

1 Sexually dimorphic repression of germline identity by the

2 X-linked intellectual disability gene KDM5C

3

4 Katherine M. Bonefas^{1,2}, Ilakkiya Venkatachalam^{2,3}, and Shigeki Iwase².

5 1. Neuroscience Graduate Program, University of Michigan Medical School, Ann Arbor, MI, 48109, USA.

6 2. Department of Human Genetics, Michigan Medicine, University of Michigan Medical School, Ann Arbor,
7 MI, 48109, USA.

8 3. Genetics and Genomics Graduate Program, University of Michigan, Ann Arbor, MI, 48109, USA.

9 Abstract

10 Mutations in numerous chromatin-modifying enzymes cause neurodevelopmental disorders (NDDs) with
11 unknown mechanisms. Loss of repressive chromatin regulators can lead to the aberrant transcription of
12 tissue-specific genes outside of their intended context, however the mechanisms and consequences of their
13 dysregulation are largely unknown. Here, we examine how the X-linked intellectual disability gene lysine
14 demethylase 5c (KDM5C), an eraser of histone 3 lysine 4 di and tri-methylation (H3K4me2/3), contributes
15 to tissue identity. We found male *Kdm5c* knockout (-KO) mice, which recapitulate key human neurological
16 phenotypes, aberrantly express many liver, muscle, ovary, and testis genes within the amygdala and
17 hippocampus. Gonad-enriched genes misexpressed in the *Kdm5c*-KO brain are unique to germ cells,
18 indicating an erosion of the soma-germline boundary. Germline genes are typically decommissioned in
19 somatic lineages in the post-implantation epiblast, yet *Kdm5c*-KO epiblast-like cells (EpiLCs) aberrantly
20 expressed key regulators of germline identity and meiosis, including *Dazl* and *Stra8*. Characterizing germline
21 gene misexpression in males and female mutants revealed germline gene repression is sexually dimorphic,
22 with female EpiLCs requiring a higher dose of KDM5C to maintain germline gene suppression. Using a
23 comprehensive list of mouse germline-enriched genes, we found KDM5C is selectively recruited to a subset
24 of germline gene promoters that contain CpG islands (CGIs) to facilitate DNA CpG methylation (CpGme)
25 during ESC to EpiLC differentiation. However, late-stage spermatogenesis genes devoid of promoter CGIs
26 can become expressed in *Kdm5c*-KO cells via ectopic activation by RFX transcription factors. Together,
27 these data demonstrate KDM5C's fundamental role in tissue identity and indicate that KDM5C acts as a
28 brake against runaway activation of germline developmental programs in somatic lineages.

29 Introduction

30 A single genome holds the instructions to generate the myriad of cell types found within an organism.
31 This is, in part, accomplished by chromatin regulators that can either promote or impede lineage-specific
32 gene expression through DNA and histone modifications^{1–5}. Human genetic studies revealed mutations in
33 chromatin regulators are a major cause of neurodevelopmental disorders (NDDs)⁶ and many studies have
34 identified their importance for regulating brain-specific transcriptional programs. Loss of chromatin regulators
35 can also result in the ectopic expression of tissue-specific genes outside of their target environment, such
36 as the misexpression of liver-specific genes within adult neurons⁷. However, the mechanisms underlying
37 ectopic gene expression and its impact upon neurodevelopment are poorly understood.

38 To elucidate the role of tissue identity in chromatin-linked NDDs, it is essential to first characterize the
39 nature of the dysregulated genes and the molecular mechanisms governing their de-repression. Here we
40 focus on the X chromosome gene lysine demethylase 5C (KDM5C, also known as SMCX or JARID1C),
41 which erases histone 3 lysine 4 di- and trimethylation (H3K4me2/3), a permissive chromatin modification
42 enriched at gene promoters⁸. Pathogenic mutations in *KDM5C* cause Intellectual Developmental Disorder,
43 X-linked, Syndromic, Claes-Jensen Type (MRXSCJ, OMIM: 300534). MRXSCJ is more common and severe
44 in males and its neurological phenotypes include intellectual disability, seizures, aberrant aggression, and
45 autistic behaviors^{9–11}. Male *Kdm5c* knockout (-KO) mice recapitulate key MRXSCJ phenotypes, including
46 hyperaggression, increased seizure propensity, social deficits, and learning impairments^{12–14}. RNA sequenc-
47 ing (RNA-seq) of the *Kdm5c*-KO hippocampus revealed ectopic expression of some germline genes within
48 the brain¹³. However, it is unclear if other tissue-specific genes are aberrantly transcribed with KDM5C loss,
49 at what point in development germline gene misexpression begins, and what mechanisms underlie their
50 dysregulation.

51 Distinguishing between germ cells and somatic cells is a key feature of multicellularity¹⁵ that occurs
52 during early embryogenesis in many metazoans¹⁶. In mammals, chromatin regulators are crucial for
53 decommissioning germline genes during the transition from naïve to primed pluripotency. Initially, germline
54 gene promoters gain repressive histone H2A lysine 119 monoubiquitination (H2AK119ub1)¹⁷ and histone
55 H3 lysine 9 trimethylation (H3K9me3)^{17,18} in embryonic stem cells (ESCs) and are then decorated with
56 DNA CpG methylation (CpGme) in post-implantation epiblast cells^{18–21}. The contribution of KDM5C to this
57 process remains unclear. Additionally, studies on germline gene repression have primarily been conducted
58 in males and focused on select marker genes important for early germ cell development, given the lack of a
59 comprehensive list for germline-enriched genes. Therefore, it is unknown if the mechanism of repression
60 differs between sexes or for different classes of germline genes, e.g. meiotic versus spermatid differentiation
61 genes.

62 To illuminate KDM5C's role in tissue identity, here we characterized the aberrant expression of tissue-
63 enriched genes within the *Kdm5c*-KO brain and epiblast-like stem cells (EpiLCs), an *in vitro* model of the

64 post-implantation embryo. We curated a list of mouse germline-enriched genes, which enabled genome-wide
65 analysis of germline gene silencing mechanisms for the first time. Additionally, we characterized germline
66 transcripts expressed in male and female *Kdm5c* mutants to illuminate the impact of sex upon germline
67 gene suppression. Based on the data presented below, we propose KDM5C plays a fundamental, sexually
68 dimorphic role in the development of tissue identity during early embryogenesis, including the establishment
69 of the soma-germline boundary.

70 Results

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

72 Previous RNA sequencing (RNA-seq) of the adult male *Kdm5c*-KO hippocampus revealed ectopic
73 expression of some germline genes unique to the testis¹³. It is currently unknown if the testis is the only
74 tissue type misexpressed in the *Kdm5c*-KO brain. We thus systematically tested whether other tissue-specific
75 genes are misexpressed in the male brain with constitutive knockout of *Kdm5c* (*Kdm5c*^{-y}, 5CKO)²² by using
76 a published list of mouse tissue-enriched genes²³.

77 We found a large proportion of significantly upregulated genes (DESeq2²⁴, log2 fold change > 0.5, q <
78 0.1) within the male *Kdm5c*-KO amygdala and hippocampus are non-brain, tissue-specific genes (Amygdala:
79 21/59 up DEGs, 35.59% ; Hippocampus: 48/183 up DEGs, 26.23%) (Figure 1A-B, Supplementary Table
80 1). For both the amygdala and hippocampus, the majority of tissue-enriched differentially expressed genes
81 (DEGs) were testis genes (Figure 1A-B). Even though the testis has the largest total number of tissue-
82 enriched genes (2,496 genes) compared to any other tissue, testis-enriched DEGs were significantly enriched
83 in both brain regions (Amygdala p = 1.83e-05, Odds Ratio = 5.13; Hippocampus p = 4.26e-11, Odds Ratio =
84 4.45, Fisher's Exact Test). An example of a testis-enriched gene misexpressed in the *Kdm5c*-KO brain is
85 *FK506 binding protein 6 (Fkbp6)*, a known regulator of PIWI-interacting RNAs (piRNAs) and meiosis^{25,26}
86 (Figure 1C).

87 Interestingly, we also observed significant enrichment of ovary-enriched genes in both the amygdala
88 and hippocampus (Amygdala p = 0.00574, Odds Ratio = 18.7; Hippocampus p = 0.048, Odds Ratio = 5.88,
89 Fisher's Exact Test) (Figure 1A-B). Ovary-enriched DEGs included *Zygotic arrest 1 (Zar1)*, which sequesters
90 mRNAs in oocytes for meiotic maturation²⁷ (Figure 1D). Given that the *Kdm5c*-KO mice we analyzed are
91 male, these data demonstrate that the ectopic expression of gonad-enriched genes is independent of
92 organismal sex.

93 Although not consistent across brain regions, we also found significant enrichment of genes biased
94 towards two non-gonadal tissues - the liver (Amygdala p = 0.04, Odds Ratio = 6.58, Fisher's Exact Test) and
95 muscle (Hippocampus p = 0.01, Odds Ratio = 6.95, Fisher's Exact Test) (Figure 1A-B). *Apolipoprotein C-I*
96 (*Apoc1*), a lipoprotein metabolism and transport gene, is among the liver-enriched DEG derepressed in both

97 the hippocampus and amygdala²⁸ and its brain overexpression has been implicated in Alzheimer's disease²⁹
98 (Figure 1E).

99 Our analysis of oligo(dT)-primed libraries²² indicates aberrantly expressed mRNAs are polyadenylated
100 and spliced into mature transcripts in the *Kdm5c*-KO brain (Figure 1C-E). Of note, we observed little to no
101 dysregulation of brain-enriched genes (Amygdala p = 1, Odds Ratio = 1.22; Hippocampus p = 0.74, Odds
102 Ratio = 1.22, Fisher's Exact Test), despite the fact these are brain samples and the brain has the second
103 highest total number of tissue-enriched genes (708 genes). Altogether, these results suggest the aberrant
104 expression of tissue-enriched genes within the brain is a major effect of KDM5C loss.

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

106 *Kdm5c*-KO brain expresses testicular germline genes¹³, however the testis also contains somatic cells that
107 support hormone production and germline functions. To determine if *Kdm5c*-KO results in ectopic expression
108 of somatic testicular genes, we first evaluated the known functions of testicular DEGs through gene ontology.
109 We found *Kdm5c*-KO testis-enriched DEGs had high enrichment of germline-relevant ontologies, including
110 spermatid development (GO: 0007286, p.adjust = 6.2e-12) and sperm axoneme assembly (GO: 0007288,
111 p.adjust = 2.45e-14) (Figure 2A, Supplementary Table 1).

112 We then evaluated *Kdm5c*-KO testicular DEG expression in wild-type testes versus testes with germ cell
113 depletion³⁰, which was accomplished by heterozygous *W* and *Wv* mutations in the enzymatic domain of *c-Kit*
114 (*Kit*^{W/Wv})³¹. Almost all *Kdm5c*-KO testis-enriched DEGs lost expression with germ cell depletion (Figure 2B).
115 We then assessed testis-enriched DEG expression in a published single cell RNA-seq dataset that identified
116 cell type-specific markers within the testis³². Some *Kdm5c*-KO testis-enriched DEGs were classified as
117 specific markers for different germ cell developmental stages (e.g. spermatogonia, spermatocytes, round
118 spermatids, and elongating spermatids), yet none marked somatic cells (Figure 2C). Together, these data
119 demonstrate that the *Kdm5c*-KO brain aberrantly expresses germline genes but not somatic testicular genes,
120 reflecting an erosion of the soma-germline boundary.

121 As of yet, research on germline gene silencing mechanisms has focused on a handful of key genes rather
122 than assessing germline gene suppression genome-wide, due to the lack of a comprehensive gene list.
123 We therefore generated a list of mouse germline-enriched genes using RNA-seq datasets of *Kit*^{W/Wv} mice
124 that included males and females at embryonic day 12, 14, and 16³³ and adult male testes³⁰. We defined
125 genes as germline-enriched if their expression met the following criteria: 1) their expression is greater than
126 1 FPKM in wild-type gonads 2) their expression in any non-gonadal tissue of adult wild type mice²³ does
127 not exceed 20% of their maximum expression in the wild-type germline, and 3) their expression in the germ
128 cell-depleted gonads, for any sex or time point, does not exceed 20% of their maximum expression in the
129 wild-type germline. These criteria yielded 1,288 germline-enriched genes (Figure 2D), which was hereafter
130 used as a resource to globally characterize germline gene misexpression with *Kdm5c* loss (Supplementary
131 Table 2).

