

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

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

27 Introduction

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

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

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

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

68 Results

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

70 Previous RNA sequencing (RNA-seq) of the adult male *Kdm5c*-KO hippocampus revealed ectopic
71 expression of some testis genes⁸. It is currently unclear if the testis is the only tissue type misexpressed in
72 the *Kdm5c*-KO brain. We thus characterized the role of KDM5C in brain tissue identity by systematically
73 assessing the dysregulation of genes enriched in 17 mouse tissues²⁴, using our published mRNA-seq
74 datasets of the male adult amygdala and hippocampus for mice with constitutive knockout of *Kdm5c*²⁵.

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

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

90 Although not consistent across brain regions, we also found significant enrichment of DEGs biased
91 towards two non-gonadal tissues - the liver (Amygdala p = 0.0398, Odds Ratio = 6.58, Fisher's Exact
92 Test) and the muscle (Hippocampus p = 0.0104, Odds Ratio = 6.95, Fisher's Exact Test) (Figure 1A-B). A
93 liver-biased DEG dysregulated in both the hippocampus and amygdala is *Apolipoprotein C-I* (*Apoc1*), a
94 lipoprotein metabolism and transport gene³⁰ that has been implicated in Alzheimer's disease³¹ (Figure 1E).

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

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

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

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

117 We then wanted to characterize germline gene misexpression with *Kdm5c* loss more deeply, but lacked
118 a comprehensive list of mouse germline genes. We therefore generated a list of germline-enriched genes
119 using RNA-seq datasets of *Kit*^{W/Wv} mice that included males and females at embryonic day 12, 14, and 16³⁵
120 and adult male testes³². We defined genes as germline-enriched if their expression met the following criteria:
121 1) their expression is greater than 1 FPKM in wild-type gonads 2) their expression in any non-gonadal tissue
122 of adult wild type mice²⁴ does not exceed 20% of their maximum expression in the wild-type germline, and
123 3) their expression in the germ cell-depleted gonads, for any sex or time point, does not exceed 20% of
124 their maximum expression in the wild-type germline. These criteria yielded 1,288 germline-enriched genes
125 (Figure 2D), which was hereafter used as a resource to globally assess germline gene misexpression with
126 *Kdm5c* loss (Supplementary table 1).

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

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

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

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

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

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

163 It is currently unknown if the misexpression of germline genes is influenced by sex, as previous studies
164 on germline gene repressors have focused on males^{16–18,50,51}. Sex is particularly pertinent in the case of
165 KDM5C because it lies on the X chromosome and partially escapes X chromosome inactivation (XCI), resulting
166 in a higher dosage in females^{20–23}. We therefore explored the impact of chromosomal sex upon germline
167 gene misexpression by comparing their dysregulation in male *Kdm5c* hemizygous knockout (XY *Kdm5c*-
168 KO), female homozygous knockout (XX *Kdm5c*-KO), and female heterozygous knockout (XX *Kdm5c*-HET)
169 EpiLCs.⁴⁵.

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

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

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

191 While many germline genes act in both the male and female germline, some display sex-biased expression
192 or have functions unique to eggs and sperm. Therefore, we wondered if *Kdm5c* mutant males were more
193 likely to express sperm genes and if mutant females would instead express egg genes. To comprehensively
194 assess whether germline gene sex corresponds with *Kdm5c* mutant sex, we filtered our list of germline-

enriched genes for egg and sperm-biased genes. We defined germ cell sex-biased genes as those whose expression in the opposite sex, at any time point, is no greater than 20% of the gene's maximum expression in a given sex. This yielded 67 egg-biased, 1,024 sperm-biased, and 197 unbiased germline-enriched genes (Figure 4H). We found egg, sperm, and unbiased germline genes were dysregulated in all *Kdm5c* mutants (Figure 4I-J). Germline genes dysregulated exclusively in *Kdm5c* mutant males and females were also not biased towards their corresponding germ cell sex (Figure 4I). This indicates sex differences in germline gene dysregulation is not due to sex-specific activation of sperm or egg transcriptional programs. Altogether, these results demonstrate that the sex of *Kdm5c* mutant cells influences the degree of germline gene, independent of germline gene sex.

• note: the edited last sentence ended with “independent of germ cell sex”, but I think if it’s read out of context that sounds like we tested KDM5C in the four core genotypes. But I’m not sure if “germline gene sex” is confusing

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

KDM5C binds to the promoters of several germline genes in embryonic stem cells (ESCs) but its binding is absent in neurons⁸. However, the lack of a comprehensive list of germline-enriched genes prohibited systematic characterization of KDM5C binding at germline gene promoters. Thus, it is unclear if KDM5C is enriched at germline gene promoters, what types of germline genes KDM5C regulates, and if its binding is maintained at any germline genes in neurons.

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

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

229 germline gene promoters in EpiLCs (Figure 5E). One notable gene that lacked KDM5C binding was *Stra8*,
230 even though its mRNA is expressed in *Kdm5c*-KO EpiLCs (Figure 5F). In EpiLCs, KDM5C was only bound
231 to about one third of *Kdm5c*-KO RNA-seq DEG promoters (EpiLC only DEGs: 36%, Brain only DEGs:
232 33.3%), but 3 out of the 4 genes dysregulated in both the brain and EpiLCs (Supplementary figure XXX). In
233 concordance with our gene ontology results, we did not observe KDM5C accumulation at any germline gene
234 promoters in PNCs (Figure 5D, Supplementary figure XXX). Together, these results demonstrate KDM5C
235 is recruited to a subset of germline genes in EpiLCs, including enrichment of meiotic genes, but does not
236 directly regulate germline genes in neurons. Furthermore, the majority of germline mRNAs expressed in
237 *Kdm5c*-KO cells are dysregulated independent of direct KDM5C binding to their promoters.

238 Many germline-specific genes are suppressed by the transcription factor heterodimers E2F/DP1 and
239 MGA/MAX, which respectively bind E2F and E-box motifs^{18,50,51,54,55}. To elucidate if KDM5C is recruited to
240 germline gene promoters by a similar mechanism, we used HOMER to identify transcription factor motifs
241 enriched at KDM5C-bound or unbound germline gene promoters⁵⁶ (TSS ± 500 bp, q-value < 0.1). MAX
242 and E2F6 binding sites were significantly enriched at germline genes bound by KDM5C in EpiLCs, but
243 not at germline genes unbound by KDM5C (MAX q-value: 0.0068, E2F6 q-value: 0.0673, E2F q-value:
244 0.0917) (Figure 5G). One third of KDM5C-bound promoters contained the consensus sequence for either
245 E2F6 (E2F, 5'-TCCCGC-3'), MGA (E-box, 5'-CACGTG-3'), or both, but only 17% of KDM5C-unbound genes
246 contained these motifs (Figure 5H). KDM5C-unbound germline genes were instead enriched for multiple RFX
247 transcription factor binding sites (RFX q-value < 0.0001, RFX2 q-value < 0.0001, RFX5 q-value < 0.0001)
248 (Figure 5I). RFX transcription factors bind X-box motifs⁵⁷ to promote ciliogenesis^{58,59} and among them is
249 RFX2, a central regulator of post-meiotic spermatogenesis^{60,61}. Interestingly, RFX2 mRNA is derepressed
250 in *Kdm5c*-KO EpiLCs (Figure 5J), however it is also not a direct target of KDM5C (Supplementary figure
251 XXX). Thus, RFX2 is a candidate transcription factor for driving the ectopic expression of KDM5C-unbound
252 germline genes in *Kdm5c*-KO cells.

253 **KDM5C is recruited to CpG islands and facilitates *de novo* DNA methylation at 254 germline gene promoters**

255 In the early embryo, germline gene promoters are initially decorated with repressive histone modifications
256 and are then silenced long-term via DNA CpG methylation (CpGme)^{16,17,44,62}. Our results above indicate
257 KDM5C also acts at germline gene promoters during this time period. However, how KDM5C interacts with
258 other germline gene silencing mechanisms is currently unclear. KDM5C is generally thought to suppress
259 transcription through erasure of histone 3 lysine 4 di- and trimethylation (H3K4me2/3)⁹, yet KDM5C's
260 catalytic activity was recently shown to be dispensable for suppressing *Dazl* in undifferentiated ESCs⁴⁹. Since
261 H3K4me3 impedes *de novo* CpGme placement^{63,64}, KDM5C's catalytic activity may instead be required
262 in the post-implantation embryo for long-term silencing of germline genes. In support of this, CpGme is

263 markedly reduced at two germline gene promoters in the *Kdm5c*-KO adult hippocampus⁸.

