

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

## 2 intellectual disability factor KDM5C

3

<sup>4</sup> Katherine M. Bonefas, Ilakkiya Venkatachalam, and Shigeki Iwase.

5 Abstract

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

24 **Introduction**

25 A single genome holds the instructions to generate the myriad of cell types found within an organism.  
26 This is, in part, accomplished by chromatin regulators that can either promote or impede lineage-specific  
27 gene expression through DNA and histone modifications<sup>1-5</sup>. Human genetic studies revealed mutations in

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

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

45 Distinguishing between germ cells and somatic cells is a key feature of multicellularity<sup>14</sup> that occurs  
46 during early embryogenesis in many metazoans<sup>15</sup>. In mammals, chromatin regulators are crucial for  
47 decommissioning germline genes during the transition from naïve to primed pluripotency. Initially, germline  
48 gene promoters gain repressive histone H2A lysine 119 monoubiquitination (H2AK119ub1)<sup>16</sup> and histone 3  
49 lysine 9 trimethylation (H3K9me3)<sup>16,17</sup> in embryonic stem cells (ESCs) and are then decorated with DNA  
50 CpG methylation (CpGme) in the post-implantation embryo<sup>17–20</sup>. The contribution of KDM5C to this process  
51 remains unclear. Furthermore, studies on germline gene repression have primarily been conducted in males  
52 and focused on marker genes important for germ cell development rather than germline genes as a whole,  
53 given the lack of a curated list for germline-enriched genes. Therefore, it is unknown if the mechanism  
54 of repression differs between sexes or for certain classes of germline genes, e.g. meiotic genes versus  
55 spermatid differentiation genes.

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

62 **Results**

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

64 Previous RNA sequencing (RNA-seq) of the adult male *Kdm5c*-KO hippocampus revealed ectopic  
65 expression of some germline genes unique to the testis<sup>13</sup>. It is currently unknown if the testis is the only  
66 tissue type misexpressed in the *Kdm5c*-KO brain. We systematically tested whether other tissue-specific  
67 genes are misexpressed in the brain with constitutive knockout of *Kdm5c*<sup>21</sup> by using a published list of mouse  
68 tissue-enriched genes<sup>22</sup>.

69 We found a large proportion of significantly upregulated genes (DESeq2<sup>23</sup>, log2 fold change > 0.5, q  
70 < 0.1) within the male *Kdm5c*-KO brain are typically enriched within non-brain tissues (Amygdala: 35%,  
71 Hippocampus: 24%) (Figure 1A-B). For both the amygdala and hippocampus, the majority of tissue-enriched  
72 differentially expressed genes (DEGs) were testis genes (Figure 1A-C). Even though the testis has the  
73 largest total number of tissue-biased genes (2,496 genes) compared to any other tissue, testis-biased DEGs  
74 were significantly enriched for both brain regions (Amygdala p = 1.83e-05, Odds Ratio = 5.13; Hippocampus  
75 p = 4.26e-11, Odds Ratio = 4.45, Fisher's Exact Test). One example of a testis-enriched gene misexpressed  
76 in the *Kdm5c*-KO brain is *FK506 binding protein 6* (*Fkbp6*), a known regulator of PIWI-interacting RNAs  
77 (piRNAs) and meiosis<sup>24,25</sup> (Figure 1C).

78 Interestingly, we also observed significant enrichment of ovary-biased DEGs in both the amygdala and  
79 hippocampus (Amygdala p = 0.00574, Odds Ratio = 18.7; Hippocampus p = 0.048, Odds Ratio = 5.88,  
80 Fisher's Exact) (Figure 1A-B). Ovary-enriched DEGs included *Zygotic arrest 1* (*Zar1*), which sequesters  
81 mRNAs in oocytes for meiotic maturation<sup>26</sup> (Figure 1D). Given that the *Kdm5c*-KO mice we analyzed are  
82 male, these data demonstrate that the ectopic expression of gonad-enriched genes is independent of  
83 organismal sex.

84 Although not consistent across brain regions, we also found significant enrichment of DEGs biased  
85 towards two non-gonadal tissues - the liver (Amygdala p = 0.04, Odds Ratio = 6.58, Fisher's Exact Test) and  
86 the muscle (Hippocampus p = 0.01, Odds Ratio = 6.95, Fisher's Exact Test) (Figure 1A-B). *Apolipoprotein*  
87 *C-1* (*Apoc1*) a lipoprotein metabolism and transport gene, is among the liver-biased DEG derepressed in both  
88 the hippocampus and amygdala<sup>27</sup> and its brain overexpression has been implicated in Alzheimer's disease<sup>28</sup>  
89 (Figure 1E).

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

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

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

103        We then evaluated testicular DEG expression in wild-type testes versus testes with germ cell depletion<sup>29</sup>,  
104 which was accomplished by heterozygous *W* and *Wv* mutations in the enzymatic domain of *c-Kit* (*Kit*<sup>W/Wv</sup>)<sup>30</sup>.  
105 Almost all *Kdm5c*-KO testis-enriched DEGs lost expression with germ cell depletion (Figure 2B). We then  
106 assessed testis-enriched DEG expression in a published single cell RNA-seq dataset that identified cell  
107 type-specific markers within the testis<sup>31</sup>. Some *Kdm5c*-KO testis-enriched DEGs were classified as specific  
108 markers for different germ cell developmental stages (e.g. spermatogonia, spermatocytes, round spermatids,  
109 and elongating spermatids), yet none marked somatic cells (Figure 2C). Together, these data demonstrate  
110 that the *Kdm5c*-KO brain aberrantly expresses germline genes, but not somatic testicular genes, reflecting  
111 an erosion of the soma-germline boundary.

112        As of yet, research on germline gene silencing mechanisms has focused on a handful of key genes  
113 rather than assessing germline gene suppression genome-wide due to the lack of a comprehensive gene list.  
114 We therefore generated a list of mouse germline-enriched genes using RNA-seq datasets of *Kit*<sup>W/Wv</sup> mice  
115 that included males and females at embryonic day 12, 14, and 16<sup>32</sup> and adult male testes<sup>29</sup>. We defined  
116 genes as germline-enriched if their expression met the following criteria: 1) their expression is greater than  
117 1 FPKM in wild-type gonads 2) their expression in any non-gonadal tissue of adult wild type mice<sup>22</sup> does  
118 not exceed 20% of their maximum expression in the wild-type germline, and 3) their expression in the germ  
119 cell-depleted gonads, for any sex or time point, does not exceed 20% of their maximum expression in the  
120 wild-type germline. These criteria yielded 1,288 germline-enriched genes (Figure 2D), which was hereafter  
121 used as a resource to globally characterize germline gene misexpression with *Kdm5c* loss (Supplementary  
122 table 1).

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

124        Germ cells are typically distinguished from somatic cells soon after the embryo implants into the uterine  
125 wall<sup>33,34</sup>, when germline genes are silenced in epiblast stem cells that will form the somatic tissues<sup>35</sup>. This  
126 developmental time point can be modeled *in vitro* through differentiation of naïve embryonic stem cells  
127 (nESCs) into epiblast-like stem cells (EpiLCs) (Figure 3A)<sup>36,37</sup>. While some germline-enriched genes are  
128 also expressed in nESCs and in the 2-cell stage<sup>38–40</sup>, they are silenced as they differentiate into EpiLCs<sup>17,18</sup>.  
129 Therefore, we tested if KDM5C was necessary for the initial silencing of germline genes in somatic lineages

130 by evaluating the impact of *Kdm5c* loss in male EpiLCs.

131        *Kdm5c*-KO cell morpholgy during ESC to EpiLC differentiation appeared normal (Figure 3B) and EpiLCs  
132 properly expressed markers of primed pluripotency, such as *Dnmt3b*, *Fgf5*, *Pou3f1*, and *Otx2* (Figure 3C). We  
133 then identified tissue-enriched DEGs in a RNA-seq dataset of wild-type and *Kdm5c*-KO EpiLCs<sup>41</sup> (DESeq2,  
134 log2 fold change > 0.5, q < 0.1). Similar to the *Kdm5c*-KO brain, we observed general dysregulation of  
135 tissue-enriched genes, with the largest number of genes belonging to the brain and testis, although they  
136 were not significantly enriched (Figure 3D). Using our list of mouse germline-enriched genes assembled  
137 above, we found 68 germline genes were misexpressed in male *Kdm5c*-KO EpiLCs.

