Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Jie Yin Date

DNA Methylation Association with Chronic Inflammation: A Twin Study

By

Jie Yin
MPH

Department of Epidemiology

_____

Yan Sun
Committee Chair

_____

Committee Member

_____

Committee Member

_____

Committee Member

DNA Methylation Association with Chronic Inflammation: A Twin Study

By

Jie Yin

Bachelor of Medicine
Nanjing Medical University
2014

Thesis Committee Chair: Yan Sun, Ph.D.

An abstract of
A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Epidemiology
2016

Abstract

DNA Methylation Association with Chronic Inflammation: A Twin Study
By Jie Yin

Inflammation plays a very important role in chronic diseases including cardiovascular disease (CVD). DNA methylation (DNAm) is a basic mechanism regulating gene expression and can be affected and modified by chronic inflammation. To understand the influence of chronic inflammation on DNA methylation, we conducted an epigenome-wide association study using a twin-specific model to investigate the epigenetic associations with eight CVD-associated inflammatory biomarkers, by measuring 409,968 DNAm sites of peripheral blood leukocytes from 217 Caucasian male twins. By conducting linear mixed models, no methylome-wide significant association was found corrected for multiple testing across all the eight inflammatory biomarkers, adjusted for age, body max index, smoking status, and cell type proportions of peripheral leukocytes. However, 45 DNAm cites reached a threshold for suggestive significance (p-value $<10^{-5}$), among which 66.7% were hypermethylated with lower levels of inflammatory biomarkers. The most Gene Ontology terms enriched among these suggestive genes were "intracellular" and "binding". In addition, four DNAm cites were associated with at least two inflammatory biomarkers. DNAm sites cg03359731 (OCIAD1) and cg14557787 (AP2A1) were related to the serum levels of both CRP and IL-6; cg17598713 (RGL2) was associated with ICAM-1 and VCAM-1; cg00459119 (SNX29) was associated with ICAM-1 and P-selectin. Although further replication and validation are needed, our results may provide evidence to investigate the mechanism of the process of inflammation affects CVDs through epigenetic modifications.

DNA Methylation Association with Chronic Inflammation: A Twin Study

By

Jie Yin

Bachelor of Medicine
Nanjing Medical University
2014

Thesis Committee Chair: Yan Sun, Ph.D.

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Epidemiology
2016

Acknowledgements

I would like to express deepest gratitude to my thesis advisor Dr. Yan Sun for his full support, expert guidance, understanding and encouragement throughout my study and research. I would also like to thank my colleagues, Xi and Yunfeng, for the discussions which helped me develop the ideas. Finally, I would like to thank my parents and friends for their emotional and spiritual support.

Table of Contents

LITERATURE REVIEW……...……1

METHODS…….....…….....……6

RESULTS……...……..11

DISCUSSION…….…….……...…….13

REFERENCES……...…….…….17

TABLES……...……...…….23

FIGURES……...……..28

LITERATURE REVIEW

Cardiovascular disease (CVD) is one of the main causes of morbidity as well as mortality worldwide [1]. In the United States, CVD is the leading cause of death, with the rate of eight million deaths every year, corresponding to one of every three deaths per year [2]. CVD is a general term classifying many pathological conditions such as hypertension, stroke, congestive heart failure, coronary artery disease (CAD), aortic stenosis, cardiomyopathy, and arrhythmias that affect the heart, blood and vasculature of the body [3]. It is believed that the main causes of the development of CVD are inflammation, impairment in endothelium, and oxidative stress [4]. Inflammation plays a very important role in CVD such as atherosclerosis and hypertension, functioning as one of the mechanisms promoting endothelial dysfunction [5]. Recent studies on atherosclerosis have focused on inflammation, giving new insights into the etiology of CVD. Atherosclerosis is characterized by a chronic disorder of inflammatory function and the innate immune response, involving biomarkers such as C-reactive protein (CRP), interleukin 6 (IL-6), tumor necrosis factor-alpha (TNF-α), and some cell adhesion molecules which are related to the onset of stroke and myocardial infarction in both healthy populations and patients with coronary disease [6]. Inflammatory biomarkers have been under interest for the ability to predicting cardiovascular disease risk, due to the involvement of inflammation in atherosclerosis. CRP is one of the most validated biomarker, as well as soluble CD40 ligand, IL-18, and matrix metalloproteinase 9 (MMP-9), contributing to cardiovascular risk prediction [5].

CRP, a member of the pentraxin family, is a plasma protein produced in the liver and secreted in the plasma and plays an important role in inflammatory reaction and immune response [7]. CRP is not only a biomarker in infection, inflammation and tissue injury, but also is now a disease biomarker of CVDs used as a screening tool[8]. IL-6 is a pro-inflammatory cytokine released from pericardial fat tissue, believed to be a contribution to coronary vessel inflammation [9]. As a key mediator in inflammatory responses, IL-6 mediates the production of some acute phase proteins such as CRP and fibrinogen [10]. It was suggested that vascular cell adhesion molecule-1 (VCAM-1) and intracellular adhesion molecule-1 (ICAM-1), the inflammatory and endothelial biomarkers, have been used as independent predictors of CVD in many human prospective studies [11]. Some studies have suggested that matrix metalloproteinases (MMPs) had association with the risk of cardiovascular or atherosclerotic events, especially for MMP-9, the most promising MMP, is found to be associated with atherosclerosis, and is efficient to predict some adverse cardiovascular events [12]. P-selectin, also a cellular adhesion molecule, has been demonstrated to play a very important role in the development of atherosclerosis, as well as the tumor necrosis factor-alpha (TGF-β) [13, 14].

Epigenetics is defined as mitotically or meiotically heritable changes in gene expression without the changes in DNA sequence [15]. The molecular mechanisms of epigenetic processes include methylation of DNA, modifications of histone, positioning of histone variants, and gene regulation by non-coding RNAs [16]. A research conducted by

Duygu et al. has found that, in the mammalian genome, the most common epigenetic basis is DNA methylation (DNAm) [17]. When DNA is methylated, it is covalently modified by adding a methyl group to nucleotide cytosines followed by guanines (CpG) that forms 5-methylcytosines. DNA methylation can modulate the gene expression by modifying the process of transcriptional regulation of DNA [18]. Animal studies have provided the evidence that DNA methylation plays a very important role in the development of cardiovascular diseases by inducing hyperproliferation and the dysfunction of cell types active in inflammatory and immune responses [19]. Although more studies are needed to identify the causal relation between DNA methylation and CVD, previous studies have suggested that the status of DNA methylation may be a novel biomarker of CVD [20].

DNA methylation plays an essential role in the regulation of transcription and normal development, and it has been demonstrated that the aberrant DNA methylation are associated with diverse inflammation related diseases and conditions. However, it is suggested that the modification of DNA methylation occurs not only early in life, but also in fully differentiated cells. Thus, DNA methylation patterns are dependent not only on early maternal influence but also on many environmental factors such as smoking status and environment toxicants [21]. Several studies have investigated the connection between chronic inflammation and DNA methylation. Stoyanov et al found that chronic liver inflammation may modify DNA methylation by inducing hypermethylation in specific CpG islands [22]. In Stenvinkel's study exploring the

impact of inflammation on DNA methylation among patients with chronic kidney disease, the researchers have found significant DNA hypermethylation among participants with high levels of hsCRP (high sensitive CRP), while the inverse association was observed between the pattern of DNAm and the level of IL-6 [23]. Given these evidence of epigenetic associations, we hypothesized that the patterns of DNA methylation are associated with the serum levels of CVD-associated inflammatory biomarkers.

Genetic epidemiology is a field that aims to understand the distribution of genomic variation in human population, and investigate the associations with disease risks and prognosis. Since the start of the Human Genome Project (HGP), the development of study design and statistical analysis for genome-wide association studies (GWAS) has made a great contribution to understanding the genetic architecture of human diseases and traits [24]. As GWAS derived from and serve for the field of genetic epidemiology, epigenome-wide association studies (EWAS) grew from and serve for the field of epigenetic epidemiology. Although the term “epigenome” also involves histone modifications and non-coding RNA regulation in addition to DNA methylation, EWAS is now more widely used to studyDNA methylation using genomic DNA extracted from peripheral blood and other specimens [25]. It has been proposed in a review that in contrast to GWAS, an EWAS is a technique of great value to identify molecular mechanisms linking environmental exposures and diseases, and identify biomarkers with potential predictive value [26].

To date, many studies have suggested that genomic functions can be influenced by environmental factors through epigenetic modifications [27-30]. Most of the phenotypes also result from the effect of interaction between environment factors and genomic variation and epigenetics may also play a very important role. Although current technologies have provided us accurate and integrated profiling of epigenome-wide patterns, understanding the epigenetic patterns under genetic and environmental influences is still a challenge to traditional methodologies of epidemiology [31]. A study has applied the twin method to identify global DNA methylation in MZ and DZ twins, and found that epigenetic regulation was significantly influenced by both environmental and genetic factors [32]. Thus, discordant twin design has unique values in epigenetic research by controlling for their genetic similarity and shared growth environment. This design ensures that the epigenetic association is adjusted for confounder such as age, genetic background and shared familial factors and guarantees that the identified epigenetic associations are influenced by environmental factors.

In this study, we conducted an EWAS using a twin-specific model to investigate the epigenetic associations with eight CVD-associated inflammatory biomarkers, by measuring 409,968 DNAm sites from 217 male twins.