264 Based on the above observations, we hypothesized KDM5C erases H3K4me3 to promote the initial
265 placement of CpGme at germline gene promoters in EpiLCs. To test this hypothesis, we first characterized
266 KDM5C's substrates, histone 3 lysine 4 di- and trimethylation (H3K4me2/3) in our previously published ChIP-
267 seq datasets of the wild type and *Kdm5c*-KO amygdala²⁵ and EpiLCs⁴⁵. In congruence with previous work in
268 the *Kdm5c*-KO hippocampus⁸, we observed aberrant accumulation of H3K4me3 around the transcription
269 start site (TSS) of germline genes in the *Kdm5c*-KO amygdala (Figure 6A). We additionally found a marked
270 increase in H3K4me2 around the TSS of germline genes in *Kdm5c*-KO EpiLCs (Figure 6B). We next evaluated
271 KDM5C's role in germline gene silencing during ESC to EpiLC differentiation. We first characterized KDM5C's
272 mRNA and protein expression at 0, 24, and 48 hours of differentiation into EpiLCs (Figure 6C). While *Kdm5c*
273 mRNA steadily decreased from 0 to 48 hours of differentiation (Figure 6D), KDM5C protein initially increased
274 from 0 to 24 hours but then decreased to near knockout levels by 48 hours (Figure 6E).

275 In wild-type cells, germline genes are known to accumulate CpG methylation (CpGme) at CpG islands
276 (CGIs) during the transition from naïve to primed pluripotency^{19,44,65}. We first identified how many germline
277 genes contained CGIs using the UCSC genome browser⁶⁶ and found out of 1,288 germline-enriched genes,
278 only 356 (27.64%) contained CGIs within their promoters (TSS ± 500 bp) (Figure 6F). CGI-containing
279 germline genes were enriched for meiotic gene ontologies, including meiotic nuclear division (GO:XXXX,
280 p.adj) and meiosis I (GO:XXXX, p.adj) (Figure 6G).

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

287 We then assessed differentially methylated regions (DMRs) at germline gene promoters in wild-type
288 versus *Kdm5c*-KO exEpiLCs. We found 28 germline promoters were significantly hypomethylated with
289 *Kdm5c* loss (Figure 6I). About half of promoter DMRs hypomethylated in *Kdm5c*-KO exEpiLCs are direct
290 targets of KDM5C in EpiLCs (13 out of 28 hypomethylated DMRs). Hypomethylated promoters included
291 genes consistently dysregulated in *Kdm5c*-KO cells, such as *D1PAs1* (methylation difference = -60.03%,
292 q-value = 3.26e-153) (Figure 6J).

- 293 • Then compared CpGme accumulation in wild-type vs *Kdm5c*-KO exEpiLCs.
- 294 – Out of the CGI genes, which had significantly reduced CpGme in 5CKO (heatmap of % methylation)
295 – Highlight interesting genes affected vs unaffected by KDM5C (especially if same/different from
296 E2F6, PRC1.6, Setdb1 targets)
297 – Although wild-type cells accumulated high levels of DNA methylation at germline gene promoters
298 over the course of ESCs to exEpiLC differentiation, DNA methylation was markedly reduced in

- 299 *Kdm5c*-KO exEpiLCs (Figure 6F).
300 * CGI vs Non-CGI germline genes, any significant differences?
301 • Altogether, this suggests KDM5C is dynamically regulated during ESC to EpiLC differentiation and
302 promotes germline gene silencing via DNA methylation at CGIs during early embryogenesis.

303 **Discussion**

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

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

330 Although expressed in naïve ESCs, *Dazl* and other “self-renewal” germline genes are silenced during
331 differentiation into epiblast stem cells/EpiLCs^{17,44}. We found that while *Kdm5c*-KO EpiLCs also expressed
332 *Dazl*, they did not express 2-cell specific genes. Misexpression of germline genes in *Kdm5c*-KO EpiLCs could

333 indicate they are differentiating into primordial germ cell-like cells (PGCLCs)^{36,37,39}. Yet, *Kdm5c*-KO EpiLCs
334 had normal cellular morphology and properly expressed markers for primed pluripotency, including *Otx2*
335 which is known to repress EpiLC differentiation into PGCs/PGCLCs⁷⁰. Altogether, these data suggest *Kdm5c*-
336 KO germline gene misexpression occurs ectopically in conjunction with typical developmental programs and
337 the 2-cell-like state observed in *Kdm5c*-KO ESCs is an indirect consequence of KDM5C's primary role in
338 germline gene repression.

339 Although eggs and sperm employ the same transcriptional programs for shared functions, e.g. PGC
340 formation, meiosis, and genome defense, some germline genes are sex specific. We found *Kdm5c* mutant
341 males and females expressed both sperm and egg-biased genes, indicating the mechanism of derepression
342 is independent of a given germline gene's sex. However, organismal sex did greatly influence the degree of
343 germline gene dysregulation, as female *Kdm5c*-KO EpiLCs had over double the number of germline-enriched
344 DEGs compared to males. X chromosome inactivation (XCI) defects could explain why *Kdm5c* knockout
345 females are more prone to germline gene dysregulation, given the X chromosome harbors a large number
346 of spermatogenesis genes^{52,53} and loss of *Kdm5c* impairs XCI⁴⁵. However, female germline DEGs were
347 not significantly biased towards the X chromosome. Sex differences in germline gene suppression may be
348 instead connected to females having a higher dose of KDM5C than males, due to its partial escape from
349 XCI²⁰⁻²³. Intriguingly, females with heterozygous loss of *Kdm5c* also had over double the number of germline
350 DEGs than hemizygous knockout males, even though their level of KDM5C should be roughly equivalent to
351 that of wild-type males. Altogether, these results suggest female EpiLCs are more prone to transitioning
352 to a germ cell-like state than males and require more KDM5C to maintain somatic cellular identity. Future
353 studies are required to illuminate if this phenomenon is unique to sexually dimorphic chromatin regulators or
354 a general feature of female cells.

355 It is yet unknown how KDM5C is recruited to germline gene promoters, given that KDM5C itself does
356 not contain domains for sequence-specific binding⁹. In HeLa cells and ESCs^{49,71}, KDM5C associates with
357 members of the polycomb repressive complex 1.6 (PRC1.6), which is recruited to germline gene promoters
358 through E2F6/DP1 and MGA/MAX heterodimers^{16,55}. While MAX and E2F6 motifs were enriched at KDM5C-
359 bound germline genes in EpiLCs, only about one-third of promoters contained their consensus sequence.
360 Additionally, some germline genes can become active in *Kdm5c*-KO cells independent of KDM5C binding to
361 their promoters. Germline genes unbound by KDM5C were primarily late-stage spermatogenesis genes and
362 significantly enriched for RFX transcription factors, including RFX2, a central regulator of spermiogenesis^{60,61}.
363 Another notable KDM5C-unbound DEG is *Stra8*, a meiotic transcription factor that is turned on in germ cells
364 with retionic acid signaling and DAZL expression^{72,73}. Misexpression of *Dazl* and *Rfx2* and their downstream
365 targets in *Kdm5c*-KO cells suggests that once activated, ectopic germline genes can then turn on other
366 aberrant germline programs to loosely mimic germ cell development.

367 Unlike the previously characterized germline gene suppressors that place repressive histone
368 modifications^{16,17,19,62,74-76}, KDM5C removes the active histone mark histone 3 lysine 4 di- and trimethylation

369 (H3K4me2/3)⁹. Yet, KDM5C's catalytic activity may not be required for germline gene silencing, as it
370 was recently found to be dispensable for repressing *Dazl* in ESCs⁴⁹. Emerging work indicates many
371 histone-modifying enzymes have non-catalytic functions that can influence gene expression, sometimes even
372 more potently than their catalytic roles^{77,78}. Although dispensable in ESCs, KDM5C's catalytic activity could
373 be necessary to silence germline genes in EpiLCs, as H3K4me3 impedes *de novo* CpGme placement^{63,64}.
374 In support of this, CpGme is significantly eroded at at least two germline promoters in the adult *Kdm5c*-KO
375 hippocampus⁸. We found *Kdm5c*-KO cells accumulate H3K4me2/3 around the TSS of germline genes
376 and have impaired CpGme placement at germline CpG islands. KDM5C's shifting role in germline gene
377 repression during the transition from naïve to primed pluripotency indicates chromatin modifiers can switch
378 between catalytic and non-catalytic gene regulatory mechanisms at the same loci over development.

