

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

3

4 Katherine M. Bonefas, Ilakkia Venkatachalam, and Shigeki Iwase. # Abstract

It is currently unclear why mutations in numerous chromatin-modifying enzymes cause neurodevelopmental disorders (NDDs). While some chromatin regulators are important for neuron-specific gene expression, loss of repressive chromatin regulators can also lead to the aberrant transcription of tissue-specific genes outside of their intended context. Because very few studies have characterized these aberrant, tissue-specific transcripts, the mechanisms behind their dysregulation and their functional consequences are largely unknown. Here, we explored this crisis in cellular identity in NDDs by investigating the ectopic expression of germline (e.g. sperm and egg) genes in mice lacking lysine demethylase 5c (KDM5C, also known as JARID1C or SMCX), an eraser of histone 3 lysine 4 di and tri-methylation (H3K4me2/3). We found male *Kdm5c* knockout (-KO) mice, which recapitulate key behavioral phenotypes of Claes-Jensen X-linked intellectual disability (MRXSCJ, OMIM: 300534), aberrantly expresses many liver, muscle, ovary, and testis genes within the amygdala and hippocampus. Many genes expressed in the *Kdm5c*-KO brain were typically unique to germ cells, indicating an erosion of the soma-germline boundary. Using germ cell-depleted RNA sequencing datasets, we curated a list of germline-enriched genes for unbiased characterization of KDM5C in germline gene repression. Germline genes are typically decommissioned in somatic lineages in the post-implantation epiblast, yet *Kdm5c*-KO epiblast-like cells (EpILCs) aberrantly expressed key regulators of germline identity and meiosis, including *Dazl* and *Stra8*. We found that while KDM5C promotes the initial placement of DNA methylation at a subset of germline gene promoters in EpILCs, germline genes can also become activated independent of direct KDM5C regulation. Intriguingly, many of the germline transcripts expressed in the *Kdm5c*-KO brain were late-stage spermatogenesis genes that are not directly bound by KDM5C. This suggests germline developmental programs can progress ectopically in the background of typical *Kdm5c*-KO development. Ultimately, this work provides novel insight into the demarcation of somatic and germline lineages in mammals while also characterizing the cellular identity crisis within a chromatin-linked neurodevelopmental disorder.

28 Introduction

29 A single genome holds the instructions to generate the myriad of cell types found within the adult organism.
30 This is, in part, accomplished by chromatin regulators that can either promote or impede lineage-specific
31 gene expression through DNA and histone modifications^{1,2}. While many chromatin regulators were initially
32 identified for their roles in shaping cellular and tissue identity^{3–5}, recent advancements in next generation
33 sequencing unexpectedly revealed many neurodevelopmental disorders (NDDs) are caused by mutations in
34 chromatin regulators⁶. Several studies have suggested this connection between chromatin regulators and
35 neurodevelopment is due to their regulation of brain-specific genes, such as orchestrating transcriptional
36 programs for synaptic maturation⁷ and transitioning between neuronal and glial fates during neural precursor
37 differentiation⁸. However, loss of some chromatin regulators can also result in the ectopic transcription
38 of tissue-specific genes outside of their target environment, such as the misexpression of heart^{9,10} and
39 liver-specific¹¹ genes within adult neurons. Very few studies have investigated this cellular identity crisis
40 in chromatin-linked NDDs^{11,12} and it is currently unknown if partial loss of brain identity contributes to
41 neurodevelopmental impairments.

42 To elucidate the role of tissue identity in chromatin-linked neurodevelopmental disorders, it is essential
43 to first characterize the types of genes dysregulated and the molecular mechanisms governing their de-
44 repression. In this study, we characterized the aberrant expression of tissue-enriched genes with loss of
45 lysine demethylase 5C (KDM5C). KDM5C, , also known as SMCX or JARID1C, is a chromatin regulator that
46 can repress gene expression through erasure of histone 3 lysine 4 di- and trimethylation (H3K4me2/3)¹³,
47 marks enriched at active gene promoters. Pathogenic mutations in *KDM5C* cause Intellectual Developmental
48 Disorder, X-linked, Syndromic, Claes-Jensen Type (MRXSCJ, OMIM: 300534), whose predominant features
49 include intellectual disability, seizures, aberrant aggression, and autistic behaviors^{14–16}. *Kdm5c* knockout
50 (-KO) mice recapitulate key MRXSCJ patient phenotypes, including hyperaggression, increased seizure
51 propensity, and learning impairments^{12,17}. Unexpectedly, RNA sequencing (RNA-seq) of the *Kdm5c*-KO
52 hippocampus revealed ectopic expression of testis genes within the brain¹². It is currently unknown what types
53 of testis genes are dysregulated, when in *Kdm5c*-KO development this begins, and if other tissue-specific
54 genes are also aberrantly expressed with KDM5C loss.

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

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

70 To illuminate KDM5C's role in tissue identity, here we characterized the aberrant transcription of tissue-enriched genes within the *Kdm5c*-KO brain and epiblast-like stem cells (EpiLCs), an *in vitro* model of the post-implantation embryo. We observed general dysregulation of tissue-enriched genes in both the *Kdm5c*-KO brain and EpiLCs, including misexpression of genes typically highly enriched within the testis, liver, muscle, and ovary. Both the *Kdm5c*-KO amygdala and hippocampus had significant enrichment of testis genes. We demonstrated testis and ovary-enriched genes are germline genes by generating a list of germline enriched-genes using germ cell-depleted RNA sequencing datasets. Using this curated list, we found *Kdm5c*-KO EpiLCs aberrantly expressed key drivers of germline identity and meiosis, including *Dazl* and *Stra8*, while the *Kdm5c*-KO brain primarily expressed germline genes important for late spermatogenesis. Additionally, KDM5C was highly enriched at germline gene promoters in EpiLCs but only bound to a subset of germline genes expressed in *Kdm5c*-KO cells, indicating germline-enriched mRNAs can be aberrantly transcribed through indirect mechanisms. Finally, we found KDM5C promotes the long-term silencing of germline genes in somatic cells by aiding the placement of DNA CpG methylation in EpiLCs. Thus, we propose KDM5C plays a fundamental role in the development of tissue identity during early embryogenesis, including the establishment of the soma-germline boundary.

85 Results

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

88 Previous RNA sequencing (RNA-seq) in the adult hippocampus revealed ectopic expression of some testis genes within the *Kdm5c* knockout (-KO) brain¹². Since tissue-enriched genes have not been systematically characterized in the *Kdm5c*-KO brain, if the testis is the only tissue type misexpressed and if loss of tissue identity a major consequence of KDM5C loss. Therefore, we globally assessed the expression of genes enriched in 17 mouse tissues²² in our published mRNA-seq datasets of the adult amygdala and hippocampus in mice with constitutive knockout of *Kdm5c*²³.

94 We found a large proportion of genes that are significantly upregulated within the *Kdm5c*-KO brain
95 (DESeq2²⁴, log2 fold change > 0.5, q < 0.1) are enriched in non-brain tissues in wild-type mice (Amygdala:

96 35%, Hippocampus: 24%) (Figure 1A-B). The majority of tissue-enriched differentially expressed genes
97 (tissue-enriched DEGs) were testis genes (Figure 1A-C). Even though the testis has the largest total number
98 of tissue-biased genes compared to any other tissue (2,496 genes), testis-biased DEGs were significantly
99 enriched for both brain regions (Amygdala $p = 1.83e-05$, Odds Ratio = 5.13; Hippocampus $p = 4.26e-11$,
100 Odds Ratio = 4.45, Fisher's Exact Test). One example of a testis-enriched DEG is *FK506 binding protein 6*
101 (*Fkbp6*), a known regulator of piRNA expression and meiosis in germ cells^{25,26} (Figure 1C).

102 In addition to the high enrichment of testis genes, we also identified aberrant expression of other tissue-
103 enriched genes within the *Kdm5c*-KO brain. Although the *Kdm5c*-KO mice we examined were male, we
104 observed significant enrichment of ovary-biased DEGs in both the amygdala and hippocampus (Amygdala
105 $p = 0.00574$, Odds Ratio = 18.7; Hippocampus $p = 0.048$, Odds Ratio = 5.88, Fisher's Exact) (Figure 1D).
106 Ovary-enriched DEGs included *Zygotic arrest 1* (*Zar1*), sequesters mRNAs in oocytes for meiotic maturation
107 and early zygote development²⁷ (Figure 1D). Although not consistent across brain regions, we also found
108 significant enrichment of DEGS biased towards two non-gonadal tissues - the liver (Amygdala $p = 0.0398$,
109 Odds Ratio = 6.58, Fisher's Exact Test) and the muscle (Hippocampus $p = 0.0104$, Odds Ratio = 6.95,
110 Fisher's Exact Test). An example of a liver-biased DEG dysregulated in both the hippocampus and amygdala
111 is *Apolipoprotein C-I* (*Apoc1*), a lipoprotein metabolism and transport gene²⁸ (Figure 1E). Testis, ovary,
112 and liver-enriched DEGs showed little to no expression in the developing and adult wild-type brain, yet our
113 mRNA-seq data indicate they are polyadenylated and spliced into mature transcripts (Figure 1C-E). Of
114 note, we did not observe enrichment of brain-enriched genes (Amygdala $p = 1$; Hippocampus $p = 0.74$,
115 Fisher's Exact), despite the fact these are brain samples and the brain has the second highest total number
116 of tissue-enriched genes (708 genes). Together, these results suggest misexpression of tissue-enriched
117 genes within the brain is a major effect of KDM5C loss.

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

119 • Intriguingly, some of the ectopic testis transcripts identified within the *Kdm5c*-KO hippocampus have
120 known functions unique to germ cells¹², suggesting KDM5C may play a role in demarcating somatic
121 versus germline identity.

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

128 To further validate if these testis DEGs are truly germline genes, we then evaluated their expression
129 in somatic versus germ cells within the testis. We first compared their expression in the testis with germ

130 cell depletion²⁹, which was accomplished by heterozygous *W* and *Wv* mutations in the enzymatic domain
131 of *c-Kit* (*Kit*^{W/Wv}) that prevent the maturation of germ cells³⁰. Almost all *Kdm5c*-KO testis-enriched DEGs
132 lost expression with germ cell depletion (Figure 2B). We then assessed testis-enriched DEG expression
133 in a published single cell RNA-seq dataset that identified cell type-specific markers within the testis³¹. We
134 found that while some testis-enriched DEGs were classified as specific markers for different germ cell types
135 (e.g. spermatogonia, spermatocytes, round spermatids, and elongating spermatids) none marked somatic
136 cells (Figure 2C). Together, these data demonstrate that the *Kdm5c*-KO brain aberrantly expresses germline
137 genes, reflecting an erosion between somatic versus germline identity.

138 We then wanted to more deeply characterize germline gene misexpression with *Kdm5c* loss, but lacked
139 a comprehensive list of mouse germline-enriched genes. To facilitate downstream analyses, we generated
140 a list of germline-enriched genes using currently available RNA-seq datasets in *Kit*^{W/Wv} mice. Wild-type
141 and *Kit*^{W/Wv} datasets included males and females at embryonic day 12, 14, and 16³², as well as adult
142 male testes²⁹. We defined genes as germline-enriched if their expression met the following criteria: 1)
143 their expression is greater than 1 FPKM in wild-type gonads 2) their expression in any adult wild-type,
144 non-gonadal tissue²² does not exceed 20% of their maximum expression in the wild-type germline, and
145 3) their expression in the germ cell-depleted germline, for any sex or time point, does not exceed 20% of
146 their maximum expression in the wild-type germline. These criteria yielded 1,288 germline-enriched genes
147 (Figure 2D), which was hereafter used as a resource to globally assess germline gene misexpression with
148 *Kdm5c* loss (Supplementary table 1).

