# Blood DNA methylation analysis reveals a distinctive epigenetic signature of vasospasm in aneurysmal subarachnoid hemorrhage.

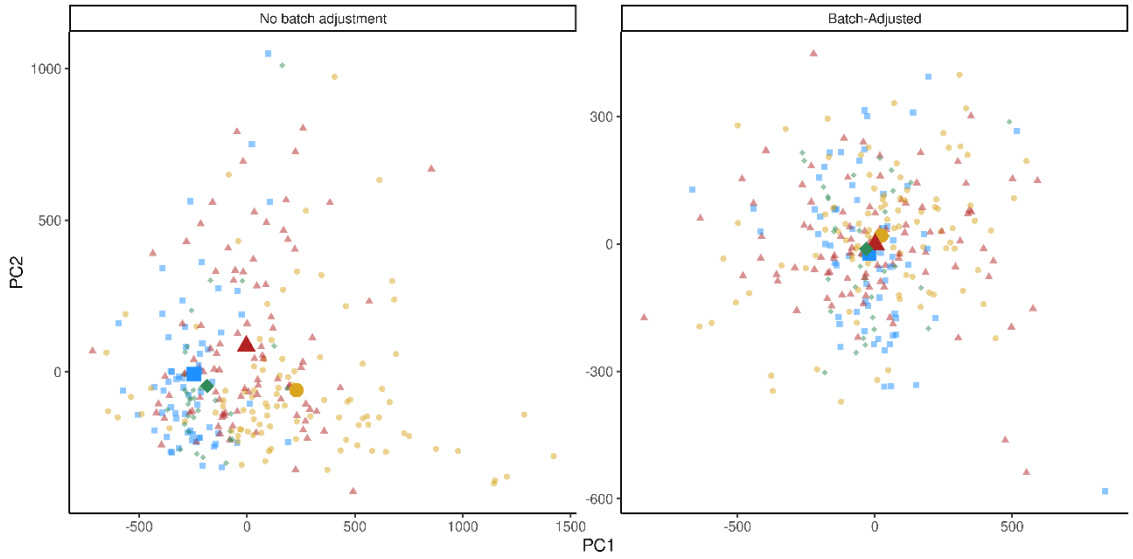
Isabel Fernández-Pérez, MD[1,2], Joan Jiménez-Balado, PhD[2], Adrià Macias-Gómez, MD[1,2], Antoni Suárez-Pérez, MD[1,2], Marta Vallverdú-Prats, PhD[2], Alberto Perez Giraldo, MD[3], Marc Viles, MD[4], Julia Peris, MD[1], Sergio Vidal, MD[1], Eva Giralt-Steinhauer, MD, PhD[1,2,5], Daniel Guisado-Alonso MD, PhD[1,2], Manel Esteller, MD, PhD[6,7], Ana Rodriguez-Campello, MD, PhD[1,2,5], Jordi Jiménez-Conde, MD, PhD[1,2,5], Angel Ois, MD, PhD*[1,2,5], Elisa Cuadrado-Godia, MD, PhD*[1,2,5]

1. Neurology Department, Hospital del Mar, Barcelona, Catalunya, Spain
2. Neurovascular Research Group, Hospital del Mar Medical Research Institute, Barcelona, Catalunya, Spain
3. Neurosurgery Department, Hospital del Mar, Barcelona, Catalunya, Spain
4. Neuroradiology Department, Hospital del Mar, Barcelona, Catalunya, Spain
5. Pompeu Fabra University, Barcelona, Catalunya, Spain
6. Cancer Epigenetics Group, Research Institute Against Leukemia Josep Carreras, Badalona, Catalunya, Spain
7. Physiological Sciences Department, School of Medicine and Health Sciences, University of Barcelona, Barcelona, Catalunya, Spain

