

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

2 intellectual disability factor KDM5C

3

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

5 Abstract

6 Mutations in numerous chromatin-modifying enzymes cause neurodevelopmental disorders (NDDs).
7 While, loss of repressive chromatin regulators can lead to the aberrant transcription of tissue-specific genes
8 outside of their intended context, the mechanisms and consequences are largely unknown. Here, we
9 explored this cellular identity crisis in NDDs by investigating the role of lysine demethylase 5c (KDM5C),
10 an eraser of histone 3 lysine 4 di and tri-methylation (H3K4me2/3). We found male *Kdm5c* knockout
11 (-KO) mice, which recapitulate key behavioral phenotypes of Claes-Jensen X-linked intellectual disability,
12 aberrantly expresses many liver, muscle, ovary, and testis genes within the amygdala and hippocampus.
13 Gonad-enriched genes expressed in the *Kdm5c*-KO brain were typically unique to germ cells, indicating an
14 erosion of the soma-germline boundary. Germline genes are typically decommissioned in somatic lineages in
15 the post-implantation epiblast, yet *Kdm5c*-KO epiblast-like cells (EpiLCs) aberrantly expressed key regulators
16 of germline identity and meiosis, including *Dazl* and *Stra8*. Germline gene dysregulation was sexually
17 dimorphic, as female *Kdm5c*-KO EpiLCs were more prone to germline gene misexpression than knockout
18 males. Using a curated list of mouse germline-enriched genes, we found KDM5C is selectively recruited to
19 a subset of germline gene promoters that contain CpG islands (CGIs) to facilitate DNA CpG methylation
20 (CpGme) during ESC to EpiLC differentiation. However, CGI-free germline gene promoters lacking KDM5C,
21 particularly late-stage spermatogenesis genes, can also become activated in *Kdm5c*-KO cells via ectopic
22 activation by RFX transcription factors. Thus, ectopic germline transcriptional programs can progress in the
23 background of typical *Kdm5c*-KO development, due to downstream activation by key drivers of germline
24 gene expression.

25 Introduction

26 A single genome holds the instructions to generate the myriad of cell types found in an organism. This is, in
27 part, accomplished by chromatin regulators that can either promote or impede lineage-specific gene expres-

28 sion through DNA and histone modifications^{1–5}. Human genetic studies revealed mutations in chromatin
29 regulators are a major cause of neurodevelopmental disorders (NDDs)⁶ and many studies have identified
30 their importance in regulating the expression of brain-specific genes. Loss of some chromatin regulators
31 can also result in the ectopic transcription of tissue-specific genes outside of their target environment, such
32 as the misexpression of liver-specific genes within adult neurons⁷. However, the mechanisms underlying
33 ectopic gene expression and its impact upon neurodevelopment are poorly understood.

34 To elucidate the role of tissue identity in chromatin-linked NDDs, it is essential to first characterize the
35 nature of the dysregulated genes and the molecular mechanisms governing their de-repression. Here we
36 focus on lysine demethylase 5C (KDM5C), also known as SMCX or JARID1C, which erases histone 3 lysine
37 4 di- and trimethylation (H3K4me2/3), a permissive chromatin modification enriched at gene promoters⁸.
38 Pathogenic mutations in *KDM5C* cause Intellectual Developmental Disorder, X-linked, Syndromic, Claes-
39 Jensen Type (MRXSCJ, OMIM: 300534). MRXSCJ is more common and severe in males and its neurological
40 phenotypes include intellectual disability, seizures, aberrant aggression, and autistic behaviors^{9–11}. Male
41 *Kdm5c* knockout (-KO) mice recapitulate key MRXSCJ phenotypes, including hyperaggression, increased
42 seizure propensity, and learning impairments^{12,13}. RNA sequencing (RNA-seq) of the *Kdm5c*-KO hippocam-
43 pus revealed ectopic expression of some germline genes within the brain¹³. However, it is unknown whether
44 other types of testis genes are dysregulated, at what point in development germline gene misexpression
45 begins, and if other tissue-specific genes are aberrantly transcribed with KDM5C loss.

46 Distinguishing between germ cells and somatic cells is a key feature of multicellularity¹⁴ that occurs
47 during early embryogenesis in many metazoans¹⁵. In mammals, chromatin regulators are crucial for
48 decommissioning germline genes in somatic cells during the transition from naïve to primed pluripotency.
49 Initially, germline gene promoters gain repressive histone H2A lysine 119 monoubiquitination (H2AK119ub1)¹⁶
50 and histone 3 lysine 9 trimethylation (H3K9me3)^{16,17} in embryonic stem cells and are then decorated with
51 DNA CpG methylation (CpGme) in the post-implantation embryo^{17–19}. The contribution of KDM5C to
52 this process remains unclear. It is additionally unknown if KDM5C-mediated germline gene repression is
53 impacted by chromosomal sex, as females have a higher KDM5C dose than males due to its escape from X
54 chromosome inactivation^{20–23}. Furthermore, studies on germline gene repression have primarily focused on
55 marker genes important for germ cell development rather than germline genes as a whole, given the lack of
56 a curated list for germline-enriched genes. Therefore, it is unknown if the mechanism of repression differs for
57 certain classes of germline genes, e.g. meiotic genes versus spermatid differentiation genes.

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

64 **Results**

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

66 Previous RNA sequencing (RNA-seq) of the adult male *Kdm5c*-KO hippocampus revealed ectopic
67 expression of some germline genes unique to the testis¹³. It is currently unclear if the testis is the only tissue
68 type misexpressed in the *Kdm5c*-KO brain. We thus characterized the role of KDM5C in brain tissue identity
69 by systematically assessing the dysregulation of genes specific to or enriched in 17 mouse tissues²⁴. We
70 analyzed our published mRNA-seq datasets of the adult amygdala and hippocampus from wild-type and
71 constitutive *Kdm5c*-KO male mice²⁵.

72 We found a large proportion significantly upregulated genes within the *Kdm5c*-KO brain (DESeq2²⁶,
73 log2 fold change > 0.5, q < 0.1) are typically enriched in non-brain tissues (Amygdala: 35%, Hippocampus:
74 24%) (Figure 1A-B). For both the amygdala and hippocampus, the majority of tissue-enriched differentially
75 expressed genes (DEGs) were testis genes (Figure 1A-C). Even though the testis has the largest total
76 number of tissue-biased genes (2,496 genes) compared to any other tissue, testis-biased DEGs were
77 significantly enriched for both brain regions (Amygdala p = 1.83e-05, Odds Ratio = 5.13; Hippocampus p =
78 4.26e-11, Odds Ratio = 4.45, Fisher's Exact Test). One example of a testis-enriched gene misexpressed
79 in the *Kdm5c*-KO brain is *FK506 binding protein 6* (*Fkbp6*), a known regulator of PIWI-interacting RNAs
80 (piRNAs) and meiosis^{27,28} (Figure 1C).

81 Interestingly, we also observed significant enrichment of ovary-biased DEGs in both the amygdala and
82 hippocampus (Amygdala p = 0.00574, Odds Ratio = 18.7; Hippocampus p = 0.048, Odds Ratio = 5.88,
83 Fisher's Exact) (Figure 1A-B). Ovary-enriched DEGs included *Zygotic arrest 1* (*Zar1*), which sequesters
84 mRNAs in oocytes for meiotic maturation²⁹ (Figure 1D). Given that the *Kdm5c*-KO mice we analyzed are male,
85 these data demonstrate that the ectopic expression of tissue-enriched genes is independent of organismal
86 sex.

87 Although not consistent across brain regions, we also found significant enrichment of DEGs biased
88 towards two non-gonadal tissues - the liver (Amygdala p = 0.04, Odds Ratio = 6.58, Fisher's Exact Test)
89 and the muscle (Hippocampus p = 0.01, Odds Ratio = 6.95, Fisher's Exact Test) (Figure 1A-B). Among
90 liver-biased DEG derepressed in both the hippocampus and amygdala is *Apolipoprotein C-I* (*Apoc1*), a
91 lipoprotein metabolism and transport gene³⁰ that has been implicated in Alzheimer's disease³¹ (Figure 1E).

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

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

98 In addition to germ cells, the testis also contains both somatic cells, such as Leydig cells that support
99 hormone production and germline functions. To determine if *Kdm5c*-KO results in ectopic expression of
100 somatic testicular genes, we first evaluated their known functions through gene ontology analysis. We found
101 *Kdm5c*-KO testis-enriched DEGs had high enrichment of germline-relevant ontologies, including spermatid
102 development (GO: 0007286, p.adjust = 6.2e-12) and sperm axoneme assembly (GO: 0007288, p.adjust =
103 2.45e-14) (Figure 2A).

104 To validate the cell-type origin of testicular DEGs, we then compared their expression in somatic and
105 germ cells within the testis. We first evaluated their expression in wild-type testes versus testes with germ cell
106 depletion³², which was accomplished by heterozygous *W* and *Wv* mutations in the enzymatic domain of *c-Kit*
107 (*Kit*^{W/Wv})³³. Almost all *Kdm5c*-KO testis-enriched DEGs lost expression with germ cell depletion (Figure 2B).
108 We then assessed testis-enriched DEG expression in a published single cell RNA-seq dataset that identified
109 cell type-specific markers within the testis³⁴. Some *Kdm5c*-KO testis-enriched DEGs were classified as
110 specific markers for different germ cell developmental stages (e.g. spermatogonia, spermatocytes, round
111 spermatids, and elongating spermatids), yet none marked somatic cells (Figure 2C). Together, these data
112 demonstrate that the *Kdm5c*-KO brain aberrantly expresses germline genes, but not somatic testicular genes,
113 reflecting an erosion of the soma-germline boundary.

114 As of yet, research on germline silencing mechanisms has focused on a handful of key genes rather than
115 assessing germline gene suppression genome-wide due to the lack of a comprehensive mouse germline
116 gene list. We therefore generated a list of germline-enriched genes using RNA-seq datasets of *Kit*^{W/Wv} mice
117 that included males and females at embryonic day 12, 14, and 16³⁵ and adult male testes³². We defined
118 genes as germline-enriched if their expression met the following criteria: 1) their expression is greater than
119 1 FPKM in wild-type gonads 2) their expression in any non-gonadal tissue of adult wild type mice²⁴ does
120 not exceed 20% of their maximum expression in the wild-type germline, and 3) their expression in the germ
121 cell-depleted gonads, for any sex or time point, does not exceed 20% of their maximum expression in the
122 wild-type germline. These criteria yielded 1,288 germline-enriched genes (Figure 2D), which was hereafter
123 used as a resource to globally characterize germline gene misexpression with *Kdm5c* loss (Supplementary
124 table 1).

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

126 Germ cells are typically distinguished from somatic cells soon after the embryo implants into the uterine
127 wall^{36,37}. A subset of epiblast stem cells become the primordial germ cells (PGCs) while the remainder
128 differentiate into the ectoderm, mesoderm, and endoderm to form the somatic tissues³⁸. This developmental
129 time point can be modeled *in vitro* through differentiation of embryonic stem cells (ESCs) into epiblast-
130 like stem cells (EpiLCs) (Figure 3A)^{39,40}. While some germline-enriched genes are also expressed in

131 naïve embryonic stem cells (ESCs) and in the 2-cell stage^{41–43}, they are silenced as they differentiate into
132 EpiLCs^{17,44}. Therefore, we tested if KDM5C was necessary for the initial silencing germline genes in somatic
133 lineages by evaluating the impact of *Kdm5c* loss in male EpiLCs.

