

1 Somatic misexpression of germline programs with loss of the X-linked
2 intellectual 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. Although initially identified for their roles in cellular identity^{1,2}, recent advancements in next
40 generation sequencing technologies unexpectedly found many neurodevelopmental disorders (NDDs) are caused by or linked
41 to mutations in chromatin regulators. This relationship is partially explained by their regulation of brain-specific genes or
42 chromatin states, such as modulating genes involved in synaptic maturation³ or the transition between neuronal and glial
43 developmental programs⁴. However, loss of some chromatin regulators can also lead to the misexpression of tissue-specific
44 genes outside of their intended environment^{1,2,5}. Currently, very few studies have explored the misexpression of non-neuronal,
45 tissue-specific genes in chromatin-linked neurodevelopmental disorders^{6,7} 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 provide the basis for identifying
49 molecular footholds that can then be exploited for future studies and potential therapeutics.

50 In this study, we characterized the misexpression of tissue-specific with loss of the chromatin regulator lysine demethylase
51 5C (KDM5C, also known as SMCX or JARID1C), a histone 3 lysine 4 demethylase. Pathogenic mutations in *KDM5C*
52 cause Intellectual Developmental Disorder, X-linked, Syndromic, Claes-Jensen Type (MRXSCJ, OMIM: 300534), whose
53 features include short stature, intellectual disability, seizures, aberrant aggression, and autistic behaviors^{8–10}. Previous work
54 has demonstrated *Kdm5c* knockout (-KO) mice recapitulate key MRXSCJ patient phenotypes, including hyperaggression
55 and learning impairments¹¹. Next generation RNA sequencing (RNA-seq) in the *Kdm5c*-KO hippocampus surprisingly
56 revealed ectopic expression of testis-enriched genes within the brain⁷. However, it is currently unclear if misexpression in the
57 *Kdm5c*-KO brain is unique to male germline genes, as other tissue-specific genes have not been systematically evaluated.

58 Interestingly, some of the ectopic testis genes identified in the *Kdm5c*-KO brain are typically expressed in germ cells, such
59 as *Cytochrome C, Testis-specific*, a sperm mitochondrial gene[cit].

- 60 • Germ cells are meiotic cells that pass on their genetic material to the next generation and are typically distinguished
61 from somatic cells very early on in embryogenesis.
- 62 • The demarcation of the germ vs soma is a key feature of multicellularity
- 63 • Germ cells, i.e. cells that undergo meiosis, are typically distinguished from somatic cells very early on in embryogenesis
64 and this distinction is a key feature of multicellularity.
- 65 • However, the testis also contains somatic cells that support germline function, and it is currently unclear if it is currently
66 unclear if all testis- (maybe too specific)
 - 67 – Chromatin regulators are very important for decommissioning germline genes and act successively the embryo
68 implants into the uterine wall
 - 69 * Most studies have focused on ESCs, which have a similar transcriptome to germ cells / 2-cells
 - 70 * recently, KDM5C was shown to repress DAZL in ESCs, independent of its catalytic activity
 - 71 * However, DNA methylation is lost in the mature 5CKO brain, DNA methylation is placed later and it's unclear if
72 it's required for long-term repression (maybe too specific, just trying to go into the fact that the mechanism is
73 partially understood but unclear)

74 Furthermore, the testis contains both somatic and germline

75 It is currently unclear if these germline genes

76 * Systematic characterization of ectopic germline genes hasn't been done

77 * unknown if other germline-enriched genes are dysregulated, including oocyte-specific genes

78 * Crucially, it's unknown if misexpression of the germline program leads to functional consequences in 5CKO

79 However, it is

80 Germ cells (meiotic cells) are typically distinguished from somatic cells very early on in embryogenesis and is a key
81 feature of multicellularity

82 * Misexpression of tissue-enriched genes hasn't been systematically characterized - Unclear if these genes are

83

84 * Interestingly, these genes (Cyct, D1pas1) typically function in the germline

85 * Germ cells (meiotic cells) are typically distinguished from somatic cells very early on in embryogenesis and

86 * Chromatin regulators are very important for decommissioning germline genes and act successively the embryo i

87 * Most studies have focused on ESCs, which have a similar transcriptome to germ cells / 2-cells

88 * recently, KDM5C was shown to repress DAZL in ESCs, independent of its catalytic activity

89 * However, DNA methylation is lost in the mature 5CKO brain, DNA methylation is placed later and it's Unclear

90 Answering these questions will help identify molecular footholds that can be exploited to test their overall contribution to
91 neuronal impairments that will lay the ground work for therapeutic intervention.

92 Loss of KDM5C can result in the misexpression of genes typically only found in the testis * Misexpression of tissue-enriched

93 genes hasn't been systematically characterized - Unclear if these genes are exceptions or if other tissue-specific genes are

94 dysregulated * Interestingly, these genes (Cyct, D1pas1) typically function in the germline * Germ cells (meiotic cells) are

95 typically distinguished from somatic cells very early on in embryogenesis and is a key feature of multicellularity * Chromatin

96 regulators are very important for decommissioning germline genes and act successively the embryo implants into the uterine

97 wall * Most studies have focused on ESCs, which have a similar transcriptome to germ cells / 2-cells * recently, KDM5C was

98 shown to repress DAZL in ESCs, independent of its catalytic activity * However, DNA methylation is lost in the mature 5CKO

99 brain, DNA methylation is placed later and it's Unclear if it's required for long-term repression (maybe too specific, just trying

100 to go into the fact that the mechanism is partially understood but unclear) * Systematic characterization of ectopic germline

101 genes hasn't been done * unknown if other germline-enriched genes are dysregulated, including oocyte-specific genes *

102 Crucially, it's unknown if misexpression of the germline program leads to functional consequences in 5CKO cells.

103 To determine the

104 • To ultimately determine their contribution, it is essential to characterize the types of genes expressed and the mechanism
105 behind their de-repression.

