

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

- 27 • Not sure if I like the last sentence, I think we should focus on what we did do versus what we could do  
28 in the future

- 29 – Add something that we deepened the characterization of germline gene classes that use different  
30 regulatory mechanisms (CpG islands vs not, meiotic vs late-stage, E2F6/MAX vs no)

## 31 Introduction

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

42 To elucidate the role of tissue identity in chromatin-linked neurodevelopmental disorders, it is essential  
43 to first characterize the nature of the dysregulated genes and the molecular mechanisms governing their  
44 de-repression. We characterized lysine demethylase 5C (KDM5C), also known as SMCX or JARID1C, which  
45 erases histone 3 lysine 4 di- and trimethylation (H3K4me2/3) found at active gene promoters<sup>9</sup>. Pathogenic  
46 mutations in *KDM5C* cause Intellectual Developmental Disorder, X-linked, Syndromic, Claes-Jensen Type  
47 (MRXSCJ, OMIM: 300534). MRXSCJ is more common and severe in males and its neurological phenotypes  
48 include intellectual disability, seizures, aberrant aggression, and autistic behaviors<sup>10–12</sup>. Male *Kdm5c*  
49 knockout (-KO) mice recapitulate key MRXSCJ phenotypes, including hyperaggression, increased seizure  
50 propensity, and learning impairments<sup>8,13</sup>. RNA sequencing (RNA-seq) of the *Kdm5c*-KO hippocampus  
51 revealed ectopic expression of testis genes within the brain<sup>8</sup>, however it is unknown what types of testis  
52 genes are dysregulated, when in *Kdm5c*-KO development testis gene dysregulation begins, and if other  
53 tissue-specific genes are also aberrantly transcribed with KDM5C loss.

54 Distinguishing between germ cells and somatic cells is a key feature of multicellularity<sup>14</sup> that occurs  
55 during early embryogenesis in many metazoans<sup>15</sup>. In mammals, chromatin regulators are crucial for  
56 decommissioning germline genes in somatic cells during the transition from naïve to primed pluripotency.  
57 Initially, germline gene promoters gain repressive histone H2A lysine 119 monoubiquitination (H2AK119ub1)<sup>16</sup>  
58 and histone 3 lysine 9 trimethylation (H3K9me3)<sup>16,17</sup> in embryonic stem cells and are then decorated with  
59 DNA CpG methylation in the post-implantation embryo<sup>17–19</sup>. The precise roles of KDM5C during this process  
60 remains unclear. Additionally, studies on germline gene repression have primarily focused on marker genes  
61 important for germ cell development rather than germline genes as a whole, given they lacked a curated  
62 list of germline-enriched genes. Therefore, it is unknown if the mechanism of repression differs for certain

63 classes of germline genes, e.g. meiotic genes versus spermatid differentiation genes.

64 It is also unknown if ectopic germline gene expression is influenced by chromosomal sex, as previous  
65 studies on germline gene repression have been exclusively in males. Sex is particularly pertinent in the case  
66 of KDM5C, which lies on the X chromosome and partially escapes X chromosome inactivation, resulting in a  
67 higher dosage in females<sup>20-23</sup>.

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

## 74 Results

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

76 Previous RNA sequencing (RNA-seq) in the adult hippocampus revealed ectopic expression of some  
77 testis genes within the male *Kdm5c* knockout (-KO) brain<sup>8</sup>. It is currently unclear if the testis is the only  
78 tissue type misexpressed in the *Kdm5c*-KO brain. We thus characterized dysregulation of *Kdm5c*-KO brain  
79 tissue identity by systematically assessing the expression of genes enriched in 17 mouse tissues<sup>24</sup>, using  
80 our published mRNA-seq datasets of the adult amygdala and hippocampus for male mice with constitutive  
81 knockout of *Kdm5c*<sup>25</sup>.

82 We found a large proportion of genes that are significantly upregulated within the *Kdm5c*-KO brain  
83 (DESeq2<sup>26</sup>, log2 fold change > 0.5, q < 0.1) are enriched in non-brain tissues in wild-type mice (Amygdala:  
84 35%, Hippocampus: 24%) (Figure 1A-B). The majority of tissue-enriched differentially expressed genes  
85 (tissue-enriched DEGs) were testis genes (Figure 1A-C). Even though the testis has the largest total number  
86 of biased genes (2,496 genes) compared to any other tissue, testis-biased DEGs were significantly enriched  
87 for both brain regions (Amygdala p = 1.83e-05, Odds Ratio = 5.13; Hippocampus p = 4.26e-11, Odds Ratio =  
88 4.45, Fisher's Exact Test). One example of a testis-enriched gene misexpressed in the *Kdm5c*-KO brain is  
89 *FK506 binding protein 6* (*Fkbp6*), a known regulator of piRNA expression and meiosis<sup>27,28</sup> (Figure 1C).

90 Interestingly, we also observed significant enrichment of ovary-biased DEGs in both the amygdala and  
91 hippocampus (Amygdala p = 0.00574, Odds Ratio = 18.7; Hippocampus p = 0.048, Odds Ratio = 5.88,  
92 Fisher's Exact) (Figure 1D). Ovary-enriched DEGs included *Zygotic arrest 1* (*Zar1*), which sequesters mRNAs  
93 in oocytes for meiotic maturation and early zygote development<sup>29</sup> (Figure 1D). Given the *Kdm5c*-KO mice  
94 we analyzed are male, this demonstrates ectopic expression of tissue-enriched genes is independent of  
95 organismal sex.

96     Although not consistent across brain regions, we also found significant enrichment of DEGs biased  
97 towards two non-gonadal tissues - the liver (Amygdala p = 0.0398, Odds Ratio = 6.58, Fisher's Exact Test)  
98 and the muscle (Hippocampus p = 0.0104, Odds Ratio = 6.95, Fisher's Exact Test). A liver-biased DEG  
99 dysregulated in both the hippocampus and amygdala is *Apolipoprotein C-I* (*Apoc1*), a lipoprotein metabolism  
100 and transport gene<sup>30</sup> that has been implicated in Alzheimer's disease<sup>31</sup> (Figure 1E). For all tissue-enriched  
101 DEGs, aberrantly expressed mRNAs are polyadenylated and spliced into mature transcripts (Figure 1C-E).  
102 Of note, we observed little to no dysregulation of brain-enriched genes (Amygdala p = 1; Hippocampus p =  
103 0.74, Fisher's Exact), despite the fact these are brain samples and the brain has the second highest total  
104 number of tissue-enriched genes (708 genes). Together, these results suggest the aberrant expression of  
105 tissue-enriched genes within the brain is a major effect of KDM5C loss.

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

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

113     To further validate if these testis DEGs are truly germline genes, we assessed their expression in somatic  
114 versus germ cells within the testis. We first compared their expression in wild-type testes to those without  
115 germ cells<sup>32</sup>, which was accomplished by heterozygous *W* and *Wv* mutations in the enzymatic domain of  
116 *c-Kit* (*Kit*<sup>W/Wv</sup>)<sup>33</sup>. Almost all *Kdm5c*-KO testis-enriched DEGs lost expression with germ cell depletion (Figure  
117 2B). We then assessed testis-enriched DEG expression in a published single cell RNA-seq dataset that  
118 identified cell type-specific markers within the testis<sup>34</sup>. Some *Kdm5c*-KO testis-enriched DEGs were classified  
119 as specific markers for different germ cell types (e.g. spermatogonia, spermatocytes, round spermatids, and  
120 elongating spermatids), yet none marked somatic cells (Figure 2C). Together, these data demonstrate that  
121 the *Kdm5c*-KO brain aberrantly expresses germline genes, but not somatic testicular genes, reflecting an  
122 erosion of the soma-germline boundary.

123     We then wanted to characterize germline gene misexpression with *Kdm5c* loss more deeply, but lacked  
124 a comprehensive list of mouse germline genes. We therefore generated a list of germline-enriched genes  
125 using RNA-seq datasets of *Kit*<sup>W/Wv</sup> mice that included males and females at embryonic day 12, 14, and 16<sup>35</sup>  
126 and adult male testes<sup>32</sup>. We defined genes as germline-enriched if their expression met the following criteria:  
127 1) their expression is greater than 1 FPKM in wild-type gonads 2) their expression in any non-gonadal tissue  
128 of adult wild type mice<sup>24</sup> does not exceed 20% of their maximum expression in the wild-type germline, and  
129 3) their expression in the germ cell-depleted gonads, for any sex or time point, does not exceed 20% of  
130 their maximum expression in the wild-type germline. These criteria yielded 1,288 germline-enriched genes

131 (Figure 2D), which was hereafter used as a resource to globally assess germline gene misexpression with  
132 *Kdm5c* loss (Supplementary table 1).

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

134 Germ cells are typically distinguished from somatic cells soon after the embryo implants into the uterine  
135 wall<sup>36,37</sup>. A subset of epiblast stem cells become the primordial germ cells (PGCs) while the remainder  
136 differentiate into the ectoderm, mesoderm, and endoderm to form the somatic tissues<sup>38</sup>. This developmental  
137 time point can be modeled *in vitro* through differentiation of embryonic stem cells (ESCs) into epiblast-like  
138 stem cells (EpiLCs) (Figure 3A, top)<sup>39,40</sup>. While some germline-enriched genes are also expressed in  
139 naïve embryonic stem cells (ESCs) and in the 2-cell stage<sup>41–43</sup>, they are silenced as they differentiate into  
140 EpiLCs<sup>17,44</sup>. Therefore, we tested if KDM5C was necessary for silencing germline genes at this developmental  
141 stage by evaluating the impact of *Kdm5c* loss in male EpiLCs.

142 We first identified *Kdm5c*-KO EpiLC DEGs in our previously published RNA-seq dataset<sup>45</sup> (DESeq2, log2  
143 fold change > 0.5, q < 0.1) and then assessed tissue-enriched gene expression. Similar to the *Kdm5c*-KO  
144 brain, we observed general dysregulation of tissue-enriched genes, with the largest number of genes  
145 belonging to the brain and testis, although they were not significantly enriched (Figure 3B). Markers of primed  
146 pluripotency, such as *Dnmt3b*, *Fgf5*, *Pou3f1*, and *Otx2*, were unaltered in *Kdm5c*-KO EpiLCs (Figure 3C).  
147 Furthermore, *Kdm5c*-KO cell morphology during ESC to EpiLC differentiation appeared normal (Figure 3D),  
148 indicating KDM5C loss does not impair EpiLC formation.

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

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

166 germline identity and meiosis.

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

169 It is currently unknown if the misexpression of germline genes is influenced by sex, as previous studies  
170 on germline gene repressors have focused on males<sup>16–18,50,51</sup>. We explored the impact of sex upon germline  
171 gene misexpression by comparing their dysregulation in male *Kdm5c* hemizygous knockout (XY *Kdm5c*-  
172 KO), female homozygous knockout (XX *Kdm5c*-KO), and female heterozygous knockout (XX *Kdm5c*-HET)  
173 EpiLCs.<sup>45</sup>.

