SUPPLEMENTARY MATERIAL

MATERIAL AND METHODS

Subjects

This cross-sectional study comprised a total of 113 Spanish individuals, including 82 biopsy-proven GCA patients and 31 age and sex-matched healthy controls. All patients fulfilled the 1990 American College of Rheumatology classification criteria for this disease[1]. Clinical and laboratory characteristics of the GCA patients at disease onset are summarized in **supplementary table 39**. Patients with GCA were consecutively selected among those newly diagnosed and those controlled at the outpatient facility of Hospital Clinic of Barcelona. Three groups of patients were defined according to their clinical disease activity status at the time of sample collection: a) active disease (n=20): newly diagnosed patients treatment naïve (n=16) or a maximum of 2 days of glucocorticoids (GCs) (n=2), or patients with a disease relapse during the follow-up (n=2); b) in remission with treatment (n=33): patients presenting disease in remission on low doses of prednisone (< 10mg/day) during a minimum of 1 month; c) in remission without treatment (n=29): patients in remission of the disease without any treatment for at least 1 month. Remission was defined as the absence of GCA-related symptoms along with normal acute phase reactants. Patients not fitting into these three categories were discarded. Healthy controls samples were obtained from healthy individuals accompanying the patients to the clinic. All samples from GCA patients and healthy controls were collected at Hospital Clinic of Barcelona and all participants signed an informed consent form in accordance with the ethical guidelines of the 1975 declaration of Helsinki. The ethics committee of the hospital approved the study.

Patients and Public involvement

Patients or the public were not involved in any of the stages of our research.

CD14+ monocytes isolation and DNA/RNA extraction

Peripheral blood mononuclear cells (PBMCs) were first obtained from whole blood by density gradient centrifugation using Ficoll-Paque (Rafer, Zaragoza, Spain). For subsequent CD14+ monocytes isolation, PBMCs were incubated with CD14-PE and

CD15-FITC conjugated antibodies (MiltenyiBiotec, Germany) in staining buffer (PBS with 2mM of EDTA and 4% FBS) for 20 minutes in the dark. Pure monocytes were next isolated by positive selection as CD14+CD15- cells using flow cytometry sorting and were pelleted and stored at -80ºC.

Genomic DNA and total RNA were extracted from the same cell pellet utilizing the AllPrep DNA/RNA/miRNA Universal kit (Qiagen, Hilden, Germany), according to manufacturer's instructions. DNA and RNA were quantified with Nanodrop ND-2100 and Qubit® DNA assay kit (invitrogen), respectively. Furthermore, RNA quality was determined by the 2100 Bioanalyzer System (Agilent, CA, USA).

DNA methylation assay and data processing

Five hundred nanograms of each DNA sample was bisulfite-converted with the EZ DNA Methylation™ kit (Zymo Research, Irvine, CA, USA), following the manufacturer's instructions. Samples were hybridized onto an Infinium MethylationEPIC Bead Chip array (Illumina, Inc., San Diego, CA, USA), which measures DNA methylation status at >850,000 CpG sites at single-nucleotide resolution, covering 99% of reference sequence (RefSeq) genes and 95% of CpG islands. To minimize batch effects, samples were randomized across arrays and slides. Samples hybridization and array scanning were performed at the Josep Carreras Research Institute Genomics Platforms (Barcelona).

DNA methylation raw data (IDAT files) were processed using ShinyÉPICo[2], a graphical pipeline based on minfi and limma R packages[3,4]. Firstly, quality controls were conducted. Samples showing poor bisulfite conversion and those with gender discordance between prediction and demographic information were filtered out from further analyses. In addition, we used the 59 single nucleotide polymorphisms (SNPs) probes included in the array to check for sample duplication. One individual from each pair of duplicates was excluded. Number of samples maintained after quality controls filters are displayed in **supplementary table 39.** Next, probes with a detection p-value < 0.01 were removed. To avoid technical and biological bias, we also excluded CpGs from X and Y chromosomes as well as those probes containing either a SNP at the CpG interrogation site or at the single base extension site. Probes were annotated using IlluminaHumanMethylationEPICmanifest v0.4.0[5]. Finally, raw methylation values were normalized using the Noob+Quantile method,

and DNA methylation was measured as beta and M values. Beta values (ranging from 0 to 1) that correspond with the intensity ratios of methylated and unmethylated probes were used for visual representation and biological interpretation. For statistical purposes, beta values were converted to M values (log2-transformed beta values) to achieve a normal distribution.

