GENETICS OF INTRACRANIAL ANEURYSMS AND RELATED DISEASES

FEMKE VAN 'T HOF

ISBN: 978-94-629-5613-1

ProefschriftMaken || www.proefschriftmaken.nl

GENETICS OF INTRACRANIAL ANEURYSMS AND RELATED DISEASES

GENETICA VAN INTRACRANIËLE ANEURYSMATA EN GERELATEERDE AANDOENINGEN

(MET EEN SAMENVATTING IN HET NEDERLANDS)

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de rector magnificus, prof.dr. G.J. van der Zwaan, ingevolge het besluit van het college voor promoties in het openbaar te verdedigen op woensdag 24 mei 2017 des middags te 4.15 uur

door

Femke Neeltje Geertruida van ’t Hof

geboren op 20 januari 1981 te Beverwijk

Promotoren: Prof. dr. G.J.E. Rinkel
Prof. dr. P.I.W. de Bakker

Copromotor: Dr. Y.M. Ruigrok

The research described in this thesis was supported by a grant of the Dutch Heart Foundation (DHF grant number 2008B004)

Financial support by the Dutch Heart Foundation and Brain Center Rudolf Magnus for the publication of this thesis is gratefully acknowledged.

TABLE OF CONTENTS

ADDENDUM

Chapter 1

Introduction

Intracranial aneurysms (IA) are dilatations of the vessel walls of cerebral arteries, and affect approximately 3% of the general population.[1] Most IA remain asymptomatic during life, but some can rupture and result in a subarachnoid hemorrhage (SAH). SAH is a subtype of stroke, with a relatively low mean age of onset (50 years). The prognosis of SAH is poor, with a mortality rate around 30%.[2] The exact pathogenesis of IA development and subsequent SAH is not completely known, but processes like hemodynamic stress, matrix degeneration and inflammation appear to be involved.[3, 4] Modifiable risk factors such as smoking, hypertension and excessive alcohol consumption are important risk factors for IA and SAH, but a positive family history of SAH is the strongest risk factor.[1] Disease risk is increased up to seven fold for first-degree relatives of SAH patients[5], and disease heritability is estimated at 40%.[6] This indicates that inherited genetic variation influences disease risk.

The first studies aiming to uncover the genetic background of IA were mostly linkage analyses or candidate gene studies. However, most of these studies are statistically underpowered to identify disease-associated genes with small effect sizes, and the few associations that were found for IA and SAH have failed to replicate in other, larger studies.[5,7, 8] More recently, genetic research of common diseases has developed towards the assessment of common genetic variants (variants with minor allele frequencies > 5%) with lower penetrance and thus smaller effect size. This approach was based on the hypothesis that common diseases are likely to be caused by the collective effect of many low-penetrance common genetic variants.[9] Microarrays were developed to genotype hundreds of thousands of common genetic variants (so-called single nucleotide polymorphisms (SNPs)) within one experiment. SNPs represented on microarrays were

selected to represent a large fraction of the common genetic variation in the human genome, either directly, or through correlation between SNPs (called linkage disequilibrium). These arrays are used to perform genome-wide association studies (GWAS), which aim to compare the genotypes of SNPs between individuals with and without a common trait or disease, and thus statistically demonstrate the association of certain genomic regions in these traits. In contrast to candidate studies, GWAS are not based on a hypothesis about which gene or genetic region would be involved, but rather interrogate SNPs across the full span of the genome. For IA, GWAS have discovered a role for risk alleles at risk loci near the genes *CDKN2A* and *CDKN2B*,[10-12] *STARD13-KL, RBBP8, SOX17, CNNM2* and *EDNRA*.[10, 11, 13] In addition, common variants at a locus encompassing *HDAC9* were associated in European IA populations,[14] and low-frequency variants (minor allele frequency <5%) near *FSTL1* and *EPM2A* were associated with IA in a Dutch and Finnish cohort.[15]

Risk alleles near *CDKN2A* and *CDKN2B* have also found to be associated with other diseases, including abdominal aortic aneurysms (AAA) and coronary artery disease.[12] This indicates that the same genetic variant can have an effect on multiple, different traits, a phenomenon called pleiotropy. Additional genetic variants with pleiotropic effects could exist for multiple types of aneurysms: because a co-occurrence of IA, AAA and thoracic aortic aneurysms (TAA) has been described,[16, 17,18, 19] mainly within families,[20, 21] some overlap of the genetic background of these diseases can be hypothesized. Similarly, IA could share some genetic risk factors with other related diseases or risk traits such as hypertension.
In parallel to studies focused on disease-associated DNA variations, gene expression studies have been developed to look for RNA changes associated with disease. Under the assumption that multiple genetic factors play a role in complex diseases like IA, genome-wide gene expression studies are performed to identify transcription differences in thousands of genes throughout the genome. For IA, these studies have mainly focused on aneurysm tissue, and differentially expressed genes between IA and control tissue have been identified in multiple studies: *BCL2, COL1A2, COL3A1, COL5A2, CXCL12, TIMP4,* and *TNC.*[22] These and other studies point to the involvement of pathophysiological pathways in IA and IA rupture, like signal transduction, protein binding, focal adhesion and extracellular matrix pathways.[23] Alternatively, measurement of gene expression alterations in blood of IA patients could be effective; blood is easier to obtain, and gene expression profiling in blood has been successful for other vascular diseases, like transient ischemic attacks (TIAs)[24] and ischemic stroke.[25]

Despite the recent discoveries of genetic factors involved in IA, much of the genetic architecture of the disease is still unknown. Previous GWAS focused on common variants and have identified 10 loci associated to disease. These variants only explain a few per cent of the heritability of IA.[13, 15] There are likely many more of such common variants that have yet to be discovered. In addition, part of the heritability could also be explained by genetic variants with lower minor allele frequencies (<5%). These variants can be investigated through sequencing, but despite recent developments in sequencing technologies, this approach is still far more expensive compared to microarray genotyping, and thus cannot easily be performed at large scale. Recently, genotyping arrays were developed for testing low-frequency variants in the exome (the coding region of the genome), allowing us to study these variants in large study cohorts.

Furthermore, the exact role of the known genetic risk variants in IA pathology is unclear: SNPs from GWAS point to a genetic region that could be involved in the disease, but the exact causal variants driving the association, and causal pathways leading to disease (from DNA to transcription products such as RNA and proteins) are mostly unknown. Additionally, we do not know if different IA subgroups harbor some differences in their genetic backgrounds; IA can be present at different locations at the circle of Willis, and some environmental risk factors appear to have different effects on IA depending on this location.[26] Similar to this clinical heterogeneity, genetic heterogeneity could also exist for IA.

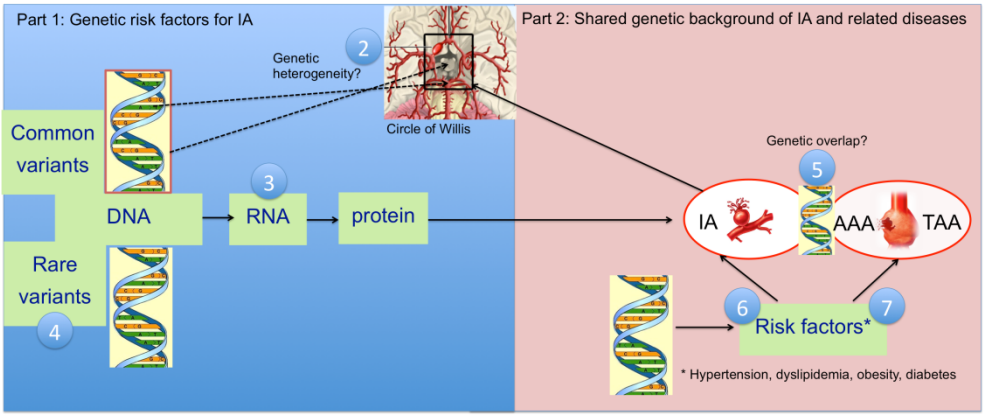


Figure 1. Thesis outline. The numbers refer to the different chapters in this thesis.

Thesis outline

This thesis is set out to elucidate the pathophysiology of IA from a genetic perspective (Figure 1). **Part 1** describes the search for additional genetic factors involved in IA, as well as the investigation of the role of known genetic factors in IA. In **Part 2**, the relation between the genetic background of IA and related traits is investigated.

Part 1: *Chapter 2* focuses on genetic heterogeneity in IA, by describing effect differences of known genetic risk variants for IA on specific IA subtypes. *Chapter 3* describes the results of a genome-wide gene expression study, in which RNA in blood of IA patients is compared with RNA of unaffected individuals. In *Chapter 4*, the search for novel genetic variants involved in IA is described, by means of a GWAS focusing on low-frequency genetic variants located in the exome.

Part 2: *Chapter 5* outlines the results of a GWAS searching for yet unknown common variants involved in aneurysm pathology, by assessing the role of disease pleiotropy and analyzing genotypes of individuals with different types of aneurysms (IA, AAA and TAA) together in one study. This chapter also describes the shared effects of known genetic risk factors of IA, AAA and TAA, and of yet unknown shared polygenic effects. *Chapter 6 and 7* assess the genetic overlap of aneurysms (IA and AAA) and related diseases. *Chapter 6* describes the association between genetic risk profiles of established risk traits for aneurysms (lipid factors, coronary artery disease and blood pressure) and both IA and AAA, while the association of genetic risk profiles of yet unknown but suspected risk traits for aneurysms (obesity and type 2 diabetes) is described in *Chapter 7*.

In *Chapter 8*, I summarize the main results and discuss the implications, challenges, and provide thoughts about future research directions on genetics of IA.

Part I

Genetic risk factors for intracranial aneurysms

Chapter 2

Genetic risk load according to the site of intracranial aneurysms

FNG van 't Hof, MI Kurki, R Kleinloog, PIW de Bakker, M von und zu Fraunberg, JE Jääskeläinen, EI Gaál, H Lehto, R Kivisaari, A Laakso, M Niemelä, J Hernesniemi, MC Brouwer, D van de Beek, GJE Rinkel, YM Ruigrok.

Neurology. 2014 Jul 1;83(1):34-9.

Abstract

Objective

We investigated whether risk alleles of single nucleotide polymorphisms (SNPs) associated with IA are enriched in patients with familial-IA, IA located at the middle cerebral artery (MCA), or IA rupture at a younger age.

Methods

In this case-only study, we calculated genetic risk scores (GRSs) for 973 Dutch and 718 Finnish IA patients by summing effect size weighted risk allele counts of seven SNPs associated with IA previously identified through genome-wide association studies. We tested the GRS for association to presence of familial-IA or IA at the MCA using logistic regression, and to age at time of IA-rupture using linear regression. We also calculated odds ratios (ORs) with 95% confidence intervals for the proportion of patients with each characteristic in the highest compared to the lowest GRS tertile.

Results

GRSs were higher in IA at the MCA in the Dutch (p=2.5x10^{-4}), Finnish (p=0.039), and combined cohort (p=4.9x10^{-5}). GRSs were not associated with familial-IA in the Dutch (p=0.34), Finnish (p=0.45) and combined cohort (p=0.98) or with age at time of IA-rupture in the Dutch (p=0.28), Finnish (p=0.86) and combined cohort (p=0.45). In the combined cohort, ORs were 0.89 (0.67-1.20) for familial-IA, 1.03 (0.79-1.34) for lower age, and 1.54 (1.20-1.98) for MCA-aneurysms.

Conclusions

Our findings suggest that genetic risk factors play a larger role in the development of IA at the MCA than at other sites, and that genetic heterogeneity should be taken into account in future genetic studies.

Introduction

A positive family history is the strongest known risk factor for subarachnoid hemorrhage (SAH) from a ruptured intracranial aneurysm (IA),[5] which suggests heritable DNA variation may influence IA susceptibility. Indeed, genome-wide association studies (GWAS) have identified strong associations between IA risk and seven single nucleotide polymorphisms (SNPs) at six genetic loci encompassing the genes *CDKN2BAS, STARD13-KL, RBBP8, SOX17*, *CNNM2* and *EDNRA*.[10, 11, 13]

It is widely accepted that there is considerable phenotypic and clinical heterogeneity among IA cases. For example, some risk factors appear to have different effects depending on the location of IA at the circle of Willis.[26] Moreover, several characteristics differ between patients with familial-IA and sporadic IA: familial-IA tend to rupture at a younger age and are more often located at the middle cerebral artery (MCA).[27-29] In general, genetic studies do not take the phenotypic heterogeneity of IA into account.[8] Explicitly accounting for such heterogeneity might be useful for future studies to gain more insight into commonalities and differences between IA subtypes.

