

KDM5C is a sex-biased brake against germline gene expression programs in somatic lineages

3

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

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

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

⁹ Correspondence should be addressed to K. Bonefas and S. Iwase (siwase@umich.edu)

10 Abstract

11 The division of labor among cellular lineages is a pivotal step in the evolution of multicellularity. In
12 mammals, the soma-germline boundary is formed during early embryogenesis, when genes that drive
13 germline identity are repressed in somatic lineages through DNA and histone modifications at promoter CpG
14 islands (CGIs). Somatic misexpression of germline genes is a signature of cancer and observed in select
15 neurodevelopmental disorders. However, it is currently unclear if all germline genes use the same repressive
16 mechanisms and if factors like development and sex influence their dysregulation. Here, we examine how
17 cellular context influences the formation of somatic tissue identity in mice lacking lysine demethylase 5c
18 (KDM5C), an X chromosome eraser of histone 3 lysine 4 di and tri-methylation (H3K4me2/3). We found male
19 *Kdm5c* knockout (-KO) mice aberrantly express many tissue-specific genes within the brain, the majority of
20 which are unique to the germline. By developing a comprehensive list of mouse germline-enriched genes,
21 we observed *Kdm5c*-KO cells aberrantly express key drivers of germline fate during early embryogenesis
22 but late-stage spermatogenesis genes within the mature brain. KDM5C binds CGIs within germline gene
23 promoters to facilitate DNA CpG methylation as embryonic stem cells differentiate into epiblast-like cells
24 (EpiLCs). However, the majority of late-stage spermatogenesis genes expressed within the *Kdm5c*-KO brain
25 did not harbor promoter CGIs. These CGI-free germline genes were not bound by KDM5C and instead
26 expressed through ectopic activation by RFX transcription factors. Furthermore, germline gene repression
27 is sexually dimorphic, as female EpiLCs require a higher dose of KDM5C to maintain germline silencing.
28 Altogether, these data revealed distinct regulatory classes of germline genes and sex-biased silencing
29 mechanisms in somatic cells.

30 Introduction

31 The separation of germline and somatic cellular identity is a pivotal step in the evolution of multicellularity
32 and sexual reproduction^{1–4}. In mammals, chromatin regulators decommission germline genes in somatic
33 lineages when the early embryo transitions from naïve to primed pluripotency. Germline gene promoters
34 initially gain repressive histone H2A lysine 119 monoubiquitination (H2AK119ub1)⁵ and histone H3 lysine
35 9 trimethylation (H3K9me3)^{5,6} in embryonic stem cells (ESCs) and are then decorated with DNA CpG
36 methylation (CpGme) at their CpG islands (CGIs) in post-implantation epiblast cells^{6–9}. While the silencing
37 mechanisms for genes that establish germline identity are well characterized, it is unclear if other types
38 of germline genes employ the same silencing mechanisms, such as those involved in the later stages
39 of oogenesis and spermatogenesis. Furthermore, because many studies have focused on the silencing
40 of key marker genes during early male embryonic development, much is unknown about how cellular
41 context (i.e. sex and tissue environment) influences the manifestation of germline gene misexpression.
42 Intriguingly, impaired soma-germline demarcation is a signature of aggressive cancers and observed in
43 select neurodevelopmental disorders (NDDs)^{10–13}. Thus, elucidating how cell context contributes to germline
44 gene dysregulation will reveal novel mechanisms governing these pathologies.

45 Here, we employed genome-wide analyses to explore the loss of tissue identity in mice lacking the
46 chromatin regulator lysine demethylase 5C (KDM5C, also known as SMCX or JARID1C). KDM5C lies on the
47 X chromosome and erases histone 3 lysine 4 di- and trimethylation (H3K4me2/3), a permissive chromatin
48 modification enriched at gene promoters¹⁴. Somatic loss of KDM5C promotes tumorigenicity in a variety of
49 cancer types^{15–17}, while pathogenic germline mutations cause the NDD Intellectual Developmental Disorder,
50 X-linked, Syndromic, Claes-Jensen Type (MRXSCJ, OMIM: 300534). MRXSCJ is more common and
51 severe in males and its neurological phenotypes include intellectual disability, seizures, aberrant aggression,
52 and autistic behaviors^{18–20}. Male *Kdm5c* knockout (-KO) mice recapitulate key MRXSCJ phenotypes,
53 including hyperaggression, increased seizure propensity, social deficits, and learning impairments^{21–23}. RNA
54 sequencing (RNA-seq) of the *Kdm5c*-KO hippocampus revealed ectopic expression of some testis germline
55 genes within the brain²². However, it is unclear if other tissue-specific genes are aberrantly transcribed with
56 KDM5C loss, at what point in development germline gene misexpression begins, what mechanisms underlie

57 their dysregulation, and how KDM5C interacts with other known germline silencing mechanisms.

58 To illuminate KDM5C's role in tissue identity, we characterized the aberrant expression of tissue-enriched
59 genes within the *Kdm5c*-KO brain and epiblast-like stem cells (EpiLCs), an *in vitro* model of the post-
60 implantation embryo. We curated a list of mouse germline-enriched genes, enabling genome-wide analysis
61 of germline gene silencing mechanisms for the first time. We identified two classes of germline genes based
62 on their promoter CpG island content, which are dysregulated with KDM5C loss by distinct mechanisms and
63 in a sex-biased manner.

64 Results

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

66 Previous RNA sequencing (RNA-seq) of the adult male *Kdm5c*-KO hippocampus revealed ectopic
67 expression of some germline genes unique to the testis²². It is currently unknown if the testis is the only
68 tissue type misexpressed in the *Kdm5c*-KO brain. We first systematically tested whether other tissue-specific
69 genes are misexpressed in the male brain with constitutive knockout of *Kdm5c* (*Kdm5c*^y, 5CKO in figures)²⁴
70 by using a published list of mouse tissue-enriched genes²⁵.

71 We found a large proportion of significantly upregulated genes (DESeq2²⁶, log2 fold change > 0.5, q <
72 0.1) within the male *Kdm5c*-KO amygdala and hippocampus are non-brain, tissue-specific genes (Amygdala:
73 0/0 up DEGs, NaN% ; Hippocampus: 0/0 up DEGs, NaN%) (Figure 1A-B, Supplementary Table 1). For both
74 the amygdala and hippocampus, the majority of tissue-enriched differentially expressed genes (DEGs) were
75 testis genes (Figure 1A-B). Even though the testis has the largest total number of tissue-enriched genes
76 (2,496 genes) compared to any other tissue, testis-enriched DEGs were significantly enriched in both brain
77 regions (Amygdala p = 1.83e-05, Odds Ratio = 5.13; Hippocampus p = 4.26e-11, Odds Ratio = 4.45, Fisher's
78 Exact Test). An example of a testis-enriched gene misexpressed in the *Kdm5c*-KO brain is *FK506 binding*
79 *protein 6 (Fkbp6)*, a known regulator of PIWI-interacting RNAs (piRNAs) and meiosis^{27,28} (Figure 1C).

80 Interestingly, we also observed significant enrichment of ovary-enriched genes in both the amygdala
81 and hippocampus (Amygdala p = 0.00574, Odds Ratio = 18.7; Hippocampus p = 0.048, Odds Ratio = 5.88,

82 Fisher's Exact Test) (Figure 1A-B). Ovary-enriched DEGs included *Zygotic arrest 1 (Zar1)*, which sequesters
83 mRNAs in oocytes for meiotic maturation²⁹ (Figure 1D). Given that the *Kdm5c*-KO mice we analyzed are
84 male, these data demonstrate that the ectopic expression of gonad-enriched genes is independent of
85 organismal sex.

86 Although not consistent across brain regions, we also found significant enrichment of genes biased
87 towards two non-gonadal tissues - the liver (Amygdala p = 0.04, Odds Ratio = 6.58, Fisher's Exact Test)
88 and muscle (Hippocampus p = 0.01, Odds Ratio = 6.95, Fisher's Exact Test) (Figure 1A-B). These include
89 *Apolipoprotein C-I (Apoc1)*, a lipoprotein metabolism and transport gene³⁰ (Figure 1E, see Discussion).

90 Our analysis of oligo(dT)-primed libraries²⁴ indicates aberrantly expressed mRNAs are polyadenylated
91 and spliced into mature transcripts in the *Kdm5c*-KO brain (Figure 1C-E). Of note, we observed little to no
92 dysregulation of brain-enriched genes (Amygdala p = 1, Odds Ratio = 1.22; Hippocampus p = 0.74, Odds
93 Ratio = 1.22, Fisher's Exact Test), despite the fact these are brain samples and the brain has the second
94 highest total number of tissue-enriched genes (708 genes). Altogether, these results suggest the aberrant
95 expression of tissue-enriched genes within the 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 genes²² (Figure 1), however the testis also contains
98 somatic cells that support hormone production and germline functions. To determine if *Kdm5c*-KO results
99 in ectopic expression of testicular somatic genes, we first evaluated the known functions of testicular
100 DEGs through gene ontology. We found *Kdm5c*-KO testis-enriched DEGs had high enrichment of germline-
101 relevant ontologies, including spermatid development (GO: 0007286, p.adjust = 6.2e-12) and sperm axoneme
102 assembly (GO: 0007288, p.adjust = 2.45e-14) (Figure 2A, Supplementary Table 1).

103 We then evaluated *Kdm5c*-KO testicular DEG expression in wild-type testes versus testes with germ cell
104 depletion³¹, which was accomplished by heterozygous *W* and *Wv* mutations in the enzymatic domain of *c-Kit*
105 (*Kit*^{W/Wv})³². Almost all *Kdm5c*-KO testis-enriched DEGs lost expression with germ cell depletion (Figure 2B).
106 We then assessed testis-enriched DEG expression in a published single cell RNA-seq dataset that identified
107 cell type-specific markers within the testis³³. Some *Kdm5c*-KO testis-enriched DEGs were classified as

108 specific markers for different germ cell developmental stages (e.g. spermatogonia, spermatocytes, round
109 spermatids, and elongating spermatids), yet none marked somatic cells (Figure 2C). Together, these data
110 demonstrate that the *Kdm5c*-KO brain aberrantly expresses germline genes but not somatic testicular genes,
111 reflecting 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 rather
113 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 16³⁴ and adult male testes³¹. 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 mice²⁵ 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 2).