106 – Necessary to first characterize the mechanism behind their derepression to identify molecular footholds into testing
107 their contribution to neuronal impairments and potential for therapeutic intervention

108 – However, loss of some chromatin regulators can also lead to ectopic expression of tissue-enriched genes

109 * Very few studies have looked at these genes and it's unclear if these genes contribute to NDD impairments.

110 – Necessary to first characterize the mechanism behind their derepression to identify molecular footholds into testing
111 their contribution to neuronal impairments and potential for therapeutic intervention

112 • For example, histone 3 lysine 4 methylation (H3K4me)

113 • ¹².

114 • This relationship is partially It is currently unc

115 • There is a strong link between H3K4me regulation and NDDs, as mutation of 11 out of 15 genes that read, write, or
116 erase H3K4 mono-, di-, and tri- methylation result in varying neurodevelopmental impairments and syndromic NDDs¹³.
117 Yet it is still unclear why regulation of H3K4 methylation is so crucial for neurodevelopment, as most of H3K4 methylation
118 regulators are ubiquitously expressed across tissues, cell types, and lifespan.

119 With the recent advancement of next generation sequencing technologies came the unexpected finding that many
120 neurodevelopmental disorders are caused by or linked to mutations in genes that regulate chromatin^{14–16}.

121 Chromatin regulators, including those that DNA and histone modification, is an integral component of cellular identity

122 Chromatin regulators, including those that DNA and histone modification, is an integral component of cellular identity

123 Cellular identity Some of the first chromatin regulators , such as

124 With the recent advancement of next generation sequencing technologies came the unexpected finding that many
125 neurodevelopmental disorders are caused by or linked to mutations in genes that regulate chromatin^{14–16}.

126 The basic subunit of chromatin is the nucleosome and is comprised of DNA wrapped around octamers of histone
127 proteins¹².

128 Histone proteins can undergo covalent chemical modifications on their N-terminal tails that can in turn influence cofactor
129 recruitment, DNA accessibility, and gene expression^{17,18}.

130 Results

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

132 RNA sequencing (RNA-seq) studies surprisingly revealed the aberrant expression of testis genes within the adult *Kdm5c*
133 knockout (-KO) hippocampus⁷. Given the high abundance of testis-enriched genes within the mouse transcriptome, it is
134 currently unclear if their misexpression in the *Kdm5c*-KO brain is indicative of random de-repression throughout the genome
135 or instead reflects an impairment in tissue identity. To distinguish these possibilities, we globally assessed the expression of
136 previously characterized tissue-enriched genes¹⁹ in our previously published mRNA-seq datasets of the *Kdm5c*-KO amygdala
137 and hippocampus²⁰

138 We found many genes significantly upregulated within the *Kdm5c*-KO brain (DESeq²¹, log2 fold change > 0.5, q < 0.1)
139 are typically enriched in non-brain tissues, with the majority of differentially expressed genes (DEGs) displaying biased
140 expression towards the testis (Figure 1A-C). Even though the testis has the largest number of tissue-biased genes compared
141 to any other tissue (2,496 genes), testis-biased DEGs were significantly enriched for both brain regions (Amygdala p =
142 1.83e-05; Hippocampus p = 4.26e-11, Fisher's Exact).

143 Even though the *Kdm5c*-KO mice are male, we also observed significant enrichment of ovary-biased DEGs within both
144 brain regions (Amygdala p = 0.00574; Hippocampus p = 0.048, Fisher's Exact) (Figure 1D). Intriguingly, many ovary and

145 testis-biased DEGs have functions specific to germ cells and have no known role in the brain. For example, the testis-biased
146 DEG *Cytochrome C, testis* (*Cyct*) is a testis-specific cytochrome c isoform found in spermatozoa mitochondria and is important
147 for flagellar beating^{22,23} (Figure 1C). The ovary-enriched DEG *Zygotic arrest 1* (*Zar1*) was recently shown to sequester
148 mRNAs in oocytes for meiotic maturation and early zygote development²⁴ (Figure 1D). Although not consistent across brain
149 regions, we also found enrichment of two non-gonadal tissues - the liver (Amygdala p = 0.0398, Fisher's Exact) and the
150 muscle (Hippocampus p = 0.0104, Fisher's Exact). One example of a liver-biased DEG is *Apolipoprotein C-I* (*Apoc1*), which
151 is involved in lipoprotein metabolism (Figure 1E). Overall, these tissue-enriched genes show little to no expression in the
152 wild-type brain, yet our mRNA-seq data indicate they are polyadenylated and spliced into mature transcripts (Figure 1C-E).

153 Of note, we did not observe enrichment of brain-enriched genes (Amygdala p = 1; Hippocampus p = 0.74, Fisher's
154 Exact), even though the brain has the second highest total number of tissue-enriched genes (708 genes). Altogether, these
155 results suggest misexpression of testis and other tissue-enriched genes within the *Kdm5c*-KO brain is not due to random
156 de-repression of all genes, but rather due to a dysregulation of tissue identity.