***contributed as last authors.***

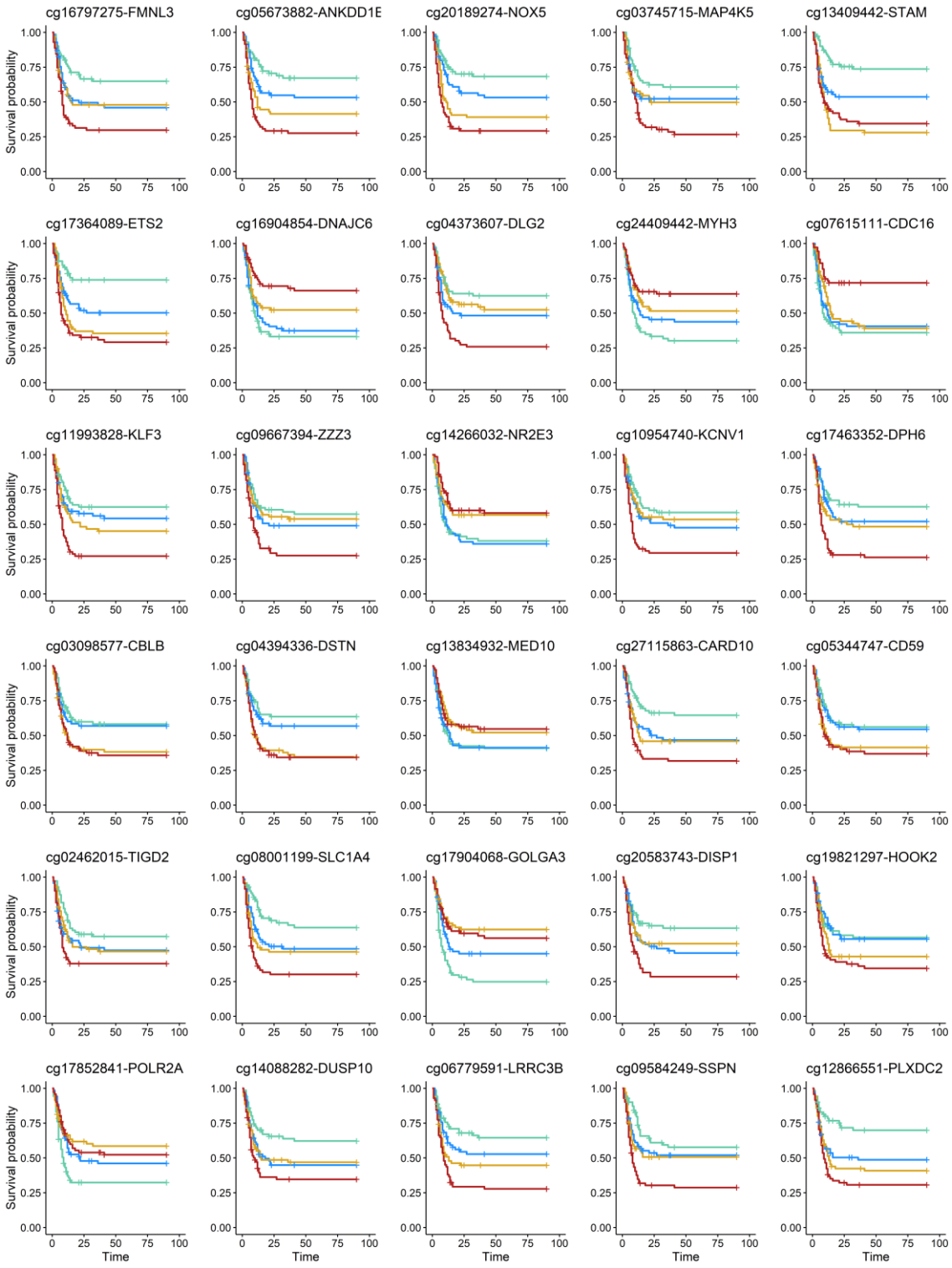
**Corresponding author:** Joan Jiménez Balado. E-mail address: jjimenez3@researchmar.net.

