Pathway-Level Transcriptomic Dysregulation in AKAP11 Mutant Mouse Prefrontal Cortex

Author: Amalya S. Murrill

Introduction

This project seeks to replicate the findings from Song et al. (2025), "Elevated synaptic PKA activity and abnormal striatal dopamine signaling in Akap11 mutant mice, a genetic model of schizophrenia and bipolar disorder" (https://doi.org/10.1038/s41467-025-66504-2). The loss-of-function in AKAP11, a protein kinase A (PKA)-binding protein, produces transcriptomic dysregulation in the brain and thus phenotypic changes characteristic of psychiatric disorders such as bipolar disorder and schizophrenia, as prior studies have shown. Given the clinical urgency of understanding the biological mechanisms behind these debilitating conditions, of which very little is known, the authors take a multi-omic, brain-wide approach to investigating the effect of AKAP11 loss-of-function, using Akap11 mutant mice as a model organism. The breadth of this study is profound, employing multiple methods across six regions of the brain, performing proteomic (e.g. coIP-MS), bulk RNA- and scRNA-seq, and even behavioral experiments for comparison to clinical presentation in humans. The experiments were also stratified by age at 4 weeks-, 12 weeks-, and 28 weeks-old to assess the degree to which neural development plays a part. In this analysis, a working bulk RNA-seq pipeline and pathway enrichment analysis for the 12 week prefrontal cortex group across three genotypes, Akap11 (-/-, KO), Akap11 (+/-, HET), and Akap11 (+/+, WT), were developed to reproduce and compare the findings found in Song. et al.

Methods

Bulk RNA-seq data was found on Gene Expression Omnibus (GEO) with the ID number GSE306677 as listed in the publication. For instructive purposes, only the extracted prefrontal cortices of 12 week-old mice were analyzed. However, all five biological replicates for each genotype (KO, HET, and WT) were selected for a total of fifteen paired-end samples. A pipeline consisting of pre-processing and quality control was performed using Nextflow (v25.04.6). Singularity container images provided by the pipeline_containers repository (BF528, n.d.) were implemented for the majority of processes, while a select few came from the BioContainers framework (da Veiga Leprevost et al., 2017). To extract the paired-end fastq files from GEO SRA accession numbers, a process employed sra-tools(v3.1.1) with the --split-files parameter. Fastq file names were renamed by sample name rather than SRA accession number (i.e. 12wk-

PFC-HET3) using an argparse-based python script (pandas v2.2.3) for downstream sample identification and overall clarity. Two further processes downloaded the GTF annotation and reference genome files (both M38 primary assembly) with the wget command using default parameters. The pre-processing pipeline began with FASTQC (v0.12.1) on a channel of the paired reads using default parameters. The genome was indexed via STAR (v2.7.11b) using default parameters and then aligned to generate unsorted BAM files and output logs with the following highlighted parameters: --outSAMtype BAM Unsorted and 2> ${name}.Log.final.out. These STAR ALIGN logs and FASTQC zip files were fed into a MULTIQC process (v1.25) using standard parameters to further assess the quality of the alignment and raw reads. Finally, a counts matrix using VERSE (v0.1.5) under default parameters was constructed via the aligned reads and GTF annotation file in order to quantify gene expression as a start to differential expression analysis.

DE analysis was performed in RStudio using BiocManager (v1.30.26) to take advantage of libraries such as DESeq2 (v1.46.0), msigdbr (v7.5.1), and fgsea (v1.32.4). The counts matrix was pre-filtered by keeping genes with at least 10 counts in at least 3 samples with the intention of removing noise from the matrix not likely to be biologically relevant. After constructing the DESeq2 object, quality control plots like MA plots assessed the level of systemic bias and quickly visualized the distribution of differentially expression genes (DEGs). Principal Components Analysis (PCA) was developed via the variance-stabilizing transformation (vst) method with its default (blind = TRUE). The subsequent plot was then constructed. Finally, FGSEA was performed, focusing on KO vs WT and HET vs WT for the analysis. Ranked lists were generated by the Wald statistic. As is in the literature, gene sets within the MsigDB pathways were collected from Gene Ontology: Biological Processes and Gene Ontology: Cellular Components (GO BP and GO CC). Genes of interest were also sourced from the paper and used in place of GWAS HPO (human phenotype) or DO (disease), as nothing of significance was found. Multilevel FGSEA was run on each set of pathways with comparisons KO vs WT and HET vs WT. A padj threshold of 0.25 was used, along with a minSize of 1 for GWAS to increase the range in order to hopefully visualize enrichment of these pathways in the data.

