SUPPLEMENTARY METHODS

DNA methylation profiling and data analysis

Genome-wide methylation profiling was performed using Illumina's Infinium MethylationEPIC BeadChip array. The obtained IDAT files were imported into R with the champ.load function from the ChAMP (v2.34.0) package [1] with method set as "minfi" and other parameters set to default. The beta values were normalized using BMIQ normalization.

Total RNA-seq and data analysis

Transcriptome profiling was performed using total RNA sequencing. Sequencing libraries were prepared using Ovation SoLo RNA-seq Library Preparation kit (Tecan Genomics, Redwood city, CA) as per the manufacturares instruction manual (library type: fr-secondstrand). Reads were sequenced on paired-end (PE) 150 bp mode on NovaSeq 6000 (Illumina, San Diego, CA). The obtained reads were assessed for quality using FastQC and trimmed for adapter sequences using Trimmomatic (v0.39) in PE mode and 'MINLEN' parameter set to 50. Only the good quality reads were considered for alignment to the human reference genome (hg19) using STAR (v2.7.4a) [2] with 'outFilterMultimapNmax' to 1. UMI-Tools [3] utility was employed in PE mode to remove the duplicate reads. The filtered reads were then assigned to genes using 'featureCount' function from Subread package (v2.0.1) [4].

Bioinformatic and Statistical analyses

All bioinformatic analyses were performed using the R statistical software (version 4.4.3) and Bioconductor (version 3.19.1). All results were visualised with ggplot2 (v3.5.1) and pheatmap

(v1.0.12) R packages. Pair effects were controlled across all omics assays using mixed- or fixed-effect models as needed. For methylation array data, intra-pair correlation was calculated using 'duplicateCorrelation', and pairs were modeled as random effects via the block parameter in limma. For RNA-seq data, pair effects were modeled as fixed effects in the design matrix of the DESeq2 analysis. The dataset was balanced for gender, and all samples were of similar age; therefore, no additional correction for these variables was required. P-values were adjusted wherever relevant using Benjamini-Hochberg method and values below 0.05 were considered significant unless mentioned otherwise. Euclidian distance and 'Ward-D2' algorithm were used in all clustering analysis.

Differential methylation analysis was performed on M-values of tumours and controls using Limma (v3.60.6) package [5]. Geneset enrichment analysis was performed using rGREAT (v2.6.0) package [6]. The enriched terms were filtered for an adjusted hypergeometric p-value below 0.05. PERMANOVA (Permutational Multivariate Analysis of Variance) tests were performed using the adonis2 function from the vegan R package (v2.6-10). The significance of group separation was evaluated using 1 000 permutations. To confirm that the observed difference was not driven by unequal within-group dispersion, beta-dispersion was calculated, followed by ANOVA using the stats package (v4.4.3) in R to test for homogeneity of variances across groups.

The expression data were processed using CPM filtering implemented in NOIseq (v2.38.0) [7] with default parameters. The filtered data was normalised using upperquartile normalisation from the EDAseq package (v2.28.0) [8]. Differential expression analysis was performed using the DESeq2 R package (v1.44.0) [9]. KEGG pathway over-representation analysis was performed using gprofiler2 package (v0.2.3) [10]. Ordinal and quantile regressions were performed using MASS (v7.3-64) and quantreg (v5.99) R packages. All visualizations were based on z-scores or log transformed values.

Spearman correlation was computed between gene expression and mean promoter (5'UTR, TSS1500, TSS200, 1stExon) methylation levels, using the correlation R package (v0.8.6) and correlations were deemed significant if the Spearman coefficient exceeded 0.45 with an adjusted p-value < 0.05.

Non-negative Matrix Factorization (NMF) was performed using NMF R package (v0.28) [11]. The optimal number of clusters was determined using cophenetic, dispersion and silhouette statistics. Nearest Template Prediction (NTP) was used to assign subtypes in the study cohort based on correlation with expression signatures defined using Kim's method in 'extractFeatures' function. Overlap between expression subtypes and clinical information was evaluated using Fisher's exact test. To evaluate the tumour microenvironment, the ESTIMATE R package (v1.0.13) was used to infer the tumour purity, with KEGG pathway annotations obtained via the msigdbr package (v7.5.1). Drug sensitivity across subtypes was predicted for commonly used WT chemotherapeutics using the pRRophetic[12], with the "tissueType" parameter set to "allSolidTumors".

Access and processing of publicly available data

Additional pediatric tumour DNA methylation datasets were downloaded from the TARGET initiative (Neuroblastoma, Osteosarcoma) and from the Gene Expression Omnibus (GEO) database (Fetal kidney: GSE69502; Hepatoblastoma: GSE132399; Adrenocortical carcinoma: GSE131350) of NCBI (National Center for Biotechnology Expression) (http://www.ncbi.nlm.nih.gov/geo/). Datasets from TCGA were accessed and pre-processed using the TCGABiolinks package (v2.32.0) [13] with default parameters. GEO datasets were processed as described above for the study cohort. The datasets were merged using common

CpGs. A gene-expression dataset of additional WT samples was accessed from St. Jude Hospital databank[14].

SUPPLEMENTARY REFERENCES

1 Morris TJ, Butcher LM, Feber A, Teschendorff AE, Chakravarthy AR, Wojdacz TK et al. ChAMP: 450k Chip Analysis Methylation Pipeline. Bioinformatics 2014; 30: 428-430.

2 Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 2013; 29: 15-21.

3 Smith T, Heger A, Sudbery I. UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. Genome Res 2017; 27: 491-499.

4 Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 2014; 30: 923-930.

5 Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 2015; 43: e47.

6 Gu Z, Hubschmann D. rGREAT: an R/bioconductor package for functional enrichment on genomic regions. Bioinformatics 2023; 39.