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

151 Germ cells are typically distinguished from somatic cells soon after the embryo implants into the uter-
152 ine wall^{33,34} when a subset of epiblast stem cells become the primordial germ cells (PGCs) while the
153 remainder differentiate into the ectoderm, mesoderm, and endoderm to form the somatic tissues³⁵. This
154 developmental time point can be modeled *in vitro* through differentiation of embryonic stem cells (ESCs) into
155 post-implantation epiblast-like stem cells (EpiLCs) (Figure 3A, top). While some germline-enriched genes
156 are also expressed in embryonic stem cells (ESCs) and in the 2-cell stage^{36–38}, they are silenced as they
157 differentiate into EpiLCs¹⁹. Therefore, we tested if KDM5C was necessary for silencing germline genes in
158 the post-implantation embryo by evaluating the impact of *Kdm5c* loss in EpiLCs.

159 We first identified *Kdm5c*-KO EpiLC DEGs in our previously published RNA-seq dataset³⁹ (DESeq2, log2
160 fold change > 0.5, q < 0.1) and assessed tissue-enriched gene expression. Similar to the *Kdm5c*-KO brain,
161 we observed general dysregulation of tissue-enriched genes, with the largest number of genes belonging
162 to the brain and testis, although they were not significantly enriched (Figure 3B). Although we observed
163 aberrant expression of many tissue-enriched genes, including 54 germline-enriched genes, we did not find

any significant difference in primed pluripotency genes, such as *Dnmt3b*, *Fgf5*, *Pou3f1*, and *Otx2* (Figure 3C). We also did not observe any gross changes in *Kdm5c*-KO cell morphology during ESC to EpiLC differentiation (Figure 3D), altogether indicating KDM5C loss does not impair EpiLC formation.

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

Interestingly, DEGs unique to *Kdm5c*-KO EpiLCs included key drivers of germline identity, such as *Stimulated by retinoic acid 8 (Stra8)* and *Deleted in azoospermia like (Dazl)* (Figure 3H). These genes are typically expressed during embryonic germ cell development to commit PGCs to the germline fate, but are also expressed later in life to trigger meiotic gene expression programs^{40–42}. Of note, some germline genes, including *Dazl*, are also expressed in the two-cell embryo^{37,43}. However, we did not see misexpression of two-cell embryo-specific genes, like *Duxf3 (Dux)* ($q = 0.337$) and *Zscan4d* ($q = 0.381$), indicating *Kdm5c*-KO in EpiLCs do not revert back to a 2-cell state (Figure 3H).

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

Previous work suggests KDM5C represses germline genes during early development, as re-expression of KDM5C in knockout neuronal cultures fails to suppress two hippocampal germline DEGs¹². There is some evidence KDM5C binds to select germline gene promoters in embryonic stem cells (ESCs) but is lost in neurons¹². However, the lack of a comprehensive list of germline-enriched genes prohibited systematic characterization of KDM5C binding at germline gene promoters. Thus, it is currently unclear if KDM5C is enriched at germline gene promoters, what types of germline genes KDM5C regulates, and if its binding is maintained at any germline genes in neurons.

To further characterize KDM5C's role in germline gene silencing, we analyzed KDM5C chromatin immunoprecipitation followed by DNA sequencing (ChIP-seq) datasets in EpiLCs³⁹ and primary neuron cultures (PNCs) from the cortex and hippocampus¹⁷. EpiLCs had a higher total number of KDM5C peaks than PNCs (EpiLCs: 5,808, PNCs: 1,276, MACS2 $q < 0.1$ and fold enrichment > 1 , removal of *Kdm5c*-KO false positive peaks) and KDM5C was primarily localized to gene promoters in both cell types (EpiLCs: 4,190, PNCs: 745 +/- 500bp from the TSS), although PNCs showed increased localization to non-promoter regions (Figure 4A).

The majority of promoters bound by KDM5C in PNCs were also bound in EpiLCs (513 shared promoters),

however a large portion of gene promoters were only bound by KDM5C in EpiLCs (3677 EpiLC only promoters) (Figure 4B). We then performed gene ontology analysis to compare the known functions of genes with KDM5C bound to their promoter in EpiLCs and PNCs. While there were no ontologies significantly enriched for genes only bound by KDM5C in PNCs, gene ontologies for genes shared between PNCs and EpiLCs were enriched for functions involving nucleic acid turnover, such as deoxyribonucleotide metabolic process (GO:0009262, p.adjust = 8.28e-05) (Figure 4C). Germline-specific ontologies were only enriched in promoters unique to EpiLCs, such as meiotic nuclear division (GO: 0007127 p.adjust = 6.77e-16) and meiotic cell cycle process (GO:1903046, p.adjust = 5.05e-16) (Figure 3C). We then evaluated KDM5C binding around the transcription start site (TSS) of all germline-enriched genes. While KDM5C was bound to a subset of germline gene promoters in EpiLCs, it was not bound to any in PNCs (Figure 4D-E). Together, these results demonstrate KDM5C is significantly enriched at a subset of germline gene promoters in EpiLCs, including meiotic genes, but does not regulate germline genes in neurons.

We then compared KDM5C binding at RNA-seq DEG promoters to determine if the germline mRNAs expressed in the *Kdm5c*-KO brain and EpiLCs are direct targets of KDM5C (Figure 4F). About one third of EpiLC-specific and brain-specific (hippocampus, amygdala, or both) germline DEGs were bound by KDM5C in EpiLCs (EpiLC only: 36%, Brain only: 33.3%). Again, we did not observe any KDM5C binding at germline gene promoters in PNCs, even for brain-specific DEGs (Figure 4H). Some notable differences in KDM5C binding for EpiLC-specific DEGs included *Dazl* and *Stra8*, two key drivers of germline identity discussed above. Although both mRNAs are expressed in *Kdm5c*-KO EpiLCs, KDM5C is only bound to *Dazl*'s 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 brain and EpiLCs are bound by KDM5C around the TSS (Figure 4G). Altogether, this suggests KDM5C decommissions germline genes in EpiLCs but is no longer recruited to germline gene promoters in neurons. Furthermore, majority of germline mRNAs expressed in *Kdm5c*-KO cells are dysregulated independent of direct KDM5C recruitment to their promoters in EpiLCs.

223 KDM5C promotes *de novo* DNA methylation at germline genes

In the early embryo, germline gene promoters are initially decorated with repressive histone modifications before long-term silencing by *de novo* DNA CpG methylation (CpGme)^{18,19,44}. Our results above suggest KDM5C also acts during this time period, however how KDM5C interacts with other germline gene silencing mechanism is currently unclear. KDM5C is generally thought to suppress transcription through erasure of histone 3 lysine 4 di- and trimethylation (H3K4me2/3), marks enriched at active gene promoters¹³. However, KDM5C's catalytic activity was recently shown to not be required for suppressing *Dazl* in ESCs⁴³. Since H3K4me3 impedes *de novo* CpGme placement^{45,46}, KDM5C's catalytic activity may instead be required for long-term germline gene silencing. In support of this, CpGme is markedly reduced at two germline gene promoters in the adult hippocampus¹². However, because KDM5C's role in germline gene repression has only been characterized at select germline genes in ESCs and in the mature brain, it is currently unclear to

234 what extent KDM5C is involved during initial CpGme placement.

235 To elucidate KDM5C's role in germline gene silencing, we first characterized KDM5C's substrates, histone
236 3 lysine 4 di- and trimethylation (H3K4me2/3) in our previously published ChIP-seq datasets in wild type and
237 *Kdm5c*-KO amygdala²³ and EpiLCs³⁹. In congruence with previous work in the *Kdm5c*-KO hippocampus¹²,
238 we observed aberrant accumulation of H3K4me3 around the transcription start site (TSS) of germline genes
239 in the *Kdm5c*-KO amygdala (Figure 5A). We additionally found a marked increase in H3K4me2 around the
240 TSS of germline genes in *Kdm5c*-KO EpiLCs (Figure 5B).

241 We then evaluated KDM5C's role in germline gene silencing during ESC to EpiLC differentiation. We
242 first characterized KDM5C's mRNA and protein expression at 0, 24, and 48 hours of differentiation (Figure
243 5C). While *Kdm5c* mRNA steadily decreased from 0 to 48 hours of differentiation (Figure 5D), KDM5C
244 protein initially increased from 0 to 24 hours but then decreased to near knockout levels by 48 hours (Figure
245 5E). To determine KDM5C's role in the initial placement of DNA methylation at germline gene promoters,
246 we performed whole genome bisulfite sequencing (WGBS) in wild-type and *Kdm5c*-KO ESCs and 96 hour
247 extended EpiLCs (exEpiLCs). Although wild-type cells accumulated high levels of DNA methylation at
248 germline gene promoters over the course of ESCs to exEpiLC differentiation, DNA methylation was markedly
249 reduced in *Kdm5c*-KO exEpiLCs (Figure 5F).

250 • Altogether, this suggests KDM5C is dynamically regulated during ESC to EpiLC differentiation and
251 promotes germline gene silencing via removal of H3K4me2/3 during early embryogenesis.

252 Discussion

253 In the above study, we demonstrate KDM5C's pivotal role in the development of tissue identity, resulting in
254 substantial misexpression of tissue-enriched genes within the adult *Kdm5c*-KO hippocampus and amygdala.
255 In addition to testis genes identified previously¹², we found significant enrichment of muscle, liver, and
256 even ovary-biased genes aberrantly expressed within the male *Kdm5c*-KO brain. Given the majority of
257 tissue-enriched DEGs are testis and ovary genes with no known brain functions, it is currently unknown if
258 they impair *Kdm5c*-KO brain function and contribute to MRXSCJ-related phenotypes like intellectual disability
259 and aberrant aggression. However, select liver and muscle-biased DEGs do have known roles within the
260 brain, such as the liver-enriched lipid metabolism gene *Apolipoprotein C-I* (*Apoc1*)²⁸ that is highly expressed
261 in the *Kdm5c*-KO amygdala and hippocampus. Intriguingly, *Apoc1* overexpression in the mouse brain can
262 impair learning and memory⁴⁷ and is implicated in Alzheimer's disease in humans⁴⁸. Thus, failure to fine-tune
263 the expression of tissue-enriched, dosage-sensitive genes like *Apoc1* could be one route by which loss of
264 brain tissue identity contributes to *Kdm5c*-KO impairments.

265 We found a substantial portion of genes upregulated in *Kdm5c*-KO brain are germline genes with no
266 known functions within the brain. Distinguishing the germline (i.e. cells that pass on their genetic material
267 to the next generation) from the soma (i.e. all other cells of the body) is a key feature of multicellularity

268 and sexual reproduction. Previous work characterizing chromatin regulators that silence germ cell-specific
269 transcription has predominantly focused on their repression of key marker genes in embryonic stem cells
270 (ESCs), such as *Dazl* and *Ddx4*^{18,19,49}. To globally characterize KDM5C's role in germline gene repression
271 throughout life, we curated a list of mouse germline-enriched genes using publicly available germ cell-
272 depleted RNA-seq datasets from *Kit^{W/Wv}* mice^{29,32}. This resource enabled us to identify 1) the extent of
273 germline gene dysregulation in *Kdm5c*-KO cells, 2) the types of germline genes misexpressed at different
274 developmental time points, and 3) the categories of germline genes directly and indirectly regulated by
275 KDM5C.

