

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 brain is a major effect of
115 KDM5C loss.

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

117 The testis contains both germ cells (meiotic cells, e.g. spermatogonia) and somatic cells (non-meiotic, e.g. Leydig cells)
118 that support hormone production and germline functions. Select testis-enriched DEGs that were characterized previously
119 had germline-specific functions⁹, suggesting *Kdm5c*-KO cells fail to demarcate between the soma and germline. To test if
120 this holds true for all *Kdm5c*-KO testis-biased DEGs, we first assed their function through gene ontology. We found high
121 enrichment of germline-relevant ontologies, including spermatid development (GO: 0007286, p.adjust = 6.2e-12) and sperm
122 axoneme assembly (GO: 0007288, p.adjust = 2.45e-14) (Figure 2A).

123 To further validate if these testis DEGs are truly germline genes, we then compared their expression in somatic versus
124 germ cells within the testis. We first compared their expression within wild-type versus germ cell-depleted testes²⁶. In this
125 study, germ cell depletion was accomplished by heterozygous *W* and *Wv* mutations in the enzymatic domain of *c-Kit* (*Kit*^{W/Wv}),
126 which prevents the maturation of germ cells and results in overall germline loss²⁷. Almost all *Kdm5c*-KO testis-enriched
127 DEGs lost expression with germ cell depletion (Figure 2B). The only testis-enriched DEG that did not show considerable
128 downregulation with germline depletion was *FK506 binding protein 6* (*Fkbp6*), the aforementioned testis gene that regulates
129 piRNA expression and meiosis in germ cells^{23,24}. We then assessed testis-enriched DEG expression in a published single
130 cell RNA-seq dataset that identified cell type-specific markers within the testis²⁸. We found that while some testis-enriched
131 DEGs were classified as specific markers for different germ cell types (e.g. spermatogonia, spermatocytes, round spermatids,
132 and elongating spermatids) none marked somatic cells (Figure 2C). Together, these data suggest the somatic *Kdm5c*-KO
133 brain aberrantly expresses germline genes.

134 We wanted to more deeply characterize germline gene misexpression with *Kdm5c* loss, but lacked a comprehensive list of
135 mouse germline-enriched genes. To facilitate downstream analyses, we generated a list of male and female germline-enriched
136 genes by evaluating gene expression in germline-depleted (*Kit*^{W/Wv}) mice. Current available *Kit*^{W/Wv} datasets included males
137 and females at embryonic day 12, 14, and 16²⁹, as well as adult male testes²⁶.

138 We defined genes as germline-enriched if their expression met the following criteria: 1) their expression is greater than 1
139 FPKM in wild-type gonads 2) their expression in any wild-type, non-gonadal tissue²⁰ does not exceed 20% of their maximum
140 expression in the wild-type germline, and 3) their expression in the germ cell-depleted germline, for any sex or time point,
141 does not exceed 20% of their maximum expression in the wild-type germline. These criteria yielded 1,288 germline-enriched
142 genes (Figure 2D), which was hereafter used as a resource for assessing misexpression of germline genes with loss of
143 *Kdm5c* (Supplementary table 1).

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

145 Misexpression of germline genes in the adult *Kdm5c*-KO brain suggests mutants fail to demarcate germline and somatic
146 cellular identity. Germ cells are typically distinguished from somatic cells soon after the embryo implants into the uterine
147 wall^{30,31} when a subset of epiblast stem cells become the primordial germ cells (PGCs) while the remainder differentiate into
148 the ectoderm, mesoderm, and endoderm to form the somatic tissues³². This developmental time point can be modeled *in*

149 *vitro* through differentiation of embryonic stem cells (ESCs) into post-implantation epiblast-like stem cells (EpiLCs) (Figure 3A,
150 top). Previous studies have demonstrated that while some germline-enriched genes are also expressed in embryonic stem
151 cells (ESCs), they are silenced as they differentiate into EpiLCs¹⁵. Therefore, we assessed if KDM5C was necessary for
152 initial germline gene silencing by evaluating the impact of *Kdm5c* loss in male EpiLCs.

153 We first identified *Kdm5c*-KO EpiLC DEGs through DESeq2 analysis of our previously published RNA-seq dataset³³
154 (\log_2 fold change > 0.5 , $q < 0.1$). Similar to the *Kdm5c*-KO brain, we observed general dysregulation of tissue-enriched
155 genes, with the largest number of genes belonging to the brain and testis, although they were not significantly enriched
156 (Figure 3B). Using our curated list, we found *Kdm5c*-KO EpiLCs aberrantly expressed 54 germline-enriched genes, including
157 the previously characterized hippocampal DEG⁹ *Cytochrome C, testis-specific (Cyct)* (Figure 3C). Although we observed
158 aberrant expression of many tissue-enriched genes, we did not observe any significant difference in primed pluripotency
159 genes (Figure 3D) or gross changes in *Kdm5c*-KO cell morphology during differentiation (Figure 3E), indicating KDM5C loss
160 does not impair EpiLC formation.

