

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

2 intellectual disability factor KDM5C

3

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

5 Abstract

6 Mutations in numerous chromatin-modifying enzymes cause neurodevelopmental disorders (NDDs). 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 modifications¹⁻⁵. Human genetic studies revealed mutations in

28 chromatin regulators are a major cause of neurodevelopmental disorders (NDDs)⁶ 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 neurons⁷. 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 promoters⁸.
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^{9–11}. Male
40 *Kdm5c* knockout (-KO) mice recapitulate key MRXSCJ phenotypes, including hyperaggression, increased
41 seizure propensity, and learning impairments^{12,13}. RNA sequencing (RNA-seq) of the *Kdm5c*-KO hippocam-
42 pus revealed ectopic expression of some germline genes within the brain¹³. 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 multicellularity¹⁴ that occurs
46 during early embryogenesis in many metazoans¹⁵. 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)¹⁶ and histone 3
49 lysine 9 trimethylation (H3K9me3)^{16,17} in embryonic stem cells (ESCs) and are then decorated with DNA
50 CpG methylation (CpGme) in the post-implantation embryo^{17–20}. 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 testis¹³. 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*²¹ by using a published list of mouse
68 tissue-enriched genes²².

69 We found a large proportion of significantly upregulated genes (DESeq2²³, log2 fold change > 0.5, q <
70 0.1) within the male *Kdm5c*-KO (5CKO) 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^{24,25} (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 maturation²⁶ (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 amygdala²⁷ and its brain overexpression has been implicated in Alzheimer's disease²⁸
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 genes¹³, 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 depletion²⁹,
104 which was accomplished by heterozygous *W* and *Wv* mutations in the enzymatic domain of *c-Kit* (*Kit*^{W/Wv})³⁰.
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 testis³¹. 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*^{W/Wv} mice
115 that included males and females at embryonic day 12, 14, and 16³² and adult male testes²⁹. 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 mice²² 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^{33,34}, when germline genes are silenced in epiblast stem cells that will form the somatic tissues³⁵. 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)^{36,37}. While some germline-enriched genes are
128 also expressed in nESCs and in the 2-cell stage^{38–40}, they are silenced as they differentiate into EpiLCs^{17,18}.
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 EpiLCs⁴¹ (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 programs⁴²⁻⁴⁴. Of note, some germline genes, including *Dazl*, are also expressed in the two-cell embryo^{39,45}.
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^{16,17,19,46,47}. 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 females⁴⁸⁻⁵¹. 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, XY 5CKO), female

164 homozygous knockout (XX *Kdm5c*-KO, XX 5CKO), and female heterozygous knockout (XX *Kdm5c*-HET, XX
165 5CHET) EpiLCs⁴¹.

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

176 The majority of germline genes expressed in both sexes were more highly dysregulated in females
177 compared to males (Figure 4D-F). This increased degree of dysregulation in females, along with the
178 increased total number of germline genes, indicates females are more sensitive to losing KDM5C-mediated
179 germline gene suppression. Female sensitivity could be due to impaired XCI in *Kdm5c* mutants⁴¹, as many
180 spermatogenesis genes lie on the X chromosome^{52,53}. However, female germline DEGs were not biased
181 towards the X chromosome and had a similar overall proportion of X chromosome DEGs compared to
182 males (XY *Kdm5c*-KO - 10.29%, XX *Kdm5c*-HET - 7.43%, XX *Kdm5c*-KO - 10.59%) (Figure 4G). The
183 majority of germline DEGs instead lie on autosomes for both male and female *Kdm5c* mutants (Figure 4G).
184 Thus, while female EpiLCs are more prone to germline gene misexpression with KDM5C loss, it is likely
185 independent of XCI defects.

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

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

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

200 KDM5C binds to the promoters of several germline genes in embryonic stem cells (ESCs) but its binding
201 is absent in neurons¹³. However, the lack of a comprehensive list of germline-enriched genes prohibited
202 genome-wide characterization of KDM5C binding at germline gene promoters. Thus, it is unclear if KDM5C
203 is enriched at germline gene promoters, what types of germline genes KDM5C regulates, and if its binding is
204 maintained at any germline genes in neurons.

205 To address these questions, we analyzed KDM5C chromatin immunoprecipitation followed by DNA
206 sequencing (ChIP-seq) datasets in EpiLCs⁴¹ and primary forebrain neuron cultures (PNCs)¹². EpiLCs had a
207 higher total number of high-confidence KDM5C peaks than PNCs (EpiLCs: 5,808, PNCs: 1,276, MACS2 q <
208 0.1 and fold enrichment > 1). KDM5C was primarily localized to gene promoters in both cell types (EpiLCs:
209 4,190, PNCs: 745 ± 500bp from the TSS), although PNCs showed increased localization to non-promoter
210 regions (Figure 5A).

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

228 Many germline-specific genes are suppressed by the polycomb repressive complex 1.6 (PRC1.6), which
229 contains transcription factor heterodimers E2F6/DP1 and MGA/MAX that respectively bind E2F and E-box
230 motifs⁵⁶. PRC1.6 members may recruit KDM5C to germline gene promoters, given their association with
231 KDM5C in HeLa cells and ESCs^{45,57}. We thus used HOMER⁵⁸ to identify transcription factor motifs enriched
232 at KDM5C-bound or unbound germline gene promoters (TSS ± 500 bp, q-value < 0.1). MAX and E2F6 binding
233 sites were significantly enriched at germline genes bound by KDM5C in EpiLCs (MAX q-value: 0.0068, E2F6
234 q-value: 0.0673, E2F q-value: 0.0917), but not at germline genes unbound by KDM5C (Figure 5G). One third

235 of KDM5C-bound promoters contained the consensus sequence for either E2F6 (E2F, 5'-TCCCGC-3'), MGA
236 (E-box, 5'-CACGTG-3'), or both, but only 17% of KDM5C-unbound genes contained these motifs (Figure
237 5H). KDM5C-unbound germline genes were instead enriched for multiple RFX transcription factor binding
238 sites (RFX q-value < 0.0001, RFX2 q-value < 0.0001, RFX5 q-value < 0.0001) (Figure 5I, Supplementary
239 figure 1D). RFX transcription factors bind X-box motifs⁵⁹ to promote ciliogenesis^{60,61} and among them is
240 RFX2, a central regulator of post-meiotic spermatogenesis^{62,63}. Interestingly, RFX2 mRNA is derepressed
241 in *Kdm5c*-KO EpiLCs (Figure 5J), however it is also not a direct target of KDM5C (Supplementary figure
242 1E). Thus, RFX2 is a candidate transcription factor for driving the ectopic expression of KDM5C-unbound
243 germline genes in *Kdm5c*-KO cells.

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

246 Previous work found two germline gene promoters have a marked reduction in DNA CpG methylation
247 (CpGme) in the *Kdm5c*-KO adult hippocampus¹³. Since histone 3 lysine 4 di- and trimethylation (H3K4me2/3)
248 impede *de novo* CpGme^{64,65}, KDM5C's removal of H3K4me2/3 may be required to suppress germline genes.
249 However, KDM5C's catalytic activity was recently shown to be dispensable for suppressing *Dazl* in ESCs⁴⁵.
250 To reconcile these observations, we hypothesized KDM5C erases H3K4me3 to promote the initial placement
251 of CpGme at germline gene promoters in EpiLCs.

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

260 Germline genes accumulate CpG methylation (CpGme) at CpG islands (CGIs) during the transition
261 from naïve to primed pluripotency^{18,20,66}. We first examined how many of our germline-enriched genes had
262 promoter CGIs (TSS ± 500 bp) using the UCSC genome browser⁶⁷. Notably, out of 1,288 germline-enriched
263 genes, only 356 (27.64%) had promoter CGIs (Figure 6F). CGI-containing germline genes had substantial
264 enrichment of meiotic gene ontologies compared to CGI-free genes, including meiotic nuclear division
265 (GO:0140013, p.adjust = 2.17e-12) and meiosis I (GO:0007127, p.adjust = 3.91e-10) (Figure 6G). Germline
266 genes with promoter CGIs were more highly expressed than CGI-free genes across spermatogenesis stages,
267 with highest expression in meiotic spermatocytes (Figure 6H). Contrastingly, CGI-free genes only displaying
268 substantial expression in post-meiotic round spermatids. Although only a minor portion of germline gene

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

272 To assess how KDM5C loss impacts initial CpGme placement at germline gene promoters, we performed
273 whole genome bisulfite sequencing (WGBS) in male wild-type and *Kdm5c*-KO ESCs and 96-hour extend
274 EpiLCs (exEpiLCs), when germline genes reach peak methylation level¹⁷ (Figure 6I). We first identified
275 which germline gene promoters significantly gained CpGme in wild-type cells during nESC to exEpiLCs
276 differentiation (methylKit⁶⁸, $q < 0.01$, $|methylation\ difference| \geq 25\%$, TSS ± 500 bp). In wild-type cells,
277 the majority of germline genes gained substantial CpGme at their promoter during differentiation (60.08%),
278 regardless if their promoter contained a CGI (Figure 6J).