METHODS

Study Population

The data of DNA methylation and phenotype were extracted from 218 Caucasian male twins from the Emory Twin Study (ETS) of cardiovascular disease [33]. The ETS included 307 middle-aged male monozygotic (MZ) and dizygotic (DZ) twin pairs born between 1946 and 1956 from the Vietnam Era Twin (VET) Registry, which is one of the largest twin registries in the United States [34]. All twins were treated in pairs and examined at the Emory University General Clinical Research Center between 2002 and 2010. The protocol of ETS was approved by the Emory University Institutional Review Board, and all participants completed an informed consent form.

Phenotypic Measurements

All the twins were provided the same diet the night before the multiple measurements, and were instructed to forbear from smoking. All assessments were performed in the morning after an overnight fast, and both participants of twin pairs were tested at the same time. All medications were suspended for about 24 hours prior to the assessments. The biochemical assays for each twin pair were processed in the same analytical run. Weight and height were measured and body mass index (BMI) were calculated with weight and height. Cigarette smoking status was categorized into current smoker (regardless the number of cigarettes per day) versus never or past smoker. Plasma and peripheral blood leukocytes (PBLs) samples were drawn from venous blood and stored at -80°C until the biomedical assay.

The inflammatory biomarkers selected for investigation were IL-6, CRP, TGF-β, P-selectin, MMP-2, MMP-9, VCAM-1 and ICAM-1. IL-6, TGF-β, P-selectin, VCAM-1, and ICAM-1 were measured using commercially available ELISA (enzyme-linked immunosorbent assay) kits from R and D Systems. The level of high-sensitivity CRP (hsCRP) was measured with the Beckman Coulter High Sensitivity assay. MMP-2 and MMP-9 concentrations were assessed using a commercially available EIA (enzyme immunoassay).

DNA methylation methods

DNA was extracted from PBLs and then 0.5 μg of genomic DNA per sample was bisulfite-converted using the EZ DNA Methylation Kit (Zymo Research, Orange CA). The bisulfite-converted DNA samples were fragmented enzymatically and whole-genome amplified with high purity. The Illumina HumanMethylation450 BeadChip containing locus-specific DNA oligomers was then used to hybridize the DNA samples in batches of 12. The main measures were the fluorescence intensities at each bead site from the arrays, which were fluorescently stained, scanned, and assessed.

The fluorescent signals were measured from the site-specific methlyated (M) and unmethylated (U) beads at each methylation site. The data pre-processing for use were beta (β) values generated by GenomeStudio software, which were continuous variables ranging from 0 to 1 representing the ratio of fluorescence intensity of the methylated and unmethylated sites. We analyzed the quantile-normalized β-values, which were used to adjust for the known technical shift between methylation signals across multiple

categories of probe. The samples with missing rate above 5% would be excluded by using the detection p-value threshold of 0.001. Fortunately, no sample was detected by the method of control probe values greater than 4 standard deviations from their mean values. The methylation sites overlapping with SNPs or not uniquely mapped to the reference genome were excluded from later analyses [35]. The final dataset used for methylome-wide association analyses was 409,786 CpG sites from 217 twin participants.

Epigenetic Association Analysis

To identify the DNAm sites related to inflammation, we used twin-specific linear mixed models with the β values of CpG sites as the dependent variable. The primary independent variables were the between-pair effect and within-pair effect, respectively. The between-pair coefficient represents the expected change in CpG site β value for a one-unit change in the twin-pair mean value of the inflammatory markers, holding the difference between the individual and the twin-pair mean value constant. The within-pair coefficient represents the change in CpG site β value for a one-unit change in the individual deviation from the twin-pair mean value of the inflammatory markers, holding all others constant [36]. Thus, the primary results summarize between-twin effect and within-twin effect. Each inflammatory marker (predictor variable) was analyzed separately, thus, we performed eight EWAS for eight inflammatory markers. Covariates adjusted in the EWAS analyses were age, smoking status (current vs. non-current smokers), BMI and the proportions of leukocyte types (i.e., CD8+T cell,

CD4+T cell, natural killer (NK) cell, B cell, Mononucleosis cell and Granulocyte). Age and BMI were treated as continuous variables, with the units of years and kg/m^2, respectively. Smoking status was defined as a dichotomous variable, where 1 (i.e., current smoker) represented "had smoked within the past year" and 0 represented who did not. The proportions were computed using an algorithm according to cell-type specific DNAm sites [37]. Random effects were modeled to adjust for the chip effects and twin relationships.

The level of hsCRP was logarithmic transformed due to skewed distribution of long tail, while the levels of the other seven biomarkers remained in the original scales. Outliers - the observations in abnormal distance from other values were detected and deleted for each biomarker specific analysis based on extreme values of the biomarker serum level that are three standard deviations away from the mean, as well as the missing values. Singletons were thus generated by removing outliers and missing values. The level of biomarkers for each singleton was treated as the mean value of the twin pair. After removing outliers and missing values, the number of participants left for analyses were 213, 207, 215, 214, 213, 212, 211 and 213 for CRP, IL-6, TGF-β, P-selectin, MMP-2, MMP-9, ICAM-1 and VCAM-1, respectively.

All statistical analyses were performed using the R statistical environment version 3.1.2 (http://www.r-project.org/). R package "nlme" was used to conduct the linear mixed effect model. In addition, functional annotation analysis was conducted for a list of suggestive inflammation-associated genes identified by multivariate analyses. An

online bioinformatics tool – the Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.7 was used to identify the enrichment Gene Ontology (GO) terms among these genes [38]. In DAVID, the p-values for gene enrichment analysis were measured by Fisher Exact test and the p-value threshold was set as 0.05 for over-represented annotation terms.

RESULTS

The demographic information was summarized in Table 1. Due to the different numbers of outliers and missing values, the sample sizes for each inflammatory biomarker ranged from 211 to 215. The average age of the 217 participants was 55.55 years old, with a standard deviation of 3.28. The mean of BMI was 29.31 (SD=4.57) kg/m2. 64 (29.49%) out of the 217 participants were current smokers. The distribution of serum CRP level was skewed with the median of 1.56 (Q1=0.61, Q3=3.14) mg/L. The average serum IL-6 level was 1.89pg/mL (SD=1.43pg/mL). The average serum levels of MMP-2, MMP-9, ICAM-1, VCAM-1, P-Selectin and TGF-β were 166.27±28.80, 325.22±155.77, 314.75±91.74, 630±219.71, 99.85±30.30 and 28.31±8.19 ng/mL, respectively.

After stringent Bonferroni correction for multiple testing in the epigenetic association analyses of the inflammatory biomarkers using the linear mixed models, no DNAm sites with significant methylome-wide association was found across all the eight inflammatory biomarkers. Table 2 presents 45 suggestive DNAm sites with a p-value $< 10^{-5}$ for either within-twin effect or between-twin effect. One site with p-value $<10^{-5}$ was found for log CRP, two sites for IL-6, three sites for MMP-2, six sites for MMP-9, and five, ten, five and thirteen for ICAM-1, VCAM-1, P-selectin and TGF-β, respectively. Among these most significant associations, 30 out of 45 (66.7%) DNAm sites showed negative correlations that hypermethylation is associated with lower level of inflammatory biomarkers. The p-values distributions ranged from moderately

deflated to moderately inflated with the inflation factor ranging from 0.81 to 1.19 across the eight biomarkers (Table 1). Figure 1A-B displays two quantile-quantile (Q-Q) plots with the lowest and highest inflation factors as example, respectively. Manhattan plots of the –log P-value for all 409,786 CpG sites, for both within-twin and between-twin effects across all the eight biomarkers are presented in Figure 2-9.

There were 31 genes annotated to 45 suggestive inflammation-associated DNAm sites and were mapped to GO database of human to determine the enrichment. Through the analysis, we found 5 "biological process" GO terms, 3 "cellular component" GO terms, and 1 "molecular function" GO term over-represented among these suggestive genes (Table 3). Moreover, "intracellular" and "binding" were identified as the major functional base that these inflammation-associated genes are involved, with 21 and 26 genes, respectively.

To identify the DNAm sites significant across the eight inflammatory biomarkers, top 30 significant DNAm sites for both within-twin and between-twin effects were examined, and four sites were associated with at least two biomarkers. The CpG site cg03359731 (OCIAD1), located on chromosome 4, and cg14557787 (AP2A1), located on chromosome 19 were related to the serum levels of both CRP and IL-6; cg17598713 (RGL2) was associated with ICAM-1 and VCAM-1; cg00459119 (SNX29) was associated with ICAM-1 and P-selectin (Table 4).

DISCUSSION

In this twin specific EWAS on a sample of 217 Caucasian male twins from the ETS, no DNAm sites were found that had significant methlyome-wide association with the eight inflammatory biomarkers. However, based on top 30 significant CpG sites for each biomarker, 4 sites were found to be associated with at least two inflammatory biomarkers (Table 4).

In table 4, the first DNA methylation site cg03359731, located on chromosome 4 in Ovarian Carcinoma Immunoreactive Antigen domain containing 1 (OCIAD1) gene, was suggestively associated with both CRP and IL-6. To date, very limited studies had revealed the biological function of OCIAD1. Based on a limited number of studies, OCIAD1 has been demonstrated to be functionally associated with the cytoskeleton-integrin-extracellular matrix interaction and to enhance cell adhesion to extracellular matrix in metastatic ovarian carcinoma induced by lysophosphatidic acid (LPA) – a stimulator in the progress of ovarian cancer [39-41].