138        We then compared EpiLC germline DEGs to those expressed in the *Kdm5c*-KO brain to determine if  
139 germline genes are constitutively dysregulated or change over the course of development. The majority of  
140 germline DEGs were unique to either EpiLCs or the brain, with only *D1Pas1* and *Cyct* shared across all  
141 tissue/cell types (Figure 3E-F). EpiLCs had particularly high enrichment of meiosis-related gene ontologies  
142 (Figure 3G), such as meiotic cell cycle process (GO:1903046, p.adjust = 2.2e-07) and meiotic nuclear  
143 division (GO:0140013, p.adjust = 1.37e-07). While there was modest enrichment of meiotic gene ontologies  
144 in both brain regions, the *Kdm5c*-KO hippocampus primarily expressed late-stage spermatogenesis genes  
145 involved in sperm axoneme assembly (GO:0007288, p.adjust = 0.00621) and sperm motility (GO:0097722,  
146 p.adjust = 0.00612).

147        Notably, DEGs unique to *Kdm5c*-KO EpiLCs included key drivers of germline identity, such as *Stimulated*  
148 *by retinoic acid 8* (*Stra8*: log2 fold change = 3.73, q = 2.17e-39) and *Deleted in azoospermia like* (*Dazl*):  
149 log2 fold change = 3.36, q = 3.19e-12) (Figure 3H). These genes are typically expressed when primordial  
150 germ cells (PGCs) are committed to the germline fate and later in life to trigger meiotic gene expression  
151 programs<sup>42-44</sup>. Of note, some germline genes, including *Dazl*, are also expressed in the two-cell embryo<sup>39,45</sup>.  
152 However, we did not see derepression of two-cell stage-specific genes, like *Duxf3* (*Dux*) (log2 fold change  
153 = -0.282, q = 0.337) and *Zscan4d* (log2 fold change = 0.25, q = 0.381) (Figure 3H), indicating *Kdm5c*-KO  
154 EpiLCs do not revert back to a 2-cell state. Altogether, *Kdm5c*-KO EpiLCs express key drivers of germline  
155 identity and meiosis while the brain primarily expresses spermiogenesis genes, indicating germline gene  
156 misexpression mirrors germline development during the progression of somatic development.

## 157 **Female epiblast-like cells have increased sensitivity to germline gene misexpression 158 with *Kdm5c* loss**

159        It is currently unknown if the misexpression of germline genes is influenced by sex, as previous studies  
160 on germline gene repressors have focused on male cells<sup>16,17,19,46,47</sup>. Sex is particularly pertinent in the case  
161 of KDM5C because it partially escapes X chromosome inactivation (XCI), resulting in a higher dosage in  
162 females<sup>48-51</sup>. We therefore explored the impact of chromosomal sex upon germline gene suppression by  
163 comparing their dysregulation in male *Kdm5c* hemizygous knockout (XY *Kdm5c*-KO), female homozygous

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

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

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

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

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

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

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

210 The majority of promoters bound by KDM5C in PNCs were also bound in EpiLCs (513 shared promoters),  
211 however a large portion of gene promoters were bound by KDM5C only in EpiLCs (3,677 EpiLC only  
212 promoters) (Figure 5B). Genes bound by KDM5C in both PNCs and EpiLCs were enriched for functions  
213 involving nucleic acid turnover, such as deoxyribonucleotide metabolic process (GO:0009262, p.adjust =  
214 8.28e-05) (Figure 5C). Germline-specific ontologies were enriched only in EpiLC-specific KDM5C-bound  
215 promoters, such as meiotic nuclear division (GO: 0007127 p.adjust = 6.77e-16) (Figure 5C). There were no  
216 ontologies significantly enriched for PNC-specific KDM5C target genes. Using our mouse germline gene list,  
217 we observed evident KDM5C signal around the TSS of many germline genes in EpiLCs, but not in PNCs  
218 (Figure 5D). Based on our ChIP-seq peak cut-off criteria, KDM5C was highly enriched at 211 germline gene  
219 promoters in EpiLCs (16.4% of all germline genes) (Figure 5E). Of note, KDM5C was only bound to about  
220 one third of *Kdm5c*-KO RNA-seq DEG promoters (EpiLC only DEGs: 34.92063%, Brain only DEGs: 30%)  
221 (Supplementary figure 1A-C). However, KDM5C did bind the promoter of 4 out of the 5 genes dysregulated  
222 in both the brain and EpiLCs. Representative examples of KDM5C-bound and unbound germline DEGs  
223 are *Dazl* and *Stra8*, respectively (Figure 5F). Together, these results demonstrate KDM5C is recruited to a  
224 subset of germline genes in EpiLCs, including meiotic genes, but does not directly regulate germline genes  
225 in neurons. Furthermore, the majority of germline mRNAs expressed in *Kdm5c*-KO cells are dysregulated  
226 independent of direct KDM5C binding to their promoters.

227 Many germline-specific genes are suppressed by the polycomb repressive complex 1.6 (PRC1.6), which  
228 contains transcription factor heterodimers E2F6/DP1 and MGA/MAX that respectively bind E2F and E-box  
229 motifs<sup>56</sup>. PRC1.6 members may recruit KDM5C to germline gene promoters, given their association with  
230 KDM5C in HeLa cells and ESCs<sup>45,57</sup>. We thus used HOMER<sup>58</sup> to identify transcription factor motifs enriched  
231 at KDM5C-bound or unbound germline gene promoters (TSS ± 500 bp, q-value < 0.1). MAX and E2F6 binding  
232 sites were significantly enriched at germline genes bound by KDM5C in EpiLCs (MAX q-value: 0.0068, E2F6  
233 q-value: 0.0673, E2F q-value: 0.0917), but not at germline genes unbound by KDM5C (Figure 5G). One third

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

243 **KDM5C is recruited to CpG islands at germline promoters to facilitate *de novo* DNA  
244 methylation**

245 Previous work found two germline gene promoters have a marked reduction in DNA CpG methylation  
246 (CpGme) in the *Kdm5c*-KO adult hippocampus<sup>13</sup>. Since histone 3 lysine 4 di- and trimethylation (H3K4me2/3)  
247 impede *de novo* CpGme<sup>64,65</sup>, KDM5C's removal of H3K4me2/3 may be required to suppress germline genes.  
248 However, KDM5C's catalytic activity was recently shown to be dispensable for suppressing *Dazl* in ESCs<sup>45</sup>.  
249 To reconcile these observations, we hypothesized KDM5C erases H3K4me3 to promote the initial placement  
250 of CpGme at germline gene promoters in EpiLCs.

251 To test this hypothesis, we first characterized KDM5C's expression as naïve ESCs differentiate into  
252 EpiLCs (Figure 6A). While *Kdm5c* mRNA steadily decreased from 0 to 48 hours of differentiation (Figure  
253 6B), KDM5C protein initially increased from 0 to 24 hours but then decreased to near knockout levels by 48  
254 hours (Figure 6C). We then characterized KDM5C's substrates (H3K4me2/3) at germline gene promoters  
255 with *Kdm5c* loss using published ChIP-seq datasets<sup>21,41</sup>. *Kdm5c*-KO samples showed a marked increase in  
256 H3K4me2 in EpiLCs (Figure 6D) and H3K4me3 in the amygdala (Figure 6E) around the TSS of germline  
257 genes. Together, these data suggest KDM5C acts during the transition between ESCs and EpiLCs to remove  
258 H3K4me2/3 at germline gene promoters.

259 Germline genes accumulate CpG methylation (CpGme) at CpG islands (CGIs) during the transition  
260 from naïve to primed pluripotency<sup>18,20,66</sup>. We first examined how many of our germline-enriched genes had  
261 promoter CGIs (TSS ± 500 bp) using the UCSC genome browser<sup>67</sup>. Notably, out of 1,288 germline-enriched  
262 genes, only 356 (27.64%) had promoter CGIs (Figure 6F). CGI-containing germline genes had substantial  
263 enrichment of meiotic gene ontologies compared to CGI-free genes, including meiotic nuclear division  
264 (GO:0140013, p.adjust = 2.17e-12) and meiosis I (GO:0007127, p.adjust = 3.91e-10) (Figure 6G). Germline  
265 genes with promoter CGIs were more highly expressed than CGI-free genes across spermatogenesis stages,  
266 with highest expression in meiotic spermatocytes (Figure 6H). Contrastingly, CGI-free genes only displaying  
267 substantial expression in post-meiotic round spermatids. Although only a minor portion of germline gene

268 promoters contained CGIs, CGIs strongly determined KDM5C's recruitment to germline genes, with 79.15%  
269 ( $p = 2.37e-67$ , Odds Ratio = 17.8, Fisher's exact test) of KDM5C-bound germline gene promoters harboring  
270 CGIs (Figure 6G).