7 Tarazona S, Garcia-Alcalde F, Dopazo J, Ferrer A, Conesa A. Differential expression in RNA-seq: a matter of depth. Genome Res 2011; 21: 2213-2223.

8 Risso D, Schwartz K, Sherlock G, Dudoit S. GC-content normalization for RNA-Seq data. BMC Bioinformatics 2011; 12: 480.

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

10 Kolberg L, Raudvere U, Kuzmin I, Vilo J, Peterson H. gprofiler2 -- an R package for gene list functional enrichment analysis and namespace conversion toolset g:Profiler. F1000Res 2020; 9.

11 Gaujoux R, Seoighe C. A flexible R package for nonnegative matrix factorization. BMC Bioinformatics 2010; 11: 367.

12 Geeleher P, Cox NJ, Huang RS. Clinical drug response can be predicted using baseline gene expression levels and in vitro drug sensitivity in cell lines. Genome Biol 2014; 15: R47.

13 Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D et al. TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. Nucleic Acids Res 2016; 44: e71.

14 McLeod C, Gout AM, Zhou X, Thrasher A, Rahbarinia D, Brady SW et al. St. Jude Cloud: A Pediatric Cancer Genomic Data-Sharing Ecosystem. Cancer Discov 2021; 11: 1082-1099.

Figure S1. Comparison of the global DNA methylation levels of WT and control

A Violin plots showing the distribution of methylation beta values across all CpG sites profiled by the Illumina EPIC array in WT (pink) and NNK (grey) samples. The y-axis represents beta values, reflecting methylation levels from 0 (unmethylated) to 1 (fully methylated). Asterisks indicate statistical significance between groups (Paired Wilcoxon): **** denotes $p < 0.0001$. **B** Heatmap showing the hierarchical clustering of DMSs associated with genes belonging to pathways in cancer.

Figure S2. Rank survey analysis and subtype-specific gene template derived from the St. Jude WT cohort's expression data

A Multi-facet line plot depicting cophenetic coefficient, dispersion and average silhouette respectively. The metrics (y-axis) were calculated for each tested factorization rank (x-axis) NMF consensus clustering. Higher cophenetic and silhouette values together with lower dispersion indicate more stable and well-separated cluster structures. **B** PCA plot of the coefficient matrix obtained from NMF with rank 3. Each point represents a sample. The separation of points along principal components reflects the distinct cluster assignments derived from the factorization. **C** Heatmap showing the top contributing genes for each of the three subtypes.

Figure S3. Correlation between gene-expression subtypes and DNA methylation clusters

UpSet plot showing overlap between gene-expression WT subtypes and methylation clusters (21 WT samples, Fisher's exact test adjusted p-value = 0.02296125). The horizontal bar chart (left) shows the sizes of each individual set, while the vertical bar chart (top) displays

intersection sizes, indicating the number of samples belonging to specific combinations of subtypes and methylation clusters. Connected dots in the matrix (bottom) represent which sets are included in each intersection.

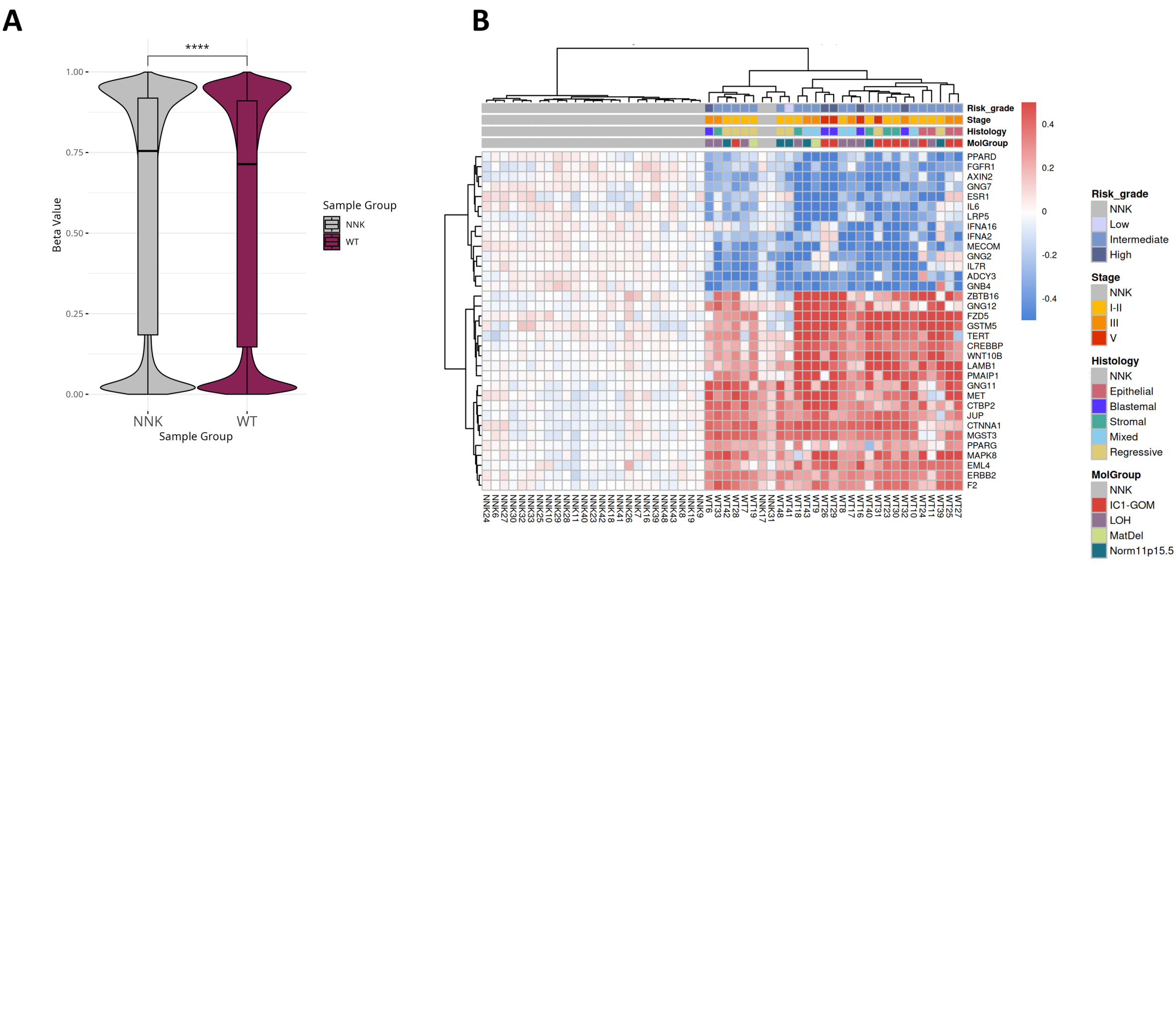
Figure S4. Examples of correlations between cancer gene-expression and promoter DNA methylation