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

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

163 To further validate if these testis DEGs are truly germline genes, we then compared their expression in a published
164 RNA-seq dataset of wild-type (WT) and germ cell-depleted (*Kit*^{W/W^v}) mouse testes²⁵. We found almost all *Kdm5c*-KO
165 testis-enriched DEGs lose their expression with germ cell depletion (Figure 2B). The only testis DEG that did not show
166 considerable downregulation with germline depletion was *FK506 binding protein 6* (*Fkbp6*), a known regulator of piRNA
167 expression and meiosis in germ cells^{26,27}. We additionally assessed testis DEG expression in a published testis single cell
168 RNA-seq dataset that identified cell type-specific markers²⁸. We found that while some testis-enriched DEGs were classified
169 as specific markers for different germ cell types (e.g. spermatogonia, spermatocytes, round spermatids, and elongating
170 spermatids) none marked somatic cells (Figure 2C). Together, these data suggest the *Kdm5c*-KO brain expresses germline
171 genes.

172 We then aimed to globally characterize germline gene misexpression in *Kdm5c*-KO mice, but lacked a comprehensive list
173 of mouse germline-enriched genes. To facilitate downstream analyses, we generated a list of male and female germline-
174 enriched genes by evaluating gene expression in wild-type and germline-depleted (*Kit*^{W/W^v}) mice. We assessed gene
175 expression in male and female mice at embryonic day 12, 14, and 16, as well as male postnatal day 6 and adult testes. We
176 defined genes as germline-enriched if their expression met the following criteria: 1) their expression is greater than 1 FPKM
177 in wild-type germline 2) their expression in any wild-type, non-gonadal tissue¹⁹ does not exceed 20% of maximum expression
178 in wild-type germline, and 3) their expression in the germ cell-depleted germline, for any sex or time point, does not exceed
179 20% of their maximum expression in the wild-type germline. These criteria yielded 1,287 germline-enriched genes.

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

181 Misexpression of germline genes in the *Kdm5c*-KO brain suggests mutants fail to demarcate germline and somatic cellular
182 identity. Germ cells are typically distinguished from somatic cells soon after the embryo implants into the uterine wall^{29,30}
183 when a subset of epiblast stem cells become the primordial germ cells (PGCs) while the remainder differentiate into the
184 ectoderm, mesoderm, and endoderm to form the body's somatic tissues³¹. This developmental time point can be modeled *in*
185 *vitro* through differentiation of embryonic stem cells (ESCs) into post-implantation epiblast-like stem cells (EpiLCs) (Figure
186 3A, top). Previous studies have demonstrated that while some germline-enriched genes are also expressed in embryonic
187 stem cells (ESCs), they are silenced as they differentiate into EpiLCs³². Therefore, we assessed if KDM5C was necessary
188 for initial germline gene silencing by assessing their expression in male *Kdm5c*-KO EpiLCs using our previously generated
189 RNA-seq dataset³³.

190 We did not observe any apparent morphology changes between wild-type and *Kdm5c*-KO cells during ESC to EpiLC
191 differentiation (Figure 3A, bottom) and found no significant difference in primed pluripotency genes that indicate proper EpiLC
192 differentiation (Figure 3B). We then identified *Kdm5c*-KO EpiLC DEGs through DESeq2²¹ (log2 fold change > 0.5, q < 0.1)
193 and again observed widespread dysregulation of tissue-enriched genes, with the largest number of genes belonging to the
194 brain and testis, although they were not significantly enriched (Figure 3C).

195 Using our curated list, we found 54 germline-enriched genes were misexpressed in *Kdm5c*-KO EpiLCs, including *Cyct*
196 (Figure 3D). To evaluate whether germline mRNAs are constitutively dysregulated or change over the course of development,
197 we then compared EpiLC germline DEGs to those expressed in the *Kdm5c*-KO brain. The majority of EpiLC germline DEGs
198 were unique to EpiLCs, with only *Cyct* shared across sequencing datasets (Figure 3E).

199 We then compared the function of EpiLC and brain germline DEGs through gene ontology and found particularly high
200 enrichment of meiosis-related gene ontologies in EpiLCs (Figure 3F), such as meiotic cell cycle (GO:0,051,321, p.adjust =
201 4.44e-07). While a few meiosis-related genes are misexpressed in the *Kdm5c*-KO brain, DEGs unique to *Kdm5c*-KO EpiLCs
202 included master regulators of germline identity, such as *Stimulated by retinoic acid 8 (Stra8)* and *Deleted in azoospermia*
203 *like (Dazl)* (Figure 3G). These genes are typically expressed during embryonic germ cell development to commit PGCs
204 to the germline fate, but are also expressed later in life to trigger meiotic gene expression programs³⁴⁻³⁶. Of note, some
205 germline genes, including *Dazl*, are also expressed in the two-cell embryo. However, we did not see misexpression of two-cell
206 embryo-specific genes, like *Zscan4d* (q = 0.381).

207 *Dazl* is an RNA-binding protein that regulates the translation of germline mRNAs and is essential for germ cell function³⁷.
208 Interestingly, a significant portion of germline transcripts expressed in *Kdm5c*-KO EpiLCs are known binding targets of
209 DAZL, including *Stra8*³⁸ (p = 1.698e-07, Fisher's Exact Test). This suggests expression of DAZL protein would enable
210 the translation of other aberrant germline transcripts, influencing their ability to impact on *Kdm5c*-KO cellular function. We
211 therefore stained *Kdm5c*-KO EpiLCs for DAZL protein expression through immunocytochemistry (Figure 3H). We found
212 about 25% of *Kdm5c*-KO EpiLCs expressed DAZL protein in their cytoplasm (p = 0.0015, Welch's t-test), consistent with the
213 subcellular localization observed when DAZL is stained in spermatogonia³⁸. Altogether these results suggest *Kdm5c*-KO
214 EpiLCs fail to decommission germline genes during early embryogenesis, including master regulators of germline identity
215 that can be translated into protein.

