

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 some chromatin regulators are important for neuron-specific gene expression, loss of repressive chromatin regulators can also lead to the aberrant transcription of tissue-specific genes outside of their intended context. Because very few studies have characterized the these aberrant, tissue-specific transcripts, the mechanisms behind their dysregulation and their functional consequences are largely unknown. Here, we explored this cellular identity crisis in NDDs by investigating the ectopic expression of germline (e.g. sperm and egg) genes in mice lacking lysine demethylase 5c (KDM5C, also known as JARID1C or SMCX), an eraser of histone 3 lysine 4 di and tri-methylation (H3K4me2/3). We found male *Kdm5c* knockout (-KO) mice, which recapitulate key behavioral phenotypes of Claes-Jensen X-linked intellectual disability (MRXSCJ, OMIM: 300534), aberrantly expresses many liver, muscle, ovary, and testis genes within the amygdala and hippocampus. Many genes expressed in the *Kdm5c*-KO brain were typically unique to germ cells, indicating an erosion of the soma-germline boundary. Using germ cell-depleted RNA sequencing datasets, we 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 is sexually dimorphic, as female *Kdm5c*-KO EpiLCs were more prone to germline gene dysregulation than knockout males. We found KDM5C acts at a subset of germline gene promoters in EpiLCs to promote the initial placement of DNA methylation at CpG islands. However, germline genes can also become activated in *Kdm5c*-KO cells independent of direct KDM5C regulation. Intriguingly, many of the germline transcripts expressed in the *Kdm5c*-KO brain were late-stage spermatogenesis genes that are not directly bound by KDM5C. This suggests germline developmental programs can progress ectopically in the background of typical *Kdm5c*-KO development. Ultimately, this work characterizes a novel suppressor of germline gene transcription and links impaired soma-vs-germline demarcation to a chromatin-linked neurodevelopmental disorder.

30 Introduction

31 A single genome holds the instructions to generate the myriad of cell types found within the adult
32 organism. This is, in part, accomplished by chromatin regulators that can either promote or impede lineage-
33 specific gene expression through DNA and histone modifications^{1,2}. Many chromatin regulators were initially
34 identified for their roles in shaping cellular and tissue identity^{3–5}. Unexpectedly, next generation sequencing
35 revealed mutations in many chromatin regulators directly cause or are associated with neurodevelopmental
36 disorders (NDDs)⁶. Several studies have suggested this connection between chromatin regulators and
37 neurodevelopment is due to their regulation of brain-specific genes, such as orchestrating transcriptional
38 programs for synaptic maturation⁷ and transitioning between neuronal and glial fates during neural precursor
39 differentiation⁸. However, loss of some chromatin regulators can also result in the ectopic transcription
40 of tissue-specific genes outside of their target environment, such as the misexpression of liver-specific
41 genes within adult neurons⁹. Very few studies have investigated this cellular identity crisis in chromatin-
42 linked NDDs^{9,10} and it is currently unknown if ectopic gene expression contributes to neurodevelopmental
43 impairments.

44 To elucidate the role of tissue identity in chromatin-linked neurodevelopmental disorders, it is essential
45 to first characterize the nature of the dysregulated genes and the molecular mechanisms governing their
46 de-repression. In this study, we characterized the aberrant expression of tissue-enriched genes with
47 loss of lysine demethylase 5C (KDM5C). KDM5C, also known as SMCX or JARID1C, is a chromatin
48 regulator that can repress gene expression through erasure of histone 3 lysine 4 di- and trimethylation
49 (H3K4me2/3)¹¹, marks enriched at active gene promoters. Pathogenic mutations in *KDM5C* cause Intellectual
50 Developmental Disorder, X-linked, Syndromic, Claes-Jensen Type (MRXSCJ, OMIM: 300534). MRXSCJ
51 is more common and often more severe in males and its neurological phenotypes include intellectual
52 disability, seizures, aberrant aggression, and autistic behaviors^{12–14}. Male *Kdm5c* knockout (-KO) mice
53 recapitulate key MRXSCJ phenotypes, including hyperaggression, increased seizure propensity, and learning
54 impairments^{10,15}. Unexpectedly, RNA sequencing (RNA-seq) of the *Kdm5c*-KO hippocampus revealed
55 ectopic expression of testis genes within the brain¹⁰. It is currently unknown what types of testis genes are
56 dysregulated, when in *Kdm5c*-KO development this begins, if this phenomenon is sexually dimorphic, and if
57 other tissue-specific genes are also aberrantly expressed with KDM5C loss.

58 The testis contains both germ cells (cells that pass on their genetic material, e.g. sperm) and somatic
59 cells (cells that perform all other bodily functions). Distinguishing between the germline and the soma is a
60 key feature of multicellularity¹⁶ that occurs during early embryogenesis in many metazoans¹⁷. In mammals,
61 chromatin regulators are crucial for decommissioning germline genes in somatic cells during the transition
62 from naïve to primed pluripotency. Initially, germline gene promoters gain repressive histone H2A lysine 119
63 monoubiquitination (H2AK119ub1)¹⁸ and histone 3 lysine 9 trimethylation (H3K9me3)^{18,19} in embryonic stem
64 cells and are then decorated with DNA CpG methylation in the post-implantation embryo^{19–21}. Currently,

most studies have focused on marker genes important for early germ cell development rather than germline genes as a whole given the lack of a curated list of germline-enriched genes. Thus, it is unclear if the mechanism of repression differs for certain classes of germline genes, e.g. meiotic genes versus spermatid differentiation genes. Furthermore, it is currently unknown if ectopic germline gene expression is influenced by chromosomal sex, as previous studies on germline gene repression have been exclusively in males. Sex is particularly pertinent in the case of sexually dimorphic chromatin regulators like KDM5C, which lies on the X chromosome and partially escapes X chromosome inactivation, resulting in 1.2-2-fold higher in females²²⁻²⁵. Thus, systematically characterizing KDM5C's role in germline gene repression during early embryogenesis, including its interaction with known silencing mechanisms, will unveil key mechanisms underlying the demarcation between soma and germline identity. These results will provide molecular footholds that can then be exploited to test the contribution of ectopic germline genes on neurodevelopment.

To illuminate KDM5C's role in tissue identity, here we characterized the aberrant transcription of tissue-enriched genes within the *Kdm5c*-KO brain and epiblast-like stem cells (EpiLCs), an *in vitro* model of the post-implantation embryo. We observed general dysregulation of tissue-enriched genes in both the *Kdm5c*-KO brain and EpiLCs, including testis, liver, muscle, and ovary-enriched genes. The *Kdm5c*-KO amygdala and hippocampus had significant enrichment of testis genes that are specific to germ cells and not somatic cells of the testis. We found *Kdm5c*-KO EpiLCs aberrantly expressed key drivers of germline identity and meiosis, including *Dazl* and *Stra8*, while the *Kdm5c*-KO brain primarily expressed germline genes important for late spermatogenesis. Female *Kdm5c*-KO EpiLCs expressed over double the number of germline genes than *Kdm5c*-KO males and germline genes dysregulated in both sexes were expressed more highly in females, demonstrating females have increased sensitivity to germline gene dysregulation. KDM5C was bound to only a subset of germline genes expressed in *Kdm5c*-KO EpiLCs, indicating germline-enriched mRNAs can be aberrantly transcribed through indirect mechanisms. Finally, we found KDM5C promotes the long-term silencing of germline genes in somatic cells by aiding the placement of DNA CpG methylation in EpiLCs. Thus, we propose KDM5C plays a fundamental role in the development of tissue identity during early embryogenesis, including the establishment of the soma-germline boundary.

Results

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

Previous RNA sequencing (RNA-seq) in the adult hippocampus revealed ectopic expression of some testis genes within the male *Kdm5c* knockout (-KO) brain¹⁰. It is currently unclear if the testis is the only tissue type misexpressed in the *Kdm5c*-KO brain. We thus characterized dysregulation of *Kdm5c*-KO brain tissue identity by systematically assessing the expression of genes enriched in 17 mouse tissues²⁶, using our published mRNA-seq datasets of the adult amygdala and hippocampus for male mice with constitutive

98 knockout of *Kdm5c*²⁷.

99 We found a large proportion of genes that are significantly upregulated within the *Kdm5c*-KO brain
100 (DESeq2²⁸, log2 fold change > 0.5, q < 0.1) are enriched in non-brain tissues in wild-type mice (Amygdala:
101 35%, Hippocampus: 24%) (Figure 1A-B). The majority of tissue-enriched differentially expressed genes
102 (tissue-enriched DEGs) were testis genes (Figure 1A-C). Even though the testis has the largest total number
103 of tissue-biased genes compared to any other tissue (2,496 genes), testis-biased DEGs were significantly
104 enriched for both brain regions (Amygdala p = 1.83e-05, Odds Ratio = 5.13; Hippocampus p = 4.26e-11,
105 Odds Ratio = 4.45, Fisher's Exact Test). One example of a testis-enriched gene misexpressed in the
106 *Kdm5c*-KO brain is *FK506 binding protein 6* (*Fkbp6*), a known regulator of piRNA expression and meiosis in
107 germ cells^{29,30} (Figure 1C).

108 Interestingly, we also observed significant enrichment of ovary-biased DEGs in both the amygdala and
109 hippocampus (Amygdala p = 0.00574, Odds Ratio = 18.7; Hippocampus p = 0.048, Odds Ratio = 5.88,
110 Fisher's Exact) (Figure 1D). Ovary-enriched DEGs included *Zygotic arrest 1* (*Zar1*), which sequesters mRNAs
111 in oocytes for meiotic maturation and early zygote development³¹ (Figure 1D). Given the *Kdm5c*-KO mice
112 we analyzed are male, this demonstrates ectopic expression of tissue-enriched genes is independent of
113 organismal sex.

114 Although not consistent across brain regions, we also found significant enrichment of DEGs biased
115 towards two non-gonadal tissues - the liver (Amygdala p = 0.0398, Odds Ratio = 6.58, Fisher's Exact Test)
116 and the muscle (Hippocampus p = 0.0104, Odds Ratio = 6.95, Fisher's Exact Test). A liver-biased DEG
117 dysregulated in both the hippocampus and amygdala is *Apolipoprotein C-I* (*Apoc1*), a lipoprotein metabolism
118 and transport gene³² that has been implicated in Alzheimer's disease³³ (Figure 1E). For all tissue-enriched
119 DEGs, aberrantly expressed mRNAs are polyadenylated and spliced into mature transcripts (Figure 1C-E).
120 Of note, we observed little to no dysregulation of brain-enriched genes (Amygdala p = 1; Hippocampus p =
121 0.74, Fisher's Exact), despite the fact these are brain samples and the brain has the second highest total
122 number of tissue-enriched genes (708 genes). Together, these results suggest the aberrant expression of
123 tissue-enriched genes within the brain is a major effect of KDM5C loss.

