DNA methylome analysis reveals distinct epigenetic patterns of ascending aortic dissection and bicuspid aortic valve

Sun Pan[1†], Hao Lai[1†], Yiru Shen[2], Charles Breeze[3], Stephan Beck[3], Tao Hong[1], Chunsheng Wang[1]*, Andrew E Teschendorff[2,4,5]*

[1]Department of Cardiac Surgery, Zhongshan Hospital, Fudan University, 180 Fenglin Road, Xuhui District, Shanghai 200032, China

[2]CAS Key Lab of Computational Biology, CAS-MPG Partner Institute for Computational Biology, Shanghai Institute for Biological Sciences, Chinese Academy of Sciences, 320 Yueyang Road, Shanghai 200031, China.

[3]Medical Genomics, Paul O'Gorman Building, UCL Cancer Institute, University College London, 72 Huntley Street, London WC1E 6BT, UK.

[4]Statistical Cancer Genomics, Paul O'Gorman Building, UCL Cancer Institute, University College London, 72 Huntley Street, London WC1E 6BT, UK.

[5]Department of Women's Cancer, University College London, Medical School Building, Room 340, 74 Huntley Street, LondonWC1E 6AU, UK.

* Corresponding author. Tel/Fax: +8621 64223006, E-mail: wangchunsheng1964@163.com (CW); a.teschendorff@ucl.ac.uk (AET)

[†]These authors contributed equally to this work.

Accession codes: The microarray data have been deposited in the GEO database under accession code GSE84274.

Abstract

Aims Epigenetics may mediate the effects of environmental risk factors on disease, including heart disease. Thus, measuring the DNA methylome offers the opportunity to identify novel disease biomarkers and novel insights into disease mechanisms. The DNA methylation landscape of ascending aortic dissection (AD) and bicuspid aortic valve (BAV) with aortic aneurysmal dilatation remain uncharacterized. The present study aimed to explore the genome-wide DNA methylation landscape underpinning these two diseases.

Methods and results We used Illumina 450k DNA methylation beadarrays to analyze 21 ascending aorta samples, including 10 cases with AD, 5 with BAV and 6 healthy controls. We adjusted for intra-sample cellular heterogeneity, providing the first unbiased genome-wide exploration of the DNA methylation landscape underpinning these two diseases. We discover that both diseases are characterized by loss of DNA methylation at non-CpG sites. We validate this non-CpG hypomethylation signature with pyrosequencing. In contrast to non-CpGs, AD and BAV exhibit distinct DNA methylation landscapes at CpG sites, with BAV characterized mainly by hypermethylation of EZH2 targets. In the case of AD, integrative DNA methylation gene expression analysis reveals that AD is characterized by a dedifferentiated smooth muscle cell phenotype. Our integrative analysis further reveals hypomethylation associated overexpression of RARA in AD, a pattern which is also seen in cells exposed to smoke toxins.

Conclusion Our data supports a model in which increased cellular proliferation in AD and BAV underpins loss of methylation at non-CpG sites. Our data further supports a model, in which AD is associated with an inflammatory vascular remodeling process, possibly mediated by the epigenome and linked to environmental risk factors such as smoking.

1. Introduction

Aortic dissection (AD) is the most frequently diagnosed lethal condition of the aorta, and is classified as Stanford type A if the ascending aorta is involved. Bicuspid aortic valve (BAV) is the most common congenital cardiac malformation and is frequently associated with an aortopathy manifested by aneurysmal dilatation of the ascending aorta. Aortic diseases are only diagnosed after a long period of subclinical development, at which point they present with a dissection or rupture, with an extremely poor prognosis. Furthermore, the overall global death rate from AD and aortic aneurysms has increased from 2.49 per 100 000 in 1999 to 2.78 per 100 000 inhabitants in 2010[4], representing an increased global health burden. Underlying this increased burden is also the increased worldwide exposure to major risk factors, including notably smoking and hypertension. Thus, while risk prediction and early detection of aortic diseases remains the outstanding challenge, there is an equally urgent need to elucidate the molecular mechanisms linking the major risk factors to AD and BAV.

The epigenome, and DNA methylation in particular, is a highly malleable entity, with DNAm alterations having been associated with all major disease risk factors including diet, smoking and age [Petronis A et al 2010, Teschendorff et al JAMA Onco 2015, Teschendorff et al Genome Res 2010]. For instance, recent studies have identified DNAm changes in the blood of smokers which may mediate the causal link to lung cancer and which are able to predict the future risk of lung cancer [BMJ paper + Fasanelli F et al Nat Comm.2015]. While the role of DNAm alterations in cardiovascular disease is also rapidly increasing [5,6], its role (if any) in the pathogenesis of AD and BAV is unclear.

Here, we decided to perform an explorative study of the DNAm landscapes underpinning AD and BAV. The comparison of AD to BAV is also of interest, as it has been proposed that BAV is not only a disorder of valvulogenesis, but also represents the co-existent abnormalities of aortic media[2,3]. Indeed, patients with BAV, including those with a haemodynamically normal valve, may have dilated aortic roots and ascending aortas. In addition, for AD we perform an integrative DNAm – mRNA expression analysis, using previous gene expression data of the ascending aorta of Stanford type A acute aortic dissection cases [1]

2. Methods

2.1 Ethics statement and samples

This study was approved by the Ethics Committee of Zhongshan Hospital, Fudan University (Approval No. B2012-001) and all patients gave written informed consent. The study conforms to the principles outlined in the Declaration of Helsinki. A total of 24 ascending aortic tissue samples were collected. The 24 samples were collected from the individuals including 12 patients with acute ascending aortic dissection, 6 patients with bicuspid aortic valve associated with aneurysmal dilatation of the ascending aorta (aortic diameter >4.5 cm) and 6 organ donors. Enrollment criteria of patients with aortic dissection and method of samples harvest were previously described[1]. Ascending aortic tissue samples from patients with bicuspid aortic valve were similarly harvested at the time of aortic valve surgery and ascending aortic replacement. The tissue specimens used for DNA isolation were free of macroscopic thrombus or blood. Normal control samples were treated in the same manner as the test samples. Detailed clinical information of the individuals enrolled in the study is shown in *table S1*.

2.2 DNA isolation and bisulfite modification of DNA

DNA was isolated from aorta tissue using DNeasy Blood and Tissue Kit (Qiagen, Hilden, Germany) and genomic DNA was treated with sodium bisulfite using the EZ DNA Methylation kit (ZymoResearch, Irvine, CA, USA) following the protocol supplied by the manufacturer.

2.3 Methylation analysis using the 450k array

DNA methylation analysis using the Infinium Human-Methylation450k BeadChip (Illumina, San Diego, CA, USA) was performed according the manufactures' instruction. Raw Illumina data files were generated for further analysis.

