

1 Erosion of somatic tissue identity with loss of the X-linked intellectual 2 disability factor KDM5C

3
4 Gaps in knowledge addressed: * Are other tissue-enriched genes dysregulated, or only testis, germline genes? * Curating
5 a robust list of male and female germline genes * Should talk about 2-cell genes vs germline genes - way to systematically
6 categorize? * Mechanism behind long-term germline gene misexpression * Recent evidence suggests loss of KDM5C
7 in ESCs express some germline genes * Unclear if catalytic activity is required for long-term silencing * Unclear if their
8 dysregulation lasts throughout life or the same between brain or not * When in development does it begin? - Recent evidence
9 suggests some germline genes expressed in 5CKO ESCs but unclear if their dysregulation lasts throughout differentiation
10 and if the identity of germline genes are different compared to the brain * Are there functional consequences to germline
11 gene misexpression?

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

- 19 • Loss of KDM5C can result in the misexpression of genes typically only found in the testis
20 – Misexpression of tissue-enriched genes hasn't been systematically characterized - Unclear if these genes are
21 exceptions or if other tissue-specific genes are dysregulated
22 – Interestingly, these genes (Cyct, D1pas1) typically function in the germline
23 – Germ cells (meiotic cells) are typically distinguished from somatic cells very early on in embryogenesis and is a
24 key feature of multicellularity
25 – Chromatin regulators are very important for decommissioning germline genes and act successively the embryo
26 implants into the uterine wall
27 * Most studies have focused on ESCs, which have a similar transcriptome to germ cells / 2-cells
28 * recently, KDM5C was shown to repress DAZL in ESCs, independent of its catalytic activity
29 * However, DNA methylation is lost in the mature 5CKO brain, DNA methylation is placed later and it's unclear if
30 it's required for long-term repression (maybe too specific, just trying to go into the fact that the mechanism is
31 partially understood but unclear)
32 – Systematic characterization of ectopic germline genes hasn't been done
33 * unknown if other germline-enriched genes are dysregulated, including oocyte-specific genes
34 * Crucially, it's unknown if misexpression of the germline program leads to functional consequences in 5CKO
35 cells.

36 Introduction

37 To form a complete organism, embryonic stem cells must differentiate into a myriad of discrete cellular identities. This is
38 in part accomplished by chromatin regulators that can either promote or impede lineage-specific gene expression through
39 histone and DNA modifications^{1,2}. Although initially identified for their roles in cellular identity^{3,4}, recent advancements in
40 next generation sequencing technologies unexpectedly found many neurodevelopmental disorders (NDDs) are caused by or
41 linked to mutations in chromatin regulators. This relationship is partially explained by their regulation of brain-specific genes
42 or chromatin states, such as modulating genes involved in synaptic maturation⁵ or the transition between neuronal and glial
43 developmental programs⁶. However, loss of some chromatin regulators can also lead to the misexpression of tissue-specific
44 genes outside of their intended environment^{3,4,7}. Currently, very few studies have explored the misexpression of non-neuronal,
45 tissue-specific genes in chromatin-linked neurodevelopmental disorders^{8,9} and it is unclear if this partial loss of brain identity
46 contributes to neurodevelopmental impairments. To elucidate their contribution to neurodevelopmental impairments, it is
47 essential to first characterize the types of genes misexpressed, the developmental time point the dysregulation begins, and
48 the molecular mechanism underlying their de-repression. Characterizing these features will enable us to identify molecular
49 footholds common between NDDs that can then be exploited for potential therapeutics.

50 In this study, we characterized the misexpression of tissue-enriched genes with loss of the chromatin regulator lysine
51 demethylase 5C (KDM5C, also known as SMCX or JARID1C), a histone 3 lysine 4 demethylase. Pathogenic mutations in
52 *KDM5C* cause Intellectual Developmental Disorder, X-linked, Syndromic, Claes-Jensen Type (MRXSCJ, OMIM: 300534),
53 whose features include short stature, intellectual disability, seizures, aberrant aggression, and autistic behaviors^{10–12}. Previous
54 work has demonstrated constitutive *Kdm5c* knockout (-KO) mice recapitulate key MRXSCJ patient phenotypes, including
55 hyperaggression and learning impairments¹³. Next generation RNA sequencing (RNA-seq) in the *Kdm5c*-KO hippocampus
56 unexpectedly revealed ectopic expression of testis-enriched genes within the brain⁹. However, it is currently unclear if
57 misexpression in the *Kdm5c*-KO brain is unique to testis genes, as other tissue-enriched genes have not been systematically
58 evaluated.

59 Interestingly, some of the ectopic testis transcripts identified in the *Kdm5c*-KO brain are typically expressed in germ cells⁹.
60 Unlike somatic cells, germ cells (e.g. sperm and eggs) undergo meiosis and pass on their genetic material to the next genera-
61 tion. The germline and the soma are typically distinguished during early embryogenesis, when germline genes are silenced
62 in epiblast stem cells soon after implantation and only reactivated in a subset to form the germline. Chromatin regulators
63 play a key role in decommissioning germline genes as the embryo transitions from naive to primed pluripotency by placing
64 repressive histone H2A lysine 119 monoubiquitination (H2AK119ub1)¹⁴, histone 3 lysine 9 trimethylation (H3K9me3)^{14,15},
65 and DNA CpG methylation^{15–17} at germline gene promoters. KDM5C may also be involved in this early decommissioning of
66 germline genes, as re-expression of KDM5C in neurons fails to suppress their dysregulation⁹. In support of this, KDM5C
67 was very recently shown to repress *Deleted in azoospermia like (Dazl)*, a key regulator of germline development, in mouse
68 embryonic stem cells (ESCs)^{18,19}. In support of this, two independent screens in mouse embryonic stem cells (ESCs) recently
69 identified KDM5C as a repressor of *Deleted in azoospermia like (Dazl)*, a key regulator of germline development. However,
70 KDM5C's role in embryonic germline gene repression is currently unclear, given that Dazl is also expressed in ESCs and in
71 the 2-cell stage and germline gene misexpression has yet to be globally characterized during *Kdm5c*-KO embryogenesis.