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

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

131 To further validate if these testis DEGs are truly germline genes, we evaluated their expression in
132 somatic versus germ cells within the testis. We first compared their expression in the testis without germ

133 cells³⁴, which was accomplished by heterozygous *W* and *Wv* mutations in the enzymatic domain of *c-Kit*
134 (*Kit*^{W/Wv})³⁵. Almost all *Kdm5c*-KO testis-enriched DEGs lost expression with germ cell depletion (Figure
135 2B). We then assessed testis-enriched DEG expression in a published single cell RNA-seq dataset that
136 identified cell type-specific markers within the testis³⁶. We found some testis-enriched DEGs were classified
137 as specific markers for different germ cell types (e.g. spermatogonia, spermatocytes, round spermatids, and
138 elongating spermatids), yet none marked somatic cells (Figure 2C). Together, these data demonstrate that
139 the *Kdm5c*-KO brain aberrantly expresses germline genes, but not somatic testicular genes, reflecting an
140 erosion between somatic versus germline identity.

141 We then wanted to characterize germline gene misexpression with *Kdm5c* loss more deeply, but lacked a
142 comprehensive list of mouse germline-enriched genes. We thus generated a list of germline-enriched genes
143 using RNA-seq datasets of *Kit*^{W/Wv} mice that include males and females at embryonic day 12, 14, and 16³⁷
144 and adult male testes³⁴. We defined genes as germline-enriched if their expression met the following criteria:
145 1) their expression is greater than 1 FPKM in wild-type gonads 2) their expression in any non-gonadal tissue
146 for adult wild type²⁶ does not exceed 20% of their maximum expression in the wild-type germline, and 3)
147 their expression in the germ cell-depleted gonads, for any sex or time point, does not exceed 20% of their
148 maximum expression in the wild-type germline. These criteria yielded 1,288 germline-enriched genes (Figure
149 2D), which was hereafter used as a resource to globally assess germline gene misexpression with *Kdm5c*
150 loss (Supplementary table 1).

151 ***Kdm5c*-KO epiblast-like cells aberrantly express master regulators of germline
152 identity**

153 Germ cells are typically distinguished from somatic cells soon after the embryo implants into the uterine
154 wall^{38,39}. A subset of epiblast stem cells become the primordial germ cells (PGCs) while the remainder
155 differentiate into the ectoderm, mesoderm, and endoderm to form the somatic tissues⁴⁰. This developmental
156 time point can be modeled *in vitro* through differentiation of embryonic stem cells (ESCs) into epiblast-like
157 stem cells (EpiLCs) (Figure 3A, top)^{41,42}. While some germline-enriched genes are also expressed in
158 embryonic stem cells (ESCs) and in the 2-cell stage^{43–45}, they are silenced as they differentiate into EpiLCs¹⁹.
159 Therefore, we tested if KDM5C was necessary for silencing germline genes at this developmental stage by
160 evaluating the impact of *Kdm5c* loss in EpiLCs.

161 We first identified *Kdm5c*-KO EpiLC DEGs in our previously published RNA-seq dataset⁴⁶ (DESeq2, log2
162 fold change > 0.5, q < 0.1) and assessed tissue-enriched gene expression. Similar to the *Kdm5c*-KO brain,
163 we observed general dysregulation of tissue-enriched genes, with the largest number of genes belonging to
164 the brain and testis, although they were not significantly enriched (Figure 3B). Primed pluripotency genes,
165 such as *Dnmt3b*, *Fgf5*, *Pou3f1*, and *Otx2*, were unaltered in *Kdm5c*-KO EpiLCs (Figure 3C). Furthermore,
166 *Kdm5c*-KO cell morphology during ESC to EpiLC differentiation (Figure 3D) was normal, indicating KDM5C

167 loss does not impair EpiLC formation.

168 To determine if germline DEGs are constitutively dysregulated or if they can change over the course
169 of development, we compared EpiLC germline DEGs to those expressed in the *Kdm5c*-KO brain. We
170 found the majority of germline DEGs were unique to either EpiLCs or the brain, with only *Cyct* shared
171 across all tissue/cell types (Figure 3E-F). We found particularly high enrichment of meiosis-related gene
172 ontologies in EpiLCs (Figure 3G), such as meiotic cell cycle process (GO: 1903046, p.adjust = 1.59e-08)
173 and meiotic nuclear division (GO: 0140013, p.adjust = 9.76e-09). While there was modest enrichment of
174 meiotic gene ontologies in both brain regions, the *Kdm5c*-KO hippocampus primarily expressed late-stage
175 spermatogenesis genes, such those involved in the sperm axoneme structure.

176 Notably, DEGs unique to *Kdm5c*-KO EpiLCs included key drivers of germline identity, such as *Stimulated*
177 *by retinoic acid 8* (*Stra8*) and *Deleted in azoospermia like* (*Dazl*) (Figure 3H). These genes are typically
178 expressed during embryonic germ cell development to commit PGCs to the germline fate, but are also
179 expressed later in life to trigger meiotic gene expression programs⁴⁷⁻⁴⁹. Of note, some germline genes,
180 including *Dazl*, are also expressed in the two-cell embryo^{44,50}. However, we did not see misexpression of
181 two-cell embryo-specific genes, like *Duxf3* (*Dux*) ($q = 0.337$) and *Zscan4d* ($q = 0.381$), indicating *Kdm5c*-KO
182 in EpiLCs do not revert back to a 2-cell state (Figure 3H).

183 **Females have increased sensitivity to germline gene misexpression with *Kdm5c* 184 loss**

185 It is currently unknown if the misexpression of germline genes is influenced by chromosomal sex, as
186 previous studies on germline gene repressors have been conducted exclusively in males. We explored the
187 impact of sex upon germline gene misexpression by comparing their dysregulation in male *Kdm5c* hemizygous
188 knockout (XY *Kdm5c*-KO), female homozygous knockout (XX *Kdm5c*-KO) and female heterozygous (XX
189 *Kdm5c*-HET) EpiLCs.⁴⁶ We first identified differentially expressed genes (DEGs) compared to sex-matched
190 wild-type controls (DESeq2²⁸, log2 fold change > 0.5, $q < 0.1$) and then filtered for germline-enriched genes.

191 Homozygous and heterozygous females expressed over double the number of germline-enriched genes
192 than hemizygous males (Figure 4A). While the majority of germline DEGs in *Kdm5c*-KO males were also
193 dysregulated in females (74%), there were also many male-specific and female-specific germline DEGs
194 (Figure 4A-B). We compared the known functions of germline genes dysregulated in all samples (shared),
195 only in females (XX only - XX *Kdm5c*-KO, XX *Kdm5c*-HET, or both), and only in males (XY only). Gene
196 ontologies uniquely enriched in female-specific germline DEGs included meiotic (meiotic cell cycle) and
197 flagellar (cilium movement) genes, while mitochondrial and cell signaling gene ontologies were enriched in
198 male-specific DEGs (protein localization to mitochondrion).

199 Germline genes dysregulated in both sexes were also enriched for meiotic ontologies (meiotic nuclear
200 division), as well as egg-specific genes (female gamete generation). The majority of these shared germline

201 DEGs had a greater log₂ fold change from wild-type in females compared to males (Figure 4D-F). The
202 increased number of germline genes and degree of dysregulation in females could be caused by improper
203 X chromosome inactivation (XCI), as the X chromosome is enriched for many testis-specific germline
204 genes[XXX]. However, both shared and female-specific germline DEGs were not biased towards the X
205 chromosome, with the majority of genes lying on autosomes instead (Figure 4G). Thus, while female EpiLCs
206 have increased sensitivity to germline gene misexpression with KDM5C loss, it is likely independent of
207 potential XCI defects.

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

209 While many germline genes have roles in both the male and female germline, some display sex-biased
210 expression or have functions unique to eggs and sperm. To comprehensively assess if germline gene
211 sex corresponds with *Kdm5c* mutant sex, we filtered our list of germline-enriched genes for egg (XX) and
212 sperm (XY) biased genes. We defined sex-biased genes as those whose expression in the opposite sex,
213 at any time point, is no greater than 20% of the gene's maximum expression in a given sex. This yielded
214 67 egg-biased, 1,024 sperm-biased, and 197 unbiased germline-enriched genes, which is consistent with
215 the testis overall having a more unique transcriptome than the ovary²⁶ (Figure 4H). We found egg, sperm,
216 and unbiased germline genes were dysregulated in all *Kdm5c* mutants (Figure 4I-L). *Kdm5c* mutant male
217 and female-specific germline DEGs were also not biased to the corresponding germ cell sex, indicating
218 differences between male and female germline gene dysregulation not due to sex-specific activation of
219 sperm or egg-specific transcriptional programs. These results demonstrate sex influences the degree of
220 gene misexpression with loss of KDM5C, but not the sex of germ cell-enriched genes.

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

222 Previous work suggests KDM5C represses germline genes during early development, as re-expression
223 of KDM5C in knockout neuronal cultures fails to suppress two hippocampal germline DEGs¹⁰. KDM5C binds
224 to several germline gene promoters in embryonic stem cells (ESCs) but is lost in neurons¹⁰. However, the
225 lack of a comprehensive list of germline-enriched genes prohibited systematic characterization of KDM5C
226 binding at germline gene promoters. Thus, it is currently unclear if KDM5C is enriched at germline gene
227 promoters, what types of germline genes KDM5C regulates, and if its binding is maintained at any germline
228 genes in neurons.

229 To address these questions, we analyzed KDM5C chromatin immunoprecipitation followed by DNA
230 sequencing (ChIP-seq) datasets in EpiLCs⁴⁶ and primary forebrain neuron cultures (PNCs)¹⁵. EpiLCs had a
231 higher total number of KDM5C peaks than PNCs (EpiLCs: 5,808, PNCs: 1,276, MACS2 q < 0.1 and fold
232 enrichment > 1, removal of *Kdm5c*-KO false positive peaks) and KDM5C was primarily localized to gene
233 promoters in both cell types (EpiLCs: 4,190, PNCs: 745 +/- 500bp from the TSS), although PNCs showed

234 increased localization to non-promoter regions (Figure 4A).

