

1 Misexpression of germline programs with loss of the X-linked intellectual
2 disability factor KDM5C

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

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

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

36 **Results**

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

38 RNA sequencing (RNA-seq) studies surprisingly revealed the aberrant expression of testis genes within the adult *Kdm5c*
39 knockout (-KO) hippocampus¹. Given the high abundance of testis-enriched genes within the mouse transcriptome, it is
40 currently unclear if their misexpression in the *Kdm5c*-KO brain is indicative of random de-repression throughout the genome
41 or instead reflects an impairment in tissue identity. To distinguish these possibilities, we globally assessed the expression of
42 previously characterized tissue-enriched genes² in our previously published mRNA-seq datasets of the *Kdm5c*-KO amygdala
43 and hippocampus³

44 We found many genes significantly upregulated within the *Kdm5c*-KO brain (DESeq2⁴, log2 fold change > 0.5, q < 0.1) are
45 typically enriched in non-brain tissues, with the majority of differentially expressed genes (DEGs) displaying biased expression
46 towards the testis (Figure 1A-C). Even though the testis has the largest number of tissue-biased genes compared to any
47 other tissue (2,496 genes), testis-biased DEGs were significantly enriched for both brain regions (Amygdala p = 1.83e-05;
48 Hippocampus p = 4.26e-11, Fisher's Exact).

49 Even though the *Kdm5c*-KO mice are male, we also observed significant enrichment of ovary-biased DEGs within both
50 brain regions (Amygdala p = 0.00574; Hippocampus p = 0.048, Fisher's Exact) (Figure 1D). Intriguingly, many ovary and
51 testis-biased DEGs have functions specific to germ cells and have no known role in the brain. For example, the testis-biased
52 DEG *Cytochrome C, testis* (*Cyct*) is a testis-specific cytochrome c isoform found in spermatozoa mitochondria and is important
53 for flagellar beating^{5,6} (Figure 1C). The ovary-enriched DEG *Zygotic arrest 1* (*Zar1*) was recently shown to sequester mRNAs
54 in oocytes for meiotic maturation and early zygote development⁷ (Figure 1D). Although not consistent across brain regions,
55 we also found enrichment of two non-gonadal tissues - the liver (Amygdala p = 0.0398, Fisher's Exact) and the muscle
56 (Hippocampus p = 0.0104, Fisher's Exact). One example of a liver-biased DEG is *Apolipoprotein C-I* (*Apoc1*), which is
57 involved in lipoprotein metabolism (Figure 1E). Overall, these tissue-enriched genes show little to no expression in the
58 wild-type brain, yet our mRNA-seq data indicate they are polyadenylated and spliced into mature transcripts (Figure 1C-E).

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

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

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

69 To further validate if these testis DEGs are truly germline genes, we then compared their expression in a published
70 RNA-seq dataset of wild-type (WT) and germ cell-depleted (*Kit^{W/Wv}*) mouse testes⁸. We found almost all *Kdm5c*-KO testis-
71 enriched DEGs lose their expression with germ cell depletion (Figure 2B). The only testis DEG that did not show considerable
72 downregulation with germline depletion was *FK506 binding protein 6* (*Fkbp6*), a known regulator of piRNA expression and

73 meiosis in germ cells^{9,10}. We additionally assessed testis DEG expression in a published testis single cell RNA-seq dataset
74 that identified cell type-specific markers¹¹. We found that while some testis-enriched DEGs were classified as specific
75 markers for different germ cell types (e.g. spermatogonia, spermatocytes, round spermatids, and elongating spermatids)
76 none marked somatic cells (Figure 2C). Together, these data suggest the *Kdm5c*-KO brain expresses germline genes.

77 We then aimed to globally characterize germline gene misexpression in *Kdm5c*-KO mice, but lacked a comprehensive list
78 of mouse germline-enriched genes. To facilitate downstream analyses, we generated a list of male and female germline-
79 enriched genes by evaluating gene expression in wild-type and germline-depleted (*Kit^{W/Wv}*) mice. We assessed gene
80 expression in male and female mice at embryonic day 12, 14, and 16 embryos, as well as male postnatal day 6 and adult
81 testes. We defined genes as germline-enriched if their expression met the following criteria: 1) their expression is greater than
82 1 FPKM in wild-type germline 2) their expression in any wild-type, non-gonadal tissue² does not exceed 20% of maximum
83 expression in wild-type germline, and 3) their expression in the germ cell-depleted germline, for any sex or time point, does
84 not exceed 20% of their maximum expression in the wild-type germline. These criteria yielded 1,287 germline-enriched
85 genes.