271 To assess how KDM5C loss impacts initial CpGme placement at germline gene promoters, we performed  
272 whole genome bisulfite sequencing (WGBS) in male wild-type and *Kdm5c*-KO ESCs and 96-hour extend  
273 EpiLCs (exEpiLCs), when germline genes reach peak methylation level<sup>17</sup> (Figure 6I). We first identified  
274 which germline gene promoters significantly gained CpGme in wild-type cells during nESC to exEpiLCs  
275 differentiation (methylKit<sup>68</sup>,  $q < 0.01$ ,  $|methylation\ difference| \geq 25\%$ , TSS  $\pm 500$  bp). In wild-type cells,  
276 the majority of germline genes gained substantial CpGme at their promoter during differentiation (60.08%),  
277 regardless if their promoter contained a CGI (Figure 6J).

278 We then identified promoters differentially methylated in wild-type versus *Kdm5c*-KO exEpiLCs (methylKit,  
279  $q < 0.01$ ,  $|methylation\ difference| > 25\%$ , TSS  $\pm 500$  bp). Of the 48,882 promoters assessed, 274 promoters  
280 were significantly hypomethylated and 377 promoters were significantly hypermethylated with KDM5C  
281 loss (Supplementary figure 2A). Hypomethylated promoters were significantly enriched for germline gene  
282 ontologies, such as meiotic nuclear division (GO:0140013, p.adjust = 0.012)(Supplementary figure 2B), with  
283 10.22% of hypomethylated promoters belonging to germline genes. Approximately half of germline promoters  
284 hypomethylated in *Kdm5c*-KO exEpiLCs are direct targets of KDM5C in EpiLCs (13 out of 28 hypomethylated  
285 promoters). Promoters that showed the most robust loss of CpGme (lowest q-values) harbored CGIs (Figure  
286 6K). CGI promoters, but not CGI-free promoters, had a significant reduction in CpGme with KDM5C loss  
287 as a whole (Figure 6L) (Non-CGI promoters  $p = 0.0846$ , CGI promoters  $p = 0.0081$ , Mann-Whitney U test).  
288 Significantly hypomethylated promoters included germline genes consistently dysregulated across multiple  
289 *Kdm5c*-KO RNA-seq datasets<sup>13</sup>, such as *Naa11* and *D1PAs1* (methylation difference = -60.03%, q-value  
290 = 3.26e-153) (Figure 6M). Surprisingly, we found only a modest reduction in CpGme at *Dazl*'s promoter  
291 (methylation difference = -6.525%, q-value = 0.0159) (Figure 6N). Altogether, these results demonstrate  
292 KDM5C is recruited to germline gene CGIs to promote CpGme at germline gene promoters. This suggests  
293 KDM5C's catalytic activity is required for germline gene repression in EpiLCs, however some loci can  
294 compensate for KDM5C loss through other silencing mechanisms, even when retaining H3K4me around  
295 the TSS.

## 296 Discussion

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

302 including *Dazl* and *Stra8*, while the adult brain primarily expresses genes important for late spermatogenesis.  
303 We demonstrated that although *Kdm5c* mutant sex did not influence whether sperm or egg-specific genes  
304 were misexpressed, female EpiLCs are more sensitive to germline gene de-repression. Germline genes  
305 can become aberrantly expressed in *Kdm5c*-KO cells via an indirect mechanism, such as activation via  
306 ectopic RFX transcription factors. Finally, we found KDM5C is dynamically regulated during ESC to EpiLC  
307 differentiation to promote long-term germline gene silencing through DNA methylation at CpG islands.  
308 Therefore, we propose KDM5C plays a fundamental role in the development of tissue identity during early  
309 embryogenesis, including the establishment of the soma-germline boundary. By systematically characterizing  
310 KDM5C's role in germline gene repression, we unveiled derepressive mechanisms governing distinct classes  
311 of germline gene in somatic lineages. Furthermore, these data provide molecular footholds which can be  
312 exploited to test the overarching contribution of ectopic germline gene expression upon neurodevelopment.

313 Although eggs and sperm employ the same transcriptional programs for shared functions, e.g. PGC  
314 formation, meiosis, and genome defense, some germline genes are sex specific. We found *Kdm5c* mutant  
315 males and females expressed both sperm and egg-biased genes, indicating the mechanism of derepression  
316 is independent of a given germline gene's sex. However, organismal sex did greatly influence the degree of  
317 germline gene dysregulation, as female *Kdm5c*-KO EpiLCs had over double the number of germline-enriched  
318 DEGs compared to males. The lack of X-linked gene enrichment in females suggests that this greater  
319 sensitivity to germline gene misexpress is not due to XCI defects previously reported in *Kdm5c*-KO females<sup>41</sup>.  
320 Intriguingly, females with heterozygous loss of *Kdm5c* also had over double the number of germline DEGs  
321 than hemizygous knockout males, even though their level of KDM5C should be roughly equivalent to that  
322 of wild-type males. Increased female sensitivity to germline gene de-repression may be related to females  
323 having a higher dose of KDM5C than males, due to its escape from XCI<sup>48–51</sup>. KDM5C's Y-chromosome  
324 homolog, KDM5D, exhibits weaker demethylase activity than KDM5C<sup>8</sup> and it is currently unknown to regulate  
325 germline gene expression. Altogether, these results suggests germline gene silencing mechanisms differ  
326 between males and females, which warrants further study to identify the biological implications and underlying  
327 mechanisms.

328 We found KDM5C is largely dispensable for promoting normal gene expression during development, yet is  
329 critical for suppressing ectopic developmental programs. It is important to note that while we highlighted  
330 KDM5C's regulation of germline genes, some germline-enriched genes are also expressed at the 2-cell  
331 stage and in naïve ESCs for their role in pluripotency and self-renewal<sup>40,45,69,70</sup>. Although expressed in  
332 naïve ESCs, “self-renewal” germline genes like *Dazl* are silenced during ESC differentiation into epiblast  
333 stem cells/EpiLCs<sup>17,18</sup>. We found that while *Kdm5c*-KO EpiLCs also express *Dazl*, they did not express  
334 2-cell-specific genes. These data suggest the 2-cell-like state reported in *Kdm5c*-KO ESCs<sup>45</sup> likely reflects  
335 KDM5C's primary role in germline gene repression. Germline gene misexpression in *Kdm5c*-KO EpiLCs may  
336 indicate they are differentiating into primordial germ cell-like cells (PGCLCs), rather than de-differentiating  
337 into 2-cell-like cells<sup>33,34,36</sup>. Yet, *Kdm5c*-KO EpiLCs had normal cellular morphology and properly expressed

338 markers for primed pluripotency, including *Otx2* which blocks EpiLC differentiation into PGCs/PGCLCs<sup>71</sup>.  
339 In addition to unimpaired EpiLC differentiation, *Kdm5c*-KO gross brain morphology is overall normal<sup>12</sup> and  
340 hardly any brain-specific genes were significantly dysregulated. Thus, ectopic germline gene expression  
341 occurs along with proper somatic development in *Kdm5c*-KO animals.

342 Our work provides novel insight into the cross-talk between H3K4me and CpGme. In EpiLCs, loss of  
343 KDM5C binding at a subset of germline gene promoters, e.g. *D1Pas1* and *Naa11*, strongly impaired CGI  
344 methylation, and resulted in their long-lasting de-repression into adulthood. Removal of H3K4me2/3 at CGIs  
345 is a plausible mechanism for KDM5C-mediated germline gene suppression<sup>13,72</sup>, given H3K4me2/3 primarily  
346 do not colocalize with CpGme<sup>73</sup> and can oppose DNMT3 activity<sup>64,65</sup>. However, emerging work indicates  
347 many histone-modifying enzymes have non-catalytic functions that influence gene expression, sometimes  
348 even more potently than their catalytic roles<sup>74,75</sup>. Indeed, KDM5C's catalytic activity was recently found to be  
349 dispensible for repressing *Dazl* in ESCs<sup>45</sup>. In our study, *Dazl*'s promoter still gained CpGme in *Kdm5c*-KO  
350 exEpiLCs, even with elevated H3K4me2. *Dazl* and a few other germline gene CGIs use multiple repressive  
351 mechanisms to facilitate CpGme<sup>16,17,46,47</sup>. Together, this suggests alternative silencing mechanisms are  
352 sufficient to recruit DNMT3s to some germline CGIs, while others may require KDM5C-mediated H3K4me  
353 removal to overcome promoter CGI escape from CpGme<sup>73,76</sup>. Furthermore, these results indicate the  
354 requirement for catalytic activity can change depending upon the locus and developmental stage, even for  
355 the same class of genes.