Scatter plots depicting gene-expression and promoter DNA methylation levels for some ncDEGs.

Figure S5. Correlation of *IGF2* expression with the DNA methylation of its regulatory regions

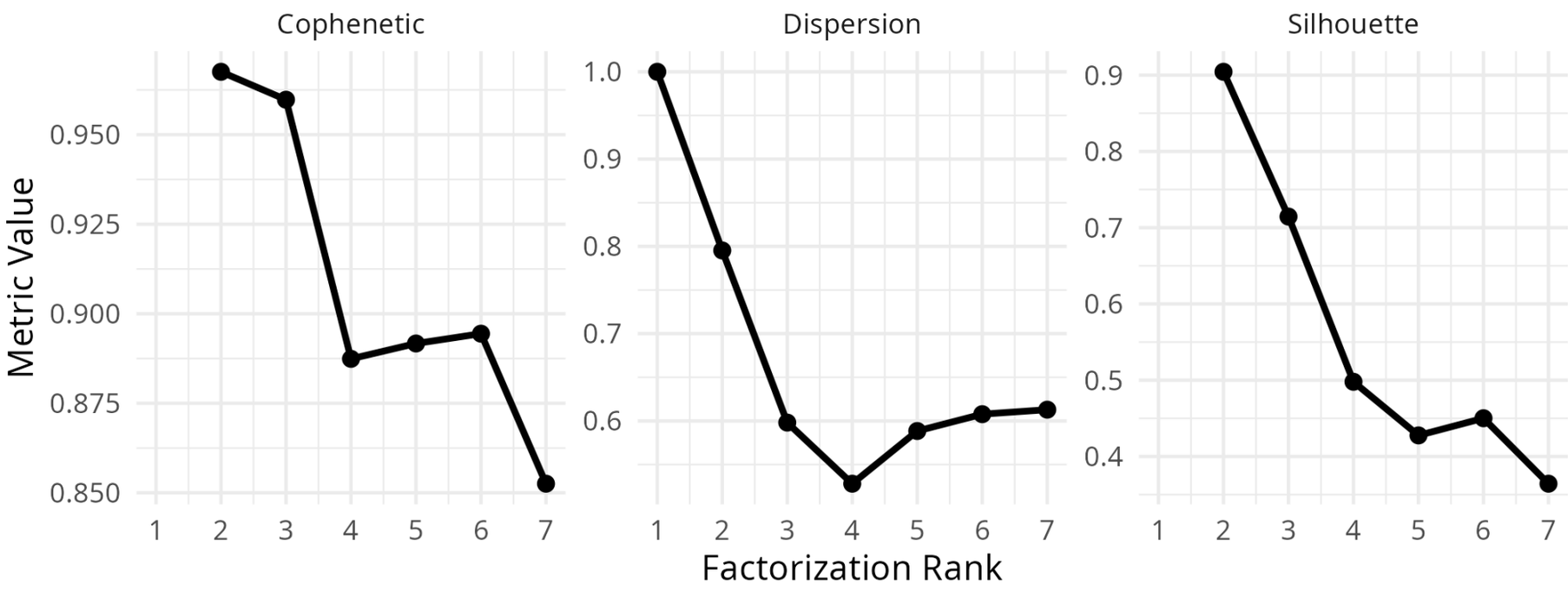
Scatter plots depicting the relationship between *IGF2* expression (log-transformed; x-axis) and DNA methylation levels (b-value; y-axis) at two key regulatory regions: IC1 (**A**) and DMR0 (**B**). Each point represents an individual sample. **C** Multi-facet line plots summarising quantile-regression analysis of *IGF2* expression versus methylation at five 11p15.5 loci (IC1, P4, P3, P2 and DMR0). The x-axis shows the expression quantiles ($\tau = 0.1 – 0.9$), and the y-axis gives the regression coefficient for each quantile. Solid lines trace the estimated coefficients and the shaded ribbons denote 95 % confidence intervals. Coefficients above zero indicate a positive association, while coefficients below zero indicate a negative association. Associations are significant where the confidence interval does not include the horizontal zero line.