72 To elucidate KDM5C's role in tissue identity, here we characterized the aberrant transcription of tissue-enriched genes
73 within the *Kdm5c*-KO brain and epiblast-like cells (EpiLCs), an *in vitro* model of the post-implantation embryo. We identified
74 general dysregulation of tissue-enriched genes in both the adult *Kdm5c*-KO brain and EpiLCs, including misexpression of

75 liver, muscle, and ovary genes. The *Kdm5c*-KO amygdala and hippocampus had significant enrichment of testis-biased
76 genes that are unique to germ cells. To better characterize germline gene misexpression, we then generated a dataset of
77 germline-enriched genes by comparing gene expression in gonads with germ cell depletion. We found *Kdm5c*-KO EpiLCs
78 primarily expressed unique germline genes compared to the mature *Kdm5c*-KO brain, including *Dazl* and *Stra8*, key drivers
79 of germline identity and meiosis. While KDM5C is directly bound to some germline gene promoters in EpiLCs, it is not directly
80 bound to many germline-enriched mRNAs expressed with *Kdm5c*-KO cells, indicating germline genes can be aberrantly
81 transcribed through indirect mechanisms. Finally, we found KDM5C loss impairs the placement of DNA methylation at
82 germline gene promoters as ESCs differentiate into EpiLCs. Therefore, we propose KDM5C plays a crucial role in the
83 development of tissue identity during early embryogenesis, including establishment of the soma-germline boundary.

84 **note: need a better conclusion sentence - work on when we know what's happening with last figure/functional
85 consequences**

86 Results

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

88 • **note: should I compare amygdala and hippocampus? scandaglia only looked at hippocampus**

89 Previous RNA sequencing (RNA-seq) in the adult hippocampus ectopic expression of some testis genes within the *Kdm5c*
90 knockout (-KO) brain⁹. Since tissue-enriched genes have not been systematically characterized in the *Kdm5c*-KO brain,
91 it is currently unclear if this erosion of brain tissue identity is a major consequence of *Kdm5c* loss and if it is unique to
92 testis-enriched genes. Therefore, we first globally assessed the expression of previously characterized mouse tissue-enriched
93 genes²⁰ in our published mRNA-seq datasets of the amygdala and hippocampus in adult mice with constitutive knockout of
94 *Kdm5c*²¹.

95 We found a large proportion of genes significantly upregulated within the *Kdm5c*-KO brain (DESeq2²², log2 fold change >
96 0.5, q < 0.1) are typically enriched in non-brain tissues (Amygdala: 35%, Hippocampus: 24%) (Figure 1A-B). The majority
97 of tissue-enriched differentially expressed genes (tissue-enriched DEGs) were testis genes (Figure 1A-C). Testis-biased
98 DEGs were significantly enriched for both brain regions (Amygdala p = 1.83e-05; Hippocampus p = 4.26e-11, Fisher's Exact
99 Test), even though the testis has the largest total number of tissue-biased genes compared to any other tissue (2,496 genes).
100 Of note, we did not observe enrichment of brain-enriched genes (Amygdala p = 1; Hippocampus p = 0.74, Fisher's Exact),
101 despite the fact these are brain samples and the brain has the second highest total number of tissue-enriched genes (708
102 genes).

103 In addition to the high enrichment of testis genes, we also identified aberrant expression of other tissue-enriched genes
104 within the *Kdm5c*-KO brain. Although the *Kdm5c*-KO mice were male, we also observed significant enrichment of ovary-biased
105 DEGs in both the amygdala and hippocampus (Amygdala p = 0.00574; Hippocampus p = 0.048, Fisher's Exact) (Figure 1D).
106 Intriguingly, many ovary and testis-biased DEGs have functions specific to germ cells and have no known role in the brain.
107 For example, the testis-biased DEG *FK506 binding protein 6* (*Fkbp6*) is a known regulator of piRNA expression and meiosis
108 in germ cells^{23,24} (Figure 1C) while the ovary-enriched DEG *Zygotic arrest 1* (*Zar1*) was recently shown to sequester mRNAs
109 in oocytes for meiotic maturation and early zygote development²⁵ (Figure 1D). Although not consistent across brain regions,
110 we also found enrichment of two non-gonadal tissues - the liver (Amygdala p = 0.0398, Fisher's Exact Test) and the muscle
111 (Hippocampus p = 0.0104, Fisher's Exact Test). An example of a liver-biased DEG is *Apolipoprotein C-I* (*Apoc1*), which is

112 involved in lipoprotein metabolism (Figure 1E). Testis, ovary, and liver-enriched DEGs showed little to no expression in the
113 wild-type brain, yet our mRNA-seq data indicate they are polyadenylated and spliced into mature transcripts (Figure 1C-E).
114 Together, these results suggest misexpression of testis and other tissue-enriched genes within the *Kdm5c*-KO brain is a
115 major effect of KDM5C loss.

116 **Germline genes are aberrantly expressed in the male *Kdm5c*-KO brain**

117 The testis contains both germ cells (e.g. spermatogonia) and somatic cells (e.g. Leydig cells) that support hormone
118 production and germline functions. We then wondered if the high enrichment of testis genes in the *Kdm5c*-KO brain reflected
119 a failure to demarcate between the soma and germline. Gene ontology analysis of testis-biased DEGs revealed enrichment
120 of germline-relevant ontologies, including spermatid development (GO:0,007,286, p.adjust = 6.2e-12) and sperm axoneme
121 assembly (GO:0,007,288, p.adjust = 2.45e-14) (Figure 2A).