279 We then identified promoters differentially methylated in wild-type versus *Kdm5c*-KO exEpiLCs (methylKit,
280 $q < 0.01$, $|methylation\ difference| > 25\%$, TSS ± 500 bp). Of the 48,882 promoters assessed, 274 promoters
281 were significantly hypomethylated and 377 promoters were significantly hypermethylated with KDM5C
282 loss (Supplementary figure 2A). Hypomethylated promoters were significantly enriched for germline gene
283 ontologies, such as meiotic nuclear division (GO:0140013, p.adjust = 0.012)(Supplementary figure 2B), with
284 10.22% of hypomethylated promoters belonging to germline genes. Approximately half of germline promoters
285 hypomethylated in *Kdm5c*-KO exEpiLCs are direct targets of KDM5C in EpiLCs (13 out of 28 hypomethylated
286 promoters). Promoters that showed the most robust loss of CpGme (lowest q-values) harbored CGIs (Figure
287 6K). CGI promoters, but not CGI-free promoters, had a significant reduction in CpGme with KDM5C loss
288 as a whole (Figure 6L) (Non-CGI promoters $p = 0.0846$, CGI promoters $p = 0.0081$, Mann-Whitney U test).
289 Significantly hypomethylated promoters included germline genes consistently dysregulated across multiple
290 *Kdm5c*-KO RNA-seq datasets¹³, such as *Naa11* and *D1PAs1* (methylation difference = -60.03%, q-value
291 = 3.26e-153) (Figure 6M). Surprisingly, we found only a modest reduction in CpGme at *Dazl*'s promoter
292 (methylation difference = -6.525%, q-value = 0.0159) (Figure 6N). Altogether, these results demonstrate
293 KDM5C is recruited to germline gene CGIs to promote CpGme at germline gene promoters. This suggests
294 KDM5C's catalytic activity is required for germline gene repression in EpiLCs, however some loci can
295 compensate for KDM5C loss through other silencing mechanisms, even when retaining H3K4me around
296 the TSS.

297 Discussion

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

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

314 Although eggs and sperm employ the same transcriptional programs for shared functions, e.g. PGC
315 formation, meiosis, and genome defense, some germline genes are sex specific. We found *Kdm5c* mutant
316 males and females expressed both sperm and egg-biased genes, indicating the mechanism of derepression
317 is independent of a given germline gene's sex. However, organismal sex did greatly influence the degree of
318 germline gene dysregulation, as female *Kdm5c*-KO EpiLCs had over double the number of germline-enriched
319 DEGs compared to males. The lack of X-linked gene enrichment in females suggests that this greater
320 sensitivity to germline gene misexpress is not due to XCI defects previously reported in *Kdm5c*-KO females⁴¹.
321 Intriguingly, females with heterozygous loss of *Kdm5c* also had over double the number of germline DEGs
322 than hemizygous knockout males, even though their level of KDM5C should be roughly equivalent to that
323 of wild-type males. Increased female sensitivity to germline gene de-repression may be related to females
324 having a higher dose of KDM5C than males, due to its escape from XCI^{48–51}. KDM5C's Y-chromosome
325 homolog, KDM5D, exhibits weaker demethylase activity than KDM5C⁸ and it is currently unknown to regulate
326 germline gene expression. Altogether, these results suggests germline gene silencing mechanisms differ
327 between males and females, which warrants further study to identify the biological implications and underlying
328 mechanisms.

329 We found KDM5C is largely dispensable for promoting normal gene expression during development, yet is
330 critical for suppressing ectopic developmental programs. It is important to note that while we highlighted
331 KDM5C's regulation of germline genes, some germline-enriched genes are also expressed at the 2-cell
332 stage and in naïve ESCs for their role in pluripotency and self-renewal^{40,45,69,70}. Although expressed in
333 naïve ESCs, “self-renewal” germline genes like *Dazl* are silenced during ESC differentiation into epiblast
334 stem cells/EpiLCs^{17,18}. We found that while *Kdm5c*-KO EpiLCs also express *Dazl*, they did not express
335 2-cell-specific genes. These data suggest the 2-cell-like state reported in *Kdm5c*-KO ESCs⁴⁵ likely reflects
336 KDM5C's primary role in germline gene repression. Germline gene misexpression in *Kdm5c*-KO EpiLCs may
337 indicate they are differentiating into primordial germ cell-like cells (PGCLCs), rather than de-differentiating
338 into 2-cell-like cells^{33,34,36}. Yet, *Kdm5c*-KO EpiLCs had normal cellular morphology and properly expressed

339 markers for primed pluripotency, including *Otx2* which blocks EpiLC differentiation into PGCs/PGCLCs⁷¹.
340 In addition to unimpaired EpiLC differentiation, *Kdm5c*-KO gross brain morphology is overall normal¹² and
341 hardly any brain-specific genes were significantly dysregulated. Thus, ectopic germline gene expression
342 occurs along with proper somatic development in *Kdm5c*-KO animals.

343 Our work provides novel insight into the cross-talk between H3K4me and CpGme. In EpiLCs, loss of
344 KDM5C binding at a subset of germline gene promoters, e.g. *D1Pas1* and *Naa11*, strongly impaired CGI
345 methylation, and resulted in their long-lasting de-repression into adulthood. Removal of H3K4me2/3 at CGIs
346 is a plausible mechanism for KDM5C-mediated germline gene suppression^{13,72}, given H3K4me2/3 primarily
347 do not colocalize with CpGme⁷³ and can oppose DNMT3 activity^{64,65}. However, emerging work indicates
348 many histone-modifying enzymes have non-catalytic functions that influence gene expression, sometimes
349 even more potently than their catalytic roles^{74,75}. Indeed, KDM5C's catalytic activity was recently found to be
350 dispensible for repressing *Dazl* in ESCs⁴⁵. In our study, *Dazl*'s promoter still gained CpGme in *Kdm5c*-KO
351 exEpiLCs, even with elevated H3K4me2. *Dazl* and a few other germline gene CGIs use multiple repressive
352 mechanisms to facilitate CpGme^{16,17,46,47}. Together, this suggests alternative silencing mechanisms are
353 sufficient to recruit DNMT3s to some germline CGIs, while others may require KDM5C-mediated H3K4me
354 removal to overcome promoter CGI escape from CpGme^{73,76}. Furthermore, these results indicate the
355 requirement for catalytic activity can change depending upon the locus and developmental stage, even for
356 the same class of genes.

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

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

375 humans²⁸ and overexpression of *Apoc1* in the mouse brain can impair learning and memory⁷⁷. KDM5C may
376 therefore be crucial for neurodevelopment by fine-tuning the expression of tissue-enriched, dosage-sensitive
377 genes like *Apoc1*. Given germline genes have no known functions within the brain, their impact upon
378 neurodevelopment is currently unknown. Ectopic testicular germline transcripts have been observed in a
379 variety of cancers^{78,79}, including brain tumors in *Drosophila* and mammals and shown to promote cancer
380 progression^{80–82}. Intriguingly, mouse and human models for other chromatin-linked neurodevelopmental
381 disorders also display impaired soma-germline demarcation^{7,83–86}, such as DNA methyltransferase 3b
382 (DNMT3B), H3K9me1/2 methyltransferases G9A/GLP, methyl-CpG -binding protein 2 (MECP2)⁸³. Thus,
383 KDM5C is among a growing cohort of chromatin-linked neurodevelopmental disorders with similar erosion of
384 the germline versus soma boundary. Further research is required to determine the impact of these germline
385 genes and the extent to which this phenomenon occurs in humans.

386 Materials and Methods

387 Classifying tissue-enriched and germline-enriched genes

388 Tissue-enriched differentially expressed genes (DEGs) were determined by their classification in a previ-
389 ously published dataset from 17 male and female mouse tissues²². This study defined tissue expression as
390 greater than 1 Fragments Per Kilobase of transcript per Million mapped read (FPKM) and tissue enrichment
391 as at least 4-fold higher expression than any other tissue.

392 We curated a list of germline-enriched genes using an RNA-seq dataset from wild-type and germline-
393 depleted (Kit^{W/W^v}) male and female mouse embryos from embryonic day 12, 14, and 16³², as well as adult
394 male testes²⁹. Germline-enriched genes met the following criteria: 1) their expression is greater than 1
395 FPKM in wild-type germline 2) their expression in any wild-type somatic tissues²² does not exceed 20%
396 of maximum expression in wild-type germline, and 3) their expression in the germ cell-depleted (Kit^{W/W^v})
397 germline, for any sex or time point, does not exceed 20% of maximum expression in wild-type germline. We
398 defined sperm and egg-biased genes as those whose expression in the opposite sex, at any time point, is no
399 greater than 20% of the gene's maximum expression in a given sex. Genes that did not meet this threshold
400 for either sex were classified as 'unbiased'.