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^{35,36}, when germline genes are silenced in epiblast stem cells that will form the somatic tissues³⁷. 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)^{38,39}. While some germline-enriched genes are
128 also expressed in nESCs and in the 2-cell stage^{40–42}, they are silenced as they differentiate into EpiLCs^{6,7}.
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 morphology 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 EpiLCs⁴³ (DESeq2,

134 log₂ fold change > 0.5, q < 0.1, Supplementary Table 3). Similar to the *Kdm5c*-KO brain, we observed
135 general dysregulation of tissue-enriched genes, with the largest number of genes belonging to the brain and
136 testis, although they were not significantly enriched (Figure 3D). Using our list of mouse germline-enriched
137 genes assembled above, we identified 68 germline genes 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). EpiLC germline DEGs had particularly high enrichment of meiosis-related
142 gene ontologies when compared to the brain (Figure 3G, Supplementary Table 3), such as meiotic cell
143 cycle process (GO:1903046, p.adjust = 2.2e-07) and meiotic nuclear division (GO:0140013, p.adjust
144 = 1.37e-07). While there was modest enrichment of meiotic gene ontologies in both brain regions, the
145 *Kdm5c*-KO hippocampus primarily expressed late-stage spermatogenesis genes involved in sperm axoneme
146 assembly (GO:0007288, p.adjust = 0.00621) and sperm motility (GO:0097722, 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 a subset
150 of epiblast stem cells become primordial germ cells (PGCs) and then again in mature germ cells to trigger
151 meiotic gene expression programs^{44–46}. Of note, some germline genes, including *Dazl*, are also expressed
152 in the two-cell embryo^{41,47}. However, we did not see derepression of two-cell stage-specific genes, like
153 *Duxf3* (*Dux*) (log₂ fold change = -0.282, q = 0.337) and *Zscan4d* (log₂ fold change = 0.25, q = 0.381) (Figure
154 3H, Supplementary Table 3), indicating *Kdm5c*-KO EpiLCs do not revert back to a 2-cell state. Altogether,
155 *Kdm5c*-KO EpiLCs express key drivers of germline identity and meiosis while the brain primarily expresses
156 spermiogenesis genes, indicating germline gene misexpression mirrors germline development during the
157 progression of somatic development.

158 **Female epiblast-like cells have heightened germline gene misexpression with *Kdm5c***

159 **loss**

160 It is currently unknown if the misexpression of germline genes is influenced by sex, as previous studies
161 on germline gene repressors have focused on male cells^{5,6,8,48,49}. Sex is particularly pertinent in the case
162 of KDM5C because it partially escapes X chromosome inactivation (XCI), resulting in a higher dosage in
163 females⁵⁰⁻⁵³. We therefore explored the impact of chromosomal sex upon germline gene suppression by
164 comparing their dysregulation in male *Kdm5c* hemizygous knockout (*Kdm5c*^{-y}, XY *Kdm5c*-KO, XY 5CKO),
165 female homozygous knockout (*Kdm5c*^{-/-}, XX *Kdm5c*-KO, XX 5CKO), and female heterozygous knockout
166 (*Kdm5c*^{+/+}, XX *Kdm5c*-HET, XX 5CHET) EpiLCs⁴³.

167 In EpiLCs, homozygous and heterozygous *Kdm5c* knockout females expressed over double the number
168 of germline-enriched genes than hemizygous males (Figure 4A, Supplementary Table 3). While the majority
169 of germline DEGs in *Kdm5c*-KO males were also dysregulated in females (74%), many were sex-specific,
170 such as *Tktl2* and *Esx1* (Figure 4B). We then compared the known functions of germline genes dysregulated
171 uniquely in males and females or misexpressed in all samples (Figure 4C, Supplementary Table 3). Female-
172 specific germline DEGs were enriched for meiotic (GO:0051321 - meiotic cell cycle, p.adjust = 7.81E-14) and
173 flagellar (GO:0003341 - cilium movement, p.adjust = 4.87E-06) functions, while male-specific DEGs had roles
174 in mitochondrial and cell signaling (GO:0070585 - protein localization to mitochondrion, p.adjust = 0.025).

175 The majority of germline genes expressed in both sexes were more highly dysregulated in females
176 compared to males (Figure 4D-F). This increased degree of dysregulation in females, along with the
177 increased total number of germline genes, indicates females are more sensitive to losing KDM5C-mediated
178 germline gene suppression. Heightened germline gene dysregulation in females could be due to impaired
179 XCI in *Kdm5c* mutants⁴³, as many spermatogenesis genes lie on the X chromosome^{54,55}. However, female
180 germline DEGs were not biased towards the X chromosome (p = 1, Odds Ratio = 0.96, Fisher's Exact Test)
181 and females had a similar overall proportion of germline DEGs belonging to the X chromosome as males
182 (XY *Kdm5c*-KO - 10.29%, XX *Kdm5c*-HET - 7.43%, XX *Kdm5c*-KO - 10.59%) (Figure 4G). The majority of
183 germline DEGs instead lie on autosomes for both male and female *Kdm5c* mutants (Figure 4G). Thus, while
184 female EpiLCs are more prone to germline gene misexpression with KDM5C loss, it is likely independent of

185 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, e.g. PGC
188 formation, meiosis, and genome defense, some have unique or sex-biased expression. Therefore, we
189 wondered if *Kdm5c* mutant males would primarily express sperm genes while mutant females would primarily
190 express egg genes. To comprehensively assess whether germline gene sex corresponds with *Kdm5c*
191 mutant sex, we first filtered our list of germline-enriched genes for egg and sperm-biased genes (Figure 4,
192 Supplementary Table 2). We defined germ cell sex-biased genes as those whose expression in the opposite
193 sex, at any time point, is no greater than 20% of the gene's maximum expression in a given sex. This
194 criteria yielded 67 egg-biased, 1,024 sperm-biased, and 197 unbiased germline-enriched genes. We found
195 regardless of sex, egg, sperm, and unbiased germline genes were dysregulated in all *Kdm5c* mutants at
196 similar proportions (Figure 4I-J). Furthermore, germline genes dysregulated exclusively in either male or
197 female mutants were also not biased towards their corresponding germ cell sex (Figure 4I). Altogether, these
198 results demonstrate sex differences in germline gene dysregulation is not due to sex-specific activation of
199 sperm or egg transcriptional programs.

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

201 KDM5C binds to the promoters of several germline genes in embryonic stem cells (ESCs) but not in
202 neurons^{22,56}. However, due to the lack of a comprehensive list of germline-enriched genes, it is unclear if
203 KDM5C is enriched at germline gene promoters, what types of germline genes KDM5C regulates, and if its
204 binding is 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 EpiLCs⁴³ and primary forebrain neuron cultures (PNCs)²¹ (MACS2 q <
207 0.1, fold enrichment > 1, and removal of false-positive *Kdm5c*-KO peaks). EpiLCs had a higher total number
208 of high-confidence KDM5C peaks than PNCs (EpiLCs: 5,808, PNCs: 1,276). KDM5C was primarily localized
209 to gene promoters in both cell types (promoters = transcription start site (TSS) ± 500 bp, EpiLCs: 4,190,

210 PNCs: 745), although PNCs showed increased localization to non-promoter 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, Supplementary Table 4). Germline ontologies were enriched only in EpiLC-specific,
216 KDM5C-bound promoters, such as meiotic nuclear division (GO: 0007127 p.adjust = 6.77e-16) (Figure 5C).
217 There were no significant ontologies for PNC-specific KDM5C target genes. Using our mouse germline gene
218 list, 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, Supplementary Table 2). Of note, KDM5C
221 was only bound to about one third of RNA-seq DEG promoters unique to EpiLCs or the brain (EpiLC only
222 DEGs: 34.9%, Brain only DEGs: 30%) (Supplementary Figure 1A-C). Representative examples of EpiLC
223 DEGs bound and unbound by KDM5C in EpiLCs are *Dazl* and *Stra8*, respectively (Figure 5F). However,
224 the four of the five germline genes dysregulated in both EpiLCs and the brain were bound by KDM5C in
225 EpiLCs (*D1Pas1*, *Hsf2bp*, *Cyct*, and *Stk31*) (Supplementary Figure 1A). Together, these results demonstrate
226 KDM5C is recruited to a subset of germline genes in EpiLCs, including meiotic genes, but does not directly
227 regulate germline genes in neurons. Furthermore, the majority of germline mRNAs expressed in *Kdm5c*-KO
228 cells are dysregulated independent of direct KDM5C recruitment to their gene promoters, however genes
229 dysregulated across *Kdm5c*-KO development are often direct KDM5C targets.

230 Many germline-specific genes are suppressed by the polycomb repressive complex 1.6 (PRC1.6), which
231 contains the transcription factor heterodimers E2F/DP1 and MGA/MAX that respectively bind E2F and
232 E-box motifs within germline gene promoters^{5,6,8,42,48,49,57-59}. PRC1.6 members may recruit KDM5C to
233 germline gene promoters²², given their association with KDM5C in HeLa cells and ESCs^{47,60}. We thus
234 used HOMER⁶¹ to identify transcription factor motifs enriched at KDM5C-bound or unbound germline gene
235 promoters (TSS ± 500 bp, q-value < 0.1, Supplementary Table 4). MAX and E2F6 binding sites were
236 significantly enriched at germline genes bound by KDM5C in EpiLCs (MAX q-value: 0.0068, E2F6 q-value:

237 0.0673, E2F q-value: 0.0917), but not at germline genes unbound by KDM5C (Figure 5G). One third of
238 KDM5C-bound promoters contained the consensus sequence for either E2F6 (E2F, 5'-TCCCGC-3'), MGA
239 (E-box, 5'-CACGTG-3'), or both, but only 17% of KDM5C-unbound genes contained these motifs (Figure 5H).
240 KDM5C-unbound germline genes were instead enriched for multiple RFX transcription factor binding sites
241 (RFX q-value < 0.0001, RFX2 q-value < 0.0001, RFX5 q-value < 0.0001) (Figure 5I, Supplementary figure
242 1D). RFX transcription factors bind X-box motifs⁶² to promote ciliogenesis^{63,64} and among them is RFX2, a
243 central regulator of post-meiotic spermatogenesis^{65,66}. Although *Rfx2* is also not a direct target of KDM5C
244 (Supplementary Figure 1E), RFX2 mRNA is derepressed in *Kdm5c*-KO EpiLCs (Figure 5J). Thus, RFX2 is a
245 candidate transcription factor for driving the ectopic expression of many KDM5C-unbound germline genes in
246 *Kdm5c*-KO cells.

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

249 Previous work found two germline gene promoters have a marked reduction in DNA CpG methylation
250 (CpGme) in the adult *Kdm5c*-KO hippocampus²². Since histone H3K4me2/3 impede *de novo* CpGme^{67,68},
251 KDM5C's removal of H3K4me2/3 may be required to suppress germline genes. However, KDM5C's catalytic
252 activity was recently shown to be dispensable for suppressing *Dazl* in undifferentiated ESCs⁴⁷. To reconcile
253 these observations, we hypothesized KDM5C erases H3K4me2/3 to promote the initial placement of CpGme
254 at germline gene promoters in EpiLCs.

255 To test this hypothesis, we first characterized KDM5C's expression as naïve ESCs differentiate into
256 EpiLCs (Figure 6A). While *Kdm5c* mRNA steadily decreased from 0 to 48 hours of differentiation (Figure
257 6B), KDM5C protein initially increased from 0 to 24 hours and then decreased to near knockout levels by 48
258 hours (Figure 6C). We then characterized KDM5C's substrates (H3K4me2/3) at germline gene promoters
259 with *Kdm5c* loss using published ChIP-seq datasets^{24,43}. *Kdm5c*-KO samples showed a marked increase in
260 H3K4me2 in EpiLCs (Figure 6D) and H3K4me3 in the amygdala (Figure 6E) around the TSS of germline
261 genes. Together, these data suggest KDM5C acts during the transition between ESCs and EpiLCs to remove
262 H3K4me2/3 at germline gene promoters.

