**Reads processing, QC and alignment**

Paired-end fastq reads files for each sample were aligned to mouse genome GRCm39 RefSeq assembly GCF\_000001635.27 (Release Feb, 2024). We used STAR1 in twopassMode to align the reads to the reference genome and sort them by coordinate. Each sample was evaluated according to a variety of both pre- and post-alignment quality control measures with FastQC2 0.12.1, Picard3 3.2.0., RSeQC4 5.0.4, MultiQC5 1.25. Reads were trimmed with Trimmomatic6 0.39. PCR duplicatess were marked with Picard tools, deduplication effect was further tested downstream. Aligned read counts were calculated by featureCounts7 from the subread 2.0.2 release. The pipeline for reads processing is stored as a [GitHub repository](https://github.com/tony-zhelonkin/RNAseq_pipelineDock)

**Downstream Analysis**

Bulk RNAseq counts data were then further analyzed in R8 4.4.1, Rstudio9 2024.9.0.375. Differential expression analysis was performed using the edgeR10,11 4.2.1 and limma-voom12 3.60.6. Picard deduplication didn`t strongly affect the gene rankings (Spearman r = 0.991), but it introduced ties in the preranked genes list statistics during gene set enrichment analysis (52.54%). We decided to stick to un-deduplicated read counts for the entire analysis, as is considered the best practice practice in bulk RNAseq analysis13, at least in cases where RNA-seq library was prepared without unique molecular identifiers. 14

Counts were filtered for low count genes and library sizes were normalised with Trimmed Mean of M-values (TMM) method15 via edgeR`s *`filterByExpr`* and *`normLibSizes`* functions. Linear model fit for the comparison of Mecr-KO samples to WT was performed with edgeR`s *`voomLmFit`* wrapper function of the limma`s *`voomWithQualityWeights`* to combine observational-level precision weights with sample-specific quality weights and increase the power of the analysis16. Volcano plots were visualized with EnhancedVolcano17 1.22.0 and interactive volcano plots were customised with ggplot218 3.5.1 plotly19 4.10.4 and shiny20 1.9.1.

**Gene set enrichment analysis**

Gene set enrichment analysis21 was performed with clusterProfiler22–24 4.12.6 using `fgsea`25 as the calculation method with no boundaries for p-value estimation. For the GSEA we included all the genes into the background ranked gene list after excluding the low counts genes with *`filterByExpr`* thus leaving the genes that have any chance to be assessed as differentially expressed26,27*.* The background gene list was ranked by the moderated t-statistic. We tested the background gene list against MSigDB`s28 via msigdbr 7.5.1 Hallmark pathways29, C5 Gene Ontology molecular functions, C2 Canonical Pathways30,31, KEGG32 pathways (excluding disease-related gene sets) and also REACTOME33 database with ReactomePA34 1.48.0 library. Pathways with a qvalue < 0.05 were deemed to be significant.

All raw and processed data, as well as scripts for analysis, are available at

*[link to GitHub Repo]*

1. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29(1):15-21. doi:10.1093/bioinformatics/bts635

2. Wingett SW, Andrews S. FastQ Screen: A tool for multi-genome mapping and quality control. Published online September 17, 2018. doi:10.12688/f1000research.15931.2

3. Picard toolkit. Published online 2019. https://broadinstitute.github.io/picard/

4. Wang L, Wang S, Li W. RSeQC: quality control of RNA-seq experiments. *Bioinformatics*. 2012;28(16):2184-2185. doi:10.1093/bioinformatics/bts356

5. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*. 2016;32(19):3047-3048. doi:10.1093/bioinformatics/btw354

6. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114-2120. doi:10.1093/bioinformatics/btu170

7. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014;30(7):923-930. doi:10.1093/bioinformatics/btt656

8. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing; 2024. https://www.R-project.org/

9. Posit team. *RStudio: Integrated Development Environment for R*. Posit Software, PBC; 2024. http://www.posit.co/

10. Chen Y, Chen L, Lun ATL, Baldoni PL, Smyth GK. edgeR 4.0: powerful differential analysis of sequencing data with expanded functionality and improved support for small counts and larger datasets. Published online January 24, 2024:2024.01.21.576131. doi:10.1101/2024.01.21.576131

11. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010;26(1):139-140. doi:10.1093/bioinformatics/btp616

12. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43(7):e47. doi:10.1093/nar/gkv007

13. Parekh S, Ziegenhain C, Vieth B, Enard W, Hellmann I. The impact of amplification on differential expression analyses by RNA-seq. *Sci Rep*. 2016;6(1):25533. doi:10.1038/srep25533

14. Zajac N, Vlachos IS, Sajibu S, et al. The impact of PCR duplication on RNAseq data generated using NovaSeq 6000, NovaSeq X, AVITI and G4 sequencers. Published online December 13, 2023:2023.12.12.571280. doi:10.1101/2023.12.12.571280

15. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology*. 2010;11(3):R25. doi:10.1186/gb-2010-11-3-r25

16. Liu R, Holik AZ, Su S, et al. Why weight? Modelling sample and observational level variability improves power in RNA-seq analyses. *Nucleic Acids Research*. 2015;43(15):e97. doi:10.1093/nar/gkv412

17. Blighe K, Rana S, Lewis M. *EnhancedVolcano: Publication-Ready Volcano Plots with Enhanced Colouring and Labeling*.; 2024. doi:10.18129/B9.bioc.EnhancedVolcano

18. Wickham H. *Ggplot2: Elegant Graphics for Data Analysis*. Springer International Publishing; 2016. doi:10.1007/978-3-319-24277-4

19. Carson Sievert. *Interactive Web-Based Data Visualization with R, Plotly, and Shiny*. Chapman & Hall/CRC; 2020.

20. Chang W, Cheng J, Allaire J, et al. *Shiny: Web Application Framework for R*.; 2024. https://CRAN.R-project.org/package=shiny

21. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles | PNAS. Accessed October 18, 2024. https://www.pnas.org/doi/10.1073/pnas.0506580102

22. Wu T, Hu E, Xu S, et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation*. 2021;2(3). doi:10.1016/j.xinn.2021.100141

23. Xu S, Hu E, Cai Y, et al. Using clusterProfiler to characterize multiomics data. *Nat Protoc*. Published online July 17, 2024:1-29. doi:10.1038/s41596-024-01020-z

24. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS: A Journal of Integrative Biology*. 2012;16(5):284-287. doi:10.1089/omi.2011.0118

25. Korotkevich G, Sukhov V, Budin N, Shpak B, Artyomov MN, Sergushichev A. Fast gene set enrichment analysis. Published online February 1, 2021:060012. doi:10.1101/060012

26. Wijesooriya K, Jadaan SA, Perera KL, Kaur T, Ziemann M. Urgent need for consistent standards in functional enrichment analysis. *PLOS Computational Biology*. 2022;18(3):e1009935. doi:10.1371/journal.pcbi.1009935

27. Timmons JA, Szkop KJ, Gallagher IJ. Multiple sources of bias confound functional enrichment analysis of global -omics data. *Genome Biology*. 2015;16(1):186. doi:10.1186/s13059-015-0761-7

28. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. *Bioinformatics*. 2011;27(12):1739-1740. doi:10.1093/bioinformatics/btr260

29. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database Hallmark Gene Set Collection. *cels*. 2015;1(6):417-425. doi:10.1016/j.cels.2015.12.004

30. Ashburner M, Ball CA, Blake JA, et al. Gene Ontology: tool for the unification of biology. *Nat Genet*. 2000;25(1):25-29. doi:10.1038/75556

31. The Gene Ontology Consortium, Aleksander SA, Balhoff J, et al. The Gene Ontology knowledgebase in 2023. *Genetics*. 2023;224(1):iyad031. doi:10.1093/genetics/iyad031

32. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*. 2000;28(1):27-30. doi:10.1093/nar/28.1.27

33. Croft D, O’Kelly G, Wu G, et al. Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Research*. 2011;39(suppl\_1):D691-D697. doi:10.1093/nar/gkq1018

34. Yu G, He QY. ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. *Mol BioSyst*. 2016;12(2):477-479. doi:10.1039/C5MB00663E