

Erosion of somatic tissue identity with loss of the X-linked intellectual disability factor KDM5C

3

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

5 Abstract

It is currently unclear why mutations in numerous chromatin-modifying enzymes cause neurodevelopmental disorders (NDDs). Loss of repressive chromatin regulators can lead to the aberrant transcription of tissue-specific genes outside of their intended context, however the mechanisms behind their dysregulation and their functional consequences are largely unknown. Here, we explored this cellular identity crisis in NDDs by investigating the role of lysine demethylase 5c (KDM5C, also known as JARID1C or SMCX), an eraser of histone 3 lysine 4 di and tri-methylation (H3K4me2/3), in tissue identity. We found male *Kdm5c* knockout (-KO) mice, which recapitulate key behavioral phenotypes of Claes-Jensen X-linked intellectual disability, aberrantly expresses many liver, muscle, ovary, and testis genes within the amygdala and hippocampus. Gonad-enriched genes expressed in the *Kdm5c*-KO brain were typically unique to germ cells, indicating an erosion of the soma-germline boundary. We then curated a list of germline-enriched genes for unbiased characterization of KDM5C in germline gene repression. Germline genes are typically decommissioned in somatic lineages in the post-implantation epiblast, yet *Kdm5c*-KO epiblast-like cells (EpiLCs) aberrantly expressed key regulators of germline identity and meiosis, including *Dazl* and *Stra8*. Germline gene dysregulation was sexually dimorphic, as female *Kdm5c*-KO EpiLCs were more prone to germline gene misexpression than knockout males. We found KDM5C represses germline-specific transcription by binding to a subset of germline gene promoters in EpiLCs to facilitate DNA methylation of CpG islands. However, germline genes, particularly late-stage spermatogenesis genes, can also become activated in *Kdm5c*-KO cells independent of direct KDM5C binding. This suggests germline developmental programs can progress ectopically in the background of typical *Kdm5c*-KO development, due to downstream activation by germline transcription factors. These results define KDM5C's role in germline gene suppression and potentially link impaired soma-vs-germline demarcation to a chromatin-based neurodevelopmental disorder.

- 27 • Not sure if I like the last sentence, I think we should focus on what we did do versus what we could do
28 in the future

- 29 – Add something that we deepened the characterization of germline gene classes that use different
30 regulatory mechanisms (CpG islands vs not, meiotic vs late-stage, E2F6/MAX vs no)

31 Introduction

32 A single genome holds the instructions to generate the myriad of cell types found within the adult organism.
33 This is, in part, accomplished by chromatin regulators that can either promote or impede lineage-specific gene
34 expression through DNA and histone modifications^{1,2}. Many chromatin regulators were initially identified
35 for their roles in shaping cellular and tissue identity^{3–5}. Unexpectedly, human genetic studies revealed
36 mutations in chromatin regulators are a major cause of neurodevelopmental disorders (NDDs)⁶. Most studies
37 investigating this relationship have explored their regulation of brain-specific genes and chromatin marks.
38 However, loss of some chromatin regulators can also result in the ectopic transcription of tissue-specific
39 genes outside of their target environment, such as the misexpression of liver-specific genes within adult
40 neurons⁷. Very few studies have investigated this severe crisis in cellular identity in chromatin-linked NDDs^{7,8}
41 and it is currently unknown if these ectopic genes contribute to neurodevelopmental impairments.

42 To elucidate the role of tissue identity in chromatin-linked neurodevelopmental disorders, it is essential
43 to first characterize the nature of the dysregulated genes and the molecular mechanisms governing their
44 de-repression. We characterized lysine demethylase 5C (KDM5C), also known as SMCX or JARID1C, which
45 erases histone 3 lysine 4 di- and trimethylation (H3K4me2/3) found at active gene promoters⁹. Pathogenic
46 mutations in *KDM5C* cause Intellectual Developmental Disorder, X-linked, Syndromic, Claes-Jensen Type
47 (MRXSCJ, OMIM: 300534). MRXSCJ is more common and severe in males and its neurological phenotypes
48 include intellectual disability, seizures, aberrant aggression, and autistic behaviors^{10–12}. Male *Kdm5c*
49 knockout (-KO) mice recapitulate key MRXSCJ phenotypes, including hyperaggression, increased seizure
50 propensity, and learning impairments^{8,13}. RNA sequencing (RNA-seq) of the *Kdm5c*-KO hippocampus
51 revealed ectopic expression of testis genes within the brain⁸, however it is unknown what types of testis
52 genes are dysregulated, when in *Kdm5c*-KO development testis gene dysregulation begins, and if other
53 tissue-specific genes are also aberrantly transcribed with KDM5C loss.

54 Distinguishing between germ cells and somatic cells is a key feature of multicellularity¹⁴ that occurs
55 during early embryogenesis in many metazoans¹⁵. In mammals, chromatin regulators are crucial for
56 decommissioning germline genes in somatic cells during the transition from naïve to primed pluripotency.
57 Initially, germline gene promoters gain repressive histone H2A lysine 119 monoubiquitination (H2AK119ub1)¹⁶
58 and histone 3 lysine 9 trimethylation (H3K9me3)^{16,17} in embryonic stem cells and are then decorated with
59 DNA CpG methylation in the post-implantation embryo^{17–19}. The precise roles of KDM5C during this process
60 remains unclear. Additionally, studies on germline gene repression have primarily focused on marker genes
61 important for germ cell development rather than germline genes as a whole, given they lacked a curated
62 list of germline-enriched genes. Therefore, it is unknown if the mechanism of repression differs for certain

63 classes of germline genes, e.g. meiotic genes versus spermatid differentiation genes.

64 It is also unknown if ectopic germline gene expression is influenced by chromosomal sex, as previous
65 studies on germline gene repression have been exclusively in males. Sex is particularly pertinent in the case
66 of KDM5C, which lies on the X chromosome and partially escapes X chromosome inactivation, resulting in a
67 higher dosage in females²⁰⁻²³.

68 To illuminate KDM5C's role in tissue identity, here we characterized the aberrant transcription of tissue-
69 enriched genes within the *Kdm5c*-KO brain and epiblast-like stem cells (EpiLCs), an *in vitro* model of
70 the post-implantation embryo. We curated list of germline-enriched genes, which enabled genome-wide
71 analysis of germline gene silencing mechanisms. Based on the data presented below, we propose KDM5C
72 plays a fundamental role in the development of tissue identity during early embryogenesis, including the
73 establishment of the soma-germline boundary.

74 Results

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

76 Previous RNA sequencing (RNA-seq) in the adult hippocampus revealed ectopic expression of some
77 testis genes within the male *Kdm5c* knockout (-KO) brain⁸. It is currently unclear if the testis is the only
78 tissue type misexpressed in the *Kdm5c*-KO brain. We thus characterized dysregulation of *Kdm5c*-KO brain
79 tissue identity by systematically assessing the expression of genes enriched in 17 mouse tissues²⁴, using
80 our published mRNA-seq datasets of the adult amygdala and hippocampus for male mice with constitutive
81 knockout of *Kdm5c*²⁵.

82 We found a large proportion of genes that are significantly upregulated within the *Kdm5c*-KO brain
83 (DESeq2²⁶, log2 fold change > 0.5, q < 0.1) are enriched in non-brain tissues in wild-type mice (Amygdala:
84 35%, Hippocampus: 24%) (Figure 1A-B). The majority of tissue-enriched differentially expressed genes
85 (tissue-enriched DEGs) were testis genes (Figure 1A-C). Even though the testis has the largest total number
86 of tissue-biased genes compared to any other tissue (2,496 genes), testis-biased DEGs were significantly
87 enriched for both brain regions (Amygdala p = 1.83e-05, Odds Ratio = 5.13; Hippocampus p = 4.26e-11,
88 Odds Ratio = 4.45, Fisher's Exact Test). One example of a testis-enriched gene misexpressed in the
89 *Kdm5c*-KO brain is *FK506 binding protein 6* (*Fkbp6*), a known regulator of piRNA expression and meiosis in
90 germ cells^{27,28} (Figure 1C).

91 Interestingly, we also observed significant enrichment of ovary-biased DEGs in both the amygdala and
92 hippocampus (Amygdala p = 0.00574, Odds Ratio = 18.7; Hippocampus p = 0.048, Odds Ratio = 5.88,
93 Fisher's Exact) (Figure 1D). Ovary-enriched DEGs included *Zygotic arrest 1* (*Zar1*), which sequesters mRNAs
94 in oocytes for meiotic maturation and early zygote development²⁹ (Figure 1D). Given the *Kdm5c*-KO mice
95 we analyzed are male, this demonstrates ectopic expression of tissue-enriched genes is independent of

96 organismal sex.

97 Although not consistent across brain regions, we also found significant enrichment of DEGs biased
98 towards two non-gonadal tissues - the liver (Amygdala p = 0.0398, Odds Ratio = 6.58, Fisher's Exact Test)
99 and the muscle (Hippocampus p = 0.0104, Odds Ratio = 6.95, Fisher's Exact Test). A liver-biased DEG
100 dysregulated in both the hippocampus and amygdala is *Apolipoprotein C-I* (*Apoc1*), a lipoprotein metabolism
101 and transport gene³⁰ that has been implicated in Alzheimer's disease³¹ (Figure 1E). For all tissue-enriched
102 DEGs, aberrantly expressed mRNAs are polyadenylated and spliced into mature transcripts (Figure 1C-E).
103 Of note, we observed little to no dysregulation of brain-enriched genes (Amygdala p = 1; Hippocampus p =
104 0.74, Fisher's Exact), despite the fact these are brain samples and the brain has the second highest total
105 number of tissue-enriched genes (708 genes). Together, these results suggest the aberrant expression of
106 tissue-enriched genes within the brain is a major effect of KDM5C loss.

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

108 The testis contains both germ cells (meiotic cells, e.g. spermatogonia) and somatic cells (non-meiotic,
109 e.g. Leydig cells) that support hormone production and germline functions. To determine if *Kdm5c*-KO
110 testis-enriched DEGs are somatic or germline genes, we first evaluated their known functions through gene
111 ontology analysis. We found *Kdm5c*-KO testis-enriched DEGs had high enrichment of germline-relevant
112 ontologies, including spermatid development (GO: 0007286, p.adjust = 6.2e-12) and sperm axoneme
113 assembly (GO: 0007288, p.adjust = 2.45e-14) (Figure 2A).

114 To further validate if these testis DEGs are truly germline genes, we then assessed their expression in
115 somatic versus germ cells within the testis. We first compared their expression in wild-type testes to those
116 without germ cells³², which was accomplished by heterozygous *W* and *Wv* mutations in the enzymatic
117 domain of *c-Kit* (*Kit*^{W/Wv})³³. Almost all *Kdm5c*-KO testis-enriched DEGs lost expression with germ cell
118 depletion (Figure 2B). We then assessed testis-enriched DEG expression in a published single cell RNA-seq
119 dataset that identified cell type-specific markers within the testis³⁴. Some *Kdm5c*-KO testis-enriched DEGs
120 were classified as specific markers for different germ cell types (e.g. spermatogonia, spermatocytes, round
121 spermatids, and elongating spermatids), yet none marked somatic cells (Figure 2C). Together, these data
122 demonstrate that the *Kdm5c*-KO brain aberrantly expresses germline genes, but not somatic testicular genes,
123 reflecting an erosion between somatic versus germline identity.

124 We then wanted to characterize germline gene misexpression with *Kdm5c* loss more deeply, but lacked
125 a comprehensive list of mouse germline genes. We therefore generated a list of germline-enriched genes
126 using RNA-seq datasets of *Kit*^{W/Wv} mice that included males and females at embryonic day 12, 14, and 16³⁵
127 and adult male testes³². We defined genes as germline-enriched if their expression met the following criteria:
128 1) their expression is greater than 1 FPKM in wild-type gonads 2) their expression in any non-gonadal tissue
129 of adult wild type mice²⁴ does not exceed 20% of their maximum expression in the wild-type germline, and
130 3) their expression in the germ cell-depleted gonads, for any sex or time point, does not exceed 20% of

131 their maximum expression in the wild-type germline. These criteria yielded 1,288 germline-enriched genes
132 (Figure 2D), which was hereafter used as a resource to globally assess germline gene misexpression with
133 *Kdm5c* loss (Supplementary table 1).

