

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

3

<sup>4</sup> 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 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*. 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 links soma-vs-germline demarcation to a chromatin-linked neurodevelopmental disorder.

## 28 Introduction

29 A single genome holds the instructions to generate the myriad of cell types found within the adult  
30 organism. This is, in part, accomplished by chromatin regulators that can either promote or impede lineage-  
31 specific gene expression through DNA and histone modifications<sup>1,2</sup>. Many chromatin regulators were initially  
32 identified for their roles in shaping cellular and tissue identity<sup>3–5</sup>. Recent advancements in next generation  
33 sequencing unexpectedly revealed mutations in many chromatin regulators cause neurodevelopmental  
34 disorders (NDDs)<sup>6</sup>. Several studies have suggested this connection between chromatin regulators and  
35 neurodevelopment is due to their regulation of brain-specific genes, such as orchestrating transcriptional  
36 programs for synaptic maturation<sup>7</sup> and transitioning between neuronal and glial fates during neural precursor  
37 differentiation<sup>8</sup>. However, loss of some chromatin regulators can also result in the ectopic transcription  
38 of tissue-specific genes outside of their target environment, such as the misexpression of liver-specific  
39 genes within adult neurons<sup>9</sup>. Very few studies have investigated this cellular identity crisis in chromatin-  
40 linked NDDs<sup>9,10</sup> and it is currently unknown if ectopic gene expression contributes to neurodevelopmental  
41 impairments.

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

55 The testis contains both germ cells (cells that pass on their genetic material, e.g. sperm) and somatic  
56 cells (cells that perform all other bodily functions). Distinguishing between the germline and the soma is a key  
57 feature of multicellularity<sup>16</sup> and that typically occurs during early embryogenesis<sup>17</sup>. In mammals, chromatin  
58 regulators play a key role in decommissioning germline genes in somatic cells during the transition from naïve  
59 to primed pluripotency. Initially, repressive histone H2A lysine 119 monoubiquitination (H2AK119ub1)<sup>18</sup>,  
60 histone 3 lysine 9 trimethylation (H3K9me3)<sup>18,19</sup>, are placed at germline gene promoters in embryonic stem  
61 cells and are then decorated with DNA CpG methylation<sup>19–21</sup> in the post-implantation embryo. How KDM5C  
62 interacts with known germline gene silencers during this time point is currently unknown. Additionally, most

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

69 To illuminate KDM5C's role in tissue identity, here we characterized the aberrant transcription of tissue-  
70 enriched genes within the *Kdm5c*-KO brain and epiblast-like stem cells (EpiLCs), an *in vitro* model of  
71 the post-implantation embryo. We observed general dysregulation of tissue-enriched genes in both the  
72 *Kdm5c*-KO brain and EpiLCs, including testis, liver, muscle, and ovary-enriched genes. Both the *Kdm5c*-KO  
73 amygdala and hippocampus had significant enrichment of testis genes. We demonstrated testis-enriched  
74 genes genes are germline genes and not somatic testis genes by. We found *Kdm5c*-KO EpiLCs aberrantly  
75 expressed key drivers of germline identity and meiosis, including *Dazl* and *Stra8*, while the *Kdm5c*-KO  
76 brain primarily expressed germline genes important for late spermatogenesis. Additionally, KDM5C was  
77 bound to only a subset of germline genes expressed in *Kdm5c*-KO EpiLCs, indicating germline-enriched  
78 mRNAs can be aberrantly transcribed through indirect mechanisms. Finally, we found KDM5C promotes the  
79 long-term silencing of germline genes in somatic cells by aiding the placement of DNA CpG methylation in  
80 EpiLCs. Thus, we propose KDM5C plays a fundamental role in the development of tissue identity during  
81 early embryogenesis, including the establishment of the soma-germline boundary.

## 82 Results

### 83 **Tissue-enriched genes, including testis genes, are aberrantly expressed in the** 84 ***Kdm5c*-KO brain**

85 Previous RNA sequencing (RNA-seq) in the adult hippocampus revealed ectopic expression of some testis  
86 genes within the *Kdm5c* knockout (-KO) brain<sup>10</sup>. Since tissue-enriched genes have not been systematically  
87 characterized in the *Kdm5c*-KO brain, if the testis is the only tissue type misexpressed. Therefore, we  
88 systematically assessed the expression of genes enriched in 17 mouse tissues<sup>22</sup> in our published mRNA-seq  
89 datasets of the adult amygdala and hippocampus in mice with constitutive knockout of *Kdm5c*<sup>23</sup>.

90 We found a large proportion of genes that are significantly upregulated within the *Kdm5c*-KO brain  
91 (DESeq2<sup>24</sup>, log2 fold change > 0.5, q < 0.1) are enriched in non-brain tissues in wild-type mice (Amygdala:  
92 35%, Hippocampus: 24%) (Figure 1A-B). The majority of tissue-enriched differentially expressed genes  
93 (tissue-enriched DEGs) were testis genes (Figure 1A-C). Even though the testis has the largest total number  
94 of tissue-biased genes compared to any other tissue (2,496 genes), testis-biased DEGs were significantly  
95 enriched for both brain regions (Amygdala p = 1.83e-05, Odds Ratio = 5.13; Hippocampus p = 4.26e-11,

96 Odds Ratio = 4.45, Fisher's Exact Test). One example of a testis-enriched DEG is *FK506 binding protein 6*  
97 (*Fkbp6*), a known regulator of piRNA expression and meiosis in germ cells<sup>25,26</sup> (Figure 1C).

98 Interestingly, we also observed significant enrichment of ovary-biased DEGs in both the amygdala and  
99 hippocampus (Amygdala p = 0.00574, Odds Ratio = 18.7; Hippocampus p = 0.048, Odds Ratio = 5.88,  
100 Fisher's Exact) (Figure 1D). Ovary-enriched DEGs included *Zygotic arrest 1 (Zar1)*, sequesters mRNAs  
101 in oocytes for meiotic maturation and early zygote development<sup>27</sup> (Figure 1D). Given the *Kdm5c*-KO mice  
102 we analyzed are male, this demonstrates ectopic expression of tissue-enriched genes is independent  
103 of organismal sex. Although not consistent across brain regions, we also found significant enrichment  
104 of DEGs biased towards two non-gonadal tissues - the liver (Amygdala p = 0.0398, Odds Ratio = 6.58,  
105 Fisher's Exact Test) and the muscle (Hippocampus p = 0.0104, Odds Ratio = 6.95, Fisher's Exact Test). An  
106 example of a liver-biased DEG dysregulated in both the hippocampus and amygdala is *Apolipoprotein C-I*  
107 (*Apoc1*), a lipoprotein metabolism and transport gene<sup>28</sup> (Figure 1E). These aberrantly expressed mRNAs are  
108 polyadenylated and spliced into mature transcripts (Figure 1C-E). Of note, we did not observe enrichment  
109 of brain-enriched genes (Amygdala p = 1; Hippocampus p = 0.74, Fisher's Exact), despite the fact these  
110 are brain samples and the brain has the second highest total number of tissue-enriched genes (708 genes).  
111 Together, these results suggest misexpression of tissue-enriched genes within the brain is a major effect of  
112 KDM5C loss.

### 113 Germline genes are misexpressed in the *Kdm5c*-KO brain

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

120 To further validate if these testis DEGs are truly germline genes, we evaluated their expression in  
121 somatic versus germ cells within the testis. We first compared their expression in the testis without germ  
122 cells<sup>29</sup>, which was accomplished by heterozygous *W* and *Wv* mutations in the enzymatic domain of *c-Kit*  
123 (*Kit<sup>W/Wv</sup>*)<sup>30</sup>. Almost all *Kdm5c*-KO testis-enriched DEGs lost expression with germ cell depletion (Figure  
124 2B). We then assessed testis-enriched DEG expression in a published single cell RNA-seq dataset that  
125 identified cell type-specific markers within the testis<sup>31</sup>. We found some testis-enriched DEGs were classified  
126 as specific markers for different germ cell types (e.g. spermatogonia, spermatocytes, round spermatids, and  
127 elongating spermatids), yet none marked somatic cells (Figure 2C). Together, these data demonstrate that  
128 the *Kdm5c*-KO brain aberrantly expresses germline genes, but not somatic testicular genes, reflecting an  
129 erosion between somatic versus germline identity.

130 We then wanted to characterize germline gene misexpression with *Kdm5c* loss more deeply, but lacked a

131 comprehensive list of mouse germline-enriched genes. We thus generated a list of germline-enriched genes  
132 using RNA-seq datasets of *Kit<sup>WWv</sup>* mice that include males and females at embryonic day 12, 14, and 16<sup>32</sup>  
133 and adult male testes<sup>29</sup>. We defined genes as germline-enriched if their expression met the following criteria:  
134 1) their expression is greater than 1 FPKM in wild-type gonads 2) their expression in any non-gonadal tissue  
135 for adult wild type<sup>22</sup> does not exceed 20% of their maximum expression in the wild-type germline, and 3)  
136 their expression in the germ cell-depleted gonads, for any sex or time point, does not exceed 20% of their  
137 maximum expression in the wild-type germline. These criteria yielded 1,288 germline-enriched genes (Figure  
138 2D), which was hereafter used as a resource to globally assess germline gene misexpression with *Kdm5c*  
139 loss (Supplementary table 1).