276 RNA-seq of epiblast-like cells (EpiLCs) revealed dysregulation of *Kdm5c*-KO tissue identity begins during
277 early embryogenesis. *In vivo*, germline genes are typically decommissioned in epiblast stem cells and
278 remain silenced as the epiblast differentiates into the body's somatic tissues³⁵. However, a small subset of
279 epiblast stem cells will receive signals to reactivate germline gene expression to become the primordial germ
280 cells (PGCs) that will ultimately form the mature germline^{33,34}. This process can be mimicked *in vitro* by
281 differentiating EpiLCs into primordial germ cell-like cells (PGCLCs)⁵⁰. Therefore, misexpression of germline
282 genes in EpiLCs might suggest *Kdm5c*-KO ESCs are progressing beyond EpiLC differentiation and instead
283 becoming PGCLCs. Yet, *Kdm5c*-KO EpiLCs properly express primed pluripotency genes, including *Otx2*
284 - an epiblast stem cell marker that is known to repress epiblast and EpiLC differentiation into PGC and
285 PGCLCs⁵¹. Furthermore, we observed no difference in cellular morphology during *Kdm5c*-KO ESC to EpiLC
286 differentiation. Proper EpiLC differentiation, together with *Kdm5c*-KO mice being viable, suggests germline
287 gene expression is occurring ectopically in conjunction with typical developmental programs, rather than a
288 complete shift towards germline identity.

289 We then globally characterized KDM5C binding at germline-enriched gene promoters through KDM5C
290 ChIP-seq in EpiLCs and primary hippocampus and cortex neuron cultures (PNCs). While we observed
291 high enrichment of KDM5C peaks at germline genes in EpiLCs, KDM5C was not bound to germline gene
292 promoters in PNCs. This suggests dysregulation of germline genes in the mature *Kdm5c*-KO brain is due
293 to loss of repression during embryogenesis, which is consistent with previous work that found introducing
294 human KDM5C into *Kdm5c*-KO PNCs does not repress two aberrant germline transcripts¹². Although
295 enriched at germline gene promoters in EpiLCs, KDM5C was only bound to about a third of the germline
296 genes aberrantly transcribed in *Kdm5c*-KO EpiLCs. One notable EpiLC DEG not bound by KDM5C is
297 *Stra8*, a transcription factor activated by retinoic acid signaling in germ cells to promote meiotic initiation^{52,53}.
298 However, retinoic acid can only activate *Stra8* expression when DAZL is present. Unlike *Stra8*, *Dazl* is a
299 direct target of KDM5C in EpiCs and is both transcribed and translated in *Kdm5c*-KO EpiLCs. Expression of
300 indirect DEGs like *Stra8* indicates germline genes can be aberrantly expressed in *Kdm5c*-KO cells through
301 downstream activation by other ectopic germline programs. These ectopic germline programs appear to
302 loosely mimic the trajectory of typical germline development, as germline genes important for early germ
303 cell development and meiosis are expressed in *Kdm5c*-KO EpiLC while late spermatogenesis genes are

304 expressed in the mature *Kdm5c*-KO brain. Altogether, this indicates once a subset of germline genes are
305 activated due to loss of direct KDM5C repression in the early embryo, these ectopic germline programs can
306 continue to progress in the background of *Kdm5c*-KO somatic development.

307 It is important to note that while we focused on KDM5C's regulation of germline genes, some germline-
308 enriched genes are also expressed in naïve ESCs and at the 2-cell stage due to their roles in pluripotency and
309 self-renewal. For example, *Dazl* is primarily known for committing PGCs to the germline fate and regulating
310 the translation of germline-specific RNAs, but is also expressed at the 2-cell stage⁵⁴, in naïve ESCs³⁶, and in
311 the inner cell mass³⁶. Very recently, two screens of *Dazl*-repressors identified KDM5C as a direct regulator in
312 ESCs^{43,55}. Interestingly, one study found *Kdm5c*-KO ESCs also aberrantly expressed 2-cell-specific genes,
313 indicating KDM5C negatively regulates totipotency⁴³. However, out of the four regulators characterized,
314 KDM5C was the only factor whose repression of *Dazl* was independent of the 2-cell specific transcription
315 factor *Dux*⁴³. Although expressed at the 2-cell stage and in ESCs, *Dazl* is typically silenced when ESCs
316 differentiate into EpiLCs¹⁹. We found KDM5C also represses *Dazl* in EpiLCs, but *Kdm5c*-KO EpiLCs did
317 not express 2-cell specific genes like *Dux* and *Zscan4c*. Together, this suggests the 2-cell-like state in
318 *Kdm5c*-KO ESCs is an indirect consequence of KDM5C's primary role in germline gene repression.

319 *Dazl* and other germline-enriched genes are initially decommissioned in ESCs by repressive histone
320 modifications and then *de novo* DNA CpG methylation (CpGme) in EpiLCs^{19–21,44,56,57}. Unlike the previously
321 characterized germline gene suppressors, KDM5C removes an active histone mark, histone 3 lysine 4 di-
322 and trimethylation (H3K4me2/3)¹³. In this study, we observed a global increase in H3K4me2/3 in *Kdm5c*-KO
323 EpiLCs and amygdala around the transcription start site (TSS) of germline-enriched genes. However,
324 KDM5C's catalytic activity may not be required for germline gene silencing, as it was recently found to be
325 dispensible for repressing *Dazl* in ESCs⁴³. Although not necessary in ESCs, KDM5C's catalytic activity be
326 required for CpGme in EpiLCs, as H3K4me3 impedes *de novo* CpGme placement^{45,46}. This is supported
327 by previous work in the *Kdm5c*-KO adult hippocampus, which found CpGme is significantly eroded at at
328 least two germline promoters¹². To elucidate the mechanism behind KDM5C-mediated silencing of germline
329 genes, we characterized KDM5C in ESCs and EpiLCs and found its expression is dynamically regulated
330 during differentiation. Furthermore, KDM5C acts during the transition from naïve to primed pluripotency to
331 promote the initial placement of CpGme at CpG islands for the long-term silencing of germline genes.

332 The above work provides the mechanistic foundation for KDM5C's role in germline gene repression
333 and offers novel insight into how dysregulation in *Kdm5c*-KO tissue identity changes over the course of
334 development. Although their impact upon neurodevelopment is currently unclear, ectopic germline transcripts
335 are also found in models of other related neurodevelopmental disorders⁵⁸, including Immunodeficiency,
336 Centromere instability, Facial anomalies syndrome 1 (ICF1 OMIM: #242860)^{59,60}, Kleefstra syndrome
337 1 (OMIM: #610253)¹¹, and MeCP2 duplication syndrome (MDS, OMIM: #300260)⁶¹. Like KDM5C, the
338 chromatin regulators underlying these conditions, DNA methyltransferase 3b (DNMT3B), H3K9me1/2
339 methyltransferases G9A/GLP, methyl-CpG -binding protein 2 (MECP2), primarily silence gene expression.

340 Thus, KDM5C is among a growing cohort of chromatin-linked neurodevelopmental disorders that may have a
341 similar underlying cause of germline versus soma dysregulation. However, further research is required to
342 determine the ultimate impact of these germline genes and the extent to which this phenomenon occurs in
343 humans.

344 Materials and Methods

345 Classifying tissue-enriched and germline-enriched genes

346 Tissue-enriched differentially expresssd genes (DEGs) were determined by their classification in a previ-
347 ously published dataset from 17 male and female mouse tissues²². This study defined tissue expression as
348 greater than 1 Fragments Per Kilobase of transcript per Million mapped read (FPKM) and tissue enrichment
349 as at least 4-fold higher expression than any other tissue.

350 We curated a list of germline-enriched genes using an RNA-seq dataset from wild-type and germline-
351 depleted (Kit^{W/W^v}) male and female mouse embryos from embryonic day 12, 14, and 16³², as well as adult
352 male testes²⁹. Germline-enriched genes met the following criteria: 1) their expression is greater than 1
353 FPKM in wild-type germline 2) their expression in any wild-type somatic tissues²² does not exceed 20%
354 of maximum expression in wild-type germline, and 3) their expression in the germ cell-depleted (Kit^{W/W^v})
355 germline, for any sex or time point, does not exceed 20% of maximum expression in wild-type germline.

356 Cell culture

357 We utilized our previously established cultures of male wild-type and *Kdm5c* knockout (-KO) embryonic
358 stem cells³⁹. Sex was confirmed by genotyping *Uba1* on the X and Y chromosomes with the following
359 primers: 5'-TGGATGGTGTGCCAATG-3', 5'-CACCTGCACGTTGCCCTT-3'. Deletion of *Kdm5c* was
360 confirmed through the primers 5'-ATGCCCATATTAAGAGTCCTG-3', 5'-TCTGCCTTGATGGGACTGTT-3',
361 and 5'-GGTTCTCAACACTCACATAGTG-3'.

362 Embryonic stem cells (ESCs) were initially cultured in primed ESC (pESC) media consisting of Knock-
363 Out DMEM (Gibco#10829-018), fetal bovine serum (Gibco#A5209501), KnockOut serum replacement
364 (Invitrogen#10828-028), Glutamax (Gibco#35050-061), Anti-Anti (Gibco#15240-062), MEM Non-essential
365 amino acids (Gibco#11140-050), and beta-mercaptoethanol (Sigma#M7522). They were then transitioned
366 into ground-state "naïve" ESCs (nESCs) by culturing for four passages in N2B27 media containing DMEM/F12
367 (Gibco#11330-032), Neurobasal media (Gibco#21103-049), Gluamax, Anti-Anti, N2 supplement (Invitro-
368 gen#17502048), and B27 supplement without vitamin A (Invitrogen#12587-010), and beta-mercaptoethanol.
369 Both pESC and nESC media were supplemented with 3 μ M GSK3 inhibitor CHIR99021 (Sigma #SML1046-
370 5MG), 1 μ M MEK inhibitor PD0325901 (Sigma #PZ0162-5MG), and 1,000 units/mL leukemia inhibitory factor
371 (LIF, Millipore#ESG1107).

372 nESCs were differentiated into epiblast-like cells (EpiLCs, 48 hours) and extendend EpiLCs (exEpiLCs,
373 96 hours) by culturing in N2B27 media containing DMEM/F12, Neurobasal media, Gluamax, Anti-Anti,
374 N2 supplement, B27 supplement (Invitrogen#17504044), beta-mercaptoethanol, fibroblast growth factor 2
375 (FGF2, R&D Biotechne 233-FB), and activin A (R&D Biotechne 338AC050CF), as previously described⁶².

376 **Immunocytochemistry (ICC)**

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

386 **RNA sequencing (RNA-seq)**

387 After ensuring read quality via FastQC (v0.11.8), reads were then mapped to the mm10 *Mus musculus*
388 genome (Gencode) using STAR (v2.5.3a), during which we removed duplicates and kept only uniquely
389 mapped reads. Count files were generated by FeatureCounts (Subread v1.5.0), and BAM files were
390 converted to bigwigs using deeptools (v3.1.3) and visualized by the UCSC genome browser. RStudio (v3.6.0)
391 was then used to analyze counts files by DESeq2 (v1.26.0)²⁴ to identify differentially expressed genes
392 (DEGs) with a q-value (p-adjusted via FDR/Benjamini–Hochberg correction) less than 0.1 and a log2 fold
393 change greater than 0.5. For all DESeq2 analyses, log2 fold changes were calculated with IfcShrink using
394 the ashR package⁶³. MA-plots were generated by ggpubr (v0.6.0), and Eulerr diagrams were generated by
395 eulerr (v6.1.1). Boxplots and scatterplots were generated by ggpubr (v0.6.0) and ggplot2 (v3.3.2). The Upset
396 plot was generated via the package UpSetR (v1.4.0)⁶⁴. Gene ontology (GO) analyses were performed by
397 the R package enrichPlot (v1.16.2) using the biological processes setting and compareCluster.