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

184 The majority of germline genes expressed in both sexes had a greater log2 fold change in females  
185 compared to males (Figure 4D). This increased degree of dysregulation in females, along with the increased  
186 total number of germline genes, indicates females are more sensitive to losing KDM5C-mediated suppression  
187 of germline genes. Female sensitivity could be due to improper X chromosome inactivation (XCI) in *Kdm5c*  
188 mutants<sup>45</sup>, as many spermatogenesis genes lie on the X chromosome<sup>52,53</sup>. However, both shared and  
189 female-specific germline DEGs were not biased towards the X chromosome and the majority of female  
190 DEGs instead lie on autosomes (Figure 4G). Thus, while female EpiLCs are more prone to germline gene  
191 misexpression with KDM5C loss, it is likely independent of potential defects in XCI.

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

193 While many germline genes act in both the male and female germline, some display sex-biased expression  
194 or have functions unique to eggs and sperm. Therefore, we wondered if *Kdm5c* mutant males express  
195 sperm genes while mutant females express egg genes. To comprehensively assess whether germline  
196 gene sex corresponds with *Kdm5c* mutant sex, we filtered our list of germline-enriched genes for egg and  
197 sperm-biased genes. We defined germ cell sex-biased genes as those whose expression in the opposite  
198 sex, at any time point, is no greater than 20% of the gene's maximum expression in a given sex. This yielded

199 67 egg-biased, 1,024 sperm-biased, and 197 unbiased germline-enriched genes (Figure 4H). We found  
200 egg, sperm, and unbiased germline genes were dysregulated in all *Kdm5c* mutants (Figure 4I-J). Germline  
201 genes dysregulated exclusively in *Kdm5c* mutant males and females were also not biased towards their  
202 corresponding germ cell sex. This indicates sex differences in germline gene dysregulation is not due to  
203 sex-specific activation of sperm or egg transcriptional programs. These results demonstrate that the sex of  
204 *Kdm5c* mutant cells influences the degree of germline gene, independent of germline gene sex.

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

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

209 KDM5C binds to the promoters of several germline genes in embryonic stem cells (ESCs) but its binding  
210 is absent in neurons<sup>8</sup>. However, the lack of a comprehensive list of germline-enriched genes prohibited  
211 systematic characterization of KDM5C binding at germline gene promoters. Thus, it is unclear if KDM5C is  
212 enriched at germline gene promoters, what types of germline genes KDM5C regulates, and if its binding is  
213 maintained at any germline genes in neurons.

214 To address these questions, we analyzed KDM5C chromatin immunoprecipitation followed by DNA  
215 sequencing (ChIP-seq) datasets in EpiLCs<sup>45</sup> and primary forebrain neuron cultures (PNCs)<sup>13</sup>. EpiLCs had a  
216 higher total number of KDM5C peaks than PNCs (EpiLCs: 5,808, PNCs: 1,276, MACS2 q < 0.1 and fold  
217 enrichment > 1, removal of *Kdm5c*-KO false positive peaks) and KDM5C was primarily localized to gene  
218 promoters in both cell types (EpiLCs: 4,190, PNCs: 745 +/- 500bp from the TSS), although PNCs showed  
219 increased localization to non-promoter regions (Figure 5A).

220 The majority of promoters bound by KDM5C in PNCs were also bound in EpiLCs (513 shared promoters),  
221 however a large portion of gene promoters were only bound by KDM5C in EpiLCs (3677 EpiLC only  
222 promoters) (Figure 5B). Genes bound by KDM5C in both PNCs and EpiLCs were enriched for functions  
223 involving nucleic acid turnover, such as deoxyribonucleotide metabolic process (GO:0009262, p.adjust =  
224 8.28e-05) (Figure 5C). Germline-specific ontologies were only enriched in promoters unique to EpiLCs, such  
225 as meiotic nuclear division (GO: 0007127 p.adjust = 6.77e-16) and meiotic cell cycle process (GO:1903046,  
226 p.adjust = 5.05e-16) (Figure 5C). There were no ontologies significantly enriched for genes only bound  
227 by KDM5C in PNCs. We next evaluated KDM5C binding around the transcription start site (TSS) of all  
228 germline-enriched genes. In EpiLCs, we observed modest KDM5C signal at about half of all germline genes  
229 (Figure 5D). Based on our ChIP-seq peak cut-off criteria, KDM5C was strongly bound to about 10% of  
230 germline gene promoters in EpiLCs (Figure 5E). One notable gene that lacked KDM5C binding was *Stra8*,  
231 even though its mRNA is expressed in *Kdm5c*-KO EpiLCs (Figure 5K). In EpiLCs, KDM5C was only bound  
232 to about one third of *Kdm5c*-KO RNA-seq DEG promoters (EpiLC only DEGs: 36%, Brain only DEGs:

233 33.3%), but 3 out of the 4 genes dysregulated in both the brain and EpiLCs (Supplementary figure XXX).  
234 In condordance with our gene ontology results, we did not observe KDM5C accumulation at any germline  
235 gene promtoers in PNCs (Figure 5D). Together, these results demonstrate KDM5C is recruited to a subset of  
236 germline genes in EpiLCs, including meiotic genes, but does not directly regulate germline genes in neurons.  
237 Furthermore, the majority of germline mRNAs expressed in *Kdm5c*-KO cells are dysregulated independent  
238 of direct KDM5C binding to their promoters.

239 Many germline-specific genes are suppressed by the transcription factor heterodimers E2F/DP1 and  
240 MGA/MAX, which respectively bind E2F and E-box motifs<sup>18,50,51,54,55</sup>. To elucidate if KDM5C is recruited to  
241 germline gene promoters by a similar mechanism, we used HOMER to identify transcription factor motifs  
242 enriched at KDM5C-bound or unbound germline gene promoters<sup>56</sup> (TSS +/- 500 bp, q-value < 0.1). MAX  
243 and E2F6 binding sites were significantly enriched at germline genes bound by KDM5C in EpiLCs, but  
244 not at germline genes unbound by KDM5C (MAX q-value: 0.0068, E2F6 q-value: 0.0673, E2F q-value:  
245 0.0917) (Figure 5F). One third of KDM5C-bound promoters contained the consensus sequence for either  
246 E2F6 (E2F, 5'-TCCCGC-3'), MGA (E-box, 5'-CACGTG-3'), or both, but only 17% of KDM5C-unbound genes  
247 contained these motifs (Figure 5G). KDM5C-unbound germline genes were intstead enriched for multiple  
248 RFX transcription factor binding sites (RFX q-value < 0.0001, RFX2 q-value < 0.0001, RFX5 q-value < 0.0001)  
249 (Figure 5H). RFX transcription factors bind X-box motifs<sup>57</sup> to promote ciliogenesis<sup>58,59</sup> and among them is  
250 RFX2, a central regulator of post-meiotic spermatogeneis<sup>60,61</sup>. Interestingly, RFX2 mRNA is derepressed in  
251 *Kdm5c*-KO EpiLCs (Figure 5I), however it is not a direct target of KDM5C (Supplementary figure XXX). Thus,  
252 RFX2 is a candidate transcription factor for driving the ectopic expression of KDM5C-unbound germline  
253 genes in *Kdm5c*-KO cells.

254 **KDM5C promotes *de novo* DNA methylation at germline genes (if there are differ-  
255 ences, say at germline gene CpG islands)**

256 In the early embryo, germline gene promoters are initially decorated with repressive histone modifications  
257 and are then silenced long-term via DNA CpG methylation (CpGme)<sup>16,17,44,62</sup>. Our results above indicate  
258 KDM5C also acts at germline gene promoters during this time period. However, how KDM5C interacts with  
259 other germline gene silencing mechanisms is currently unclear. KDM5C is generally thought to supress  
260 transcription through erasure of histone 3 lysine 4 di- and trimethylation (H3K4me2/3)<sup>9</sup>, yet KDM5C's  
261 catalytic activity was recently shown to be dispensible for suppressing *Dazl* in undifferentiated ESCs<sup>49</sup>. Since  
262 H3K4me3 impedes *de novo* CpGme placement<sup>63,64</sup>, KDM5C's catalytic activity may instead be required later  
263 in development for long-term silencing of germline genes. In support of this, CpGme is markedly reduced at  
264 two germline gene promoters in the *Kdm5c*-KO adult hippocampus<sup>8</sup>.

265 Based on the above observations, we hypothesized KDM5C erases H3K4me3 to promote the initial  
266 placement of CpGme at germline gene promoters in EpiLCs. To test this hypothesis, we first characterized

267 KDM5C's substrates, histone 3 lysine 4 di- and trimethylation (H3K4me2/3) in our previously published ChIP-  
268 seq datasets of the wild type and *Kdm5c*-KO amygdala<sup>25</sup> and EpiLCs<sup>45</sup>. In congruence with previous work in  
269 the *Kdm5c*-KO hippocampus<sup>8</sup>, we observed aberrant accumulation of H3K4me3 around the transcription  
270 start site (TSS) of germline genes in the *Kdm5c*-KO amygdala (Figure 6A). We additionally found a marked  
271 increase in H3K4me2 around the TSS of germline genes in *Kdm5c*-KO EpiLCs (Figure 6B).

272 We then evaluated KDM5C's role in germline gene silencing during ESC to EpiLC differentiation. We first  
273 characterized KDM5C's mRNA and protein expression at 0, 24, and 48 hours of differentiation into EpiLCs  
274 (Figure 6C). While *Kdm5c* mRNA steadily decreased from 0 to 48 hours of differentiation (Figure 6D), KDM5C  
275 protein initially increased from 0 to 24 hours but then decreased to near knockout levels by 48 hours (Figure  
276 6E).

277 To determine KDM5C's role in the initial placement of DNA methylation at germline gene promoters,  
278 we performed whole genome bisulfite sequencing (WGBS) in wild-type and *Kdm5c*-KO ESCs and 96 hour  
279 extended EpiLCs (exEpiLCs).

- 280 • Germline genes are known to accumulate CpGme at (CGIs) during the transition from naive to primed  
281 pluripotency.
- 282 • We first identified the number of germline genes with CGIs neighboring their promoter.
  - 283 – Found XXX% of germline genes had CGIs, XXX% of which were direct KDM5C targets in EpiLCs  
284 (Eulerr).
- 285 • We then curated a list of germline genes that significantly gained CpGme in wild-type exEpiLCs  
286 compared to ESCs.
  - 287 – Example gene bedgraph
  - 288 – Majority of CGI germline genes?
- 289 • Out of the CGI genes, which had significantly reduced CpGme in 5CKO (heatmap of % methylation)
  - 290 – Highlight interesting genes affected vs unaffected by KDM5C (especially if same/different from  
291 E2F6, PRC1.6, Setdb1 targets)
  - 292 – CGIs that never gain DNAme in WT
  - 293 – Although wild-type cells accumulated high levels of DNA methylation at germline gene promoters  
294 over the course of ESCs to exEpiLC differentiation, DNA methylation was markedly reduced in  
295 *Kdm5c*-KO exEpiLCs (Figure 6F).
- 296 • Non-CGI germline genes, any significant changes at their promoter? (Heatmap again?)
- 297 • Altogether, this suggests KDM5C is dynamically regulated during ESC to EpiLC differentiation and  
298 promotes germline gene silencing via DNA methylation during early embryogenesis.