140 ***Kdm5c*-KO epiblast-like cells aberrantly express master regulators of germline  
141 identity**

142 Germ cells are typically distinguished from somatic cells soon after the embryo implants into the uterine  
143 wall<sup>33,34</sup>. A subset of epiblast stem cells become the primordial germ cells (PGCs) while the remainder  
144 differentiate into the ectoderm, mesoderm, and endoderm to form the somatic tissues<sup>35</sup>. This developmental  
145 time point can be modeled *in vitro* through differentiation of embryonic stem cells (ESCs) into epiblast-like  
146 stem cells (EpiLCs) (Figure 3A, top)<sup>36,37</sup>. While some germline-enriched genes are also expressed in  
147 embryonic stem cells (ESCs) and in the 2-cell stage<sup>38–40</sup>, they are silenced as they differentiate into EpiLCs<sup>19</sup>.  
148 Therefore, we tested if KDM5C was necessary for silencing germline genes at this developmental stage by  
149 evaluating the impact of *Kdm5c* loss in EpiLCs.

150 We first identified *Kdm5c*-KO EpiLC DEGs in our previously published RNA-seq dataset<sup>41</sup> (DESeq2, log2  
151 fold change > 0.5, q < 0.1) and assessed tissue-enriched gene expression. Similar to the *Kdm5c*-KO brain,  
152 we observed general dysregulation of tissue-enriched genes, with the largest number of genes belonging to  
153 the brain and testis, although they were not significantly enriched (Figure 3B). Primed pluripotency genes,  
154 such as *Dnmt3b*, *Fgf5*, *Pou3f1*, and *Otx2*, were unaltered in *Kdm5c*-KO EpiLCs (Figure 3C). Furthermore,  
155 *Kdm5c*-KO cell morphology during ESC to EpiLC differentiation (Figure 3D) was normal, indicating KDM5C  
156 loss does not impair EpiLC formation.

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

165 Notably, DEGs unique to *Kdm5c*-KO EpiLCs included key drivers of germline identity, such as *Stimulated*  
166 *by retinoic acid 8 (Stra8)* and *Deleted in azoospermia like (Dazl)* (Figure 3H). These genes are typically  
167 expressed during embryonic germ cell development to commit PGCs to the germline fate, but are also  
168 expressed later in life to trigger meiotic gene expression programs<sup>42-44</sup>. Of note, some germline genes,  
169 including *Dazl*, are also expressed in the two-cell embryo<sup>39,45</sup>. However, we did not see misexpression of  
170 two-cell embryo-specific genes, like *Duxf3 (Dux)* ( $q = 0.337$ ) and *Zscan4d* ( $q = 0.381$ ), indicating *Kdm5c*-KO  
171 in EpiLCs do not revert back to a 2-cell state (Figure 3H).

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

173 Previous work suggests KDM5C represses germline genes during early development, as re-expression  
174 of KDM5C in knockout neuronal cultures fails to suppress two hippocampal germline DEGs<sup>10</sup>. KDM5C binds  
175 to several germline gene promoters in embryonic stem cells (ESCs) but is lost in neurons<sup>10</sup>. However, the  
176 lack of a comprehensive list of germline-enriched genes prohibited systematic characterization of KDM5C  
177 binding at germline gene promoters. Thus, it is currently unclear if KDM5C is enriched at germline gene  
178 promoters, what types of germline genes KDM5C regulates, and if its binding is maintained at any germline  
179 genes in neurons.

180 To address these questions, we analyzed KDM5C chromatin immunoprecipitation followed by DNA  
181 sequencing (ChIP-seq) datasets in EpiLCs<sup>41</sup> and primary forebrain neuron cultures (PNCs)<sup>15</sup>. EpiLCs had a  
182 higher total number of KDM5C peaks than PNCs (EpiLCs: 5,808, PNCs: 1,276, MACS2  $q < 0.1$  and fold  
183 enrichment  $> 1$ , removal of *Kdm5c*-KO false positive peaks) and KDM5C was primarily localized to gene  
184 promoters in both cell types (EpiLCs: 4,190, PNCs: 745 +/- 500bp from the TSS), although PNCs showed  
185 increased localization to non-promoter regions (Figure 4A).

186 The majority of promoters bound by KDM5C in PNCs were also bound in EpiLCs (513 shared promoters),  
187 however a large portion of gene promoters were only bound by KDM5C in EpiLCs (3677 EpiLC only  
188 promoters) (Figure 4B). We then performed gene ontology analysis to compare the known functions of genes  
189 with KDM5C bound to their promoter in EpiLCs and PNCs. While there were no ontologies significantly  
190 enriched for genes only bound by KDM5C in PNCs, gene ontologies for genes shared between PNCs and  
191 EpiLCs were enriched for functions involving nucleic acid turnover, such as deoxyribonucleotide metabolic  
192 process (GO:0009262,  $p.adjust = 8.28e-05$ ) (Figure 4C). Germline-specific ontologies were only enriched in  
193 promoters unique to EpiLCs, such as meiotic nuclear division (GO: 0007127  $p.adjust = 6.77e-16$ ) and meiotic  
194 cell cycle process (GO:1903046,  $p.adjust = 5.05e-16$ ) (Figure 3C). We then evaluated KDM5C binding  
195 around the transcription start site (TSS) of all germline-enriched genes. In EpiLCs, We observed modest  
196 KDM5C signal at about half of all germline genes (Figure 4D). Based on our ChIP-seq peak criteria, KDM5C  
197 was strongly bound to about 10% of germline gene promoters in EpiLCs (Figure 4E). In condordance with our  
198 gene ontology results, we did not observe KDM5C accumulation at any germline gene promoters in PNCs  
199 (Figure 4D). Together, these results demonstrate KDM5C is significantly enriched at a subset of germline

200 gene promoters in EpiLCs, including meiotic genes, but does not directly regulate germline genes in neurons.  
201 Many germline-specific genes, including meiotic genes, are suppressed by the transcription factor  
202 heterodimers E2F6/DP1 and MGA/MAX, which respectively bind E2F and E-box motifs<sup>20,46–49</sup>. Thus, we  
203 identified transcription factor motifs enriched at KDM5C-bound or unbound germline gene promoters using  
204 HOMER<sup>50</sup> (TSS +/- 500 bp, q-value < 0.1). MAX and E2F6 binding sites were significantly enriched at  
205 germline genes bound by KDM5C in EpiLCs, but not at germline genes unbound by KDM5C (MAX q-value: ,  
206 E2F6 q-value: , E2F q-value: ) (Figure 4). One third of KDM5C-bound promoters contained the consensus  
207 sequence for either E2F6 (E2F, 5'-TCCCGC-3'), MGA (E-box, 5'-CACGTG-3'), or both, but only 17% of  
208 KDM5C-unbound genes contained these motifs (Figure 4). KDM5C-unbound germline genes were instead  
209 enriched for multiple RFX transcription factor binding sites. RFX transcription factors bind X-box motifs<sup>51</sup>  
210 to promote ciliogenesis<sup>52,53</sup>. Enriched RFX transcription factors included RFX2, a central regulator of post-  
211 meiotic spermatogenesis<sup>54,55</sup>. Interestingly, RFX2 mRNA is upregulated in *Kdm5c*-KO EpiLCs, but is also not  
212 a direct target of KDM5C (Figure 4). Thus, distinct transcription factor programs regulate KDM5C-bound and  
213 unbound germline genes.

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

## 226 **KDM5C promotes *de novo* DNA methylation at germline genes**

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

228 In the early embryo, germline gene promoters are initially silenced by repressive histone modifications,  
229 including histone 2A lysine 119 monoubiquitination (H2AK119ub1) and histone 3 lysine 9 trimethylation  
230 (H3K9me3), and are then silenced long-term via DNA CpG methylation (CpGme)<sup>18,19,56</sup>. Our results above  
231 suggest KDM5C also acts during this time period, however how KDM5C interacts with other germline gene  
232 silencing mechanism is currently unclear. KDM5C is generally thought to suppress transcription through  
233 erasure of histone 3 lysine 4 di- and trimethylation (H3K4me2/3)<sup>11</sup>. However, KDM5C's catalytic activity was

234 recently shown to not be required for suppressing *Dazl* in undifferentiated ESCs<sup>45</sup>. Since H3K4me3 impedes  
235 *de novo* CpGme placement<sup>57,58</sup>, KDM5C's catalytic activity may instead be required for CpGme-mediated,  
236 long-term silencing of germline genes. In support of this, CpGme is markedly reduced at two germline gene  
237 promoters in the *Kdm5c*-KO adult hippocampus<sup>10</sup>. Based on these observations, we hypothesized KDM5C  
238 erases H3K4me3 to promote the initial placement of CpGme at germline gene promoters in EpiLCs.

239 To test this hypothesis, we first characterized KDM5C's substrates, histone 3 lysine 4 di- and trimethylation  
240 (H3K4me2/3) in our previously published ChIP-seq datasets of the wild type and *Kdm5c*-KO amygdala<sup>23</sup>  
241 and EpiLCs<sup>41</sup>. In congruence with previous work in the *Kdm5c*-KO hippocampus<sup>10</sup>, we observed aberrant  
242 accumulation of H3K4me3 around the transcription start site (TSS) of germline genes in the *Kdm5c*-KO  
243 amygdala (Figure 5A). We additionally found a marked increase in H3K4me2 around the TSS of germline  
244 genes in *Kdm5c*-KO EpiLCs (Figure 5B).