To characterize the clinical heterogeneity of IA, we studied whether risk alleles of validated GWAS SNPs associated with IA are enriched in patients with familial-IA or with familial-IA-associated characteristics: a young age at time of SAH or IA located at the MCA. We performed a case-only study in two separate cohorts of IA patients, one from the Netherlands and another from Finland.

Methods

Patients

Patients in the Dutch cohort were admitted to the University Medical Center Utrecht, the Netherlands between 1997 and 2011. The total study population consisted of 1080 IA patients. The Finnish cohort consisted of 790 IA patients treated at the Helsinki and Kuopio University Hospitals. Both cohorts included patients with ruptured and unruptured IA. Ruptured IA cases were defined by symptoms suggestive of SAH combined with subarachnoid blood on a computed tomography (CT) scan and a proven IA at angiography (conventional angiogram, CT- or magnetic resonance (MR)-angiogram). Unruptured IA cases were identified by CT or MR angiography or conventional angiography in the absence of clinical or radiological signs of SAH. Patients with fusiform IA, possible traumatic SAH, and polycystic

kidney disease were excluded. Further details of these study populations have been described previously.[10]

Phenotypic data

Aneurysm characteristics (total number and location of IA) and patient characteristics (gender, age at time of blood sampling and age at time of SAH, if applicable) were retrieved from the databases of patients with SAH and patients with unruptured IA (Dutch cohort) or from the medical reports (Finnish cohort). Familial-IA was defined as having one or more first-degree relative(s) with SAH or IA, according to a self-reported family history. Based on the phenotypic data, patients were divided into subgroups according to presence or absence of familial-IA, and according to presence or absence of at least one IA located at the MCA.

Genotypic data

We retrieved the genotypes of SNPs with an established genome-wide significant association with IA in patients with a European ancestry, according to previous GWAS. These seven SNPs were rs6841581 (locus 4q31.23),[13] rs10958409 (8q11.23), rs9298506 (locus 8q12.1), rs1333040 (locus 9p21.3), rs12413409 (locus 10q24.32), rs9315204 (locus 13q13.1) and rs11661542 (locus 18q11.2).[10, 11] The two SNPs at locus 8q were independent (in low linkage disequilibrium; r^2=0.058).

All patients in the Finnish cohort and 785 of the 1080 patients of the Dutch cohort were already genotyped for six of the risk SNPs (rs10958409, rs9298506, rs1333040, rs12413409, rs9315204 and rs11661542) in previous GWAS.[10, 11] We performed quality control of SNPs and patients based on GWAS data, and imputed genotypes of the SNP rs6841581 that was not genotyped in previous GWAS (e-Methods). The remaining 295 study patients in the Dutch cohort were genotyped only for these seven SNPs (e-Methods). In this group, patients were excluded when the genotype of at least one of the seven SNPs was missing.

With the genotypes of the seven risk SNPs, we calculated a genetic risk score (GRS) for each study patient as follows. For each SNP, the number of genotyped risk alleles (0, 1 or 2) or the estimated allele dosage after imputation (between 0 and 2) was multiplied by its effect size, defined as the natural log of the published odds ratio (OR) for that risk allele.[11, 13] We summed up these effect size weighted allele counts across all seven SNPs to generate a GRS in each study participant.

Association analyses

We tested the association between GRS and each characteristic in the Dutch and Finnish cohort separately, and in the two cohorts combined. Patients with no information about family history were excluded from the analysis on familial-IA. For the characteristics 'familial-IA' and 'IA located at MCA', we tested the association between presence of the characteristic and GRS using logistic regression. For the characteristic 'age at time of SAH', we tested the association between age and GRS using linear regression. Adjustment for population (Dutch or Finnish) was applied to the analyses of the two cohorts combined. Associations with $p<0.05$ were considered significant. R version 2.15.2[30] was used for all statistical analyses.

For each statistically significant association, we performed an additional analysis with adjustment for gender and age (or only gender, when analyzing the characteristic 'age at time of SAH'). We also tested the association between each individual SNP from the GRS and the characteristic, in order to investigate whether the association of the GRS was driven by one or more SNPs.

To evaluate the effect size of each genetic score on each characteristic,we performed an additional analysis dividing the study population into tertiles based on the respective GRS distribution. We then calculated ORs and corresponding 95% confidence intervals (CI) for the proportion of patients with each characteristic in the highest compared to the lowest GRS tertile. For the characteristic 'age at time of SAH', we tested the proportion of patients with an age below the median age at time of SAH in the highest compared to the lowest GRS tertile.

Sensitivity analyses

As the accuracy of self-reporting of a familial history of IA is known to be modest,[31] we performed an additional analysis for the characteristic 'familial-IA'. Patients were marked as having definite familial-IA when the episodes of SAH or IA in the first-degree relatives were confirmed by medical reports (Dutch cohort only). We performed the analyses described above after dividing patients into subgroups according to presence of definite familial-IA, or absence of familial-IA. Also for the characteristic 'IA at MCA' we performed two analyses: 1. by including all patients with single and multiple IA, in order to keep maximal sample sizes in the subgroups, and 2. by including only patients with a single IA, in order to obtain more homogeneous subgroups.

Population stratification

Principal component analysis was initially applied to detect population stratification outliers (e-Methods). To check whether population structure influenced the association analyses, we repeated the analysis of any significant findings both in the Dutch and the Finnish cohort, including the first four principal components as covariates. For the Dutch cohort, this additional analysis was restricted to the patients who were part of the GWAS cohort.

Standard protocol approvals, registrations, and patient consents

For both the Dutch and the Finnish cohort, the local ethics committees approved the study and all patients gave written informed consent.

Results

Quality control and baseline characteristics

From the total group of 1080 Dutch patients, 71 patients were excluded for not passing GWAS quality control, and 36 patients because of missing genotypes; thus, 973 patients were available for the analysis. From the Finnish replication cohort of 880 patients, 120 GWAS quality control outliers and 42 patients without phenotype information were excluded; thus, 718 patients were available for the analysis. Table 1 shows the baseline characteristics of both study cohorts. The distributions of the GRS in patients with and without presence of familial-IA, age at the time of SAH below the median age, and IA located at the MCA in the Dutch and the Finnish cohorts are shown in figure e-1 and e-2.

Association analyses

Table 2 shows the association between GRS or GRS tertile and the presence of familial-IA, age at time of SAH, and presence of IA located at the MCA in the Dutch and Finnish cohort, and in both cohorts combined. We did not observe an association between GRS and familial-IA and between GRS and age at time of SAH in both the Dutch cohort and the Finnish cohort. In contrast, a higher GRS was associated with IA located at the MCA in both cohorts. Adjustment for sex and age did not qualitatively change the results in the Dutch cohort (p=2.7 x 10^{-4}) and the Finnish cohort (p=0.067). Adjustment for the first four principal components did not change the association results in the restricted Dutch GWAS cohort (p=0.016, compared to p=0.020 without adjustment) and in the Finnish cohort (p=0.031).

When testing the individual IA risk alleles for association to IA at the MCA, we observed that multiple alleles contributed to the association of the GRS and this characteristic; all seven risk alleles had an OR ≥ 1.0 in the Dutch cohort, and six out of seven risk alleles had an OR ≥ 1.0 in the Finnish cohort (Table 3).

Sensitivity analyses

When restricting the analysis on 'familial-IA' in the Dutch cohort to 56 patients with definite familial-IA compared to 739 patients without familial-IA, we found a trend between higher GRS and definite familial IA (p=0.087). The OR for the proportion of patients with definite familial-IA in the highest GRS tertile compared to the lowest tertile was 1.73 (95% CI 0.82-3.80), which was higher than the OR of 1.14 (95% CI 0.66-1.97) for familial-IA based on self-reporting of family history.

For the analysis on 'IA at MCA' restricted to patients with single IA only, we still observed a higher GRS in patients having IA located at the MCA in the Dutch cohort (p=2.3 x 10^{-3} for 158 patients with vs. 618 patients without IA at the MCA), a trend towards association in the Finnish cohort (p=0.053 for 189 patients with vs. 299 patients without IA at the MCA), and an association in both cohorts combined (p=3.8 x 10^{-4}). The OR for the proportion of patients with single IA located at the MCA in the highest GRS tertile compared to the lowest tertile was 1.82 (95% CI 1.16-2.88) in the Dutch cohort, 1.49 (95% CI 0.93-2.39) in the Finnish cohort, and 1.56 (95% CI 1.14-2.14) in both cohorts combined. Thus, the association results remained essentially unchanged when restricting the analysis to patients with single IA only.

Discussion

Established genetic risk variants for IA are enriched in patients with IA located at the MCA. This finding was significant in the Dutch cohort and in the Finnish cohort of IA patients, which confirms the validity of this observation. Multiple IA risk SNPs included in the GRS contributed to this effect. In contrast, no significant genetic enrichment was observed in patients with familial-IA and in patients with a young age at time of SAH.

Recently, the FIA study group reported that the concordance of the site of IA is high between familial-IA patients from the same family. This was found for IA at the internal carotid artery, MCA and vertebrobasilar system.[32] This observation supports the notion that genetic factors are involved in determining IA location.

The genetic heterogeneity between IA locations adds to previous findings of clinical differences according to IA location. Some risk factors for IA predispose to IA at different locations.[26, 33-35] For example, associations were described between younger age and MCA aneurysms,[26] female gender and internal carotid artery aneurysms, and male gender and anterior cerebral artery aneurysms.[34, 35] Also, certain clinical features, such as risk of rupture, differ between IA locations.[15, 16] One previous study also reported genetic heterogeneity according to IA location. In a Japanese cohort, risk variants at locus 9p21 were more frequent in patients with IA in the posterior circulation as compared to other locations.[36] However, as our study findings are generalized to all known IA SNPs instead of only one locus, and as different classifications of IA locations were used, the results of this Japanese study are not comparable to our findings.

The absence of an association between GRS and familial-IA in our study may be caused by misclassification in the familial-IA subgroup, as the accuracy of self-reporting of a familial history of IA is known to be modest. In a previous Dutch study, the sensitivity of reporting presence of SAH in first-degree family members was only 0.75.[31] Indeed, we observed a stronger association between GRS and familial-IA after restricting the analysis to patients in which familial-IA was confirmed by medical reports. However, the association did not reach statistical significance, possibly due to limited statistical power in this small subgroup.

In contrast to our hypothesis, we did not observe an association between high GRS and low age at time of SAH. This finding implies that the age at which aneurysms rupture is mainly determined by other (genetic or environmental) factors than the known genetic risk factors for IA.

There are several strengths of this study. First, we tested the association of the characteristics independently in two large study cohorts with a different ancestry. The significant genetic enrichment in patients with IA at the MCA was observed in both the Dutch and the Finnish study cohort. This is a strong argument against population-specificity of this association, especially given the known differences in IA characteristics between the two patient populations. For example, IA located at the MCA, male sex and multiple IA are more common in Finnish than in Dutch SAH patients.[37] This could be due to population differences in prevalence of genetic risk factors, as the Finnish population is a genetic isolate that harbours a higher prevalence of some variants that are rare outside Finland.[38]

Second, we captured the cumulative effect of all IA risk alleles in a GRS, which is an effective method in the genetic differentiation of disease subgroups.[39, 40] Third, for the association between GRS and IA at the MCA, we excluded the possibility that heterogeneity in the location subgroups, caused by inclusion of patients with multiple IA at different locations, influenced our findings; narrowing the analysis to patients without multiple IA resulted in comparable association results. In addition, we did not observe confounding effects of age, sex or population stratification on this association.

A limitation of this study is the moderate power of the GRS, which is limited by the number of validated SNP associations for IA identified through GWAS. As the IA risk loci known to date explain only 4% of the estimated heritability of the disease in European samples,[11] the GRS consequently represents only a fraction of the heritable risk for IA. We found strong evidence of an association between genetic risk and IA at the MCA despite this limitation, but discovery of more risk loci in future genetic studies can lead to more accurate genetic risk estimations.

Overall, we found that common genetic variants associated with IA are enriched in patients with IA located at the MCA. This might suggest that the seven risk loci identified to date play a more important role in the development of IA at the MCA than at other sites. This difference in genetic load according to IA location also implies that IA is not only clinically, but also genetically a heterogeneous disease. How an increased genetic burden of IA risk alleles predisposes to IA located at MCA will require a more complete understanding of the underlying mechanisms involved. Future genetic studies should add subgroup analyses of IA at different locations to gain more insight into the complex relation between genetic variation and risk of IA.