- 379 • DNAme and CpG islands
- 380 – Loss of many chromatin regulators affects CpGme (Multiple, non-redundant silencing mechanisms) - 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.

404 KDM5C may therefore be crucial for neurodevelopment by fine-tuning the expression of tissue-enriched,
405 dosage-sensitive genes like *Apoc1*. Given germline genes have no known functions within the brain, their
406 impact upon neurodevelopment is currently unknown. Ectopic germline transcripts have been observed in a
407 variety of cancers^{80,81} and can drive brain tumor formation in *Drosophila*⁸², indicating their dysregulation
408 may promote genome instability and cellular de-differentiation. Intriguingly, some models for other chromatin-
409 linked neurodevelopmental disorders also display impaired soma-germline demarcation^{7,83–86}. Like KDM5C,
410 the chromatin regulators underlying these conditions - DNA methyltransferase 3b (DNMT3B), H3K9me1/2
411 methyltransferases G9A/GLP, methyl-CpG -binding protein 2 (MECP2) - primarily silence gene expression.
412 Thus, KDM5C is among a growing cohort of chromatin-linked neurodevelopmental disorders that have a
413 similar underlying phenotype of germline versus soma dysregulation. However, further research is required to
414 determine the impact of these germline genes and the extent to which this phenomenon occurs in humans.

415 Materials and Methods

416 Classifying tissue-enriched and germline-enriched genes

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

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

427 Cell culture

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

433 Embryonic stem cells (ESCs) and epiblast-like cells were cultured using previously established
434 methods⁴⁰. Briefly, ESCs were initially cultured in primed ESC (pESC) media consisting of KnockOut
435 DMEM (Gibco#10829–018), fetal bovine serum (Gibco#A5209501), KnockOut serum replacement

436 (Invitrogen#10828–028), Glutamax (Gibco#35050-061), Anti-Anti (Gibco#15240-062), MEM Non-essential
437 amino acids (Gibco#11140-050), and beta-mercaptoethanol (Sigma#M7522). They were then transitioned
438 into ground-state, “naïve” ESCs (nESCs) by culturing for four passages in N2B27 media containing
439 DMEM/F12 (Gibco#11330–032), Neurobasal media (Gibco#21103–049), Gluamax, Anti-Anti, N2 sup-
440 plement (Invitrogen#17502048), and B27 supplement without vitamin A (Invitrogen#12587-010), and
441 beta-mercaptoethanol. Both pESC and nESC media were supplemented with 3 μ M GSK3 inhibitor
442 CHIR99021 (Sigma #SML1046-5MG), 1 μ M MEK inhibitor PD0325901 (Sigma #PZ0162-5MG), and 1,000
443 units/mL leukemia inhibitory factor (LIF, Millipore#ESG1107).

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

448 **Immunocytochemistry (ICC)**

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

458 **RNA sequencing (RNA-seq)**

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

469 the R package enrichPlot (v1.16.2) using the biological processes setting and compareCluster.

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

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

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

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

486 **Data availability**

487 **Published datasets**

488 All published datasets are available at the Gene Expression Omnibus (GEO) <https://www.ncbi.nlm.nih>
489 .gov/geo. Published RNA-seq datasets analyzed in this study included the male wild-type and *Kdm5c*-KO
490 adult amygdala and hippocampus²⁵ (available at GEO: GSE127722) and male wild-type and *Kdm5c*-KO
491 EpiLCs⁴⁵ (available at GEO: GSE96797).

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

497 **Data analysis**

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

499 XXX

500 **Acknowledgements**

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

506 **References**

- 507 1. Strahl, B.D., and Allis, C.D. (2000). The language of covalent histone modifications. *Nature* **403**,
508 41–45. <https://doi.org/10.1038/47412>.
- 509 2. Jenuwein, T., and Allis, C.D. (2001). Translating the Histone Code. *Science* **293**, 1074–1080.
510 <https://doi.org/10.1126/science.1063127>.
- 511 3. Lewis, E.B. (1978). A gene complex controlling segmentation in *Drosophila*. *Nature* **276**, 565–570.
512 <https://doi.org/10.1038/276565a0>.
- 513 4. Kennison, J.A., and Tamkun, J.W. (1988). Dosage-dependent modifiers of polycomb and antennapedia
514 mutations in *Drosophila*. *Proc. Natl. Acad. Sci. U.S.A.* **85**, 8136–8140. <https://doi.org/10.1073/pnas.8>
5.21.8136.
- 515 5. Kassis, J.A., Kennison, J.A., and Tamkun, J.W. (2017). Polycomb and Trithorax Group Genes in
516 *Drosophila*. *Genetics* **206**, 1699–1725. <https://doi.org/10.1534/genetics.115.185116>.
- 517 6. Gabriele, M., Lopez Tobon, A., D'Agostino, G., and Testa, G. (2018). The chromatin basis of
518 neurodevelopmental disorders: Rethinking dysfunction along the molecular and temporal axes. *Prog
Neuropsychopharmacol Biol Psychiatry* **84**, 306–327. <https://doi.org/10.1016/j.pnpbp.2017.12.013>.
- 519 7. Schaefer, A., Sampath, S.C., Intrator, A., Min, A., Gertler, T.S., Surmeier, D.J., Tarakhovsky, A.,
520 and Greengard, P. (2009). Control of cognition and adaptive behavior by the GLP/G9a epigenetic
suppressor complex. *Neuron* **64**, 678–691. <https://doi.org/10.1016/j.neuron.2009.11.019>.

- 521 8. Scandaglia, M., Lopez-Atalaya, J.P., Medrano-Fernandez, A., Lopez-Cascales, M.T., Del Blanco, B.,
Lipinski, M., Benito, E., Olivares, R., Iwase, S., Shi, Y., et al. (2017). Loss of Kdm5c Causes Spurious
Transcription and Prevents the Fine-Tuning of Activity-Regulated Enhancers in Neurons. *Cell Rep* **21**,
47–59. <https://doi.org/10.1016/j.celrep.2017.09.014>.
- 522 9. Iwase, S., Lan, F., Bayliss, P., De La Torre-Ubieta, L., Huarte, M., Qi, H.H., Whetstine, J.R., Bonni, A.,
Roberts, T.M., and Shi, Y. (2007). The X-Linked Mental Retardation Gene SMCX/JARID1C Defines a
Family of Histone H3 Lysine 4 Demethylases. *Cell* **128**, 1077–1088. <https://doi.org/10.1016/j.cell.2007.02.017>.
- 523 10. Claes, S., Devriendt, K., Van Goethem, G., Roelen, L., Meireleire, J., Raeymaekers, P., Cassiman,
J.J., and Fryns, J.P. (2000). Novel syndromic form of X-linked complicated spastic paraplegia. *Am J
Med Genet* **94**, 1–4.
- 524 11. Jensen, L.R., Amende, M., Gurok, U., Moser, B., Gimmel, V., Tschach, A., Janecke, A.R., Tariverdian,
G., Chelly, J., Fryns, J.P., et al. (2005). Mutations in the JARID1C gene, which is involved in
transcriptional regulation and chromatin remodeling, cause X-linked mental retardation. *Am J Hum
Genet* **76**, 227–236. <https://doi.org/10.1086/427563>.
- 525 12. Carmignac, V., Nambot, S., Lehalle, D., Callier, P., Moortgat, S., Benoit, V., Ghoumid, J., Delobel,
B., Smol, T., Thuillier, C., et al. (2020). Further delineation of the female phenotype with KDM5C
disease causing variants: 19 new individuals and review of the literature. *Clin Genet* **98**, 43–55.
<https://doi.org/10.1111/cge.13755>.
- 526 13. Iwase, S., Brookes, E., Agarwal, S., Badeaux, A.I., Ito, H., Vallianatos, C.N., Tomassy, G.S., Kasza, T.,
Lin, G., Thompson, A., et al. (2016). A Mouse Model of X-linked Intellectual Disability Associated with
Impaired Removal of Histone Methylation. *Cell Reports* **14**, 1000–1009. <https://doi.org/10.1016/j.celrep.2015.12.091>.
- 527 14. Devlin, D.K., Ganley, A.R.D., and Takeuchi, N. (2023). A pan-metazoan view of germline-soma
distinction challenges our understanding of how the metazoan germline evolves. *Current Opinion in
Systems Biology* **36**, 100486. <https://doi.org/10.1016/j.coisb.2023.100486>.
- 528 15. Lehmann, R. (2012). Germline Stem Cells: Origin and Destiny. *Cell Stem Cell* **10**, 729–739.
<https://doi.org/10.1016/j.stem.2012.05.016>.
- 529 16. Endoh, M., Endo, T.A., Shinga, J., Hayashi, K., Farcas, A., Ma, K.W., Ito, S., Sharif, J., Endoh, T.,
Onaga, N., et al. (2017). PCGF6-PRC1 suppresses premature differentiation of mouse embryonic
stem cells by regulating germ cell-related genes. *eLife* **6**. <https://doi.org/10.7554/eLife.21064>.
- 530 17. Mochizuki, K., Sharif, J., Shirane, K., Uranishi, K., Bogutz, A.B., Janssen, S.M., Suzuki, A., Okuda,
A., Koseki, H., and Lorincz, M.C. (2021). Repression of germline genes by PRC1.6 and SETDB1
in the early embryo precedes DNA methylation-mediated silencing. *Nat Commun* **12**, 7020. <https://doi.org/10.1038/s41467-021-27345-x>.