235 The majority of promoters bound by KDM5C in PNCs were also bound in EpiLCs (513 shared promoters),
236 however a large portion of gene promoters were only bound by KDM5C in EpiLCs (3677 EpiLC only
237 promoters) (Figure 4B). We then performed gene ontology analysis to compare the known functions of genes
238 with KDM5C bound to their promoter in EpiLCs and PNCs. While there were no ontologies significantly
239 enriched for genes only bound by KDM5C in PNCs, gene ontologies for genes shared between PNCs and
240 EpiLCs were enriched for functions involving nucleic acid turnover, such as deoxyribonucleotide metabolic
241 process (GO:0009262, p.adjust = 8.28e-05) (Figure 4C). Germline-specific ontologies were only enriched in
242 promoters unique to EpiLCs, such as meiotic nuclear division (GO: 0007127 p.adjust = 6.77e-16) and meiotic
243 cell cycle process (GO:1903046, p.adjust = 5.05e-16) (Figure 3C). We then evaluated KDM5C binding
244 around the transcription start site (TSS) of all germline-enriched genes. In EpiLCs, We observed modest
245 KDM5C signal at about half of all germline genes (Figure 4D). Based on our ChIP-seq peak criteria, KDM5C
246 was strongly bound to about 10% of germline gene promoters in EpiLCs (Figure 4E). In condordance with our
247 gene ontology results, we did not observe KDM5C accumulation at any germline gene promoters in PNCs
248 (Figure 4D). Together, these results demonstrate KDM5C is significantly enriched at a subset of germline
249 gene promoters in EpiLCs, including meiotic genes, but does not directly regulate germline genes in neurons.

250 Many germline-specific genes, including meiotic genes, are suppressed by the transcription factor
251 heterodimers E2F6/DP1 and MGA/MAX, which respectively bind E2F and E-box motifs^{20,51-54}. Thus, we
252 identified transcription factor motifs enriched at KDM5C-bound or unbound germline gene promoters using
253 HOMER⁵⁵ (TSS +/- 500 bp, q-value < 0.1). MAX and E2F6 binding sites were significantly enriched at
254 germline genes bound by KDM5C in EpiLCs, but not at germline genes unbound by KDM5C (MAX q-value: ,
255 E2F6 q-value: , E2F q-value:) (Figure 4). One third of KDM5C-bound promoters contained the consensus
256 sequence for either E2F6 (E2F, 5'-TCCCGC-3'), MGA (E-box, 5'-CACGTG-3'), or both, but only 17% of
257 KDM5C-unbound genes contained these motifs (Figure 4). KDM5C-unbound germline genes were instead
258 enriched for multiple RFX transcription factor binding sites. RFX transcription factors bind X-box motifs⁵⁶
259 to promote ciliogenesis^{57,58}. Enriched RFX transcription factors included RFX2, a central regulator of post-
260 meiotic spermatogenesis^{59,60}. Interestingly, RFX2 mRNA is upregulated in *Kdm5c*-KO EpiLCs, but is also not
261 a direct target of KDM5C (Figure 4). Thus, distinct transcription factor programs regulate KDM5C-bound and
262 unbound germline genes.

263 Finally, we compared KDM5C binding at promoters of RNA-seq DEGs to determine if the germline
264 mRNAs de-repressed in *Kdm5c*-KO EpiLCs and brain are direct targets of KDM5C (Figure 4F). In EpiLCs,
265 KDM5C was bound to about one third of EpiLC-specific and brain-specific germline DEGs (EpiLC only: 36%,
266 Brain only: 33.3%). Again, we did not observe any KDM5C binding at germline gene promoters in PNCs,
267 even for brain-specific DEGs (Figure 4H). Some notable differences in KDM5C binding for EpiLC-specific
268 DEGs included *Dazl* and *Stra8*, two key drivers of germline identity discussed above. Although both mRNAs
269 are expressed in *Kdm5c*-KO EpiLCs, KDM5C is only bound to *Dazl*'s promoter and not *Stra8*'s in EpiLCs

270 (Figure 4F). In contrast to the unique DEGs, 3 out of the 4 genes dysregulated in both the brain and
271 EpiLCs are bound by KDM5C around the TSS (Figure 4G). Altogether, this suggests KDM5C decommissions
272 germline genes in EpiLCs but is no longer recruited to germline gene promoters in neurons. Furthermore, the
273 majority of germline mRNAs expressed in *Kdm5c*-KO cells are dysregulated independent of direct KDM5C
274 recruitment to their promoters.

275 **KDM5C promotes *de novo* DNA methylation at germline genes**

276 • if there are differences, say at germline gene CpG islands

277 In the early embryo, germline gene promoters are initially silenced by repressive histone modifications,
278 including histone 2A lysine 119 monoubiquitination (H2AK119ub1) and histone 3 lysine 9 trimethylation
279 (H3K9me3), and are then silenced long-term via DNA CpG methylation (CpGme)^{18,19,61}. Our results above
280 suggest KDM5C also acts during this time period, however how KDM5C interacts with other germline gene
281 silencing mechanism is currently unclear. KDM5C is generally thought to suppress transcription through
282 erasure of histone 3 lysine 4 di- and trimethylation (H3K4me2/3)¹¹. However, KDM5C's catalytic activity was
283 recently shown to not be required for suppressing *Dazl* in undifferentiated ESCs⁵⁰. Since H3K4me3 impedes
284 *de novo* CpGme placement^{62,63}, KDM5C's catalytic activity may instead be required for CpGme-mediated,
285 long-term silencing of germline genes. In support of this, CpGme is markedly reduced at two germline gene
286 promoters in the *Kdm5c*-KO adult hippocampus¹⁰. Based on these observations, we hypothesized KDM5C
287 erases H3K4me3 to promote the initial placement of CpGme at germline gene promoters in EpiLCs.

288 To test this hypothesis, we first characterized KDM5C's substrates, histone 3 lysine 4 di- and trimethylation
289 (H3K4me2/3) in our previously published ChIP-seq datasets of the wild type and *Kdm5c*-KO amygdala²⁷
290 and EpiLCs⁴⁶. In congruence with previous work in the *Kdm5c*-KO hippocampus¹⁰, we observed aberrant
291 accumulation of H3K4me3 around the transcription start site (TSS) of germline genes in the *Kdm5c*-KO
292 amygdala (Figure 5A). We additionally found a marked increase in H3K4me2 around the TSS of germline
293 genes in *Kdm5c*-KO EpiLCs (Figure 5B).

294 We then evaluated KDM5C's role in germline gene silencing during ESC to EpiLC differentiation. We first
295 characterized KDM5C's mRNA and protein expression at 0, 24, and 48 hours of differentiation (Figure 5C).
296 While *Kdm5c* mRNA steadily decreased from 0 to 48 hours of differentiation (Figure 5D), KDM5C protein
297 initially increased from 0 to 24 hours but then decreased to near knockout levels by 48 hours (Figure 5E).

298 • Germline genes are known to accumulate CpGme at (CGIs).

299 – What percentage of germline genes have CGIs

300 • Do differential methylation analysis for WT ESCs to WT EpiLCs

301 – What percentage of germline genes significantly gain methylation (at CGI or at promoter)

302 – Out of the ones that gain methylation, which are significantly reduced

303 To determine KDM5C's role in the initial placement of DNA methylation at germline gene promoters,
304 we performed whole genome bisulfite sequencing (WGBS) in wild-type and *Kdm5c*-KO ESCs and 96 hour
305 extended EpiLCs (exEpiLCs). Although wild-type cells accumulated high levels of DNA methylation at
306 germline gene promoters over the course of ESCs to exEpiLC differentiation, DNA methylation was markedly
307 reduced in *Kdm5c*-KO exEpiLCs (Figure 5F).

- 308 • discussion
309 • – However, CPG islands greatly determine KDM5C recruitment - KDM5C is known to be enriched at
310 CGIs.)
311 • Altogether, this suggests KDM5C is dynamically regulated during ESC to EpiLC differentiation and
312 promotes germline gene silencing via removal of H3K4me2/3 during early embryogenesis.

313 Discussion

314 In the above study, we demonstrate KDM5C's pivotal role in the development of tissue identity, resulting in
315 substantial misexpression of tissue-enriched genes within the adult *Kdm5c*-KO hippocampus and amygdala.
316 In addition to testis genes identified previously¹⁰, we found significant enrichment of muscle, liver, and even
317 ovary-biased genes aberrantly expressed within the male *Kdm5c*-KO brain. Deeper characterization of
318 testis-enriched DEGs revealed they were not somatic testis genes, but germline genes, thus demonstrating
319 KDM5C's crucial role in establishing the soma-germline boundary. Given the majority of tissue-enriched
320 DEGs are testis and ovary genes with no known brain functions, it is currently unknown if they impair
321 *Kdm5c*-KO brain function and contribute to MRXSCJ-related phenotypes like intellectual disability and
322 aberrant aggression. However, select liver and muscle-biased DEGs do have known roles within the brain,
323 such as the liver-enriched lipid metabolism gene *Apolipoprotein C-I* (*Apoc1*)³² that is highly expressed in the
324 *Kdm5c*-KO amygdala and hippocampus. Intriguingly, *Apoc1* overexpression in the mouse brain can impair
325 learning and memory⁶⁴ and is implicated in Alzheimer's disease in humans³³.

326 We found a substantial portion of genes upregulated in *Kdm5c*-KO brain are germline genes with no
327 known functions within the brain. Distinguishing the germline from the soma is a key feature of multicellularity
328 and sexual reproduction¹⁶. Previous work characterizing chromatin regulators that silence germ cell-specific
329 transcription has predominantly focused on their repression of key marker genes in embryonic stem cells
330 (ESCs), such as *Dazl* and *Ddx4*^{18,19,54}. To characterize KDM5C's role in germline gene repression at a
331 genome-wide level throughout life, we curated a list of mouse germline-enriched genes using publicly
332 available germ cell-depleted RNA-seq datasets from *Kit^{W/W^v}* mice^{34,37}. This resource enabled us to identify 1)
333 the extent of germline gene dysregulation in *Kdm5c*-KO cells, 2) the types of germline genes misexpressed
334 at different developmental time points, and 3) which groups of germline genes are directly and indirectly

335 regulated by KDM5C. This list of germline-enriched genes serves as a resource for future studies to
336 systematically assess soma-germline dysregulation.

337 RNA-seq of epiblast-like cells (EpiLCs) revealed dysregulation of *Kdm5c*-KO tissue identity begins during
338 early embryogenesis. *In vivo*, germline genes are typically decommissioned in epiblast stem cells and remain
339 silenced as the epiblast differentiates into somatic tissues⁴⁰. However, a small subset of epiblast stem cells
340 will receive signals to reactivate germline gene expression to become the primordial germ cells (PGCs) that
341 will ultimately form the mature germline^{38,39}. This process can be mimicked *in vitro* by differentiating EpiLCs
342 into primordial germ cell-like cells (PGCLCs)⁴¹. Therefore, misexpression of germline genes in EpiLCs might
343 suggest *Kdm5c*-KO ESCs are progressing beyond EpiLC differentiation and instead becoming PGCLCs. Yet,
344 *Kdm5c*-KO EpiLCs properly express primed pluripotency genes, including *Otx2* - an epiblast stem cell marker
345 that is known to repress differentiation into PGCs/PGCLCs⁶⁵. Furthermore, we observed no difference in
346 cellular morphology during *Kdm5c*-KO ESC to EpiLC differentiation. Proper EpiLC differentiation, together
347 with *Kdm5c*-KO mice being viable, suggests germline gene expression is occurring ectopically in conjunction
348 with typical developmental programs, rather than a complete shift to towards germline identity.