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

133 Germ cells are typically distinguished from somatic cells soon after the embryo implants into the uterine
134 wall^{34,35}, when germline genes are silenced in epiblast stem cells that will form the somatic tissues³⁶. This
135 developmental time point can be modeled *in vitro* through differentiation of naïve embryonic stem cells
136 (nESCs) into epiblast-like stem cells (EpiLCs) (Figure 3A)^{37,38}. While some germline-enriched genes are
137 also expressed in nESCs and in the 2-cell stage^{39–41}, they are silenced as they differentiate into EpiLCs^{18,19}.
138 Therefore, we tested if KDM5C was necessary for the initial silencing of germline genes in somatic lineages
139 by evaluating the impact of *Kdm5c* loss in male EpiLCs.

140 *Kdm5c*-KO cell morphology during ESC to EpiLC differentiation appeared normal (Figure 3B) and EpiLCs
141 properly expressed markers of primed pluripotency, such as *Dnmt3b*, *Fgf5*, *Pou3f1*, and *Otx2* (Figure 3C). We
142 then identified tissue-enriched DEGs in a RNA-seq dataset of wild-type and *Kdm5c*-KO EpiLCs⁴² (DESeq2,
143 log₂ fold change > 0.5, q < 0.1, Supplementary Table 3). Similar to the *Kdm5c*-KO brain, we observed
144 general dysregulation of tissue-enriched genes, with the largest number of genes belonging to the brain and
145 testis, although they were not significantly enriched (Figure 3D). Using our list of mouse germline-enriched
146 genes assembled above, we identified 68 germline genes misexpressed in male *Kdm5c*-KO EpiLCs.

147 We then compared EpiLC germline DEGs to those expressed in the *Kdm5c*-KO brain to determine if
148 germline genes are constitutively dysregulated or change over the course of development. The majority of
149 germline DEGs were unique to either EpiLCs or the brain, with only *D1Pas1* and *Cyct* shared across all
150 tissue/cell types (Figure 3E-F). EpiLC germline DEGs had particularly high enrichment of meiosis-related
151 gene ontologies when compared to the brain (Figure 3G, Supplementary Table 3), such as meiotic cell
152 cycle process (GO:1903046, p.adjust = 2.2e-07) and meiotic nuclear division (GO:0140013, p.adjust
153 = 1.37e-07). While there was modest enrichment of meiotic gene ontologies in both brain regions, the
154 *Kdm5c*-KO hippocampus primarily expressed late-stage spermatogenesis genes involved in sperm axoneme
155 assembly (GO:0007288, p.adjust = 0.00621) and sperm motility (GO:0097722, p.adjust = 0.00612).

156 Notably, DEGs unique to *Kdm5c*-KO EpiLCs included key drivers of germline identity, such as *Stimulated*
157 *by retinoic acid 8* (*Stra8*: log₂ fold change = 3.73, q = 2.17e-39) and *Deleted in azoospermia like* (*Dazl*:
158 log₂ fold change = 3.36, q = 3.19e-12) (Figure 3H). These genes are typically expressed when a subset
159 of epiblast stem cells become primordial germ cells (PGCs) and then again in mature germ cells to trigger
160 meiotic gene expression programs^{43–45}. Of note, some germline genes, including *Dazl*, are also expressed
161 in the two-cell embryo^{40,46}. However, we did not see derepression of two-cell stage-specific genes, like
162 *Duxf3* (*Dux*) (log₂ fold change = -0.282, q = 0.337) and *Zscan4d* (log₂ fold change = 0.25, q = 0.381) (Figure
163 3H, Supplementary Table 3), indicating *Kdm5c*-KO EpiLCs do not revert back to a 2-cell state. Altogether,
164 *Kdm5c*-KO EpiLCs express key drivers of germline identity and meiosis while the brain primarily expresses
165 spermiogenesis genes, indicating germline gene misexpression mirrors germline development during the
166 progression of somatic development.

167 **Female epiblast-like cells have heightened germline gene misexpression with *Kdm5c***
168 **loss**

169 It is currently unknown if the misexpression of germline genes is influenced by sex, as previous studies
170 on germline gene repressors have focused on male cells^{17,18,20,47,48}. Sex is particularly pertinent in the case
171 of KDM5C because it partially escapes X chromosome inactivation (XCI), resulting in a higher dosage in
172 females^{49–52}. We therefore explored the impact of chromosomal sex upon germline gene suppression by
173 comparing their dysregulation in male *Kdm5c* hemizygous knockout (*Kdm5c*^{-y}, XY *Kdm5c*-KO, XY 5CKO),
174 female homozygous knockout (*Kdm5c*^{-/-}, XX *Kdm5c*-KO, XX 5CKO), and female heterozygous knockout
175 (*Kdm5c*^{-/+}, XX *Kdm5c*-HET, XX 5CHET) EpiLCs⁴².

176 In EpiLCs, homozygous and heterozygous *Kdm5c* knockout females expressed over double the number
177 of germline-enriched genes than hemizygous males (Figure 4A, Supplementary Table 3). While the majority
178 of germline DEGs in *Kdm5c*-KO males were also dysregulated in females (74%), many were sex-specific,
179 such as *Tktl2* and *Esx1* (Figure 4B). We then compared the known functions of germline genes dysregulated
180 uniquely in males and females or misexpressed in all samples (Figure 4C, Supplementary Table 3). Female-
181 specific germline DEGs were enriched for meiotic (GO:0051321 - meiotic cell cycle) and flagellar (GO:
182 0003341 - cilium movement) functions, while male-specific DEGs had roles in mitochondrial and cell signaling
183 (GO:0070585 - protein localization to mitochondrion).

184 The majority of germline genes expressed in both sexes were more highly dysregulated in females
185 compared to males (Figure 4D-F). This increased degree of dysregulation in females, along with the
186 increased total number of germline genes, indicates females are more sensitive to losing KDM5C-mediated
187 germline gene suppression. Heightened germline gene dysregulation in females could be due to impaired
188 XCI in *Kdm5c* mutants⁴², as many spermatogenesis genes lie on the X chromosome^{53,54}. However, female
189 germline DEGs were not biased towards the X chromosome and females had a similar overall proportion
190 of germline DEGs belonging to the X chromosome as males (XY *Kdm5c*-KO - 10.29%, XX *Kdm5c*-HET -
191 7.43%, XX *Kdm5c*-KO - 10.59%) (Figure 4G). The majority of germline DEGs instead lie on autosomes for
192 both male and female *Kdm5c* mutants (Figure 4G). Thus, while female EpiLCs are more prone to germline
193 gene misexpression with KDM5C loss, it is likely independent of XCI defects.

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

195 Although many germline genes have shared functions in the male and female germline, e.g. PGC
196 formation, meiosis, and genome defense, some have unique or sex-biased expression. Therefore, we
197 wondered if *Kdm5c* mutant males would primarily express sperm genes while mutant females would primarily
198 express egg genes. To comprehensively assess whether germline gene sex corresponds with *Kdm5c*
199 mutant sex, we first filtered our list of germline-enriched genes for egg and sperm-biased genes (Figure 4,
200 Supplementary Table 2). We defined germ cell sex-biased genes as those whose expression in the opposite

201 sex, at any time point, is no greater than 20% of the gene's maximum expression in a given sex. This
202 criteria yielded 67 egg-biased, 1,024 sperm-biased, and 197 unbiased germline-enriched genes. We found
203 regardless of sex, egg, sperm, and unbiased germline genes were dysregulated in all *Kdm5c* mutants at
204 similar proportions (Figure 4I-J). Furthermore, germline genes dysregulated exclusively in either male or
205 female mutants were also not biased towards their corresponding germ cell sex (Figure 4I). Altogether, these
206 results demonstrate sex differences in germline gene dysregulation is not due to sex-specific activation of
207 sperm or egg transcriptional programs.

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

209 KDM5C binds to the promoters of several germline genes in embryonic stem cells (ESCs) but not in
210 neurons^{13,55}. However, the lack of a comprehensive list of germline-enriched genes prohibited genome-wide
211 characterization of KDM5C binding at germline gene promoters. Thus, it is unclear if KDM5C is enriched at
212 germline gene promoters, what types of germline genes KDM5C regulates, and if its binding is maintained at
213 any germline genes in neurons.

214 To address these questions, we analyzed KDM5C chromatin immunoprecipitation followed by DNA
215 sequencing (ChIP-seq) datasets in EpiLCs⁴² and primary forebrain neuron cultures (PNCs)¹² (MACS2 q <
216 0.1, fold enrichment > 1, and removal of false-positive *Kdm5c*-KO peaks). EpiLCs had a higher total number
217 of high-confidence KDM5C peaks than PNCs (EpiLCs: 5,808, PNCs: 1,276). KDM5C was primarily localized
218 to gene promoters in both cell types (promoters = transcription start site (TSS) ± 500 bp, EpiLCs: 4,190,
219 PNCs: 745), although PNCs showed increased localization to non-promoter regions (Figure 5A).

220 The majority of promoters bound by KDM5C in PNCs were also bound in EpiLCs (513 shared promoters),
221 however a large portion of gene promoters were bound by KDM5C only in EpiLCs (3,677 EpiLC only
222 promoters) (Figure 5B). Genes bound by KDM5C in both PNCs and EpiLCs were enriched for functions
223 involving nucleic acid turnover, such as deoxyribonucleotide metabolic process (GO:0009262, p.adjust =
224 8.28e-05) (Figure 5C, Supplementary Table 4). Germline ontologies were enriched only in EpiLC-specific,
225 KDM5C-bound promoters, such as meiotic nuclear division (GO: 0007127 p.adjust = 6.77e-16) (Figure 5C).
226 There were no significant ontologies for PNC-specific KDM5C target genes. Using our mouse germline gene
227 list, we observed evident KDM5C signal around the TSS of many germline genes in EpiLCs, but not in PNCs
228 (Figure 5D). Based on our ChIP-seq peak cut-off criteria, KDM5C was highly enriched at 211 germline gene
229 promoters in EpiLCs (16.4% of all germline genes) (Figure 5E, Supplementary Table 2). Of note, KDM5C
230 was only bound to about one third of RNA-seq DEG promoters unique to EpiLCs or the brain (EpiLC only
231 DEGs: 34.9%, Brain only DEGs: 30%) (Supplementary Figure 1A-C). Representative examples of EpiLC
232 DEGs bound and unbound by KDM5C in EpiLCs are *Dazl* and *Stra8*, respectively (Figure 5F). However,
233 the four of the five germline genes dysregulated in both EpiLCs and the brain were bound by KDM5C in
234 EpiLCs (*D1Pas1*, *Hsf2bp*, *Cyct*, and *Stk31*) (Supplementary Figure 1A). Together, these results demonstrate
235 KDM5C is recruited to a subset of germline genes in EpiLCs, including meiotic genes, but does not directly

236 regulate germline genes in neurons. Furthermore, the majority of germline mRNAs expressed in *Kdm5c*-KO
237 cells are dysregulated independent of direct KDM5C recruitment to their gene promoters, however genes
238 dysregulated across *Kdm5c*-KO development are often direct KDM5C targets.

239 Many germline-specific genes are suppressed by the polycomb repressive complex 1.6 (PRC1.6), which
240 contains the transcription factor heterodimers E2F/DP1 and MGA/MAX that respectively bind E2F and
241 E-box motifs within germline gene promoters^{17,18,20,41,47,48,56–58}. PRC1.6 members may recruit KDM5C to
242 germline gene promoters¹³, given their association with KDM5C in HeLa cells and ESCs^{46,59}. We thus
243 used HOMER⁶⁰ to identify transcription factor motifs enriched at KDM5C-bound or unbound germline gene
244 promoters (TSS ± 500 bp, q-value < 0.1, Supplementary Table 4). MAX and E2F6 binding sites were
245 significantly enriched at germline genes bound by KDM5C in EpiLCs (MAX q-value: 0.0068, E2F6 q-value:
246 0.0673, E2F q-value: 0.0917), but not at germline genes unbound by KDM5C (Figure 5G). One third of
247 KDM5C-bound promoters contained the consensus sequence for either E2F6 (E2F, 5'-TCCCGC-3'), MGA
248 (E-box, 5'-CACGTG-3'), or both, but only 17% of KDM5C-unbound genes contained these motifs (Figure 5H).
249 KDM5C-unbound germline genes were instead enriched for multiple RFX transcription factor binding sites
250 (RFX q-value < 0.0001, RFX2 q-value < 0.0001, RFX5 q-value < 0.0001) (Figure 5I, Supplementary figure
251 1D). RFX transcription factors bind X-box motifs⁶¹ to promote ciliogenesis^{62,63} and among them is RFX2, a
252 central regulator of post-meiotic spermatogenesis^{64,65}. Although *Rfx2* is also not a direct target of KDM5C
253 (Supplementary Figure 1E), RFX2 mRNA is derepressed in *Kdm5c*-KO EpiLCs (Figure 5J). Thus, RFX2 is a
254 candidate transcription factor for driving the ectopic expression of many KDM5C-unbound germline genes in
255 *Kdm5c*-KO cells.