- 540
- 541 18. Velasco, G., Hubé, F., Rollin, J., Neuillet, D., Philippe, C., Bouzinba-Segard, H., Galvani, A., Viegas-Péquignot, E., and Francastel, C. (2010). Dnmt3b recruitment through E2F6 transcriptional repressor mediates germ-line gene silencing in murine somatic tissues. *Proc Natl Acad Sci U S A* **107**, 9281–9286. <https://doi.org/10.1073/pnas.1000473107>.
- 542
- 543 19. Hackett, J.A., Reddington, J.P., Nestor, C.E., Duncan, D.S., Branco, M.R., Reichmann, J., Reik, W., Surani, M.A., Adams, I.R., and Meehan, R.R. (2012). Promoter DNA methylation couples genome-defence mechanisms to epigenetic reprogramming in the mouse germline. *Development* **139**, 3623–3632. <https://doi.org/10.1242/dev.081661>.
- 544
- 545 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>.
- 546
- 547 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>.
- 548
- 549 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>.
- 550
- 551 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>.
- 552
- 553 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>.
- 554
- 555 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>.
- 556
- 557 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>.
- 558
- 559 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>.
- 560

- 561 28. Xiol, J., Cora, E., Koglgruber, R., Chuma, S., Subramanian, S., Hosokawa, M., Reuter, M., Yang, Z.,
Berninger, P., Palencia, A., et al. (2012). A Role for Fkbp6 and the Chaperone Machinery in piRNA
Amplification and Transposon Silencing. *Molecular Cell* 47, 970–979. <https://doi.org/10.1016/j.molcel.2012.07.019>.
- 562
- 563 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>.
- 564
- 565 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>.
- 566
- 567 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>.
- 568
- 569 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>.
- 570
- 571 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/biolreprod.20.5.1031>.
- 572
- 573 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>.
- 574
- 575 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>.
- 576
- 577 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>.
- 578
- 579 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>.
- 580
- 581 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>.
- 582

- 583 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>.
- 584
- 585 40. Samanta, M., and Kalantry, S. (2020). Generating primed pluripotent epiblast stem cells: A methodol-
ogy chapter. *Curr Top Dev Biol* **138**, 139–174. <https://doi.org/10.1016/bs.ctdb.2020.01.005>.
- 586
- 587 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.
EMBO Reports **16**, 791–802. <https://doi.org/10.15252/embr.201540538>.
- 588
- 589 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
activity. *Nature* **487**, 57–63. <https://doi.org/10.1038/nature11244>.
- 590
- 591 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
germline stem cells. *Nat Commun* **7**, 11056. <https://doi.org/10.1038/ncomms11056>.
- 592
- 593 44. Borgel, J., Guibert, S., Li, Y., Chiba, H., Schübler, D., Sasaki, H., Forné, T., and Weber, M. (2010).
Targets and dynamics of promoter DNA methylation during early mouse development. *Nat Genet* **42**,
1093–1100. <https://doi.org/10.1038/ng.708>.
- 594
- 595 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
conserved function of KDM5C demethylase. *Nat Commun* **13**, 2602. <https://doi.org/10.1038/s41467-022-30352-1>.
- 596
- 597 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**,
2474–2479. <https://doi.org/10.1073/pnas.0510813103>.
- 598
- 599 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
- 601 48. Endo, T., Mikedis, M.M., Nicholls, P.K., Page, D.C., and De Rooij, D.G. (2019). Retinoic Acid and Germ
Cell Development in the Ovary and Testis. *Biomolecules* **9**, 775. <https://doi.org/10.3390/biom9120775>.
- 602
- 603 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.
Nat Struct Mol Biol. <https://doi.org/10.1038/s41594-023-01038-z>.
- 604

- 605 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-Specific Gene Expression. *Current Biology* *15*, 1051–1057. <https://doi.org/10.1016/j.cub.2005.04.060>.
- 606
607 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 during embryonic development. *Nat Commun* *12*, 3582. <https://doi.org/10.1038/s41467-021-23596-w>.
- 608
609 52. Wang, P.J., McCarrey, J.R., Yang, F., and Page, D.C. (2001). An abundance of X-linked genes expressed in spermatogonia. *Nat Genet* *27*, 422–426. <https://doi.org/10.1038/86927>.
- 610
611 53. Khil, P.P., Smirnova, N.A., Romanienko, P.J., and Camerini-Otero, R.D. (2004). The mouse X chromosome is enriched for sex-biased genes not subject to selection by meiotic sex chromosome inactivation. *Nat Genet* *36*, 642–646. <https://doi.org/10.1038/ng1368>.
- 612
613 54. Hurlin, P.J. (1999). Mga, a dual-specificity transcription factor that interacts with Max and contains a T-domain DNA-binding motif. *The EMBO Journal* *18*, 7019–7028. <https://doi.org/10.1093/emboj/18.2.4.7019>.
- 614
615 55. Stielow, B., Finkernagel, F., Stiewe, T., Nist, A., and Suske, G. (2018). MGA, L3MBTL2 and E2F6 determine genomic binding of the non-canonical Polycomb repressive complex PRC1.6. *PLoS Genet* *14*, e1007193. <https://doi.org/10.1371/journal.pgen.1007193>.
- 616
617 56. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Molecular Cell* *38*, 576–589. <https://doi.org/10.1016/j.molcel.2010.05.004>.
- 618
619 57. Gajiwala, K.S., Chen, H., Cornille, F., Roques, B.P., Reith, W., Mach, B., and Burley, S.K. (2000). Structure of the winged-helix protein hRFX1 reveals a new mode of DNA binding. *Nature* *403*, 916–921. <https://doi.org/10.1038/35002634>.
- 620
621 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
623 59. Ashique, A.M., Choe, Y., Karlen, M., May, S.R., Phamluong, K., Solloway, M.J., Ericson, J., and Peterson, A.S. (2009). The Rfx4 Transcription Factor Modulates Shh Signaling by Regional Control of Ciliogenesis. *Sci. Signal.* *2*. <https://doi.org/10.1126/scisignal.2000602>.
- 624
625 60. Kistler, W.S., Baas, D., Lemeille, S., Paschaki, M., Seguin-Estevez, Q., Barras, E., Ma, W., Duteyrat, J.-L., Morlé, L., Durand, B., et al. (2015). RFX2 Is a Major Transcriptional Regulator of Spermiogenesis. *PLoS Genet* *11*, e1005368. <https://doi.org/10.1371/journal.pgen.1005368>.
- 626