Tables

Table 1: Baseline characteristics of the study populations

Characteristics	Dutch population (n=973)	Finnish population (n=718)
Mean age in years	55 (range 16 – 89)	45 (range 11 – 89)
Mean age at time of SAH in years	51 (range 10 – 88)	45 (range 11 – 89)
Women (N/T)	672 / 973 (70%)	417 / 718 (58%)
Familial-IA (N/T)	114 / 854 (13%)	281 / 718 (39%)
SAH (N/T)	880 / 973 (90%)	535 / 718 (75%)
Presence of multiple IA (N/T)	183 / 966 (19%)	242 / 718 (34%)
Presence of IA located at MCA (N/T)	254 / 973 (26%)	358 / 718 (50%)

SAH indicates subarachnoid hemorrhage; N, number of patients with characteristic; T, total number of patients with data on characteristic; IA, intracranial aneurysm; MCA, middle cerebral artery

Table 2: Association between characteristics of IA patients and GRS and in the Dutch and Finnish cohorts

	Characteristic		
	Familial-IA	Age at time of SAH	IA at MCA
Dutch cohort (n=973)			
Characteristic + / - (n)	114 / 740	51[*]	254 / 719
Characteristic vs. GRS (p-value) [†]	0.34	0.28	2.5×10^{-4}
Characteristic in GRS T3 / T1 (n) [‡]	36 / 31	151 / 138	103 / 67
Characteristic in GRS T3 vs. T1 (OR (95% CI)) [§]	1.14 (0.66 – 1.97)	1.24 (0.89 - 1.73)	1.80 (1.24 - 2.62)
Finnish cohort (n=718)			
Characteristic + / - (n)	281 / 437	45[*]	360 / 112
Characteristic vs. GRS (p-value) [†]	0.45	0.86	0.039
Characteristic in GRS T3 / T1 (n) [‡]	85 / 101	85 / 93	128 / 112
Characteristic in GRS T3 vs. T1 (OR (95% CI)) [§]	0.81 (0.55 - 1.19)	0.75 (0.48 - 1.17)	1.42 (0.98 - 2.07)
Dutch and Finnish cohorts combined (n=1691)			
Characteristic + / - (n)	395 / 1177	50[*]	612 / 1079
Characteristic vs. GRS (p-value) [†]	0.98	0.45	4.9×10^{-5}
Characteristic in GRS T3 / T1 (n) [‡]	121 / 132	236 / 231	231 / 179
Characteristic in GRS T3 vs. T1 (OR (95% CI)) [§]	0.89 (0.67 - 1.20)	1.03 (0.79 - 1.34)	1.54 (1.20 – 1.98)

IA indicates intracranial aneurysm; GRS, genetic risk score; SAH, subarachnoid hemorrhage; MCA, middle cerebral artery; T, Tertile; OR, odds ratio; CI, confidence interval; NA, not applicable.

[*] *mean age at time of SAH*

† *p-value from logistic regression for association between GRS and presence of characteristic ('Familial-IA' or 'IA at MCA'), or p-value from linear regression for association between GRS and age at time of SAH.*
‡ *Number of patients with the characteristic (or with an age below the median, for the characteristic 'Age at time of SAH') in the highest versus the lowest tertile*
§ *Proportion of patients with the characteristic (or with an age below the median, for the characteristic 'Age at time of SAH') in the highest versus the lowest GRS tertile*

Table 3: Association between individual IA risk SNPs and IA located at MCA in the Dutch and Finnish cohorts

			Dutch cohort				Finnish cohort			
			RAF in MCA subgroups		Risk alleles in MCA+ vs MCA-		RAF in MCA subgroups		Risk alleles in MCA+ vs MCA-	
SNP	Locus	Risk allele	MCA+	MCA-	OR	95% CI	MCA+	MCA-	OR	95% CI
rs6841581	4q31.23	G	0.91	0.87	1.08	1.02 - 1.14	0.90	0.88	1.23	1.08 - 1.40
rs10958409	8q11.23	A	0.20	0.18	1.02	0.97 - 1.07	0.24	0.20	1.24	1.11 - 1.36
rs9298506	8q12.1	A	0.85	0.85	1.00	0.94 - 1.05	0.80	0.80	1.05	0.92 - 1.17
rs1333040	9p21.3	T	0.65	0.60	1.05	1.00 - 1.09	0.53	0.51	1.07	0.96 - 1.17
rs12413409	10q24.32	G	0.94	0.93	1.02	0.95 - 1.11	0.93	0.93	0.95	0.75 - 1.16
rs9315204	13q13.1	T	0.25	0.23	1.02	0.98 - 1.07	0.41	0.36	1.21	1.10 - 1.31
rs11661542	18q11.2	C	0.60	0.54	1.05	1.01 - 1.09	0.48	0.47	1.07	0.97 - 1.17

SNP indicates single nucleotide polymorphism; RAF, risk allele frequency; MCA, middle cerebral artery; OR, odds ratio; CI, confidence interval.

Supplemental data
e-Methods, e-Tables and e-Figures are available with the online version of this paper.

Chapter 3

Whole blood gene expression profiles of patients with a past aneurysmal subarachnoid hemorrhage

FNG van 't Hof, YM Ruigrok, J Medic, B Sanjabi, P van der Vlies, GJE Rinkel, JH Veldink.

PLoS One. 2015 Oct 6;10(10):e0139352.

Abstract

Background

The pathogenesis of development and rupture of intracranial aneurysms (IA) is largely unknown. Also, screening for IA to prevent aneurysmal subarachnoid hemorrhage (aSAH) is inefficient, as disease markers are lacking. We investigated gene expression profiles in blood of previous aSAH patients, who are still at risk for future IA, aiming to gain insight into the pathogenesis of IA and aSAH, and to make a first step towards improvement of aSAH risk prediction.

Methods and Results

We collected peripheral blood of 119 patients with aSAH at least two years prior, and 118 controls. We determined gene expression profiles using Illumina HumanHT-12v4 BeadChips. After quality control, we divided the dataset in a discovery (2/3) and replication set (1/3), identified differentially expressed genes, and applied (co-)differential co-expression to identify disease-related gene networks. No genes with a significant (false-discovery rate <5%) differential expression were observed. We detected one gene network with significant differential co-expression, but did not find biologically meaningful gene networks related to a history of aSAH. Next, we applied prediction analysis of microarrays to find a gene set that optimally predicts absence or presence of a history of aSAH. We found no gene sets with a correct disease state prediction higher than 40%.

Conclusions

No gene expression differences were present in blood of previous aSAH patients compared to controls, besides one differentially co-expressed gene network without a clear relevant biological function. Our findings suggest that gene expression profiles, as detected in blood of previous aSAH patients, do not reveal the pathogenesis of IA and aSAH, and cannot be used for aSAH risk prediction.

Introduction

Subarachnoid hemorrhage (SAH) from a ruptured intracranial aneurysm (IA) is a severe subtype of stroke, occurring in relatively young people (mean age 50 years), of whom a third dies as a consequence of the aneurysmal SAH (aSAH).[2] It is known that both environmental exposures and genetic predisposition play a role in susceptibility of aSAH,[30] with an estimated heritability of around 40%.[6] The exact pathogenesis of IA development and subsequent aSAH is not exactly known, but processes like hemodynamic stress, matrix degeneration and inflammation appear to play a role.[3, 4]

Around 10% of the aSAH patients has one or more first degree relatives with aSAH, and unaffected first degree relatives are at increased risk of developing an aneurysms and having an aSAH.[41] IA are generally asymptomatic before rupture, and therefore have to be detected by screening. Magnetic resonance angiography (MRA) is currently the standard screening method for individuals at high risk for IA development and subsequent rupture, but screening has disadvantages in terms of costs and negative consequences.[42, 43] Moreover, screening is inefficient in first-degree relatives if only one relative is affected, although they have an increased life time risk of aSAH.[41, 44] Thus, we need tools to better detect persons with high risk of aneurysm development or rupture. Gene expression profiling in blood of previous aSAH patients may help to identify individuals who are at high risk for aSAH. Given the high long-term risk for developing new aneurysms in previous aSAH patients, aSAH seems to be a continuous disease of the vessel wall, making these patients suitable subjects for studying ongoing pathophysiologic processes involved in IA.[45]

Therefore, in this study, we compared gene expression profiles in blood between individuals who had survived an episode of aSAH and healthy controls. We aimed to gain more insight into the pathogenesis of IA and aSAH, and also to see whether these gene expression profiles can improve identification of individuals with an increased risk of aSAH.

Materials and methods

Study design and subjects

Between August 2010 and January 2011, we sampled blood from 119 persons who had been treated for aSAH in the University Medical Center Utrecht (UMCU), the Netherlands. All patients visited the outpatient clinic at the UMCU for blood sampling. We included only patients who had the last

episode of aSAH at least two years (median 7.5 years; range 2-23 years) before the blood sample collection to minimize the chance of detecting direct effects of the bleeding on gene expression profiles. Aneurysmal SAH was defined by symptoms indicative of SAH combined with subarachnoid blood on a computed tomography (CT) scan and a proven aneurysm at angiography (conventional angiogram, CT- or magnetic resonance (MR)-angiogram). Ruptured IA were treated by operative clipping or by coiling. In a subgroup of 16 patients, one or multiple unruptured IA were found in addition to the ruptured IA. Of these, five patients had an IA that was left untreated. The controls were genetically unrelated individuals accompanying the patient to the outpatient clinic (mostly spouses of the aSAH patients). When such unrelated individuals were unavailable, spouses of other patients visiting the neurology outpatient clinic served as controls. In total, we included 118 controls. For all participants, we obtained information about age, smoking history, hypertension (defined as a self-reported history of hypertension and/or use of antihypertensive medication) and presence of familial IA (defined as having one or more first-degree relative(s) with SAH or IA). All controls confirmed a negative history of SAH or IA.

The study was approved by the Medical Ethics Committee at the University Medical Center Utrecht, and all participants provided written informed consent.

Blood sample collection and processing

Blood samples were obtained in the morning after overnight fasting. In each participant (i.e. cases and controls), we collected two PAXgene tubes (Qiagen) for genome-wide gene expression, and an EDTA tube for measurement of leukocyte differential counts. PAXgene tubes were frozen at -20 °C after two hours at room temperature, until RNA was isolated using PAXgene extraction kits (Qiagen). We excluded five cases and three controls with low RNA quality or quantity, defined as an RNA integrity number (RIN) value below 6. We also excluded one case with low RNA quantity after RNA amplification, defined as a 260/280 ratio below 1.8, measured using nanodrop (www.nanodrop.com). The remaining samples were hybridized to Illumina HumanHT-12v4 Expression BeadChips.

Data quality control

R version 2.15.2 was used for quality control and statistical analysis of gene expression data.[46] After calculating principal components (PCs), we excluded nine samples identified as outliers based on visual inspection of PC plots (S1 Fig.). Seven samples showing inconsistency between reported gender and expression data based on at least two out of eight

non-pseudoautosomal sex chromosome transcripts, and two duplicate samples (>99% gene expression correlation, measured using Pearson's correlation coefficient) were excluded. In total, ten cases and eight controls were excluded after quality control. All probe sequences were aligned to the NCBI build 36 reference genome using UCSC's Genome Browser function BLAT.[47] We removed non-specific probes, defined as no or multiple hits with a sequence homology >95%, and non-autosomal probes (n = 13 188). Probes mapping to transcripts designated as 'retired' according to RefSeq (updated on 27 September 2010) and UniGene (build #228, release data 29 October 2010) databases were also excluded. After exclusion of sample and probe outliers, the raw dataset was again quantile normalized and $\log_2$ transformed before further analyses.

Data analysis

After exclusion of samples that did not surpass the quality control, we divided the remaining samples in a discovery set (2/3 of the dataset) and a replication set (1/3 of the dataset), with an equal distribution of cases and controls in each set. We performed four different analyses to investigate the gene expression profiles. Analyses 1 to 3 were aimed at gaining more insight in the pathogenesis of IA and aSAH and analysis 4 at exploring the possible use of these profiles in prediction of aSAH risk.

1. Differential expression

We calculated case-control differences in expression of all probes in the discovery set, using logistic regression. To eliminate expression heterogeneity caused by known and unknown technical and biological background, data were normalized applying surrogate variable analysis (SVA). This produces surrogate variables for which expression levels can be corrected by calculating residuals in a linear regression model. This reduces batch specific background noise thereby increasing the ability to detect biologically meaningful signals.[48] As covariables, we included the 14 variables as specified by the SVA procedure. To correct for multiple testing, we calculated Benjamini Hochberg false discovery rates (FDR). All probes with a FDR-corrected p-value (p_{FDR}) below 0.05 were tested for case-control difference in the replication set. Probes with $p_{FDR} < 0.05$ in the replication set were marked as significant.

Next, we created a list of 69 genes with a previously described association with IA (S1 Table), consisting of all genes in significant loci from previous genome-wide association studies (GWAS),[10, 11, 13] and genes associated with IA in at least three gene expression studies in IA tissue.[22] We checked the p-value from the differential expression analysis for probes mapping to these genes. Probes were significant if p-values were below 0.05 after

Bonferroni correction for the number of tested probes.