122 To further validate if these testis DEGs are truly germline genes, we then evaluated their expression in germ cell-depleted
123 testes via *c-Kit* mutation (*Kit*^{W/Wv})²⁶. Heterozygous *W* and *Wv* mutations in the enzymatic domain of *c-Kit* prevents the
124 maturation of germ cells, resulting in overall germ cell depletion^{handelSertoliCellDifferentiation1979?}. We found almost all *Kdm5c*-KO
125 testis-enriched DEGs lose their expression with germ cell depletion (Figure 2B). The only testis DEG that did not show
126 considerable downregulation with germline depletion was *FK506 binding protein 6* (*Fkbp6*), the aforementioned testis
127 gene that regulates piRNA expression and meiosis in germ cells^{23,24}. We additionally assessed testis DEG expression
128 in a published testis single cell RNA-seq dataset that identified cell type-specific markers²⁷. We found that while some
129 testis-enriched DEGs were classified as specific markers for different germ cell types (e.g. spermatogonia, spermatocytes,
130 round spermatids, and elongating spermatids) none marked somatic cells (Figure 2C). Together, these data suggest the
131 *Kdm5c*-KO brain expresses germline genes.

132 We then aimed to globally characterize germline gene misexpression in *Kdm5c*-KO mice, but lacked a comprehensive list
133 of mouse germline-enriched genes. To facilitate downstream analyses, we generated a list of male and female germline-
134 enriched genes by evaluating gene expression in wild-type and germline-depleted (*Kit*^{W/Wv}) mice. We assessed gene
135 expression in male and female mice at embryonic day 12, 14, and 16, as well as male postnatal day 6 and adult testes. We
136 defined genes as germline-enriched if their expression met the following criteria: 1) their expression is greater than 1 FPKM
137 in wild-type germline 2) their expression in any wild-type, non-gonadal tissue²⁰ does not exceed 20% of maximum expression
138 in wild-type germline, and 3) their expression in the germ cell-depleted germline, for any sex or time point, does not exceed
139 20% of their maximum expression in the wild-type germline. These criteria yielded 1,288 germline-enriched genes (Figure
140 2D), which was hereafter used as a resource for assessing misexpression of germline genes (Supplementary table 1).

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

142 Misexpression of germline genes in the *Kdm5c*-KO brain suggests mutants fail to demarcate germline and somatic cellular
143 identity. Germ cells are typically distinguished from somatic cells soon after the embryo implants into the uterine wall^{28,29}
144 when a subset of epiblast stem cells become the primordial germ cells (PGCs) while the remainder differentiate into the
145 ectoderm, mesoderm, and endoderm to form the body's somatic tissues³⁰. This developmental time point can be modeled *in*
146 *vitro* through differentiation of embryonic stem cells (ESCs) into post-implantation epiblast-like stem cells (EpiLCs) (Figure
147 3A, top). Previous studies have demonstrated that while some germline-enriched genes are also expressed in embryonic
148 stem cells (ESCs), they are silenced as they differentiate into EpiLCs¹⁵. Therefore, we assessed if KDM5C was necessary

149 for initial germline gene silencing by assessing their expression in male *Kdm5c*-KO EpiLCs using our previously generated
150 RNA-seq dataset³¹.

151 We did not observe any apparent morphology changes between wild-type and *Kdm5c*-KO cells during ESC to EpiLC
152 differentiation (Figure 3A, bottom) and found no significant difference in primed pluripotency genes that indicate proper EpiLC
153 differentiation (Figure 3B). We then identified *Kdm5c*-KO EpiLC DEGs through DESeq2²² (log2 fold change > 0.5, q < 0.1)
154 and again observed general dysregulation of tissue-enriched genes, with the largest number of genes belonging to the brain
155 and testis, although they were not significantly enriched (Figure 3C).

156 Using our curated list, we found 54 germline-enriched genes were misexpressed in *Kdm5c*-KO EpiLCs, including *Cyct*
157 (Figure 3D). To evaluate whether germline mRNAs are constitutively dysregulated or change over the course of development,
158 we then compared EpiLC germline DEGs to those expressed in the *Kdm5c*-KO brain. The majority of EpiLC germline
159 DEGs were unique to EpiLCs, with only *Cyct* shared across sequencing datasets (Figure 3E). We then compared the
160 function of EpiLC and brain germline DEGs through gene ontology and found particularly high enrichment of meiosis-related
161 gene ontologies in EpiLCs (Figure 3F), such as meiotic cell cycle (GO:0,051,321, p.adjust = 4.44e-07). Contrastingly, the
162 *Kdm5c*-KO hippocampus instead expressed genes involved in late-stage sperm development, those involved in the sperm
163 axoneme. This suggests the germline developmental program may be occurring ectopically as *Kdm5c*-KO cells progresses
164 through somatic tissue development (**note: placement is awkward. Even put here or talk about in dicussion because**
165 **ChIPseq could help support this? - KDM5C not bound to a lot of brain DEGs.**)

166 While a few meiosis-related genes are misexpressed in the *Kdm5c*-KO brain, DEGs unique to *Kdm5c*-KO EpiLCs included
167 key drivers of germline identity, such as *Stimulated by retinoic acid 8 (Stra8)* and *Deleted in azoospermia like (Dazl)* (Figure
168 3G). These genes are typically expressed during embryonic germ cell development to commit PGCs to the germline fate, but
169 are also expressed later in life to trigger meiotic gene expression programs^{32–34}. Of note, some germline genes, including
170 *Dazl*, are also expressed in the two-cell embryo. However, we did not see misexpression of two-cell embryo-specific genes,
171 like *Zscan4d* (q = 0.381), indicating *Kdm5c*-KO in EpiLCs does not revert cells back to a 2-cell state (Figure 3G).

172 *Dazl* is an RNA-binding protein that regulates the translation of germline mRNAs and is essential for germ cell function³⁵.
173 Interestingly, a significant portion of germline transcripts expressed in *Kdm5c*-KO EpiLCs are known binding targets of
174 DAZL, including *Stra8*³⁶ (p = 1.698e-07, Fisher's Exact Test). This suggests expression of DAZL protein would enable
175 the translation of other aberrant germline transcripts, influencing their ability to impact on *Kdm5c*-KO cellular function. We
176 therefore stained *Kdm5c*-KO EpiLCs for DAZL protein expression through immunocytochemistry (Figure 3H). We found
177 about 25% of *Kdm5c*-KO EpiLCs expressed DAZL protein in their cytoplasm (p = 0.0015, Welch's t-test), consistent with the
178 subcellular localization observed when DAZL is stained in spermatogonia³⁶. Altogether these results suggest *Kdm5c*-KO
179 EpiLCs fail to decommission germline genes during early embryogenesis, including key drivers of germline identity that can
180 be translated into protein.