- 627 61. Wu, Y., Hu, X., Li, Z., Wang, M., Li, S., Wang, X., Lin, X., Liao, S., Zhang, Z., Feng, X., et al.
628 (2016). Transcription Factor RFX2 Is a Key Regulator of Mouse Spermiogenesis. *Sci Rep* 6, 20435.
<https://doi.org/10.1038/srep20435>.
- 629 62. Liu, M., Zhu, Y., Xing, F., Liu, S., Xia, Y., Jiang, Q., and Qin, J. (2020). The polycomb group protein
630 PCGF6 mediates germline gene silencing by recruiting histone-modifying proteins to target gene
631 promoters. *J Biol Chem* 295, 9712–9724. <https://doi.org/10.1074/jbc.RA119.012121>.
- 631 63. Otani, J., Nankumo, T., Arita, K., Inamoto, S., Ariyoshi, M., and Shirakawa, M. (2009). Structural basis
632 for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX–DNMT3–DNMT3L
633 domain. *EMBO Reports* 10, 1235–1241. <https://doi.org/10.1038/embor.2009.218>.
- 633 64. Guo, X., Wang, L., Li, J., Ding, Z., Xiao, J., Yin, X., He, S., Shi, P., Dong, L., Li, G., et al. (2015).
634 Structural insight into autoinhibition and histone H3-induced activation of DNMT3A. *Nature* 517,
640–644. <https://doi.org/10.1038/nature13899>.
- 635 65. Meissner, A., Mikkelsen, T.S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein,
636 B.E., Nusbaum, C., Jaffe, D.B., et al. (2008). Genome-scale DNA methylation maps of pluripotent and
637 differentiated cells. *Nature* 454, 766–770. <https://doi.org/10.1038/nature07107>.
- 637 66. Nassar, L.R., Barber, G.P., Benet-Pagès, A., Casper, J., Clawson, H., Diekhans, M., Fischer, C.,
638 Gonzalez, J.N., Hinrichs, A.S., Lee, B.T., et al. (2023). The UCSC Genome Browser database: 2023
639 update. *Nucleic Acids Research* 51, D1188–D1195. <https://doi.org/10.1093/nar/gkac1072>.
- 639 67. Akalin, A., Kormaksson, M., Li, S., Garrett-Bakelman, F.E., Figueiroa, M.E., Melnick, A., and Mason,
640 C.E. (2012). methylKit: A comprehensive R package for the analysis of genome-wide DNA methylation
641 profiles. *Genome Biol* 13, R87. <https://doi.org/10.1186/gb-2012-13-10-r87>.
- 641 68. Yang, M., Yu, H., Yu, X., Liang, S., Hu, Y., Luo, Y., Izsvák, Z., Sun, C., and Wang, J. (2022). Chemical-
642 induced chromatin remodeling reprograms mouse ESCs to totipotent-like stem cells. *Cell Stem Cell*
643 29, 400–418.e13. <https://doi.org/10.1016/j.stem.2022.01.010>.
- 643 69. Al Adhami, H., Vallet, J., Schaal, C., Schumacher, P., Bardet, A.F., Dumas, M., Chicher, J., Hammann,
644 P., Daujat, S., and Weber, M. (2023). Systematic identification of factors involved in the silencing
645 of germline genes in mouse embryonic stem cells. *Nucleic Acids Research* 51, 3130–3149. <https://doi.org/10.1093/nar/gkad071>.
- 645 70. Zhang, J., Zhang, M., Acampora, D., Vojtek, M., Yuan, D., Simeone, A., and Chambers, I. (2018).
646 OTX2 restricts entry to the mouse germline. *Nature* 562, 595–599. <https://doi.org/10.1038/s41586-018-0581-5>.
- 647 71. Tahiliani, M., Mei, P., Fang, R., Leonor, T., Rutenberg, M., Shimizu, F., Li, J., Rao, A., and Shi, Y.
648 (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>.

- 649 72. Endo, T., Romer, K.A., Anderson, E.L., Baltus, A.E., De Rooij, D.G., and Page, D.C. (2015). Periodic retinoic acid–STRA8 signaling intersects with periodic germ-cell competencies to regulate
650 spermatogenesis. *Proc. Natl. Acad. Sci. U.S.A.* *112*. <https://doi.org/10.1073/pnas.1505683112>.
- 651 73. Bowles, J., Knight, D., Smith, C., Wilhelm, D., Richman, J., Mamiya, S., Yashiro, K., Chawengsak-
652 sophak, K., Wilson, M.J., Rossant, J., et al. (2006). Retinoid Signaling Determines Germ Cell Fate in
Mice. *Science* *312*, 596–600. <https://doi.org/10.1126/science.1125691>.
- 653 74. Karimi, M.M., Goyal, P., Maksakova, I.A., Bilenky, M., Leung, D., Tang, J.X., Shinkai, Y., Mager, D.L.,
654 Jones, S., Hirst, M., et al. (2011). DNA Methylation and SETDB1/H3K9me3 Regulate Predominantly
Distinct Sets of Genes, Retroelements, and Chimeric Transcripts in mESCs. *Cell Stem Cell* *8*,
676–687. <https://doi.org/10.1016/j.stem.2011.04.004>.
- 655 75. Auclair, G., Borgel, J., Sanz, L.A., Vallet, J., Guibert, S., Dumas, M., Cavelier, P., Girardot, M., Forné,
656 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>.
- 657 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.
658 <https://doi.org/10.1371/journal.pone.0205969>.
- 659 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>.
- 660 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.
662 <https://doi.org/10.1038/s41588-020-00736-4>.
- 663 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>.
- 665 80. Nielsen, A.Y., and Gjerstorff, M.F. (2016). Ectopic Expression of Testis Germ Cell Proteins in Cancer
666 and Its Potential Role in Genomic Instability. *Int J Mol Sci* *17*. <https://doi.org/10.3390/ijms17060890>.
- 667 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
668 Sciences* *20*. <https://doi.org/10.22038/ijbms.2017.9259>.
- 669 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.
<https://doi.org/10.1126/science.1195481>.

- 670
- 671 83. Bonefas, K.M., and Iwase, S. (2021). Soma-to-germline transformation in chromatin-linked neurode-
672 velopmental disorders? *FEBS J.* <https://doi.org/10.1111/febs.16196>.
- 673 84. Velasco, G., Walton, E.L., Sterlin, D., Hédouin, S., Nitta, H., Ito, Y., Fouyssac, F., Mégarbané, A.,
674 Sasaki, H., Picard, C., et al. (2014). Germline genes hypomethylation and expression define a
molecular signature in peripheral blood of ICF patients: Implications for diagnosis and etiology.
Orphanet J Rare Dis 9, 56. <https://doi.org/10.1186/1750-1172-9-56>.
- 675 85. Walton, E.L., Francastel, C., and Velasco, G. (2014). Dnmt3b Prefers Germ Line Genes and Cen-
676 tromeric Regions: Lessons from the ICF Syndrome and Cancer and Implications for Diseases. *Biology*
(Basel) 3, 578–605. <https://doi.org/10.3390/biology3030578>.
- 677 86. Samaco, R.C., Mandel-Brehm, C., McGraw, C.M., Shaw, C.A., McGill, B.E., and Zoghbi, H.Y. (2012).
Crh and Oprm1 mediate anxiety-related behavior and social approach in a mouse model of MECP2
678 duplication syndrome. *Nat Genet* 44, 206–211. <https://doi.org/10.1038/ng.1066>.
- 679 87. Stephens, M. (2016). False discovery rates: A new deal. *Biostat*, kxw041. <https://doi.org/10.1093/biostatistics/kxw041>.
- 680 88. Conway, J.R., Lex, A., and Gehlenborg, N. (2017). UpSetR: An R package for the visualization of
681 intersecting sets and their properties. *Bioinformatics* 33, 2938–2940. <https://doi.org/10.1093/bioinformatics/btx364>.
- 682 89. Li, H., Liang, Z., Yang, J., Wang, D., Wang, H., Zhu, M., Geng, B., and Xu, E.Y. (2019). DAZL
is a master translational regulator of murine spermatogenesis. *Natl Sci Rev* 6, 455–468. <https://doi.org/10.1093/nsr/nwy163>.
- 683 90. Mikedis, M.M., Fan, Y., Nicholls, P.K., Endo, T., Jackson, E.K., Cobb, S.A., De Rooij, D.G., and Page,
684 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>.