161 To evaluate if all germline DEGs are constitutively dysregulated or change over the course of development, we then
162 compared EpiLC germline DEGs to those expressed in the *Kdm5c*-KO brain. The majority of EpiLC germline DEGs were
163 unique to EpiLCs, with only *Cyct* shared across all sequencing datasets (Figure 3F). We then compared the function of
164 EpiLC and brain germline DEGs and found particularly high enrichment of meiosis-related gene ontologies in EpiLCs (Figure
165 3G), such as meiotic cell cycle (GO: 0051321, $p.adjust = 4.44e-07$). While there was modest enrichment of meiotic gene
166 ontologies in both brain regions, the *Kdm5c*-KO hippocampus showed strong enrichment of late-stage sperm genes, such
167 as those involved in the sperm axoneme. This shift from meiotic genes to later spermatogenesis genes in the hippocampus
168 suggests the germline developmental program could occur ectopically as *Kdm5c*-KO cells progresses through somatic
169 tissue development. **note: this is strengthened by the ChIP-seq data since KDM5C is not directly bound to many**
170 **brain/flagellar DEGs. This point might be stronger in the ChIPseq figure**

171 Interestingly, DEGs unique to *Kdm5c*-KO EpiLCs included key drivers of germline identity, such as *Stimulated by retinoic*
172 *acid 8 (Stra8)* and *Deleted in azoospermia like (Dazl)* (Figure 3H). These genes are typically expressed during embryonic
173 germ cell development to commit PGCs to the germline fate, but are also expressed later in life to trigger meiotic gene
174 expression programs^{34–36}. Of note, some germline genes, including *Dazl*, are also expressed in the two-cell embryo. However,
175 we did not see misexpression of two-cell embryo-specific genes, like *Zscan4d* ($q = 0.381$), indicating *Kdm5c*-KO in EpiLCs do
176 not revert back to a 2-cell state (Figure 3H).

177 *Dazl* is an RNA-binding protein that regulates the translation of germline mRNAs and is essential for germ cell function³⁷.
178 A significant portion of germline transcripts misexpressed in *Kdm5c*-KO EpiLCs are known binding targets of DAZL, including
179 *Stra8*³⁸ ($p = 1.698e-07$, Fisher's Exact Test). This suggests expression of DAZL protein could promote the translation of other
180 aberrant germline transcripts, influencing their ability to impact *Kdm5c*-KO cellular function. We thus tested DAZL protein
181 expression in *Kdm5c*-KO EpiLCs through immunocytochemistry (Figure 3H). We observed about 25% of *Kdm5c*-KO EpiLCs
182 expressed DAZL protein and it was localized to the cytoplasm ($p = 0.0015$, Welch's t-test), consistent with the pattern of DAZL
183 expression in spermatogonia³⁸. Altogether these results suggest tissue-specific genes are misexpressed during *Kdm5c*-KO
184 embryogenesis, including key drivers of germline identity that can be translated into protein.

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

- 186 • **note: do Direct vs indirect DEGs motif analysis**
187 • However, it is currently unclear if KDM5C binds to all germline DEGs and if its binding is maintained at any germline

188 genes in neurons.

189 Previous work suggests KDM5C represses germline genes during early development, as re-expression of KDM5C in
190 knockout neuronal cultures fails to suppress hippocampal germline DEGs and KDM5C is not bound to their promoters
191 in neurons⁹. There is some evidence KDM5C binds to select germline gene promoters in ESCs⁹, including two recent
192 independent screens that found KDM5C binds to Dazl's promoter^{18,19}. As KDM5C's binding at germline gene promoters has
193 not been systematically characterized, it is currently unclear what types of germline genes KDM5C regulates and if its binding
194 is maintained at any germline genes in neurons.

195 To further characterize the mechanism behind *Kdm5c*-KO germline gene misexpression, we analyzed KDM5C chromatin
196 immunoprecipitation followed by DNA sequencing (ChIP-seq) datasets in EpiLCs³³ and primary neuron cultures (PNCs) from
197 the cortex and hippocampus¹³. EpiLCs had a higher total number of KDM5C peaks than PNCs (EpiLCs: 5,808, PNCs: 1,276,
198 MACS2 q < 0.1 and fold enrichment > 1) and KDM5C was primarily localized to gene promoters in both cell types (EpiLCs:
199 4,190, PNCs: 745 +/- 500kb from TSS), although PNCs showed increased localization to non-promoter regions (Figure 4A).