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

- 249 • Germline genes are known to accumulate CpGme at (CGIs).
- 250     – What percentage of germline genes have CGIs
- 251 • Do differential methylation analysis for WT ESCs to WT EpiLCs
- 252     – What percentage of germline genes significantly gain methylation (at CGI or at promoter)
- 253     – Out of the ones that gain methylation, which are significantly reduced

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

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

264 **Discussion**

265 In the above study, we demonstrate KDM5C's pivotal role in the development of tissue identity, resulting in  
266 substantial misexpression of tissue-enriched genes within the adult *Kdm5c*-KO hippocampus and amygdala.  
267 In addition to testis genes identified previously<sup>10</sup>, we found significant enrichment of muscle, liver, and even  
268 ovary-biased genes aberrantly expressed within the male *Kdm5c*-KO brain. Deeper characterization of  
269 testis-enriched DEGs revealed they were not somatic testis genes, but germline genes, thus demonstrating  
270 KDM5C's crucial role in establishing the soma-germline boundary. Given the majority of tissue-enriched  
271 DEGs are testis and ovary genes with no known brain functions, it is currently unknown if they impair  
272 *Kdm5c*-KO brain function and contribute to MRXSCJ-related phenotypes like intellectual disability and  
273 aberrant aggression. However, select liver and muscle-biased DEGs do have known roles within the brain,  
274 such as the liver-enriched lipid metabolism gene *Apolipoprotein C-I (Apoc1)*<sup>28</sup> that is highly expressed in the  
275 *Kdm5c*-KO amygdala and hippocampus. Intriguingly, *Apoc1* overexpression in the mouse brain can impair  
276 learning and memory<sup>59</sup> and is implicated in Alzheimer's disease in humans<sup>60</sup>.

277 We found a substantial portion of genes upregulated in *Kdm5c*-KO brain are germline genes with no  
278 known fucntions within the brain. Distinguishing the germline from the soma is a key feature of multicellularity  
279 and sexual reproduction<sup>16</sup>. Previous work characterizing chromatin regualtors that silence germ cell-specific  
280 transcription has predominatly focused on their repression of key marker genes in embryonic stem cells  
281 (ESCs), such as *Dazl* and *Ddx4*<sup>18,19,49</sup>. To characterize KDM5C's role in germline gene repression at a  
282 genome-wide level throughout life, we curated a list of mouse germline-enriched genes using publically  
283 available germ cell-depleted RNA-seq datasets from *Kit<sup>W/Wv</sup>* mice<sup>29,32</sup>. This resource enabled us to identify 1)  
284 the extent of germline gene dysregulation in *Kdm5c*-KO cells, 2) the types of germline genes misexpressed  
285 at different developmental time points, and 3) which groups of germline genes are directly and indirectly  
286 regulated by KDM5C. This list of germline-enriched genes serves as a resource for future studies to  
287 systematically assess soma-germline dysregulation.

288 RNA-seq of epiblast-like cells (EpiLCs) revealed dysregulation of *Kdm5c*-KO tissue identity begins during  
289 early emrbyogenesis. *In vivo*, germline genes are typically decommissioned in epiblast stem cells and remain  
290 silenced as the epiblast differentiates into somatic tissues<sup>35</sup>. However, a small subset of epiblast stem cells  
291 will receive signals to reactivate germline gene expression to become the primordial germ cells (PGCs) that  
292 will ultimately form the mature germline<sup>33,34</sup>. This process can be mimicked *in vitro* by differentiating EpiLCs  
293 into primordial germ cell-like cells (PGCLCs)<sup>36</sup>. Therefore, misexpression of germline genes in EpiLCs might  
294 suggest *Kdm5c*-KO ESCs are progressing beyond EpiLC differentiation and instead becoming PGCLCs. Yet,  
295 *Kdm5c*-KO EpiLCs properly express primed pluripotency genes, including *Otx2* - an epiblast stem cell marker  
296 that is known to repress differentiation into PGCs/PGCLCs<sup>61</sup>. Furthermore, we observed no difference in  
297 cellular morphology during *Kdm5c*-KO ESC to EpiLC differentiation. Proper EpiLC differentiation, together  
298 with *Kdm5c*-KO mice being viable, suggests germline gene expression is occuring ectopically in conjunction

299 with typical developmental programs, rather than a complete shift to towards germline identity.

300 • Motif analysis

301 – only 1/3 had max/e2f - Unlikely to be the major recruitment mechanism for KDM5C

302 – KDM5C does not contain motif-specific binding

303 – controls cilia in multiple cell types, initially identified in neurons. Shared axis that could contribute  
304 to NDDs

305 \* RFX Cilia in NDDs - <https://pubmed.ncbi.nlm.nih.gov/33658631/>

306 – Enriched RFX transcription factors included the spermiogenesis regulator RFX2, whose mRNA is  
307 also significantly upregulated in *Kdm5c*-KO EpiLCs (Figure 4).

308 – <https://www.nature.com/articles/srep20435>

309 – <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4498915/>

310 – <https://pubmed.ncbi.nlm.nih.gov/32725242/>

311 We then globally characterized KDM5C binding at germline-enriched gene promoters through KDM5C

312 ChIP-seq in EpiLCs and primary hippocampus and cortex neuron cultures (PNCs).

313 Our analysis of KDM5C ChIP-seq data provided novel insights into direct and indirect mechanisms by

314 which germline genes can be misexpressed in *Kdm5c*-KO cells.

315 While we observed high enrichment of KDM5C peaks at germline genes in EpiLCs, KDM5C was not

316 bound to germline gene promoters in PNCs. This suggests dysregulation of germline genes in the mature

317 *Kdm5c*-KO brain is due to loss of repression during embryogenesis, which is consistent with previous

318 work that found introducing human KDM5C into *Kdm5c*-KO PNCs does not repress two aberrant germline

319 transcripts<sup>10</sup>. Although enriched at germline gene promoters in EpiLCs, KDM5C was only bound to about a

320 third of the germline genes aberrantly transcribed in *Kdm5c*-KO EpiLCs. One notable EpiLC DEG not bound

321 by KDM5C is *Stra8*, a transcription factor activated by retinoic acid signaling in germ cells to promote meiotic

322 initiation<sup>62,63</sup>. However, retinoic acid can only activate *Stra8* expression when DAZL is present. Unlike *Stra8*,

323 *Dazl* is a direct target of KDM5C in EpiLCs and is both transcribed and translated in *Kdm5c*-KO EpiLCs.

324 Expression of indirect DEGs like *Stra8* indicates germline genes can be aberrantly expressed in *Kdm5c*-KO

325 cells through downstream activation by other ectopic germline programs. These ectopic germline programs

326 appear to loosely mimic the trajectory of typical germline development, as germline genes important for early

327 germ cell development and meiosis are expressed in *Kdm5c*-KO EpiLC while late spermatogenesis genes

328 are expressed in the mature *Kdm5c*-KO brain. Altogether, this indicates once a subset of germline genes

329 are activated due to loss of direct KDM5C repression in the early embryo, these ectopic germline programs

330 can continue to progress in the background of *Kdm5c*-KO somatic development.

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

332 enriched genes are also expressed in naïve ESCs and at the 2-cell stage due to their roles in pluripotency and

333 self-renewal. For example, *Dazl* is primarily known for committing PGCs to the germline fate and regulating

334 the translation of germline-specific RNAs, but is also expressed at the 2-cell stage<sup>64</sup>, in naïve ESCs<sup>38</sup>, and in  
335 the inner cell mass<sup>38</sup>. Very recently, two screens of *Dazl*-repressors identified KDM5C as a direct regulator in  
336 ESCs<sup>45,65</sup>. Interestingly, one study found *Kdm5c*-KO ESCs also aberrantly expressed 2-cell-specific genes,  
337 indicating KDM5C negatively regulates totipotency<sup>45</sup>. However, out of the four regulators characterized,  
338 KDM5C was the only factor whose repression of *Dazl* was independent of the 2-cell specific transcription  
339 factor *Dux*<sup>45</sup>. Although expressed at the 2-cell stage and in ESCs, *Dazl* is typically silenced when ESCs  
340 differentiate into EpiLCs<sup>19</sup>. We found KDM5C also represses *Dazl* in EpiLCs, but *Kdm5c*-KO EpiLCs did  
341 not express 2-cell specific genes like *Dux* and *Zscan4c*. Together, this suggests the 2-cell-like state in  
342 *Kdm5c*-KO ESCs is an indirect consequence of KDM5C's primary role in germline gene repression.

343 *Dazl* and other germline-enriched genes are initially decommissioned in ESCs by repressive histone  
344 modifications and then *de novo* DNA CpG methylation (CpGme) in EpiLCs<sup>19–21,56,66,67</sup>. Unlike the previously  
345 characterized germline gene suppressors, KDM5C removes an active histone mark, histone 3 lysine 4 di-  
346 and trimethylation (H3K4me2/3)<sup>11</sup>. In this study, we observed a global increase in H3K4me2/3 in *Kdm5c*-KO  
347 EpiLCs and amygdala around the transcription start site (TSS) of germline-enriched genes. However,  
348 KDM5C's catalytic activity may not be required for germline gene silencing, as it was recently found to be  
349 dispensible for repressing *Dazl* in ESCs<sup>45</sup>. Although not necessary in ESCs, KDM5C's catalytic activity be  
350 required for CpGme in EpiLCs, as H3K4me3 impedes *de novo* CpGme placement<sup>57,58</sup>. This is supported  
351 by previous work in the *Kdm5c*-KO adult hippocampus, which found CpGme is significantly eroded at at  
352 least two germline promoters<sup>10</sup>. To elucidate the mechanism behind KDM5C-mediated silencing of germline  
353 genes, we characterized KDM5C in ESCs and EpiLCs and found its expression is dynamically regulated  
354 during differentiation. Furthermore, KDM5C acts during the transition from naïve to primed pluripotency to  
355 promote the initial placement of CpGme at CpG islands for the long-term silencing of germline genes.