181 **KDM5C-mediated removal of H3K4me3 during early embryogenesis promotes long-term repression of germline
182 genes**

183 **Figure 4: Loss of KDM5C's catalytic activity impairs DNAme placement and long-term silencing of germline genes**
184 A) KDM5C binding in EpiLCs vs PNCs - number of peaks and percentage of bound genomic regions B) KDM5C binding in
185 EpiIcs vs pncs germline genes - GO C) Left - bar graph of germline DEGs KDM5C is bound to in EpiLCs Right D) Example
186 bigwigs * Brain and EpiLC shared, EpiLC-specific, Brain-specific, E) motif analysis of KDM5C-bound vs unbound genes

187 --- separate into two figures
188 D) Bigwigs of Increase in H3K4me3 in Kdm5c-KO amygdala at germline genes
189 E) Increase in H3K4me2 in EpiLCs at germline genes (left - bigwigs, right average signal at germline gene TSS?)
190 * Previous studies only looked at ESCs, unknown if catalytic activity is required for long-term repression, es
191 F) KDM5C expression ESC --> EpiLC (increasing then decreasing) Left - RNA right - protein
192 G) RNA expression of germline genes with catalytic dead rescue (Ilakkiya)
193 H) DNA methylation in WT and 5CKO EpiLCs (Ilakkiya)
194
195 Direct vs indirect binding motif analysis
196 Previous work suggests KDM5C represses germline genes during early development, since KDM5C binding is gradually
197 lost at hippocampal germline DEG promoters over development⁹ and re-expression of KDM5C in neuronal cultures fails to
198 suppress their transcription. However, we identified many germline genes that were only dysregulated in *Kdm5c*-KO EpiLCs
199 and not in the mature brain, including key regulators of meiosis and germline identity. While KDM5C was recently shown to
200 bind to *Dazl* in ESCs^{18,19}, it's unclear if KDM5C binds to other germline drivers that are expressed in *Kdm5c*-KO EpiLCs and
201 if this binding is maintained as cells differentiate.
202 To better elucidate the mechanism behind KDM5C-mediated germline gene silencing, we analyzed KDM5C chromatin
203 immunoprecipitation followed by DNA sequencing (ChIP-seq) datasets we previously generated in EpiLCs³¹ and primary
204 hippocampal and cortical neuron cultures (PNCs)¹³. We identified a higher number of KDM5C peaks in EpiLCs overall
205 (EpiLCs: 5,808, PNCs: 1,276, MACS2 q < 0.1 and fold enrichment > 1) and while KDM5C was primarily localized to gene
206 promoters in both cell types (EpiLCs: 4,190, PNCs: 745 +/- 500kb from TSS), PNCs showed increased localization to distal
207 intergenic regions (EpiLCs: , PNCs:) (Figure 4A). Gene ontology analysis of KDM5C-bound promoters in EpiLCs revealed
208 high enrichment of germline-related ontologies, including meiotic nuclear division (GO:0140013) and synaptonemal complex
209 assembly (GO:0007130) (Figure 4B). Contrastingly, germline ontologies are not enriched in PNCs and KDM5C-bound genes
210 instead perform non-germline-specific functions, such as heterocycle synthesis (GO: 0018130) and pyrimidine metabolism
211 (GO: 0072527).
212 We then assessed KDM5C binding at germline genes expressed *Kdm5c*-KO cells to assess if their dysregulation is due to
213 direct loss of KDM5C binding during embryogenesis.
214 * KDM5C was bound to the promoter of many genes dysregulated in both EpiLCs and brain datasets, including _D1P
215 * KDM5C was not bound to many of the brain-specific germline DEGs, such as XXX.
216 * While KDM5C binds to many germline DEGs in EpiLCs, including _Dazl_, many were also unbound. This included t
217 * All genes lacked binding in PNCs.
218 * Altogether, these data indicate that while some germline genes are misexpressed due to direct loss of KDM5C
219

220 Although KDM5C is generally thought to suppress transcription through erasure of histone 3 lysine 4 di- and trimethylation
221 (H3K4me2/3), recent studies in ESCs have suggested KDM5C's repression *Dazl* is independent of its catalytic activity¹⁸. It
222 is currently unclear if KDM5C's catalytic activity aids in the long-term repression of germline genes, such as enabling the
223 placement of DNA methylation.
224 • In congruence with previous work in the *Kdm5c*-KO hippocampus⁹, we observed ectopic H3K4me3 peaks in the
225 *Kdm5c*-KO amygdala at the transcription start site (TSS) of germline genes (Figure 4).

- 226 • We additionally observed a marked increase in H3K4me2 at germline gene promoters in *Kdm5c*-KO EpiLCs.
227 – This increase was greatest at germline genes directly bound by KDM5C (?)

228 **Discussion**

- 229 • The demarcation of the germ vs soma is a key feature of multicellularity
230 • Other H3K4me regulators, anything known about tissue-biased gene expression?
231 • This suggests the germline developmental program is occurring ectopically as *Kdm5c*-KOs progresses through somatic
232 tissue development
233 • tissue-biased gene expression:
234 – However unlike the gonadal-biased DEGs, many liver and muscle-biased DEGs have a known involvement in brain
235 function. For example, the liver-biased DEG is *Apolipoprotein C-I* (*Apoc1*), is important for lipoprotein metabolism
236 but has also been shown to influence learning and memory (Figure 1E).
237 • Otx2 is properly expressed in EpiLCs and prevents pgc identity <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6485399/>
238 • Papers to read/reference:
239 – Reconstitution of the Mouse Germ Cell Specification Pathway in Culture by Pluripotent Stem Cells: [https://www.cell.com/fulltext/S0092-8674\(11\)00771-9](https://www.cell.com/fulltext/S0092-8674(11)00771-9)
240 – 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/>