200 The majority of promoters bound by KDM5C in PNCs were also bound in EpiLCs (513 shared promoters), however a
201 large portion of gene promoters were only bound by KDM5C in EpiLCs (3677 EpiLC only promoters) (Figure 4B). We then
202 performed gene ontology analysis to compare the functions of genes bound by KDM5C at the promoter in different cell
203 types. While there were no significant ontologies for genes with KDM5C only bound in PNCs, gene ontologies for peaks
204 shared between PNCs and EpiLCs were enriched for functions involving nucleic acid turnover, such as deoxyribonucleotide
205 metabolic process (GO:0009262, p.adjust = 8.28e-05) (Figure 4C). Germline-specific ontologies were only enriched in
206 peaks unique to EpiLCs, such as meiotic nuclear division (GO: 0007127 p.adjust = 6.77e-16) and meiotic cell cycle process
207 (GO:1903046, p.adjust = 5.05e-16) (Figure 3C). When comparing KDM5C binding at all germline gene promoters, KDM5C
208 was only bound to a subset of germline gene promoters in EpiLCs and was not bound to any in PNCs (Figure 4D). Together,
209 this suggests KDM5C is recruited to a subset of germline gene promoters in EpiLCs, including meiotic genes, but does not
210 regulate germline genes in neurons.

211 To determine if the germline mRNAs expressed in the *Kdm5c*-KO brain and EpiLCs are direct targets of KDM5C, we then
212 compared KDM5C binding at RNA-seq DEG promoters (Figure 4E). About one third of EpiLC specific and brain-specific
213 (hippocampus, amygdala, or both) germline DEGs were bound by KDM5C in EpiLCs (EpiLC only: 36%, Brain only: 33.3%).
214 Some notable differences in KDM5C binding for EpiLC-specific DEGs included *Dazl* and *Stra8*, two key drivers of germline
215 identity discussed above. Although both mRNAs are expressed in *Kdm5c*-KO EpiLCs, KDM5C is only bound to *Dazl*'s
216 promoter and not *Stra8*'s in EpiLCs (Figure 4F). In contrast to the unique DEGs, 3 out of the 4 genes dysregulated in both the
217 brain and EpiLCs bound by KDM5C (Figure 4G). Again, we did not observe any KDM5C binding at germline gene promoters
218 in PNCs, even for brain-specific DEGs (Figure 4H). Altogether, this suggests the majority of germline mRNAs expressed in
219 *Kdm5c*-KO cells are dysregulated independent of direct KDM5C recruitment to their promoters during embryogenesis.

220 **KDM5C erases H3K4me3 to promote long-term repression of germline genes via DNA methylation**

221 Although KDM5C is generally thought to suppress transcription through erasure of histone 3 lysine 4 di- and trimethylation
222 (H3K4me2/3)³⁹, recent studies in ESCs have suggested KDM5C's repression *Dazl* is independent of its catalytic activity¹⁸.
223 Somatic repression of germline genes is typically established during the transition between naive and primed pluripotency,
224 which modeled in vitro as ESCs to EpiLC differentiation. In ESCs, chromatin regulators place repressive histone modifications
225 at germline gene promoters (**note: check**), including H2AK119ub1 and H3K9me3. Germline genes are then silenced

226 long-term by CpG DNA methylation in EpiLCs germline genes. As KDM5C's embryonic role in germline gene repression has
227 only been characterized in ESCs, it is currently unclear if KDM5C is involved in the transition between ESCs and EpiLCs.

228 It is currently unclear if KDM5C's catalytic activity aids in the long-term repression of germline genes, such as enabling
229 the placement of DNA methylation.

230 • In congruence with previous work in the *Kdm5c*-KO hippocampus⁹, we observed ectopic H3K4me3 peaks in the
231 *Kdm5c*-KO amygdala at the transcription start site (TSS) of germline genes (Figure 4).

232 • We additionally observed a marked increase in H3K4me2 at germline gene promoters in *Kdm5c*-KO EpiLCs.

233 – This increase was greatest at germline genes directly bound by KDM5C (?)

234 Discussion

235 • The demarcation of the germ vs soma is a key feature of multicellularity

236 • Other H3K4me regulators, anything known about tissue-biased gene expression?

237 • This suggests the germline developmental program is occurring ectopically as *Kdm5c*-KOs progresses through somatic
238 tissue development

239 • tissue-biased gene expression:

240 – However unlike the gonadal-biased DEGs, many liver and muscle-biased DEGs have a known involvement in brain
241 function. For example, the liver-biased DEG is *Apolipoprotein C-I (Apoc1)*, is important for lipoprotein metabolism
242 but has also been shown to influence learning and memory (Figure 1E).

243 • Otx2 is properly expressed in *Kdm5c*-KO EpiLCs and is known to prevent pgc identity <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6485399/>

245 • Altogether, these data indicate that while some germline genes are misexpressed due to direct loss of KDM5C binding
246 during emryogenesis, secondary downstream mechanisms can also promote their aberrant transcription.

247 • Papers to read/reference:

248 – 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)

250 – 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/>

252 References

- 253 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>.
- 254 2. Jenuwein, T., and Allis, C.D. (2001). Translating the Histone Code. *Science* *293*, 1074–1080. <https://doi.org/10.1126/science.1063127>.