401 Cell culture

402 We utilized our previously established cultures of male wild-type and *Kdm5c* knockout (-KO) embryonic
403 stem cells⁴¹. Sex was confirmed by genotyping *Uba1/Uba1y* on the X and Y chromosomes with the following
404 primers: 5'-TGGATGGTGTGCCAATG-3', 5'-CACCTGCACGTTGCCCTT-3'. Deletion of *Kdm5c* was
405 confirmed through the primers 5'-ATGCCCATATTAAGAGTCCTG-3', 5'-TCTGCCTTGATGGGACTGTT-3',
406 and 5'-GGTTCTCAACACTCACATAGTG-3'.

407 Embryonic stem cells (ESCs) and epiblast-like cells were cultured using previously established
408 methods³⁷. Briefly, ESCs were initially cultured in primed ESC (pESC) media consisting of KnockOut
409 DMEM (Gibco#10829–018), fetal bovine serum (Gibco#A5209501), KnockOut serum replacement
410 (Invitrogen#10828–028), Glutamax (Gibco#35050-061), Anti-Anti (Gibco#15240-062), MEM Non-essential
411 amino acids (Gibco#11140-050), and beta-mercaptoethanol (Sigma#M7522). They were then transitioned
412 into ground-state, “naïve” ESCs (nESCs) by culturing for four passages in N2B27 media containing
413 DMEM/F12 (Gibco#11330–032), Neurobasal media (Gibco#21103–049), Gluamax, Anti-Anti, N2 sup-
414 plement (Invitrogen#17502048), and B27 supplement without vitamin A (Invitrogen#12587-010), and
415 beta-mercaptoethanol. Both pESC and nESC media were supplemented with 3 μ M GSK3 inhibitor
416 CHIR99021 (Sigma #SML1046-5MG), 1 μ M MEK inhibitor PD0325901 (Sigma #PZ0162-5MG), and 1,000
417 units/mL leukemia inhibitory factor (LIF, Millipore#ESG1107).

418 nESCs were differentiated into epiblast-like cells (EpiLCs, 48 hours) and extendend EpiLCs (exEpiLCs,
419 96 hours) by culturing in N2B27 media containing DMEM/F12, Neurobasal media, Gluamax, Anti-Anti,
420 N2 supplement, B27 supplement (Invitrogen#17504044), beta-mercaptoethanol, fibroblast growth factor 2
421 (FGF2, R&D Biotechne 233-FB), and activin A (R&D Biotechne 338AC050CF), as previously described³⁷.

422 Real time quantitative PCR (RT-qPCR)

423 nESCs were differentiated into EpiLCs as described above. Cells were lysed with Tri-reagent BD (Sigma
424 #T3809) at 0, 24, and 48 hours of differentiation. RNA was phase separated with 0.1 uL/uL 1-bromo-3-
425 chloropropane (Sigma #B9673) and then precipitated with isopropanol (Sigma #I9516). For each sample,
426 2 ug of RNA was reverse transcribed using the ProtoScript II Reverse transcriptase kit from New England
427 Biolabs (NEB #M0368S) and primed with oligo dT. Expression of *Kdm5c* was detected using the primers
428 5'-CCCATGGAGGCCAGAGAATAAG-3' 5'-CTCAGCGATAAGAGAATTGCTAC-3' and normalized to TBP
429 with the Power SYBR™ Green PCR Master Mix (ThermoFisher #4367659).

430 Western Blot

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

436 RNA sequencing (RNA-seq) data analysis

437 After ensuring read quality via FastQC (v0.11.8), reads were then mapped to the mm10 *Mus musculus*
438 genome (Gencode) using STAR (v2.5.3a), during which we removed duplicates and kept only uniquely

439 mapped reads. Count files were generated by FeatureCounts (Subread v1.5.0), and BAM files were
440 converted to bigwigs using deeptools (v3.1.3) and visualized by the UCSC genome browser. RStudio (v3.6.0)
441 was then used to analyze counts files by DESeq2 (v1.26.0)²³ to identify differentially expressed genes
442 (DEGs) with a q-value (p-adjusted via FDR/Benjamini–Hochberg correction) less than 0.1 and a log2 fold
443 change greater than 0.5. For all DESeq2 analyses, log2 fold changes were calculated with IfcShrink using
444 the ashr package⁸⁷. MA-plots were generated by ggpahr (v0.6.0), and Eulerr diagrams were generated by
445 eulerr (v6.1.1). Boxplots and scatterplots were generated by ggpahr (v0.6.0) and ggplot2 (v3.3.2). The Upset
446 plot was generated via the package UpSetR (v1.4.0)⁸⁸. Gene ontology (GO) analyses were performed by
447 the R package enrichPlot (v1.16.2) using the biological processes setting and compareCluster.

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

449 ChIP-seq reads were aligned to mm10 using Bowtie1 (v1.1.2) allowing up to two mismatches. Only
450 uniquely mapped reads were used for analysis. Peaks were called using MACS2 software (v2.2.9.1) using
451 input BAM files for normalization, with filters for a q-value < 0.1 and a fold enrichment > 1. We removed
452 “black-listed” genomic regions that often give aberrant signals. Common peak sets were obtained in R via
453 DiffBind[estrogen@ross-innesDifferentialOestrogenReceptor2012] (v3.6.5). In the case of KDM5C ChIP-seq,
454 *Kdm5c*-KO peaks were then subtracted from wild-type samples using bedtools (v2.25.0). Peak proximity
455 to genome annotations was determined by ChIPSeeker⁸⁹ (v1.32.1). Gene ontology (GO) analyses were
456 performed by the R package enrichPlot (v1.16.2) using the biological processes setting and compareCluster.
457 Enriched motifs were identified using HOMER⁵⁸. Average binding across the genome was visualized using
458 deeptools (v3.1.3). Bigwigs were visualized using the UCSC genome browser.

459 **CpG island (CGI) analysis**

460 Locations of CpG islands were determined through the mm10 UCSC genome browser CpG island track⁶⁷,
461 which classified CGIs as regions that have greater than 50% GC content, are larger than 200 base pairs,
462 and have a ratio of CG dinucleotides observed over the expected amount greater than 0.6. CGI genomic
463 coordinates were then annotated using ChIPseeker⁸⁹ (v1.32.1) and filtered for ones that lie within promoters
464 of our germline-enriched genes (TSS ± 500).

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

466 Genomic DNA (gDNA) from naïve ESCs and extended EpiLCs was extracted using the Wizard Genomic
467 DNA Purification Kit (Promega A1120), following the instructions for Tissue Culture Cells. gDNA from
468 two wild-types and two *Kdm5c*-KOs of each cell type was sent to Novogene for WGBS using the Illumina
469 NovaSeq X Plus platform and sequenced for 150bp paired-end reads (PE150). All samples had greater
470 than 99% bisulfite conversion rates. Reads were adapter and quality trimmed with Trim Galore (v0.6.10)

471 and aligned to the mm10 genome using Bismark⁹⁰ (v0.22.1). Analysis of differential methylation at germline
472 gene promoters was performed using methylKit⁶⁸ (v1.28.0) with a minimum coverage of 3 paired reads, a
473 percentage cut-off of 25%, and q-value of 0.01. methylKit was also used to calculate average percentage
474 methylation at germline gene promoters. Methylation bedgraph tracks were generated via Bismark and
475 visualized using the UCSC genome browser.

476 **Data availability**

477 **Published datasets**

478 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
479 adult amygdala and hippocampus²¹ (available at GEO: GSE127722) and male wild-type and *Kdm5c*-KO
480 EpiLCs⁴¹ (available at GEO: GSE96797).

482 Previously published ChIP-seq experiments included KDM5C in wild-type and *Kdm5c*-KO EpiLCs⁴¹ (avail-
483 able at GEO: GSE96797) and mouse primary neuron cultures (PNCs) from the cortex and hippocampus¹²
484 (available at GEO: GSE61036). ChIP-seq of histone 3 lysine 4 dimethylation in male wild-type and *Kdm5c*-KO
485 EpiLCs⁴¹ is also available at GEO: GSE96797. ChIP-seq of histone 3 lysine 4 trimethylation in wild-type and
486 *Kdm5c*-KO male amygdala²¹ are available at GEO: GSE127817.