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

87 Misexpression of germline genes in the *Kdm5c*-KO brain suggests mutants fail to demarcate germline and somatic cellular
88 identity. Germ cells are typically distinguished from somatic cells soon after the embryo implants into the uterine wall^{12,13}
89 when a subset of epiblast stem cells become the primordial germ cells (PGCs) while the remainder differentiate into the
90 ectoderm, mesoderm, and endoderm to form the body's somatic tissues¹⁴. This developmental time point can be modeled *in*
91 *vitro* through differentiation of embryonic stem cells (ESCs) into post-implantation epiblast-like stem cells (EpiLCs) (Figure
92 3A, top). Previous studies have demonstrated that while some germline-enriched genes are also expressed in embryonic
93 stem cells (ESCs), they are silenced as they differentiate into EpiLCs¹⁵. Therefore, we assessed if KDM5C was necessary
94 for initial germline gene silencing by assessing their expression in male *Kdm5c*-KO EpiLCs using our previously generated
95 RNA-seq dataset¹⁶.

96 We did not observe any apparent morphology changes between wild-type and *Kdm5c*-KO cells during ESC to EpiLC
97 differentiation (Figure 3A, bottom) and found no significant difference in primed pluripotency genes that indicate proper EpiLC
98 differentiation (Figure 3B). We then identified *Kdm5c*-KO EpiLC DEGs through DESeq2⁴ (log2 fold change > 0.5, q < 0.1) and
99 again observed widespread dysregulation of tissue-enriched genes, with the largest number of genes belonging to the brain
100 and testis, although they were not significantly enriched (Figure 3C).

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

105 We then compared the function of EpiLC and brain germline DEGs through gene ontology and found particularly high
106 enrichment of meiosis-related gene ontologies in EpiLCs (Figure 3F), such as meiotic cell cycle (GO:0,051,321, p.adjust =
107 4.44e-07). While a few meiosis-related genes are misexpressed in the *Kdm5c*-KO brain, DEGs unique to *Kdm5c*-KO EpiLCs
108 included master regulators of germline identity, such as *Stimulated by retinoic acid 8 (Stra8)* and *Deleted in azoospermia*
109 *like (Dazl)* (Figure 3G). These genes are typically expressed during embryonic germ cell development to commit PGCs
110 to the germline fate, but are also expressed later in life to trigger meiotic gene expression programs¹⁷⁻¹⁹. Of note, some
111 germline genes, including *Dazl*, are also expressed in the two-cell embryo. However, we did not see misexpression of two-cell

112 embryo-specific genes, like *Zscan4d* ($q = 0.381$).

113 *Dazl* is an RNA-binding protein that regulates the translation of germline mRNAs and is essential for germ cell function²⁰.
114 Interestingly, a significant portion of germline transcripts expressed in *Kdm5c*-KO EpiLCs are known binding targets of
115 DAZL, including *Stra8*²¹ ($p = 1.698e-07$, Fisher's Exact Test). This suggests expression of DAZL protein would enable
116 the translation of other aberrant germline transcripts, influencing their ability to impact on *Kdm5c*-KO cellular function. We
117 therefore stained *Kdm5c*-KO EpiLCs for DAZL protein expression through immunocytochemistry (Figure 3H). We found
118 about 25% of *Kdm5c*-KO EpiLCs expressed DAZL protein in their cytoplasm ($p = 0.0015$, Welch's t-test), consistent with the
119 subcellular localization observed when DAZL is stained in spermatogonia²¹. Altogether these results suggest *Kdm5c*-KO
120 EpiLCs fail to decommission germline genes during early embryogenesis, including master regulators of germline identity
121 that can be translated into protein.