356 The above work provides the mechanistic foundation for KDM5C's role in germline gene repression  
357 and offers novel insight into how dysregulation in *Kdm5c*-KO tissue identity changes over the course of  
358 development. Although their impact upon neurodevelopment is currently unclear, ectopic germline transcripts  
359 are also found in models of other related neurodevelopmental disorders<sup>68</sup>, including Immunodeficiency,  
360 Centromere instability, Facial anomalies syndrome 1 (ICF1 OMIM: #242860)<sup>69,70</sup>, Kleefstra syndrome  
361 1 (OMIM: #610253)<sup>9</sup>, and MeCP2 duplication syndrome (MDS, OMIM: #300260)<sup>71</sup>. Like KDM5C, the  
362 chromatin regulators underlying these conditions, DNA methyltransferase 3b (DNMT3B), H3K9me1/2  
363 methyltransferases G9A/GLP, methyl-CpG -binding protein 2 (MECP2), primarily silence gene expression.  
364 Thus, KDM5C is among a growing cohort of chromatin-linked neurodevelopmental disorders that may have a  
365 similar underlying cause of germline versus soma dysregulation. However, further research is required to  
366 determine the ultimate impact of these germline genes and the extent to which this phenomenon occurs in  
367 humans.

368 • Last paragraph

369 – Thus, failure to fine-tune the expression of tissue-enriched, dosage-sensitive genes like *Apoc1*

370 could be one route by which loss of brain tissue identity contributes to *Kdm5c*-KO impairments.

## 371 Materials and Methods

### 372 Classifying tissue-enriched and germline-enriched genes

373 Tissue-enriched differentially expressed genes (DEGs) were determined by their classification in a previ-  
374 ously published dataset from 17 male and female mouse tissues<sup>22</sup>. This study defined tissue expression as  
375 greater than 1 Fragments Per Kilobase of transcript per Million mapped read (FPKM) and tissue enrichment  
376 as at least 4-fold higher expression than any other tissue.

377 We curated a list of germline-enriched genes using an RNA-seq dataset from wild-type and germline-  
378 depleted (*Kit*<sup>W/Wv</sup>) male and female mouse embryos from embryonic day 12, 14, and 16<sup>32</sup>, as well as adult  
379 male testes<sup>29</sup>. Germline-enriched genes met the following criteria: 1) their expression is greater than 1  
380 FPKM in wild-type germline 2) their expression in any wild-type somatic tissues<sup>22</sup> does not exceed 20%  
381 of maximum expression in wild-type germline, and 3) their expression in the germ cell-depleted (*Kit*<sup>W/Wv</sup>)  
382 germline, for any sex or time point, does not exceed 20% of maximum expression in wild-type germline.

### 383 Cell culture

384 We utilized our previously established cultures of male wild-type and *Kdm5c* knockout (-KO) embryonic  
385 stem cells<sup>41</sup>. Sex was confirmed by genotyping *Uba1/Uba1y* on the X and Y chromosomes with the following  
386 primers: 5'-TGGATGGTGTGCCAATG-3', 5'-CACCTGCACGTTGCCCTT-3'. Deletion of *Kdm5c* was  
387 confirmed through the primers 5'-ATGCCCATATTAAAGAGTCCTG-3', 5'-TCTGCCTTGATGGGACTGTT-3',  
388 and 5'-GGTTCTCAACACTCACATAGTG-3'.

389 Embryonic stem cells (ESCs) and epiblast-like cells were cultured using previously established  
390 methods<sup>37</sup>. Briefly, ESCs were initially cultured in primed ESC (pESC) media consisting of KnockOut  
391 DMEM (Gibco#10829–018), fetal bovine serum (Gibco#A5209501), KnockOut serum replacement  
392 (Invitrogen#10828–028), Glutamax (Gibco#35050-061), Anti-Anti (Gibco#15240-062), MEM Non-essential  
393 amino acids (Gibco#11140-050), and beta-mercaptoethanol (Sigma#M7522). They were then transitioned  
394 into ground-state, “naïve” ESCs (nESCs) by culturing for four passages in N2B27 media containing  
395 DMEM/F12 (Gibco#11330–032), Neurobasal media (Gibco#21103–049), Gluamax, Anti-Anti, N2 sup-  
396 plement (Invitrogen#17502048), and B27 supplement without vitamin A (Invitrogen#12587-010), and  
397 beta-mercaptoethanol. Both pESC and nESC media were supplemented with 3  $\mu$ M GSK3 inhibitor  
398 CHIR99021 (Sigma #SML1046-5MG), 1  $\mu$ M MEK inhibitor PD0325901 (Sigma #PZ0162-5MG), and 1,000  
399 units/mL leukemia inhibitory factor (LIF, Millipore#ESG1107).

400 nESCs were differentiated into epiblast-like cells (EpiLCs, 48 hours) and extended EpiLCs (exEpiLCs,  
401 96 hours) by culturing in N2B27 media containing DMEM/F12, Neurobasal media, Gluamax, Anti-Anti,

402 N2 supplement, B27 supplement (Invitrogen#17504044), beta-mercaptoethanol, fibroblast growth factor 2  
403 (FGF2, R&D Biotechne 233-FB), and activin A (R&D Biotechne 338AC050CF), as previously described<sup>37</sup>.

#### 404 **Immunocytochemistry (ICC)**

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

#### 414 **RNA sequencing (RNA-seq)**

415 After ensuring read quality via FastQC (v0.11.8), reads were then mapped to the mm10 *Mus musculus*  
416 genome (Gencode) using STAR (v2.5.3a), during which we removed duplicates and kept only uniquely  
417 mapped reads. Count files were generated by FeatureCounts (Subread v1.5.0), and BAM files were  
418 converted to bigwigs using deeptools (v3.1.3) and visualized by the UCSC genome browser. RStudio (v3.6.0)  
419 was then used to analyze counts files by DESeq2 (v1.26.0)<sup>24</sup> to identify differentially expressed genes  
420 (DEGs) with a q-value (p-adjusted via FDR/Benjamini–Hochberg correction) less than 0.1 and a log2 fold  
421 change greater than 0.5. For all DESeq2 analyses, log2 fold changes were calculated with IfcShrink using  
422 the ashR package<sup>72</sup>. MA-plots were generated by ggpqr (v0.6.0), and Eulerr diagrams were generated by  
423 eulerr (v6.1.1). Boxplots and scatterplots were generated by ggpqr (v0.6.0) and ggplot2 (v3.3.2). The Upset  
424 plot was generated via the package UpSetR (v1.4.0)<sup>73</sup>. Gene ontology (GO) analyses were performed by  
425 the R package enrichPlot (v1.16.2) using the biological processes setting and compareCluster.

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

427 ChIP-seq reads were aligned to mm10 using Bowtie1 (v1.1.2) allowing up to two mismatches. Only  
428 uniquely mapped reads were used for analysis. Peaks were called using MACS2 software (v2.2.9.1) using  
429 input BAM files for normalization, with filters for a q-value < 0.1 and a fold enrichment > 1. We removed  
430 “black-listed” genomic regions that often give aberrant signals. Common peak sets were obtained in R via  
431 DiffBind (v3.6.5). In the case of KDM5C ChIP-seq, *Kdm5c*-KO peaks were then subtracted from wild-type  
432 samples using bedtools (v2.25.0). Peak proximity to genome annotations was determined by ChIPSeeker  
433 (v1.32.1). Gene ontology (GO) analyses were performed by the R package enrichPlot (v1.16.2) using the

434 biological processes setting and compareCluster. Enriched motifs were identified using HOMER<sup>50</sup>. Average  
435 binding across the genome was visualized using deeptools (v3.1.3). Bigwigs were visualized using the  
436 UCSC genome browser.

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

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

442 **Data availability**

443 **Published datasets**

444 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  
445 adult amygdala and hippocampus<sup>23</sup> (available at GEO: GSE127722) and male wild-type and *Kdm5c*-KO  
446 EpiLCs<sup>41</sup> (available at GEO: GSE96797).

447 Previously published ChIP-seq experiments included KDM5C in wild-type and *Kdm5c*-KO EpiLCs<sup>41</sup> (avail-  
448 able at GEO: GSE96797) and mouse primary neuron cultures (PNCs) from the cortex and hippocampus<sup>15</sup>  
449 (available at GEO: GSE61036). ChIP-seq of histone 3 lysine 4 dimethylation in male wild-type and *Kdm5c*-KO  
450 EpiLCs<sup>41</sup> is also available at GEO: GSE96797. ChIP-seq of histone 3 lysine 4 trimethylation in wild-type and  
451 *Kdm5c*-KO male amygdala<sup>23</sup> are available at GEO: GSE127817.