256 **KDM5C is recruited to CpG islands at germline promoters to facilitate *de novo* DNA 257 methylation**

258 Previous work found two germline gene promoters have a marked reduction in DNA CpG methylation
259 (CpGme) in the adult *Kdm5c*-KO hippocampus¹³. Since histone 3 lysine 4 di- and trimethylation (H3K4me2/3)
260 impede *de novo* CpGme^{66,67}, KDM5C's removal of H3K4me2/3 may be required to suppress germline
261 genes. However, KDM5C's catalytic activity was recently shown to be dispensable for suppressing *Dazl* in
262 undifferentiated ESCs⁴⁶. To reconcile these observations, we hypothesized KDM5C erases H3K4me2/3 to
263 promote the initial placement of CpGme at germline gene promoters in EpiLCs.

264 To test this hypothesis, we first characterized KDM5C's expression as naïve ESCs differentiate into
265 EpiLCs (Figure 6A). While *Kdm5c* mRNA steadily decreased from 0 to 48 hours of differentiation (Figure
266 6B), KDM5C protein initially increased from 0 to 24 hours and then decreased to near knockout levels by 48
267 hours (Figure 6C). We then characterized KDM5C's substrates (H3K4me2/3) at germline gene promoters
268 with *Kdm5c* loss using published ChIP-seq datasets^{22,42}. *Kdm5c*-KO samples showed a marked increase in
269 H3K4me2 in EpiLCs (Figure 6D) and H3K4me3 in the amygdala (Figure 6E) around the TSS of germline

270 genes. Together, these data suggest KDM5C acts during the transition between ESCs and EpiLCs to remove
271 H3K4me2/3 at germline gene promoters.

272 Germline genes accumulate CpG methylation (CpGme) at CpG islands (CGIs) during the transition
273 from naïve to primed pluripotency^{19,21,68}. We first examined how many of our germline-enriched genes had
274 promoter CGIs (TSS ± 500 bp) using the UCSC genome browser⁶⁹. Notably, out of 1,288 germline-enriched
275 genes, only 356 (27.64%) had promoter CGIs (Figure 6F, Supplementary Table 2). CGI-containing germline
276 genes had higher enrichment of meiotic gene ontologies compared to CGI-free genes, including meiotic
277 nuclear division (GO:0140013, p.adjust = 2.17e-12) and meiosis I (GO:0007127, p.adjust = 3.91e-10)
278 (Figure 6G, Supplementary Table 5). Germline genes with promoter CGIs were more highly expressed than
279 CGI-free genes across spermatogenesis stages, with highest expression in meiotic spermatocytes (Figure
280 6H). Contrastingly, CGI-free genes only displayed substantial expression in post-meiotic round spermatids
281 (Figure 6H). Although only a minor portion of germline gene promoters contained CGIs, CGIs strongly
282 determined KDM5C's recruitment to germline genes ($p = 2.37e-67$, Odds Ratio = 17.8, Fisher's Exact Test),
283 with 79.15% of KDM5C-bound germline gene promoters harboring CGIs (Figure 6F).

284 To assess how KDM5C loss impacts initial CpGme placement at germline gene promoters, we performed
285 whole genome bisulfite sequencing (WGBS) in male wild-type and *Kdm5c*-KO ESCs and 96-hour extend
286 EpiLCs (exEpiLCs), when germline genes reach peak methylation levels¹⁸ (Figure 6I). We first identified
287 which germline gene promoters significantly gained CpGme in wild-type cells during nESC to exEpiLCs
288 differentiation (methylKit⁷⁰, $q < 0.01$, $|methylation\ difference| > 25\%$, TSS ± 500 bp). In wild-type cells, the
289 majority of germline genes gained substantial CpGme at their promoter during differentiation (60.08%),
290 regardless if their promoter contained a CGI (Figure 6J, Supplementary Table 5).

291 We then identified promoters differentially methylated in wild-type versus *Kdm5c*-KO exEpiLCs (methylKit,
292 $q < 0.01$, $|methylation\ difference| > 25\%$, TSS ± 500 bp, Supplementary Table 5). Of the 48,882 promoters
293 assessed, 274 promoters were significantly hypomethylated and 377 promoters were significantly hyper-
294 methylated with KDM5C loss (Supplementary Figure 2A). Many promoters hyper- and hypomethylated
295 in *Kdm5c*-KO exEpiLCs belonged to genes with unknown functions. However, 10.22% of hypomethyl-
296 ated promoters belonged to germline genes and germline-relevant ontologies like meiotic nuclear division
297 (GO:0140013, p.adjust = 0.012) are significantly enriched (Supplementary Figure 2B, Supplementary Table
298 5). Approximately half of all germline gene promoters hypomethylated in *Kdm5c*-KO exEpiLCs are direct
299 targets of KDM5C in EpiLCs (13 out of 28 hypomethylated promoters).

300 Promoters that showed the most robust loss of CpGme in *Kdm5c*-KO exEpiLCs (lowest q-values) harbored
301 CGIs (Figure 6K). CGI promoters, but not CGI-free promoters, had a significant reduction in CpGme with
302 KDM5C loss as a whole (Figure 6L) (Non-CGI promoters $p = 0.0846$, CGI promoters $p = 0.0081$, Mann-
303 Whitney U test). Significantly hypomethylated promoters included germline genes consistently dysregulated
304 across multiple *Kdm5c*-KO RNA-seq datasets¹³, such as *D1Pas1* (methylation difference = -60.03%, q-value
305 = 3.26e-153) and *Naa11* (methylation difference = -42.45%, q-value = 1.44e-38) (Figure 6M). Surprisingly,

306 we observed only a modest reduction in CpGme at *Dazl*'s promoter (methylation difference = -6.525%,
307 q-value = 0.0159) (Figure 6N). Altogether, these results demonstrate KDM5C is recruited to germline gene
308 CGIs in EpiLCs to promote CpGme at germline gene promoters. Furthermore, this suggests while KDM5C's
309 catalytic activity is required for the repression of some germline genes, CpGme can be placed at others even
310 with elevated H3K4me2/3 around the TSS.

311 Discussion

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

328 By comparing *Kdm5c* mutant males and females, we revealed germline gene suppression is sexually
329 dimorphic. Female EpiLCs are more severely impacted by loss of KDM5C-mediated germline gene sup-
330 pression, yet this difference is not due to the large number of germline genes on the X chromosome^{53,54}.
331 Increased germline gene misexpression in females may be related to females having a higher dose of
332 KDM5C than males, due to its escape from XCI^{49–52}. Intriguingly, heterozygous knockout females (*Kdm5c*^{-/+})
333 also had over double the number of germline DEGs than hemizygous knockout males (*Kdm5c*^{-/Y}), even
334 though their expression of KDM5C should be roughly equivalent to that of wild-type males (*Kdm5c*^{+/Y}). Males
335 could partially compensate for KDM5C's loss via the Y-chromosome homolog, KDM5D. However, KDM5D
336 exhibits weaker demethylase activity than KDM5C⁸ and has not been reported to regulate germline gene
337 expression. Nevertheless, these results demonstrate germline gene silencing mechanisms differ between
338 males and females, which warrants further study to elucidate the biological ramifications and underlying
339 mechanisms.

340 We found KDM5C is largely dispensable for promoting normal gene expression during development, yet
341 is critical for suppressing ectopic developmental programs. It is important to note that while we highlighted
342 KDM5C's repression of germline genes, some germline-enriched genes like *Dazl* are also expressed at the 2-
343 cell stage and in the inner cell mass/naïve ESCs for their role in pluripotency and self-renewal^{41,46,71,72}. These
344 "self-renewal" germline genes are then silenced during ESC differentiation into epiblast stem cells/EpiLCs^{18,19}.
345 We found that while *Kdm5c*-KO EpiLCs express *Dazl*, they did not express 2-cell-specific genes like *Zscan4c*.
346 These data suggest the 2-cell-like state reported in *Kdm5c*-KO ESCs⁴⁶ likely reflects KDM5C's primary
347 role in germline gene repression. Germline gene misexpression in *Kdm5c*-KO EpiLCs may indicate they
348 are differentiating into primordial germ cell-like cells (PGCLCs)^{34,35,37}. Yet, *Kdm5c*-KO EpiLCs had normal
349 cellular morphology and properly expressed markers for primed pluripotency, including *Otx2* which blocks
350 EpiLC differentiation into PGCs/PGCLCs⁷³. In addition to unimpaired EpiLC differentiation, *Kdm5c*-KO gross
351 brain morphology is overall normal¹² and hardly any brain-specific genes were significantly dysregulated in
352 the amygdala and hippocampus. Thus, ectopic germline gene expression occurs in conjunction with overall
353 proper somatic differentiation in *Kdm5c*-KO animals.

354 Our work provides novel insight into the cross-talk between H3K4me2/3 and CpGme, which are gen-
355 erally mutually exclusive⁷⁴. In EpiLCs, loss of KDM5C binding at a subset of germline gene promoters,
356 e.g. *D1Pas1*, strongly impaired promoter CGI methylation and resulted in their long-lasting de-repression into
357 adulthood. Removal of H3K4me2/3 at CGIs is a plausible mechanism for KDM5C-mediated germline gene
358 suppression^{13,55}, given H3K4me2/3 can oppose DNMT3 activity^{66,67}. However, emerging work indicates
359 many histone-modifying enzymes have non-catalytic functions that influence gene expression, sometimes
360 even more potently than their catalytic roles^{75,76}. Indeed, KDM5C's catalytic activity was recently found to be
361 dispensible for repressing *Dazl* in ESCs⁴⁶. In our study, *Dazl*'s promoter still gained CpGme in *Kdm5c*-KO
362 exEpiLCs, even with elevated H3K4me2. *Dazl* and a few other germline genes employ multiple repressive
363 mechanisms to facilitate CpGme, such as DNMT3A/B recruitment via E2F6 and MGA^{17,18,47,48}. This suggests
364 alternative silencing mechanisms are sufficient to recruit DNMT3s to some germline CGIs, while others may
365 require KDM5C-mediated H3K4me removal to overcome promoter CGI escape from CpGme^{74,77}. These
366 results also suggest the requirement for KDM5C's catalytic activity can change depending upon the locus
367 and developmental stage, even for the same class of genes. However, further experiments are required to
368 determine if catalytically inactive KDM5C can suppress germline genes at later developmental stages.

369 By generating a comprehensive list of mouse germline-enriched genes, we were able to reveal distinct
370 derepressive mechanisms governing early versus late-stage germline developmental programs. Previous
371 work on germline gene silencing has focused on genes with promoter CGIs^{19,74}, and indeed the major-
372 ity of KDM5C targets in EpiLCs were germ cell identity genes harboring CGIs. However, over 70% of
373 germline-enriched gene promoters lacked CGIs, including the many KDM5C-unbound germline genes
374 that are de-repressed in *Kdm5c*-KO cells. CGI-free, KDM5C-unbound germline genes were primarily
375 late-stage spermatogenesis genes and significantly enriched for RFX2 binding sites, a central regulator

376 of spermiogenesis^{64,65}. These data suggest that once activated during early embryogenesis, drivers of
377 germline identity like *Rfx2*, *Stra8*, and *Dazl* turn on downstream germline programs, ultimately culminating in
378 the expression of spermiogenesis genes in the adult *Kdm5c*-KO brain. Therefore, we propose KDM5C is
379 recruited via promoter CGIs to act as a brake against runaway activation of germline-specific programs.

