PPM1A Methylation Is Associated With Vascular Recurrence in Aspirin-Treated Patients

Cristina Gallego-Fabrega, MSc; Caty Carrera, MD, MSc; Jean-Luc Reny, MD, PhD; Pierre Fontana, MD, PhD; Agnieszka Slowik, MD, PhD; Joanna Pera, MD, PhD; Alessandro Pezzini, MD; Gemma Serrano-Heras, PhD; Tomáš Segura, MD, PhD; Abdul-Aziz A. Bin Dukhyil, PhD; Joan Martí-Fabregas, MD, PhD; Elena Muñoz, MD; Natalia Cullell, MSc; Joan Montaner, MD, PhD; Jerzy Krupinski, MD, PhD; Israel Fernandez-Cadenas, PhD

Background and Purpose—Despite great efforts by pharmacogenetic studies, the causes of aspirin failure to prevent the recurrence of ischemic events remain unclear. Our aim was to study whether epigenetics could be associated with the risk of vascular recurrence in aspirin-treated stroke patients.

Methods—We performed an epigenetic joint analysis study in 327 patients treated with aspirin. In the discovery stage, we performed a nested case–control study in 38 matched ischemic stroke patients in whom 450,000 methylation sites were analyzed. Nineteen patients presented vascular recurrence after stroke, and 19 matched patients did not present vascular recurrence during the first year of follow-up. In a second stage, 289 new patients were analyzed by EpiTYPER.

Results—The following 3 differentially methylated candidate CpG sites, were identified in the discovery stage and analyzed in the second stage: cg26039762 (P=9.69×10^{-6}, RAI), cg04985020 (P=3.47×10^{-3}, PPM1A), and cg08419850 (P=3.47×10^{-3}, KCNQ1). Joint analysis identified an epigenome-wide association for cg04985020 (PPM1A; P=1.78×10^{-4}), with vascular recurrence in patients treated with aspirin.

Conclusions—The pattern of differential methylation in PPM1A is associated with vascular recurrence in aspirin-treated stroke patients. (Stroke. 2016;47:1926-1929. DOI: 10.1161/STROKEAHA.116.013340.)

Key Words: aspirin □ methylation □ phenotype □ platelet aggregation □ stroke

Ischemic stroke patients are at high risk of having new cardio-vascular events. The cumulative risk of vascular recurrence after a first stroke reaches up to 48% at 10 years. Acetylsalicylic acid (aspirin) and clopidogrel are the most widely used antiplatelet agents in secondary prevention of stroke. Aspirin has been associated with a relative risk reduction of 25% in secondary cardiovascular events and 13% to 15% in secondary prevention of stroke. A high platelet reactivity phenotype can be present in ≤60% of aspirin-treated patients and has been associated with an increased risk of recurrent ischemic events, and although it is a substantially heritable phenotype, very few genetic determinant have been identified. Furthermore, the causes of aspirin failure goes beyond high platelet reactivity and are still largely unknown.

Only polymorphisms in PEAR1 have been confirmed as genetic risk factors of aspirin failure. The rs12041331 polymorphism is associated with platelet aggregation, increased platelet reactivity, and vascular recurrence in aspirin-treated patients, although these polymorphisms do not explain the whole variability observed in aspirin resistance or in vascular responsiveness. Other genomic regulations, such as epigenetics modifications could be associated with the risk of vascular recurrence in patients treated with aspirin.

We aimed to analyze the whole epigenome of stroke patients treated with aspirin, to find altered methylation sites associated with vascular recurrence.
Materials and Methods

Study Patients
Thirty-eight subjects from a cohort of 1900 ischemic stroke patients recruited prospectively at Vall d’Hebron Hospital (Spain), who started aspirin treatment after a first ischemic stroke, were analyzed. Nineteen participants presented a new vascular event (ischemic stroke, myocardial infarction, peripheral vascular disease, or cardiovascular death) during the first year of follow-up and were matched one-to-one (age, sex, TOAST [Trial of Org 10172 in Acute Stroke Treatment]) with 19 participants without a vascular event. Second-stage analysis was performed in 289 prospective enrolled participants from 4 independent cohorts with ischemic stroke or ischemic atherothrombotic disease; 29 presented a vascular recurrence within 1 year of follow-up (Table I in the online-only Data Supplement).

Discovery: HumanMethylation450 Assay
DNA was extracted from blood samples using standard methods. Samples were obtained during the first 24 hours after stroke onset, before initial aspirin administration.