Quality Control Evaluation

MultiQC

With the collection of FastQC zip files and STAR logs, the findings of MultiQC presented an overall positive summary of the quality of the raw reads and alignments. General statistics showed a high alignment rate, no lower than 90.4% across all samples, with the maximum being 97.0%. This is indicative of a successful mapping, demonstrating quality reads and a strategic selection of a concurrent reference genome. The read

depth was consistent with basic gene expression studies at an average of 56.65 million read across these 15 samples. The mismatch and indel rates were each low to zero, with 0.1% mismatch and 0% indels. PHRED scores remained high (~ 35-36) and there was a relatively consistent normal distribution of GC content percentage across all samples. The sequence duplication level at approximately 25% of the library having duplicates greater than 10 was flagged by MultiQC; this, however, is expected for RNA-seq, given the abundance of transcripts that are sequenced multiple times. There were two overrepresented sequences: a string of Gs and a sequence from the genome of SARS-CoV-2 (confirmed by BLAST search). The former likely represents the poly-G tail found as a technical artifact of sequencing. It could be recommended to trim the reads using Trimmomatic. However, this tail was 0.0042% of all reads - understandably ignored by the authors. Based on these metrics, one can confidently continue with downstream analysis.

MA Plots

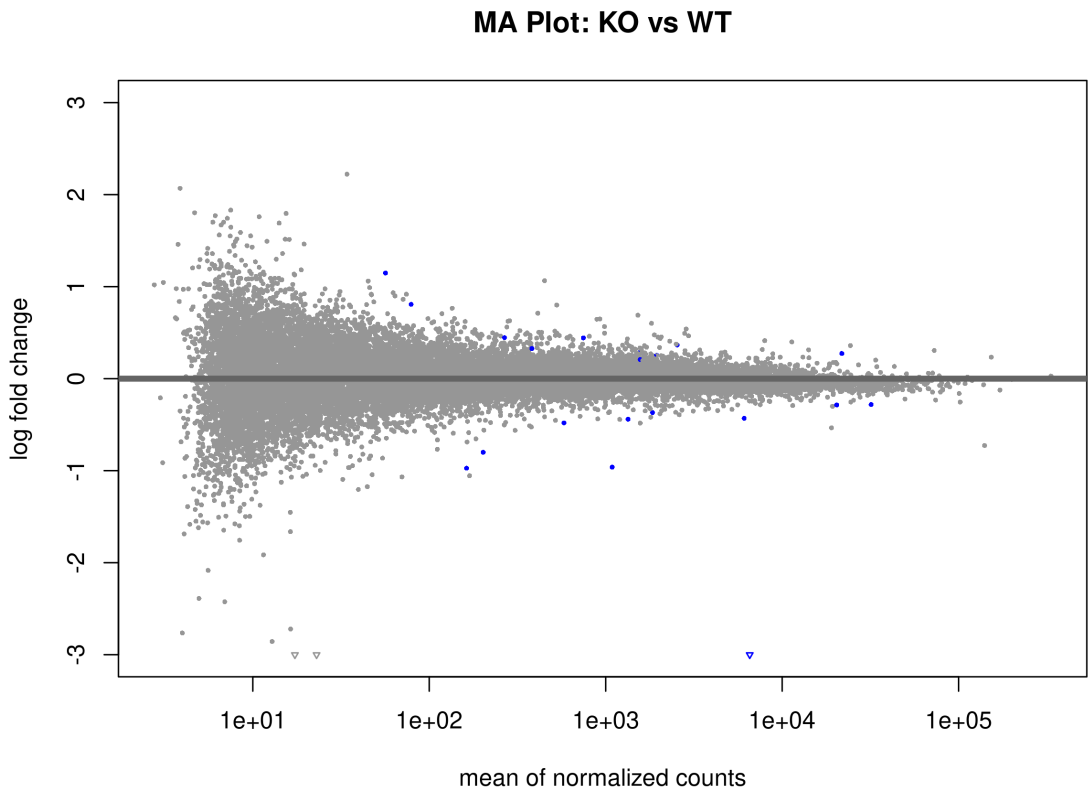

Figure 1. MA plot of KO versus WT samples. Blue dots represent statistically significant differentially expressed genes. Triangles show values for genes beyond the scope of the plot.

An MA plot was used to visualize the relationship between mean expression and $\log_2$ fold change in the KO vs WT samples. As expected from RNA-seq differential expression data, there is an abundance of stably expressed genes (grey), with a few scattered

genes presenting with higher expression levels. Low average expression showed greater variability in fold change (left), while highly expressed genes clustered tightly around zero (right). As a quality metic, there is no evidence of expression bias, exhibiting appropriate normalization and stable expression across samples.

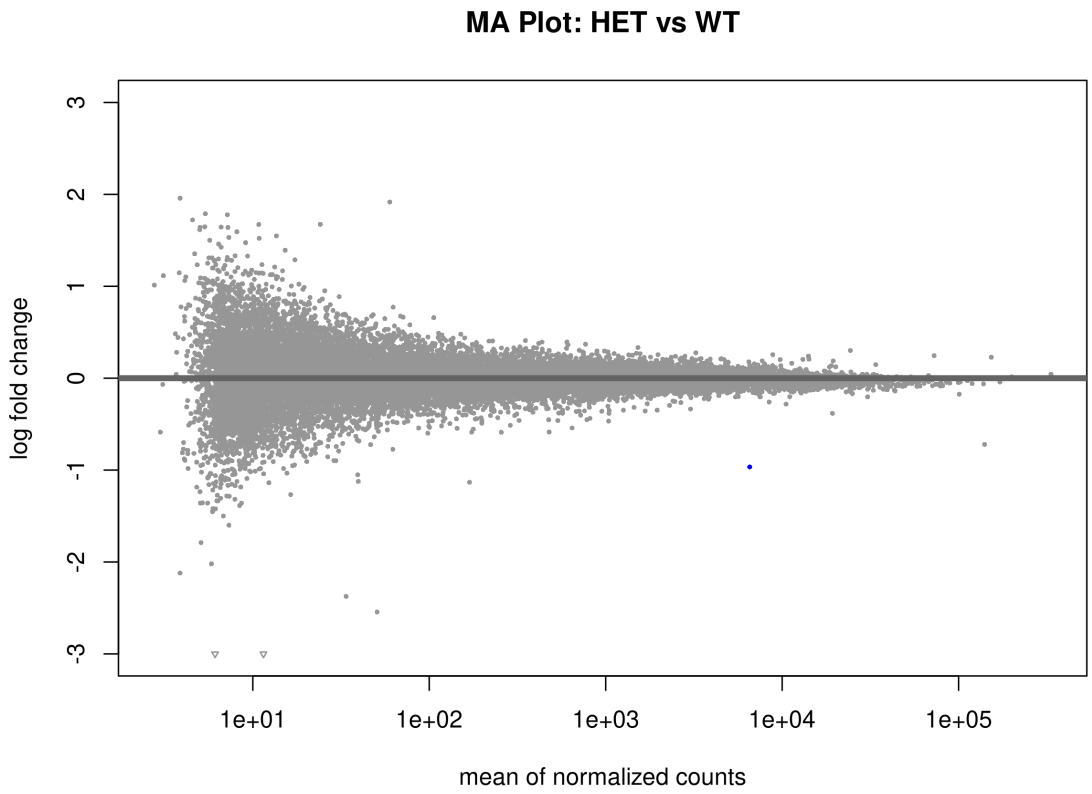

Figure 2. MA plot of HET versus WT samples. Blue dots represent statistically significant differentially expressed genes. Triangles show values for genes beyond the scope of the plot.