380 The above work provides the mechanistic foundation for KDM5C-mediated repression of tissue and
381 germline-specific genes. However, the contribution of these ectopic, tissue-specific genes towards *Kdm5c*-
382 KO neurological impairments is still unknown. In addition to germline genes, we also identified significant
383 enrichment of muscle and liver-enriched transcripts within the *Kdm5c*-KO brain. Intriguingly, select liver and
384 muscle-enriched DEGs do have known roles within the brain, such as the liver-enriched lipid metabolism
385 gene *Apolipoprotein C-I (Apoc1)*²⁸. *APOC1* dysregulation is implicated in Alzheimer's disease in humans²⁹
386 and overexpression of *Apoc1* in the mouse brain can impair learning and memory⁷⁸. KDM5C may therefore
387 be crucial for neurodevelopment by fine-tuning the expression of tissue-enriched, dosage-sensitive genes
388 like *Apoc1*.

389 Given that germline genes have no known functions within the brain, their impact upon neurodevelopment
390 is currently unknown. In *C. elegans*, somatic misexpression of germline genes via loss of *Retinoblastoma*
391 (*Rb*) homologs results in enhanced piRNA signaling and ectopic P granule formation in neurons^{79,80}. Ectopic
392 testicular germline transcripts have also been observed in a variety of cancers, including brain tumors in
393 *Drosophila* and mammals^{81,82} and shown to promote cancer progression⁸³⁻⁸⁵. Intriguingly, mouse models
394 and human cells for other chromatin-linked NDDs also display impaired soma-germline demarcation⁸⁶⁻⁸⁸,
395 such as mutations in DNA methyltransferase 3b (DNMT3B), H3K9me1/2 methyltransferases G9A/GLP,
396 and methyl-CpG -binding protein 2 (MECP2). Recently, the transcription factor ZMYM2 (ZNF198), whose
397 mutation causes neurodevelopmental-craniofacial syndrome with variable renal and cardiac abnormalities
398 (OMIM #619522), was also shown to repress germline genes by promoting H3K4 methylation removal and
399 CpGme⁸⁹. Thus, KDM5C is among a growing cohort of neurodevelopmental disorders that have erosion of
400 the germline-soma boundary. Further research is required to determine the impact of these germline genes
401 upon neuronal functions and the extent to which this phenomenon occurs in humans.

402 Materials and Methods

403 Classifying tissue-enriched and germline-enriched genes

404 Tissue-enriched differentially expressd genes (DEGs) were determined by their classification in a previ-
405 ously published dataset from 17 male and female mouse tissues²³. This study defined tissue expression as
406 greater than 1 Fragments Per Kilobase of transcript per Million mapped read (FPKM) and tissue enrichment
407 as at least 4-fold higher expression than any other tissue.

408 We curated a list of germline-enriched genes using an RNA-seq dataset from wild-type and germline-

409 depleted (*Kit*^{W/W^v}) male and female mouse embryos from embryonic day 12, 14, and 16³³, as well as adult
410 male testes³⁰. Germline-enriched genes met the following criteria: 1) their expression is greater than 1
411 FPKM in wild-type germline 2) their expression in any wild-type somatic tissues²³ does not exceed 20%
412 of maximum expression in wild-type germline, and 3) their expression in the germ cell-depleted (*Kit*^{W/W^v})
413 germline, for any sex or time point, does not exceed 20% of maximum expression in wild-type germline. We
414 defined sperm and egg-biased genes as those whose expression in the opposite sex, at any time point, is no
415 greater than 20% of the gene's maximum expression in a given sex. Genes that did not meet this threshold
416 for either sex were classified as 'unbiased'.

417 Cell culture

418 We utilized our previously established cultures of male wild-type and *Kdm5c* knockout (-KO)
419 embryonic stem cells⁴². Sex was confirmed by genotyping *Uba1/Uba1y* on the X and Y chromo-
420 somes with the following primers: 5'-TGGATGGTGTGCCAATG-3', 5'-CACCTGCACGTTGCCCTT-
421 3'. Deletion of *Kdm5c* exons 11 and 12, which destabilize KDM5C protein¹², was confirmed
422 through the primers 5'-ATGCCCATATTAAGAGTCCCTG-3', 5'-TCTGCCTTGATGGACTGTT-3', and
423 5'-GGTTCTAACACTCACATAGTG-3'.

424 Embryonic stem cells (ESCs) and epiblast-like cells were cultured using previously established
425 methods³⁸. Briefly, ESCs were initially cultured in primed ESC (pESC) media consisting of KnockOut
426 DMEM (Gibco#10829–018), fetal bovine serum (Gibco#A5209501), KnockOut serum replacement
427 (Invitrogen#10828–028), Glutamax (Gibco#35050-061), Anti-Anti (Gibco#15240-062), MEM Non-essential
428 amino acids (Gibco#11140-050), and beta-mercaptoethanol (Sigma#M7522). They were then transitioned
429 into ground-state, "naïve" ESCs (nESCs) by culturing for four passages in N2B27 media containing
430 DMEM/F12 (Gibco#11330–032), Neurobasal media (Gibco#21103–049), Gluamax (Gibco#35050-061),
431 Anti-Anti (Gibco#15240-062), N2 supplement (Invitrogen#17502048), and B27 supplement without vitamin
432 A (Invitrogen#12587-010), and beta-mercaptoethanol (Sigma#M7522). Both pESC and nESC media
433 were supplemented with 3 μ M GSK3 inhibitor CHIR99021 (Sigma #SML1046-5MG), 1 μ M MEK inhibitor
434 PD0325901 (Sigma #PZ0162-5MG), and 1,000 units/mL leukemia inhibitory factor (LIF, Millipore#ESG1107).

435 nESCs were differentiated into epiblast-like cells (EpiLCs, 48 hours) and extendend EpiLCs (exEpiLCs,
436 96 hours) by culturing in N2B27 media containing DMEM/F12, Neurobasal media, Gluamax, Anti-Anti, N2
437 supplement, B27 supplement (Invitrogen#17504044), and beta-mercaptoethanol supplemented with 10
438 ng/mL fibroblast growth factor 2 (FGF2, R&D Biotechne 233-FB), and 20 ng/mL activin A (R&D Biotechne
439 338AC050CF), as previously described³⁸.

440 Real time quantitative PCR (RT-qPCR)

441 nESCs were differentiated into EpiLCs as described above. Cells were lysed with Tri-reagent BD (Sigma
442 #T3809) at 0, 24, and 48 hours of differentiation. RNA was phase separated with 0.1 uL/uL 1-bromo-3-
443 chloropropane (Sigma #B9673) and then precipitated with isopropanol (Sigma #I9516) and ethanol puri-
444 fied. For each sample, 2 μ g of RNA was reverse transcribed using the ProtoScript II Reverse transcriptase kit
445 from New England Biolabs (NEB #M0368S) and primed with oligo dT. Expression of *Kdm5c* was detected us-
446 ing the primers 5'-CCCATGGAGGCCAGAGAATAAG-3' 5'-CTCAGCGGATAAGAGAATTGCTAC-3' and nor-
447 malized to TBP using the primers 5'-TTCAGAGGATGCTCTAGGGAAGA-3' 5'-CTGTGGAGTAAGTCCTGTGCC-
448 3' with the Power SYBR™ Green PCR Master Mix (ThermoFisher #4367659).

449 Western Blot

450 Total protein was extracted during nESC to EpiLC differentiation at 0, 24, and 48 hours by sonicating cells
451 at 20% amplitude for 15 seconds in 2X SDS sample buffer, then boiling at 100 °C for 10 minutes. Proteins
452 were separated on a 7.5% SDS page gel, transferred overnight onto a fluorescent membrane, blotted for
453 rabbit anti-KDM5C (in house, 1:500) and mouse anti-DAXX (Santa Cruz #(H-7): sc-8043, 1:500) and imaged
454 using the LiCor Odyssey CLx system. Band intensity was quantified using ImageJ.

455 RNA sequencing (RNA-seq) data analysis

456 After ensuring read quality via FastQC (v0.11.8), reads were then mapped to the mm10 *Mus musculus*
457 genome (Gencode) using STAR (v2.5.3a), during which we removed duplicates and kept only uniquely
458 mapped reads. Count files were generated by FeatureCounts (Subread v1.5.0), and BAM files were
459 converted to bigwigs using deeptools (v3.1.3) and visualized by the UCSC genome browser⁶⁹. RStudio
460 (v3.6.0) was then used to analyze counts files by DESeq2 (v1.26.0)²⁴ to identify differentially expressed
461 genes (DEGs) with a q-value (p-adjusted via FDR/Benjamini–Hochberg correction) less than 0.1 and a log2
462 fold change greater than 0.5. For all DESeq2 analyses, log2 fold changes were calculated with IfcShrink
463 using the ashR package⁹⁰. MA-plots were generated by ggpibr (v0.6.0), and Eulerr diagrams were generated
464 by eulerr (v6.1.1). Boxplots and scatterplots were generated by ggpibr (v0.6.0) and ggplot2 (v3.3.2). The
465 Upset plot was generated via the package UpSetR (v1.4.0)⁹¹. Gene ontology (GO) analyses were performed
466 by the R package enrichPlot (v1.16.2) using the biological processes setting and compareCluster.

467 Chromatin immunoprecipitation followed by DNA sequencing (ChIP-seq) data analysis

468 ChIP-seq reads were aligned to mm10 using Bowtie1 (v1.1.2) allowing up to two mismatches. Only
469 uniquely mapped reads were used for analysis. Peaks were called using MACS2 software (v2.2.9.1) using
470 input BAM files for normalization, with filters for a q-value < 0.1 and a fold enrichment > 1. We removed

471 “black-listed” genomic regions that often give aberrant signals. Common peak sets were obtained in R via
472 DiffBind⁹² (v3.6.5). In the case of KDM5C ChIP-seq, *Kdm5c*-KO peaks false-positive peaks were then
473 removed from wild-type samples using bedtools (v2.25.0). Peak proximity to genomic loci was determined
474 by ChIPSeeker⁹³ (v1.32.1). Gene ontology (GO) analyses were performed by the R package enrichPlot
475 (v1.16.2) using the biological processes setting and compareCluster. Enriched motifs were identified using
476 HOMER⁶⁰ to search for known motifs within 500 base pairs of the transcription start site. Average binding
477 across genes was visualized using deeptools (v3.1.3). Bigwigs were visualized using the UCSC genome
478 browser⁶⁹.

479 **CpG island (CGI) analysis**

480 Locations of CpG islands were determined through the mm10 UCSC genome browser CpG island track⁶⁹,
481 which classified CGIs as regions that have greater than 50% GC content, are larger than 200 base pairs,
482 and have a ratio of CG dinucleotides observed over the expected amount greater than 0.6. CGI genomic
483 coordinates were then annotated using ChIPseeker⁹³ (v1.32.1) and filtered for ones that lie within promoters
484 of germline-enriched genes (TSS ± 500).

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

486 Genomic DNA (gDNA) from naïve ESCs and extended EpiLCs was extracted using the Wizard Genomic
487 DNA Purification Kit (Promega A1120), following the instructions for Tissue Culture Cells. gDNA from
488 two wild-types and two *Kdm5c*-KOs of each cell type was sent to Novogene for WGBS using the Illumina
489 NovaSeq X Plus platform and sequenced for 150bp paired-end reads (PE150). All samples had greater
490 than 99% bisulfite conversion rates. Reads were adapter and quality trimmed with Trim Galore (v0.6.10)
491 and aligned to the mm10 genome using Bismark⁹⁴ (v0.22.1). Analysis of differential methylation at germline
492 gene promoters was performed using methylKit⁷⁰ (v1.28.0) with a minimum coverage of 3 paired reads, a
493 percentage greater than 25% or less than -25%, and q-value less than 0.01. methylKit was also used to
494 calculate average percentage methylation at germline gene promoters. Methylation bedgraph tracks were
495 generated via Bismark and visualized using the UCSC genome browser⁶⁹.

496 **Data availability**

497 **WGBS in wild-type and *Kdm5c*-KO ESCs and exEpiLCs**

498 Raw fastq files are deposited in the Sequence Read Archive (SRA) <https://www.ncbi.nlm.nih.gov/sra>
499 under the bioProject XXX

500 **Published datasets**

501 All published datasets are available at the Gene Expression Omnibus (GEO) <https://www.ncbi.nlm.nih.gov/geo>. Published RNA-seq datasets analyzed in this study included the male wild-type and *Kdm5c*-KO
502 adult amygdala and hippocampus²² (available at GEO: GSE127722). Male and female wild-type, *Kdm5c*-KO,
503 and *Kdm5c*-HET EpiLCs⁴² are available at GEO: GSE96797.