2.4 Methylation analysis using pyrosequencing

Aliquots of the same genomic DNA as for microarray experiment were converted by bisulfite as previously mentioned. PCR reactions amplifying bisulfite-treated DNA for subsequent pyrosequencing analysis are performed using PyroMark PCR Kit (Qiagen, Hilden, Germany). Pyrosequencing reactions were performed using PyroMark Gold Q96 Reagent Kit (Qiagen) in PyroMark Q96 ID System (Qiagen, Hilden, Germany). PCR primers and Pyrosequencing primers were designed for 10 non-CpG loci using PyroMark Assay Design Software 2.0 (Qiagen, Hilden,

Germany). The sequence of primers are shown in *table S2*. All experimental processes were carried out according to the manufacturer's protocol.

2.5 Preprocessing, quality control and normalization

We used the "preprocessRaw" function in the R package "minfi"[7] to convert raw Red/Green channel signals (idat files) into methylation signals. The "detectionP" function from the same package was used to determine coverage per probe and sample using a detection P-value threshold of 0.05. We then ran BMIQ[8] to correct for type-2 probe bias.

2.6 Reference DNA methylation centroid construction and cell-type deconvolution

In order to obtain approximate estimates of the cellular proportions in our samples, we used Illumina 450k data from the ENCODE project[9]. The ENCODE 450k data was normalized with BMIQ. To identify the most relevant ENCODE cell-lines we estimated partial correlation coefficients between the DNAm profiles of our 21 aorta samples and each of the 63 ENCODE cell-lines. Partial correlations assess the similarity of each of our 21 aorta samples to each of the 63 ENCODE cell-lines taking into account the correlation to all other ENCODE cell-lines. Thus a significant positive partial correlation between a sample and a given cell-line means that the cell-line's DNAm profile is highly similar to that of the sample, and that this similarity can't be explained by the correlation to another ENCODE cell-line. ENCODE cell-lines were then ranked according to the average partial correlation over the 21 samples. In line with the expectation that aorta samples are made up primarily of aortic smooth muscle cells, fibroblasts and endothelial cells, the top 3 ranked cell-lines represented these cell-types. Specifically, the top-ranked cell-lines represent models for progenitor fibroblasts, aortic smooth muscle cells and human umbilical vein endothelial cells. Other highly ranked cell-lines represent other types of fibroblasts but were excluded due to highly similar profiles with progenitor fibroblasts. In order to construct the reference DNAm profiles (the "centroid"), we identified high-confidence differentially methylated CpGs between each pair of cell-types (6 pairwise comparisons) by ranking probes according to their difference in methylation and picking the *n* top-ranked probes, where *n=min(*50, #probes with $|\Delta\beta|>0.7$). Thus, for each comparison we picked the number of probes where the difference in methylation was larger than 0.7 in absolute terms, or the top-ranked 50, whichever number was the smallest. This resulted in a centroid DNAm data matrix of 131 unique probes and 3-cell types.

With this reference centroid, and for an independent sample with a 450k DNAm profile, the proportions of the underlying cell-types was estimated using Houseman's CP algorithm[10].

2.7 Unsupervised analysis using SVD

Random Matrix Theory (RMT)[11] and Singular Value Decomposition (SVD) was used to assess the number and nature of the significant components of variation in the data. Significant components of variation were then correlated with biological phenotypes, including age, disease status, and the cell type proportions estimated using the Houseman CP algorithm. Age and cell-type proportions were treated as continuous variables and linear regression was used, whereas disease status (H, AD, BAV) was treated as categorical and so a Kruskal-Wallis test was used. This unsupervised analysis was performed in both the beta and M-value ($M=\log_2(\beta/(1-\beta))$ basis.

2.8 Differential DNA methylation analysis

Due to the small sample size of our study, and therefore the need to use empirical Bayes methods for calling differential methylation[12], beta values were converted to M-values, since M-values are less heteroscedastic and therefore conform better to the Gaussian assumption underlying the empirical Bayes model. Differential methylation was called at the probe-level on the M-valued PC1-adjusted data using an empirical Bayesian framework as implemented in the R package "Limma"[13]. This allowed us to detect differentially methylated CpGs (DMCs) between every pair of phenotypic comparisons. False Discovery Rate (FDR) was used to correct P-values for multiple testing and a threshold of FDR<0.15 was used to declare statistical significance. We note that we relaxed the threshold of significance since in some cases no DMC passed a threshold of 0.05. This is still acceptable since in our experience FDR thresholds of even <0.3 can lead to molecular signatures that can be validated in external data[14]. In our case, an FDR<0.15 means that a DMC has an approximately 15% change of being a false positive, so an 85% change of being a true positive.

2.10 Integration of DNA methylation and mRNA expression

Our previous study reported 1152 differentially expressed genes (DEGs) between AD cases and healthy controls[1]. For each of these DEGs, we selected all differentially methylated probes between AD and H (at FDR < 0.3) that mapped to this gene.

2.11 Enrichment of ChIP-Seq histone signals and transcription factor binding site analysis

Fully processed Roadmap epigenomics histone mark data were downloaded from http://egg2.wustl.edu/roadmap/data/byFileType/peaks/consolidated/broadPeak. We used bedtools[15] to evaluate overlap with 450k array probes. For each probe overlap with a given genomic element was coded as one, and no overlap was coded as zero. Extended documentation on how to rebuild the database from scratch, as well as the code used, is available in https://github.com/charlesbreeze/eFORGE/tree/master/database.

For the transcription factor binding site analysis, we followed the same procedure as in our previous publication[16]. For a given list of DMCs, these were split into hypermethylated and hypomethylated subsets, and enrichment for transcription factor binding sites or for histone marks determined using a one-tailed Fisher exact test.

2.12 Enrichment analysis against age-DMCs

To test for enrichment of AD and BAV associated DMCs for sites undergoing differential methylation with age, we identified age-DMCs from a large (n>560 samples) Illumina 450k EWAS for aging conducted in whole blood [Hannum et al Mol.Cell.2014]. The age-DMCs were derived using a very stringent procedure which adjusted for sex, ethnicity, plate effects and changes in blood-cell type composition [Yuan T , Teschendorff AE PLoS Genetics 2015]. A total of 70,249 CpGs passed an FDR<0.05, of which 31,217 were hypermethyated with age, and 39,032 were hypomethylated with age. For the given set of AD-DMCs (or BAV-DMCs) we asked how many of these were significantly associated with age in Hannum et al, taking into account directionality of methylation change, which is important to consider since AD (or BAV) cases are older than controls. Thus, for AD (and separately for BAV) we obtain a 2 x 2 matrix of counts, representing the number of hypermethylated and hypomethylated AD-DMCs (or BAV-DMCs) which are hypermethylated or hypomethylated with age. Odds ratio and P-value of enrichment was then computed using a one-tailed Fisher's exact test.