299 **Discussion**

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

317 It is important to note that while we highlighted KDM5C's regulation of germline genes, some germline-  
318 enriched genes are also expressed at the 2-cell stage and in naïve ESCs for their role in pluripotency and  
319 self-renewal. For example, *Dazl* is primarily known for committing PGCs to the germline fate and regulating  
320 the translation of germline-specific mRNAs, but is also expressed at the 2-cell stage<sup>65</sup>, in naïve ESCs<sup>41</sup>, and  
321 in the inner cell mass<sup>41</sup>. KDM5C may therefore negatively regulate totipotency, since KDM5C suppresses  
322 *Dazl* in ESCs<sup>49,66</sup> and *Kdm5c*-KO ESCs aberrantly express 2-cell-specific genes like *Zscan4c*<sup>49</sup>. However,  
323 *Kdm5c*-KO misexpression of *Dazl* was independent of the 2-cell specific transcription factor *Dux*<sup>49</sup> and,  
324 unlike *Dazl*, KDM5C does not bind to the *Dux* promoter. Although expressed in naïve ESCs, *Dazl* and  
325 other "self-renewal" germline genes are silenced during differentiation into EpiLCs<sup>17</sup>. We found *Kdm5c*-KO  
326 EpiLCs expressed *Dazl* but did not express 2-cell specific genes, suggesting the 2-cell-like state observed  
327 in *Kdm5c*-KO ESCs is an indirect consequence of KDM5C's primary role in germline gene repression.  
328 Misexpression of germline genes in *Kdm5c*-KO EpiLCs could indicate they are differentiating into primordial  
329 germ cell-like cells (PGCLCs)<sup>36,37,39</sup>. Yet, *Kdm5c*-KO EpiLCs had normal cellular morphology and properly  
330 expressed markers for primed pluripotency, including *Otx2* which is known to repress EpiLC differentiation  
331 into PGCs/PGCLCs<sup>67</sup>. Altogether, these results demonstrate KDM5C represses germline identity during  
332 embryogenesis and *Kdm5c*-KO germline gene misexpression occurs ectopically in conjunction with typical  
333 developmental programs.

334     Although eggs and sperm initially express very similar sets of germline genes for PGC formation and  
335 meiosis, some germline genes are sexually dimorphic. We found *Kdm5c* mutant males and females  
336 expressed both sperm and egg-biased genes, indicating the mechanism of repression is independent of  
337 a given germline gene's sex. However, organismal sex did greatly influence the manifestation of germline  
338 gene dysregulation, as female *Kdm5c*-KO EpiLCs had over double the number of germline-enriched DEGs  
339 compared to males. X chromosome inactivation (XCI) defects could explain why homozygous *Kdm5c*  
340 knockout females are more prone to germline gene dysregulation, given the X chromosome harbors a large  
341 number of spermatogenesis genes<sup>52,53</sup> and loss of *Kdm5c* impairs XCI<sup>45</sup>. However, female germline DEGs  
342 were not significantly biased towards the X chromosome. Sex differences in germline gene suppression may  
343 be instead connected to females having a higher dose of KDM5C than males, due to its partial escape from  
344 XCI<sup>20-23</sup>. Intriguingly, females with heterozygous loss of *Kdm5c* also had over double the number of germline  
345 DEGs than males, even though their level of KDM5C should be roughly equivalent to that of wild-type males.  
346 Altogether, these results suggest female EpiLCs are more prone to transitioning to a germ cell-like state  
347 than males and require more KDM5C to maintain somatic cellular identity. Future studies are required to  
348 illuminate if this phenomenon is unique to sexually dimorphic chromatin regulators or a general feature of  
349 female cells.

350     It is yet unknown how KDM5C is recruited to germline gene promoters, given that KDM5C itself does  
351 not contain domains for sequence-specific binding. In HeLa cells and ESCs<sup>49,68</sup>, KDM5C associates with  
352 members of the polycomb repressive complex 1.6 (PRC1.6), which is recruited to germline gene promoters  
353 through E2F6/DP1 and MGA/MAX heterodimers<sup>16,55</sup>. While MAX and E2F6 motifs were enriched at KDM5C-  
354 bound germline promoters in EpiLCs, only about one-third of promoters contained their consensus sequence.  
355 Additionally, some germline genes can become active in *Kdm5c*-KO cells independent of KDM5C binding to  
356 their promoters. Germline genes unbound by KDM5C were primarily late-stage spermatogenesis genes and  
357 significantly enriched for RFX transcription factors, including RFX2, a central regulator of spermiogenesis<sup>60,61</sup>.  
358 Another notable KDM5C-unbound DEG is *Stra8*, a meiotic transcription factor that is turned on in germ  
359 cells with retionic acid signaling and DAZL expression<sup>69,70</sup>. Together, this suggests once germline genes  
360 like *Dazl* and *Rfx2* are activated in *Kdm5c*-KO cells, they can then turn on other ectopic germline programs  
361 downstream and loosely mimic germ cell development.

362     Unlike the previously characterized germline gene suppressors that place repressive histone  
363 modifications<sup>16,17,19,62,71-73</sup>, KDM5C removes the active histone mark histone 3 lysine 4 di- and trimethylation  
364 (H3K4me2/3)<sup>9</sup>. Yet, KDM5C's catalytic activity may not be required for germline gene silencing, as it  
365 was recently found to be dispensable for repressing *Dazl* in ESCs<sup>49</sup>. Emerging work indicates many  
366 histone-modifying enzymes have non-catalytic functions that can influence gene expression even more  
367 potently as their catalytic roles<sup>74,75</sup>. Although dispensable in ESCs, KDM5C's catalytic activity could be  
368 necessary to silence germline genes in EpiLCs, as H3K4me3 impedes *de novo* CpGme placement<sup>63,64</sup>. In  
369 support of this, CpGme is significantly eroded at at least two germline promoters in the adult *Kdm5c*-KO

370 hippocampus<sup>8</sup>. Additionally, we found *Kdm5c*-KOs accumulate H3K4me2/3 around the TSS of germline  
371 genes and impaired CpGme placement at CpG islands. KDM5C's shifting mode of germline gene repression  
372 as cells transition from naïve to primed pluripotency indicates chromatin regulators can employ both catalytic  
373 and non-catalytic repressive mechanisms at the same loci depending upon the developmental and cellular  
374 context.

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

391 The above work provides the mechanistic foundation for KDM5C's role in germline gene repression  
392 and offers novel insight into how dysregulation in *Kdm5c*-KO tissue identity changes over the course  
393 of development. However, the contribution of these ectopic, tissue-specific genes towards *Kdm5c*-KO  
394 neurological impairments is still unknown. In addition to germline genes, we also identified significant  
395 enrichment of muscle, liver, and even ovary-biased transcripts within the male *Kdm5c*-KO brain. Intriguingly,  
396 select liver and muscle-biased DEGs do have known roles within the brain, such as the liver-enriched lipid  
397 metabolism gene *Apolipoprotein C-I* (*Apoc1*)<sup>30</sup> that can impair learning and memory when overexpressed  
398 in the mouse brain<sup>76</sup> and is implicated in Alzheimer's disease in humans<sup>31</sup>. Thus, KDM5C may be crucial  
399 for neurodevelopment by fine-tuning the expression of tissue-enriched, dosage-sensitive genes like *Apoc1*.  
400 Given germline genes have no known functions within the brain, their impact upon neurodevelopment is  
401 currently unknown. Ectopic germline transcripts have been observed in a variety of cancers<sup>77,78</sup> and can  
402 drive brain tumor formation in *Drosophila*<sup>79</sup>, indicating they may promote genome instability and cellular  
403 de-differentiation. Intriguingly, select models of other chromatin-linked neurodevelopmental disorders also  
404 display impaired soma-germline demarcation<sup>7,80-83</sup>. Like KDM5C, the chromatin regulators underlying these

405 conditions - DNA methyltransferase 3b (DNMT3B), H3K9me1/2 methyltransferases G9A/GLP, methyl-CpG  
406 -binding protein 2 (MECP2) - primarily silence gene expression. Thus, KDM5C is among a growing cohort of  
407 chromatin-linked neurodevelopmental disorders that have a similar underlying phenotype of germline versus  
408 soma dysregulation. However, further research is required to determine the impact of these germline genes  
409 and the extent to which this phenomenon occurs in humans.

## 410 Materials and Methods

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

412 Tissue-enriched differentially expressed genes (DEGs) were determined by their classification in a previ-  
413 ously published dataset from 17 male and female mouse tissues<sup>24</sup>. This study defined tissue expression as  
414 greater than 1 Fragments Per Kilobase of transcript per Million mapped read (FPKM) and tissue enrichment  
415 as at least 4-fold higher expression than any other tissue.

416 We curated a list of germline-enriched genes using an RNA-seq dataset from wild-type and germline-  
417 depleted (Kit<sup>W/Wv</sup>) male and female mouse embryos from embryonic day 12, 14, and 16<sup>35</sup>, as well as adult  
418 male testes<sup>32</sup>. Germline-enriched genes met the following criteria: 1) their expression is greater than 1  
419 FPKM in wild-type germline 2) their expression in any wild-type somatic tissues<sup>24</sup> does not exceed 20%  
420 of maximum expression in wild-type germline, and 3) their expression in the germ cell-depleted (Kit<sup>W/Wv</sup>)  
421 germline, for any sex or time point, does not exceed 20% of maximum expression in wild-type germline.

### 422 Cell culture

423 We utilized our previously established cultures of male wild-type and *Kdm5c* knockout (-KO) embryonic  
424 stem cells<sup>45</sup>. Sex was confirmed by genotyping *Uba1/Uba1y* on the X and Y chromosomes with the following  
425 primers: 5'-TGGATGGTGTGCCAATG-3', 5'-CACCTGCACGTTGCCCTT-3'. Deletion of *Kdm5c* was  
426 confirmed through the primers 5'-ATGCCCATATTAAAGAGTCCCTG-3', 5'-TCTGCCTTGATGGGACTGTT-3',  
427 and 5'-GGTTCTCAACACTCACATAGTG-3'.