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

135 Germ cells are typically distinguished from somatic cells soon after the embryo implants into the uterine
136 wall^{36,37}. A subset of epiblast stem cells become the primordial germ cells (PGCs) while the remainder
137 differentiate into the ectoderm, mesoderm, and endoderm to form the somatic tissues³⁸. This developmental
138 time point can be modeled *in vitro* through differentiation of embryonic stem cells (ESCs) into epiblast-like
139 stem cells (EpiLCs) (Figure 3A, top)^{39,40}. While some germline-enriched genes are also expressed in
140 naïve embryonic stem cells (ESCs) and in the 2-cell stage^{41–43}, they are silenced as they differentiate into
141 EpiLCs^{17,44}. Therefore, we tested if KDM5C was necessary for silencing germline genes at this developmental
142 stage by evaluating the impact of *Kdm5c* loss in male EpiLCs.

143 We first identified *Kdm5c*-KO EpiLC DEGs in our previously published RNA-seq dataset⁴⁵ (DESeq2, log2
144 fold change > 0.5, q < 0.1) and then assessed tissue-enriched gene expression. Similar to the *Kdm5c*-KO
145 brain, we observed general dysregulation of tissue-enriched genes, with the largest number of genes
146 belonging to the brain and testis, although they were not significantly enriched (Figure 3B). Markers of primed
147 pluripotency, such as *Dnmt3b*, *Fgf5*, *Pou3f1*, and *Otx2*, were unaltered in *Kdm5c*-KO EpiLCs (Figure 3C).
148 Furthermore, *Kdm5c*-KO cell morphology during ESC to EpiLC differentiation (Figure 3D) appeared normal,
149 indicating KDM5C loss does not impair EpiLC formation.

150 To determine if germline DEGs are constitutively dysregulated or if they can change over the course of
151 development, we next compared EpiLC germline DEGs to those expressed in the *Kdm5c*-KO brain. We found
152 the majority of germline DEGs were unique to either EpiLCs or the brain, with only *Cyct* shared across all
153 tissue/cell types (Figure 3E-F). EpiLCs had particularly high enrichment of meiosis-related gene ontologies
154 (Figure 3G), such as meiotic cell cycle process (GO:1903046, p.adjust = 1.59e-08) and meiotic nuclear
155 division (GO:0140013, p.adjust = 9.76e-09). While there was modest enrichment of meiotic gene ontologies
156 in both brain regions, the *Kdm5c*-KO hippocampus primarily expressed late-stage spermatogenesis genes,
157 such those involved in sperm axoneme assembly (GO:0007288, p.adjust = 0.00621) and sperm motility
158 (GO:0097722, p.adjust = 0.00612).

159 Notably, DEGs unique to *Kdm5c*-KO EpiLCs included key drivers of germline identity, such as *Stimulated*
160 *by retinoic acid 8* (*Stra8*) and *Deleted in azoospermia like* (*Dazl*) (Figure 3H). These genes are typically
161 expressed during embryonic germ cell development to commit PGCs to the germline fate, but are also
162 expressed later in life to trigger meiotic gene expression programs^{46–48}. Of note, some germline genes,
163 including *Dazl*, are also expressed in the two-cell embryo^{42,49}. However, we did not see misexpression of
164 two-cell embryo-specific genes, like *Duxf3* (*Dux*) (q = 0.337) and *Zscan4d* (q = 0.381), indicating *Kdm5c*-KO
165 in EpiLCs do not revert back to a 2-cell-like state (Figure 3H). Altogether, these results demonstrate that while

166 the *Kdm5c*-KO brain primarily expresses spermatogenesis genes, *Kdm5c*-KO EpiLCs express key drivers of
167 germline identity and meiosis.

168 **Female epiblast-like cells have increased sensitivity to germline gene misexpression
169 with *Kdm5c* loss**

170 It is currently unknown if the misexpression of germline genes is influenced by chromosomal sex, as
171 previous studies on germline gene repressors have focused on their regulation in males^{16–18,50,51}. We
172 explored the impact of sex upon germline gene misexpression by comparing their dysregulation in male
173 *Kdm5c* hemizygous knockout (XY *Kdm5c*-KO), female homozygous knockout (XX *Kdm5c*-KO) and female
174 heterozygous (XX *Kdm5c*-HET) EpiLCs.⁴⁵ We first identified differentially expressed genes (DEGs) com-
175 pared to sex-matched wild-type controls (DESeq2²⁶, log2 fold change > 0.5, q < 0.1) and then filtered for
176 germline-enriched genes.

177 Homozygous and heterozygous females expressed over double the number of germline-enriched genes
178 than hemizygous males (Figure 4A). While the majority of germline DEGs in *Kdm5c*-KO males were also
179 dysregulated in females (74%), there were also many male-specific and female-specific germline DEGs
180 (Figure 4B). We compared the known functions of germline genes dysregulated only in females (XX
181 only - XX *Kdm5c*-KO, XX *Kdm5c*-HET, or both), only in males (XY only), or in all samples (shared). Female-
182 specific germline DEGs were enriched for meiotic (GO:XXX meiotic cell cycle) and flagellar (GO:XXX cilium
183 movement) functions, male-specific DEGs had mitochondrial and cell signaling gene functions (GO:XXX
184 protein localization to mitochondrion). Germline transcripts expressed in both sexes were enriched for meiotic
185 (meiotic nuclear division) and egg-specific functions (female gamete generation).

186 The majority of germline genes expressed in both sexes had a greater log2 fold change from wild-type
187 in females compared to males (Figure 4D). This increased degree of dysregulation in females, along with
188 the increased total number of germline genes, indicates females are more sensitive to the loss of KDM5C-
189 mediated suppression of germline genes. Female sensitivity could be due to improper X chromosome
190 inactivation (XCI) in *Kdm5c* mutants, as many spermatogenesis genes lie on the X chromosome^{52,53}.
191 However, both shared and female-specific germline DEGs were not biased towards the X chromosome, with
192 the majority of genes instead lying on autosomes (Figure 4G). Thus, while female EpiLCs are more prone to
193 germline gene misexpression with KDM5C loss, it is likely independent of potential defects in XCI.

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

195 While many germline genes act in both the male and female germline, some display sex-biased expression
196 or have functions unique to eggs and sperm. To comprehensively assess whether germline gene sex
197 corresponds with *Kdm5c* mutant sex, we filtered our list of germline-enriched genes for egg (XX) and
198 sperm (XY) biased genes. We defined sex-biased genes as those whose expression in the opposite sex,

199 at any time point, is no greater than 20% of the gene's maximum expression in a given sex. This yielded
200 67 egg-biased, 1,024 sperm-biased, and 197 unbiased germline-enriched genes (Figure 4H), which is
201 consistent with the testis overall having a more unique transcriptome than the ovary²⁴. We found egg,
202 sperm, and unbiased germline genes were dysregulated in all *Kdm5c* mutants (Figure 4I-J). Germline genes
203 dysregulated exclusively in males and females were also not biased towards their corresponding germ
204 cell sex. This indicates differences between male and female germline gene dysregulation is not due to
205 sex-specific activation of sperm or egg-specific transcriptional programs. These results demonstrate that
206 sex influences the degree of germline gene misexpression with KDM5C loss, but not the germ cell sex from
207 which the germline genes originate.

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

209 Previous work suggests KDM5C represses germline genes during early development, as re-expression
210 of KDM5C in knockout neuronal cultures fails to suppress two hippocampal germline DEGs⁸. KDM5C binds
211 to several germline gene promoters in embryonic stem cells (ESCs) but is lost in neurons⁸. However, the lack
212 of a comprehensive list of germline-enriched genes prohibited systematic characterization of KDM5C binding
213 at germline gene promoters. It is currently unclear if KDM5C is enriched at germline gene promoters, what
214 types of germline genes KDM5C regulates, and if its binding is maintained at any germline genes in neurons.

215 To address these questions, we analyzed KDM5C chromatin immunoprecipitation followed by DNA
216 sequencing (ChIP-seq) datasets in EpiLCs⁴⁵ and primary forebrain neuron cultures (PNCs)¹³. EpiLCs had a
217 higher total number of KDM5C peaks than PNCs (EpiLCs: 5,808, PNCs: 1,276, MACS2 q < 0.1 and fold
218 enrichment > 1, removal of *Kdm5c*-KO false positive peaks) and KDM5C was primarily localized to gene
219 promoters in both cell types (EpiLCs: 4,190, PNCs: 745 +/- 500bp from the TSS), although PNCs showed
220 increased localization to non-promoter regions (Figure 5A).

221 The majority of promoters bound by KDM5C in PNCs were also bound in EpiLCs (513 shared promoters),
222 however a large portion of gene promoters were only bound by KDM5C in EpiLCs (3677 EpiLC only
223 promoters) (Figure 5B). Genes bound by KDM5C in both PNCs and EpiLCs were enriched for functions
224 involving nucleic acid turnover, such as deoxyribonucleotide metabolic process (GO:0009262, p.adjust =
225 8.28e-05) (Figure 5C). Germline-specific ontologies were only enriched in promoters unique to EpiLCs, such
226 as meiotic nuclear division (GO: 0007127 p.adjust = 6.77e-16) and meiotic cell cycle process (GO:1903046,
227 p.adjust = 5.05e-16) (Figure 5C). There were no ontologies significantly enriched for genes only bound
228 by KDM5C in PNCs. We next evaluated KDM5C binding around the transcription start site (TSS) of all
229 germline-enriched genes. In EpiLCs, We observed modest KDM5C signal at about half of all germline genes
230 (Figure 5D). Based on our ChIP-seq peak cut-off criteria, KDM5C was strongly bound to about 10% of
231 germline gene promoters in EpiLCs (Figure 5E). In condordance with our gene ontology results, we did not
232 observe KDM5C accumulation at any germline gene promoters in PNCs (Figure 5D). Together, these results
233 demonstrate KDM5C is significantly enriched at a subset of germline gene promoters in EpiLCs, including

234 meiotic genes, but does not directly regulate germline genes in neurons.

235 Many germline-specific genes are suppressed by the transcription factor heterodimers E2F6/DP1 and
236 MGA/MAX, which respectively bind E2F and E-box motifs^{18,50,51,54,55}. To elucidate if KDM5C is recruited to
237 germline gene promoters by a similar mechanism, we used HOMER to identify transcription factor motifs
238 enriched at KDM5C-bound or unbound germline gene promoters⁵⁶ (TSS +/- 500 bp, q-value < 0.1). MAX
239 and E2F6 binding sites were significantly enriched at germline genes bound by KDM5C in EpiLCs, but
240 not at germline genes unbound by KDM5C (MAX q-value: 0.0068, E2F6 q-value: 0.0673, E2F q-value:
241 0.0917) (Figure 5F). One third of KDM5C-bound promoters contained the consensus sequence for either
242 E2F6 (E2F, 5'-TCCCGC-3'), MGA (E-box, 5'-CACGTG-3'), or both, but only 17% of KDM5C-unbound genes
243 contained these motifs (Figure 5G). KDM5C-unbound germline genes were instead enriched for multiple
244 RFX transcription factor binding sites (RFX q-value < 0.0001, Rfx2 q-value < 0.0001, Rfx5 q-value < 0.0001)
245 (Figure 5H). RFX transcription factors bind X-box motifs⁵⁷ to promote ciliogenesis^{58,59}. Enriched RFX
246 transcription factors included RFX2, a central regulator of post-meiotic spermatogenesis^{60,61}. Interestingly,
247 RFX2 mRNA is upregulated in *Kdm5c*-KO EpiLCs (Figure 5I), but is not a direct target of KDM5C. Thus,
248 distinct transcription factor programs regulate the expression of KDM5C-bound and unbound germline genes.

249 Finally, we compared KDM5C binding at the promoters of RNA-seq DEGs to determine if the germline
250 genes transcribed in *Kdm5c*-KO cells are direct targets of KDM5C (Figure 5J-L). In EpiLCs, KDM5C was
251 bound to about one third of EpiLC-specific and brain-specific germline DEG promoters (EpiLC only: 36%,
252 Brain only: 33.3%). Again, we did not observe any KDM5C binding at germline gene promoters in PNCs,
253 even for brain-specific DEGs. Some notable differences in KDM5C binding for EpiLC-specific DEGs included
254 *Dazl* and *Stra8*, two key drivers of germline identity discussed above. Although both mRNAs are expressed
255 in *Kdm5c*-KO EpiLCs, KDM5C is only bound to *Dazl*'s promoter and not *Stra8*'s in EpiLCs (Figure 5K). In
256 contrast to the unique DEGs, 3 out of the 4 genes dysregulated in both the brain and EpiLCs are bound
257 by KDM5C around the TSS. Altogether, this suggests KDM5C decommissions a subset of germline genes
258 in EpiLCs but is no longer recruited to germline gene promoters in neurons. Furthermore, the majority of
259 germline mRNAs expressed in *Kdm5c*-KO cells are dysregulated independent of direct KDM5C recruitment
260 to their promoters.