487 **Data analysis**

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

490 **Acknowledgements**

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

502 **Author contributions**

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

506 **References**

- 507 1. Strahl, B.D., and Allis, C.D. (2000). The language of covalent histone modifications. *Nature* **403**,
508 41–45. <https://doi.org/10.1038/47412>.
- 509 2. Jenuwein, T., and Allis, C.D. (2001). Translating the Histone Code. *Science* **293**, 1074–1080.
510 <https://doi.org/10.1126/science.1063127>.
- 511 3. Lewis, E.B. (1978). A gene complex controlling segmentation in *Drosophila*. *Nature* **276**, 565–570.
512 <https://doi.org/10.1038/276565a0>.
- 513 4. Kennison, J.A., and Tamkun, J.W. (1988). Dosage-dependent modifiers of polycomb and antennapedia
514 mutations in *Drosophila*. *Proc. Natl. Acad. Sci. U.S.A.* **85**, 8136–8140. <https://doi.org/10.1073/pnas.8>
5.21.8136.
- 515 5. Kassis, J.A., Kennison, J.A., and Tamkun, J.W. (2017). Polycomb and Trithorax Group Genes in
516 *Drosophila*. *Genetics* **206**, 1699–1725. <https://doi.org/10.1534/genetics.115.185116>.
- 517 6. Gabriele, M., Lopez Tobon, A., D'Agostino, G., and Testa, G. (2018). The chromatin basis of
518 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>.
- 519 7. Schaefer, A., Sampath, S.C., Intrator, A., Min, A., Gertler, T.S., Surmeier, D.J., Tarakhovsky, A.,
520 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>.
- 521 8. Iwase, S., Lan, F., Bayliss, P., De La Torre-Ubieta, L., Huarte, M., Qi, H.H., Whetstine, J.R., Bonni, A.,
Roberts, T.M., and Shi, Y. (2007). The X-Linked Mental Retardation Gene SMCX/JARID1C Defines a
Family of Histone H3 Lysine 4 Demethylases. *Cell* **128**, 1077–1088. <https://doi.org/10.1016/j.cell.200>
7.02.017.
- 523 9. Claes, S., Devriendt, K., Van Goethem, G., Roelen, L., Meireleire, J., Raeymaekers, P., Cassiman,
J.J., and Fryns, J.P. (2000). Novel syndromic form of X-linked complicated spastic paraparesia. *Am J
Med Genet* **94**, 1–4.

- 525 10. Jensen, L.R., Amende, M., Gurok, U., Moser, B., Gimmel, V., Tschach, A., Janecke, A.R., Tariverdian,
526 G., Chelly, J., Fryns, J.P., et al. (2005). Mutations in the JARID1C gene, which is involved in
transcriptional regulation and chromatin remodeling, cause X-linked mental retardation. *Am J Hum
Genet* 76, 227–236. <https://doi.org/10.1086/427563>.
- 527 11. Carmignac, V., Nambot, S., Lehalle, D., Callier, P., Moortgat, S., Benoit, V., Ghoumid, J., Delobel,
B., Smol, T., Thuillier, C., et al. (2020). Further delineation of the female phenotype with KDM5C
disease causing variants: 19 new individuals and review of the literature. *Clin Genet* 98, 43–55.
<https://doi.org/10.1111/cge.13755>.
- 528 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.celr
ep.2015.12.091](https://doi.org/10.1016/j.celr
ep.2015.12.091).
- 529 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,
530 47–59. <https://doi.org/10.1016/j.celrep.2017.09.014>.
- 531 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>.
- 532 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>.
- 533 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
534 stem cells by regulating germ cell-related genes. *eLife* 6. <https://doi.org/10.7554/eLife.21064>.
- 535 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
536 in the early embryo precedes DNA methylation-mediated silencing. *Nat Commun* 12, 7020. <https://doi.org/10.1038/s41467-021-27345-x>.
- 537 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,
538 1093–1100. <https://doi.org/10.1038/ng.708>.
- 539 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
540 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>.

- 544
- 545 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>.
- 546
- 547 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>.
- 548
- 549 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>.
- 550
- 551 23. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* *15*, 550. <https://doi.org/10.1186/s13059-014-0550-8>.
- 552
- 553 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 Chromosome Pairing in Meiosis. *Science* *300*, 1291–1295. <https://doi.org/10.1126/science.1083022>.
- 554
- 555 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>.
- 556
- 557 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 membraneless compartment. *Science* *378*, eabq4835. <https://doi.org/10.1126/science.abq4835>.
- 558
- 559 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. *Cardiovasc Diabetol* *21*, 272. <https://doi.org/10.1186/s12933-022-01703-5>.
- 560
- 561 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>.
- 562
- 563 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 chromosomes for the male germ line. *Nat Genet* *45*, 1083–1087. <https://doi.org/10.1038/ng.2705>.
- 564

- 565 30. Handel, M.A., and Eppig, J.J. (1979). Sertoli Cell Differentiation in the Testes of Mice Genetically
566 Deficient in Germ Cells. *Biology of Reproduction* *20*, 1031–1038. <https://doi.org/10.1095/biolreprod20.5.1031>.
- 567 31. Green, C.D., Ma, Q., Manske, G.L., Shami, A.N., Zheng, X., Marini, S., Moritz, L., Sultan, C.,
568 Gurczynski, S.J., Moore, B.B., et al. (2018). A Comprehensive Roadmap of Murine Spermatogenesis
569 Defined by Single-Cell RNA-Seq. *Dev Cell* *46*, 651–667.e10. <https://doi.org/10.1016/j.devcel.2018.07.025>.
- 570 32. Soh, Y.Q., Junker, J.P., Gill, M.E., Mueller, J.L., van Oudenaarden, A., and Page, D.C. (2015). A
571 Gene Regulatory Program for Meiotic Prophase in the Fetal Ovary. *PLoS Genet* *11*, e1005531.
572 <https://doi.org/10.1371/journal.pgen.1005531>.
- 573 33. Magnúsdóttir, E., and Surani, M.A. (2014). How to make a primordial germ cell. *Development* *141*,
574 245–252. <https://doi.org/10.1242/dev.098269>.
- 575 34. Günesdogan, U., Magnúsdóttir, E., and Surani, M.A. (2014). Primordial germ cell specification: A
576 context-dependent cellular differentiation event [corrected]. *Philos Trans R Soc Lond B Biol Sci* *369*.
<https://doi.org/10.1098/rstb.2013.0543>.
- 577 35. Bardot, E.S., and Hadjantonakis, A.-K. (2020). Mouse gastrulation: Coordination of tissue patterning,
578 specification and diversification of cell fate. *Mechanisms of Development* *163*, 103617. <https://doi.org/10.1016/j.mod.2020.103617>.
- 579 36. Hayashi, K., Ohta, H., Kurimoto, K., Aramaki, S., and Saitou, M. (2011). Reconstitution of the
580 mouse germ cell specification pathway in culture by pluripotent stem cells. *Cell* *146*, 519–532.
581 <https://doi.org/10.1016/j.cell.2011.06.052>.
- 582 37. Samanta, M., and Kalantry, S. (2020). Generating primed pluripotent epiblast stem cells: A methodol-
583 ogy chapter. *Curr Top Dev Biol* *138*, 139–174. <https://doi.org/10.1016/bs.ctdb.2020.01.005>.
- 584 38. Welling, M., Chen, H., Muñoz, J., Musheev, M.U., Kester, L., Junker, J.P., Mischerikow, N., Arbab, M.,
585 Kuijk, E., Silberstein, L., et al. (2015). DAZL regulates Tet1 translation in murine embryonic stem cells.
586 *EMBO Reports* *16*, 791–802. <https://doi.org/10.15252/embr.201540538>.
- 587 39. Macfarlan, T.S., Gifford, W.D., Driscoll, S., Lettieri, K., Rowe, H.M., Bonanomi, D., Firth, A., Singer, O.,
588 Trono, D., and Pfaff, S.L. (2012). Embryonic stem cell potency fluctuates with endogenous retrovirus
589 activity. *Nature* *487*, 57–63. <https://doi.org/10.1038/nature11244>.
- 590 40. Suzuki, A., Hirasaki, M., Hishida, T., Wu, J., Okamura, D., Ueda, A., Nishimoto, M., Nakachi, Y.,
591 Mizuno, Y., Okazaki, Y., et al. (2016). Loss of MAX results in meiotic entry in mouse embryonic and
592 germline stem cells. *Nat Commun* *7*, 11056. <https://doi.org/10.1038/ncomms11056>.

