), Anti-Anti (Gibco#15240-062), MEM Non-essential  
416 amino acids (Gibco#11140-050), and beta-mercaptoethanol (Sigma#M7522). They were then transitioned  
417 into ground-state, "naïve" ESCs (nESCs) by culturing for four passages in N2B27 media containing  
418 DMEM/F12 (Gibco#11330-032), Neurobasal media (Gibco#21103-049), Gluamax, Anti-Anti, N2 sup-  
419 plement (Invitrogen#17502048), and B27 supplement without vitamin A (Invitrogen#12587-010), and  
420 beta-mercaptoethanol. Both pESC and nESC media were supplemented with 3 μM GSK3 inhibitor  
421 CHIR99021 (Sigma #SML1046-5MG), 1 μM MEK inhibitor PD0325901 (Sigma #PZ0162-5MG), and 1,000  
422 units/mL leukemia inhibitory factor (LIF, Millipore#ESG1107).

423 nESCs were differentiated into epiblast-like cells (EpiLCs, 48 hours) and extendend EpiLCs (exEpiLCs,  
424 96 hours) by culturing in N2B27 media containing DMEM/F12, Neurobasal media, Gluamax, Anti-Anti,  
425 N2 supplement, B27 supplement (Invitrogen#17504044), beta-mercaptoethanol, fibroblast growth factor 2  
426 (FGF2, R&D Biotechne 233-FB), and activin A (R&D Biotechne 338AC050CF), as previously described<sup>36</sup>.

427 **RNA sequencing (RNA-seq)**

428 After ensuring read quality via FastQC (v0.11.8), reads were then mapped to the mm10 *Mus musculus*  
429 genome (Gencode) using STAR (v2.5.3a), during which we removed duplicates and kept only uniquely  
430 mapped reads. Count files were generated by FeatureCounts (Subread v1.5.0), and BAM files were  
431 converted to bigwigs using deeptools (v3.1.3) and visualized by the UCSC genome browser. RStudio (v3.6.0)  
432 was then used to analyze counts files by DESeq2 (v1.26.0)<sup>22</sup> to identify differentially expressed genes  
433 (DEGs) with a q-value (p-adjusted via FDR/Benjamini–Hochberg correction) less than 0.1 and a log2 fold  
434 change greater than 0.5. For all DESeq2 analyses, log2 fold changes were calculated with IfcShrink using  
435 the ashR package<sup>87</sup>. MA-plots were generated by ggpahr (v0.6.0), and Eulerr diagrams were generated by  
436 eulerr (v6.1.1). Boxplots and scatterplots were generated by ggpahr (v0.6.0) and ggplot2 (v3.3.2). The Upset  
437 plot was generated via the package UpSetR (v1.4.0)<sup>88</sup>. Gene ontology (GO) analyses were performed by  
438 the R package enrichPlot (v1.16.2) using the biological processes setting and compareCluster.

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

440 ChIP-seq reads were aligned to mm10 using Bowtie1 (v1.1.2) allowing up to two mismatches. Only  
441 uniquely mapped reads were used for analysis. Peaks were called using MACS2 software (v2.2.9.1) using  
442 input BAM files for normalization, with filters for a q-value < 0.1 and a fold enrichment > 1. We removed  
443 “black-listed” genomic regions that often give aberrant signals. Common peak sets were obtained in R via  
444 DiffBind (v3.6.5). In the case of KDM5C ChIP-seq, *Kdm5c*-KO peaks were then subtracted from wild-type  
445 samples using bedtools (v2.25.0). Peak proximity to genome annotations was determined by ChIPSeeker  
446 (v1.32.1). Gene ontology (GO) analyses were performed by the R package enrichPlot (v1.16.2) using the  
447 biological processes setting and compareCluster. Enriched motifs were identified using HOMER<sup>58</sup>. Average  
448 binding across the genome was visualized using deeptools (v3.1.3). Bigwigs were visualized using the  
449 UCSC genome browser.

450 **CpG island (CGI) analysis**

451 Locations of CpG islands were determined through the mm10 UCSC genome browser CpG island track.  
452 CGI coordinates were then annotated using ChIPseeker (v1.32.1) and filtered for promoters of germline  
453 genes (TSS ± 500).

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

455 Genomic DNA (gDNA) from naïve ESCs and extended EpiLCs was extracted using the Wizard Genomic  
456 DNA Purification Kit (Promega A1120), following the instructions for Tissue Culture Cells. gDNA from  
457 two wild-types and two *Kdm5c*-KOs of each cell type was sent to Novogene for WGBS using the Illumina  
458 NovaSeq X Plus platform and sequenced for 150bp paired-end reads (PE150). Reads were adapter and  
459 quality trimmed with Trim Galore (v0.6.10) and aligned to the mm10 genome using Bismark (v0.22.1).  
460 Analysis of differential methylation at germline gene promoters was performed using methylKit (v1.28.0) with a  
461 minimum coverage of 3 paired reads, a percentage cut-off of 25%, and q-value of 0.01. Average percentage  
462 methylation at germline gene promoters was determined via methylKit (v1.28.0). Methylation bedgraph  
463 tracks were generated via Bismark and visualized using the UCSC genome browser.

464 **Data availability**

465 **Published datasets**

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

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

475 **Data analysis**

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

478 **Acknowledgements**

479 We thank Drs. Sundeep Kalantry, Milan Samanta, and Rebecca Malcore for providing protocols and  
480 expertise in culturing mouse ESCs and EpiLCs, as well as providing wild-type and *Kdm5c*-KO ESCs used in  
481 this study. We thank Dr. Jacob Mueller for his insight in germline gene regulation and directing us to the  
482 germline-depleted mouse models. We also thank Drs. Stephanie Bielas, Michael Sutton, Donna Martin, and  
483 the members of the Iwase, Sutton, Bielas, and Martin labs for helpful discussions and critiques of the data.  
484 We thank members of the University of Michigan Reproductive Sciences Program for providing feedback  
485 throughout the development of this work. This work was supported by grants from the National Institutes  
486 of Health (NIH) (National Institute of Neurological Disorders and Stroke: NS089896, 5R21NS104774, and  
487 NS116008 to S.I.), Farrehi Family Foundation Grant (to S.I.), the University of Michigan Career Training in  
488 Reproductive Biology (NIH T32HD079342, to K.M.B), and the NIH Early Stage Training in the Neurosciences  
489 Training Grant (T32-NS076401 to K.M.B).

490 **Author contributions**

491 K.M.B. and S.I. conceived the study and designed the experiments. I.V. generated the ESC and exEpiLC  
492 WGBS data. K.M.B performed the data analysis and all other experiments. K.M.B and S.I. wrote and edited  
493 the manuscript.

494 **References**

- 495 1. Strahl, B.D., and Allis, C.D. (2000). The language of covalent histone modifications. *Nature* **403**,  
496 41–45. <https://doi.org/10.1038/47412>.