263 Germline genes accumulate CpG methylation (CpGme) at CpG islands (CGIs) during the transition
264 from naïve to primed pluripotency^{7,9,69}. We first examined how many of our germline-enriched genes had
265 promoter CGIs (TSS ± 500 bp) using the UCSC genome browser⁷⁰. Notably, out of 1,288 germline-enriched
266 genes, only 356 (27.64%) had promoter CGIs (Figure 6F, Supplementary Table 2). CGI-containing germline
267 genes had higher enrichment of meiotic gene ontologies compared to CGI-free genes, including meiotic
268 nuclear division (GO:0140013, p.adjust = 2.17e-12) and meiosis I (GO:0007127, p.adjust = 3.91e-10)
269 (Figure 6G, Supplementary Table 5). Germline genes with promoter CGIs were more highly expressed than
270 CGI-free genes across spermatogenesis stages, with highest expression in meiotic spermatocytes (Figure
271 6H). Contrastingly, CGI-free genes only displayed substantial expression in post-meiotic round spermatids
272 (Figure 6H). Although only a minor portion of germline gene promoters contained CGIs, CGIs strongly
273 determined KDM5C's recruitment to germline genes ($p = 2.37e-67$, Odds Ratio = 17.8, Fisher's Exact Test),
274 with 79.15% of KDM5C-bound germline gene promoters harboring CGIs (Figure 6F).

275 To assess how KDM5C loss impacts initial CpGme placement at germline gene promoters, we performed
276 whole genome bisulfite sequencing (WGBS) in male wild-type and *Kdm5c*-KO ESCs and 96-hour extend
277 EpiLCs (exEpiLCs), when germline genes reach peak methylation levels⁶ (Figure 6I). We first identified
278 which germline gene promoters significantly gained CpGme in wild-type cells during nESC to exEpiLCs
279 differentiation (methylKit⁷¹, $q < 0.01$, $|methylation\ difference| > 25\%$, TSS ± 500 bp). In wild-type cells, the
280 majority of germline genes gained substantial CpGme at their promoter during differentiation (60.08%),
281 regardless if their promoter contained a CGI (Figure 6J, Supplementary Table 5).

282 We then identified promoters differentially methylated in wild-type versus *Kdm5c*-KO exEpiLCs (methylKit,
283 $q < 0.01$, $|methylation\ difference| > 25\%$, TSS ± 500 bp, Supplementary Table 5). Of the 48,882 promoters
284 assessed, 274 promoters were significantly hypomethylated and 377 promoters were significantly hyper-
285 methylated with KDM5C loss (Supplementary Figure 2A). Many promoters hyper- and hypomethylated
286 in *Kdm5c*-KO exEpiLCs belonged to genes with unknown functions. However, 10.22% of hypomethyl-
287 lated promoters belonged to germline genes and germline-relevant ontologies like meiotic nuclear division
288 (GO:0140013, p.adjust = 0.012) are significantly enriched (Supplementary Figure 2B, Supplementary Table
289 5). Approximately half of all germline gene promoters hypomethylated in *Kdm5c*-KO exEpiLCs are direct

290 targets of KDM5C in EpiLCs (13 out of 28 hypomethylated promoters).
291 Promoters that showed the most robust loss of CpGme in *Kdm5c*-KO exEpiLCs (lowest q-values) harbored
292 CGIs (Figure 6K). CGI promoters, but not CGI-free promoters, had a significant reduction in CpGme with
293 KDM5C loss as a whole (Figure 6L) (Non-CGI promoters p = 0.0846, CGI promoters p = 0.0081, Mann-
294 Whitney U test). Significantly hypomethylated promoters included germline genes consistently dysregulated
295 across multiple *Kdm5c*-KO RNA-seq datasets²², such as *D1Pas1* (methylation difference = -60.03%, q-value
296 = 3.26e-153) and *Naa11* (methylation difference = -42.45%, q-value = 1.44e-38) (Figure 6M). Unexpectedly,
297 we observed only a modest reduction in CpGme at *Dazl*'s promoter (methylation difference = -6.525%,
298 q-value = 0.0159) (Figure 6N). Altogether, these results demonstrate KDM5C is recruited to germline gene
299 CGIs in EpiLCs to promote CpGme at those promoters. Furthermore, our data suggest while KDM5C's
300 catalytic activity is required for the repression of some germline genes, CpGme can be placed at others even
301 with elevated H3K4me2/3 around the TSS.

302 Discussion

303 In the above study, we demonstrate KDM5C's pivotal role in the development of tissue identity. We first
304 characterized tissue-enriched genes expressed within the mouse *Kdm5c*-KO brain and identified substantial
305 derepression of testis, liver, muscle, and ovary-enriched genes. Testis genes significantly enriched within the
306 *Kdm5c*-KO amygdala and hippocampus are specific to the germline and absent in somatic cells. *Kdm5c*-
307 KO epiblast-like cells (EpiLCs) aberrantly express key drivers of germline identity and meiosis, including
308 *Dazl* and *Stra8*, while the adult brain primarily expresses genes important for late spermatogenesis. We
309 demonstrated that although sex did not influence whether sperm or egg-specific genes were misexpressed,
310 female EpiLCs have heightened germline gene de-repression with KDM5C loss. Germline genes can become
311 aberrantly expressed in *Kdm5c*-KO cells via indirect mechanisms, such as activation through ectopic RFX
312 transcription factors. Finally, we found KDM5C is dynamically regulated during ESC to EpiLC differentiation
313 to promote long-term germline gene silencing through CGI DNA methylation. Therefore, we propose KDM5C
314 plays a fundamental role in the development of tissue identity during early embryogenesis, including the

315 establishment of the soma-germline boundary. By systematically characterizing KDM5C's role in germline
316 gene repression, we unveiled divergent mechanisms governing the misexpression of distinct germline gene
317 classes in somatic lineages.

318 By comparing *Kdm5c* mutant males and females, we revealed germline gene suppression is sexually
319 dimorphic. Female EpiLCs are more severely impacted by loss of KDM5C-mediated germline gene sup-
320 pression, yet this difference is not due to the large number of germline genes on the X chromosome^{54,55}.
321 Heightened germline gene misexpression in females may be related to females having a higher dose of
322 KDM5C than males, due to its escape from XCI⁵⁰⁻⁵³. Intriguingly, heterozygous knockout females (*Kdm5c*^{-/+})
323 also had over double the number of germline DEGs than hemizygous knockout males (*Kdm5c*^{+/y}), even
324 though their expression of KDM5C should be roughly equivalent to that of wild-type males (*Kdm5c*^{+/y}). Males
325 could partially compensate for KDM5C's loss via the Y-chromosome homolog, KDM5D¹⁴. However, KDM5D
326 has not been reported to regulate germline gene expression. Nevertheless, these results demonstrate
327 germline gene silencing mechanisms differ between males and females, which warrants further study to
328 elucidate the biological ramifications and underlying mechanisms.

329 We found KDM5C is largely dispensable for promoting normal gene expression during development, yet
330 is critical for suppressing ectopic developmental programs. While some germline genes, such as *Dazl*, are
331 also expressed in the 2-cell stage, the inner cell mass, and naïve ESCs, they are silenced in epiblast stem
332 cells/EpiLCs^{6,42,47,72,73}. Our data suggest the 2-cell-like state reported in *Kdm5c*-KO ESCs⁴⁷ likely reflects
333 KDM5C's primary role in germline gene repression (Figure 3). Germline gene misexpression in *Kdm5c*-
334 KO EpiLCs may indicate they are differentiating into primordial germ cell-like cells (PGCLCs)^{35,36,38}. Yet,
335 *Kdm5c*-KO EpiLCs had normal cellular morphology and properly expressed markers for primed pluripotency,
336 including *Otx2* which blocks EpiLC differentiation into PGCs/PGCLCs⁷⁴. In addition to unimpaired EpiLC
337 differentiation, *Kdm5c*-KO gross brain morphology is overall normal²¹ and hardly any brain-specific genes
338 were significantly dysregulated in the amygdala and hippocampus (Figure 1). Thus, ectopic germline gene
339 expression occurs in conjunction with overall proper somatic differentiation in *Kdm5c*-KO animals.

340 Our work provides novel insight into the cross-talk between H3K4me2/3 and CpGme, which are gen-
341 erally mutually exclusive⁷⁵. In EpiLCs, loss of KDM5C binding at a subset of germline gene promoters,

342 e.g. *D1Pas1*, strongly impaired promoter CGI methylation and resulted in their long-lasting de-repression
343 into adulthood. Removal of H3K4me2/3 at CGIs is a plausible mechanism for KDM5C-mediated germline
344 gene suppression^{22,56}, given H3K4me2/3 repell DNMT3 activity^{67,68}. However, emerging work indicates
345 many histone-modifying enzymes have non-catalytic functions that influence gene expression, sometimes
346 even more potently than their catalytic roles^{76,77}. Indeed, KDM5C's catalytic activity was recently found to be
347 dispensible for repressing *Dazl* in ESCs⁴⁷. In our study, *Dazl*'s promoter still gained CpGme in *Kdm5c*-KO
348 exEpiLCs, even with elevated H3K4me2. *Dazl* and a few other germline genes employ multiple repressive
349 mechanisms to facilitate CpGme, such as DNMT3A/B recruitment via E2F6 and MGA^{5,6,48,49}. Thus, while
350 some germline CGIs require KDM5C-mediated H3K4me removal to overcome promoter CGI escape from
351 CpGme^{75,78}, others do not. These results also suggest the requirement for KDM5C's catalytic activity can
352 change depending upon the locus and developmental stage. Further experiments are required to determine
353 if catalytically inactive KDM5C can suppress germline genes at later developmental stages.

354 By generating a comprehensive list of mouse germline-enriched genes, we revealed distinct derepressive
355 mechanisms governing early versus late-stage germline programs. Previous work on germline gene silencing
356 has focused on genes with promoter CGIs^{7,75}, and indeed the majority of KDM5C targets in EpiLCs were
357 germ cell identity genes harboring CGIs. However, over 70% of germline-enriched gene promoters lacked
358 CGIs, including the many KDM5C-unbound germline genes that are de-repressed in *Kdm5c*-KO cells. CGI-
359 free, KDM5C-unbound germline genes were primarily late-stage spermatogenesis genes and significantly
360 enriched for RFX2 binding sites, a central regulator of spermiogenesis^{65,66}. These data suggest that once
361 activated during early embryogenesis, drivers of germline gene expression like *Rfx2*, *Stra8*, and *Dazl* turn
362 on downstream germline programs, ultimately culminating in the expression of spermiogenesis genes in
363 the adult *Kdm5c*-KO brain. Therefore, we propose KDM5C is recruited via promoter CGIs to act as a brake
364 against runaway activation of germline-specific programs. Future studies should address how KDM5C is
365 targeted to CGIs.