- 257 3. Lewis, E.B. (1978). A gene complex controlling segmentation in *Drosophila*. *Nature* *276*, 565–570. <https://doi.org/10.1038/276565a0>.
- 258 4. Kennison, J.A., and Tamkun, J.W. (1988). Dosage-dependent modifiers of polycomb and antennapedia mutations in
259 *Drosophila*. *Proc. Natl. Acad. Sci. U.S.A.* *85*, 8136–8140. <https://doi.org/10.1073/pnas.85.21.8136>.
- 260 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
261 al. (2006). Brain-specific phosphorylation of MeCP2 regulates activity-dependent Bdnf transcription, dendritic growth,
262 and spine maturation. *Neuron* *52*, 255–269. <https://doi.org/10.1016/j.neuron.2006.09.037>.
- 263 6. Hirabayashi, Y., Suzuki, N., Tsuboi, M., Endo, T.A., Toyoda, T., Shinga, J., Koseki, H., Vidal, M., and Gotoh, Y. (2009).
264 Polycomb Limits the Neurogenic Competence of Neural Precursor Cells to Promote Astrogenic Fate Transition. *Neuron* *63*, 600–613. <https://doi.org/10.1016/j.neuron.2009.08.021>.
- 265 7. Kassis, J.A., Kennison, J.A., and Tamkun, J.W. (2017). Polycomb and Trithorax Group Genes in *Drosophila*. *Genetics* *206*, 1699–1725. <https://doi.org/10.1534/genetics.115.185116>.
- 266 8. Schaefer, A., Sampath, S.C., Intrator, A., Min, A., Gertler, T.S., Surmeier, D.J., Tarakhovsky, A., and Greengard, P.
267 (2009). Control of cognition and adaptive behavior by the GLP/G9a epigenetic suppressor complex. *Neuron* *64*,
268 678–691. <https://doi.org/10.1016/j.neuron.2009.11.019>.
- 269 9. Scandaglia, M., Lopez-Atalaya, J.P., Medrano-Fernandez, A., Lopez-Cascales, M.T., Del Blanco, B., Lipinski, M., Benito,
270 E., Olivares, R., Iwase, S., Shi, Y., et al. (2017). Loss of Kdm5c Causes Spurious Transcription and Prevents the Fine-
271 Tuning of Activity-Regulated Enhancers in Neurons. *Cell Rep* *21*, 47–59. <https://doi.org/10.1016/j.celrep.2017.09.014>.
- 272 10. Claes, S., Devriendt, K., Van Goethem, G., Roelen, L., Meireleire, J., Raeymaekers, P., Cassiman, J.J., and Fryns, J.P.
273 (2000). Novel syndromic form of X-linked complicated spastic paraplegia. *Am J Med Genet* *94*, 1–4.
- 274 11. Jensen, L.R., Amende, M., Gurok, U., Moser, B., Gimmel, V., Tzschach, A., Janecke, A.R., Tariverdian, G., Chelly, J.,
275 Fryns, J.P., et al. (2005). Mutations in the JARID1C gene, which is involved in transcriptional regulation and chromatin
276 remodeling, cause X-linked mental retardation. *Am J Hum Genet* *76*, 227–236. <https://doi.org/10.1086/427563>.
- 277 12. Carmignac, V., Nambot, S., Lehalle, D., Callier, P., Moortgat, S., Benoit, V., Ghoumid, J., Delobel, B., Smol, T., Thuillier,
278 C., et al. (2020). Further delineation of the female phenotype with KDM5C disease causing variants: 19 new individuals
279 and review of the literature. *Clin Genet* *98*, 43–55. <https://doi.org/10.1111/cge.13755>.
- 280 13. Iwase, S., Brookes, E., Agarwal, S., Badeaux, A.I., Ito, H., Vallianatos, C.N., Tomassy, G.S., Kasza, T., Lin, G.,
281 Thompson, A., et al. (2016). A Mouse Model of X-linked Intellectual Disability Associated with Impaired Removal of
282 Histone Methylation. *Cell Reports* *14*, 1000–1009. <https://doi.org/10.1016/j.celrep.2015.12.091>.
- 283 14. Endoh, M., Endo, T.A., Shinga, J., Hayashi, K., Farcas, A., Ma, K.W., Ito, S., Sharif, J., Endoh, T., Onaga, N., et
284 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>.
- 285 15. Mochizuki, K., Sharif, J., Shirane, K., Uranishi, K., Bogutz, A.B., Janssen, S.M., Suzuki, A., Okuda, A., Koseki, H.,
286 and Lorincz, M.C. (2021). Repression of germline genes by PRC1.6 and SETDB1 in the early embryo precedes DNA
287 methylation-mediated silencing. *Nat Commun* *12*, 7020. <https://doi.org/10.1038/s41467-021-27345-x>.
- 288 16. Velasco, G., Hubé, F., Rollin, J., Neuillet, D., Philippe, C., Bouzinba-Segard, H., Galvani, A., Viegas-Péquignot, E., and
289 Francastel, C. (2010). Dnmt3b recruitment through E2F6 transcriptional repressor mediates germ-line gene silencing
290 in murine somatic tissues. *Proc Natl Acad Sci U S A* *107*, 9281–9286. <https://doi.org/10.1073/pnas.1000473107>.