504 Previously published ChIP-seq experiments included KDM5C binding in wild-type and *Kdm5c*-KO
505 EpiLCs⁴² (available at GEO: GSE96797) and mouse primary neuron cultures (PNCs) from the cortex
506 and hippocampus¹² (available at GEO: GSE61036). ChIP-seq of histone 3 lysine 4 dimethylation (H3K4me2)
507 in male wild-type and *Kdm5c*-KO EpiLCs⁴² is also available at GEO: GSE96797. ChIP-seq of histone 3 lysine
508 4 trimethylation (H3K4me3) in wild-type and *Kdm5c*-KO male amygdala²² are available at GEO: GSE127817.
509

510 **Data analysis**

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

513 **Acknowledgements**

514 We thank Drs. Sundeep Kalantry, Milan Samanta, and Rebecca Malcore for providing protocols and
515 expertise in culturing mouse ESCs and EpiLCs, as well as providing the wild-type and *Kdm5c*-KO ESCs
516 used in this study. We thank Dr. Jacob Mueller for his insight in germline gene regulation and directing us to
517 the germline-depleted mouse models. We also thank Drs. Gabriel Corfas, Kenneth Kwan, Natalie Tronson,
518 Michael Sutton, Stephanie Bielas, Donna Martin, and the members of the Iwase, Sutton, Bielas, and Martin
519 labs for helpful discussions and critiques of the data. We thank members of the University of Michigan
520 Reproductive Sciences Program for providing feedback throughout the development of this work. This work
521 was supported by grants from the National Institutes of Health (NIH) (National Institute of Neurological
522 Disorders and Stroke: NS089896, 5R21NS104774, and NS116008 to S.I.), Farrehi Family Foundation Grant
523 (to S.I.), the University of Michigan Career Training in Reproductive Biology (NIH T32HD079342, to K.M.B.),
524 the NIH Early Stage Training in the Neurosciences Training Grant (NIH T32NS076401 to K.M.B.), and the
525 Michigan Predoctoral Training in Genetics Grant (NIH T32GM007544, to I.V.)

526 **Author contributions**

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

530 **References**

- 531 1. Strahl, B.D., and Allis, C.D. (2000). The language of covalent histone modifications. *Nature* **403**,
532 41–45. <https://doi.org/10.1038/47412>.
- 533 2. Jenuwein, T., and Allis, C.D. (2001). Translating the Histone Code. *Science* **293**, 1074–1080.
534 <https://doi.org/10.1126/science.1063127>.
- 535 3. Lewis, E.B. (1978). A gene complex controlling segmentation in *Drosophila*. *Nature* **276**, 565–570.
536 <https://doi.org/10.1038/276565a0>.
- 537 4. Kennison, J.A., and Tamkun, J.W. (1988). Dosage-dependent modifiers of polycomb and antennapedia
538 mutations in *Drosophila*. *Proc. Natl. Acad. Sci. U.S.A.* **85**, 8136–8140. <https://doi.org/10.1073/pnas.8>
538 5.21.8136.
- 539 5. Kassis, J.A., Kennison, J.A., and Tamkun, J.W. (2017). Polycomb and Trithorax Group Genes in
540 *Drosophila*. *Genetics* **206**, 1699–1725. <https://doi.org/10.1534/genetics.115.185116>.
- 541 6. Gabriele, M., Lopez Tobon, A., D'Agostino, G., and Testa, G. (2018). The chromatin basis of
542 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>.
- 543 7. Schaefer, A., Sampath, S.C., Intrator, A., Min, A., Gertler, T.S., Surmeier, D.J., Tarakhovsky, A.,
544 and Greengard, P. (2009). Control of cognition and adaptive behavior by the GLP/G9a epigenetic
544 suppressor complex. *Neuron* **64**, 678–691. <https://doi.org/10.1016/j.neuron.2009.11.019>.
- 545 8. Iwase, S., Lan, F., Bayliss, P., De La Torre-Ubieta, L., Huarte, M., Qi, H.H., Whetstine, J.R., Bonni, A.,
546 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>
546 7.02.017.
- 547 9. Claes, S., Devriendt, K., Van Goethem, G., Roelen, L., Meireleire, J., Raeymaekers, P., Cassiman,
548 J.J., and Fryns, J.P. (2000). Novel syndromic form of X-linked complicated spastic paraparesia. *Am J
Med Genet* **94**, 1–4.
- 549 10. Jensen, L.R., Amende, M., Gurok, U., Moser, B., Gimmel, V., Tzschach, A., Janecke, A.R., Tariverdian,
550 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>.
- 551 11. Carmignac, V., Nambot, S., Lehalle, D., Callier, P., Moortgat, S., Benoit, V., Ghoumid, J., Delobel,
552 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>.

- 553 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>.
- 554
- 555 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>.
- 556
- 557 14. Bonefas, K.M., Vallianatos, C.N., Raines, B., Tronson, N.C., and Iwase, S. (2023). Sexually Dimorphic
Alterations in the Transcriptome and Behavior with Loss of Histone Demethylase KDM5C. *Cells* *12*,
637. <https://doi.org/10.3390/cells12040637>.
- 558
- 559 15. 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>.
- 560
- 561 16. 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>.
- 562
- 563 17. 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>.
- 564
- 565 18. 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>.
- 566
- 567 19. 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>.
- 568
- 569 20. 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>.
- 570
- 571 21. 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>.
- 572

- 573 22. 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>.
- 574
- 575 23. 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>.
- 576
- 577 24. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for
578 RNA-seq data with DESeq2. *Genome Biol* 15, 550. <https://doi.org/10.1186/s13059-014-0550-8>.
- 579 25. 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
580 Chromosome Pairing in Meiosis. *Science* 300, 1291–1295. <https://doi.org/10.1126/science.1083022>.
- 581 26. 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
582 Amplification and Transposon Silencing. *Molecular Cell* 47, 970–979. <https://doi.org/10.1016/j.molcel.2012.07.019>.
- 583 27. 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
584 membraneless compartment. *Science* 378, eabq4835. <https://doi.org/10.1126/science.abq4835>.
- 585 28. Rouland, A., Masson, D., Lagrost, L., Vergès, B., Gautier, T., and Bouillet, B. (2022). Role of
586 apolipoprotein C1 in lipoprotein metabolism, atherosclerosis and diabetes: A systematic review.
Cardiovasc Diabetol 21, 272. <https://doi.org/10.1186/s12933-022-01703-5>.
- 587 29. Leduc, V., Jasmin-Bélanger, S., and Poirier, J. (2010). APOE and cholesterol homeostasis in
588 Alzheimer's disease. *Trends in Molecular Medicine* 16, 469–477. <https://doi.org/10.1016/j.molmed.2010.07.008>.
- 589 30. Mueller, J.L., Skaletsky, H., Brown, L.G., Zaghlul, S., Rock, S., Graves, T., Auger, K., Warren,
590 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>.
- 591 31. Handel, M.A., and Eppig, J.J. (1979). Sertoli Cell Differentiation in the Testes of Mice Genetically
592 Deficient in Germ Cells. *Biology of Reproduction* 20, 1031–1038. <https://doi.org/10.1095/biolreprod20.5.1031>.
- 593 32. 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>.
- 594

- 595 33. Soh, Y.Q., Junker, J.P., Gill, M.E., Mueller, J.L., van Oudenaarden, A., and Page, D.C. (2015). A
596 Gene Regulatory Program for Meiotic Prophase in the Fetal Ovary. *PLoS Genet* **11**, e1005531.
<https://doi.org/10.1371/journal.pgen.1005531>.
- 597 34. Magnúsdóttir, E., and Surani, M.A. (2014). How to make a primordial germ cell. *Development* **141**,
598 245–252. <https://doi.org/10.1242/dev.098269>.
- 599 35. Günesdogan, U., Magnúsdóttir, E., and Surani, M.A. (2014). Primordial germ cell specification: A
600 context-dependent cellular differentiation event [corrected]. *Philos Trans R Soc Lond B Biol Sci* **369**.
<https://doi.org/10.1098/rstb.2013.0543>.
- 601 36. 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>.
- 603 37. Hayashi, K., Ohta, H., Kurimoto, K., Aramaki, S., and Saitou, M. (2011). Reconstitution of the
604 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>.
- 605 38. Samanta, M., and Kalantry, S. (2020). Generating primed pluripotent epiblast stem cells: A methodol-
606 ogy chapter. *Curr Top Dev Biol* **138**, 139–174. <https://doi.org/10.1016/bs.ctdb.2020.01.005>.
- 607 39. 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.
608 *EMBO Reports* **16**, 791–802. <https://doi.org/10.15252/embr.201540538>.
- 609 40. 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
610 activity. *Nature* **487**, 57–63. <https://doi.org/10.1038/nature11244>.
- 611 41. 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
612 germline stem cells. *Nat Commun* **7**, 11056. <https://doi.org/10.1038/ncomms11056>.
- 613 42. 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>.
- 615 43. 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**,
616 2474–2479. <https://doi.org/10.1073/pnas.0510813103>.

- 617 44. Lin, Y., Gill, M.E., Koubova, J., and Page, D.C. (2008). Germ Cell-Intrinsic and -Extrinsic Factors
618 Govern Meiotic Initiation in Mouse Embryos. *Science* *322*, 1685–1687. <https://doi.org/10.1126/science.1166340>.
- 619 45. Endo, T., Mikedis, M.M., Nicholls, P.K., Page, D.C., and De Rooij, D.G. (2019). Retinoic Acid and Germ
620 Cell Development in the Ovary and Testis. *Biomolecules* *9*, 775. <https://doi.org/10.3390/biom9120775>.
- 621 46. Gupta, N., Yakhou, L., Albert, J.R., Azogui, A., Ferry, L., Kirsh, O., Miura, F., Battault, S., Yamaguchi,
622 K., Laisné, M., et al. (2023). A genome-wide screen reveals new regulators of the 2-cell-like cell state.
Nat Struct Mol Biol. <https://doi.org/10.1038/s41594-023-01038-z>.
- 623 47. Pohlers, M., Truss, M., Frede, U., Scholz, A., Strehle, M., Kuban, R.-J., Hoffmann, B., Morkel, M.,
Birchmeier, C., and Hagemeier, C. (2005). A Role for E2F6 in the Restriction of Male-Germ-Cell-
624 Specific Gene Expression. *Current Biology* *15*, 1051–1057. <https://doi.org/10.1016/j.cub.2005.04.060>.
- 625 48. Dahlet, T., Truss, M., Frede, U., Al Adhami, H., Bardet, A.F., Dumas, M., Vallet, J., Chicher, J.,
Hammann, P., Kottnik, S., et al. (2021). E2F6 initiates stable epigenetic silencing of germline genes
626 during embryonic development. *Nat Commun* *12*, 3582. <https://doi.org/10.1038/s41467-021-23596-w>.
- 627 49. Agulnik, A.I., Mitchell, M.J., Mattei, M.G., Borsani, G., Avner, P.A., Lerner, J.L., and Bishop, C.E.
(1994). A novel X gene with a widely transcribed Y-linked homologue escapes X-inactivation in mouse
628 and human. *Hum Mol Genet* *3*, 879–884. <https://doi.org/10.1093/hmg/3.6.879>.
- 629 50. 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>.
- 630 51. 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*
631 *5*, 1355–1360. <https://doi.org/10.1093/hmg/5.9.1355>.
- 632 52. 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>.
- 633 53. Wang, P.J., McCarrey, J.R., Yang, F., and Page, D.C. (2001). An abundance of X-linked genes
expressed in spermatogonia. *Nat Genet* *27*, 422–426. <https://doi.org/10.1038/86927>.
- 634 54. 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
inactivation. *Nat Genet* *36*, 642–646. <https://doi.org/10.1038/ng1368>.