261 **KDM5C promotes *de novo* DNA methylation at germline genes (if there are differ-
262 ences, say at germline gene CpG islands)**

263 In the early embryo, germline gene promoters are initially decorated with repressive histone modifications
264 and are then silenced long-term via DNA CpG methylation (CpGme)^{16,17,44,62}. Our results above indicate
265 KDM5C also acts at germline gene promoters during this time period. However, how KDM5C interacts with
266 other germline gene silencing mechanisms is currently unclear. KDM5C is generally thought to suppress
267 transcription through erasure of histone 3 lysine 4 di- and trimethylation (H3K4me2/3)⁹, yet KDM5C's

268 catalytic activity was recently shown to be dispensable for suppressing *Dazl* in undifferentiated ESCs⁴⁹. Since
269 H3K4me3 impedes *de novo* CpGme placement^{63,64}, KDM5C's catalytic activity may instead be required later
270 in development for long-term silencing of germline genes. In support of this, CpGme is markedly reduced at
271 two germline gene promoters in the *Kdm5c*-KO adult hippocampus⁸.

272 Based on the above observations, we hypothesized KDM5C erases H3K4me3 to promote the initial
273 placement of CpGme at germline gene promoters in EpiLCs. To test this hypothesis, we first characterized
274 KDM5C's substrates, histone 3 lysine 4 di- and trimethylation (H3K4me2/3) in our previously published ChIP-
275 seq datasets of the wild type and *Kdm5c*-KO amygdala²⁵ and EpiLCs⁴⁵. In congruence with previous work in
276 the *Kdm5c*-KO hippocampus⁸, we observed aberrant accumulation of H3K4me3 around the transcription
277 start site (TSS) of germline genes in the *Kdm5c*-KO amygdala (Figure 6A). We additionally found a marked
278 increase in H3K4me2 around the TSS of germline genes in *Kdm5c*-KO EpiLCs (Figure 6B).

279 We then evaluated KDM5C's role in germline gene silencing during ESC to EpiLC differentiation. We first
280 characterized KDM5C's mRNA and protein expression at 0, 24, and 48 hours of differentiation into EpiLCs
281 (Figure 6C). While *Kdm5c* mRNA steadily decreased from 0 to 48 hours of differentiation (Figure 6D), KDM5C
282 protein initially increased from 0 to 24 hours but then decreased to near knockout levels by 48 hours (Figure
283 6E).

284 To determine KDM5C's role in the initial placement of DNA methylation at germline gene promoters,
285 we performed whole genome bisulfite sequencing (WGBS) in wild-type and *Kdm5c*-KO ESCs and 96 hour
286 extended EpiLCs (exEpiLCs).

- 287 • Germline genes are known to accumulate CpGme at (CGIs) during the transition from naive to primed
288 pluripotency.
- 289 • We first identified the number of germline genes with CGIs neighboring their promoter.
 - 290 – Found XXX% of germline genes had CGIs, XXX% of which were direct KDM5C targets in EpiLCs
291 (Eulerr).
- 292 • We then curated a list of germline genes that significantly gained CpGme in wild-type exEpiLCs
293 compared to ESCs.
 - 294 – Example gene bedgraph
 - 295 – Majority of CGI germline genes?
- 296 • Out of the CGI genes, which had significantly reduced CpGme in 5CKO (heatmap of % methylation)
 - 297 – Highlight interesting genes affected vs unaffected by KDM5C (especially if same/different from
298 E2F6, PRC1.6, Setdb1 targets)
 - 299 – CGIs that never gain DNAme in WT

- 300 – Although wild-type cells accumulated high levels of DNA methylation at germline gene promoters
301 over the course of ESCs to exEpiLC differentiation, DNA methylation was markedly reduced in
302 *Kdm5c*-KO exEpiLCs (Figure 6F).
- 303 • Non-CGI germline genes, any significant changes at their promoter? (Heatmap again?)
304 • Altogether, this suggests KDM5C is dynamically regulated during ESC to EpiLC differentiation and
305 promotes germline gene silencing via DNA methylation during early embryogenesis.

306 Discussion

307 We observed general dysregulation of tissue-enriched genes in both the *Kdm5c*-KO brain and EpiLCs,
308 including testis, liver, muscle, and ovary-enriched genes. The *Kdm5c*-KO amygdala and hippocampus had
309 significant enrichment of testis genes that are specific to germ cells and not somatic cells of the testis. We
310 found *Kdm5c*-KO EpiLCs aberrantly expressed key drivers of germline identity and meiosis, including *Dazl*
311 and *Stra8*, while the *Kdm5c*-KO brain primarily expressed germline genes important for late spermatogenesis.
312 Female *Kdm5c*-KO EpiLCs expressed over double the number of germline genes than *Kdm5c*-KO males and
313 germline genes dysregulated in both sexes were expressed more highly in females, demonstrating females
314 have increased sensitivity to germline gene dysregulation. KDM5C was bound to only a subset of germline
315 genes expressed in *Kdm5c*-KO EpiLCs, indicating germline-enriched mRNAs can be aberrantly transcribed
316 through indirect mechanisms. Finally, we found KDM5C promotes the long-term silencing of germline genes
317 in somatic cells by aiding the placement of DNA CpG methylation in EpiLCs. - Change: we propose KDM5C
318 plays a fundamental role in the development of tissue identity during early embryogenesis, including the
319 establishment of the soma-germline boundary.

320 • Thus, systematically characterizing KDM5C's role in germline gene repression during early embryogen-
321 esis, including its interaction with known silencing mechanisms, will unveil key mechanisms underlying
322 the demarcation between soma and germline identity. These results will provide molecular footholds
323 that can then be exploited to test the contribution of ectopic germline genes on neurodevelopment.

324 In the above study, we demonstrate KDM5C's pivotal role in the development of tissue identity, resulting in
325 substantial misexpression of tissue-enriched genes within the adult *Kdm5c*-KO hippocampus and amygdala.
326 In addition to testis genes identified previously⁸, we found significant enrichment of muscle, liver, and
327 even ovary-biased genes aberrantly expressed within the male *Kdm5c*-KO brain. Given the majority of
328 tissue-enriched DEGs have no known brain functions, it is currently unknown if they impair *Kdm5c*-KO
329 neurodevelopment and contribute to MRXSCJ-related phenotypes like intellectual disability and aberrant
330 aggression. However, select liver and muscle-biased DEGs do have known roles within the brain, such as the
331 liver-enriched lipid metabolism gene *Apolipoprotein C-I (Apoc1)*³⁰ that is highly expressed in the *Kdm5c*-KO

332 amygdala and hippocampus. Intriguingly, *Apoc1* overexpression in the mouse brain can impair learning and
333 memory⁶⁵ and is implicated in Alzheimer's disease in humans³¹, however further investigation is required to
334 determine its impact upon *Kdm5c*-KO phenotypes.

335 Deeper characterization of testis-enriched DEGs revealed they were not somatic testis genes but instead
336 germline genes, thus demonstrating KDM5C's crucial role in establishing the soma-germline boundary.
337 Distinguishing the germline from the soma is a key feature of multicellularity and sexual reproduction¹⁴.
338 Previous work characterizing repressors of germ cell-specific transcription has predominately focused on
339 their repression of key marker genes in embryonic stem cells (ESCs), such as *Dazl* and *Ddx4*^{16,17,51}. To
340 characterize KDM5C's role in germline gene repression throughout life at a genome-wide level, we curated a
341 list of germline-enriched genes using publically available germ cell-depleted RNA-seq datasets from Kit^{W/Wv}
342 mice^{32,35}. This resource enabled us to identify 1) the extent of germline gene dysregulation in *Kdm5c*-KO
343 cells, 2) the types of germline genes misexpressed at different developmental time points, 3) the impact of
344 germ cell sex upon germline gene misexpression, and 4) which groups of germline genes are directly and
345 indirectly regulated by KDM5C. This list of germline-enriched genes serves as a resource for future studies
346 to systematically assess soma-germline dysregulation.

347 Analysis of epiblast-like cells (EpiLCs) revealed dysregulation of *Kdm5c*-KO tissue identity begins during
348 early emrbyogenesis. *In vivo*, germline genes are typically fully decommissioned in epiblast stem cells and
349 remain silenced as the epiblast differentiates into somatic tissues³⁸. However, a small subset of epiblast
350 stem cells will reactivate germline genes to become the primordial germ cells (PGCs) that will ultimately form
351 the mature germline^{36,37}. This process can be mimicked *in vitro* by differentiating EpiLCs into primordial
352 germ cell-like cells (PGCLCs)³⁹. Thus, misexpression of germline genes in *Kdm5c*-KO EpiLCs might suggest
353 ESCs are progressing beyond EpiLC differentiation and instead becoming PGCLCs. Yet, *Kdm5c*-KO EpiLCs
354 properly express genes for primed pluripotency, including *Otx2* - an epiblast stem cell marker that is known to
355 repress differentiation into PGCs/PGCLCs⁶⁶. Furthermore, we observed no difference in cellular morphology
356 during *Kdm5c*-KO ESC to EpiLC differentiation. Proper EpiLC differentiation, coupled with *Kdm5c*-KO
357 mice being viable, suggests germline gene expression is occurring ectopically in conjunction with typical
358 developmental programs, rather than a complete shift to towards germline identity.

359 While many genes involved in early germline development and meiosis are common between males
360 and females, some germline genes display sex-biased expression or perform functions unique to eggs and
361 sperm. We found both males and females expressed both sperm and egg biased genes, indicating the
362 mechanisms underlying germline gene repression are independent of germ cell sex. However, we found
363 organismal sex did greatly influence the manifestation of germline gene dysregulation, as *Kdm5c*-KO females
364 had over double the number of germline-enriched DEGs compared to male EpiLCs. Knockout of KDM5C in
365 females is embryonic lethal, likely due to impaired X chromosome inactivation (XCI)⁴⁵. XCI defects could
366 explain why *Kdm5c*-KO females are more prone to germline gene dysregulation, given the X chromosome
367 harbors a large number of spermatogenesis genes^{52,53}. However, ectopic germline transcripts, including

368 those unique to females, were not significantly biased towards the X chromosome. Sex differences in germline
369 gene suppression may be instead connected to females having a higher dose of KDM5C than males, due to
370 its partial escape from XCI^{20–23}. Intriguingly, females heterozygous for *Kdm5c* loss also had over double the
371 number of germline DEGs than males, even though their XCI is largely preserved and their expression of
372 KDM5C should be roughly equivalent to that of wild-type males. Altogether, this suggests female EpiLCs are
373 more prone to transitioning to a germ cell-like state than males and require more KDM5C to maintain somatic
374 cellular identity. Future studies investigating sex differences in germline gene repression are required to
375 illuminate if this phenomenon is unique to sexually dimorphic chromatin regulators like KDM5C or a general
376 feature of female cells.

377 Our analysis of KDM5C ChIP-seq data provided novel insights into direct and indirect mechanisms by
378 which germline genes can be misexpressed in *Kdm5c*-KO cells. While we observed high enrichment of
379 KDM5C peaks at germline genes in EpiLCs, KDM5C was not bound to germline gene promoters in PNCs.
380 This suggests dysregulation of germline genes in the mature *Kdm5c*-KO brain is due to loss of repression
381 during embryogenesis, which is consistent with previous work that found introducing human KDM5C into
382 *Kdm5c*-KO PNCs does not repress two aberrant germline transcripts⁸. It is unclear how KDM5C is recruited
383 to germline genes given that KDM5C itself does not contain domains for sequence-specific binding. In HeLa
384 cells and ESCs^{49,67}, KDM5C associates with members of the polycomb repressive complex 1.6 (PRC1.6),
385 which is recruited to germline gene promoters through E2F6/DP1 and MGA/MAX heterodimers^{16,55}. While
386 MAX and E2F6 motifs were enriched at KDM5C-bound promoters in EpiLCs, only about one third of KDM5C-
387 bound germline genes contained their consensus sequence at their promoter. Thus, other, unknown factors
388 likely facilitate KDM5C's recruitment to germline genes.