- 285 17. Hackett, J.A., Reddington, J.P., Nestor, C.E., Duncan, D.S., Branco, M.R., Reichmann, J., Reik, W., Surani, M.A.,
286 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>.
- 287 18. Gupta, N., Yakhou, L., Albert, J.R., Azogui, A., Ferry, L., Kirsh, O., Miura, F., Battault, S., Yamaguchi, K., Laisné,
288 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>.
- 289 19. Al Adhami, H., Vallet, J., Schaal, C., Schumacher, P., Bardet, A.F., Dumas, M., Chicher, J., Hammann, P., Daujat,
290 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>.
- 291 20. Li, B., Qing, T., Zhu, J., Wen, Z., Yu, Y., Fukumura, R., Zheng, Y., Gondo, Y., and Shi, L. (2017). A Comprehensive
Mouse Transcriptomic BodyMap across 17 Tissues by RNA-seq. *Sci Rep* *7*, 4200. <https://doi.org/10.1038/s41598-017-04520-z>.
- 293 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.,
294 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>.
- 295 22. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data
296 with DESeq2. *Genome Biol* *15*, 550. <https://doi.org/10.1186/s13059-014-0550-8>.
- 297 23. Crackower, M.A., Kolas, N.K., Noguchi, J., Sarao, R., Kikuchi, K., Kaneko, H., Kobayashi, E., Kawai, Y., Kozieradzki, I.,
298 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>.
- 299 24. Xiol, J., Cora, E., Koglgruber, R., Chuma, S., Subramanian, S., Hosokawa, M., Reuter, M., Yang, Z., Berninger, P.,
300 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>.
- 301 25. Cheng, S., Altmeppen, G., So, C., Welp, L.M., Penir, S., Ruhwedel, T., Menelaou, K., Harasimov, K., Stützer, A.,
302 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>.
- 303 26. Mueller, J.L., Skaletsky, H., Brown, L.G., Zaghlul, S., Rock, S., Graves, T., Auger, K., Warren, W.C., Wilson, R.K., and
304 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>.
- 305 27. Handel, M.A., and Eppig, J.J. (1979). Sertoli Cell Differentiation in the Testes of Mice Genetically Deficient in Germ
306 Cells. *Biology of Reproduction* *20*, 1031–1038. <https://doi.org/10.1095/biolreprod20.5.1031>.
- 307 28. Green, C.D., Ma, Q., Manske, G.L., Shami, A.N., Zheng, X., Marini, S., Moritz, L., Sultan, C., Gurczynski, S.J., Moore,
308 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>.
- 309 29. 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>.

- 311 30. Magnúsdóttir, E., and Surani, M.A. (2014). How to make a primordial germ cell. *Development* *141*, 245–252.
312 https://doi.org/10.1242/dev.098269.
- 313 31. Günesdogan, U., Magnúsdóttir, E., and Surani, M.A. (2014). Primordial germ cell specification: A context-dependent
314 cellular differentiation event [corrected]. *Philos Trans R Soc Lond B Biol Sci* *369*. https://doi.org/10.1098/rstb.2013.0543.
- 315 32. Bardot, E.S., and Hadjantonakis, A.-K. (2020). Mouse gastrulation: Coordination of tissue patterning, specification and
316 diversification of cell fate. *Mechanisms of Development* *163*, 103617. https://doi.org/10.1016/j.mod.2020.103617.
- 317 33. Samanta, M.K., Gayen, S., Harris, C., Maclary, E., Murata-Nakamura, Y., Malcore, R.M., Porter, R.S., Garay, P.M.,
318 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.
- 319 34. Koubova, J., Menke, D.B., Zhou, Q., Capel, B., Griswold, M.D., and Page, D.C. (2006). Retinoic acid regulates
320 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.
- 321 35. Lin, Y., Gill, M.E., Koubova, J., and Page, D.C. (2008). Germ Cell-Intrinsic and -Extrinsic Factors Govern Meiotic
322 Initiation in Mouse Embryos. *Science* *322*, 1685–1687. https://doi.org/10.1126/science.1166340.
- 323 36. Endo, T., Mikedis, M.M., Nicholls, P.K., Page, D.C., and De Rooij, D.G. (2019). Retinoic Acid and Germ Cell
324 Development in the Ovary and Testis. *Biomolecules* *9*, 775. https://doi.org/10.3390/biom9120775.
- 325 37. Li, H., Liang, Z., Yang, J., Wang, D., Wang, H., Zhu, M., Geng, B., and Xu, E.Y. (2019). DAZL is a master translational
326 regulator of murine spermatogenesis. *Natl Sci Rev* *6*, 455–468. https://doi.org/10.1093/nsr/nwy163.
- 327 38. Mikedis, M.M., Fan, Y., Nicholls, P.K., Endo, T., Jackson, E.K., Cobb, S.A., De Rooij, D.G., and Page, D.C. (2020).
DAZL mediates a broad translational program regulating expansion and differentiation of spermatogonial progenitors.
328 *eLife* *9*, e56523. https://doi.org/10.7554/eLife.56523.
- 329 39. Iwase, S., Lan, F., Bayliss, P., De La Torre-Ubieta, L., Huarte, M., Qi, H.H., Whetstine, J.R., Bonni, A., Roberts, T.M.,
and Shi, Y. (2007). The X-Linked Mental Retardation Gene SMCX/JARID1C Defines a Family of Histone H3 Lysine 4
330 Demethylases. *Cell* *128*, 1077–1088. https://doi.org/10.1016/j.cell.2007.02.017.