## SUPPLEMENTARY FIGURES



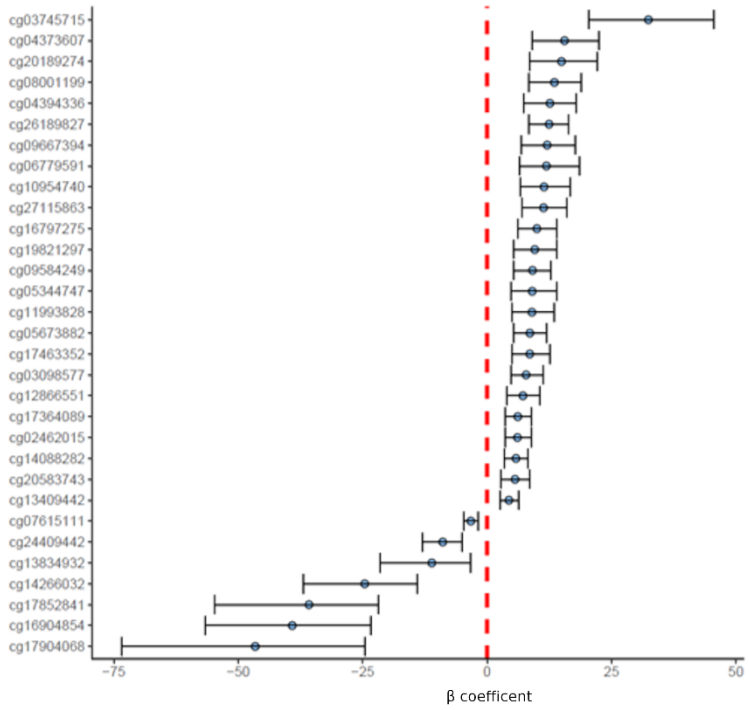

**Supplementary figure 1: Adjustment for batch effect.** We used principal components analysis to reduce the dimensions of DNAm dataset. In this plot we show the first dimension (PC1) against the second one (PC2) before (left panel) and after (right panel) adjusting for the batch effect. Each small dot represents a patient labeled (color and shape) according to the technical

run. Big dots represent the centroid of each technical run. As observed, batch adjustment reduced systematic differences between technical runs (centroids look closer).

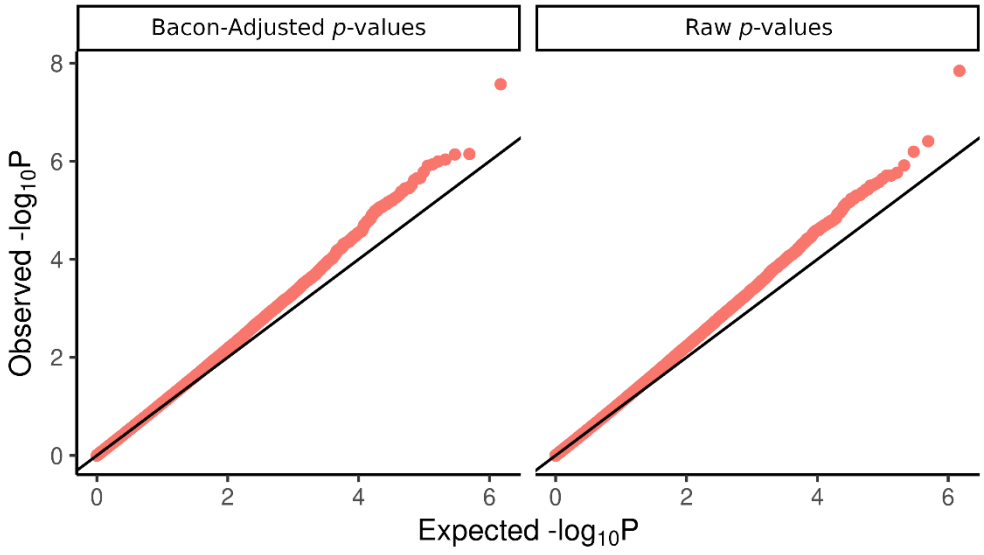



**Supplementary figure 2: Effect of CpG candidates on incidence of vasospasm.** In this figure we show the Kapplan-Meier curves of nominally siognificant DMPss ($p$-value<$10^{-5}$). Labels at the top of each plot indicate the CpG labels (Illumina) and annotated gene (GREAT software). The X axis represents time in days, while the Y axis represents the survival function. Groups have

