

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

3

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

5 Abstract

It is currently unclear why mutations in numerous chromatin-modifying enzymes cause neurodevelopmental disorders (NDDs). 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 these aberrant, tissue-specific transcripts, the mechanisms behind their dysregulation and their functional consequences are largely unknown. Here, we explored this crisis in cellular identity 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 express 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*. We found that while KDM5C promotes the initial placement of DNA methylation at a subset of germline gene promoters in EpiLCs, germline genes can also become activated 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 provides novel insight into the demarcation of somatic and germline lineages in mammals while also characterizing the cellular identity crisis within a chromatin-linked neurodevelopmental disorder.

29 Introduction

30 A single genome holds the instructions to generate the myriad of cell types found within the adult organism.
31 This is, in part, accomplished by chromatin regulators that can either promote or impede lineage-specific
32 gene expression through DNA and histone modifications^{1,2}. While many chromatin regulators were initially
33 identified for their roles in shaping cellular and tissue identity^{3–5}, recent advancements in next generation
34 sequencing unexpectedly revealed many neurodevelopmental disorders (NDDs) are caused by mutations in
35 chromatin regulators⁶. Several studies have suggested this connection between chromatin regulators and
36 neurodevelopment is due to their regulation of brain-specific genes, such as orchestrating transcriptional
37 programs for synaptic maturation⁷ and transitioning between neuronal and glial fates during neural precursor
38 differentiation⁸. However, loss of some chromatin regulators can also result in the ectopic transcription
39 of tissue-specific genes outside of their target environment, such as the misexpression of heart^{9,10} and
40 liver-specific¹¹ genes within adult neurons. Very few studies have investigated this cellular identity crisis
41 in chromatin-linked NDDs^{11,12} and it is currently unknown if partial loss of brain identity contributes to
42 neurodevelopmental impairments.

43 To elucidate the role of tissue identity in chromatin-linked neurodevelopmental disorders, it is essential
44 to first characterize the types of genes dysregulated and the molecular mechanisms governing their de-
45 repression. In this study, we characterized the aberrant expression of tissue-enriched genes with loss of
46 lysine demethylase 5C (KDM5C). KDM5C, , also known as SMCX or JARID1C, is a chromatin regulator that
47 can repress gene expression through erasure of histone 3 lysine 4 di- and trimethylation (H3K4me2/3)¹³,
48 marks enriched at active gene promoters. Pathogenic mutations in *KDM5C* cause Intellectual Developmental
49 Disorder, X-linked, Syndromic, Claes-Jensen Type (MRXSCJ, OMIM: 300534), whose predominant features
50 include intellectual disability, seizures, aberrant aggression, and autistic behaviors^{14–16}. *Kdm5c* knockout
51 (-KO) mice recapitulate key MRXSCJ patient phenotypes, including hyperaggression, increased seizure
52 propensity, and learning impairments^{12,17}. Unexpectedly, RNA sequencing (RNA-seq) of the *Kdm5c*-KO
53 hippocampus revealed ectopic expression of testis genes within the brain¹². It is currently unknown what types
54 of testis genes are dysregulated, when in *Kdm5c*-KO development this begins, and if other tissue-specific
55 genes are also aberrantly expressed with KDM5C loss.

56 The testis contains both germ cells (cells that pass on their genetic material, e.g. sperm) and somatic
57 cells (cells that perform all other bodily functions). Distinguishing between the germline and the soma is a
58 key feature of multicellularity and that typically occurs during early embryogenesis. In mammals, chromatin
59 regulators play a key role in decommissioning germline genes in somatic cells during the transition from naïve
60 to primed pluripotency. Initially, repressive histone H2A lysine 119 monoubiquitination (H2AK119ub1)¹⁸,
61 histone 3 lysine 9 trimethylation (H3K9me3)^{18,19}, are placed at germline gene promoters in embryonic stem
62 cells and are then decorated with DNA CpG methylation^{19–21} in the post-implantation embryo. How KDM5C
63 interacts with known germline gene silencers during this time point is currently unknown. Additionally, most

64 studies have focused on select genes important for early germ cell development rather than germline-enriched genes as a whole given the lack of a curated list of germline-enriched genes. Thus, it is unclear
65 if the mechanism of repression differs for certain classes of germline genes, e.g. meiotic genes versus
66 spermatid differentiation genes. Systematically characterizing KDM5C's role in germline gene repression
67 during early embryogenesis will unveil key mechanisms underlying the demarcation between soma and
68 germline identity and while also providing molecular footholds to test the impact of ectopic germline genes
69 on neurodevelopment.

70 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 misexpression of genes typically highly enriched within the testis, liver, muscle, and ovary. Both the *Kdm5c*-KO amygdala and hippocampus had significant enrichment of testis genes. We demonstrated testis and ovary-enriched genes are germline genes by generating a list of germline enriched-genes using germ cell-depleted RNA sequencing datasets. Using this curated list, 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. Additionally, KDM5C was highly enriched at germline gene promoters in EpiLCs but only bound to a subset of germline genes expressed in *Kdm5c*-KO cells, 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.

86 Results

87 **Tissue-enriched genes, including testis genes, are aberrantly expressed in the 88 *Kdm5c*-KO brain**

89 Previous RNA sequencing (RNA-seq) in the adult hippocampus revealed ectopic expression of some testis genes within the *Kdm5c* knockout (-KO) brain¹². Since tissue-enriched genes have not been systematically characterized in the *Kdm5c*-KO brain, if the testis is the only tissue type misexpressed and if loss of tissue identity a major consequence of KDM5C loss. Therefore, we globally assessed the expression of genes enriched in 17 mouse tissues²² in our published mRNA-seq datasets of the adult amygdala and hippocampus in mice with constitutive knockout of *Kdm5c*²³.

90 We found a large proportion of genes that are significantly upregulated within the *Kdm5c*-KO brain (DESeq2²⁴, log2 fold change > 0.5, q < 0.1) are enriched in non-brain tissues in wild-type mice (Amygdala:

97 35%, Hippocampus: 24%) (Figure 1A-B). The majority of tissue-enriched differentially expressed genes
98 (tissue-enriched DEGs) were testis genes (Figure 1A-C). Even though the testis has the largest total number
99 of tissue-biased genes compared to any other tissue (2,496 genes), testis-biased DEGs were significantly
100 enriched for both brain regions (Amygdala $p = 1.83e-05$, Odds Ratio = 5.13; Hippocampus $p = 4.26e-11$,
101 Odds Ratio = 4.45, Fisher's Exact Test). One example of a testis-enriched DEG is *FK506 binding protein 6*
102 (*Fkbp6*), a known regulator of piRNA expression and meiosis in germ cells^{25,26} (Figure 1C).

103 In addition to the high enrichment of testis genes, we also identified aberrant expression of other tissue-
104 enriched genes within the *Kdm5c*-KO brain. Although the *Kdm5c*-KO mice we examined were male, we
105 observed significant enrichment of ovary-biased DEGs in both the amygdala and hippocampus (Amygdala
106 $p = 0.00574$, Odds Ratio = 18.7; Hippocampus $p = 0.048$, Odds Ratio = 5.88, Fisher's Exact) (Figure 1D).
107 Ovary-enriched DEGs included *Zygotic arrest 1* (*Zar1*), sequesters mRNAs in oocytes for meiotic maturation
108 and early zygote development²⁷ (Figure 1D). Although not consistent across brain regions, we also found
109 significant enrichment of DEGS biased towards two non-gonadal tissues - the liver (Amygdala $p = 0.0398$,
110 Odds Ratio = 6.58, Fisher's Exact Test) and the muscle (Hippocampus $p = 0.0104$, Odds Ratio = 6.95,
111 Fisher's Exact Test). An example of a liver-biased DEG dysregulated in both the hippocampus and amygdala
112 is *Apolipoprotein C-I* (*Apoc1*), a lipoprotein metabolism and transport gene²⁸ (Figure 1E). Testis, ovary,
113 and liver-enriched DEGs showed little to no expression in the developing and adult wild-type brain, yet our
114 mRNA-seq data indicate they are polyadenylated and spliced into mature transcripts (Figure 1C-E). Of
115 note, we did not observe enrichment of brain-enriched genes (Amygdala $p = 1$; Hippocampus $p = 0.74$,
116 Fisher's Exact), despite the fact these are brain samples and the brain has the second highest total number
117 of tissue-enriched genes (708 genes). Together, these results suggest misexpression of tissue-enriched
118 genes within the brain is a major effect of KDM5C loss.

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

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

126 To further validate if these testis DEGs are truly germline genes, we then evaluated their expression
127 in somatic versus germ cells within the testis. We first compared their expression in the testis with germ
128 cell depletion²⁹, which was accomplished by heterozygous *W* and *Wv* mutations in the enzymatic domain
129 of *c-Kit* (*Kit^{W/Wv}*) that prevent the maturation of germ cells³⁰. Almost all *Kdm5c*-KO testis-enriched DEGs
130 lost expression with germ cell depletion (Figure 2B). We then assessed testis-enriched DEG expression
131 in a published single cell RNA-seq dataset that identified cell type-specific markers within the testis³¹. We

132 found that while some testis-enriched DEGs were classified as specific markers for different germ cell types
133 (e.g. spermatogonia, spermatocytes, round spermatids, and elongating spermatids) none marked somatic
134 cells (Figure 2C). Together, these data demonstrate that the *Kdm5c*-KO brain aberrantly expresses germline
135 genes, reflecting an erosion between somatic versus germline identity.