428 Embryonic stem cells (ESCs) and epiblast-like cells were cultured using previously established  
429 methods<sup>40</sup>. Briefly, ESCs were initially cultured in primed ESC (pESC) media consisting of KnockOut  
430 DMEM (Gibco#10829-018), fetal bovine serum (Gibco#A5209501), KnockOut serum replacement  
431 (Invitrogen#10828-028), Glutamax (Gibco#35050-061), Anti-Anti (Gibco#15240-062), MEM Non-essential  
432 amino acids (Gibco#11140-050), and beta-mercaptoethanol (Sigma#M7522). They were then transitioned  
433 into ground-state, “naïve” ESCs (nESCs) by culturing for four passages in N2B27 media containing  
434 DMEM/F12 (Gibco#11330-032), Neurobasal media (Gibco#21103-049), Gluamax, Anti-Anti, N2 sup-  
435 plement (Invitrogen#17502048), and B27 supplement without vitamin A (Invitrogen#12587-010), and  
436 beta-mercaptoethanol. Both pESC and nESC media were supplemented with 3 μM GSK3 inhibitor

437 CHIR99021 (Sigma #SML1046-5MG), 1  $\mu$ M MEK inhibitor PD0325901 (Sigma #PZ0162-5MG), and 1,000  
438 units/mL leukemia inhibitory factor (LIF, Millipore#ESG1107).

439 nESCs were differentiated into epiblast-like cells (EpiLCs, 48 hours) and extendend EpiLCs (exEpiLCs,  
440 96 hours) by culturing in N2B27 media containing DMEM/F12, Neurobasal media, Gluamax, Anti-Anti,  
441 N2 supplement, B27 supplement (Invitrogen#17504044), beta-mercaptoethanol, fibroblast growth factor 2  
442 (FGF2, R&D Biotechne 233-FB), and activin A (R&D Biotechne 338AC050CF), as previously described<sup>40</sup>.

#### 443 **Immunocytochemistry (ICC)**

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

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

454 After ensuring read quality via FastQC (v0.11.8), reads were then mapped to the mm10 *Mus musculus*  
455 genome (Gencode) using STAR (v2.5.3a), during which we removed duplicates and kept only uniquely  
456 mapped reads. Count files were generated by FeatureCounts (Subread v1.5.0), and BAM files were  
457 converted to bigwigs using deeptools (v3.1.3) and visualized by the UCSC genome browser. RStudio (v3.6.0)  
458 was then used to analyze counts files by DESeq2 (v1.26.0)<sup>26</sup> to identify differentially expressed genes  
459 (DEGs) with a q-value (p-adjusted via FDR/Benjamini–Hochberg correction) less than 0.1 and a log2 fold  
460 change greater than 0.5. For all DESeq2 analyses, log2 fold changes were calculated with IfcShrink using  
461 the ashr package<sup>84</sup>. MA-plots were generated by ggpubr (v0.6.0), and Eulerr diagrams were generated by  
462 eulerr (v6.1.1). Boxplots and scatterplots were generated by ggpubr (v0.6.0) and ggplot2 (v3.3.2). The Upset  
463 plot was generated via the package UpSetR (v1.4.0)<sup>85</sup>. Gene ontology (GO) analyses were performed by  
464 the R package enrichPlot (v1.16.2) using the biological processes setting and compareCluster.

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

466 ChIP-seq reads were aligned to mm10 using Bowtie1 (v1.1.2) allowing up to two mismatches. Only  
467 uniquely mapped reads were used for analysis. Peaks were called using MACS2 software (v2.2.9.1) using  
468 input BAM files for normalization, with filters for a q-value < 0.1 and a fold enrichment > 1. We removed

469 “black-listed” genomic regions that often give aberrant signals. Common peak sets were obtained in R via  
470 DiffBind (v3.6.5). In the case of KDM5C ChIP-seq, *Kdm5c*-KO peaks were then subtracted from wild-type  
471 samples using bedtools (v2.25.0). Peak proximity to genome annotations was determined by ChIPSeeker  
472 (v1.32.1). Gene ontology (GO) analyses were performed by the R package enrichPlot (v1.16.2) using the  
473 biological processes setting and compareCluster. Enriched motifs were identified using HOMER<sup>56</sup>. Average  
474 binding across the genome was visualized using deeptools (v3.1.3). Bigwigs were visualized using the  
475 UCSC genome browser.

## 476 **Whole genome bisulfite sequencing (WGBS)**

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

## 481 **Data availability**

### 482 **Published datasets**

483 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  
484 adult amygdala and hippocampus<sup>25</sup> (available at GEO: GSE127722) and male wild-type and *Kdm5c*-KO  
485 EpiLCs<sup>45</sup> (available at GEO: GSE96797).

486 Previously published ChIP-seq experiments included KDM5C in wild-type and *Kdm5c*-KO EpiLCs<sup>45</sup> (avail-  
487 able at GEO: GSE96797) and mouse primary neuron cultures (PNCs) from the cortex and hippocampus<sup>13</sup>  
488 (available at GEO: GSE61036). ChIP-seq of histone 3 lysine 4 dimethylation in male wild-type and *Kdm5c*-KO  
489 EpiLCs<sup>45</sup> is also available at GEO: GSE96797. ChIP-seq of histone 3 lysine 4 trimethylation in wild-type and  
490 *Kdm5c*-KO male amygdala<sup>25</sup> are available at GEO: GSE127817.

### 492 **Data analysis**

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

## 495 **Acknowledgements**

496 • Jacob Mueller for providing insight in germline gene regulation.

- 497     • Sundeep Kalantry for providing reagents and expertise in culturing mouse embryonic stem cells and  
498        epiblast-like cells  
499     • Ilakkiya  
500     • Funding sources

501 **References**

- 502 1. Strahl, B.D., and Allis, C.D. (2000). The language of covalent histone modifications. *Nature* **403**,  
503     41–45. <https://doi.org/10.1038/47412>.
- 504 2. Jenuwein, T., and Allis, C.D. (2001). Translating the Histone Code. *Science* **293**, 1074–1080.  
505     <https://doi.org/10.1126/science.1063127>.
- 506 3. Lewis, E.B. (1978). A gene complex controlling segmentation in *Drosophila*. *Nature* **276**, 565–570.  
507     <https://doi.org/10.1038/276565a0>.
- 508 4. Kennison, J.A., and Tamkun, J.W. (1988). Dosage-dependent modifiers of polycomb and antennapedia  
509        mutations in *Drosophila*. *Proc. Natl. Acad. Sci. U.S.A.* **85**, 8136–8140. <https://doi.org/10.1073/pnas.8>  
510        5.21.8136.
- 511 5. Kassis, J.A., Kennison, J.A., and Tamkun, J.W. (2017). Polycomb and Trithorax Group Genes in  
512        *Drosophila*. *Genetics* **206**, 1699–1725. <https://doi.org/10.1534/genetics.115.185116>.
- 513 6. Gabriele, M., Lopez Tobon, A., D'Agostino, G., and Testa, G. (2018). The chromatin basis of  
514        neurodevelopmental disorders: Rethinking dysfunction along the molecular and temporal axes. *Prog  
515        Neuropsychopharmacol Biol Psychiatry* **84**, 306–327. <https://doi.org/10.1016/j.pnpbp.2017.12.013>.
- 516 7. Schaefer, A., Sampath, S.C., Intrator, A., Min, A., Gertler, T.S., Surmeier, D.J., Tarakhovsky, A.,  
517        and Greengard, P. (2009). Control of cognition and adaptive behavior by the GLP/G9a epigenetic  
518        suppressor complex. *Neuron* **64**, 678–691. <https://doi.org/10.1016/j.neuron.2009.11.019>.
- 519 8. Scandaglia, M., Lopez-Atalaya, J.P., Medrano-Fernandez, A., Lopez-Cascales, M.T., Del Blanco, B.,  
520        Lipinski, M., Benito, E., Olivares, R., Iwase, S., Shi, Y., et al. (2017). Loss of Kdm5c Causes Spurious  
521        Transcription and Prevents the Fine-Tuning of Activity-Regulated Enhancers in Neurons. *Cell Rep* **21**,  
522        47–59. <https://doi.org/10.1016/j.celrep.2017.09.014>.
- 523 9. Iwase, S., Lan, F., Bayliss, P., De La Torre-Ubieta, L., Huarte, M., Qi, H.H., Whetstine, J.R., Bonni, A.,  
524        Roberts, T.M., and Shi, Y. (2007). The X-Linked Mental Retardation Gene SMCX/JARID1C Defines a  
525        Family of Histone H3 Lysine 4 Demethylases. *Cell* **128**, 1077–1088. <https://doi.org/10.1016/j.cell.200>  
526        7.02.017.
- 527 10. Claes, S., Devriendt, K., Van Goethem, G., Roelen, L., Meireleire, J., Raeymaekers, P., Cassiman,  
528        J.J., and Fryns, J.P. (2000). Novel syndromic form of X-linked complicated spastic paraparesia. *Am J  
529        Med Genet* **94**, 1–4.

- 521
- 522 11. 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>.
- 523
- 524 12. Carmignac, V., Nambot, S., Lehalle, D., Callier, P., Moortgat, S., Benoit, V., Ghoumid, J., Delobel,  
B., Smol, T., Thuillier, C., et al. (2020). Further delineation of the female phenotype with KDM5C  
disease causing variants: 19 new individuals and review of the literature. *Clin Genet* *98*, 43–55.  
<https://doi.org/10.1111/cge.13755>.
- 525
- 526 13. Iwase, S., Brookes, E., Agarwal, S., Badeaux, A.I., Ito, H., Vallianatos, C.N., Tomassy, G.S., Kasza, T.,  
Lin, G., Thompson, A., et al. (2016). A Mouse Model of X-linked Intellectual Disability Associated with  
Impaired Removal of Histone Methylation. *Cell Reports* *14*, 1000–1009. <https://doi.org/10.1016/j.celrep.2015.12.091>.
- 527
- 528 14. Devlin, D.K., Ganley, A.R.D., and Takeuchi, N. (2023). A pan-metazoan view of germline-soma  
distinction challenges our understanding of how the metazoan germline evolves. *Current Opinion in  
Systems Biology* *36*, 100486. <https://doi.org/10.1016/j.coisb.2023.100486>.
- 529
- 530 15. Lehmann, R. (2012). Germline Stem Cells: Origin and Destiny. *Cell Stem Cell* *10*, 729–739.  
<https://doi.org/10.1016/j.stem.2012.05.016>.
- 531
- 532 16. Endoh, M., Endo, T.A., Shinga, J., Hayashi, K., Farcas, A., Ma, K.W., Ito, S., Sharif, J., Endoh, T.,  
Onaga, N., et al. (2017). PCGF6-PRC1 suppresses premature differentiation of mouse embryonic  
stem cells by regulating germ cell-related genes. *Elife* *6*. <https://doi.org/10.7554/eLife.21064>.
- 533
- 534 17. Mochizuki, K., Sharif, J., Shirane, K., Uranishi, K., Bogutz, A.B., Janssen, S.M., Suzuki, A., Okuda,  
A., Koseki, H., and Lorincz, M.C. (2021). Repression of germline genes by PRC1.6 and SETDB1  
in the early embryo precedes DNA methylation-mediated silencing. *Nat Commun* *12*, 7020. <https://doi.org/10.1038/s41467-021-27345-x>.
- 535
- 536 18. Velasco, G., Hubé, F., Rollin, J., Neuillet, D., Philippe, C., Bouzinba-Segard, H., Galvani, A., Viegas-  
Péquignot, E., and Francastel, C. (2010). Dnmt3b recruitment through E2F6 transcriptional repressor  
mediates germ-line gene silencing in murine somatic tissues. *Proc Natl Acad Sci U S A* *107*, 9281–  
9286. <https://doi.org/10.1073/pnas.1000473107>.
- 537
- 538 19. Hackett, J.A., Reddington, J.P., Nestor, C.E., Dunican, D.S., Branco, M.R., Reichmann, J., Reik,  
W., Surani, M.A., Adams, I.R., and Meehan, R.R. (2012). Promoter DNA methylation couples  
genome-defence mechanisms to epigenetic reprogramming in the mouse germline. *Development*  
*139*, 3623–3632. <https://doi.org/10.1242/dev.081661>.
- 539