687 **Figures and Tables**

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

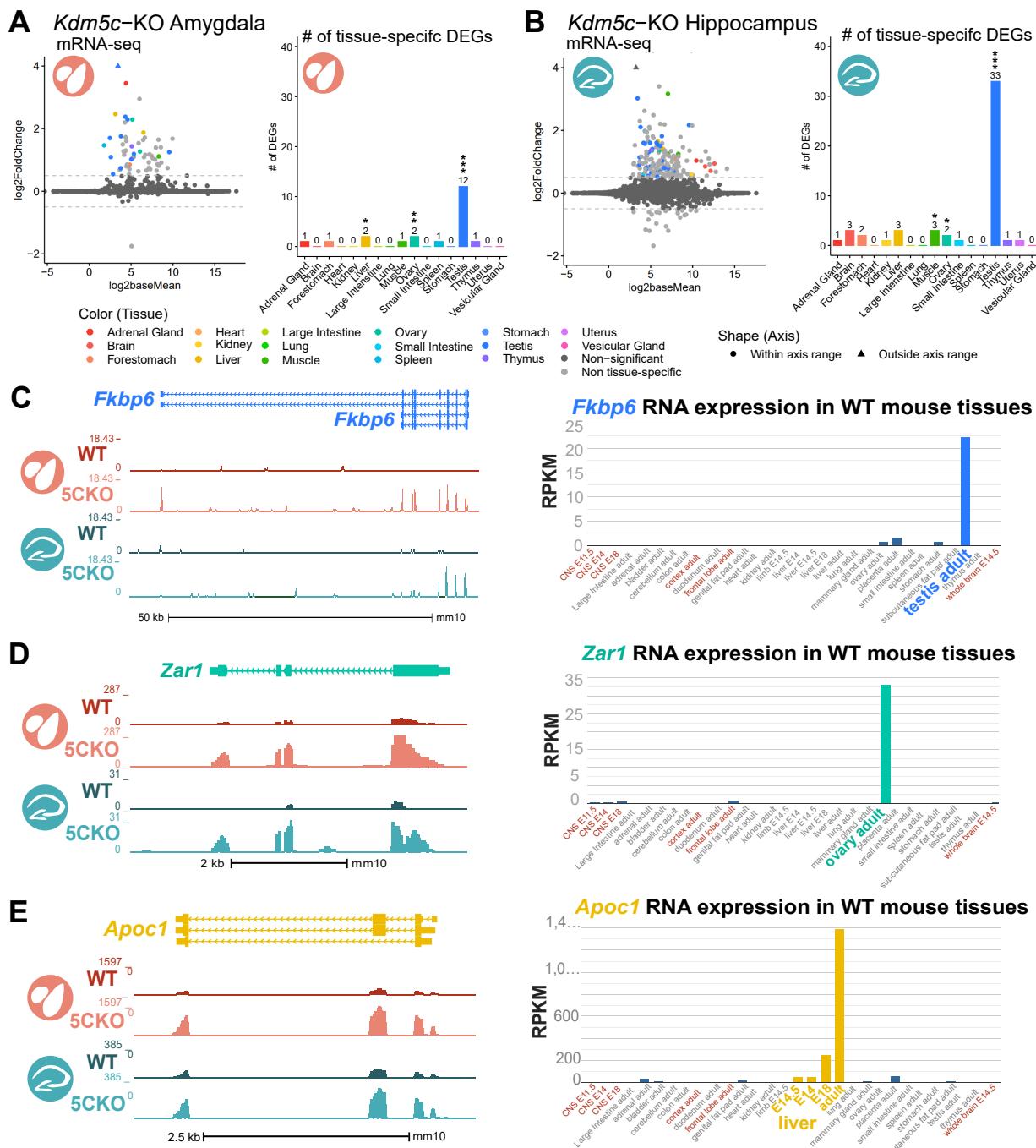


Figure 1: Tissue-enriched genes are misexpressed in the *Kdm5c*-KO brain. **A-B.** Expression of tissue-enriched genes (Li et al 2017) in the male *Kdm5c*-KO amygdala (A) and hippocampus (B). Left - MA plot of mRNA-sequencing. Right - Number of tissue-enriched differentially expressed genes (DEGs). *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

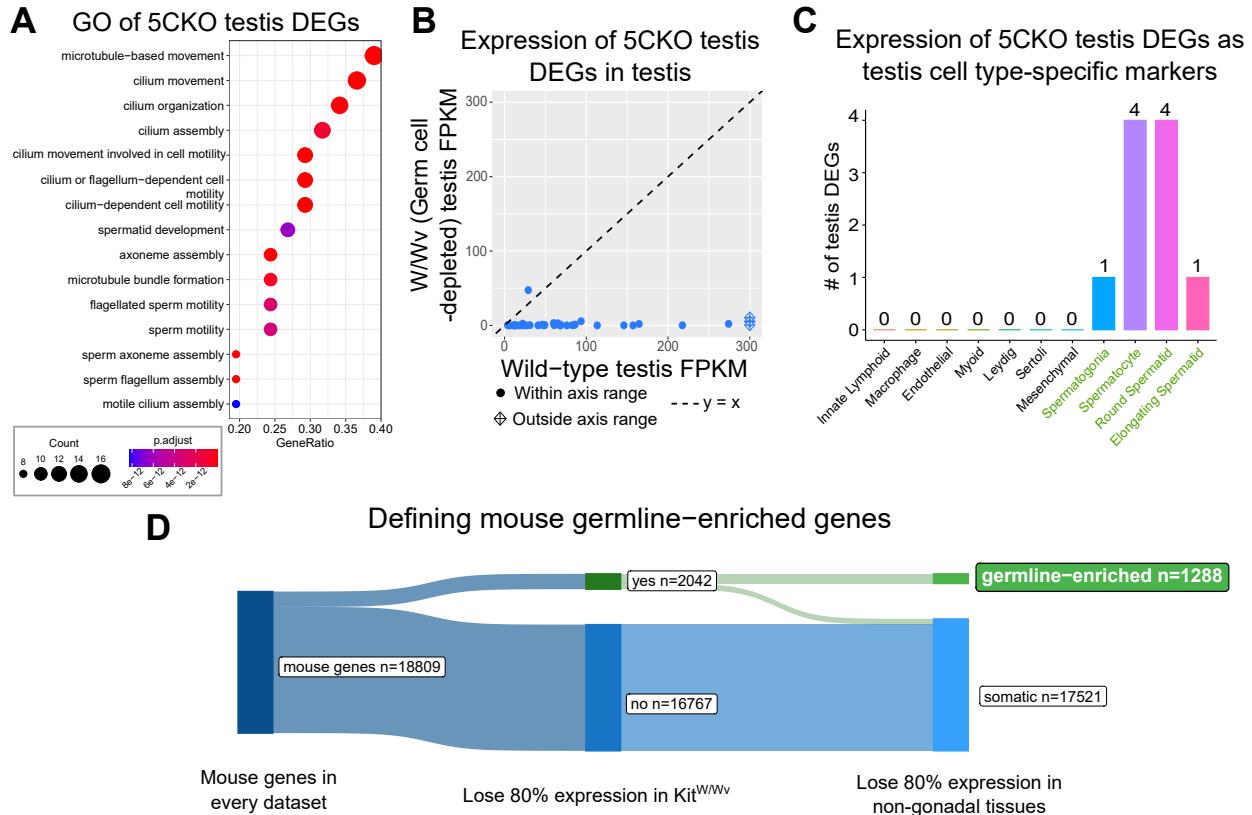


Figure 2: Aberrant transcription of germline genes in the *Kdm5c*-KO in the brain. **A.** enrichPlot gene ontology (GO) of *Kdm5c*-KO amygdala and hippocampus testis-enriched DEGs **B.** Expression of testis DEGs in wild-type (WT) testis versus germ cell-depleted (W/W_v) testis (Mueller et al 2013).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