- 497 2. Jenuwein, T., and Allis, C.D. (2001). Translating the Histone Code. *Science* *293*, 1074–1080.  
498 <https://doi.org/10.1126/science.1063127>.
- 499 3. Lewis, E.B. (1978). A gene complex controlling segmentation in *Drosophila*. *Nature* *276*, 565–570.  
500 <https://doi.org/10.1038/276565a0>.
- 501 4. Kennison, J.A., and Tamkun, J.W. (1988). Dosage-dependent modifiers of polycomb and antennapedia  
502 mutations in *Drosophila*. *Proc. Natl. Acad. Sci. U.S.A.* *85*, 8136–8140. <https://doi.org/10.1073/pnas.85.21.8136>.
- 503 5. Kassis, J.A., Kennison, J.A., and Tamkun, J.W. (2017). Polycomb and Trithorax Group Genes in  
504 *Drosophila*. *Genetics* *206*, 1699–1725. <https://doi.org/10.1534/genetics.115.185116>.
- 505 6. Gabriele, M., Lopez Tobon, A., D'Agostino, G., and Testa, G. (2018). The chromatin basis of  
506 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>.
- 507 7. Schaefer, A., Sampath, S.C., Intrator, A., Min, A., Gertler, T.S., Surmeier, D.J., Tarakhovsky, A.,  
508 and Greengard, P. (2009). Control of cognition and adaptive behavior by the GLP/G9a epigenetic  
509 suppressor complex. *Neuron* *64*, 678–691. <https://doi.org/10.1016/j.neuron.2009.11.019>.
- 510 8. Iwase, S., Lan, F., Bayliss, P., De La Torre-Ubieta, L., Huarte, M., Qi, H.H., Whetstine, J.R., Bonni, A.,  
511 Roberts, T.M., and Shi, Y. (2007). The X-Linked Mental Retardation Gene SMCX/JARID1C Defines a  
512 Family of Histone H3 Lysine 4 Demethylases. *Cell* *128*, 1077–1088. <https://doi.org/10.1016/j.cell.2007.02.017>.
- 513 9. Claes, S., Devriendt, K., Van Goethem, G., Roelen, L., Meireleire, J., Raeymaekers, P., Cassiman,  
514 J.J., and Fryns, J.P. (2000). Novel syndromic form of X-linked complicated spastic paraparesis. *Am J  
Med Genet* *94*, 1–4.
- 515 10. Jensen, L.R., Amende, M., Gurok, U., Moser, B., Gimmel, V., Tzschach, A., Janecke, A.R., Tariverdian,  
516 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>.
- 517 11. Carmignac, V., Nambot, S., Lehalle, D., Callier, P., Moortgat, S., Benoit, V., Ghoumid, J., Delobel,  
518 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>.
- 517 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>.

- 519 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>.
- 520
- 521 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>.
- 522
- 523 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>.
- 524
- 525 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>.
- 526
- 527 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>.
- 528
- 529 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>.
- 530
- 531 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>.
- 532
- 533 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>.
- 534
- 535 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>.
- 536
- 537 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>.
- 538

- 539 23. Crackower, M.A., Kolas, N.K., Noguchi, J., Sarao, R., Kikuchi, K., Kaneko, H., Kobayashi, E., Kawai, Y.,  
540 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>.
- 541 24. Xiol, J., Cora, E., Koglgruber, R., Chuma, S., Subramanian, S., Hosokawa, M., Reuter, M., Yang, Z.,  
542 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>.
- 543 25. Cheng, S., Altmeppen, G., So, C., Welp, L.M., Penir, S., Ruhwedel, T., Menelaou, K., Harasimov, K.,  
544 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>.
- 545 26. Rouland, A., Masson, D., Lagrost, L., Vergès, B., Gautier, T., and Bouillet, B. (2022). Role of  
546 apolipoprotein C1 in lipoprotein metabolism, atherosclerosis and diabetes: A systematic review.  
*Cardiovasc Diabetol* *21*, 272. <https://doi.org/10.1186/s12933-022-01703-5>.
- 547 27. Leduc, V., Jasmin-Bélanger, S., and Poirier, J. (2010). APOE and cholesterol homeostasis in  
548 Alzheimer's disease. *Trends in Molecular Medicine* *16*, 469–477. <https://doi.org/10.1016/j.molmed.2010.07.008>.
- 549 28. Mueller, J.L., Skaletsky, H., Brown, L.G., Zaghlul, S., Rock, S., Graves, T., Auger, K., Warren,  
550 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>.
- 551 29. Handel, M.A., and Eppig, J.J. (1979). Sertoli Cell Differentiation in the Testes of Mice Genetically  
552 Deficient in Germ Cells. *Biology of Reproduction* *20*, 1031–1038. <https://doi.org/10.1095/biolreprod20.5.1031>.
- 553 30. Green, C.D., Ma, Q., Manske, G.L., Shami, A.N., Zheng, X., Marini, S., Moritz, L., Sultan, C.,  
554 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>.
- 555 31. Soh, Y.Q., Junker, J.P., Gill, M.E., Mueller, J.L., van Oudenaarden, A., and Page, D.C. (2015). A  
556 Gene Regulatory Program for Meiotic Prophase in the Fetal Ovary. *PLoS Genet* *11*, e1005531.  
<https://doi.org/10.1371/journal.pgen.1005531>.
- 557 32. Magnúsdóttir, E., and Surani, M.A. (2014). How to make a primordial germ cell. *Development* *141*,  
558 245–252. <https://doi.org/10.1242/dev.098269>.
- 559 33. Günesdogan, U., Magnúsdóttir, E., and Surani, M.A. (2014). Primordial germ cell specification: A  
560 context-dependent cellular differentiation event [corrected]. *Philos Trans R Soc Lond B Biol Sci* *369*.  
<https://doi.org/10.1098/rstb.2013.0543>.

- 561 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>.
- 562
- 563 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.  
<https://doi.org/10.1016/j.cell.2011.06.052>.
- 564
- 565 36. Samanta, M., and Kalantry, S. (2020). Generating primed pluripotent epiblast stem cells: A methodol-  
ogy chapter. *Curr Top Dev Biol* *138*, 139–174. <https://doi.org/10.1016/bs.ctdb.2020.01.005>.
- 566
- 567 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.  
*EMBO Reports* *16*, 791–802. <https://doi.org/10.15252/embr.201540538>.
- 568
- 569 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  
activity. *Nature* *487*, 57–63. <https://doi.org/10.1038/nature11244>.
- 570
- 571 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>.
- 572
- 573 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>.
- 574
- 575 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>.
- 576
- 577 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>.
- 578
- 579 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>.
- 580
- 581 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>.
- 582