To identify associations between DNA methylation levels and GCA disease state or clinical status along the genome, we applied a eBayes moderated t-test with trend and robust options enabled. We used a false discovery rate (FDR) to correct for multiple testing and those probes with a FDR < 0.05 were considered as differentially methylated positions (DMPs). Sensitivity analyses were realized for each comparison to examine the contribution of all potential confounders (sex, age, slide, array and sample plate). Pearson correlation or Wilcoxon signed-rank test was applied depending on whether the variable of interest was continuous or categorical. Variables with a p-value < 0.05 were considered to significantly contribute to DNA methylation and were therefore included as covariates in the model (**supplementary figure 3**).

RNA-seq and gene expression analysis

For library construction, 1 µg of excellent quality RNA (RNA integrity number >7) was subjected to the TruSeq Stranded mRNA Library Prep Kit (Illumina, CA, USA) according to the manufacturer's protocol. The paired-end sequencing was conducted on a HiSeq sequencer (Illumina, CA, USA) producing 34.6x2 M raw paired reads sample on average. We used miARma-Seq pipeline to process the data and perform the analyses[7]. First, raw data were evaluated using FastQC software to analyse the quality of the reads, and afterwards, sequences were aligned to the GRCh38 reference genome with STAR[8,9].

In addition, principal component analysis (PCA) and hierarchical clustering of normalized samples were performed to check the similarity of RNA-sequencing samples (**supplementary figure 4**)[10]. Samples identified as outliers were filtered out from further analyses as well as duplicated samples and those presenting gender discrepancies. Number of samples maintained after quality controls filters are displayed in **supplementary table 40.**

Differential expression analyses were carried out with the edgeR package[11]. Low expressed genes were removed (CPM< 1) and the remaining genes were normalized by the trimmed mean of M-values (TMM) method. Furthermore, we calculated reads per kilobase per million mapped reads (RPKM) and counts per million (CPM) and log2-counts per million (log-CPM) per gene on each sample[12]. All genes with a FDR < 0.05 were statistically significant and designated as differentially expressed genes (DEGs). log2FC was used to assess expression changes among groups of comparison. To evaluate the contribution of potential confounders such as sex and age, we performed sensitivity analyses. We applied Pearson correlation or Wilcoxon signed-rank test depending on whether the variable of interest was continuous or categorical. Variables with a p-value < 0.05 were considered to significantly contribute to gene expression and were therefore included as covariates in the model **(**supplementary figure 5).

Enrichment analysis

Gene ontology (GO) enrichment analyses were assessed for biological process, molecular function, and cellular complex ontology terms. For DMPs, we used the GREAT online tool v4.0.4 in which CpGs were annotated to the nearest gene[13]. In this analysis, annotated CpGs in the EPIC array were selected as background. GO terms with a p-value < 0.01 and fold enrichment > 2 were considered as significantly enriched. We also carried out a transcription factor binding motif enrichment analysis by using HOMER motif discovery software v4.5, where a 250 bp-window upstream and downstream of the DMPs was applied. For background correction, CpGs annotated in the EPIC array were used[14].

For DEGs, GO analyses were carried out with the online tool DAVID under functional annotation settings[15]. GO categories with a p-value < 0.01 and with a minimum count of 3 genes were considered significant. Kyoto Encyclopedia of genes and genomes (KEGG) pathway enrichment was also calculated from DEGs. For data visualization, results were plotted with the ggplot2 R package[16].