331 **Figures and Tables**

332 —

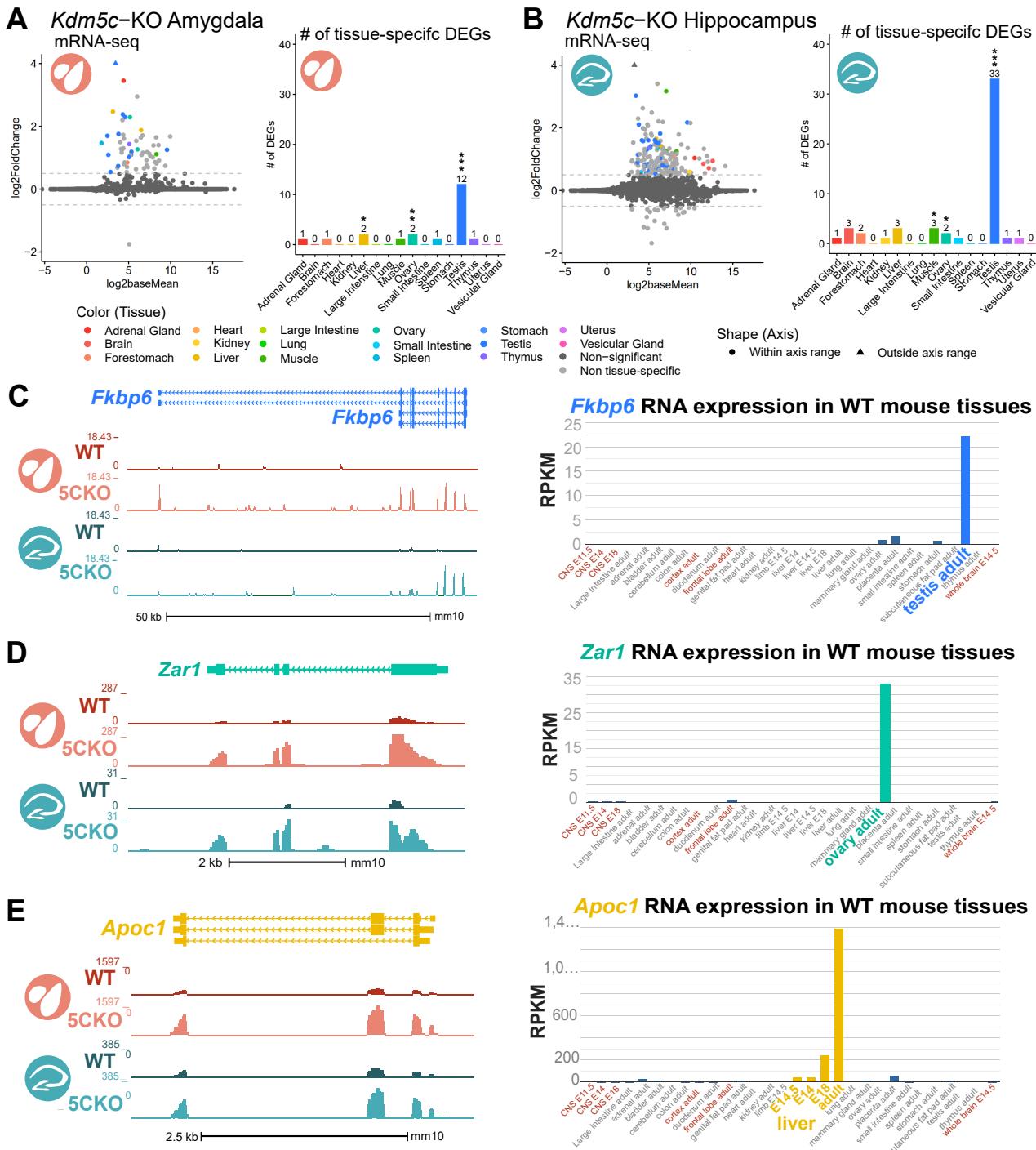


Figure 1: Tissue-enriched genes are misexpressed in the *Kdm5c*-KO brain. **A.** Expression of tissue-enriched genes in the male *Kdm5c*-KO amygdala. Left - MA plot of mRNA-sequencing. Right - Number of tissue-enriched differentially expressed genes (DEGs). * $p<0.05$, ** $p<0.01$, *** $p<0.001$, Fisher's 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 (Cyclt)* in the wild-type (WT) and *Kdm5c*-KO (5CKO) amygdala (red) and hippocampus (teal). Right - Expression of *Cyclt* 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.