243 **References**

- 244 1. Strahl, B.D., and Allis, C.D. (2000). The language of covalent histone modifications. *Nature* *403*, 41–45. <https://doi.org/10.1038/47412>.
- 245 2. Jenuwein, T., and Allis, C.D. (2001). Translating the Histone Code. *Science* *293*, 1074–1080. <https://doi.org/10.1126/science.1063127>.
- 246 3. Lewis, E.B. (1978). A gene complex controlling segmentation in *Drosophila*. *Nature* *276*, 565–570. <https://doi.org/10.1038/276565a0>.
- 247 4. Kennison, J.A., and Tamkun, J.W. (1988). Dosage-dependent modifiers of polycomb and antennapedia mutations in *Drosophila*. *Proc. Natl. Acad. Sci. U.S.A.* *85*, 8136–8140. <https://doi.org/10.1073/pnas.85.21.8136>.
- 248 5. Zhou, Z., Hong, E.J., Cohen, S., Zhao, W.-N., Ho, H.-Y.H., Schmidt, L., Chen, W.G., Lin, Y., Savner, E., Griffith, E.C., et al. (2006). Brain-specific phosphorylation of MeCP2 regulates activity-dependent Bdnf transcription, dendritic growth, and spine maturation. *Neuron* *52*, 255–269. <https://doi.org/10.1016/j.neuron.2006.09.037>.
- 249 6. Hirabayashi, Y., Suzuki, N., Tsuboi, M., Endo, T.A., Toyoda, T., Shinga, J., Koseki, H., Vidal, M., and Gotoh, Y. (2009). Polycomb Limits the Neurogenic Competence of Neural Precursor Cells to Promote Astrocytic Fate Transition. *Neuron* *63*, 600–613. <https://doi.org/10.1016/j.neuron.2009.08.021>.

- 256 7. Kassis, J.A., Kennison, J.A., and Tamkun, J.W. (2017). Polycomb and Trithorax Group Genes in *Drosophila*. *Genetics*
257 206, 1699–1725. <https://doi.org/10.1534/genetics.115.185116>.
- 258 8. Schaefer, A., Sampath, S.C., Intrator, A., Min, A., Gertler, T.S., Surmeier, D.J., Tarakhovsky, A., and Greengard, P.
259 (2009). Control of cognition and adaptive behavior by the GLP/G9a epigenetic suppressor complex. *Neuron* 64,
678–691. <https://doi.org/10.1016/j.neuron.2009.11.019>.
- 260 9. Scandaglia, M., Lopez-Atalaya, J.P., Medrano-Fernandez, A., Lopez-Cascales, M.T., Del Blanco, B., Lipinski, M., Benito,
261 E., Olivares, R., Iwase, S., Shi, Y., et al. (2017). Loss of Kdm5c Causes Spurious Transcription and Prevents the Fine-
Tuning of Activity-Regulated Enhancers in Neurons. *Cell Rep* 21, 47–59. <https://doi.org/10.1016/j.celrep.2017.09.014>.
- 262 10. Claes, S., Devriendt, K., Van Goethem, G., Roelen, L., Meireleire, J., Raeymaekers, P., Cassiman, J.J., and Fryns, J.P.
263 (2000). Novel syndromic form of X-linked complicated spastic paraplegia. *Am J Med Genet* 94, 1–4.
- 264 11. Jensen, L.R., Amende, M., Gurok, U., Moser, B., Gimmel, V., Tzschach, A., Janecke, A.R., Tariverdian, G., Chelly, J.,
265 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>.
- 266 12. Carmignac, V., Nambot, S., Lehalle, D., Callier, P., Moortgat, S., Benoit, V., Ghoumid, J., Delobel, B., Smol, T., Thuillier,
267 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>.
- 268 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
269 Histone Methylation. *Cell Reports* 14, 1000–1009. <https://doi.org/10.1016/j.celrep.2015.12.091>.
- 270 14. Endoh, M., Endo, T.A., Shinga, J., Hayashi, K., Farcas, A., Ma, K.W., Ito, S., Sharif, J., Endoh, T., Onaga, N., et
271 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>.
- 272 15. Mochizuki, K., Sharif, J., Shirane, K., Uranishi, K., Bogutz, A.B., Janssen, S.M., Suzuki, A., Okuda, A., Koseki, H.,
273 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>.
- 274 16. Velasco, G., Hubé, F., Rollin, J., Neuillet, D., Philippe, C., Bouzinba-Segard, H., Galvani, A., Viegas-Péquignot, E., and
275 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>.
- 276 17. 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
277 reprogramming in the mouse germline. *Development* 139, 3623–3632. <https://doi.org/10.1242/dev.081661>.
- 278 18. Gupta, N., Yakhou, L., Albert, J.R., Azogui, A., Ferry, L., Kirsh, O., Miura, F., Battault, S., Yamaguchi, K., Laisné,
279 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>.
- 280 19. 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>.