The second DNAm site cg17598713, which is located on chromosome 6 in Ral Guanine Nucleotide Dissociation Stimulator-Like 2 (RGL2) gene, was suggestively related to serum level of ICAM-1 and VCAM-1. RGL2 encodes the protein RGL2, a member of the family of candidate Ral guanine nucleotide exchange factors (GEFs), which activate Ras by stimulating the exchange of GDP to GTP. In addition, the domain of RGL2 binding with Ras can also block the signaling and transformation of oncogenic Ras.

Ras is a signaling molecule that plays an important role in cell growth and cell differentialtion [42]. Rap is a kind of protein functioning as a Ras antagonist [43]. RGL2 interaction with Rap is also suggested to contribute to the antagonism of Ras function. Thus, RGL2 is identified as a candidate effector for both Ras and Rap [44].

DNAm site cg00459119, which is located on chromosome 16 in Sorting Nexin 29 (SNX29) gene, suggestively related to ICAM-1 and P-selectin. SNX29 belongs to the sorting nexin family, which includes a huge group of cellular trafficking proteins localized in the cytoplasm, sharing the same domain – the PX domain, a phospholipid-binding motif. These proteins usually bind to specific phospholipids and then form protein complexes, which have the function of membrane trafficking, especially protein sorting [45].

Another DNAm site suggestively correlated with CRP and IL-6 is cg14557787, located on chromosome 19 in the Adaptor-related Protein complex 2 alpha 1 (AP2A1) gene. AP2A1 gene encodes the subunit alpha 1 adaptin of the adaptor-related protein complex 2 (AP-2 complex), which contributes to the formation of clathrin coated vesicles at plasma membrane. The AP-2 complex also has the ability to regulate brassinosteroid receptor (BR) endocytosis and mediated BR signaling by blocking the internalization of brassinosteroid insensitive 1 [46]. In addition, a study showed that AP2A1 is involved in the regulation of human immunodeficiency virus type 1 (HIV-1) replication, by promoting the transport of viral genome [47].

Several limitations of this study need to be addressed in future studies. Ideally, the significant EWAS findings would be replicated in a separate and independent sample to decrease false positive results. However, we were unable to directly replicate the identified epigenetic associations, because of the unique twin design and the availability of inflammatory biomarkers. Without further replication studies in independent populations, the suggestive and significant epigenetic associations with inflammatory biomarkers may not be generalizable to other populations. In addition, the sample size of this study is relatively small, which may make it statistically challenging to identify potential associations with smaller effects.

Epigenomic profiles are cell type-specific [48]. DNAm profiles are commonly studied in PBLs. Since PBLs involve diverse types of leukocytes, the identified inflammation-related DNAm changes may only have an impact on a subtype of PBLs. We were unable to investigate the DNA methylation profiles in all of the cell types in this study, thus we treated PBLs as an aggregated summary of multiple subtypes and also treated the cell type proportions as confounders in the model.

In conclusion, this study on Caucasian male twins presents some suggestive results offering candidate genes that contain multiple inflammation-related CpG sites. The results highlight the genes with the functions of intracellular and molecular binding involved in the regulation of chronic inflammation. Although we found suggestive DNAm association with several inflammatory biomarkers, in terms of DNAm modification and inflammation, the causal directionality is still questionable. Thus,

further researches are needed to confirm the association between inflammation and DNA methylation, and to investigate the exact mechanisms underlying the causal directionality between them. By confirming this, future studies may provide more evidence to investigate the predicting value of inflammation-related DNAm sites to cardiovascular diseases.

REFERENCES

1. Butler, D., *UN targets top killers.* Nature, 2011. **477**(7364): p. 260-1.
2. Go, A.S., et al., *Executive summary: heart disease and stroke statistics--2014 update: a report from the American Heart Association.* Circulation, 2014. **129**(3): p. 399-410.
3. John, J.M. and D.L. Bhatt, *Emerging risk factors for atherosclerosis.* Indian Heart J, 2007. **59**(1): p. 28-37.
4. Reverri, E.J., et al., *Inflammation, oxidative stress, and cardiovascular disease risk factors in adults with cystic fibrosis.* Free Radic Biol Med, 2014. **76**: p. 261-77.
5. Guarner, V. and M.E. Rubio-Ruiz, *Low-grade systemic inflammation connects aging, metabolic syndrome and cardiovascular disease.* Interdiscip Top Gerontol, 2015. **40**: p. 99-106.
6. Packard, R.R. and P. Libby, *Inflammation in atherosclerosis: from vascular biology to biomarker discovery and risk prediction.* Clin Chem, 2008. **54**(1): p. 24-38.
7. Tillett, W.S. and T. Francis, *SEROLOGICAL REACTIONS IN PNEUMONIA WITH A NON-PROTEIN SOMATIC FRACTION OF PNEUMOCOCCUS.* J Exp Med, 1930. **52**(4): p. 561-71.
8. Ansar, W. and S. Ghosh, *C-reactive protein and the biology of disease.* Immunol Res, 2013. **56**(1): p. 131-42.

9. Lim, S. and J.B. Meigs, *Ectopic fat and cardiometabolic and vascular risk.* Int J Cardiol, 2013. **169**(3): p. 166-76.

10. Ong, K.L., et al., *Relationship of pericardial fat with biomarkers of inflammation and hemostasis, and cardiovascular disease: the Multi-Ethnic Study of Atherosclerosis.* Atherosclerosis, 2015. **239**(2): p. 386-92.

11. Malik, I., et al., *Soluble adhesion molecules and prediction of coronary heart disease: a prospective study and meta-analysis.* Lancet, 2001. **358**(9286): p. 971-6.

12. Sundstrom, J. and R.S. Vasan, *Circulating biomarkers of extracellular matrix remodeling and risk of atherosclerotic events.* Curr Opin Lipidol, 2006. **17**(1): p. 45-53.

13. Bielinski, S.J., et al., *P-selectin and subclinical and clinical atherosclerosis: the Multi-Ethnic Study of Atherosclerosis (MESA).* Atherosclerosis, 2015. **240**(1): p. 3-9.

14. Cugno, M., et al., *Potential effect of anti-tumour necrosis factor-alpha treatment on reducing the cardiovascular risk related to rheumatoid arthritis.* Curr Vasc Pharmacol, 2010. **8**(2): p. 285-92.

15. Bernstein, B.E., A. Meissner, and E.S. Lander, *The mammalian epigenome.* Cell, 2007. **128**(4): p. 669-81.

16. Sarkar, D., et al., *Epigenetic regulation in human melanoma: past and future.* Epigenetics, 2015. **10**(2): p. 103-21.

17. Duygu, B., E.M. Poels, and P.A. da Costa Martins, *Genetics and epigenetics of arrhythmia and heart failure.* Front Genet, 2013. **4**: p. 219.

18. Feinberg, A.P., *Epigenetics at the epicenter of modern medicine.* Jama, 2008. **299**(11): p. 1345-50.

19. Baccarelli, A., M. Rienstra, and E.J. Benjamin, *Cardiovascular epigenetics: basic concepts and results from animal and human studies.* Circ Cardiovasc Genet, 2010. **3**(6): p. 567-73.

20. Udali, S., et al., *Cardiovascular epigenetics: from DNA methylation to microRNAs.* Mol Aspects Med, 2013. **34**(4): p. 883-901.

21. Horsburgh, S., et al., *Exercise and inflammation-related epigenetic modifications: focus on DNA methylation.* Exerc Immunol Rev, 2015. **21**: p. 26-41.

22. Stoyanov, E., et al., *Chronic liver inflammation modifies DNA methylation at the precancerous stage of murine hepatocarcinogenesis.* Oncotarget, 2015. **6**(13): p. 11047-60.

23. Stenvinkel, P., et al., *Impact of inflammation on epigenetic DNA methylation - a novel risk factor for cardiovascular disease?* J Intern Med, 2007. **261**(5): p. 488-99.

24. Hindorff LA, M.J., Morales J, Junkins HA, Hall PN, Klemm AK, and Manolio TA. *A Catalog of Published Genome-Wide Association Studies*. 2015 [cited 2016 Mar 29]; Available from: http://www.genome.gov/gwastudies.

25. Flanagan, J.M., *Epigenome-wide association studies (EWAS): past, present, and future.* Methods Mol Biol, 2015. **1238**: p. 51-63.

26. Paul, D.S. and S. Beck, *Advances in epigenome-wide association studies for common diseases.* Trends Mol Med, 2014. **20**(10): p. 541-3.

27. Fraga, M.F., et al., *Epigenetic differences arise during the lifetime of monozygotic twins.* Proc Natl Acad Sci U S A, 2005. **102**(30): p. 10604-9.

28. Ling, C. and L. Groop, *Epigenetics: a molecular link between environmental factors and type 2 diabetes.* Diabetes, 2009. **58**(12): p. 2718-25.

29. Petronis, A., *Epigenetics as a unifying principle in the aetiology of complex traits and diseases.* Nature, 2010. **465**(7299): p. 721-7.

30. Szyf, M., P. McGowan, and M.J. Meaney, *The social environment and the epigenome.* Environ Mol Mutagen, 2008. **49**(1): p. 46-60.

31. Tan, Q., et al., *Twins for epigenetic studies of human aging and development.* Ageing Res Rev, 2013. **12**(1): p. 182-7.