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

218 **Figure 4: Loss of KDM5C's catalytic activity impairs DNAm placement and long-term silencing of germline genes**

219 A) KDM5C binding in EpiLCs vs PNCs - number of peaks and percentage of bound genomic regions B) KDM5C binding in
220 EpiLCs vs pncs germline genes - GO C) Left - bar graph of germline DEGs KDM5C is bound to in EpiLCs Right D) Example
221 bigwigs * Brain and EpiLC shared, EpiLC-specific, Brain-specific, E) motif analysis of KDM5C-bound vs unbound genes

222 --- separate into two figures

223 D) Bigwigs of Increase in H3K4me3 in *Kdm5c*-KO amygdala at germline genes

224 E) Increase in H3K4me2 in EpiLCs at germline genes (left - bigwigs, right average signal at germline gene TSS?)

225 * Previous studies only looked at ESCs, unknown if catalytic activity is required for long-term repression, es

226 F) KDM5C expression ESC --> EpiLC (increasing then decreasing) Left - RNA right - protein

227 G) RNA expression of germline genes with catalytic dead rescue (Ilakkia)

228 H) DNA methylation in WT and 5CKO EpiLCs (Ilakkia)

229

230 Direct vs indirect binding motif analysis

231 Previous work suggests KDM5C represses germline genes during early development, since KDM5C binding is gradually
232 lost at hippocampal germline DEG promoters over development⁷ and re-expression of KDM5C in neuronal cultures fails to
233 suppress their transcription. However, we identified many germline genes that were only dysregulated in *Kdm5c*-KO EpiLCs
234 and not in the mature brain, including master regulators of meiosis and germline identity. While KDM5C was recently shown
235 to bind to *Dazl* in ESCs^{39,40}, it's unclear if KDM5C binds to other master regulators that are expressed in *Kdm5c*-KO EpiLCs
236 and if this binding is maintained as cells differentiate.

237 To better elucidate the mechanism behind KDM5C-mediated germline gene silencing, we analyzed KDM5C chromatin
238 immunoprecipitation followed by DNA sequencing (ChIP-seq) datasets we previously generated in EpiLCs³³ and primary
239 hippocampal and cortical neuron cultures (PNCs)¹¹. We identified a higher number of KDM5C peaks in EpiLCs overall
240 (EpiLCs: 2,437, PNCs: 1,610, MACS2 q < 0.1 and fold enrichment > 1) and while KDM5C was primarily localized to gene
241 promoters in both cell types (EpiLCs: , PNCs: , +/- kb TSS), PNCs showed increased localization to distal intergenic regions
242 (EpiLCs: , PNCs:) (Figure 4A). Gene ontology analysis of KDM5C-bound promoters in EpiLCs revealed high enrichment
243 of germline-related ontologies, including meiotic nuclear division (GO:0140013) and synaptonemal complex assembly
244 (GO:0007130) (Figure 4B). Contrastingly, germline ontologies are not enriched in PNCs and KDM5C-bound genes instead
245 perform non-germline-specific functions, such as heterocycle synthesis (GO: 0018130) and pyrimidine metabolism (GO:
246 0072527).

247 We then assessed KDM5C binding at germline genes expressed *Kdm5c*-KO cells to assess if their dysregulation is due to
248 direct loss of KDM5C binding during embryogenesis.

249 * KDM5C was bound to the promoter of many genes dysregulated in both EpiLCs and brain datasets, including _D1P

250 * KDM5C was not bound to many of the brain-specific germline DEGs, such as XXX.

251 * While KDM5C binds to many germline DEGs in EpiLCs, including _Dazl_, many were also unbound. This included t

252 * All genes lacked binding in PNCs.

253 * Altogether, these data indicate that while some germline genes are misexpressed due to direct loss of KDM5C

254

255 Although KDM5C is generally thought to suppress transcription through erasure of histone 3 lysine 4 di- and trimethylation
 256 (H3K4me2/3), recent studies in ESCs have suggested KDM5C's repression *Dazl* is independent of its catalytic activity³⁹. It
 257 is currently unclear if KDM5C's catalytic activity aids in the long-term repression of germline genes, such as enabling the
 258 placement of DNA methylation.

259 • In congruence with previous work in the *Kdm5c*-KO hippocampus⁷, we observed ectopic H3K4me3 peaks in the
 260 *Kdm5c*-KO amygdala at the transcription start site (TSS) of germline genes (Figure 4).
 261 • We additionally observed a marked increase in H3K4me2 at germline gene promoters in *Kdm5c*-KO EpiLCs.
 262 – This increase was greatest at germline genes directly bound by KDM5C (?)