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

399 ChIP-seq reads were aligned to mm10 using Bowtie1 (v1.1.2) allowing up to two mismatches. Only
400 uniquely mapped reads were used for analysis. Peaks were called using MACS2 software (v2.2.9.1) using
401 input BAM files for normalization, with filters for a q-value < 0.1 and a fold enrichment > 1. We removed
402 “black-listed” genomic regions that often give aberrant signals. Common peak sets were obtained in R via
403 DiffBind (v3.6.5). In the case of KDM5C ChIP-seq, *Kdm5c*-KO peaks were then subtracted from wild-type

404 samples using bedtools (v2.25.0). Peak proximity to genome annotations was determined by ChIPSeeker
405 (v1.32.1). Gene ontology (GO) analyses were performed by the R package enrichPlot (v1.16.2) using the
406 biological processes setting and compareCluster. Enriched motifs were identified using HOMER⁶⁵. Average
407 binding across the genome was visualized using deeptools (v3.1.3). Bigwigs were visualized using the
408 UCSC genome browser.

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

410 **Data availability**

411 **Published datasets**

412 All published datasets are available at the Gene Expression Omnibus (GEO) <https://www.ncbi.nlm.nih.gov/geo>. Published RNA-seq datasets analyzed in this study included the male wild-type and *Kdm5c*-KO
413 adult amygdala and hippocampus²³ (available at GEO: GSE127722) and male wild-type and *Kdm5c*-KO
414 EpiLCs³⁹ (available at GEO: GSE96797).

416 Previously published ChIP-seq experiments included KDM5C in wild-type and *Kdm5c*-KO EpiLCs³⁹ (avail-
417 able at GEO: GSE96797) and mouse primary neuron cultures (PNCs) from the cortex and hippocampus¹⁷
418 (available at GEO: GSE61036). ChIP-seq of histone 3 lysine 4 dimethylation in male wild-type and *Kdm5c*-KO
419 EpiLCs³⁹ is also available at GEO: GSE96797. ChIP-seq of histone 3 lysine 4 trimethylation in wild-type and
420 *Kdm5c*-KO male amygdala²³ are available at GEO: GSE127817.

421 **Acknowledgements**

- 422 • Jacob Mueller for providing insight in germline gene regulation.
423 • Sundeep Kalantry for providing reagents and expertise in culturing mouse embryonic stem cells and
424 epiblast like cells
425 • Funding sources

426 **References**

- 427 1. Strahl, B.D., and Allis, C.D. (2000). The language of covalent histone modifications. *Nature* 403,
428 41–45. <https://doi.org/10.1038/47412>.
- 429 2. Jenuwein, T., and Allis, C.D. (2001). Translating the Histone Code. *Science* 293, 1074–1080.
430 <https://doi.org/10.1126/science.1063127>.
- 431 3. Lewis, E.B. (1978). A gene complex controlling segmentation in Drosophila. *Nature* 276, 565–570.
432 <https://doi.org/10.1038/276565a0>.

- 433 4. Kennison, J.A., and Tamkun, J.W. (1988). Dosage-dependent modifiers of polycomb and antennapedia
434 mutations in *Drosophila*. Proc. Natl. Acad. Sci. U.S.A. 85, 8136–8140. <https://doi.org/10.1073/pnas.85.21.8136>.
- 435 5. Kassis, J.A., Kennison, J.A., and Tamkun, J.W. (2017). Polycomb and Trithorax Group Genes in
436 *Drosophila*. Genetics 206, 1699–1725. <https://doi.org/10.1534/genetics.115.185116>.
- 437 6. Gabriele, M., Lopez Tobon, A., D'Agostino, G., and Testa, G. (2018). The chromatin basis of
438 neurodevelopmental disorders: Rethinking dysfunction along the molecular and temporal axes. Prog
439 Neuropsychopharmacol Biol Psychiatry 84, 306–327. <https://doi.org/10.1016/j.pnpbp.2017.12.013>.
- 440 7. Zhou, Z., Hong, E.J., Cohen, S., Zhao, W.-N., Ho, H.-Y.H., Schmidt, L., Chen, W.G., Lin, Y., Savner,
441 E., Griffith, E.C., et al. (2006). Brain-specific phosphorylation of MeCP2 regulates activity-dependent
442 Bdnf transcription, dendritic growth, and spine maturation. Neuron 52, 255–269. <https://doi.org/10.1016/j.neuron.2006.09.037>.
- 443 8. Hirabayashi, Y., Suzuki, N., Tsuboi, M., Endo, T.A., Toyoda, T., Shinga, J., Koseki, H., Vidal, M., and
444 Gotoh, Y. (2009). Polycomb Limits the Neurogenic Competence of Neural Precursor Cells to Promote
445 Astrogenic Fate Transition. Neuron 63, 600–613. <https://doi.org/10.1016/j.neuron.2009.08.021>.
- 446 9. Mall, M., Karefa, M.S., Chanda, S., Ahlenius, H., Perotti, N., Zhou, B., Grieder, S.D., Ge, X., Drake,
447 S., Euong Ang, C., et al. (2017). Myt1l safeguards neuronal identity by actively repressing many
448 non-neuronal fates. Nature 544, 245–249. <https://doi.org/10.1038/nature21722>.
- 449 10. Weigel, B., Tegethoff, J.F., Grieder, S.D., Lim, B., Nagarajan, B., Liu, Y.-C., Truberg, J., Papageorgiou,
450 D., Adrian-Segarra, J.M., Schmidt, L.K., et al. (2023). MYT1L haploinsufficiency in human neurons
451 and mice causes autism-associated phenotypes that can be reversed by genetic and pharmacologic
452 intervention. Mol Psychiatry 28, 2122–2135. <https://doi.org/10.1038/s41380-023-01959-7>.
- 453 11. Schaefer, A., Sampath, S.C., Intrator, A., Min, A., Gertler, T.S., Surmeier, D.J., Tarakhovsky, A.,
454 and Greengard, P. (2009). Control of cognition and adaptive behavior by the GLP/G9a epigenetic
455 suppressor complex. Neuron 64, 678–691. <https://doi.org/10.1016/j.neuron.2009.11.019>.
- 456 12. Scandaglia, M., Lopez-Atalaya, J.P., Medrano-Fernandez, A., Lopez-Cascales, M.T., Del Blanco, B.,
457 Lipinski, M., Benito, E., Olivares, R., Iwase, S., Shi, Y., et al. (2017). Loss of Kdm5c Causes Spurious
458 Transcription and Prevents the Fine-Tuning of Activity-Regulated Enhancers in Neurons. Cell Rep 21,
459 47–59. <https://doi.org/10.1016/j.celrep.2017.09.014>.
- 460 13. Iwase, S., Lan, F., Bayliss, P., De La Torre-Ubieta, L., Huarte, M., Qi, H.H., Whetstone, J.R., Bonni, A.,
461 Roberts, T.M., and Shi, Y. (2007). The X-Linked Mental Retardation Gene SMCX/JARID1C Defines a
462 Family of Histone H3 Lysine 4 Demethylases. Cell 128, 1077–1088. <https://doi.org/10.1016/j.cell.2007.02.017>.

- 453 14. 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.
- 454
- 455 15. Jensen, L.R., Amende, M., Gurok, U., Moser, B., Gimmel, V., Tschach, A., Janecke, A.R., Tariverdian,
G., Chelly, J., Fryns, J.P., et al. (2005). Mutations in the JARID1C gene, which is involved in
transcriptional regulation and chromatin remodeling, cause X-linked mental retardation. *Am J Hum
Genet* 76, 227–236. <https://doi.org/10.1086/427563>.
- 456
- 457 16. 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>.
- 458
- 459 17. 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.celr
ep.2015.12.091](https://doi.org/10.1016/j.celr
ep.2015.12.091).
- 460
- 461 18. Endoh, M., Endo, T.A., Shinga, J., Hayashi, K., Farcas, A., Ma, K.W., Ito, S., Sharif, J., Endoh, T.,
Onaga, N., et al. (2017). PCGF6-PRC1 suppresses premature differentiation of mouse embryonic
stem cells by regulating germ cell-related genes. *eLife* 6. <https://doi.org/10.7554/eLife.21064>.
- 462
- 463 19. Mochizuki, K., Sharif, J., Shirane, K., Uranishi, K., Bogutz, A.B., Janssen, S.M., Suzuki, A., Okuda,
A., Koseki, H., and Lorincz, M.C. (2021). Repression of germline genes by PRC1.6 and SETDB1
in the early embryo precedes DNA methylation-mediated silencing. *Nat Commun* 12, 7020. <https://doi.org/10.1038/s41467-021-27345-x>.
- 464
- 465 20. Velasco, G., Hubé, F., Rollin, J., Neuillet, D., Philippe, C., Bouzinba-Segard, H., Galvani, A., Viegas-
Péquignot, E., and Francastel, C. (2010). Dnmt3b recruitment through E2F6 transcriptional repressor
mediates germ-line gene silencing in murine somatic tissues. *Proc Natl Acad Sci U S A* 107, 9281–
9286. <https://doi.org/10.1073/pnas.1000473107>.
- 466
- 467 21. Hackett, J.A., Reddington, J.P., Nestor, C.E., Duncan, D.S., Branco, M.R., Reichmann, J., Reik,
W., Surani, M.A., Adams, I.R., and Meehan, R.R. (2012). Promoter DNA methylation couples
genome-defence mechanisms to epigenetic reprogramming in the mouse germline. *Development*
139, 3623–3632. <https://doi.org/10.1242/dev.081661>.
- 468
- 469 22. Li, B., Qing, T., Zhu, J., Wen, Z., Yu, Y., Fukumura, R., Zheng, Y., Gondo, Y., and Shi, L. (2017). A
Comprehensive Mouse Transcriptomic BodyMap across 17 Tissues by RNA-seq. *Sci Rep* 7, 4200.
<https://doi.org/10.1038/s41598-017-04520-z>.
- 470