- 540 20. Agulnik, A.I., Mitchell, M.J., Mattei, M.G., Borsani, G., Avner, P.A., Lerner, J.L., and Bishop, C.E.  
541 (1994). A novel X gene with a widely transcribed Y-linked homologue escapes X-inactivation in mouse  
and human. *Hum Mol Genet* 3, 879–884. <https://doi.org/10.1093/hmg/3.6.879>.
- 542 21. Carrel, L., Hunt, P.A., and Willard, H.F. (1996). Tissue and lineage-specific variation in inactive  
543 X chromosome expression of the murine Smcx gene. *Hum Mol Genet* 5, 1361–1366. <https://doi.org/10.1093/hmg/5.9.1361>.
- 544 22. Sheardown, S., Norris, D., Fisher, A., and Brockdorff, N. (1996). The mouse Smcx gene exhibits  
545 developmental and tissue specific variation in degree of escape from X inactivation. *Hum Mol Genet*  
5, 1355–1360. <https://doi.org/10.1093/hmg/5.9.1355>.
- 546 23. Xu, J., Deng, X., and Disteche, C.M. (2008). Sex-Specific Expression of the X-Linked Histone  
Demethylase Gene Jarid1c in Brain. *PLoS ONE* 3, e2553. <https://doi.org/10.1371/journal.pone.0002553>.
- 547 24. Li, B., Qing, T., Zhu, J., Wen, Z., Yu, Y., Fukumura, R., Zheng, Y., Gondo, Y., and Shi, L. (2017). A  
Comprehensive Mouse Transcriptomic BodyMap across 17 Tissues by RNA-seq. *Sci Rep* 7, 4200.  
549 <https://doi.org/10.1038/s41598-017-04520-z>.
- 550 25. Vallianatos, C.N., Raines, B., Porter, R.S., Bonefas, K.M., Wu, M.C., Garay, P.M., Collette, K.M.,  
Seo, Y.A., Dou, Y., Keegan, C.E., et al. (2020). Mutually suppressive roles of KMT2A and KDM5C  
in behaviour, neuronal structure, and histone H3K4 methylation. *Commun Biol* 3, 278. <https://doi.org/10.1038/s42003-020-1001-6>.
- 552 26. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for  
553 RNA-seq data with DESeq2. *Genome Biol* 15, 550. <https://doi.org/10.1186/s13059-014-0550-8>.
- 554 27. Crackower, M.A., Kolas, N.K., Noguchi, J., Sarao, R., Kikuchi, K., Kaneko, H., Kobayashi, E., Kawai, Y.,  
Kozieradzki, I., Landers, R., et al. (2003). Essential Role of Fkbp6 in Male Fertility and Homologous  
555 Chromosome Pairing in Meiosis. *Science* 300, 1291–1295. <https://doi.org/10.1126/science.1083022>.
- 556 28. Xiòl, J., Corà, E., Koglgruber, R., Chuma, S., Subramanian, S., Hosokawa, M., Reuter, M., Yang, Z.,  
Berninger, P., Palencia, A., et al. (2012). A Role for Fkbp6 and the Chaperone Machinery in piRNA  
Amplification and Transposon Silencing. *Molecular Cell* 47, 970–979. <https://doi.org/10.1016/j.molcel.2012.07.019>.
- 558 29. Cheng, S., Altmeppen, G., So, C., Welp, L.M., Penir, S., Ruhwedel, T., Menelaou, K., Harasimov, K.,  
Stützer, A., Blayney, M., et al. (2022). Mammalian oocytes store mRNAs in a mitochondria-associated  
559 membraneless compartment. *Science* 378, eabq4835. <https://doi.org/10.1126/science.abq4835>.
- 560 30. Rouland, A., Masson, D., Lagrost, L., Vergès, B., Gautier, T., and Bouillet, B. (2022). Role of  
apolipoprotein C1 in lipoprotein metabolism, atherosclerosis and diabetes: A systematic review.  
561 *Cardiovasc Diabetol* 21, 272. <https://doi.org/10.1186/s12933-022-01703-5>.

- 562 31. Leduc, V., Jasmin-Bélanger, S., and Poirier, J. (2010). APOE and cholesterol homeostasis in  
Alzheimer's disease. *Trends in Molecular Medicine* *16*, 469–477. <https://doi.org/10.1016/j.molmed.2010.07.008>.
- 563 32. Mueller, J.L., Skaletsky, H., Brown, L.G., Zaghlul, S., Rock, S., Graves, T., Auger, K., Warren,  
W.C., Wilson, R.K., and Page, D.C. (2013). Independent specialization of the human and mouse X  
chromosomes for the male germ line. *Nat Genet* *45*, 1083–1087. <https://doi.org/10.1038/ng.2705>.
- 564 33. Handel, M.A., and Eppig, J.J. (1979). Sertoli Cell Differentiation in the Testes of Mice Genetically  
Deficient in Germ Cells. *Biology of Reproduction* *20*, 1031–1038. <https://doi.org/10.1095/biolreprod20.5.1031>.
- 565 34. Green, C.D., Ma, Q., Manske, G.L., Shami, A.N., Zheng, X., Marini, S., Moritz, L., Sultan, C.,  
Gurczynski, S.J., Moore, B.B., et al. (2018). A Comprehensive Roadmap of Murine Spermatogenesis  
Defined by Single-Cell RNA-Seq. *Dev Cell* *46*, 651–667.e10. <https://doi.org/10.1016/j.devcel.2018.07.025>.
- 566 35. Soh, Y.Q., Junker, J.P., Gill, M.E., Mueller, J.L., van Oudenaarden, A., and Page, D.C. (2015). A  
Gene Regulatory Program for Meiotic Prophase in the Fetal Ovary. *PLoS Genet* *11*, e1005531.  
<https://doi.org/10.1371/journal.pgen.1005531>.
- 567 36. Magnúsdóttir, E., and Surani, M.A. (2014). How to make a primordial germ cell. *Development* *141*,  
245–252. <https://doi.org/10.1242/dev.098269>.
- 568 37. Günesdogan, U., Magnúsdóttir, E., and Surani, M.A. (2014). Primordial germ cell specification: A  
context-dependent cellular differentiation event [corrected]. *Philos Trans R Soc Lond B Biol Sci* *369*.  
<https://doi.org/10.1098/rstb.2013.0543>.
- 569 38. Bardot, E.S., and Hadjantonakis, A.-K. (2020). Mouse gastrulation: Coordination of tissue patterning,  
specification and diversification of cell fate. *Mechanisms of Development* *163*, 103617. <https://doi.org/10.1016/j.mod.2020.103617>.
- 570 39. Hayashi, K., Ohta, H., Kurimoto, K., Aramaki, S., and Saitou, M. (2011). Reconstitution of the  
mouse germ cell specification pathway in culture by pluripotent stem cells. *Cell* *146*, 519–532.  
<https://doi.org/10.1016/j.cell.2011.06.052>.
- 571 40. Samanta, M., and Kalantry, S. (2020). Generating primed pluripotent epiblast stem cells: A methodol-  
ogy chapter. *Curr Top Dev Biol* *138*, 139–174. <https://doi.org/10.1016/bs.ctdb.2020.01.005>.
- 572 41. Welling, M., Chen, H., Muñoz, J., Musheev, M.U., Kester, L., Junker, J.P., Mischerikow, N., Arbab, M.,  
Kuijk, E., Silberstein, L., et al. (2015). DAZL regulates Tet1 translation in murine embryonic stem cells.  
*EMBO Reports* *16*, 791–802. <https://doi.org/10.15252/embr.201540538>.
- 573 42. Macfarlan, T.S., Gifford, W.D., Driscoll, S., Lettieri, K., Rowe, H.M., Bonanomi, D., Firth, A., Singer, O.,  
Trono, D., and Pfaff, S.L. (2012). Embryonic stem cell potency fluctuates with endogenous retrovirus  
activity. *Nature* *487*, 57–63. <https://doi.org/10.1038/nature11244>.