136 We then wanted to more deeply characterize germline gene misexpression with *Kdm5c* loss, but lacked
137 a comprehensive list of mouse germline-enriched genes. To facilitate downstream analyses, we generated
138 a list of germline-enriched genes using currently available RNA-seq datasets in Kit^{W/Wv} mice. Wild-type
139 and Kit^{W/Wv} datasets included males and females at embryonic day 12, 14, and 16³², as well as adult
140 male testes²⁹. We defined genes as germline-enriched if their expression met the following criteria: 1)
141 their expression is greater than 1 FPKM in wild-type gonads 2) their expression in any adult wild-type,
142 non-gonadal tissue²² does not exceed 20% of their maximum expression in the wild-type germline, and
143 3) their expression in the germ cell-depleted germline, for any sex or time point, does not exceed 20% of
144 their maximum expression in the wild-type germline. These criteria yielded 1,288 germline-enriched genes
145 (Figure 2D), which was hereafter used as a resource to globally assess germline gene misexpression with
146 *Kdm5c* loss (Supplementary table 1).

147 ***Kdm5c*-KO epiblast-like cells aberrantly express master regulators of germline
148 identity**

149 Germ cells are typically distinguished from somatic cells soon after the embryo implants into the uter-
150 ine wall^{33,34} when a subset of epiblast stem cells become the primordial germ cells (PGCs) while the
151 remainder differentiate into the ectoderm, mesoderm, and endoderm to form the somatic tissues³⁵. This
152 developmental time point can be modeled *in vitro* through differentiation of embryonic stem cells (ESCs) into
153 post-implantation epiblast-like stem cells (EpiLCs) (Figure 3A, top). While some germline-enriched genes
154 are also expressed in embryonic stem cells (ESCs) and in the 2-cell stage^{36–38}, they are silenced as they
155 differentiate into EpiLCs¹⁹. Therefore, we tested if KDM5C was necessary for silencing germline genes in
156 the post-implantation embryo by evaluating the impact of *Kdm5c* loss in EpiLCs.

157 We first identified *Kdm5c*-KO EpiLC DEGs in our previously published RNA-seq dataset³⁹ (DESeq2, log2
158 fold change > 0.5, q < 0.1) and assessed tissue-enriched gene expression. Similar to the *Kdm5c*-KO brain,
159 we observed general dysregulation of tissue-enriched genes, with the largest number of genes belonging
160 to the brain and testis, although they were not significantly enriched (Figure 3B). Although we observed
161 aberrant expression of many tissue-enriched genes, including 54 germline-enriched genes, we did not find
162 any significant difference in primed pluripotency genes, such as *Dnmt3b*, *Fgf5*, *Pou3f1*, and *Otx2* (Figure 3C).
163 We also did not observe any gross changes in *Kdm5c*-KO cell morphology during ESC to EpiLC differentiation
164 (Figure 3D), altogether indicating KDM5C loss does not impair EpiLC formation.

165 To determine if germline DEGs are constitutively dysregulated or if they can change over the course of

166 development, we compared compared EpiLC germline DEGs to those expressed in the *Kdm5c*-KO brain.
167 We found the majority of germline DEGs expressed in EpiLCs and the brain were unique, with only *Cyct*
168 shared across all sequencing datasets (Figure 3E-F). We then evaluated the known functions of EpiLC and
169 brain germline DEGs and found particularly high enrichment of meiosis-related gene ontologies in EpiLCs
170 (Figure 3G), such as meiotic cell cycle process (GO: 1903046, p.adjust = 1.59e-08) and meiotic nuclear
171 division (GO: 0140013, p.adjust = 9.76e-09). While there was modest enrichment of meiotic gene ontologies
172 in both brain regions, the *Kdm5c*-KO hippocampus primarily expressed late-stage spermatogenesis genes,
173 such those involved in the sperm axoneme structure.

174 Interestingly, DEGs unique to *Kdm5c*-KO EpiLCs included key drivers of germline identity, such as
175 *Stimulated by retinoic acid 8 (Stra8)* and *Deleted in azoospermia like (Dazl)* (Figure 3H). These genes are
176 typically expressed during embryonic germ cell development to commit PGCs to the germline fate, but are
177 also expressed later in life to trigger meiotic gene expression programs^{40–42}. Of note, some germline genes,
178 including *Dazl*, are also expressed in the two-cell embryo^{37,43}. However, we did not see misexpression of
179 two-cell embryo-specific genes, like *Duxf3 (Dux)* ($q = 0.337$) and *Zscan4d* ($q = 0.381$), indicating *Kdm5c*-KO
180 in EpiLCs do not revert back to a 2-cell state (Figure 3H).

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

182 Previous work suggests KDM5C represses germline genes during early development, as re-expression
183 of KDM5C in knockout neuronal cultures fails to suppress two hippocampal germline DEGs¹². There is some
184 evidence KDM5C binds to select germline gene promoters in embryonic stem cells (ESCs) but is lost in
185 neurons¹². However, the lack of a comprehensive list of germline-enriched genes prohibited systematic
186 characterization of KDM5C binding at germline gene promoters. Thus, it is currently unclear if KDM5C is
187 enriched at germline gene promoters, what types of germline genes KDM5C regulates, and if its binding is
188 maintained at any germline genes in neurons.

189 To further characterize KDM5C's role in germline gene silencing, we analyzed KDM5C chromatin
190 immunoprecipitation followed by DNA sequencing (ChIP-seq) datasets in EpiLCs³⁹ and primary neuron
191 cultures (PNCs) from the cortex and hippocampus¹⁷. EpiLCs had a higher total number of KDM5C peaks
192 than PNCs (EpiLCs: 5,808, PNCs: 1,276, MACS2 $q < 0.1$ and fold enrichment > 1 , removal of *Kdm5c*-KO
193 false positive peaks) and KDM5C was primarily localized to gene promoters in both cell types (EpiLCs: 4,190,
194 PNCs: 745 +/- 500bp from the TSS), although PNCs showed increased localization to non-promoter regions
195 (Figure 4A).

196 The majority of promoters bound by KDM5C in PNCs were also bound in EpiLCs (513 shared promoters),
197 however a large portion of gene promoters were only bound by KDM5C in EpiLCs (3677 EpiLC only
198 promoters) (Figure 4B). We then performed gene ontology analysis to compare the known functions of genes
199 with KDM5C bound to their promoter in EpiLCs and PNCs. While there were no ontologies significantly
200 enriched for genes only bound by KDM5C in PNCs, gene ontologies for genes shared between PNCs and

201 EpiLCs were enriched for functions involving nucleic acid turnover, such as deoxyribonucleotide metabolic
202 process (GO:0009262, p.adjust = 8.28e-05) (Figure 4C). Germline-specific ontologies were only enriched
203 in promoters unique to EpiLCs, such as meiotic nuclear division (GO: 0007127 p.adjust = 6.77e-16) and
204 meiotic cell cycle process (GO:1903046, p.adjust = 5.05e-16) (Figure 3C). We then evaluated KDM5C
205 binding around the transcription start site (TSS) of all germline-enriched genes. While KDM5C was bound to
206 a subset of germline gene promoters in EpiLCs, it was not bound to any in PNCs (Figure 4D-E). Together,
207 these results demonstrate KDM5C is significantly enriched at a subset of germline gene promoters in EpiLCs,
208 including meiotic genes, but does not regulate germline genes in neurons.

209 We then compared KDM5C binding at RNA-seq DEG promoters to determine if the germline mRNAs
210 expressed in the *Kdm5c*-KO brain and EpiLCs are direct targets of KDM5C (Figure 4F). About one third of
211 EpiLC-specific and brain-specific (hippocampus, amygdala, or both) germline DEGs were bound by KDM5C
212 in EpiLCs (EpiLC only: 36%, Brain only: 33.3%). Again, we did not observe any KDM5C binding at germline
213 gene promoters in PNCs, even for brain-specific DEGs (Figure 4H). Some notable differences in KDM5C
214 binding for EpiLC-specific DEGs included *Dazl* and *Stra8*, two key drivers of germline identity discussed
215 above. Although both mRNAs are expressed in *Kdm5c*-KO EpiLCs, KDM5C is only bound to *Dazl*'s promoter
216 and not *Stra8*'s in EpiLCs (Figure 4F). In contrast to the unique DEGs, 3 out of the 4 genes dysregulated
217 in both the brain and EpiLCs are bound by KDM5C around the TSS (Figure 4G). Altogether, this suggests
218 KDM5C decommissions germline genes in EpiLCs but is no longer recruited to germline gene promoters
219 in neurons. Furthermore, majority of germline mRNAs expressed in *Kdm5c*-KO cells are dysregulated
220 independent of direct KDM5C recruitment to their promoters in EpiLCs.

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

222 In the early embryo, germline gene promoters are initially decorated with repressive histone modifications
223 before long-term silencing by *de novo* DNA CpG methylation (CpGme)^{18,19,44}. Our results above suggest
224 KDM5C also acts during this time period, however how KDM5C interacts with other germline gene silencing
225 mechanism is currently unclear. KDM5C is generally thought to suppress transcription through erasure of
226 histone 3 lysine 4 di- and trimethylation (H3K4me2/3), marks enriched at active gene promoters¹³. However,
227 KDM5C's catalytic activity was recently shown to not be required for suppressing *Dazl* in ESCs⁴³. Since
228 H3K4me3 impedes *de novo* CpGme placement^{45,46}, KDM5C's catalytic activity may instead be required for
229 long-term germline gene silencing. In support of this, CpGme is markedly reduced at two germline gene
230 promoters in the adult hippocampus¹². However, because KDM5C's role in germline gene repression has
231 only been characterized at select germline genes in ESCs and in the mature brain, it is currently unclear to
232 what extent KDM5C is involved during initial CpGme placement.