Genome-wide DNA methylation (DNAm) was assessed using the HumanMethylation450 (Illumina) assay. All samples (n=38) were processed in a single working batch. Quality Control and differentially methylated CpG (DMC) sites analysis were performed as described elsewhere, using the R computing environment (3.1.3 version; Table II in the online-only Data Supplement).

Rs120411331 polymorphism (PEAR1) was checked to discover whether it could be a confounding factor in the EWAS (Epigenome-Wide Association Study) analysis.

CpGs Selection
One DMC with epigenome-wide significance (RAF1; P<10^-6) was selected for further replication. Candidate DMCs for second-stage analysis were defined as: top 100 significant DMCs from the univariate analysis, which also appear in the top 100 DMCs after multivariable analysis. Multivariable analyses were adjusted for principal components and DNAm potential covariates (age, sex, and current smoking). From a total of 36 candidate CpGs, those mapped on genes previously associated with cardiovascular disease were considered for further analysis. A PubMed search (http://www.ncbi.nlm.nih.gov/pubmed) was undertaken using the following key words: aspirin, atherosclerosis, cardiovascular, stroke, and vascular. Finally, from 4 candidate DMCs, only 2 CpGs (PPM1A and KCNQ1) were located in regions suitable for EpiTYPER analysis (Figure IV in the online-only Data Supplement).

Figure. A, Manhattan plot for the EWAS (Epigenome-Wide Association Study) of aspirin-treated stroke patients. x axis: chromosome position; y axis: −log10 of P values. B, Hierarchical cluster analysis of most significant differentially methylated CpG associated (Table) with vascular recurrence. CA indicates recurrence patients; and CO, nonrecurrence patients.
Second-Stage EpiTYPER Assay
An averaged 400-bp region was sequenced using EpiTYPER (Sequenom) around each CpG selected from the discovery stage in 289 new patients.

Statistical Analysis
DMC sites were calculated using the Mann–Whitney U test and multivariable generalized linear analysis. Sample size calculation indicated that 19 subjects per condition were needed to achieve a $10^{-05}$ significance level and 80% statistical power. A joint analysis strategy using the DerSimonian–Laid test was performed to increase the statistical power in the 2-stage study as previously suggested.

Results

Discovery
DNAm levels were obtained for 485 577 CpGs sites (Figure [A]). After data preprocessing and quality control analysis one DMC was associated with vascular recurrence at epigenome-wide level ($P<10^{-46}$), covariant adjustment analysis was also performed (Table; Table IV in the online-only Data Supplement). Table IV in the online-only Data Supplement shows the 36 candidate CpG sites which were best ranked in both univariate and covari-ante analyses, including cg04985020 (PPM1A) and cg08419850 (KCNQ1). Figure (B) shows a hierarchical cluster analysis, which was able to distinguish vascular recurrence patients from nonvascular patients, with only 2 samples misclassified.

Second-Stage Analysis
Second-stage analysis was performed for the epigenome-wide significant CpG, cg26039762 (RAF1; $P=9.69 \times 10^{-49}$), and 2 additional candidate CpG sites, cg04985020 (PPM1A; $P=3.46 \times 10^{-40}$) and cg08419850 (KCNQ1; $P=3.46 \times 10^{-44}$). All 3 CpGs map genes involved in atherosclerosis and vascular process. The cg04985020 (PPM1A) was associated with vascular recurrence ($P=0.027$). The same cg04985020 methylation pattern was observed in each of the 4 validation-stage cohorts and when nonstroke and cardioembolic patients were excluded from the analyses (Materials and Methods section in the online-only Data Supplement). Additionally, the following 3 CpG sites surrounding cg26039762 (RAF1) were de novo associated with vascular recurrence: RAR1_10 ($P=0.0015$), RAF1_6 ($P=0.0043$), and RAF1_3 ($P=0.01$; Table VI in the online-only Data Supplement). However, cg26039762 (RAF1) and cg08419850 (KCNQ1) were not associated in this second stage.

Joint analysis of discovery and replication stage revealed an epigenome-wide statistically significant meta $P$ value for PPM1A (meta $P=1.78 \times 10^{-40}$) but not for RAF1. PPM1A was independently associated with vascular recurrence when analyzed possible DNAm confounding factors (age, sex, and current smoking; $P=5.74 \times 10^{-30}$) and also PEAR1 polymorphism (rs12041331; $P=9.82 \times 10^{-40}$). Additionally, PPM1A methylation levels were not influenced by cell-type proportions.