- 587 41. Samanta, M.K., Gayen, S., Harris, C., Maclary, E., Murata-Nakamura, Y., Malcore, R.M., Porter, R.S.,
Garay, P.M., Vallianatos, C.N., Samollow, P.B., et al. (2022). Activation of Xist by an evolutionarily
conserved function of KDM5C demethylase. *Nat Commun* *13*, 2602. <https://doi.org/10.1038/s41467-022-30352-1>.
- 588
- 589 42. Koubova, J., Menke, D.B., Zhou, Q., Capel, B., Griswold, M.D., and Page, D.C. (2006). Retinoic
acid regulates sex-specific timing of meiotic initiation in mice. *Proc. Natl. Acad. Sci. U.S.A.* *103*,
590 2474–2479. <https://doi.org/10.1073/pnas.0510813103>.
- 591 43. Lin, Y., Gill, M.E., Koubova, J., and Page, D.C. (2008). Germ Cell-Intrinsic and -Extrinsic Factors
Govern Meiotic Initiation in Mouse Embryos. *Science* *322*, 1685–1687. <https://doi.org/10.1126/science.1166340>.
- 592
- 593 44. Endo, T., Mikedis, M.M., Nicholls, P.K., Page, D.C., and De Rooij, D.G. (2019). Retinoic Acid and Germ
594 Cell Development in the Ovary and Testis. *Biomolecules* *9*, 775. <https://doi.org/10.3390/biom9120775>.
- 595 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.
596 *Nat Struct Mol Biol*. <https://doi.org/10.1038/s41594-023-01038-z>.
- 597 46. Pohlers, M., Truss, M., Frede, U., Scholz, A., Strehle, M., Kuban, R.-J., Hoffmann, B., Morkel, M.,
Birchmeier, C., and Hagemeier, C. (2005). A Role for E2F6 in the Restriction of Male-Germ-Cell-
598 Specific Gene Expression. *Current Biology* *15*, 1051–1057. <https://doi.org/10.1016/j.cub.2005.04.060>.
- 599 47. Dahlet, T., Truss, M., Frede, U., Al Adhami, H., Bardet, A.F., Dumas, M., Vallet, J., Chicher, J.,
Hammann, P., Kottnik, S., et al. (2021). E2F6 initiates stable epigenetic silencing of germline genes
600 during embryonic development. *Nat Commun* *12*, 3582. <https://doi.org/10.1038/s41467-021-23596-w>.
- 601 48. Agulnik, A.I., Mitchell, M.J., Mattei, M.G., Borsani, G., Avner, P.A., Lerner, J.L., and Bishop, C.E.
(1994). A novel X gene with a widely transcribed Y-linked homologue escapes X-inactivation in mouse
602 and human. *Hum Mol Genet* *3*, 879–884. <https://doi.org/10.1093/hmg/3.6.879>.
- 603 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>.
- 604
- 605 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*
606 *5*, 1355–1360. <https://doi.org/10.1093/hmg/5.9.1355>.
- 607 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>.
- 608
- 609 52. Wang, P.J., McCarrey, J.R., Yang, F., and Page, D.C. (2001). An abundance of X-linked genes
expressed in spermatogonia. *Nat Genet* *27*, 422–426. <https://doi.org/10.1038/86927>.

- 610
- 611 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
inactivation. *Nat Genet* 36, 642–646. <https://doi.org/10.1038/ng1368>.
- 612
- 613 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/18.24.7019>.
- 614
- 615 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
repression of germ cell-related genes in mouse embryonic stem cells. *PLoS ONE* 13, e0205969.
<https://doi.org/10.1371/journal.pone.0205969>.
- 616
- 617 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*
14, e1007193. <https://doi.org/10.1371/journal.pgen.1007193>.
- 618
- 619 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.
Nature 447, 601–605. <https://doi.org/10.1038/nature05823>.
- 620
- 621 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.
<https://doi.org/10.1016/j.molcel.2010.05.004>.
- 622
- 623 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,
916–921. <https://doi.org/10.1038/35002634>.
- 624
- 625 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).
- 626
- 627 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
Ciliogenesis. *Sci. Signal.* 2. <https://doi.org/10.1126/scisignal.2000602>.
- 628
- 629 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.
PLoS Genet 11, e1005368. <https://doi.org/10.1371/journal.pgen.1005368>.
- 630

- 631 63. Wu, Y., Hu, X., Li, Z., Wang, M., Li, S., Wang, X., Lin, X., Liao, S., Zhang, Z., Feng, X., et al.
632 (2016). Transcription Factor RFX2 Is a Key Regulator of Mouse Spermiogenesis. *Sci Rep* 6, 20435.
<https://doi.org/10.1038/srep20435>.
- 633 64. Otani, J., Nankumo, T., Arita, K., Inamoto, S., Ariyoshi, M., and Shirakawa, M. (2009). Structural basis
634 for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX–DNMT3–DNMT3L
domain. *EMBO Reports* 10, 1235–1241. <https://doi.org/10.1038/embor.2009.218>.
- 635 65. Guo, X., Wang, L., Li, J., Ding, Z., Xiao, J., Yin, X., He, S., Shi, P., Dong, L., Li, G., et al. (2015).
636 Structural insight into autoinhibition and histone H3-induced activation of DNMT3A. *Nature* 517,
640–644. <https://doi.org/10.1038/nature13899>.
- 637 66. Meissner, A., Mikkelsen, T.S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein,
638 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>.
- 639 67. Nassar, L.R., Barber, G.P., Benet-Pagès, A., Casper, J., Clawson, H., Diekhans, M., Fischer, C.,
640 Gonzalez, J.N., Hinrichs, A.S., Lee, B.T., et al. (2023). The UCSC Genome Browser database: 2023
update. *Nucleic Acids Research* 51, D1188–D1195. <https://doi.org/10.1093/nar/gkac1072>.
- 641 68. Akalin, A., Kormaksson, M., Li, S., Garrett-Bakelman, F.E., Figueiroa, M.E., Melnick, A., and Mason,
642 C.E. (2012). methylKit: A comprehensive R package for the analysis of genome-wide DNA methylation
profiles. *Genome Biol* 13, R87. <https://doi.org/10.1186/gb-2012-13-10-r87>.
- 643 69. Torres-Padilla, M.-E. (2020). On transposons and totipotency. *Philos Trans R Soc Lond B Biol Sci*
644 375, 20190339. <https://doi.org/10.1098/rstb.2019.0339>.
- 645 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*
646 29, 400–418.e13. <https://doi.org/10.1016/j.stem.2022.01.010>.
- 647 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. [018-0581-5](https://doi.org/10.1038/s41586-
648 018-0581-5).
- 649 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
of germline genes in mouse embryonic stem cells. *Nucleic Acids Research* 51, 3130–3149. <https://doi.org/10.1093/nar/gkad071>.
- 650 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
652 genome. *Nat Genet* 39, 457–466. <https://doi.org/10.1038/ng1990>.

- 653 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>.
- 654 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.
<https://doi.org/10.1038/s41588-020-00736-4>.
- 655 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,
6693–6706. <https://doi.org/10.1093/nar/gkw258>.
- 656 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>.
- 657 78. Nielsen, A.Y., and Gjerstorff, M.F. (2016). Ectopic Expression of Testis Germ Cell Proteins in Cancer
and Its Potential Role in Genomic Instability. *Int J Mol Sci* 17. <https://doi.org/10.3390/ijms17060890>.
- 658 79. Adebayo Babatunde, K., Najafi, A., Salehipour, P., Modarressi, M.H., and Mobasher, M.B. (2017).
Cancer/Testis genes in relation to sperm biology and function. *Iranian Journal of Basic Medical
Sciences* 20. <https://doi.org/10.22038/ijbms.2017.9259>.
- 659 80. Janic, A., Mendizabal, L., Llamazares, S., Rossell, D., and Gonzalez, C. (2010). Ectopic expression
of germline genes drives malignant brain tumor growth in *Drosophila*. *Science* 330, 1824–1827.
<https://doi.org/10.1126/science.1195481>.
- 660 81. Ghafouri-Fard, S., and Modarressi, M.-H. (2012). Expression of Cancer–Testis Genes in Brain Tumors:
Implications for Cancer Immunotherapy. *Immunotherapy* 4, 59–75. <https://doi.org/10.2217/imt.11.145>.
- 661 82. Nin, D.S., and Deng, L.-W. (2023). Biology of Cancer-Testis Antigens and Their Therapeutic Implica-
tions in Cancer. *Cells* 12, 926. <https://doi.org/10.3390/cells12060926>.
- 662 83. Bonefas, K.M., and Iwase, S. (2021). Soma-to-germline transformation in chromatin-linked neurode-
velopmental disorders? *FEBS J.* <https://doi.org/10.1111/febs.16196>.
- 663 84. Velasco, G., Walton, E.L., Sterlin, D., Héduouin, S., Nitta, H., Ito, Y., Fouyssac, F., Mégarbané, A.,
Sasaki, H., Picard, C., et al. (2014). Germline genes hypomethylation and expression define a
molecular signature in peripheral blood of ICF patients: Implications for diagnosis and etiology.
Orphanet J Rare Dis 9, 56. <https://doi.org/10.1186/1750-1172-9-56>.
- 664 85. Walton, E.L., Francastel, C., and Velasco, G. (2014). Dnmt3b Prefers Germ Line Genes and Cen-
tromeric Regions: Lessons from the ICF Syndrome and Cancer and Implications for Diseases. *Biology
(Basel)* 3, 578–605. <https://doi.org/10.3390/biology3030578>.