453 **Data analysis**

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

456 **Acknowledgements**

- 457 • Jacob Mueller for providing insight in germline gene regulation.  
458 • Sundeep Kalantry for providing reagents and expertise in culturing mouse embryonic stem cells and  
459 epiblast-like cells  
460 • Funding sources

461 **References**

- 462 1. Strahl, B.D., and Allis, C.D. (2000). The language of covalent histone modifications. *Nature* **403**,  
463 41–45. <https://doi.org/10.1038/47412>.
- 464 2. Jenuwein, T., and Allis, C.D. (2001). Translating the Histone Code. *Science* **293**, 1074–1080.  
465 <https://doi.org/10.1126/science.1063127>.
- 466 3. Lewis, E.B. (1978). A gene complex controlling segmentation in *Drosophila*. *Nature* **276**, 565–570.  
467 <https://doi.org/10.1038/276565a0>.
- 468 4. Kennison, J.A., and Tamkun, J.W. (1988). Dosage-dependent modifiers of polycomb and antennapedia  
469 mutations in *Drosophila*. *Proc. Natl. Acad. Sci. U.S.A.* **85**, 8136–8140. <https://doi.org/10.1073/pnas.8>  
5.21.8136.
- 470 5. Kassis, J.A., Kennison, J.A., and Tamkun, J.W. (2017). Polycomb and Trithorax Group Genes in  
471 *Drosophila*. *Genetics* **206**, 1699–1725. <https://doi.org/10.1534/genetics.115.185116>.
- 472 6. Gabriele, M., Lopez Tobon, A., D'Agostino, G., and Testa, G. (2018). The chromatin basis of  
473 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>.
- 474 7. Zhou, Z., Hong, E.J., Cohen, S., Zhao, W.-N., Ho, H.-Y.H., Schmidt, L., Chen, W.G., Lin, Y., Savner,  
475 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>.
- 476 8. Hirabayashi, Y., Suzuki, N., Tsuboi, M., Endo, T.A., Toyoda, T., Shinga, J., Koseki, H., Vidal, M., and  
477 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>.
- 478 9. Schaefer, A., Sampath, S.C., Intrator, A., Min, A., Gertler, T.S., Surmeier, D.J., Tarakhovsky, A.,  
479 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>.
- 480 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  
481 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>.
- 482 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  
Family of Histone H3 Lysine 4 Demethylases. *Cell* **128**, 1077–1088. [7.02.017](https://doi.org/10.1016/j.cell.200)  
483 [7.02.017](https://doi.org/10.1016/j.cell.200).

- 484 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 paraparesis. *Am J Med Genet* *94*, 1–4.
- 485 13. Jensen, L.R., Amende, M., Gurok, U., Moser, B., Gimmel, V., Tschach, A., Janecke, A.R., Tariverdian, G., Chelly, J., Fryns, J.P., et al. (2005). Mutations in the JARID1C gene, which is involved in transcriptional regulation and chromatin remodeling, cause X-linked mental retardation. *Am J Hum Genet* *76*, 227–236. <https://doi.org/10.1086/427563>.
- 486 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. <https://doi.org/10.1111/cge.13755>.
- 487 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>.
- 488 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>.
- 489 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>.
- 490 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>.
- 491 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>.
- 492 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>.
- 493 21. Hackett, J.A., Reddington, J.P., Nestor, C.E., Duncan, D.S., Branco, M.R., Reichmann, J., Reik, W., Surani, M.A., Adams, I.R., and Meehan, R.R. (2012). Promoter DNA methylation couples genome-defence mechanisms to epigenetic reprogramming in the mouse germline. *Development* *139*, 3623–3632. <https://doi.org/10.1242/dev.081661>.

- 503
- 504 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>.
- 505
- 506 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>.
- 507
- 508 24. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for  
509 RNA-seq data with DESeq2. *Genome Biol* 15, 550. <https://doi.org/10.1186/s13059-014-0550-8>.
- 510 25. Crackower, M.A., Kolas, N.K., Noguchi, J., Sarao, R., Kikuchi, K., Kaneko, H., Kobayashi, E., Kawai, Y.,  
Kozieradzki, I., Landers, R., et al. (2003). Essential Role of Fkbp6 in Male Fertility and Homologous  
511 Chromosome Pairing in Meiosis. *Science* 300, 1291–1295. <https://doi.org/10.1126/science.1083022>.
- 512 26. Xiol, J., Cora, E., Koglgruber, R., Chuma, S., Subramanian, S., Hosokawa, M., Reuter, M., Yang, Z.,  
Berninger, P., Palencia, A., et al. (2012). A Role for Fkbp6 and the Chaperone Machinery in piRNA  
513 Amplification and Transposon Silencing. *Molecular Cell* 47, 970–979. <https://doi.org/10.1016/j.molcel.2012.07.019>.
- 514 27. Cheng, S., Altmeppen, G., So, C., Welp, L.M., Penir, S., Ruhwedel, T., Menelaou, K., Harasimov, K.,  
Stützer, A., Blayney, M., et al. (2022). Mammalian oocytes store mRNAs in a mitochondria-associated  
515 membraneless compartment. *Science* 378, eabq4835. <https://doi.org/10.1126/science.abq4835>.
- 516 28. Rouland, A., Masson, D., Lagrost, L., Vergès, B., Gautier, T., and Bouillet, B. (2022). Role of  
517 apolipoprotein C1 in lipoprotein metabolism, atherosclerosis and diabetes: A systematic review.  
*Cardiovasc Diabetol* 21, 272. <https://doi.org/10.1186/s12933-022-01703-5>.
- 518 29. Mueller, J.L., Skaletsky, H., Brown, L.G., Zaghloul, S., Rock, S., Graves, T., Auger, K., Warren,  
W.C., Wilson, R.K., and Page, D.C. (2013). Independent specialization of the human and mouse X  
519 chromosomes for the male germ line. *Nat Genet* 45, 1083–1087. <https://doi.org/10.1038/ng.2705>.
- 520 30. Handel, M.A., and Eppig, J.J. (1979). Sertoli Cell Differentiation in the Testes of Mice Genetically  
Deficient in Germ Cells. *Biology of Reproduction* 20, 1031–1038. <https://doi.org/10.1095/biolreprod20.5.1031>.
- 521
- 522 31. 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>.
- 523

- 524 32. 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>.
- 525
- 526 33. 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>.
- 527
- 528 34. 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>.
- 529
- 530 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>.
- 531
- 532 36. 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>.
- 533
- 534 37. 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>.
- 535
- 536 38. 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>.
- 537
- 538 39. 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>.
- 539
- 540 40. 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>.
- 541
- 542 41. 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>.
- 543
- 544 42. Koubova, J., Menke, D.B., Zhou, Q., Capel, B., Griswold, M.D., and Page, D.C. (2006). Retinoic  
acid regulates sex-specific timing of meiotic initiation in mice. *Proc. Natl. Acad. Sci. U.S.A.* *103*,  
2474–2479. <https://doi.org/10.1073/pnas.0510813103>.
- 545

- 546 43. Lin, Y., Gill, M.E., Koubova, J., and Page, D.C. (2008). Germ Cell-Intrinsic and -Extrinsic Factors  
547 Govern Meiotic Initiation in Mouse Embryos. *Science* *322*, 1685–1687. <https://doi.org/10.1126/science.1166340>.
- 548 44. Endo, T., Mikedis, M.M., Nicholls, P.K., Page, D.C., and De Rooij, D.G. (2019). Retinoic Acid and Germ  
549 Cell Development in the Ovary and Testis. *Biomolecules* *9*, 775. <https://doi.org/10.3390/biom9120775>.
- 550 45. Gupta, N., Yakhou, L., Albert, J.R., Azogui, A., Ferry, L., Kirsh, O., Miura, F., Battault, S., Yamaguchi,  
551 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>.
- 552 46. Hurlin, P.J. (1999). Mga, a dual-specificity transcription factor that interacts with Max and contains a  
553 T-domain DNA-binding motif. *The EMBO Journal* *18*, 7019–7028. <https://doi.org/10.1093/emboj/18.2.4.7019>.
- 554 47. 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-  
555 Specific Gene Expression. *Current Biology* *15*, 1051–1057. <https://doi.org/10.1016/j.cub.2005.04.060>.
- 556 48. Stielow, B., Finkernagel, F., Stiewe, T., Nist, A., and Suske, G. (2018). MGA, L3MBTL2 and E2F6  
557 determine genomic binding of the non-canonical Polycomb repressive complex PRC1.6. *PLoS Genet*  
*14*, e1007193. <https://doi.org/10.1371/journal.pgen.1007193>.
- 558 49. Dahlet, T., Truss, M., Frede, U., Al Adhami, H., Bardet, A.F., Dumas, M., Vallet, J., Chicher, J.,  
Hamann, P., Kottnik, S., et al. (2021). E2F6 initiates stable epigenetic silencing of germline genes  
559 during embryonic development. *Nat Commun* *12*, 3582. <https://doi.org/10.1038/s41467-021-23596-w>.
- 560 50. 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.  
561 <https://doi.org/10.1016/j.molcel.2010.05.004>.
- 562 51. Gajiwala, K.S., Chen, H., Cornille, F., Roques, B.P., Reith, W., Mach, B., and Burley, S.K. (2000).  
563 Structure of the winged-helix protein hRFX1 reveals a new mode of DNA binding. *Nature* *403*,  
916–921. <https://doi.org/10.1038/35002634>.
- 564 52. Swoboda, P., Adler, H.T., and Thomas, J.H. (2000). The RFX-Type Transcription Factor DAF-19  
565 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).
- 566 53. 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  
567 Ciliogenesis. *Sci. Signal.* *2*. <https://doi.org/10.1126/scisignal.2000602>.