- 471 23. Vallianatos, C.N., Raines, B., Porter, R.S., Bonefas, K.M., Wu, M.C., Garay, P.M., Collette, K.M.,
Seo, Y.A., Dou, Y., Keegan, C.E., et al. (2020). Mutually suppressive roles of KMT2A and KDM5C
in behaviour, neuronal structure, and histone H3K4 methylation. *Commun Biol* 3, 278. <https://doi.org/10.1038/s42003-020-1001-6>.
- 472
- 473 24. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for
474 RNA-seq data with DESeq2. *Genome Biol* 15, 550. <https://doi.org/10.1186/s13059-014-0550-8>.
- 475 25. Crackower, M.A., Kolas, N.K., Noguchi, J., Sarao, R., Kikuchi, K., Kaneko, H., Kobayashi, E., Kawai, Y.,
Kozieradzki, I., Landers, R., et al. (2003). Essential Role of Fkbp6 in Male Fertility and Homologous
476 Chromosome Pairing in Meiosis. *Science* 300, 1291–1295. <https://doi.org/10.1126/science.1083022>.
- 477 26. Xiol, J., Cora, E., Koglgruber, R., Chuma, S., Subramanian, S., Hosokawa, M., Reuter, M., Yang, Z.,
Berninger, P., Palencia, A., et al. (2012). A Role for Fkbp6 and the Chaperone Machinery in piRNA
478 Amplification and Transposon Silencing. *Molecular Cell* 47, 970–979. <https://doi.org/10.1016/j.molcel.2012.07.019>.
- 479 27. Cheng, S., Altmeppen, G., So, C., Welp, L.M., Penir, S., Ruhwedel, T., Menelaou, K., Harasimov, K.,
Stützer, A., Blayney, M., et al. (2022). Mammalian oocytes store mRNAs in a mitochondria-associated
480 membraneless compartment. *Science* 378, eabq4835. <https://doi.org/10.1126/science.abq4835>.
- 481 28. Rouland, A., Masson, D., Lagrost, L., Vergès, B., Gautier, T., and Bouillet, B. (2022). Role of
482 apolipoprotein C1 in lipoprotein metabolism, atherosclerosis and diabetes: A systematic review.
Cardiovasc Diabetol 21, 272. <https://doi.org/10.1186/s12933-022-01703-5>.
- 483 29. Mueller, J.L., Skaletsky, H., Brown, L.G., Zaghlul, S., Rock, S., Graves, T., Auger, K., Warren,
W.C., Wilson, R.K., and Page, D.C. (2013). Independent specialization of the human and mouse X
484 chromosomes for the male germ line. *Nat Genet* 45, 1083–1087. <https://doi.org/10.1038/ng.2705>.
- 485 30. Handel, M.A., and Eppig, J.J. (1979). Sertoli Cell Differentiation in the Testes of Mice Genetically
486 Deficient in Germ Cells. *Biology of Reproduction* 20, 1031–1038. <https://doi.org/10.1095/biolreprod.20.5.1031>.
- 487 31. Green, C.D., Ma, Q., Manske, G.L., Shami, A.N., Zheng, X., Marini, S., Moritz, L., Sultan, C.,
Gurczynski, S.J., Moore, B.B., et al. (2018). A Comprehensive Roadmap of Murine Spermatogenesis
Defined by Single-Cell RNA-Seq. *Dev Cell* 46, 651–667.e10. <https://doi.org/10.1016/j.devcel.2018.07.025>.
- 488
- 489 32. Soh, Y.Q., Junker, J.P., Gill, M.E., Mueller, J.L., van Oudenaarden, A., and Page, D.C. (2015). A
490 Gene Regulatory Program for Meiotic Prophase in the Fetal Ovary. *PLoS Genet* 11, e1005531.
<https://doi.org/10.1371/journal.pgen.1005531>.
- 491 33. Magnúsdóttir, E., and Surani, M.A. (2014). How to make a primordial germ cell. *Development* 141,
492 245–252. <https://doi.org/10.1242/dev.098269>.

- 493 34. Günesdogan, U., Magnúsdóttir, E., and Surani, M.A. (2014). Primordial germ cell specification: A
494 context-dependent cellular differentiation event [corrected]. *Philos Trans R Soc Lond B Biol Sci* **369**.
<https://doi.org/10.1098/rstb.2013.0543>.
- 495 35. Bardot, E.S., and Hadjantonakis, A.-K. (2020). Mouse gastrulation: Coordination of tissue patterning,
496 specification and diversification of cell fate. *Mechanisms of Development* **163**, 103617. <https://doi.org/10.1016/j.mod.2020.103617>.
- 497 36. Welling, M., Chen, H., Muñoz, J., Musheev, M.U., Kester, L., Junker, J.P., Mischerikow, N., Arbab, M.,
498 Kuijk, E., Silberstein, L., et al. (2015). DAZL regulates Tet1 translation in murine embryonic stem cells.
EMBO Reports **16**, 791–802. <https://doi.org/10.15252/embr.201540538>.
- 499 37. Macfarlan, T.S., Gifford, W.D., Driscoll, S., Lettieri, K., Rowe, H.M., Bonanomi, D., Firth, A., Singer, O.,
500 Trono, D., and Pfaff, S.L. (2012). Embryonic stem cell potency fluctuates with endogenous retrovirus
activity. *Nature* **487**, 57–63. <https://doi.org/10.1038/nature11244>.
- 501 38. Suzuki, A., Hirasaki, M., Hishida, T., Wu, J., Okamura, D., Ueda, A., Nishimoto, M., Nakachi, Y.,
502 Mizuno, Y., Okazaki, Y., et al. (2016). Loss of MAX results in meiotic entry in mouse embryonic and
germline stem cells. *Nat Commun* **7**, 11056. <https://doi.org/10.1038/ncomms11056>.
- 503 39. Samanta, M.K., Gayen, S., Harris, C., Maclary, E., Murata-Nakamura, Y., Malcore, R.M., Porter, R.S.,
504 Garay, P.M., Vallianatos, C.N., Samollow, P.B., et al. (2022). Activation of Xist by an evolutionarily
conserved function of KDM5C demethylase. *Nat Commun* **13**, 2602. <https://doi.org/10.1038/s41467-022-30352-1>.
- 505 40. Koubova, J., Menke, D.B., Zhou, Q., Capel, B., Griswold, M.D., and Page, D.C. (2006). Retinoic
506 acid regulates sex-specific timing of meiotic initiation in mice. *Proc. Natl. Acad. Sci. U.S.A.* **103**,
2474–2479. <https://doi.org/10.1073/pnas.0510813103>.
- 507 41. Lin, Y., Gill, M.E., Koubova, J., and Page, D.C. (2008). Germ Cell-Intrinsic and -Extrinsic Factors
508 Govern Meiotic Initiation in Mouse Embryos. *Science* **322**, 1685–1687. <https://doi.org/10.1126/science.1166340>.
- 509 42. Endo, T., Mikedis, M.M., Nicholls, P.K., Page, D.C., and De Rooij, D.G. (2019). Retinoic Acid and Germ
510 Cell Development in the Ovary and Testis. *Biomolecules* **9**, 775. <https://doi.org/10.3390/biom9120775>.
- 511 43. Gupta, N., Yakhou, L., Albert, J.R., Azogui, A., Ferry, L., Kirsh, O., Miura, F., Battault, S., Yamaguchi,
512 K., Laisné, M., et al. (2023). A genome-wide screen reveals new regulators of the 2-cell-like cell state.
Nat Struct Mol Biol. <https://doi.org/10.1038/s41594-023-01038-z>.
- 513 44. Liu, M., Zhu, Y., Xing, F., Liu, S., Xia, Y., Jiang, Q., and Qin, J. (2020). The polycomb group protein
514 PCGF6 mediates germline gene silencing by recruiting histone-modifying proteins to target gene
promoters. *J Biol Chem* **295**, 9712–9724. <https://doi.org/10.1074/jbc.RA119.012121>.

- 515 45. Otani, J., Nankumo, T., Arita, K., Inamoto, S., Ariyoshi, M., and Shirakawa, M. (2009). Structural basis
516 for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX–DNMT3–DNMT3L
domain. *EMBO Reports* *10*, 1235–1241. <https://doi.org/10.1038/embor.2009.218>.
- 517 46. Guo, X., Wang, L., Li, J., Ding, Z., Xiao, J., Yin, X., He, S., Shi, P., Dong, L., Li, G., et al. (2015).
518 Structural insight into autoinhibition and histone H3-induced activation of DNMT3A. *Nature* *517*,
640–644. <https://doi.org/10.1038/nature13899>.
- 519 47. Abildayeva, K., Berbée, J.F.P., Blokland, A., Jansen, P.J., Hoek, F.J., Meijer, O., Lütjohann, D., Gautier,
520 T., Pillot, T., De Vente, J., et al. (2008). Human apolipoprotein C-I expression in mice impairs learning
and memory functions. *Journal of Lipid Research* *49*, 856–869. <https://doi.org/10.1194/jlr.M700518-JLR200>.
- 521 48. Leduc, V., Jasmin-Bélanger, S., and Poirier, J. (2010). APOE and cholesterol homeostasis in
522 Alzheimer's disease. *Trends in Molecular Medicine* *16*, 469–477. <https://doi.org/10.1016/j.molmed.2010.07.008>.
- 523 49. Dahlet, T., Truss, M., Frede, U., Al Adhami, H., Bardet, A.F., Dumas, M., Vallet, J., Chicher, J.,
524 Hammann, P., Kotnik, S., et al. (2021). E2F6 initiates stable epigenetic silencing of germline genes
during embryonic development. *Nat Commun* *12*, 3582. <https://doi.org/10.1038/s41467-021-23596-w>.
- 525 50. Hayashi, K., Ohta, H., Kurimoto, K., Aramaki, S., and Saitou, M. (2011). Reconstitution of the
526 mouse germ cell specification pathway in culture by pluripotent stem cells. *Cell* *146*, 519–532.
<https://doi.org/10.1016/j.cell.2011.06.052>.
- 527 51. Zhang, J., Zhang, M., Acampora, D., Vojtek, M., Yuan, D., Simeone, A., and Chambers, I. (2018).
528 OTX2 restricts entry to the mouse germline. *Nature* *562*, 595–599. <https://doi.org/10.1038/s41586-018-0581-5>.
- 529 52. Endo, T., Romer, K.A., Anderson, E.L., Baltus, A.E., De Rooij, D.G., and Page, D.C. (2015). Pe-
530 riodic retinoic acid–STRA8 signaling intersects with periodic germ-cell competencies to regulate
spermatogenesis. *Proc. Natl. Acad. Sci. U.S.A.* *112*. <https://doi.org/10.1073/pnas.1505683112>.
- 531 53. Bowles, J., Knight, D., Smith, C., Wilhelm, D., Richman, J., Mamiya, S., Yashiro, K., Chawengsak-
532 sophak, K., Wilson, M.J., Rossant, J., et al. (2006). Retinoid Signaling Determines Germ Cell Fate in
Mice. *Science* *312*, 596–600. <https://doi.org/10.1126/science.1125691>.
- 533 54. Yang, M., Yu, H., Yu, X., Liang, S., Hu, Y., Luo, Y., Izsvák, Z., Sun, C., and Wang, J. (2022). Chemical-
534 induced chromatin remodeling reprograms mouse ESCs to totipotent-like stem cells. *Cell Stem Cell*
29, 400–418.e13. <https://doi.org/10.1016/j.stem.2022.01.010>.
- 535 55. 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>.