389 We also found germline genes can become activated in *Kdm5c*-KO cells independent of direct loss of
390 KDM5C suppression, given that KDM5C was only bound to about a third of germline-enriched DEGs. One
391 notable EpiLC DEG not bound by KDM5C is *Stra8*, a transcription factor activated by retinoic acid signaling in
392 germ cells to promote meiotic initiation^{68,69}. However, retinoic acid can only activate *Stra8* expression when
393 DAZL is present. Unlike *Stra8*, *Dazl* is a direct target of KDM5C in EpiLCs and is transcribed *Kdm5c*-KO
394 EpiLCs. Expression of indirect DEGs like *Stra8* indicates germline genes can be aberrantly expressed
395 in *Kdm5c*-KO cells through downstream activation by other ectopic germline programs. In support of
396 this, germline genes unbound by KDM5C were significantly enriched for RFX transcription factors. RFX
397 transcription factors bind to X-box motifs⁵⁷ to promote the transcription of cilia and flagellar genes^{58,59}. One
398 of the top RFX members enriched included RFX2, a central regulator of spermiogenesis^{60,61} whose mRNA
399 is also expressed in *Kdm5c*-KO EpiLCs. Intriguingly, *Kdm5c*-KO EpiLCs primarily expressed germline genes
400 involved in early germ cell development and meiosis while the mature *Kdm5c*-KO brain expressed late-stage
401 spermatogenesis genes, including known RFX2 targets. Altogether, these data suggest once activated due
402 to loss of direct KDM5C repression, ectopic germline programs can loosely mimic germline development and
403 progress in the background of *Kdm5c*-KO somatic development.

404 It is important to note that while we focused on KDM5C's regulation of germline genes, some germline-
405 enriched genes are also expressed in naïve ESCs and at the 2-cell stage due to their roles in pluripotency and
406 self-renewal. For example, *Dazl* is primarily known for committing PGCs to the germline fate and regulating
407 the translation of germline-specific RNAs, but is also expressed at the 2-cell stage⁷⁰, in naïve ESCs⁴¹, and in
408 the inner cell mass⁴¹. Very recently, two screens of *Dazl*-repressors identified KDM5C as a direct regulator in
409 ESCs^{49,71}. Interestingly, one study found *Kdm5c*-KO ESCs also aberrantly expressed 2-cell-specific genes,
410 indicating KDM5C negatively regulates totipotency⁴⁹. However, out of the four regulators characterized,
411 KDM5C was the only factor whose repression of *Dazl* was independent of the 2-cell specific transcription
412 factor *Dux*⁴⁹. Although expressed at the 2-cell stage and in ESCs, *Dazl* is typically silenced when ESCs
413 differentiate into EpiLCs¹⁷. We found KDM5C also represses *Dazl* in EpiLCs, but *Kdm5c*-KO EpiLCs did
414 not express 2-cell specific genes like *Dux* and *Zscan4c*. Together, this suggests the 2-cell-like state in
415 *Kdm5c*-KO ESCs is an indirect consequence of KDM5C's primary role in germline gene repression.

416 *Dazl* and other germline-enriched genes are initially decommissioned in ESCs by repressive histone
417 modifications and then *de novo* DNA CpG methylation (CpGme) in EpiLCs^{17–19,62,72,73}. Unlike the previously
418 characterized germline gene suppressors, KDM5C removes an active histone mark, histone 3 lysine 4 di-
419 and trimethylation (H3K4me2/3)⁹. In this study, we observed a global increase in H3K4me2/3 in *Kdm5c*-KO
420 EpiLCs and amygdala around the transcription start site (TSS) of germline-enriched genes. However,
421 KDM5C's catalytic activity may not be required for germline gene silencing, as it was recently found to be
422 dispensible for repressing *Dazl* in ESCs⁴⁹. Although not necessary in ESCs, KDM5C's catalytic activity be
423 required for CpGme in EpiLCs, as H3K4me3 impedes *de novo* CpGme placement^{63,64}. This is supported
424 by previous work in the *Kdm5c*-KO adult hippocampus, which found CpGme is significantly eroded at at
425 least two germline promoters⁸. To elucidate the mechanism behind KDM5C-mediated silencing of germline
426 genes, we characterized KDM5C in ESCs and EpiLCs and found its expression is dynamically regulated
427 during differentiation. Furthermore, KDM5C acts during the transition from naïve to primed pluripotency to
428 promote the initial placement of CpGme at CpG islands for the long-term silencing of germline genes.

429 • CPG islands greatly determine KDM5C recruitment - KDM5C is previously known to be enriched at
430 CGIs. CGIs not typically methylated, germline genes exception. KDM5C promotes this process.

431 The above work provides the mechanistic foundation for KDM5C's role in germline gene repression
432 and offers novel insight into how dysregulation in *Kdm5c*-KO tissue identity changes over the course of
433 development. However, the contribution of ectopic, tissue-specific genes towards *Kdm5c*-KO neurological
434 impairments is still unknown. KDM5C may be crucial for neurodevelopment by fine-tuning the expression of
435 tissue-enriched, dosage-sensitive genes, such as *Apoc1*. KDM5C could also ensure tissue-specific timing of
436 transcription factor family members, like RFX genes that are broadly required for cilia and flagella formation,
437 including neuronal cilia^{58,59,74,75}. While their impact upon neurodevelopment is currently unclear, ectopic
438 germline transcripts are also found in models of other related neurodevelopmental disorders⁷⁶, including

439 Immunodeficiency, Centromere instability, Facial anomalies syndrome 1 (ICF1 OMIM: #242860)^{77,78}, Kleefstra
440 syndrome 1 (OMIM: #610253)⁷, and MeCP2 duplication syndrome (MDS, OMIM: #300260)⁷⁹. Like KDM5C,
441 the chromatin regulators underlying these conditions, DNA methyltransferase 3b (DNMT3B), H3K9me1/2
442 methyltransferases G9A/GLP, methyl-CpG -binding protein 2 (MECP2), primarily silence gene expression.
443 Thus, KDM5C is among a growing cohort of chromatin-linked neurodevelopmental disorders that have
444 a similar underlying phenotype of germline versus soma dysregulation. Further research is required to
445 determine the ultimate impact of these germline genes and the extent to which this phenomenon occurs in
446 humans.

447 Materials and Methods

448 Classifying tissue-enriched and germline-enriched genes

449 Tissue-enriched differentially expresssd genes (DEGs) were determined by their classification in a previ-
450 ously published dataset from 17 male and female mouse tissues²⁴. This study defined tissue expression as
451 greater than 1 Fragments Per Kilobase of transcript per Million mapped read (FPKM) and tissue enrichment
452 as at least 4-fold higher expression than any other tissue.

453 We curated a list of germline-enriched genes using an RNA-seq dataset from wild-type and germline-
454 depleted (Kit^{W/W^v}) male and female mouse embryos from embryonic day 12, 14, and 16³⁵, as well as adult
455 male testes³². Germline-enriched genes met the following criteria: 1) their expression is greater than 1
456 FPKM in wild-type germline 2) their expression in any wild-type somatic tissues²⁴ does not exceed 20%
457 of maximum expression in wild-type germline, and 3) their expression in the germ cell-depleted (Kit^{W/W^v})
458 germline, for any sex or time point, does not exceed 20% of maximum expression in wild-type germline.

459 Cell culture

460 We utilized our previously established cultures of male wild-type and *Kdm5c* knockout (-KO) embryonic
461 stem cells⁴⁵. Sex was confirmed by genotyping *Uba1/Uba1y* on the X and Y chromosomes with the following
462 primers: 5'-TGGATGGTGTGCCAATG-3', 5'-CACCTGCACGTTGCCCTT-3'. Deletion of *Kdm5c* was
463 confirmed through the primers 5'-ATGCCCATATTAAAGAGTCCTG-3', 5'-TCTGCCTTGATGGACTGTT-3',
464 and 5'-GGTTCTCAACACTCACATAGTG-3'.

465 Embryonic stem cells (ESCs) and epiblast-like cells were cultured using previously established
466 methods⁴⁰. Briefly, ESCs were initially cultured in primed ESC (pESC) media consisting of KnockOut
467 DMEM (Gibco#10829–018), fetal bovine serum (Gibco#A5209501), KnockOut serum replacement
468 (Invitrogen#10828–028), Glutamax (Gibco#35050-061), Anti-Anti (Gibco#15240-062), MEM Non-essential
469 amino acids (Gibco#11140-050), and beta-mercaptoethanol (Sigma#M7522). They were then transitioned
470 into ground-state, “naïve” ESCs (nESCs) by culturing for four passages in N2B27 media containing

471 DMEM/F12 (Gibco#11330–032), Neurobasal media (Gibco#21103–049), Gluamax, Anti-Anti, N2 sup-
472 plement (Invitrogen#17502048), and B27 supplement without vitamin A (Invitrogen#12587-010), and
473 beta-mercaptoethanol. Both pESC and nESC media were supplemented with 3 μ M GSK3 inhibitor
474 CHIR99021 (Sigma #SML1046-5MG), 1 μ M MEK inhibitor PD0325901 (Sigma #PZ0162-5MG), and 1,000
475 units/mL leukemia inhibitory factor (LIF, Millipore#ESG1107).

476 nESCs were differentiated into epiblast-like cells (EpiLCs, 48 hours) and extendend EpiLCs (exEpiLCs,
477 96 hours) by culturing in N2B27 media containing DMEM/F12, Neurobasal media, Gluamax, Anti-Anti,
478 N2 supplement, B27 supplement (Invitrogen#17504044), beta-mercaptoethanol, fibroblast growth factor 2
479 (FGF2, R&D Biotechne 233-FB), and activin A (R&D Biotechne 338AC050CF), as previously described⁴⁰.

480 **Immunocytochemistry (ICC)**

481 ICC of DAZL in EpiLCs was performed by first growing cells on fibronectin-coated coverslips. Cells were
482 then washed thrice with phosphobuffered saline (PBS), fixed in 4% paraformaldehyde, washed thrice in PBS,
483 and blacked in PBS containing 0.3% Triton X-100, and 5% fetal bovine serum for 1 hour. They were then
484 washed thrice with PBS and incubating in primary antibody (Rabbit anti DAZL, abcam ab34139, 1:200) in
485 the blocking solution overnight at 4°C with gentle rocking. The next day, cells were rinsed thrice with PBS,
486 and incubated in secondary antibody (Alexaflouor 488 Invitrogen #710369, 1:1,000) with DAPI (1:1,000) in
487 blocking buffer for 1 hour at room temperature. Coverslips were then rinsed thrice in PBS, and mounted onto
488 slides using Prolong Gold (Invitrogen #P36930). Images were taken blinded for genotype, chosen based on
489 similar levels of DAPI signal, and quantified via ImageJ before unblinding.

490 **RNA sequencing (RNA-seq)**

491 After ensuring read quality via FastQC (v0.11.8), reads were then mapped to the mm10 *Mus musculus*
492 genome (Gencode) using STAR (v2.5.3a), during which we removed duplicates and kept only uniquely
493 mapped reads. Count files were generated by FeatureCounts (Subread v1.5.0), and BAM files were
494 converted to bigwigs using deeptools (v3.1.3) and visualized by the UCSC genome browser. RStudio (v3.6.0)
495 was then used to analyze counts files by DESeq2 (v1.26.0)²⁶ to identify differentially expressed genes
496 (DEGs) with a q-value (p-adjusted via FDR/Benjamini–Hochberg correction) less than 0.1 and a log2 fold
497 change greater than 0.5. For all DESeq2 analyses, log2 fold changes were calculated with IfcShrink using
498 the ashR package⁸⁰. MA-plots were generated by ggpubr (v0.6.0), and Eulerr diagrams were generated by
499 eulerr (v6.1.1). Boxplots and scatterplots were generated by ggpubr (v0.6.0) and ggplot2 (v3.3.2). The Upset
500 plot was generated via the package UpSetR (v1.4.0)⁸¹. Gene ontology (GO) analyses were performed by
501 the R package enrichPlot (v1.16.2) using the biological processes setting and compareCluster.

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

503 ChIP-seq reads were aligned to mm10 using Bowtie1 (v1.1.2) allowing up to two mismatches. Only
504 uniquely mapped reads were used for analysis. Peaks were called using MACS2 software (v2.2.9.1) using
505 input BAM files for normalization, with filters for a q-value < 0.1 and a fold enrichment > 1. We removed
506 “black-listed” genomic regions that often give aberrant signals. Common peak sets were obtained in R via
507 DiffBind (v3.6.5). In the case of KDM5C ChIP-seq, *Kdm5c*-KO peaks were then subtracted from wild-type
508 samples using bedtools (v2.25.0). Peak proximity to genome annotations was determined by ChIPSeeker
509 (v1.32.1). Gene ontology (GO) analyses were performed by the R package enrichPlot (v1.16.2) using the
510 biological processes setting and compareCluster. Enriched motifs were identified using HOMER⁵⁶. Average
511 binding across the genome was visualized using deeptools (v3.1.3). Bigwigs were visualized using the
512 UCSC genome browser.