356 By generating a comprehensive list of mouse germline-enriched genes, we were able to reveal distinct  
357 derepressive mechanisms governing early versus late-stage germline developmental programs. Previous  
358 work on germline gene silencing has focused on genes with promoter CGIs<sup>18,73</sup>, and indeed the major-  
359 ity of KDM5C targets in EpiLCs were germ cell identity genes harboring CGIs. However, over 70% of  
360 germline-enriched gene promoters lacked CGIs, including the many KDM5C-unbound germline genes  
361 that were de-repressed in *Kdm5c*-KO cells. CGI-free, KDM5C-unbound germline genes were primarily  
362 late-stage spermatogenesis genes and significantly enriched for RFX2 binding sites, a central regulator  
363 of spermiogenesis<sup>62,63</sup>. These data suggest that once activated during early embryogenesis, drivers of  
364 germline identity like *Rfx2*, *Stra8*, and *Dazl* turn on downstream germline programs, ultimately culminating in  
365 the expression of spermiogenesis genes in the adult *Kdm5c*-KO brain. Therefore, we propose KDM5C is  
366 recruited via promoter CGIs to genes that shape germ cell formation and acts as break against runaway  
367 activation of germline-specific programs.

368 The above work provides the mechanistic foundation for KDM5C-mediated repression of tissue and  
369 germline-specific genes. However, the contribution of these ectopic, tissue-specific genes towards *Kdm5c*-  
370 KO neurological impairments is still unknown. In addition to germline genes, we also identified significant  
371 enrichment of muscle, liver, and even ovary-biased transcripts within the male *Kdm5c*-KO brain. Intriguingly,  
372 select liver and muscle-biased DEGs do have known roles within the brain, such as the liver-enriched lipid  
373 metabolism gene *Apolipoprotein C-I (Apoc1)*<sup>27</sup>. *APOC1* dysregulation is implicated in Alzheimer's disease in

374 humans<sup>28</sup> and overexpression of *Apoc1* in the mouse brain can impair learning and memory<sup>77</sup>. KDM5C may  
375 therefore be crucial for neurodevelopment by fine-tuning the expression of tissue-enriched, dosage-sensitive  
376 genes like *Apoc1*. Given germline genes have no known functions within the brain, their impact upon  
377 neurodevelopment is currently unknown. Ectopic testicular germline transcripts have been observed in a  
378 variety of cancers<sup>78,79</sup>, including brain tumors in *Drosophila* and mammals and shown to promote cancer  
379 progression<sup>80,81,ninBiologyCancerTestisAntigens2023?</sup>. Intriguingly, mouse and human models for other chromatin-  
380 linked neurodevelopmental disorders also display impaired soma-germline demarcation<sup>7,82–85</sup>, such as DNA  
381 methyltransferase 3b (DNMT3B), H3K9me1/2 methyltransferases G9A/GLP, methyl-CpG -binding protein 2  
382 (MECP2)<sup>82</sup>. Thus, KDM5C is among a growing cohort of chromatin-linked neurodevelopmental disorders  
383 with similar erosion of the germline versus soma boundary. Further research is required to determine the  
384 impact of these germline genes and the extent to which this phenomenon occurs in humans.

## 385 Materials and Methods

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

387 Tissue-enriched differentially expressed genes (DEGs) were determined by their classification in a previ-  
388 ously published dataset from 17 male and female mouse tissues<sup>22</sup>. This study defined tissue expression as  
389 greater than 1 Fragments Per Kilobase of transcript per Million mapped read (FPKM) and tissue enrichment  
390 as at least 4-fold higher expression than any other tissue.

391 We curated a list of germline-enriched genes using an RNA-seq dataset from wild-type and germline-  
392 depleted (Kit<sup>W/Wv</sup>) male and female mouse embryos from embryonic day 12, 14, and 16<sup>32</sup>, as well as adult  
393 male testes<sup>29</sup>. Germline-enriched genes met the following criteria: 1) their expression is greater than 1  
394 FPKM in wild-type germline 2) their expression in any wild-type somatic tissues<sup>22</sup> does not exceed 20%  
395 of maximum expression in wild-type germline, and 3) their expression in the germ cell-depleted (Kit<sup>W/Wv</sup>)  
396 germline, for any sex or time point, does not exceed 20% of maximum expression in wild-type germline. We  
397 defined sperm and egg-biased genes as those whose expression in the opposite sex, at any time point, is no  
398 greater than 20% of the gene's maximum expression in a given sex. Genes that did not meet this threshold  
399 for either sex were classified as 'unbiased'.

### 400 Cell culture

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

406      Embryonic stem cells (ESCs) and epiblast-like cells were cultured using previously established  
407 methods<sup>37</sup>. Briefly, ESCs were initially cultured in primed ESC (pESC) media consisting of KnockOut  
408 DMEM (Gibco#10829–018), fetal bovine serum (Gibco#A5209501), KnockOut serum replacement  
409 (Invitrogen#10828–028), Glutamax (Gibco#35050-061), Anti-Anti (Gibco#15240-062), MEM Non-essential  
410 amino acids (Gibco#11140-050), and beta-mercaptoethanol (Sigma#M7522). They were then transitioned  
411 into ground-state, “naïve” ESCs (nESCs) by culturing for four passages in N2B27 media containing  
412 DMEM/F12 (Gibco#11330–032), Neurobasal media (Gibco#21103–049), Gluamax, Anti-Anti, N2 sup-  
413 plement (Invitrogen#17502048), and B27 supplement without vitamin A (Invitrogen#12587-010), and  
414 beta-mercaptoethanol. Both pESC and nESC media were supplemented with 3  $\mu$ M GSK3 inhibitor  
415 CHIR99021 (Sigma #SML1046-5MG), 1  $\mu$ M MEK inhibitor PD0325901 (Sigma #PZ0162-5MG), and 1,000  
416 units/mL leukemia inhibitory factor (LIF, Millipore#ESG1107).

417      nESCs were differentiated into epiblast-like cells (EpiLCs, 48 hours) and extendend EpiLCs (exEpiLCs,  
418 96 hours) by culturing in N2B27 media containing DMEM/F12, Neurobasal media, Gluamax, Anti-Anti,  
419 N2 supplement, B27 supplement (Invitrogen#17504044), beta-mercaptoethanol, fibroblast growth factor 2  
420 (FGF2, R&D Biotechne 233-FB), and activin A (R&D Biotechne 338AC050CF), as previously described<sup>37</sup>.

## 421 **RT-qPCR**

422      XXX

## 423 **Western Blot**

424      XXX

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

426      After ensuring read quality via FastQC (v0.11.8), reads were then mapped to the mm10 *Mus musculus*  
427 genome (Gencode) using STAR (v2.5.3a), during which we removed duplicates and kept only uniquely  
428 mapped reads. Count files were generated by FeatureCounts (Subread v1.5.0), and BAM files were  
429 converted to bigwigs using deeptools (v3.1.3) and visualized by the UCSC genome browser. RStudio (v3.6.0)  
430 was then used to analyze counts files by DESeq2 (v1.26.0)<sup>23</sup> to identify differentially expressed genes  
431 (DEGs) with a q-value (p-adjusted via FDR/Benjamini–Hochberg correction) less than 0.1 and a log2 fold  
432 change greater than 0.5. For all DESeq2 analyses, log2 fold changes were calculated with IfcShrink using  
433 the ashR package<sup>86</sup>. MA-plots were generated by ggpubr (v0.6.0), and Eulerr diagrams were generated by  
434 eulerr (v6.1.1). Boxplots and scatterplots were generated by ggpubr (v0.6.0) and ggplot2 (v3.3.2). The Upset  
435 plot was generated via the package UpSetR (v1.4.0)<sup>87</sup>. Gene ontology (GO) analyses were performed by  
436 the R package enrichPlot (v1.16.2) using the biological processes setting and compareCluster.