- 282 20. Li, B., Qing, T., Zhu, J., Wen, Z., Yu, Y., Fukumura, R., Zheng, Y., Gondo, Y., and Shi, L. (2017). A Comprehensive
283 Mouse Transcriptomic BodyMap across 17 Tissues by RNA-seq. *Sci Rep* 7, 4200. <https://doi.org/10.1038/s41598-017-04520-z>.
- 284 21. 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.,
285 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>.
- 286 22. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data
287 with DESeq2. *Genome Biol* 15, 550. <https://doi.org/10.1186/s13059-014-0550-8>.
- 288 23. Crackower, M.A., Kolas, N.K., Noguchi, J., Sarao, R., Kikuchi, K., Kaneko, H., Kobayashi, E., Kawai, Y., Kozieradzki, I.,
289 Landers, R., et al. (2003). Essential Role of Fkbp6 in Male Fertility and Homologous Chromosome Pairing in Meiosis.
Science 300, 1291–1295. <https://doi.org/10.1126/science.1083022>.
- 290 24. Xiol, J., Cora, E., Koglgruber, R., Chuma, S., Subramanian, S., Hosokawa, M., Reuter, M., Yang, Z., Berninger, P.,
291 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>.
- 292 25. Cheng, S., Altmeppen, G., So, C., Welp, L.M., Penir, S., Ruhwedel, T., Menelaou, K., Harasimov, K., Stützer, A.,
293 Blayney, M., et al. (2022). Mammalian oocytes store mRNAs in a mitochondria-associated membraneless compartment.
Science 378, eabq4835. <https://doi.org/10.1126/science.abq4835>.
- 294 26. Mueller, J.L., Skaletsky, H., Brown, L.G., Zaghlul, S., Rock, S., Graves, T., Auger, K., Warren, W.C., Wilson, R.K., and
295 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>.
- 296 27. Green, C.D., Ma, Q., Manske, G.L., Shami, A.N., Zheng, X., Marini, S., Moritz, L., Sultan, C., Gurczynski, S.J., Moore,
297 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>.
- 298 28. Magnúsdóttir, E., and Surani, M.A. (2014). How to make a primordial germ cell. *Development* 141, 245–252.
299 <https://doi.org/10.1242/dev.098269>.
- 300 29. Günesdogan, U., Magnúsdóttir, E., and Surani, M.A. (2014). Primordial germ cell specification: A context-dependent
301 cellular differentiation event [corrected]. *Philos Trans R Soc Lond B Biol Sci* 369. <https://doi.org/10.1098/rstb.2013.0543>.
- 302 30. Bardot, E.S., and Hadjantonakis, A.-K. (2020). Mouse gastrulation: Coordination of tissue patterning, specification and
303 diversification of cell fate. *Mechanisms of Development* 163, 103617. <https://doi.org/10.1016/j.mod.2020.103617>.
- 304 31. Samanta, M.K., Gayen, S., Harris, C., Maclary, E., Murata-Nakamura, Y., Malcore, R.M., Porter, R.S., Garay, P.M.,
305 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>.
- 306 32. Koubova, J., Menke, D.B., Zhou, Q., Capel, B., Griswold, M.D., and Page, D.C. (2006). Retinoic acid regulates
307 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>.
- 308 33. Lin, Y., Gill, M.E., Koubova, J., and Page, D.C. (2008). Germ Cell-Intrinsic and -Extrinsic Factors Govern Meiotic
309 Initiation in Mouse Embryos. *Science* 322, 1685–1687. <https://doi.org/10.1126/science.1166340>.

- 310 34. Endo, T., Mikedis, M.M., Nicholls, P.K., Page, D.C., and De Rooij, D.G. (2019). Retinoic Acid and Germ Cell
311 Development in the Ovary and Testis. *Biomolecules* 9, 775. <https://doi.org/10.3390/biom9120775>.
- 312 35. Li, H., Liang, Z., Yang, J., Wang, D., Wang, H., Zhu, M., Geng, B., and Xu, E.Y. (2019). DAZL is a master translational
313 regulator of murine spermatogenesis. *Natl Sci Rev* 6, 455–468. <https://doi.org/10.1093/nsr/nwy163>.
- 314 36. Mikedis, M.M., Fan, Y., Nicholls, P.K., Endo, T., Jackson, E.K., Cobb, S.A., De Rooij, D.G., and Page, D.C. (2020).
315 DAZL mediates a broad translational program regulating expansion and differentiation of spermatogonial progenitors.
eLife 9, e56523. <https://doi.org/10.7554/eLife.56523>.