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

514 Genomic DNA (gDNA) from naïve ESCs and extended EpiLCs was extracted using the Wizard Genomic
515 DNA Purification Kit (Promega A1120), following the instructions for Tissue Culture Cells. gDNA was sent
516 to Novogene for WGBS using the Illumina NovaSeq X Plus platform and sequenced for 150bp paired-end
517 reads (PE150).

518 **Data availability**

519 **Published datasets**

520 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
521 adult amygdala and hippocampus²⁵ (available at GEO: GSE127722) and male wild-type and *Kdm5c*-KO
522 EpiLCs⁴⁵ (available at GEO: GSE96797).

523 Previously published ChIP-seq experiments included KDM5C in wild-type and *Kdm5c*-KO EpiLCs⁴⁵ (avail-
524 able at GEO: GSE96797) and mouse primary neuron cultures (PNCs) from the cortex and hippocampus¹³
525 (available at GEO: GSE61036). ChIP-seq of histone 3 lysine 4 dimethylation in male wild-type and *Kdm5c*-KO
526 EpiLCs⁴⁵ is also available at GEO: GSE96797. ChIP-seq of histone 3 lysine 4 trimethylation in wild-type and
527 *Kdm5c*-KO male amygdala²⁵ are available at GEO: GSE127817.

529 **Data analysis**

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

532 **Acknowledgements**

- 533 • Jacob Mueller for providing insight in germline gene regulation.
534 • Sundeep Kalantry for providing reagents and expertise in culturing mouse embryonic stem cells and
535 epiblast-like cells
536 • Ilakkia
537 • Funding sources

538 **References**

- 539 1. Strahl, B.D., and Allis, C.D. (2000). The language of covalent histone modifications. *Nature* **403**,
540 41–45. <https://doi.org/10.1038/47412>.
- 541 2. Jenuwein, T., and Allis, C.D. (2001). Translating the Histone Code. *Science* **293**, 1074–1080.
542 <https://doi.org/10.1126/science.1063127>.
- 543 3. Lewis, E.B. (1978). A gene complex controlling segmentation in *Drosophila*. *Nature* **276**, 565–570.
544 <https://doi.org/10.1038/276565a0>.
- 545 4. Kennison, J.A., and Tamkun, J.W. (1988). Dosage-dependent modifiers of polycomb and antennapedia
546 mutations in *Drosophila*. *Proc. Natl. Acad. Sci. U.S.A.* **85**, 8136–8140. <https://doi.org/10.1073/pnas.8>
5.21.8136.
- 547 5. Kassis, J.A., Kennison, J.A., and Tamkun, J.W. (2017). Polycomb and Trithorax Group Genes in
548 *Drosophila*. *Genetics* **206**, 1699–1725. <https://doi.org/10.1534/genetics.115.185116>.
- 549 6. Gabriele, M., Lopez Tobon, A., D'Agostino, G., and Testa, G. (2018). The chromatin basis of
550 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>.
- 551 7. Schaefer, A., Sampath, S.C., Intrator, A., Min, A., Gertler, T.S., Surmeier, D.J., Tarakhovsky, A.,
552 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>.
- 553 8. Scandaglia, M., Lopez-Atalaya, J.P., Medrano-Fernandez, A., Lopez-Cascales, M.T., Del Blanco, B.,
554 Lipinski, M., Benito, E., Olivares, R., Iwase, S., Shi, Y., et al. (2017). Loss of Kdm5c Causes Spurious
Transcription and Prevents the Fine-Tuning of Activity-Regulated Enhancers in Neurons. *Cell Rep* **21**,
47–59. <https://doi.org/10.1016/j.celrep.2017.09.014>.
- 555 9. 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.

- 557 10. 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
Med Genet* 94, 1–4.
- 558
- 559 11. 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
Genet* 76, 227–236. <https://doi.org/10.1086/427563>.
- 560
- 561 12. 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>.
- 562
- 563 13. 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).
- 564
- 565 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>.
- 566
- 567 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>.
- 568
- 569 16. Endoh, M., Endo, T.A., Shinga, J., Hayashi, K., Farcas, A., Ma, K.W., Ito, S., Sharif, J., Endoh, T.,
Onaga, N., et al. (2017). PCGF6-PRC1 suppresses premature differentiation of mouse embryonic
stem cells by regulating germ cell-related genes. *eLife* 6. <https://doi.org/10.7554/eLife.21064>.
- 570
- 571 17. Mochizuki, K., Sharif, J., Shirane, K., Uranishi, K., Bogutz, A.B., Janssen, S.M., Suzuki, A., Okuda,
A., Koseki, H., and Lorincz, M.C. (2021). Repression of germline genes by PRC1.6 and SETDB1
in the early embryo precedes DNA methylation-mediated silencing. *Nat Commun* 12, 7020. <https://doi.org/10.1038/s41467-021-27345-x>.
- 572
- 573 18. Velasco, G., Hubé, F., Rollin, J., Neuillet, D., Philippe, C., Bouzinba-Segard, H., Galvani, A., Viegas-
Péquignot, E., and Francastel, C. (2010). Dnmt3b recruitment through E2F6 transcriptional repressor
mediates germ-line gene silencing in murine somatic tissues. *Proc Natl Acad Sci U S A* 107, 9281–
9286. <https://doi.org/10.1073/pnas.1000473107>.
- 574
- 575 19. Hackett, J.A., Reddington, J.P., Nestor, C.E., Duncan, 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>.

- 576
- 577 20. 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
578 and human. *Hum Mol Genet* 3, 879–884. <https://doi.org/10.1093/hmg/3.6.879>.
- 579 21. Carrel, L., Hunt, P.A., and Willard, H.F. (1996). Tissue and lineage-specific variation in inactive
580 X chromosome expression of the murine Smcx gene. *Hum Mol Genet* 5, 1361–1366. <https://doi.org/10.1093/hmg/5.9.1361>.
- 581 22. Sheardown, S., Norris, D., Fisher, A., and Brockdorff, N. (1996). The mouse Smcx gene exhibits
582 developmental and tissue specific variation in degree of escape from X inactivation. *Hum Mol Genet*
5, 1355–1360. <https://doi.org/10.1093/hmg/5.9.1355>.
- 583 23. 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>.
- 584 24. 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.
586 <https://doi.org/10.1038/s41598-017-04520-z>.
- 587 25. 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>.
- 589 26. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for
590 RNA-seq data with DESeq2. *Genome Biol* 15, 550. <https://doi.org/10.1186/s13059-014-0550-8>.
- 591 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
592 Chromosome Pairing in Meiosis. *Science* 300, 1291–1295. <https://doi.org/10.1126/science.1083022>.
- 593 28. Xioli, 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>.
- 595 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
596 membraneless compartment. *Science* 378, eabq4835. <https://doi.org/10.1126/science.abq4835>.
- 597 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.
Cardiovasc Diabetol 21, 272. <https://doi.org/10.1186/s12933-022-01703-5>.

598

- 599 31. Leduc, V., Jasmin-Bélanger, S., and Poirier, J. (2010). APOE and cholesterol homeostasis in
600 Alzheimer's disease. *Trends in Molecular Medicine* *16*, 469–477. <https://doi.org/10.1016/j.molmed.2010.07.008>.
- 601 32. Mueller, J.L., Skaletsky, H., Brown, L.G., Zaghlul, S., Rock, S., Graves, T., Auger, K., Warren,
602 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>.
- 603 33. 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>.
- 604 34. Green, C.D., Ma, Q., Manske, G.L., Shami, A.N., Zheng, X., Marini, S., Moritz, L., Sultan, C.,
605 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>.
- 606 35. 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.
607 <https://doi.org/10.1371/journal.pgen.1005531>.
- 608 36. Magnúsdóttir, E., and Surani, M.A. (2014). How to make a primordial germ cell. *Development* *141*,
609 245–252. <https://doi.org/10.1242/dev.098269>.
- 610 37. 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*.
611 <https://doi.org/10.1098/rstb.2013.0543>.
- 612 38. 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>.
- 613 39. 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.
614 <https://doi.org/10.1016/j.cell.2011.06.052>.
- 615 40. Samanta, M., and Kalantry, S. (2020). Generating primed pluripotent epiblast stem cells: A methodol-
616 ogy chapter. *Curr Top Dev Biol* *138*, 139–174. <https://doi.org/10.1016/bs.ctdb.2020.01.005>.
- 617 41. Welling, M., Chen, H., Muñoz, J., Musheev, M.U., Kester, L., Junker, J.P., Mischerikow, N., Arbab, M.,
618 Kuijk, E., Silberstein, L., et al. (2015). DAZL regulates Tet1 translation in murine embryonic stem cells.
619 *EMBO Reports* *16*, 791–802. <https://doi.org/10.15252/embr.201540538>.
- 620

- 621 42. Macfarlan, T.S., Gifford, W.D., Driscoll, S., Lettieri, K., Rowe, H.M., Bonanomi, D., Firth, A., Singer, O.,
622 Trono, D., and Pfaff, S.L. (2012). Embryonic stem cell potency fluctuates with endogenous retrovirus
activity. *Nature* *487*, 57–63. <https://doi.org/10.1038/nature11244>.
- 623 43. Suzuki, A., Hirasaki, M., Hishida, T., Wu, J., Okamura, D., Ueda, A., Nishimoto, M., Nakachi, Y.,
624 Mizuno, Y., Okazaki, Y., et al. (2016). Loss of MAX results in meiotic entry in mouse embryonic and
germline stem cells. *Nat Commun* *7*, 11056. <https://doi.org/10.1038/ncomms11056>.
- 625 44. Borgel, J., Guibert, S., Li, Y., Chiba, H., Schübeler, D., Sasaki, H., Forné, T., and Weber, M. (2010).
626 Targets and dynamics of promoter DNA methylation during early mouse development. *Nat Genet* *42*,
1093–1100. <https://doi.org/10.1038/ng.708>.
- 627 45. Samanta, M.K., Gayen, S., Harris, C., Maclary, E., Murata-Nakamura, Y., Malcore, R.M., Porter, R.S.,
628 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>.
- 629 46. 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*,
630 2474–2479. <https://doi.org/10.1073/pnas.0510813103>.
- 631 47. Lin, Y., Gill, M.E., Koubova, J., and Page, D.C. (2008). Germ Cell-Intrinsic and -Extrinsic Factors
632 Govern Meiotic Initiation in Mouse Embryos. *Science* *322*, 1685–1687. <https://doi.org/10.1126/science.1166340>.
- 633 48. Endo, T., Mikedis, M.M., Nicholls, P.K., Page, D.C., and De Rooij, D.G. (2019). Retinoic Acid and Germ
634 Cell Development in the Ovary and Testis. *Biomolecules* *9*, 775. <https://doi.org/10.3390/biom9120775>.
- 635 49. 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.
636 *Nat Struct Mol Biol*. <https://doi.org/10.1038/s41594-023-01038-z>.
- 637 50. 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-
638 Specific Gene Expression. *Current Biology* *15*, 1051–1057. <https://doi.org/10.1016/j.cub.2005.04.060>.
- 639 51. 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
640 during embryonic development. *Nat Commun* *12*, 3582. <https://doi.org/10.1038/s41467-021-23596-w>.
- 641 52. Wang, P.J., McCarrey, J.R., Yang, F., and Page, D.C. (2001). An abundance of X-linked genes
642 expressed in spermatogonia. *Nat Genet* *27*, 422–426. <https://doi.org/10.1038/86927>.