Discussion
Patients with vascular recurrence presented higher methylation levels of cg04985020 CpG (PPM1A) compared with nonrecurrent patients. Despite cg04985020 not being significant on the epigenome-wide level, in the discovery stage ($P=3.46 \times 10^{-40}$), the joint analysis revealed a meta $P$ value of $1.78 \times 10^{-40}$ that was significant on the epigenome-wide level.7 These results suggest an association between DNAm levels of PPM1A and vascular recurrence during on-aspirin treatment, regardless of the disease (cardiovascular or stroke). The association was not confounded by common DNAm confounding factors or PEAR1 polymorphism. The cg26039762 (RAF1) was not replicated, but 3 surrounding CpGs were associated de novo with vascular recurrence. Patients with vascular recurrence presented lower methylation levels of these 3 surrounding CpGs compared with nonvascular recurrence patients, similar to cg26039762 (RAF1) in the discovery stage. Additionally, hierarchical cluster analysis showed a potential role of DMCs as a model to classify patients within highest or lowest risk of vascular recurrence.

Protein phosphatase magnesium dependent 1A (PPM1A) is involved in the regulation of transforming growth factor (TGF)-β1 signaling and plasminogen activator inhibitor-1 transcrip- tion.8 Plasminogen activator inhibitor-1 levels are associated with increased risk of cardiovascular events, particularly in the context of elevated tissue TGF-β1.10 Raf1 proto-oncogene, serine/threonine kinase (RAF1), has been described in many vascular diseases, such as vascular dementia, angiogenesis processes, or cardiovascular events.11 RAF1 has also been described as taking part in signal transduction in the TGF-β1/Smad pathway.12 Furthermore, aspirin administration has been associated with decreasing TGF-β1 serum levels in hypercholesterolemic rats.13 We hypothesize that PPM1A- and RAF1-altered methylation could be associated with vascular...
recurrence because of the regulation of the TGF-ß pathway; however, further studies are needed to confirm this hypothesis.

Pharmacoepigenomics is a growing field that could be used in the future to assist in personalized antiplatelet therapy or to find potential new drug targets. Our study suggests a remarkable role for epigenetics in the modulation of aspirin response.

Limitations
The small sample size in the discovery stage impeded the achievement of more significant values; however, joint analysis was useful to improve the results of our study. mRNA studies are needed to confirm the biological significance of our results.

Sources of Funding
This study was supported by Miguel Servet grant (CP12/03298, Instituto de Salud Carlos III and Fondo Europeo de Desarrollo Regional), Sheikh Abdullah Bin Abdul Mohsen Al Tuwaijri Chair for Applied Research in Stroke, Majmaah University, Saudi Arabia, and SEDMAN Project. J. Pera is supported by the Jagiellonian University Medical College by grant no. KZDS/003844. J. Martí-Fabregas is supported by Redes Temáticas de Investigación Cooperativa en Salud RD12-0014-0002.

Disclosures
P. Fontana discloses honoraria from Bayer. The other authors report no conflicts.

References
PPM1A Methylation Is Associated With Vascular Recurrence in Aspirin-Treated Patients
Cristina Gallego-Fabrega, Caty Carrera, Jean-Luc Reny, Pierre Fontana, Agnieszka Slowik, Joanna Pera, Alessandro Pezzini, Gemma Serrano-Heras, Tomás Segura, Abdul-Aziz A. Bin Dukhyil, Joan Martí-Fàbregas, Elena Muño, Natalia Cullell, Joan Montaner, Jerzy Krupinski and Israel Fernandez-Cadenas

Stroke. 2016;47:1926-1929; originally published online June 14, 2016;
doi: 10.1161/STROKEAHA.116.013340
Stroke is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2016 American Heart Association, Inc. All rights reserved.
Print ISSN: 0039-2499. Online ISSN: 1524-4628

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://stroke.ahajournals.org/content/47/7/1926

Data Supplement (unedited) at:
http://stroke.ahajournals.org/content/suppl/2016/06/27/STROKEAHA.116.013340.DC1

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Stroke can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to Stroke is online at:
http://stroke.ahajournals.org/subscriptions/
Supplemental Material

Materials and Methods

Clinical protocol

From a cohort of 1,900 stroke patients from Vall d’Hebron University Hospital (Barcelona, Spain), 38 subjects were selected. 19 ischemic stroke patients treated with aspirin with a new vascular event (new ischemic stroke, myocardial infarction, peripheral vascular disease and cardiovascular death) and 19 ischemic stroke patients treated with aspirin without a new vascular event were matched one-to-one by age (±7 years), sex, and TOAST\(^1\) classification (Table I).