Integrative analysis of transcriptome and DNA methylation

In order to identify interactions between DNA methylation changes and gene expression alterations, we applied a Pearson correlation test by using the MatrixEQTL R package[17]. A maximum distance of 1 Mb between CpG sites and genes was defined. Those CpG-gene expression interactions with a FDR < 0.05 were considered significant. To focus on the CpG-gene expression interactions that could be relevant in the pathophysiology of GCA and/or in the clinical status, we further selected those interaction pairs in which both methylation and expression levels were independently significantly associated. We applied this filter in each of the comparisons performed, resulting in both CpG-gene expression interactions significantly associated in an unique comparison and CpG-gene expression interactions significantly associated in more than one comparison.

Validation of expression levels by real-time quantitative reverse-transcribed polymerase chain reaction (qRT-PCR)

300 ng of total RNA were reverse-transcribed to cDNA with Transcriptor First Strand cDNA Synthesis Kit (Roche) following manufacturer's instructions. qRT-PCR was performed in technical triplicate for each biological replicate, using LightCycler® 480 SYBR Green Mix (Roche), and 5 ng of cDNA per reaction. The average value from each technical replicate was obtained. Then, the standard double-delta Ct method was used to determine the relative quantities of target genes, and values were normalized against the control gene *RPL38*. Custom primers were designed to analyze genes of interest (supplementary table 41). Student's t test was performed to compare the difference in mean ΔΔCt values between the comparison groups, all analyses with a p-value < 0.05 were considered significant. In addition, a correlation analysis between ΔCt values obtained by qPCR and corresponding normalized intensities (TMM values) from the RNA-seq were estimated using the Spearman correlation coefficient in R Software.

References

1. Hunder, G. G. et al. The American College of Rheumatology 1990 criteria for the classification of giant cell arteritis. Arthritis Rheum. 33, 1122–1128 (1990).
2. Morante-Palacios, O. & Ballestar, E. shinyÉPICo: A graphical pipeline to analyze Illumina DNA methylation arrays. Bioinformatics 37, 257–259 (2021).
3. Aryee, M. J. et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics 30, 1363–1369 (2014).
4. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 43, e47–e47 (2015).
5. Hansen KD. IlluminaHumanMethylationEPICmanifest: Manifest for Illumina's EPIC methylation arrays. R package version 0.3.0, https://bitbucket.com/kasperdanielhansen/Illumina_EPIC. https://bioconductor.org/packages/release/data/annotation/html/IlluminaHumanMethylationEPICmanifest.html (2016).
6. Teschendorff, A. E. et al. DNA methylation outliers in normal breast tissue identify field defects that are enriched in cancer. Nat. Commun. 2016 71 7, 1–12 (2016).
7. Andrés-León, E., Núñez-Torres, R. & Rojas, A. M. miARma-Seq: a comprehensive tool for miRNA, mRNA and circRNA analysis. Sci. Reports 2016 61 6, 1–8 (2016).
8. Andrews S. FastQC: a quality control tool for high throughput sequence data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (2010).
9. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21 (2013).
10. Reeb, P. D., Bramardi, S. J. & Steibel, J. P. Assessing Dissimilarity Measures for Sample-Based Hierarchical Clustering of RNA Sequencing Data Using Plasmode Datasets. PLoS One 10, e0132310 (2015).
11. Nikolayeva, O. & Robinson, M. D. edgeR for differential RNA-seq and ChIP-seq analysis: an application to stem cell biology. Methods Mol. Biol. 1150, 45–79 (2014).
12. Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. 11, (2010).
13. McLean, C. Y. et al. GREAT improves functional interpretation of cis-regulatory regions. Nat. Biotechnol. 28, 495–501 (2010).
14. Heinz, S., Benner, C., Spann, N., Bertolino, E., et al. Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. Mol Cell. 28;38(4):576-589 (2010).
15. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. 2009 41 4, 44–57 (2008).
16. Wickham H. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. ISBN 978-3-319-24277-4. https://ggplot2.tidyverse.org/reference/index.html (2016).
17. Shabalin, A. A. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. Bioinformatics 28, 1353–1358 (2012).