- 583 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.  
584 Nat Struct Mol Biol. <https://doi.org/10.1038/s41594-023-01038-z>.
- 585 46. Pohlers, M., Truss, M., Frede, U., Scholz, A., Strehle, M., Kuban, R.-J., Hoffmann, B., Morkel, M.,  
Birchmeier, C., and Hagemann, C. (2005). A Role for E2F6 in the Restriction of Male-Germ-Cell-  
586 Specific Gene Expression. Current Biology 15, 1051–1057. <https://doi.org/10.1016/j.cub.2005.04.060>.
- 587 47. Dahlet, T., Truss, M., Frede, U., Al Adhami, H., Bardet, A.F., Dumas, M., Vallet, J., Chicher, J.,  
Hamann, P., Kottnik, S., et al. (2021). E2F6 initiates stable epigenetic silencing of germline genes  
588 during embryonic development. Nat Commun 12, 3582. <https://doi.org/10.1038/s41467-021-23596-w>.
- 589 48. Agulnik, A.I., Mitchell, M.J., Mattei, M.G., Borsani, G., Avner, P.A., Lerner, J.L., and Bishop, C.E.  
590 (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>.
- 591 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>.
- 592 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  
593 5, 1355–1360. <https://doi.org/10.1093/hmg/5.9.1355>.
- 594 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.0002553>.
- 595 52. Wang, P.J., McCarrey, J.R., Yang, F., and Page, D.C. (2001). An abundance of X-linked genes  
596 expressed in spermatogonia. Nat Genet 27, 422–426. <https://doi.org/10.1038/86927>.
- 597 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  
598 inactivation. Nat Genet 36, 642–646. <https://doi.org/10.1038/ng1368>.
- 599 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/18.2.7019>.
- 600 55. Tatsumi, D., Hayashi, Y., Endo, M., Kobayashi, H., Yoshioka, T., Kiso, K., Kanno, S., Nakai, Y., Maeda,  
I., Mochizuki, K., et al. (2018). DNMTs and SETDB1 function as co-repressors in MAX-mediated  
601 repression of germ cell-related genes in mouse embryonic stem cells. PLoS ONE 13, e0205969.  
602 <https://doi.org/10.1371/journal.pone.0205969>.

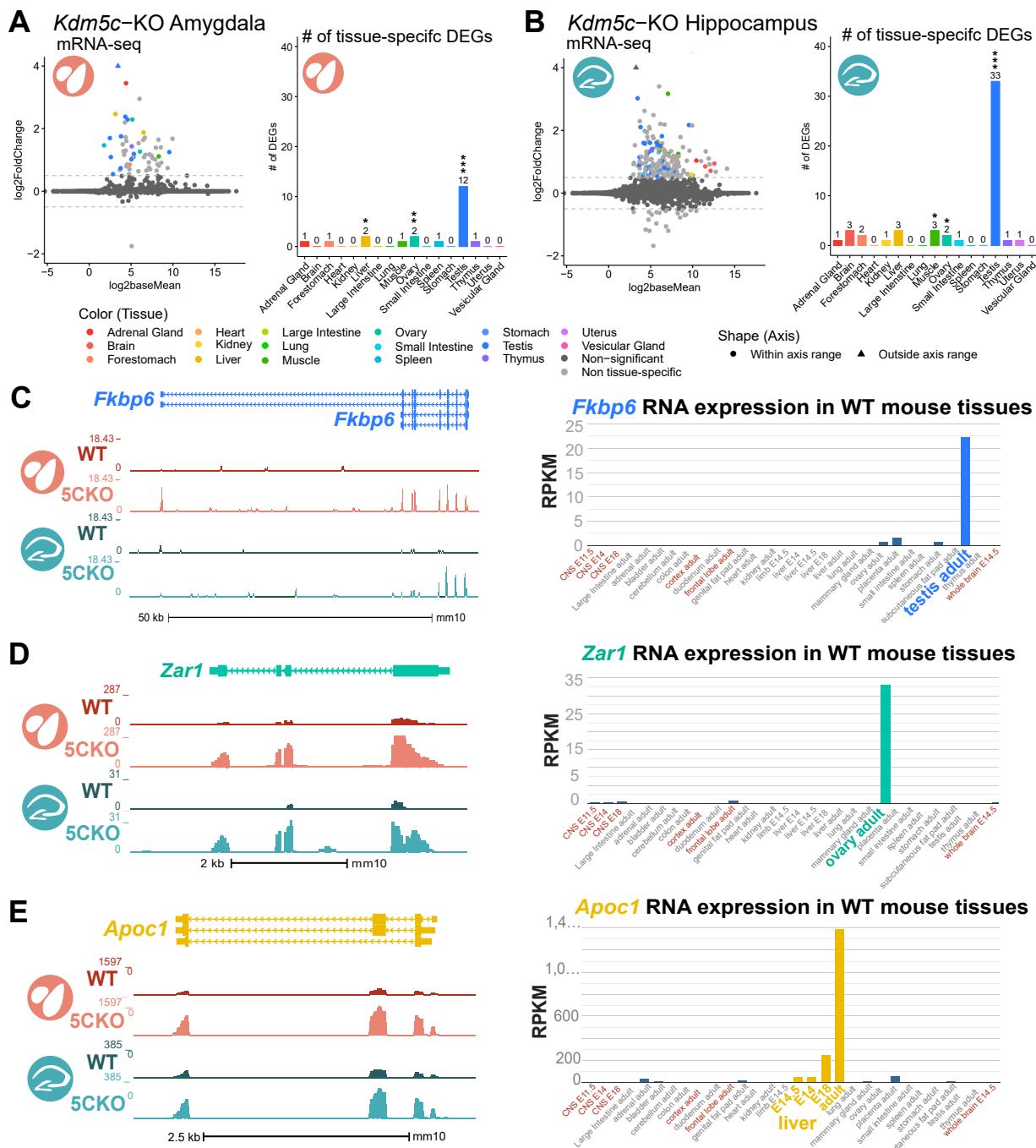
- 605 56. Stielow, B., Finkernagel, F., Stiewe, T., Nist, A., and Suske, G. (2018). MGA, L3MBTL2 and E2F6  
determine genomic binding of the non-canonical Polycomb repressive complex PRC1.6. *PLoS Genet*  
606 14, e1007193. <https://doi.org/10.1371/journal.pgen.1007193>.
- 607 57. Tahiliani, M., Mei, 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.  
608 *Nature* 447, 601–605. <https://doi.org/10.1038/nature05823>.
- 609 58. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H.,  
and Glass, C.K. (2010). Simple Combinations of Lineage-Determining Transcription Factors Prime  
cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Molecular Cell* 38, 576–589.  
610 <https://doi.org/10.1016/j.molcel.2010.05.004>.
- 611 59. 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,  
612 916–921. <https://doi.org/10.1038/35002634>.
- 613 60. 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).
- 614 61. 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  
615 Ciliogenesis. *Sci. Signal.* 2. <https://doi.org/10.1126/scisignal.2000602>.
- 616 62. 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.  
617 *PLoS Genet* 11, e1005368. <https://doi.org/10.1371/journal.pgen.1005368>.
- 618 63. Wu, Y., Hu, X., Li, Z., Wang, M., Li, S., Wang, X., Lin, X., Liao, S., Zhang, Z., Feng, X., et al.  
(2016). Transcription Factor RFX2 Is a Key Regulator of Mouse Spermiogenesis. *Sci Rep* 6, 20435.  
619 <https://doi.org/10.1038/srep20435>.
- 620 64. Liu, M., Zhu, Y., Xing, F., Liu, S., Xia, Y., Jiang, Q., and Qin, J. (2020). The polycomb group protein  
PCGF6 mediates germline gene silencing by recruiting histone-modifying proteins to target gene  
621 promoters. *J Biol Chem* 295, 9712–9724. <https://doi.org/10.1074/jbc.RA119.012121>.
- 622 65. Otani, J., Nankumo, T., Arita, K., Inamoto, S., Ariyoshi, M., and Shirakawa, M. (2009). Structural basis  
for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX–DNMT3–DNMT3L  
623 domain. *EMBO Reports* 10, 1235–1241. <https://doi.org/10.1038/embor.2009.218>.
- 624 66. Guo, X., Wang, L., Li, J., Ding, Z., Xiao, J., Yin, X., He, S., Shi, P., Dong, L., Li, G., et al. (2015).  
Structural insight into autoinhibition and histone H3-induced activation of DNMT3A. *Nature* 517,  
625 640–644. <https://doi.org/10.1038/nature13899>.