134 *Kdm5c*-KO cell morpholgy during ESC to EpiLC differentiation appeared normal (Figure 3A) and EpiLCs
135 properly expressed markers of primed pluripotency, such as *Dnmt3b*, *Fgf5*, *Pou3f1*, and *Otx2* (Figure 3B).
136 We then identified tissue-enriched DEGs in our previously published RNA-seq dataset of wild-type and
137 *Kdm5c*-KO EpiLCs⁴⁵ (DESeq2, log2 fold change > 0.5, q < 0.1). Similar to the *Kdm5c*-KO brain, we observed
138 general dysregulation of tissue-enriched genes, with the largest number of genes belonging to the brain and
139 testis, although they were not significantly enriched (Figure 3B).

140 Using the list of mouse germline-enriched genes curated above, we found 54 were misexpressed in male
141 *Kdm5c*-KO EpiLCs. We then compared EpiLC germline DEGs to those expressed in the *Kdm5c*-KO brain to
142 determine if germline genes are constitutively dysregulated or change over the course of development. We
143 found the majority of germline DEGs were unique to either EpiLCs or the brain, with only *Cyct* shared across
144 all tissue/cell types (Figure 3E-F). EpiLCs had particularly high enrichment of meiosis-related gene ontologies
145 (Figure 3G), such as meiotic cell cycle process (GO:1903046, p.adjust = 1.59e-08) and meiotic nuclear
146 division (GO:0140013, p.adjust = 9.76e-09). While there was modest enrichment of meiotic gene ontologies
147 in both brain regions, the *Kdm5c*-KO hippocampus primarily expressed late-stage spermatogenesis genes
148 involved in sperm axoneme assembly (GO:0007288, p.adjust = 0.00621) and sperm motility (GO:0097722,
149 p.adjust = 0.00612).

150 Notably, DEGs unique to *Kdm5c*-KO EpiLCs included key drivers of germline identity, such as *Stimulated*
151 *by retinoic acid 8* (*Stra8*) and *Deleted in azoospermia like* (*Dazl*) (Figure 3H). These genes are typically
152 expressed when PGCs commit to the germline fate, but are also expressed later in life to trigger meiotic gene
153 expression programs^{46–48}. Of note, some germline genes, including *Dazl*, are also expressed in the two-cell
154 embryo^{42,49}. However, we did not see derepression of two-cell stage-specific genes, like *Duxf3* (*Dux*) (q =
155 0.337) and *Zscan4d* (q = 0.381) (Figure 3H), indicating *Kdm5c*-KO EpiLCs do not revert back to a 2-cell-like
156 state. Altogether, these results demonstrate *Kdm5c*-KO EpiLCs express key drivers of germline identity and
157 meiosis while the *Kdm5c*-KO brain primarily expresses spermatogenesis genes.

158 **Female epiblast-like cells have increased sensitivity to germline gene misexpression 159 with *Kdm5c* loss**

160 It is currently unknown if the misexpression of germline genes is influenced by sex, as previous studies
161 on germline gene repressors have focused on males^{16–18,50,51}. Sex is particularly pertinent in the case of
162 KDM5C because it lies on the X chromosome and partially escapes X chromosome inactivation (XCI), resulting
163 in a higher dosage in females^{20–23}. We therefore explored the impact of chromosomal sex upon germline
164 gene misexpression by comparing their dysregulation in male *Kdm5c* hemizygous knockout (XY *Kdm5c*-

165 KO), female homozygous knockout (XX *Kdm5c*-KO), and female heterozygous knockout (XX *Kdm5c*-HET)
166 EpiLCs.⁴⁵.

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

177 The majority of germline genes expressed in both sexes had a greater log2 fold change in females
178 compared to males (Figure 4D). This increased degree of dysregulation in females, along with the increased
179 total number of germline genes, indicates females are more sensitive to losing KDM5C-mediated germline
180 gene suppression. Female sensitivity could be due to impaired XCI in *Kdm5c* mutants⁴⁵, as many spermato-
181 genesis genes lie on the X chromosome^{52,53}. However, female germline DEGs were not biased towards
182 the X chromosome and had a similar overall proportion of X chromosome DEGs compared to males
183 (XY *Kdm5c*-KO - 10.29%, XX *Kdm5c*-HET - 7.43%, XX *Kdm5c*-KO - 10.59%) (Figure 4G). The majority of
184 germline DEGs instead lie on autosomes for both male and female *Kdm5c* mutants (Figure 4G). Thus, while
185 female EpiLCs are more prone to germline gene misexpression with KDM5C loss, it is likely independent of
186 XCI defects.

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

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

200 genes.

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

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

207 To address these questions, we analyzed KDM5C chromatin immunoprecipitation followed by DNA
208 sequencing (ChIP-seq) datasets in EpiLCs⁴⁵ and primary forebrain neuron cultures (PNCs)¹². EpiLCs had a
209 higher total number of KDM5C peaks than PNCs (EpiLCs: 5,808, PNCs: 1,276, MACS2 q < 0.1 and fold
210 enrichment > 1, removal of *Kdm5c*-KO false positive peaks) and KDM5C was primarily localized to gene
211 promoters in both cell types (EpiLCs: 4,190, PNCs: 745 ± 500bp from the TSS), although PNCs showed
212 increased localization to non-promoter regions (Figure 5A).

213 The majority of promoters bound by KDM5C in PNCs were also bound in EpiLCs (513 shared promoters),
214 however a large portion of gene promoters were only bound by KDM5C in EpiLCs (3677 EpiLC only
215 promoters) (Figure 5B). Genes bound by KDM5C in both PNCs and EpiLCs were enriched for functions
216 involving nucleic acid turnover, such as deoxyribonucleotide metabolic process (GO:0009262, p.adjust =
217 8.28e-05) (Figure 5C). Germline-specific ontologies were only enriched in KDM5C-bound promoters unique
218 to EpiLCs, such as meiotic nuclear division (GO: 0007127 p.adjust = 6.77e-16) and meiotic cell cycle process
219 (GO:1903046, p.adjust = 5.05e-16) (Figure 5C). There were no ontologies significantly enriched for genes
220 bound by KDM5C only in PNCs. We next evaluated KDM5C binding around the transcription start site
221 (TSS) of all germline-enriched genes. KDM5C binding around the TSS of many germline genes is evident in
222 EpiLCs, but absent in PNCs (Figure 5D). Based on our ChIP-seq peak cut-off criteria, KDM5C was strongly
223 bound to about 10% of germline gene promoters in EpiLCs (Figure 5E).

224 In EpiLCs, KDM5C was only bound to about one third of *Kdm5c*-KO RNA-seq DEG promoters (EpiLC
225 only DEGs: 36%, Brain only DEGs: 33.3%). However, KDM5C was bound to the promoter at 3 out of the 4
226 genes dysregulated in both the brain and EpiLCs (Supplementary figure XXX). Representative examples of
227 KDM5C-bound and unbound germline DEGs are *Dazl* and *Stra8*, respectively (Figure 5F). Together, these
228 results demonstrate KDM5C is recruited to a subset of germline genes in EpiLCs, including enrichment
229 of meiotic genes, but does not directly regulate germline genes in neurons. Furthermore, the majority of
230 germline mRNAs expressed in *Kdm5c*-KO cells are dysregulated independent of direct KDM5C binding to
231 their promoters.

232 One notable gene that lacked KDM5C binding was *Stra8*, even though its mRNA is expressed in
233 *Kdm5c*-KO EpiLCs ().

234 Many germline-specific genes are suppressed by the transcription factor heterodimers E2F6/DP1 and

235 MGA/MAX, which respectively bind E2F and E-box motifs^{18,50,51,54,55}. To elucidate if KDM5C is recruited to
236 germline gene promoters by a similar mechanism, we used HOMER to identify transcription factor motifs
237 enriched at KDM5C-bound or unbound germline gene promoters⁵⁶ (TSS ± 500 bp, q-value < 0.1). MAX
238 and E2F6 binding sites were significantly enriched at germline genes bound by KDM5C in EpiLCs, but
239 not at germline genes unbound by KDM5C (MAX q-value: 0.0068, E2F6 q-value: 0.0673, E2F q-value:
240 0.0917) (Figure 5G). One third of KDM5C-bound promoters contained the consensus sequence for either
241 E2F6 (E2F, 5'-TCCCGC-3'), MGA (E-box, 5'-CACGTG-3'), or both, but only 17% of KDM5C-unbound genes
242 contained these motifs (Figure 5H). KDM5C-unbound germline genes were instead enriched for multiple RFX
243 transcription factor binding sites (RFX q-value < 0.0001, RFX2 q-value < 0.0001, RFX5 q-value < 0.0001)
244 (Figure 5I). RFX transcription factors bind X-box motifs⁵⁷ to promote ciliogenesis^{58,59} and among them is
245 RFX2, a central regulator of post-meiotic spermatogenesis^{60,61}. Interestingly, RFX2 mRNA is derepressed
246 in *Kdm5c*-KO EpiLCs (Figure 5J), however it is also not a direct target of KDM5C (Supplementary figure
247 XXX). Thus, RFX2 is a candidate transcription factor for driving the ectopic expression of KDM5C-unbound
248 germline genes in *Kdm5c*-KO cells.

249 **KDM5C is recruited to germline gene promoters harboring CpG islands to facilitate
250 *de novo* DNA methylation**

251 In the early embryo, germline gene promoters are initially decorated with repressive histone modifications
252 and are then silenced long-term via DNA CpG methylation (CpGme)^{16,17,44,62}. Our results above indicate
253 KDM5C also acts at germline gene promoters during this time period. However, how KDM5C interacts with
254 other germline gene silencing mechanisms is currently unclear. KDM5C is generally thought to suppress
255 transcription through erasure of histone 3 lysine 4 di- and trimethylation (H3K4me2/3)⁸, yet KDM5C's
256 catalytic activity was recently shown to be dispensable for suppressing *Dazl* in undifferentiated ESCs⁴⁹. Since
257 H3K4me3 impedes *de novo* CpGme placement^{63,64}, KDM5C's catalytic activity may instead be required
258 in the post-implantation embryo for long-term silencing of germline genes. In support of this, CpGme is
259 markedly reduced at two germline gene promoters in the *Kdm5c*-KO adult hippocampus¹³.

260 Based on the above observations, we hypothesized KDM5C erases H3K4me3 to promote the initial
261 placement of CpGme at germline gene promoters in EpiLCs. To test this hypothesis, we first characterized
262 KDM5C's substrates, histone 3 lysine 4 di- and trimethylation (H3K4me2/3), at germline gene promoters in
263 our previously published ChIP-seq datasets in male wild type and *Kdm5c*-KO amygdala²⁵ and EpiLCs⁴⁵. In
264 congruence with previous work in the *Kdm5c*-KO hippocampus¹³, we observed aberrant accumulation of
265 H3K4me3 around the transcription start site (TSS) of germline genes in the *Kdm5c*-KO amygdala (Figure
266 6A). We additionally found a marked increase in H3K4me2 around the TSS of germline genes in *Kdm5c*-KO
267 EpiLCs (Figure 6B). To elucidate KDM5C's embryonic role, we then characterized KDM5C's mRNA and protein
268 expression during male ESC to EpiLC differentiation (Figure 6C). While *Kdm5c* mRNA steadily decreased

269 from 0 to 48 hours of differentiation (Figure 6D), KDM5C protein initially increased from 0 to 24 hours but
270 then decreased to near knockout levels by 48 hours (Figure 6E). Together, these data suggest KDM5C acts
271 during the transition between ESCs and EpiLCs to remove H3K4me at germline gene promoters.