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

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

448 **CpG island (CGI) analysis**

449 Locations of CpG islands were determined through the mm10 UCSC genome browser CpG island track<sup>67</sup>,  
450 which classified CGIs as regions that have greater than 50% GC content, are larger than 200 base pairs,  
451 and have a ratio of CG dinucleotides observed over the expected amount greater than 0.6. CGI genomic  
452 coordinates were then annotated using ChIPseeker (v1.32.1) and filtered for ones that lie within promoters of  
453 our germline-enriched 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.8>  
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  
suppressor complex. *Neuron* **64**, 678–691. <https://doi.org/10.1016/j.neuron.2009.11.019>.
- 509 8. Iwase, S., Lan, F., Bayliss, P., De La Torre-Ubieta, L., Huarte, M., Qi, H.H., Whetstine, J.R., Bonni, A.,  
510 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.200>  
5. Claes, S., Devriendt, K., Van Goethem, G., Roelen, L., Meireleire, J., Raeymaekers, P., Cassiman,  
512 J.J., and Fryns, J.P. (2000). Novel syndromic form of X-linked complicated spastic paraparesia. *Am J  
Med Genet* **94**, 1–4.
- 513 10. Jensen, L.R., Amende, M., Gurok, U., Moser, B., Gimmel, V., Tzschach, A., Janecke, A.R., Tariverdian,  
514 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>.
- 515 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  
516 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>.
- 518
- 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. 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>.
- 530
- 531 19. 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>.
- 532
- 533 20. 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>.
- 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. 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>.
- 538
- 539 23. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for  
540 RNA-seq data with DESeq2. *Genome Biol* 15, 550. <https://doi.org/10.1186/s13059-014-0550-8>.
- 541 24. 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  
542 Chromosome Pairing in Meiosis. *Science* 300, 1291–1295. <https://doi.org/10.1126/science.1083022>.
- 543 25. 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>.
- 544
- 545 26. 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  
546 membraneless compartment. *Science* 378, eabq4835. <https://doi.org/10.1126/science.abq4835>.
- 547 27. 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.  
548 *Cardiovasc Diabetol* 21, 272. <https://doi.org/10.1186/s12933-022-01703-5>.
- 549 28. Leduc, V., Jasmin-Bélanger, S., and Poirier, J. (2010). APOE and cholesterol homeostasis in  
Alzheimer's disease. *Trends in Molecular Medicine* 16, 469–477. <https://doi.org/10.1016/j.molmed.2010.07.008>.
- 550
- 551 29. Mueller, J.L., Skaletsky, H., Brown, L.G., Zaghlul, S., Rock, S., Graves, T., Auger, K., Warren,  
W.C., Wilson, R.K., and Page, D.C. (2013). Independent specialization of the human and mouse X  
552 chromosomes for the male germ line. *Nat Genet* 45, 1083–1087. <https://doi.org/10.1038/ng.2705>.
- 553 30. Handel, M.A., and Eppig, J.J. (1979). Sertoli Cell Differentiation in the Testes of Mice Genetically  
Deficient in Germ Cells. *Biology of Reproduction* 20, 1031–1038. <https://doi.org/10.1095/biolreprod20.5.1031>.
- 554
- 555 31. 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>.
- 556
- 557 32. 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>.
- 558

- 559 33. Magnúsdóttir, E., and Surani, M.A. (2014). How to make a primordial germ cell. *Development* *141*,  
560 245–252. <https://doi.org/10.1242/dev.098269>.
- 561 34. Günesdogan, U., Magnúsdóttir, E., and Surani, M.A. (2014). Primordial germ cell specification: A  
562 context-dependent cellular differentiation event [corrected]. *Philos Trans R Soc Lond B Biol Sci* *369*.  
<https://doi.org/10.1098/rstb.2013.0543>.
- 563 35. Bardot, E.S., and Hadjantonakis, A.-K. (2020). Mouse gastrulation: Coordination of tissue patterning,  
564 specification and diversification of cell fate. *Mechanisms of Development* *163*, 103617. <https://doi.org/10.1016/j.mod.2020.103617>.
- 565 36. Hayashi, K., Ohta, H., Kurimoto, K., Aramaki, S., and Saitou, M. (2011). Reconstitution of the  
566 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>.
- 567 37. Samanta, M., and Kalantry, S. (2020). Generating primed pluripotent epiblast stem cells: A methodol-  
568 ogy chapter. *Curr Top Dev Biol* *138*, 139–174. <https://doi.org/10.1016/bs.ctdb.2020.01.005>.
- 569 38. Welling, M., Chen, H., Muñoz, J., Musheev, M.U., Kester, L., Junker, J.P., Mischerikow, N., Arbab, M.,  
570 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>.
- 571 39. Macfarlan, T.S., Gifford, W.D., Driscoll, S., Lettieri, K., Rowe, H.M., Bonanomi, D., Firth, A., Singer, O.,  
572 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>.
- 573 40. Suzuki, A., Hirasaki, M., Hishida, T., Wu, J., Okamura, D., Ueda, A., Nishimoto, M., Nakachi, Y.,  
574 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>.
- 575 41. Samanta, M.K., Gayen, S., Harris, C., Maclary, E., Murata-Nakamura, Y., Malcore, R.M., Porter, R.S.,  
576 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>.
- 577 42. Koubova, J., Menke, D.B., Zhou, Q., Capel, B., Griswold, M.D., and Page, D.C. (2006). Retinoic  
578 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>.
- 579 43. Lin, Y., Gill, M.E., Koubova, J., and Page, D.C. (2008). Germ Cell-Intrinsic and -Extrinsic Factors  
580 Govern Meiotic Initiation in Mouse Embryos. *Science* *322*, 1685–1687. <https://doi.org/10.1126/science.1166340>.
- 581 44. Endo, T., Mikedis, M.M., Nicholls, P.K., Page, D.C., and De Rooij, D.G. (2019). Retinoic Acid and Germ  
582 Cell Development in the Ovary and Testis. *Biomolecules* *9*, 775. <https://doi.org/10.3390/biom9120775>.

- 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>.

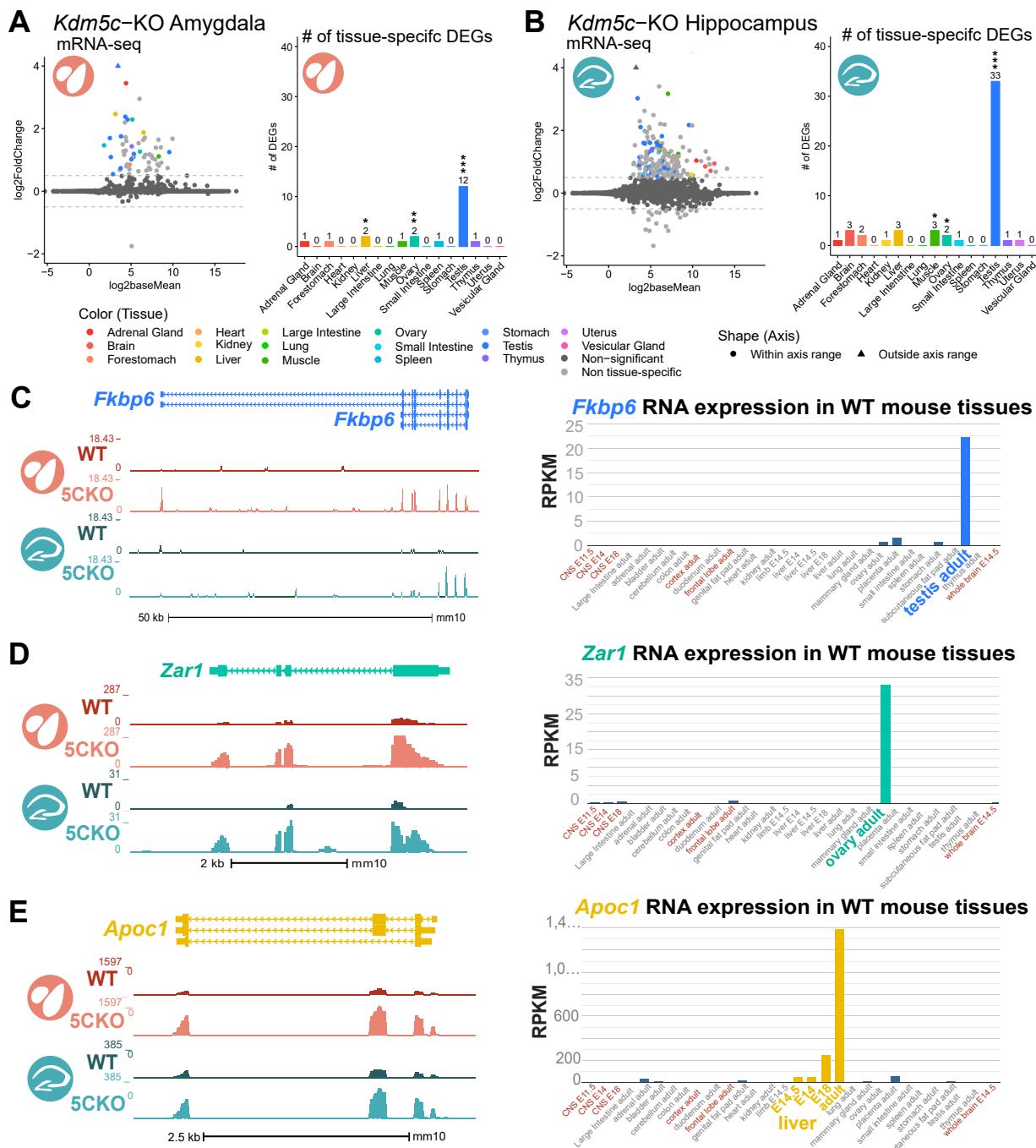
- 605 56. Stielow, B., Finkernagel, F., Stiewe, T., Nist, A., and Suske, G. (2018). MGA, L3MBTL2 and E2F6  
determine genomic binding of the non-canonical Polycomb repressive complex PRC1.6. *PLoS Genet*  
606 14, e1007193. <https://doi.org/10.1371/journal.pgen.1007193>.
- 607 57. Tahiliani, M., Mei, 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. 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  
621 domain. *EMBO Reports* 10, 1235–1241. <https://doi.org/10.1038/embor.2009.218>.
- 622 65. 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,  
623 640–644. <https://doi.org/10.1038/nature13899>.
- 624 66. 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  
625 differentiated cells. *Nature* 454, 766–770. <https://doi.org/10.1038/nature07107>.