Supplementary table legends

Supplementary table 1. Results from the methylation analysis between GCA patients and controls. List of the significant hypermethylated DMPs.

Supplementary table 2. Gene Ontology results from the hypermethylated DMPs between GCA patients and controls.

Supplementary table 3. Results from the methylation analysis between GCA patients and controls. List of the significant hypomethylated DMPs.

Supplementary table 4. Gene Ontology results from the hypomethylated DMPs between GCA patients and controls.

Supplementary table 5. Results from the gene expression analysis between GCA patients and controls. List of the significant upregulated DEGs.

Supplementary table 6. Results from the gene expression analysis between GCA patients and controls. List of the significant downregulated DEGs.

Supplementary table 7. Results from the methylation analysis between patients with active disease and controls. List of the significant hypermethylated DMPs.

Supplementary table 8. Results from the methylation analysis between patients with active disease and controls. List of the significant hypomethylated DMPs.

Supplementary table 9. Gene Ontology results from the hypermethylated DMPs between patients with active disease and controls.

Supplementary table 10. Gene Ontology results from the hypomethylated DMPs between patients with active disease and controls.

Supplementary table 11. Results from the methylation analysis between patients with active disease and patients in remission without treatment. List of the significant hypomethylated DMPs.

Supplementary table 12. Results from the methylation analysis between patients with active disease and patients in remission without treatment. List of the significant hypermethylated DMPs.

Supplementary table 13. Gene Ontology results from the hypomethylated DMPs between patients with active disease and patients in remission without treatment.

Supplementary table 14. Gene Ontology results from the hypermethylated DMPs between patients with active disease and patients in remission without treatment.

Supplementary table 15. Results from the methylation analysis between patients with active disease and patients in remission with treatment. List of the significant hypomethylated DMPs.

Supplementary table 16. Results from the methylation analysis between patients with active disease and patients in remission with treatment. List of the significant hypermethylated DMPs.

Supplementary table 17. Gene Ontology results from the hypomethylated DMPs between patients with active disease and patients in remission with treatment.

Supplementary table 18. Gene Ontology results from the hypermethylated DMPs between patients with active disease and patients in remission with treatment.

Supplementary table 19. Results from the methylation analysis between patients in remission with treatment and patients in remission without treatment. List of the significant DMPs.

Supplementary table 20. Gene Ontology results from the hypomethylated DMPs between patients in remission with treatment and patients in remission without treatment.

Supplementary table 21. Validation of the expression levels by real-time quantitative reverse-transcribed polymerase chain reaction (qRT-PCR).

Supplementary table 22. Results from the gene expression analysis between patients with active disease and controls. List of the significant upregulated DEGs.

Supplementary table 23. Results from the gene expression analysis between patients with active disease and controls. List of the significant downregulated DEGs.

Supplementary table 24. Gene Ontology results from the upregulated DEGs between patients with active disease and controls.

Supplementary table 25. Gene Ontology results from the downregulated DEGs between patients with active disease and controls.

Supplementary table 26. Results from the gene expression analysis between patients with active disease and patients in remission without treatment. List of the significant upregulated DEGs.

Supplementary table 27. Results from the gene expression analysis between patients with active disease and patients in remission without treatment. List of the significant downregulated DEGs.

Supplementary table 28. Gene Ontology results from the upregulated DEGs between patients with active disease and patients in remission without treatment.

Supplementary table 29. Gene Ontology results from the downregulated DEGs between patients with active disease and patients in remission without treatment.

Supplementary table 30. Results from the gene expression analysis between patients with active disease and patients in remission with treatment. List of the significant upregulated DEGs.

Supplementary table 31. Results from the gene expression analysis between patients with active disease and patients in remission with treatment. List of the significant downregulated DEGs.