2. Co-differential co-expression (CDC)

Differentially expressed genes can be interacting with other genes to generate their effect on diseases. We tried to identify a differential gene regulating network involved in IA, based on co-differential co-expression.[49, 50] Therefore, we investigated which genes have a similar pattern of differential expression in IA cases and controls. First, we calculated residuals of gene expression levels after adjustment for known risk factors for SAH: age, sex, hypertension, smoking (ever or never) and familial IA. Leukocyte differential counts were not included as co-variables in the CDC analysis, because no significant differences (threshold $p<0.05$ using two-sample Wilcoxon test) between cases and controls were observed (S2 Table).

For the co-expression analyses, we used Weighted Gene Co-expression Network Analysis.[51] An adjacency matrix was defined between all genes under study, based on pair-wise correlations in all subjects in the discovery set. We used the Spearman rank correlations, to avoid the leverage of influential outliers.[52] On this co-expression matrix, standard hierarchical clustering with average linkage was applied, followed by gene group extraction from the resulting dendrogram, using a fixed cut height (0.96).[52] We created modules of co-expression based on the adjacency matrix, and for each gene we matched the significance level of differential expression to the module assignment. Next, we calculated the average gene significance level in each module based on analysis 1, and tested whether this level was significantly higher than expected by chance ($p<0.05$).

To test the reproducibility of these significant modules, we calculated preservation statistics for each module in the replication set, as described previously.[53] Specifically, the Z-summary statistic was investigated. This statistic captures the density and connectivity statistics, and adjusts for module size. DC modules with a Z-summary statistic > 10 (threshold for strong preservation evidence[53]) were considered as significantly preserved.

3. Differential co-expression (DC)[52, 54]

Standard differential expression mainly investigates regulatory genetic variation that leads to expression level changes between cases and controls. Known disease genes, however, are often not differentially expressed because mutations in the coding region can affect the interaction of the gene with other genes, which will affect the co-expression pattern of the gene. Networks from gene expression data can be inferred by

calculating all pair-wise correlations of the genes in a diseased and a control state and compare these networks based on differential co-expression.[52]

First, we created an adjacency matrix as described in paragraph 2 ('Co-differential co-expression'), but this time by investigating the cases and controls separately, instead of all subjects in the discovery set together. Then the co-expression changes were computed from the difference in adjacency matrices. Modules of differentially co-expressed genes were extracted from this matrix, and these modules were randomly color-labeled. The statistical significance of differential co-expression of gene groups was assessed using 10 000 permutations of the data to generate a null distribution of the dispersion statistic, followed by a Bonferroni correction of the empirical p-values.[52] The dispersion statistic (D_s) is a measure of correlation change for groups of genes.[24]

We selected modules with a Bonferroni-corrected p-value (p_{Bonf}) below 0.05 for preservation testing in the replication set as described above. Next, we created correlation heatmaps of differentially co-expressed gene modules for the discovery set and replication set separately, using hierarchical cluster analysis based on gene correlation values.

Biological relevance of gene modules

We investigated the biological meaning of preserved modules based on (co-)differential co-expression in two ways. First, we tested the enrichment of genes involved in biological pathways in each module, using the database for annotation, visualization and integrated discovery (DAVID).[55] Biological pathways with $p_{Bonf} < 0.05$ were considered as significantly enriched. Second, we determined hub nodes in each module. Hub nodes or "hubs" take a central position in a network. They can be easily reached by most of the nodes of the network due to their central position.[56, 57] Several metrics can be used to identify hubs. Hub nodes generally display an above average high number of connections to other nodes in the network, a high level of (betweenness) centrality, a short average distance towards the other nodes of the network, and a low clustering coefficient.[56, 57] We identified hubs by computing a level of "hubness" for each node determining whether a node belonged to: (1) the top 20% of nodes showing the highest level of connectivity; (2) the top 20% of nodes showing the highest level centrality; (3) the top 20% of nodes showing the lowest path length; and/or (4) the top 20% nodes showing the lowest clustering coefficient.[58] Each node was assigned a score between 0 and 4, determined by the total number of hub criteria fulfilled. Regions showing a

hub-score of 2 or higher were marked as hub nodes.
We looked up the function and published disease associations for the most important hub nodes at the website of the National Center of Biotechnology Information (http://www.ncbi.nlm.nih.gov). Next, we investigated whether there was any overlap between these hub genes and the list of 69 genes with a previously described association with IA.

4. Prediction analysis of microarrays (PAM)

With this analysis, which uses the nearest shrunken centroid method,[59] we aimed to find the smallest set of genes in our data that can accurately classify samples as cases or controls. We computed a nearest shrunken centroid classifier for the SVA-normalized discovery dataset, and determined the amount of shrinkage by cross-validation. Next, we estimated FDRs for this classifier in the replication set. The optimal classifier gene set was determined by calculating sample misclassification errors for gene sets at different shrinkage thresholds.

Results

A total number of 210 samples (103 cases and 107 controls) and 34 135 probes were included after quality control. Table 1 shows the baseline characteristics of the study population.

1. Differential expression

We observed no probes with a significant ($p_{FDR} < 0.05$) case-control difference in expression in the discovery set. Consequently, no probes were selected for further analysis in the replication set.

For IA genes from the literature, we listed the p-values of differential expression in the discovery set (S1 Table). In total, 87 probes that mapped to these 69 IA genes were tested in our study. No probes were differentially expressed between patients and controls with $p_{Bonf} < 0.05$.

2. Co-differential co-expression

We identified nine different modules of co-expressed genes in the discovery set of cases and controls combined. In none of these modules, the average gene significance level of differential expression of the genes was significantly higher than expected by chance, with p-values ranging from 0.23 to 0.48.

3. Differential co-expression and biological relevance of gene modules

After creating a dendrogram of DC genes in the discovery set, we observed six different modules of DC genes (S2 Fig.). After 10 000 permutations, three modules remained significant (Table 2): the blue (p_{Bonf} = 0.024), turquoise (p_{Bonf} = 0.045) and yellow module (p_{Bonf} = 0.047).

The correlation heatmaps of differentially co-expressed gene modules showed that the pattern of case-control differences in gene correlation per module, as observed in the discovery set, did not look similar in the replication set (S3 Fig.). We formally tested the preservation of the three significant modules in the replication set, and observed a significant preservation of only the yellow module (Z-summary = 11.3; Table 2). This module consisted of 1818 probes. Pathway analysis in DAVID revealed that only Gene Ontology pathways involved in processes in the vacuole and lysosome were significantly enriched in this module ($p_{Bonf} = 5 \times 10^{-3}$). We did not find any publications on a role of such processes in IA.

We identified a total number of 129 genes from the yellow module as hub genes. No overlap was found between the 69 IA genes and the list of hub genes. The most important hub gene of the yellow module was *CLCN6* (chloride channel, voltage-sensitive 6) at locus 1p36.22. This gene encodes a member of the voltage-dependent chloride channel protein family. An association with diastolic blood pressure has been described for a polymorphism in the promoter region of this gene.[60] For the remaining genes in the 'top 10' of hub genes, no functions or disease associations with a known relation to IA were found. The ten most important hub genes are described in Table 3.

4. Prediction analysis of microarrays (PAM)

The gene sets created with PAM in the discovery set all had a low predictive ability for classifying samples in the replication set as cases and controls, with misclassification rates of 40% or higher. The smallest gene set with a misclassification rate of 40% consisted of 2388 genes. This gene set did not cluster cases and controls in separate groups in the replication set (Fig. 1), with a sensitivity for detecting SAH cases of 66%, and a specificity of 64%.

Discussion

We found no significantly differentially expressed genes in peripheral blood between patients with a history of aSAH and controls. We identified only

one group of genes with a significant and preservated case-control difference in co-expression, which can indicate the presence of coding changes in one or more genes present in that group. However, it remains unclear whether these genes are biologically relevant. First, we could not ascribe biological meaning either to this group of genes, or to the most connected genes within this network, the so-called 'hub' genes, according to pathway analysis. Specifically, the gene groups were not involved in pathophysiological processes known to be involved in IA formation and rupture, like vessel wall degeneration and inflammation.[3, 4] Second, these genes were not present in a list of 69 genes with a previously described association to IA based on GWAS[10, 11, 13] and gene expression studies.[22] In addition, we did not find a set of differentially expressed genes with the ability to predict disease status.

Previous studies on gene expression in IA mainly investigated gene expression in IA tissue, and some genes were differentially expressed between IA and control tissue in multiple studies: *BCL2, COL1A2, COL3A1, COL5A2, CXCL12, TIMP4,* and *TNC.*[22] In our study, these genes were not differentially expressed in blood (S1 Table) and did also not play an important role in DC. It should be noted that many of the previous IA tissue gene expression studies had limitations, in particular most were small. Moreover, their study designs were heterogeneous. These factors could limit the number of genes that were differentially expressed in multiple studies. One relatively large gene expression study in tissue of IA at the middle cerebral artery discovered upregulation of several biological processes in ruptured compared to unruptured IA: response to turbulent blood flow, chemotaxis, leukocyte migration, oxidative stress, vascular remodeling, and extracellular matrix degradation.[61] These processes were also not found in our pathway analysis. The most likely explanation for the discrepancies between the IA tissue studies and our study findings is that the published genes could be expressed specifically in IA tissue and not in blood. Additionally, some of these genes[61] could be specifically involved in IA rupture and not in IA development. Our motivation to study gene expression in blood is the higher clinical relevance compared to IA tissue studies: because blood is accessible and easy to obtain, gene expression differences present in blood could potentially lead to development of a biomarker.

Recently, two studies on gene expression in blood of aSAH patients have been published. One study included aSAH patients in the acute phase and found expression differences of genes related to T lymphocytes, monocytes and neutrophils.[62] This study is not comparable to our study, in which we

explicitly wanted to avoid including patients in the acute phase, since these acute consequences probably do not tell anything about the pathogenesis of development and rupture of IA. In a second study, in which gene expression in blood was compared between 30 patients with ruptured and unruptured IA and 15 controls, expression differences were found for genes coding for matrix metalloproteinases, extracellular matrix and cytoskeleton proteins and of genes related to apoptosis.[63] However, no correction for multiple testing and no replication experiments were performed. The timing of blood sampling in relation to aSAH (in case of ruptured IA) and to treatment of the IA is not described, thus we cannot tell whether the differences found are secondary to rupture or treatment, or are related to IA development.

Gene expression profiling has already been proven to characterize disease status in several vascular diseases, including transient ischemic attacks (TIAs)[24] and ischemic stroke.[25] Gene expression studies in blood have also been performed for abdominal aortic aneurysms (AAA)[64] and thoracic aortic aneurysms (TAA).[65] In contrast to the present study, these studies did reveal gene expression differences. In AAA, some disease-specific gene expression differences were found both in blood and in tissue, while in TAA it was even possible to predict disease state with gene expression profiles. However, a direct comparison with our study is not straightforward, because it is not clearly stated for these studies in which stage of the disease the blood samples were taken (before or after aneurysm rupture, and before or after treatment). Secondary effects of aneurysm rupture and operative treatment on gene expression are therefore not excluded.

A strength of this study is the fact that we obtained blood samples several years after aSAH. In this way, we minimized secondary effects of IA rupture on gene expression. To our knowledge, the sample size of this study is the largest compared to previous gene expression studies in blood of aneurysms and of other cardiovascular diseases.[66] Furthermore, to minimise confounding we matched cases and controls for factors including age and lifestyle by including mainly spouses of aSAH patients as controls, and blood samples were obtained at the same time and under the same circumstances for all participants. We also adjusted for known risk factors of IA and aSAH. In this way, we minimized influence of factors not related to the disease, with a known influence on gene expression.[48]

Our retrospective study design, using previous aSAH patients, could lead to limitations. First, disease-specific gene expression differences can theoretically be undetectable after IA treatment. Ideally, we would also have

studied patients with unruptured IA who are followed-up for growth of their IA, using growth as a surrogate risk factor of rupture,[67] and compare gene expression between IA that grow to those IA that remain stable over time. However, such a study is difficult to perform given the low availability of such patients. Also, the usage of previous aSAH patients to study IA pathogenesis can be justified, as IA appears to be a continuous disease process in the intracranial vessel wall: previous aSAH patients are still at risk of developing new aneurysms, and of growth of already present unruptured IA.[45]
Second, the study design precludes patients who deceased after aSAH. This selection bias could theoretically have influenced our results, but it is unlikely that disease-specific gene expression differences are only detectable in patients with a poor outcome.