- 585
- 586 43. Suzuki, A., Hirasaki, M., Hishida, T., Wu, J., Okamura, D., Ueda, A., Nishimoto, M., Nakachi, Y., Mizuno, Y., Okazaki, Y., et al. (2016). Loss of MAX results in meiotic entry in mouse embryonic and germline stem cells. *Nat Commun* 7, 11056. <https://doi.org/10.1038/ncomms11056>.
- 587
- 588 44. Borgel, J., Guibert, S., Li, Y., Chiba, H., Schübeler, D., Sasaki, H., Forné, T., and Weber, M. (2010). Targets and dynamics of promoter DNA methylation during early mouse development. *Nat Genet* 42, 1093–1100. <https://doi.org/10.1038/ng.708>.
- 589
- 590 45. Samanta, M.K., Gayen, S., Harris, C., Maclary, E., Murata-Nakamura, Y., Malcore, R.M., Porter, R.S., Garay, P.M., Vallianatos, C.N., Samollow, P.B., et al. (2022). Activation of Xist by an evolutionarily conserved function of KDM5C demethylase. *Nat Commun* 13, 2602. <https://doi.org/10.1038/s41467-022-30352-1>.
- 591
- 592 46. Koubova, J., Menke, D.B., Zhou, Q., Capel, B., Griswold, M.D., and Page, D.C. (2006). Retinoic acid regulates sex-specific timing of meiotic initiation in mice. *Proc. Natl. Acad. Sci. U.S.A.* 103, 2474–2479. <https://doi.org/10.1073/pnas.0510813103>.
- 593
- 594 47. Lin, Y., Gill, M.E., Koubova, J., and Page, D.C. (2008). Germ Cell-Intrinsic and -Extrinsic Factors Govern Meiotic Initiation in Mouse Embryos. *Science* 322, 1685–1687. <https://doi.org/10.1126/science.1166340>.
- 595
- 596 48. Endo, T., Mikedis, M.M., Nicholls, P.K., Page, D.C., and De Rooij, D.G. (2019). Retinoic Acid and Germ Cell Development in the Ovary and Testis. *Biomolecules* 9, 775. <https://doi.org/10.3390/biom9120775>.
- 597
- 598 49. Gupta, N., Yakhou, L., Albert, J.R., Azogui, A., Ferry, L., Kirsh, O., Miura, F., Battault, S., Yamaguchi, K., Laisné, M., et al. (2023). A genome-wide screen reveals new regulators of the 2-cell-like cell state. *Nat Struct Mol Biol*. <https://doi.org/10.1038/s41594-023-01038-z>.
- 599
- 600 50. Pohlers, M., Truss, M., Frede, U., Scholz, A., Strehle, M., Kuban, R.-J., Hoffmann, B., Morkel, M., Birchmeier, C., and Hagemeier, C. (2005). A Role for E2F6 in the Restriction of Male-Germ-Cell-Specific Gene Expression. *Current Biology* 15, 1051–1057. <https://doi.org/10.1016/j.cub.2005.04.060>.
- 601
- 602 51. Dahlet, T., Truss, M., Frede, U., Al Adhami, H., Bardet, A.F., Dumas, M., Vallet, J., Chicher, J., Hammann, P., Kottnik, S., et al. (2021). E2F6 initiates stable epigenetic silencing of germline genes during embryonic development. *Nat Commun* 12, 3582. <https://doi.org/10.1038/s41467-021-23596-w>.
- 603
- 604 52. Wang, P.J., McCarrey, J.R., Yang, F., and Page, D.C. (2001). An abundance of X-linked genes expressed in spermatogonia. *Nat Genet* 27, 422–426. <https://doi.org/10.1038/86927>.
- 605
- 606 53. Khil, P.P., Smirnova, N.A., Romanienko, P.J., and Camerini-Otero, R.D. (2004). The mouse X chromosome is enriched for sex-biased genes not subject to selection by meiotic sex chromosome inactivation. *Nat Genet* 36, 642–646. <https://doi.org/10.1038/ng1368>.
- 607

- 608 54. Hurlin, P.J. (1999). Mga, a dual-specificity transcription factor that interacts with Max and contains a  
609 T-domain DNA-binding motif. *The EMBO Journal* *18*, 7019–7028. <https://doi.org/10.1093/emboj/18.2.4.7019>.
- 610 55. Stielow, B., Finkernagel, F., Stiewe, T., Nist, A., and Suske, G. (2018). MGA, L3MBTL2 and E2F6  
611 determine genomic binding of the non-canonical Polycomb repressive complex PRC1.6. *PLoS Genet*  
612 *14*, e1007193. <https://doi.org/10.1371/journal.pgen.1007193>.
- 613 56. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H.,  
and Glass, C.K. (2010). Simple Combinations of Lineage-Determining Transcription Factors Prime  
cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Molecular Cell* *38*, 576–589.  
<https://doi.org/10.1016/j.molcel.2010.05.004>.
- 614 57. Gajiwala, K.S., Chen, H., Cornille, F., Roques, B.P., Reith, W., Mach, B., and Burley, S.K. (2000).  
Structure of the winged-helix protein hRFX1 reveals a new mode of DNA binding. *Nature* *403*,  
615 916–921. <https://doi.org/10.1038/35002634>.
- 616 58. Swoboda, P., Adler, H.T., and Thomas, J.H. (2000). The RFX-Type Transcription Factor DAF-19  
Regulates Sensory Neuron Cilium Formation in *C. elegans*. *Molecular Cell* *5*, 411–421. [https://doi.org/10.1016/S1097-2765\(00\)80436-0](https://doi.org/10.1016/S1097-2765(00)80436-0).
- 617 59. Ashique, A.M., Choe, Y., Karlen, M., May, S.R., Phamluong, K., Solloway, M.J., Ericson, J., and  
Peterson, A.S. (2009). The Rfx4 Transcription Factor Modulates Shh Signaling by Regional Control of  
619 Ciliogenesis. *Sci. Signal.* *2*. <https://doi.org/10.1126/scisignal.2000602>.
- 620 60. Kistler, W.S., Baas, D., Lemeille, S., Paschaki, M., Seguin-Estevez, Q., Barras, E., Ma, W., Duteyrat, J.-  
L., Morlé, L., Durand, B., et al. (2015). RFX2 Is a Major Transcriptional Regulator of Spermiogenesis.  
621 *PLoS Genet* *11*, e1005368. <https://doi.org/10.1371/journal.pgen.1005368>.
- 622 61. Wu, Y., Hu, X., Li, Z., Wang, M., Li, S., Wang, X., Lin, X., Liao, S., Zhang, Z., Feng, X., et al.  
(2016). Transcription Factor RFX2 Is a Key Regulator of Mouse Spermiogenesis. *Sci Rep* *6*, 20435.  
623 <https://doi.org/10.1038/srep20435>.
- 624 62. Liu, M., Zhu, Y., Xing, F., Liu, S., Xia, Y., Jiang, Q., and Qin, J. (2020). The polycomb group protein  
PCGF6 mediates germline gene silencing by recruiting histone-modifying proteins to target gene  
625 promoters. *J Biol Chem* *295*, 9712–9724. <https://doi.org/10.1074/jbc.RA119.012121>.
- 626 63. Otani, J., Nankumo, T., Arita, K., Inamoto, S., Ariyoshi, M., and Shirakawa, M. (2009). Structural basis  
for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX–DNMT3–DNMT3L  
627 domain. *EMBO Reports* *10*, 1235–1241. <https://doi.org/10.1038/embor.2009.218>.
- 628 64. Guo, X., Wang, L., Li, J., Ding, Z., Xiao, J., Yin, X., He, S., Shi, P., Dong, L., Li, G., et al. (2015).  
Structural insight into autoinhibition and histone H3-induced activation of DNMT3A. *Nature* *517*,  
629 640–644. <https://doi.org/10.1038/nature13899>.

- 630 65. 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>.
- 631
- 632 66. 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>.
- 633
- 634 67. 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>.
- 635
- 636 68. Tahiliani, M., Mei, P., Fang, R., Leonor, T., Rutenberg, M., Shimizu, F., Li, J., Rao, A., and Shi, Y. (2007). The histone H3K4 demethylase SMCX links REST target genes to X-linked mental retardation. *Nature* 447, 601–605. <https://doi.org/10.1038/nature05823>.
- 637
- 638 69. 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>.
- 639
- 640 70. 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>.
- 641
- 642 71. 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>.
- 643
- 644 72. Auclair, G., Borgel, J., Sanz, L.A., Vallet, J., Guibert, S., Dumas, M., Cavelier, P., Girardot, M., Forné, T., Feil, R., et al. (2016). EHMT2 directs DNA methylation for efficient gene silencing in mouse embryos. *Genome Res.* 26, 192–202. <https://doi.org/10.1101/gr.198291.115>.
- 645
- 646 73. 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>.
- 647
- 648 74. Aubert, Y., Egolf, S., and Capell, B.C. (2019). The Unexpected Noncatalytic Roles of Histone Modifiers in Development and Disease. *Trends in Genetics* 35, 645–657. <https://doi.org/10.1016/j.tig.2019.06.004>.
- 649

- 650 75. Morgan, M.A.J., and Shilatifard, A. (2020). Reevaluating the roles of histone-modifying enzymes  
and their associated chromatin modifications in transcriptional regulation. *Nat Genet* 52, 1271–1281.  
<https://doi.org/10.1038/s41588-020-00736-4>.
- 651
- 652 76. 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>.
- 653
- 654 77. Nielsen, A.Y., and Gjerstorff, M.F. (2016). Ectopic Expression of Testis Germ Cell Proteins in Cancer  
and Its Potential Role in Genomic Instability. *Int J Mol Sci* 17. <https://doi.org/10.3390/ijms17060890>.
- 655
- 656 78. Adebayo Babatunde, K., Najafi, A., Salehipour, P., Modarressi, M.H., and Mobasher, M.B. (2017).  
Cancer/Testis genes in relation to sperm biology and function. *Iranian Journal of Basic Medical  
Sciences* 20. <https://doi.org/10.22038/ijbms.2017.9259>.
- 657
- 658 79. Janic, A., Mendizabal, L., Llamazares, S., Rossell, D., and Gonzalez, C. (2010). Ectopic expression  
of germline genes drives malignant brain tumor growth in Drosophila. *Science* 330, 1824–1827.  
<https://doi.org/10.1126/science.1195481>.
- 659
- 660 80. Bonefas, K.M., and Iwase, S. (2021). Soma-to-germline transformation in chromatin-linked neurode-  
velopmental disorders? *FEBS J*. <https://doi.org/10.1111/febs.16196>.
- 661
- 662 81. Velasco, G., Walton, E.L., Sterlin, D., Hédonin, 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>.
- 663
- 664 82. 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  
(Basel)* 3, 578–605. <https://doi.org/10.3390/biology3030578>.
- 665
- 666 83. Samaco, R.C., Mandel-Brehm, C., McGraw, C.M., Shaw, C.A., McGill, B.E., and Zoghbi, H.Y. (2012).  
Crh and Oprm1 mediate anxiety-related behavior and social approach in a mouse model of MECP2  
duplication syndrome. *Nat Genet* 44, 206–211. <https://doi.org/10.1038/ng.1066>.
- 667
- 668 84. Stephens, M. (2016). False discovery rates: A new deal. *Biostat*, kxw041. <https://doi.org/10.1093/biostatistics/kxw041>.
- 669
- 670 85. 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>.
- 671
- 672 86. 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>.

673

- 674 87. Mikedis, M.M., Fan, Y., Nicholls, P.K., Endo, T., Jackson, E.K., Cobb, S.A., De Rooij, D.G., and Page,  
675 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>.