366 The above work provides the mechanistic foundation for KDM5C-mediated repression of tissue and
367 germline-specific genes. However, the contribution of these ectopic, tissue-specific genes towards neurolog-
368 ical impairments is still unknown. In addition to germline genes, we also identified significant enrichment

369 of muscle and liver-enriched transcripts within the *Kdm5c*-KO brain. Intriguingly, select liver and muscle-
370 enriched DEGs do have known roles within the brain, such as the liver-enriched lipid metabolism gene
371 *Apolipoprotein C-I (Apoc1)*³⁰. *APOC1* dysregulation is implicated in Alzheimer's disease in humans⁷⁹ and
372 overexpression of *Apoc1* in the mouse brain can impair learning and memory⁸⁰. KDM5C may therefore be
373 crucial for neurodevelopment by fine-tuning the expression of tissue-enriched, dosage-sensitive genes like
374 *Apoc1*.

375 Given that germline genes have no known functions within the brain, their impact upon neurodevelopment
376 is currently unknown. In *C. elegans*, somatic misexpression of germline genes via loss of *Retinoblastoma*
377 (*Rb*) homologs results in enhanced piRNA signaling and ectopic P granule formation in neurons^{81,82}. Ectopic
378 testicular germline transcripts have also been observed in a variety of cancers, including brain tumors in
379 *Drosophila* and mammals and shown to promote cancer progression^{10,11,83–85}. Intriguingly, mouse models
380 and human cells for other chromatin-linked NDDs also display impaired soma-germline demarcation^{13,86,87},
381 such as mutations in DNA methyltransferase 3b (DNMT3B), H3K9me1/2 methyltransferases G9A/GLP,
382 and methyl-CpG -binding protein 2 (MECP2). Recently, the transcription factor ZMYM2 (ZNF198), whose
383 mutation causes a NDD (OMIM #619522), was also shown to repress germline genes by promoting H3K4me
384 removal and CpGme⁸⁸. Thus, KDM5C is among a growing cohort of neurodevelopmental disorders with
385 erosion of the germline-soma boundary. Further research is required to determine the impact of these
386 germline genes upon neuronal functions and the extent to which this phenomenon occurs in humans.

387 Materials and Methods

388 Classifying tissue-enriched and germline-enriched genes

389 Tissue-enriched differentially expresssd genes (DEGs) were determined by their classification in a previ-
390 ously published dataset from 17 male and female mouse tissues²⁵. This study defined tissue expression as
391 greater than 1 Fragments Per Kilobase of transcript per Million mapped read (FPKM) and tissue enrichment
392 as at least 4-fold higher expression than any other tissue.

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

394 depleted (*Kit*^{W/W^v}) male and female mouse embryos from embryonic day 12, 14, and 16³⁴, as well as adult
395 male testes³¹. Germline-enriched genes met the following criteria: 1) their expression is greater than 1
396 FPKM in wild-type germline 2) their expression in any wild-type somatic tissues²⁵ does not exceed 20%
397 of maximum expression in wild-type germline, and 3) their expression in the germ cell-depleted (*Kit*^{W/W^v})
398 germline, for any sex or time point, does not exceed 20% of maximum expression in wild-type germline. We
399 defined sperm and egg-biased genes as those whose expression in the opposite sex, at any time point, is no
400 greater than 20% of the gene's maximum expression in a given sex. Genes that did not meet this threshold
401 for either sex were classified as 'unbiased'.

402 Cell culture

403 We utilized our previously established cultures of male wild-type and *Kdm5c* knockout (-KO)
404 embryonic stem cells⁴³. Sex was confirmed by genotyping *Uba1/Uba1y* on the X and Y chromo-
405 somes with the following primers: 5'-TGGATGGTGTGCCAATG-3', 5'-CACCTGCACGTTGCCCTT-
406 3'. Deletion of *Kdm5c* exons 11 and 12, which destabilize KDM5C protein²¹, was confirmed
407 through the primers 5'-ATGCCCATATTAAGAGTCCCTG-3', 5'-TCTGCCTTGATGGGACTGTT-3', and
408 5'-GGTTCTAACACTCACATAGTG-3'.

409 Embryonic stem cells (ESCs) and epiblast-like cells were cultured using previously established
410 methods³⁹. Briefly, ESCs were initially cultured in primed ESC (pESC) media consisting of KnockOut
411 DMEM (Gibco#10829–018), fetal bovine serum (Gibco#A5209501), KnockOut serum replacement
412 (Invitrogen#10828–028), Glutamax (Gibco#35050-061), Anti-Anti (Gibco#15240-062), MEM Non-essential
413 amino acids (Gibco#11140-050), and beta-mercaptoethanol (Sigma#M7522). They were then transitioned
414 into ground-state, "naïve" ESCs (nESCs) by culturing for four passages in N2B27 media containing
415 DMEM/F12 (Gibco#11330–032), Neurobasal media (Gibco#21103–049), Gluamax (Gibco#35050-061),
416 Anti-Anti (Gibco#15240-062), N2 supplement (Invitrogen#17502048), and B27 supplement without vitamin
417 A (Invitrogen#12587-010), and beta-mercaptoethanol (Sigma#M7522). Both pESC and nESC media
418 were supplemented with 3 μ M GSK3 inhibitor CHIR99021 (Sigma #SML1046-5MG), 1 μ M MEK inhibitor
419 PD0325901 (Sigma #PZ0162-5MG), and 1,000 units/mL leukemia inhibitory factor (LIF, Millipore#ESG1107).

420 nESCs were differentiated into epiblast-like cells (EpiLCs, 48 hours) and extendend EpiLCs (exEpiLCs,
421 96 hours) by culturing in N2B27 media containing DMEM/F12, Neurobasal media, Gluamax, Anti-Anti, N2
422 supplement, B27 supplement (Invitrogen#17504044), and beta-mercaptoethanol supplemented with 10
423 ng/mL fibroblast growth factor 2 (FGF2, R&D Biotechne 233-FB), and 20 ng/mL activin A (R&D Biotechne
424 338AC050CF), as previously described³⁹.

425 **Real time quantitative PCR (RT-qPCR)**

426 nESCs were differentiated into EpiLCs as described above. Cells were lysed with Tri-reagent BD (Sigma
427 #T3809) at 0, 24, and 48 hours of differentiation. RNA was phase separated with 0.1 μ L/ μ L 1-bromo-3-
428 chloropropane (Sigma #B9673) and then precipitated with isopropanol (Sigma #I9516) and ethanol puri-
429 fied. For each sample, 2 μ g of RNA was reverse transcribed using the ProtoScript II Reverse transcriptase kit
430 from New England Biolabs (NEB #M0368S) and primed with oligo dT. Expression of *Kdm5c* was detected us-
431 ing the primers 5'-CCCATGGAGGCCAGAGAATAAG-3' 5'-CTCAGCGGATAAGAGAATTGCTAC-3' and nor-
432 malized to TBP using the primers 5'-TTCAGAGGATGCTCTAGGGAAGA-3' 5'-CTGTGGAGTAAGTCCTGTGCC-
433 3' with the Power SYBR™ Green PCR Master Mix (ThermoFisher #4367659).

434 **Western Blot**

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

440 **RNA sequencing (RNA-seq) data analysis**

441 After ensuring read quality via FastQC (v0.11.8), reads were then mapped to the mm10 *Mus musculus*
442 genome (Gencode) using STAR (v2.5.3a), during which we removed duplicates and kept only uniquely
443 mapped reads. Count files were generated by FeatureCounts (Subread v1.5.0), and BAM files were

444 converted to bigwigs using deeptools (v3.1.3) and visualized by the UCSC genome browser⁷⁰. RStudio
445 (v3.6.0) was then used to analyze counts files by DESeq2 (v1.26.0)²⁶ to identify differentially expressed
446 genes (DEGs) with a q-value (p-adjusted via FDR/Benjamini–Hochberg correction) less than 0.1 and a log2
447 fold change greater than 0.5. For all DESeq2 analyses, log2 fold changes were calculated with IfcShrink
448 using the ashr package⁸⁹. MA-plots were generated by ggpubr (v0.6.0), and Eulerr diagrams were generated
449 by eulerr (v6.1.1). Boxplots and scatterplots were generated by ggpubr (v0.6.0) and ggplot2 (v3.3.2). The
450 Upset plot was generated via the package UpSetR (v1.4.0)⁹⁰. Gene ontology (GO) analyses were performed
451 by the R package enrichPlot (v1.16.2) using the biological processes setting and compareCluster.

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

453 ChIP-seq reads were aligned to mm10 using Bowtie1 (v1.1.2) allowing up to two mismatches. Only
454 uniquely mapped reads were used for analysis. Peaks were called using MACS2 software (v2.2.9.1) using
455 input BAM files for normalization, with filters for a q-value < 0.1 and a fold enrichment > 1. We removed
456 “black-listed” genomic regions that often give aberrant signals. Common peak sets were obtained in R via
457 DiffBind⁹¹ (v3.6.5). In the case of KDM5C ChIP-seq, *Kdm5c*-KO false-positive peaks were then removed from
458 wild-type samples using bedtools (v2.25.0). Peak proximity to genomic loci was determined by ChIPSeeker⁹²
459 (v1.32.1). Gene ontology (GO) analyses were performed by the R package enrichPlot (v1.16.2) using the
460 biological processes setting and compareCluster. Enriched motifs were identified using HOMER⁶¹ to search
461 for known motifs within 500 base pairs up and downstream of the transcription start site. Average binding
462 across genes was visualized using deeptools (v3.1.3). Bigwigs were visualized using the UCSC genome
463 browser⁷⁰.

464 CpG island (CGI) analysis

465 Locations of CpG islands were determined through the mm10 UCSC genome browser CpG island track⁷⁰,
466 which classified CGIs as regions that have greater than 50% GC content, are larger than 200 base pairs,
467 and have a ratio of CG dinucleotides observed over the expected amount greater than 0.6. CGI genomic
468 coordinates were then annotated using ChIPseeker⁹² (v1.32.1) and filtered for ones that lie within promoters

469 of germline-enriched genes (TSS ± 500).

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

471 Genomic DNA (gDNA) from male naïve ESCs and extended EpiLCs was extracted using the Wizard
472 Genomic DNA Purification Kit (Promega A1120), following the instructions for Tissue Culture Cells. gDNA
473 from two wild-types and two *Kdm5c*-KOs of each cell type was sent to Novogene for WGBS using the
474 Illumina NovaSeq X Plus platform and sequenced for 150 bp paired-end reads (PE150). All samples had
475 greater than 99% bisulfite conversion rates. Reads were adapter and quality trimmed with Trim Galore
476 (v0.6.10) and aligned to the mm10 genome using Bismark⁹³ (v0.22.1). Analysis of differential methylation at
477 gene promoters was performed using methylKit⁷¹ (v1.28.0) with a minimum coverage of 3 paired reads, a
478 percentage greater than 25% or less than -25%, and q-value less than 0.01. methylKit was also used to
479 calculate average percentage methylation at germline gene promoters. Methylation bedgraph tracks were
480 generated via Bismark and visualized using the UCSC genome browser⁷⁰.