- 536
- 537 56. Karimi, M.M., Goyal, P., Maksakova, I.A., Bilenky, M., Leung, D., Tang, J.X., Shinkai, Y., Mager, D.L., Jones, S., Hirst, M., et al. (2011). DNA Methylation and SETDB1/H3K9me3 Regulate Predominantly Distinct Sets of Genes, Retroelements, and Chimeric Transcripts in mESCs. *Cell Stem Cell* *8*, 676–687. <https://doi.org/10.1016/j.stem.2011.04.004>.
- 538
- 539 57. Tatsumi, D., Hayashi, Y., Endo, M., Kobayashi, H., Yoshioka, T., Kiso, K., Kanno, S., Nakai, Y., Maeda, I., Mochizuki, K., et al. (2018). DNMTs and SETDB1 function as co-repressors in MAX-mediated repression of germ cell-related genes in mouse embryonic stem cells. *PLoS ONE* *13*, e0205969. <https://doi.org/10.1371/journal.pone.0205969>.
- 540
- 541 58. Bonefas, K.M., and Iwase, S. (2021). Soma-to-germline transformation in chromatin-linked neurode-
542 velopmental disorders? *FEBS J.* <https://doi.org/10.1111/febs.16196>.
- 543 59. Velasco, G., Walton, E.L., Sterlin, D., Hédonin, S., Nitta, H., Ito, Y., Fouyssac, F., Mégarbané, A.,
Sasaki, H., Picard, C., et al. (2014). Germline genes hypomethylation and expression define a
molecular signature in peripheral blood of ICF patients: Implications for diagnosis and etiology.
544 *Orphanet J Rare Dis* *9*, 56. <https://doi.org/10.1186/1750-1172-9-56>.
- 545 60. Walton, E.L., Francastel, C., and Velasco, G. (2014). Dnmt3b Prefers Germ Line Genes and Cen-
tromeric Regions: Lessons from the ICF Syndrome and Cancer and Implications for Diseases. *Biology*
546 (*Basel*) *3*, 578–605. <https://doi.org/10.3390/biology3030578>.
- 547 61. Samaco, R.C., Mandel-Brehm, C., McGraw, C.M., Shaw, C.A., McGill, B.E., and Zoghbi, H.Y. (2012).
Crh and Oprm1 mediate anxiety-related behavior and social approach in a mouse model of MECP2
548 duplication syndrome. *Nat Genet* *44*, 206–211. <https://doi.org/10.1038/ng.1066>.
- 549 62. Samanta, M., and Kalantry, S. (2020). Generating primed pluripotent epiblast stem cells: A methodol-
550 ogy chapter. *Curr Top Dev Biol* *138*, 139–174. <https://doi.org/10.1016/bs.ctdb.2020.01.005>.
- 551 63. Stephens, M. (2016). False discovery rates: A new deal. *Biostat*, kxw041. <https://doi.org/10.1093/biostatistics/kxw041>.
- 552
- 553 64. Conway, J.R., Lex, A., and Gehlenborg, N. (2017). UpSetR: An R package for the visualization of
intersecting sets and their properties. *Bioinformatics* *33*, 2938–2940. <https://doi.org/10.1093/bioinformatics/btx364>.
- 554
- 555 65. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H.,
and Glass, C.K. (2010). Simple Combinations of Lineage-Determining Transcription Factors Prime
cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Molecular Cell* *38*, 576–589.
556 <https://doi.org/10.1016/j.molcel.2010.05.004>.
- 557 66. Li, H., Liang, Z., Yang, J., Wang, D., Wang, H., Zhu, M., Geng, B., and Xu, E.Y. (2019). DAZL
is a master translational regulator of murine spermatogenesis. *Natl Sci Rev* *6*, 455–468. <https://doi.org/10.1093/nsr/nwy163>.

558

- 559 67. Mikedis, M.M., Fan, Y., Nicholls, P.K., Endo, T., Jackson, E.K., Cobb, S.A., De Rooij, D.G., and Page,
560 D.C. (2020). DAZL mediates a broad translational program regulating expansion and differentiation of
spermatogonial progenitors. *eLife* 9, e56523. <https://doi.org/10.7554/eLife.56523>.

561 **Figures and Tables**

- 562 • Supplementary table 1: list of all germline genes.
- 563 – Columns to include:
- 564 * KDM5C bound vs not
- 565 * DEG in EpiLC, brain, both, neither (separate columns?)

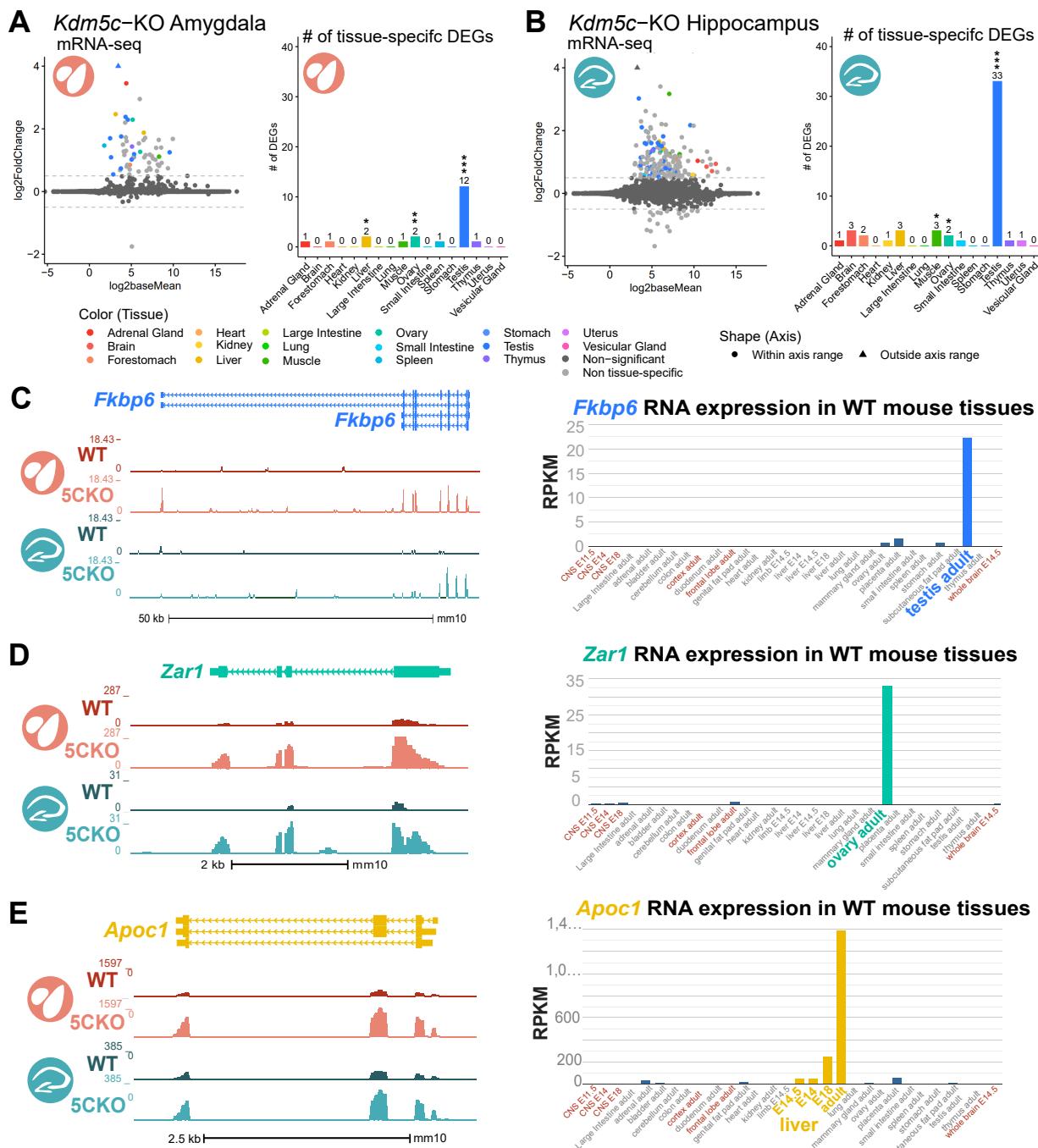


Figure 1: Tissue-enriched genes are misexpressed in the *Kdm5c*-KO brain. **A-B.** Expression of tissue-enriched genes (Li et al 2017) in the male *Kdm5c*-KO amygdala (A) and hippocampus (B). Left - MA plot of mRNA-seq. Right - Number of tissue-enriched differentially expressed genes (DEGs). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, Fisher's exact test. **C.** Left - Average bigwigs of an example aberrantly expressed testis-enriched DEG, *FK506 binding protein 6* (*Fkbp6*) in the wild-type (WT) and *Kdm5c*-KO (5CKO) amygdala (red) and hippocampus (teal). Right - Expression of *Cyct* in wild-type tissues from NCBI Gene, with testis highlighted in blue and brain tissues highlighted in red. **D.** Left - Average bigwigs of an example ovary-enriched DEG, *Zygotic arrest 1* (*Zar1*). Right - Expression of *Zar1* in wild-type tissues from NCBI Gene, with ovary highlighted in teal and brain tissues highlighted in red. **E.** Left - Average bigwigs of an example liver-enriched DEG, *Apolipoprotein C-I* (*Apoc1*). Right - Expression of *Apoc1* in wild-type tissues from NCBI Gene, with liver highlighted in orange and brain tissues highlighted in red.

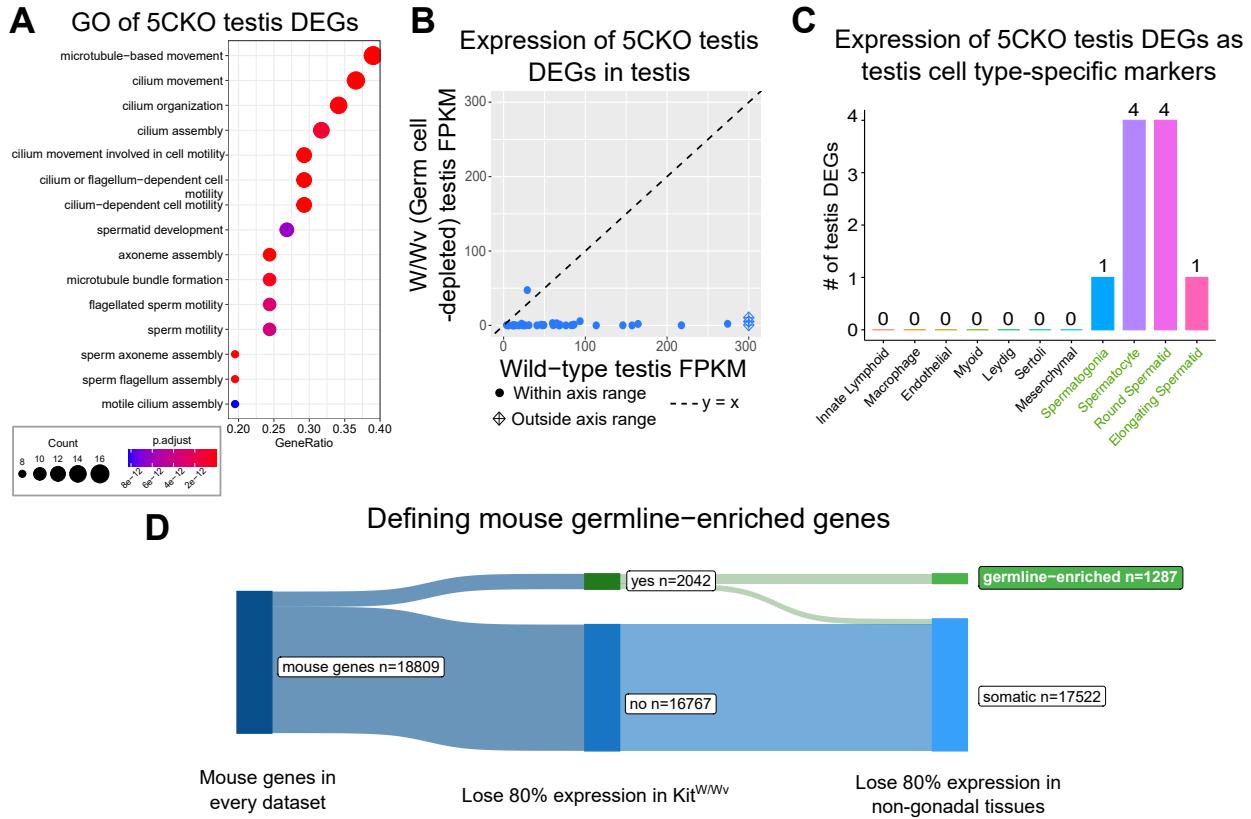


Figure 2: Aberrant transcription of germline genes in the *Kdm5c*-KO in the brain. **A.** enrichPlot gene ontology (GO) of *Kdm5c*-KO amygdala and hippocampus testis-enriched DEGs **B.** Expression of testis DEGs in wild-type (WT) testis versus germ cell-depleted (W/Wv) testis (Mueller et al 2013). Expression is in Fragments Per Kilobase of transcript per Million mapped reads (FPKM). **C.** Number of testis DEGs that were classified as cell-type specific markers in a single cell RNA-seq dataset of the testis (Green et al 2018). Germline cell types are highlighted in green, somatic cell types in black. **D.** Sankey diagram of mouse genes filtered for germline enrichment based on their expression in wild-type and germline-depleted mice and in adult mouse non-gonadal tissues (Li et al 2017).