272 In wild-type cells, germline genes accumulate CpG methylation (CpGme) at CpG islands (CGIs) during
273 the transition from naïve to primed pluripotency^{19,44,65}, reaching peak methylation levels when differentiated
274 into EpiLCs for 96 hours (extended EpiLCs, exEpiLCs)¹⁷. We first identified how many germline genes
275 contained CGIs using the UCSC genome browser⁶⁶ and found out of 1,288 germline-enriched genes, only
276 356 (27.64%) contained CGIs within their promoters (TSS ± 500 bp) (Figure 6F). CGI-containing germline
277 genes were enriched for meiotic gene ontologies, including meiotic nuclear division (GO:XXXX, p.adj) and
278 meiosis I (GO:XXXX, p.adj) when compared to CGI-free genes (Figure 6G). Although a minor portion of
279 germline gene promoters contained CGIs, CGIs strongly determined KDM5C's recruitment to germline gene
280 promoters (FISHER'S XXXX), with 79% of KDM5C-bound germline genes containing CGIs (Figure 6G).

281 We then performed whole genome bisulfite sequencing (WGBS) in male wild-type and *Kdm5c*-KO ESCs
282 and 96 hour extended EpiLCs (exEpiLCs) to assess how KDM5C loss impacted initial CpGme placement
283 at germline gene promoters (Figure 6H). We first identified which germline gene promoters significantly
284 gained CpGme in wild-type cells during ESC to exEpiLCs differentiation (methylKit⁶⁷, q < 0.01, |methylation
285 difference| >= 25%, TSS ± 500 bp). In wild-type cells, the majority of germline genes gained substantial
286 CpGme at their promoter during differentiation (60.08%), regardless if their promoter contained a CGI (Figure
287 6I).

288 We then identified germline gene promoters differentially methylated in wild-type versus *Kdm5c*-KO
289 exEpiLCs (methylKit, q < 0.01, |methylation difference| >= 25%, TSS ± 500 bp) and found 28 germline
290 promoters were significantly hypomethylated with *Kdm5c* loss (Figure 6J). Approximately half of germline
291 promoters hypomethylated in *Kdm5c*-KO exEpiLCs are direct targets of KDM5C in EpiLCs (13 out of 28
292 hypomethylated DMRs). We then evaluated promoter CpGme at germline genes ectopically transcribed in
293 either *Kdm5c*-KO EpiLCs or within the brain and found promoter CpGme was substantially reduced in about
294 half of germline DEGs (Figure 6K). Significantly hypomethylated promoters included genes consistently
295 dysregulated across multiple *Kdm5c*-KO RNA-seq datasets, such as *D1Pas1* (methylation difference =
296 -60.03%, q-value = 3.26e-153) (Figure 6L). Surprisingly, we only observed a modest reduction in CpGme
297 at *Dazl*'s promoter (methylation difference = -6.525%, q-value = 0.0159) (Figure 6M). Altogether, these
298 results demonstrate KDM5C is recruited to germline gene CGIs to promote CpGme at select germline gene
299 promoters during early embryogenesis, however other germline gene silencing mechanism can sufficiently
300 compensate for KDM5C's loss at select germline gene promoters.

301 **Discussion**

302 In the above study, we demonstrate KDM5C's pivotal role in the development of tissue identity. We
303 first characterized the misexpression of tissue-enriched genes within the *Kdm5c*-KO brain and identified
304 substantial dysregulation of testis, liver, muscle, and ovary-enriched genes. Testis genes significantly
305 enriched within the *Kdm5c*-KO amygdala and hippocampus are specific to germ cells and not expressed
306 within testis somatic cells. *Kdm5c*-KO epiblast-like cells (EpiLCs) aberrantly express key drivers of germline
307 identity and meiosis, including *Dazl* and *Stra8*, while the *Kdm5c*-KO brain primarily expresses genes
308 important for late spermatogenesis. We demonstrated that while *Kdm5c* mutant sex did not influence
309 whether sperm or egg-specific genes were misexpressed, female EpiLCs are more sensitive to germline
310 gene de-repression. Germline-enriched mRNAs can become aberrantly expressed in *Kdm5c*-KO cells
311 indirect of KDM5C, as although KDM5C is enriched at germline gene promoters in EpiLCs, it is only bound
312 to a subset of germline-enriched DEGs. Finally, we found KDM5C is dynamically regulated during ESC to
313 EpiLC differentiation and promotes long-term germline gene silencing through DNA methylation at CpG
314 islands. Therefore, we propose KDM5C plays a fundamental role in the development of tissue identity
315 during early embryogenesis, including the establishment of the soma-germline boundary. By systematically
316 characterizing KDM5C's role in germline gene repression, including its interaction with known silencing
317 mechanisms, we unveiled unique repressive mechanisms governing distinct classes of germline gene in
318 somatic lineages. Furthermore, these data provide molecular footholds that can then be exploited to test the
319 ultimate contribution of ectopic germline gene expression upon neurodevelopment.

320 It is important to note that while we highlighted KDM5C's regulation of germline genes, some germline-
321 enriched genes are also expressed at the 2-cell stage and in naïve ESCs for their role in pluripotency and
322 self-renewal. For example, *Dazl* is primarily known for committing PGCs to the germline fate and regulating
323 the translation of germline-specific mRNAs, but is also expressed at the 2-cell stage⁶⁸, in naïve ESCs⁴¹,
324 and in the inner cell mass⁴¹. KDM5C may therefore negatively regulate totipotency, as KDM5C suppresses
325 *Dazl* in ESCs^{49,69} and *Kdm5c*-KO ESCs aberrantly express 2-cell-specific genes like *Zscan4c*⁴⁹. However,
326 misexpression of *Dazl* in *Kdm5c*-KO ESCs was independent of the 2-cell specific transcription factor *Dux*⁴⁹
327 and, unlike *Dazl*, KDM5C does not bind to the *Dux* promoter.

328 Although expressed in naïve ESCs, *Dazl* and other "self-renewal" germline genes are silenced during
329 differentiation into epiblast stem cells/EpiLCs^{17,44}. We found that while *Kdm5c*-KO EpiLCs also expressed
330 *Dazl*, they did not express 2-cell specific genes. Misexpression of germline genes in *Kdm5c*-KO EpiLCs could
331 indicate they are differentiating into primordial germ cell-like cells (PGCLCs)^{36,37,39}. Yet, *Kdm5c*-KO EpiLCs
332 had normal cellular morphology and properly expressed markers for primed pluripotency, including *Otx2*
333 which is known to repress EpiLC differentiation into PGCs/PGCLCs⁷⁰. Altogether, these data suggest *Kdm5c*-
334 KO germline gene misexpression occurs ectopically in conjunction with typical developmental programs and
335 the 2-cell-like state observed in *Kdm5c*-KO ESCs is an indirect consequence of KDM5C's primary role in

336 germline gene repression.

337 Although eggs and sperm employ the same transcriptional programs for shared functions, e.g. PGC
338 formation, meiosis, and genome defense, some germline genes are sex specific. We found *Kdm5c* mutant
339 males and females expressed both sperm and egg-biased genes, indicating the mechanism of derepression
340 is independent of a given germline gene's sex. However, organismal sex did greatly influence the degree of
341 germline gene dysregulation, as female *Kdm5c*-KO EpiLCs had over double the number of germline-enriched
342 DEGs compared to males. X chromosome inactivation (XCI) defects could explain why *Kdm5c* knockout
343 females are more prone to germline gene dysregulation, given the X chromosome harbors a large number
344 of spermatogenesis genes^{52,53} and loss of *Kdm5c* impairs XCI⁴⁵. However, female germline DEGs were
345 not significantly biased towards the X chromosome. Sex differences in germline gene suppression may be
346 instead connected to females having a higher dose of KDM5C than males, due to its partial escape from
347 XCI^{20–23}. Intriguingly, females with heterozygous loss of *Kdm5c* also had over double the number of germline
348 DEGs than hemizygous knockout males, even though their level of KDM5C should be roughly equivalent to
349 that of wild-type males. Altogether, these results suggest female EpiLCs are more prone to transitioning
350 to a germ cell-like state than males and require more KDM5C to maintain somatic cellular identity. Future
351 studies are required to illuminate if this phenomenon is unique to sexually dimorphic chromatin regulators or
352 a general feature of female cells.

353 It is yet unknown how KDM5C is recruited to germline gene promoters, given that KDM5C itself does
354 not contain domains for sequence-specific binding⁸. In HeLa cells and ESCs^{49,71}, KDM5C associates with
355 members of the polycomb repressive complex 1.6 (PRC1.6), which is recruited to germline gene promoters
356 through E2F6/DP1 and MGA/MAX heterodimers^{16,55}. While MAX and E2F6 motifs were enriched at KDM5C-
357 bound germline genes in EpiLCs, only about one-third of promoters contained their consensus sequence.
358 Additionally, some germline genes can become active in *Kdm5c*-KO cells independent of KDM5C binding to
359 their promoters. Germline genes unbound by KDM5C were primarily late-stage spermatogenesis genes and
360 significantly enriched for RFX transcription factors, including RFX2, a central regulator of spermiogenesis^{60,61}.
361 Another notable KDM5C-unbound DEG is *Stra8*, a meiotic transcription factor that is turned on in germ cells
362 with retionic acid signaling and DAZL expression^{72,73}. Misexpression of *Dazl* and *Rfx2* and their downstream
363 targets in *Kdm5c*-KO cells suggests that once activated, ectopic germline genes can then turn on other
364 aberrant germline programs to loosely mimic germ cell development.

365 Unlike the previously characterized germline gene suppressors that place repressive histone
366 modifications^{16,17,19,62,74–76}, KDM5C removes the active histone mark histone 3 lysine 4 di- and trimethylation
367 (H3K4me2/3)⁸. Yet, KDM5C's catalytic activity may not be required for germline gene silencing, as it
368 was recently found to be dispensible for repressing *Dazl* in ESCs⁴⁹. Emerging work indicates many
369 histone-modifying enzymes have non-catalytic functions that can influence gene expression, sometimes even
370 more potently than their catalytic roles^{77,78}. Although dispensible in ESCs, KDM5C's catalytic activity could
371 be necessary to silence germline genes in EpiLCs, as H3K4me3 impedes *de novo* CpGme placement^{63,64}.

372 In support of this, CpGme is significantly eroded at at least two germline promoters in the adult *Kdm5c*-KO
373 hippocampus¹³. We found *Kdm5c*-KO cells accumulate H3K4me2/3 around the TSS of germline genes and
374 have impaired CpGme placement at CGIs for select germline gene promoters. KDM5C's shifting role in
375 germline gene repression during the transition from naïve to primed pluripotency indicates the requirement
376 of catalytic and non-catalytic gene regulatory mechanisms can change depending upon the locus and
377 developmental stage.