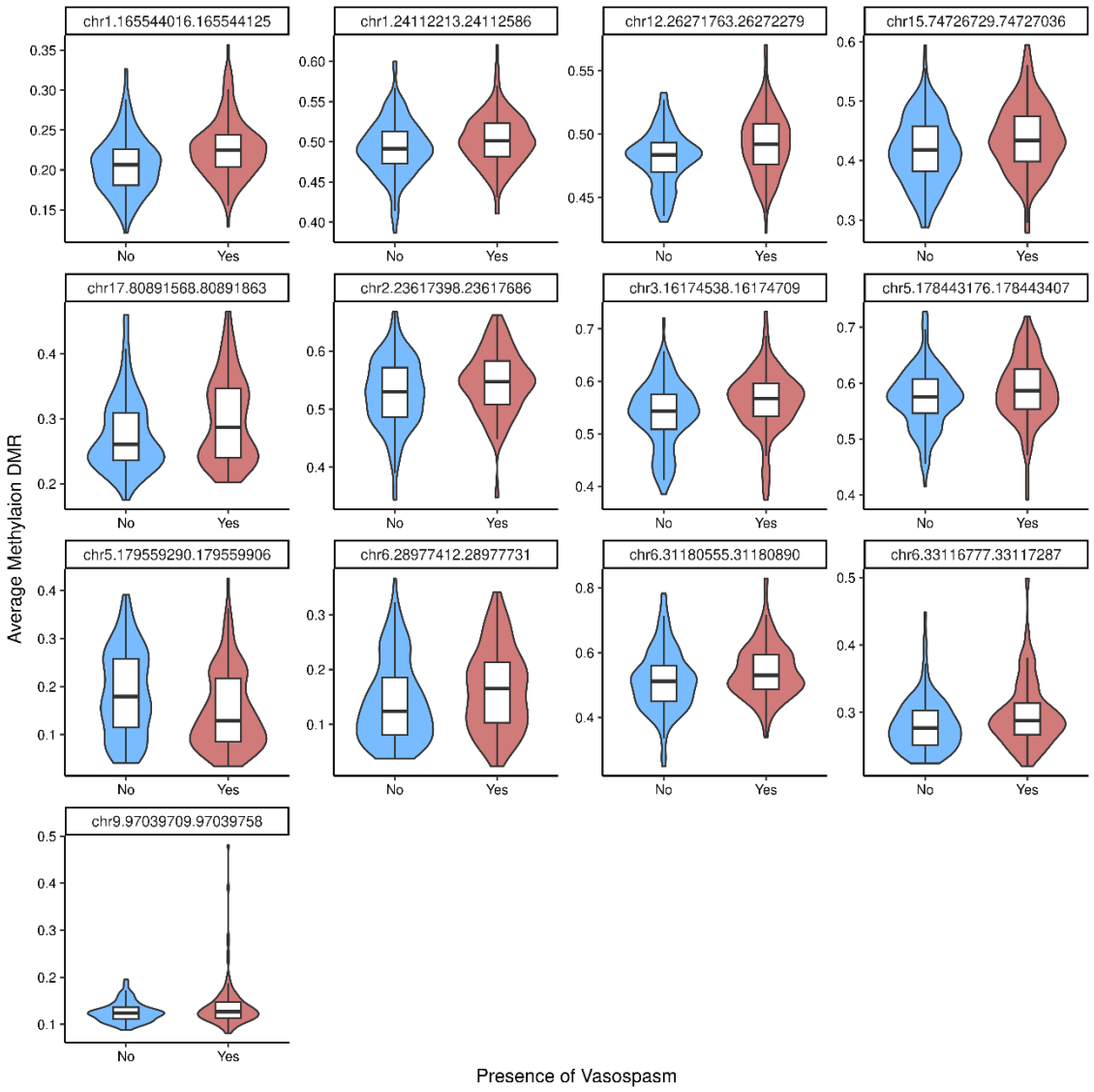
been obtained by discretizing methylation (β values) for each CpG according to quartiles distribution (light blue: first quartile; blue: second quartile; yellow: third quartile; red: fourth quartile).