3. Results

3.1 Unsupervised analysis captures DNAm variation associated with AD and BAV

We performed Illumina 450k DNAm profiling on a total of 6 ascending aorta samples from healthy individuals, 12 samples from ascending aortic dissection (AD) cases and another 6 samples from patients with bicuspid aortic valve (BAV) associated with ascending aortic aneurysmal dilatation. Data underwent a stringent quality control (QC) procedure, including normalization for type-2 probe design bias, as performed by us in previous studies[17], resulting in 484,724 usable probes. All 6 healthy samples were from males, with the 12 AD cases coming from 10 males and 2 females, whereas 5 out of 6 BAV cases were from males (*Table S1*). Hence, in order to avoid confounding by sex, we only retained the 21 male samples for further analysis. Singular Value Decomposition of the 484,724 x 21 data matrix, and using permutations to estimate the number of significant components, revealed 4 significantly variable singular vectors (SVs) (or principal components-PCs) (*Figure 1A*).The top PC accounted for over 35% of the total data variation, with PC-2 and PC-3 accounting for approximately 7-8% of total data variation (*Figure 1A*). PC-1 did not correlate with disease status or age (*Figure 1C*), but we hypothesized that it might correlate with intra-sample cellular heterogeneity[18]. Since aorta samples are expected to be made up mainly of smooth muscle cells, fibroblasts and endothelial cells, we used Illumina 450k data of representative cell-lines from ENCODE[9] to construct a reference DNA methylation centroid from which we then estimated cell-type fractions in individual samples using the Houseman CP algorithm[10]. To identify the most relevant ENCODE cell-lines, we computed partial correlations of each sample's DNAm profile to the corresponding DNAm profile of each of 63 ENCODE cell-lines (*Methods*). This showed that a progenitor fibroblast (ProgFib), an aortic smooth muscle cell (AoSMC), and human umbilical vein endothelial cells (HUVEC) were the most representative cell-lines for modelling aortic smooth muscle cells, fibroblasts and endothelial cells present in our samples (*Figure 1B*). Confirming our expectation, estimated fractions for these 3 cell-types correlated strongly with PC-1 (and only with PC-1) (*Figure 1C*). Specifically, we observed that the proportion of AoSMC-like cells decreased in AD cases, whereas the endothelial cell-like proportion increased (*Figure S1***)**. PC-2 correlated marginally with disease status and age (*Figure 1C*). Since AD and BAV cases were significantly older (*Figure S2*) than the healthy controls, we interpret PC-2 mainly as an age-driven component. Attesting to the quality of our data, Horvath's DNAm-Age[19] correlated significantly with chronological age (PCC=0.62, P=0.003), despite the relatively small sample size (*Figure S3*). Interestingly, all samples except one normal sample, exhibited age-acceleration, but with AD and BAV cases however exhibiting less age-acceleration than the healthy samples (*Figure S3*). PC-3 correlated only with disease status, and was specially prominent discriminating AD from BAV cases, although interestingly it also discriminated both types of disease from healthy controls (*Figure 1D*). Given that PC-1 captures variation associated

with cellular heterogeneity, and that age is an important predictor of outcome in AD and BAV[2,20], we decided to adjust the data for PC-1 only, by regressing this component out of the data prior to the supervised analysis.

3.2 Supervised analysis reveals a non-CpG hypomethylation signature associated with AD and BAV

Applying an empirical Bayesian framework, which works optimally in a small sample size setting[12], to the PC1-adjusted data matrix, we inferred a total of 706 differentially methylated cytosines (DMCs) between AD and H, 3775 between BAV and H, and a total of 12817 DMCs between BAV and AD (*Figure 2A, table S3*). Although AD cases were notably older than controls (*table S1*), among the corresponding DMCs we did not observe an enrichment for age-associated DMCs (Methods, [Hannum et al Mol Cell 2014]), in contrast to BAV-associated DMCs which did exhibit such an enrichment (*Figure S4*). Of the 706 DMCs between AD and H, 396 (56%) were hypermethylated in AD compared to H. Among the 3775 DMCs between BAV and H, 1979 (52%) were hypermethylated in BAV compared to H (*Figure 2B*). Over 75% of the 12817 DMCs between BAV and AD were hypomethylated in BAV compared to AD. Intriguingly, in the AD-H comparison, we observed a 34-fold enrichment of non-CpGs (n=122, Fisher-test, P<1e-100) among the 706 DMCs, with this non-CpG overenrichment being less significant in the case of BAV-H and non-existent between BAV and AD (*Figure 2C*). A heatmap of relative methylation values over the 122 non-CpG DMCs between AD and H revealed that effectively all of these sites lost methylation in AD cases compared to healthy controls (*Figure 2D*). Of note, these sites also lost methylation in BAV cases (*Figure 2D*).

In order to shed light on the nature of this non-CpG hypomethylation signature, we asked if there was a specific bias in terms of the sequence context of the non-CpGs[21]. Comparing the relative occurrence of [CA]C vs [CA]G sequence among our 122 non-CpGs, we observed a striking enrichment for the [CA]C context (*Table 1*).

Table 1 Sequence context enrichment table of significantly hypomethylated non-CpGs for each of the three comparisons: AD vs Healthy, BAV vs Healthy and BAV vs AD.

	[CA]C				[CA]G			
	Count	Exp. Count	OR	P-value	Count	Exp. Count	OR	P-value
AD vs. Healthy	39	6.79	9.99(6.39,15.44)	<2E-16	76	102.88	0.29(0.19,0.43)	1.79E-009
BAV vs. Healthy	21	6.9	3.8(2.19,6.32)	2.73E-006	98	104.57	0.69(0.44,1.12)	0.1

BAV vs. AD	0	1.45	0(0.00,2.59)	0.4	21	21.92	0.78(0.28,2.66)	0.59

OR and P-value estimated by Fisher's Exact Test.

3.3 Technical validation of the non-CpG hypomethylation signature using pyrosequencing

In order to further test the reliability of the data, we decided to validate the non-CpG hypomethylation signature in AD cases using the gold-standard procedure of pyrosequencing. We randomly picked 10 of the top ranked non-CpG probes exhibiting hypomethylation in AD cases vs healthy controls (H) according to the Illumina 450k assay, and for these loci we assessed DNA methylation using pyrosequencing in a subset of 6 H and 6 AD cases (a subset of the original samples used in the discovery). All 10 non-CpG loci exhibited significant hypomethylation in AD cases, thus validating the Illumina results (*Figure 3A*). Further attesting to the quality of the data, we observed very strong correlations between the DNAm values obtained using Illumina 450k and pyrosequencing when assessed in the 6 healthy controls (*Figure 3B, Figure S5*)