- 378 • DNAm and CpG islands
- 379 – Loss of many chromatin regulators affects CpGme (Multiple, non-redundant silencing mechan-
380 isms) - seems to be the crux point of germline gene silencing
- 381 – Germline genes are methylated at promoter CGIs, which are typically not methylated for other
382 types of genes
- 383 – Because CGIs are typically resistant to CpGme (accurate?), germline CGIs may require a highly
384 repressive histone landscape to recruit sufficient DNMTs to these loci
- 385 – We found only 30% of germline-enriched genes have CGIs, but the majority of CGI-free germline
386 genes still gain CpGme around the TSS. Unclear what the repressive mechanism is for these
387 genes.
- 388 – CGIs highly determine KDM5C recruitment to germline gene promoters. - KDM5C is known to be
389 enriched at CGIs (in neurons? are these methylated? or is its germline CGI function different from
390 its somatic CGI function?).
- 391 – KDM5C loss impacts germline gene CGI methylation, but not really the non-CGI TSS methylation
- 392 * Other studies on germline gene repressors have shown they are important for CGIme, unclear
393 if they participate in non-CGI TSS CpGme
- 394 – Altogether demonstrates different classes of germline genes use distinct silencing mechanisms

395 The above work provides the mechanistic foundation for KDM5C's repression of germline-specific
396 transcription and offers novel insight into how dysregulation of *Kdm5c*-KO tissue identity changes over the
397 course of development. However, the contribution of these ectopic, tissue-specific genes towards *Kdm5c*-KO
398 neurological impairments is still unknown. In addition to germline genes, we also identified significant
399 enrichment of muscle, liver, and even ovary-biased transcripts within the male *Kdm5c*-KO brain. Intriguingly,
400 select liver and muscle-biased DEGs do have known roles within the brain, such as the liver-enriched
401 lipid metabolism gene *Apolipoprotein C-I* (*Apoc1*)³⁰. *APOC1* dysregulation is implicated in Alzheimer's
402 disease in humans³¹ and overexpression of *Apoc1* in the mouse brain can impair learning and memory⁷⁹.
403 KDM5C may therefore be crucial for neurodevelopment by fine-tuning the expression of tissue-enriched,
404 dosage-sensitive genes like *Apoc1*. Given germline genes have no known functions within the brain, their
405 impact upon neurodevelopment is currently unknown. Ectopic germline transcripts have been observed in a
406 variety of cancers^{80,81} and can drive brain tumor formation in *Drosophila*⁸², indicating their dysregulation

407 may promote genome instability and cellular de-differentiation. Intriguingly, some models for other chromatin-
408 linked neurodevelopmental disorders also display impaired soma-germline demarcation^{7,83–86}. Like KDM5C,
409 the chromatin regulators underlying these conditions - DNA methyltransferase 3b (DNMT3B), H3K9me1/2
410 methyltransferases G9A/GLP, methyl-CpG -binding protein 2 (MECP2) - primarily silence gene expression.
411 Thus, KDM5C is among a growing cohort of chromatin-linked neurodevelopmental disorders that have a
412 similar underlying phenotype of germline versus soma dysregulation However, further research is required to
413 determine the impact of these germline genes and the extent to which this phenomenon occurs in humans.

414 Materials and Methods

415 Classifying tissue-enriched and germline-enriched genes

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

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

426 Cell culture

427 We utilized our previously established cultures of male wild-type and *Kdm5c* knockout (-KO) embryonic
428 stem cells⁴⁵. Sex was confirmed by genotyping *Uba1/Uba1y* on the X and Y chromosomes with the following
429 primers: 5'-TGGATGGTGTGCCAATG-3', 5'-CACCTGCACGTTGCCCTT-3'. Deletion of *Kdm5c* was
430 confirmed through the primers 5'-ATGCCCATATTAAGAGTCCCTG-3', 5'-TCTGCCTTGATGGGACTGTT-3',
431 and 5'-GGTTCTCAACACTCACATAGTG-3'.

432 Embryonic stem cells (ESCs) and epiblast-like cells were cultured using previously established
433 methods⁴⁰. Briefly, ESCs were initially cultured in primed ESC (pESC) media consisting of KnockOut
434 DMEM (Gibco#10829–018), fetal bovine serum (Gibco#A5209501), KnockOut serum replacement
435 (Invitrogen#10828–028), Glutamax (Gibco#35050-061), Anti-Anti (Gibco#15240-062), MEM Non-essential
436 amino acids (Gibco#11140-050), and beta-mercaptoethanol (Sigma#M7522). They were then transitioned
437 into ground-state, “naïve” ESCs (nESCs) by culturing for four passages in N2B27 media containing
438 DMEM/F12 (Gibco#11330–032), Neurobasal media (Gibco#21103–049), Gluamax, Anti-Anti, N2 sup-

439 plement (Invitrogen#17502048), and B27 supplement without vitamin A (Invitrogen#12587-010), and
440 beta-mercaptoethanol. Both pESC and nESC media were supplemented with 3 μ M GSK3 inhibitor
441 CHIR99021 (Sigma #SML1046-5MG), 1 μ M MEK inhibitor PD0325901 (Sigma #PZ0162-5MG), and 1,000
442 units/mL leukemia inhibitory factor (LIF, Millipore#ESG1107).

443 nESCs were differentiated into epiblast-like cells (EpiLCs, 48 hours) and extendend EpiLCs (exEpiLCs,
444 96 hours) by culturing in N2B27 media containing DMEM/F12, Neurobasal media, Gluamax, Anti-Anti,
445 N2 supplement, B27 supplement (Invitrogen#17504044), beta-mercaptoethanol, fibroblast growth factor 2
446 (FGF2, R&D Biotechne 233-FB), and activin A (R&D Biotechne 338AC050CF), as previously described⁴⁰.

447 **Immunocytochemistry (ICC)**

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

457 **RNA sequencing (RNA-seq)**

458 After ensuring read quality via FastQC (v0.11.8), reads were then mapped to the mm10 *Mus musculus*
459 genome (Gencode) using STAR (v2.5.3a), during which we removed duplicates and kept only uniquely
460 mapped reads. Count files were generated by FeatureCounts (Subread v1.5.0), and BAM files were
461 converted to bigwigs using deeptools (v3.1.3) and visualized by the UCSC genome browser. RStudio (v3.6.0)
462 was then used to analyze counts files by DESeq2 (v1.26.0)²⁶ to identify differentially expressed genes
463 (DEGs) with a q-value (p-adjusted via FDR/Benjamini–Hochberg correction) less than 0.1 and a log2 fold
464 change greater than 0.5. For all DESeq2 analyses, log2 fold changes were calculated with IfcShrink using
465 the ashR package⁸⁷. MA-plots were generated by ggpibr (v0.6.0), and Eulerr diagrams were generated by
466 eulerr (v6.1.1). Boxplots and scatterplots were generated by ggpibr (v0.6.0) and ggplot2 (v3.3.2). The Upset
467 plot was generated via the package UpSetR (v1.4.0)⁸⁸. Gene ontology (GO) analyses were performed by
468 the R package enrichPlot (v1.16.2) using the biological processes setting and compareCluster.

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

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

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

481 Genomic DNA (gDNA) from naïve ESCs and extended EpiLCs was extracted using the Wizard Genomic
482 DNA Purification Kit (Promega A1120), following the instructions for Tissue Culture Cells. gDNA from
483 two wild-types and two *Kdm5c*-KOs of each cell type was sent to Novogene for WGBS using the Illumina
484 NovaSeq X Plus platform and sequenced for 150bp paired-end reads (PE150). - bismark - Methylkit

485 **Data availability**

486 **Published datasets**

487 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
488 adult amygdala and hippocampus²⁵ (available at GEO: GSE127722) and male wild-type and *Kdm5c*-KO
489 EpiLCs⁴⁵ (available at GEO: GSE96797).

490 Previously published ChIP-seq experiments included KDM5C in wild-type and *Kdm5c*-KO EpiLCs⁴⁵ (avail-
491 able at GEO: GSE96797) and mouse primary neuron cultures (PNCs) from the cortex and hippocampus¹²
492 (available at GEO: GSE61036). ChIP-seq of histone 3 lysine 4 dimethylation in male wild-type and *Kdm5c*-KO
493 EpiLCs⁴⁵ is also available at GEO: GSE96797. ChIP-seq of histone 3 lysine 4 trimethylation in wild-type and
494 *Kdm5c*-KO male amygdala²⁵ are available at GEO: GSE127817.

496 **Data analysis**

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

499 **Acknowledgements**

- 500 • Jacob Mueller for providing insight in germline gene regulation.
501 • Sundeep Kalantry for providing reagents and expertise in culturing mouse embryonic stem cells and
502 epiblast-like cells
503 • Ilakkiya
504 • Funding sources

505 **References**

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

- 524 10. Jensen, L.R., Amende, M., Gurok, U., Moser, B., Gimmel, V., Tschach, A., Janecke, A.R., Tariverdian,
525 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>.
- 526 11. Carmignac, V., Nambot, S., Lehalle, D., Callier, P., Moortgat, S., Benoit, V., Ghoumid, J., Delobel,
B., Smol, T., Thuillier, C., et al. (2020). Further delineation of the female phenotype with KDM5C
disease causing variants: 19 new individuals and review of the literature. *Clin Genet* *98*, 43–55.
527 <https://doi.org/10.1111/cge.13755>.
- 528 12. Iwase, S., Brookes, E., Agarwal, S., Badeaux, A.I., Ito, H., Vallianatos, C.N., Tomassy, G.S., Kasza, T.,
Lin, G., Thompson, A., et al. (2016). A Mouse Model of X-linked Intellectual Disability Associated with
Impaired Removal of Histone Methylation. *Cell Reports* *14*, 1000–1009. [https://doi.org/10.1016/j.celr.ep.2015.12.091](https://doi.org/10.1016/j.celr
ep.2015.12.091).
- 529 13. Scandaglia, M., Lopez-Atalaya, J.P., Medrano-Fernandez, A., Lopez-Cascales, M.T., Del Blanco, B.,
Lipinski, M., Benito, E., Olivares, R., Iwase, S., Shi, Y., et al. (2017). Loss of Kdm5c Causes Spurious
Transcription and Prevents the Fine-Tuning of Activity-Regulated Enhancers in Neurons. *Cell Rep* *21*,
530 47–59. <https://doi.org/10.1016/j.celrep.2017.09.014>.
- 531 14. Devlin, D.K., Ganley, A.R.D., and Takeuchi, N. (2023). A pan-metazoan view of germline-soma
distinction challenges our understanding of how the metazoan germline evolves. *Current Opinion in
532 Systems Biology* *36*, 100486. <https://doi.org/10.1016/j.coisb.2023.100486>.
- 533 15. Lehmann, R. (2012). Germline Stem Cells: Origin and Destiny. *Cell Stem Cell* *10*, 729–739.
534 <https://doi.org/10.1016/j.stem.2012.05.016>.
- 535 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
536 stem cells by regulating germ cell-related genes. *Elife* *6*. <https://doi.org/10.7554/eLife.21064>.
- 537 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
538 in the early embryo precedes DNA methylation-mediated silencing. *Nat Commun* *12*, 7020. [//doi.org/10.1038/s41467-021-27345-x](https:
539 //doi.org/10.1038/s41467-021-27345-x).
- 540 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–
541 9286. <https://doi.org/10.1073/pnas.1000473107>.