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

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

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

128 --- separate into two figures

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

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

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

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

133 G) RNA expression of germline genes with catalytic dead rescue (Ilakkiya)

134 H) DNA methylation in WT and 5CKO EpiLCs (Ilakkiya)

135

136 Direct vs indirect binding motif analysis

137 Previous work suggests KDM5C represses germline genes during early development, since KDM5C binding is gradually
138 lost at hippocampal germline DEG promoters over development¹ and re-expression of KDM5C in neuronal cultures fails to
139 suppress their transcription. However, we identified many germline genes that were only dysregulated in *Kdm5c*-KO EpiLCs
140 and not in the mature brain, including master regulators of meiosis and germline identity. While KDM5C was recently shown
141 to bind to *Dazl* in ESCs^{22,23}, it's unclear if KDM5C binds to other master regulators that are expressed in *Kdm5c*-KO EpiLCs
142 and if this binding is maintained as cells differentiate.

143 To better elucidate the mechanism behind KDM5C-mediated germline gene silencing, we analyzed KDM5C chromatin
144 immunoprecipitation followed by DNA sequencing (ChIP-seq) datasets we previously generated in EpiLCs¹⁶ and primary
145 hippocampal and cortical neuron cultures (PNCs)²⁴. We identified a higher number of KDM5C peaks in EpiLCs overall
146 (EpiLCs: 2,437, PNCs: 1,610, MACS2 $q < 0.1$ and fold enrichment > 1) and while KDM5C was primarily localized to gene
147 promoters in both cell types (EpiLCs: , PNCs: , +/- kb TSS), PNCs showed increased localization to distal intergenic regions
148 (EpiLCs: , PNCs:) (Figure 4A). Gene ontology analysis of KDM5C-bound promoters in EpiLCs revealed high enrichment
149 of germline-related ontologies, including meiotic nuclear division (GO:0140013) and synaptonemal complex assembly

150 (GO:0007130) (Figure 4B). Contrastingly, germline ontologies are not enriched in PNCs and KDM5C-bound genes instead
151 perform non-germline-specific functions, such as heterocycle synthesis (GO: 0018130) and pyrimidine metabolism (GO:
152 0072527).

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

155 * KDM5C was bound to the promoter of many genes dysregulated in both EpiLCs and brain datasets, including _D1P
156 * KDM5C was not bound to many of the brain-specific germline DEGs, such as XXX.
157 * While KDM5C binds to many germline DEGs in EpiLCs, including _Dazl_, many were also unbound. This included t.
158 * All genes lacked binding in PNCs.
159 * Altogether, these data indicate that while some germline genes are misexpressed due to direct loss of KDM5C

160

161 Although KDM5C is generally thought to suppress transcription through erasure of histone 3 lysine 4 di- and trimethylation
162 (H3K4me2/3), recent studies in ESCs have suggested KDM5C's repression *Dazl* is independent of its catalytic activity²². It
163 is currently unclear if KDM5C's catalytic activity aids in the long-term repression of germline genes, such as enabling the
164 placement of DNA methylation.

165 • In congruence with previous work in the *Kdm5c*-KO hippocampus¹, we observed ectopic H3K4me3 peaks in the
166 *Kdm5c*-KO amygdala at the transcription start site (TSS) of germline genes (Figure 4).
167 • We additionally observed a marked increase in H3K4me2 at germline gene promoters in *Kdm5c*-KO EpiLCs.
168 – This increase was greatest at germline genes directly bound by KDM5C (?)

169 Discussion

170 • tissue-biased gene expression:

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

174 • Otx2 is properly expressed in EpiLCs and prevents pgc identity <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6485399/>

175 • Papers to read/reference:

176 – 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)