349 • XX vs XY

350 • While many germline genes have roles in both the male and female germline, some display sex-biased
351 expression or have functions unique to eggs and sperm. Many of the germline genes characterized in
352 the male *Kdm5c*-KO brain have sperm-specific functions. It is currently unclear if sperm genes are also
353 dysregulated in female *Kdm5c* mutants, or if they would instead misexpress egg-biased genes.

354 • Motif analysis

355 – only 1/3 had max/e2f - Unlikely to be the major recruitment mechanism for KDM5C
356 – KDM5C does not contain motif-specific binding
357 – controls cilia in multiple cell types, initially identified in neurons. Shared axis that could contribute
358 to NDDs

359 * RFX Cilia in NDDs - <https://pubmed.ncbi.nlm.nih.gov/33658631/>
360 – Enriched RFX transcription factors included the spermiogenesis regulator RFX2, whose mRNA is
361 also significantly upregulated in *Kdm5c*-KO EpiLCs (Figure 4).
362 – <https://www.nature.com/articles/srep20435>
363 – <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4498915/>
364 – <https://pubmed.ncbi.nlm.nih.gov/32725242/>

365 We then globally characterized KDM5C binding at germline-enriched gene promoters through KDM5C
366 ChIP-seq in EpiLCs and primary hippocampus and cortex neuron cultures (PNCs).

367 Our analysis of KDM5C ChIP-seq data provided novel insights into direct and indirect mechanisms by
368 which germline genes can be misexpressed in *Kdm5c*-KO cells.

369 While we observed high enrichment of KDM5C peaks at germline genes in EpiLCs, KDM5C was not
370 bound to germline gene promoters in PNCs. This suggests dysregulation of germline genes in the mature
371 *Kdm5c*-KO brain is due to loss of repression during embryogenesis, which is consistent with previous
372 work that found introducing human KDM5C into *Kdm5c*-KO PNCs does not repress two aberrant germline
373 transcripts¹⁰. Although enriched at germline gene promoters in EpiLCs, KDM5C was only bound to about a
374 third of the germline genes aberrantly transcribed in *Kdm5c*-KO EpiLCs. One notable EpiLC DEG not bound
375 by KDM5C is *Stra8*, a transcription factor activated by retinoic acid signaling in germ cells to promote meiotic
376 initiation^{66,67}. However, retinoic acid can only activate *Stra8* expression when DAZL is present. Unlike *Stra8*,
377 *Dazl* is a direct target of KDM5C in EpiCs and is both transcribed and translated in *Kdm5c*-KO EpiLCs.
378 Expression of indirect DEGs like *Stra8* indicates germline genes can be aberrantly expressed in *Kdm5c*-KO
379 cells through downstream activation by other ectopic germline programs. These ectopic germline programs
380 appear to loosely mimic the trajectory of typical germline development, as germline genes important for early
381 germ cell development and meiosis are expressed in *Kdm5c*-KO EpiLC while late spermatogenesis genes
382 are expressed in the mature *Kdm5c*-KO brain. Altogether, this indicates once a subset of germline genes
383 are activated due to loss of direct KDM5C repression in the early embryo, these ectopic germline programs
384 can continue to progress in the background of *Kdm5c*-KO somatic development.

385 It is important to note that while we focused on KDM5C's regulation of germline genes, some germline-
386 enriched genes are also expressed in naïve ESCs and at the 2-cell stage due to their roles in pluripotency and
387 self-renewal. For example, *Dazl* is primarily known for committing PGCs to the germline fate and regulating
388 the translation of germline-specific RNAs, but is also expressed at the 2-cell stage⁶⁸, in naïve ESCs⁴³, and in
389 the inner cell mass⁴³. Very recently, two screens of *Dazl*-repressors identified KDM5C as a direct regulator in
390 ESCs^{50,69}. Interestingly, one study found *Kdm5c*-KO ESCs also aberrantly expressed 2-cell-specific genes,
391 indicating KDM5C negatively regulates totipotency⁵⁰. However, out of the four regulators characterized,
392 KDM5C was the only factor whose repression of *Dazl* was independent of the 2-cell specific transcription
393 factor *Dux*⁵⁰. Although expressed at the 2-cell stage and in ESCs, *Dazl* is typically silenced when ESCs
394 differentiate into EpiLCs¹⁹. We found KDM5C also represses *Dazl* in EpiLCs, but *Kdm5c*-KO EpiLCs did
395 not express 2-cell specific genes like *Dux* and *Zscan4c*. Together, this suggests the 2-cell-like state in
396 *Kdm5c*-KO ESCs is an indirect consequence of KDM5C's primary role in germline gene repression.

397 *Dazl* and other germline-enriched genes are initially decommissioned in ESCs by repressive histone
398 modifications and then *de novo* DNA CpG methylation (CpGme) in EpiLCs^{19–21,61,70,71}. Unlike the previously
399 characterized germline gene suppressors, KDM5C removes an active histone mark, histone 3 lysine 4 di-
400 and trimethylation (H3K4me2/3)¹¹. In this study, we observed a global increase in H3K4me2/3 in *Kdm5c*-KO
401 EpiLCs and amygdala around the transcription start site (TSS) of germline-enriched genes. However,
402 KDM5C's catalytic activity may not be required for germline gene silencing, as it was recently found to be
403 dispensible for repressing *Dazl* in ESCs⁵⁰. Although not necessary in ESCs, KDM5C's catalytic activity be
404 required for CpGme in EpiLCs, as H3K4me3 impedes *de novo* CpGme placement^{62,63}. This is supported

405 by previous work in the *Kdm5C*-KO adult hippocampus, which found CpGme is significantly eroded at at
406 least two germline promoters¹⁰. To elucidate the mechanism behind KDM5C-mediated silencing of germline
407 genes, we characterized KDM5C in ESCs and EpiLCs and found its expression is dynamically regulated
408 during differentiation. Furthermore, KDM5C acts during the transition from naïve to primed pluripotency to
409 promote the initial placement of CpGme at CpG islands for the long-term silencing of germline genes.

410 The above work provides the mechanistic foundation for KDM5C's role in germline gene repression
411 and offers novel insight into how dysregulation in *Kdm5c*-KO tissue identity changes over the course of
412 development. Although their impact upon neurodevelopment is currently unclear, ectopic germline transcripts
413 are also found in models of other related neurodevelopmental disorders⁷², including Immunodeficiency,
414 Centromere instability, Facial anomalies syndrome 1 (ICF1 OMIM: #242860)^{73,74}, Kleefstra syndrome
415 1 (OMIM: #610253)⁹, and MeCP2 duplication syndrome (MDS, OMIM: #300260)⁷⁵. Like KDM5C, the
416 chromatin regulators underlying these conditions, DNA methyltransferase 3b (DNMT3B), H3K9me1/2
417 methyltransferases G9A/GLP, methyl-CpG -binding protein 2 (MECP2), primarily silence gene expression.
418 Thus, KDM5C is among a growing cohort of chromatin-linked neurodevelopmental disorders that may have a
419 similar underlying cause of germline versus soma dysregulation. However, further research is required to
420 determine the ultimate impact of these germline genes and the extent to which this phenomenon occurs in
421 humans.

- 422 • Last paragraph
- 423 – Thus, failure to fine-tune the expression of tissue-enriched, dosage-sensitive genes like *Apoc1*
424 could be one route by which loss of brain tissue identity contributes to *Kdm5c*-KO impairments.

425 Materials and Methods

426 Classifying tissue-enriched and germline-enriched genes

427 Tissue-enriched differentially expresssd genes (DEGs) were determined by their classification in a previ-
428 ously published dataset from 17 male and female mouse tissues²⁶. This study defined tissue expression as
429 greater than 1 Fragments Per Kilobase of transcript per Million mapped read (FPKM) and tissue enrichment
430 as at least 4-fold higher expression than any other tissue.

431 We curated a list of germline-enriched genes using an RNA-seq dataset from wild-type and germline-
432 depleted (*Kit^{W/Wv}*) male and female mouse embryos from embryonic day 12, 14, and 16³⁷, as well as adult
433 male testes³⁴. Germline-enriched genes met the following criteria: 1) their expression is greater than 1
434 FPKM in wild-type germline 2) their expression in any wild-type somatic tissues²⁶ does not exceed 20%
435 of maximum expression in wild-type germline, and 3) their expression in the germ cell-depleted (*Kit^{W/Wv}*)
436 germline, for any sex or time point, does not exceed 20% of maximum expression in wild-type germline.

437 **Cell culture**

438 We utilized our previously established cultures of male wild-type and *Kdm5c* knockout (-KO) embryonic
439 stem cells⁴⁶. Sex was confirmed by genotyping *Uba1/Uba1y* on the X and Y chromosomes with the following
440 primers: 5'-TGGATGGTGTGCCAATG-3', 5'-CACCTGCACGTTGCCCT-3'. Deletion of *Kdm5c* was
441 confirmed through the primers 5'-ATGCCCATATTAAGAGTCCTG-3', 5'-TCTGCCTTGATGGGACTGTT-3',
442 and 5'-GGTTCTAACACTCACATAGTG-3'.

443 Embryonic stem cells (ESCs) and epiblast-like cells were cultured using previously established
444 methods⁴². Briefly, ESCs were initially cultured in primed ESC (pESC) media consisting of KnockOut
445 DMEM (Gibco#10829-018), fetal bovine serum (Gibco#A5209501), KnockOut serum replacement
446 (Invitrogen#10828-028), Glutamax (Gibco#35050-061), Anti-Anti (Gibco#15240-062), MEM Non-essential
447 amino acids (Gibco#11140-050), and beta-mercaptoethanol (Sigma#M7522). They were then transitioned
448 into ground-state, "naïve" ESCs (nESCs) by culturing for four passages in N2B27 media containing
449 DMEM/F12 (Gibco#11330-032), Neurobasal media (Gibco#21103-049), Gluamax, Anti-Anti, N2 sup-
450 plement (Invitrogen#17502048), and B27 supplement without vitamin A (Invitrogen#12587-010), and
451 beta-mercaptoethanol. Both pESC and nESC media were supplemented with 3 μM GSK3 inhibitor
452 CHIR99021 (Sigma #SML1046-5MG), 1 μM MEK inhibitor PD0325901 (Sigma #PZ0162-5MG), and 1,000
453 units/mL leukemia inhibitory factor (LIF, Millipore#ESG1107).