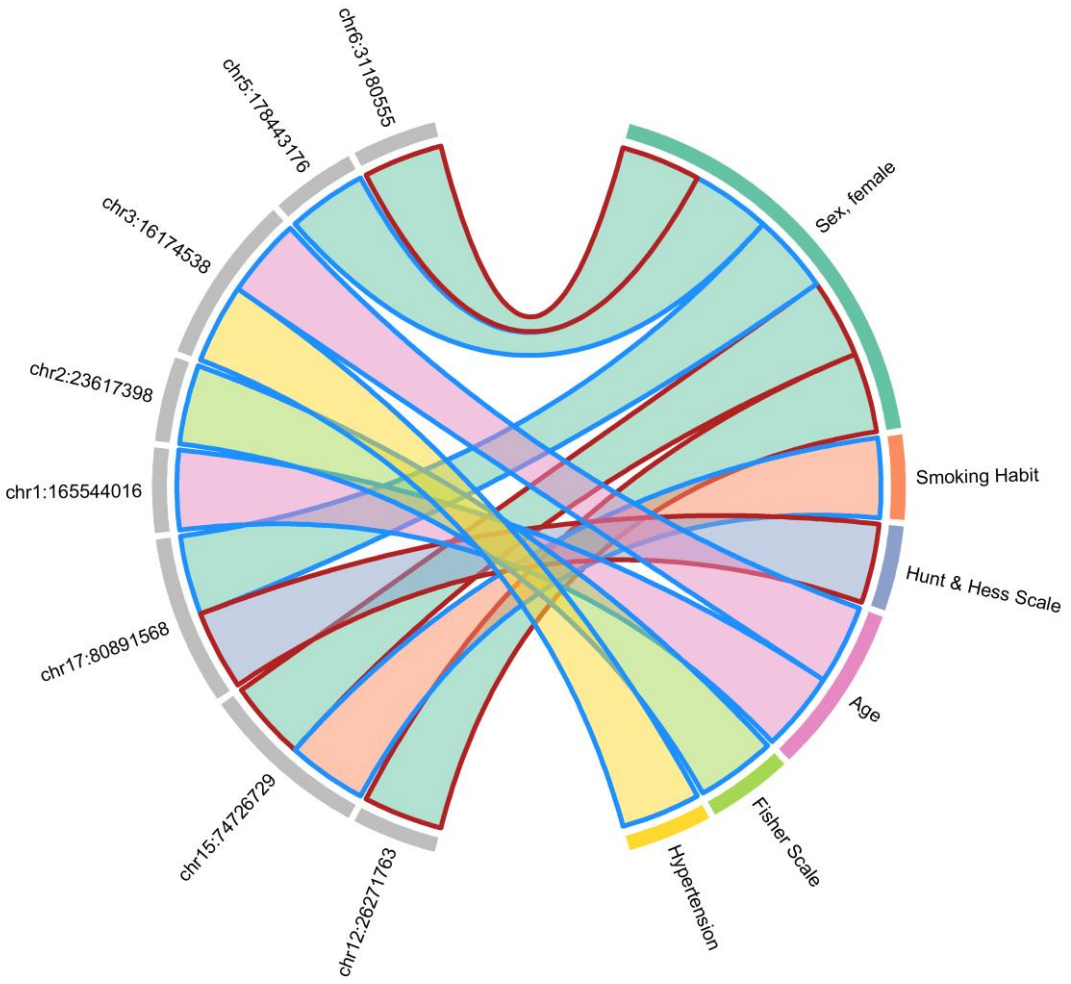

**Supplementary figure 3: Bootstrap of significant DMPs found in the EWAS** (table 2). We plotted the original β-coefficient of each CpG (dots) and the 99.9% confidence interval obtained after conducting the bootstrap (error bars). The vertical red dashed line corresponds to the absence of effect. Those CpGs crossing this line are not significant after doing the bootstrap, although no CpG was lost.