The same MA plot structure was used to visualize the relationship between mean expression and $\log_2$ fold change in the HET vs WT samples. Similar clustering is seen as in the KO vs WT case. However, there is only one significant DEG shown. Regardless, quality-wise, this too indicates no bias and correct normalization.

Summarized, both MA plots show no cause for concern, yet the results insist upon performing pathway-based differential expression analysis rather than gene-wise analysis, as will be shown later.

PCA Plot

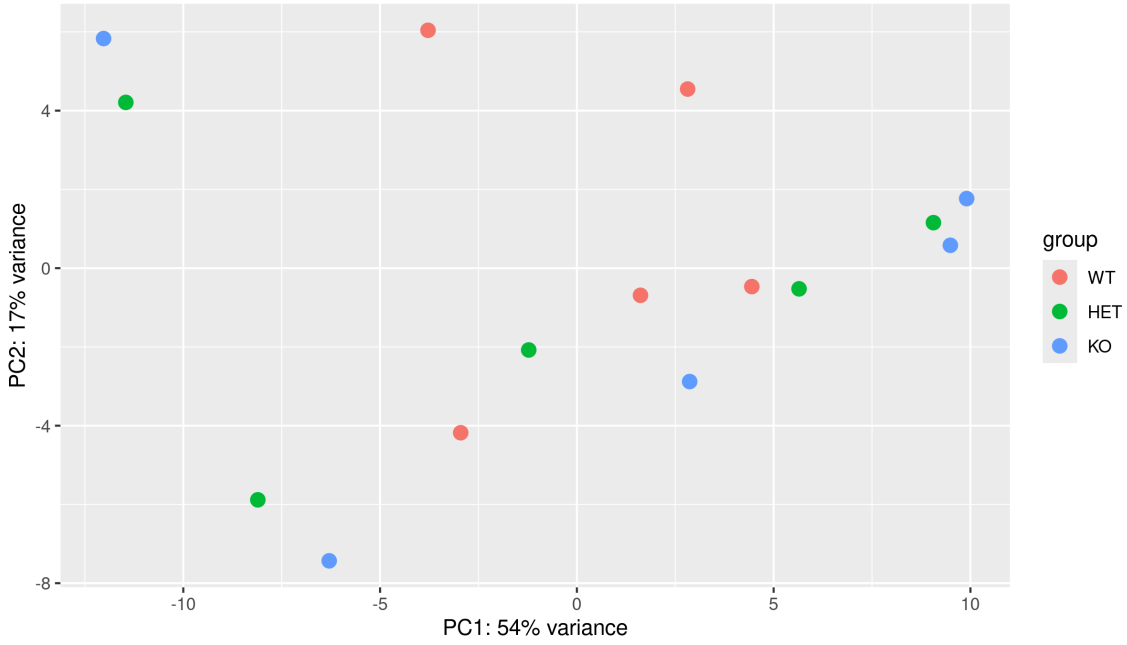

Figure 3. Principal Component Analysis (PCA) plot across three genotypes.

PC1 (54% variance) shows moderate separation in which the KO samples trend right. The WT samples exhibit more spread but rather, trend left. The presence of batch effects seems unlikely, as the samples are not clustering by litter and show no clear pattern. As cited by Song et. al, "...heterozygous mice were crossed with each other to produce WT (WT, +/+), heterozygous (Het, +/-) and knockout (KO, -/-) littermates used for experiments." Thus, the scheme is most likely due to biological variation between individual mice.