233 To elucidate KDM5C's role in germline gene silencing, we first characterized KDM5C's substrates, histone
234 3 lysine 4 di- and trimethylation (H3K4me2/3) in our previously published ChIP-seq datasets in wild type and
235 *Kdm5c*-KO amygdala²³ and EpiLCs³⁹. In congruence with previous work in the *Kdm5c*-KO hippocampus¹²,

236 we observed aberrant accumulation of H3K4me3 around the transcription start site (TSS) of germline genes
237 in the *Kdm5c*-KO amygdala (Figure 5A). We additionally found a marked increase in H3K4me2 around the
238 TSS of germline genes in *Kdm5c*-KO EpiLCs (Figure 5B).

239 We then evaluated KDM5C's role in germline gene silencing during ESC to EpiLC differentiation. We
240 first characterized KDM5C's mRNA and protein expression at 0, 24, and 48 hours of differentiation (Figure
241 5C). While *Kdm5c* mRNA steadily decreased from 0 to 48 hours of differentiation (Figure 5D), KDM5C
242 protein initially increased from 0 to 24 hours but then decreased to near knockout levels by 48 hours (Figure
243 5E). To determine KDM5C's role in the initial placement of DNA methylation at germline gene promoters,
244 we performed whole genome bisulfite sequencing (WGBS) in wild-type and *Kdm5c*-KO ESCs and 96 hour
245 extended EpiLCs (exEpiLCs). Although wild-type cells accumulated high levels of DNA methylation at
246 germline gene promoters over the course of ESCs to exEpiLC differentiation, DNA methylation was markedly
247 reduced in *Kdm5c*-KO exEpiLCs (Figure 5F).

- 248 • Altogether, this suggests KDM5C is dynamically regulated during ESC to EpiLC differentiation and
249 promotes germline gene silencing via removal of H3K4me2/3 during early embryogenesis.

250 Discussion

251 In the above study, we demonstrate KDM5C's pivotal role in the development of tissue identity, resulting in
252 substantial misexpression of tissue-enriched genes within the adult *Kdm5c*-KO hippocampus and amygdala.
253 In addition to testis genes identified previously¹², we found significant enrichment of muscle, liver, and
254 even ovary-biased genes aberrantly expressed within the male *Kdm5c*-KO brain. Given the majority of
255 tissue-enriched DEGs are testis and ovary genes with no known brain functions, it is currently unknown if
256 they impair *Kdm5c*-KO brain function and contribute to MRXSCJ-related phenotypes like intellectual disability
257 and aberrant aggression. However, select liver and muscle-biased DEGs do have known roles within the
258 brain, such as the liver-enriched lipid metabolism gene *Apolipoprotein C-I* (*Apoc1*)²⁸ that is highly expressed
259 in the *Kdm5c*-KO amygdala and hippocampus. Intriguingly, *Apoc1* overexpression in the mouse brain can
260 impair learning and memory⁴⁷ and is implicated in Alzheimer's disease in humans⁴⁸. Thus, failure to fine-tune
261 the expression of tissue-enriched, dosage-sensitive genes like *Apoc1* could be one route by which loss of
262 brain tissue identity contributes to *Kdm5c*-KO impairments.

263 We found a substantial portion of genes upregulated in *Kdm5c*-KO brain are germline genes with no
264 known functions within the brain. Distinguishing the germline (i.e. cells that pass on their genetic material
265 to the next generation) from the soma (i.e. all other cells of the body) is a key feature of multicellularity
266 and sexual reproduction. Previous work characterizing chromatin regulators that silence germ cell-specific
267 transcription has predominantly focused on their repression of key marker genes in embryonic stem cells
268 (ESCs), such as *Dazl* and *Ddx4*^{18,19,49}. To globally characterize KDM5C's role in germline gene repression
269 throughout life, we curated a list of mouse germline-enriched genes using publicly available germ cell-

270 depleted RNA-seq datasets from *Kit^{W/Wv}* mice^{29,32}. This resource enabled us to identify 1) the extent of
271 germline gene dysregulation in *Kdm5c*-KO cells, 2) the types of germline genes misexpressed at different
272 developmental time points, and 3) the categories of germline genes directly and indirectly regulated by
273 KDM5C.

274 RNA-seq of epiblast-like cells (EpiLCs) revealed dysregulation of *Kdm5c*-KO tissue identity begins during
275 early embryogenesis. *In vivo*, germline genes are typically decommissioned in epiblast stem cells and
276 remain silenced as the epiblast differentiates into the body's somatic tissues³⁵. However, a small subset of
277 epiblast stem cells will receive signals to reactivate germline gene expression to become the primordial germ
278 cells (PGCs) that will ultimately form the mature germline^{33,34}. This process can be mimicked *in vitro* by
279 differentiating EpiLCs into primordial germ cell-like cells (PGCLCs)⁵⁰. Therefore, misexpression of germline
280 genes in EpiLCs might suggest *Kdm5c*-KO ESCs are progressing beyond EpiLC differentiation and instead
281 becoming PGCLCs. Yet, *Kdm5c*-KO EpiLCs properly express primed pluripotency genes, including *Otx2*
282 - an epiblast stem cell marker that is known to repress epiblast and EpiLC differentiation into PGC and
283 PGCLCs⁵¹. Furthermore, we observed no difference in cellular morphology during *Kdm5c*-KO ESC to EpiLC
284 differentiation. Proper EpiLC differentiation, together with *Kdm5c*-KO mice being viable, suggests germline
285 gene expression is occurring ectopically in conjunction with typical developmental programs, rather than a
286 complete shift to towards germline identity.

287 We then globally characterized KDM5C binding at germline-enriched gene promoters through KDM5C
288 ChIP-seq in EpiLCs and primary hippocampus and cortex neuron cultures (PNCs). While we observed
289 high enrichment of KDM5C peaks at germline genes in EpiLCs, KDM5C was not bound to germline gene
290 promoters in PNCs. This suggests dysregulation of germline genes in the mature *Kdm5c*-KO brain is due
291 to loss of repression during embryogenesis, which is consistent with previous work that found introducing
292 human KDM5C into *Kdm5c*-KO PNCs does not repress two aberrant germline transcripts¹². Although
293 enriched at germline gene promoters in EpiLCs, KDM5C was only bound to about a third of the germline
294 genes aberrantly transcribed in *Kdm5c*-KO EpiLCs. One notable EpiLC DEG not bound by KDM5C is
295 *Stra8*, a transcription factor activated by retinoic acid signaling in germ cells to promote meiotic initiation^{52,53}.
296 However, retinoic acid can only activate *Stra8* expression when DAZL is present. Unlike *Stra8*, *Dazl* is a
297 direct target of KDM5C in EpiLCs and is both transcribed and translated in *Kdm5c*-KO EpiLCs. Expression of
298 indirect DEGs like *Stra8* indicates germline genes can be aberrantly expressed in *Kdm5c*-KO cells through
299 downstream activation by other ectopic germline programs. These ectopic germline programs appear to
300 loosely mimic the trajectory of typical germline development, as germline genes important for early germ
301 cell development and meiosis are expressed in *Kdm5c*-KO EpiLC while late spermatogenesis genes are
302 expressed in the mature *Kdm5c*-KO brain. Altogether, this indicates once a subset of germline genes are
303 activated due to loss of direct KDM5C repression in the early embryo, these ectopic germline programs can
304 continue to progress in the background of *Kdm5c*-KO somatic development.

305 It is important to note that while we focused on KDM5C's regulation of germline genes, some germline-

enriched genes are also expressed in naïve ESCs and at the 2-cell stage due to their roles in pluripotency and self-renewal. For example, *Dazl* is primarily known for committing PGCs to the germline fate and regulating the translation of germline-specific RNAs, but is also expressed at the 2-cell stage⁵⁴, in naïve ESCs³⁶, and in the inner cell mass³⁶. Very recently, two screens of *Dazl*-repressors identified KDM5C as a direct regulator in ESCs^{43,55}. Interestingly, one study found *Kdm5c*-KO ESCs also aberrantly expressed 2-cell-specific genes, indicating KDM5C negatively regulates totipotency⁴³. However, out of the four regulators characterized, KDM5C was the only factor whose repression of *Dazl* was independent of the 2-cell specific transcription factor *Dux*⁴³. Although expressed at the 2-cell stage and in ESCs, *Dazl* is typically silenced when ESCs differentiate into EpiLCs¹⁹. We found KDM5C also represses *Dazl* in EpiLCs, but *Kdm5c*-KO EpiLCs did not express 2-cell specific genes like *Dux* and *Zscan4c*. Together, this suggests the 2-cell-like state in *Kdm5c*-KO ESCs is an indirect consequence of KDM5C's primary role in germline gene repression.

Dazl and other germline-enriched genes are initially decommissioned in ESCs by repressive histone modifications and then *de novo* DNA CpG methylation (CpGme) in EpiLCs^{19–21,44,56,57}. Unlike the previously characterized germline gene suppressors, KDM5C removes an active histone mark, histone 3 lysine 4 di- and trimethylation (H3K4me2/3)¹³. In this study, we observed a global increase in H3K4me2/3 in *Kdm5c*-KO EpiLCs and amygdala around the transcription start site (TSS) of germline-enriched genes. However, KDM5C's catalytic activity may not be required for germline gene silencing, as it was recently found to be dispensable for repressing *Dazl* in ESCs⁴³. Although not necessary in ESCs, KDM5C's catalytic activity is required for CpGme in EpiLCs, as H3K4me3 impedes *de novo* CpGme placement^{45,46}. This is supported by previous work in the *Kdm5c*-KO adult hippocampus, which found CpGme is significantly eroded at least two germline promoters¹². To elucidate the mechanism behind KDM5C-mediated silencing of germline genes, we characterized KDM5C in ESCs and EpiLCs and found its expression is dynamically regulated during differentiation. Furthermore, KDM5C acts during the transition from naïve to primed pluripotency to promote the initial placement of CpGme at CpG islands for the long-term silencing of germline genes.