Supplementary Figure S1

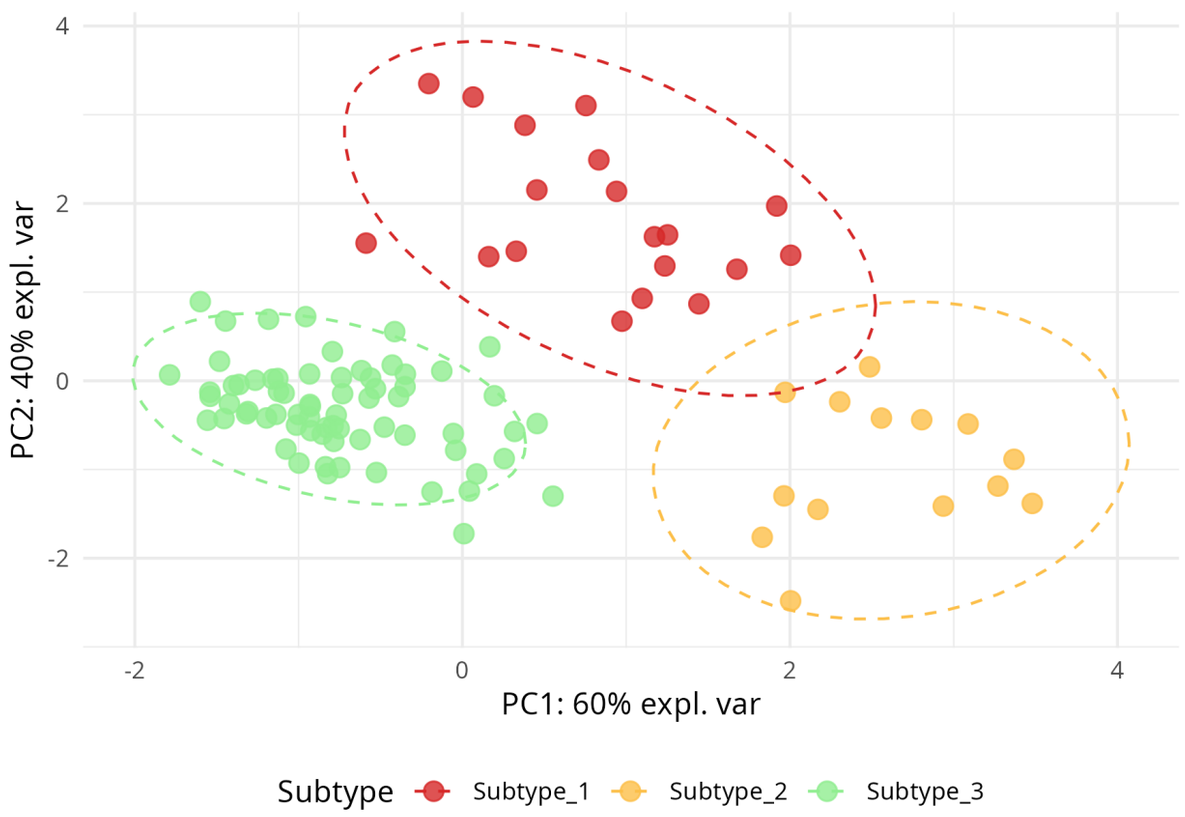


Supplementary Figure S2

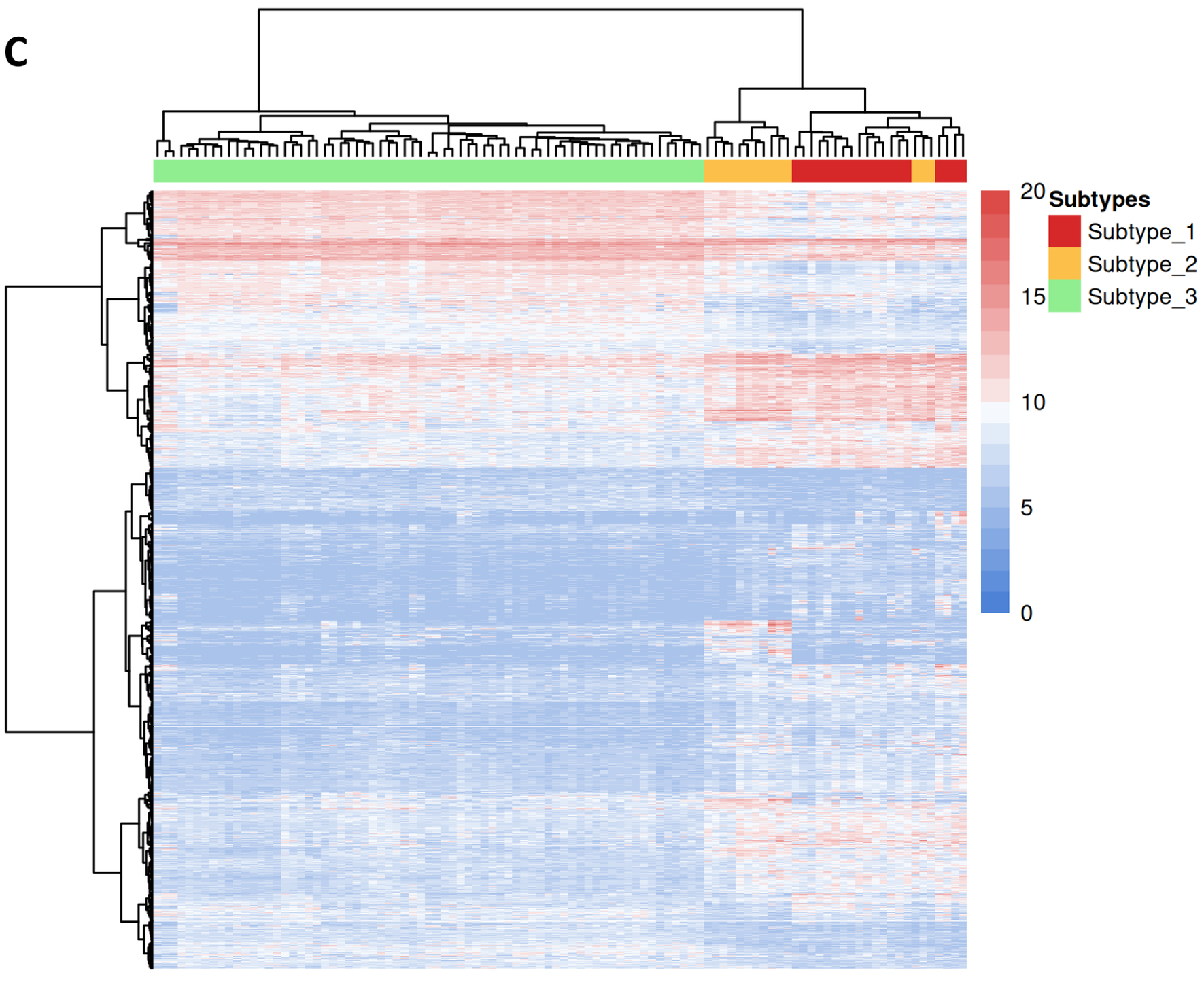