177 – 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/>

180 **References**

- 181 1. 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>.
- 182 2. 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>.
- 183 3. 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>.
- 184 4. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15, 550. <https://doi.org/10.1186/s13059-014-0550-8>.
- 185 5. Goldberg, E., Sberna, D., Wheat, T.E., Urbanski, G.J., and Margoliash, E. (1977). Cytochrome c: Immunofluorescent Localization of the Testis-Specific Form. *Science* 196, 1010–1012. <https://doi.org/10.1126/science.193188>.
- 186 6. Narisawa, S., Hecht, N.B., Goldberg, E., Boatright, K.M., Reed, J.C., and Millán, J.L. (2002). Testis-Specific Cytochrome 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>.
- 187 7. 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 membraneless compartment. *Science* 378, eabq4835. <https://doi.org/10.1126/science.abq4835>.
- 188 8. Mueller, J.L., Skaletsky, H., Brown, L.G., Zaghloul, S., Rock, S., Graves, T., Auger, K., Warren, W.C., Wilson, R.K., and Page, D.C. (2013). Independent specialization of the human and mouse X chromosomes for the male germ line. *Nat Genet* 45, 1083–1087. <https://doi.org/10.1038/ng.2705>.
- 189 9. 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. *Science* 300, 1291–1295. <https://doi.org/10.1126/science.1083022>.
- 190 10. 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 Silencing. *Molecular Cell* 47, 970–979. <https://doi.org/10.1016/j.molcel.2012.07.019>.
- 191 11. 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>.
- 192 12. Magnúsdóttir, E., and Surani, M.A. (2014). How to make a primordial germ cell. *Development* 141, 245–252. <https://doi.org/10.1242/dev.098269>.
- 193 13. Günesdogan, U., Magnúsdóttir, E., and Surani, M.A. (2014). Primordial germ cell specification: A context-dependent cellular differentiation event [corrected]. *Philos Trans R Soc Lond B Biol Sci* 369. <https://doi.org/10.1098/rstb.2013.0543>.

- 207 14. Bardot, E.S., and Hadjantonakis, A.-K. (2020). Mouse gastrulation: Coordination of tissue patterning, specification and
208 diversification of cell fate. *Mechanisms of Development* *163*, 103617. <https://doi.org/10.1016/j.mod.2020.103617>.
- 209 15. Mochizuki, K., Sharif, J., Shirane, K., Uranishi, K., Bogutz, A.B., Janssen, S.M., Suzuki, A., Okuda, A., Koseki, H.,
210 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>.
- 211 16. Samanta, M.K., Gayen, S., Harris, C., Maclary, E., Murata-Nakamura, Y., Malcore, R.M., Porter, R.S., Garay, P.M.,
212 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>.
- 213 17. Koubova, J., Menke, D.B., Zhou, Q., Capel, B., Griswold, M.D., and Page, D.C. (2006). Retinoic acid regulates
214 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>.
- 215 18. Lin, Y., Gill, M.E., Koubova, J., and Page, D.C. (2008). Germ Cell-Intrinsic and -Extrinsic Factors Govern Meiotic
216 Initiation in Mouse Embryos. *Science* *322*, 1685–1687. <https://doi.org/10.1126/science.1166340>.
- 217 19. Endo, T., Mikedis, M.M., Nicholls, P.K., Page, D.C., and De Rooij, D.G. (2019). Retinoic Acid and Germ Cell
218 Development in the Ovary and Testis. *Biomolecules* *9*, 775. <https://doi.org/10.3390/biom9120775>.
- 219 20. Li, H., Liang, Z., Yang, J., Wang, D., Wang, H., Zhu, M., Geng, B., and Xu, E.Y. (2019). DAZL is a master translational
220 regulator of murine spermatogenesis. *Natl Sci Rev* *6*, 455–468. <https://doi.org/10.1093/nsr/nwy163>.
- 221 21. Mikedis, M.M., Fan, Y., Nicholls, P.K., Endo, T., Jackson, E.K., Cobb, S.A., De Rooij, D.G., and Page, D.C. (2020).
DAZL mediates a broad translational program regulating expansion and differentiation of spermatogonial progenitors.
222 *eLife* *9*, e56523. <https://doi.org/10.7554/eLife.56523>.
- 223 22. Gupta, N., Yakhou, L., Albert, J.R., Azogui, A., Ferry, L., Kirsh, O., Miura, F., Battault, S., Yamaguchi, K., Laisné,
224 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>.
- 225 23. 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
226 embryonic stem cells. *Nucleic Acids Research* *51*, 3130–3149. <https://doi.org/10.1093/nar/gkad071>.
- 227 24. 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
228 Histone Methylation. *Cell Reports* *14*, 1000–1009. <https://doi.org/10.1016/j.celrep.2015.12.091>.