- 639 55. Al Adhami, H., Vallet, J., Schaal, C., Schumacher, P., Bardet, A.F., Dumas, M., Chicher, J., Hammann,
640 P., Daujat, S., and Weber, M. (2023). Systematic identification of factors involved in the silencing
of germline genes in mouse embryonic stem cells. *Nucleic Acids Research* *51*, 3130–3149. <https://doi.org/10.1093/nar/gkad071>.
- 641 56. Hurlin, P.J. (1999). Mga, a dual-specificity transcription factor that interacts with Max and contains a
642 T-domain DNA-binding motif. *The EMBO Journal* *18*, 7019–7028. <https://doi.org/10.1093/emboj/18.2.4.7019>.
- 643 57. Stielow, B., Finkernagel, F., Stiewe, T., Nist, A., and Suske, G. (2018). MGA, L3MBTL2 and E2F6
644 determine genomic binding of the non-canonical Polycomb repressive complex PRC1.6. *PLoS Genet*
14, e1007193. <https://doi.org/10.1371/journal.pgen.1007193>.
- 645 58. 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
646 repression of germ cell-related genes in mouse embryonic stem cells. *PLoS ONE* *13*, e0205969.
<https://doi.org/10.1371/journal.pone.0205969>.
- 647 59. 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.
648 *Nature* *447*, 601–605. <https://doi.org/10.1038/nature05823>.
- 649 60. 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
650 cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Molecular Cell* *38*, 576–589.
<https://doi.org/10.1016/j.molcel.2010.05.004>.
- 651 61. 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*,
652 916–921. <https://doi.org/10.1038/35002634>.
- 653 62. 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).
- 655 63. 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
656 Ciliogenesis. *Sci. Signal.* *2*. <https://doi.org/10.1126/scisignal.2000602>.
- 657 64. 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.
658 *PLoS Genet* *11*, e1005368. <https://doi.org/10.1371/journal.pgen.1005368>.

- 659 65. Wu, Y., Hu, X., Li, Z., Wang, M., Li, S., Wang, X., Lin, X., Liao, S., Zhang, Z., Feng, X., et al.
660 (2016). Transcription Factor RFX2 Is a Key Regulator of Mouse Spermiogenesis. *Sci Rep* 6, 20435.
<https://doi.org/10.1038/srep20435>.
- 661 66. Otani, J., Nankumo, T., Arita, K., Inamoto, S., Ariyoshi, M., and Shirakawa, M. (2009). Structural basis
662 for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX–DNMT3–DNMT3L
domain. *EMBO Reports* 10, 1235–1241. <https://doi.org/10.1038/embor.2009.218>.
- 663 67. Guo, X., Wang, L., Li, J., Ding, Z., Xiao, J., Yin, X., He, S., Shi, P., Dong, L., Li, G., et al. (2015).
664 Structural insight into autoinhibition and histone H3-induced activation of DNMT3A. *Nature* 517,
640–644. <https://doi.org/10.1038/nature13899>.
- 665 68. Meissner, A., Mikkelsen, T.S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein,
666 B.E., Nusbaum, C., Jaffe, D.B., et al. (2008). Genome-scale DNA methylation maps of pluripotent and
differentiated cells. *Nature* 454, 766–770. <https://doi.org/10.1038/nature07107>.
- 667 69. 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
668 update. *Nucleic Acids Research* 51, D1188–D1195. <https://doi.org/10.1093/nar/gkac1072>.
- 669 70. 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
670 profiles. *Genome Biol* 13, R87. <https://doi.org/10.1186/gb-2012-13-10-r87>.
- 671 71. Torres-Padilla, M.-E. (2020). On transposons and totipotency. *Philos Trans R Soc Lond B Biol Sci*
672 375, 20190339. <https://doi.org/10.1098/rstb.2019.0339>.
- 673 72. 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*
674 29, 400–418.e13. <https://doi.org/10.1016/j.stem.2022.01.010>.
- 675 73. 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. [018-0581-5](https://doi.org/10.1038/s41586-
676 018-0581-5).
- 677 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
678 genome. *Nat Genet* 39, 457–466. <https://doi.org/10.1038/ng1990>.
- 679 75. 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>.

- 681 76. Morgan, M.A.J., and Shilatifard, A. (2020). Reevaluating the roles of histone-modifying enzymes
682 and their associated chromatin modifications in transcriptional regulation. *Nat Genet* 52, 1271–1281.
<https://doi.org/10.1038/s41588-020-00736-4>.
- 683 77. Long, H.K., King, H.W., Patient, R.K., Odom, D.T., and Klose, R.J. (2016). Protection of CpG
684 islands from DNA methylation is DNA-encoded and evolutionarily conserved. *Nucleic Acids Res* 44,
6693–6706. <https://doi.org/10.1093/nar/gkw258>.
- 685 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>.
- 686 79. Wang, D., Kennedy, S., Conte, D., Kim, J.K., Gabel, H.W., Kamath, R.S., Mello, C.C., and Ruvkun,
G. (2005). Somatic misexpression of germline P granules and enhanced RNA interference in
688 retinoblastoma pathway mutants. *Nature* 436, 593–597. <https://doi.org/10.1038/nature04010>.
- 689 80. Wu, X., Shi, Z., Cui, M., Han, M., and Ruvkun, G. (2012). Repression of germline RNAi pathways
in somatic cells by retinoblastoma pathway chromatin complexes. *PLoS Genet* 8, e1002542. <https://doi.org/10.1371/journal.pgen.1002542>.
- 690 81. Nielsen, A.Y., and Gjerstorff, M.F. (2016). Ectopic Expression of Testis Germ Cell Proteins in Cancer
692 and Its Potential Role in Genomic Instability. *Int J Mol Sci* 17. <https://doi.org/10.3390/ijms17060890>.
- 693 82. Adebayo Babatunde, K., Najafi, A., Salehipour, P., Modarressi, M.H., and Mobasheri, M.B. (2017).
Cancer/Testis genes in relation to sperm biology and function. *Iranian Journal of Basic Medical
694 Sciences* 20. <https://doi.org/10.22038/ijbms.2017.9259>.
- 695 83. 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.
696 <https://doi.org/10.1126/science.1195481>.
- 697 84. Ghafouri-Fard, S., and Modarressi, M.-H. (2012). Expression of Cancer–Testis Genes in Brain Tumors:
698 Implications for Cancer Immunotherapy. *Immunotherapy* 4, 59–75. <https://doi.org/10.2217/imt.11.145>.
- 699 85. Nin, D.S., and Deng, L.-W. (2023). Biology of Cancer-Testis Antigens and Their Therapeutic Implications
700 in Cancer. *Cells* 12, 926. <https://doi.org/10.3390/cells12060926>.
- 701 86. Bonefas, K.M., and Iwase, S. (2021). Soma-to-germline transformation in chromatin-linked neurode-
702 velopmental disorders? *FEBS J.* <https://doi.org/10.1111/febs.16196>.
- 703 87. Velasco, G., Walton, E.L., Sterlin, D., Hédouin, S., Nitta, H., Ito, Y., Fouyssac, F., Mégarbané, A.,
Sasaki, H., Picard, C., et al. (2014). Germline genes hypomethylation and expression define a
molecular signature in peripheral blood of ICF patients: Implications for diagnosis and etiology.
704 *Orphanet J Rare Dis* 9, 56. <https://doi.org/10.1186/1750-1172-9-56>.

- 705 88. Walton, E.L., Francastel, C., and Velasco, G. (2014). Dnmt3b Prefers Germ Line Genes and Cen-
706 tromeric Regions: Lessons from the ICF Syndrome and Cancer and Implications for Diseases. *Biology*
(Basel) 3, 578–605. <https://doi.org/10.3390/biology3030578>.
- 707 89. Graham-Paquin, A.-L., Saini, D., Sirois, J., Hossain, I., Katz, M.S., Zhuang, Q.K.-W., Kwon, S.Y.,
Yamanaka, Y., Bourque, G., Bouchard, M., et al. (2023). ZMYM2 is essential for methylation of
germline genes and active transposons in embryonic development. *Nucleic Acids Research*, gkad540.
<https://doi.org/10.1093/nar/gkad540>.
- 708 90. Stephens, M. (2016). False discovery rates: A new deal. *Biostat*, kxw041. <https://doi.org/10.1093/biostatistics/kxw041>.
- 709 91. Conway, J.R., Lex, A., and Gehlenborg, N. (2017). UpSetR: An R package for the visualization of
710 intersecting sets and their properties. *Bioinformatics* 33, 2938–2940. <https://doi.org/10.1093/bioinformatics/btx364>.
- 711 92. Ross-Innes, C.S., Stark, R., Teschendorff, A.E., Holmes, K.A., Ali, H.R., Dunning, M.J., Brown, G.D.,
712 Gojis, O., Ellis, I.O., Green, A.R., et al. (2012). Differential oestrogen receptor binding is associated
with clinical outcome in breast cancer. *Nature* 481, 389–393. <https://doi.org/10.1038/nature10730>.
- 713 93. Yu, G., Wang, L.G., and He, Q.Y. (2015). ChIPseeker: An R/Bioconductor package for ChIP peak
714 annotation, comparison and visualization. *Bioinformatics* 31, 2382–2383. <https://doi.org/10.1093/bioinformatics/btv145>.
- 715 94. Krueger, F., and Andrews, S.R. (2011). Bismark: A flexible aligner and methylation caller for Bisulfite-
716 Seq applications. *Bioinformatics* 27, 1571–1572. <https://doi.org/10.1093/bioinformatics/btr167>.

719 **Figures and Tables**

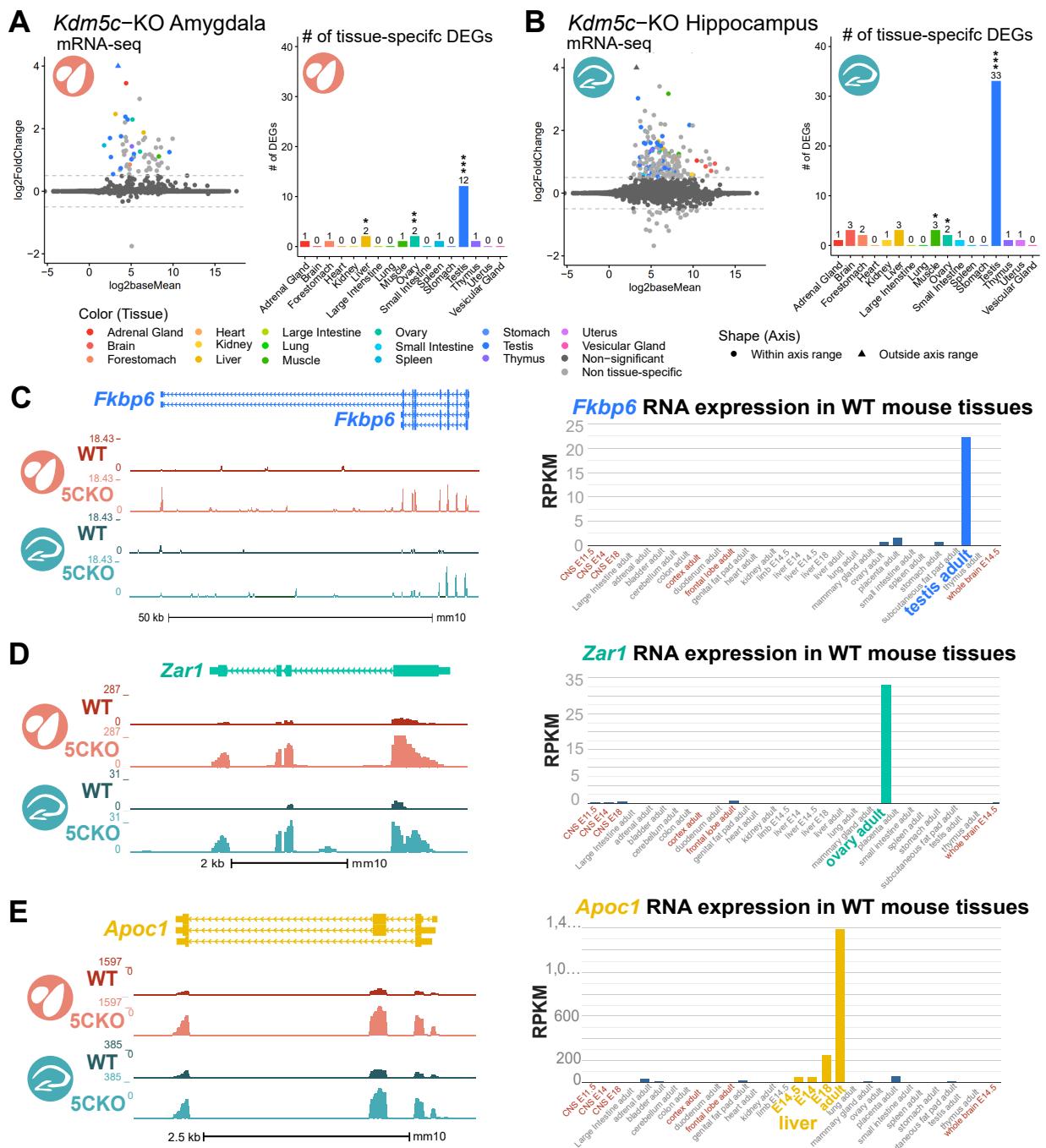


Figure 1: Tissue-enriched genes are misexpressed in the *Kdm5c*-KO brain. **A-B.** Expression of tissue-enriched genes (Li et al 2017) in the male *Kdm5c*-KO amygdala (A) and hippocampus (B). Left - MA plot of mRNA-sequencing. Right - Number of tissue-enriched differentially expressed genes (DEGs). * p<0.05, ** p<0.01, *** p<0.001, Fisher's Exact Test. **C.** Left - UCSC browser view 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) (Average, n = 4). Right - Expression of *Cyct* in wild-type tissues from NCBI Gene, with testis highlighted in blue and brain tissues highlighted in red. **D.** Left - UCSC browser view 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 - UCSC browser view 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.