- 627 67. Meissner, A., Mikkelsen, T.S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B.E., Nusbaum, C., Jaffe, D.B., et al. (2008). Genome-scale DNA methylation maps of pluripotent and  
628 differentiated cells. *Nature* **454**, 766–770. <https://doi.org/10.1038/nature07107>.
- 629 68. Nassar, L.R., Barber, G.P., Benet-Pagès, A., Casper, J., Clawson, H., Diekhans, M., Fischer, C., Gonzalez, J.N., Hinrichs, A.S., Lee, B.T., et al. (2023). The UCSC Genome Browser database: 2023  
630 update. *Nucleic Acids Research* **51**, D1188–D1195. <https://doi.org/10.1093/nar/gkac1072>.
- 631 69. Akalin, A., Kormaksson, M., Li, S., Garrett-Bakelman, F.E., Figueiroa, M.E., Melnick, A., and Mason, C.E. (2012). methylKit: A comprehensive R package for the analysis of genome-wide DNA methylation  
632 profiles. *Genome Biol* **13**, R87. <https://doi.org/10.1186/gb-2012-13-10-r87>.
- 633 70. Torres-Padilla, M.-E. (2020). On transposons and totipotency. *Philos Trans R Soc Lond B Biol Sci*  
634 **375**, 20190339. <https://doi.org/10.1098/rstb.2019.0339>.
- 635 71. Yang, M., Yu, H., Yu, X., Liang, S., Hu, Y., Luo, Y., Izsvák, Z., Sun, C., and Wang, J. (2022). Chemical-  
induced chromatin remodeling reprograms mouse ESCs to totipotent-like stem cells. *Cell Stem Cell*  
636 **29**, 400–418.e13. <https://doi.org/10.1016/j.stem.2022.01.010>.
- 637 72. 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>.
- 639 73. Long, H.K., King, H.W., Patient, R.K., Odom, D.T., and Klose, R.J. (2016). Protection of CpG  
islands from DNA methylation is DNA-encoded and evolutionarily conserved. *Nucleic Acids Res* **44**,  
640 6693–6706. <https://doi.org/10.1093/nar/gkw258>.
- 641 74. Weber, M., Hellmann, I., Stadler, M.B., Ramos, L., Pääbo, S., Rebhan, M., and Schübeler, D. (2007). Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human  
642 genome. *Nat Genet* **39**, 457–466. <https://doi.org/10.1038/ng1990>.
- 643 75. 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  
644 of germline genes in mouse embryonic stem cells. *Nucleic Acids Research* **51**, 3130–3149. <https://doi.org/10.1093/nar/gkad071>.
- 645 76. Aubert, Y., Egolf, S., and Capell, B.C. (2019). The Unexpected Noncatalytic Roles of Histone Modifiers  
in Development and Disease. *Trends in Genetics* **35**, 645–657. <https://doi.org/10.1016/j.tig.2019.06.004>.
- 646 77. Morgan, M.A.J., and Shilatifard, A. (2020). Reevaluating the roles of histone-modifying enzymes  
and their associated chromatin modifications in transcriptional regulation. *Nat Genet* **52**, 1271–1281.  
648 <https://doi.org/10.1038/s41588-020-00736-4>.

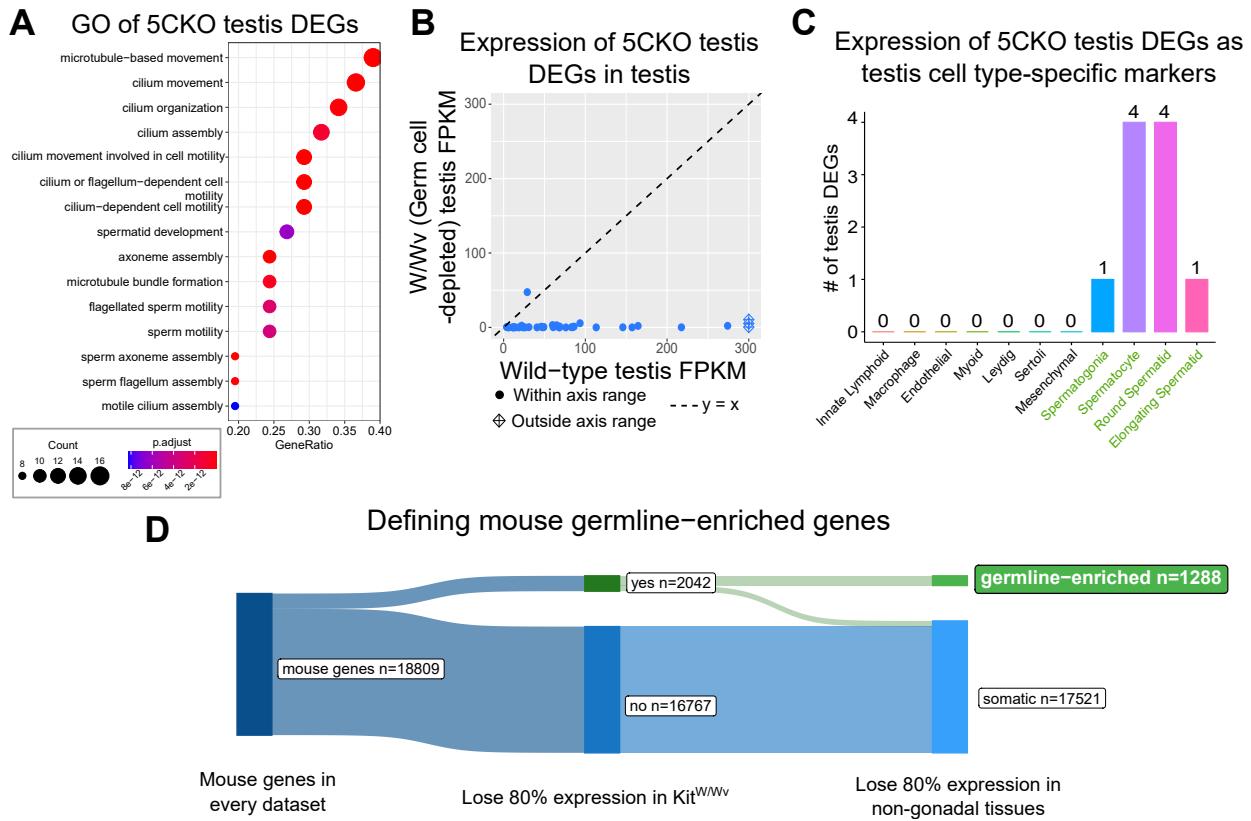
- 649 78. Abildayeva, K., Berbée, J.F.P., Blokland, A., Jansen, P.J., Hoek, F.J., Meijer, O., Lütjohann, D., Gautier, T., Pillot, T., De Vente, J., et al. (2008). Human apolipoprotein C-I expression in mice impairs learning and memory functions. *Journal of Lipid Research* 49, 856–869. <https://doi.org/10.1194/jlr.M700518JLR200>.
- 650
- 651 79. 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>.
- 652
- 653 80. 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>.
- 654
- 655 81. 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>.
- 656
- 657 82. Ghafouri-Fard, S., and Modarressi, M.-H. (2012). Expression of Cancer–Testis Genes in Brain Tumors: Implications for Cancer Immunotherapy. *Immunotherapy* 4, 59–75. <https://doi.org/10.2217/imt.11.145>.
- 658
- 659 83. 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>.
- 660
- 661 84. Velasco, G., Walton, E.L., Sterlin, D., Hédonin, 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>.
- 662
- 663 85. 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>.
- 664
- 665 86. 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>.
- 666
- 667 87. Stephens, M. (2016). False discovery rates: A new deal. *Biostat*, kxw041. <https://doi.org/10.1093/biostatistics/kxw041>.
- 668
- 669 88. 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>.
- 670