Cases were defined as ischemic stroke patients under aspirin treatment which had vascular recurrence in the first year of follow up with a good adherence to the treatment measured by the Morisky-Green test\(^2\), controls were defined as ischemic stroke patients under aspirin treatment with a good adherence to the treatment measured using the Morisky-Green test and without vascular recurrence during the first year of follow up. Vascular recurrence was described as new ischemic stroke, myocardial infarction, peripheral vascular disease or cardiovascular death and was detected by phone calls every three months or direct clinical reporting.

Second stage analysis was performed on 289 new samples from 4 international cohorts, three ischemic stroke patients’ cohorts and one cardiovascular disease patients’ cohort. The second stage cohort included 29 patients with a new vascular event and 261 patients without a new vascular event during the first year of follow up (Table II). The Italian, Spanish and Polish cohorts consist of consecutively recruited patients that started aspirin treatment after the first ischemic stroke. Vascular recurrence information was available from all patients. The Geneva cohort consists of consecutively recruited patients with symptomatic documented ischemic atherothrombotic disease (coronary artery disease [CAD], ischemic cerebrovascular disease and/or peripheral artery disease) treated with aspirin and/or clopidogrel for < 5 years. Information about clinical ischemic events over an ongoing 3-year follow-up was available from all patients. Only aspirin treated patients were selected from this cohort.

The local ethical committee approved the study (PR(AG) 03/2007).

Discovery, HumanMethylation450 assay

Genome-wide DNA methylation was assessed using the Infinium HumanMethylation450 BeadChip (Illumina Inc., San Diego Ca). All samples were processed in a single working batch.
All pre-processing, correction and normalization steps were implemented using the R statistical computing environment (3.1.3 version) with Bioconductor packages, (Table III). Plots were also produced using R functions. Quality control metrics were examined to determine the success of the bisulphite conversion and subsequent array hybridization. Fluorescence intensities were imported from GenomeStudio, then probe filtering was performed to remove probes that have failed to hybridize (detection p-value>0.05) and that were not represented by a minimum of 3 beads on the array, as described elsewhere. CpG sites containing documented single nucleotide polymorphisms (SNPs) were also excluded. Multidimensional scaling (MDS) plots were used to evaluate gender outliers based on chromosome X data. MDS and PC were also used to check unknown population structures within the sample. Then, probes mapping to sex chromosomes were removed. We also checked the white cell count (neutrophils, lymphocytes and monocytes) as a possible confounding factor. Finally, a subset quantile normalization was performed using a background adjustment between-array normalization and a dye bias correction, following previous recommendations.

The methylation level of each cytosine was expressed as a Beta value (β-value), ranging between 0 and 1, unmethylated to completely methylated respectively. Differentially methylated CpG sites (DMCs) were analyzed using the non-parametric Mann-Whitney U test for independent samples, p-values<10^-06 were selected as statistically significant and p-values<10^-05 as having nominal association. Multivariable generalized linear analyses adjusting for Principal Components and DNA methylation potential covariates (age, sex and current smoking) were also used.

Second stage, EpiTYPER assay.

Quantitative DNA methylation analysis was performed using the MassARRAY EpiTYPER (Sequenom, San Diego, CA, USA), on selected CpGs. An averaged 400 pb region was sequenced around each selected CpG, containing multiple CpG sites and aggregated CpG sites. Target-specific primers were designed using the online Epidesigner software (http://www.epidesigner.com), table V. The quantitative methylation data obtained for each CpG site, or aggregates of multiple CpG sites, were analyzed with the EpiTYPER software (Sequenom). Statistical analyses were performed using R (3.1.3 version). P-values<0.05 were considered as statistically significant, after the Mann-Whitney U test.

PEAR1 analysis.

OmniQuad Human 1M array (Illumina) was used to genotype the samples. Data was analyzed following WTCC recommendation and the specific methodological guidelines from the Broad Institute and Harvard University. Quality controls (QC) were performed before the
analyses: Missings (-0.3) and checksex QC_s were performed for each individual. Only individuals with Caucasian ethnicity were included. We also tested for population stratifications by analyzing the individual IBS, performing MDS plots and adjusting for principal components (PCs). Hardy-Weinberg (p-value>10^{-08}), missingness (0.01), Mishap (p-value>10^{-09}), were analyzed for each SNP.

After the QC process, the genotypic results were analyzed by Plink, Haploview, STATA, SNPtest and GTOOL software solutions.

Statistical analysis.

Sample size calculation was performed using the pwr package (version: 1.1-2) from Bioconductor (www.bioconductor.org). Nineteen subjects per condition were needed in order to achieve a 10^{-05} significance level and 80% statistical power, considering a Cohen's effect size = 0.9.