A



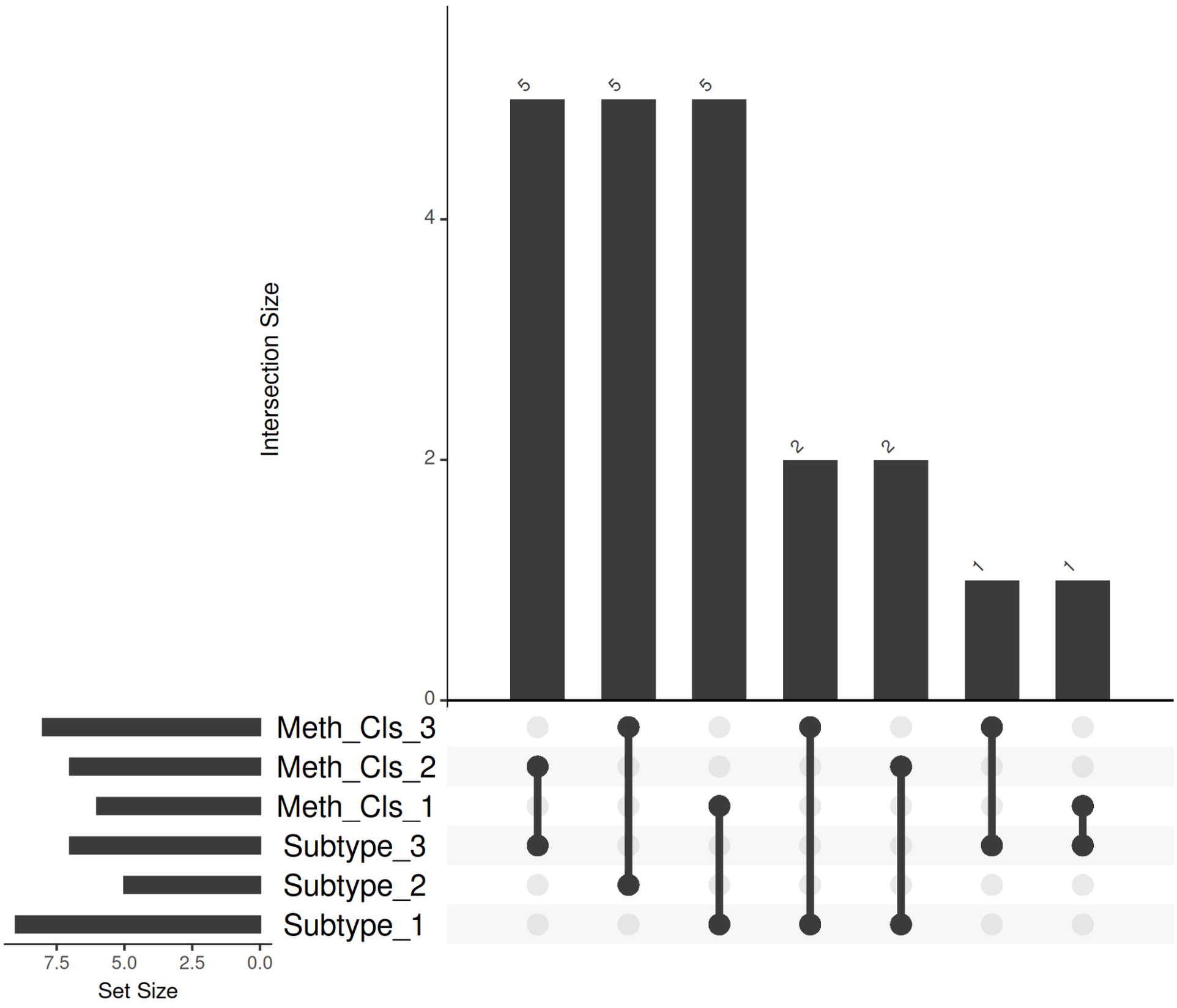
B



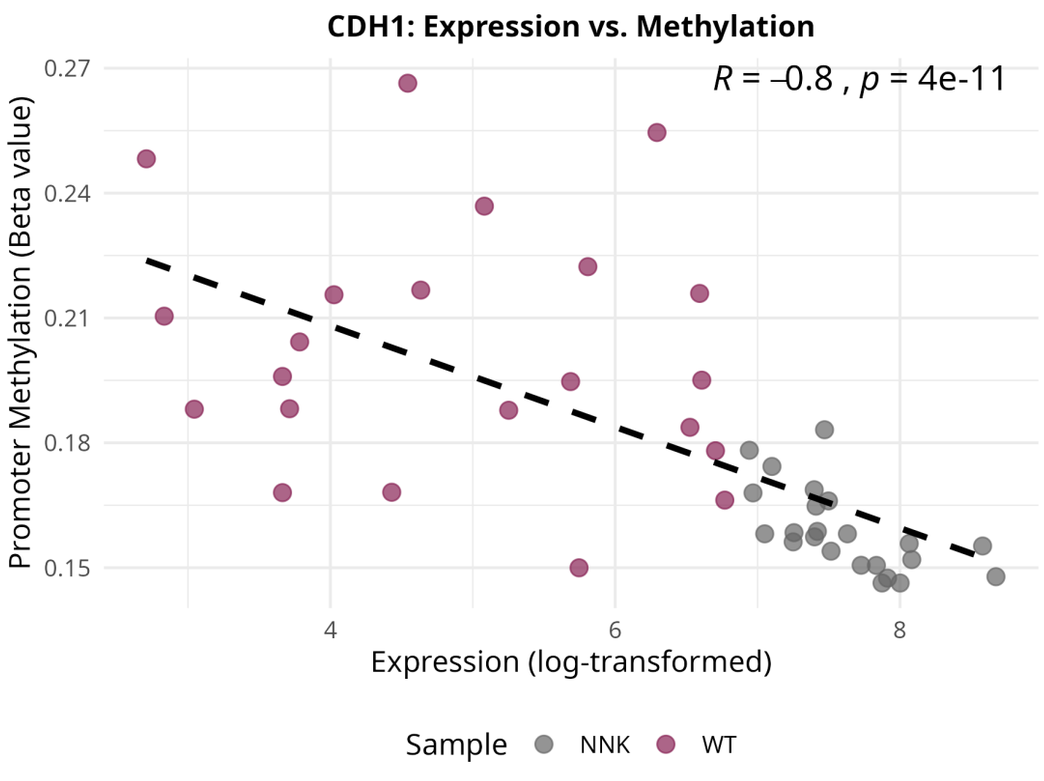
C