The above work provides the mechanistic foundation for KDM5C's role in germline gene repression and offers novel insight into how dysregulation in *Kdm5c*-KO tissue identity changes over the course of development. Although their impact upon neurodevelopment is currently unclear, ectopic germline transcripts are also found in models of other related neurodevelopmental disorders⁵⁸, including Immunodeficiency, Centromere instability, Facial anomalies syndrome 1 (ICF1 OMIM: #242860)^{59,60}, Kleefstra syndrome 1 (OMIM: #610253)¹¹, and MeCP2 duplication syndrome (MDS, OMIM: #300260)⁶¹. Like KDM5C, the chromatin regulators underlying these conditions, DNA methyltransferase 3b (DNMT3B), H3K9me1/2 methyltransferases G9A/GLP, methyl-CpG -binding protein 2 (MECP2), primarily silence gene expression. Thus, KDM5C is among a growing cohort of chromatin-linked neurodevelopmental disorders that may have a similar underlying cause of germline versus soma dysregulation. However, further research is required to determine the ultimate impact of these germline genes and the extent to which this phenomenon occurs in humans.

342 **Materials and Methods**

343 **Classifying tissue-enriched and germline-enriched genes**

344 Tissue-enriched differentially expresssd genes (DEGs) were determined by their classification in a previ-
345 ously published dataset from 17 male and female mouse tissues²². This study defined tissue expression as
346 greater than 1 Fragments Per Kilobase of transcript per Million mapped read (FPKM) and tissue enrichment
347 as at least 4-fold higher expression than any other tissue.

348 We curated a list of germline-enriched genes using an RNA-seq dataset from wild-type and germline-
349 depleted (Kit^{W/Wv}) male and female mouse embryos from embryonic day 12, 14, and 16³², as well as adult
350 male testes²⁹. Germline-enriched genes met the following criteria: 1) their expression is greater than 1
351 FPKM in wild-type germline 2) their expression in any wild-type somatic tissues²² does not exceed 20%
352 of maximum expression in wild-type germline, and 3) their expression in the germ cell-depleted (Kit^{W/Wv})
353 germline, for any sex or time point, does not exceed 20% of maximum expression in wild-type germline.

354 **Cell culture**

355 We utilized our previously established cultures of male wild-type and *Kdm5c* knockout (-KO) embryonic
356 stem cells³⁹. Sex was confirmed by genotyping *Uba1* on the X and Y chromosomes with the following
357 primers: 5'-TGGATGGTGTGCCAATG-3', 5'-CACCTGCACGTTGCCCTT-3'. Deletion of *Kdm5c* was
358 confirmed through the primers 5'-ATGCCCATATTAAGAGTCCTG-3', 5'-TCTGCCTTGATGGACTGTT-3',
359 and 5'-GGTTCTCAACACTCACATAGTG-3'.

360 Embryonic stem cells (ESCs) were initially cultured in primed ESC (pESC) media consisting of Knock-
361 Out DMEM (Gibco#10829–018), fetal bovine serum (Gibco#A5209501), KnockOut serum replacement
362 (Invitrogen#10828–028), Glutamax (Gibco#35050-061), Anti-Anti (Gibco#15240-062), MEM Non-essential
363 amino acids (Gibco#11140-050), and beta-mercaptoethanol (Sigma#M7522). They were then transitioned
364 into ground-state “naïve” ESCs (nESCs) by culturing for four passages in N2B27 media containing DMEM/F12
365 (Gibco#11330–032), Neurobasal media (Gibco#21103–049), Gluamax, Anti-Anti, N2 supplement (Invitro-
366 gen#17502048), and B27 supplement without vitamin A (Invitrogen#12587-010), and beta-mercaptoethanol.
367 Both pESC and nESC media were supplemented with 3 μ M GSK3 inhibitor CHIR99021 (Sigma #SML1046-
368 5MG), 1 μ M MEK inhibitor PD0325901 (Sigma #PZ0162-5MG), and 1,000 units/mL leukemia inhibitory factor
369 (LIF, Millipore#ESG1107).

370 nESCs were differentiated into epiblast-like cells (EpiLCs, 48 hours) and extendend EpiLCs (exEpiLCs,
371 96 hours) by culturing in N2B27 media containing DMEM/F12, Neurobasal media, Gluamax, Anti-Anti,
372 N2 supplement, B27 supplement (Invitrogen#17504044), beta-mercaptoethanol, fibroblast growth factor 2
373 (FGF2, R&D Biotechne 233-FB), and activin A (R&D Biotechne 338AC050CF), as previously described⁶².

374 **Immunocytochemistry (ICC)**

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

384 **RNA sequencing (RNA-seq)**

385 After ensuring read quality via FastQC (v0.11.8), reads were then mapped to the mm10 *Mus musculus*
386 genome (Gencode) using STAR (v2.5.3a), during which we removed duplicates and kept only uniquely
387 mapped reads. Count files were generated by FeatureCounts (Subread v1.5.0), and BAM files were
388 converted to bigwigs using deeptools (v3.1.3) and visualized by the UCSC genome browser. RStudio (v3.6.0)
389 was then used to analyze counts files by DESeq2 (v1.26.0)²⁴ to identify differentially expressed genes
390 (DEGs) with a q-value (p-adjusted via FDR/Benjamini–Hochberg correction) less than 0.1 and a log2 fold
391 change greater than 0.5. For all DESeq2 analyses, log2 fold changes were calculated with IfcShrink using
392 the ashr package⁶³. MA-plots were generated by ggpibr (v0.6.0), and Eulerr diagrams were generated by
393 eulerr (v6.1.1). Boxplots and scatterplots were generated by ggpibr (v0.6.0) and ggplot2 (v3.3.2). The Upset
394 plot was generated via the package UpSetR (v1.4.0)⁶⁴. Gene ontology (GO) analyses were performed by
395 the R package enrichPlot (v1.16.2) using the biological processes setting and compareCluster.

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

397 ChIP-seq reads were aligned to mm10 using Bowtie1 (v1.1.2) allowing up to two mismatches. Only
398 uniquely mapped reads were used for analysis. Peaks were called using MACS2 software (v2.2.9.1) using
399 input BAM files for normalization, with filters for a q-value < 0.1 and a fold enrichment > 1. We removed
400 “black-listed” genomic regions that often give aberrant signals. Common peak sets were obtained in R via
401 DiffBind (v3.6.5). In the case of KDM5C ChIP-seq, *Kdm5c*-KO peaks were then subtracted from wild-type
402 samples using bedtools (v2.25.0). Peak proximity to genome annotations was determined by ChIPSeeker
403 (v1.32.1). Gene ontology (GO) analyses were performed by the R package enrichPlot (v1.16.2) using the
404 biological processes setting and compareCluster. Enriched motifs were identified using HOMER⁶⁵. Average
405 binding across the genome was visualized using deeptools (v3.1.3). Bigwigs were visualized using the
406 UCSC genome browser.

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

408 **Data availability**

409 **Published datasets**

410 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
411 adult amygdala and hippocampus²³ (available at GEO: GSE127722) and male wild-type and *Kdm5c*-KO
412 EpiLCs³⁹ (available at GEO: GSE96797).

413 Previously published ChIP-seq experiments included KDM5C in wild-type and *Kdm5c*-KO EpiLCs³⁹ (avail-
414 able at GEO: GSE96797) and mouse primary neuron cultures (PNCs) from the cortex and hippocampus¹⁷
415 (available at GEO: GSE61036). ChIP-seq of histone 3 lysine 4 dimethylation in male wild-type and *Kdm5c*-KO
416 EpiLCs³⁹ is also available at GEO: GSE96797. ChIP-seq of histone 3 lysine 4 trimethylation in wild-type and
417 *Kdm5c*-KO male amygdala²³ are available at GEO: GSE127817.

419 **Acknowledgements**

- 420 • Jacob Mueller for providing insight in germline gene regulation.
421 • Sundeep Kalantry for providing reagents and expertise in culturing mouse embryonic stem cells and
422 epiblast like cells
423 • Funding sources

424 **References**

- 425 1. Strahl, B.D., and Allis, C.D. (2000). The language of covalent histone modifications. *Nature* **403**,
426 41–45. <https://doi.org/10.1038/47412>.
- 427 2. Jenuwein, T., and Allis, C.D. (2001). Translating the Histone Code. *Science* **293**, 1074–1080.
428 <https://doi.org/10.1126/science.1063127>.
- 429 3. Lewis, E.B. (1978). A gene complex controlling segmentation in *Drosophila*. *Nature* **276**, 565–570.
430 <https://doi.org/10.1038/276565a0>.
- 431 4. Kennison, J.A., and Tamkun, J.W. (1988). Dosage-dependent modifiers of polycomb and antennapedia
432 mutations in *Drosophila*. *Proc. Natl. Acad. Sci. U.S.A.* **85**, 8136–8140. <https://doi.org/10.1073/pnas.85.21.8136>.
- 433 5. Kassis, J.A., Kennison, J.A., and Tamkun, J.W. (2017). Polycomb and Trithorax Group Genes in
434 *Drosophila*. *Genetics* **206**, 1699–1725. <https://doi.org/10.1534/genetics.115.185116>.