- 542 19. Hackett, J.A., Reddington, J.P., Nestor, C.E., Dunican, D.S., Branco, M.R., Reichmann, J., Reik, W., Surani, M.A., Adams, I.R., and Meehan, R.R. (2012). Promoter DNA methylation couples genome-defence mechanisms to epigenetic reprogramming in the mouse germline. *Development* 139, 3623–3632. <https://doi.org/10.1242/dev.081661>.
- 543
- 544 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 and human. *Hum Mol Genet* 3, 879–884. <https://doi.org/10.1093/hmg/3.6.879>.
- 545
- 546 21. Carrel, L., Hunt, P.A., and Willard, H.F. (1996). Tissue and lineage-specific variation in inactive X chromosome expression of the murine Smcx gene. *Hum Mol Genet* 5, 1361–1366. <https://doi.org/10.1093/hmg/5.9.1361>.
- 547
- 548 22. Sheardown, S., Norris, D., Fisher, A., and Brockdorff, N. (1996). The mouse Smcx gene exhibits developmental and tissue specific variation in degree of escape from X inactivation. *Hum Mol Genet* 5, 1355–1360. <https://doi.org/10.1093/hmg/5.9.1355>.
- 549
- 550 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>.
- 551
- 552 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. <https://doi.org/10.1038/s41598-017-04520-z>.
- 553
- 554 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>.
- 555
- 556 26. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15, 550. <https://doi.org/10.1186/s13059-014-0550-8>.
- 557
- 558 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 Chromosome Pairing in Meiosis. *Science* 300, 1291–1295. <https://doi.org/10.1126/science.1083022>.
- 559
- 560 28. Xiòl, 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>.
- 561
- 562 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 membraneless compartment. *Science* 378, eabq4835. <https://doi.org/10.1126/science.abq4835>.

- 563
- 564 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>.
- 565
- 566 31. Leduc, V., Jasmin-Bélanger, S., and Poirier, J. (2010). APOE and cholesterol homeostasis in Alzheimer's disease. *Trends in Molecular Medicine* *16*, 469–477. <https://doi.org/10.1016/j.molmed.2010.07.008>.
- 567
- 568 32. Mueller, J.L., Skaletsky, H., Brown, L.G., Zaghlul, S., Rock, S., Graves, T., Auger, K., Warren, W.C., Wilson, R.K., and Page, D.C. (2013). Independent specialization of the human and mouse X chromosomes for the male germ line. *Nat Genet* *45*, 1083–1087. <https://doi.org/10.1038/ng.2705>.
- 569
- 570 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>.
- 571
- 572 34. Green, C.D., Ma, Q., Manske, G.L., Shami, A.N., Zheng, X., Marini, S., Moritz, L., Sultan, C., Gurczynski, S.J., Moore, B.B., et al. (2018). A Comprehensive Roadmap of Murine Spermatogenesis Defined by Single-Cell RNA-Seq. *Dev Cell* *46*, 651–667.e10. <https://doi.org/10.1016/j.devcel.2018.07.025>.
- 573
- 574 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. <https://doi.org/10.1371/journal.pgen.1005531>.
- 575
- 576 36. Magnúsdóttir, E., and Surani, M.A. (2014). How to make a primordial germ cell. *Development* *141*, 245–252. <https://doi.org/10.1242/dev.098269>.
- 577
- 578 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*. <https://doi.org/10.1098/rstb.2013.0543>.
- 579
- 580 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>.
- 581
- 582 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. <https://doi.org/10.1016/j.cell.2011.06.052>.
- 583
- 584 40. Samanta, M., and Kalantry, S. (2020). Generating primed pluripotent epiblast stem cells: A methodology chapter. *Curr Top Dev Biol* *138*, 139–174. <https://doi.org/10.1016/bs.ctdb.2020.01.005>.
- 585

- 586 41. Welling, M., Chen, H., Muñoz, J., Musheev, M.U., Kester, L., Junker, J.P., Mischerikow, N., Arbab, M.,
Kuijk, E., Silberstein, L., et al. (2015). DAZL regulates Tet1 translation in murine embryonic stem cells.
587 EMBO Reports 16, 791–802. <https://doi.org/10.15252/embr.201540538>.
- 588 42. Macfarlan, T.S., Gifford, W.D., Driscoll, S., Lettieri, K., Rowe, H.M., Bonanomi, D., Firth, A., Singer, O.,
Trono, D., and Pfaff, S.L. (2012). Embryonic stem cell potency fluctuates with endogenous retrovirus
589 activity. Nature 487, 57–63. <https://doi.org/10.1038/nature11244>.
- 590 43. Suzuki, A., Hirasaki, M., Hishida, T., Wu, J., Okamura, D., Ueda, A., Nishimoto, M., Nakachi, Y.,
Mizuno, Y., Okazaki, Y., et al. (2016). Loss of MAX results in meiotic entry in mouse embryonic and
591 germline stem cells. Nat Commun 7, 11056. <https://doi.org/10.1038/ncomms11056>.
- 592 44. Borgel, J., Guibert, S., Li, Y., Chiba, H., Schübeler, D., Sasaki, H., Forné, T., and Weber, M. (2010).
Targets and dynamics of promoter DNA methylation during early mouse development. Nat Genet 42,
593 1093–1100. <https://doi.org/10.1038/ng.708>.
- 594 45. Samanta, M.K., Gayen, S., Harris, C., Maclary, E., Murata-Nakamura, Y., Malcore, R.M., Porter, R.S.,
Garay, P.M., Vallianatos, C.N., Samollow, P.B., et al. (2022). Activation of Xist by an evolutionarily
595 conserved function of KDM5C demethylase. Nat Commun 13, 2602. <https://doi.org/10.1038/s41467-022-30352-1>.
- 596 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,
597 2474–2479. <https://doi.org/10.1073/pnas.0510813103>.
- 598 47. Lin, Y., Gill, M.E., Koubova, J., and Page, D.C. (2008). Germ Cell-Intrinsic and -Extrinsic Factors
Govern Meiotic Initiation in Mouse Embryos. Science 322, 1685–1687. <https://doi.org/10.1126/science.1166340>.
- 600 48. Endo, T., Mikedis, M.M., Nicholls, P.K., Page, D.C., and De Rooij, D.G. (2019). Retinoic Acid and Germ
601 Cell Development in the Ovary and Testis. Biomolecules 9, 775. <https://doi.org/10.3390/biom9120775>.
- 602 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.
603 Nat Struct Mol Biol. <https://doi.org/10.1038/s41594-023-01038-z>.
- 604 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-
605 Specific Gene Expression. Current Biology 15, 1051–1057. <https://doi.org/10.1016/j.cub.2005.04.060>.
- 606 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
607 during embryonic development. Nat Commun 12, 3582. <https://doi.org/10.1038/s41467-021-23596-w>.

- 608 52. Wang, P.J., McCarrey, J.R., Yang, F., and Page, D.C. (2001). An abundance of X-linked genes
609 expressed in spermatogonia. *Nat Genet* 27, 422–426. <https://doi.org/10.1038/86927>.
- 610 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
611 inactivation. *Nat Genet* 36, 642–646. <https://doi.org/10.1038/ng1368>.
- 612 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>.
- 614 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*
615 14, e1007193. <https://doi.org/10.1371/journal.pgen.1007193>.
- 616 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.
617 <https://doi.org/10.1016/j.molcel.2010.05.004>.
- 618 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,
619 916–921. <https://doi.org/10.1038/35002634>.
- 620 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).
- 622 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
623 Ciliogenesis. *Sci. Signal.* 2. <https://doi.org/10.1126/scisignal.2000602>.
- 624 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.
625 *PLoS Genet* 11, e1005368. <https://doi.org/10.1371/journal.pgen.1005368>.
- 626 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.
627 <https://doi.org/10.1038/srep20435>.
- 628 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>.

- 630 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>.
- 631
- 632 64. Guo, X., Wang, L., Li, J., Ding, Z., Xiao, J., Yin, X., He, S., Shi, P., Dong, L., Li, G., et al. (2015).
Structural insight into autoinhibition and histone H3-induced activation of DNMT3A. *Nature* *517*,
633 640–644. <https://doi.org/10.1038/nature13899>.
- 634 65. Meissner, A., Mikkelsen, T.S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein,
B.E., Nusbaum, C., Jaffe, D.B., et al. (2008). Genome-scale DNA methylation maps of pluripotent and
635 differentiated cells. *Nature* *454*, 766–770. <https://doi.org/10.1038/nature07107>.
- 636 66. Nassar, L.R., Barber, G.P., Benet-Pagès, A., Casper, J., Clawson, H., Diekhans, M., Fischer, C.,
Gonzalez, J.N., Hinrichs, A.S., Lee, B.T., et al. (2023). The UCSC Genome Browser database: 2023
637 update. *Nucleic Acids Research* *51*, D1188–D1195. <https://doi.org/10.1093/nar/gkac1072>.
- 638 67. Akalin, A., Kormaksson, M., Li, S., Garrett-Bakelman, F.E., Figueiroa, M.E., Melnick, A., and Mason,
C.E. (2012). methylKit: A comprehensive R package for the analysis of genome-wide DNA methylation
639 profiles. *Genome Biol* *13*, R87. <https://doi.org/10.1186/gb-2012-13-10-r87>.
- 640 68. Yang, M., Yu, H., Yu, X., Liang, S., Hu, Y., Luo, Y., Izsvák, Z., Sun, C., and Wang, J. (2022). Chemical-
induced chromatin remodeling reprograms mouse ESCs to totipotent-like stem cells. *Cell Stem Cell*
641 29, 400–418.e13. <https://doi.org/10.1016/j.stem.2022.01.010>.
- 642 69. Al Adhami, H., Vallet, J., Schaal, C., Schumacher, P., Bardet, A.F., Dumas, M., Chicher, J., Hammann,
P., Daujat, S., and Weber, M. (2023). Systematic identification of factors involved in the silencing
of germline genes in mouse embryonic stem cells. *Nucleic Acids Research* *51*, 3130–3149. <https://doi.org/10.1093/nar/gkad071>.
- 643
- 644 70. Zhang, J., Zhang, M., Acampora, D., Vojtek, M., Yuan, D., Simeone, A., and Chambers, I. (2018).
OTX2 restricts entry to the mouse germline. *Nature* *562*, 595–599. <https://doi.org/10.1038/s41586-018-0581-5>.
- 645
- 646 71. Tahiliani, M., Mei, P., Fang, R., Leonor, T., Rutenberg, M., Shimizu, F., Li, J., Rao, A., and Shi, Y.
(2007). The histone H3K4 demethylase SMCX links REST target genes to X-linked mental retardation.
Nature *447*, 601–605. <https://doi.org/10.1038/nature05823>.
- 647
- 648 72. Endo, T., Romer, K.A., Anderson, E.L., Baltus, A.E., De Rooij, D.G., and Page, D.C. (2015). Pe-
riodic retinoic acid–STRA8 signaling intersects with periodic germ-cell competencies to regulate
649 spermatogenesis. *Proc. Natl. Acad. Sci. U.S.A.* *112*. <https://doi.org/10.1073/pnas.1505683112>.
- 650 73. Bowles, J., Knight, D., Smith, C., Wilhelm, D., Richman, J., Mamiya, S., Yashiro, K., Chawengsak-
sophak, K., Wilson, M.J., Rossant, J., et al. (2006). Retinoid Signaling Determines Germ Cell Fate in
651 Mice. *Science* *312*, 596–600. <https://doi.org/10.1126/science.1125691>.