- 677 86. Samaco, R.C., Mandel-Brehm, C., McGraw, C.M., Shaw, C.A., McGill, B.E., and Zoghbi, H.Y. (2012).
678 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>.
- 679 87. Stephens, M. (2016). False discovery rates: A new deal. *Biostat*, kxw041. <https://doi.org/10.1093/biostatistics/kxw041>.
- 680 88. Conway, J.R., Lex, A., and Gehlenborg, N. (2017). UpSetR: An R package for the visualization of
681 intersecting sets and their properties. *Bioinformatics* 33, 2938–2940. <https://doi.org/10.1093/bioinformatics/btx364>.
- 682 89. Yu, G., Wang, L.G., and He, Q.Y. (2015). ChIPseeker: An R/Bioconductor package for ChIP peak
annotation, comparison and visualization. *Bioinformatics* 31, 2382–2383. <https://doi.org/10.1093/bioinformatics/btv145>.
- 683 90. Krueger, F., and Andrews, S.R. (2011). Bismark: A flexible aligner and methylation caller for Bisulfite-
684 Seq applications. *Bioinformatics* 27, 1571–1572. <https://doi.org/10.1093/bioinformatics/btr167>.

687 **Figures and Tables**

- 688 • Supplementary table 1: list of all germline genes.
- 689 – Columns to include:
- 690 * KDM5C bound vs not
- 691 * Log2fc in EpiLC, brain (separate columns?)
- 692 – CGI vs non

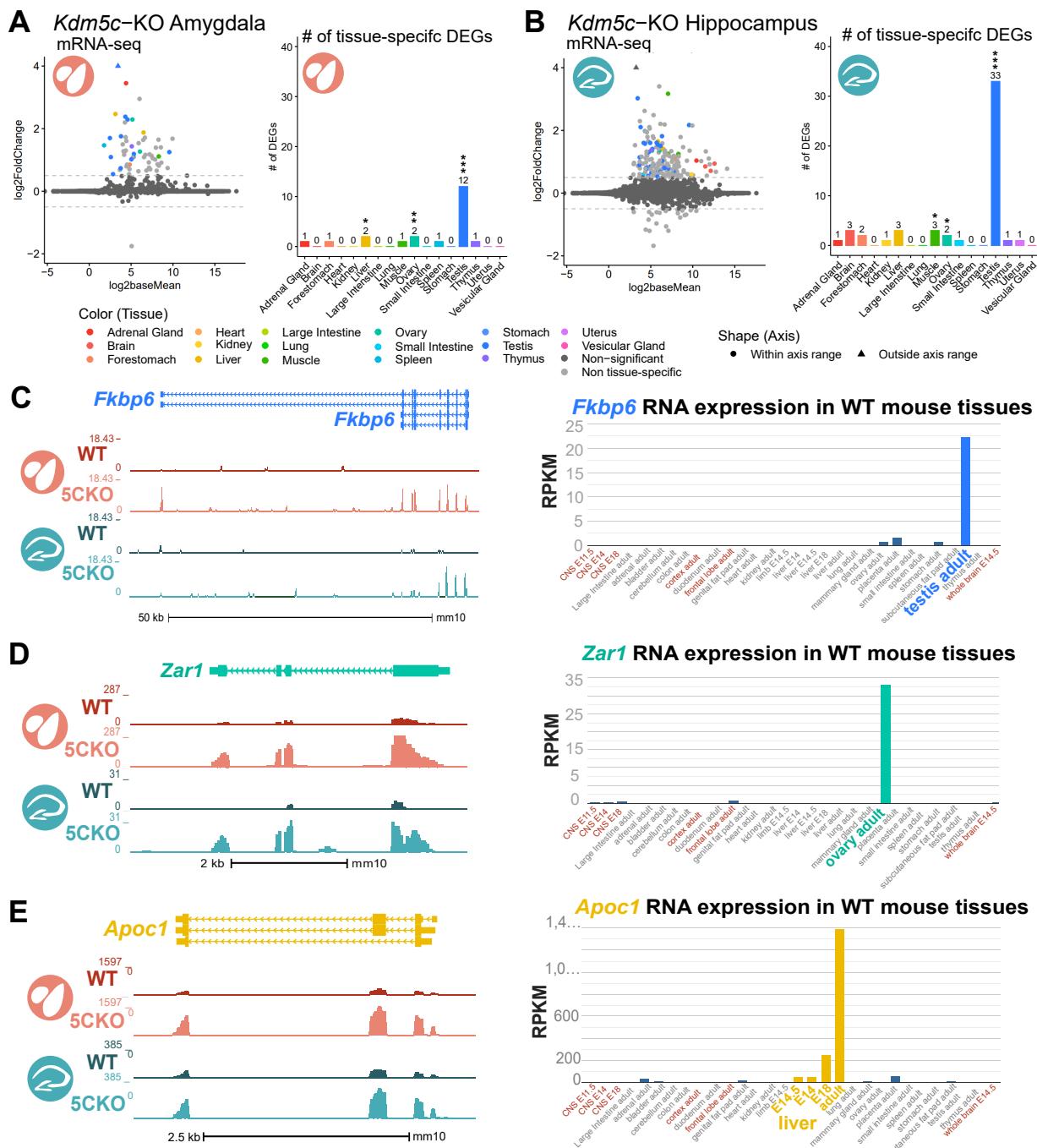


Figure 1: Tissue-enriched genes are misexpressed in the *Kdm5c*-KO brain. **A-B.** Expression of tissue-enriched genes (Li et al 2017) in the male *Kdm5c*-KO amygdala (A) and hippocampus (B). Left - MA plot of mRNA-sequencing. Right - Number of tissue-enriched differentially expressed genes (DEGs). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, Fisher's exact test. **C.** Left - 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.

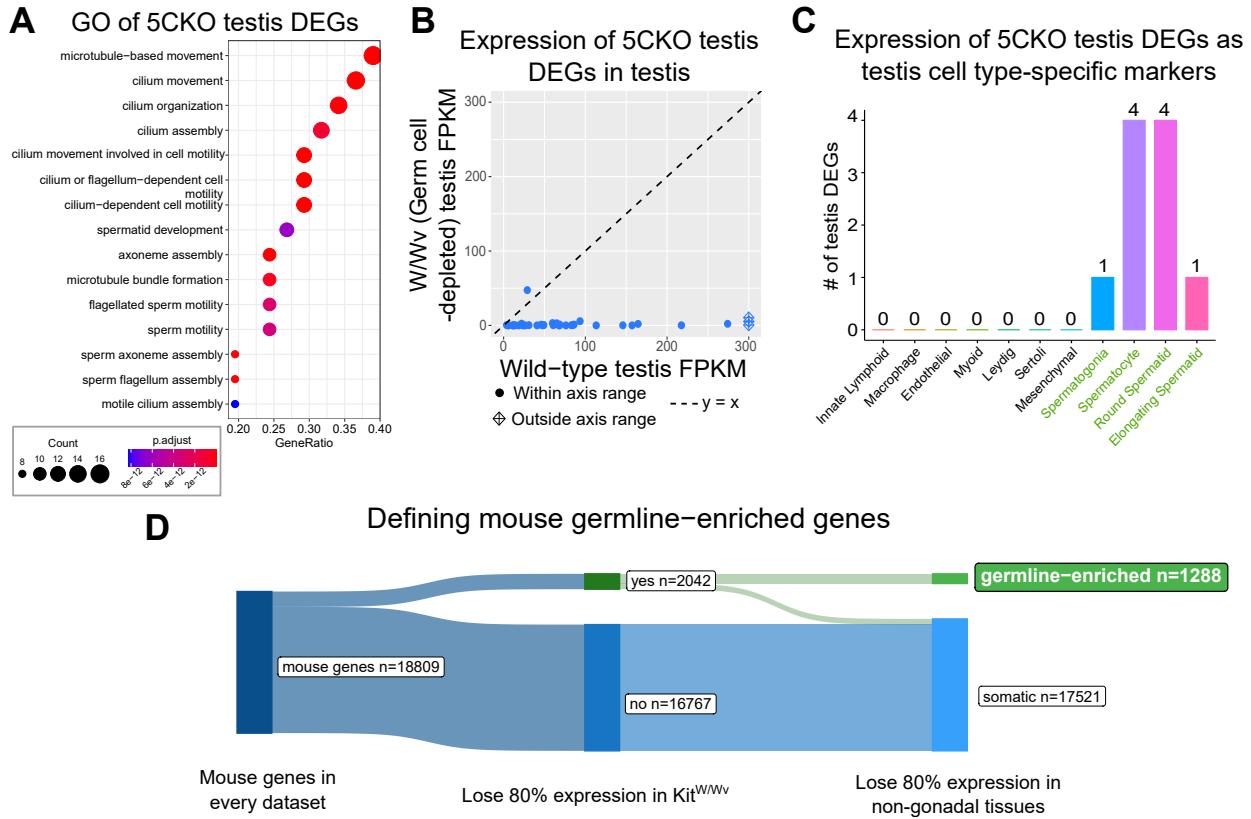


Figure 2: Aberrant transcription of germline genes in the *Kdm5c*-KO in the brain. **A.** enrichPlot gene ontology (GO) of *Kdm5c*-KO amygdala and hippocampus testis-enriched DEGs **B.** Expression of testis DEGs in wild-type (WT) testis versus germ cell-depleted (W/Wv) testis (Mueller et al 2013). Expression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