- 435 6. Gabriele, M., Lopez Tobon, A., D'Agostino, G., and Testa, G. (2018). The chromatin basis of
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>.
- 436
- 437 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>.
- 438
- 439 8. Hirabayashi, Y., Suzki, N., Tsuboi, M., Endo, T.A., Toyoda, T., Shinga, J., Koseki, H., Vidal, M., and
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>.
- 440
- 441 9. Mall, M., Karetta, M.S., Chanda, S., Ahlenius, H., Perotti, N., Zhou, B., Grieder, S.D., Ge, X., Drake,
S., Euong Ang, C., et al. (2017). Myt1l safeguards neuronal identity by actively repressing many
non-neuronal fates. *Nature* 544, 245–249. <https://doi.org/10.1038/nature21722>.
- 442
- 443 10. Weigel, B., Tegethoff, J.F., Grieder, S.D., Lim, B., Nagarajan, B., Liu, Y.-C., Truberg, J., Papageorgiou,
D., Adrian-Segarra, J.M., Schmidt, L.K., et al. (2023). MYT1L haploinsufficiency in human neurons
and mice causes autism-associated phenotypes that can be reversed by genetic and pharmacologic
intervention. *Mol Psychiatry* 28, 2122–2135. <https://doi.org/10.1038/s41380-023-01959-7>.
- 444
- 445 11. Schaefer, A., Sampath, S.C., Intrator, A., Min, A., Gertler, T.S., Surmeier, D.J., Tarakhovsky, A.,
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>.
- 446
- 447 12. Scandaglia, M., Lopez-Atalaya, J.P., Medrano-Fernandez, A., Lopez-Cascales, M.T., Del Blanco, B.,
Lipinski, M., Benito, E., Olivares, R., Iwase, S., Shi, Y., et al. (2017). Loss of Kdm5c Causes Spurious
Transcription and Prevents the Fine-Tuning of Activity-Regulated Enhancers in Neurons. *Cell Rep* 21,
47–59. <https://doi.org/10.1016/j.celrep.2017.09.014>.
- 448
- 449 13. Iwase, S., Lan, F., Bayliss, P., De La Torre-Ubieta, L., Huarte, M., Qi, H.H., Whetstine, J.R., Bonni, A.,
Roberts, T.M., and Shi, Y. (2007). The X-Linked Mental Retardation Gene SMCX/JARID1C Defines a
Family of Histone H3 Lysine 4 Demethylases. *Cell* 128, 1077–1088. <https://doi.org/10.1016/j.cell.2007.02.017>.
- 450
- 451 14. Claes, S., Devriendt, K., Van Goethem, G., Roelen, L., Meireleire, J., Raeymaekers, P., Cassiman,
J.J., and Fryns, J.P. (2000). Novel syndromic form of X-linked complicated spastic paraparesis. *Am J
Med Genet* 94, 1–4.
- 452
- 453 15. 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
Genet* 76, 227–236. <https://doi.org/10.1086/427563>.

- 454
- 455 16. Carmignac, V., Nambot, S., Lehalle, D., Callier, P., Moortgat, S., Benoit, V., Ghoumid, J., Delobel, B., Smol, T., Thuillier, C., et al. (2020). Further delineation of the female phenotype with KDM5C disease causing variants: 19 new individuals and review of the literature. *Clin Genet* **98**, 43–55. <https://doi.org/10.1111/cge.13755>.
- 456
- 457 17. 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>.
- 458
- 459 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>.
- 460
- 461 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>.
- 462
- 463 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>.
- 464
- 465 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>.
- 466
- 467 22. Li, B., Qing, T., Zhu, J., Wen, Z., Yu, Y., Fukumura, R., Zheng, Y., Gondo, Y., and Shi, L. (2017). A Comprehensive Mouse Transcriptomic BodyMap across 17 Tissues by RNA-seq. *Sci Rep* **7**, 4200. <https://doi.org/10.1038/s41598-017-04520-z>.
- 468
- 469 23. Vallianatos, C.N., Raines, B., Porter, R.S., Bonefas, K.M., Wu, M.C., Garay, P.M., Collette, K.M., Seo, Y.A., Dou, Y., Keegan, C.E., et al. (2020). Mutually suppressive roles of KMT2A and KDM5C in behaviour, neuronal structure, and histone H3K4 methylation. *Commun Biol* **3**, 278. <https://doi.org/10.1038/s42003-020-1001-6>.
- 470
- 471 24. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550. <https://doi.org/10.1186/s13059-014-0550-8>.
- 472

- 473 25. Crackower, M.A., Kolas, N.K., Noguchi, J., Sarao, R., Kikuchi, K., Kaneko, H., Kobayashi, E., Kawai, Y.,
474 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>.
- 475 26. Xiol, J., Cora, E., Koglgruber, R., Chuma, S., Subramanian, S., Hosokawa, M., Reuter, M., Yang, Z.,
476 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>.
- 477 27. Cheng, S., Altmeppen, G., So, C., Welp, L.M., Penir, S., Ruhwedel, T., Menelaou, K., Harasimov, K.,
478 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>.
- 479 28. Rouland, A., Masson, D., Lagrost, L., Vergès, B., Gautier, T., and Bouillet, B. (2022). Role of
480 apolipoprotein C1 in lipoprotein metabolism, atherosclerosis and diabetes: A systematic review.
Cardiovasc Diabetol *21*, 272. <https://doi.org/10.1186/s12933-022-01703-5>.
- 481 29. Mueller, J.L., Skaletsky, H., Brown, L.G., Zaghlul, S., Rock, S., Graves, T., Auger, K., Warren,
482 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>.
- 483 30. Handel, M.A., and Eppig, J.J. (1979). Sertoli Cell Differentiation in the Testes of Mice Genetically
484 Deficient in Germ Cells. *Biology of Reproduction* *20*, 1031–1038. <https://doi.org/10.1095/biolreprod20.5.1031>.
- 485 31. Green, C.D., Ma, Q., Manske, G.L., Shami, A.N., Zheng, X., Marini, S., Moritz, L., Sultan, C.,
486 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>.
- 487 32. Soh, Y.Q., Junker, J.P., Gill, M.E., Mueller, J.L., van Oudenaarden, A., and Page, D.C. (2015). A
488 Gene Regulatory Program for Meiotic Prophase in the Fetal Ovary. *PLoS Genet* *11*, e1005531.
<https://doi.org/10.1371/journal.pgen.1005531>.
- 489 33. Magnúsdóttir, E., and Surani, M.A. (2014). How to make a primordial germ cell. *Development* *141*,
490 245–252. <https://doi.org/10.1242/dev.098269>.
- 491 34. Günesdogan, U., Magnúsdóttir, E., and Surani, M.A. (2014). Primordial germ cell specification: A
492 context-dependent cellular differentiation event [corrected]. *Philos Trans R Soc Lond B Biol Sci* *369*.
<https://doi.org/10.1098/rstb.2013.0543>.
- 493 35. 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>.

- 495 36. 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.
496 EMBO Reports 16, 791–802. <https://doi.org/10.15252/embr.201540538>.
- 497 37. 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
498 activity. Nature 487, 57–63. <https://doi.org/10.1038/nature11244>.
- 499 38. Suzuki, A., Hirasaki, M., Hishida, T., Wu, J., Okamura, D., Ueda, A., Nishimoto, M., Nakachi, Y.,
500 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>.
- 501 39. 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
502 conserved function of KDM5C demethylase. Nat Commun 13, 2602. <https://doi.org/10.1038/s41467-022-30352-1>.
- 503 40. 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,
504 2474–2479. <https://doi.org/10.1073/pnas.0510813103>.
- 505 41. Lin, Y., Gill, M.E., Koubova, J., and Page, D.C. (2008). Germ Cell-Intrinsic and -Extrinsic Factors
506 Govern Meiotic Initiation in Mouse Embryos. Science 322, 1685–1687. <https://doi.org/10.1126/science.1166340>.
- 507 42. Endo, T., Mikedis, M.M., Nicholls, P.K., Page, D.C., and De Rooij, D.G. (2019). Retinoic Acid and Germ
508 Cell Development in the Ovary and Testis. Biomolecules 9, 775. <https://doi.org/10.3390/biom9120775>.
- 509 43. Gupta, N., Yakhou, L., Albert, J.R., Azogui, A., Ferry, L., Kirsh, O., Miura, F., Battault, S., Yamaguchi,
510 K., Laisné, M., et al. (2023). A genome-wide screen reveals new regulators of the 2-cell-like cell state.
Nat Struct Mol Biol. <https://doi.org/10.1038/s41594-023-01038-z>.
- 511 44. Liu, M., Zhu, Y., Xing, F., Liu, S., Xia, Y., Jiang, Q., and Qin, J. (2020). The polycomb group protein
512 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>.
- 513 45. Otani, J., Nankumo, T., Arita, K., Inamoto, S., Ariyoshi, M., and Shirakawa, M. (2009). Structural basis
514 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>.
- 515 46. Guo, X., Wang, L., Li, J., Ding, Z., Xiao, J., Yin, X., He, S., Shi, P., Dong, L., Li, G., et al. (2015).
516 Structural insight into autoinhibition and histone H3-induced activation of DNMT3A. Nature 517,
640–644. <https://doi.org/10.1038/nature13899>.