- 652 74. Karimi, M.M., Goyal, P., Maksakova, I.A., Bilenky, M., Leung, D., Tang, J.X., Shinkai, Y., Mager, D.L.,
653 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,
654 676–687. <https://doi.org/10.1016/j.stem.2011.04.004>.
- 655 75. Auclair, G., Borgel, J., Sanz, L.A., Vallet, J., Guibert, S., Dumas, M., Cavelier, P., Girardot, M., Forné,
T., Feil, R., et al. (2016). EHMT2 directs DNA methylation for efficient gene silencing in mouse
embryos. *Genome Res.* 26, 192–202. <https://doi.org/10.1101/gr.198291.115>.
- 656 76. Tatsumi, D., Hayashi, Y., Endo, M., Kobayashi, H., Yoshioka, T., Kiso, K., Kanno, S., Nakai, Y., Maeda,
I., Mochizuki, K., et al. (2018). DNMTs and SETDB1 function as co-repressors in MAX-mediated
repression of germ cell-related genes in mouse embryonic stem cells. *PLoS ONE* 13, e0205969.
657 <https://doi.org/10.1371/journal.pone.0205969>.
- 658 77. Aubert, Y., Egolf, S., and Capell, B.C. (2019). The Unexpected Noncatalytic Roles of Histone Modifiers
in Development and Disease. *Trends in Genetics* 35, 645–657. <https://doi.org/10.1016/j.tig.2019.06.004>.
- 659 78. Morgan, M.A.J., and Shilatifard, A. (2020). Reevaluating the roles of histone-modifying enzymes
and their associated chromatin modifications in transcriptional regulation. *Nat Genet* 52, 1271–1281.
660 <https://doi.org/10.1038/s41588-020-00736-4>.
- 661 79. Abildayeva, K., Berbée, J.F.P., Blokland, A., Jansen, P.J., Hoek, F.J., Meijer, O., Lütjohann, D., Gautier,
T., Pillot, T., De Vente, J., et al. (2008). Human apolipoprotein C-I expression in mice impairs learning
and memory functions. *Journal of Lipid Research* 49, 856–869. <https://doi.org/10.1194/jlr.M700518JLR200>.
- 662 80. Nielsen, A.Y., and Gjerstorff, M.F. (2016). Ectopic Expression of Testis Germ Cell Proteins in Cancer
663 and Its Potential Role in Genomic Instability. *Int J Mol Sci* 17. <https://doi.org/10.3390/ijms17060890>.
- 664 81. Adebayo Babatunde, K., Najafi, A., Salehipour, P., Modarressi, M.H., and Mobasher, M.B. (2017).
Cancer/Testis genes in relation to sperm biology and function. *Iranian Journal of Basic Medical
665 Sciences* 20. <https://doi.org/10.22038/ijbms.2017.9259>.
- 666 82. Janic, A., Mendizabal, L., Llamazares, S., Rossell, D., and Gonzalez, C. (2010). Ectopic expression
of germline genes drives malignant brain tumor growth in *Drosophila*. *Science* 330, 1824–1827.
667 <https://doi.org/10.1126/science.1195481>.
- 668 83. Bonefas, K.M., and Iwase, S. (2021). Soma-to-germline transformation in chromatin-linked neurode-
669 velopmental disorders? *FEBS J.* <https://doi.org/10.1111/febs.16196>.
- 670 84. Velasco, G., Walton, E.L., Sterlin, D., Hédouin, S., Nitta, H., Ito, Y., Fouyssac, F., Mégarbané, A.,
Sasaki, H., Picard, C., et al. (2014). Germline genes hypomethylation and expression define a
molecular signature in peripheral blood of ICF patients: Implications for diagnosis and etiology.
671 *Orphanet J Rare Dis* 9, 56. <https://doi.org/10.1186/1750-1172-9-56>.

- 673
- 674 85. Walton, E.L., Francastel, C., and Velasco, G. (2014). Dnmt3b Prefers Germ Line Genes and Cen-
675 tromeric Regions: Lessons from the ICF Syndrome and Cancer and Implications for Diseases. *Biology*
(Basel) 3, 578–605. <https://doi.org/10.3390/biology3030578>.
- 676 86. Samaco, R.C., Mandel-Brehm, C., McGraw, C.M., Shaw, C.A., McGill, B.E., and Zoghbi, H.Y. (2012).
677 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>.
- 678 87. Stephens, M. (2016). False discovery rates: A new deal. *Biostat*, kxw041. <https://doi.org/10.1093/biostatistics/kxw041>.
- 680 88. Conway, J.R., Lex, A., and Gehlenborg, N. (2017). UpSetR: An R package for the visualization of
681 intersecting sets and their properties. *Bioinformatics* 33, 2938–2940. <https://doi.org/10.1093/bioinformatics/btx364>.
- 682 89. Li, H., Liang, Z., Yang, J., Wang, D., Wang, H., Zhu, M., Geng, B., and Xu, E.Y. (2019). DAZL
683 is a master translational regulator of murine spermatogenesis. *Natl Sci Rev* 6, 455–468. <https://doi.org/10.1093/nsr/nwy163>.
- 684 90. Mikedis, M.M., Fan, Y., Nicholls, P.K., Endo, T., Jackson, E.K., Cobb, S.A., De Rooij, D.G., and Page,
685 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>.

686 **Figures and Tables**

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

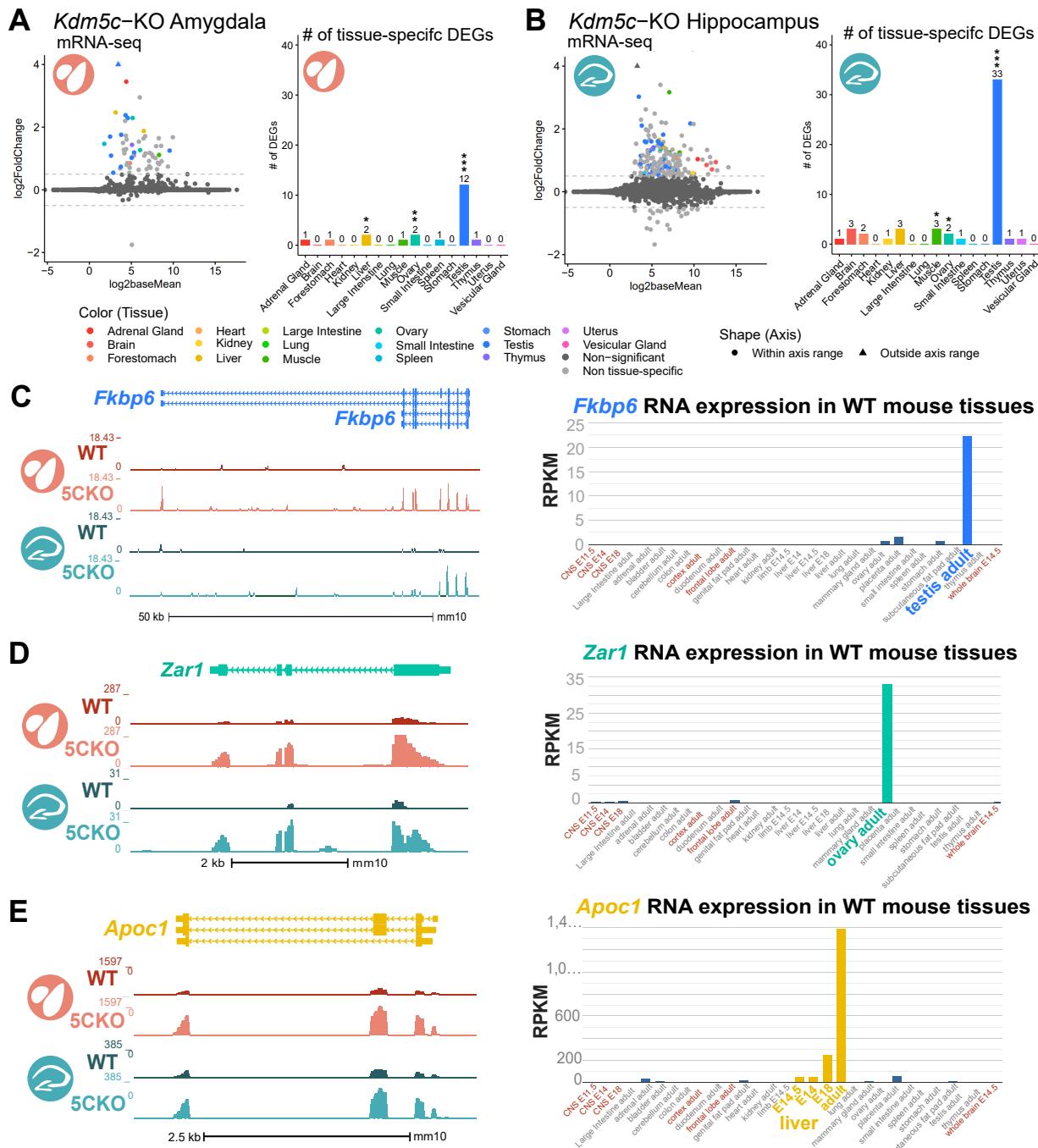


Figure 1: Tissue-enriched genes are misexpressed in the *Kdm5c*-KO brain. **A-B.** Expression of tissue-enriched genes (Li et al 2017) in the male *Kdm5c*-KO amygdala (A) and hippocampus (B). Left - MA plot of mRNA-sequencing. Right - Number of tissue-enriched differentially expressed genes (DEGs). *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