Results

Epigenome-wide analysis.

DNA methylation levels were obtained for 485.577 CpG sites across the whole genome. After pre-processing and QC analysis, 34.999 probes and three sample were removed from further analysis: 1.209 CpGs with detection p-value>0.05, 1.765 CpGs with a beadcount lower than 3 in more than 5% of the samples, 20.935 CpGs overlapping with SNPs and 11.090 CpGs located on sex chromosomes. One samples was removed because 1% of the CpG sites had detection p-values>0.05 (Figure I.A) and two more samples were removed because they showed sex discrepancies (Figure I.B).

A heatmap of the 16 differentially methylated CpGs is shown in the supplemental material, Figure II.

Second stage analysis.

Higher methylation levels of cg04985020 (PPM1A) were observed in vascular recurrence patients treated with aspirin in each of the four validation-stage cohorts. The same pattern was observed after excluding the Geneva's non-stroke patients and the cardioembolic stroke patients, although the association was not statistically significant due to a reduction of the sample size (Figure III). Indicating that cg04985020 (PPM1A) association with vascular recurrence may be independent of the cardiovascular disease.

Further analysis
Frequency distribution of the rs12041331 risk allele was equally distributed between patients (6 vascular recurrent patients) and controls (2 non-vascular recurrence patients) p-value=0.076.

The Laboratory of Stroke Pharmacogenomics and Genetics is part of the International Stroke Genetics Consortium (ISGC, www.strokegenetics.com) and coordinates the Spanish Stroke Genetics Consortium (Genestroke, www.genestroke.com). I. F-C. is supported by the Miguel Servet programme (CP12/03298), Instituto de Salud Carlos III. We thank Dr. Raid S Al Baradie for providing the necessary support.
References


Tables.

**Table I:** Descriptive characteristics of the study population. Discovery cohort n=38, validation cohort n=289.

<table>
<thead>
<tr>
<th></th>
<th>Vascular Recurrence</th>
<th>Non-Vascular Recurrence</th>
<th>Vascular Recurrence</th>
<th>Non-Vascular Recurrence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Discovery Stage Cohort (n= 38)</td>
<td>Second Stage Cohort (n= 289)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>N</strong></td>
<td>19 (50%)</td>
<td>19 (50%)</td>
<td>29 (10%)</td>
<td>260 (90%)</td>
</tr>
<tr>
<td><strong>Age (years)</strong></td>
<td>69.2±12.9</td>
<td>68.9±12.8</td>
<td>69±10</td>
<td>65±12</td>
</tr>
<tr>
<td><strong>Male</strong></td>
<td>17 (89.4%)</td>
<td>17 (89.4%)</td>
<td>17 (58.5%)</td>
<td>183 (69.5%)</td>
</tr>
<tr>
<td><strong>Female</strong></td>
<td>2 (10.5%)</td>
<td>2 (10.5%)</td>
<td>12 (41.3%)</td>
<td>77 (29.3%)</td>
</tr>
<tr>
<td><strong>TOAST</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Atherothrombotic</em></td>
<td>7 (21%)</td>
<td>7 (21%)</td>
<td>1 (3.4%)</td>
<td>12 (4.6%)</td>
</tr>
<tr>
<td><em>Lacunar</em></td>
<td>3 (15.8%)</td>
<td>3 (15.8%)</td>
<td>1 (3.4%)</td>
<td>10 (3.8%)</td>
</tr>
<tr>
<td><em>Cardioembolic</em></td>
<td>-</td>
<td>-</td>
<td>6 (20.6%)</td>
<td>20 (7.6%)</td>
</tr>
<tr>
<td><em>Undetermined</em></td>
<td>9 (47.3%)</td>
<td>9 (47.3%)</td>
<td>10 (34.4%)</td>
<td>38 (14.4%)</td>
</tr>
<tr>
<td><em>Other</em></td>
<td></td>
<td></td>
<td>1 (3.4%)</td>
<td>9 (3.4%)</td>
</tr>
<tr>
<td><strong>Dyslipidemia</strong></td>
<td>11 (57.9%)</td>
<td>6 (31.6%)</td>
<td>6 (20.6%)</td>
<td>90 (34.2%)</td>
</tr>
<tr>
<td><strong>Diabetes Mellitus</strong></td>
<td>8 (42.1%)</td>
<td>3 (15.8%)</td>
<td>12 (41.3%)</td>
<td>65 (24.7%)</td>
</tr>
<tr>
<td><strong>Hypertension</strong></td>
<td>11 (57.9%)</td>
<td>11 (57.9%)</td>
<td>21 (72.2%)</td>
<td>154 (58.5%)</td>
</tr>
<tr>
<td><strong>Current Smoker</strong></td>
<td>4 (21%)</td>
<td>5 (31.6%)</td>
<td>8 (27.5%)</td>
<td>67 (25.5%)</td>
</tr>
<tr>
<td><strong>Alcohol intake</strong></td>
<td>7 (21%)</td>
<td>4 (21%)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table II: Descriptive characteristics of each international cohort of the second stage analysis.