481 **Data access**

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

483 Raw fastq files are deposited in the Sequence Read Archive (SRA) <https://www.ncbi.nlm.nih.gov/sra>
484 under the bioProject PRJNA1165148. <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1165148>

485 **Published datasets**

486 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
487 adult amygdala and hippocampus²⁴, available at GEO: GSE127722. Male and female wild-type, *Kdm5c*-KO,
488 and *Kdm5c*-HET EpiLCs⁴³ are available at GEO: GSE96797.

490 Previously published ChIP-seq experiments included KDM5C binding in wild-type and *Kdm5c*-KO
491 EpiLCs⁴³ (available at GEO: GSE96797) and mouse primary neuron cultures (PNCs) from the cortex
492 and hippocampus²¹ (available at GEO: GSE61036). ChIP-seq of histone 3 lysine 4 dimethylation (H3K4me2)

493 in male wild-type and *Kdm5c*-KO EpiLCs⁴³ is also available at GEO: GSE96797. ChIP-seq of histone 3 lysine
494 4 trimethylation (H3K4me3) in wild-type and *Kdm5c*-KO male amygdala²⁴ are available at GEO: GSE127817.

495 **Data analysis**

496 Scripts used to generate the results, tables, and figures of this study are available via the GitHub
497 repository: https://github.com/kbonefas/KDM5C_Germ_Mechanism

498 **Competing Interest**

499 S.I. is a member of the Scientific Advisory Board of KDM5C Advocacy, Research, Education & Support
500 (KARES). All other authors declare no conflict of interest.

501 **Acknowledgements**

502 We thank Drs. Sundeep Kalantry, Milan Samanta, and Rebecca Malcore for providing protocols and
503 expertise in culturing mouse ESCs and EpiLCs, as well as providing the wild-type and *Kdm5c*-KO ESCs
504 used in this study. We thank Dr. Jacob Mueller for his insight in germline gene regulation and directing
505 us to the germline-depleted mouse models. We also thank Drs. Gabriel Corfas, Kenneth Kwan, Natalie
506 Tronson, Michael Sutton, Stephanie Bielas, Donna Martin, and the members of the Iwase, Sutton, Bielas,
507 and Martin labs for helpful discussions and critiques of the data. We thank members of the University
508 of Michigan Reproductive Sciences Program for providing feedback throughout the development of this
509 work. This work was supported by grants from the National Institutes of Health (NIH) National Institute of
510 Neurological Disorders and Stroke (NS089896, 5R21NS104774, and NS116008 to S.I.), National Institute
511 of Mental Health (1R21MH135290 to S.I.), the Simons Foundation Autism Research Initiative (SFARI, SFI-
512 AN-AR-Pilot-00005721 to S.I.), the Farrehi Family Foundation Grant (to S.I.), the University of Michigan
513 Career Training in Reproductive Biology (NIH T32HD079342, to K.M.B.), the NIH Early Stage Training in
514 the Neurosciences Training Grant (NIH T32NS076401 to K.M.B.), and the Michigan Predoctoral Training in
515 Genetics Grant (NIH T32GM007544, to I.V.)

516 **Author Contributions**

517 K.M.B. and S.I. conceived the study and designed the experiments. I.V. generated the ESC and exEpiLC
518 WGBS data. K.M.B performed all data analysis and all other experiments. The manuscript was written by
519 K.M.B and S.I. and edited by K.M.B, S.I., and I.V.

520 **Declaration of Interest**

521 S.I. is a member of the Scientific Advisory Board of KDM5C Advocacy, Research, Education & Support
522 (KARES). Other authors declare no conflict of interest.

523 **References**

- 524 1. Hanschen, E.R., Herron, M.D., Wiens, J.J., Nozaki, H., and Michod, R.E. (2018). Multicellularity
525 Drives the Evolution of Sexual Traits. *Am Nat* **192**, E93–E105. <https://doi.org/10.1086/698301>.
- 526 2. Michod, R.E. (2011). Evolutionary Transitions in Individuality: Multicellularity and Sex. In *The Major
527 Transitions in Evolution Revisited*, B. Calcott and K. Sterelny, eds. (The MIT Press), pp. 169–198.
<https://doi.org/10.7551/mitpress/8775.003.0015>.
- 528 3. Chen, L., and Wiens, J.J. (2021). Multicellularity and sex helped shape the Tree of Life. *Proc. R. Soc.
529 B.* **288**, 20211265. <https://doi.org/10.1098/rspb.2021.1265>.
- 530 4. Devlin, D.K., Ganley, A.R.D., and Takeuchi, N. (2023). A pan-metazoan view of germline-soma
531 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 5. 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
533 stem cells by regulating germ cell-related genes. *Elife* **6**. <https://doi.org/10.7554/eLife.21064>.

- 534 6. 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>.
- 535
- 536 7. 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**,
537 1093–1100. <https://doi.org/10.1038/ng.708>.
- 538 8. 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–
539 9286. <https://doi.org/10.1073/pnas.1000473107>.
- 540 9. 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*
541 139, 3623–3632. <https://doi.org/10.1242/dev.081661>.
- 542 10. Nielsen, A.Y., and Gjerstorff, M.F. (2016). Ectopic Expression of Testis Germ Cell Proteins in Cancer
543 and Its Potential Role in Genomic Instability. *Int J Mol Sci* **17**. <https://doi.org/10.3390/ijms17060890>.
- 544 11. 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
545 Sciences* **20**. <https://doi.org/10.22038/ijbms.2017.9259>.
- 546 12. Berdasco, M., and Esteller, M. (2010). Aberrant Epigenetic Landscape in Cancer: How Cellular
547 Identity Goes Awry. *Developmental Cell* **19**, 698–711. <https://doi.org/10.1016/j.devcel.2010.10.005>.
- 548 13. Bonefas, K.M., and Iwase, S. (2021). Soma-to-germline transformation in chromatin-linked neurode-
549 velopmental disorders? *FEBS J.* <https://doi.org/10.1111/febs.16196>.

- 550 14. 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>
- 551 7.02.017.
- 552 15. Shen, H., Xu, W., Guo, R., Rong, B., Gu, L., Wang, Z., He, C., Zheng, L., Hu, X., Hu, Z., et al.
(2016). Suppression of Enhancer Overactivation by a RACK7-Histone Demethylase Complex. *Cell*
553 *165*, 331–342. <https://doi.org/10.1016/j.cell.2016.02.064>.
- 554 16. Chen, X., Loo, J.X., Shi, X., Xiong, W., Guo, Y., Ke, H., Yang, M., Jiang, Y., Xia, S., Zhao, M., et al.
(2018). E6 Protein Expressed by High-Risk HPV Activates Super-Enhancers of the *EGFR* and *c-MET*
Oncogenes by Destabilizing the Histone Demethylase KDM5C. *Cancer Research* *78*, 1418–1430.
555 <https://doi.org/10.1158/0008-5472.CAN-17-2118>.
- 556 17. Zheng, Q., Li, P., Zhou, X., Qiang, Y., Fan, J., Lin, Y., Chen, Y., Guo, J., Wang, F., Xue, H., et al. (2021).
Deficiency of the X-inactivation escaping gene *KDM5C* in clear cell renal cell carcinoma promotes
tumorigenicity by reprogramming glycogen metabolism and inhibiting ferroptosis. *Theranostics* *11*,
557 8674–8691. <https://doi.org/10.7150/thno.60233>.
- 558 18. Claes, S., Devriendt, K., Van Goethem, G., Roelen, L., Meireleire, J., Raeymaekers, P., Cassiman,
J.J., and Fryns, J.P. (2000). Novel syndromic form of X-linked complicated spastic paraparesis. *Am J
559 Med Genet* *94*, 1–4.
- 560 19. Jensen, L.R., Amende, M., Gurok, U., Moser, B., Gimmel, V., Tschach, A., Janecke, A.R., Tariverdian,
G., Chelly, J., Fryns, J.P., et al. (2005). Mutations in the JARID1C gene, which is involved in
transcriptional regulation and chromatin remodeling, cause X-linked mental retardation. *Am J Hum
561 Genet* *76*, 227–236. <https://doi.org/10.1086/427563>.
- 562 20. 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.
563 <https://doi.org/10.1111/cge.13755>.

- 564 21. 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>.
- 565
- 566 22. 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,
567 47–59. <https://doi.org/10.1016/j.celrep.2017.09.014>.
- 568 23. 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,
569 637. <https://doi.org/10.3390/cells12040637>.
- 570 24. 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>.
- 571
- 572 25. 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.
573 <https://doi.org/10.1038/s41598-017-04520-z>.
- 574 26. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for
575 RNA-seq data with DESeq2. *Genome Biol* 15, 550. <https://doi.org/10.1186/s13059-014-0550-8>.
- 576 27. 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
577 Chromosome Pairing in Meiosis. *Science* 300, 1291–1295. <https://doi.org/10.1126/science.1083022>.

- 578 28. 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>.
- 580 29. 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
581 membraneless compartment. *Science* 378, eabq4835. <https://doi.org/10.1126/science.abq4835>.
- 582 30. 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.
583 *Cardiovasc Diabetol* 21, 272. <https://doi.org/10.1186/s12933-022-01703-5>.
- 584 31. 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
585 chromosomes for the male germ line. *Nat Genet* 45, 1083–1087. <https://doi.org/10.1038/ng.2705>.
- 586 32. Handel, M.A., and Eppig, J.J. (1979). Sertoli Cell Differentiation in the Testes of Mice Genetically
Deficient in Germ Cells. *Biology of Reproduction* 20, 1031–1038. <https://doi.org/10.1095/biolreprod20.5.1031>.
- 588 33. 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>.
- 590 34. Soh, Y.Q., Junker, J.P., Gill, M.E., Mueller, J.L., van Oudenaarden, A., and Page, D.C. (2015). A
Gene Regulatory Program for Meiotic Prophase in the Fetal Ovary. *PLoS Genet* 11, e1005531.
591 <https://doi.org/10.1371/journal.pgen.1005531>.
- 592 35. Magnúsdóttir, E., and Surani, M.A. (2014). How to make a primordial germ cell. *Development* 141,
593 245–252. <https://doi.org/10.1242/dev.098269>.

- 594 36. Günesdogan, U., Magnúsdóttir, E., and Surani, M.A. (2014). Primordial germ cell specification: A
context-dependent cellular differentiation event [corrected]. *Philos Trans R Soc Lond B Biol Sci* **369**.
595 <https://doi.org/10.1098/rstb.2013.0543>.
- 596 37. 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>.
- 597 38. Hayashi, K., Ohta, H., Kurimoto, K., Aramaki, S., and Saitou, M. (2011). Reconstitution of the
mouse germ cell specification pathway in culture by pluripotent stem cells. *Cell* **146**, 519–532.
599 <https://doi.org/10.1016/j.cell.2011.06.052>.
- 600 39. Samanta, M., and Kalantry, S. (2020). Generating primed pluripotent epiblast stem cells: A methodol-
601 ogy chapter. *Curr Top Dev Biol* **138**, 139–174. <https://doi.org/10.1016/bs.ctdb.2020.01.005>.
- 602 40. 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.
603 *EMBO Reports* **16**, 791–802. <https://doi.org/10.15252/embr.201540538>.
- 604 41. 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
605 activity. *Nature* **487**, 57–63. <https://doi.org/10.1038/nature11244>.
- 606 42. 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
607 germline stem cells. *Nat Commun* **7**, 11056. <https://doi.org/10.1038/ncomms11056>.
- 608 43. 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](https://doi.org/10.1038/s41467-
609 022-30352-1).