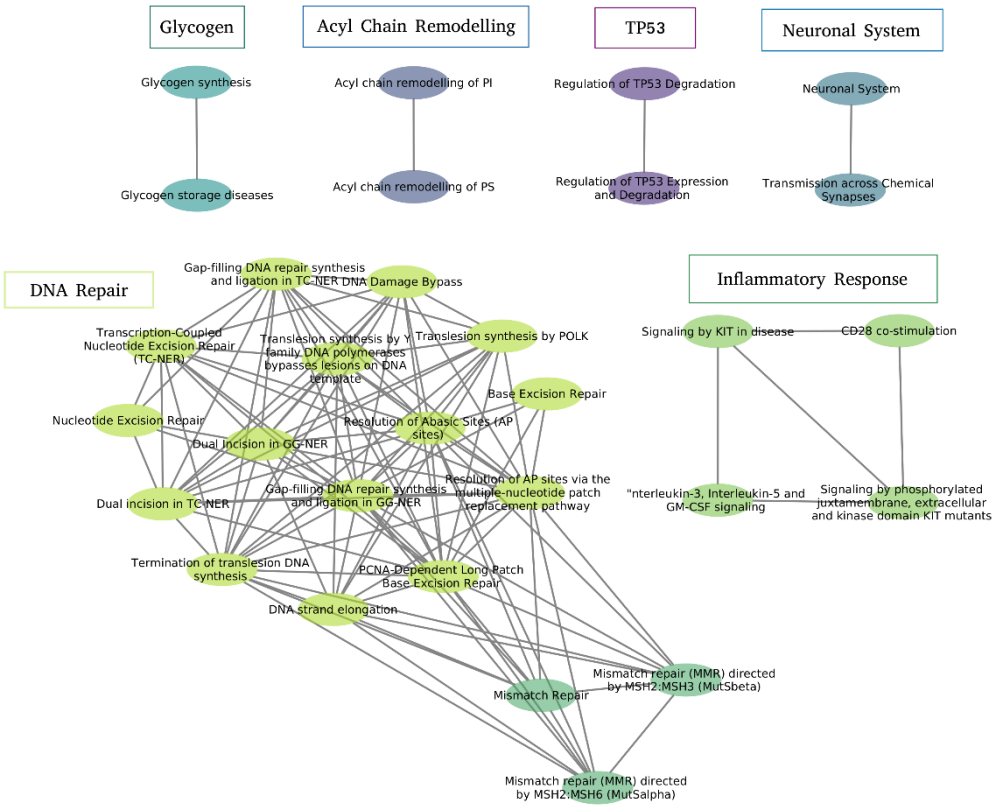
**Supplementary figure 4: Adjustment for test-statistic inflation.** Q-Q plot before and after doing test-statistic inflation adjustment (*Bacon*).




**Supplementary figure 5: Average methylation at significant DMRs by presence of vasospasm.** Violin plots comparing averaged methylation levels for each significant DMR between patients with and without vasospasm.




**Supplementary figure 6: Relationship between average methylation of significant DMRs (left-side) and main risk factors and complications for vasospasm (right-side).** DMRs are annotated according to the chromosome and the initial base-pair (table 3). In this figure we summarize significant relationships after adjusting for multiple testing (links). Border colors indicate a positive (red) or negative (blue) correlation. Filled colors represent distinct clinical factors.




**Supplementary figure 7: Gene set enrichment analysis.** We calculated the genes in common between all the possible combinations of significant pathways found in the gene set enrichment analysis (supplementary table 2). We then plotted the similarity of these sets as an acyclic graph. In this graph nodes represent pathways and edges that have been weighted according to the similarity coefficient (see the methods section). We filtered those nodes and edges having a similarity higher than 0.2. We finally clustered the graph using the autoannotate function (Cytoscape software). Each color represents an individual cluster.