- 643 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>.
- 644
- 645 54. Hurlin, P.J. (1999). Mga, a dual-specificity transcription factor that interacts with Max and contains a
T-domain DNA-binding motif. *The EMBO Journal* 18, 7019–7028. <https://doi.org/10.1093/emboj/18.2.4.7019>.
- 646
- 647 55. 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>.
- 648
- 649 56. 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>.
- 650
- 651 57. 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>.
- 652
- 653 58. 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).
- 654
- 655 59. 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>.
- 656
- 657 60. 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>.
- 658
- 659 61. 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.
<https://doi.org/10.1038/srep20435>.
- 660
- 661 62. Liu, M., Zhu, Y., Xing, F., Liu, S., Xia, Y., Jiang, Q., and Qin, J. (2020). The polycomb group protein
PCGF6 mediates germline gene silencing by recruiting histone-modifying proteins to target gene
promoters. *J Biol Chem* 295, 9712–9724. <https://doi.org/10.1074/jbc.RA119.012121>.
- 662
- 663 63. 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
domain. *EMBO Reports* 10, 1235–1241. <https://doi.org/10.1038/embor.2009.218>.
- 664

- 665 64. Guo, X., Wang, L., Li, J., Ding, Z., Xiao, J., Yin, X., He, S., Shi, P., Dong, L., Li, G., et al. (2015).
666 Structural insight into autoinhibition and histone H3-induced activation of DNMT3A. *Nature* **517**,
640–644. <https://doi.org/10.1038/nature13899>.
- 667 65. Abildayeva, K., Berbée, J.F.P., Blokland, A., Jansen, P.J., Hoek, F.J., Meijer, O., Lütjohann, D., Gautier,
668 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>.
- 669 66. Zhang, J., Zhang, M., Acampora, D., Vojtek, M., Yuan, D., Simeone, A., and Chambers, I. (2018).
670 OTX2 restricts entry to the mouse germline. *Nature* **562**, 595–599. <https://doi.org/10.1038/s41586-018-0581-5>.
- 671 67. Tahiliani, M., Mei, P., Fang, R., Leonor, T., Rutenberg, M., Shimizu, F., Li, J., Rao, A., and Shi, Y.
672 (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>.
- 673 68. Endo, T., Romer, K.A., Anderson, E.L., Baltus, A.E., De Rooij, D.G., and Page, D.C. (2015). Pe-
674 riodic retinoic acid–STRA8 signaling intersects with periodic germ-cell competencies to regulate
spermatogenesis. *Proc. Natl. Acad. Sci. U.S.A.* **112**. <https://doi.org/10.1073/pnas.1505683112>.
- 675 69. Bowles, J., Knight, D., Smith, C., Wilhelm, D., Richman, J., Mamiya, S., Yashiro, K., Chawengsak-
676 sophak, K., Wilson, M.J., Rossant, J., et al. (2006). Retinoid Signaling Determines Germ Cell Fate in
Mice. *Science* **312**, 596–600. <https://doi.org/10.1126/science.1125691>.
- 677 70. Yang, M., Yu, H., Yu, X., Liang, S., Hu, Y., Luo, Y., Izsvák, Z., Sun, C., and Wang, J. (2022). Chemical-
678 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>.
- 679 71. Al Adhami, H., Vallet, J., Schaal, C., Schumacher, P., Bardet, A.F., Dumas, M., Chicher, J., Hammann,
680 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>.
- 681 72. Karimi, M.M., Goyal, P., Maksakova, I.A., Bilenky, M., Leung, D., Tang, J.X., Shinkai, Y., Mager, D.L.,
682 Jones, S., Hirst, M., et al. (2011). DNA Methylation and SETDB1/H3K9me3 Regulate Predominantly
Distinct Sets of Genes, Retroelements, and Chimeric Transcripts in mESCs. *Cell Stem Cell* **8**,
676–687. <https://doi.org/10.1016/j.stem.2011.04.004>.
- 683 73. Tatsumi, D., Hayashi, Y., Endo, M., Kobayashi, H., Yoshioka, T., Kiso, K., Kanno, S., Nakai, Y., Maeda,
684 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>.

- 685 74. Chung, M.-I., Peyrot, S.M., LeBoeuf, S., Park, T.J., McGary, K.L., Marcotte, E.M., and Wallingford,
686 J.B. (2012). RFX2 is broadly required for ciliogenesis during vertebrate development. *Developmental
Biology* 363, 155–165. <https://doi.org/10.1016/j.ydbio.2011.12.029>.
- 687 75. Coschiera, A., Yoshihara, M., Lauter, G., Ezer, S., Pucci, M., Li, H., Kavšek, A., Riedel, C.G., Kere, J.,
688 and Swoboda, P. (2024). Primary cilia promote the differentiation of human neurons through the WNT
signaling pathway. *BMC Biol* 22, 48. <https://doi.org/10.1186/s12915-024-01845-w>.
- 689 76. Bonefas, K.M., and Iwase, S. (2021). Soma-to-germline transformation in chromatin-linked neurode-
690 velopmental disorders? *FEBS J*. <https://doi.org/10.1111/febs.16196>.
- 691 77. Velasco, G., Walton, E.L., Sterlin, D., Hédouin, S., Nitta, H., Ito, Y., Fouyssac, F., Mégarbané, A.,
692 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>.
- 693 78. Walton, E.L., Francastel, C., and Velasco, G. (2014). Dnmt3b Prefers Germ Line Genes and Cen-
694 tromeric Regions: Lessons from the ICF Syndrome and Cancer and Implications for Diseases. *Biology
(Basel)* 3, 578–605. <https://doi.org/10.3390/biology3030578>.
- 695 79. Samaco, R.C., Mandel-Brehm, C., McGraw, C.M., Shaw, C.A., McGill, B.E., and Zoghbi, H.Y. (2012).
696 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>.
- 697 80. Stephens, M. (2016). False discovery rates: A new deal. *Biostat*, kxw041. [https://doi.org/10.1093/biostatistics/kxw041](https://doi.org/10.1093/bi
698 ostatistics/kxw041).
- 699 81. Conway, J.R., Lex, A., and Gehlenborg, N. (2017). UpSetR: An R package for the visualization of
700 intersecting sets and their properties. *Bioinformatics* 33, 2938–2940. [https://doi.org/10.1093/bioinformatics/btx364](https://doi.org/10.1093/bioinfor
matics/btx364).
- 701 82. Li, H., Liang, Z., Yang, J., Wang, D., Wang, H., Zhu, M., Geng, B., and Xu, E.Y. (2019). DAZL
702 is a master translational regulator of murine spermatogenesis. *Natl Sci Rev* 6, 455–468. [https:
//doi.org/10.1093/nsr/nwy163](https://doi.org/10.1093/nsr/nwy163).
- 703 83. Mikedis, M.M., Fan, Y., Nicholls, P.K., Endo, T., Jackson, E.K., Cobb, S.A., De Rooij, D.G., and Page,
704 D.C. (2020). DAZL mediates a broad translational program regulating expansion and differentiation of
spermatogonial progenitors. *eLife* 9, e56523. <https://doi.org/10.7554/eLife.56523>.

705 **Figures and Tables**

- 706 • Supplementary table 1: list of all germline genes.
- 707 – Columns to include:
- 708 * KDM5C bound vs not
- 709 * DEG in EpiLC, brain, both, neither (separate columns?)



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



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

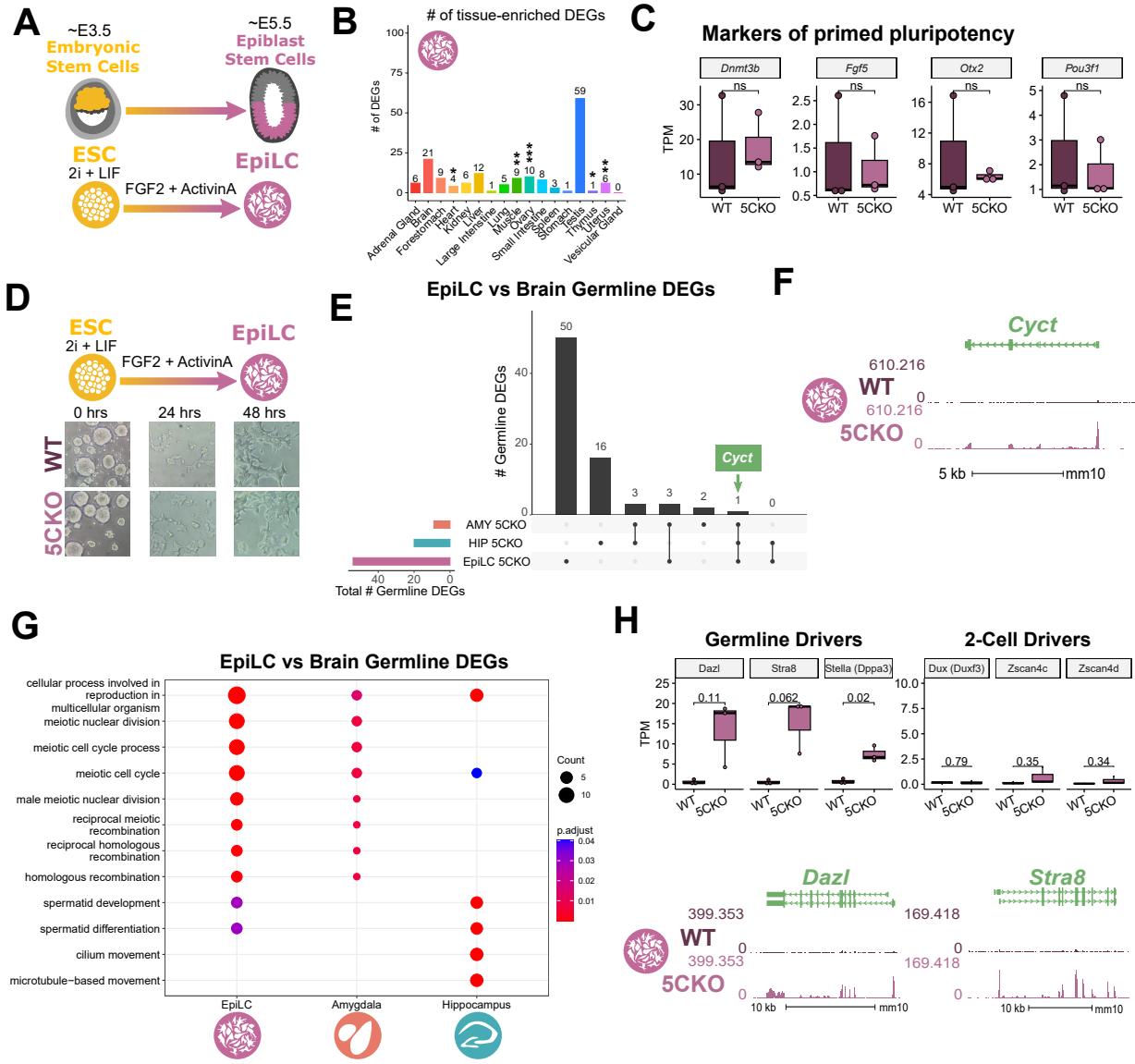


Figure 3: *Kdm5c*-KO epiblast-like cells express key drivers of germline identity

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

B. 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.

C. Average bigwigs of an example germline gene, *Cyct*, that is dysregulated *Kdm5c*-KO EpiLCs.

D. 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).

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

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

G. enrichPlot comparing enriched biological process gene ontologies for *Kdm5c*-KO EpiLC, amygdala, and hippocampus germline DEGs.

H. Top left - Example germline identity DEGs unique to EpiLCs. Top right - Example 2-cell genes that are not dysregulated in *Kdm5c*-KO EpiLCs. p-values for Welch's t-test. Bottom - Average bigwigs of *Dazl* and *Stra8* expression in wild-type and *Kdm5c*-KO EpiLCs. Immunocytochemistry of DAZL in male wild-type (WT) and *Kdm5c*-KO (5CKO) EpiLCs. Percentage of DAZL-positive cells normalized to number of DAPI-positive cells, p-value for Welch's t-test.