In conclusion, this study revealed no structural gene expression differences in blood of previous aSAH patients compared to controls. Also, gene expression profiles in blood of previous aSAH patients, as detected by expression arrays, do not seem promising for development of clinically useful biomarkers. Future studies could aim at detecting changes in gene expression profiles in patients with yet unruptured IA that grow over time, or at the identification of other potential disease markers, including microRNAs or using proteomics.

Figures

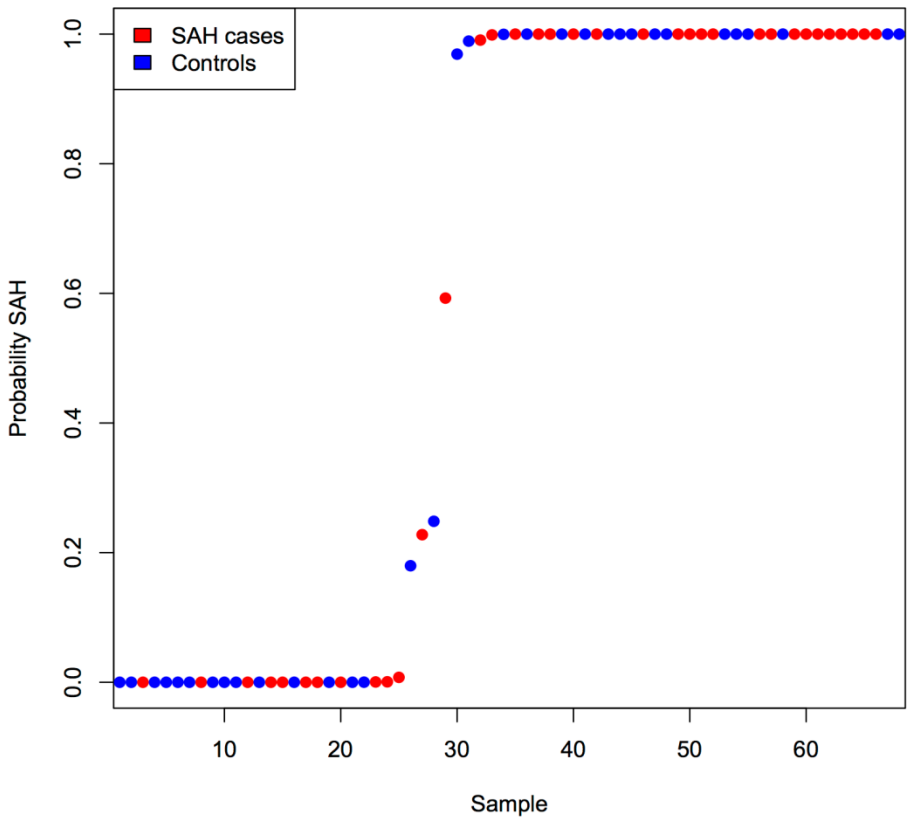


Figure 1. Predicted SAH probability in subjects from replication set, using 2388 probes selected with prediction analysis of microarrays.

This figure shows the probability of being a SAH case for each subject in the replication set, based on prediction analysis of microarrays (PAM). We used PAM to define a group of probes in the discovery set with the highest predictive value to identify cases and controls in the replication set. As a result, a group of 2388 probes was selected, with a relatively high misclassification rate of 40%. The figure shows that this group of probes does not divide cases and controls in two separate groups.

Tables

Table 1. Baseline characteristics of the study population.

Characteristics	Cases	Controls
Total number	103	107
Women *N* (%)	86 (83)	43 (40)
FIA *N* (%)	4 (4)	0 (0)
History of smoking *N* (%)	85 (83)	83 (78)
Hypertension *N* (%)	63 (61)	30 (28)
Mean time from SAH to study (range)	8 (2-23)	NA
Cases with additional aneurysms *N* (%)	14 (14)	NA

FIA indicates familial intracranial aneurysm, N: number, SAH: subarachnoid hemorrhage, NA: not applicable.

Table 2. Modules of differentially co-expressed genes: permutation and preservation results.

Module	Number of genes	Permutation index	Permutation p_{Bonf}	Preservation Zsummary
Blue	2007	40	0.024	6.06
Black	2837	553	0.33	NA
Brown	1822	103	0.062	NA
Green	1598	329	0.20	NA
Turquoise	3106	75	0.045	-0.74
Yellow	1818	81	0.047	11.32

p_{Bonf} indicates Bonferroni-corrected p-value, NA: not applicable (module not tested for preservation). Modules were randomly color-labeled.

Table 3. Top 10 hub genes in yellow module.

Rank	Gene	Locus	Product / function	Disease associations
1	*CLCN6*	1p36	Voltage-gated chloride channel proteins	Blood pressure
2	*SAG*	2q37	Cellular responses in retina and pineal gland	Oguchi disease
3	*AOC2*	17q21	Copper binding	Ocular diseases
4	*LILRA6*	19q13	Leukocyte immunoglobulin-like receptor	-
5	*PCDHA5*	5q31	Neural cell adhesion proteins	-
6	*LYPD6*	2q23	Disulfide-bonding proteins	-
7	*LHX6*	9q33	Development of neural and lymphoid cells	Lung cancer, cervical cancer
8	*LOC390531*	15q11	Non-functional	-
9	*CISH*	3p21	Regulation of cytokine-signaling	Infectious diseases
10	*HNMT*	2q22	Metabolization of histamine	Parkinson's disease

Supplemental data

Supplemental Tables and Figures are available with the online version of this paper.

Chapter 4

Exome-wide association analysis in intracranial aneurysms

FNG van ‘t Hof, J van Setten, M Bots, I Vaartjes, GJE Rinkel, PIW de Bakker, YM Ruigrok.

Manuscript in preparation.

Abstract

Background
Genetic variation contributes to the risk of intracranial aneurysms (IA), but the largest part of the heritability is still unexplained. We hypothesize that low-frequency genetic variants (with frequencies <5%) influence IA risk. Therefore, we performed an exome-wide association analysis using the exome chip in IA patients and controls.

Methods
1056 IA patients and 2097 population-based controls from the Netherlands were genotyped using the HumanExome BeadChip. After quality control (QC) of samples and single nucleotide variants (SNVs), we conducted a single variant analysis using the Fisher exact test, and three gene-based analyses (T1 (which includes SNVs with minor allele frequencies < 0.01), variable threshold (VT), and the sequence kernel association test (SKAT)) to test whether multiple variants within the same gene are associated with IA when tested on aggregate.

Results
After QC, 995 IA patients and 2080 controls remained for further analysis. The single variant analysis comprising 46,534 SNVs did not identify significant loci at the genome-wide level. The gene-based tests showed associations around the level of nominal significance for *FBLN2*, both when including SNVs with minor allele counts (MAC) >5 per cohort (strongest association for T1 with $p=7x10^{-7}$) and after adding variants with MAC≥1 (strongest association for SKAT with $p=8x10^{-6}$).

Conclusion
We did not identify significant associations between low-frequency genetic variants and IA risk. However, gene-based tests indicated that variants in *FBLN2*, which plays a role in blood vessel formation, are associated with IA. Replication in an independent cohort should clarify the robustness of this association.

Introduction

Intracranial aneurysms (IA) have a strong familial component.[68] Genome-wide association studies (GWAS) detected several risk loci for IA: common genetic variants (defined as a minor allele frequency (MAF) > 5%) near the genes *STARD13-KL*, *RBBP8*, *SOX17, CDKN2BAS, CNNM2* and *EDNRA* were associated in European and Japanese cohorts.[10, 11, 13] In addition, common variants at a locus encompassing *HDAC9* were associated in European IA populations,[14] and low-frequency variants (MAF < 5%) near *FSTL1* and *EPM2A* were associated with IA in a Dutch and Finnish cohort.[15] However, all these variants explain only a small part of the estimated heritability of IA,[13, 15] suggesting that additional genetic variation can contribute to IA.

Whole-genome and whole-exome sequencing studies have showed that low-frequency genetic variants play a role in complex diseases.[69] As these studies are expensive to perform on large scale, exome chips have been designed as an affordable alternative to detect low-frequency genetic variants in coding regions of the genome.[70, 71] Using this method, low-frequency variants have recently been detected for several common diseases and traits, like insulin processing and secretion,[72] macular degeneration,[73] cholesterol levels, and myocardial infarction.[74]

The overall aim of this study was to test whether low-frequency variants in the exome are involved in susceptibility to IA disease, by performing an exome-wide association analysis using the exome chip in IA patients and controls. In addition, we looked up the association of low-frequency variants in established IA loci,[10, 11, 13] to search for additional variants that could be responsible for the GWAS associations of common variants.[75] Finally, we applied gene-based association analyses to test whether multiple variants within the same gene may be collectively associated with IA.

Methods

Study population

Dutch IA patients (n=1056) were admitted to the University Medical Center Utrecht, the Netherlands between 1997 and 2011. The population consisted mainly of patients with ruptured IA, but also with unruptured IA. Ruptured IA cases were defined by symptoms suggestive of SAH combined with subarachnoid blood on a computed tomography (CT) scan and a proven IA at angiography (conventional angiogram, CT- or magnetic resonance (MR)-angiogram). Unruptured IA cases were identified by CT or

MR angiography or conventional angiography in the absence of clinical or radiological signs of SAH. Patients with fusiform IA, possible traumatic SAH, and polycystic kidney disease were excluded.

As controls, we retrieved a subgroup of 2097 subjects from the Utrecht Health Project (UHP). The UHP is an ongoing dynamic population-based cohort that includes residents of the large residential area Leidsche Rijn, part of the city Utrecht, the Netherlands. An extensive overview of the study has been published previously.[76] All controls were aged >18, were from European descent, and were not genetically related to other people in the cohort.

The study has been approved by the Medical Ethics Committee of the University Medical Centre Utrecht. All participants gave written informed consent.

Genotyping

DNA was prepared from peripheral blood, which was drawn during inclusion of study subjects. Cases and controls were genotyped together, as part of the Netherlands ExomeChip Project (www.bbmri.nl), on the Illumina HumanExome Beadchip v1.1. In total, 242,901 SNVs were genotyped. Genotypes were subsequently called using the GenomeStudio software from Illumina. However, this software is mostly used for calling genotypes of common SNVs, and is not designed to call low-frequency variants. We therefore used zCall[77] to call genotypes of variants that could not be called by GenomeStudio.

Quality control

We performed quality control (QC) in the case and control cohorts separately, and after merging cases and controls. Plink version 1.07[78] was used for sample and SNV QC.

Samples with a call rate <95% were removed. Using a subset of common, independent, high-quality SNVs (as defined by SNVs without deviation from Hardy-Weinberg equilibrium (HWE) ($p>0.001$), with MAF >5%, rate of missing genotypes <1%, and linkage disequilibrium (LD) $r^2<0.05$), subjects were removed based on the following four criteria: discordant sex, heterozygosity (subjects were excluded if the inbreeding coefficient deviated more than 4 standard deviations from the mean) and cryptic relatedness (by calculating identity-by-descent (IBD) for each pair of individuals). In each pair with an IBD proportion of >20%, a subject was excluded, if it exhibited distant relatedness with more than one individual.

For case-control pairs, we removed the control subject. In the case-case or control-control pairs, the subject with the lowest call rate was excluded.

Using the set of common, independent, high-quality SNVs, we performed principal components (PC) analysis using EIGENSTRAT[79] on the remaining study subjects and HapMap-CEU subjects. We created PC plots with the first four PCs, using R version 2.11.[46] Based on visual inspection of these plots, we excluded subjects that appeared to be outliers with respect to the CEU or the study population. A PC plot of the first two PCs after outlier removal is shown in Supplementary Figure 1. PCs were re-calculated for only the study subjects, after outlier removal, and tested for association with case-control status (logistic regression) to be used in the single variant analysis (see below).

After sample QC, we excluded SNVs that were monomorphic, SNVs with >5% missing genotypes, missing genotype rate higher than MAF, HWE deviation ($p<0.001$), or a differential degree of missing genotypes between cases and controls ($p<1x10^{-5}$; chi-squared test) were also removed.

Single variant association analysis

For this analysis, we selected all SNVs with a minor allele count (MAC) < 5 in either of the two cohorts. Association testing was carried out in PLINK, using Fisher's exact test. We calculated a genomic inflation factor (λ_{GC}) for the single variant analysis.[80] After Bonferroni correction for the number of SNVs tested, SNV associations with $p<1.07x10^{-6}$ (0.05 / 46,534) were considered as statistically significant. For significantly associated SNVs, we visually inspected the cluster plots. If the cluster assignment of GenomeStudio did not match the visual cluster assignment, the SNV was removed from further analyses. For remaining significant SNVs, we repeated the association analysis using logistic regression, with the first four PCs and sex as covariates, to check whether the associations were determined by these factors.