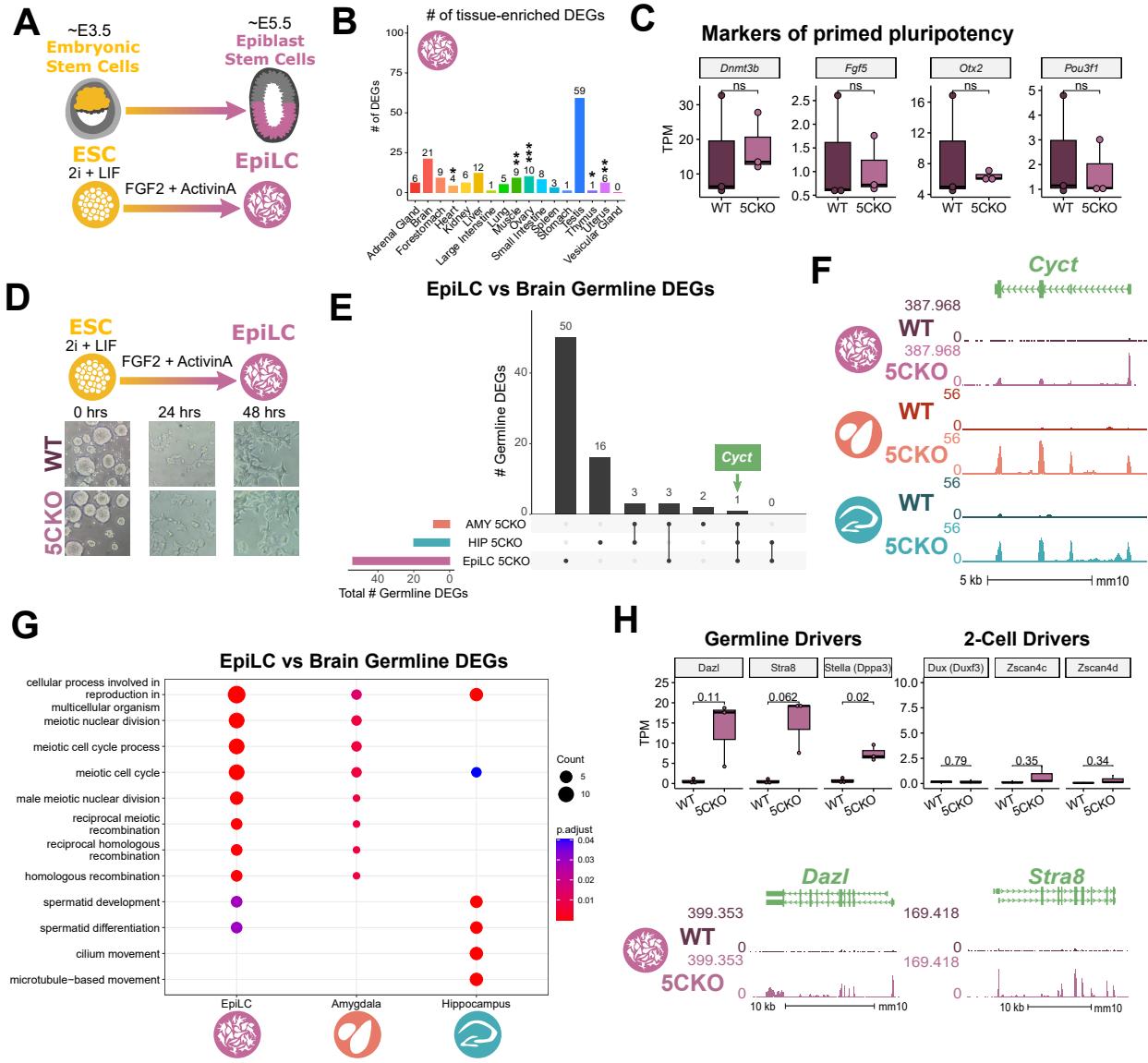


Figure 3: *Kdm5c*-KO epiblast-like cells express key drivers of germline identity **A.** Top - Diagram of *in vivo* differentiation of embryonic stem cells (ESCs) of the inner cell mass into epiblast stem cells. Bottom - *in vitro* differentiation of ESCs into epiblast-like cells (EpiLCs). **B.** Number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

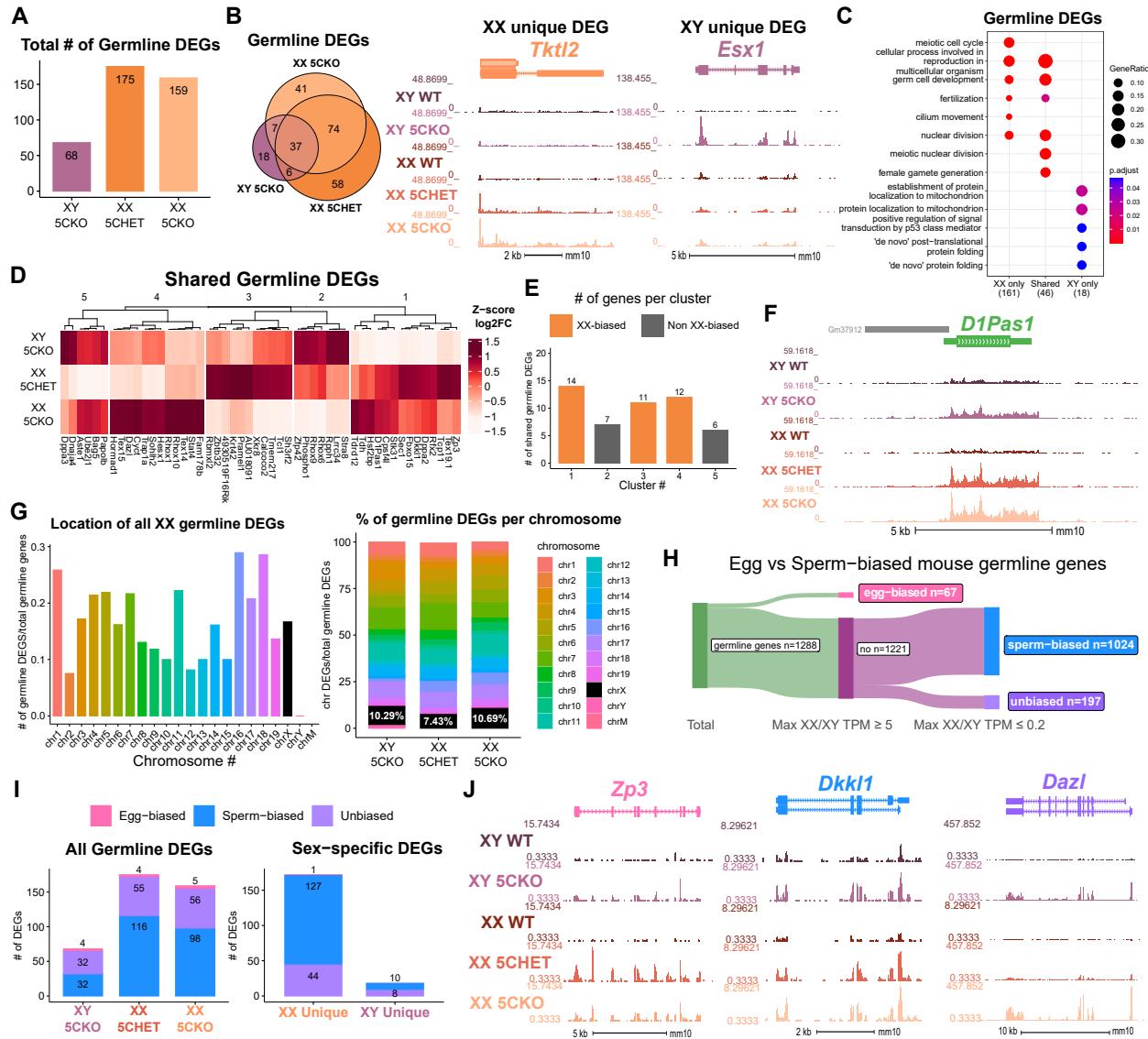


Figure 4: Chromosomal sex influences *Kdm5c*-KO germline gene misexpression. **A-B.** Expression of tissue-enriched genes (Li et al 2017) in the male *Kdm5c*-KO amygdala (A) and hippocampus (B). Left - MA plot of mRNA-seq. Right - Number of tissue-enriched differentially expressed genes (DEGs). * p<0.05,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