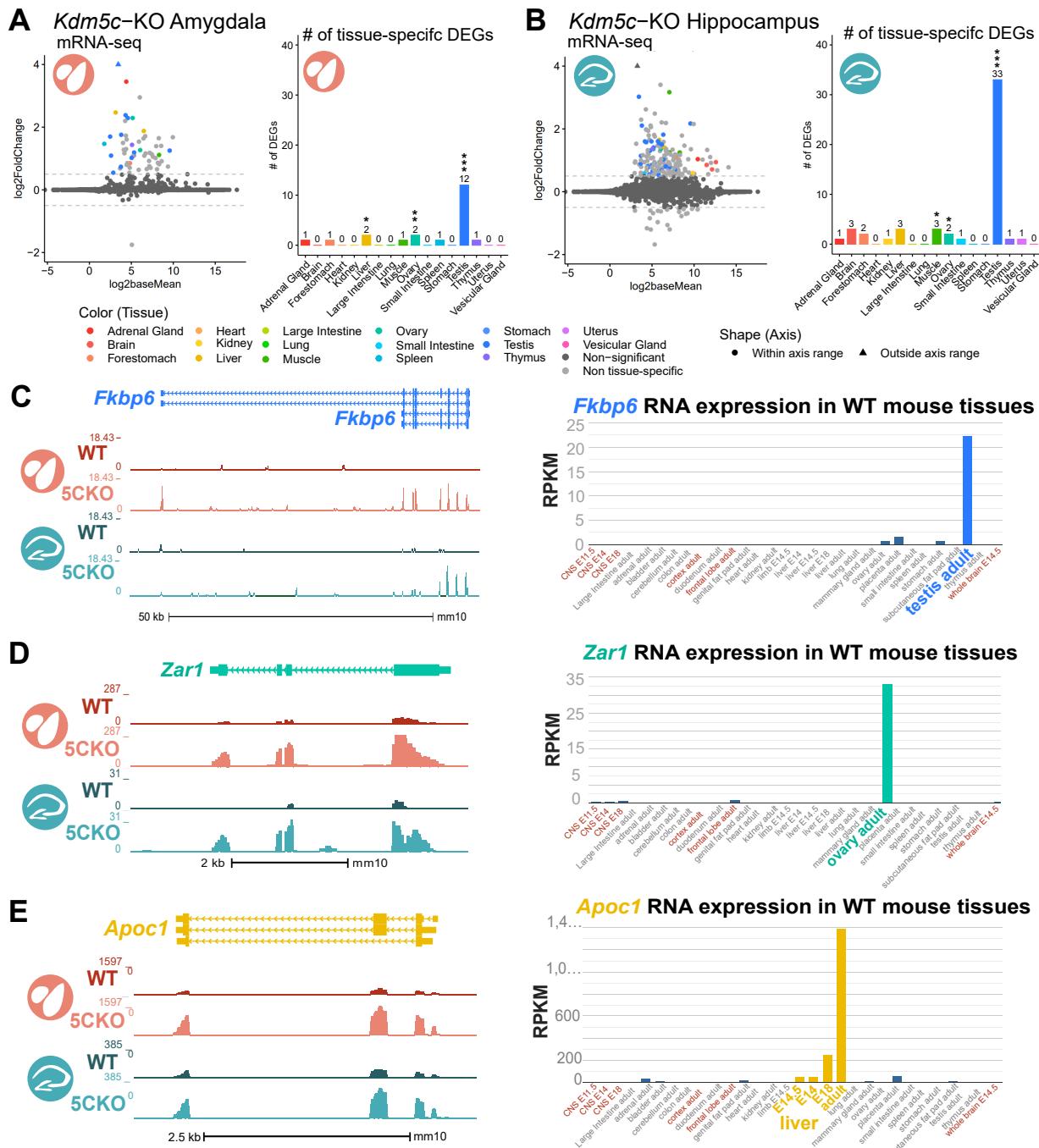
- 568 54. Kistler, W.S., Baas, D., Lemeille, S., Paschaki, M., Seguin-Estevez, Q., Barras, E., Ma, W., Duteyrat, J.-L., Morlé, L., Durand, B., et al. (2015). RFX2 Is a Major Transcriptional Regulator of Spermiogenesis. PLoS Genet 11, e1005368. <https://doi.org/10.1371/journal.pgen.1005368>.
- 569
- 570 55. 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>.
- 571
- 572 56. 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>.
- 573
- 574 57. 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>.
- 575
- 576 58. 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>.
- 577
- 578 59. 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>.
- 579
- 580 60. 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>.
- 581
- 582 61. 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>.
- 583
- 584 62. 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>.
- 585
- 586 63. 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>.
- 587
- 588 64. 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>.
- 589

- 590 65. 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>.
- 591
- 592 66. 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>.
- 593
- 594 67. 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>.
- 595
- 596 68. Bonefas, K.M., and Iwase, S. (2021). Soma-to-germline transformation in chromatin-linked neurode-597  
velopmental disorders? *FEBS J.* <https://doi.org/10.1111/febs.16196>.
- 598 69. 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>.
- 599
- 600 70. Walton, E.L., Francastel, C., and Velasco, G. (2014). Dnmt3b Prefers Germ Line Genes and Cen-601  
tromeric Regions: Lessons from the ICF Syndrome and Cancer and Implications for Diseases. *Biology* (Basel) *3*, 578–605. <https://doi.org/10.3390/biology3030578>.
- 602 71. 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 603  
duplication syndrome. *Nat Genet* *44*, 206–211. <https://doi.org/10.1038/ng.1066>.
- 604 72. Stephens, M. (2016). False discovery rates: A new deal. *Biostat*, kxw041. <https://doi.org/10.1093/biostatistics/kxw041>.
- 605
- 606 73. Conway, J.R., Lex, A., and Gehlenborg, N. (2017). UpSetR: An R package for the visualization of 607  
intersecting sets and their properties. *Bioinformatics* *33*, 2938–2940. <https://doi.org/10.1093/bioinformatics/btx364>.
- 608 74. Li, H., Liang, Z., Yang, J., Wang, D., Wang, H., Zhu, M., Geng, B., and Xu, E.Y. (2019). DAZL 609  
is a master translational regulator of murine spermatogenesis. *Natl Sci Rev* *6*, 455–468. <https://doi.org/10.1093/nsr/nwy163>.
- 610 75. 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>.