3.4 Different chromatin enrichment patterns for AD and BAV

In order to gain further insight into putative epigenetic mechanisms underlying BAV or AD disease, we asked if probes hypermethylated or hypomethylated in AD/BAV are enriched for specific histone marks. We obtained ChIP-Seq histone mark profiles for 5 major marks (H3K27me3, H3K4me3, H3K4me1, H3K36me3 and H3K9me3) in a surrogate cell-type (fetal heart) from the NIH Epigenomics Roadmap. For the enrichment analysis, we selected the top 5000 hypermethylated and 5000 hypomethylated CpGs for each of the 3 pairwise comparisons (AD vs H, BAV vs H and BAV vs AD), which all passed a FDR threshold of 0.3. Among hypermethylated probes we observed a massive enrichment of the repressive H3K27me3 mark, which was specially prominent in BAV disease compared to either healthy controls or AD cases (*Figure 4A*). In contrast, the most striking enrichment when comparing AD cases to controls was seen for the H3K4me1 and H3K36me3 marks among probes hypomethylated in AD cases. Since the histone methyltransferase EZH2 catalyzes H3K27me3 and also acts as a recruitment platform for DNA methyltransferases (DNMTs), these results suggest that BAV disease may be characterized by increased activity of EZH2. To check this we used ChIP-Seq from ENCODE for a total of 58 TFs, albeit in a different cell-type (hESCs). Confirming our hypothesis, we observed strong enrichment (Fisher test P<1e-50) of EZH2, SUZ12 (another member of the PRC2 complex) and CtBP2 among CpGs hypermethylated in BAV disease compared to controls, but no such enrichment among

hypomethylated CpGs (*Figure 4B*). Comparing AD to H, we only observed enrichment (Fisher test P<1e-6) for two TFs (BCL11A and POU5F1) among hypermethylated CpGs (*Figure 4B*), with no enrichment among hypomethylated sites, suggesting that binding of BCL11A and POU5F1 may be disrupted in AD.

3.5 Integration of DNA methylation and mRNA expression reveals downregulation of smooth muscle genes and targets of smooth muscle differentiation factors in AD

We previously performed mRNA expression profiling of 5 healthy individuals and 7 AD cases[1]. Thus, we aimed to identify genes showing both significant differential methylation and differential gene expression between AD and H. Anchoring the analysis on 1152 differentially expressed genes (DEGs) at FDR < 0.05, we identified a total of 254 unique DEGs with at least 1 probe exhibiting significant differential methylation (at FDR < 0.3) (*Table S4*). Of these 254 unique DEGs, 138 were overexpressed in AD compared to H, and 116 underexpressed. We performed GSEA separately on these over and underexpressed genes. While genes overexpressed in AD were enriched for cellular proliferation, genes underexpressed in AD were enriched for many biological terms highly relevant to AD disease (*Table S5*). For instance, we observed many genes (e.g. *CALD1, MRVI1, ADCY9, PLCB4, ACTG2, RAMP1, ADRA1B*) implicated in vascular smooth muscle contraction. Also, many of these genes, as well as other genes (e.g. *MBNL1, DACT3, LDB3, DMPK, LPP*) are targets of SRF, a well-known differentiation factor for smooth muscle cells[22]. Downregulated SRF targets (e.g. *CALD1* or *DACT3*) had probes near their TSS which exhibited hypermethylation, although this pattern was not evident for all (*Figure S6*). Likewise, we observed enrichment of many targets of a MYOD TF binding motif, implicating downregulation of MYOD1 targets (e.g. *MEF2D, GRK5, FAM107B*) in AD. In addition, we observed enrichment of 4 genes (*ADCY9, HRK5, FAM129A and CRIM1*) which have been reported to be also downregulated in unstable atherosclerotic plaque[23], 3 additional smooth muscle genes (*MYOZ2, DES and MYOM1*) and enrichment of 8 genes (*LDB3, KANK1, FAM129A, SORBS2, LATS2, ZBTB20, FOXN3, ZNF295*) which have been previously shown to be underexpressed in samples with systolic heart failure[24]. Furthermore, we observed that MYH11, MYOCD and SRF, all implicated in specifying a differentiated contractile SMC phenotype[25], were all significantly downregulated in AD cases compared to healthy controls (*Figure 5A*). Confirming this, we observed a concomitant increase of signaling entropy[26], a molecular correlate of dedifferentiation and cellular plasticity, in AD cases (*Figure 5B*).

To further test whether AD represents a departure from a normal differentiated SMC phenotype, we compared the DNAm profile of our samples to those of normal AoSMCs, as profiled by ENCODE and the NIH Epigenomics Roadmap. This confirmed that AD cases deviated more from AoSMCs than the normal samples (*Figure 5C*). Interestingly, however, BAV cases did not show significant DNAm deviations from AoSMCs (*Figure 5C*). Thus, even though there were significantly more DMCs between BAV and H than between AD and H (*Figure 2A*), when comparing AD and BAV to AoSMCs, only AD showed significant DNAm deviations (*Figure 5C*).

3.6 Hypomethylation of RARA in AD and smoking

Smoking is a major risk factor for AD and a recent meta-EWAS has identified a number of gene loci reproducibly associated with smoking exposure in blood [Gao X, Jia M, Zhang Y, Breitling LP and Brenner H. Clin Epigenetics 2015 Oct 16;7:113]. Thus, we asked if any of our AD-associated DMCs for which the linked gene also exhibits differential expression, were among gene loci where DNAm has been associated with smoking. Notably, this revealed two specific probes which map to the retinoic acid receptor alpha (RARA) gene (*Table S4*), which has been shown to undergo differential methylation in response to smoking in several EWAS conducted in blood. Specifically, we identified two probes hypomethylated in AD cases (*Table S4*) which also exhibit hypomethylation in cells exposed to smoke toxins [Gao X, Jia M, Zhang Y, Breitling LP and Brenner H. Clin Epigenetics 2015 Oct 16;7:113] . One probe mapped to within 200bp of the TSS of RARA, while the other probe mapped to the 5'UTR. Although none of the 2 probes correlated with smoking status in our AD cases and controls (*Figure S7*), when we tested these 2 probes in 3 large EWAS studies of smoking, one conducted in buccal epithelium[17] and two conducted in blood[27,28], we did observe that the probe mapping to the 5'UTR exhibited significant hypomethylation in smokers compared to non-smokers in all 3 studies (*Figure S8*). Thus, this constitutes the first report of a common molecular alteration (DNA hypomethylation) which is seen in relation to both smoking and AD.

3.7 Genes implicated in BAV exhibit more frequent differential methylation in BAV

Genes found mutated in BAV have previously been reported[29]. We asked if differential methylation around these sites is more frequently observed in BAV compared to a random set of sites. For the 9 genes (NOTCH1, AXIN1, EGFR, ENG, GATA5, NKX2-5, NOS3, PDIA2, and

TGFBR2) implicated in BAV, we identified a total of 333 CpGs mapping to them. We observed that these 333 CpGs exhibited significantly larger absolute t-statistics of differential methylation as compared to CpGs mapping to a randomly selected set of 500 genes (excluding BAV-related genes) (P< 0.0001 from a Wilcoxon-rank sum test, *Figure S9*). In fact, we observed almost twice as many DMCs mapping to BAV-related genes than what would have been expected by random chance (Binomial test P<1e-5, *Figure S9*).

