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

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

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

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

Results

Tissue-specific genes, including testis genes, are enriched within the *Kdm5c*-KO brain

note: the 5CKO DEGs table we should make a column for if the gene is tissue-specific and which tissue it's in

see if any liver or muscle genes are dysregulated in both tissues, could add as a bigwig

Previous RNA sequencing studies performed in the adult male *Kdm5c* knockout (-KO) hippocampus identified aberrant transcription of genes that are typically only expressed in the testis[1]. Given the high abundance of testis-enriched genes within the mouse transcriptome, it is currently unclear if their misexpression in the *Kdm5c*-KO brain is indicative of random de-repression throughout the genome or instead reflect widespread impairment of tissue identity. To distinguish these possibilities, we globally assessed if tissue-specific genes were significantly enriched in our previous mRNA-seq dataset of the *Kdm5c*-KO amygdala and hippocampus[2]. We classified differentially expressed genes (DEGs) as tissue-specific if they were significantly upregulated in the *Kdm5c*-KO brain compared to wild-type controls (DESeq2[3], log2 fold change > 0.5, q < 0.1) and were previously identified as uniquely enriched within a given mouse tissue[4].

- Based on these criteria, we observed widespread misexpression of tissue-enriched genes within the *Kdm5c*-KO amygdala and hippocampus, with the majority of genes belonging to the testis (Figure 1A-C).
- Even though the testis has the largest number of tissue-specific genes compared to any other tissue (2496 genes), testis genes were significantly enriched for both brain regions (Amygdala p = 1.83×10^{-5}; Hippocampus p = 4.26×10^{-11}, Fisher's Exact).
- This dysregulation of tissue identity is not unique to the testis, as we also observed significant misexpression of ovary, liver, and muscle genes within the *Kdm5c*-KO brain (Figure 1A-E).

Importantly, these tissue-specific genes show little to no expression in the wild-type brain and have no known brain functions, yet our mRNA-seq data indicates they are polyadenylated and spliced into mature transcripts (Figure 1C,G).

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

Male and female germline genes are aberrantly expressed in the male *Kdm5c*-KO brain

Intriguingly, many *Kdm5c*-KO testis-enriched DEGs have functions specific to male germ cells, such as *Cytochrome C, testis (Cyct)*, a (Figure 1C), a testis-specific cytochrome c isoform found in spermatozoa mitochondria and is important for flagellar beating[5,6]. We therefore wondered if the high enrichment of testis genes within the *Kdm5c*-KO brain reflected a failure to demarcate the soma and germline, as the testis contains both germ cells (e.g. spermatogonia) as well as somatic cells (e.g. Leydig cells) that support hormone production and germline functions. Gene ontology analysis of testis DEGs revealed enrichment of germline-relevant ontologies in the *Kdm5c*-KO amygdala and hippocampus, including cellular process involved in reproduction in multicellular organism (GO:0022412), spermatid development (GO:0007286), and sperm axoneme assembly (GO:0007288) (Figure 1A).

To validate if these testis DEGs are truly germline genes, we then compared their expression in a published RNA-seq dataset of wild-type (WT) and germ cell-depleted ($Kit^{W/Wv}$) mouse testes[7]. We found almost all *Kdm5c*-KO testis-enriched DEGs lose their expression with germ cell depletion (Figure 1B). We additionally assessed testis DEG expression in a published single cell RNA-seq dataset that identified cell type-specific markers within the testis[8]. We found that while some testis-enriched DEGs were classified as specific markers for multiple stages of germ cell development - including spermatogonia, spermatocytes, round spermatids, and elongating spermatids, none marked somatic cells (Figure 1C). Altogether, these data suggest testis genes expressed within the *Kdm5c*-KO brain are actually male germ cell genes.

Interestingly, we also observed significant enrichment of ovary genes within the male *Kdm5c*-KO brain (Amygdala p = 0.00574; Hippocampus p = 0.048, Fisher's Exact). Although there are relatively few ovary-enriched DEGs compared to testis-DEGs (Amygdala: 12 testis DEGs, 2 ovary DEGs; Hippocampus: 33 testis DEGs, 2 ovary DEGs), they also have known germline-specific functions. For example, *Zygotic arrest 1 (Zar1)* was recently shown to regulate oocyte translation by sequestering mRNAs[9] (Figure 1D).

- To assess if ovary-enriched DEGs were also germline specific, we assessed their expression in wild-type and germline-depleted ($Kit^{W/Wv}$) female embryos, when female germ cells undergo the initial stages of development and meiosis.
  - We found XYZ.
  - Altogether, these results indicate male *Kdm5c*-KO mice express both male and female germline genes within the amygdala and hippocampus.
- One difficulty in globally characterizing the misexpression of germline genes is a comprehensive list of mouse germline-specific genes is currently lacking.
- To facilitate downstream analyses, we generated a list of male and female germline-enriched genes by evaluating expression in wild-type and germline-depleted ($Kit^{W/Wv}$) male and female mouse embryos from embryonic day 12, 14, and 16.
  - **note: need to figure out why these time points. We have P6 and adult for males but not females so it would be biased towards males if we included the later ages**

We defined genes as germline-enriched if their expression met the following criteria: 1) their expression is greater than 1 FPKM in wild-type germline 2) their expression in any wild-type, non-gonadal tissue[4] does not exceed 20% of maximum expression in wild-type germline, and 3) their expression in the germ cell-depleted germline, for any sex or time point, does not exceed 20% of their maximum expression in the wild-type germline. These criteria yielded 1287 germline-enriched genes. To more comprehensively assess the impact of sex on germline gene misexpression, we additionally categorized if germline-enriched genes displayed biased expression in the wild-type female (XX) or male germline. We defined sex-biased genes as those whose expression in the opposite sex, at any time point, is no greater than 20% of the gene's maximum expression in a given sex. This yielded 65 XX-biased, 1023 XY-biased, and 199 unbiased germline-enriched genes, which is consistent with the testis overall having a more unique transcriptome than the ovary[4].