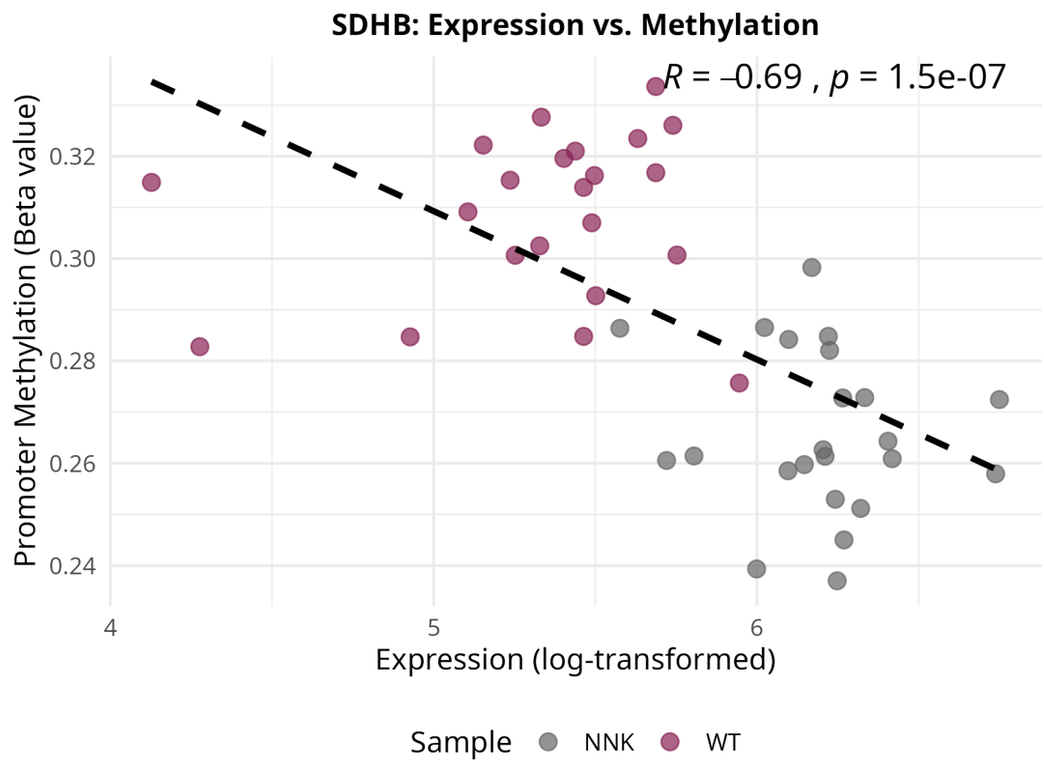


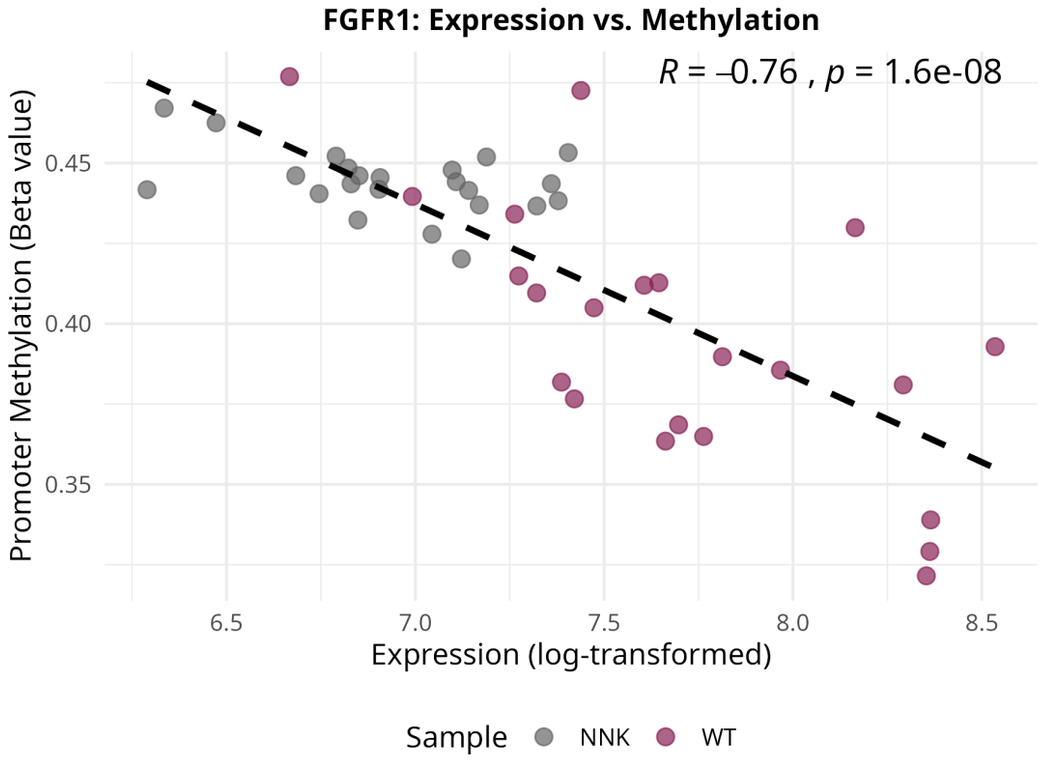
Supplementary Figure S3