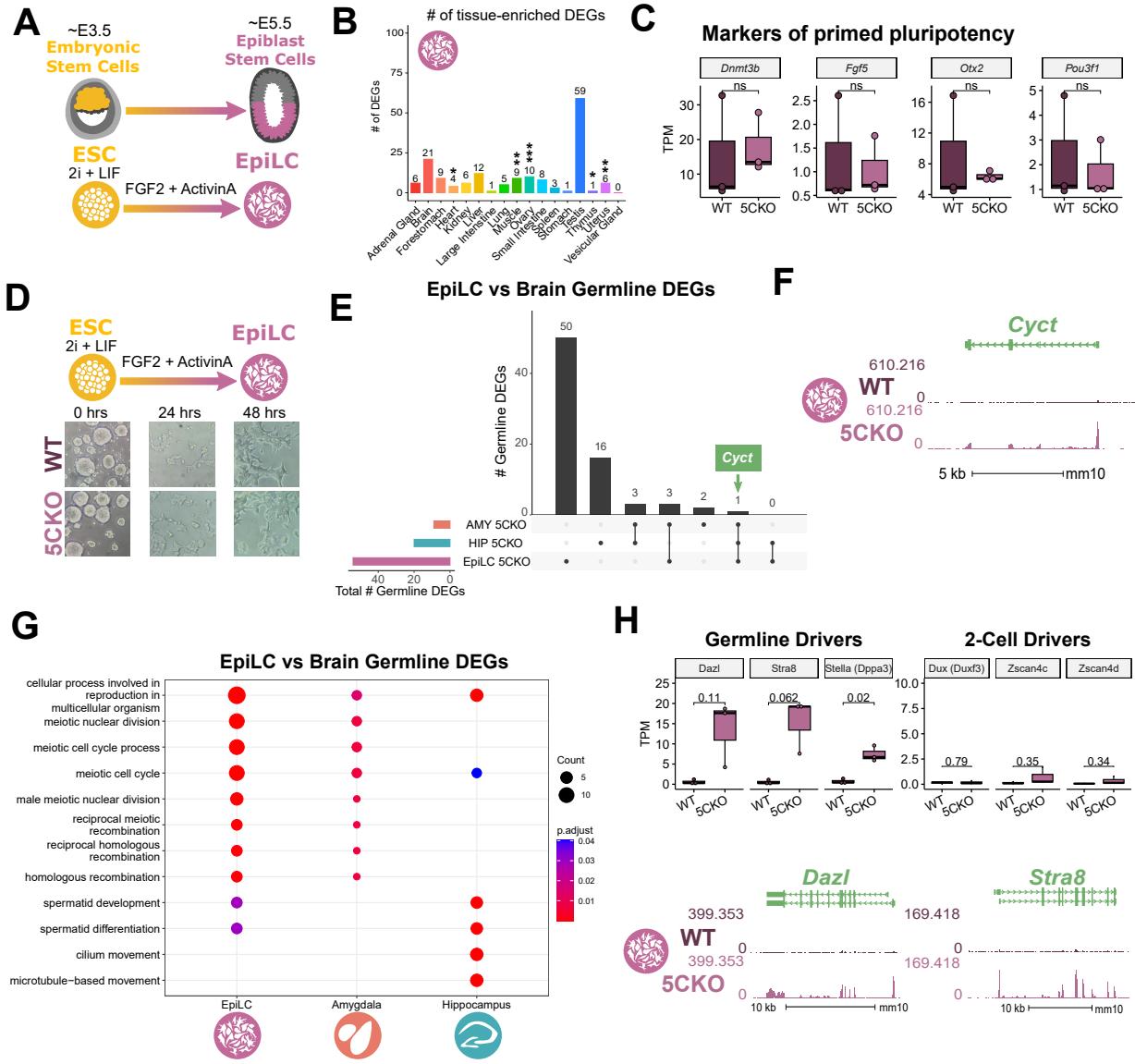


Figure 3: *Kdm5c*-KO epiblast-like cells express key drivers of germline identity

A. Top - Diagram of *in vivo* differentiation of embryonic stem cells (ESCs) of the inner cell mass into epiblast stem cells. Bottom - *in vitro* differentiation of ESCs into epiblast-like cells (EpiLCs).

B. Number of tissue-enriched differentially expressed genes (DEGs) in *Kdm5c*-KO EpiLCs. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, Fisher's exact test.

C. Average bigwigs of an example germline gene, *Cyct*, that is dysregulated *Kdm5c*-KO EpiLCs.

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

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

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

G. enrichPlot comparing enriched biological process gene ontologies for *Kdm5c*-KO EpiLC, amygdala, and hippocampus germline DEGs.

H. Top left - Example germline identity DEGs unique to EpiLCs. Top right - Example 2-cell genes that are not dysregulated in *Kdm5c*-KO EpiLCs. p-values for Welch's t-test. Bottom - Average bigwigs of *Dazl* and *Stra8* expression in wild-type and *Kdm5c*-KO EpiLCs. Immunocytochemistry of DAZL in male wild-type (WT) and *Kdm5c*-KO (5CKO) EpiLCs. Percentage of DAZL-positive cells normalized to number of DAPI-positive cells, p-value for Welch's t-test.

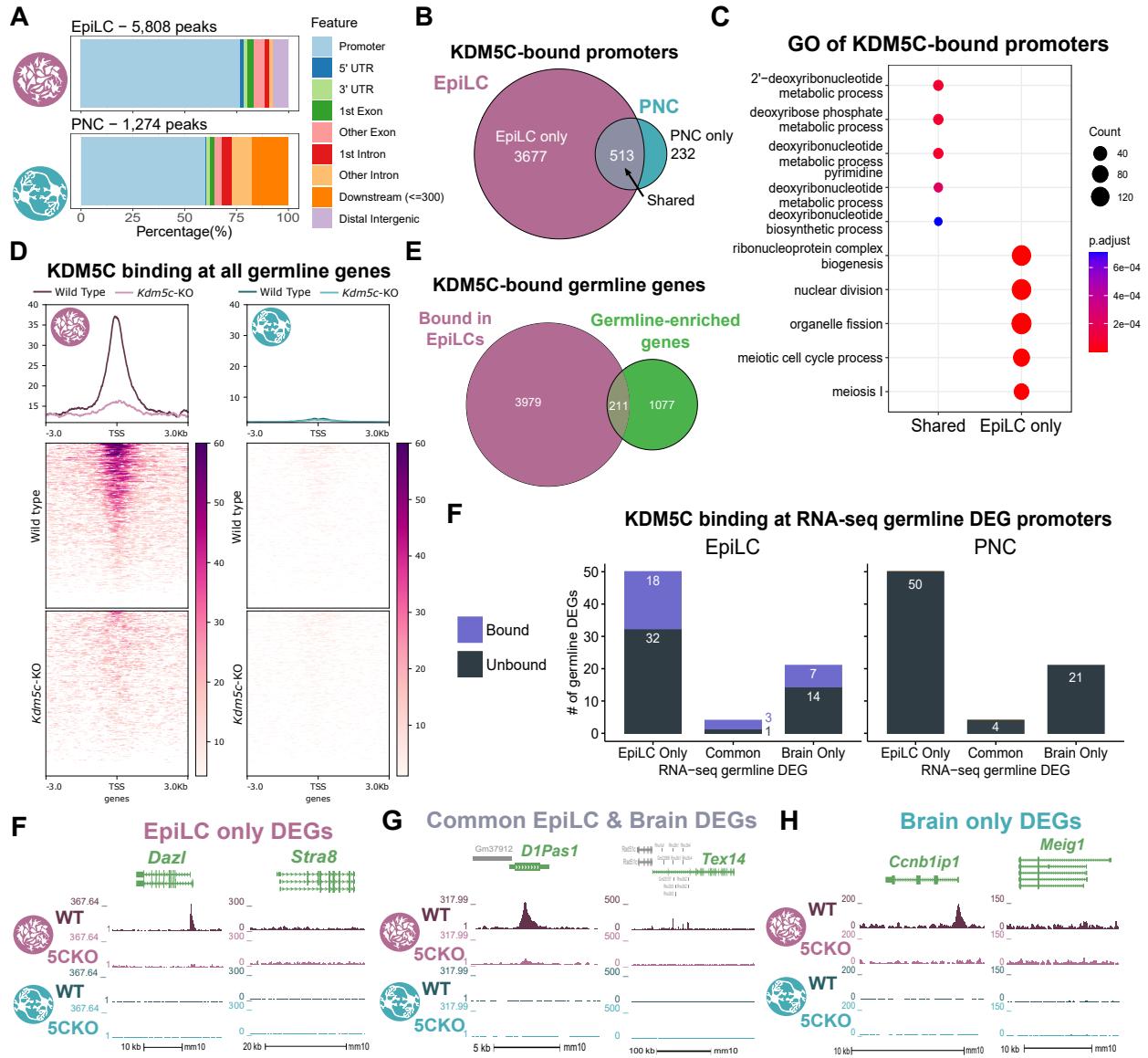
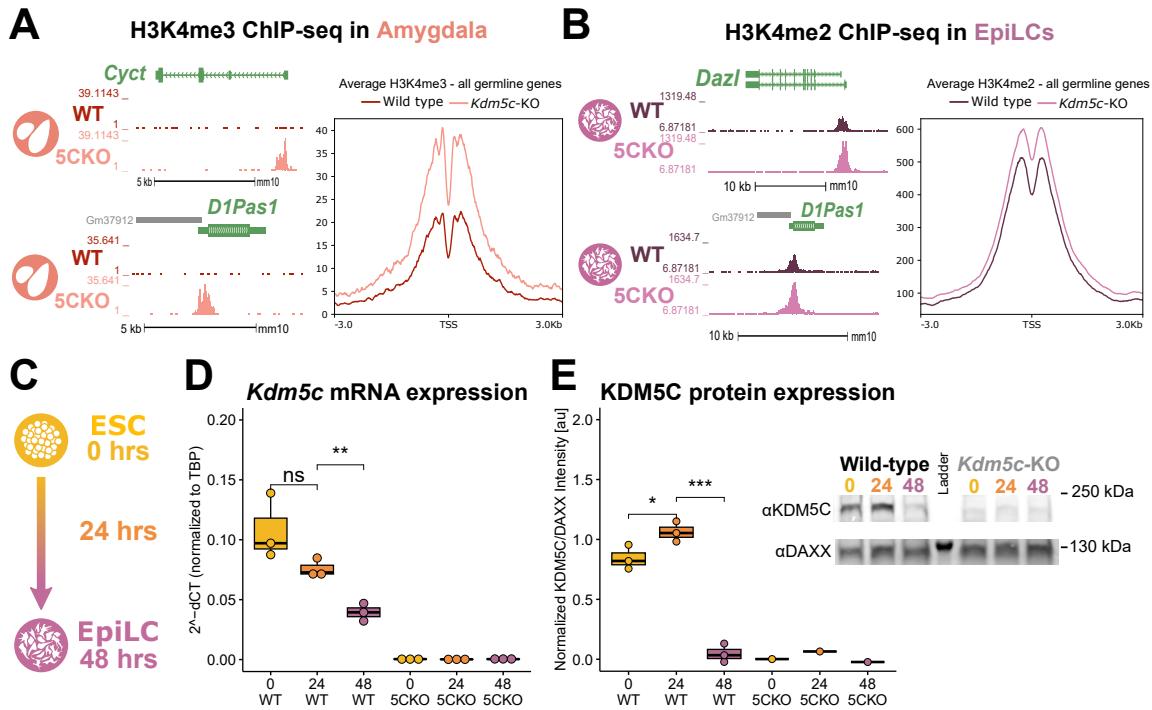