263 Discussion

264 • tissue-biased gene expression:
 265 – However unlike the gonadal-biased DEGs, many liver and muscle-biased DEGs have a known involvement in brain
 266 function. For example, the liver-biased DEG is *Apolipoprotein C-I* (*Apoc1*), is important for lipoprotein metabolism
 267 but has also been shown to influence learning and memory (Figure 1E).
 268 • Otx2 is properly expressed in EpiLCs and prevents pgc identity <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6485399/>
 269 • Papers to read/reference:
 270 – 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)
 271 – 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/>

274 References

- 275 1. Lewis, E.B. (1978). A gene complex controlling segmentation in Drosophila. *Nature* **276**, 565–570. <https://doi.org/10.1038/276565a0>.
- 276 2. 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>.
- 277 3. 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>.
- 278 4. 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>.

- 283 5. 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>.
- 284
- 285 6. Schaefer, A., Sampath, S.C., Intrator, A., Min, A., Gertler, T.S., Surmeier, D.J., Tarakhovsky, A., and Greengard, P. (2009). Control of cognition and adaptive behavior by the GLP/G9a epigenetic suppressor complex. *Neuron* 64, 678–691. <https://doi.org/10.1016/j.neuron.2009.11.019>.
- 286
- 287 7. Scandaglia, M., Lopez-Atalaya, J.P., Medrano-Fernandez, A., Lopez-Cascales, M.T., Del Blanco, B., Lipinski, M., Benito, E., Olivares, R., Iwase, S., Shi, Y., et al. (2017). Loss of Kdm5c Causes Spurious Transcription and Prevents the Fine-Tuning of Activity-Regulated Enhancers in Neurons. *Cell Rep* 21, 47–59. <https://doi.org/10.1016/j.celrep.2017.09.014>.
- 288
- 289 8. Claes, S., Devriendt, K., Van Goethem, G., Roelen, L., Meireleire, J., Raeymaekers, P., Cassiman, J.J., and Fryns, J.P. (2000). Novel syndromic form of X-linked complicated spastic paraplegia. *Am J Med Genet* 94, 1–4.
- 290
- 291 9. 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>.
- 292
- 293 10. 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>.
- 294
- 295 11. 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>.
- 296
- 297 12. Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* 389, 251–260. <https://doi.org/10.1038/38444>.
- 298
- 299 13. Vallianatos, C.N., and Iwase, S. (2015). Disrupted intricacy of histone H3K4 methylation in neurodevelopmental disorders. *Epigenomics* 7, 503–519. <https://doi.org/10.2217/epi.15.1>.
- 300
- 301 14. De Rubeis, S., He, X., Goldberg, A.P., Poultnay, C.S., Samocha, K., Cicek, A.E., Kou, Y., Liu, L., Fromer, M., Walker, S., et al. (2014). Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* 515, 209–215. <https://doi.org/10.1038/nature13772>.
- 302
- 303 15. Gabriele, M., Lopez Tobon, A., D'Agostino, G., and Testa, G. (2018). The chromatin basis of neurodevelopmental disorders: Rethinking dysfunction along the molecular and temporal axes. *Prog Neuropsychopharmacol Biol Psychiatry* 84, 306–327. <https://doi.org/10.1016/j.pnpbp.2017.12.013>.
- 304
- 305 16. Iwase, S., Bérubé, N.G., Zhou, Z., Kasri, N.N., Battaglioli, E., Scandaglia, M., and Barco, A. (2017). Epigenetic Etiology of Intellectual Disability. *J. Neurosci.* 37, 10773–10782. <https://doi.org/10.1523/JNEUROSCI.1840-17.2017>.
- 306
- 307 17. Strahl, B.D., and Allis, C.D. (2000). The language of covalent histone modifications. *Nature* 403, 41–45. <https://doi.org/10.1038/47412>.
- 308
- 309 18. Jenuwein, T., and Allis, C.D. (2001). Translating the Histone Code. *Science* 293, 1074–1080. <https://doi.org/10.1126/science.1063127>.
- 310
- 311 19. 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>.
- 312

- 313 20. 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.,
314 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>.
- 315 21. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data
316 with DESeq2. *Genome Biol* 15, 550. <https://doi.org/10.1186/s13059-014-0550-8>.
- 317 22. Goldberg, E., Sberna, D., Wheat, T.E., Urbanski, G.J., and Margoliash, E. (1977). Cytochrome c: Immunofluorescent
318 Localization of the Testis-Specific Form. *Science* 196, 1010–1012. <https://doi.org/10.1126/science.193188>.
- 319 23. Narisawa, S., Hecht, N.B., Goldberg, E., Boatright, K.M., Reed, J.C., and Millán, J.L. (2002). Testis-Specific Cytochrome
320 c -Null Mice Produce Functional Sperm but Undergo Early Testicular Atrophy. *Molecular and Cellular Biology* 22,
5554–5562. <https://doi.org/10.1128/MCB.22.15.5554-5562.2002>.
- 321 24. Cheng, S., Altmeppen, G., So, C., Welp, L.M., Penir, S., Ruhwedel, T., Menelaou, K., Harasimov, K., Stützer, A.,
Blayne, M., et al. (2022). Mammalian oocytes store mRNAs in a mitochondria-associated membraneless compartment.
322 *Science* 378, eabq4835. <https://doi.org/10.1126/science.abq4835>.
- 323 25. Mueller, J.L., Skaletsky, H., Brown, L.G., Zaghlul, S., Rock, S., Graves, T., Auger, K., Warren, W.C., Wilson, R.K., and
324 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>.
- 325 26. 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 Chromosome Pairing in Meiosis.
326 *Science* 300, 1291–1295. <https://doi.org/10.1126/science.1083022>.
- 327 27. Xiol, J., Cora, E., Koglgruber, R., Chuma, S., Subramanian, S., Hosokawa, M., Reuter, M., Yang, Z., Berninger, P.,
Palencia, A., et al. (2012). A Role for Fkbp6 and the Chaperone Machinery in piRNA Amplification and Transposon
328 Silencing. *Molecular Cell* 47, 970–979. <https://doi.org/10.1016/j.molcel.2012.07.019>.
- 329 28. 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
330 Cell* 46, 651–667.e10. <https://doi.org/10.1016/j.devcel.2018.07.025>.
- 331 29. Magnúsdóttir, E., and Surani, M.A. (2014). How to make a primordial germ cell. *Development* 141, 245–252.
332 <https://doi.org/10.1242/dev.098269>.
- 333 30. Günesdogan, U., Magnúsdóttir, E., and Surani, M.A. (2014). Primordial germ cell specification: A context-dependent
334 cellular differentiation event [corrected]. *Philos Trans R Soc Lond B Biol Sci* 369. <https://doi.org/10.1098/rstb.2013.0543>.
- 335 31. Bardot, E.S., and Hadjantonakis, A.-K. (2020). Mouse gastrulation: Coordination of tissue patterning, specification and
336 diversification of cell fate. *Mechanisms of Development* 163, 103617. <https://doi.org/10.1016/j.mod.2020.103617>.
- 337 32. 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
338 methylation-mediated silencing. *Nat Commun* 12, 7020. <https://doi.org/10.1038/s41467-021-27345-x>.
- 339 33. 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
340 demethylase. *Nat Commun* 13, 2602. <https://doi.org/10.1038/s41467-022-30352-1>.