We compared the allele frequencies of the significantly associated variants between our study samples and the European population reported by the Exome Aggregation Consortium (ExAC) database.[81]

Gene-based association analysis

We tested the collective effect of multiple low-frequency variants within the same gene on IA risk, using SCORE-Seq.[82] All SNVs that passed QC and with MAF < 0.05 were included in this analysis. In a second analysis, we also included SNVs with lower frequencies (MAC ≥1 in both cohorts). In

contrast to single SNV association tests, gene-based tests are expected to have more statistical power to detect genes including these SNVs with lower frequencies.

For both analyses, with MAC thresholds of 5 and 1 respectively, we applied three different gene-based tests. The first test is T1 gene-based association method,[83] which includes SNVs with MAF < 0.01. Second, we applied the variable threshold (VT) test,[84] which includes SNVs at different MAF thresholds below 0.05. The third test is the sequence kernel association test (SKAT),[85] a weighted sum of individual score statistics. This test includes all SNVs but weighs them based on their MAF; SNVs with lower MAF obtain a greater weight.

In the first analysis including SNVs with MAC>5, genes with $p<7x10^{-6}$ (after Bonferroni-correction for the number of tested genes: 0.05 / 7110) were defined as statistically significant. In the second analysis with MAC ≥1, the significance level was set at $p<3x10^{-6}$, after adjustment for 14,648 tested genes.

Look-up of known IA loci

We selected SNVs and genes in 11 known IA risk loci (Supplementary Table 3),[10, 11, 13-15] and looked up their association results in our single variant and gene-based analyses. The selection of IA risk loci was based on genome-wide significance ($p<5x10^{-8}$) in previous GWAS in European populations. Genotyped SNVs within 500 kilobases distance from the selected index SNV were included. We were particularly interested in low-frequent SNVs (MAF<0.05) associated with IA with $p<8x10^{-4}$ (after Bonferroni-correction for the number of low-frequent SNVs: 0.05 / 64), in order to detect low-frequency causal variants responsible for the associations of these risk loci.

Power calculation

Given our sample size of 995 cases and 2080 controls (after QC), we have >80% power to detect genetic variants with MAF=1% with an effect size of 22.2.[86]

Results

Study population and QC

After QC, there were 995 IA patients, 2080 controls and 46,534 SNVs available for further analysis. The baseline characteristics of the patients and controls are shown in Table 1. The number of samples and SNVs

removed at each QC step are shown in Supplementary Table 1. PC plots after removal of outliers showed well-matched cases and controls (Supplementary Figure 1). None of the first ten PCs was significantly associated with case-control status (p≥0.09 in logistic regression).

We performed visual inspection of all SNVs that showed statistically significant associations with IA in the single variant association analysis. In total, 20 SNVs were associated with p-values ranging from $6.5x10^{-7}$ to $2.8x10^{-56}$ (Supplementary Table 2). However we observed that the genotypes of all 20 SNVs were incorrectly clustered and therefore assigned to the wrong genotype.[71] Supplementary Figure 2 shows an example of one of these cluster plots. After manual adjustment of the clusters, the associations of any of the 20 SNVs were no longer statistically significant. Consequently, all these associations were interpreted as false positive, and the SNVs were excluded from further analyses.

Single variant association analysis

In total, 46,514 SNVs were tested for association with IA. We did not observe inflation of the test statistic distribution to the expected median (λ_{GC} = 0.92; Supplementary Figure 3), and we did not observe any statistically significant associations (other than the false positive associations that were removed during QC). Figure 1 shows the results of the association analysis.

Gene-based association analysis

We found one statistically significant association of gene *FBLN2* on chromosome 3, with p-values of $7x10^{-7}$ (T1 test), $1x10^{-6}$ (VT test), and $6x10^{-6}$ (SKAT). In the first analysis including SNVs with MAC>5, only two SNVs in this gene were analyzed in the gene-based tests: exm292082 and exm292166. In a second analysis, which also included SNVs with MAC between 1 and 5, associations for *FBLN2* were slightly weaker, with p-values of $8x10^{-6}$ (SKAT), 0.10 (T1 test), and 0.26 (VT test), based on a total number of 13 SNVs in this region. After visual inspection of the cluster plots of all 13 SNVs, we did not detect incorrect cluster assignment by GenomeStudio (an example is shown in Supplementary Figure 4). No other statistically significant associated genes were found in this second analysis, after removal of SNVs with incorrect clustering.

In the single variant analysis, three of 13 *FBLN2* SNVs showed a suggestive level of association (p<0.05) with IA: exm292082 / rs111389908 ($p=2.0x10^{-3}$; OR=2.6 for A allele), exm292166 / rs201160150 ($p=6.3x10^{-4}$; OR=3.4 for T allele), and exm292103 / rs113265853 ($p=1.8x10^{-3}$; OR=4.2

for allele C). According to the ExAC database (http://exac.broadinstitute.org),[87] exm292082 is a missense variant with MAF=0.011 in the European population. The variants exm292166 (MAF=0.004) and exm292103 (MAF=0.005) are both located at the splicing region. Further details about the associations of all 13 SNVs in *FBLN2* are shown in Table 2.

Look-up of known IA loci

The association results of SNVs and genes in 11 established IA risk loci are shown in Supplementary Table 3. We observed suggestive associations of the following previously reported common risk SNVs: rs1333040 at 9p21 ($p=9.3x10^{-4}$), rs10958409 at 8q11 ($p=1.7x10^{-4}$), and exm2269203 (alias rs10931779; proxy of risk SNV rs919433 with $r^2=0.92$) at 2q33 ($p=6.9x10^{-5}$). However, we observed no significant associations of low-frequent SNVs in these loci ($p \geq 0.02$).

Discussion

We found no statistically significant associations of individual low-frequency genetic variants with IA, but a gene-based test indicated that variants in *FBLN2* were collectively associated with IA in the Dutch population. We did not find associations of low-frequency variants in established IA loci[10, 11, 13] that could be responsible for the GWAS associations of common variants.

These study results indicate a possible role of *FBLN2* in IA. The *FBLN2* (fibulin 2) gene is an interesting candidate gene for IA, because it encodes an extracellular matrix protein that belongs to the fibulin family.[88] It plays a role in the formation of the elastic lamina of blood vessels.[89] Mutations in this gene have not yet been linked to aneurysm formation, according to small studies in abdominal aortic aneurysms[90] and thoracic aortic aneurysms.[91] The gene-based tests in our study were statistically significant when including only 2 SNVs with MAC >5, and the association of one test (SKAT) was borderline significant after adding 11 SNVs with MAC between 1 and 5. Three SNVs in these analyses were suggestively associated with IA on an individual level. These SNVs were all missense variants or located at the splicing region, which indicates that these variants could alter protein function and thus be functionally relevant.

The absence of significant associations in the single variant analysis can well be due to lack of statistical power to detect SNVs with lower effect sizes, because of the limited sample size of the available cohorts with exome chip data of IA patients used in this study. The lack of findings with

this sample size is in line with other genome-wide studies of low-frequency variants in common diseases. A recent study in intracerebral hemorrhage, with a smaller sample size (n=1553), did also not reveal new associations.92 In contrast, sample sizes were much larger (n>8000) in studies on insulin secretion,72 myocardial infarction and cholesterol levels,74 and blood pressure,93 for which significant associations of single low-frequency variants were found.

Our gene-based tests did reveal a suggestive association. Similarly, another study on low-frequency variants in meningitis, with a comparable sample size to our study, also found an association in the gene-based analysis and not in the single-variant analysis.[94] As gene-based analyses test the effect of different SNVs in one gene collectively, they can provide a more powerful way to analyze the involvement of low-frequency variants, which is particularly useful for studies with relatively small datasets, like our study.

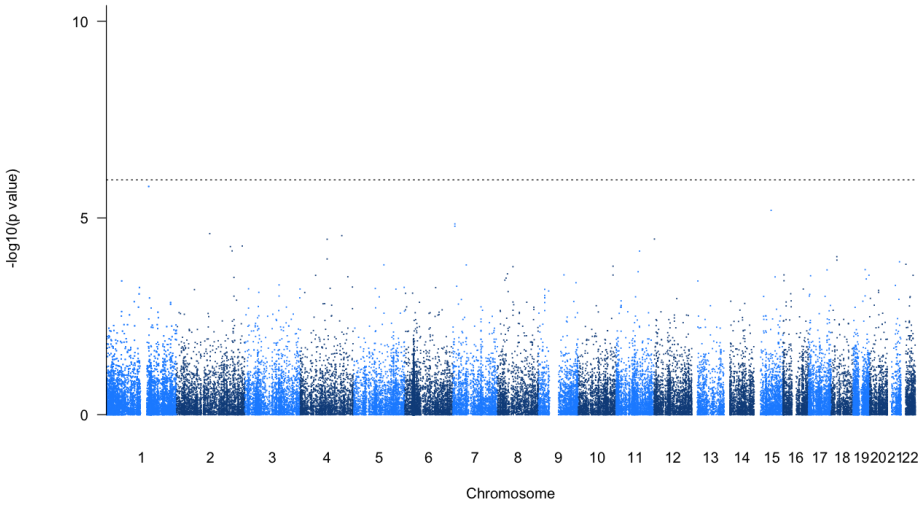
The exome chip has the advantage that low-frequency variants in coding regions of the genome can be tested on a large scale. However, variants in noncoding regions remain undetected, although these can also play a role in complex traits, like has recently been shown for bone density.[95] Also, the exome chip only contains predetermined variants, so novel variants remain undetected. In contrast, sequencing studies do not have this limitation, but these are more expensive to perform on large scale.

Another possible drawback of this exome chip study is the genotype calling of low-frequency variants, which can be inefficient because the minor allele clusters are very small.[71] Although we used appropriate methods for calling of these variants, we still observed some mistakes in cluster assignment. We were able to distinguish true from false positive associations by visual inspection of the cluster plots of the associated SNVs, but theoretically, true associations may have remained undetected due to clustering errors.

In conclusion, this study indicates a possible role of the *FBLN2* gene in IA pathology, based on the collective association of coding low-frequency variants in this gene. Genotyping of variants in the *FBLN2* gene in an independent cohort would be needed to confirm these findings. In addition, future studies focusing on discovery of low-frequency coding variants involved in IA should either contain much larger sample sizes of IA patients, or focus on sequencing, for example in specific subgroups, like familial IA or population isolates.

Figures

Figure 1. Manhattan plot of the single variant association analysis.
This Manhattan plot shows the p-values for association with IA for all 46,514 SNVs in the single variant analysis (after exclusion of false positive associations). P-values on the y-axis are presented on an inverse log scale. The x-axis represents the genomic position of each SNV. The dashed line represents $p=1.07x10^{-6}$, the cut-off value for genome-wide association. No significantly associated SNVs were observed.



Tables

Table 1. Baseline characteristics of the study population.