- 517 47. 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>.
- 518
- 519 48. Leduc, V., Jasmin-Bélanger, S., and Poirier, J. (2010). APOE and cholesterol homeostasis in Alzheimer's disease. *Trends in Molecular Medicine* *16*, 469–477. <https://doi.org/10.1016/j.molmed.2010.07.008>.
- 520
- 521 49. Dahlet, T., Truss, M., Frede, U., Al Adhami, H., Bardet, A.F., Dumas, M., Vallet, J., Chicher, J., Hammann, P., Kottnik, S., et al. (2021). E2F6 initiates stable epigenetic silencing of germline genes during embryonic development. *Nat Commun* *12*, 3582. <https://doi.org/10.1038/s41467-021-23596-w>.
- 522
- 523 50. 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>.
- 524
- 525 51. 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>.
- 526
- 527 52. 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>.
- 528
- 529 53. 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>.
- 530
- 531 54. 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>.
- 532
- 533 55. 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>.
- 534
- 535 56. 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>.
- 536

- 537 57. 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>.
- 538
- 539 58. Bonefas, K.M., and Iwase, S. (2021). Soma-to-germline transformation in chromatin-linked neurode-
540 developmental disorders? *FEBS J.* <https://doi.org/10.1111/febs.16196>.
- 541 59. 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.
542 *Orphanet J Rare Dis* *9*, 56. <https://doi.org/10.1186/1750-1172-9-56>.
- 543 60. Walton, E.L., Francastel, C., and Velasco, G. (2014). Dnmt3b Prefers Germ Line Genes and Cen-
tromeric Regions: Lessons from the ICF Syndrome and Cancer and Implications for Diseases. *Biology*
544 (*Basel*) *3*, 578–605. <https://doi.org/10.3390/biology3030578>.
- 545 61. 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
546 duplication syndrome. *Nat Genet* *44*, 206–211. <https://doi.org/10.1038/ng.1066>.
- 547 62. Samanta, M., and Kalantry, S. (2020). Generating primed pluripotent epiblast stem cells: A methodol-
548 ogy chapter. *Curr Top Dev Biol* *138*, 139–174. <https://doi.org/10.1016/bs.ctdb.2020.01.005>.
- 549 63. Stephens, M. (2016). False discovery rates: A new deal. Biostat, kxw041. <https://doi.org/10.1093/biostatistics/kxw041>.
- 550
- 551 64. 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>.
- 552
- 553 65. 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.
554 <https://doi.org/10.1016/j.molcel.2010.05.004>.
- 555 66. 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>.
- 556
- 557 67. 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
558 spermatogonial progenitors. *eLife* *9*, e56523. <https://doi.org/10.7554/eLife.56523>.

559 **Figures and Tables**

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

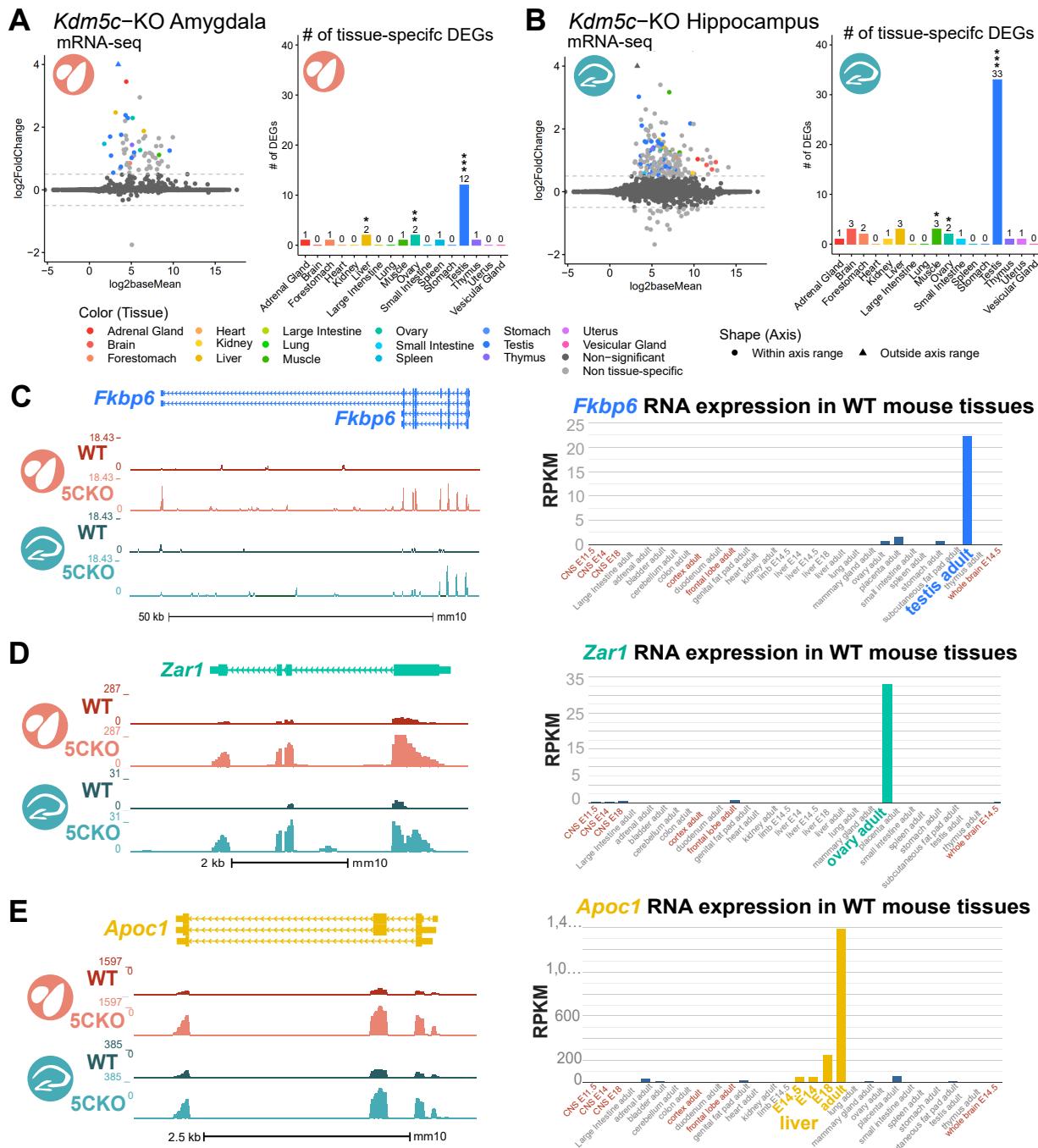


Figure 1: Tissue-enriched genes are misexpressed in the *Kdm5c*-KO brain. **A-B.** Expression of tissue-enriched genes (Li et al 2017) in the male *Kdm5c*-KO amygdala (A) and hippocampus (B). Left - MA plot of mRNA-sequencing. Right - Number of tissue-enriched differentially expressed genes (DEGs). *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

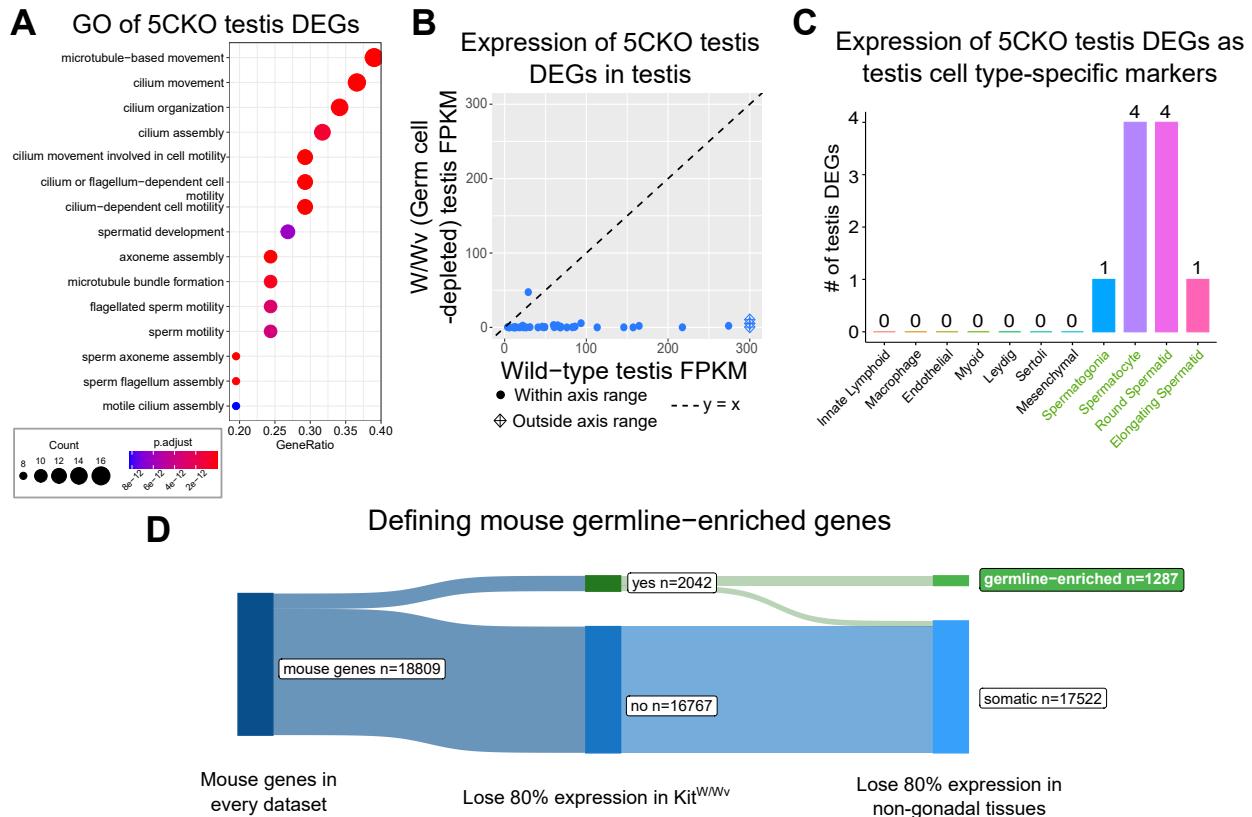


Figure 2: Aberrant transcription of germline genes in the *Kdm5c*-KO in the brain. **A.** enrichPlot gene ontology (GO) of *Kdm5c*-KO amygdala and hippocampus testis-enriched DEGs **B.** Expression of testis DEGs in wild-type (WT) testis versus germ cell-depleted (W/W_v) testis (Mueller et al 2013).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