- 341 34. Koubova, J., Menke, D.B., Zhou, Q., Capel, B., Griswold, M.D., and Page, D.C. (2006). Retinoic acid regulates
342 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>.
- 343 35. Lin, Y., Gill, M.E., Koubova, J., and Page, D.C. (2008). Germ Cell-Intrinsic and -Extrinsic Factors Govern Meiotic
344 Initiation in Mouse Embryos. *Science* **322**, 1685–1687. <https://doi.org/10.1126/science.1166340>.
- 345 36. Endo, T., Mikedis, M.M., Nicholls, P.K., Page, D.C., and De Rooij, D.G. (2019). Retinoic Acid and Germ Cell
346 Development in the Ovary and Testis. *Biomolecules* **9**, 775. <https://doi.org/10.3390/biom9120775>.
- 347 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
348 regulator of murine spermatogenesis. *Natl Sci Rev* **6**, 455–468. <https://doi.org/10.1093/nsr/nwy163>.
- 349 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).
350 DAZL mediates a broad translational program regulating expansion and differentiation of spermatogonial progenitors.
eLife **9**, e56523. <https://doi.org/10.7554/eLife.56523>.
- 351 39. Gupta, N., Yakhou, L., Albert, J.R., Azogui, A., Ferry, L., Kirsh, O., Miura, F., Battault, S., Yamaguchi, K., Laisné,
352 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>.
- 353 40. Al Adhami, H., Vallet, J., Schaal, C., Schumacher, P., Bardet, A.F., Dumas, M., Chicher, J., Hammann, P., Daujat,
354 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>.