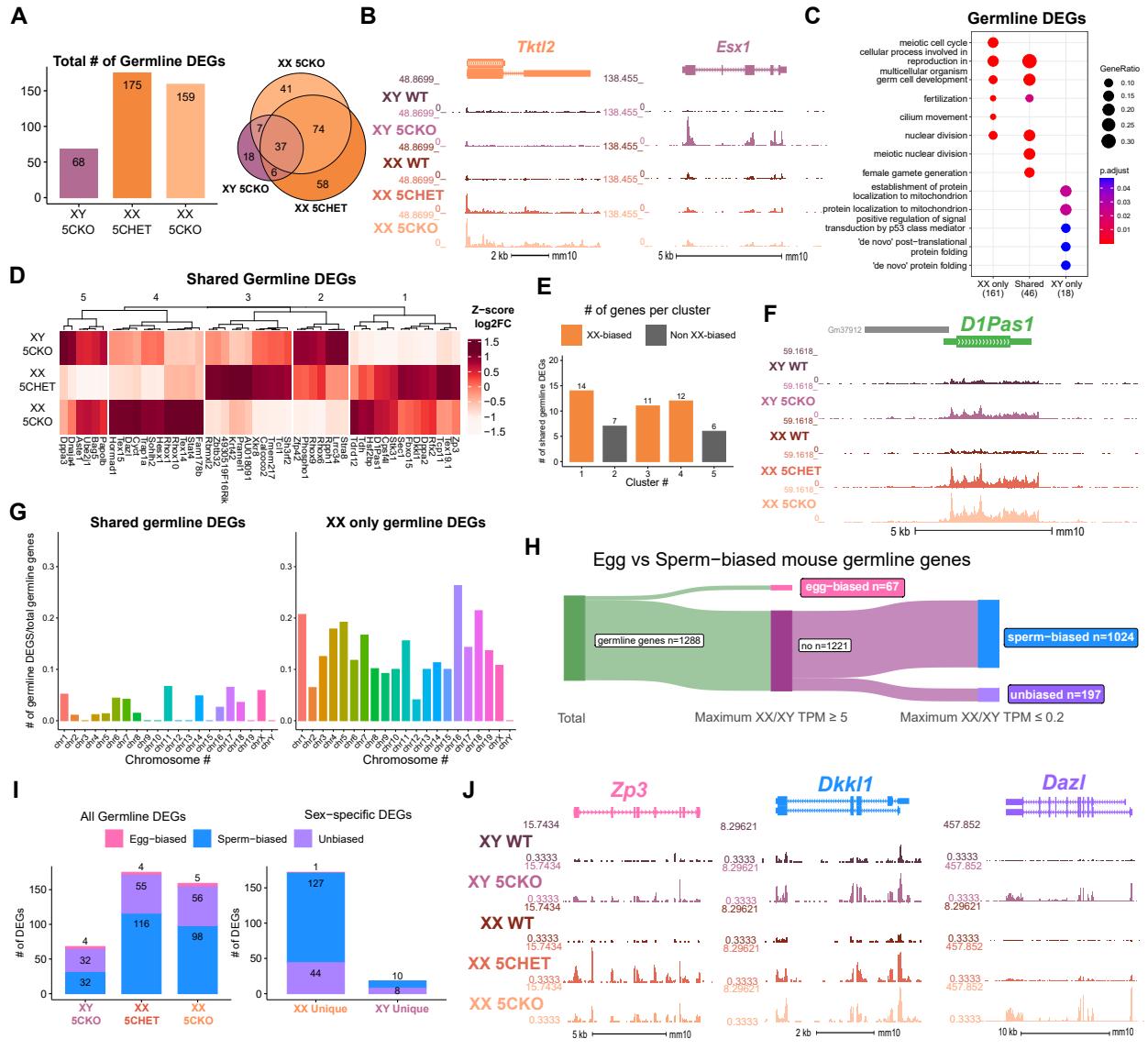


Figure 4: Chromosomal sex influences *Kdm5c*-KO germline gene misexpression. **A-B.** Expression of tissue-enriched genes (Li et al 2017) in the male *Kdm5c*-KO amygdala (A) and hippocampus (B). Left - MA plot of mRNA-seq. Right - Number of tissue-enriched differentially expressed genes (DEGs). * $p<0.05$, ** $p<0.01$, *** $p<0.001$, Fisher's exact test. **C.** Left - Average bigwigs of an example aberrantly expressed testis-enriched DEG, *FK506 binding protein 6* (*Fkbp6*) in the wild-type (WT) and *Kdm5c*-KO (5CKO) amygdala (red) and hippocampus (teal). Right - Expression of *Cyct* in wild-type tissues from NCBI Gene, with testis highlighted in blue and brain tissues highlighted in red. **D.** Left - Average bigwigs of an example ovary-enriched DEG, *Zygotic arrest 1* (*Zar1*). Right - Expression of *Zar1* in wild-type tissues from NCBI Gene, with ovary highlighted in teal and brain tissues highlighted in red. **E.** Left - Average bigwigs of an example liver-enriched DEG, *Apolipoprotein C-I* (*Apoc1*). Right - Expression of *Apoc1* in wild-type tissues from NCBI Gene, with liver highlighted in orange and brain tissues highlighted in red.

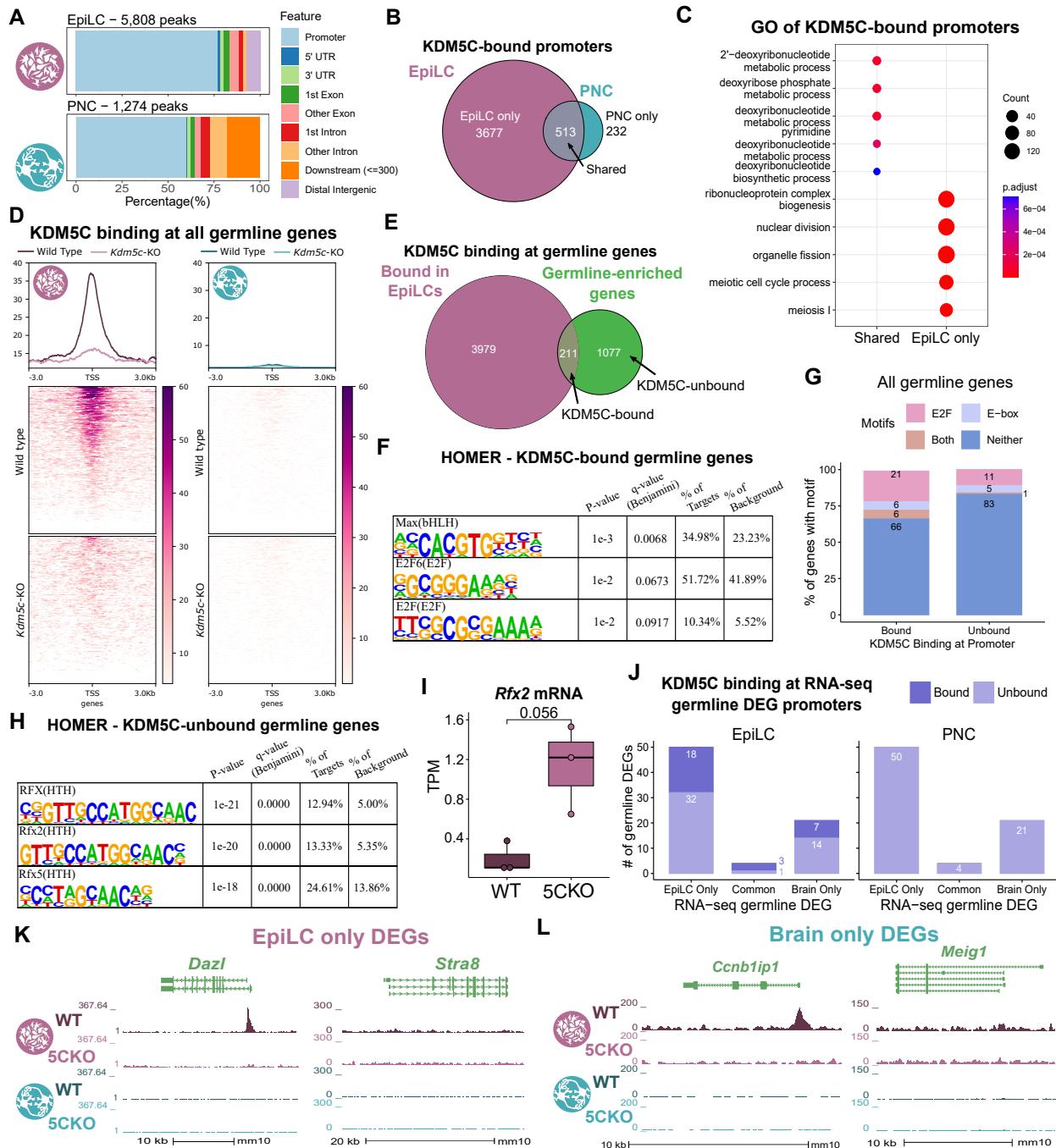
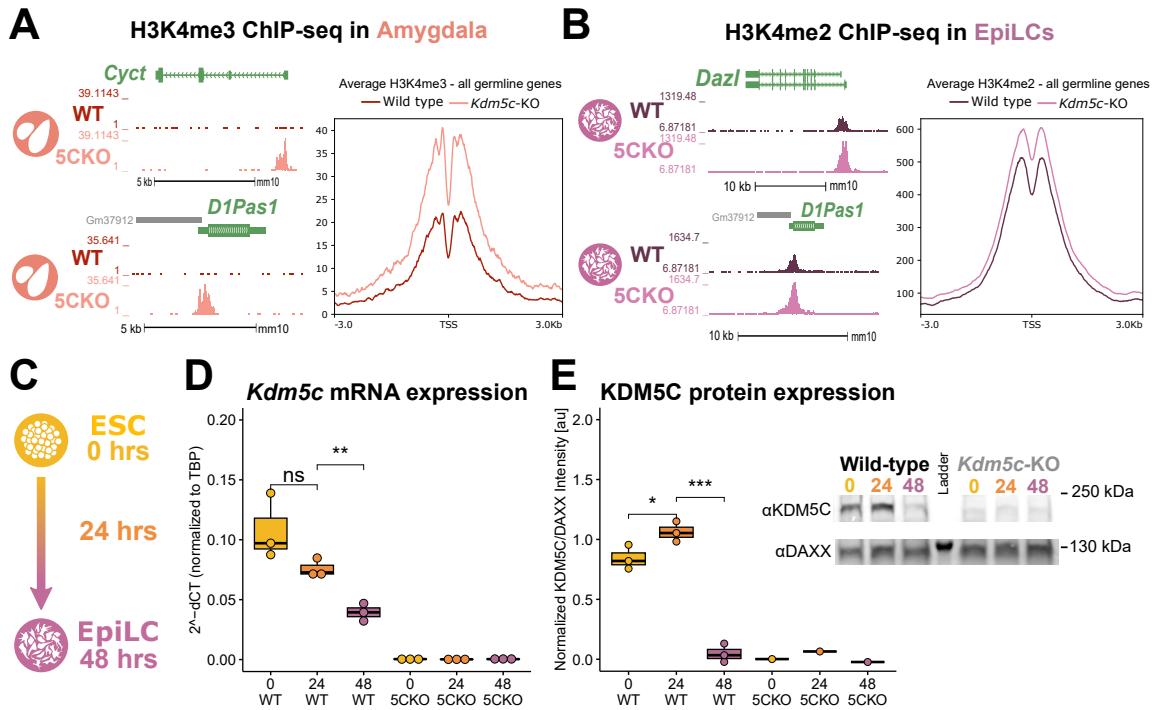


Figure 5: KDM5C binds to a subset of germline gene promoters during early embryogenesis **A.** ChIPseeker localization of KDM5C peaks at different genomic regions in EpiLCs (top) and hippocampal and cortex primary neuron cultures (PNCs, bottom). **B.** Overlap of genes with KDM5C bound to their promoters 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.** KDM5C binding at the promoters of RNA-seq germline DEGs. Genes were classified as either only dysregulated in EpiLCs (EpiLC only), genes dysregulated in the hippocampus or amygdala but not EpiLCs (brain only), or genes dysregulated in both EpiLCs and the brain (common). (Legend continued on next page.)

Figure 5: KDM5C binds to a subset of germline gene promoters during early embryogenesis. (Legend continued.) **F.** Example KDM5C ChIP-seq bigwigs of DEGs unique to EpiLCs. Although both are expressed in *Kdm5c*-KO EpiLCs, KDM5C is only bound to the *Dazl* promoter and not the *Stra8* promoter in EpiLCs. **G.** Example KDM5C ChIP-seq bigwigs of RNA-seq DEGs common between the brain and EpiLCs. KDM5C is bound to the *D1Pas1* promoter but not the *XXX* promoter in EpiLCs. **H.** Example KDM5C ChIP-seq bigwigs of RNA-seq DEGs unique to the brain. KDM5C is bound to the *XXX* promoter but not the *Meig1* promoter



F

G

Figure 6: KDM5C's catalytic activity promotes long-term silencing of germline genes via DNA methylation. **A.** Left - Bigwigs of representative histone 3 lysine 4 trimethylation (H3K4me3) ChIP-seq peaks at two germline genes in the wild-type (WT) and *Kdm5c*-KO (5CKO) adult amygdala. Right - Average H3K4me3 at the transcripton start site (TSS) of all germline-enriched genes in wild-type (dark red) and *Kdm5c*-KO (light red) amygdala. **B.** Left - Bigwigs of representative histone 3 lysine 4 dimethylation (H3K4me2) ChIP-seq peaks at representative germline genes in wild-type and *Kdm5c*-KO EpiLCs. Right - Average H3K4me2 at the TSS of all germline-enriched genes in wild-type (dark purple) and *Kdm5c*-KO (light purple) EpiLCs. **C.** Diagram of embryonic stem cell (ESC) to epiblast-like cell (EpiLC) differentiation protocol and collection time points for RNA and protein. **D.** Real time quantitative PCR (RT-qPCR) of *Kdm5c* RNA expression, calculated in comparision to TBP expression ($2^{-\Delta\Delta CT}$). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, Welch's t-test. **E.** KDM5C protein expression normalized to DAXX. Quantified intensity using ImageJ (artificial units - au). Right - representative lanes of Western blot for KDM5C and DAXX. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, Welch's t-test **F.** XXX **G.** XXX

710 Notes

711 Things to do

- 712 • Move dazl to new figure if other staining works
- 713 • Use Ddx4 staining (with dazl?). If not should add RNA to diagram because its a key germline gene.
- 714 • Motif analysis
 - 715 – Discussion - talk about motifs

716 Dazl

717 We were particularly interested in the aberrant transcription of *Dazl*, since it is essential for germ cell
718 development and promotes the translation of germline mRNAs⁸². A significant portion of germline transcripts
719 misexpressed in *Kdm5c*-KO EpiLCs are known binding targets of DAZL, including *Stra8*⁸³ ($p = 1.698e-07$,
720 Fisher's Exact Test). This suggests expression of DAZL protein could promote the translation of other
721 aberrant germline transcripts, influencing their ability to impact *Kdm5c*-KO cellular function. We thus tested
722 DAZL protein expression in *Kdm5c*-KO EpiLCs through immunocytochemistry (Figure 3I). We observed
723 about 25% of *Kdm5c*-KO EpiLCs expressed DAZL protein and it was localized to the cytoplasm ($p = 0.0015$,
724 Welch's t-test), consistent with the pattern of DAZL expression in spermatogonia⁸³. Altogether these results
725 suggest tissue-specific genes are misexpressed during *Kdm5c*-KO embryogenesis, including key drivers of
726 germline identity that can be translated into protein.

