**Processing of RNA data and differential expression analysis:**

RNA sequencing data is processed using trimmomatic (v0.39.0) with the following options: adapter removal, leading and trailing base with low quality (3) removal, read trimming when average per base quality is below 15, and removal of reads that are shorter than 36 bases [1]. RNA data are then aligned and quantified using Salmon (v) with automatic library type option (-l A), and with selective alignment for better sensitivity and specificity of mapping (--validateMappings) [2]. Then, custom TPM matrix is made using custom awk commands to prepare for DESeq2 (v1.46.0). RNA processing step can be found as RNAseq.sh script on github. After processing the RNA data, DESeq2 is used to normalize gene expressions [3]. All samples are separated by cohort and differential expressions are analyzed by survival status using DESeq2. DESeq2 step can be found on github as deseq.r.

**Gene set enrichment analysis (GSEA for differentially expressed genes (DE genes):**

Gene data and annotations are extracted using AnnotationHub (v3.14.0) [4], biomaRt (v2.62.0) [5,6], BSgenome.Hsapiens.UCSC.hg38 (v1.4.5) [7], and EnsDb.Hsapiens.v86 (v2.99.0) [8] packages in R [9]. A list of all main-categories and sub-categories from MSigDB is also created using msigdbr (v7.5.1) [10] package in R. Genes are then prepared for GSEA with column names being gene names and log2FoldChange being used instead of gene expression. GSEA is then done using fgsea (v1.32.2) [11] package in R and the MSigDB list with minimum and maximum size of pathway being 0 and 8000 respectively.

**Methylation data processing, beta value normalization, and LUMP score calculation using RnBeads:**

Methylation data processing is done using RnBeads (v2.24.0) [12,13] package in R to obtain immune cell content (LUMP), and normalized beta value (methylation value) for downstream analysis. RnBeads script can be found on our github with the following options: filtering.low.coverage.masking = TRUE, filtering.greedycut = FALSE, filtering.missing.value.quantile = 0.5, filtering.high.coverage.outliers = TRUE, inference=TRUE, and assembly = “hg19”. Produced normalized beta matrix has chromosome position of hg19 converted to hg38 using UCSC genome browser LiftOver tool [14].

**Immune cell type inference and immune cell content testing:**

Normalized beta value (methylation values) matrixes (separated by cohort A and B) are also used to infer the percentage of each immune cell types using MethylResolver [15] package in R (options : alpha = 0.5, absolute = FALSE). Then, ANOVA is used to determine the statistical significance of each immune cell percentage per cohort by survival status.

**Identification of differentially methylated CpG (DM CpG) with Mann-Whitney U test (Wilcoxon test):**

Samples are separated by cohort. Mann-Whitney U test (wilcox.test from stats package in R) is done for every CpG by survival status to identify CpG with differentially methylated status between Over and Under survival status. A custom script is also used to further filter DM CpGs by their distance (set at +- 5kb) to DE genes.

1. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30: 2114–2120.

2. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods. 2017;14: 417–419.

3. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15: 550.

4. Morgan M, Shepherd L. AnnotationHub: Client to access AnnotationHub resources. 2024.

5. Durinck S, Spellman PT, Birney E, Huber W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat Protoc. 2009;4: 1184–1191.

6. Durinck S, Moreau Y, Kasprzyk A, Davis S, De Moor B, Brazma A, et al. BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. Bioinformatics. 2005;21: 3439–3440.

7. Team T. BSgenome.Hsapiens.UCSC.hg38: Full genomic sequences for Homo sapiens (UCSC genome hg38). 2023.

8. Rainer J. EnsDb.Hsapiens.v86. Bioconductor; 2017. doi:10.18129/B9.BIOC.ENSDB.HSAPIENS.V86

9. The R project for statistical computing. [cited 15 Mar 2025]. Available: https://www.R-project.org/.

10. [cited 15 Mar 2025]. Available: https://davislaboratory.github.io/msigdb.

11. Korotkevich G, Sukhov V, Budin N, Shpak B, Artyomov MN, Sergushichev A. Fast gene set enrichment analysis. bioRxiv. bioRxiv; 2016. p. 060012. doi:10.1101/060012

12. [cited 15 Mar 2025]. Available: https://rnbeads.org.

13. [cited 15 Mar 2025]. Available: https://rnbeads.org.

14. Hinrichs AS, Karolchik D, Baertsch R, Barber GP, Bejerano G, Clawson H, et al. The UCSC Genome Browser Database: update 2006. Nucleic Acids Res. 2006;34: D590-8.

15. Arneson D, Yang X, Wang K. MethylResolver-a method for deconvoluting bulk DNA methylation profiles into known and unknown cell contents. Commun Biol. 2020;3: 422.