4. Discussion

4.1 Significant non-CpG methylation in the ascending aorta

The first important finding of our study is the significant non-CpG methylation within the ascending aorta and the subsequent loss of methylation at these sites in AD and BAV. The first observation is consistent with a recent study reporting detectable levels of non-CpG methylation in 2 donor aorta samples[31]. Interestingly, among non-CpGs with high methylation levels in normal aorta, we observed an enrichment for a [CA]C context, which is similar to that seen in adult brain tissues[32,33] .This is noteworthy given that previous studies have revealed non-CpG methylation to be abundant only in pluripotent cells and brain cells [30]. Importantly, it has been demonstrated that during development of the mammalian cardiovascular system, the smooth muscle of the ascending aorta derives from the cardiac neural crest[34]. This contribution of the neural crest to the ascending aortic smooth muscle is unusual as most smooth muscle is derived from the mesoderm, yet it clearly indicates a developmental link between this specific area of the aorta and the neural system, which may explain the observed non-CpG methylation in our aorta samples. Furthermore, for non-CpG methylation to be maintained, it would need to be re-established *de novo* after each cell division, yet there is no known maintenance mechanism for DNAm at non-CpG sites. Thus, in most cell types non-CpG methylation is rapidly lost following cell division, except in infrequently dividing cells such as neurons[21]. Like neurons, differentiated SMCs in adult blood vessels proliferate at an extremely low rate. Thus, the loss of methylation at non-CpGs observed in AD and BAV could be due to abnormally proliferating SMCs, consistent with the observed higher expression of cell-proliferation genes. Some reports have also provided evidence that non-CpG methylation could have a functional role in biological and pathological processes, such as genomic imprinting[32], somatic cell reprogramming[35], brain development[36], Rett syndrome[37], diabetes[38] and obesity[39]. Although we don't have any data to support that the observed hypomethylation at non-CpGs is of functional consequence, it will be interesting for

future studies to investigate if the non-CpG methylation in the ascending aorta has a direct functional consequence in aortopathy. Regardless of a functional effect or not, our finding of a strong non-CpG hypomethylation signature in ascending aortic dissection and aortic aneurysmal dilatation with BAV hints at a potential future application of non-CpG methylation as an epigenetic biomarker.

4.2 AD is characterized by a dedifferentiated smooth muscle cell phenotype

Our second important finding is that of a dedifferentiated smooth muscle cell phenotype, as a key feature of AD. Smooth muscle cells are thought to be the major cell type in the aorta and display a remarkable plasticity undergoing phenotype changes in response to environmental cues. Differentiated SMCs express contractile marker genes such as MYH11, MYOCD and SRF. In our study, the integration of DNA methylation and mRNA expression in AD revealed downregulation of smooth muscle genes and targets of smooth muscle differentiation factors (e.g. SRF), while genes overexpressed in AD were enriched for cellular proliferation, suggesting that SMCs in AD underwent dedifferentiation. We note that all of these results were obtained after correction for cell-type composition changes, strongly supporting the view of a dedifferentiated SMC phenotype as the most likely mechanism underlying the observed DNAm changes in AD. That is, even though the proportion of AoSMC-like cells decreased in AD, this by itself does not seem to explain all observed patterns of DNAm alteration in AD. Although the phenotypic plasticity exhibited by mature SMCs confers an advantage during repair of vascular injury, this plasticity can also induce adverse phenotypic switching and contribute to the development and progression of vascular diseases[40-42]. Our integrated data is indicative of such a kind of adverse phenotypic switching of SMCs affecting the contractile function in AD. In advanced atherosclerotic plaques, SMCs may play either a beneficial role or a detrimental role in determining plaque stability, depending on the phenotypic state[43]. The downregulation of genes in AD which are also downregulated in unstable atherosclerotic plaque further suggests that phenotypic changes in AD may be contributing to the instability of the aortic wall and the end-stage disease event of dissection. Other studies have associated vascular inflammatory response with vascular dysfunction and disease, with inflammatory cytokines interacting with SMCs through specific receptors to promote cell growth and migration, which impacts on vascular smooth muscle reactivity[44,45]. Given that our previous mRNA expression study revealed a vascular inflammatory process characterized by overexpressed cytokines and receptors in AD, this supports a model of interaction between inflammatory response and vascular function in the disease.

4.3 Epigenome mediates phenotypic alteration linking to environmental risk factors such as smoking in AD

Smoking has been identified to be a critical risk factor for acute aortic dissection[20] and has been associated with durable alterations in vascular smooth muscle cell and inflammatory cell function[46,47]. Interestingly, RARA has previously been associated with smoking-associated differential methylation in blood[27,28] and was also among the genes exhibiting significant differential methylation and differential expression in AD. We further demonstrated that a specific probe undergoing hypomethylation in AD also undergoes smoking-associated hypomethylation in blood and buccal tissue, suggesting that this smoking-associated hypomethylation may be valid in any cell which comes into direct exposure with smoke toxins. Of note, the observed hypomethylation in our AD cases could not be attributed to their increased smoking exposure, as smokers and non-smokers exhibited similar levels of RARA methylation in both AD cases and controls. Thus, our study demonstrates a common molecular alteration in smoking and AD. In summary, our integrative DNAm-mRNA expression based approach suggests that AD is defined by a dedifferentiated phenotypic alteration in SMCs, probably associated with an impaired contractile function of SMCs and weakening of the aortic wall, itself suggestive of a vascular pathological process that occurs in response to environmental cues such as smoking (*Figure 6*).

4.4 AD and BAV with aortic aneurysmal dilatation exhibit distinct DNA methylomes

Our study has further demonstrated that BAV and AD exhibit different epigenetic profiles, supporting the view that these represent two very different pathological conditions. Indeed, there is ample evidence that BAV associated with aortic insufficiency has a genetic origin and a higher risk of adverse aortic complications irrespective of the extent of valvular disease[48]. However, the underlying genetic origins and epigenetic pathways predisposing to aortopathy remain to be demonstrated. Our results revealed that while BAV with aortic aneurysmal dilatation had more DMCs than AD, that its global DNAm profile did not deviate appreciably from normal AoSMCs. Interestingly, we observed that many of the hypermethylated DMCs characterizing BAV appear to occur at PRC2/EZH2 binding sites suggesting increased DNMT and repressor activity. Of note, we observed that CpGs mapping to genes previously found mutated in BAV, were almost twice as likely to be differentially methylated in BAV than a random set of CpGs. Although we did not assess here whether these DNAm changes were functional, it will be exciting to explore this further and assess whether epigenetically mediated dysfunction of these genes provides an alternative pathway to BAV pathogenesis. In summary, our data points towards widely different

altered epigenetic landscapes underlying BAV and AD, although the diseases themselves may exhibit similar complications such as aortic rupture.