32. Kaminsky, Z.A., et al., *DNA methylation profiles in monozygotic and dizygotic twins.* Nat Genet, 2009. **41**(2): p. 240-5.

33. Vaccarino, V., et al., *Association of major depressive disorder with serum myeloperoxidase and other markers of inflammation: a twin study.* Biol Psychiatry, 2008. **64**(6): p. 476-83.

34. Goldberg, J., et al., *The Vietnam Era Twin Registry.* Twin Res, 2002. **5**(5): p. 476-81.

35. Chen, Y.A., et al., *Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray.* Epigenetics, 2013. **8**(2): p. 203-9.

36. Carlin, J.B., et al., *Regression models for twin studies: a critical review.* Int J Epidemiol, 2005. **34**(5): p. 1089-99.

37. Houseman, E.A., et al., *DNA methylation arrays as surrogate measures of cell mixture distribution.* BMC Bioinformatics, 2012. **13**: p. 86.

38. Dennis, G., Jr., et al., *DAVID: Database for Annotation, Visualization, and Integrated Discovery.* Genome Biol, 2003. **4**(5): p. P3.

39. Sengupta, S., et al., *Ovarian cancer immuno-reactive antigen domain containing 1 (OCIAD1), a key player in ovarian cancer cell adhesion.* Gynecol Oncol, 2008. **109**(2): p. 226-33.

40. Wang, C., et al., *Role of the 18:1 lysophosphatidic acid-ovarian cancer immunoreactive antigen domain containing 1 (OCIAD1)-integrin axis in generating late-stage ovarian cancer.* Mol Cancer Ther, 2010. **9**(6): p. 1709-18.

41. Yang, A.H., et al., *Expression of NCAM and OCIAD1 in well-differentiated thyroid carcinoma: correlation with the risk of distant metastasis.* J Clin Pathol, 2012. **65**(3): p. 206-12.

42. Barbacid, M., *ras genes.* Annu Rev Biochem, 1987. **56**: p. 779-827.

43. Campa, M.J., et al., *Inhibition of ras-induced germinal vesicle breakdown in*

Xenopus oocytes by rap-1B. Biochem Biophys Res Commun, 1991. **174**(1): p. 1-5.

44. Peterson, S.N., et al., *Identification of a novel RalGDS-related protein as a candidate effector for Ras and Rap1.* J Biol Chem, 1996. **271**(47): p. 29903-8.

45. Worby, C.A. and J.E. Dixon, *Sorting out the cellular functions of sorting nexins.* Nat Rev Mol Cell Biol, 2002. **3**(12): p. 919-31.

46. Di Rubbo, S., et al., *The clathrin adaptor complex AP-2 mediates endocytosis of brassinosteroid insensitive1 in Arabidopsis.* Plant Cell, 2013. **25**(8): p. 2986-97.

47. Sapsutthipas, S., et al., *Viral factors involved in adapter-related protein complex 2 alpha 1 subunit-mediated regulation of human immunodeficiency virus type 1 replication.* Southeast Asian J Trop Med Public Health, 2011. **42**(2): p. 311-9.

48. Sun, Y.V., et al., *Comparison of the DNA methylation profiles of human peripheral blood cells and transformed B-lymphocytes.* Hum Genet, 2010. **127**(6): p. 651-8.

TABLES

Table 1. Summary of demographic information for each inflammatory biomarkers under analysis

	Mean (SD) / N(%)	N for final analysis	Inflation Factor
Age (yr.)	55.55 (3.28)	217	
BMI (kg/m^2)	29.31 (4.57)	217	
Current smoker		217	
Yes	64 (29.49%)		
CRP (mg/L) *	2.40 (2.64)	213	
lnCRP (mg/L)	0.37(1.10)	213	0.864
IL-6 (pg/mL)	1.89 (1.43)	207	0.978
MMP-2 (ng/mL)	166.27 (28.80)	213	0.814
MMP-9 (ng/mL)	325.22 (155.77)	212	0.903
ICAM-1 (ng/mL)	314.75 (91.74)	211	0.951
VCAM-1 (ng/mL)	630.15 (219.71)	213	0.916
P-selectin (ng/mL)	99.85 (30.20)	214	1.130
TGF-β (ng/mL)	28.31 (8.19)	215	1.190

*The distribution of CRP is skewed. (median=1.56, Q1=0.61, Q3=3.14)

Table 2. Top significant DNAm cites for all eight inflammatory biomarkers

DNAm	Gene Name	Chr.[a]	MAPINFO[b]	Within-β[c]	Between-β[d]	Within-P[e]	Between-P[f]
log CRP							
cg20335008	SRBD1	2	45838293	-0.0045	-0.0099	0.0636	3.49E-06
IL-6							
cg12371169	RAP1GDS1	4	99182555	-0.0021	0.0005	8.80E-07	0.1719
cg06652011		16	83968316	0.0033	-0.0324	0.2152	5.78E-06
MMP-2							
cg14782672	TWIST1	7	19157634	-0.0002	-9.84E-06	1.42E-06	0.7675
cg15213081		19	523300	0.0012	-0.0001	2.34E-06	0.6913
cg02627352	SRGAP3	3	9267381	5.93E-05	-0.0001	0.1648	3.78E-06
MMP-9							
cg15151929	KIAA1751	1	1896429	4.75E-05	-1.42E-05	4.30E-07	0.1902
cg24450312	RASSF5	1	206681158	-0.0001	1.27E-05	1.19E-06	0.5622
cg05133398	ZNF311	6	28969259	-4.40E-05	-3.52E-06	5.71E-06	0.7539
cg15783800	ALPK2	18	56246997	3.09E-05	-3.10E-06	7.68E-06	0.6276
cg25253705	SLAIN1	13	78314465	-0.0001	-4.71E-05	8.04E-06	0.1632
cg08951271	DDR1	6	30850543	-2.26E-05	-6.75E-05	0.1756	9.52E-06
ICAM-1							
cg19642505		6	30224156	-0.0002	-1.76E-05	1.01E-06	0.5964
cg17598713	RGL2	6	33265534	-3.13E-05	-8.12E-05	0.1183	4.53E-07
cg04648382	RAD52	12	1022780	-4.79E-06	0.0001	0.8818	3.84E-06
cg01098656	IDI1	10	1095069	-9.43E-06	-8.71E-05	0.6692	5.26E-06
cg03005124	OBSL1	2	220426848	-8.61E-06	-0.0003	0.7003	7.62E-06
VCAM-1							
cg13148921	ARNT2	15	80853140	-0.0002	2.13E-05	2.07E-07	0.7717
cg24968869		10	85321586	-0.0001	-1.46E-05	5.98E-07	0.7024
cg13838276		6	118158769	0.0002	-0.0002	3.13E-06	0.0137
cg23633700	HUS1	7	48016735	0.0001	-3.61E-05	5.55E-06	0.3622
cg09357589		6	31148552	0.0002	-5.20E-05	6.49E-06	0.2631
cg13854012		2	162103682	-0.0001	4.18E-05	7.08E-06	0.3453
cg04015962		1	10949192	-0.0001	-1.13E-05	8.04E-06	0.7725
cg14855931		3	138631456	-0.0001	-7.98E-05	8.26E-06	0.0100
cg12805491	PAQR7	1	26198721	7.33E-05	-0.0001	8.41E-05	2.09E-06
cg08638044		15	27819924	6.68E-06	8.72E-05	0.7455	8.93E-06
P-selectin							
cg19347782	MICAL2	11	12159762	-0.0013	0.0001	8.94E-07	0.7858
cg00456395	ELK3	12	96641089	0.0002	1.57E-05	2.48E-06	0.6012
cg15103180	GLRA1	5	151296785	-0.0003	0.0002	6.19E-06	0.1803
cg01938825		7	1563708	0.0004	0.0001	7.74E-06	0.2244
cg14005217	PKMYT1	16	3029753	-6.93E-05	0.0005	0.6575	7.71E-06
TGF-β							

cg21911021	ZIK1	19	58095011	-2.21E-06	9.17E-07	4.11E-07	0.1711
cg03630736	SRP68	17	74069290	-2.83E-06	9.48E-07	6.65E-07	0.1779
cg25854303	SYNPR	3	63264142	-8.64E-07	1.18E-07	1.21E-06	0.5723
cg08253296		5	74907592	-7.90E-07	1.89E-08	1.83E-06	0.9178
cg07201717		1	25031947	-1.03E-06	1.34E-08	3.25E-06	0.9553
cg05348982	MIR618;LIN7A	12	81331012	-6.01E-07	2.82E-07	5.86E-06	0.0856
cg14407437	FABP3	1	31845923	-1.25E-06	-2.70E-07	6.24E-06	0.3986
cg10803098		10	130064137	-1.36E-06	5.43E-07	6.28E-06	0.2842
cg06618097	PNMAL2	19	46999444	-1.58E-06	8.40E-07	8.89E-06	0.0469
cg06496344		19	23253834	-2.04E-07	1.34E-06	0.4271	5.10E-06
cg26225694	LHX4	1	180205253	-5.31E-07	1.16E-06	0.0769	5.49E-06
cg22193726	PTPN11	12	112857626	-3.39E-09	1.10E-06	0.9889	6.51E-06
cg10573018	RSPO1	1	38100837	-8.89E-07	2.11E-06	0.0275	8.63E-06

a. Chromosome number; b. Chromosomal location is based on NCBI build 36.1; c. Coefficient-β for within-twin effect; d. Coefficient-β for between-twin effect; e. P-value for within-twin effect; f. P-value for between-twin effect.