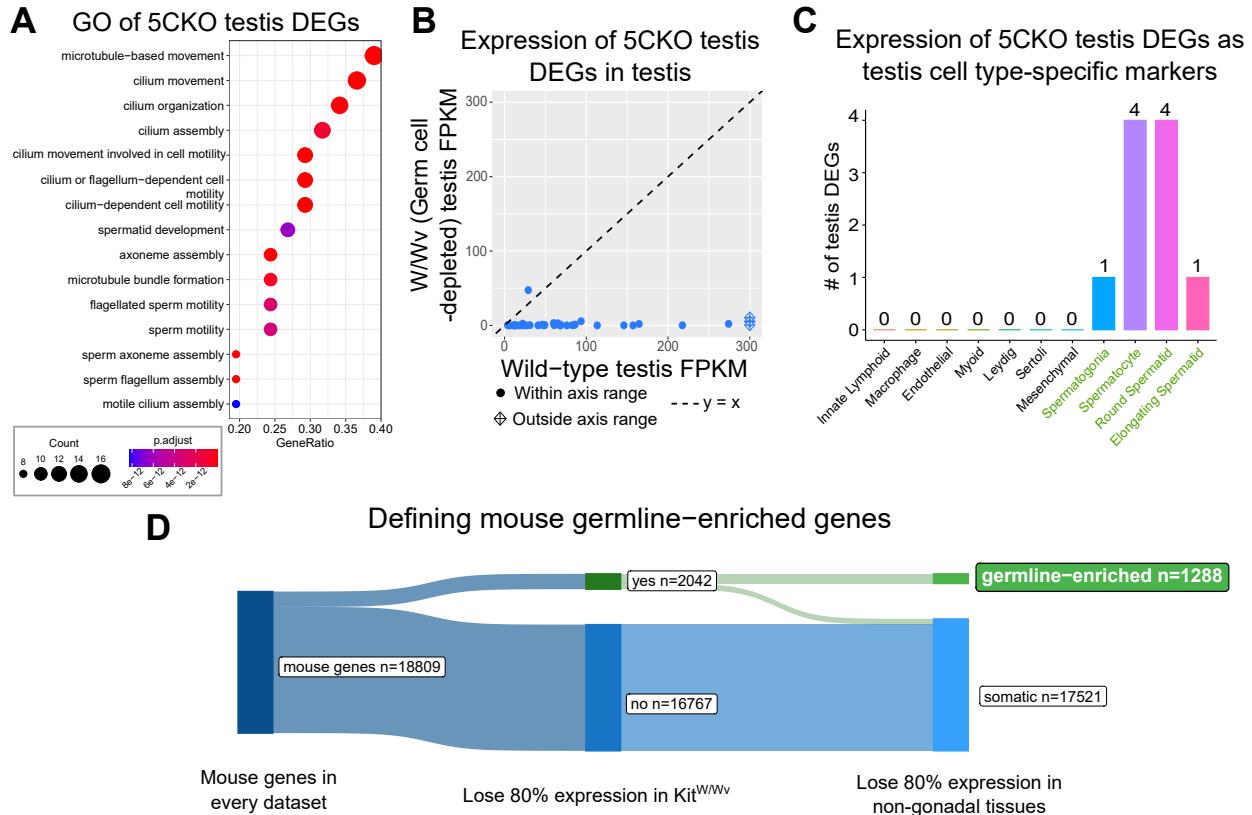


Figure 2: Aberrant transcription of germline genes in the *Kdm5c*-KO in the brain. **A.** enrichPlot gene ontology (GO) of *Kdm5c*-KO amygdala and hippocampus testis-enriched DEGs **B.** Expression of testis DEGs in wild-type (WT) testis versus germ cell-depleted (W/W_v) testis (Mueller et al 2013).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

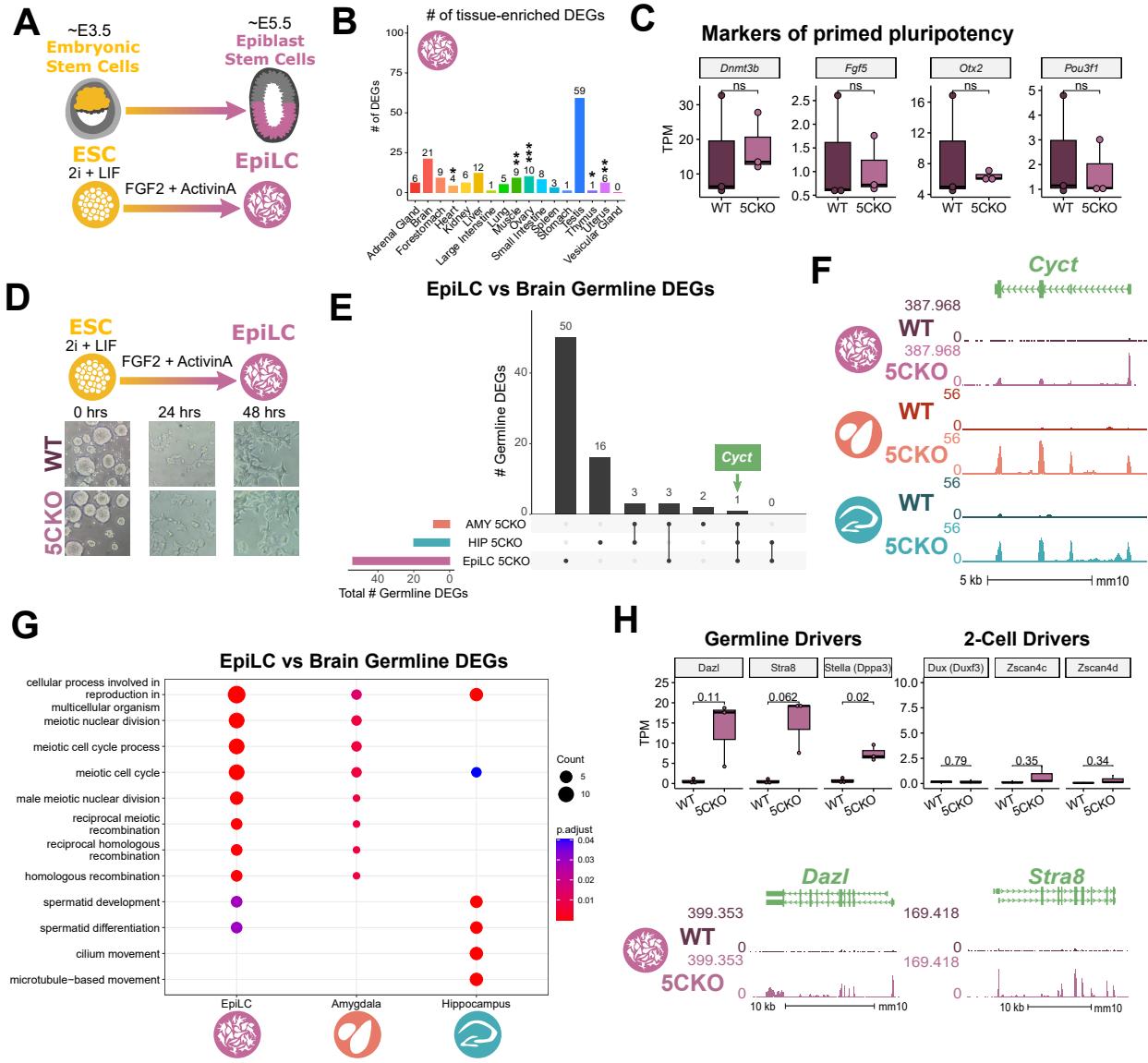


Figure 3: *Kdm5c*-KO epiblast-like cells express key drivers of germline identity **A.** Top - Diagram of *in vivo* differentiation of embryonic stem cells (ESCs) of the inner cell mass into epiblast stem cells. Bottom - *in vitro* differentiation of ESCs into epiblast-like cells (EpiLCs). **B.** Number 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, *CytC*, 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. **I.** 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.

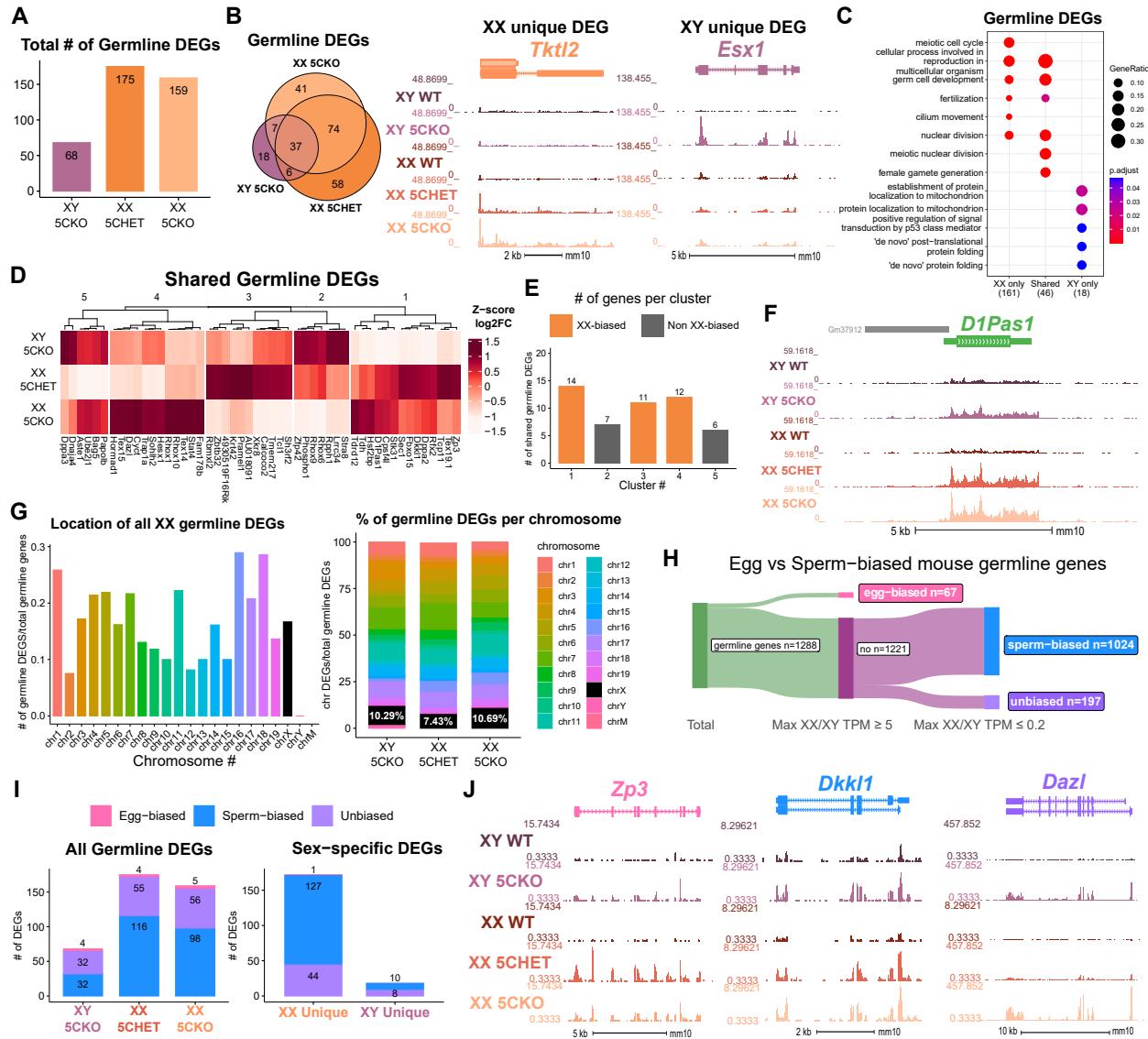


Figure 4: Chromosomal sex influences *Kdm5c*-KO germline gene misexpression. **A-B.** Expression of tissue-enriched genes (Li et al 2017) in the male *Kdm5c*-KO amygdala (A) and hippocampus (B). Left - MA plot of mRNA-seq. Right - Number of tissue-enriched differentially expressed genes (DEGs). * p<0.05, ** 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.