- 727 • We additionally found *Kdm5c*-KO EpiLCs ectopically express DAZL protein that is localized to the
728 cytoplasm, similar to its morphology in spermatogonia⁸³. **note: maybe just put in results.** Could
729 move around depending upon if I get pheno working.

730 Discussion notes

- 731 • For other paper:
 - 732 – for methods: Heatmaps of gene expression were generated using the base R functions scale and
733 hclust and visualized using the R package ComplexHeatmap (v2.12.1).
 - 734 – * Might be good to look at for retinoic acid paper (WT germ expression): <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0005654>
- 736 • Talk about different motifs if we do see differences there and if it explains direct vs indirect dysregulation
 - 737 – Could also mention PRC1.6/KDM5C binds pcgf6 in ESCs
- 738 • Maybe talk about other regulators/if anything is known about them long term? Or save for 2nd paper.

- 739 – another germline repressor: <https://pubmed.ncbi.nlm.nih.gov/35345626/>
- 740 • end with something like: this indicates kdm5c not only modulates pluripotency and self-renewal in
741 ESCs, but also has a role in long-term silencing of germline genes
- 742 – then transition into the long term silencing mechanism paragraph
- 743 • KDM5C is dynamically regulated during the window of embryonic germline gene silencing, our EpiLC
744 ChIP-seq is likely catching the tail end of KDM5C's main involvement.
- 745 • Our results indicate KDM5C is a pivotal repressor of germline programs in somatic cells.
- 746 • Studies on other chromatin regulators have focused on key marker genes to identify germline gene
747 misexpression, such as *Dazl*.
- 748 • Using a germline-depleted mouse model, we curated a list of mouse germline-enriched genes to
749 globally assess germline gene dysregulation.
- 750 • Many of the genes dysregulated in the mature *Kdm5c*-KO brain were important for late stages of
751 spermatogenesis, such as those important for sperm flagellar structure. Contrastingly, *Kdm5c*-KO
752 EpiLCs aberrantly expressed key drivers of germ cell identity, including *Dazl* and *Stra8*.
- 753 • While a significant portion of KDM5C-bound promoters are germline genes, many germline genes
754 expressed during *Kdm5c*-KO embryogenesis are not directly bound by kdm5c.
- 755 • One notable example is *Stra8*, a transcription factor important for germ cell specific transcription and
756 meiotic initiation
- 757 • The including the demarcation between soma and germline fates.
- 758 the *Kdm5c*-KO brain also Previous work identified testis-enriched genes in *Kdm5c* genes
- 759 –
- 760 – However unlike the gonadal-biased DEGs,
- 761 • The demarcation of the germline versus soma is a key feature of multicellularity and sexually dimorphic
762 reproduction
- 763 • Anything known about tissue-biased gene expression in other H3K4me regulators?
- 764 • Our data suggests the germline developmental program is occurring ectopically as *Kdm5c*-KOs pro-
765 gresses through somatic tissue development
- 766 • tissue-biased gene expression:

- 767 • Altogether, these data indicate that while some germline genes are misexpressed due to direct loss of
768 KDM5C binding during emryogenesis, secondary downstream mechanisms can also promote their
769 aberrant transcription.
- 770 • Papers to read/reference:
771 – Reconstitution of the Mouse Germ Cell Specification Pathway in Culture by Pluripotent Stem Cells:
772 [https://www.cell.com/fulltext/S0092-8674\(11\)00771-9](https://www.cell.com/fulltext/S0092-8674(11)00771-9)
773 – two cell gene list used by Suzuki et al Max paper is based on 2 cell sequencing: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3395470/>

775 **Figure outline:**

776 **Figure 1: Misexpression of tissue-specific genes in the *Kdm5c*-KO brain** * MA-plot and bar graphs of
777 tissue-enriched genes * Example testis-specific genes (NCBI and bigwigs) * An example ovary tissue-specific
778 gene * An example muscle/liver tissue-specific gene (NCBI and bigwigs)

779 **Figure 2: The male *Kdm5c*-KO brain expresses male and female germline-enriched genes** * Gene
780 ontology of testis DEGs in the amygdala and hippocampus - ontologies are germline ontologies * Expression
781 of testis DEGs in germline-depleted testis (this is adult testis data) * scRNASeq of testis - # of testis DEGs that
782 are germline-specific markers * Although far fewer, 5CKO brain also expresses ovary-enriched genes(NCBI
783 and bigwigs of Zar1) * These ovary enriched genes are also germline specific (NCBI/Li tissues are in adult
784 ovary, so it would be best to show they're oocyte-specific in adult ovary. But I don't think there's a published
785 adult female W/Wv dataset. Could try looking at scRNASeq or just do TPM in embryonic W/Wv data since
786 oocytes are developed at this point? Or both?) * Defining what is/isn't a germline gene, and which are
787 male/female biased using embryonic W/Wv data

788 **Figure 3: *Kdm5c*-KO epiblast-like cells express key drivers regulators of germline identity** * A) ESC
789 to EpiLC differentiation Left - Morphology is unchanged, * B) 5CKO EpiLCs express EpiLC differentiation
790 genes similar to WT lvs * C) Male EpiLCs express germline genes (example Cyct again) * Overlap between
791 brain and EpiLC germline genes - show they're mostly unique * GO of Brain and EpiLC germline genes
792 (meiotic enriched) * Bigwigs or TPM of master regulators * Show that while some are also 2-cell genes
793 (Dazl), 2-cell specific genes aren't dysregulated (Zscan4). Important point because published KDM5C dazl
794 paper is saying KDM5C is a 2-cell regulator, but as far as I can tell only genes shared between germline and
795 2-cell are dysregulated.

796 Staining of Dazl (+ Stra8 if I can get it to work)

797 **Figure 4: Loss of KDM5C's catalytic activity impairs DNAme placement and long-term silencing of
798 germline genes** * Increase in H3K4me3 in *Kdm5c*-KO amygdala at germline genes * Increase in H3K4me2
799 in EpiLCs at germline genes * *Kdm5c* binding in EpiLCs vs PNCs to show that germline repression is
800 happening in early embryo * Previous studies only looked at ESCs, unknown if catalytic activity is required

801 for long-term repression, especially since DNA methylation is placed later). KDM5C RNA and protein ESC →
802 EpiLC (increasing then decreasing) * RNA expression of germline genes with catalytic dead rescue (Ilakkia)
803 * DNA methylation in WT and 5CKO EpiLCs (Ilakkia)

804 **Figure 5: Ectopic, germline-like phenotypes in Kdm5c-KO ESCs/EpiLCs** * Sycp3 staining * DDX4
805 staining and repression of retrotransposons * Cilia??

806 Gaps in knowledge addressed: * Are other tissue-enriched genes dysregulated, or only testis, germline
807 genes? * Curating a robust list of male and female germline genes * Should talk about 2-cell genes
808 vs germline genes - way to systematically categorize? * Mechanism behind long-term germline gene
809 misexpression * Recent evidence suggests loss of KDM5C in ESCs express some germline genes * Unclear
810 if catalytic activity is required for long-term silencing * Unclear if their dysregulation lasts throughout life or
811 the same between brain or not * When in development does it begin? - Recent evidence suggests some
812 germline genes expressed in 5CKO ESCs but unclear if their dysregulation lasts throughout differentiation
813 and if the identity of germline genes are different compared to the brain * Are there functional consequences
814 to germline gene misexpression?

815 Introduction: * Chromatin regulators are important for cellular identity * H3K4me1-3 linked to active
816 gene promoters and enhancers * Surprisingly, mutations in Chromatin regulators lead to many NDDs
817 (including many H3K4 regulators) * Recent studies have shown some chromatin regulators are important
818 for regulating neuron-specific gene expression/chromatin stat_compare_means * However, loss of some
819 chromatin regulators can also lead to ectopic expression of tissue-enriched genes * Very few studies have
820 looked at these genes and it's unclear if these genes contribute to NDD impairments. * Necessary to
821 first characterize the mechanism behind their derepression to identify molecular footholds into testing their
822 contribution to neuronal impairments and potential for therapeutic intervention

- 823 • Loss of KDM5C can result in the misexpression of genes typically only found in the testis
- 824 – Misexpression of tissue-enriched genes hasn't been systematically characterized - Unclear if
825 these genes are exceptions or if other tissue-specific genes are dysregulated
- 826 – Interestingly, these genes (Cyct, D1pas1) typically function in the germline
- 827 – Germ cells (meiotic cells) are typically distinguished from somatic cells very early on in embryogen-
828 esis and is a key feature of multicellularity
- 829 – Chromatin regulators are very important for decommissioning germline genes and act successively
830 the embryo implants into the uterine wall
- 831 – Most studies have focused on ESCs, which have a similar transcriptome to germ cells / 2-cells
- 832 * recently, KDM5C was shown to repress DAZL in ESCs, independent of its catalytic activity
- 833 * However, DNA methylation is lost in the mature 5CKO brain, DNA methylation is placed later
834 and it's Unclear if it's required for long-term repression (maybe too specific, just trying to go
835 into the fact that the mechanism is partially understood but unclear)

- 836 – Systematic characterization of ectopic germline genes hasn't been done
837 * unknown if other germline-enriched genes are dysregulated, including oocyte-specific genes
838 * Crucially, it's unknown if misexpression of the germline program leads to functional conse-
839 quences in 5CKO cells.

840 **Germline gene repression background:**

841 Interestingly, some of the ectopic testis transcripts identified in the *Kdm5c*-KO brain are typically ex-
842 pressed in germ cells⁸. Unlike somatic cells, germ cells (e.g. sperm and eggs) undergo meiosis and pass on
843 their genetic material to the next generation. The germline and the soma are typically distinguished during
844 early embryogenesis, when germline genes are silenced in epiblast stem cells soon after implantation and
845 only reactivated in a subset to form the germline. Chromatin regulators play a key role in decommissioning
846 germline genes as the embryo transitions from naïve to primed pluripotency by placing repressive histone
847 H2A lysine 119 monoubiquitination (H2AK119ub1)¹⁶, histone 3 lysine 9 trimethylation (H3K9me3)^{16,17}, and
848 DNA CpG methylation^{17–19} at germline gene promoters. KDM5C may also be involved in this early decom-
849 missioning of germline genes, as re-expression of KDM5C in neurons fails to suppress their dysregulation⁸.
850 In support of this, KDM5C was very recently shown to repress *Deleted in azoospermia like (Dazl)*, a key
851 regulator of germline development, in mouse embryonic stem cells (ESCs)^{49,71}. In support of this, two
852 independent screens in mouse embryonic stem cells (ESCs) recently identified KDM5C as a repressor of
853 *Deleted in azoospermia like (Dazl)*, a key regulator of germline development. However, KDM5C's role in
854 embryonic germline gene repression is currently unclear, given that Dazl is also expressed in ESCs and in
855 the 2-cell stage and germline gene misexpression has yet to be globally characterized during *Kdm5c*-KO
856 embryogenesis.