Enrichment Analysis

Figure 4. Dot plot of GSEA results from MsigDB-selected pathways, a replication of Figure 6c from Song et al. Dots are sized based on padj and colored by regulation direction (red shows upregulation, blue shows downregulation).

Discussion

The decisions made throughout the workflow, whether to follow the authors exactly or to deviate from some of the methods, ultimately led to the general replication of their

findings, with minute differences possibly due to changes in unmentioned filtering strategy,alignment versus pseudo-alignment (STAR vs Salmon), or simply working with a subset as opposed to all of the samples at once. No removal of samples were recorded in the PFC set, so an assumption will be made that the reads and alignments passed the common quality checks as aforementioned. Ultimately, the results of this pathway enrichment analysis generally confirm the major findings as is in Song et al., despite these deviations. Here, GO: Biological Processes and GO: Cellular Components were highlighted as the main pathways of focus, given that the GWAS gene list for HPO and DO did not appear as significant.

In dissecting Figure 4, replicated from Figure 6c in Song et al., the major points of comparison between the two analyses lie in the difference in what was identified as significant. Despite this, many pathways in GO: CC in this analysis matched the author's results. Downregulation of "postsynaptic membrane" in both HET and KO, "gabaergic synapse" in HET, and "neuron to neuron synapse" in KO can be grouped within the overall downregulation of "Synapse" in their 12wk PFC samples. Interestingly, this analysis caught a specific enrichment: the upregulation of "intrinsic component of postsynaptic membrane" in KO, not explicitly shown in their "Synapse" pathway enrichment. This demonstrates that occasionally the expression of one cellular component may not be homogeneous in nature. There are several upregulated cell cycle/division related pathways; however they were likely filtered out as less neurobiologically relevant.

The GO: BP matches retained synaptic influence: upregulated "receptor localization to synapse" and "protein localization to synapse" in KO. The former could relate more specifically to their "Neuropeptide Signaling Pathway" given receptor positioning/signaling. Most accurate is the downregulation of "regulation of homotypic cell-cell adhesion" in HET, also found in their "Cell Adhesion" across 12wk PFC samples.

As shown by the MA plots above, there were few DEGs in the KO and HET cases, with HET showing only one distinct DEG. However, an interesting finding is the evidence of greater and more diverse pathway enrichment in HET vs WT than in KO vs WT, despite having fewer DEGs. The supplementary article Farhangdoost et al. (2025), "Transcriptomic and epigenomic consequences of heterozygous loss-of-function mutations in AKAP11, a shared risk gene for bipolar disorder and schizophrenia", specifically analyzes the effects of haploinsufficiency in Akap11 mutant mice, also finding strong evidence of dysregulation of genes involved in cell adhesion. In Song et al., they see upregulation of cell adhesion in the PFC at 4 weeks, and downregulation at 12 weeks. Remarkably, Farhangdoost et al. sees the same upregulation in their 1 month- (4 week-) old mouse model. Initial upregulation of cell adhesion pathways most likely represents the compensation for the lack of AKAP11's role in synaptic stabilization. AKAP11's localization of PKA for phosphorylation of cytoskeletal and adhesion proteins is integral to the maintenance of synaptic integrity. Thus, loss of AKAP11 disrupts this

regulation, leading to downregulation of synaptic components while simultaneously increasing cell adhesion molecule expression. This demonstrates the need for heterozygous test samples despite lower DEGs relative to WT, as HET vs WT pathway enrichment analysis can help visualize the biological consequences of the regulatory balancing act seen in haploinsufficient models.

This analysis successfully replicated the main transcriptomic findings of Song et al., demonstrating that AKAP11 loss-of-function propagates synaptic dysfunction and compensatory cellular responses. The pathway-level approach proved more adept in detecting coordinated transcriptional changes that gene-level analysis would obscure, highlighting the value of GSEA for understanding the intricacies of psychiatric disease mechanisms.