229 **Figures and Tables**

230 —

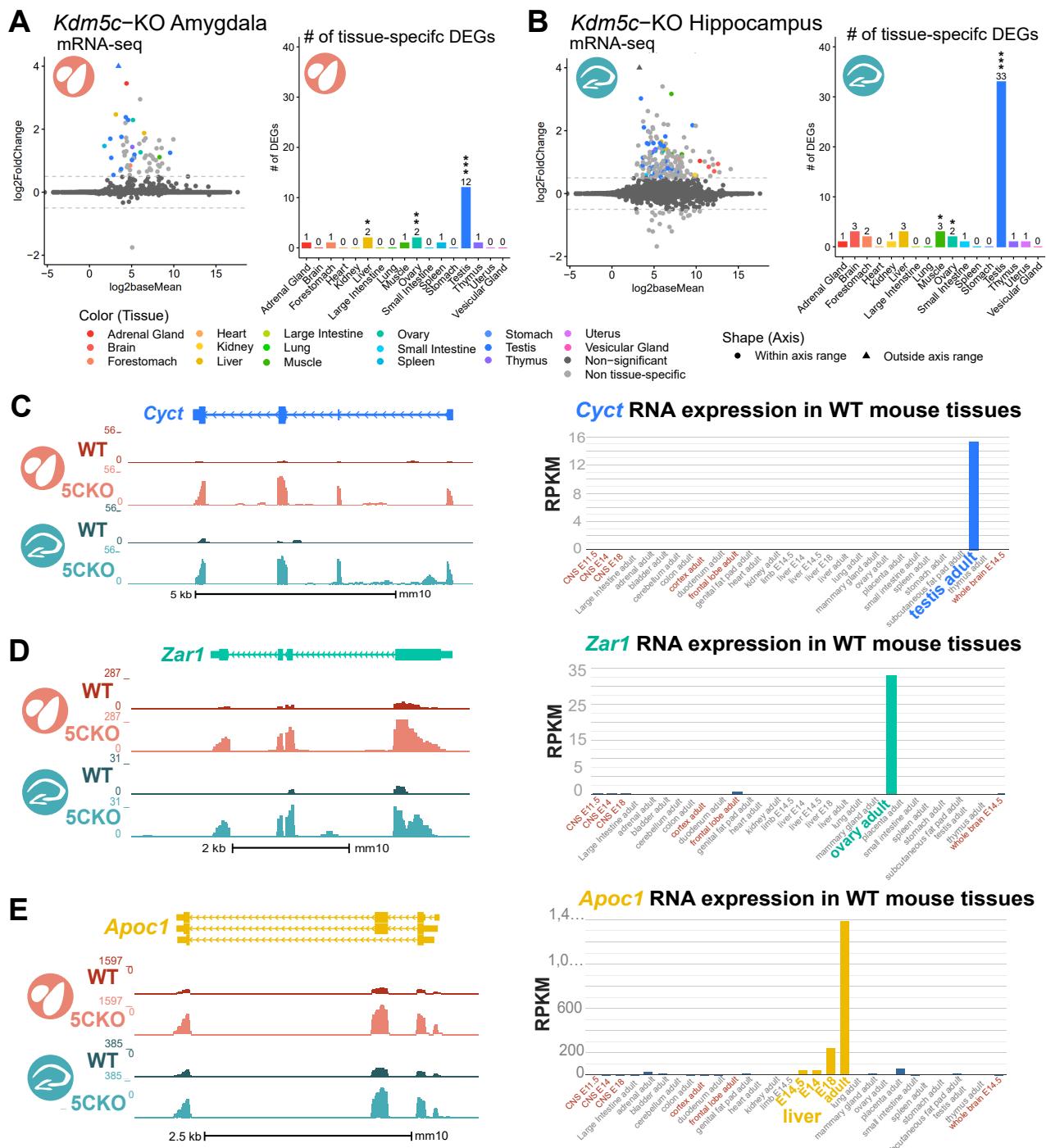


Figure 1: Tissue-enriched genes are misexpressed in the *Kdm5c*-KO brain. **A.** Expression of tissue-enriched genes in the male *Kdm5c*-KO amygdala. Left - MA plot of mRNA-seq. Right - Number of tissue-enriched differentially expressed genes (DEGs). * $p<0.05$, ** $p<0.01$, *** $p<0.001$, Fisher's exact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