671 **Figures and Tables**

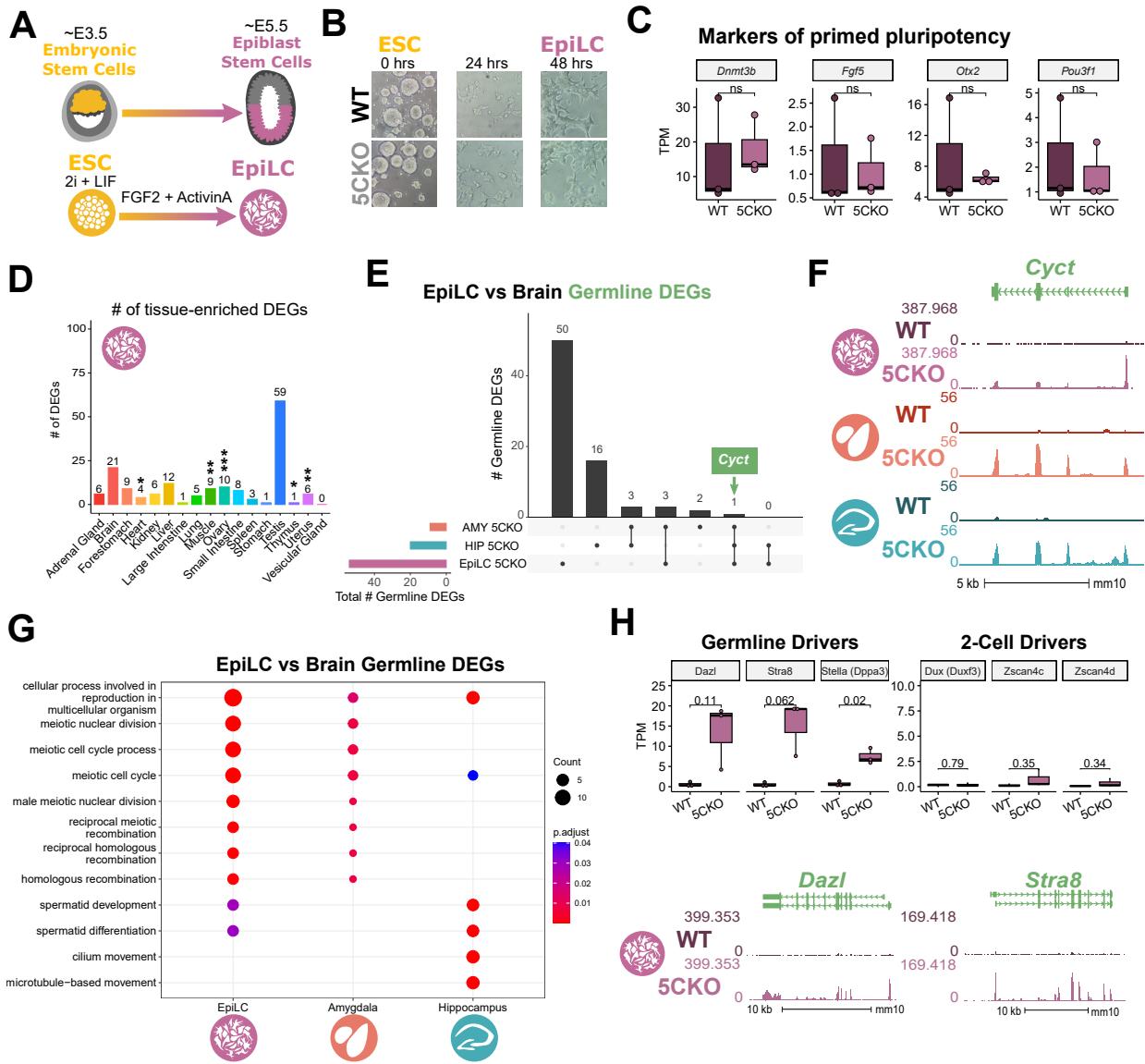
- 672     • Supplementary table 1: list of all germline genes.
- 673       – Columns to include:
- 674           \* KDM5C bound vs not
- 675           \* DEG in EpiLC, brain, both, neither (separate columns?)



**Figure 1: Tissue-enriched genes are misexpressed in the *Kdm5c*-KO brain.** **A-B.** Expression of tissue-enriched genes (Li et al 2017) in the male *Kdm5c*-KO amygdala (A) and hippocampus (B). Left - MA plot of mRNA-sequencing. Right - Number of tissue-enriched differentially expressed genes (DEGs). \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ , Fisher's exact test. **C.** Left - Average bigwigs of an example aberrantly expressed testis-enriched DEG, *FK506 binding protein 6* (*Fkbp6*) in the wild-type (WT) and *Kdm5c*-KO (5CKO) amygdala (red) and hippocampus (teal). Right - Expression of *Cyct* in wild-type tissues from NCBI Gene, with testis highlighted in blue and brain tissues highlighted in red. **D.** Left - Average bigwigs of an example ovary-enriched DEG, *Zygotic arrest 1* (*Zar1*). Right - Expression of *Zar1* in wild-type tissues from NCBI Gene, with ovary highlighted in teal and brain tissues highlighted in red. **E.** Left - Average bigwigs of an example liver-enriched DEG, *Apolipoprotein C-I* (*Apoc1*). Right - Expression of *Apoc1* in wild-type tissues from NCBI Gene, with liver highlighted in orange and brain tissues highlighted in red.