- 627 67. 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  
628 update. *Nucleic Acids Research* 51, D1188–D1195. <https://doi.org/10.1093/nar/gkac1072>.
- 629 68. Akalin, A., Kormaksson, M., Li, S., Garrett-Bakelman, F.E., Figueroa, M.E., Melnick, A., and Mason,  
C.E. (2012). methylKit: A comprehensive R package for the analysis of genome-wide DNA methylation  
630 profiles. *Genome Biol* 13, R87. <https://doi.org/10.1186/gb-2012-13-10-r87>.
- 631 69. Torres-Padilla, M.-E. (2020). On transposons and totipotency. *Philos Trans R Soc Lond B Biol Sci*  
632 375, 20190339. <https://doi.org/10.1098/rstb.2019.0339>.
- 633 70. 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*  
634 29, 400–418.e13. <https://doi.org/10.1016/j.stem.2022.01.010>.
- 635 71. 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>.
- 637 72. 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  
638 of germline genes in mouse embryonic stem cells. *Nucleic Acids Research* 51, 3130–3149. <https://doi.org/10.1093/nar/gkad071>.
- 639 73. 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  
640 genome. *Nat Genet* 39, 457–466. <https://doi.org/10.1038/ng1990>.
- 641 74. 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>.
- 643 75. 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.  
644 <https://doi.org/10.1038/s41588-020-00736-4>.
- 645 76. 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,  
646 6693–6706. <https://doi.org/10.1093/nar/gkw258>.
- 647 77. 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>.

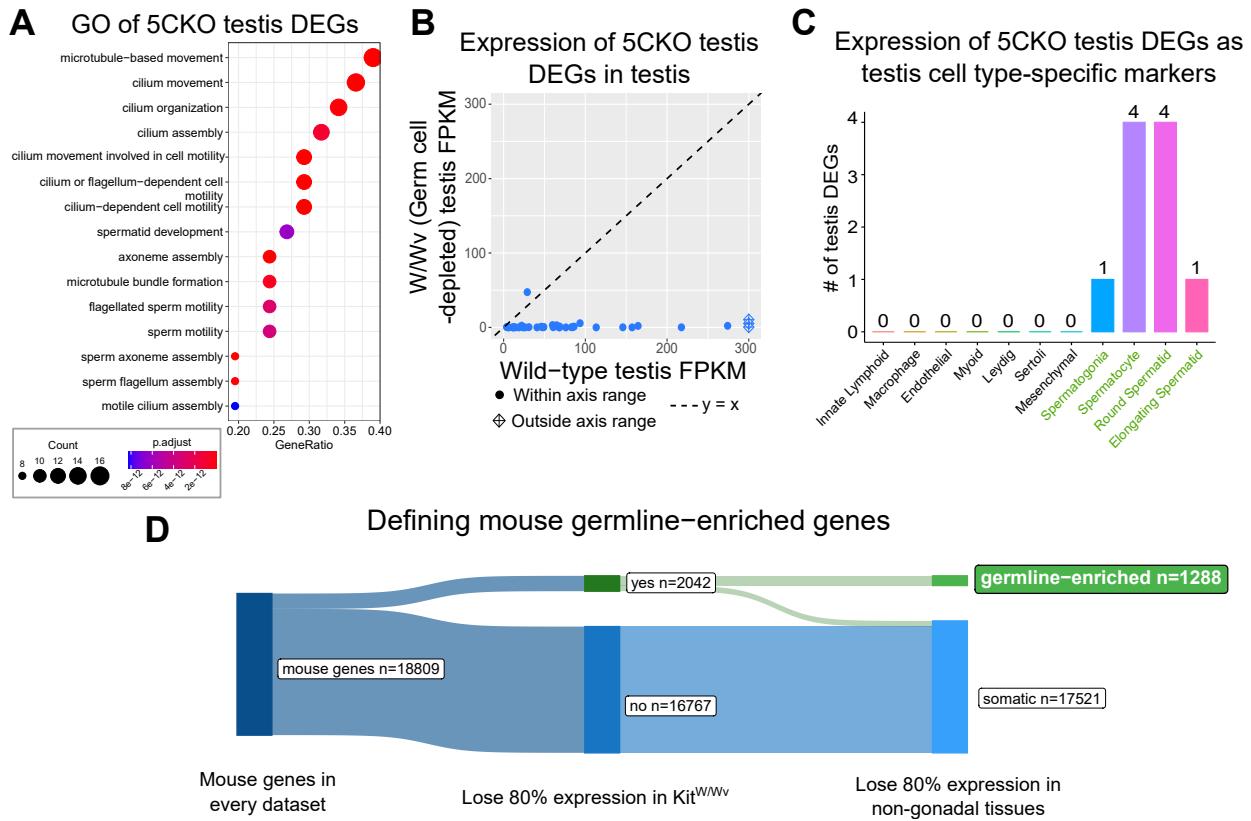
- 649 78. Nielsen, A.Y., and Gjerstorff, M.F. (2016). Ectopic Expression of Testis Germ Cell Proteins in Cancer  
650 and Its Potential Role in Genomic Instability. *Int J Mol Sci* **17**. <https://doi.org/10.3390/ijms17060890>.
- 651 79. Adebayo Babatunde, K., Najafi, A., Salehipour, P., Modarressi, M.H., and Mobasher, M.B. (2017).  
652 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>.
- 653 80. Janic, A., Mendizabal, L., Llamazares, S., Rossell, D., and Gonzalez, C. (2010). Ectopic expression  
654 of germline genes drives malignant brain tumor growth in *Drosophila*. *Science* **330**, 1824–1827.  
<https://doi.org/10.1126/science.1195481>.
- 655 81. Ghafouri-Fard, S., and Modarressi, M.-H. (2012). Expression of Cancer–Testis Genes in Brain Tumors:  
656 Implications for Cancer Immunotherapy. *Immunotherapy* **4**, 59–75. <https://doi.org/10.2217/imt.11.145>.
- 657 82. Bonefas, K.M., and Iwase, S. (2021). Soma-to-germline transformation in chromatin-linked neurode-  
658 velopmental disorders? *FEBS J*. <https://doi.org/10.1111/febs.16196>.
- 659 83. Velasco, G., Walton, E.L., Sterlin, D., Hédon, S., Nitta, H., Ito, Y., Fouyssac, F., Mégarbané, A.,  
660 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>.
- 661 84. Walton, E.L., Francastel, C., and Velasco, G. (2014). Dnmt3b Prefers Germ Line Genes and Cen-  
662 tromeric Regions: Lessons from the ICF Syndrome and Cancer and Implications for Diseases. *Biology*  
(Basel) **3**, 578–605. <https://doi.org/10.3390/biology3030578>.
- 663 85. Samaco, R.C., Mandel-Brehm, C., McGraw, C.M., Shaw, C.A., McGill, B.E., and Zoghbi, H.Y. (2012).  
664 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>.
- 665 86. Stephens, M. (2016). False discovery rates: A new deal. *Biostat*, kxw041. <https://doi.org/10.1093/biostatistics/kxw041>.
- 666 87. Conway, J.R., Lex, A., and Gehlenborg, N. (2017). UpSetR: An R package for the visualization of  
667 intersecting sets and their properties. *Bioinformatics* **33**, 2938–2940. [matics/btx364](https://doi.org/10.1093/bioinfor-<br/>matics/btx364).