355 **Figures and Tables**

356 —

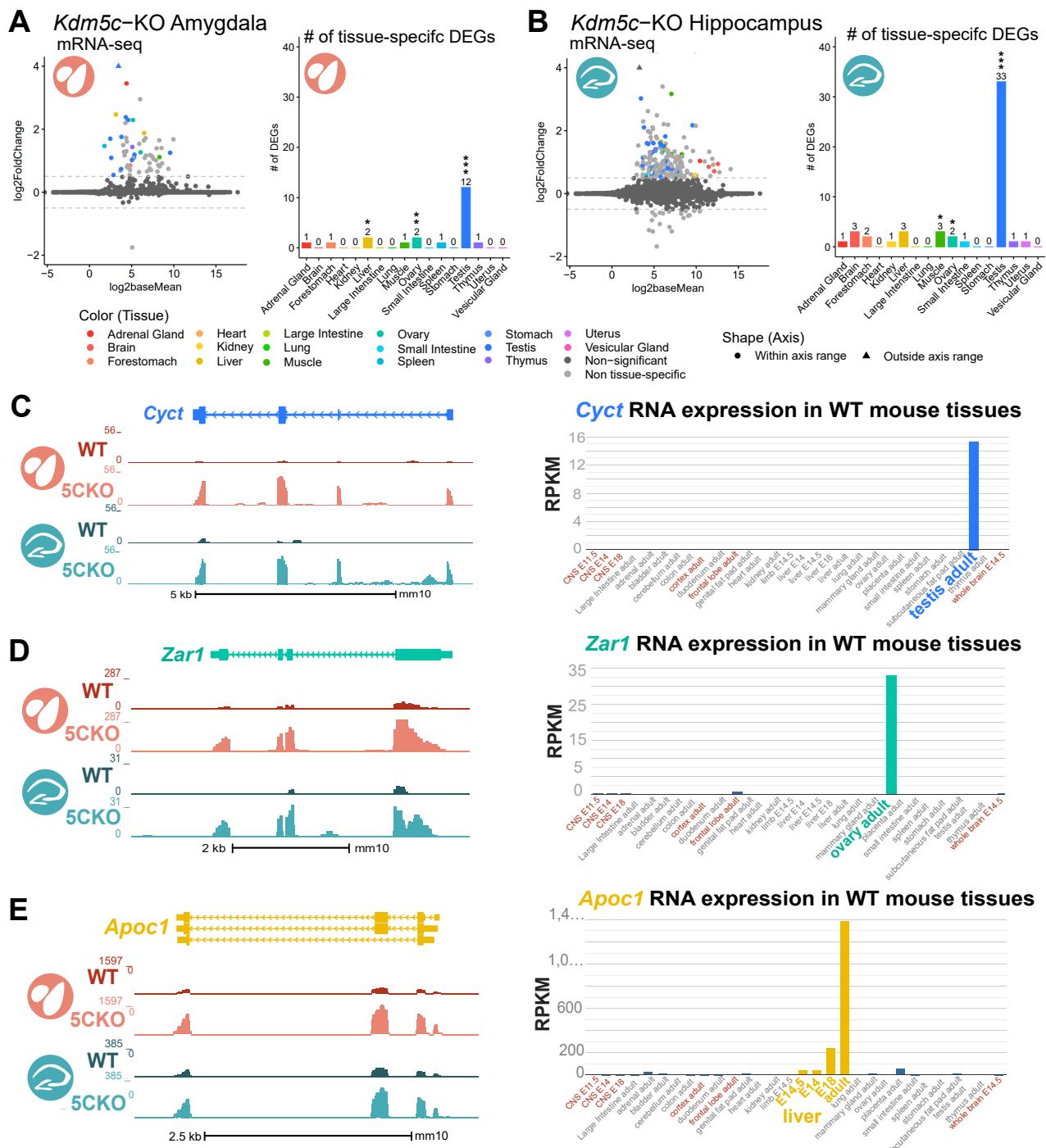


Figure 1: Tissue-enriched genes are misexpressed in the *Kdm5c*-KO brain. **A.** Expression of tissue-enriched genes in the male *Kdm5c*-KO amygdala. Left - MA plot of mRNA-sequencing. Right - Number of tissue-enriched differentially expressed genes (DEGs). * $p<0.05$, ** $p<0.01$, *** $p<0.001$, Fisher's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

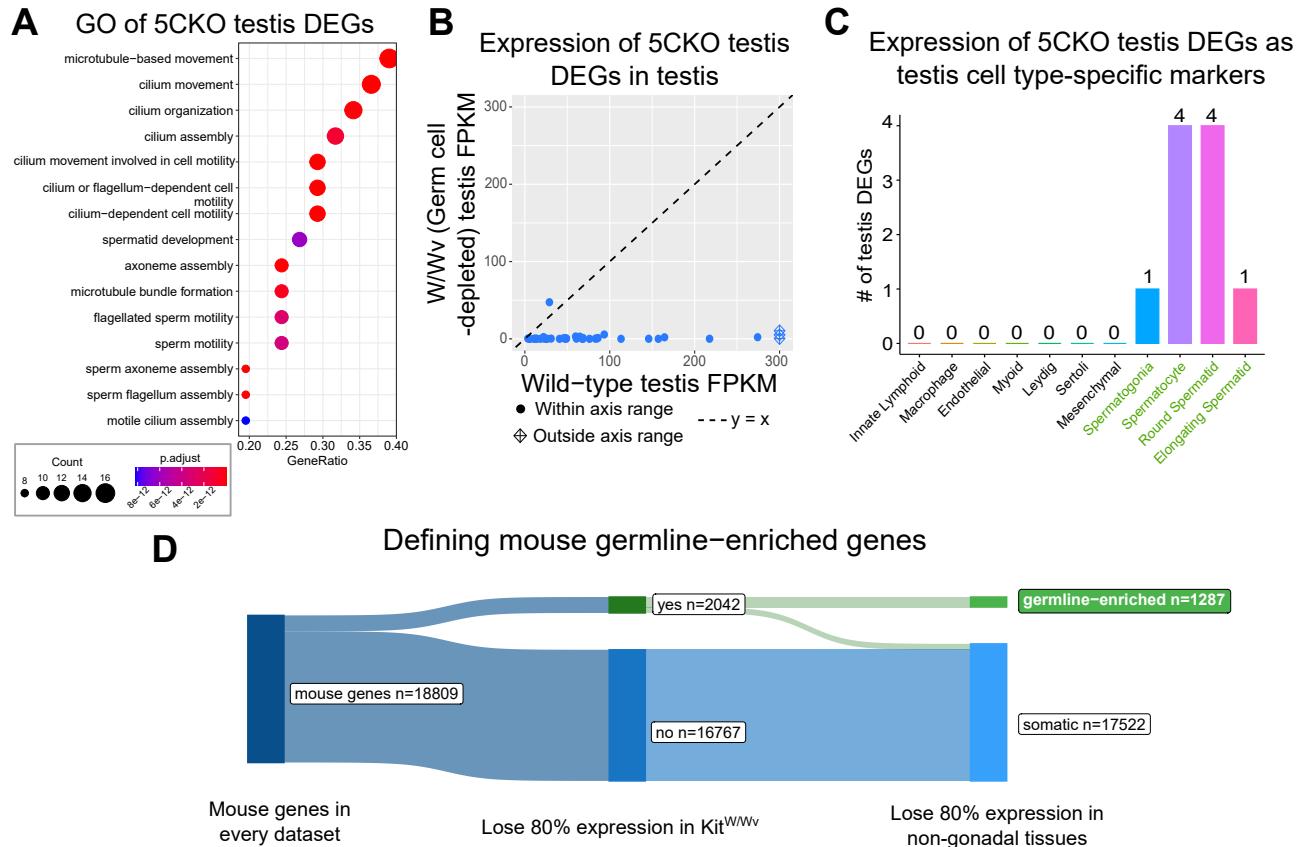


Figure 2: Aberrant transcription of germline genes in the *Kdm5c*-KO in the brain. **A.** enrichPlot gene ontology (GO) of *Kdm5c*-KO amygdala and hippocampus testis-enriched DEGs **B.** Expression of testis DEGs in wild-type (WT) testis versus germ cell-depleted (W/Wv) testis. Expression is in Fragments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