**Figure 2: Aberrant transcription of germline genes in the *Kdm5c*-KO in the brain.** **A.** enrichPlot gene ontology (GO) of *Kdm5c*-KO amygdala and hippocampus testis-enriched DEGs **B.** Expression of testis DEGs in wild-type (WT) testis versus germ cell-depleted (W/Wv) testis (Mueller et al 2013). Expression is in Fragments Per Kilobase of transcript per Million mapped reads (FPKM). **C.** Number of testis DEGs that were classified as cell-type specific markers in a single cell RNA-seq dataset of the testis (Green et al 2018). Germline cell types are highlighted in green, somatic cell types in black. **D.** Sankey diagram of mouse genes filtered for germline enrichment based on their expression in wild-type and germline-depleted mice and in adult mouse non-gonadal tissues (Li et al 2017).



**Figure 3: *Kdm5c*-KO epiblast-like cells express key drivers of germline identity**

**A.** Top - Diagram of *in vivo* differentiation of embryonic stem cells (ESCs) of the inner cell mass into epiblast stem cells. Bottom - *in vitro* differentiation of ESCs into epiblast-like cells (EpiLCs).

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

**C.** 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).

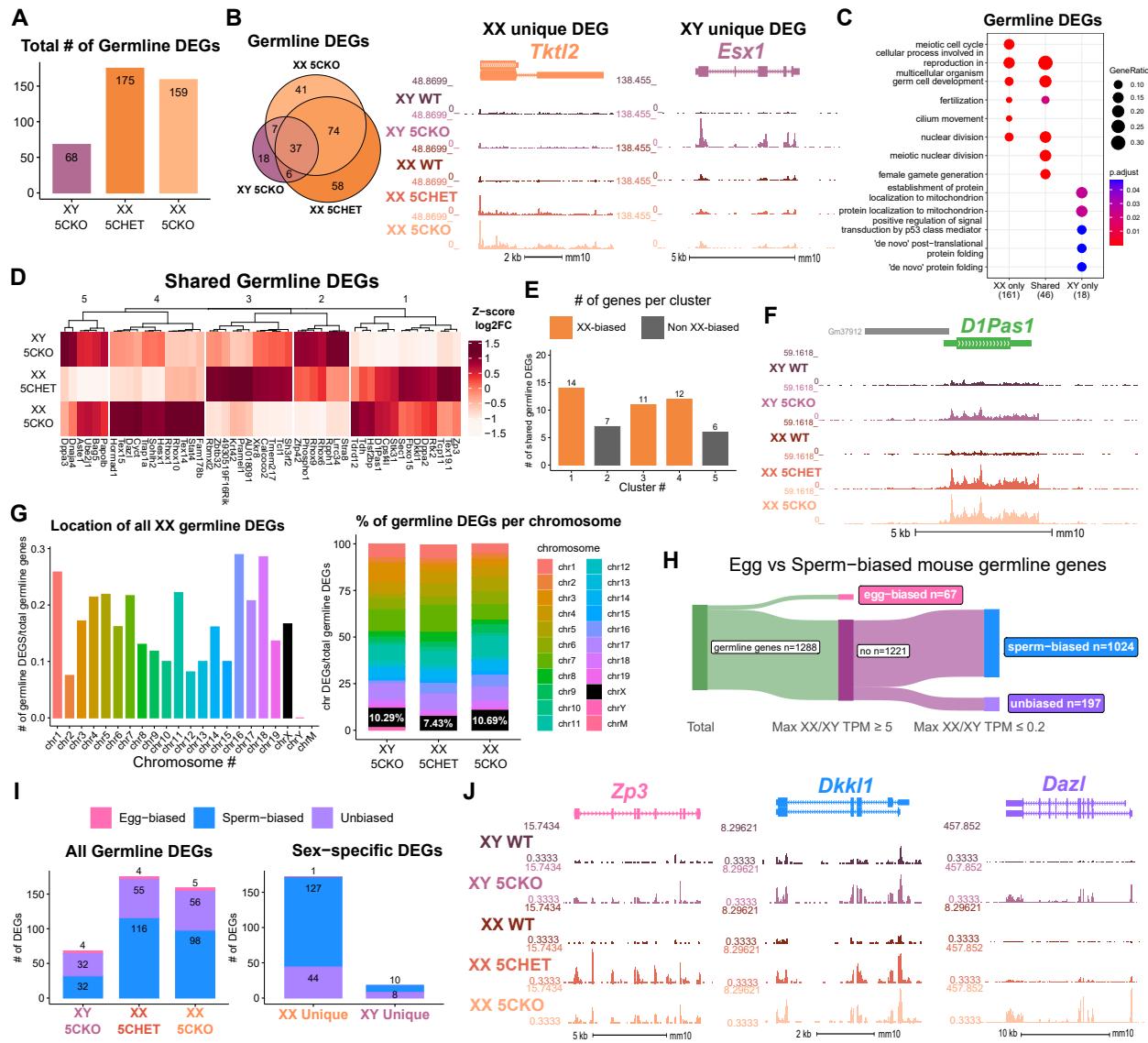
**D.** Number of tissue-enriched differentially expressed genes (DEGs) in *Kdm5c*-KO EpiLCs. \*  $p<0.05$ , \*\*  $p<0.01$ , \*\*\*  $p<0.001$ , Fisher's exact test.

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

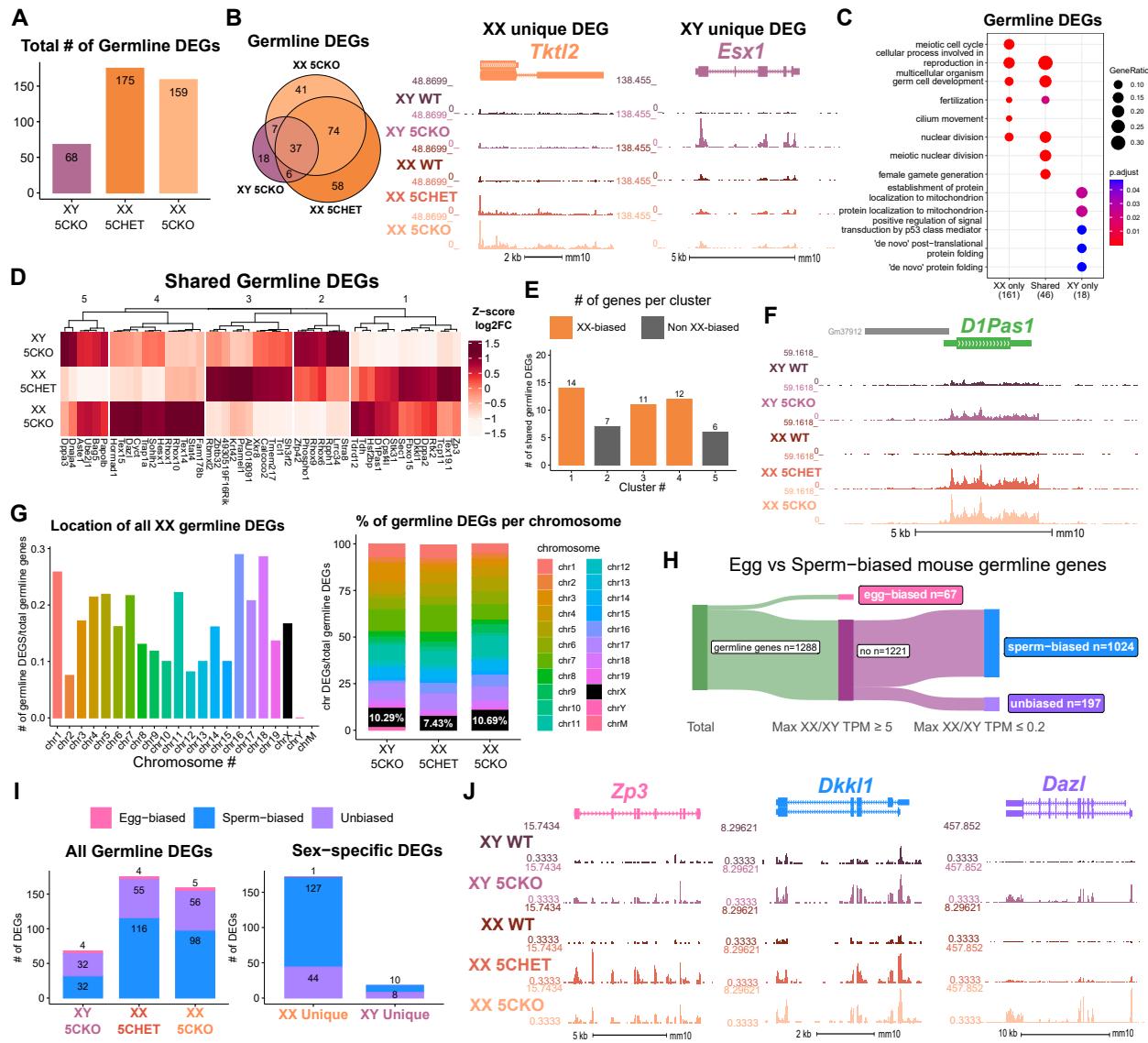
**F.** Average bigwigs of an example germline gene, *Cyct*, that is dysregulated *Kdm5c*-KO EpiLCs (top, purple), amygdala (middle, red), and hippocampus (bottom, blue).