Supplementary table 32. Gene Ontology results from the upregulated DEGs between patients with active disease and patients in remission with treatment.

Supplementary table 33. Gene Ontology results from the downregulated DEGs between patients with active disease and patients in remission with treatment.

Supplementary table 34. Results from the gene expression analysis between patients in remission with treatment and patients in remission without treatment. List of the significant upregulated DEGs.

Supplementary table 35. Results from the gene expression analysis between patients in remission with treatment and patients in remission without treatment. List of the significant downregulated DEGs.

Supplementary table 36. Gene Ontology results from the upregulated DEGs between patients in remission with treatment and patients in remission without treatment.

Supplementary table 37. Gene Ontology results from the downregulated DEGs between patients in remission with treatment and patients in remission without treatment.

Supplementary table 38. Results of the integrative analysis between DNA methylation and gene expression data in GCA. List of significant CpG-gene expression interactions.

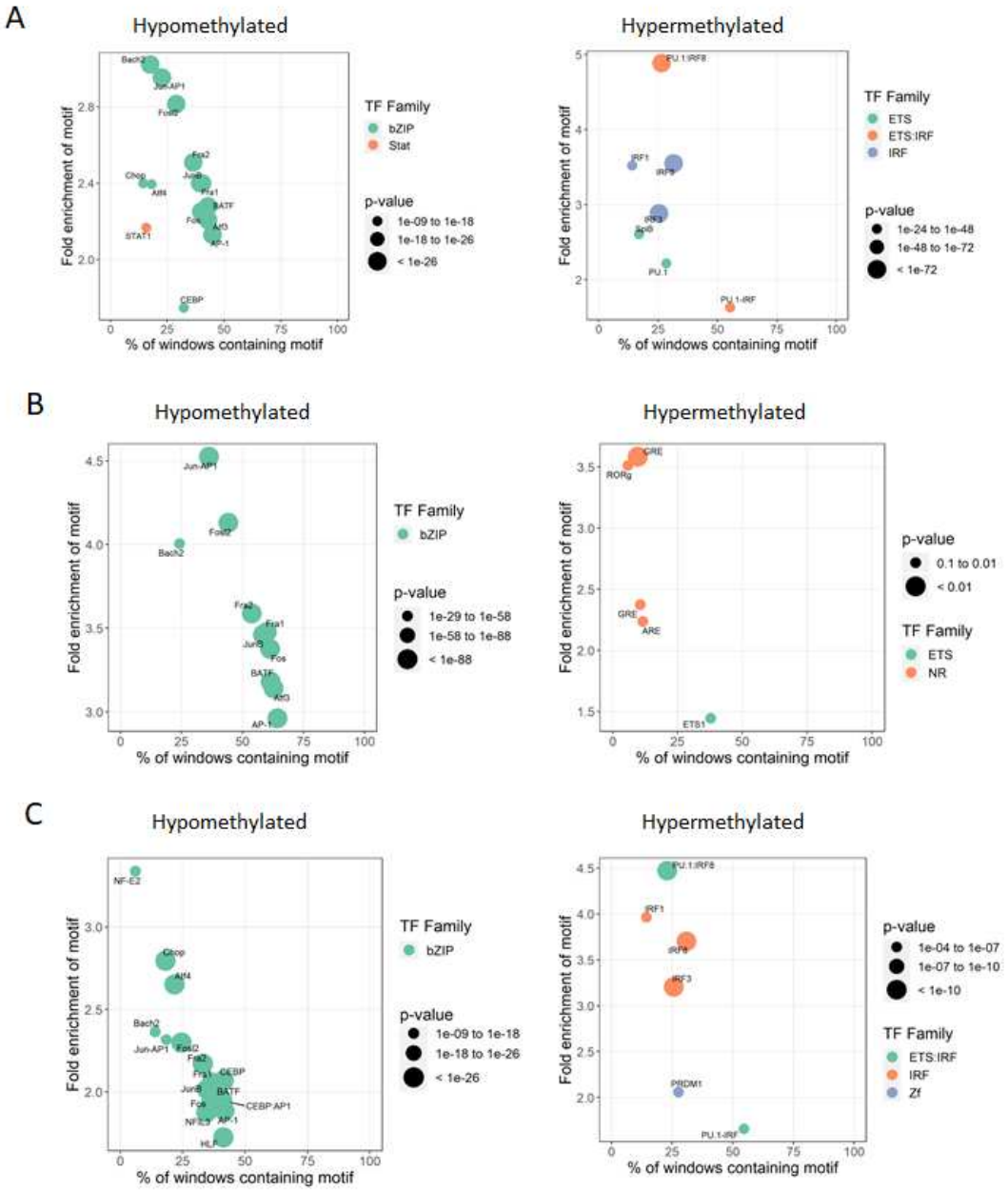
Supplementary table 39. Clinical and laboratory findings of GCA patients at disease onset.