676 **Figures and Tables**

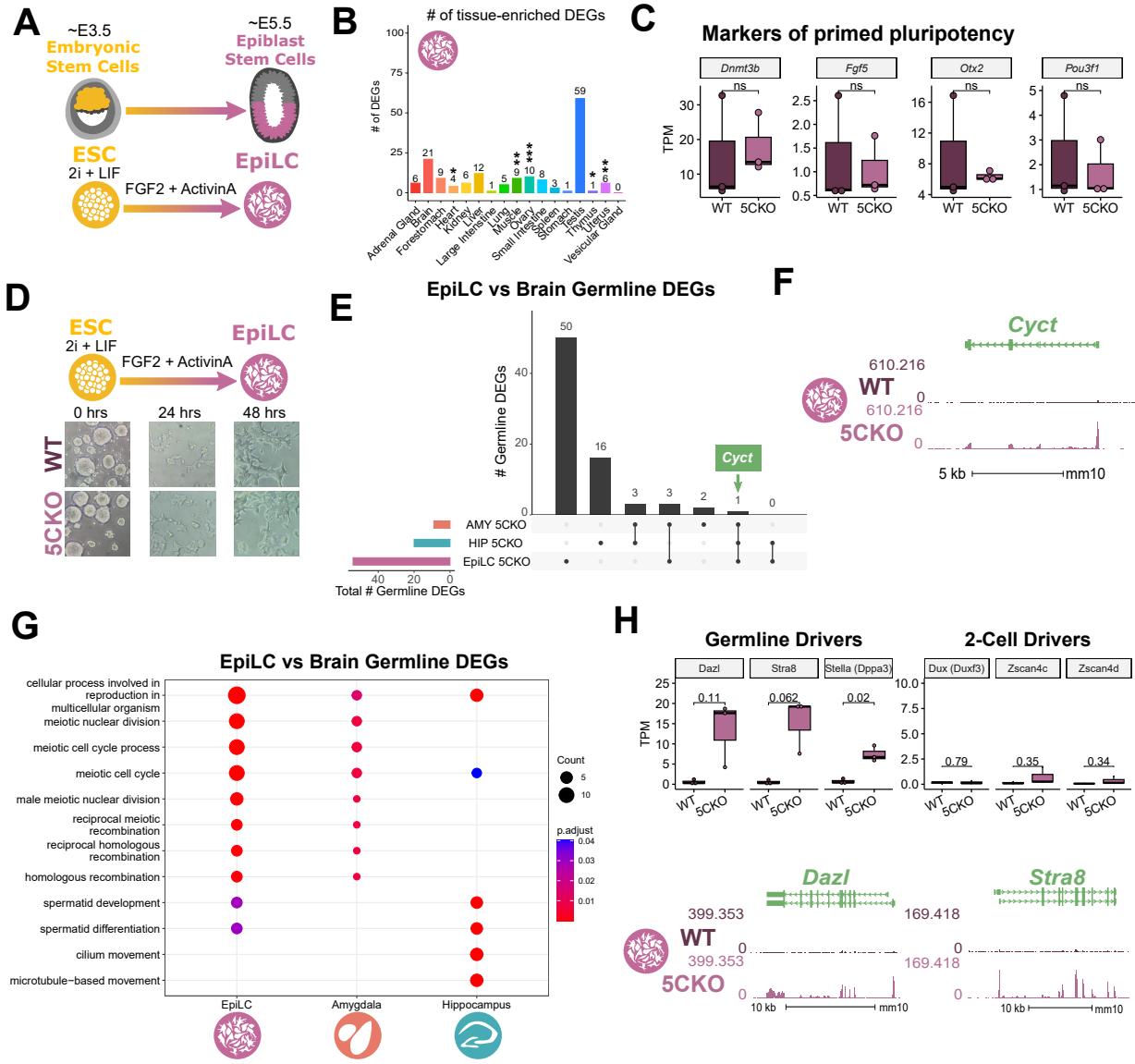
- 677     • Supplementary table 1: list of all germline genes.
- 678       – Columns to include:
- 679           \* KDM5C bound vs not
- 680           \* DEG in EpiLC, brain, both, neither (separate columns?)



**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<sub>v</sub>) 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.

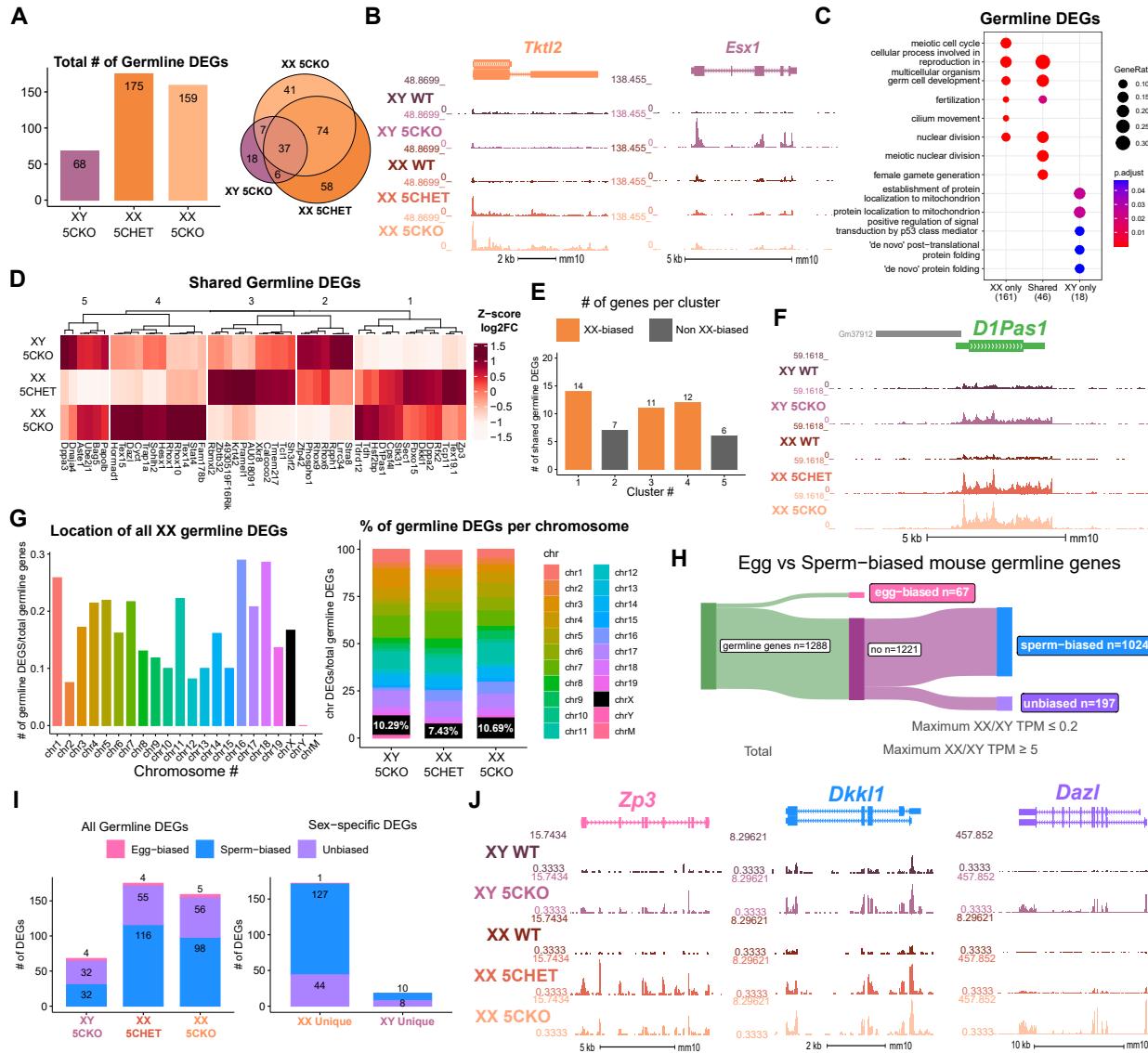
**D.** No significant difference in primed pluripotency marker expression in wild-type versus *Kdm5c*-KO EpiLCs. Welch's t-test, expression in transcripts per million (TPM).

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

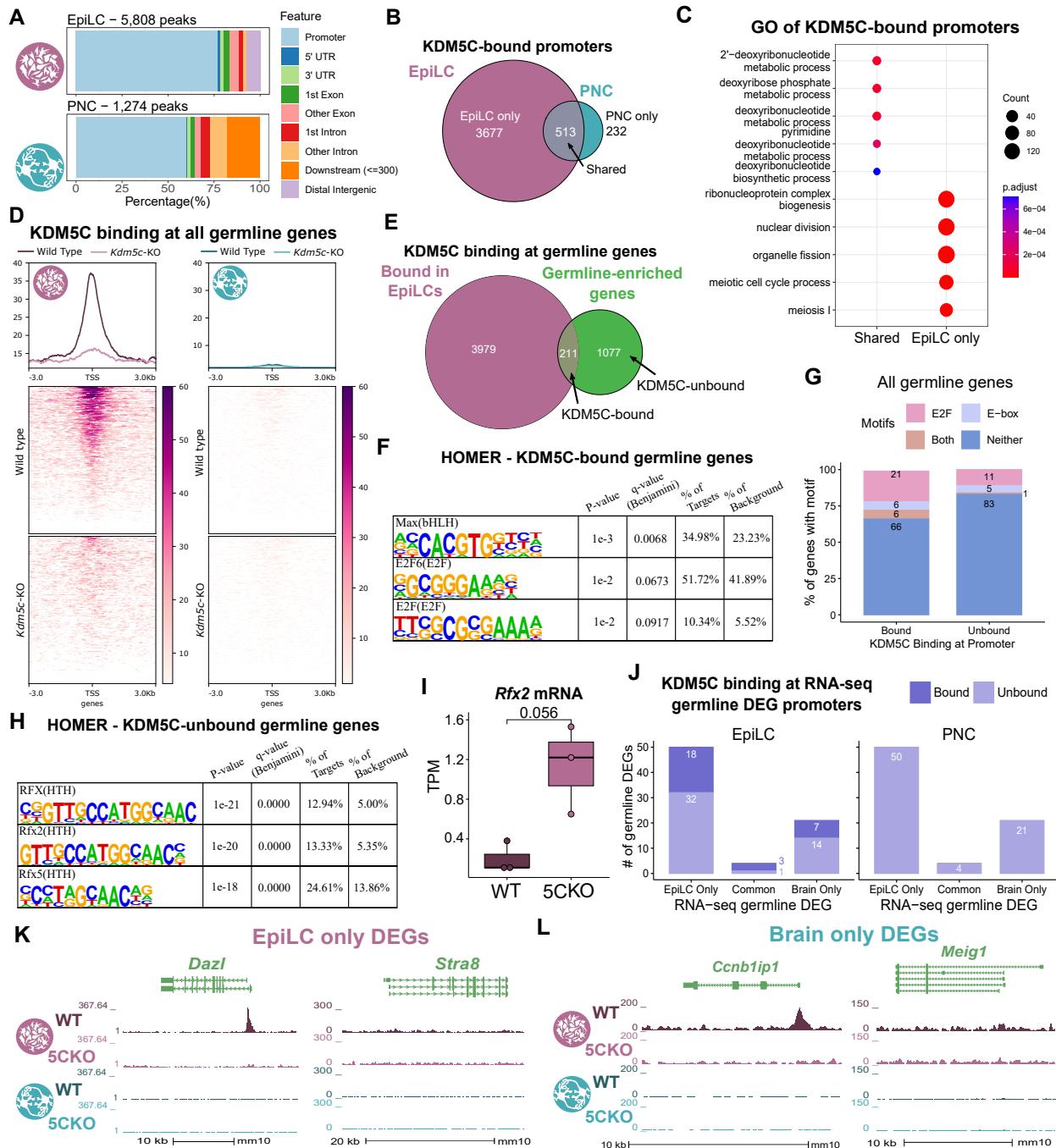
**F.** Upset plot displaying the overlap of germline DEGs expressed in *Kdm5c*-KO EpiLCs, amygdala (AMY), and hippocampus (HIP) RNA-seq datasets.