- 610 44. 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**,
611 2474–2479. <https://doi.org/10.1073/pnas.0510813103>.
- 612 45. 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>.
- 613 46. Endo, T., Mikedis, M.M., Nicholls, P.K., Page, D.C., and De Rooij, D.G. (2019). Retinoic Acid and Germ
Cell Development in the Ovary and Testis. *Biomolecules* **9**, 775. <https://doi.org/10.3390/biom9120775>.
- 614 47. 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.
615 *Nat Struct Mol Biol*. <https://doi.org/10.1038/s41594-023-01038-z>.
- 616 48. 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-
617 Specific Gene Expression. *Current Biology* **15**, 1051–1057. <https://doi.org/10.1016/j.cub.2005.04.060>.
- 618 49. Dahlet, T., Truss, M., Frede, U., Al Adhami, H., Bardet, A.F., Dumas, M., Vallet, J., Chicher, J.,
Hammann, P., Kotnik, S., et al. (2021). E2F6 initiates stable epigenetic silencing of germline genes
619 during embryonic development. *Nat Commun* **12**, 3582. <https://doi.org/10.1038/s41467-021-23596-w>.
- 620 50. 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
621 and human. *Hum Mol Genet* **3**, 879–884. <https://doi.org/10.1093/hmg/3.6.879>.
- 622 51. 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>.
- 623 52. 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*
624 **5**, 1355–1360. <https://doi.org/10.1093/hmg/5.9.1355>.
- 625
- 626
- 627

- 628 53. Xu, J., Deng, X., and Disteche, C.M. (2008). Sex-Specific Expression of the X-Linked Histone
629 Demethylase Gene Jarid1c in Brain. PLoS ONE 3, e2553. <https://doi.org/10.1371/journal.pone.0002553>.
- 630 54. Wang, P.J., McCarrey, J.R., Yang, F., and Page, D.C. (2001). An abundance of X-linked genes
631 expressed in spermatogonia. Nat Genet 27, 422–426. <https://doi.org/10.1038/86927>.
- 632 55. 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
633 inactivation. Nat Genet 36, 642–646. <https://doi.org/10.1038/ng1368>.
- 634 56. 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>.
- 635 57. Hurlin, P.J. (1999). Mga, a dual-specificity transcription factor that interacts with Max and contains a
T-domain DNA-binding motif. The EMBO Journal 18, 7019–7028. <https://doi.org/10.1093/emboj/18.2.7019>.
- 636 58. 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
637 14, e1007193. <https://doi.org/10.1371/journal.pgen.1007193>.
- 638 59. 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.
641 <https://doi.org/10.1371/journal.pone.0205969>.
- 642 60. 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.
643 Nature 447, 601–605. <https://doi.org/10.1038/nature05823>.

- 644 61. 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.
645 <https://doi.org/10.1016/j.molcel.2010.05.004>.
- 646 62. 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*,
647 916–921. <https://doi.org/10.1038/35002634>.
- 648 63. 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).
649
- 650 64. 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
651 Ciliogenesis. *Sci. Signal.* *2*. <https://doi.org/10.1126/scisignal.2000602>.
- 652 65. 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.
653 *PLoS Genet* *11*, e1005368. <https://doi.org/10.1371/journal.pgen.1005368>.
- 654 66. Wu, Y., Hu, X., Li, Z., Wang, M., Li, S., Wang, X., Lin, X., Liao, S., Zhang, Z., Feng, X., et al.
(2016). Transcription Factor RFX2 Is a Key Regulator of Mouse Spermiogenesis. *Sci Rep* *6*, 20435.
655 <https://doi.org/10.1038/srep20435>.
- 656 67. Otani, J., Nankumo, T., Arita, K., Inamoto, S., Ariyoshi, M., and Shirakawa, M. (2009). Structural basis
for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX–DNMT3–DNMT3L
657 domain. *EMBO Reports* *10*, 1235–1241. <https://doi.org/10.1038/embor.2009.218>.
- 658 68. Guo, X., Wang, L., Li, J., Ding, Z., Xiao, J., Yin, X., He, S., Shi, P., Dong, L., Li, G., et al. (2015).
Structural insight into autoinhibition and histone H3-induced activation of DNMT3A. *Nature* *517*,
659 640–644. <https://doi.org/10.1038/nature13899>.

- 660 69. Meissner, A., Mikkelsen, T.S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B.E., Nusbaum, C., Jaffe, D.B., et al. (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* *454*, 766–770. <https://doi.org/10.1038/nature07107>.
- 661
- 662 70. 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 update. *Nucleic Acids Research* *51*, D1188–D1195. <https://doi.org/10.1093/nar/gkac1072>.
- 663
- 664 71. Akalin, A., Kormaksson, M., Li, S., Garrett-Bakelman, F.E., Figueroa, M.E., Melnick, A., and Mason, C.E. (2012). methylKit: A comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol* *13*, R87. <https://doi.org/10.1186/gb-2012-13-10-r87>.
- 665
- 666 72. Torres-Padilla, M.-E. (2020). On transposons and totipotency. *Philos Trans R Soc Lond B Biol Sci* *375*, 20190339. <https://doi.org/10.1098/rstb.2019.0339>.
- 667
- 668 73. 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* *29*, 400–418.e13. <https://doi.org/10.1016/j.stem.2022.01.010>.
- 669
- 670 74. Zhang, J., Zhang, M., Acampora, D., Vojtek, M., Yuan, D., Simeone, A., and Chambers, I. (2018). OTX2 restricts entry to the mouse germline. *Nature* *562*, 595–599. <https://doi.org/10.1038/s41586-018-0581-5>.
- 671
- 672 75. 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 genome. *Nat Genet* *39*, 457–466. <https://doi.org/10.1038/ng1990>.
- 673
- 674 76. Aubert, Y., Egolf, S., and Capell, B.C. (2019). The Unexpected Noncatalytic Roles of Histone Modifiers in Development and Disease. *Trends in Genetics* *35*, 645–657. <https://doi.org/10.1016/j.tig.2019.06.004>.
- 675
- 676 77. Morgan, M.A.J., and Shilatifard, A. (2020). Reevaluating the roles of histone-modifying enzymes and their associated chromatin modifications in transcriptional regulation. *Nat Genet* *52*, 1271–1281. <https://doi.org/10.1038/s41588-020-00736-4>.
- 677

- 678 78. 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**,
679 6693–6706. <https://doi.org/10.1093/nar/gkw258>.
- 680 79. 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>.
- 682 80. 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>.
- 684 81. 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
685 retinoblastoma pathway mutants. *Nature* **436**, 593–597. <https://doi.org/10.1038/nature04010>.
- 686 82. 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>.
- 688 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.
689 <https://doi.org/10.1126/science.1195481>.
- 690 84. Ghafouri-Fard, S., and Modarressi, M.-H. (2012). Expression of Cancer–Testis Genes in Brain Tumors:
691 Implications for Cancer Immunotherapy. *Immunotherapy* **4**, 59–75. <https://doi.org/10.2217/imt.11.145>.
- 692 85. Nin, D.S., and Deng, L.-W. (2023). Biology of Cancer-Testis Antigens and Their Therapeutic Implica-
693 tions in Cancer. *Cells* **12**, 926. <https://doi.org/10.3390/cells12060926>.

- 694 86. 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.
695 Orphanet J Rare Dis 9, 56. <https://doi.org/10.1186/1750-1172-9-56>.
- 696 87. 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
697 (Basel) 3, 578–605. <https://doi.org/10.3390/biology3030578>.
- 698 88. 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.
699 <https://doi.org/10.1093/nar/gkad540>.
- 700 89. Stephens, M. (2016). False discovery rates: A new deal. Biostat, kxw041. <https://doi.org/10.1093/biostatistics/kxw041>.
- 701 90. Conway, J.R., Lex, A., and Gehlenborg, N. (2017). UpSetR: An R package for the visualization of
intersecting sets and their properties. Bioinformatics 33, 2938–2940. <https://doi.org/10.1093/bioinformatics/btx364>.
- 702 91. Ross-Innes, C.S., Stark, R., Teschendorff, A.E., Holmes, K.A., Ali, H.R., Dunning, M.J., Brown, G.D.,
Gojis, O., Ellis, I.O., Green, A.R., et al. (2012). Differential oestrogen receptor binding is associated
703 with clinical outcome in breast cancer. Nature 481, 389–393. <https://doi.org/10.1038/nature10730>.
- 704 92. 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>.
- 705 93. Krueger, F., and Andrews, S.R. (2011). Bismark: A flexible aligner and methylation caller for Bisulfite-
706 Seq applications. Bioinformatics 27, 1571–1572. <https://doi.org/10.1093/bioinformatics/btr167>.