Number of patients	995
Mean age in years (range)	55 (15-89)
Women (%)	69
Number of controls	2080
Mean age in years (range)	40 (18-91)
Women (%)	53

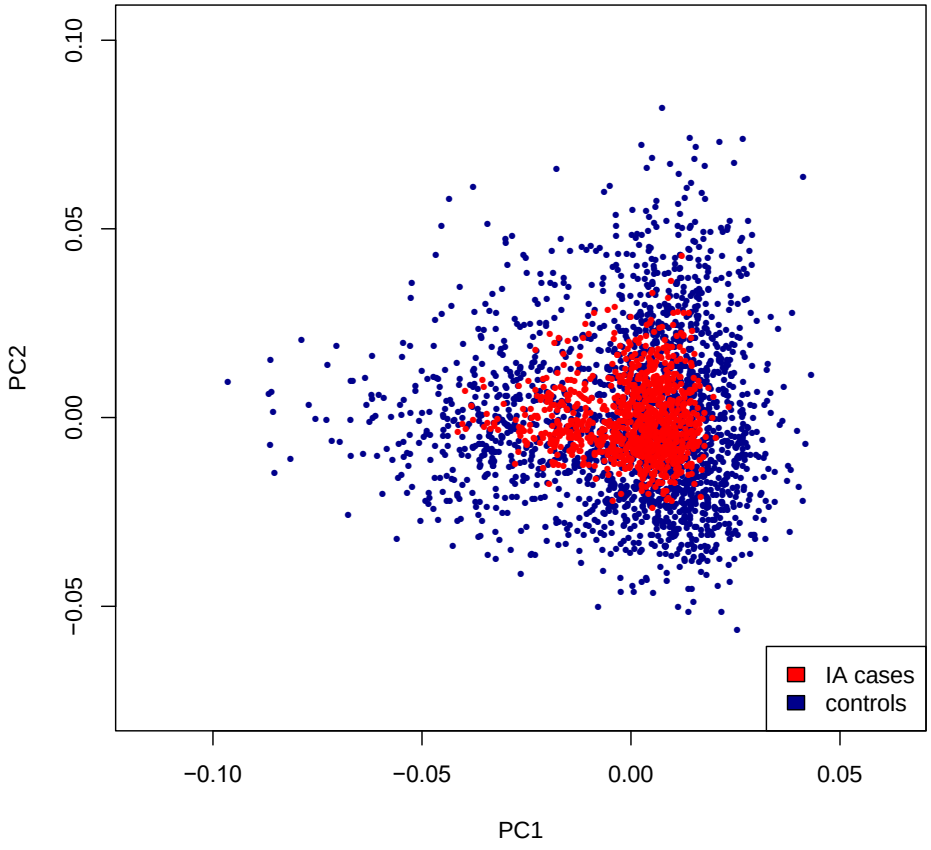
Table 2. Association results of 13 SNVs in the FBLN2 gene on chromosome 3, in the study cohort of 995 cases and 2080 controls

SNV	BP	A1/A2	MAF cases	MAF controls	MAF ExAC*	p-value	OR
exm292050	13612224	T/G	0.00050	0.00072	0.0016	1	0.70
exm292057	13612319	T/C	0.0010	0.0012	0.0012	1	0.84
exm292070	13612459	C/G	0.0020	0.0019	0.0027	1	1.05
exm292082	13612786	A/G	0.0126	0.0048	0.0110	$2.0x10^{-3}$	2.63
exm292091	13612921	A/G	0.0010	0.0012	0.0013	1	0.84
exm292097	13613015	C/A	0.0025	0.0026	0.0081	1	0.95
exm292103	13613081	T/C	0.0020	0.0085	0.0045	$1.8x10^{-3}$	0.24
exm292123	13655548	G/A	0.0010	0.00048	$7.5x10^{-5}$	0.60	2.09
exm292152	13661319	A/G	0.00050	0.00024	0.00047	0.54	2.09
exm292158	13663365	T/C	0.00050	0.00072	0.0011	1	0.70
exm292166	13669333	T/C	0.0106	0.0031	0.0036	$6.3x10^{-4}$	3.40
exm292194	13670791	T/C	0.0010	0.0005	0.0043	0.60	2.09
exm292217	13677946	T/C	0.0005	0.0005	0.0022	1	1.05

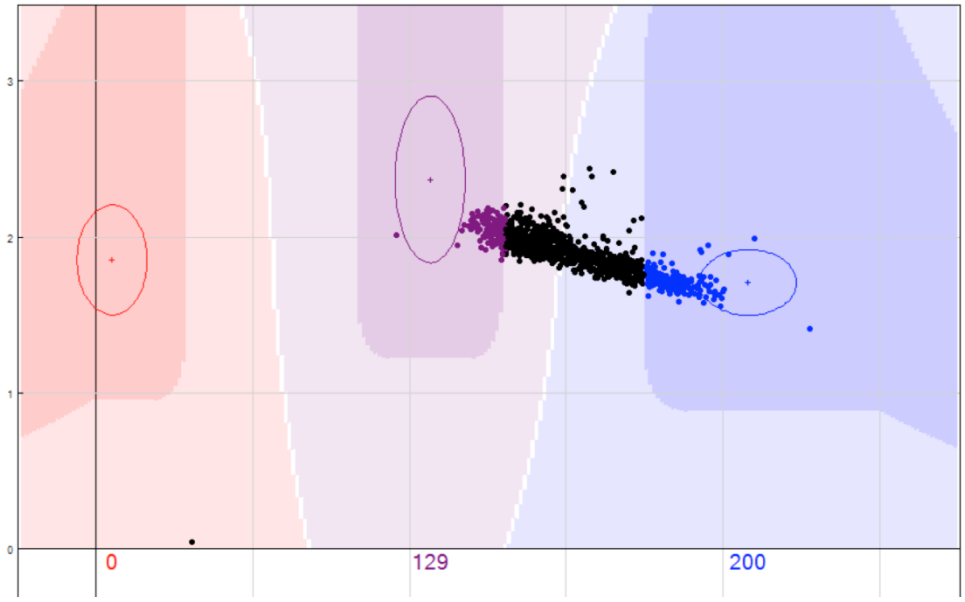
SNV indicates single nucleotide polymorphism; BP, basepair position; A1, allele 1 (minor allele); A2, allele 2 (other allele); MAF, minor allele frequency; OR, odds ratio.
* MAF in the European population according to the ExAC (Exome Aggregation Consortium) database

Supplementary Figures

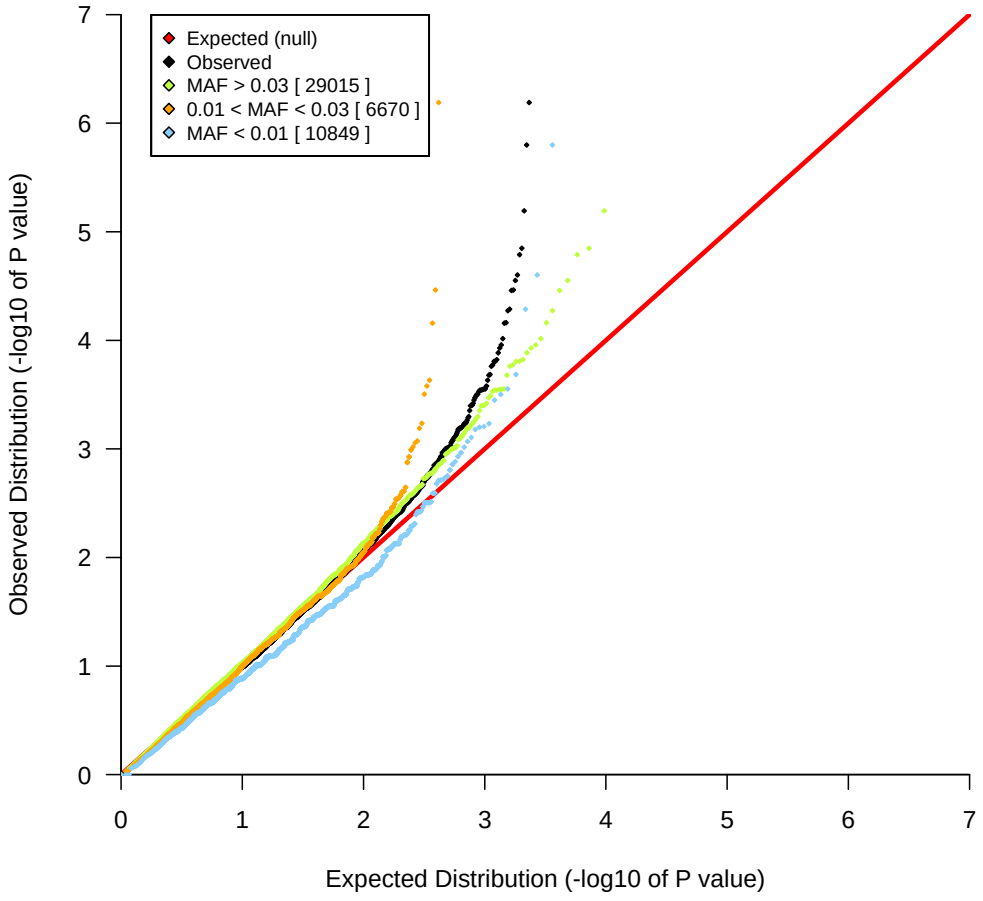
Supplementary Figure 1. Principal component analysis (PCA) plots of IA cases and controls.
This figure shows PC2 values plotted against PC1 values for each IA case and control, after removal of outliers from quality control.



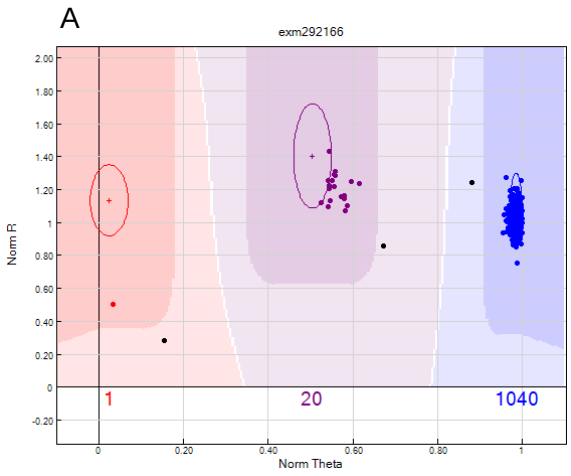
Supplementary Figure 2. Cluster plot of exm806588 in IA cases.
This figure shows the cluster plot of exm806588 (rs200909708), one of the 20 significantly associated SNV in the single variant analysis, in IA cases. The clusters were defined after genotype calling in GenomeStudio. Each dot represents one case. Blue dots are cases assigned to the cluster of major allele homozygotes (GG, n=200), purple dots are assigned to the heterozygote cluster (AG, n=129), and black dots are unassigned cases. No minor allele homozygotes (AA) are detected (red cluster, n=0). The final association analysis was done after clustering with zCall, which assigns every subject to a cluster. This resulted in 832 major allele homozygotes, 163 heterozygotes and 0 minor allele homozygotes, and a minor allele frequency (MAF) of 0.08. However, the cluster plot below shows one large cluster, probably existing of mainly major allele homozygotes. The true MAF would therefore be much lower, as would be expected based on the low MAF of 0.00022 in European samples, according to the ExAC database (http://exac.broadinstitute.org).[87] The same overestimation of heterozygotes was observed after inspection of the cluster plot in controls. Consequently, this SNV, as well as the other 19 significant SNVs, were removed from the analysis due to incorrect clustering.

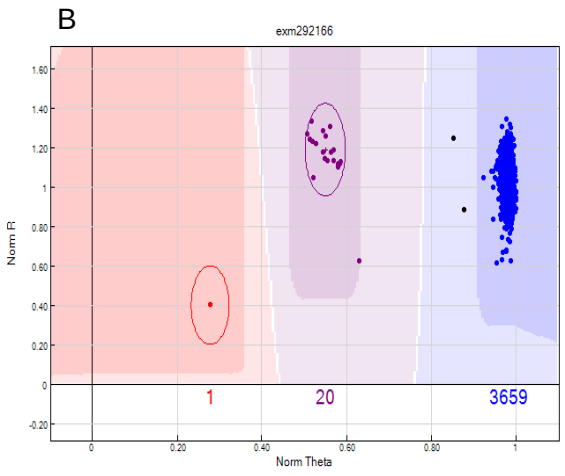


Supplementary Figure 3. Quantile-quantile (QQ) plot of the single variant association analysis.
These QQ plots show the observed distribution of p-values of all SNPs (black dots) plotted against the expected distribution of p-values (red line) on a negative log10 scale. The genomic inflation factor is 0.92, indicating no inflation of the observed compared to the expected test statistics distribution. A division of observed p-values according to different minor allele frequencies (MAF), represented by green, orange and blue dots, shows no substantial differences in p-value distribution between rare and common SNPs.



Supplementary Figure 4. Cluster plot of a SNP located in *FBLN2*, for IA cases (Figure A) and controls (Figure B).
This figure shows the cluster plot of exm292166 (rs201160150), located in the gene *FBLN2*, which was associated with IA in the gene-based test. In contrast to Supplementary Figure 2, these plots do not show an overestimation of heterozygotes, but clearly separated clusters of homozygote or heterozygote samples.





Supplementary Tables

Supplementary Table 1. Removal of samples and SNPs during quality control (QC).

Sample QC	*Cases (n)*	*Controls (n)*
Before QC	*1056*	*2097*
Missing genotypes >5%	7	NA
PCA	31	NA
Heterozygosity (IBD >4 SD)	3	3
Cryptic relatedness	12	14
Sex discordance	11	NA
After QC	*995*	*2080*
SNP QC	*SNPs (n)*	
Before QC	*242901*	
Monomorphic SNPs	156198	
MAC<5	56342	
HWE deviation (p<0.001)	709	
Missing genotypes >5%	13	
Differential missingness ($p<1x10^{-5}$)	318	
Missingness > MAF	3340	
After QC	*46534*	

SNP indicates single nucleotide polymorphism; n, number; PCA, principal component analysis; IBD, identity by descent; SD, standard deviation; MAC, minor allele count; HWE, Hardy-Weinberg equilibrium; MAF, minor allele frequency; NA, not applicable.

Supplementary Table 2. Association results of 20 significantly associated SNPs in the single variant analysis.