454 nESCs were differentiated into epiblast-like cells (EpiLCs, 48 hours) and extendend EpiLCs (exEpiLCs,
455 96 hours) by culturing in N2B27 media containing DMEM/F12, Neurobasal media, Gluamax, Anti-Anti,
456 N2 supplement, B27 supplement (Invitrogen#17504044), beta-mercaptoethanol, fibroblast growth factor 2
457 (FGF2, R&D Biotechne 233-FB), and activin A (R&D Biotechne 338AC050CF), as previously described⁴².

458 **Immunocytochemistry (ICC)**

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

468 **RNA sequencing (RNA-seq)**

469 After ensuring read quality via FastQC (v0.11.8), reads were then mapped to the mm10 *Mus musculus*
470 genome (Gencode) using STAR (v2.5.3a), during which we removed duplicates and kept only uniquely
471 mapped reads. Count files were generated by FeatureCounts (Subread v1.5.0), and BAM files were
472 converted to bigwigs using deeptools (v3.1.3) and visualized by the UCSC genome browser. RStudio (v3.6.0)
473 was then used to analyze counts files by DESeq2 (v1.26.0)²⁸ to identify differentially expressed genes
474 (DEGs) with a q-value (p-adjusted via FDR/Benjamini–Hochberg correction) less than 0.1 and a log2 fold
475 change greater than 0.5. For all DESeq2 analyses, log2 fold changes were calculated with IfcShrink using
476 the ashR package⁷⁶. MA-plots were generated by ggpibr (v0.6.0), and Eulerr diagrams were generated by
477 eulerr (v6.1.1). Boxplots and scatterplots were generated by ggpibr (v0.6.0) and ggplot2 (v3.3.2). The Upset
478 plot was generated via the package UpSetR (v1.4.0)⁷⁷. Gene ontology (GO) analyses were performed by
479 the R package enrichPlot (v1.16.2) using the biological processes setting and compareCluster.

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

481 ChIP-seq reads were aligned to mm10 using Bowtie1 (v1.1.2) allowing up to two mismatches. Only
482 uniquely mapped reads were used for analysis. Peaks were called using MACS2 software (v2.2.9.1) using
483 input BAM files for normalization, with filters for a q-value < 0.1 and a fold enrichment > 1. We removed
484 “black-listed” genomic regions that often give aberrant signals. Common peak sets were obtained in R via
485 DiffBind (v3.6.5). In the case of KDM5C ChIP-seq, *Kdm5c*-KO peaks were then subtracted from wild-type
486 samples using bedtools (v2.25.0). Peak proximity to genome annotations was determined by ChIPSeeker
487 (v1.32.1). Gene ontology (GO) analyses were performed by the R package enrichPlot (v1.16.2) using the
488 biological processes setting and compareCluster. Enriched motifs were identified using HOMER⁵⁵. Average
489 binding across the genome was visualized using deeptools (v3.1.3). Bigwigs were visualized using the
490 UCSC genome browser.

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

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

496 **Data availability**

497 **Published datasets**

498 All published datasets are available at the Gene Expression Omnibus (GEO) <https://www.ncbi.nlm.nih.gov/geo>. Published RNA-seq datasets analyzed in this study included the male wild-type and *Kdm5c*-KO
499 adult amygdala and hippocampus²⁷ (available at GEO: GSE127722) and male wild-type and *Kdm5c*-KO
500 EpiLCs⁴⁶ (available at GEO: GSE96797).

502 Previously published ChIP-seq experiments included KDM5C in wild-type and *Kdm5c*-KO EpiLCs⁴⁶ (avail-
503 able at GEO: GSE96797) and mouse primary neuron cultures (PNCs) from the cortex and hippocampus¹⁵
504 (available at GEO: GSE61036). ChIP-seq of histone 3 lysine 4 dimethylation in male wild-type and *Kdm5c*-KO
505 EpiLCs⁴⁶ is also available at GEO: GSE96797. ChIP-seq of histone 3 lysine 4 trimethylation in wild-type and
506 *Kdm5c*-KO male amygdala²⁷ are available at GEO: GSE127817.

507 **Data analysis**

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

510 **Acknowledgements**

- 511 • Jacob Mueller for providing insight in germline gene regulation.
512 • Sundeep Kalantry for providing reagents and expertise in culturing mouse embryonic stem cells and
513 epiblast-like cells
514 • Funding sources

515 **References**

- 516 1. Strahl, B.D., and Allis, C.D. (2000). The language of covalent histone modifications. *Nature* **403**,
517 41–45. <https://doi.org/10.1038/47412>.
- 518 2. Jenuwein, T., and Allis, C.D. (2001). Translating the Histone Code. *Science* **293**, 1074–1080.
519 <https://doi.org/10.1126/science.1063127>.
- 520 3. Lewis, E.B. (1978). A gene complex controlling segmentation in *Drosophila*. *Nature* **276**, 565–570.
521 <https://doi.org/10.1038/276565a0>.
- 522 4. Kennison, J.A., and Tamkun, J.W. (1988). Dosage-dependent modifiers of polycomb and antennapedia
523 mutations in *Drosophila*. *Proc. Natl. Acad. Sci. U.S.A.* **85**, 8136–8140. <https://doi.org/10.1073/pnas.85.21.8136>.

- 524 5. Kassis, J.A., Kennison, J.A., and Tamkun, J.W. (2017). Polycomb and Trithorax Group Genes in
525 *Drosophila*. *Genetics* *206*, 1699–1725. <https://doi.org/10.1534/genetics.115.185116>.
- 526 6. Gabriele, M., Lopez Tobon, A., D'Agostino, G., and Testa, G. (2018). The chromatin basis of
527 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>.
- 528 7. Zhou, Z., Hong, E.J., Cohen, S., Zhao, W.-N., Ho, H.-Y.H., Schmidt, L., Chen, W.G., Lin, Y., Savner,
E., Griffith, E.C., et al. (2006). Brain-specific phosphorylation of MeCP2 regulates activity-dependent
Bdnf transcription, dendritic growth, and spine maturation. *Neuron* *52*, 255–269. <https://doi.org/10.1016/j.neuron.2006.09.037>.
- 530 8. Hirabayashi, Y., Suzuki, N., Tsuboi, M., Endo, T.A., Toyoda, T., Shinga, J., Koseki, H., Vidal, M., and
531 Gotoh, Y. (2009). Polycomb Limits the Neurogenic Competence of Neural Precursor Cells to Promote
Astrogenic Fate Transition. *Neuron* *63*, 600–613. <https://doi.org/10.1016/j.neuron.2009.08.021>.
- 532 9. Schaefer, A., Sampath, S.C., Intrator, A., Min, A., Gertler, T.S., Surmeier, D.J., Tarakhovsky, A.,
533 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>.
- 534 10. 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
535 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>.
- 536 11. 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
537 Family of Histone H3 Lysine 4 Demethylases. *Cell* *128*, 1077–1088. <https://doi.org/10.1016/j.cell.2007.02.017>.
- 538 12. 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
539 Med Genet* *94*, 1–4.
- 540 13. Jensen, L.R., Amende, M., Gurok, U., Moser, B., Gimmel, V., Tzschach, 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
541 Genet* *76*, 227–236. <https://doi.org/10.1086/427563>.
- 542 14. 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.
543 <https://doi.org/10.1111/cge.13755>.

- 544 15. 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>.
- 545
- 546 16. 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>.
- 547
- 548 17. 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>.
- 549
- 550 18. 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>.
- 551
- 552 19. 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>.
- 553
- 554 20. 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>.
- 555
- 556 21. Hackett, J.A., Reddington, J.P., Nestor, C.E., Dunican, D.S., Branco, M.R., Reichmann, J., Reik,
W., Surani, M.A., Adams, I.R., and Meehan, R.R. (2012). Promoter DNA methylation couples
genome-defence mechanisms to epigenetic reprogramming in the mouse germline. *Development*
139, 3623–3632. <https://doi.org/10.1242/dev.081661>.
- 557
- 558 22. 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>.
- 559
- 560 23. 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>.
- 561
- 562 24. 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>.
- 563

- 564 25. Xu, J., Deng, X., and Disteche, C.M. (2008). Sex-Specific Expression of the X-Linked Histone
565 Demethylase Gene Jarid1c in Brain. PLoS ONE 3, e2553. <https://doi.org/10.1371/journal.pone.0002553>
- 566 26. Li, B., Qing, T., Zhu, J., Wen, Z., Yu, Y., Fukumura, R., Zheng, Y., Gondo, Y., and Shi, L. (2017). A
567 Comprehensive Mouse Transcriptomic BodyMap across 17 Tissues by RNA-seq. Sci Rep 7, 4200.
<https://doi.org/10.1038/s41598-017-04520-z>.
- 568 27. Vallianatos, C.N., Raines, B., Porter, R.S., Bonefas, K.M., Wu, M.C., Garay, P.M., Collette, K.M.,
569 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>.
- 570 28. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for
571 RNA-seq data with DESeq2. Genome Biol 15, 550. <https://doi.org/10.1186/s13059-014-0550-8>.
- 572 29. Crackower, M.A., Kolas, N.K., Noguchi, J., Sarao, R., Kikuchi, K., Kaneko, H., Kobayashi, E., Kawai, Y.,
573 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>.
- 574 30. Xiòl, J., Corà, E., Koglgruber, R., Chuma, S., Subramanian, S., Hosokawa, M., Reuter, M., Yang, Z.,
575 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>.
- 576 31. Cheng, S., Altmeppen, G., So, C., Welp, L.M., Penir, S., Ruhwedel, T., Menelaou, K., Harasimov, K.,
577 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>.
- 578 32. Rouland, A., Masson, D., Lagrost, L., Vergès, B., Gautier, T., and Bouillet, B. (2022). Role of
579 apolipoprotein C1 in lipoprotein metabolism, atherosclerosis and diabetes: A systematic review.
Cardiovasc Diabetol 21, 272. <https://doi.org/10.1186/s12933-022-01703-5>.
- 580 33. Leduc, V., Jasmin-Bélanger, S., and Poirier, J. (2010). APOE and cholesterol homeostasis in
581 Alzheimer's disease. Trends in Molecular Medicine 16, 469–477. <https://doi.org/10.1016/j.molmed.2010.07.008>.
- 582 34. Mueller, J.L., Skaletsky, H., Brown, L.G., Zaghlul, S., Rock, S., Graves, T., Auger, K., Warren,
583 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>.
- 584 35. Handel, M.A., and Eppig, J.J. (1979). Sertoli Cell Differentiation in the Testes of Mice Genetically
585 Deficient in Germ Cells. Biology of Reproduction 20, 1031–1038. <https://doi.org/10.1095/biolreprod20.5.1031>.

- 586 36. 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>.
- 587 37. 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>.
- 589 38. 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>.
- 590 39. 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>.
- 593 40. 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>.
- 595 41. 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>.
- 597 42. 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>.
- 600 43. 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>.
- 601 44. 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>.
- 603 45. 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>.
- 604 46. 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>.