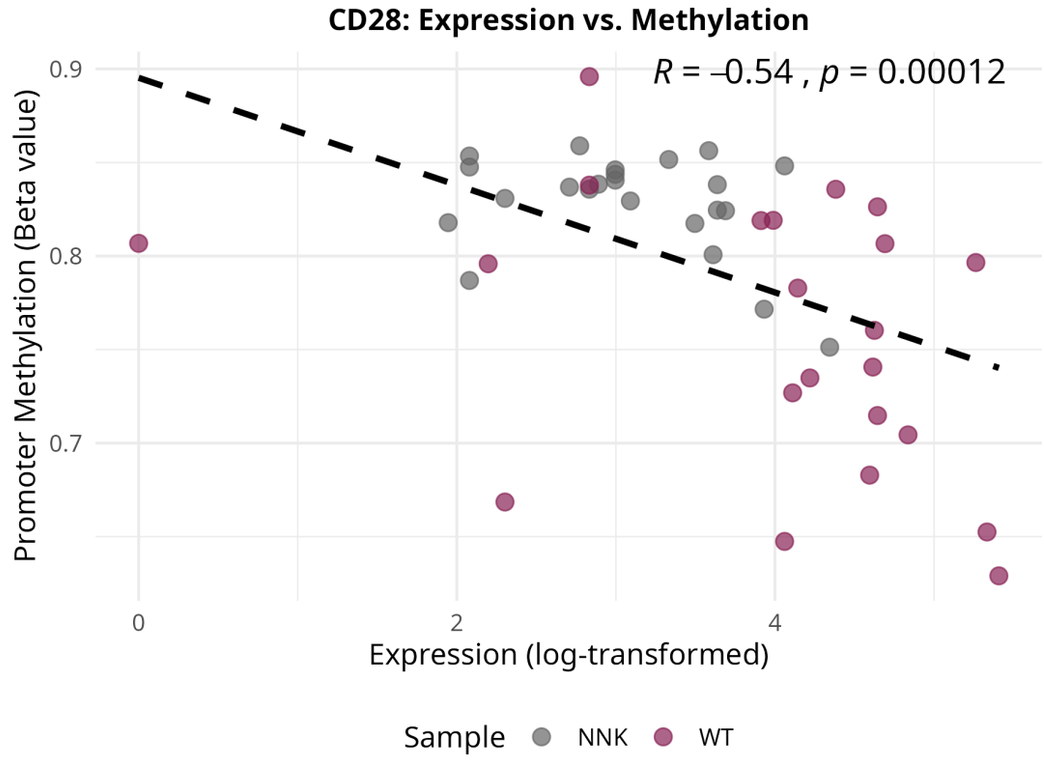


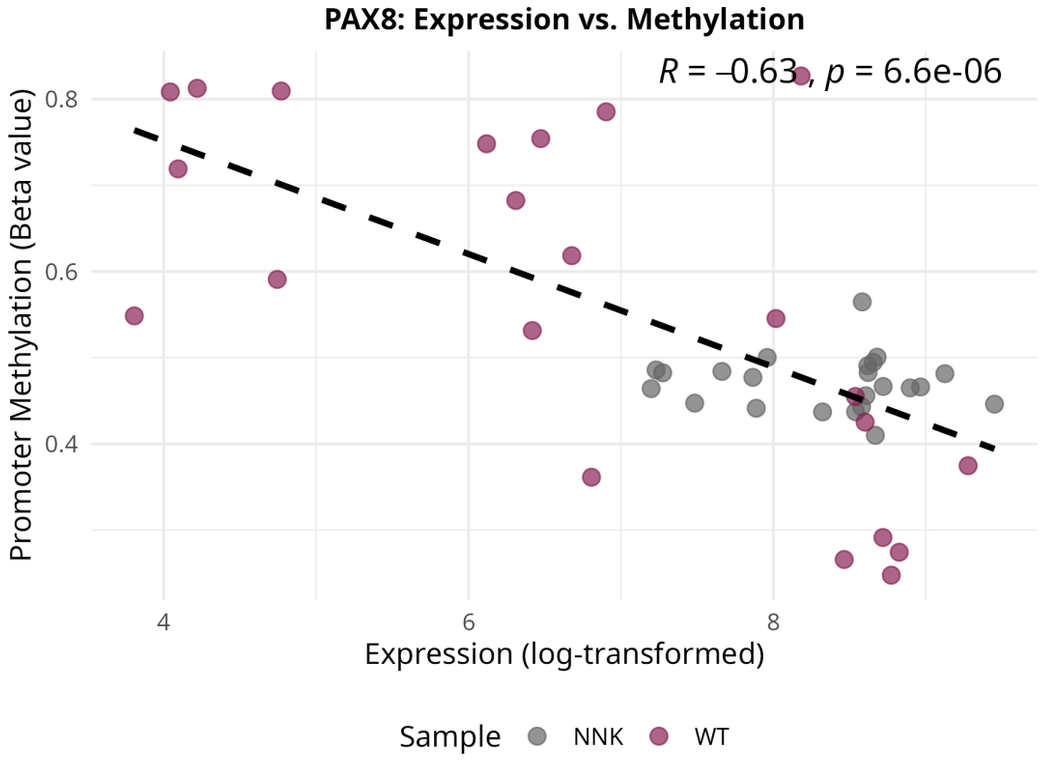
Supplementary Figure S4