Supplementary table 40. Summary of the sample sizes included in each analysis.

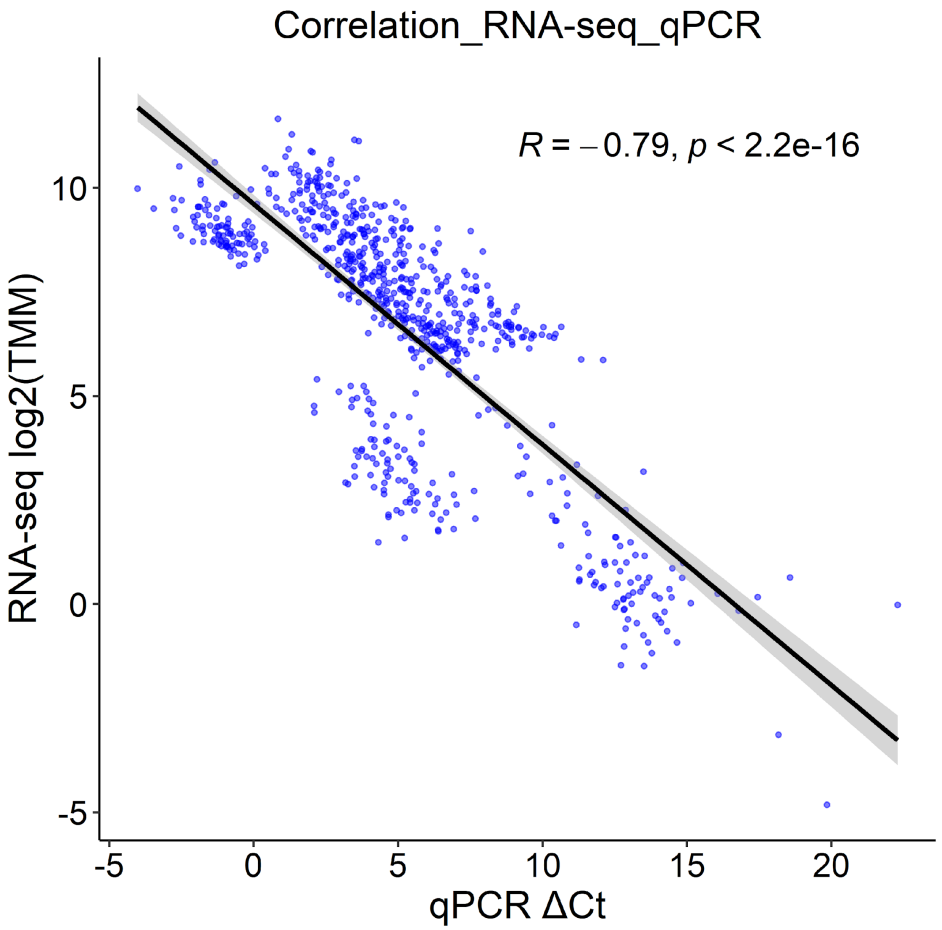
Supplementary table 41. Custom primers for real-time quantitative reverse-transcribed polymerase chain reaction (qRT-PCR).