669 **Figures and Tables**

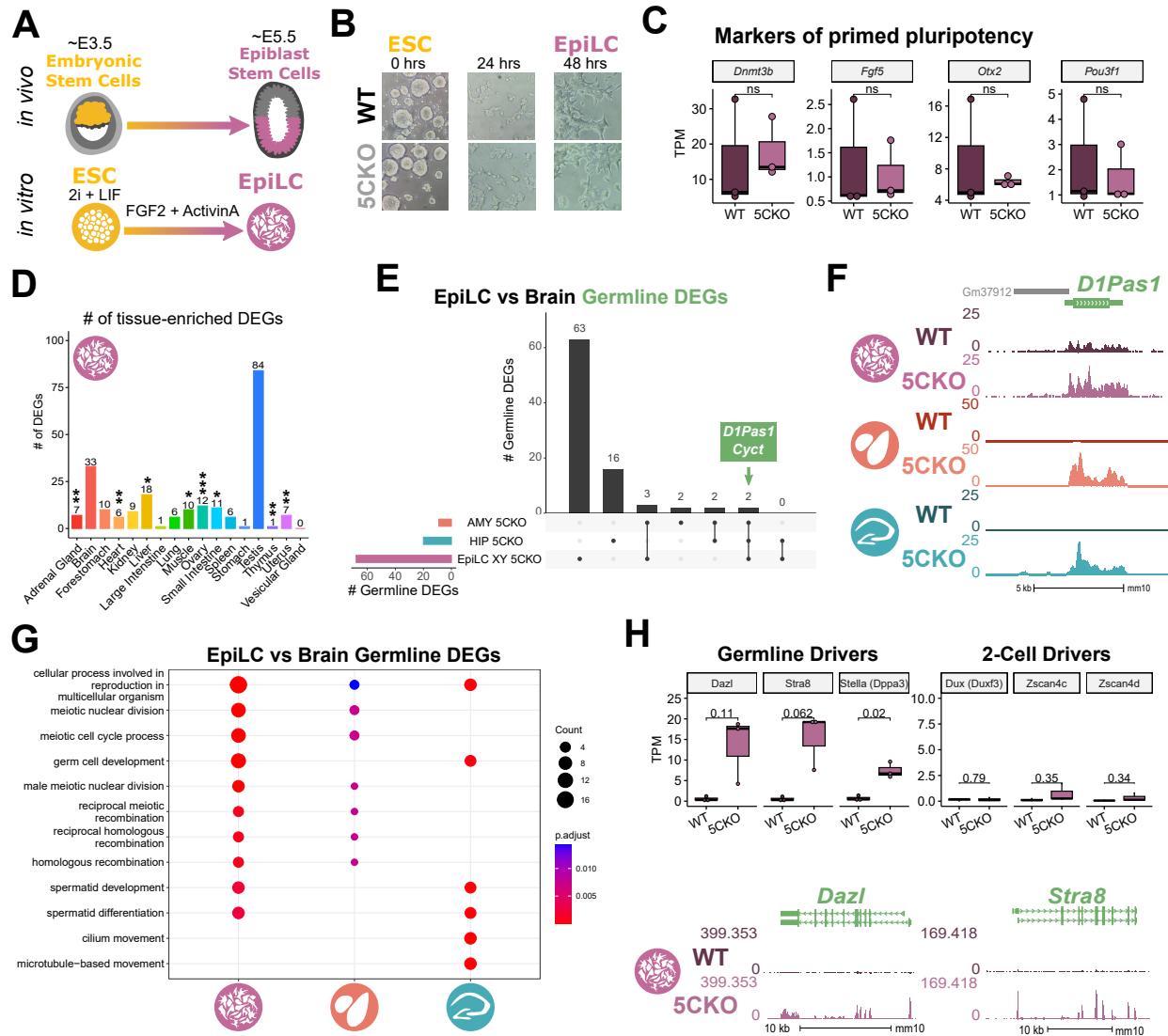
- 670     • Supplementary table 1: list of all germline genes.
- 671       – Columns to include:
- 672           \* KDM5C bound vs not
- 673           \* Log2fc in EpiLC, brain (separate columns?)
- 674       – CGI vs non



**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.** 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).

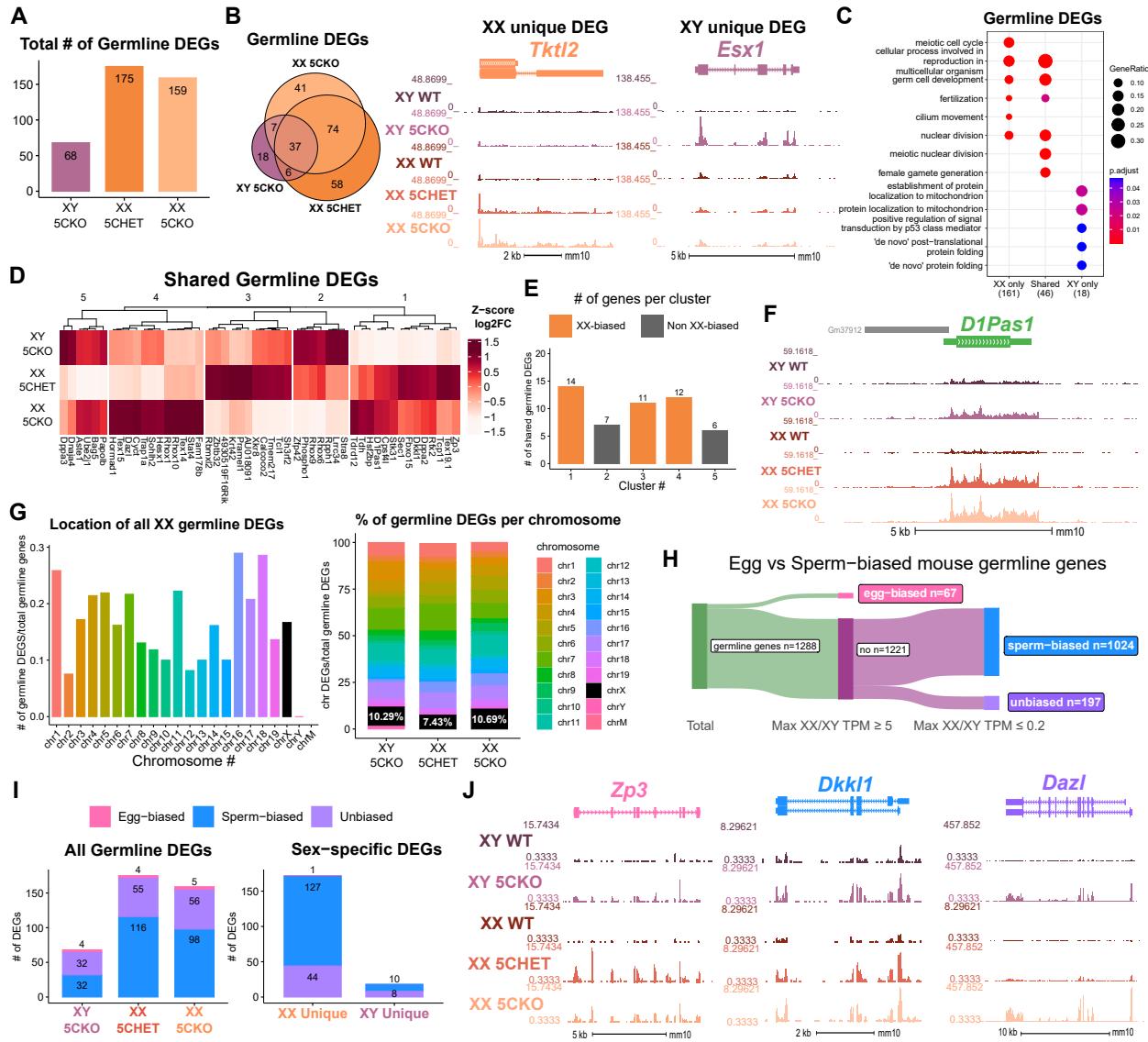
**D.** Number of tissue-enriched differentially expressed genes (DEGs) in *Kdm5c*-KO EpiLCs. \* p<0.05, \*\* p<0.01, \*\*\* p<0.001, Fisher's exact test.