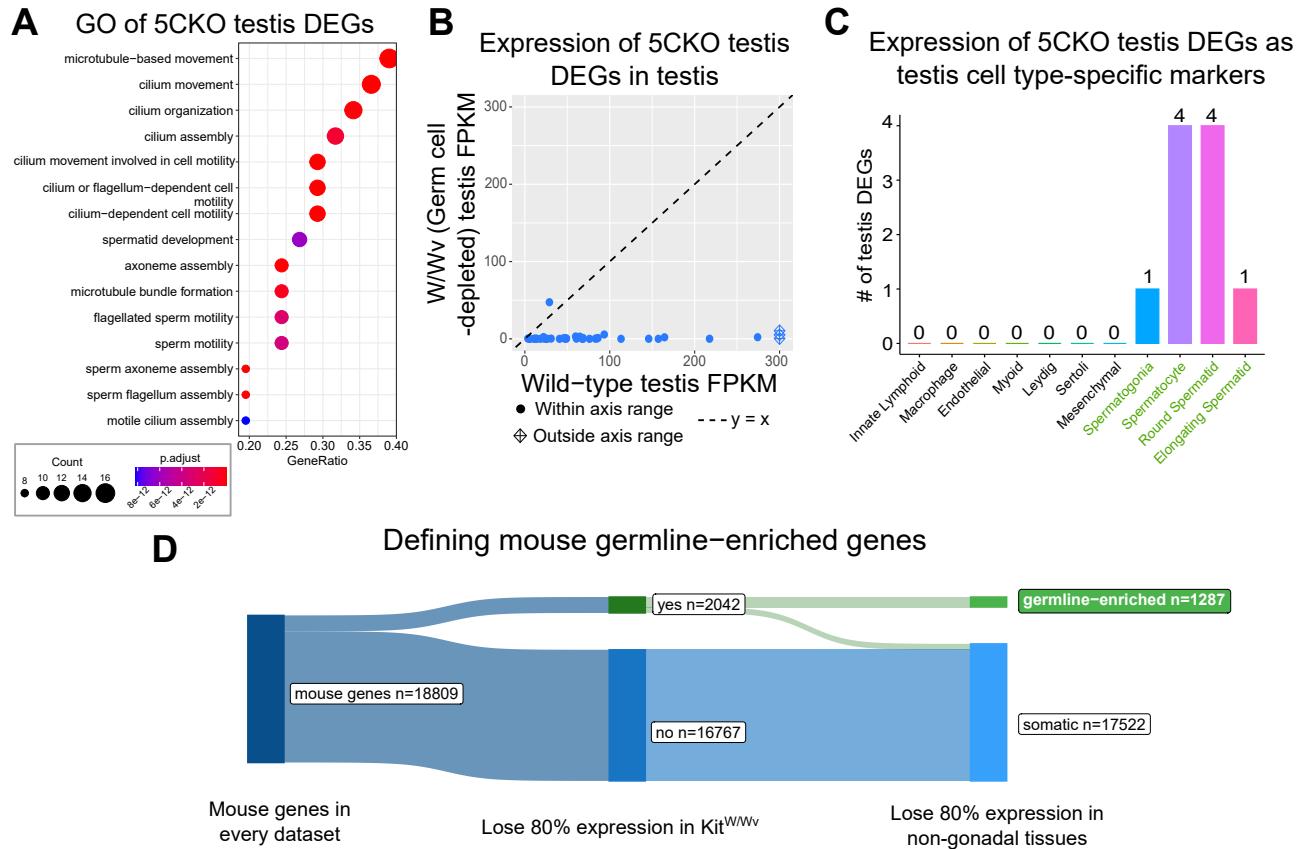


Figure 2: Aberrant transcription of germline genes in the *Kdm5c*-KO in the brain. **A.** enrichPlot gene ontology (GO) of *Kdm5c*-KO amygdala and hippocampus testis-enriched DEGs **B.** Expression of testis DEGs in wild-type (WT) testis versus germ cell-depleted (W/Wv) testis. Expression is in Fragments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