Supplementary figures

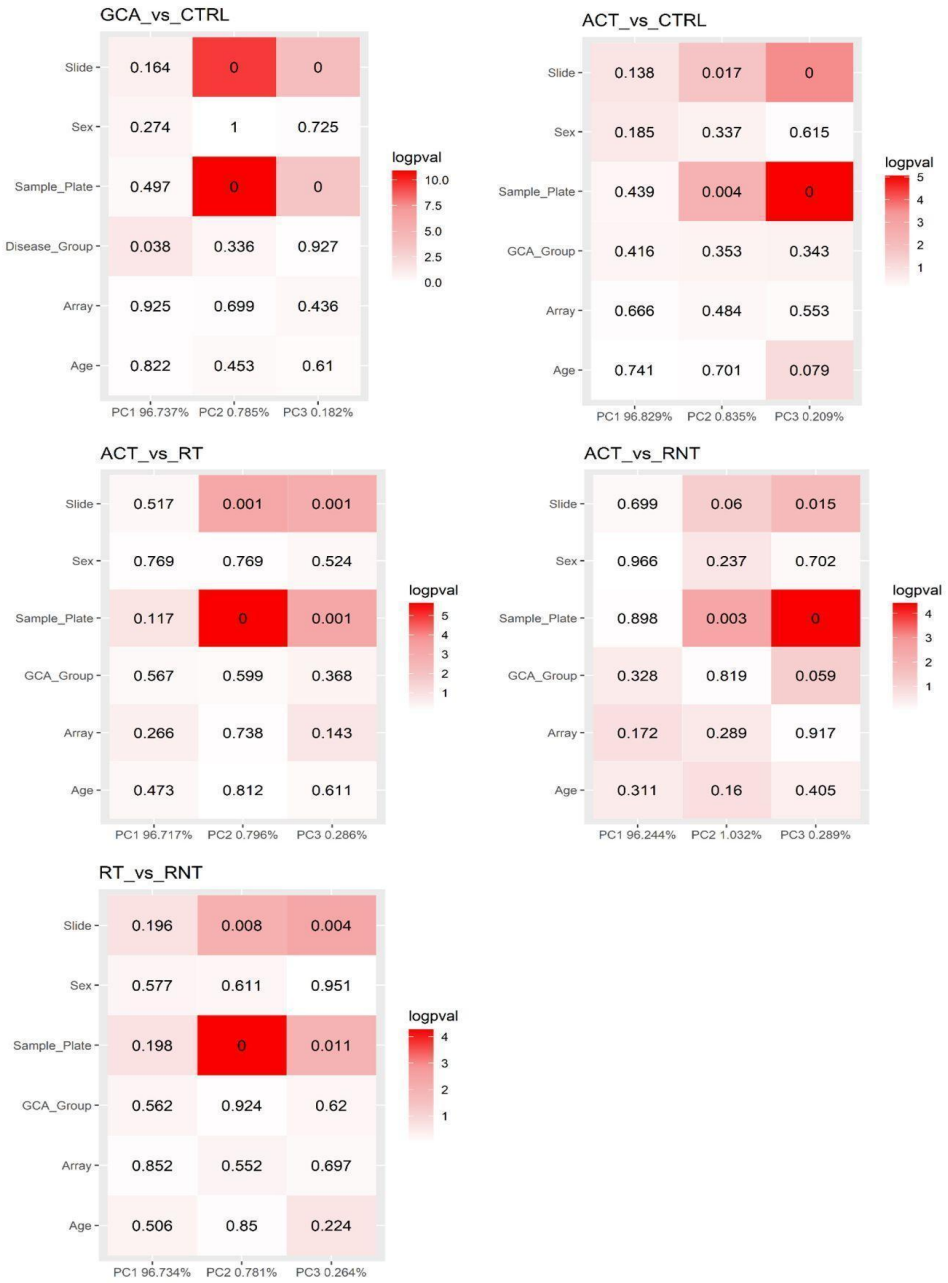
Supplementary figure 1. HOMER motif enrichment of hypo- and hypermethylated CpGs, utilizing CpGs annotated in the EPIC array background. A window of 250 bp-window upstream and downstream of each DMPs was applied. A) HOMER motif enrichment of DMPs from the comparison between active patients and healthy controls. B) HOMER motif enrichment of DMPs from the comparison between active patients and patients in remission with treatment. C) HOMER motif enrichment of DMPs from the comparison between active patients and patients in remission without treatment.