<table>
<thead>
<tr>
<th></th>
<th>Geneva's Cohort (178)</th>
<th>Polish's Cohort (59)</th>
<th>Italy's Cohort (26)</th>
<th>Spanish's Cohort (27)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Rec</td>
<td>Non-Rec</td>
<td>Rec</td>
<td>Non-Rec</td>
</tr>
<tr>
<td>N</td>
<td>10 (5.6%)</td>
<td>168 (94.4%)</td>
<td>13 (21.9%)</td>
<td>46 (77.7%)</td>
</tr>
<tr>
<td>Age (years)</td>
<td>69±9</td>
<td>64±11</td>
<td>70±10</td>
<td>66±12</td>
</tr>
<tr>
<td>Male</td>
<td>8 (80%)</td>
<td>127 (74.9%)</td>
<td>7 (53.8%)</td>
<td>24 (52.2%)</td>
</tr>
<tr>
<td>Female</td>
<td>2 (20%)</td>
<td>41 (24.2%)</td>
<td>6 (46.1%)</td>
<td>22 (47.7%)</td>
</tr>
<tr>
<td>TOAST</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aterothrombotic</td>
<td>- 32 (19%)</td>
<td>- 2</td>
<td>1 7 (50%)</td>
<td>29.1%</td>
</tr>
<tr>
<td>Lacunar</td>
<td>- 1</td>
<td>5 (7.7%)</td>
<td>10,9%</td>
<td>- 12.5%</td>
</tr>
<tr>
<td>Cardioembolic</td>
<td>-</td>
<td>5 19 (38.5%)</td>
<td>41,23%</td>
<td></td>
</tr>
<tr>
<td>Undetermined</td>
<td>-</td>
<td>7 19 (53.8%)</td>
<td>41,23%</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>-</td>
<td>1 (2.27%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dyslipidemia</td>
<td>5 88 (50%)</td>
<td>1 39 (51.9%)</td>
<td>- 8 20</td>
<td>2 3 (25%)</td>
</tr>
<tr>
<td>Diabetes Mellitus</td>
<td>7 101 (10%)</td>
<td>101 (61.5%)</td>
<td>(43.4%)</td>
<td>(100%)</td>
</tr>
<tr>
<td>Hypertension</td>
<td>(70%) 59.66%</td>
<td>69.2% (75.9%)</td>
<td>(100%)</td>
<td>(45.8%)</td>
</tr>
<tr>
<td>Current Smoker</td>
<td>4 42 (40%)</td>
<td>12 (23.1%)</td>
<td>12 (26%)</td>
<td></td>
</tr>
</tbody>
</table>

Table III: Bioconductor packages for the processing and analysis of array-based DNA methylation data.

<table>
<thead>
<tr>
<th>DNA methylation processing/analysis step</th>
<th>Bioconductor packages</th>
<th>Commands</th>
</tr>
</thead>
<tbody>
<tr>
<td>Methylation data loading</td>
<td>MethyLumi</td>
<td>methylumiR()</td>
</tr>
<tr>
<td>Quality control sample/probe</td>
<td>waterRmelon, minfi</td>
<td>pfilter()</td>
</tr>
<tr>
<td></td>
<td></td>
<td>mdsPlot()</td>
</tr>
<tr>
<td>Normalization and background correction</td>
<td>waterRmelon</td>
<td>dasen()</td>
</tr>
</tbody>
</table>
**Table IV**: DMC sites associated with vascular recurrence (p-value<10^{-5}). List of the 36 candidate DMC sites which were best ranked in both, raw and covariate analysis. Six CpGs, highlighted in bold letters, were initially selected for further analysis, even though finally only RAF1, PPM1A and KCQ1 CpG sites were suitable for MassARRAY EpiTYPER analysis.