Figure 4: KDM5C binds to a subset of germline gene promoters during early embryogenesis **A.** ChIPseeker localization of KDM5C peaks at different genomic regions in EpiLCs (top) and hippocampal and cortex primary neuron cultures (PNCs, bottom). **B.** Overlap of genes with KDM5C bound to their promoters in EpiLCs (purple) and PNCs (blue). **C.** Gene ontology (GO) comparison of genes with KDM5C bound to their promoter in EpiLCs and PNCs. Genes were classified as either bound in EpiLCs only (EpiLC only), unique to PNCs (PNC only, no significant ontologies) or bound in both PNCs and EpiLCs (shared). **D.** Average KDM5C binding around the transcription start site (TSS) of all germline-enriched genes in EpiLCs (left) and PNCs (right). **E.** KDM5C binding at the promoters of RNA-seq germline DEGs. Genes were classified as either only dysregulated in EpiLCs (EpiLC only), genes dysregulated in the hippocampus or amygdala but not EpiLCs (brain only), or genes dysregulated in both EpiLCs and the brain (common). (Legend continued on next page.)

Figure 4: KDM5C binds to a subset of germline gene promoters during early embryogenesis. (Legend continued.) **F.** Example KDM5C ChIP-seq bigwigs of DEGs unique to EpiLCs. Although both are expressed in *Kdm5c*-KO EpiLCs, KDM5C is only bound to the *Dazl* promoter and not the *Stra8* promoter in EpiLCs. **G.** Example KDM5C ChIP-seq bigwigs of RNA-seq DEGs common between the brain and EpiLCs. KDM5C is bound to the *D1Pas1* promoter but not the *XXX* promoter in EpiLCs. **H.** Example KDM5C ChIP-seq bigwigs of RNA-seq DEGs unique to the brain. KDM5C is bound to the *XXX* promoter but not the *Meig1* promoter



F

G

Figure 5: KDM5C's catalytic activity promotes long-term silencing of germline genes via DNA methylation. **A.** Left - Bigwigs of representative histone 3 lysine 4 trimethylation (H3K4me3) ChIP-seq peaks at two germline genes in the wild-type (WT) and *Kdm5c*-KO (5CKO) adult amygdala. Right - Average H3K4me3 at the transcripton start site (TSS) of all germline-enriched genes in wild-type (dark red) and *Kdm5c*-KO (light red) amygdala. **B.** Left - Bigwigs of representative histone 3 lysine 4 dimethylation (H3K4me2) ChIP-seq peaks at representative germline genes in wild-type and *Kdm5c*-KO EpiLCs. Right - Average H3K4me2 at the TSS of all germline-enriched genes in wild-type (dark purple) and *Kdm5c*-KO (light purple) EpiLCs. **C.** Diagram of embryonic stem cell (ESC) to epiblast-like cell (EpiLC) differentiation protocol and collection time points for RNA and protein. **D.** Real time quantitative PCR (RT-qPCR) of *Kdm5c* RNA expression, calculated in comparision to TBP expression ($2^{-\Delta\Delta CT}$). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, Welch's t-test. **E.** KDM5C protein expression normalized to DAXX. Quantified intensity using ImageJ (artificial units - au). Right - representative lanes of Western blot for KDM5C and DAXX. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, Welch's t-test **F.** XXX **G.** XXX

566 Notes

567 Things to do

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

572 Dazl

573 We were particularly interested in the aberrant transcription of *Dazl*, since it is essential for germ cell
574 development and promotes the translation of germline mRNAs⁶⁶. A significant portion of germline transcripts
575 misexpressed in *Kdm5c*-KO EpiLCs are known binding targets of DAZL, including *Stra8*⁶⁷ ($p = 1.698e-07$,
576 Fisher's Exact Test). This suggests expression of DAZL protein could promote the translation of other
577 aberrant germline transcripts, influencing their ability to impact *Kdm5c*-KO cellular function. We thus tested
578 DAZL protein expression in *Kdm5c*-KO EpiLCs through immunocytochemistry (Figure 3I). We observed
579 about 25% of *Kdm5c*-KO EpiLCs expressed DAZL protein and it was localized to the cytoplasm ($p = 0.0015$,
580 Welch's t-test), consistent with the pattern of DAZL expression in spermatogonia⁶⁷. Altogether these results
581 suggest tissue-specific genes are misexpressed during *Kdm5c*-KO embryogenesis, including key drivers of
582 germline identity that can be translated into protein.

- 583 • We additionally found *Kdm5c*-KO EpiLCs ectopically express DAZL protein that is localized to the
584 cytoplasm, similar to its morphology in spermatogonia⁶⁷. **note: maybe just put in results.** Could
585 move around depending upon if I get pheno working.

586 Discussion notes

- 587 • For other paper:
 - 588 – for methods: Heatmaps of gene expression were generated using the base R functions scale and
589 hclust and visualized using the R package ComplexHeatmap (v2.12.1).
 - 590 – * Might be good to look at for retinoic acid paper (WT germ expression): <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0005654>
- 592 • Talk about different motifs if we do see differences there and if it explains direct vs indirect dysregulation
 - 593 – Could also mention PRC1.6/KDM5C binds pcgf6 in ESCs
- 594 • Maybe talk about other regulators/if anything is known about them long term? Or save for 2nd paper.

- 595 – another germline repressor: <https://pubmed.ncbi.nlm.nih.gov/35345626/>
- 596 • end with something like: this indicates kdm5c not only modulates pluripotency and self-renewal in
597 ESCs, but also has a role in long-term silencing of germline genes
- 598 – then transition into the long term silencing mechanism paragraph
- 599 • KDM5C is dynamically regulated during the window of embryonic germline gene silencing, our EpiLC
600 ChIP-seq is likely catching the tail end of KDM5C's main involvement.
- 601 • Our results indicate KDM5C is a pivotal repressor of germline programs in somatic cells.
- 602 • Studies on other chromatin regulators have focused on key marker genes to identify germline gene
603 misexpression, such as *Dazl*.
- 604 • Using a germline-depleted mouse model, we curated a list of mouse germline-enriched genes to
605 globally assess germline gene dysregulation.
- 606 • Many of the genes dysregulated in the mature *Kdm5c*-KO brain were important for late stages of
607 spermatogenesis, such as those important for sperm flagellar structure. Contrastingly, *Kdm5c*-KO
608 EpiLCs aberrantly expressed key drivers of germ cell identity, including *Dazl* and *Stra8*.
- 609 • While a significant portion of KDM5C-bound promoters are germline genes, many germline genes
610 expressed during *Kdm5c*-KO embryogenesis are not directly bound by kdm5c.
- 611 • One notable example is *Stra8*, a transcription factor important for germ cell specific transcription and
612 meiotic initiation
- 613 • The including the demarcation between soma and germline fates.
- 614 the *Kdm5c*-KO brain also Previous work identified testis-enriched genes in *Kdm5c* genes
- 615 –
- 616 – However unlike the gonadal-biased DEGs,
- 617 • The demarcation of the germline versus soma is a key feature of multicellularity and sexually dimorphic
618 reproduction
- 619 • Anything known about tissue-biased gene expression in other H3K4me regulators?
- 620 • Our data suggests the germline developmental program is occurring ectopically as *Kdm5c*-KOs pro-
621 gresses through somatic tissue development
- 622 • tissue-biased gene expression:

- 623 • Altogether, these data indicate that while some germline genes are misexpressed due to direct loss of
624 KDM5C binding during emryogenesis, secondary downstream mechanisms can also promote their
625 aberrant transcription.
- 626 • Papers to read/reference:
627 – Reconstitution of the Mouse Germ Cell Specification Pathway in Culture by Pluripotent Stem Cells:
628 [https://www.cell.com/fulltext/S0092-8674\(11\)00771-9](https://www.cell.com/fulltext/S0092-8674(11)00771-9)
629 – 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/>

631 **Figure outline:**

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

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

644 **Figure 3: *Kdm5c*-KO epiblast-like cells express key drivers regulators of germline identity** * A) ESC
645 to EpiLC differentiation Left - Morphology is unchanged, * B) 5CKO EpiLCs express EpiLC differentiation
646 genes similar to WT lvs * C) Male EpiLCs express germline genes (example Cyct again) * Overlap between
647 brain and EpiLC germline genes - show they're mostly unique * GO of Brain and EpiLC germline genes
648 (meiotic enriched) * Bigwigs or TPM of master regulators * Show that while some are also 2-cell genes
649 (Dazl), 2-cell specific genes aren't dysregulated (Zscan4). Important point because published KDM5C dazl
650 paper is saying KDM5C is a 2-cell regulator, but as far as I can tell only genes shared between germline and
651 2-cell are dysregulated.

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

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

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

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

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

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

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

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

696 **Germline gene repression background:**

697 Interestingly, some of the ectopic testis transcripts identified in the *Kdm5c*-KO brain are typically ex-
698 pressed in germ cells¹². Unlike somatic cells, germ cells (e.g. sperm and eggs) undergo meiosis and pass
699 on their genetic material to the next generation. The germline and the soma are typically distinguished during
700 early embryogenesis, when germline genes are silenced in epiblast stem cells soon after implantation and
701 only reactivated in a subset to form the germline. Chromatin regulators play a key role in decommissioning
702 germline genes as the embryo transitions from naïve to primed pluripotency by placing repressive histone
703 H2A lysine 119 monoubiquitination (H2AK119ub1)¹⁸, histone 3 lysine 9 trimethylation (H3K9me3)^{18,19}, and
704 DNA CpG methylation^{19–21} at germline gene promoters. KDM5C may also be involved in this early decom-
705 missioning of germline genes, as re-expression of KDM5C in neurons fails to suppress their dysregulation¹².
706 In support of this, KDM5C was very recently shown to repress *Deleted in azoospermia like (Dazl)*, a key
707 regulator of germline development, in mouse embryonic stem cells (ESCs)^{43,55}. In support of this, two
708 independent screens in mouse embryonic stem cells (ESCs) recently identified KDM5C as a repressor of
709 *Deleted in azoospermia like (Dazl)*, a key regulator of germline development. However, KDM5C's role in
710 embryonic germline gene repression is currently unclear, given that Dazl is also expressed in ESCs and in
711 the 2-cell stage and germline gene misexpression has yet to be globally characterized during *Kdm5c*-KO
712 embryogenesis.