**G.** enrichPlot gene ontology analysis comparing enriched biological processes for *Kdm5c*-KO EpiLC, amygdala, and hippocampus germline DEGs.

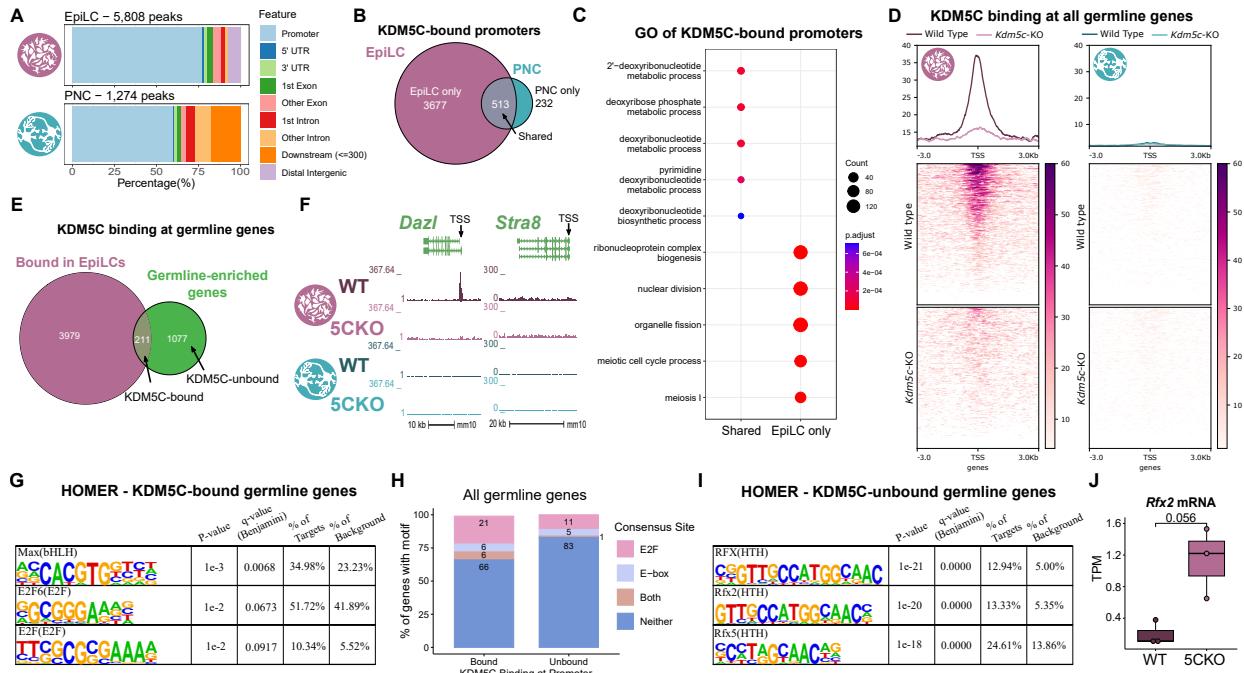
**H.** Top left - Example germline identity DEGs unique to EpiLCs. Top right - Example 2-cell genes that are not dysregulated in *Kdm5c*-KO EpiLCs. P-values for Welch's t-test. Bottom - Average bigwigs of *Dazl* and *Stra8* expression in wild-type and *Kdm5c*-KO EpiLCs.