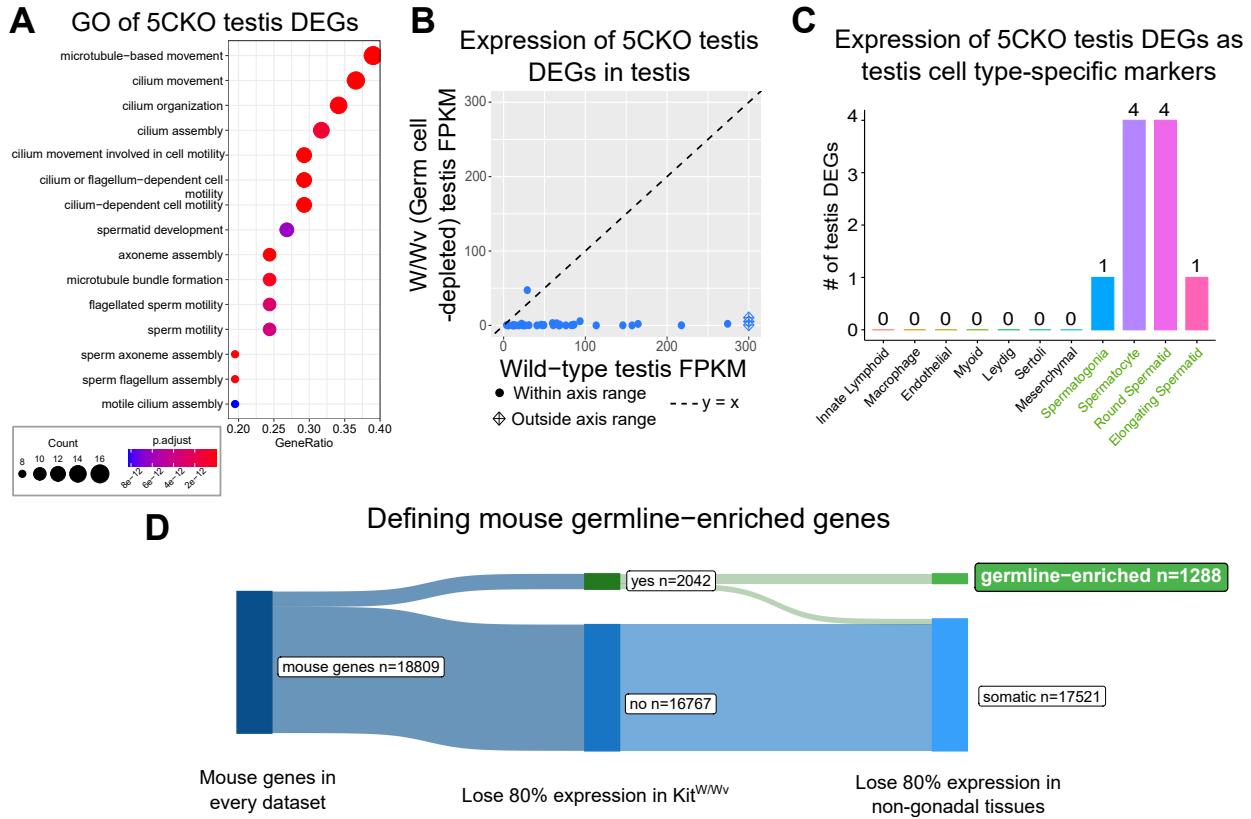
611

612 **Figures and Tables**

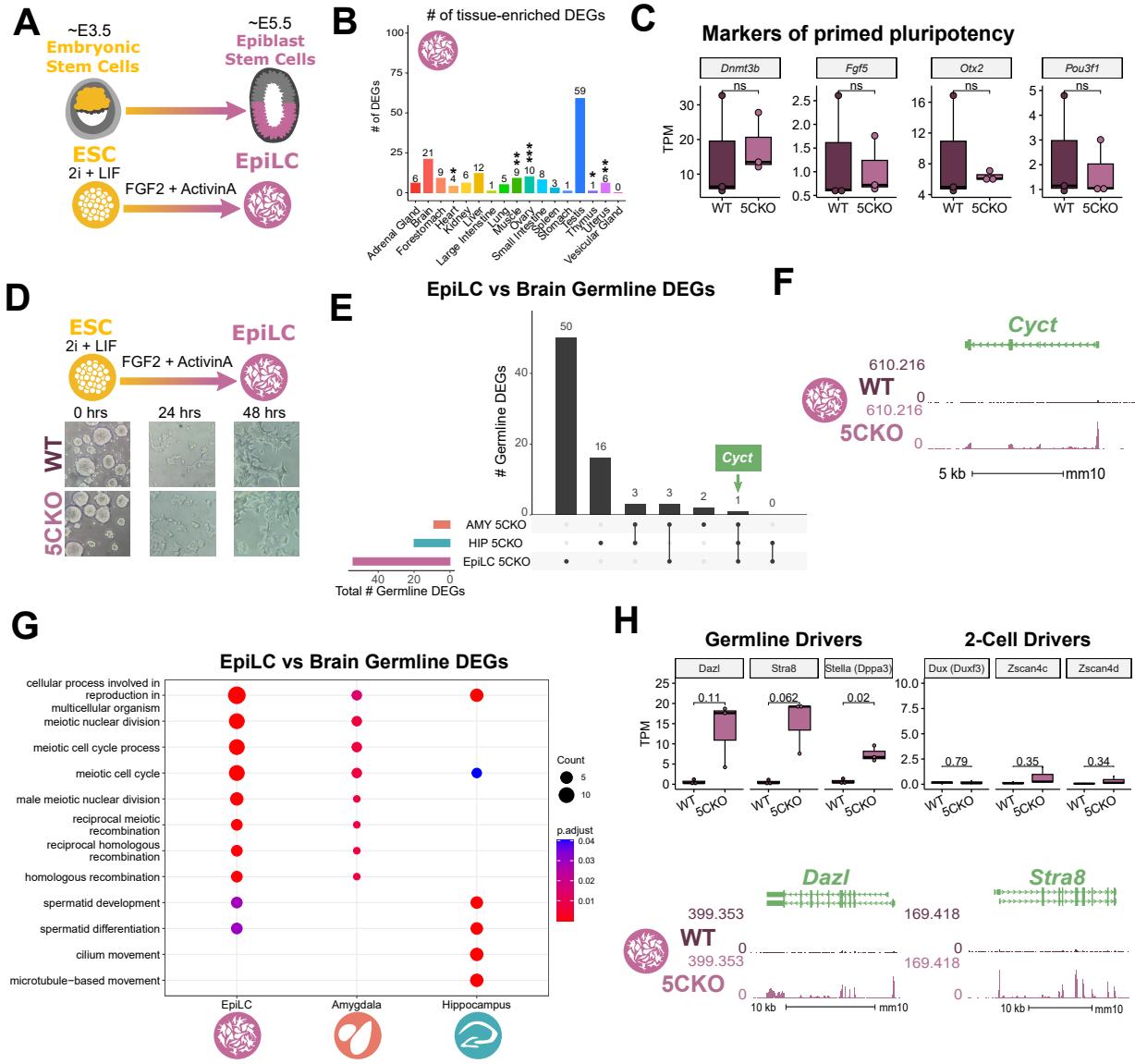
- 613     • Supplementary table 1: list of all germline genes.
- 614       – Columns to include:
- 615           \* KDM5C bound vs not
- 616           \* DEG in EpiLC, brain, both, neither (separate columns?)



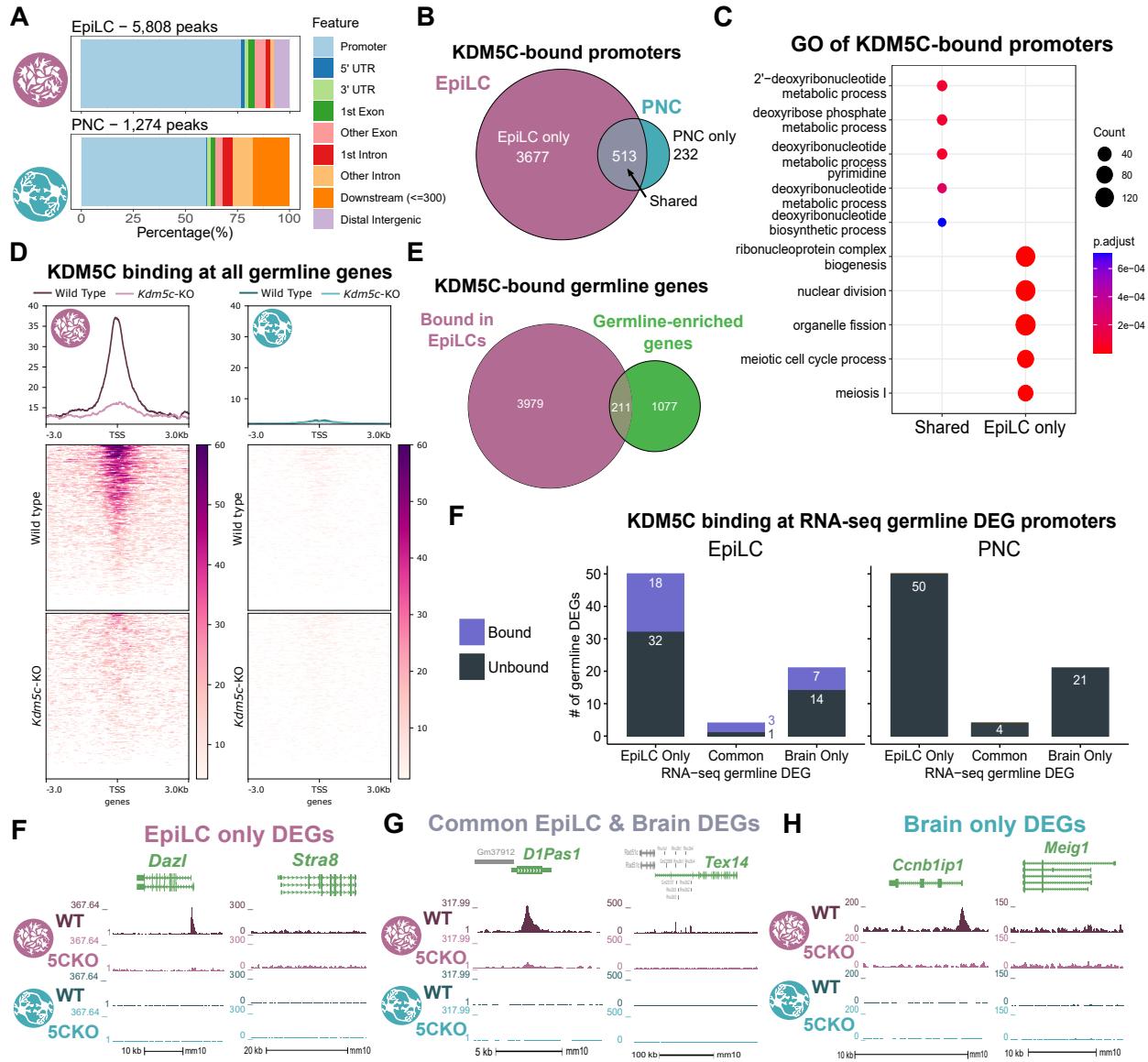
**Figure 1: Tissue-enriched genes are misexpressed in the *Kdm5c*-KO brain.** **A-B.** Expression of tissue-enriched genes (Li et al 2017) in the male *Kdm5c*-KO amygdala (A) and hippocampus (B). Left - MA plot of mRNA-sequencing. Right - Number of tissue-enriched differentially expressed genes (DEGs).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