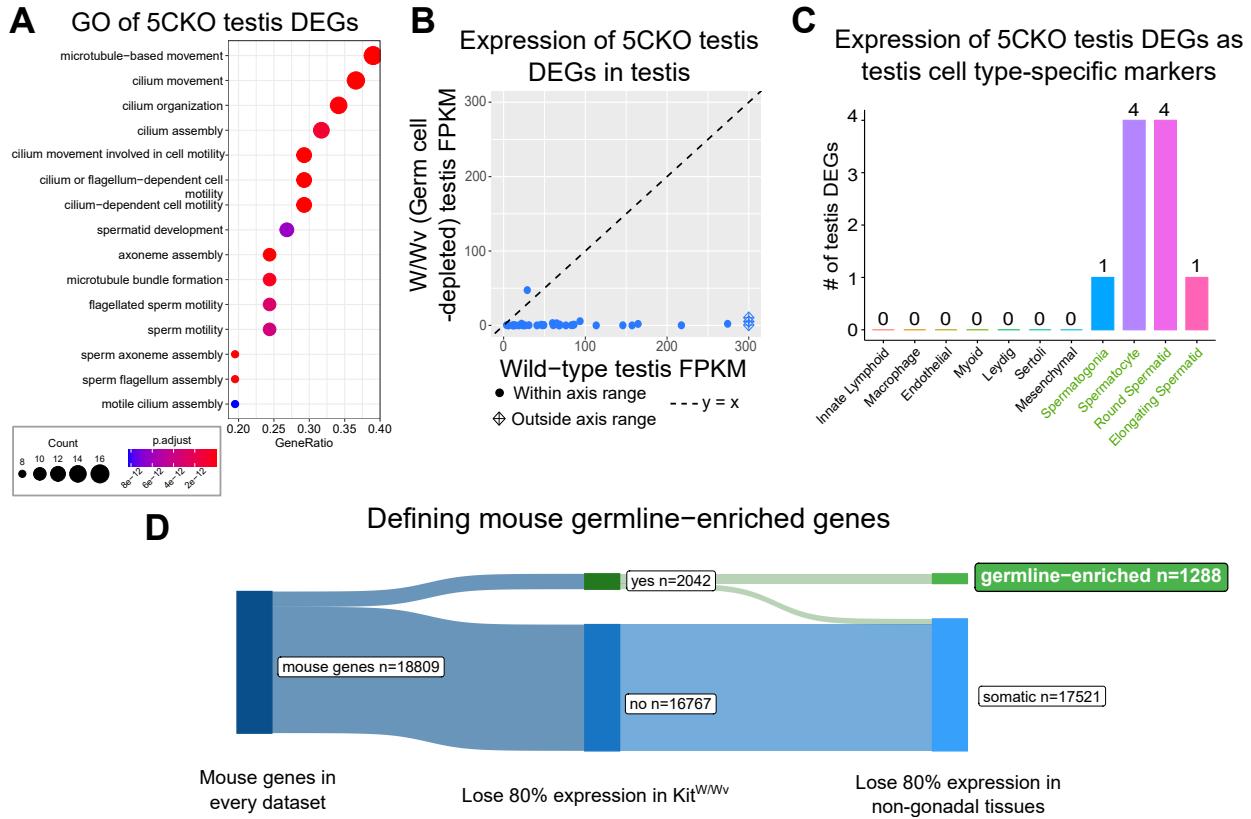


Figure 2: Aberrant transcription of germline genes in the *Kdm5c*-KO in the brain. **A.** enrichPlot gene ontology (GO) of *Kdm5c*-KO amygdala and hippocampus testis-enriched DEGs **B.** Expression of testis DEGs in wild-type (WT) testis versus germ cell-depleted (W/Wv) testis (Mueller et al 2013). Expression 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 W/Wv mice and in adult mouse non-gonadal tissues (Li et al 2017).

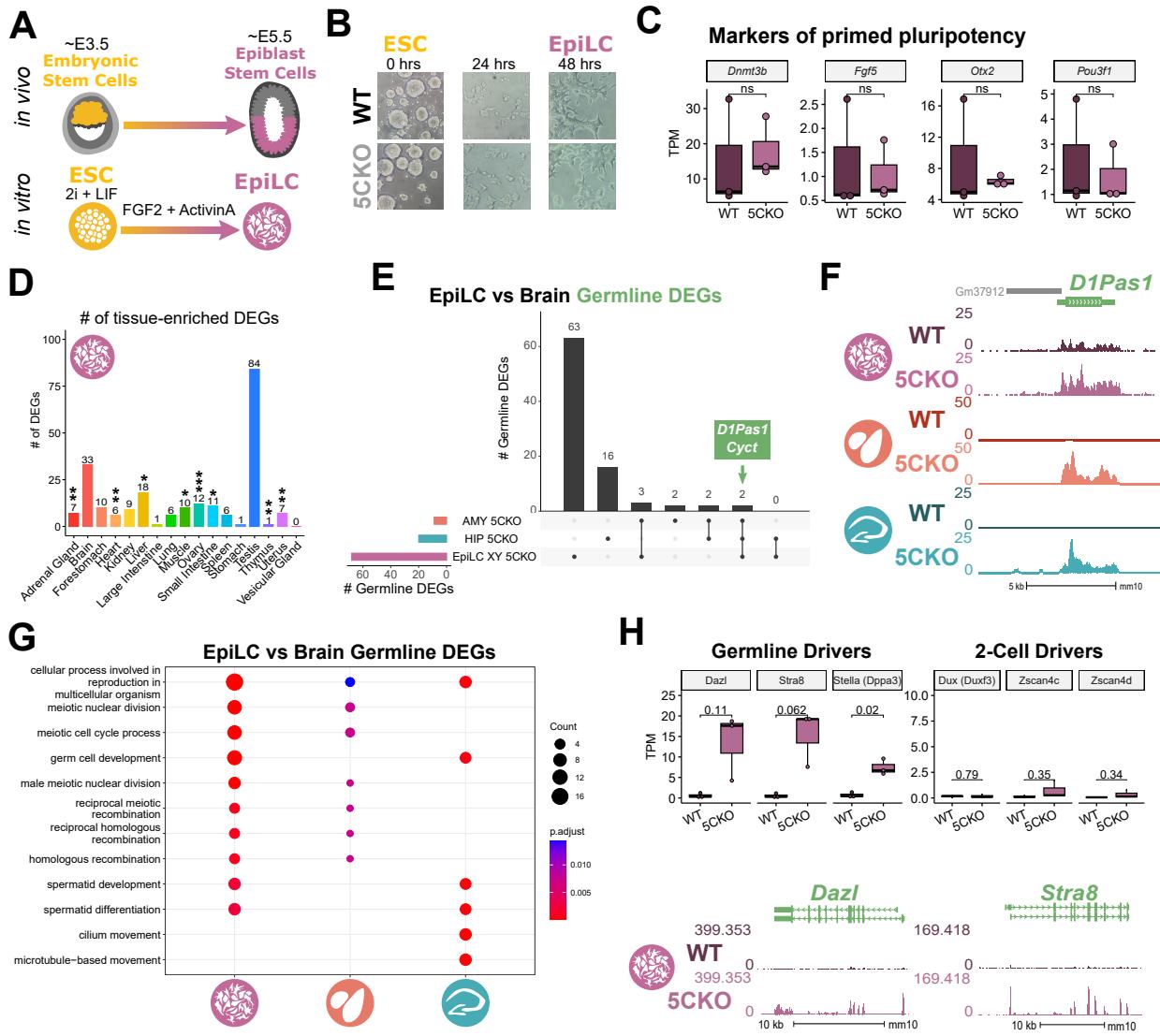


Figure 3: *Kdm5c*-KO epiblast-like cells express key drivers of germline identity

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 male wild-type (WT) and *Kdm5c*-KO (5CKO) 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, amygdala (AMY), and hippocampus (HIP) RNA-seq datasets. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, 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. UCSC browser view of an example germline gene, *D1Pas1*, that is dysregulated *Kdm5c*-KO EpiLCs (top, purple. Average, $n = 3$), amygdala (middle, red. Average, $n = 4$), and hippocampus (bottom, blue. Average, $n = 4$).

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 - UCSC browser view of *Dazl* and *Stra8* expression in wild-type and *Kdm5c*-KO EpiLCs (Average, $n = 3$).

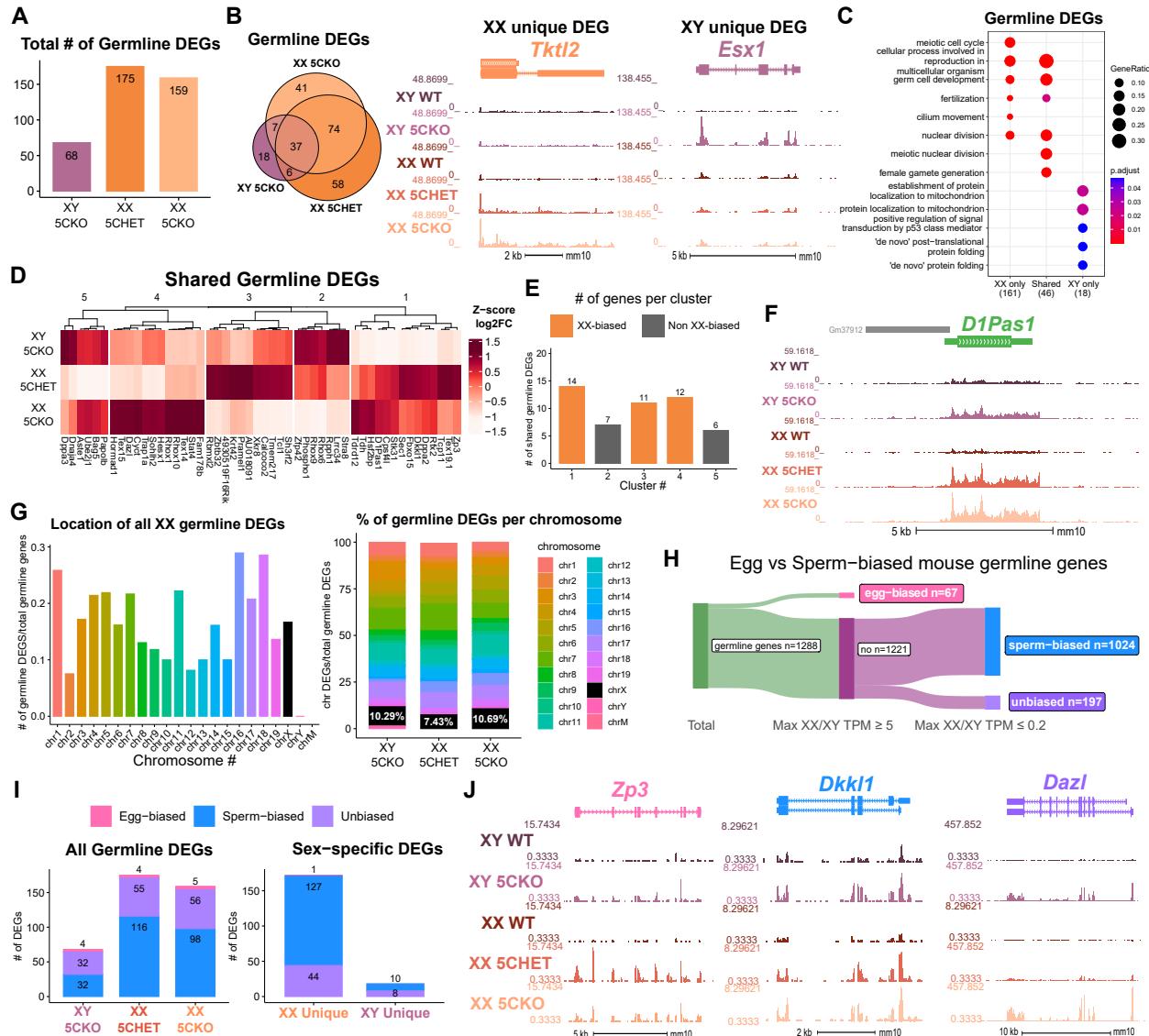


Figure 4: Chromosomal sex influences *Kdm5c*-KO germline gene misexpression. **A.** Total number of germline-enriched RNA-seq DEGs for male hemizygous *Kdm5c* knockout EpilCs (XY 5CKO, purple), female heterozygous *Kdm5c* knockout (XX 5CHET, orange), female homozygous *Kdm5c* knockout (XX 5CKO, light 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, 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 log 2 fold change from sex-matched wild-type, z-scored across rows. **E.** Number of genes within each cluster from D. Clusters with higher expression in females compared to males (XX-biased) highlighted in orange. **F.** UCSC browser view of a male and female shared germline DEG *D1Pas1* that is more highly expressed in female mutants (Average, n = 3). **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.** UCSC browser view of egg-biased (*Zp3*), sperm-biased (*Dkk1*), and unbiased (*Dazl*) germline genes dysregulated in both male and female *Kdm5c* mutants (Average of n = 3).