## SUPPLEMENTARY TABLES

### Supplementary table 1

| DNA methylation Quality Controls in the Sample | |
|---|---|
| **Sample QCs (N=288)** | |
| Sex mismatch | 3 |
| Call rate < 98% | 0 |
| Deviated median intensities | 0 |
| **Final Sample size** | **285 (98.9%)** |
| **CpGs QCs (N=865918)** | |
| 1% samples detection *p*-value < .05 | 2940 |
| Beadcount < 3 in 5% samples | 3765 |
| Non-CpG probes | 2920 |
| SNP positions | 97039 |
| Multi-hit CpGs | 20 |
| CpGs in X and Y chromosomes | 16762 |
| **Final number of CpGs** | **742472** |

**Supplementary table 1:** In this table we summarize sample and CpG QCs. After applying all QCs, we kept 285 samples and 742,472 CpGs for the analyses.

Keywords: SNP, single-nucleotide polymorphism; QC, quality control.

## Supplementary table 2

### Gene set enrichment analysis of the epigenome-wide association study of vasospasm

| Database ID | Description | Path size | Normalized enrichment score | *p*-value | Q value |
|---|---|---|---|---|---|
| | **Gene Ontologies database** | | | | |
| GO:0031941 | Filamentous actin | 30 | 1.41 | 2.29E-06 | 0.0188 |
| GO:0032092 | Positive regulation of protein binding | 82 | 1.21 | 1.49E-05 | 0.0343 |
| GO:0045773 | Positive regulation of axon extension | 38 | 1.34 | 1.75E-05 | 0.0343 |
| GO:0030155 | Regulation of cell adhesion | 701 | 1.06 | 2.16E-05 | 0.0343 |
| GO:0005244 | Voltage-gated ion channel activity | 189 | 1.12 | 3.00E-05 | 0.0343 |
| GO:0022832 | Voltage-gated channel activity | 189 | 1.12 | 3.00E-05 | 0.0343 |
| GO:0045785 | Positive regulation of cell adhesion | 414 | 1.08 | 3.00E-05 | 0.0343 |
| GO:0048538 | Thymus development | 45 | 1.26 | 3.34E-05 | 0.0343 |
| GO:0030010 | Establishment of cell polarity | 135 | 1.14 | 4.46E-05 | 0.0407 |
| | **Reactome database** | | | | |
| R-HSA-110313 | Translesion synthesis by Y family DNA polymerases bypasses lesions on DNA template | 37 | 1.33 | 1.35E-05 | 0.0180 |
| R-HSA-9725371 | Nuclear events stimulated by ALK signaling in cancer | 18 | 1.44 | 2.39E-05 | 0.0180 |
| R-HSA-73933 | Resolution of Abasic Sites (AP sites) | 38 | 1.33 | 4.13E-05 | 0.0207 |
| R-HSA-73893 | DNA Damage Bypass | 45 | 1.26 | 5.58E-05 | 0.0207 |
| R-HSA-69190 | DNA strand elongation | 31 | 1.33 | 8.07E-05 | 0.0207 |
| R-HSA-5696397 | Gap-filling DNA repair synthesis and ligation in GG-NER | 25 | 1.35 | 8.29E-05 | 0.0207 |
| R-HSA-73884 | Base Excision Repair | 45 | 1.25 | 0.000106602 | 0.0207 |
| R-HSA-5656169 | Termination of translesion DNA synthesis | 31 | 1.32 | 0.000112242 | 0.0207 |
| R-HSA-5696400 | Dual Incision in GG-NER | 39 | 1.27 | 0.000123567 | 0.0207 |
| R-HSA-71291 | Metabolism of amino acids and derivatives | 338 | 1.08 | 0.00016211 | 0.0240 |
| R-HSA-6782135 | Dual incision in TC-NER | 63 | 1.20 | 0.000175713 | 0.0240 |
| R-HSA-9669938 | Signaling by KIT in disease | 20 | 1.34 | 0.000303037 | 0.0340 |
| R-HSA-9670439 | Signaling by phosphorylated juxtamembrane, extracellular and kinase domain KIT mutants | 20 | 1.34 | 0.000303037 | 0.0340 |
| R-HSA-422475 | Axon guidance | 520 | 1.06 | 0.000316707 | 0.0340 |
| R-HSA-6782210 | Gap-filling DNA repair synthesis and ligation in TC-NER | 62 | 1.20 | 0.000357719 | 0.0349 |