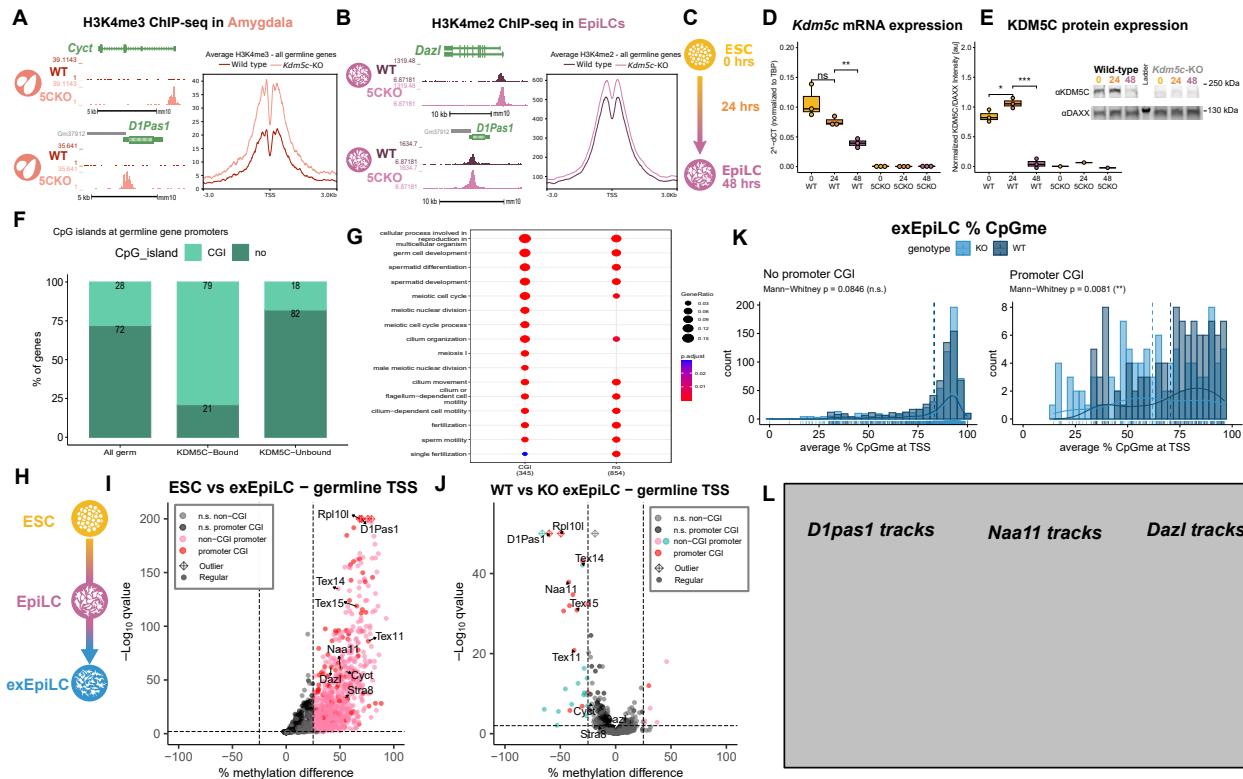
**Figure 4: Chromosomal sex influences *Kdm5c*-KO germline gene misexpression.** **A.** Total number of germline-enriched RNA-seq DEGs for male hemizygous *Kdm5c* knockout EpilCs (XY 5CKO, purple), female heterozygous *Kdm5c* knockout (XX 5CHET, orange), female homozygous *Kdm5c* knockout (XX 5CKO, light orange) EpilCs. **B.** Left - Eulerr overlap of *Kdm5c* mutant male and female EpilC germline DEGs. Right - Example of germline DEGs unique to females or males, *Tktl2* and *Esx1*. **C.** enrichPlot gene ontology analysis comparing enriched biological processes for germline DEGs shared between *Kdm5c* mutant males and females (Shared), or unique to one sex (XX only or XY only). **D.** Heatmap of germline DEGs shared between male and female mutants. Color is the average log2 fold change from sex-matched wild-type. **E.** Number of genes within each cluster from D. Clusters with higher expression in females compared to males (XX-biased) highlighted in orange. **F.** Example average bigwigs of a male and female shared germline DEG *D1Pas1* that is more highly expressed in female mutants. **G.** Left - Number of all female germline DEGs located on each chromosome over the total number of germline-enriched genes on that chromosome. Right - Percentage of germline DEGs that lie on each chromosome for each *Kdm5c* mutant. X chromosome highlighted in black. **H.** Sankey diagram classifying egg-biased (pink) and sperm-biased (blue) and unbiased (purple) mouse germline-enriched genes. **I.** Number of egg, sperm, or unbiased germline DEGs for male and female *Kdm5c* mutants. **J.** Example bigwigs of egg-biased (*Zp3*), sperm-biased (*Dkk1*), and unbiased (*Dazl*) dysregulated in both male and female *Kdm5c* mutants.



**Figure 5: Chromosomal sex influences *Kdm5c*-KO germline gene misexpression.** **A.** Total number of germline-enriched RNA-seq DEGs for male hemizygous *Kdm5c* knockout EpilCs (XY 5CKO, purple), female heterozygous *Kdm5c* knockout (XX 5CHET, orange), female homozygous *Kdm5c* knockout (XX 5CKO, light orange) EpilCs. **B.** Left - Eulerr overlap of *Kdm5c* mutant male and female EpilC germline DEGs. Right - Example of germline DEGs unique to females or males, *Tktl2* and *Esx1*. **C.** enrichPlot gene ontology analysis comparing enriched biological processes for germline DEGs shared between *Kdm5c* mutant males and females (Shared), or unique to one sex (XX only or XY only). **D.** Heatmap of germline DEGs shared between male and female mutants. Color is the average log2 fold change from sex-matched wild-type. **E.** Number of genes within each cluster from D. Clusters with higher expression in females compared to males (XX-biased) highlighted in orange. **F.** Example average bigwigs of a male and female shared germline DEG *D1Pas1* that is more highly expressed in female mutants. **G.** Left - Number of all female germline DEGs located on each chromosome over the total number of germline-enriched genes on that chromosome. Right - Percentage of germline DEGs that lie on each chromosome for each *Kdm5c* mutant. X chromosome highlighted in black. **H.** Sankey diagram classifying egg-biased (pink) and sperm-biased (blue) and unbiased (purple) mouse germline-enriched genes. **I.** Number of egg, sperm, or unbiased germline DEGs for male and female *Kdm5c* mutants. **J.** Example bigwigs of egg-biased (*Zp3*), sperm-biased (*Dkk1*), and unbiased (*Dazl*) dysregulated in both male and female *Kdm5c* mutants.



**Figure 6: 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 ( $TSS \pm 500$ ) 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.** Eulerr of number of germline-enriched genes (green) with significant KDM5C binding at their promoter in EpiLCs (purple). **F.** Example KDM5C ChIP-seq bigwigs of KDM5C binding at the *Dazl* TSS but not *Stra8* in EpiLCs. **G.** HOMER motif analysis of all KDM5C-bound germline gene promoters, highlighting significant enrichment of MAX, E2F6, and E2F motifs. **H.** Number of all gene promoters bound or unbound by KDM5C with instances of the E2F or E-box consensus sequence. **I.** HOMER motif analysis of all KDM5C-unbound germline gene promoters, highlighting significant enrichment of RFX family transcription factor motifs. **J.** Expression of RNA-seq DEG *Rfx2* in wild-type and *Kdm5c*-KO EpiLCs. P-value of Welch's t-test, expression in transcripts per million (TPM).



**Figure 7: 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* mRNA 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.** Percentage of germline genes that harbor CpG islands (CGIs) in their promoters, based on UCSC annotation. Comparing all germline-enriched genes, KDM5C-bound germline genes, or KDM5C-unbound germline genes. **G.** enrichPlot gene ontology analysis of CGI-promoter versus non-CGI promoter germline genes. **H.** Diagram of ESC to extended EpiLC (exEpiLC) differentiation. **I.** Volcano plot of whole genome bisulfite sequencing (WGBS) comparing CpG methylation at germline gene promoters ( $TSS \pm 500$ ) in wild-type ESCs versus exEpiLCs. Promoter CGI genes highlighted in red, hypermethylated genes lacking a promoter CGI in pink. **J.** Volcano plot of WGBS of wild-type versus *Kdm5c*-KO exEpiLCs. Promoter CGI genes highlighted in red, hypermethylated genes lacking a promoter CGI in blue, hypomethylated genes lacking a promoter CGI in blue. **K.** Example UCSC browser shots of germline genes of CpG methylation (CpGMe) in wild-type and *Kdm5c*-KO ESCs and exEpiLCs. **L.** Histogram of average percent CpGMe at the promoter for germline genes with or without promoter CGIs. Wild-type in navy and *Kdm5c*-KO in light blue. Dashed lines are average methylation for each genotype, p-values for Mann-Whitney U test.