- 608 47. 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**,
609 2474–2479. <https://doi.org/10.1073/pnas.0510813103>.
- 610 48. 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>.
- 612 49. Endo, T., Mikedis, M.M., Nicholls, P.K., Page, D.C., and De Rooij, D.G. (2019). Retinoic Acid and Germ
613 Cell Development in the Ovary and Testis. *Biomolecules* **9**, 775. <https://doi.org/10.3390/biom9120775>.
- 614 50. Gupta, N., Yakhou, L., Albert, J.R., Azogui, A., Ferry, L., Kirsh, O., Miura, F., Battault, S., Yamaguchi,
K., Laisné, M., et al. (2023). A genome-wide screen reveals new regulators of the 2-cell-like cell state.
615 *Nat Struct Mol Biol*. <https://doi.org/10.1038/s41594-023-01038-z>.
- 616 51. 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>.
- 618 52. 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-
619 Specific Gene Expression. *Current Biology* **15**, 1051–1057. <https://doi.org/10.1016/j.cub.2005.04.060>.
- 620 53. 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*
621 **14**, e1007193. <https://doi.org/10.1371/journal.pgen.1007193>.
- 622 54. 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
623 during embryonic development. *Nat Commun* **12**, 3582. <https://doi.org/10.1038/s41467-021-23596-w>.
- 624 55. 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.
625 <https://doi.org/10.1016/j.molcel.2010.05.004>.
- 626 56. 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**,
627 916–921. <https://doi.org/10.1038/35002634>.
- 628 57. 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).

- 630 58. 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>.
- 631
- 632 59. 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>.
- 633
- 634 60. Wu, Y., Hu, X., Li, Z., Wang, M., Li, S., Wang, X., Lin, X., Liao, S., Zhang, Z., Feng, X., et al. (2016). Transcription Factor RFX2 Is a Key Regulator of Mouse Spermiogenesis. *Sci Rep* 6, 20435. <https://doi.org/10.1038/srep20435>.
- 635
- 636 61. Liu, M., Zhu, Y., Xing, F., Liu, S., Xia, Y., Jiang, Q., and Qin, J. (2020). The polycomb group protein PCGF6 mediates germline gene silencing by recruiting histone-modifying proteins to target gene promoters. *J Biol Chem* 295, 9712–9724. <https://doi.org/10.1074/jbc.RA119.012121>.
- 637
- 638 62. Otani, J., Nankumo, T., Arita, K., Inamoto, S., Ariyoshi, M., and Shirakawa, M. (2009). Structural basis for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX–DNMT3–DNMT3L domain. *EMBO Reports* 10, 1235–1241. <https://doi.org/10.1038/embor.2009.218>.
- 639
- 640 63. Guo, X., Wang, L., Li, J., Ding, Z., Xiao, J., Yin, X., He, S., Shi, P., Dong, L., Li, G., et al. (2015). Structural insight into autoinhibition and histone H3-induced activation of DNMT3A. *Nature* 517, 640–644. <https://doi.org/10.1038/nature13899>.
- 641
- 642 64. 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>.
- 643
- 644 65. Zhang, J., Zhang, M., Acampora, D., Vojtek, M., Yuan, D., Simeone, A., and Chambers, I. (2018). OTX2 restricts entry to the mouse germline. *Nature* 562, 595–599. <https://doi.org/10.1038/s41586-018-0581-5>.
- 645
- 646 66. 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 spermatogenesis. *Proc. Natl. Acad. Sci. U.S.A.* 112. <https://doi.org/10.1073/pnas.1505683112>.
- 647
- 648 67. Bowles, J., Knight, D., Smith, C., Wilhelm, D., Richman, J., Mamiya, S., Yashiro, K., Chawengsaksophak, 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>.
- 649
- 650 68. Yang, M., Yu, H., Yu, X., Liang, S., Hu, Y., Luo, Y., Izsvák, Z., Sun, C., and Wang, J. (2022). Chemical-induced chromatin remodeling reprograms mouse ESCs to totipotent-like stem cells. *Cell Stem Cell* 29, 400–418.e13. <https://doi.org/10.1016/j.stem.2022.01.010>.
- 651

- 652 69. Al Adhami, H., Vallet, J., Schaal, C., Schumacher, P., Bardet, A.F., Dumas, M., Chicher, J., Hammann, P., Daujat, S., and Weber, M. (2023). Systematic identification of factors involved in the silencing of germline genes in mouse embryonic stem cells. *Nucleic Acids Research* *51*, 3130–3149. <https://doi.org/10.1093/nar/gkad071>.
- 653
- 654 70. Karimi, M.M., Goyal, P., Maksakova, I.A., Bilenky, M., Leung, D., Tang, J.X., Shinkai, Y., Mager, D.L., 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
- 656 71. Tatsumi, D., Hayashi, Y., Endo, M., Kobayashi, H., Yoshioka, T., Kiso, K., Kanno, S., Nakai, Y., Maeda, I., Mochizuki, K., et al. (2018). DNMTs and SETDB1 function as co-repressors in MAX-mediated repression of germ cell-related genes in mouse embryonic stem cells. *PLoS ONE* *13*, e0205969. <https://doi.org/10.1371/journal.pone.0205969>.
- 657
- 658 72. Bonefas, K.M., and Iwase, S. (2021). Soma-to-germline transformation in chromatin-linked neurode-659
660 73. Velasco, G., Walton, E.L., Sterlin, D., Hédouin, S., Nitta, H., Ito, Y., Fouyssac, F., Mégarbané, A., Sasaki, H., Picard, C., et al. (2014). Germline genes hypomethylation and expression define a molecular signature in peripheral blood of ICF patients: Implications for diagnosis and etiology. *Orphanet J Rare Dis* *9*, 56. <https://doi.org/10.1186/1750-1172-9-56>.
- 661
- 662 74. Walton, E.L., Francastel, C., and Velasco, G. (2014). Dnmt3b Prefers Germ Line Genes and Cen-663
664 75. 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 duplication syndrome. *Nat Genet* *44*, 206–211. <https://doi.org/10.1038/ng.1066>.
- 665
- 666 76. Stephens, M. (2016). False discovery rates: A new deal. *Biostat*, kxw041. <https://doi.org/10.1093/biostatistics/kxw041>.
- 667
- 668 77. Conway, J.R., Lex, A., and Gehlenborg, N. (2017). UpSetR: An R package for the visualization of intersecting sets and their properties. *Bioinformatics* *33*, 2938–2940. <https://doi.org/10.1093/bioinformatics/btx364>.
- 669
- 670 78. 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>.
- 671
- 672 79. Mikedis, M.M., Fan, Y., Nicholls, P.K., Endo, T., Jackson, E.K., Cobb, S.A., De Rooij, D.G., and Page, 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>.

673

674 **Figures and Tables**

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

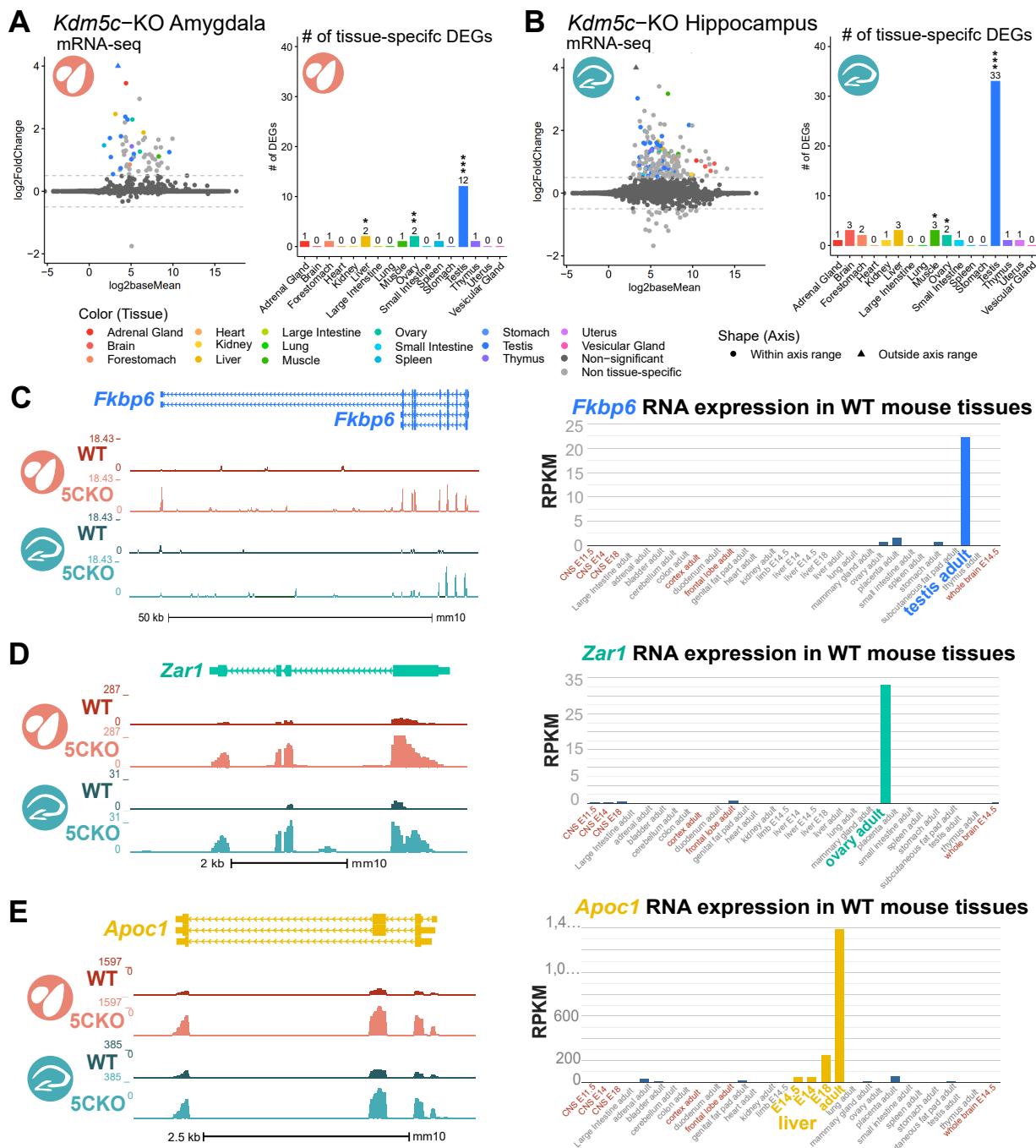


Figure 1: Tissue-enriched genes are misexpressed in the *Kdm5c*-KO brain. **A-B.** Expression of tissue-enriched genes (Li et al 2017) in the male *Kdm5c*-KO amygdala (A) and hippocampus (B). Left - MA plot of mRNA-sequencing. Right - Number of tissue-enriched differentially expressed genes (DEGs). *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