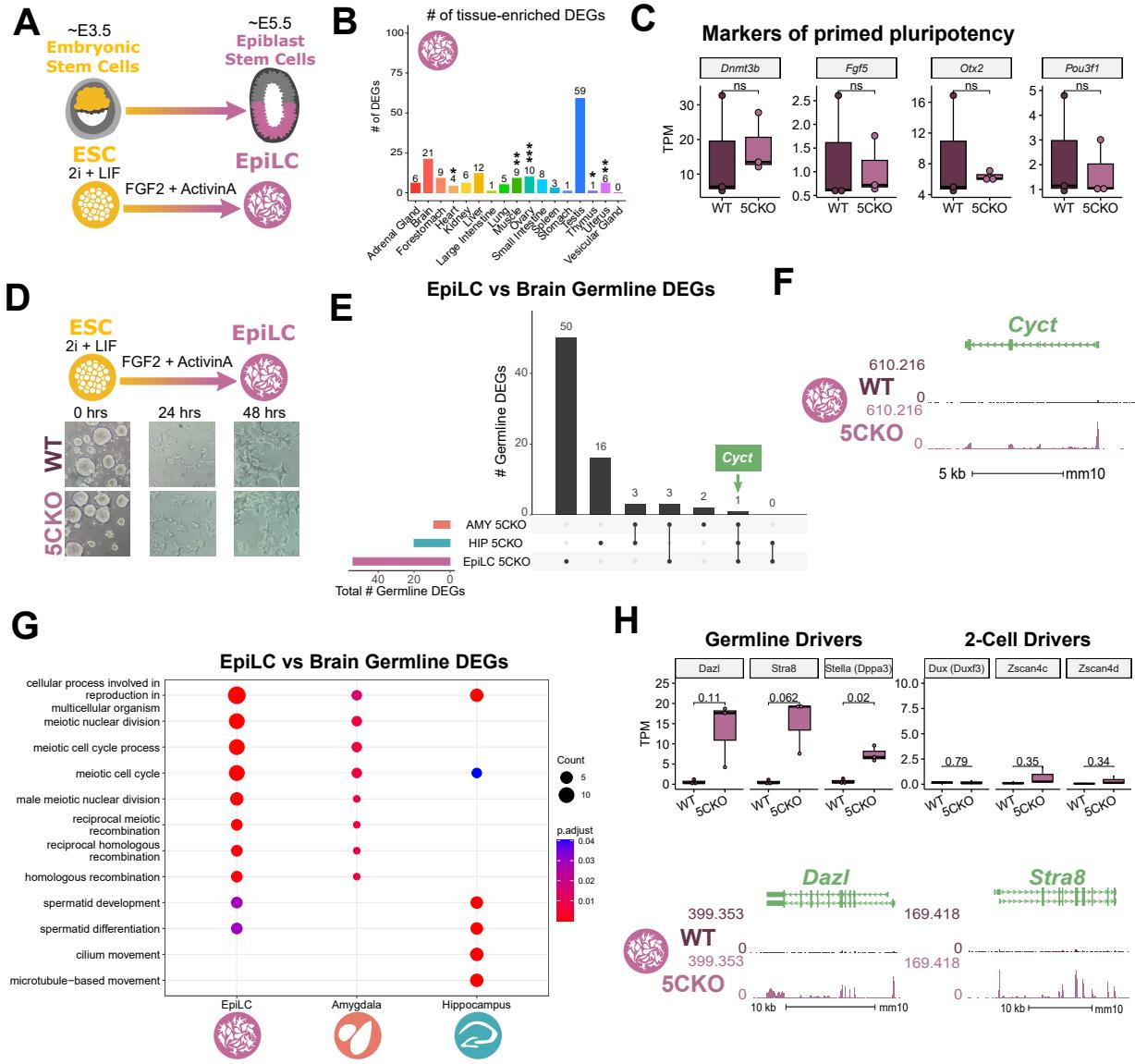


Figure 3: *Kdm5c*-KO epiblast-like cells express key drivers of germline identity

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

B. Number of tissue-enriched differentially expressed genes (DEGs) in *Kdm5c*-KO EpiLCs. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, Fisher's exact test.

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

D. No significant difference in primed pluripotency marker expression in wild-type versus *Kdm5c*-KO EpiLCs. Welch's t-test, expression in transcripts per million (TPM).

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

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

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

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

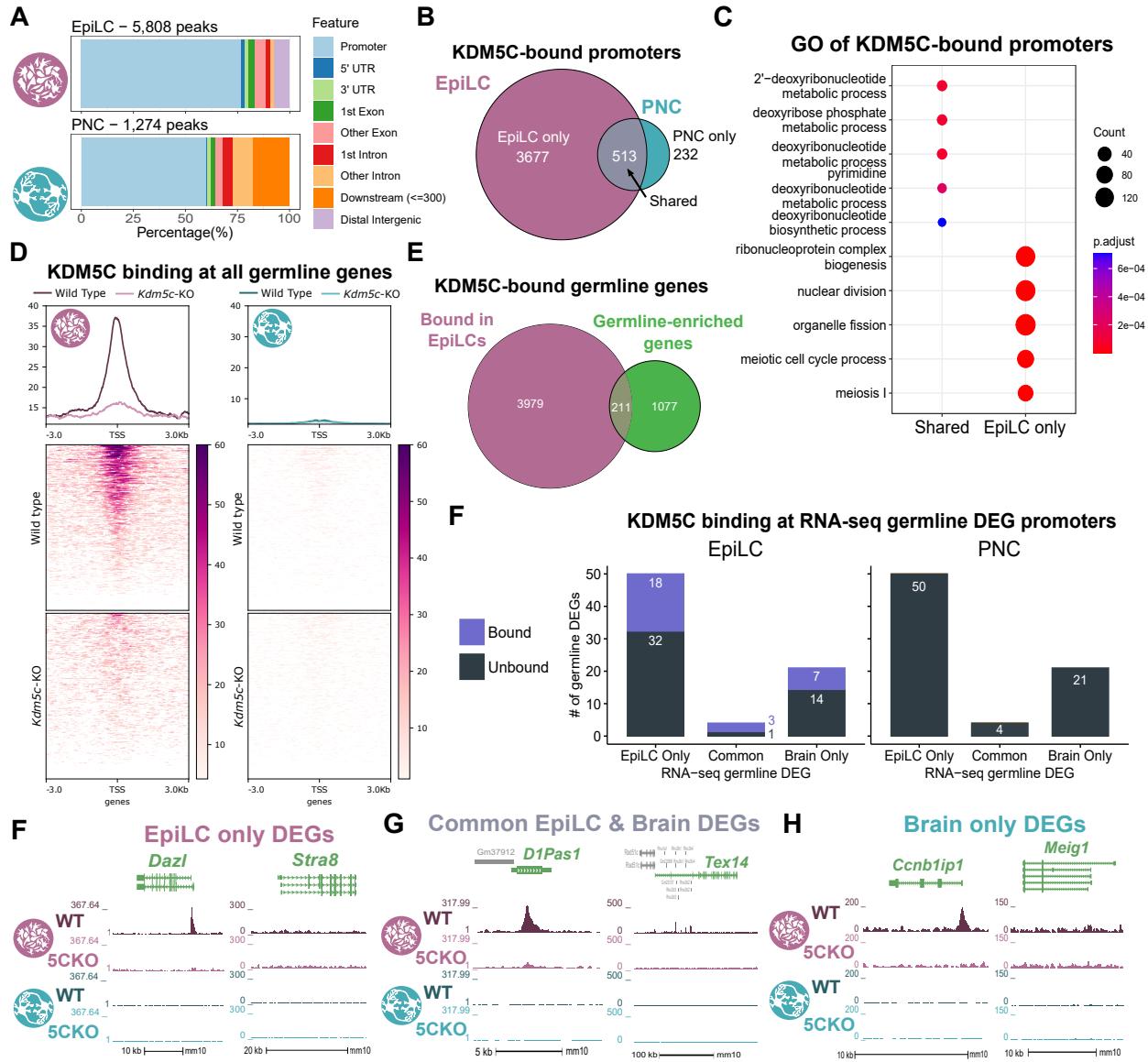
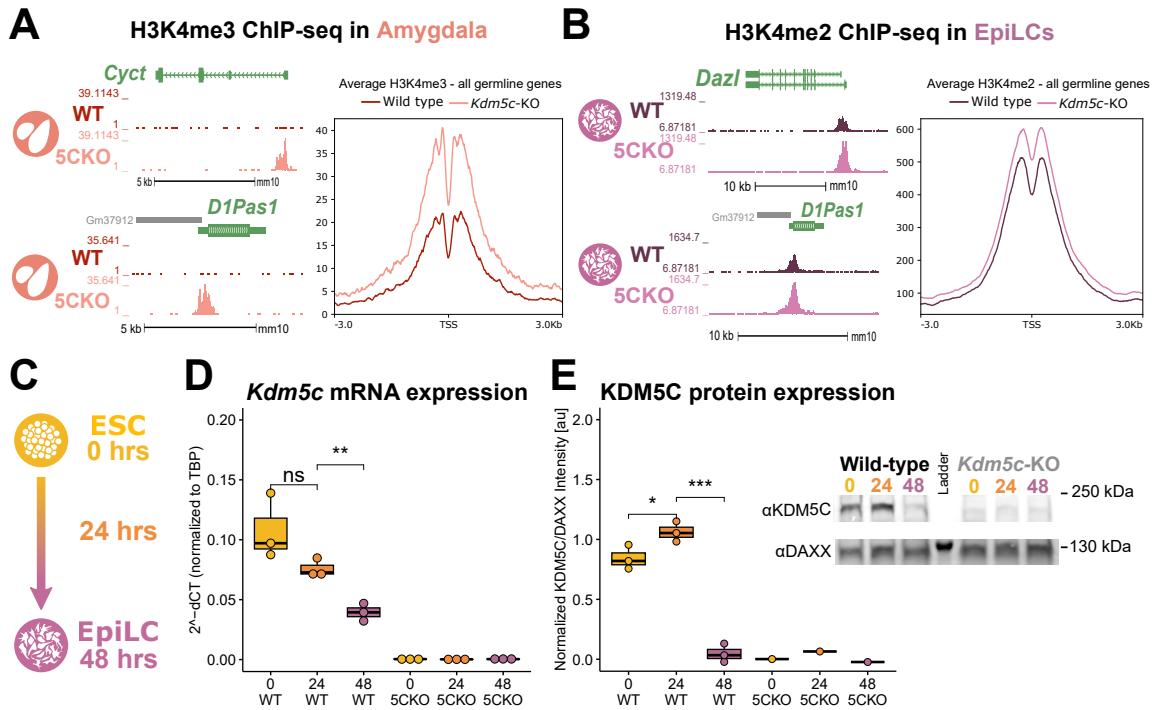