Discussion

- Papers to read/reference:

- Reconstitution of the Mouse Germ Cell Specification Pathway in Culture by Pluripotent Stem Cells: https://www.cell.com/fulltext/S0092-8674(11)00771-9
- 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/

References

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. 10.1016/j.celrep.2017.09.014.
2. 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. 10.1038/s42003-020-1001-6.
3. 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. 10.1186/s13059-014-0550-8.
4. 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. 10.1038/s41598-017-04520-z.
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. 10.1126/science.193188.
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. 10.1128/MCB.22.15.5554-5562.2002.
7. 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 chromosomes for the male germ line. Nat Genet *45*, 1083–1087. 10.1038/ng.2705.
8. 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. 10.1016/j.devcel.2018.07.025.
9. 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. 10.1126/science.abq4835.

Figures and Tables

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

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

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

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

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

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

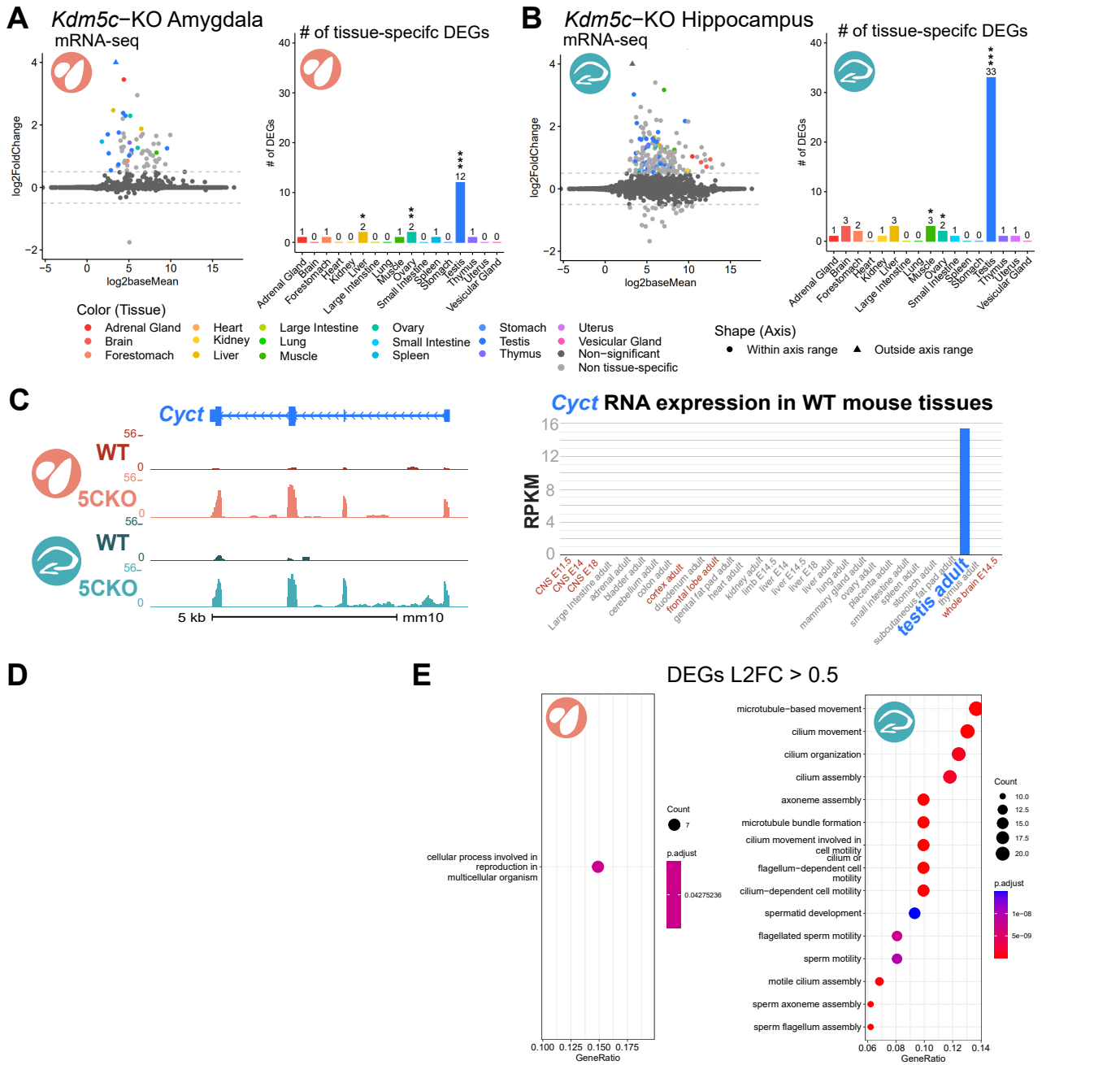

Figure 1: Misexpression of tissue-specific genes in the *Kdm5c*-KO in the brain A. Expression of tissue-enriched genes in the male *Kdm5c*-KO amygdala. Left - MA plot of mRNA-sequencing. Right - The number of tissue-enriched differentially expressed genes (DEGs). * p<0.05, ** p<0.01, *** p<0.001, Fisher's exact test **B.** Expression of tissue-enriched genes in the male *Kdm5c*-KO hippocampus. **C.** Left - Bigwigs of an example aberrantly expressed testis-enriched gene, *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 - Bigwigs of an example ovary-enriched germline 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.