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

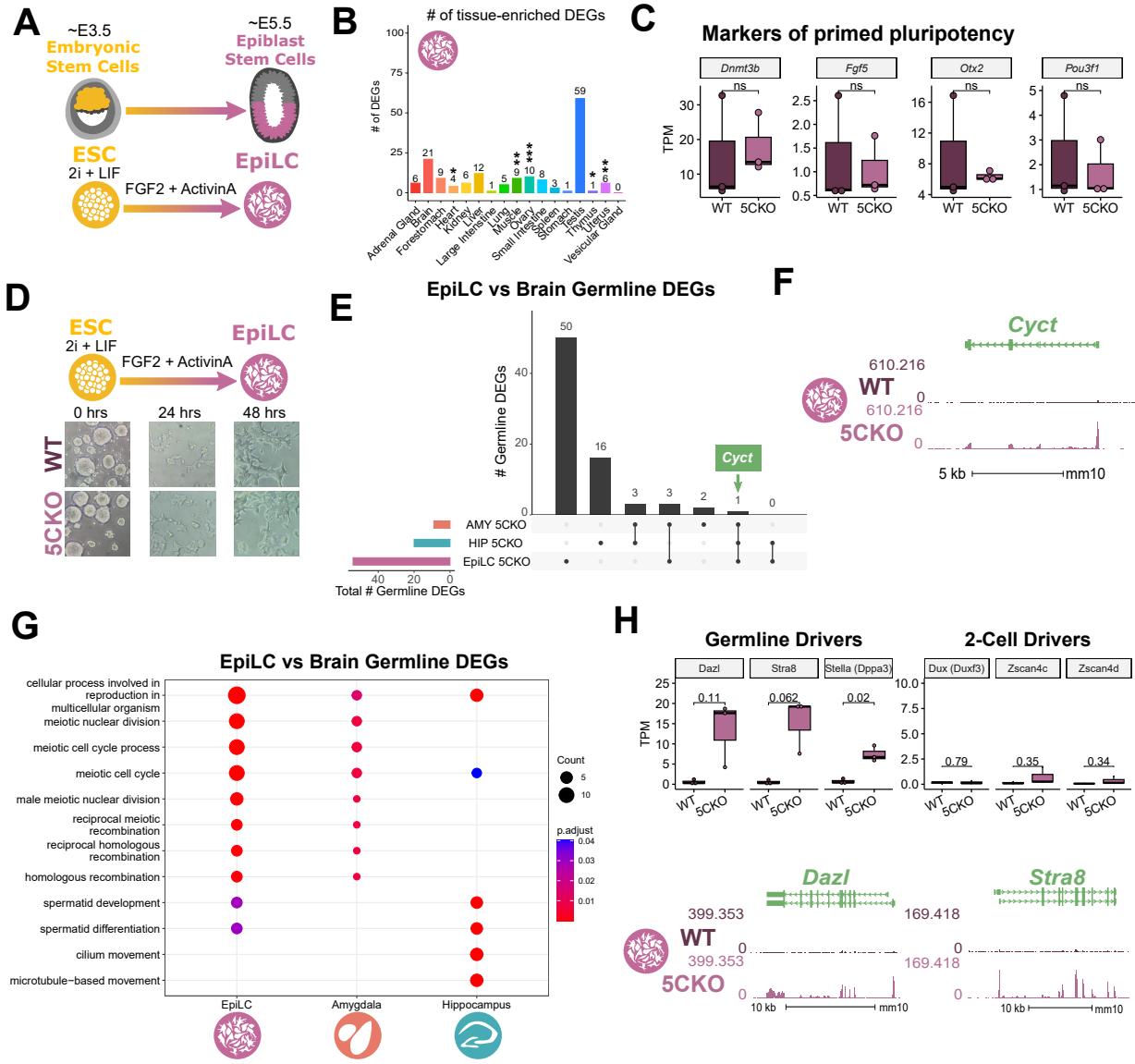


Figure 3: Kdm5c-KO epiblast-like cells express key drivers of germline identity **A.** Top - Diagram of *in vivo* differentiation of embryonic stem cells (ESCs) of the inner cell mass into epiblast stem cells. Bottom - *in vitro* differentiation of ESCs into epiblast-like cells (EpiLCs). **B.** Number of tissue-enriched differentially expressed genes (DEGs) in *Kdm5c*-KO EpiLCs. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, Fisher's exact test. **C.** Average bigwigs of an example germline gene, *Cyct*, that is dysregulated *Kdm5c*-KO EpiLCs. **D.** No significant difference in primed pluripotency marker expression in wild-type versus *Kdm5c*-KO EpiLCs. Welch's t-test, expression in transcripts per million (TPM). **E.** Representative images of wild-type (WT) and *Kdm5c*-KO cells during ESC to EpiLC differentiation. Brightfield images taken at 20X. **F.** Upset plot displaying the overlap of germline DEGs expressed in *Kdm5c*-KO EpiLCs, amygdala (AMY), and hippocampus (HIP) RNA-seq datasets. **G.** enrichPlot comparing enriched biological process gene ontologies for *Kdm5c*-KO EpiLC, amygdala, and hippocampus germline DEGs. **H.** Top left - Example germline identity DEGs unique to EpiLCs. Top right - Example 2-cell genes that are not dysregulated in *Kdm5c*-KO EpiLCs. p-values for Welch's t-test. Bottom - Average bigwigs of *Dazl* and *Stra8* expression in wild-type and *Kdm5c*-KO EpiLCs. **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.

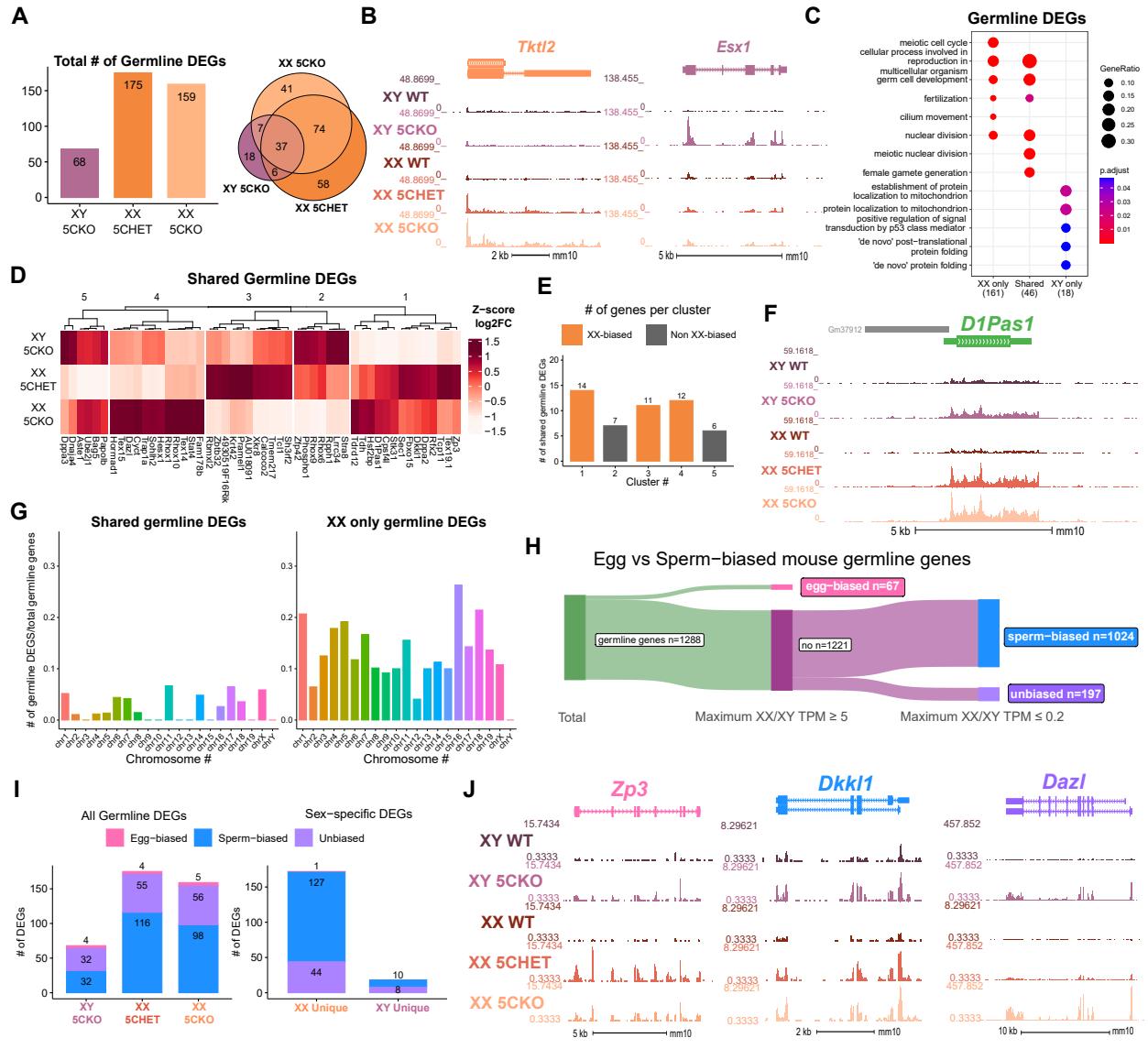


Figure 4: Chromosomal sex influences *Kdm5c*-KO germline gene misexpression. A-B. Expression of tissue-enriched genes (Li et al 2017) in the male *Kdm5c*-KO amygdala (A) and hippocampus (B). Left - MA plot of mRNA-seq. Right - Number of tissue-enriched differentially expressed genes (DEGs). * $p<0.05$, ** $p<0.01$, *** $p<0.001$, 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.