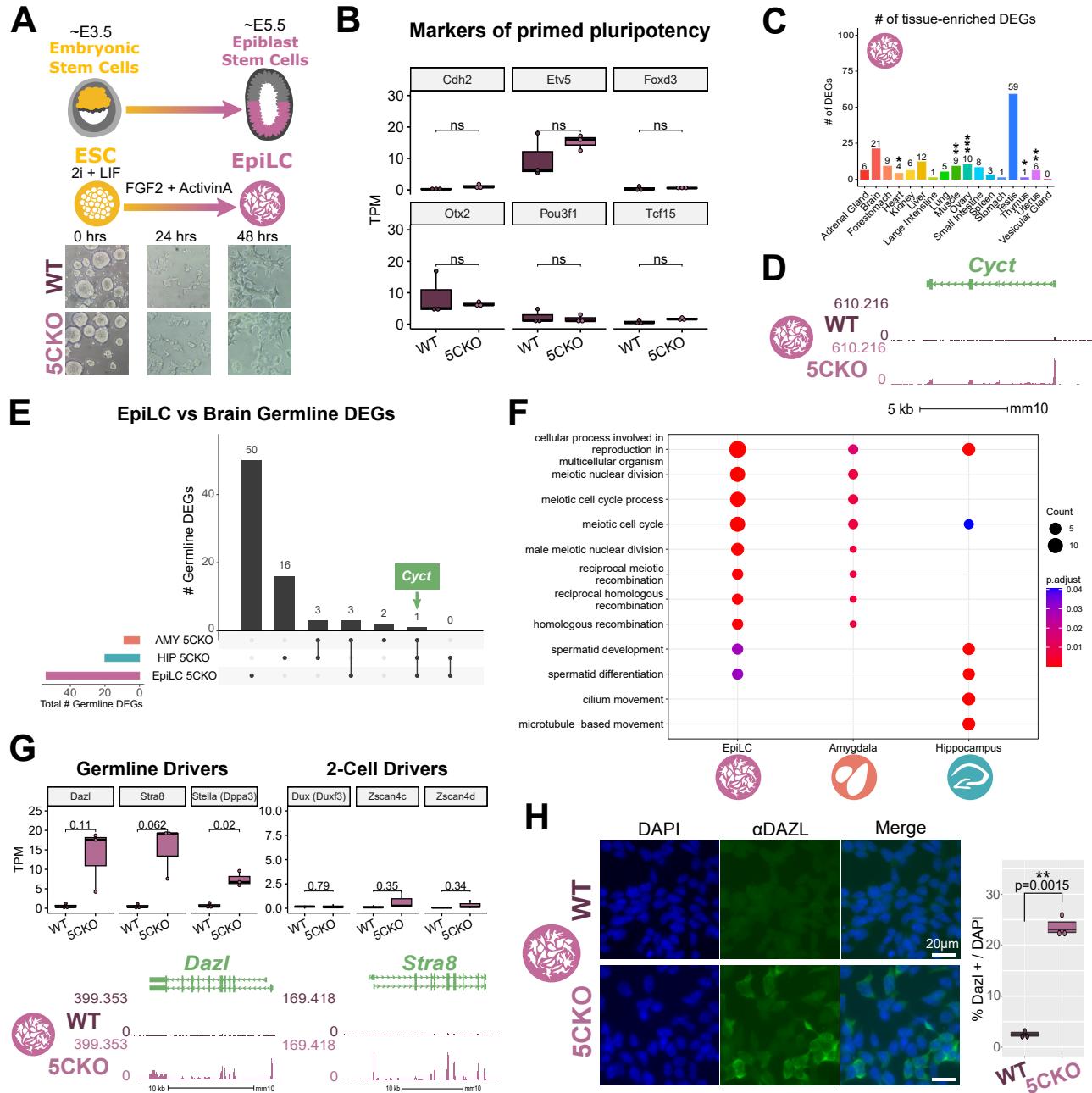


Figure 3: *Kdm5c*-KO epiblast-like cells express master regulators of germline identity

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

357 **Figure outline:**

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

361 **Figure 2: The male *Kdm5c*-KO brain expresses male and female germline-enriched genes** * Gene ontology of
362 testis DEGs in the amygdala and hippocampus - ontologies are germline ontologies * Expression of testis DEGs in germline-
363 depleted testis (this is adult testis data) * scRNASeq of testis - # of testis DEGs that are germline-specific markers * Although
364 far fewer, 5CKO brain also expresses ovary-enriched genes(NCBI and bigwigs of Zar1) * These ovary enriched genes are
365 also germline specific (NCBI/Li tissues are in adult ovary, so it would be best to show they're oocyte-specific in adult ovary.
366 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
367 W/Wv data since oocytes are developed at this point? Or both?) * Defining what is/isn't a germline gene, and which are
368 male/female biased using embryonic W/Wv data

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

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

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

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