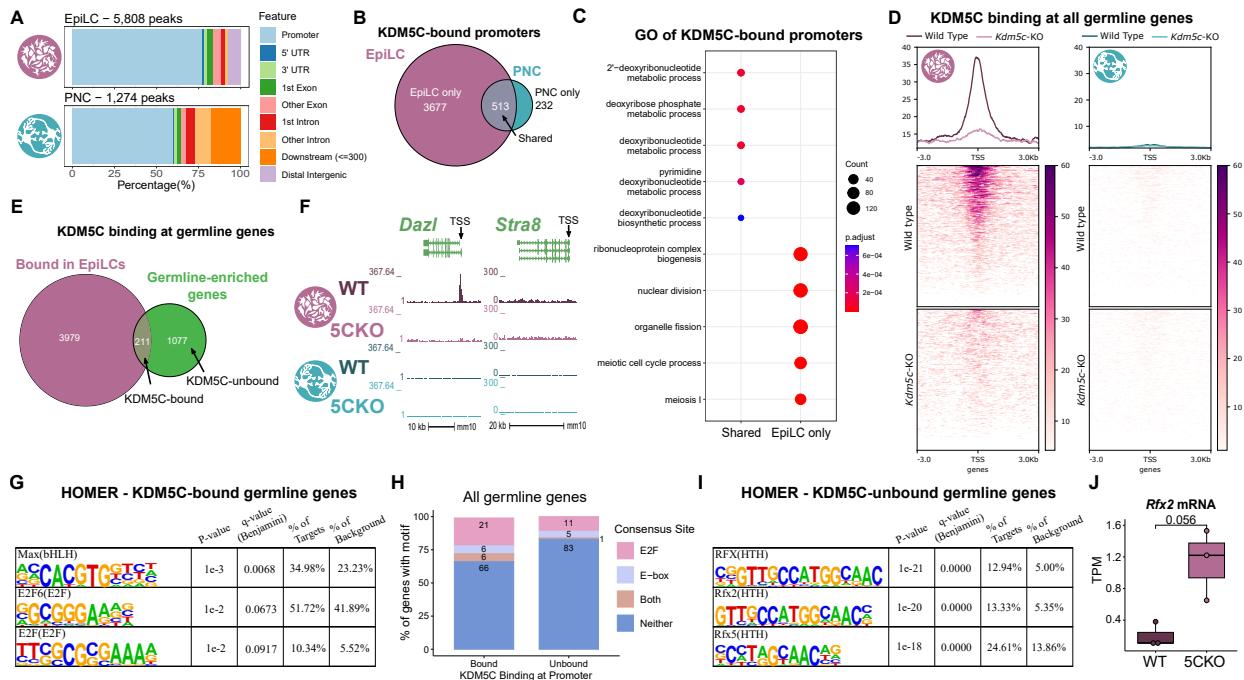


Figure 5: KDM5C binds to a subset of germline gene promoters during early embryogenesis **A.** ChIPseeker localization of KDM5C peaks at different genomic regions in EpiLCs (top) and hippocampal and cortex primary neuron cultures (PNCs, bottom). **B.** Overlap of genes with KDM5C bound to their promoters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

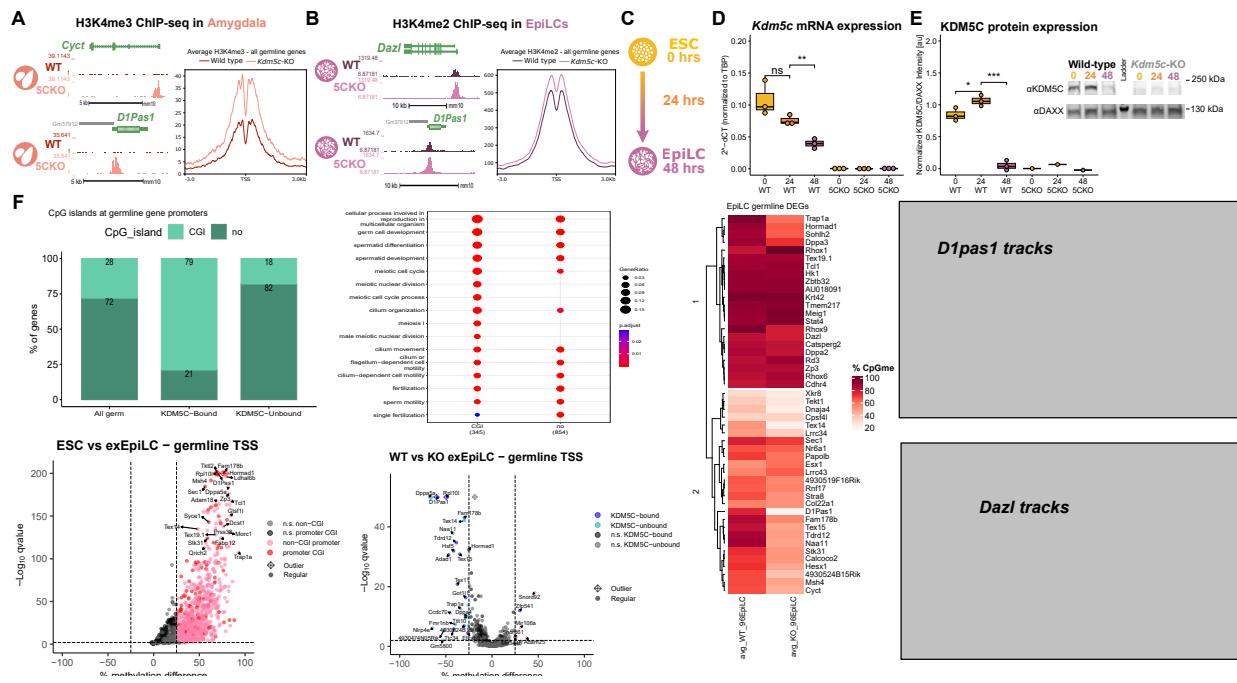


Figure 6: KDM5C's catalytic activity promotes long-term silencing of germline genes via DNA methylation. **A.** Left - Bigwigs of representative histone 3 lysine 4 trimethylation (H3K4me3) ChIP-seq peaks at two germline genes in the wild-type (WT) and *Kdm5c*-KO (5CKO) adult amygdala. Right - Average 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.

692 Notes

693 Things to do

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

698 Dazl

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

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

712 Discussion notes

- 713 • For other paper:
 - 714 – for methods: Heatmaps of gene expression were generated using the base R functions scale and
715 hclust and visualized using the R package ComplexHeatmap (v2.12.1).
 - 716 – * 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>
- 718 • Talk about different motifs if we do see differences there and if it explains direct vs indirect dysregulation
 - 719 – Could also mention PRC1.6/KDM5C binds pcgf6 in ESCs
- 720 • Maybe talk about other regulators/if anything is known about them long term? Or save for 2nd paper.

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

- 749 • Altogether, these data indicate that while some germline genes are misexpressed due to direct loss of
750 KDM5C binding during emryogenesis, secondary downstream mechanisms can also promote their
751 aberrant transcription.
- 752 • Papers to read/reference:
753 – Reconstitution of the Mouse Germ Cell Specification Pathway in Culture by Pluripotent Stem Cells:
754 [https://www.cell.com/fulltext/S0092-8674\(11\)00771-9](https://www.cell.com/fulltext/S0092-8674(11)00771-9)
755 – 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/>

757 **Figure outline:**

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

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

770 **Figure 3: *Kdm5c*-KO epiblast-like cells express key drivers regulators of germline identity** * A) ESC
771 to EpiLC differentiation Left - Morphology is unchanged, * B) 5CKO EpiLCs express EpiLC differentiation
772 genes similar to WT lvs * C) Male EpiLCs express germline genes (example Cyct again) * Overlap between
773 brain and EpiLC germline genes - show they're mostly unique * GO of Brain and EpiLC germline genes
774 (meiotic enriched) * Bigwigs or TPM of master regulators * Show that while some are also 2-cell genes
775 (Dazl), 2-cell specific genes aren't dysregulated (Zscan4). Important point because published KDM5C dazl
776 paper is saying KDM5C is a 2-cell regulator, but as far as I can tell only genes shared between germline and
777 2-cell are dysregulated.

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

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

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

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

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

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

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

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

822 **Germline gene repression background:**

823 Interestingly, some of the ectopic testis transcripts identified in the *Kdm5c*-KO brain are typically ex-
824 pressed in germ cells⁸. Unlike somatic cells, germ cells (e.g. sperm and eggs) undergo meiosis and pass on
825 their genetic material to the next generation. The germline and the soma are typically distinguished during
826 early embryogenesis, when germline genes are silenced in epiblast stem cells soon after implantation and
827 only reactivated in a subset to form the germline. Chromatin regulators play a key role in decommissioning
828 germline genes as the embryo transitions from naïve to primed pluripotency by placing repressive histone
829 H2A lysine 119 monoubiquitination (H2AK119ub1)¹⁶, histone 3 lysine 9 trimethylation (H3K9me3)^{16,17}, and
830 DNA CpG methylation^{17–19} at germline gene promoters. KDM5C may also be involved in this early decom-
831 missioning of germline genes, as re-expression of KDM5C in neurons fails to suppress their dysregulation⁸.
832 In support of this, KDM5C was very recently shown to repress *Deleted in azoospermia like (Dazl)*, a key
833 regulator of germline development, in mouse embryonic stem cells (ESCs)^{49,69}. In support of this, two
834 independent screens in mouse embryonic stem cells (ESCs) recently identified KDM5C as a repressor of
835 *Deleted in azoospermia like (Dazl)*, a key regulator of germline development. However, KDM5C's role in
836 embryonic germline gene repression is currently unclear, given that Dazl is also expressed in ESCs and in
837 the 2-cell stage and germline gene misexpression has yet to be globally characterized during *Kdm5c*-KO
838 embryogenesis.