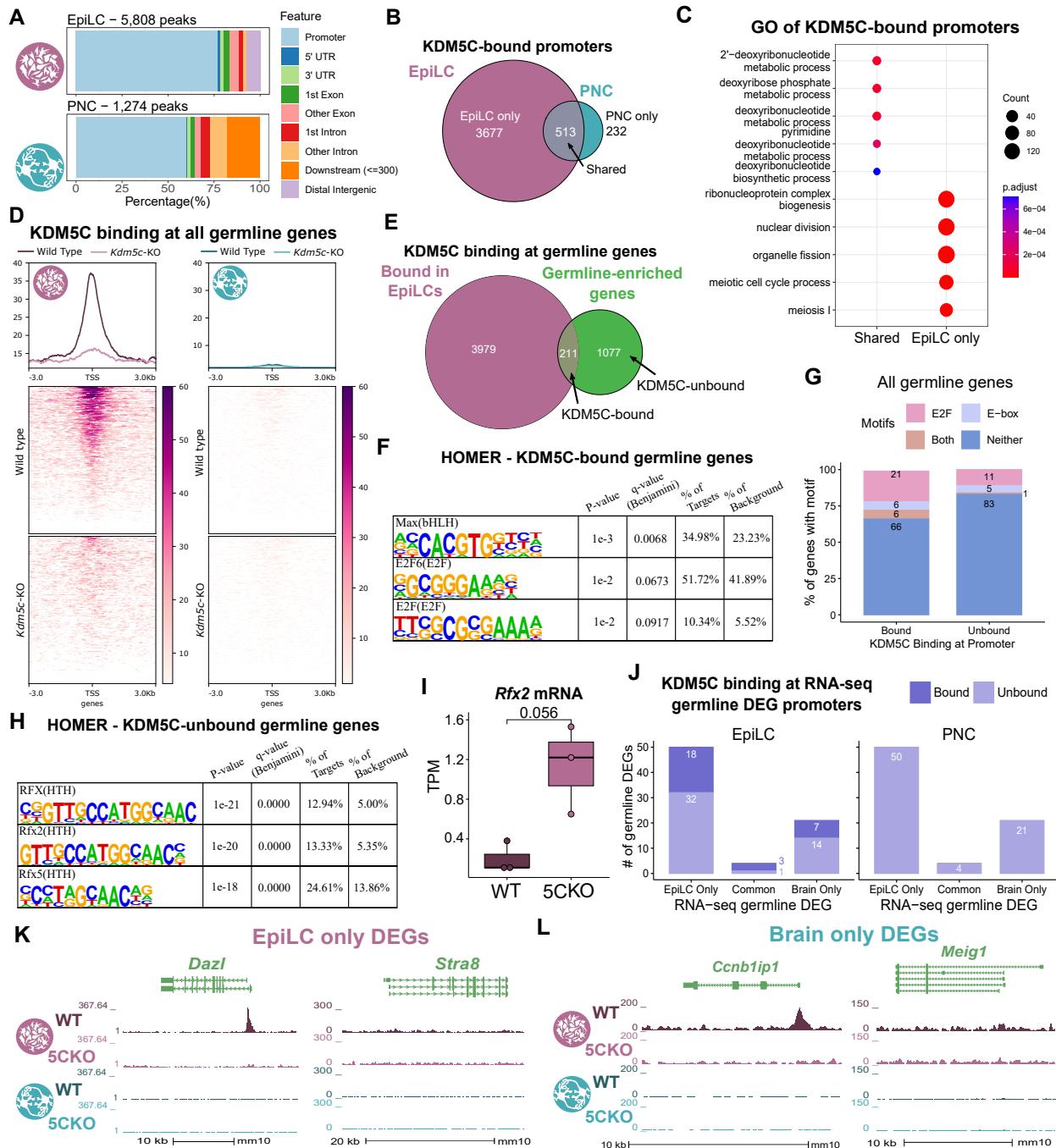
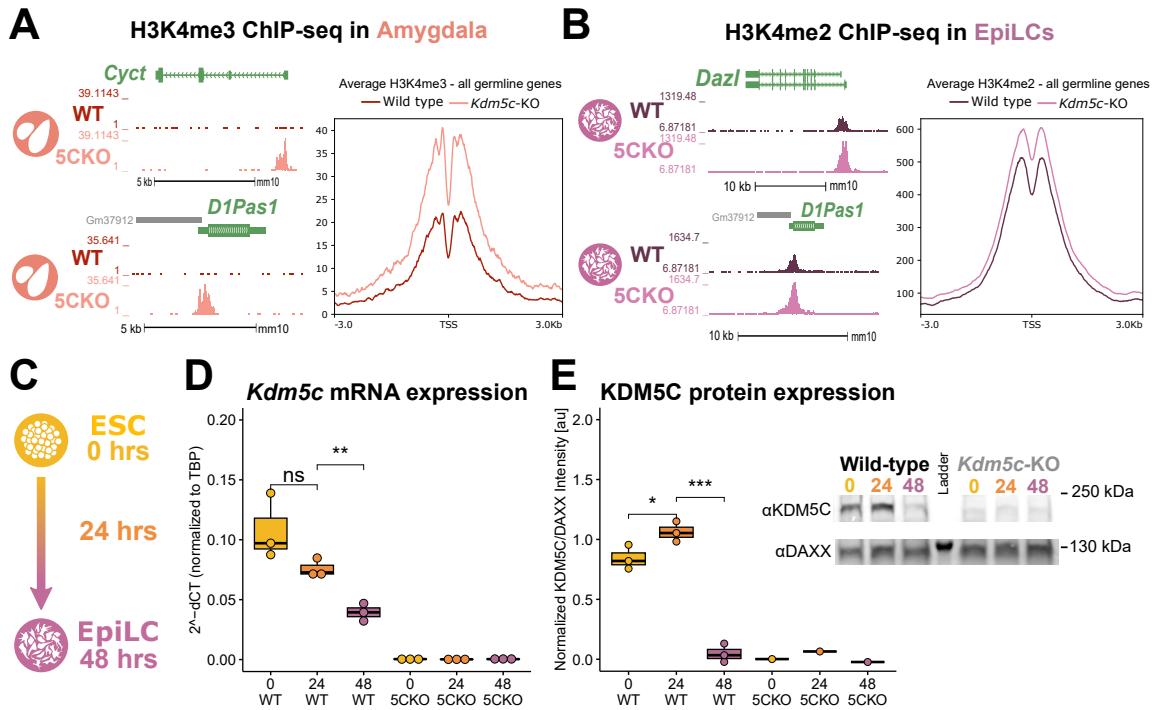


Figure 5: KDM5C binds to a subset of germline gene promoters during early embryogenesis **A.** ChIPseeker localization of KDM5C peaks at different genomic regions in EpiLCs (top) and hippocampal and cortex primary neuron cultures (PNCs, bottom). **B.** Overlap of genes with KDM5C bound to their promoters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



F

G

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

679 Notes

680 Things to do

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

685 Dazl

686 We were particularly interested in the aberrant transcription of *Dazl*, since it is essential for germ cell
687 development and promotes the translation of germline mRNAs⁷⁸. A significant portion of germline transcripts
688 misexpressed in *Kdm5c*-KO EpiLCs are known binding targets of DAZL, including *Stra8*⁷⁹ ($p = 1.698e-07$,
689 Fisher's Exact Test). This suggests expression of DAZL protein could promote the translation of other
690 aberrant germline transcripts, influencing their ability to impact *Kdm5c*-KO cellular function. We thus tested
691 DAZL protein expression in *Kdm5c*-KO EpiLCs through immunocytochemistry (Figure 3I). We observed
692 about 25% of *Kdm5c*-KO EpiLCs expressed DAZL protein and it was localized to the cytoplasm ($p = 0.0015$,
693 Welch's t-test), consistent with the pattern of DAZL expression in spermatogonia⁷⁹. Altogether these results
694 suggest tissue-specific genes are misexpressed during *Kdm5c*-KO embryogenesis, including key drivers of
695 germline identity that can be translated into protein.

- 696 • We additionally found *Kdm5c*-KO EpiLCs ectopically express DAZL protein that is localized to the
697 cytoplasm, similar to its morphology in spermatogonia⁷⁹. **note: maybe just put in results.** Could
698 move around depending upon if I get pheno working.

699 Discussion notes

- 700 • For other paper:
 - 701 – for methods: Heatmaps of gene expression were generated using the base R functions scale and
702 hclust and visualized using the R package ComplexHeatmap (v2.12.1).
 - 703 – * 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>
- 705 • Talk about different motifs if we do see differences there and if it explains direct vs indirect dysregulation
 - 706 – Could also mention PRC1.6/KDM5C binds pcgf6 in ESCs
- 707 • Maybe talk about other regulators/if anything is known about them long term? Or save for 2nd paper.

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

- 736 • Altogether, these data indicate that while some germline genes are misexpressed due to direct loss of
737 KDM5C binding during emryogenesis, secondary downstream mechanisms can also promote their
738 aberrant transcription.
- 739 • Papers to read/reference:
- 740 – Reconstitution of the Mouse Germ Cell Specification Pathway in Culture by Pluripotent Stem Cells:
741 [https://www.cell.com/fulltext/S0092-8674\(11\)00771-9](https://www.cell.com/fulltext/S0092-8674(11)00771-9)
- 742 – 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/>

744 **Figure outline:**

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

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

757 **Figure 3: *Kdm5c*-KO epiblast-like cells express key drivers regulators of germline identity** * A) ESC
758 to EpiLC differentiation Left - Morphology is unchanged, * B) 5CKO EpiLCs express EpiLC differentiation
759 genes similar to WT lvs * C) Male EpiLCs express germline genes (example Cyct again) * Overlap between
760 brain and EpiLC germline genes - show they're mostly unique * GO of Brain and EpiLC germline genes
761 (meiotic enriched) * Bigwigs or TPM of master regulators * Show that while some are also 2-cell genes
762 (Dazl), 2-cell specific genes aren't dysregulated (Zscan4). Important point because published KDM5C dazl
763 paper is saying KDM5C is a 2-cell regulator, but as far as I can tell only genes shared between germline and
764 2-cell are dysregulated.

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

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

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

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

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

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

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

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

809 **Germline gene repression background:**

810 Interestingly, some of the ectopic testis transcripts identified in the *Kdm5c*-KO brain are typically ex-
811 pressed in germ cells¹⁰. Unlike somatic cells, germ cells (e.g. sperm and eggs) undergo meiosis and pass
812 on their genetic material to the next generation. The germline and the soma are typically distinguished during
813 early embryogenesis, when germline genes are silenced in epiblast stem cells soon after implantation and
814 only reactivated in a subset to form the germline. Chromatin regulators play a key role in decommissioning
815 germline genes as the embryo transitions from naïve to primed pluripotency by placing repressive histone
816 H2A lysine 119 monoubiquitination (H2AK119ub1)¹⁸, histone 3 lysine 9 trimethylation (H3K9me3)^{18,19}, and
817 DNA CpG methylation^{19–21} at germline gene promoters. KDM5C may also be involved in this early decom-
818 missioning of germline genes, as re-expression of KDM5C in neurons fails to suppress their dysregulation¹⁰.
819 In support of this, KDM5C was very recently shown to repress *Deleted in azoospermia like (Dazl)*, a key
820 regulator of germline development, in mouse embryonic stem cells (ESCs)^{50,69}. In support of this, two
821 independent screens in mouse embryonic stem cells (ESCs) recently identified KDM5C as a repressor of
822 *Deleted in azoospermia like (Dazl)*, a key regulator of germline development. However, KDM5C's role in
823 embryonic germline gene repression is currently unclear, given that Dazl is also expressed in ESCs and in
824 the 2-cell stage and germline gene misexpression has yet to be globally characterized during *Kdm5c*-KO
825 embryogenesis.