316 **Figures and Tables**

317 —

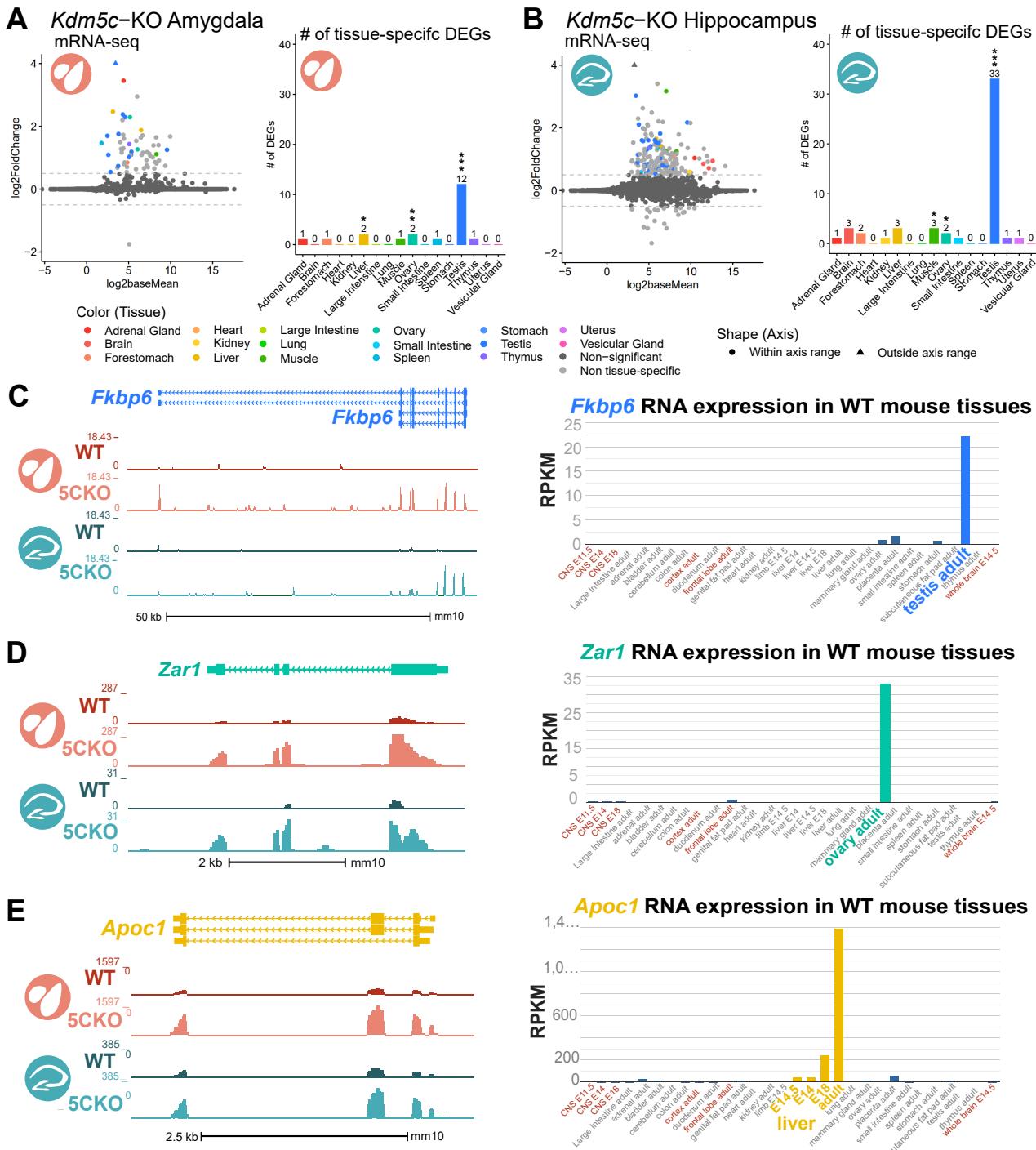


Figure 1: Tissue-enriched genes are misexpressed in the *Kdm5c*-KO brain. **A.** Expression of tissue-enriched genes in the male *Kdm5c*-KO amygdala. Left - MA plot of mRNA-seq. Right - Number of tissue-enriched differentially expressed genes (DEGs). * $p<0.05$, ** $p<0.01$, *** $p<0.001$, Fisher's exact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

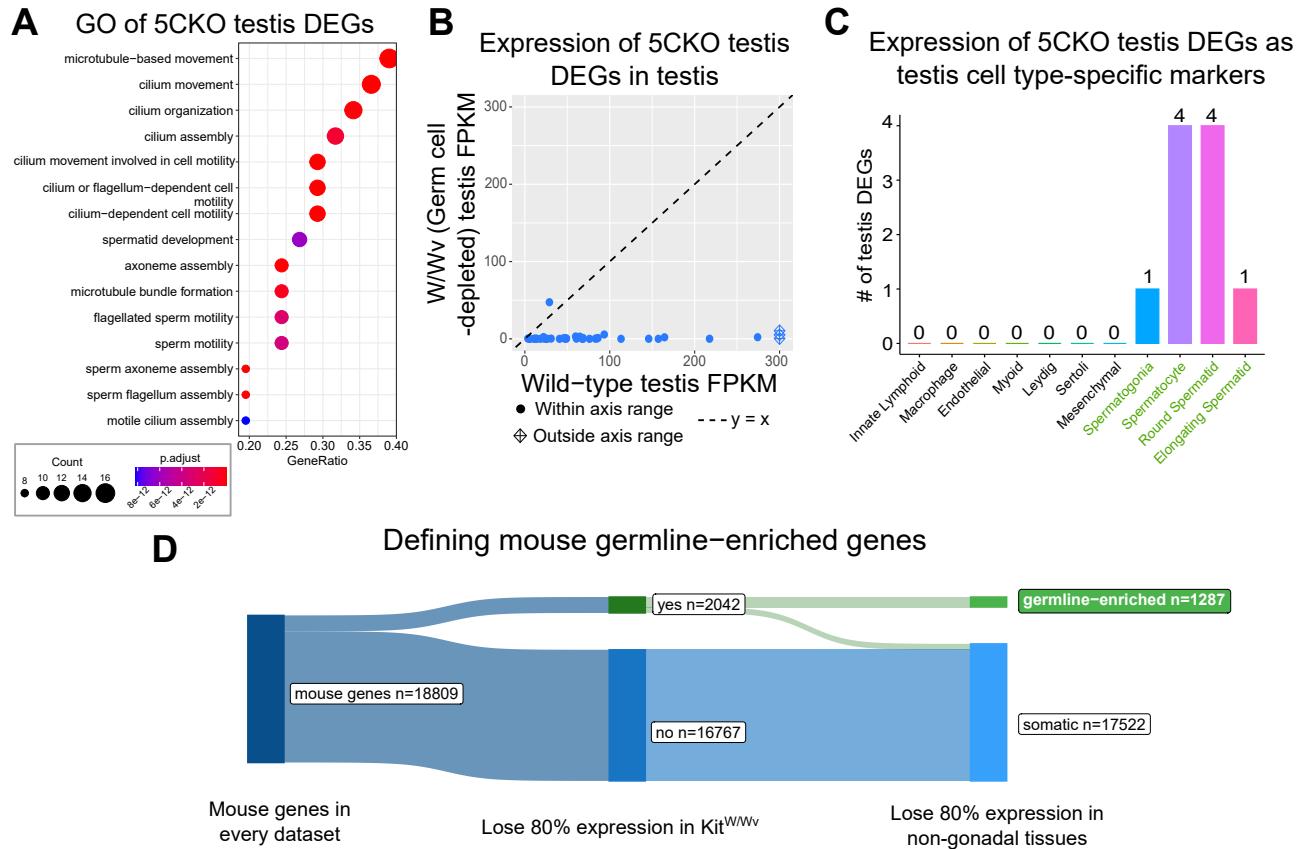


Figure 2: Aberrant transcription of germline genes in the *Kdm5c*-KO in the brain. **A.** enrichPlot gene ontology (GO) of *Kdm5c*-KO amygdala and hippocampus testis-enriched DEGs **B.** Expression of testis DEGs in wild-type (WT) testis versus germ cell-depleted (W/Wv) testis. Expression is in Fragments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