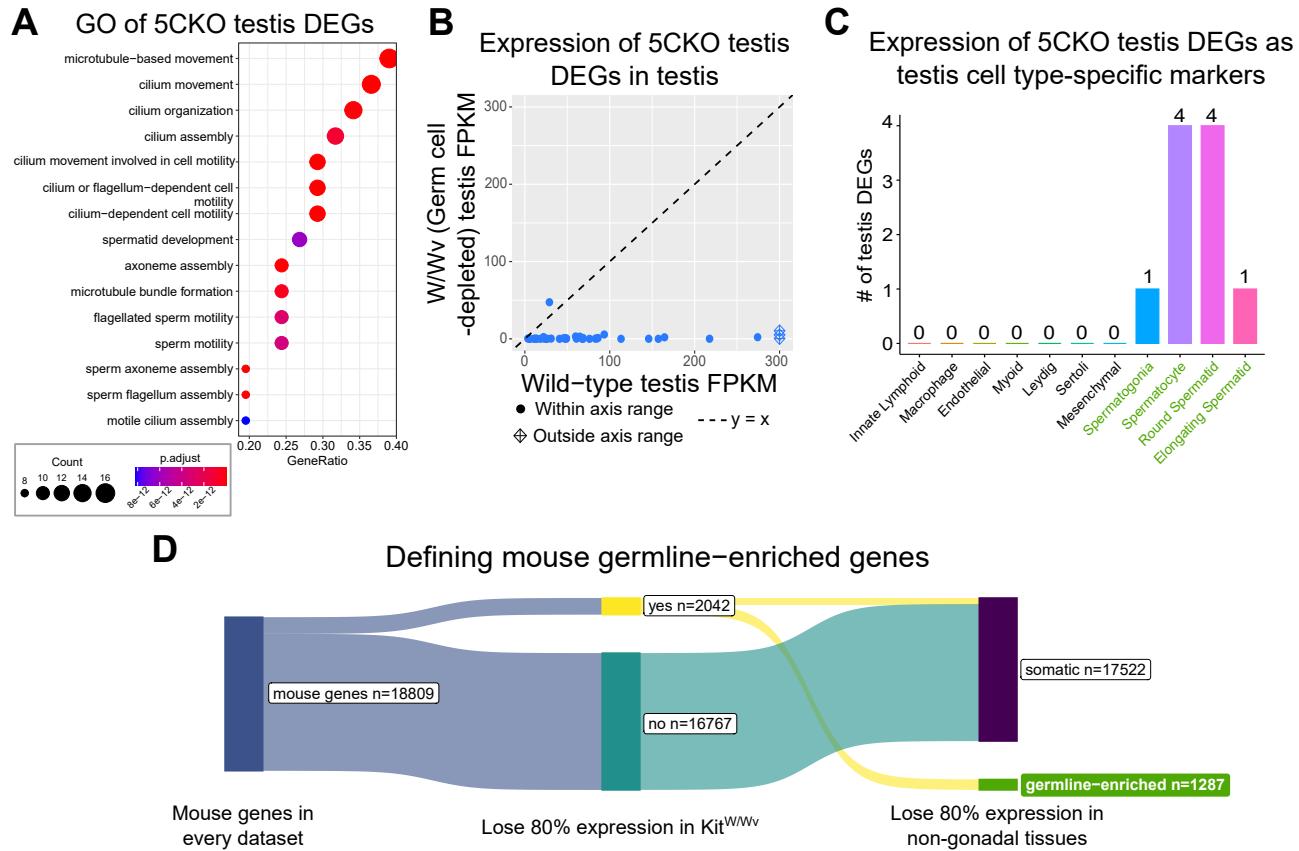


Figure 2: Aberrant transcription of germline genes in the *Kdm5c*-KO in the brain. **A.** enrichPlot gene ontology (GO) of *Kdm5c*-KO amygdala and hippocampus testis-enriched DEGs **B.** Expression of testis DEGs in wild-type (WT) testis versus germ cell-depleted (W/Wv) testis. Expression is in Fragments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