Figure 4: KDM5C binds to a subset of germline gene promoters during early embryogenesis **A.** ChIPseeker localization of KDM5C peaks at different genomic regions in EpiLCs (top) and hippocampal and cortex primary neuron cultures (PNCs, bottom). **B.** Overlap of genes with KDM5C bound to their promoters 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 4: 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 5: 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

564 Notes

565 Things to do

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

570 Dazl

571 We were particularly interested in the aberrant transcription of *Dazl*, since it is essential for germ cell
572 development and promotes the translation of germline mRNAs⁶⁶. A significant portion of germline transcripts
573 misexpressed in *Kdm5c*-KO EpiLCs are known binding targets of DAZL, including *Stra8*⁶⁷ ($p = 1.698e-07$,
574 Fisher's Exact Test). This suggests expression of DAZL protein could promote the translation of other
575 aberrant germline transcripts, influencing their ability to impact *Kdm5c*-KO cellular function. We thus tested
576 DAZL protein expression in *Kdm5c*-KO EpiLCs through immunocytochemistry (Figure 3I). We observed
577 about 25% of *Kdm5c*-KO EpiLCs expressed DAZL protein and it was localized to the cytoplasm ($p = 0.0015$,
578 Welch's t-test), consistent with the pattern of DAZL expression in spermatogonia⁶⁷. Altogether these results
579 suggest tissue-specific genes are misexpressed during *Kdm5c*-KO embryogenesis, including key drivers of
580 germline identity that can be translated into protein.

- 581 • We additionally found *Kdm5c*-KO EpiLCs ectopically express DAZL protein that is localized to the
582 cytoplasm, similar to its morphology in spermatogonia⁶⁷. **note: maybe just put in results.** Could
583 move around depending upon if I get pheno working.

584 Discussion notes

- 585 • For other paper:
 - 586 – for methods: Heatmaps of gene expression were generated using the base R functions scale and
587 hclust and visualized using the R package ComplexHeatmap (v2.12.1).
 - 588 – * 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>
- 590 • Talk about different motifs if we do see differences there and if it explains direct vs indirect dysregulation
 - 591 – Could also mention PRC1.6/KDM5C binds pcgf6 in ESCs
- 592 • Maybe talk about other regulators/if anything is known about them long term? Or save for 2nd paper.

- 593 – another germline repressor: <https://pubmed.ncbi.nlm.nih.gov/35345626/>
- 594 • end with something like: this indicates kdm5c not only modulates pluripotency and self-renewal in
595 ESCs, but also has a role in long-term silencing of germline genes
- 596 – then transition into the long term silencing mechanism paragraph
- 597 • KDM5C is dynamically regulated during the window of embryonic germline gene silencing, our EpiLC
598 ChIP-seq is likely catching the tail end of KDM5C's main involvement.
- 599 • Our results indicate KDM5C is a pivotal repressor of germline programs in somatic cells.
- 600 • Studies on other chromatin regulators have focused on key marker genes to identify germline gene
601 misexpression, such as *Dazl*.
- 602 • Using a germline-depleted mouse model, we curated a list of mouse germline-enriched genes to
603 globally assess germline gene dysregulation.
- 604 • Many of the genes dysregulated in the mature *Kdm5c*-KO brain were important for late stages of
605 spermatogenesis, such as those important for sperm flagellar structure. Contrastingly, *Kdm5c*-KO
606 EpiLCs aberrantly expressed key drivers of germ cell identity, including *Dazl* and *Stra8*.
- 607 • While a significant portion of KDM5C-bound promoters are germline genes, many germline genes
608 expressed during *Kdm5c*-KO embryogenesis are not directly bound by kdm5c.
- 609 • One notable example is *Stra8*, a transcription factor important for germ cell specific transcription and
610 meiotic initiation
- 611 • The including the demarcation between soma and germline fates.
- 612 the *Kdm5c*-KO brain also Previous work identified testis-enriched genes in *Kdm5c* genes
- 613 –
- 614 – However unlike the gonadal-biased DEGs,
- 615 • The demarcation of the germline versus soma is a key feature of multicellularity and sexually dimorphic
616 reproduction
- 617 • Anything known about tissue-biased gene expression in other H3K4me regulators?
- 618 • Our data suggests the germline developmental program is occurring ectopically as *Kdm5c*-KOs pro-
619 gresses through somatic tissue development
- 620 • tissue-biased gene expression:

- 621 • Altogether, these data indicate that while some germline genes are misexpressed due to direct loss of
622 KDM5C binding during emryogenesis, secondary downstream mechanisms can also promote their
623 aberrant transcription.
- 624 • Papers to read/reference:
625 – Reconstitution of the Mouse Germ Cell Specification Pathway in Culture by Pluripotent Stem Cells:
626 [https://www.cell.com/fulltext/S0092-8674\(11\)00771-9](https://www.cell.com/fulltext/S0092-8674(11)00771-9)
627 – 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/>

629 **Figure outline:**

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

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

642 **Figure 3: *Kdm5c*-KO epiblast-like cells express key drivers regulators of germline identity** * A) ESC
643 to EpiLC differentiation Left - Morphology is unchanged, * B) 5CKO EpiLCs express EpiLC differentiation
644 genes similar to WT lvs * C) Male EpiLCs express germline genes (example Cyct again) * Overlap between
645 brain and EpiLC germline genes - show they're mostly unique * GO of Brain and EpiLC germline genes
646 (meiotic enriched) * Bigwigs or TPM of master regulators * Show that while some are also 2-cell genes
647 (Dazl), 2-cell specific genes aren't dysregulated (Zscan4). Important point because published KDM5C dazl
648 paper is saying KDM5C is a 2-cell regulator, but as far as I can tell only genes shared between germline and
649 2-cell are dysregulated.

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

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

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

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

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

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

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

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

694 **Germline gene repression background:**

695 Interestingly, some of the ectopic testis transcripts identified in the *Kdm5c*-KO brain are typically ex-
696 pressed in germ cells¹². Unlike somatic cells, germ cells (e.g. sperm and eggs) undergo meiosis and pass
697 on their genetic material to the next generation. The germline and the soma are typically distinguished during
698 early embryogenesis, when germline genes are silenced in epiblast stem cells soon after implantation and
699 only reactivated in a subset to form the germline. Chromatin regulators play a key role in decommissioning
700 germline genes as the embryo transitions from naïve to primed pluripotency by placing repressive histone
701 H2A lysine 119 monoubiquitination (H2AK119ub1)¹⁸, histone 3 lysine 9 trimethylation (H3K9me3)^{18,19}, and
702 DNA CpG methylation^{19–21} at germline gene promoters. KDM5C may also be involved in this early decom-
703 missioning of germline genes, as re-expression of KDM5C in neurons fails to suppress their dysregulation¹².
704 In support of this, KDM5C was very recently shown to repress *Deleted in azoospermia like (Dazl)*, a key
705 regulator of germline development, in mouse embryonic stem cells (ESCs)^{43,55}. In support of this, two
706 independent screens in mouse embryonic stem cells (ESCs) recently identified KDM5C as a repressor of
707 *Deleted in azoospermia like (Dazl)*, a key regulator of germline development. However, KDM5C's role in
708 embryonic germline gene repression is currently unclear, given that Dazl is also expressed in ESCs and in
709 the 2-cell stage and germline gene misexpression has yet to be globally characterized during *Kdm5c*-KO
710 embryogenesis.