Table 3. GO terms for inflammation-associated genes (p-value threshold=0.05)

GO Term ID	GO Term Name	GO Domain*	Number of Genes	P-Value
GO:0048523	negative regulation of cellular process	BP	9	0.0062
GO:0048519	negative regulation of biological process	BP	9	0.0104
GO:0006468	protein amino acid phosphorylation	BP	5	0.0283
GO:0000725	recombinational repair	BP	2	0.0331
GO:0000724	double-strand break repair via homologous recombination	BP	2	0.0331
GO:0030054	cell junction	CC	4	0.0375
GO:0005622	intracellular	CC	21	0.0448
GO:0044456	synapse part	CC	3	0.0487
GO:0005488	binding	MF	26	0.0466

* BP: Biological Process; CC: Cellular Component; MF: Molecular Function

Table 4. Common significant DNAm cites across all eight inflammatory biomarkers

Biomarker	DNAm	Gene	CHR	MAPINFO	Within_Beta	Between_Beta	Within_P	Between_P
hsCRP	cg03359731	OCIAD1	4	48833388	-0.0047	-0.0013	0.0001	0.1852
IL-6	cg03359731	OCIAD1	4	48833388	-0.0042	0.0007	1.08E-05	0.3458
ICAM-1	cg17598713	RGL2	6	33265534	-3.13E-05	-8.12E-05	0.1183	4.53E-07
VCAM-1	cg17598713	RGL2	6	33265534	-1.43E-05	-2.40E-05	0.0557	0.0002
ICAM-1	cg00459119	SNX29	16	12172834	0.0001	4.89E-05	7.60E-05	0.1623
P-selectin	cg00459119	SNX29	16	12172834	0.0004	3.89E-05	1.65E-05	0.6867
hsCRP	cg14557787	AP2A1	19	50305342	-0.0038	0.0006	8.50E-05	0.5101
IL-6	cg14557787	AP2A1	19	50305342	-0.0030	-0.0007	4.27E-05	0.3583

FIGURES

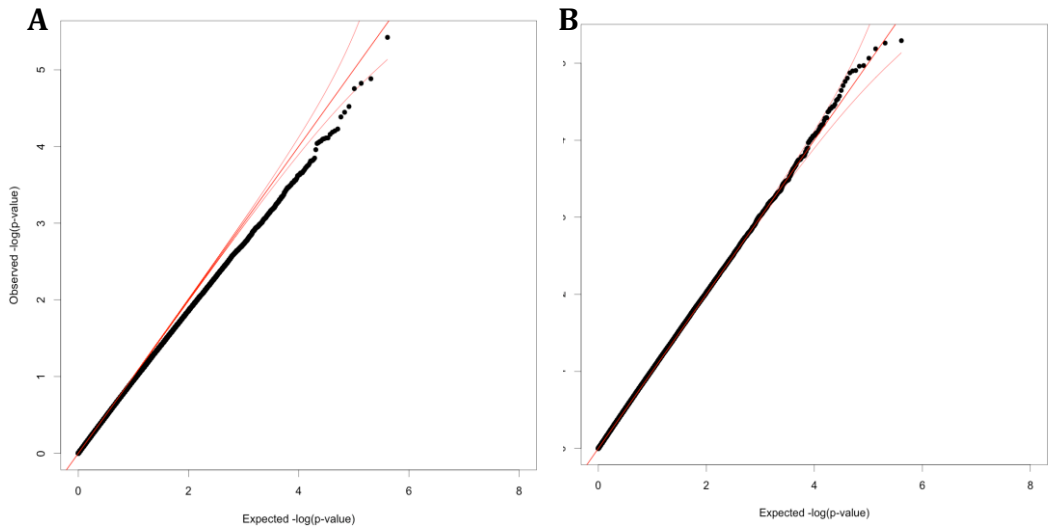

Figure 1. Quantile-quantiel plots of DNAm association with inflammatory biomarkers adjusted for age, BMI, cigarette smoking and cell type proportions of peripheral leukocytes.

Panel A presents the QQ plot for the phenotype, MMP-2 for between-twin efffect. Panel B presents the QQ plot for the phenotype, TGF-β for between-twin efffect.

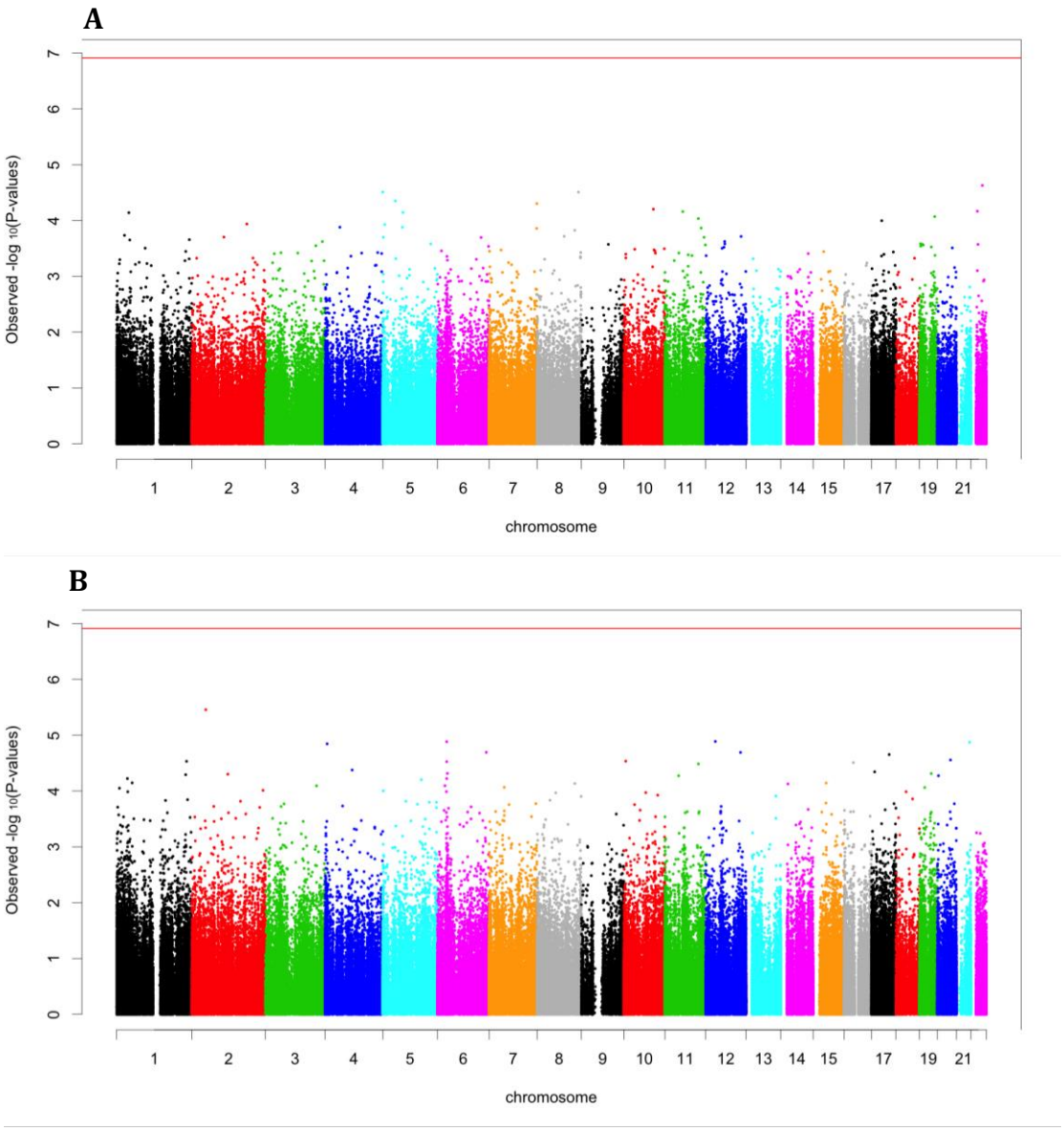

Figure 2. Manhattan plots of methylome-wide association for log CRP, adjusted for age, BMI, cigarette smoking and cell type proportions of peripheral leukocytes.

Panel A presents the Manhattan plot for log CRP, for within-twin effect.
Panel B presents the Manhattan plot for log CRP, for between-twin effect.