<table>
<thead>
<tr>
<th>CpG ID</th>
<th>Chr</th>
<th>Position</th>
<th>Gene</th>
<th>p-value</th>
<th>Covar. p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>cg26039762</td>
<td>3</td>
<td>13177788</td>
<td>RAF1</td>
<td>9,69E-06</td>
<td>0,01815912</td>
</tr>
<tr>
<td>cg0094487</td>
<td>15</td>
<td>65255928</td>
<td>SPG21</td>
<td>1,20E-05</td>
<td>0,00493098</td>
</tr>
<tr>
<td>cg26568880</td>
<td>6</td>
<td>31926569</td>
<td>RDBP; SKIV2L</td>
<td>1,47E-05</td>
<td>0,00253664</td>
</tr>
<tr>
<td>cg20702204</td>
<td>12</td>
<td>133430077</td>
<td>CHFR</td>
<td>2,19E-05</td>
<td>0,00274526</td>
</tr>
<tr>
<td>cg22352818</td>
<td>2</td>
<td>97193289</td>
<td></td>
<td>2,19E-05</td>
<td>0,00345245</td>
</tr>
<tr>
<td>cg01112035</td>
<td>2</td>
<td>105200688</td>
<td></td>
<td>2,19E-05</td>
<td>0,08986818</td>
</tr>
<tr>
<td>cg16966962</td>
<td>1</td>
<td>201915989</td>
<td>LMOD1</td>
<td>3,23E-05</td>
<td>0,00815474</td>
</tr>
<tr>
<td>cg09747456</td>
<td>10</td>
<td>91405300</td>
<td>PANK1</td>
<td>3,90E-05</td>
<td>0,0031413</td>
</tr>
<tr>
<td>cg07580707</td>
<td>3</td>
<td>14220061</td>
<td>XPC; LSM3</td>
<td>4,69E-05</td>
<td>0,00633959</td>
</tr>
<tr>
<td>cg02533998</td>
<td>12</td>
<td>101540053</td>
<td></td>
<td>5,62E-05</td>
<td>0,29100396</td>
</tr>
<tr>
<td>cg00932677</td>
<td>4</td>
<td>187776068</td>
<td></td>
<td>6,72E-05</td>
<td>0,00305652</td>
</tr>
<tr>
<td>cg08859247</td>
<td>17</td>
<td>17286854</td>
<td></td>
<td>6,72E-05</td>
<td>0,00848049</td>
</tr>
<tr>
<td>cg00608660</td>
<td>16</td>
<td>8658622</td>
<td>MTHFS; FLJ30679</td>
<td>6,72E-05</td>
<td>0,08698691</td>
</tr>
<tr>
<td>cg16933664</td>
<td>11</td>
<td>67085326</td>
<td>LOC100130987</td>
<td>8,02E-05</td>
<td>0,00328628</td>
</tr>
<tr>
<td>cg08088677</td>
<td>2</td>
<td>6911141</td>
<td></td>
<td>8,02E-05</td>
<td>0,00630309</td>
</tr>
<tr>
<td>cg26577201</td>
<td>2</td>
<td>219857793</td>
<td>CRYBA2</td>
<td>8,02E-05</td>
<td>0,00745067</td>
</tr>
</tbody>
</table>