4.5 Limitations

It is important to emphasize the main limitations of our study. First, the small sample size of our study and the lack of an independent validation set, means that our results must be interpreted with caution. Nevertheless, many of our results (e.g. the hypomethylation at non-CpGs in AD) are strongly consistent with known biology and were validated with an independent platform (i.e. pyrosequencing), indicating that the DNAm changes seen in this study are not artifacts. Second, cases and controls were not age-matched, and even though they were matched for smoking status, the differences in age means that cases and controls may have had different lifetime exposures to smoking. Nevertheless, we did not observe any evidence of confounding by age or smoking in AD, as AD-associated DMCs were not enriched for age-associated or smoking-associated DMCs. Moreover, the observed hypomethylation at non-CpGs is clearly not an age-associated or smoking-associated signature as no study has reported such a signature in relation to age [Teschendorff AE et al Hum Mol Genet 2013] or smoking [Gao X, Jia M, Zhang Y, Breitling LP and Brenner H. Clin Epigenetics 2015 Oct 16;7:113]. In contrast, for BAV we did observe an enrichment for age-associated DMCs, which may partly explain the larger number of DMCs between cases and controls, and the observed enrichment for hypermethylated at repressive chromatin marks including H3K27me3, EZH2 and SUZ12.

It could be argued that a third limitation of our study is the use of (ENCODE) cell-lines to adjust for cell-type composition. Although it is clear that cell-lines are limited as models of representative cell-types *in-vivo*, it is worth pointing out that DNAm profiles of such cell-lines have already been successfully used for performing cell-type deconvolution in other complex tissues such as breast[49,50]. The reason why cell-lines may indeed be appropriate for cell-type deconvolution is that the deconvolution itself is only performed using sites which exhibit large differences in DNAm between the underlying cell-types (typically over 80% changes in DNAm). Thus, although cell-lines are subject to cell-culture *in-vitro* effects, which undoubtedly change the DNAm landscape, it is unlikely however that these *in-vitro* effects would cause massive i.e. over 80% changes in DNAm. Thus, reference DNAm profiles derived from cell-line models provides a reasonable approach to estimate cell-type fractions in complex tissues. A key priority for future studies however, will be the generation of DNAm profiles of purified primary cell populations representing the relevant cell-types in aorta samples.

4.6 Conclusions and perspectives

To conclude, we have performed the first explorative study of the DNAm landscape underpinning AD and BAV. Both AD and BAV are characterized by a non-CpG hypomethylation signature, which we posit reflects the increased cellular proliferation seen in both diseases. However, in general, both diseases exhibit widely different DNAm landscapes, with BAV characterized mainly by hypermethylation at sites marked by repressive chromatin, while AD is characterized by a dedifferentiated smooth muscle cell phenotype. Future studies will need to determine the causes of this phenotype switch in AD and whether DNAm alterations contribute to it. Of particular interest will be to investigate the role of DNA methylation alterations as a causal link between smoking and AD.

Conflict of interest: none declared.

Funding

This work was supported by Shanghai Municipal Commission of Health and Family Planning (2014ZYJB0402 to S.P. and C.W.); National Natural Science Foundation of China (31571359 to A.E.T.); Royal Society Newton Advanced Fellowship (164914 to A.E.T.); EU-FP7 Projects EpiTrain and BLUEPRINT (316758, 282510 to C.B. and S.B.)

References

1. Pan S, Wu D, Teschendorff AE, Hong T, Wang L, Qian M, Wang C, Wang X. JAK2-centered interactome hotspot identified by an integrative network algorithm in acute Stanford type A aortic dissection. *PLoS One* 2014;**9**:e89406.
2. Siu SC, Silversides CK. Bicuspid aortic valve disease. *J Am Coll Cardiol* 2010;**55**:2789-2800.
3. Ikonomidis JS, Jones JA, Barbour JR, Stroud RE, Clark LL, Kaplan BS, Zeeshan A, Bavaria JE, Gorman JR, Spinale FG, Gorman RC. Expression of matrix metalloproteinases and endogenous inhibitors within ascending aortic aneurysms of patients with bicuspid or tricuspid aortic valves. *J Thorac Cardiovasc Surg* 2007;**133**:1028-1036.
4. Sampson UK, Norman PE, Fowkes FG, Aboyans V, Yanna S, Harrell FJ, Forouzanfar MH, Naghavi M, Denenberg JO, McDermott MM, Criqui MH, Mensah GA, Ezzati M, Murray C. Global and regional burden of aortic dissection