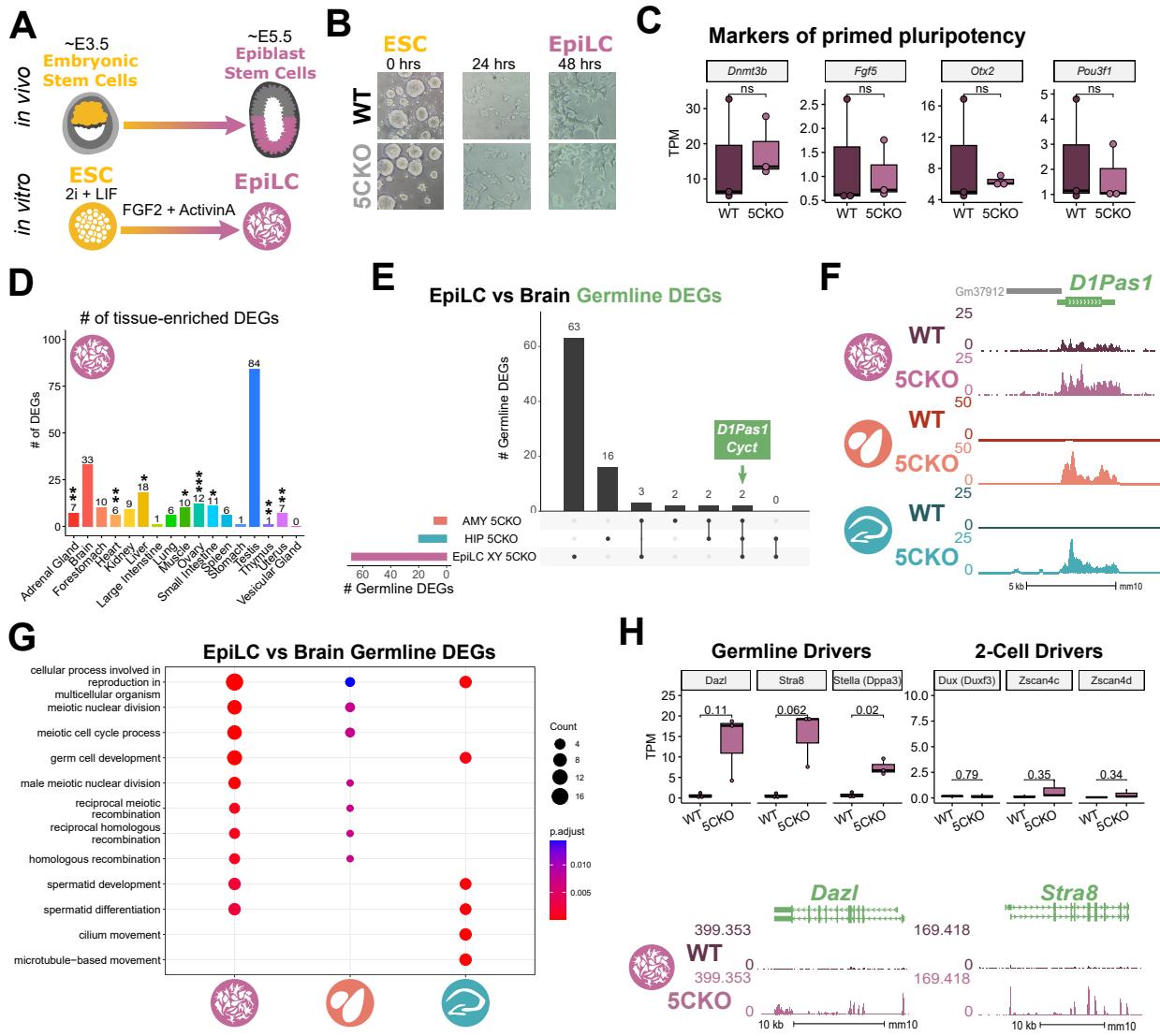


Figure 3: Kdm5c-KO epiblast-like cells express key drivers of germline identity **A.** Top - Diagram of *in vivo* differentiation of embryonic stem cells (ESCs) of the inner cell mass into epiblast stem cells. Bottom - *in vitro* differentiation of ESCs into epiblast-like cells (EpiLCs). **B.** Representative 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. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, 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, $n = 3$).

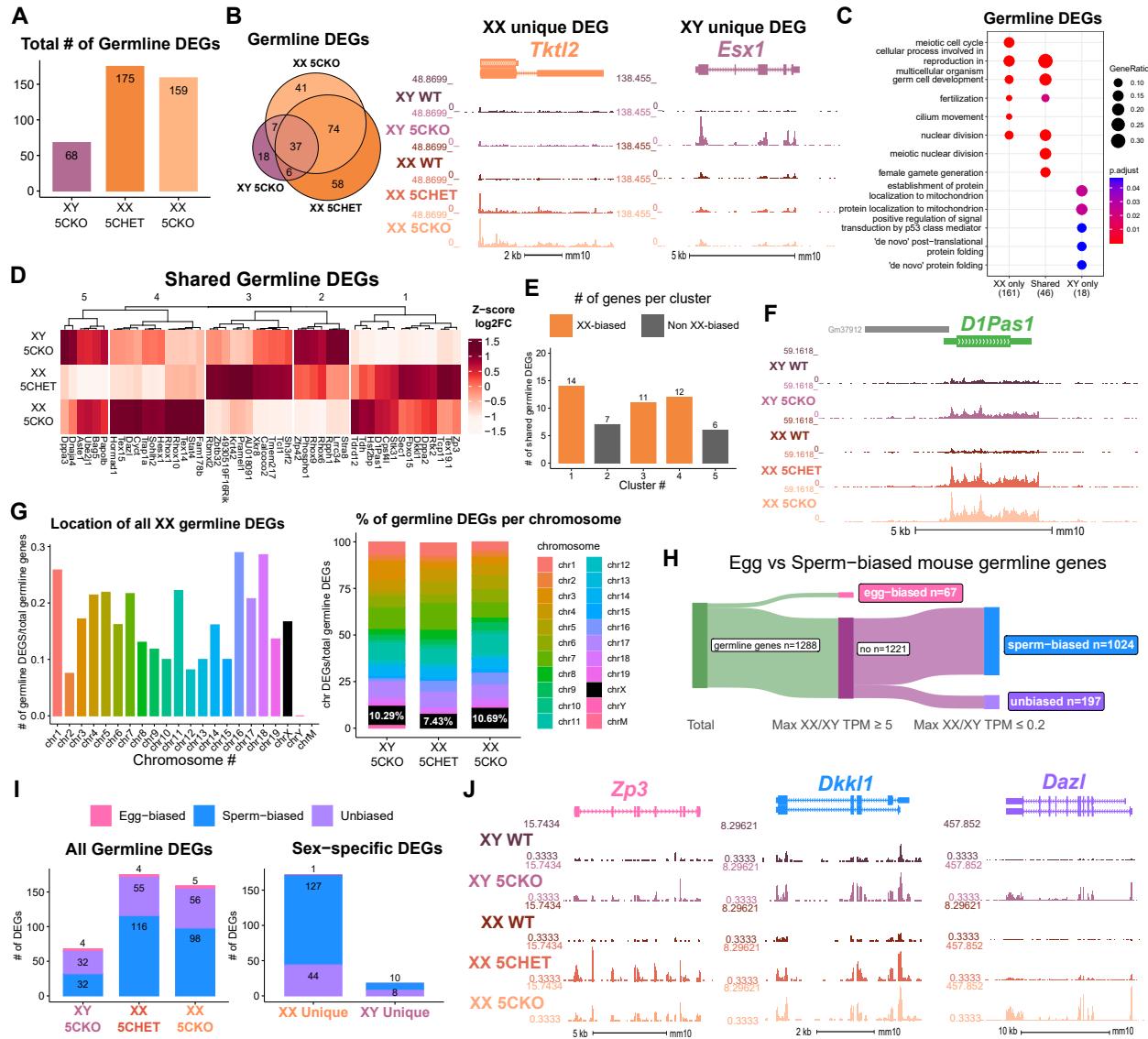


Figure 4: Chromosomal sex influences *Kdm5c*-KO germline gene misexpression. **A.** Total number of germline-enriched RNA-seq DEGs for male hemizygous *Kdm5c* knockout EpilCs (XY 5CKO, purple), female heterozygous *Kdm5c* knockout (XX 5CHET, orange), female homozygous *Kdm5c* knockout (XX 5CKO, light 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. **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.** Example bigwigs of egg-biased (*Zp3*), sperm-biased (*Dkk1*), and unbiased (*Dazl*) germline genes dysregulated in both male and female *Kdm5c* mutants.