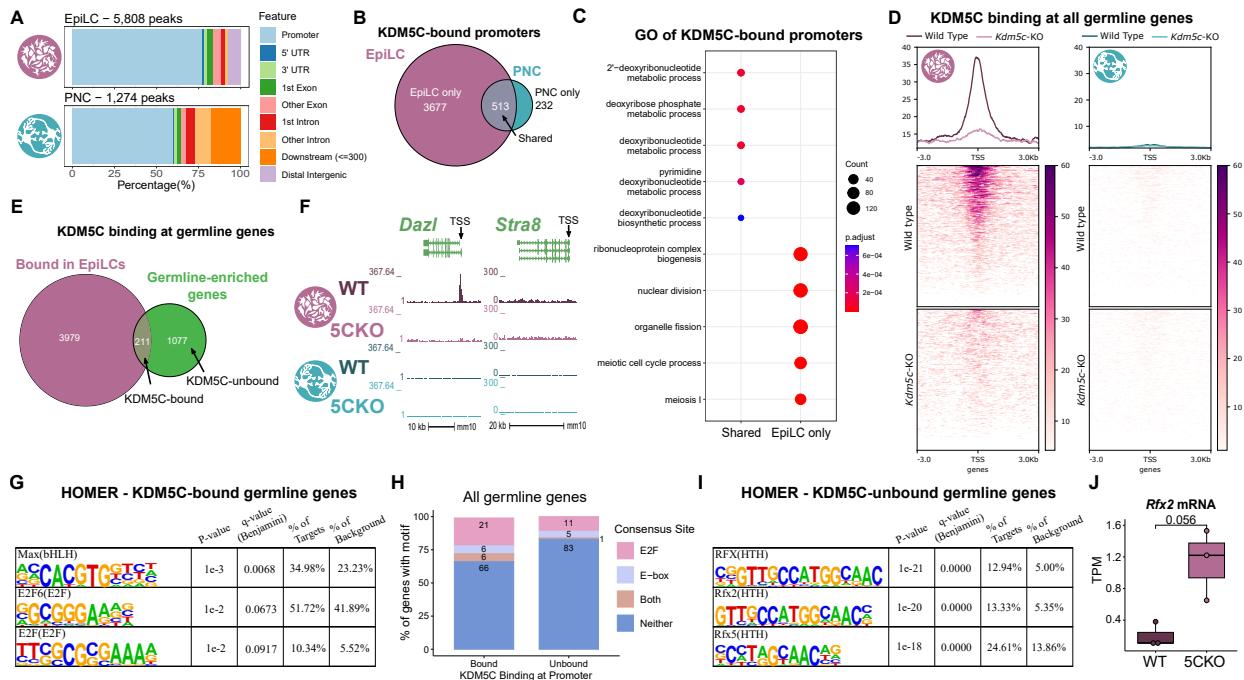


Figure 5: KDM5C binds to a subset of germline gene promoters during early embryogenesis **A.** ChIPseeker localization of KDM5C peaks at different genomic regions in EpiLCs (top) and hippocampal and cortex primary neuron cultures (PNCs, bottom). **B.** Overlap of genes with KDM5C bound to their promoters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

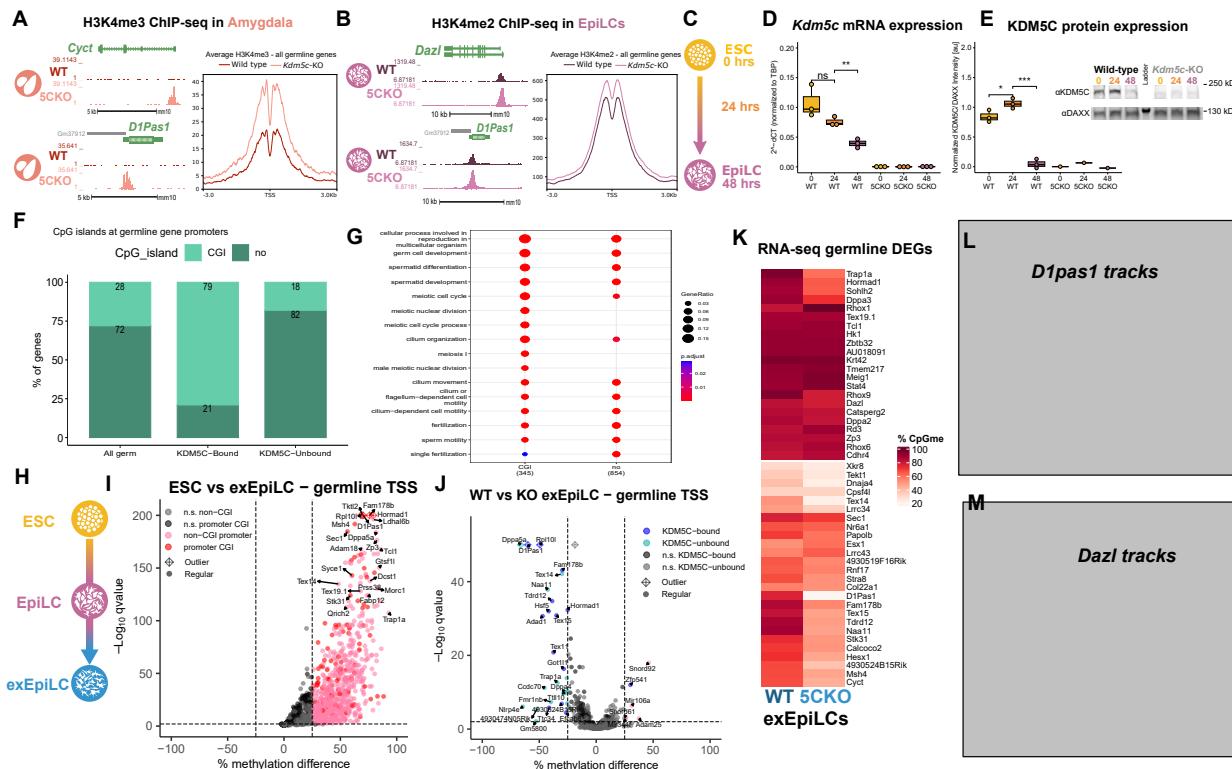


Figure 6: KDM5C's catalytic activity promotes long-term silencing of germline genes via DNA methylation. **A.** Left - Bigwigs of representative histone 3 lysine 4 trimethylation (H3K4me3) ChIP-seq peaks at two germline genes in the wild-type (WT) and *Kdm5c*-KO (5CKO) adult amygdala. Right - Average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

691 Notes

692 Things to do

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

697 Dazl

698 We were particularly interested in the aberrant transcription of *Dazl*, since it is essential for germ cell
699 development and promotes the translation of germline mRNAs⁸⁹. A significant portion of germline transcripts
700 misexpressed in *Kdm5c*-KO EpiLCs are known binding targets of DAZL, including *Stra8*⁹⁰ ($p = 1.698e-07$,
701 Fisher's Exact Test). This suggests expression of DAZL protein could promote the translation of other
702 aberrant germline transcripts, influencing their ability to impact *Kdm5c*-KO cellular function. We thus tested
703 DAZL protein expression in *Kdm5c*-KO EpiLCs through immunocytochemistry (Figure 3I). We observed
704 about 25% of *Kdm5c*-KO EpiLCs expressed DAZL protein and it was localized to the cytoplasm ($p = 0.0015$,
705 Welch's t-test), consistent with the pattern of DAZL expression in spermatogonia⁹⁰. Altogether these results
706 suggest tissue-specific genes are misexpressed during *Kdm5c*-KO embryogenesis, including key drivers of
707 germline identity that can be translated into protein.

- 708 • We additionally found *Kdm5c*-KO EpiLCs ectopically express DAZL protein that is localized to the
709 cytoplasm, similar to its morphology in spermatogonia⁹⁰. **note: maybe just put in results.** Could
710 move around depending upon if I get pheno working.

711 Discussion notes

- 712 • For other paper:
 - 713 – for methods: Heatmaps of gene expression were generated using the base R functions scale and
714 hclust and visualized using the R package ComplexHeatmap (v2.12.1).
 - 715 – * 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>
- 717 • Talk about different motifs if we do see differences there and if it explains direct vs indirect dysregulation
 - 718 – Could also mention PRC1.6/KDM5C binds pcgf6 in ESCs
- 719 • Maybe talk about other regulators/if anything is known about them long term? Or save for 2nd paper.

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

- 748 • Altogether, these data indicate that while some germline genes are misexpressed due to direct loss of
749 KDM5C binding during emryogenesis, secondary downstream mechanisms can also promote their
750 aberrant transcription.
- 751 • Papers to read/reference:
752 – Reconstitution of the Mouse Germ Cell Specification Pathway in Culture by Pluripotent Stem Cells:
753 [https://www.cell.com/fulltext/S0092-8674\(11\)00771-9](https://www.cell.com/fulltext/S0092-8674(11)00771-9)
754 – 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/>

756 **Figure outline:**

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

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

769 **Figure 3: *Kdm5c*-KO epiblast-like cells express key drivers regulators of germline identity** * A) ESC
770 to EpiLC differentiation Left - Morphology is unchanged, * B) 5CKO EpiLCs express EpiLC differentiation
771 genes similar to WT lvs * C) Male EpiLCs express germline genes (example Cyct again) * Overlap between
772 brain and EpiLC germline genes - show they're mostly unique * GO of Brain and EpiLC germline genes
773 (meiotic enriched) * Bigwigs or TPM of master regulators * Show that while some are also 2-cell genes
774 (Dazl), 2-cell specific genes aren't dysregulated (Zscan4). Important point because published KDM5C dazl
775 paper is saying KDM5C is a 2-cell regulator, but as far as I can tell only genes shared between germline and
776 2-cell are dysregulated.

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

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

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

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

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

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

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

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

821 **Germline gene repression background:**

822 Interestingly, some of the ectopic testis transcripts identified in the *Kdm5c*-KO brain are typically ex-
823 pressed in germ cells¹³. Unlike somatic cells, germ cells (e.g. sperm and eggs) undergo meiosis and pass
824 on their genetic material to the next generation. The germline and the soma are typically distinguished during
825 early embryogenesis, when germline genes are silenced in epiblast stem cells soon after implantation and
826 only reactivated in a subset to form the germline. Chromatin regulators play a key role in decommissioning
827 germline genes as the embryo transitions from naïve to primed pluripotency by placing repressive histone
828 H2A lysine 119 monoubiquitination (H2AK119ub1)¹⁶, histone 3 lysine 9 trimethylation (H3K9me3)^{16,17}, and
829 DNA CpG methylation^{17–19} at germline gene promoters. KDM5C may also be involved in this early decom-
830 missioning of germline genes, as re-expression of KDM5C in neurons fails to suppress their dysregulation¹³.
831 In support of this, KDM5C was very recently shown to repress *Deleted in azoospermia like (Dazl)*, a key
832 regulator of germline development, in mouse embryonic stem cells (ESCs)^{49,69}. In support of this, two
833 independent screens in mouse embryonic stem cells (ESCs) recently identified KDM5C as a repressor of
834 *Deleted in azoospermia like (Dazl)*, a key regulator of germline development. However, KDM5C's role in
835 embryonic germline gene repression is currently unclear, given that Dazl is also expressed in ESCs and in
836 the 2-cell stage and germline gene misexpression has yet to be globally characterized during *Kdm5c*-KO
837 embryogenesis.