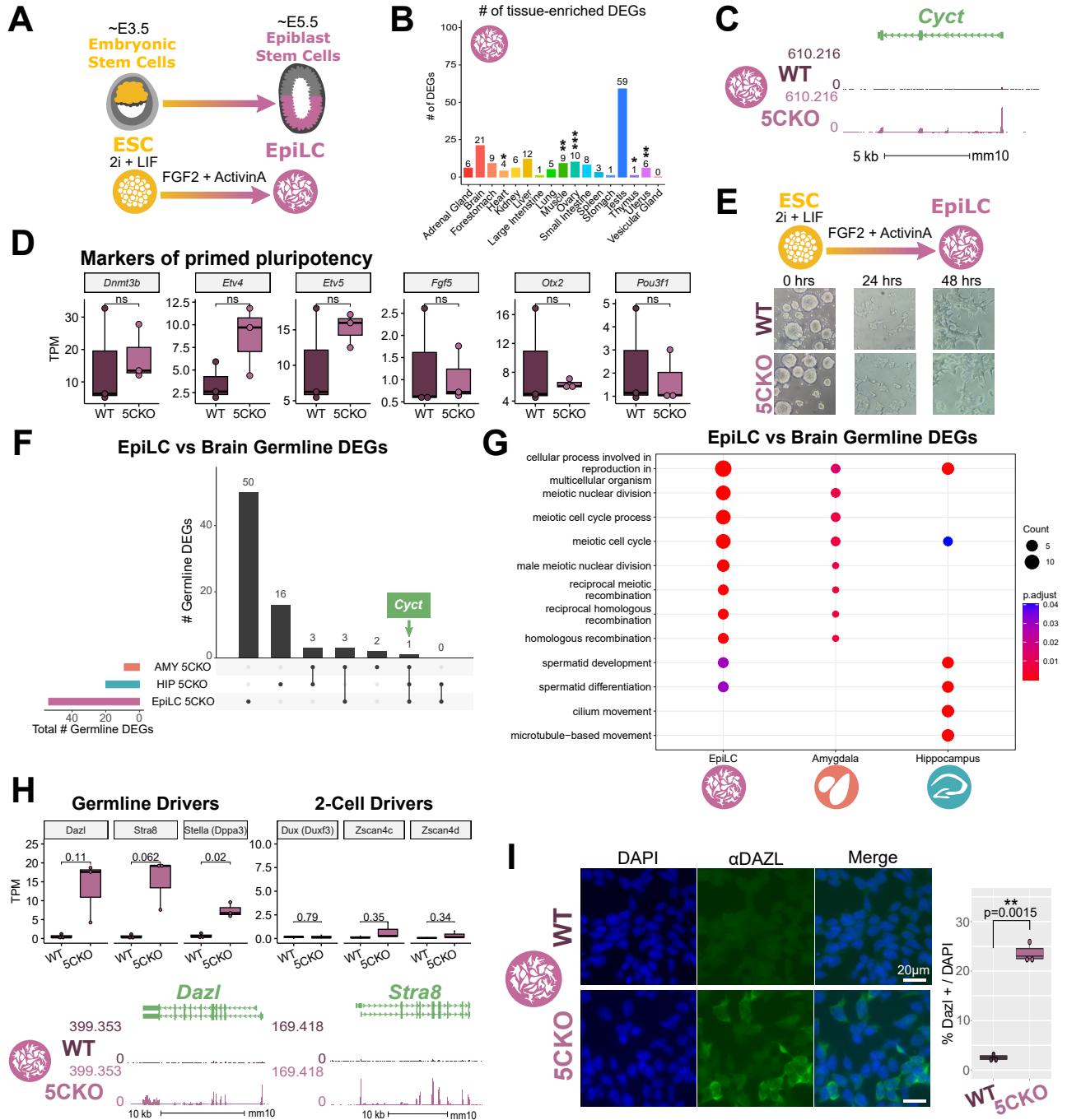


Figure 3: Kdm5c-KO epiblast-like cells express key drivers regulators of germline identity **A.** Top - Diagram of *in vivo* differentiation of embryonic stem cells (ESCs) of the inner cell mass into epiblast stem cells. Middle - *in vitro* differentiation of ESCs into epiblast-like cells (EpiLCs). Bottom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

333 **Figure outline:**

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

337 **Figure 2: The male *Kdm5c*-KO brain expresses male and female germline-enriched genes** * Gene ontology of
338 testis DEGs in the amygdala and hippocampus - ontologies are germline ontologies * Expression of testis DEGs in germline-
339 depleted testis (this is adult testis data) * scRNASeq of testis - # of testis DEGs that are germline-specific markers * Although
340 far fewer, 5CKO brain also expresses ovary-enriched genes(NCBI and bigwigs of Zar1) * These ovary enriched genes are
341 also germline specific (NCBI/Li tissues are in adult ovary, so it would be best to show they're oocyte-specific in adult ovary.
342 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
343 W/Wv data since oocytes are developed at this point? Or both?) * Defining what is/isn't a germline gene, and which are
344 male/female biased using embryonic W/Wv data

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

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

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

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