**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.

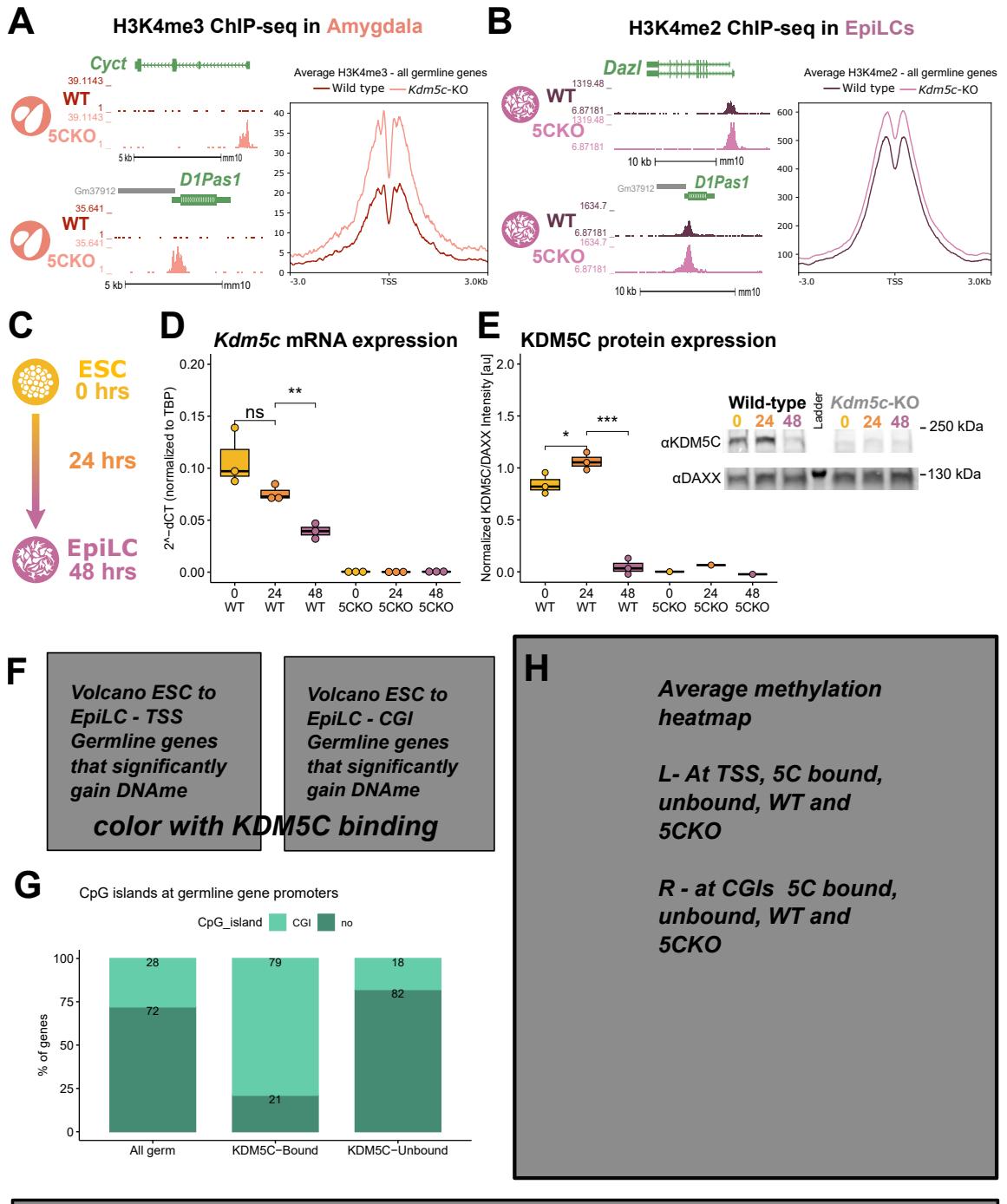


**Figure 4: Chromosomal sex influences *Kdm5c*-KO germline gene misexpression.** **A-B.** Expression of tissue-enriched genes (Li et al 2017) in the male *Kdm5c*-KO amygdala (A) and hippocampus (B). Left - MA plot of mRNA-sequencing. Right - Number of tissue-enriched differentially expressed genes (DEGs). \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ , 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 5: KDM5C binds to a subset of germline gene promoters during early embryogenesis** **A.** ChIPseeker localization of KDM5C peaks at different genomic regions in EpiLCs (top) and hippocampal and cortex primary neuron cultures (PNCs, bottom). **B.** Overlap of genes with KDM5C bound to their promoters in EpiLCs (purple) and PNCs (blue). **C.** Gene ontology (GO) comparison of genes with KDM5C bound to their promoter in EpiLCs and PNCs. Genes were classified as either bound in EpiLCs only (EpiLC only), unique to PNCs (PNC only, no significant ontologies) or bound in both PNCs and EpiLCs (shared). **D.** Average KDM5C binding around the transcription start site (TSS) of all germline-enriched genes in EpiLCs (left) and PNCs (right). **E.** KDM5C binding at the promoters of RNA-seq germline DEGs. Genes were classified as either only dysregulated in EpiLCs (EpiLC only), genes dysregulated in the hippocampus or amygdala but not EpiLCs (brain only), or genes dysregulated in both EpiLCs and the brain (common). (Legend continued on next page.)

**Figure 5: KDM5C binds to a subset of germline gene promoters during early embryogenesis.** (Legend continued.) **F.** Example KDM5C ChIP-seq bigwigs of DEGs unique to EpiLCs. Although both are expressed in *Kdm5c*-KO EpiLCs, KDM5C is only bound to the *Dazl* promoter and not the *Stra8* promoter in EpiLCs. **G.** Example KDM5C ChIP-seq bigwigs of RNA-seq DEGs common between the brain and EpiLCs. KDM5C is bound to the *D1Pas1* promoter but not the *XXX* promoter in EpiLCs. **H.** Example KDM5C ChIP-seq bigwigs of RNA-seq DEGs unique to the brain. KDM5C is bound to the *XXX* promoter but not the *Meig1* promoter



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

## 681 Notes

### 682 Things to do

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

### 687 **Dazl**

688     We were particularly interested in the aberrant transcription of *Dazl*, since it is essential for germ cell  
689 development and promotes the translation of germline mRNAs<sup>86</sup>. A significant portion of germline transcripts  
690 misexpressed in *Kdm5c*-KO EpiLCs are known binding targets of DAZL, including *Stra8*<sup>87</sup> ( $p = 1.698e-07$ ,  
691 Fisher's Exact Test). This suggests expression of DAZL protein could promote the translation of other  
692 aberrant germline transcripts, influencing their ability to impact *Kdm5c*-KO cellular function. We thus tested  
693 DAZL protein expression in *Kdm5c*-KO EpiLCs through immunocytochemistry (Figure 3I). We observed  
694 about 25% of *Kdm5c*-KO EpiLCs expressed DAZL protein and it was localized to the cytoplasm ( $p = 0.0015$ ,  
695 Welch's t-test), consistent with the pattern of DAZL expression in spermatogonia<sup>87</sup>. Altogether these results  
696 suggest tissue-specific genes are misexpressed during *Kdm5c*-KO embryogenesis, including key drivers of  
697 germline identity that can be translated into protein.

- 698     • We additionally found *Kdm5c*-KO EpiLCs ectopically express DAZL protein that is localized to the  
699         cytoplasm, similar to its morphology in spermatogonia<sup>87</sup>. **note: maybe just put in results.** Could  
700         move around depending upon if I get pheno working.

### 701 Discussion notes

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

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

- 738 • Altogether, these data indicate that while some germline genes are misexpressed due to direct loss of  
739 KDM5C binding during emryogenesis, secondary downstream mechanisms can also promote their  
740 aberrant transcription.
- 741 • Papers to read/reference:  
742 – Reconstitution of the Mouse Germ Cell Specification Pathway in Culture by Pluripotent Stem Cells:  
743 [https://www.cell.com/fulltext/S0092-8674\(11\)00771-9](https://www.cell.com/fulltext/S0092-8674(11)00771-9)  
744 – 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/>

746 **Figure outline:**

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

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

759 **Figure 3: *Kdm5c*-KO epiblast-like cells express key drivers regulators of germline identity** \* A) ESC  
760 to EpiLC differentiation Left - Morphology is unchanged, \* B) 5CKO EpiLCs express EpiLC differentiation  
761 genes similar to WT lvs \* C) Male EpiLCs express germline genes (example Cyct again) \* Overlap between  
762 brain and EpiLC germline genes - show they're mostly unique \* GO of Brain and EpiLC germline genes  
763 (meiotic enriched) \* Bigwigs or TPM of master regulators \* Show that while some are also 2-cell genes  
764 (Dazl), 2-cell specific genes aren't dysregulated (Zscan4). Important point because published KDM5C dazl  
765 paper is saying KDM5C is a 2-cell regulator, but as far as I can tell only genes shared between germline and  
766 2-cell are dysregulated.

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

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

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

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

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

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

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

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

811 **Germline gene repression background:**

812 Interestingly, some of the ectopic testis transcripts identified in the *Kdm5c*-KO brain are typically ex-  
813 pressed in germ cells<sup>8</sup>. Unlike somatic cells, germ cells (e.g. sperm and eggs) undergo meiosis and pass on  
814 their genetic material to the next generation. The germline and the soma are typically distinguished during  
815 early embryogenesis, when germline genes are silenced in epiblast stem cells soon after implantation and  
816 only reactivated in a subset to form the germline. Chromatin regulators play a key role in decommissioning  
817 germline genes as the embryo transitions from naïve to primed pluripotency by placing repressive histone  
818 H2A lysine 119 monoubiquitination (H2AK119ub1)<sup>16</sup>, histone 3 lysine 9 trimethylation (H3K9me3)<sup>16,17</sup>, and  
819 DNA CpG methylation<sup>17–19</sup> at germline gene promoters. KDM5C may also be involved in this early decom-  
820 missioning of germline genes, as re-expression of KDM5C in neurons fails to suppress their dysregulation<sup>8</sup>.  
821 In support of this, KDM5C was very recently shown to repress *Deleted in azoospermia like (Dazl)*, a key  
822 regulator of germline development, in mouse embryonic stem cells (ESCs)<sup>49,66</sup>. In support of this, two  
823 independent screens in mouse embryonic stem cells (ESCs) recently identified KDM5C as a repressor of  
824 *Deleted in azoospermia like (Dazl)*, a key regulator of germline development. However, KDM5C's role in  
825 embryonic germline gene repression is currently unclear, given that Dazl is also expressed in ESCs and in  
826 the 2-cell stage and germline gene misexpression has yet to be globally characterized during *Kdm5c*-KO  
827 embryogenesis.