and aneurysms: mortality trends in 21 world regions, 1990 to 2010. *Glob Heart* 2014;**9**:171-180.
5. Movassagh M, Choy MK, Knowles DA, Cordeddu L, Haider S, Down T, Siggens L, Vujic A, Simeoni I, Penkett C, Goddard M, Lio P, Bennett MR, Foo RS. Distinct epigenomic features in end-stage failing human hearts. *Circulation* 2011;**124**:2411-2422.
6. Zaina S, Heyn H, Carmona FJ, Varol N, Sayols S, Condom E, Ramirez-Ruz J, Gomez A, Goncalves I, Moran S, Esteller M. DNA methylation map of human atherosclerosis. *Circ Cardiovasc Genet* 2014;**7**:692-700.
7. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, Irizarry RA. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* 2014;**30**:1363-1369.
8. Teschendorff AE, Marabita F, Lechner M, Bartlett T, Tegner J, Gomez-Cabrero D, Beck S. A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics* 2013;**29**:189-196.
9. Gerstein M. Genomics: ENCODE leads the way on big data. *Nature* 2012;**489**:208.
10. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, Wiencke JK, Kelsey KT. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics* 2012;**13**:86.
11. Teschendorff AE, Zhuang J, Widschwendter M. Independent surrogate variable analysis to deconvolve confounding factors in large-scale microarray profiling studies. *Bioinformatics* 2011;**27**:1496-1505.
12. Zhuang J, Widschwendter M, Teschendorff AE. A comparison of feature selection and classification methods in DNA methylation studies using the Illumina Infinium platform. *BMC Bioinformatics* 2012;**13**:59.
13. Wettenhall JM, Smyth GK. limmaGUI: a graphical user interface for linear modeling of microarray data. *Bioinformatics* 2004;**20**:3705-3706.
14. Naderi A, Teschendorff AE, Barbosa-Morais NL, Pinder SE, Green AR, Powe DG, Robertson JF, Aparicio S, Ellis IO, Brenton JD, Caldas C. A gene-expression signature to predict survival in breast cancer across independent data sets. *Oncogene* 2007;**26**:1507-1516.
15. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 2010;**26**:841-842.
16. Yuan T, Jiao Y, de Jong S, Ophoff RA, Beck S, Teschendorff AE. An integrative multi-scale analysis of the dynamic DNA methylation landscape in aging. *PLoS Genet* 2015;**11**:e1004996.
17. Teschendorff AE, Yang Z, Wong A, Pipinikas CP, Jiao Y, Jones A, Anjum S, Hardy R, Salvesen HB, Thirlwell C, Janes SM, Kuh D, Widschwendter M. Correlation of Smoking-Associated DNA Methylation Changes in Buccal Cells With DNA Methylation Changes in Epithelial Cancer. *JAMA Oncol* 2015;**1**:476-485.
18. Jaffe AE, Irizarry RA. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome Biol* 2014;**15**:R31.
19. Horvath S. DNA methylation age of human tissues and cell types. *Genome Biol* 2013;**14**:R115.
20. Nienaber CA, Clough RE. Management of acute aortic dissection. *Lancet* 2015;**385**:800-811.
21. Patil V, Ward RL, Hesson LB. The evidence for functional non-CpG methylation in mammalian cells. *Epigenetics* 2014;**9**:823-828.
22. Kaplan-Albuquerque N, Van Putten V, Weiser-Evans MC, Nemenoff RA. Depletion of serum response factor by RNA interference mimics the mitogenic effects of platelet derived growth factor-BB in vascular smooth muscle cells. *Circ Res* 2005;**97**:427-433.
23. Papaspyridonos M, Smith A, Burnand KG, Taylor P, Padayachee S, Suckling KE, James CH, Greaves DR, Patel L. Novel candidate genes in unstable areas of human atherosclerotic plaques. *Arterioscler Thromb Vasc Biol* 2006;**26**:1837-1844.
24. Thum T, Galuppo P, Wolf C, Fiedler J, Kneitz S, van Laake LW, Doevendans PA, Mummery CL, Borlak J, Haverich A, Gross C, Engelhardt S, Ertl G, Bauersachs J. MicroRNAs in the human heart: a clue to fetal gene reprogramming in heart failure. *Circulation* 2007;**116**:258-267.
25. Liu R, Jin Y, Tang WH, Qin L, Zhang X, Tellides G, Hwa J, Yu J, Martin KA. Ten-eleven translocation-2 (TET2) is a master regulator of smooth muscle cell plasticity. *Circulation* 2013;**128**:2047-2057.
26. Banerji CR, Miranda-Saavedra D, Severini S, Widschwendter M, Enver T, Zhou JX, Teschendorff AE. Cellular network entropy as the energy potential in Waddington's differentiation landscape. *Sci Rep* 2013;**3**:3039.
27. Tsaprouni LG, Yang TP, Bell J, Dick KJ, Kanoni S, Nisbet J, Vinuela A, Grundberg E, Nelson CP, Meduri E, Buil A, Cambien F, Hengstenberg C, Erdmann J, Schunkert H, Goodall AH, Ouwehand WH, Dermitzakis E, Spector TD, Samani NJ, Deloukas P. Cigarette smoking reduces DNA methylation levels at multiple genomic loci but the effect is partially reversible upon cessation. *Epigenetics* 2014;**9**:1382-1396.
28. Liu Y, Aryee MJ, Padyukov L, Fallin MD, Hesselberg E, Runarsson A, Reinius L, Acevedo N, Taub M, Ronninger M, Shchetynsky K, Scheynius A, Kere J, Alfredsson L, Klareskog L, Ekstrom TJ, Feinberg AP. Epigenome-wide

association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. *Nat Biotechnol* 2013;**31**:142-147.
29. Dargis N, Lamontagne M, Gaudreault N, Sbarra L, Henry C, Pibarot P, Mathieu P, Bosse Y. Identification of Gender-Specific Genetic Variants in Patients With Bicuspid Aortic Valve. *Am J Cardiol* 2016;**117**:420-426.
30. He Y, Ecker JR. Non-CG Methylation in the Human Genome. *Annu Rev Genomics Hum Genet* 2015;**16**:55-77.
31. Schultz MD, He Y, Whitaker JW, Hariharan M, Mukamel EA, Leung D, Rajagopal N, Nery JR, Urich MA, Chen H, Lin S, Lin Y, Jung I, Schmitt AD, Selvaraj S, Ren B, Sejnowski TJ, Wang W, Ecker JR. Human body epigenome maps reveal noncanonical DNA methylation variation. *Nature* 2015;**523**:212-216.
32. Xie W, Barr CL, Kim A, Yue F, Lee AY, Eubanks J, Dempster EL, Ren B. Base-resolution analyses of sequence and parent-of-origin dependent DNA methylation in the mouse genome. *Cell* 2012;**148**:816-831.
33. Guo JU, Su Y, Shin JH, Shin J, Li H, Xie B, Zhong C, Hu S, Le T, Fan G, Zhu H, Chang Q, Gao Y, Ming GL, Song H. Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain. *Nat Neurosci* 2014;**17**:215-222.
34. Jiang X, Rowitch DH, Soriano P, McMahon AP, Sucov HM. Fate of the mammalian cardiac neural crest. *Development* 2000;**127**:1607-1616.
35. Butcher LM, Ito M, Brimpari M, Morris TJ, Soares FA, Ahrlund-Richter L, Carey N, Vallier L, Ferguson-Smith AC, Beck S. Non-CG DNA methylation is a biomarker for assessing endodermal differentiation capacity in pluripotent stem cells. *Nat Commun* 2016;**7**:10458.
36. Lister R, Mukamel EA, Nery JR, Urich M, Puddifoot CA, Johnson ND, Lucero J, Huang Y, Dwork AJ, Schultz MD, Yu M, Tonti-Filippini J, Heyn H, Hu S, Wu JC, Rao A, Esteller M, He C, Haghighi FG, Sejnowski TJ, Behrens MM, Ecker JR. Global epigenomic reconfiguration during mammalian brain development. *Science* 2013;**341**:1237905.
37. Gabel HW, Kinde B, Stroud H, Gilbert CS, Harmin DA, Kastan NR, Hemberg M, Ebert DH, Greenberg ME. Disruption of DNA-methylation-dependent long gene repression in Rett syndrome. *Nature* 2015;**522**:89-93.
38. Barres R, Osler ME, Yan J, Rune A, Fritz T, Caidahl K, Krook A, Zierath JR. Non-CpG methylation of the PGC-1alpha promoter through DNMT3B controls mitochondrial density. *Cell Metab* 2009;**10**:189-198.
39. Barres R, Kirchner H, Rasmussen M, Yan J, Kantor FR, Krook A, Naslund E, Zierath JR. Weight loss after gastric bypass surgery in human obesity remodels promoter methylation. *Cell Rep* 2013;**3**:1020-1027.
40. Alexander MR, Owens GK. Epigenetic control of smooth muscle cell differentiation and phenotypic switching in vascular development and disease. *Annu Rev Physiol* 2012;**74**:13-40.
41. Chaabane C, Otsuka F, Virmani R, Bochaton-Piallat ML. Biological responses in stented arteries. *Cardiovasc Res* 2013;**99**:353-363.
42. Ailawadi G, Moehle CW, Pei H, Walton SP, Yang Z, Kron IL, Lau CL, Owens GK. Smooth muscle phenotypic modulation is an early event in aortic aneurysms. *J Thorac Cardiovasc Surg* 2009;**138**:1392-1399.
43. Galis ZS, Khatri JJ. Matrix metalloproteinases in vascular remodeling and atherogenesis: the good, the bad, and the ugly. *Circ Res* 2002;**90**:251-262.
44. Sprague AH, Khalil RA. Inflammatory cytokines in vascular dysfunction and vascular disease. *Biochem Pharmacol* 2009;**78**:539-552.
45. Satoh K, Matoba T, Suzuki J, O'Dell MR, Nigro P, Cui Z, Mohan A, Pan S, Li L, Jin ZG, Yan C, Abe J, Berk BC. Cyclophilin A mediates vascular remodeling by promoting inflammation and vascular smooth muscle cell proliferation. *Circulation* 2008;**117**:3088-3098.
46. McEvoy JW, Nasir K, DeFilippis AP, Lima JA, Bluemke DA, Hundley WG, Barr RG, Budoff MJ, Szklo M, Navas-Acien A, Polak JF, Blumenthal RS, Post WS, Blaha MJ. Relationship of cigarette smoking with inflammation and subclinical vascular disease: the Multi-Ethnic Study of Atherosclerosis. *Arterioscler Thromb Vasc Biol* 2015;**35**:1002-1010.
47. Starke RM, Ali MS, Jabbour PM, Tjoumakaris SI, Gonzalez F, Hasan DM, Rosenwasser RH, Owens GK, Koch WJ, Dumont AS. Cigarette smoke modulates vascular smooth muscle phenotype: implications for carotid and cerebrovascular disease. *PLoS One* 2013;**8**:e71954.
48. Girdauskas E, Borger MA. Bicuspid aortic valve and associated aortopathy: an update. *Semin Thorac Cardiovasc Surg* 2013;**25**:310-316.
49. Onuchic V, Hartmaier RJ, Boone DN, Samuels ML, Patel RY, White WM, Garovic VD, Oesterreich S, Roth ME, Lee AV, Milosavljevic A. Epigenomic Deconvolution of Breast Tumors Reveals Metabolic Coupling between Constituent Cell Types. *Cell Rep* 2016;**17**:2075-2086.
50. Teschendorff AE, Gao Y, Jones A, Ruebner M, Beckmann MW, Wachter DL, Fasching PA, Widschwendter M. DNA methylation outliers in normal breast tissue identify field defects that are enriched in cancer. *Nat Commun* 2016;**7**:10478.