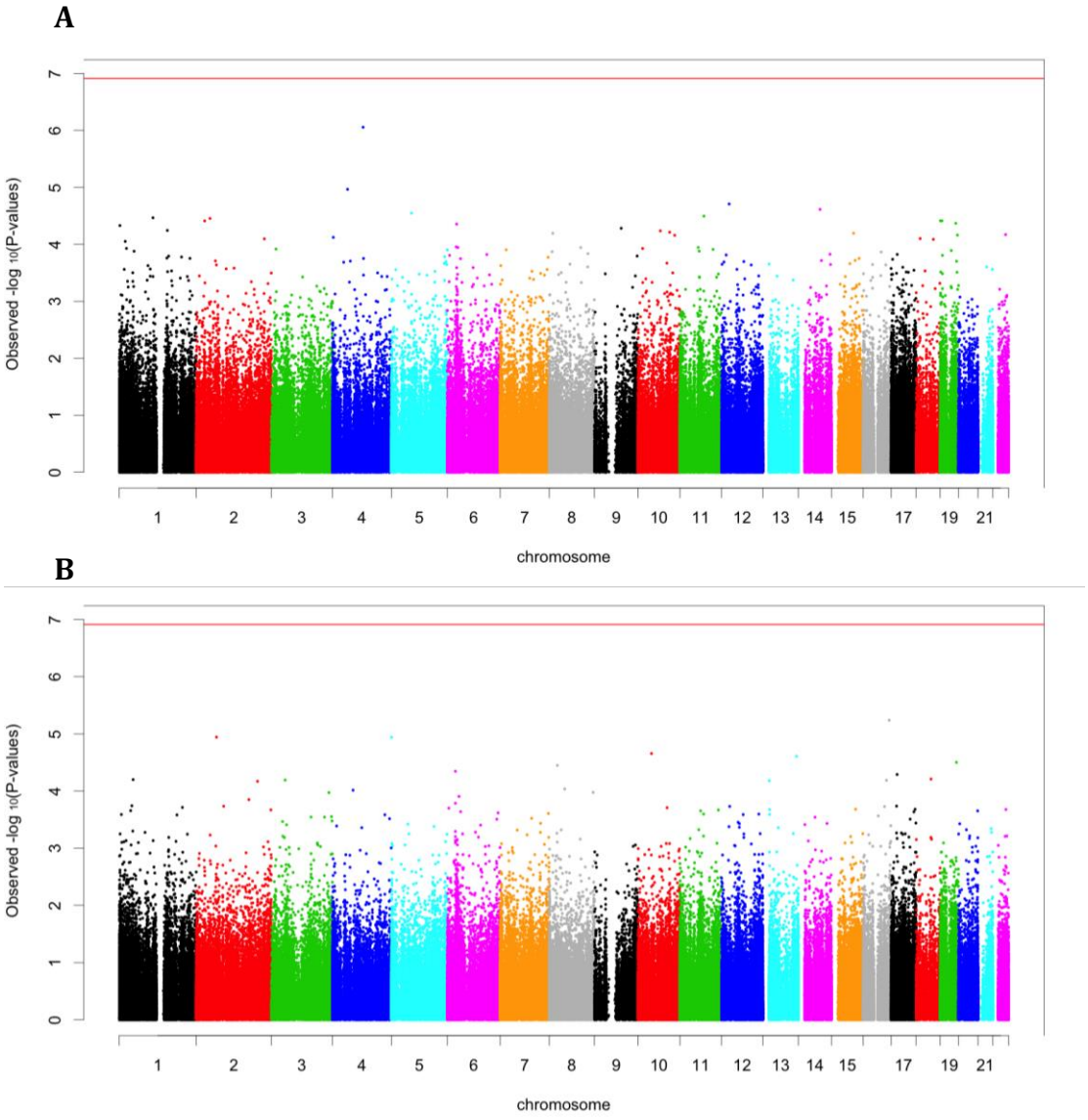

Figure 3. Manhattan plots of methylome-wide association for IL-6, adjusted for age, BMI, cigarette smoking and cell type proportions of peripheral leukocytes.

Panel A presents the Manhattan plot for IL-6, for within-twin effect.

Panel B presents the Manhattan plot for IL-6, for between-twin effect.

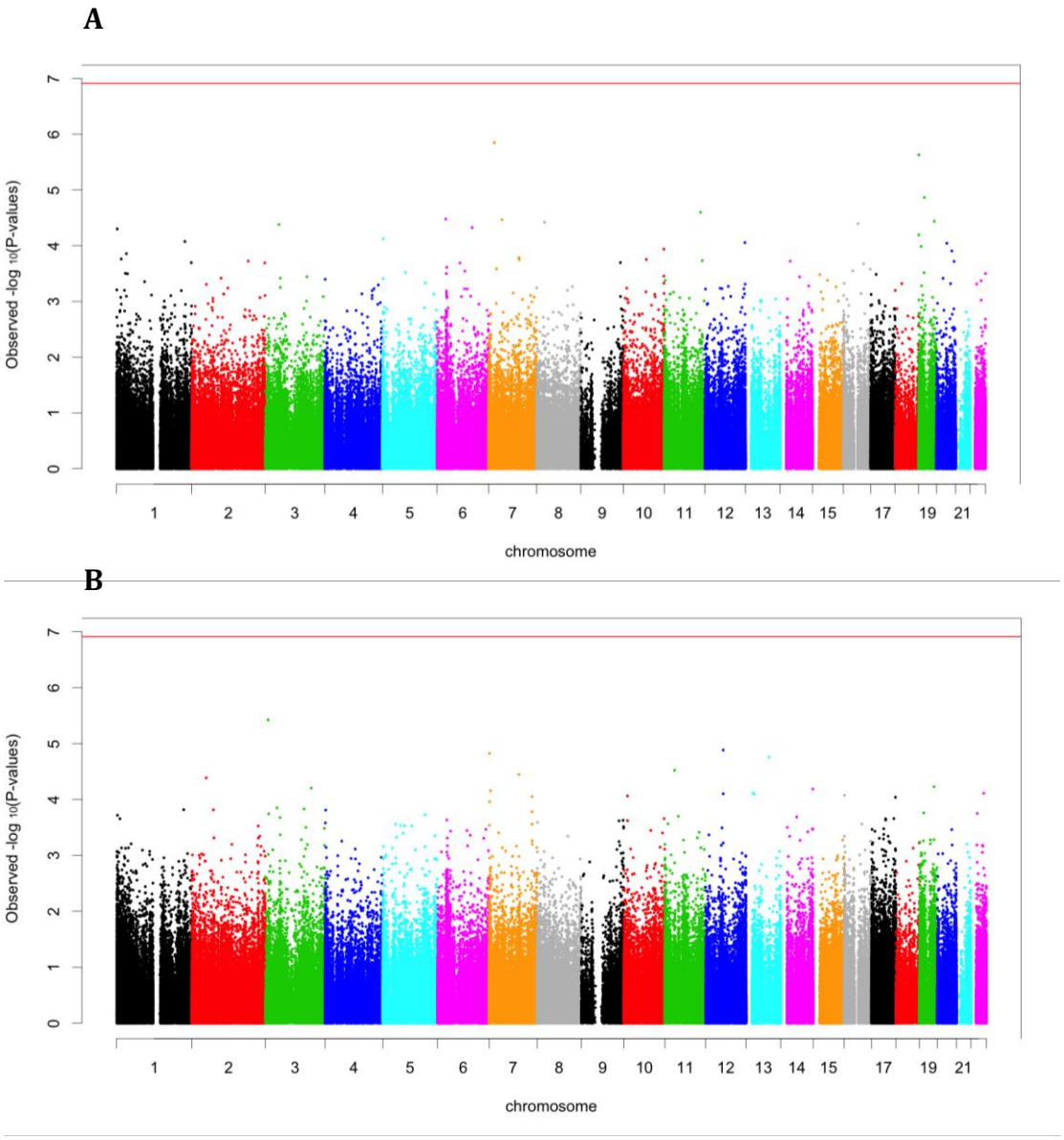

Figure 4. Manhattan plots of methylome-wide association for MMP-2, adjusted for age, BMI, cigarette smoking and cell type proportions of peripheral leukocytes.

Panel A presents the Manhattan plot for MMP-2, for within-twin effect.

Panel B presents the Manhattan plot for MMP-2, for between-twin effect.