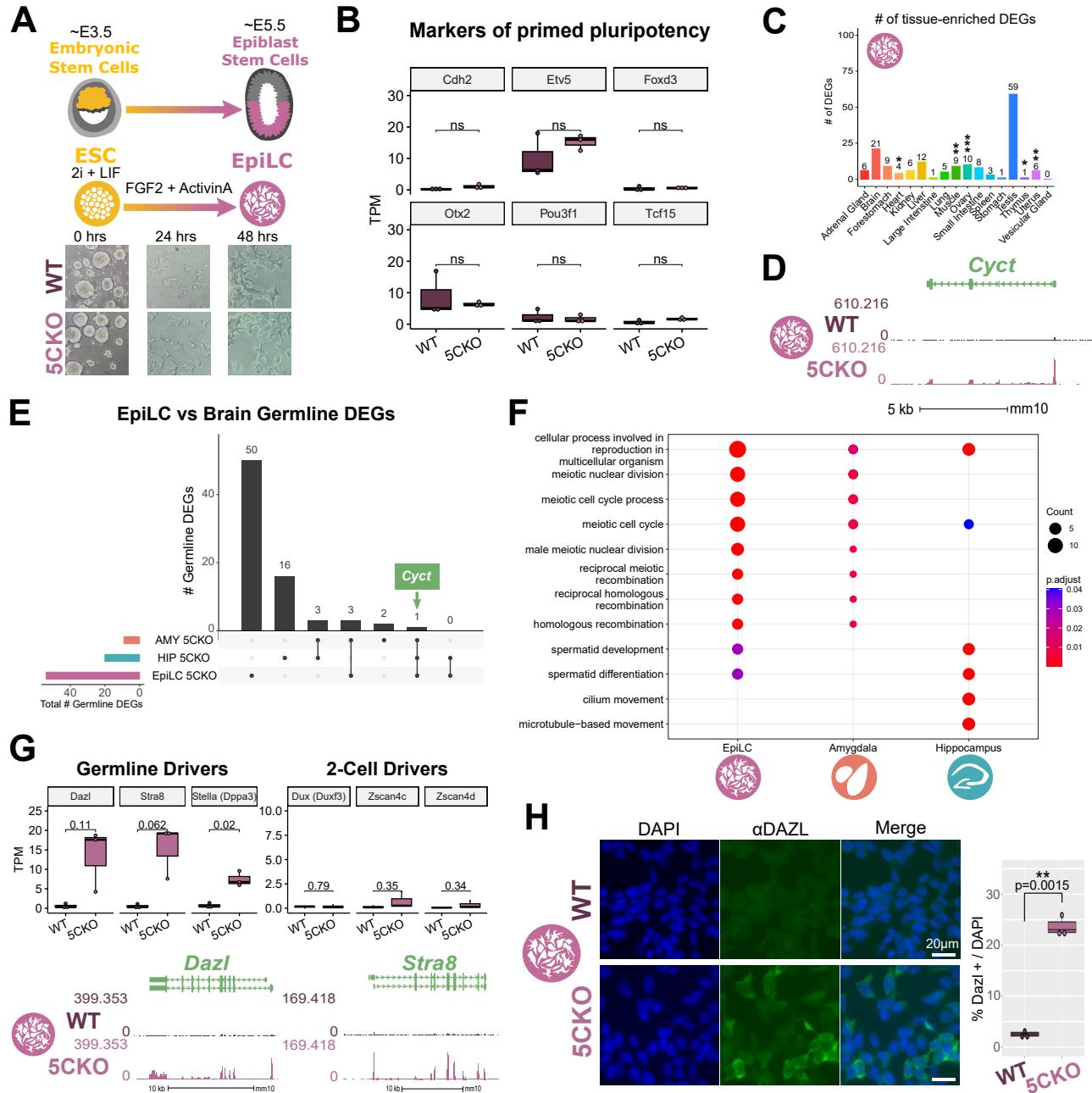


Figure 3: *Kdm5c*-KO epiblast-like cells express master regulators of germline identity

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

231 **Figure outline:**

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

235 **Figure 2: The male *Kdm5c*-KO brain expresses male and female germline-enriched genes** * Gene ontology of
236 testis DEGs in the amygdala and hippocampus - ontologies are germline ontologies * Expression of testis DEGs in germline-
237 depleted testis (this is adult testis data) * scRNAseq of testis - # of testis DEGs that are germline-specific markers * Although
238 far fewer, 5CKO brain also expresses ovary-enriched genes(NCBI and bigwigs of Zar1) * These ovary enriched genes are
239 also germline specific (NCBI/Li tissues are in adult ovary, so it would be best to show they're oocyte-specific in adult ovary.
240 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
241 W/Wv data since oocytes are developed at this point? Or both?) * Defining what is/isn't a germline gene, and which are
242 male/female biased using embryonic W/Wv data

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

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

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

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