710 **Figures and Tables**

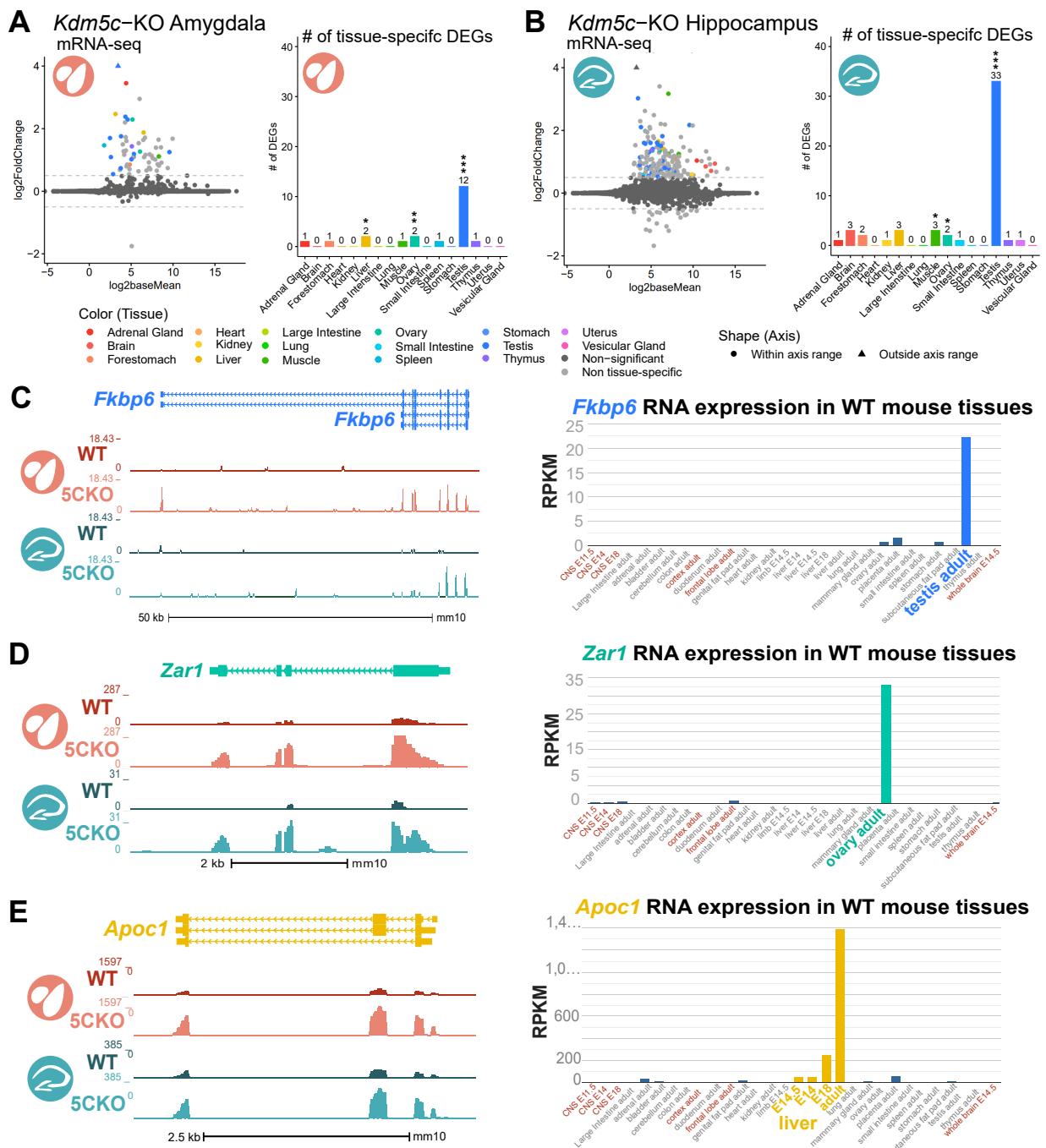


Figure 1: Tissue-enriched genes are misexpressed in the *Kdm5c*-KO brain. **A-B.** Expression of tissue-enriched genes (Li et al 2017) in the male *Kdm5c*-KO amygdala (A) and hippocampus (B). Left - MA plot of mRNA-sequencing. Right - Number of tissue-enriched differentially expressed genes (DEGs). * 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.

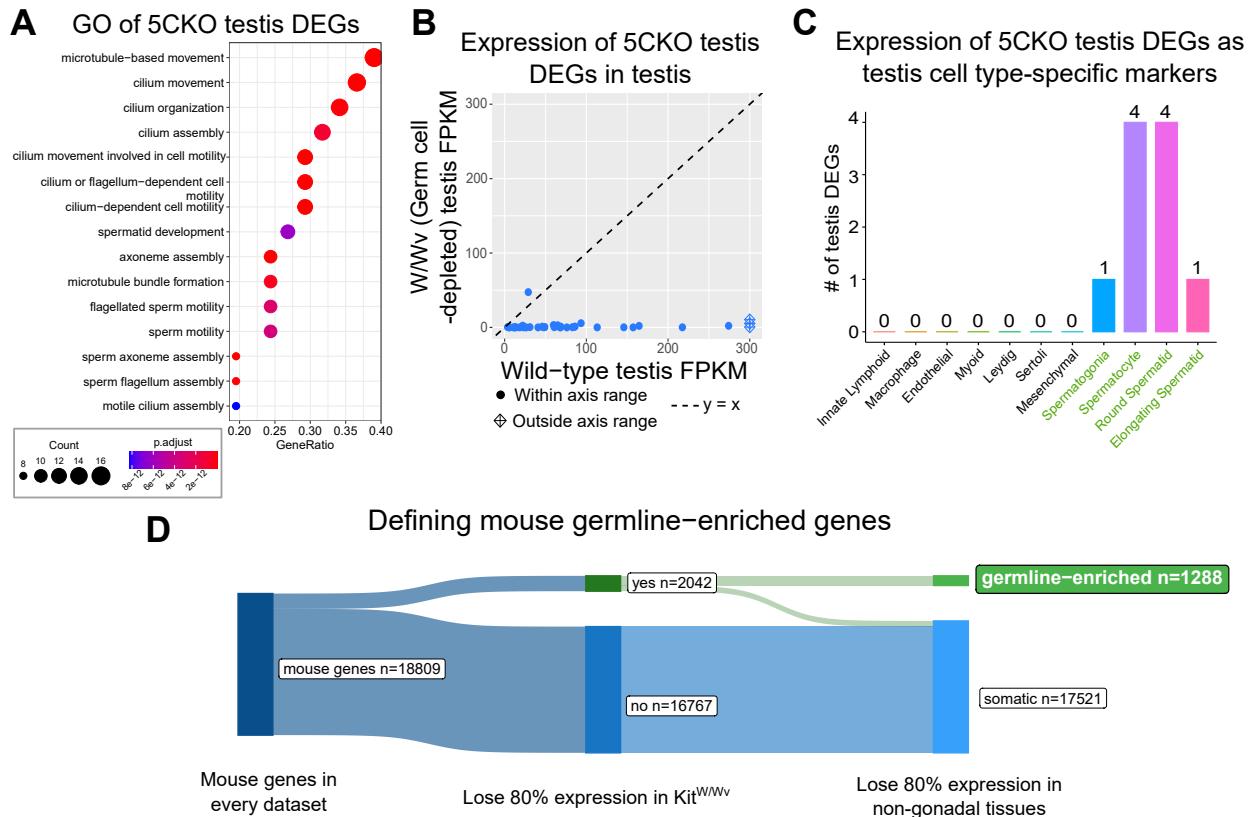


Figure 2: Aberrant transcription of germline genes in the *Kdm5c*-KO in the brain. **A.** enrichPlot gene ontology (GO) of *Kdm5c*-KO amygdala and hippocampus testis-enriched DEGs **B.** Expression of testis DEGs in wild-type (WT) testis versus germ cell-depleted (W/Wv) testis (Mueller et al 2013). Expression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

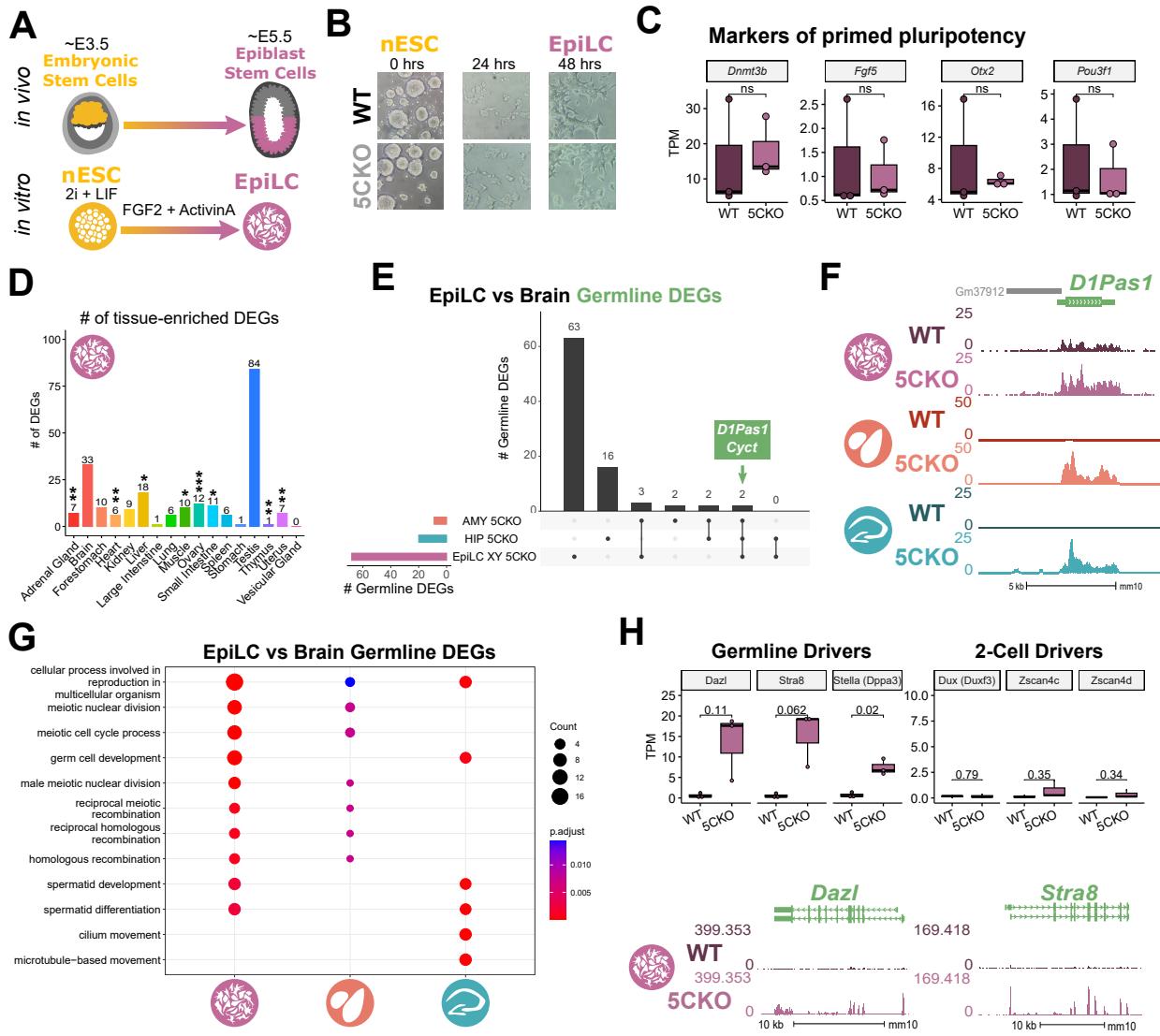


Figure 3: *Kdm5c*-KO epiblast-like cells express key drivers of germline identity

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. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, 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, $n = 3$).