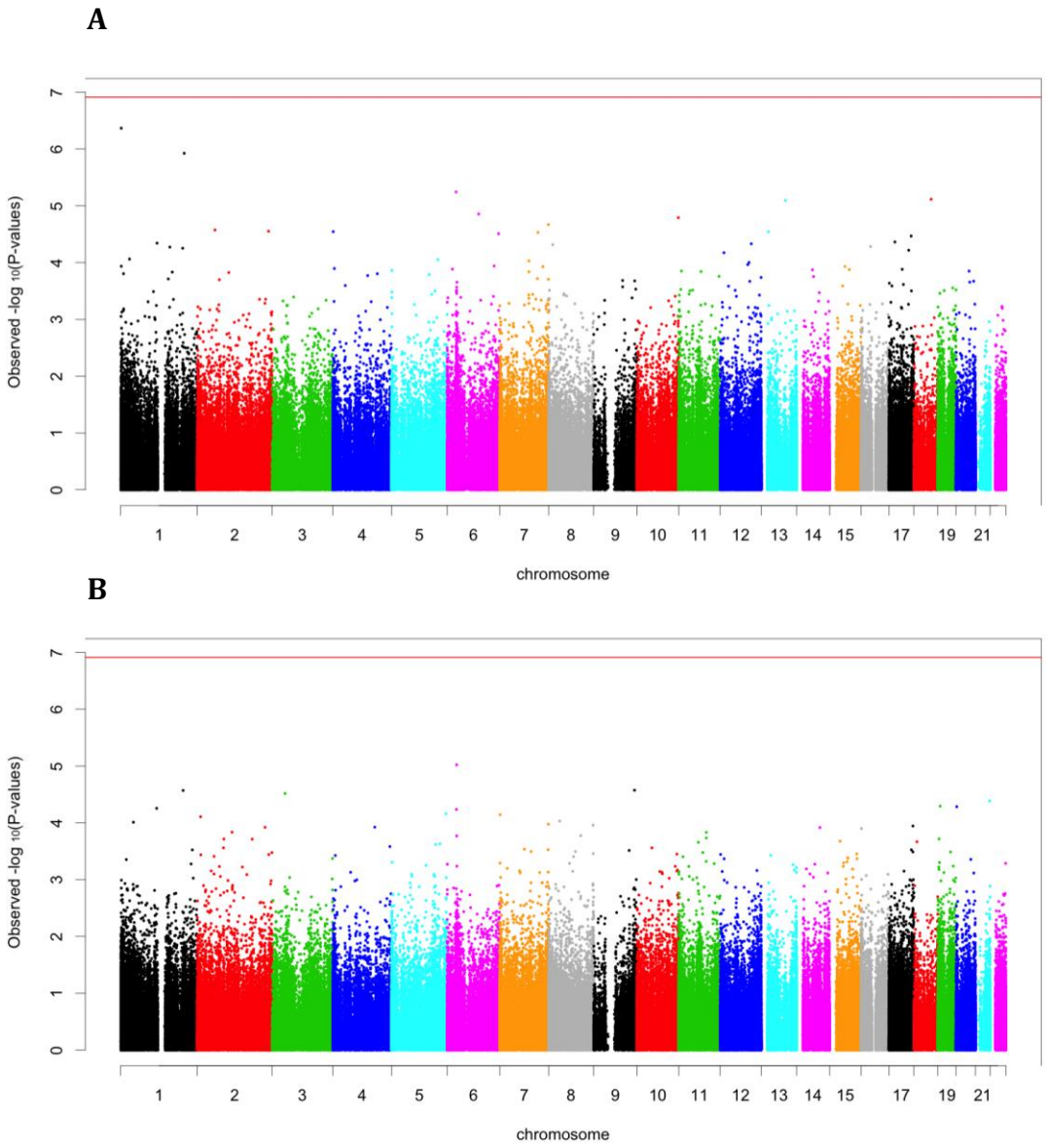

Figure 5. Manhattan plots of methylome-wide association for MMP-9, adjusted for age, BMI, cigarette smoking and cell type proportions of peripheral leukocytes.

Panel A presents the Manhattan plot for MMP-9, for within-twin effect.

Panel B presents the Manhattan plot for MMP-9, for between-twin effect.

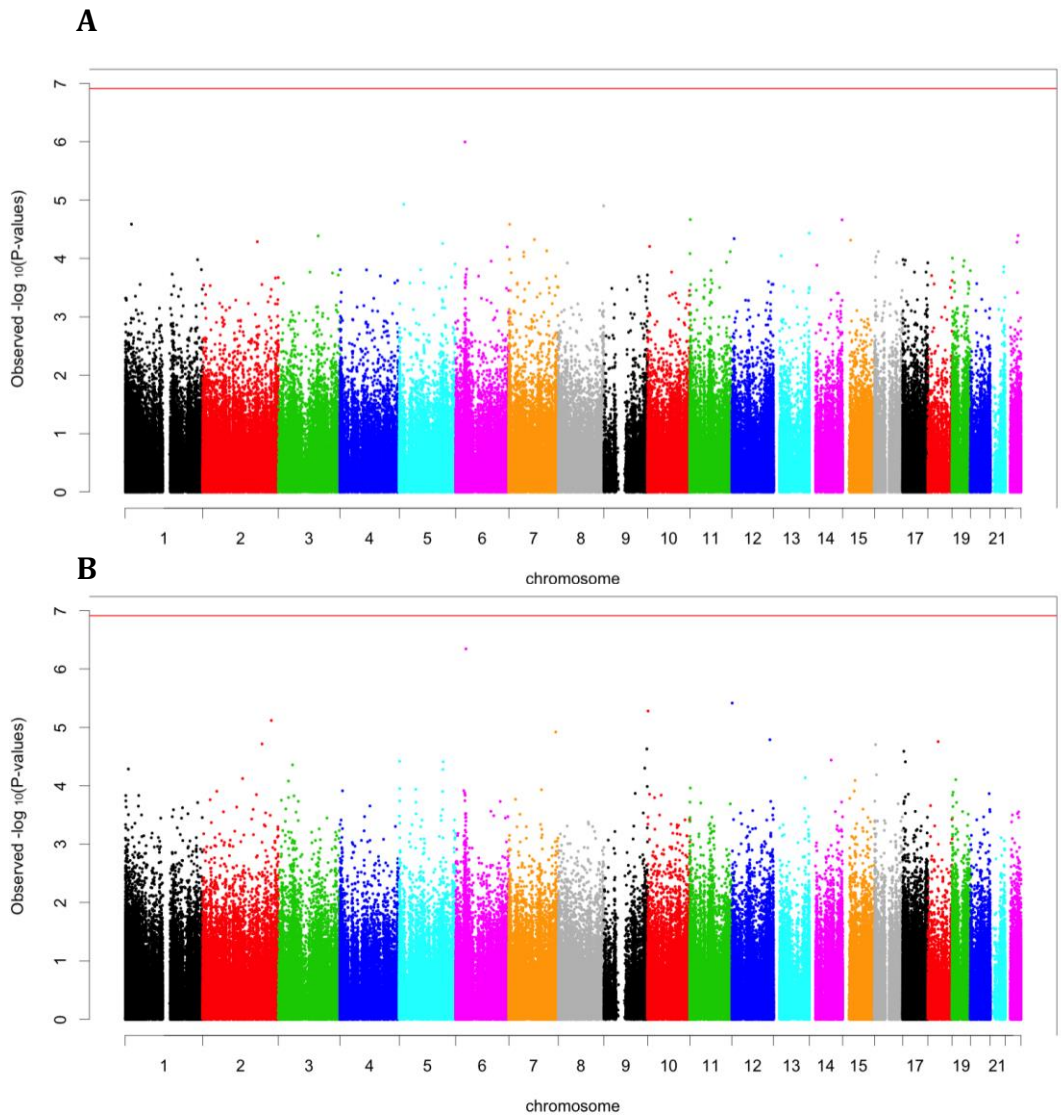

Figure 6. Manhattan plots of methylome-wide association for ICAM-1, adjusted for age, BMI, cigarette smoking and cell type proportions of peripheral leukocytes.

Panel A presents the Manhattan plot for ICAM-1, for within-twin effect.

Panel B presents the Manhattan plot for ICAM-1, for between-twin effect.

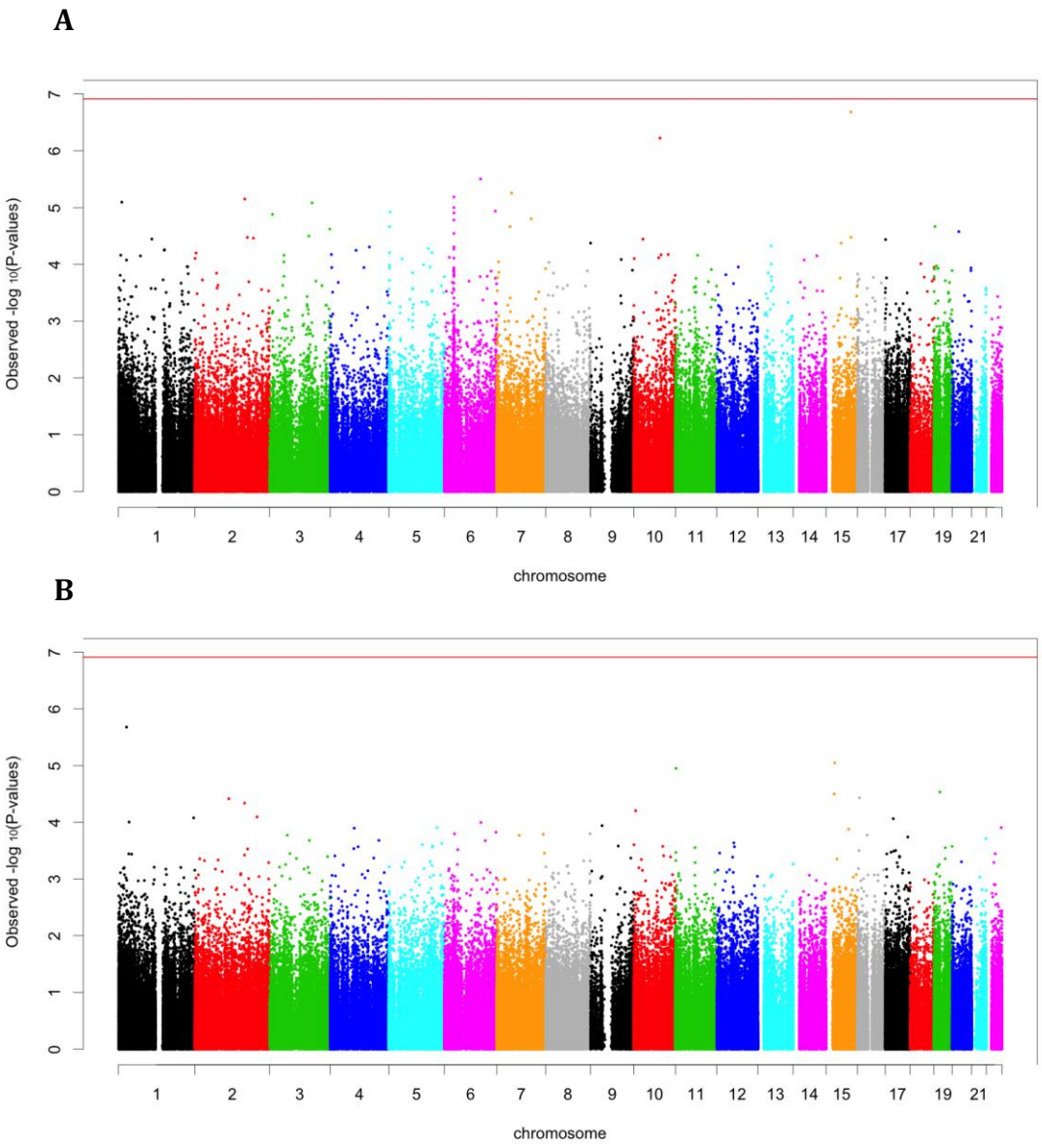

Figure 7. Manhattan plots of methylome-wide association for VCAM-1, adjusted for age, BMI, cigarette smoking and cell type proportions of peripheral leukocytes.