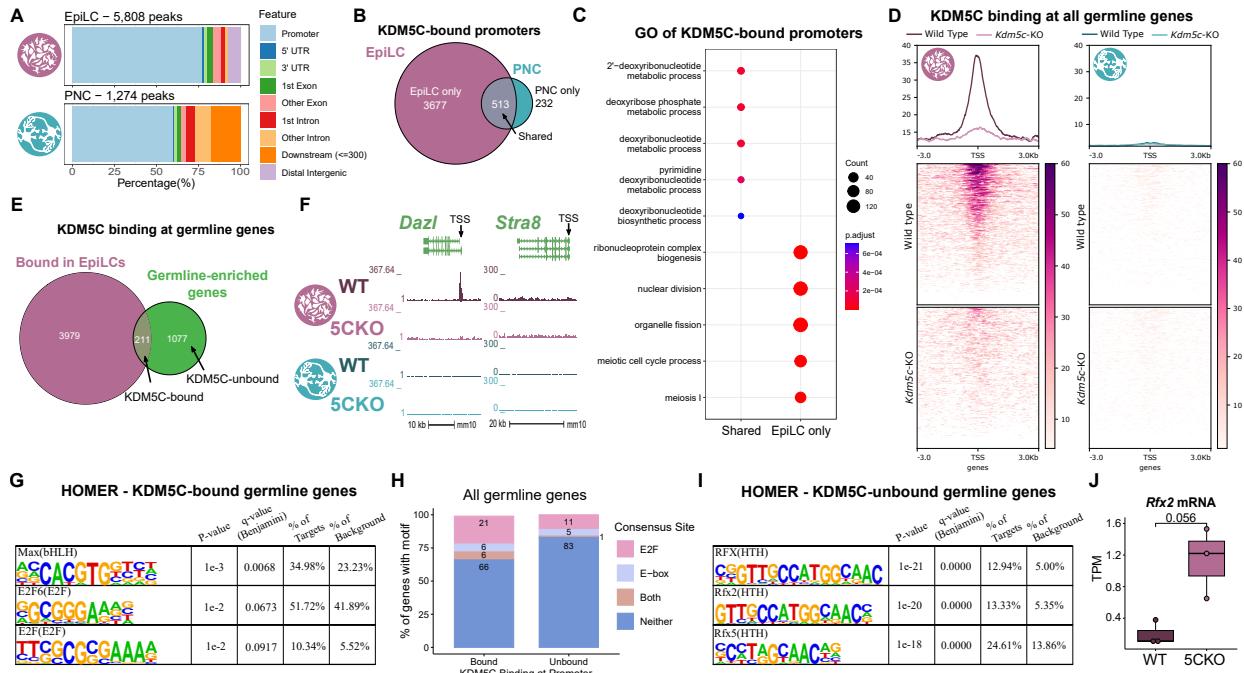


Figure 5: KDM5C binds to a subset of germline gene promoters during early embryogenesis. **A.** ChIPseeker localization of KDM5C peaks at different genomic regions in EpiLCs (top) and hippocampal and cortex primary neuron cultures (PNCs, bottom). **B.** Overlap of genes with KDM5C bound to their promoters ($TSS \pm 500$) 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 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).

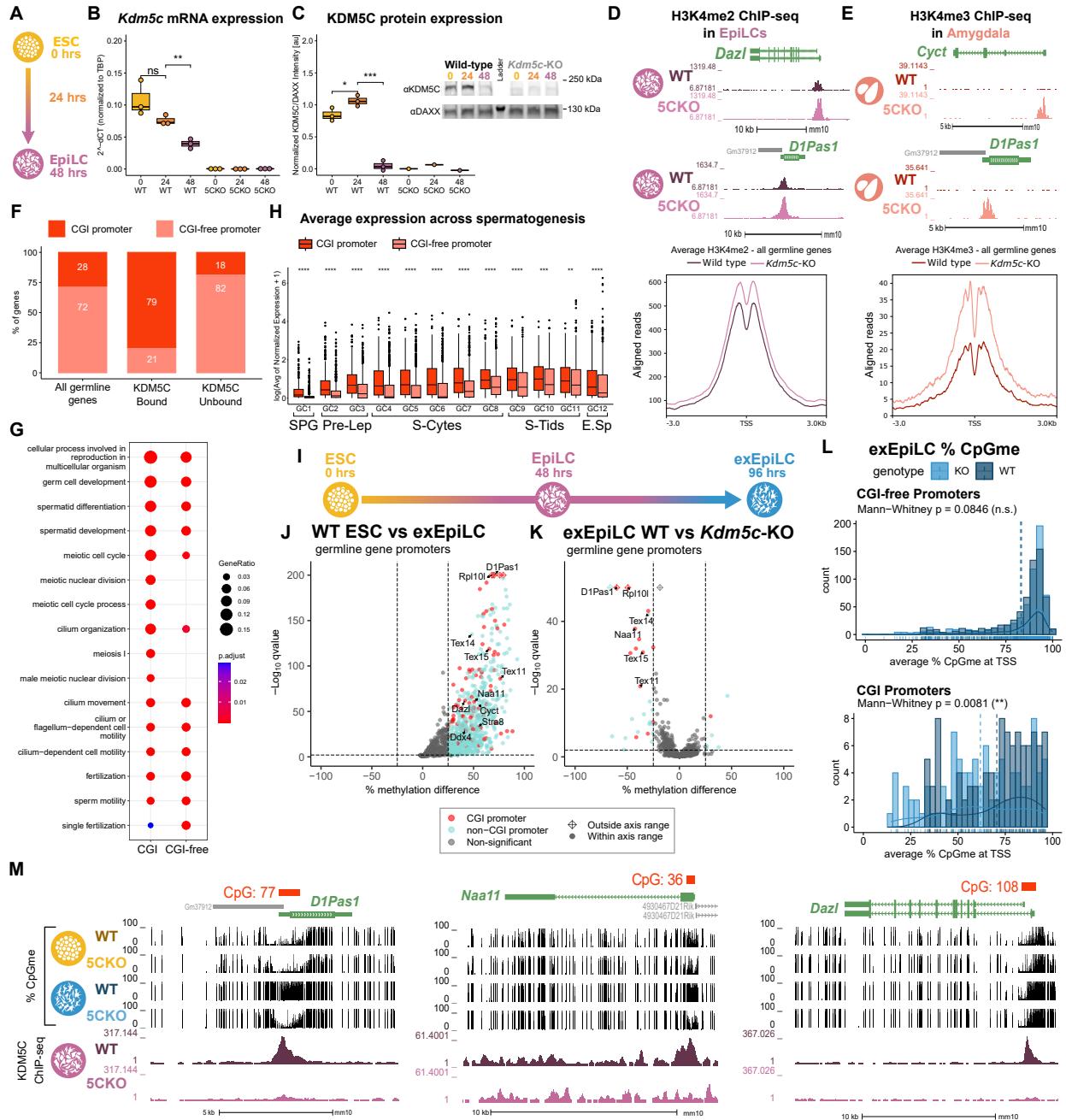


Figure 6: KDM5C promotes long-term silencing of germline genes via DNA methylation at CpG islands. **A.** Diagram of embryonic stem cell (ESC) to epiblast-like cell (EpiLC) differentiation and collection time points. **B.** Real time quantitative PCR (RT-qPCR) of *Kdm5c* mRNA expression in wild-type (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 ($q < 0.01$, $|methyl\text{ation difference}| > 25\%$) 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 around the transcription start site (TSS) of germline genes, showing UCSC annotated CGI, representative CpGme in wild-type (WT) and *Kdm5c*-KO (5CKO) ESCs and exEpiLCs.