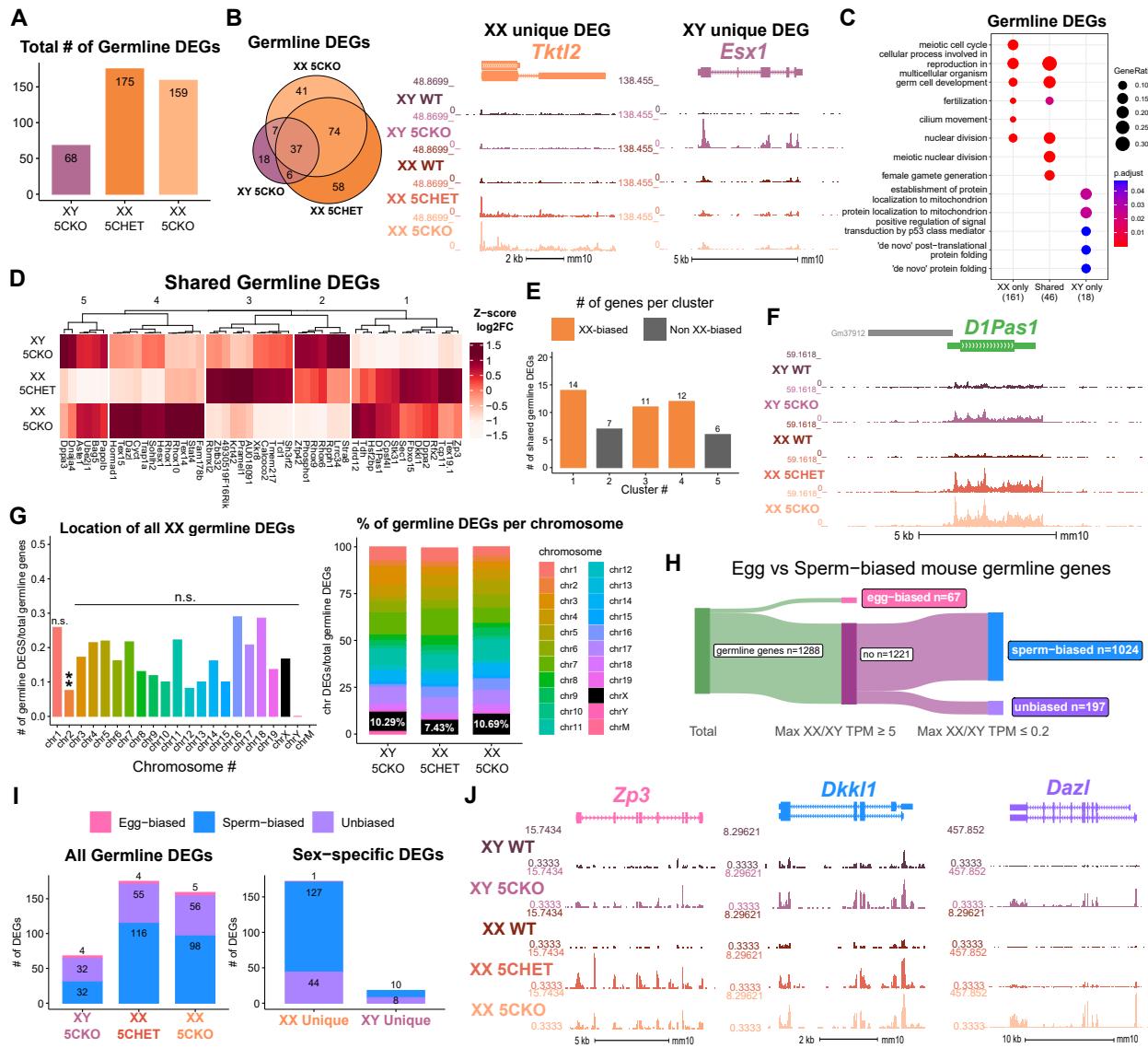


Figure 4: Chromosomal sex influences *Kdm5c*-KO germline gene misexpression. **A.** Total number of germline-enriched RNA-seq DEGs for male hemizygous *Kdm5c* knockout EpiLCs (XY 5CKO, purple), female heterozygous *Kdm5c* knockout (XX 5CHET, orange), female homozygous *Kdm5c* knockout (XX 5CKO, light 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 (Shared), or unique to one sex (XX only or XY only). **D.** Heatmap of germline DEGs shared between male and female mutants. Color is the average 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. P-values for Fisher Exact Test, ** p < 0.01, n.s. non-significant. Germline DEGs were only significant for chromosome 2, in which they were significantly depleted. 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).

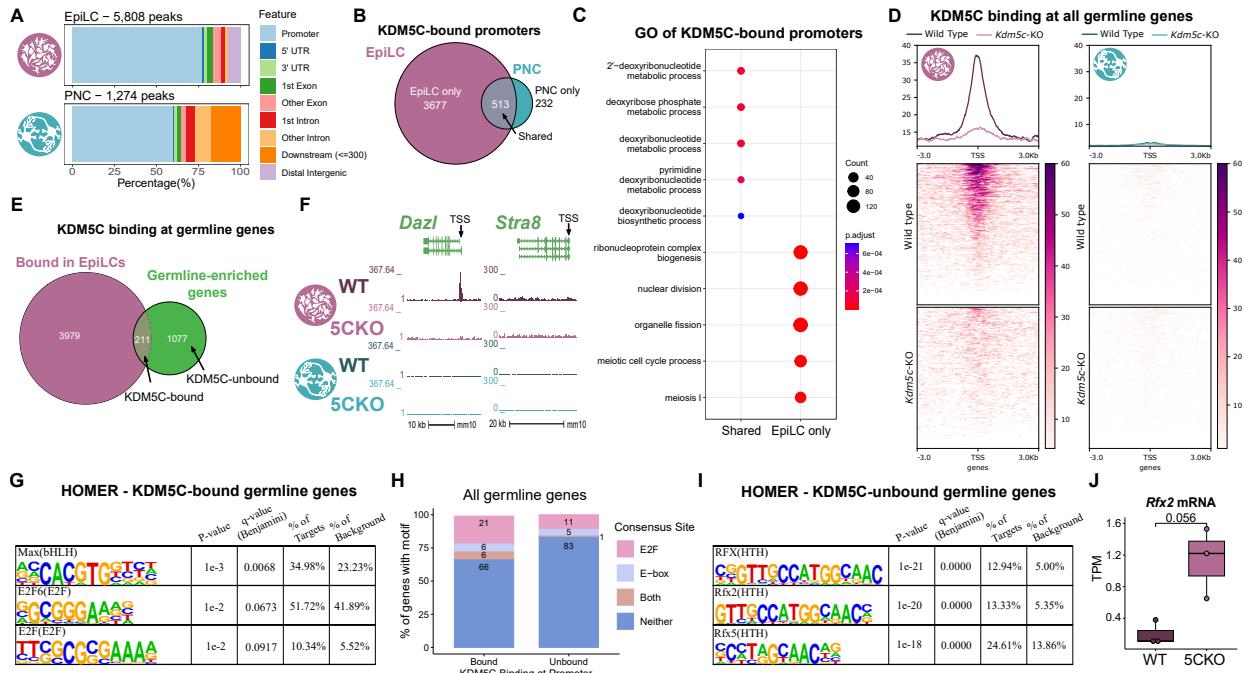


Figure 5: KDM5C binds to a subset of germline gene promoters during early embryogenesis. **A.** ChIPseeker localization of KDM5C peaks at different genomic regions in EpiLCs (top) and hippocampal and cortex primary neuron cultures (PNCs, bottom). **B.** Overlap of genes with KDM5C bound to their promoters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

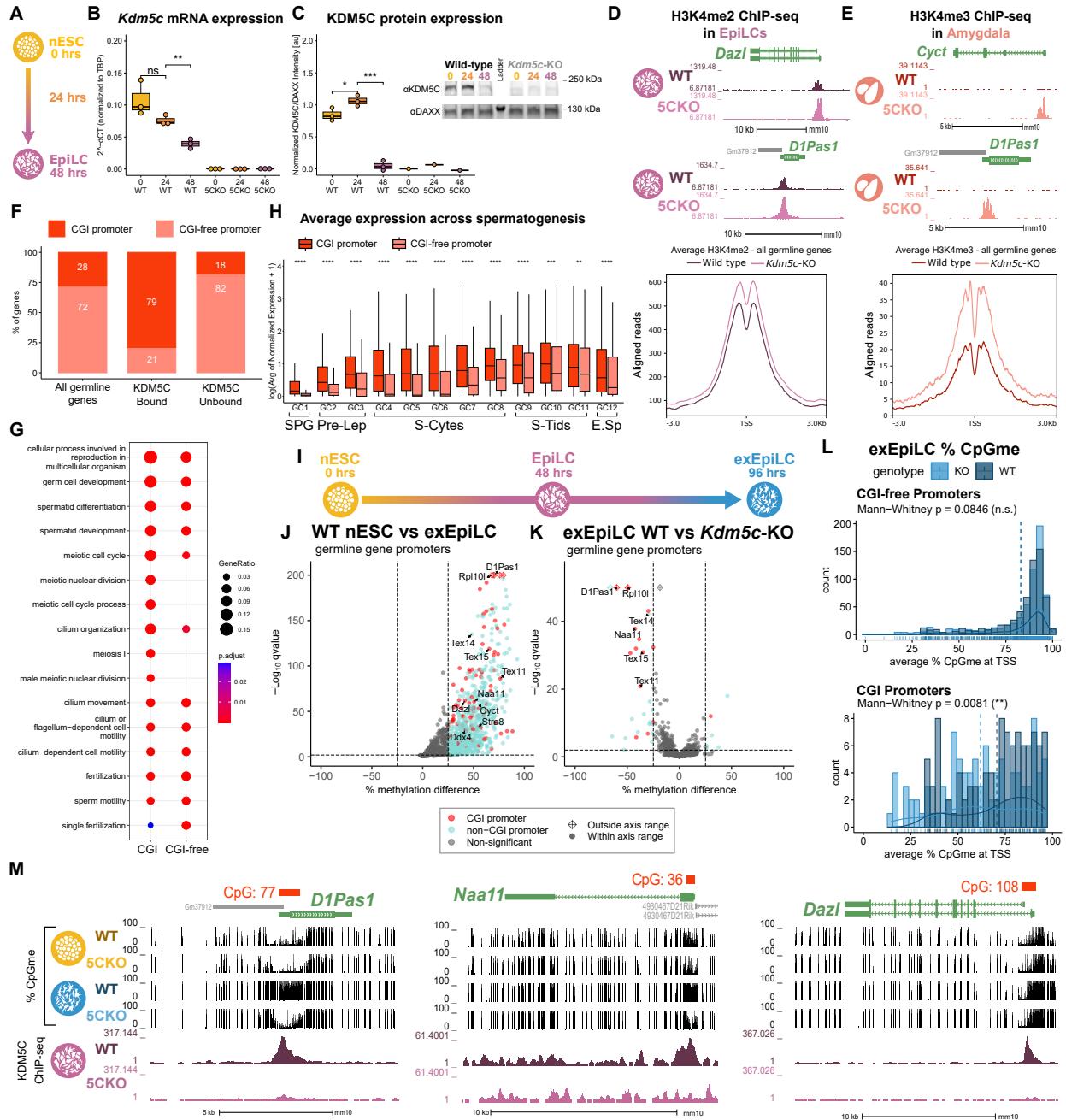


Figure 6: KDM5C promotes long-term silencing of germline genes via DNA methylation at CpG islands. **A.** Diagram of embryonic stem cell (ESC) to epiblast-like cell (EpiLC) differentiation and collection time points. **B.** Real time quantitative PCR (RT-qPCR) of *Kdm5c* mRNA expression in wild-type (WT) and *Kdm5c*-KO (5CKO) ESCs at 0, 24, and 48 hours of differentiation into EpiLCs. Expression calculated in comparison to TBP mRNA expression ($2^{-\Delta CT}$). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, 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.