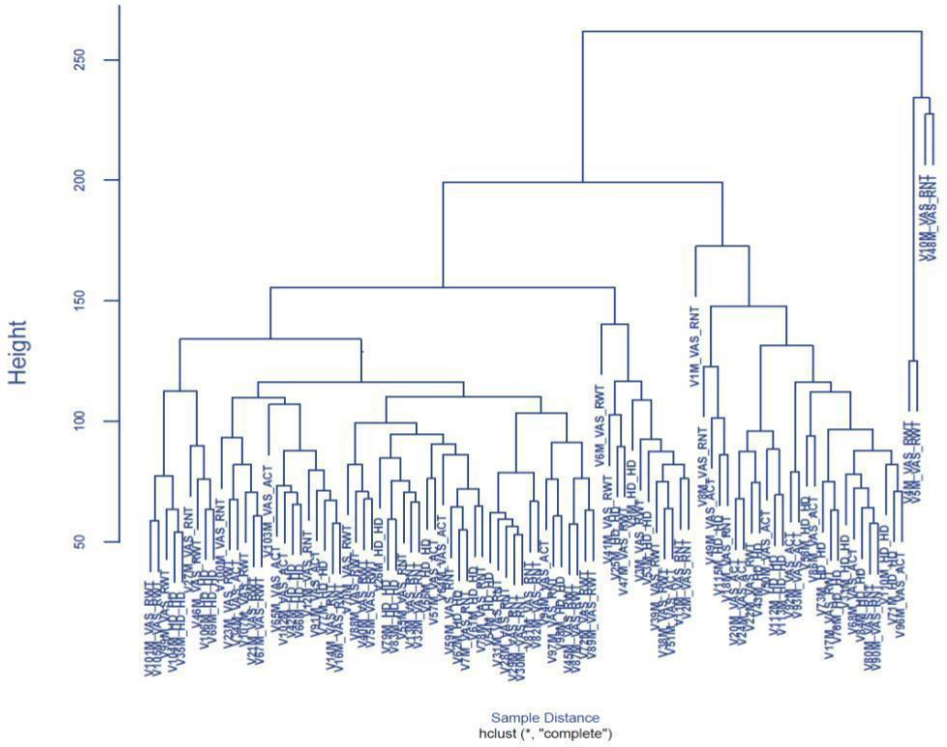
Supplementary figure 2. Real time quantitative PCR validation of RNA-seq results for eight deregulated genes. The plot represents the correlation among the log2 of normalized RNA-seq intensities (TMM) (y-axis) and cycle threshold values (ΔCt) from qRT-PCR (x-axis). TMM = trimmed mean of M-values.



Supplementary figure 3. Contribution of covariates to DNA methylation in each contrast performed in our study. GCA = giant cell arteritis patients, CTRL = controls, ACT = active disease patients, RT = remission treatment patients, RNT = remission no treatment patients.



Supplementary figure 4. Hierarchical Clustering from RPKM (normalized data) of all samples including in the RNA-seq analysis.



Supplementary figure 5. Contribution of covariates to gene expression in each contrast performed in our study. GCA = giant cell arteritis patients, CTRL = controls, ACT = active disease patients, RT = remission treatment patients, RNT = remission no treatment patients.