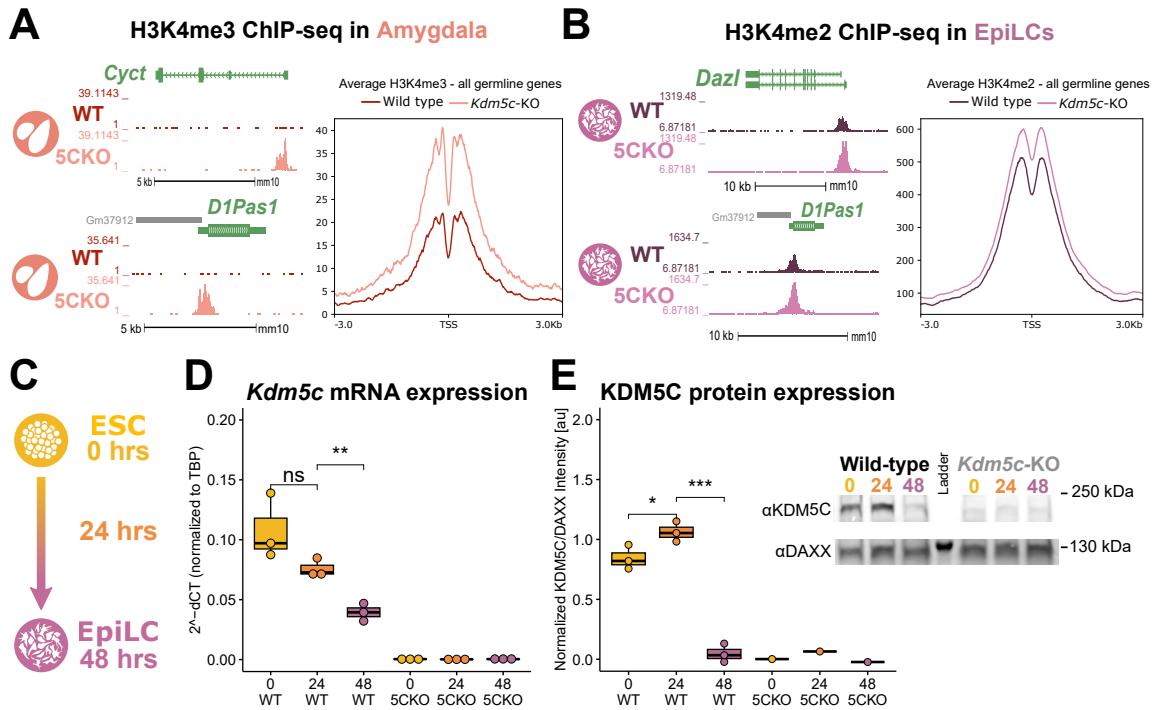


**Figure 3: Kdm5c-KO epiblast-like cells express key drivers of germline identity** **A.** Top - Diagram of *in vivo* differentiation of embryonic stem cells (ESCs) of the inner cell mass into epiblast stem cells. Bottom - *in vitro* differentiation of ESCs into epiblast-like cells (EpiLCs). **B.** Number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

## 617 Notes

### 618 Things to do

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

### 623 **Dazl**

624     We were particularly interested in the aberrant transcription of *Dazl*, since it is essential for germ cell  
625 development and promotes the translation of germline mRNAs<sup>74</sup>. A significant portion of germline transcripts  
626 misexpressed in *Kdm5c*-KO EpiLCs are known binding targets of DAZL, including *Stra8*<sup>75</sup> ( $p = 1.698e-07$ ,  
627 Fisher's Exact Test). This suggests expression of DAZL protein could promote the translation of other  
628 aberrant germline transcripts, influencing their ability to impact *Kdm5c*-KO cellular function. We thus tested  
629 DAZL protein expression in *Kdm5c*-KO EpiLCs through immunocytochemistry (Figure 3I). We observed  
630 about 25% of *Kdm5c*-KO EpiLCs expressed DAZL protein and it was localized to the cytoplasm ( $p = 0.0015$ ,  
631 Welch's t-test), consistent with the pattern of DAZL expression in spermatogonia<sup>75</sup>. Altogether these results  
632 suggest tissue-specific genes are misexpressed during *Kdm5c*-KO embryogenesis, including key drivers of  
633 germline identity that can be translated into protein.

- 634     • We additionally found *Kdm5c*-KO EpiLCs ectopically express DAZL protein that is localized to the  
635         cytoplasm, similar to its morphology in spermatogonia<sup>75</sup>. **note: maybe just put in results.** Could  
636         move around depending upon if I get pheno working.

### 637 Discussion notes

- 638     • For other paper:
  - 639         – for methods: Heatmaps of gene expression were generated using the base R functions scale and  
640             hclust and visualized using the R package ComplexHeatmap (v2.12.1).
  - 641         – \* 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>
- 643     • Talk about different motifs if we do see differences there and if it explains direct vs indirect dysregulation
  - 644         – Could also mention PRC1.6/KDM5C binds pcgf6 in ESCs
- 645     • Maybe talk about other regulators/if anything is known about them long term? Or save for 2nd paper.

- 646           – another germline repressor: <https://pubmed.ncbi.nlm.nih.gov/35345626/>
- 647       • end with something like: this indicates kdm5c not only modulates pluripotency and self-renewal in  
648        ESCs, but also has a role in long-term silencing of germline genes
- 649           – then transition into the long term silencing mechanism paragraph
- 650       • KDM5C is dynamically regulated during the window of embryonic germline gene silencing, our EpiLC  
651        ChIP-seq is likely catching the tail end of KDM5C's main involvement.
- 652       • Our results indicate KDM5C is a pivotal repressor of germline programs in somatic cells.
- 653       • Studies on other chromatin regulators have focused on key marker genes to identify germline gene  
654        misexpression, such as *Dazl*.
- 655       • Using a germline-depleted mouse model, we curated a list of mouse germline-enriched genes to  
656        globally assess germline gene dysregulation.
- 657       • Many of the genes dysregulated in the mature *Kdm5c*-KO brain were important for late stages of  
658        spermatogenesis, such as those important for sperm flagellar structure. Contrastingly, *Kdm5c*-KO  
659        EpiLCs aberrantly expressed key drivers of germ cell identity, including *Dazl* and *Stra8*.
- 660       • While a significant portion of KDM5C-bound promoters are germline genes, many germline genes  
661        expressed during *Kdm5c*-KO embryogenesis are not directly bound by kdm5c.
- 662       • One notable example is *Stra8*, a transcription factor important for germ cell specific transcription and  
663        meiotic initiation
- 664       • The including the demarcation between soma and germline fates.
- 665       the *Kdm5c*-KO brain also Previous work identified testis-enriched genes in *Kdm5c* genes
- 666           –
- 667           – However unlike the gonadal-biased DEGs,
- 668       • The demarcation of the germline versus soma is a key feature of multicellularity and sexually dimorphic  
669        reproduction
- 670       • Anything known about tissue-biased gene expression in other H3K4me regulators?
- 671       • Our data suggests the germline developmental program is occurring ectopically as *Kdm5c*-KOs pro-  
672        gresses through somatic tissue development
- 673       • tissue-biased gene expression:

- 674     • Altogether, these data indicate that while some germline genes are misexpressed due to direct loss of  
675       KDM5C binding during emryogenesis, secondary downstream mechanisms can also promote their  
676       aberrant transcription.
- 677     • Papers to read/reference:  
678       – Reconstitution of the Mouse Germ Cell Specification Pathway in Culture by Pluripotent Stem Cells:  
679       [https://www.cell.com/fulltext/S0092-8674\(11\)00771-9](https://www.cell.com/fulltext/S0092-8674(11)00771-9)  
680       – 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/>

682 **Figure outline:**

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

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

695     **Figure 3: Kdm5c-KO epiblast-like cells express key drivers regulators of germline identity** \* A) ESC  
696       to EpiLC differentiation Left - Morphology is unchanged, \* B) 5CKO EpiLCs express EpiLC differentiation  
697       genes similar to WT lvs \* C) Male EpiLCs express germline genes (example Cyct again) \* Overlap between  
698       brain and EpiLC germline genes - show they're mostly unique \* GO of Brain and EpiLC germline genes  
699       (meiotic enriched) \* Bigwigs or TPM of master regulators \* Show that while some are also 2-cell genes  
700       (Dazl), 2-cell specific genes aren't dysregulated (Zscan4). Important point because published KDM5C dazl  
701       paper is saying KDM5C is a 2-cell regulator, but as far as I can tell only genes shared between germline and  
702       2-cell are dysregulated.

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

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

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

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

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

722 Introduction: \* Chromatin regulators are important for cellular identity \* H3K4me1-3 linked to active  
723 gene promoters and enhancers \* Surprisingly, mutations in Chromatin regulators lead to many NDDs  
724 (including many H3K4 regulators) \* Recent studies have shown some chromatin regulators are important  
725 for regulating neuron-specific gene expression/chromatin stat\_compare\_means \* However, loss of some  
726 chromatin regulators can also lead to ectopic expression of tissue-enriched genes \* Very few studies have  
727 looked at these genes and it's unclear if these genes contribute to NDD impairments. \* Necessary to  
728 first characterize the mechanism behind their derepression to identify molecular footholds into testing their  
729 contribution to neuronal impairments and potential for therapeutic intervention

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

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

747 **Germline gene repression background:**

748 Interestingly, some of the ectopic testis transcripts identified in the *Kdm5c*-KO brain are typically ex-  
749 pressed in germ cells<sup>10</sup>. Unlike somatic cells, germ cells (e.g. sperm and eggs) undergo meiosis and pass  
750 on their genetic material to the next generation. The germline and the soma are typically distinguished during  
751 early embryogenesis, when germline genes are silenced in epiblast stem cells soon after implantation and  
752 only reactivated in a subset to form the germline. Chromatin regulators play a key role in decommissioning  
753 germline genes as the embryo transitions from naïve to primed pluripotency by placing repressive histone  
754 H2A lysine 119 monoubiquitination (H2AK119ub1)<sup>18</sup>, histone 3 lysine 9 trimethylation (H3K9me3)<sup>18,19</sup>, and  
755 DNA CpG methylation<sup>19–21</sup> at germline gene promoters. KDM5C may also be involved in this early decom-  
756 missioning of germline genes, as re-expression of KDM5C in neurons fails to suppress their dysregulation<sup>10</sup>.  
757 In support of this, KDM5C was very recently shown to repress *Deleted in azoospermia like (Dazl)*, a key  
758 regulator of germline development, in mouse embryonic stem cells (ESCs)<sup>45,65</sup>. In support of this, two  
759 independent screens in mouse embryonic stem cells (ESCs) recently identified KDM5C as a repressor of  
760 *Deleted in azoospermia like (Dazl)*, a key regulator of germline development. However, KDM5C's role in  
761 embryonic germline gene repression is currently unclear, given that Dazl is also expressed in ESCs and in  
762 the 2-cell stage and germline gene misexpression has yet to be globally characterized during *Kdm5c*-KO  
763 embryogenesis.