**36 candidate DMCs best ranked in both, univariate and covariate analysis**

<table>
<thead>
<tr>
<th>CpG ID</th>
<th>Chr</th>
<th>Position</th>
<th>Gene</th>
<th>p-value</th>
<th>Covar. p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>cg0094487</td>
<td>15</td>
<td>65255928</td>
<td>SPG21</td>
<td>1,20E-05</td>
<td>0,00493098</td>
</tr>
<tr>
<td>cg26568880</td>
<td>6</td>
<td>31926569</td>
<td>RDBP</td>
<td>1,47E-05</td>
<td>0,00253664</td>
</tr>
<tr>
<td>cg20702204</td>
<td>12</td>
<td>133430077</td>
<td>CHFR</td>
<td>2,19E-05</td>
<td>0,00274526</td>
</tr>
<tr>
<td>cg22352818</td>
<td>2</td>
<td>97193289</td>
<td></td>
<td>2,19E-05</td>
<td>0,00345245</td>
</tr>
<tr>
<td>cg09747456</td>
<td>10</td>
<td>91405300</td>
<td>PANK1</td>
<td>3,90E-05</td>
<td>0,0031413</td>
</tr>
<tr>
<td>cg07580707</td>
<td>3</td>
<td>14220061</td>
<td>XPC; LSM3</td>
<td>4,69E-05</td>
<td>0,00633959</td>
</tr>
<tr>
<td>cg02533998</td>
<td>12</td>
<td>101540053</td>
<td></td>
<td>5,62E-05</td>
<td>0,29100396</td>
</tr>
<tr>
<td>cg00932677</td>
<td>4</td>
<td>187776068</td>
<td></td>
<td>6,72E-05</td>
<td>0,00305652</td>
</tr>
<tr>
<td>cg08859247</td>
<td>17</td>
<td>17286854</td>
<td></td>
<td>6,72E-05</td>
<td>0,00848049</td>
</tr>
<tr>
<td>cg00608660</td>
<td>16</td>
<td>8658622</td>
<td>MTHFS; FLJ30679</td>
<td>6,72E-05</td>
<td>0,08698691</td>
</tr>
<tr>
<td>cg16933664</td>
<td>11</td>
<td>67085326</td>
<td>LOC100130987</td>
<td>8,02E-05</td>
<td>0,00328628</td>
</tr>
<tr>
<td>cg08088677</td>
<td>2</td>
<td>6911141</td>
<td></td>
<td>8,02E-05</td>
<td>0,00630309</td>
</tr>
<tr>
<td>cg26577201</td>
<td>2</td>
<td>219857793</td>
<td>CRYBA2</td>
<td>8,02E-05</td>
<td>0,00745067</td>
</tr>
<tr>
<td>Gene</td>
<td>Primer</td>
<td>Size</td>
<td>Sequence</td>
<td>Product Size</td>
<td>Nº of CpG's</td>
</tr>
<tr>
<td>--------------</td>
<td>--------</td>
<td>------</td>
<td>-----------------------------------------------</td>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td><strong>PPM1A</strong></td>
<td>LP</td>
<td>25</td>
<td>GTAGGTAGGGTGTAGGGTGTTAGAT</td>
<td>484</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>RP</td>
<td>25</td>
<td>CTTACAACCCAAAAACAAATTCAAC</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>KCNQ1</strong></td>
<td>LP</td>
<td>25</td>
<td>TTGTTGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT</td>
<td>499</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>RP</td>
<td>25</td>
<td>CATTATACACAACCTAAACACCCCC</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>RAF1</strong></td>
<td>LP</td>
<td>25</td>
<td>TATGTTTTGGTTATAGGTAGGTATG</td>
<td>335</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>RP</td>
<td>25</td>
<td>ACCCAATTTTAAAAAATAATTCCC</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* LP, Left Primer; RP, Right Primer.

**Table V:** Sequences of primers used in the MassARRAY EpiTYPER second stage analysis.
Table VI: Differentially methylated CpG sites identified *de novo* in the EpiTYPER second stage analysis.

<table>
<thead>
<tr>
<th>CpG ID</th>
<th>Chr</th>
<th>Gene</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>PPM1A_4</td>
<td>14</td>
<td>PPM1A</td>
<td>0.02733</td>
</tr>
<tr>
<td>RAF1_3</td>
<td>3</td>
<td>RAF1</td>
<td>0.01049</td>
</tr>
<tr>
<td>RAF1_6</td>
<td>3</td>
<td>RAF1</td>
<td>0.00432</td>
</tr>
<tr>
<td>RAF1_10</td>
<td>3</td>
<td>RAF1</td>
<td>0.00153</td>
</tr>
</tbody>
</table>
Figures.

Figure I: Quality control analysis. (a) Density plot showing the methylation levels distribution in each sample. The same bimodal distribution was observed in all samples except for one sample (removed from further analysis). X-axis indicates methylation β-values and Y-axis frequency, each line represents one sample. (b) Multidimensional Scaling plot showing two clusters, male samples and female samples. Two samples (indicated by arrow) were group in the wrong cluster (removed from further analysis). X-axis indicates the first dimension of variance and Y-axis indicates the second dimension of variance.
**Figure II:** Clustering heatmap of the 16 differentially methylated CpG sites identified by 450k analysis. Each column represents a sample and each horizontal line represents the methylation levels of a given CpG across samples. Methylation levels are expressed as 0-1 β-values (green and red, unmethylated and completely methylated, respectively).
Figure III: PPM1A methylation levels on stroke and cardiovascular patients, only stroke patients, and no-cardioembolic stroke patients from validation cohorts. All three analysis show higher methylation levels of vascular recurrent patients compared with non-vascular recurrent patients, regardless of the pathology (stroke or cardiovascular disease).
**Figure IV**: Diagram of candidate CpG sites selected for second stage study.

- EWAS
  - 475,500 CpG sites analyzed
  - Top 100 Differentially Methylated CpG sites
  - Top 100 Differentially Methylated CpG sites including covariates
  - 36 coincident CpG sites
  - 4 CpG sites previously associated with atherosclerotic process
  - 2 CpG sites suitable for EpiTYPER analysis