Figure Legends

Figure 1 DNA methylation variation correlates with AD and BAV. (*A*) Plot of the fraction of variation (fVAR, y-axis) explained by the 21 PCs from a SVD on the DNAm data matrix (red-points). The corresponding fraction of variation explained by PCs inferred from a scrambled-up DNAm data matrix, representing the null distribution, is shown in green. There are 4 components with more observed variation than expected by random chance. (*B*) Heatmap of partial correlations of DNAm profiles between the 21 samples (y-axis) and each of 63 ENCODE cell-lines (x-axis) with cell-lines sorted according to their average partial correlation (pCor). Absolute partial correlation values larger than 0.05 are statistically significant. (*C*) Heatmap of P-values of association between the 4 significant PCs and biological factors, including Age, Disease (BAV, AD and H) Status (Status) and estimates of cell-type proportions using aortic smooth muscle cell (AoSMC), progenitor fibroblast (ProgFib) and human umbilical vein endothelial cell (HUVEC). All P-values are estimated with an ANOVA linear model. (*D*) Boxplot of the weight in PC3 versus Disease Status. P-value is from a Kruskal-Wallis test.

Figure 2 Supervised analysis reveals DNAm signatures associated with AD and BAV. (*A*) Histograms of P-values (from moderated t-tests) for the 3 comparisons (AD - H, BAV - H, BAV-AD). Number of DMCs passing a FDR < 0.15 are given. (*B*) Fraction of DMCs identified in A) which exhibit hypermethylation and hypomethylation, for each of the 3 comparisons. For instance, there are almost 60% DMCs hypermethylated in AD compared to H. (*C*) Fraction of DMCs mapping to non-CpG sites for each comparison. Observed (ObsF) versus expected (ExpF) fractions are shown, together with the odds ratio (OR) and Fisher-test P-value. (*D*) Heatmap of relative, standardized methylation values for the 122 non-CpG DMCs between AD and H, across the 21 samples, grouped according to their disease status.

Figure 3 Technical validation of hypomethylated non-CpGs in AD cases. (*A*) Plots of selected top 10 non-CpGs DNA methylation values obtained using pyrosequencing between 6 AD cases and 6 healthy controls (H). P-values are from a one-tailed Wilcoxon rank sum test. (*B*) Scatterplot of the DNA methylation value obtained using pyrosequencing against the Illumina 450K value for all 10 selected non-CpGs, indicated in different colors. For each non-CpG we show the 6 values in the healthy controls. Average R-squared value for each non-CpG from Pearson's Correlation is provided. P-value is from a combined Fisher-test meta-analysis over all 10 non-CpGs.

Figure 4 Enrichment analysis of histone marks and transcription factor binding sites. (*A*) Odds Ratios (OR) of enrichment of histone marks among the top 5000 hypermethylated and 5000 hypomethylated CpGs (FDR < 0.3) for each of the 3 comparisons: AD vs H, BAV vs H, and BAV vs AD. For instance, for AD-H comparison, hypermethylated DMCs have higher methylation in AD vs H. Those ORs which were highly statistically significant are indicated with Fisher-test P-

values. (*B*) As A), but now for ChIP-Seq TF binding sites for TFs which were strongly enriched in any of 3 comparisons, as indicated. Enrichment P-values < 1e-6 are indicated.

Figure 5 AD associates with a loss of smooth muscle cell phenotype. (*A*) Boxplots comparing mRNA expression levels of 3 key genes specifying a differentiated contractile SMC phenotype, between AD cases and healthy controls. P-values are from a one-tailed Wilcoxon rank sum test. (*B*) Boxplot of the signaling entropy rate (SR/maxSR) between AD cases and healthy controls (H). P-value is from a one-tailed Wilcoxon rank sum test. (*C*) Genome-wide similarity of the DNAm profile of the samples with the DNAm profile of AoSMCs, with samples grouped according to disease status. P-value is from a one-tailed Wilcoxon rank sum test between H and AD (red), and between H and BAV (blue). Left panel is for a similarity measure derived using the Manhattan Distance between the DNAm profiles of the samples and the profile of AoSMCs. Right panel is for the Pearson Correlation Coefficient (PCC) between the DNAm profiles of the samples and that of AoSMCs.

Figure 6 Epigenome mediates dedifferentiated SMC phenotype alteration in AD in response to environmental risk factors such as smoking. Environmental risk factors such as smoking links to inflammatory vascular remodeling process with increased pathological cell proliferation underpinning the loss of non-CpG methylation and a dedifferentiated SMC phenotype associated with impaired contractile function. Genes hypermethylated/downregulated are enriched in Vascular Smooth Muscle Contraction Pathway in AD.