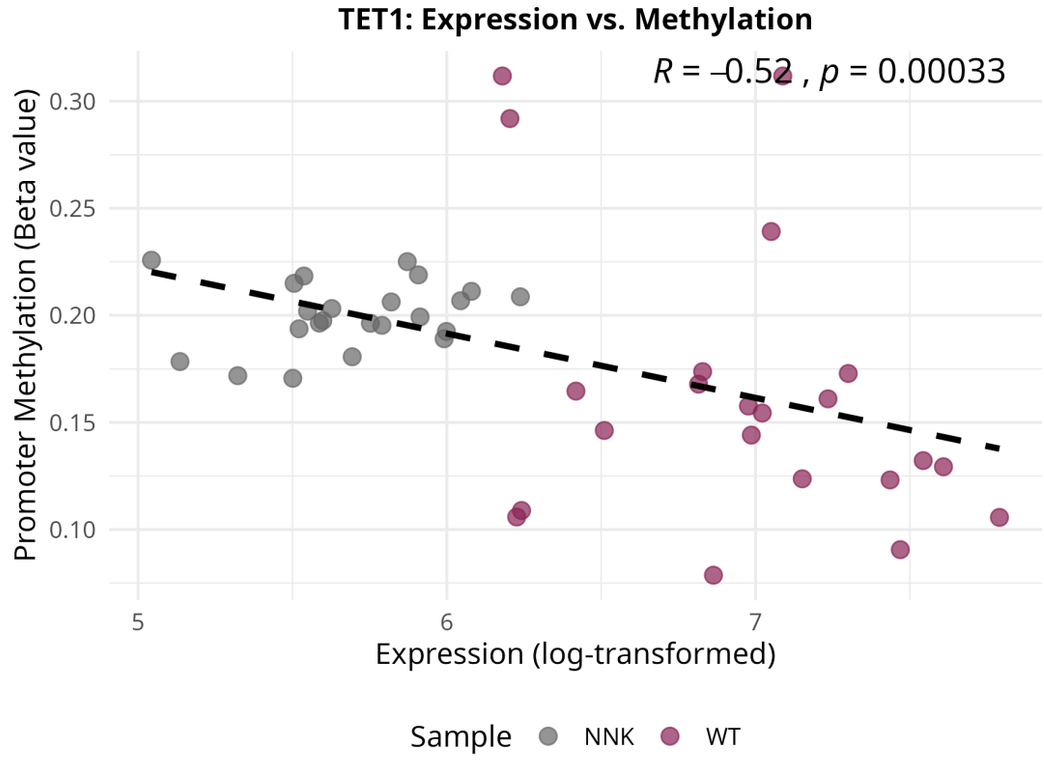






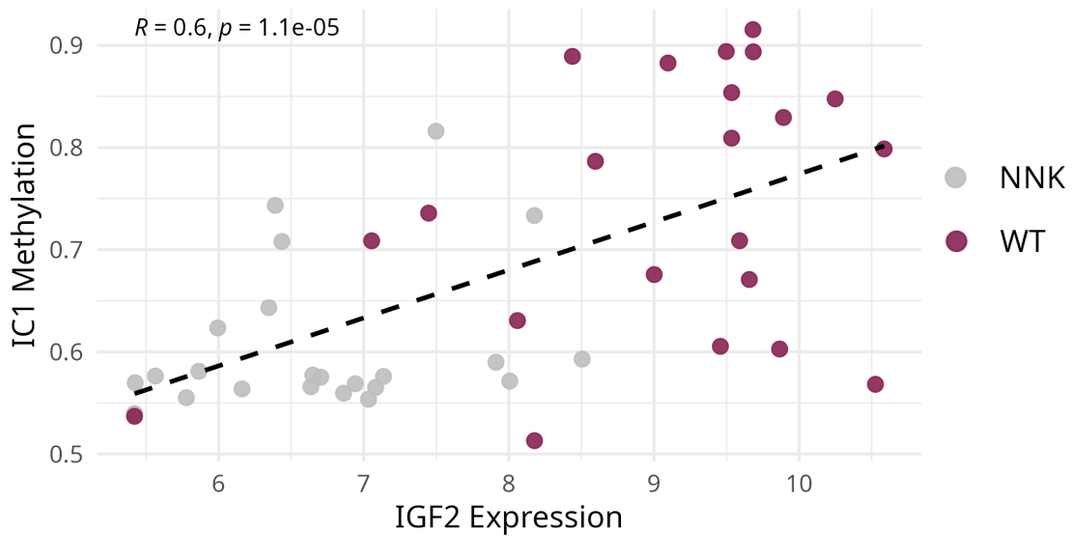






Supplementary Figure S5

A



B

C