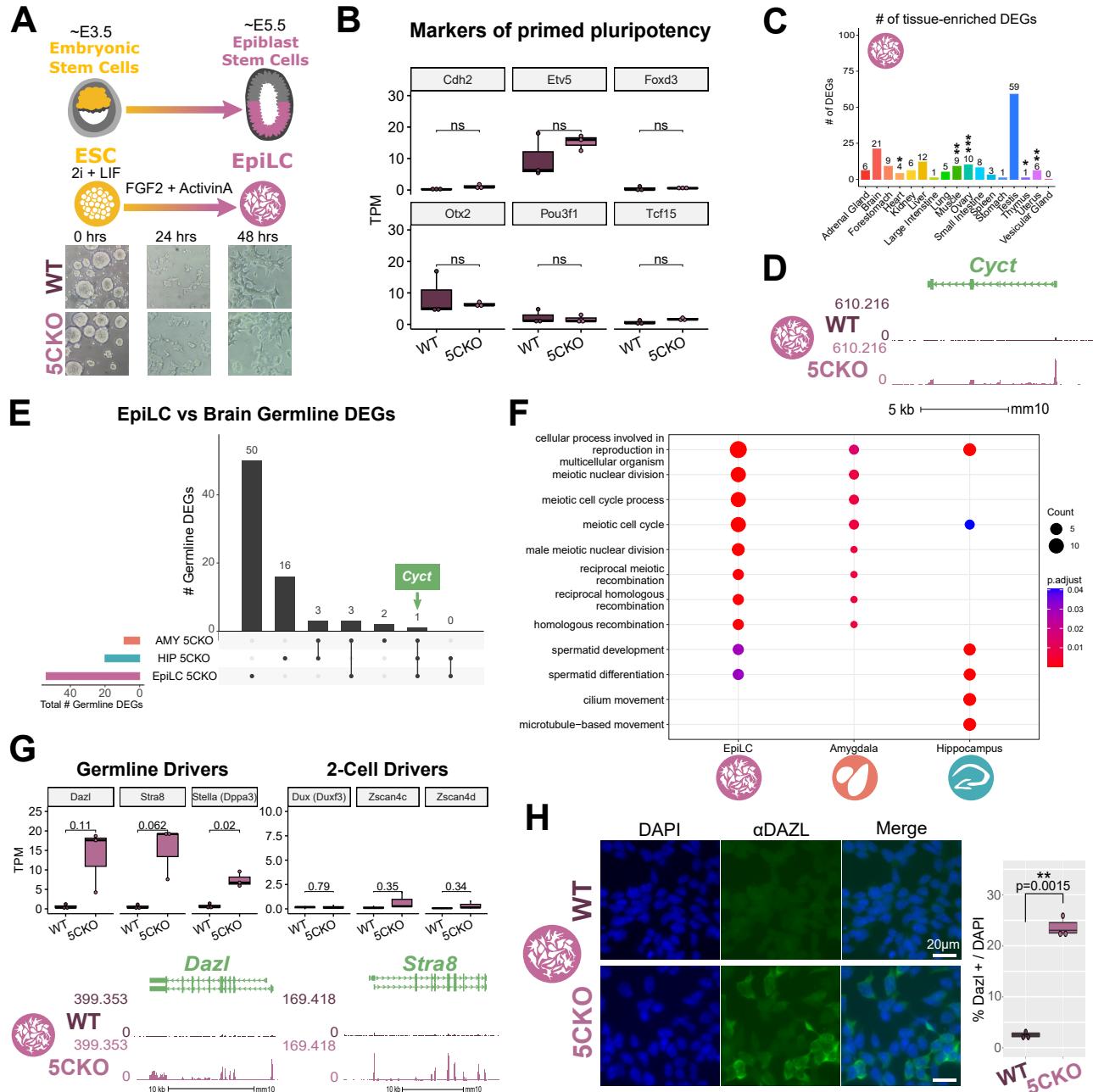


Figure 3: *Kdm5c*-KO epiblast-like cells express key drivers regulators of germline identity

A. Top - Diagram of *in vivo* differentiation of embryonic stem cells (ESCs) of the inner cell mass into epiblast stem cells. Middle - *in vitro* differentiation of ESCs into epiblast-like cells (EpiLCs). Bottom - representative images of wild-type (WT) and *Kdm5c*-KO ESC to EpiLC differentiation. Brightfield images taken at 20X. **B.** 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). **C.** Number of tissue-enriched differentially expressed genes (DEGs). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, Fisher's exact test. **D.** Average bigwigs of an example germline gene, *Cyct*, that is dysregulated *Kdm5c*-KO EpiLCs. **E.** Upset plot displaying the overlap of germline DEGs expressed in *Kdm5c*-KO EpiLCs, amygdala (AMY), and hippocampus (HIP) RNA-seq datasets. **F.** enrichPlot comparing enriched gene ontologies for *Kdm5c*-KO EpiLCs, amygdala, and hippocampus germline DEGs. **G.** Top - Example germline identity DEGs unique to EpiLCs, p-values for Welch's t-test. Bottom - Average bigwigs of *Dazl* and *Stra8* expression in wild-type and *Kdm5c*-KO EpiLCs. **H.** Immunocytochemistry of DAZL in male wild-type (WT) and *Kdm5c*-KO (5CKO) EpiLCs. Percentage of DAZL-positive cells normalized to DAPI, p-value for Welch's t-test

318 **Figure outline:**

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

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

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

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

338 **Figure 4: Loss of KDM5C's catalytic activity impairs DNAm placement and long-term silencing of germline
339 genes** * Increase in H3K4me3 in *Kdm5c*-KO amygdala at germline genes * Increase in H3K4me2 in EpiLCs at germline
340 genes * *Kdm5c* binding in EpiLCs vs PNCs to show that germline repression is happening in early embryo * Previous studies
341 only looked at ESCs, unknown if catalytic activity is required for long-term repression, especially since DNA methylation is
342 placed later). KDM5C RNA and protein ESC → EpiLC (increasing then decreasing) * RNA expression of germline genes with
343 catalytic dead rescue (Ilakkiya) * DNA methylation in WT and 5CKO EpiLCs (Ilakkiya)

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