Panel A presents the Manhattan plot for VCAM-1, for within-twin effect.

Panel B presents the Manhattan plot for VCAM-1, for between-twin effect.

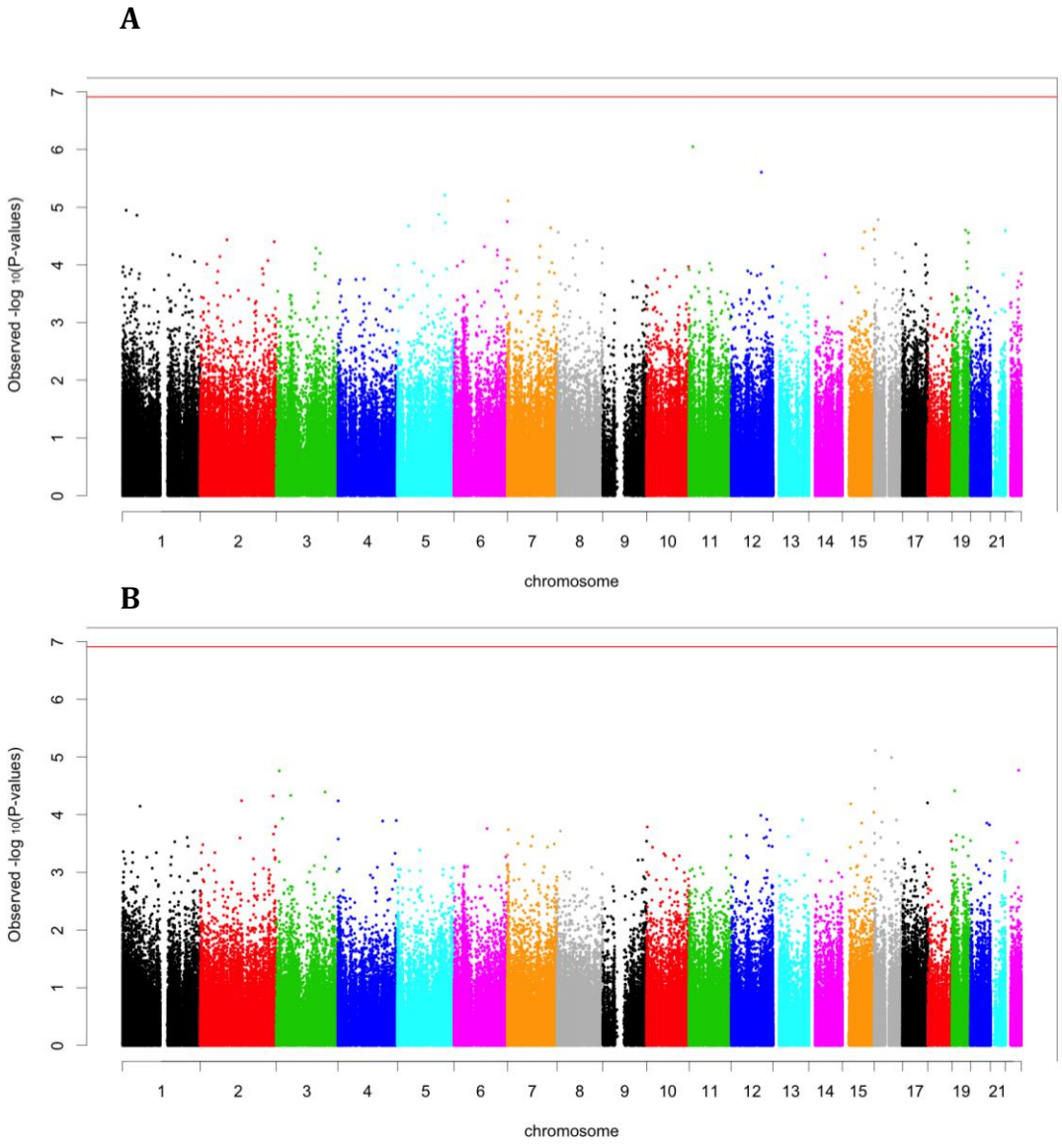

Figure 8. Manhattan plots of methylome-wide association for P-selectin, adjusted for age, BMI, cigarette smoking and cell type proportions of peripheral leukocytes.

Panel A presents the Manhattan plot for P-selectin, for within-twin effect.

Panel B presents the Manhattan plot for P-selectin, for between-twin effect.

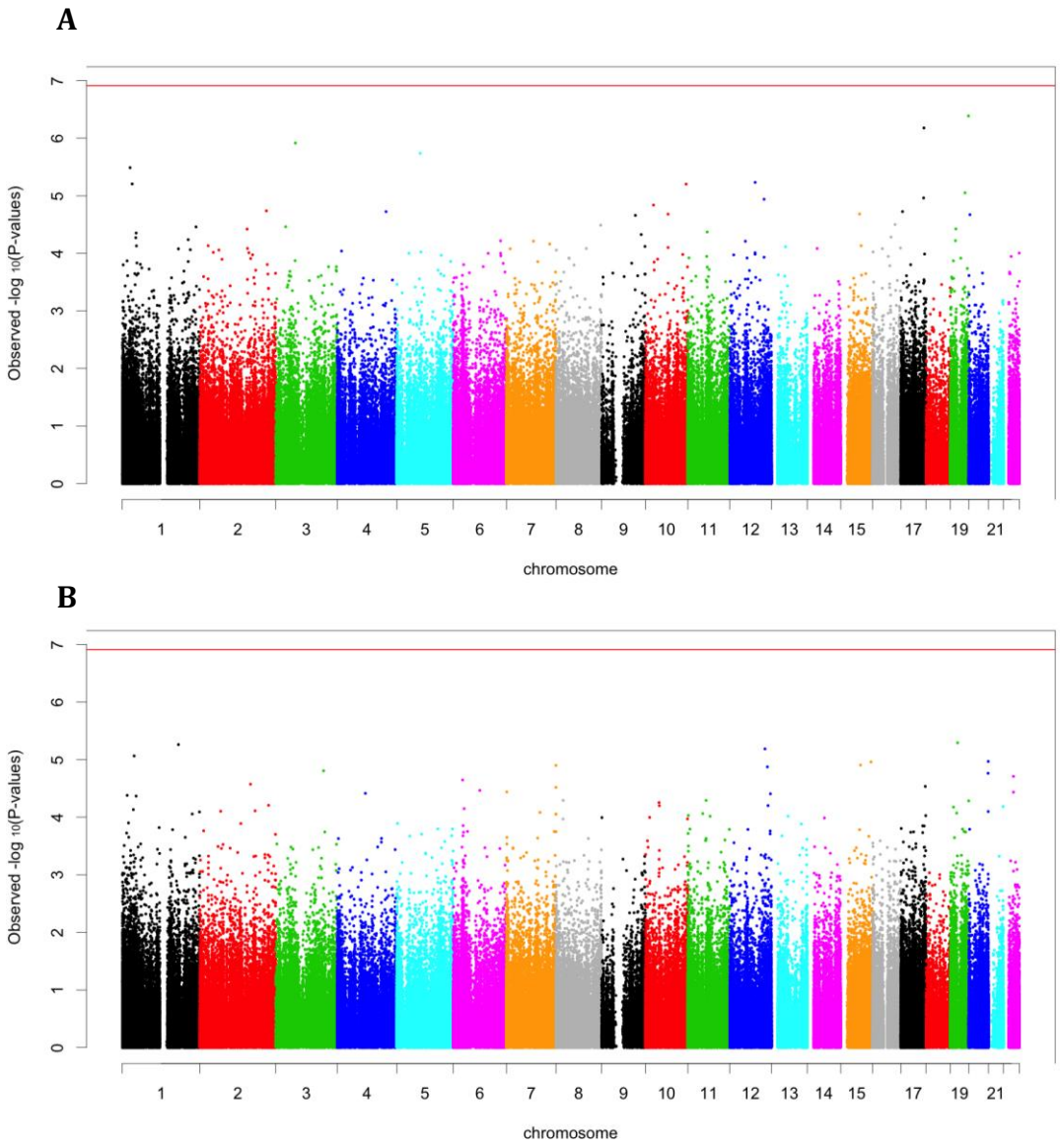

Figure 9. Manhattan plots of methylome-wide association for TGF-β, adjusted for age, BMI, cigarette smoking and cell type proportions of peripheral leukocytes.

Panel A presents the Manhattan plot for TGF-β, for within-twin effect.

Panel B presents the Manhattan plot for TGF-β, for between-twin effect.