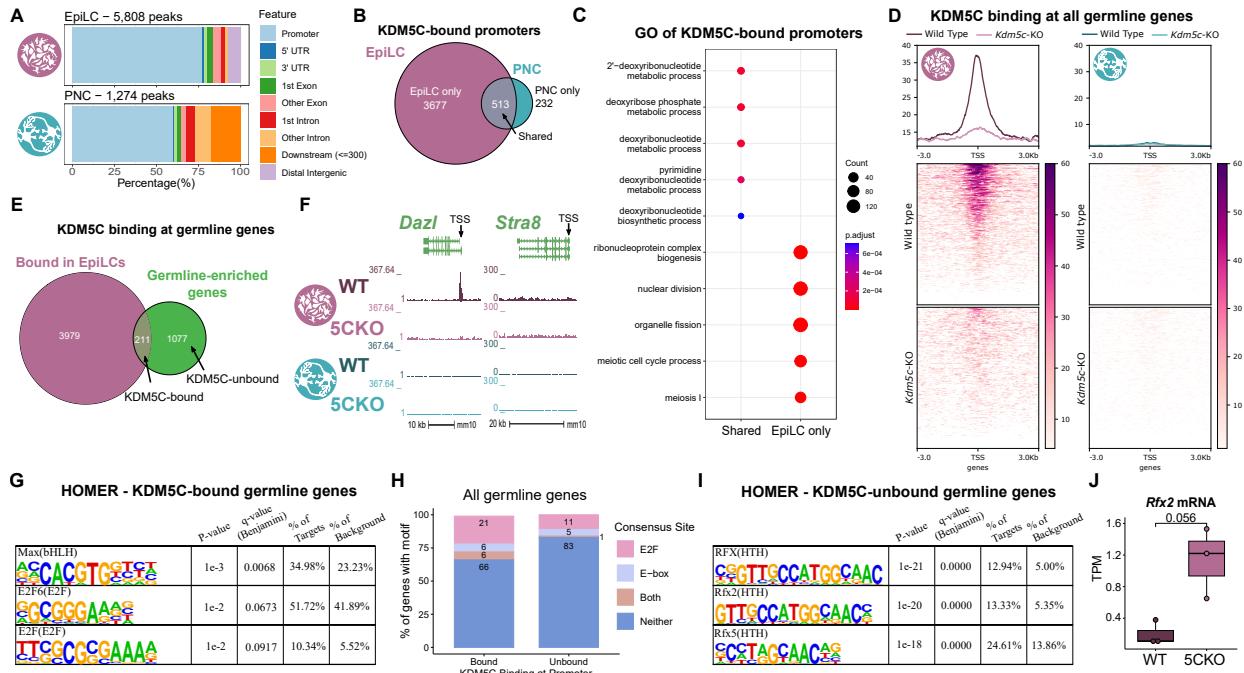


Figure 5: KDM5C binds to a subset of germline gene promoters during early embryogenesis. **A.** ChIPseeker localization of KDM5C peaks at different genomic regions in EpiLCs (top) and hippocampal and cortex primary neuron cultures (PNCs, bottom). **B.** Overlap of genes with KDM5C bound to their promoters ($TSS \pm 500$) 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 overlap of germline-enriched genes (green) with significant KDM5C binding at their promoter in EpiLCs (purple). **F.** Example KDM5C ChIP-seq signal around 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).

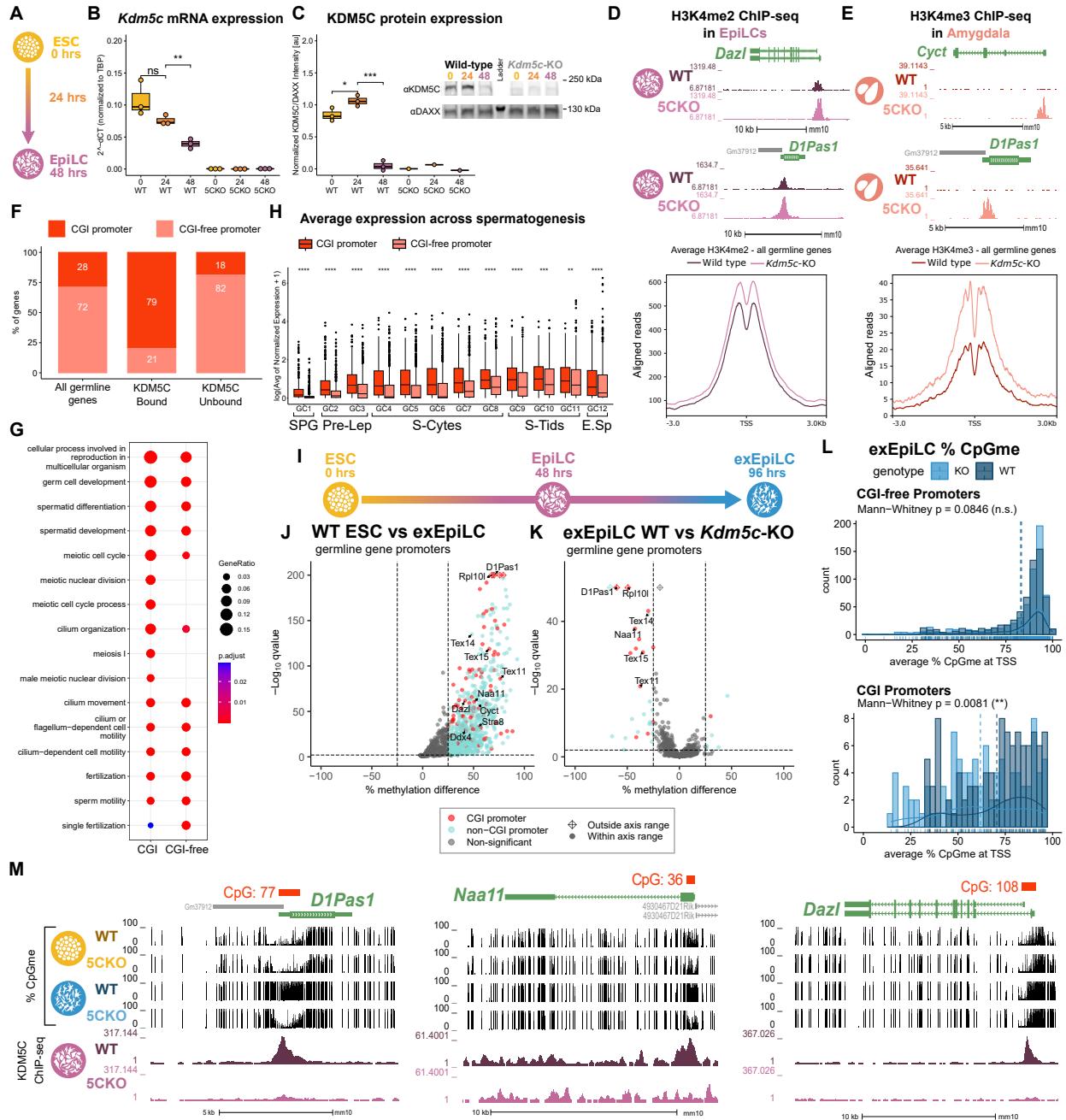


Figure 6: KDM5C promotes long-term silencing of germline genes via DNA methylation at CpG islands. **A.** Diagram of embryonic stem cell (ESC) to epiblast-like cell (EpiLC) differentiation and collection time points. **B.** Real time quantitative PCR (RT-qPCR) of *Kdm5c* mRNA expression in wild-type (WT) and *Kdm5c*-KO (5CKO) ESCs at 0, 24, and 48 hours of differentiation into EpiLCs. Expression calculated in comparison to TBP mRNA expression ($2^{-\Delta\Delta CT}$). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, Welch's t-test. **C.** 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. **D.** Top - Representative UCSC browser view of histone 3 lysine 4 dimethylation (H3K4me2) ChIP-seq signal at two germline genes in wild-type and *Kdm5c*-KO EpiLCs. Bottom - Average H3K4me2 signal at the TSS of all germline-enriched genes in wild-type (dark purple) and *Kdm5c*-KO (light purple) EpiLCs. **E.** Top - Representative UCSC browser view of histone 3 lysine 4 trimethylation (H3K4me3) ChIP-seq signal at two germline genes in the wild-type (WT) and *Kdm5c*-KO (5CKO) adult amygdala. Bottom - Average H3K4me3 signal at the transcription start site (TSS) of all germline-enriched genes in wild-type (dark red) and *Kdm5c*-KO (light red) amygdala. **F.** Percentage of germline genes that harbor CpG islands (CGIs) in their promoters (CGI promoter, red), based on UCSC annotation. Left - percentage of all germline-enriched genes, middle - KDM5C-bound germline genes, and right - KDM5C-unbound germline genes. (Legend continued on next page.)

Figure 6: KDM5C promotes long-term silencing of germline genes via DNA methylation at CpG islands. (Legend continued.) **G.** enrichPlot gene ontology analysis of germline genes with (CGI-promoter) or without (CGI-free) CGIs in their promoter. **H.** Expression of germline genes with (CGI-promoter, red) or without (CGI-free, salmon) CGIs in their promoter across stages of spermatogenesis from Green et al 2018. SPG - spermatogonia, Pre-Lep - preleptotene spermatocytes, S-Cytes - meiotic spermatocytes, S-Tids - post-meiotic haploid round spermatids, E.Sp - elongating spermatids. Wilcoxon test, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$. **I.** Diagram of ESC to extended EpiLC (exEpiLC) differentiation. **J.** Volcano plot of whole genome bisulfite sequencing (WGBS) comparing CpG methylation (CpGme) at germline gene promoters ($TSS \pm 500$) in wild-type (WT) ESCs versus exEpiLCs. Significantly differentially methylated promoters ($q < 0.01$, $|methylated difference| > 25\%$) with CGIs in red, CGI-free promoters in light blue **K.** Volcano plot of WGBS of wild-type (WT) versus *Kdm5c*-KO exEpiLCs for germline gene promoters. Promoters with CGIs in red, CGI-free promoters in light blue. **L.** Histogram of average percent CpGme at the promoter of germline genes with or without CGIs. Wild-type in navy and *Kdm5c*-KO (KO) in light blue. Dashed lines are average methylation for each genotype, p-values for Mann-Whitney U test. **M.** UCSC browser view of germline genes, showing UCSC annotated CGI with number of CpGs, representative CpGme in wild-type (WT) and *Kdm5c*-KO (5CKO) ESCs and exEpiLCs, and KDM5C ChIP-seq signal in EpiLCs.