| | | | | | |
|---|---|---|---|---|---|
| R-HSA-110373 | Resolution of AP sites via the multiple-nucleotide patch replacement pathway | 25 | 1.31 | 0.00037139 | 0.0349 |
| R-HSA-389356 | CD28 co-stimulation | 32 | 1.27 | 0.000426072 | 0.0352 |
| R-HSA-202433 | Generation of second messenger molecules | 29 | 1.28 | 0.000462527 | 0.0352 |
| R-HSA-8943724 | Regulation of PTEN gene transcription | 59 | 1.19 | 0.000471686 | 0.0352 |
| R-HSA-5696398 | Nucleotide Excision Repair | 107 | 1.13 | 0.000490003 | 0.0352 |
| R-HSA-1482922 | Acyl chain remodelling of PI | 16 | 1.37 | 0.0005455 | 0.0352 |
| R-HSA-5358508 | Mismatch Repair | 15 | 1.37 | 0.000554668 | 0.0352 |
| R-HSA-112316 | Neuronal System | 387 | 1.06 | 0.00058159 | 0.0352 |
| R-HSA-5655862 | Translesion synthesis by POLK | 17 | 1.37 | 0.000609676 | 0.0352 |
| R-HSA-5358565 | Mismatch repair (MMR) directed by MSH2:MSH6 (MutSalpha) | 14 | 1.36 | 0.000618843 | 0.0352 |
| R-HSA-5358606 | Mismatch repair (MMR) directed by MSH2:MSH3 (MutSbeta) | 14 | 1.36 | 0.000618843 | 0.0352 |
| R-HSA-6804757 | Regulation of TP53 Degradation | 36 | 1.24 | 0.000645701 | 0.0352 |
| R-HSA-6806003 | Regulation of TP53 Expression and Degradation | 37 | 1.24 | 0.00065486 | 0.0352 |
| R-HSA-1474244 | Extracellular matrix organization | 292 | 1.07 | 0.000755606 | 0.0380 |
| R-HSA-450604 | KSRP (KHSRP) binds and destabilizes mRNA | 17 | 1.36 | 0.00076553 | 0.0380 |
| R-HSA-1170546 | Prolactin receptor signaling | 13 | 1.38 | 0.000783866 | 0.0380 |
| R-HSA-3322077 | Glycogen synthesis | 15 | 1.35 | 0.000811369 | 0.0382 |
| R-HSA-112315 | Transmission across Chemical Synapses | 255 | 1.07 | 0.001076885 | 0.0471 |
| R-HSA-3229121 | Glycogen storage diseases | 14 | 1.35 | 0.001077962 | 0.0471 |
| R-HSA-5651801 | PCNA-Dependent Long Patch Base Excision Repair | 21 | 1.29 | 0.001095293 | 0.0471 |
| R-HSA-1482801 | Acyl chain remodelling of PS | 22 | 1.31 | 0.001168925 | 0.0477 |
| R-HSA-6781827 | Transcription-Coupled Nucleotide Excision Repair (TC-NER) | 76 | 1.15 | 0.001187332 | 0.0477 |
| R-HSA-512988 | Interleukin-3, Interleukin-5 and GM-CSF signaling | 45 | 1.19 | 0.00120574 | 0.0477 |
| R-HSA-71387 | Metabolism of carbohydrates | 278 | 1.07 | 0.001242556 | 0.0479 |

**Supplementary table 2**. Using the EWAS summary statistics (table 2) we studied the gene-set enrichment for several databases: Gene-Ontologies, Reactome and KEGG. In this table we filtered those gene-sets with a Q-value < 0.05. For each gene-set (rows) we show the name, the number of annotated gene, the enrichment score, *p*-value and Q-value.