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

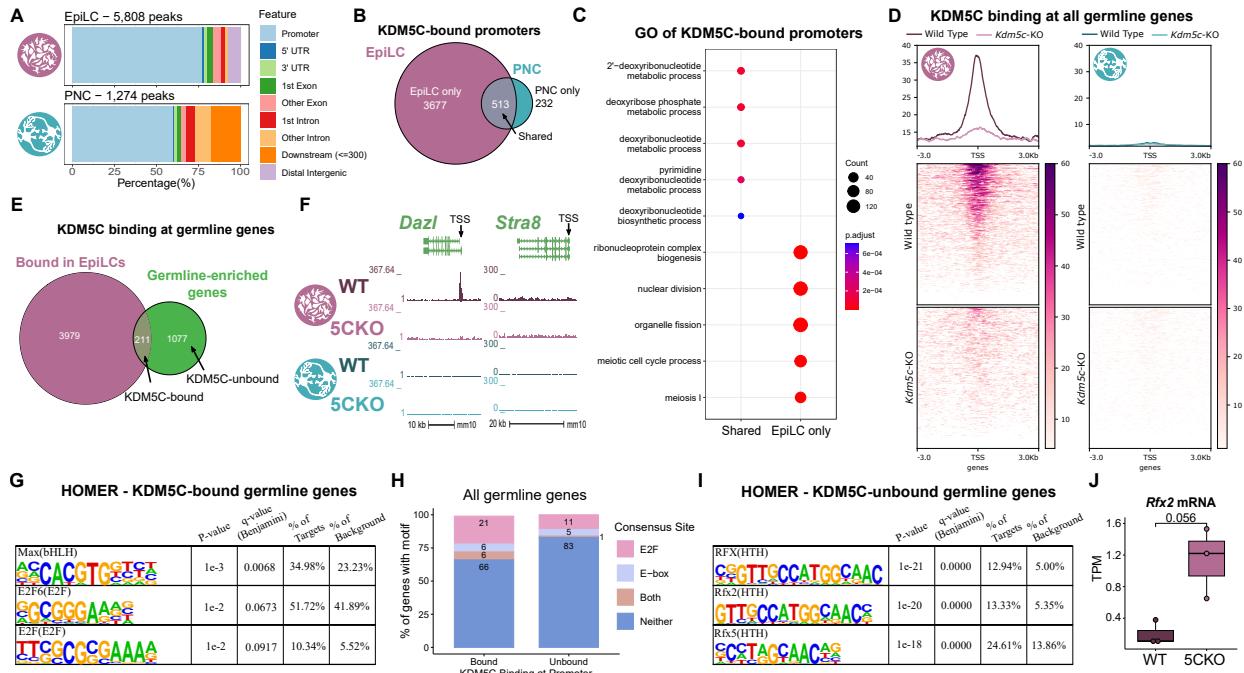
**F.** Average bigwigs of an example germline gene, *D1Pas1*, that is dysregulated *Kdm5c*-KO EpiLCs (top, purple), amygdala (middle, red), and hippocampus (bottom, blue).

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

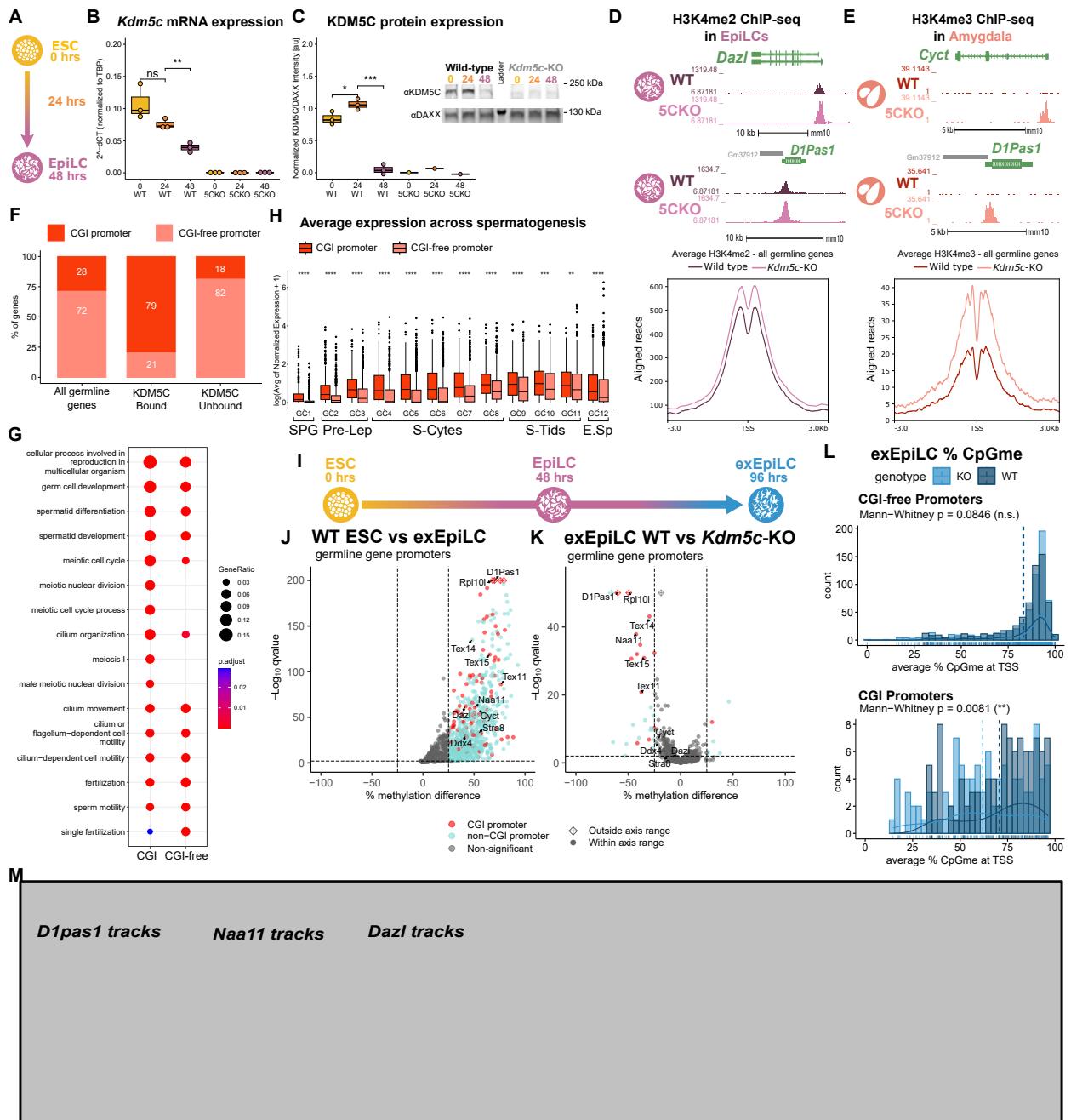
**H.** Top left - Example germline identity DEGs unique to EpiLCs. Top right - Example 2-cell genes that are not dysregulated in *Kdm5c*-KO EpiLCs. Bottom - Average bigwigs of *Dazl* and *Stra8* expression in wild-type and *Kdm5c*-KO EpiLCs.



**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*) germline genes dysregulated in both male and female *Kdm5c* mutants.



**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 ( $TSS \pm 500$ ) 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 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* 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) comprising CpG methylation at germline gene promoter (TSS ± 500) in wild-type ESCs versus exEpiLCs. Promoter CGI genes highlighted in red, hypermethylated genes lacking a promoter CGI in pink, hypomethylated genes lacking a promoter CGI in blue. **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 pink, hypomethylated genes lacking a promoter CGI in blue. **K.** Example UCSC browser shots of germline genes of CpG methylation (CpGm) in wild-type and *Kdm5c*-KO ESCs and exEpiLCs. **L.** Histogram of average percent CpGmethylated (CpGm) at the promoter for germline genes with or without promoter CGIs. Wild type vs 5CKO Mann-Whitney p = 0.0081 (\*\*). **M.** UCSC browser tracks for *D1pas1*, *Naa11*, and *Dazl* genes.