CHR	SNP	BP*	A1	A2	F A1 cases	F A1 controls	P	OR
19	exm1467381	40540697	A	G	0.0794	0.005048	2.8e-56	17.0
16	exm1239717	50659417	G	A	0.0603	0.001443	7.3e-52	44.4
10	exm806588	5011017	A	G	0.08191	0.007452	1.4e-51	11.9
19	exm1505393	55106741	T	C	0.07789	0.00625	2.5e-51	13.4
4	exm414908	99982306	T	C	0.07638	0.00625	5.3e-50	13.2
17	exm1346357	62854928	A	G	0.05276	0.001683	1.4e-43	33.1
12	exm995213	42512844	C	A	0.04673	0.003365	4.1e-32	14.5
11	exm896606	21594816	A	G	0.05126	0.005288	7.0e-31	10.2
10	exm826163	60148436	G	C	0.03819	0.003365	1.0e-24	11.8
1	exm146854	211956700	C	T	0.03317	0.003606	6.5e-20	9.5
16	exm1217630	11272474	T	C	0.04322	0.008654	5.4e-18	5.2
19	exm1445457	18572428	A	G	0.0196	0.001442	5.2e-14	13.8
7	exm660582	137776608	C	T	0.05126	0.01659	1.2e-13	3.2
6	exm593151	161143565	T	C	0.02111	0.002163	2.1e-13	9.9
9	exm775023	117093130	A	G	0.04824	0.01683	9.6e-12	3.0
12	exm1054158	133306290	C	G	0.05126	0.01923	2.9e-11	2.8
1	exm117861	161298206	T	C	0.05528	0.02282	2.2e-10	2.5
12	exm1022749	75902185	A	G	0.03317	0.01082	3.8e-09	3.1
6	exm580663	136589448	A	C	0.04623	0.01903	5.9e-09	2.5
19	exm1436192	15053181	G	A	0.02462	0.008173	6.5e-07	3.1

SNP indicates single nucleotide polymorphism; BP, basepair position; A1, allele 1; A2, allele 2; F, frequency; P, p-value; OR, odds ratio; NA, not available.
** Genomic coordinates are based on the reference assembly GRCh37 from the Genomic Reference Consortium.*

Supplementary Table 3. Association results of SNPs in established IA risk loci. [10, 11, 13]

Locus	SNP	BP**	A1	A2	F A1 cases	F A1 controls	P	OR
4q31	exm-rs1429141	148288067	C	T	0.2005	0.1724	0.007986	1.204
	exm427501	148559892	T	C	0.003015	0.003365	1	0.8956
	exm427504	148560193	T	C	0.03317	0.03317	1	0.9998
	exm427572	148589711	A	G	0.007035	0.008413	0.6472	0.835
	rs6841581*	148620640	NA	NA	NA	NA	NA	NA
	exm2269978	148763687	C	T	0.1291	0.1418	0.179	0.8973
	exm-rs6845865	148974602	C	T	0.1704	0.1719	0.9136	0.9893
	exm427742	148984321	G	A	0.07085	0.07596	0.4995	0.9276
	exm2256398	149063523	T	C	0.3859	0.3964	0.4348	0.957
9p21	exm-rs4636294	21747803	A	G	0.4724	0.4928	0.1409	0.9214
	exm-rs7023329	21816528	A	G	0.4628	0.4882	0.06363	0.9031
	exm743465	21816758	A	G	0.4226	0.4108	0.3912	1.05
	exm-rs4977749	21927327	G	A	0.1839	0.1716	0.2374	1.088
	exm743529	21970916	T	C	0.03668	0.03053	0.2188	1.209
	exm743562	21974640	G	C	0.003518	0.003606	1	0.9755
	exm-rs3217992	22003223	A	G	0.396	0.3695	0.04601	1.119
	exm-rs1063192	22003367	C	T	0.4352	0.445	0.4756	0.9611
	exm-rs564398	22029547	G	A	0.4151	0.4216	0.6388	0.9734
	exm-rs7865618	22031005	G	A	0.4266	0.4334	0.6205	0.9727
	exm-rs2157719	22033366	G	A	0.4241	0.4334	0.5089	0.9628
	exm-rs1412829	22043926	C	T	0.4124	0.4211	0.5249	0.9649
	exm-rs1011970	22062134	T	G	0.1608	0.1707	0.3428	0.9311
	exm-rs1333040*	22083404	C	T	0.3925	0.437	0.0009317	0.8322

	exm-rs4977574	22098574	G	A	0.5176	0.4709	0.0006521	1.205
	exm-rs2383207	22115959	A	G	0.4598	0.5014	0.002264	0.8463
	exm-rs10757278	22124477	G	A	0.5101	0.4654	0.001063	1.196
	exm-rs1333048	22125347	C	A	0.5256	0.4808	0.00107	1.197
	exm-rs1333049	22125503	C	G	0.5121	0.4683	0.001415	1.192
	exm-rs7020996	22129579	T	C	0.1437	0.1394	0.6665	1.036
	exm-rs2383208	22132076	G	A	0.197	0.1906	0.557	1.042
	exm-rs10965250	22133284	A	G	0.1905	0.1779	0.2437	1.087
	exm2259290	22134172	T	C	0.4201	0.4406	0.1305	0.9197
	exm-rs1333051	22136489	T	A	0.1518	0.1404	0.2443	1.096
	exm-rs961831	22362104	C	A	0.1111	0.09663	0.08601	1.168
	exm743621	22447071	T	C	0.005528	0.005048	0.8503	1.096
8q11-12	exm701266	55049131	G	A	0.00603	0.005048	0.7093	1.196
	exm-rs1504749	55310711	G	T	0.2427	0.2065	0.001372	1.232
	exm-rs10958409*	55327091	A	G	0.1819	0.1442	0.0001734	1.319
	exm-rs10102164	55421614	A	G	0.1523	0.1839	0.002218	0.7971
	exm-rs9298506	55437524	G	A	0.1503	0.1808	0.002977	0.8013
	exm701397	55537560	T	C	0.01558	0.01034	0.08106	1.515
	exm701437	55539057	A	G	0.2879	0.2781	0.4307	1.05
	exm701450	55539395	T	A	0.3915	0.4151	0.08057	0.9062
	exm701478	55540268	T	C	0.003015	0.001683	0.3727	1.794
	exm701517	55541226	A	G	0.01508	0.01611	0.8272	0.935
	exm701530	55541450	A	G	0.2729	0.2625	0.3879	1.054
	exm701576	55542540	A	G	0.391	0.4142	0.08542	0.9079
	rs9298506*	55600077	NA	NA	NA	NA	NA	NA
	exm2266623	55630615	G	A	0.3432	0.3262	0.1927	1.079

	exm2262501	55914515	T	C	0.0603	0.05745	0.6841	1.053
10q24	exm2259513	104217503	T	G	0.06382	0.06755	0.6222	0.941
	exm-rs2281880	104269217	G	A	0.4729	0.4637	0.5121	1.037
	exm852396	104359297	T	G	0.004523	0.009135	0.0597	0.4928
	exm2273468	104436641	T	C	0.07789	0.08606	0.3006	0.8971
	exm2271344	104500659	T	C	0.4347	0.4714	0.007357	0.8622
	exm852626	104572963	T	C	0.3834	0.413	0.02819	0.8839
	exm852633	104573017	T	G	0.204	0.1858	0.08996	1.123
	exm-rs17115100	104591393	T	G	0.07437	0.0851	0.1633	0.8638
	exm-rs1004467	104594507	C	T	0.07437	0.08534	0.1493	0.8612
	exm-rs7096169	104618695	G	A	0.3216	0.3466	0.05368	0.8936
	exm852742	104638723	C	T	0.09799	0.1161	0.03666	0.827
	exm-rs11191447	104652323	T	C	0.06784	0.07885	0.1341	0.8502
	exm-rs11191454	104660004	G	A	0.06583	0.07668	0.1294	0.8485
	exm-rs12413409*	104719096	A	G	0.06533	0.07668	0.116	0.8416
	exm-rs7914558	104775908	A	G	0.3678	0.4036	0.007366	0.8598
	exm-rs11191548	104846178	C	T	0.06432	0.07428	0.1678	0.8567
	exm-rs11191580	104906211	C	T	0.06533	0.07548	0.1553	0.8561
	exm852895	104934709	C	T	0.1462	0.1442	0.8465	1.016
	exm852901	105005926	A	G	0.01759	0.01562	0.5904	1.128
	exm852902	105005931	T	C	0.007035	0.008894	0.5482	0.7895
	exm852936	105048217	C	A	0.02563	0.03317	0.1146	0.7666
	exm853029	105139477	A	G	0.004523	0.003606	0.6626	1.255
	exm853093	105160184	G	A	0.00402	0.004327	1	0.9288
	exm853108	105166468	T	C	0.003015	0.002404	0.7896	1.255
	exm853150	105176350	T	C	0.004523	0.003606	0.6626	1.255

	exm853161	105177645	T	G	0.07337	0.06947	0.5951	1.061
	exm853174	105178380	A	G	0.005528	0.006971	0.6121	0.7918
	exm853210	105183348	C	T	0.01156	0.01322	0.628	0.8727
	exm853248	105194086	T	C	0.3829	0.3858	0.8446	0.9878
	exm853285	105201595	C	A	0.004523	0.004567	1	0.9902
	exm2249010	105203785	T	C	0.01407	0.01803	0.2888	0.7773
	exm853326	105203785	T	C	0.01407	0.01803	0.2888	0.7773
	exm853347	105205176	G	A	0.005528	0.006731	0.7316	0.8203
	exm853352	105205260	G	A	0.00603	0.003606	0.2153	1.676
	exm853355	105205302	C	A	0.3638	0.3632	0.9774	1.003
	exm853379	105207301	T	C	0.00402	0.005769	0.4512	0.6956
13q13	exm2272002	32301038	G	T	0.2121	0.199	0.2361	1.083
	exm1061355	32351535	C	A	0.008543	0.007212	0.6389	1.186
	exm1061358	32352714	G	A	0.004523	0.004327	1	1.045
	exm1061405	32371361	G	A	0.1407	0.1315	0.3377	1.082
	exm1061408	32371409	A	G	0.006533	0.003606	0.1543	1.817
	rs9315204*	32591837	NA	NA	NA	NA	NA	NA
	exm2271918	32755834	T	C	0.4905	0.4928	0.8701	0.9907
	exm1061564	32776604	A	G	0.004523	0.009861	0.03271	0.4562
	exm1061566	32776616	A	T	0.04673	0.04135	0.3474	1.137
	exm1061652	32811607	A	G	0.3407	0.3397	0.9541	1.005
	exm1061662	32811974	T	C	0.01608	0.01082	0.08686	1.494
	exm1061861	32906729	C	A	0.3101	0.3014	0.4958	1.041
	exm1061931	32911463	G	A	0.02362	0.02452	0.8598	0.9624
	exm1061964	32912750	T	G	0.007538	0.008413	0.7642	0.8951
	exm1062021	32914236	T	C	0.02814	0.02644	0.7366	1.066

	exm1062034	32914592	T	C	0.006533	0.004567	0.3447	1.433
	exm1062116	32937488	T	G	0.00402	0.002404	0.314	1.675
	exm1062132	32945172	C	A	0.003015	0.002885	1	1.045
	exm-rs4942486	32953388	T	C	0.4729	0.4846	0.3978	0.954
	exm1062140	32953549	T	G	0.003015	0.002163	0.583	1.395
	exm1062141	32953550	A	G	0.004523	0.005048	0.8476	0.8954
	exm1062190	32972626	T	A	0.01055	0.005769	0.0535	1.838
	exm1062237	33016660	T	C	0.0196	0.01827	0.7629	1.074
	exm1062246	33017043	C	T	0.1739	0.1714	0.8285	1.017
	exm1062250	33017158	T	G	0.0196	0.01803	0.6865	1.089
	exm2251404	33045639	G	A	0.397	0.3906	0.6353	1.027
18q11	**rs11661542***	18477693	NA	NA	NA	NA	NA	NA
	exm1377410	18546896	G	T	0.003518	0.003125	0.813	1.126
	exm1377435	18564483	C	G	0.01407	0.01322	0.8135	1.065
	exm1377493	18964286	A	G	0.003015	0.003846	0.8198	0.7833
2q33	exm254984	197735679	C	G	0.006533	0.00649	1	1.007
	exm255118	197943489	G	T	0.003015	0.003606	0.8184	0.8357
	exm255160	197990741	C	A	0.07337	0.07716	0.6434	0.9469
	exm2265382	198143381	A	G	0.4246	0.4329	0.5449	0.9666
	rs919433*	198166565	NA	NA	NA	NA	NA	NA
	exm2269203	198166825	C	T	0.4196	0.3666	6.884e-05	1.249
	exm255429	198436771	C	A	0.007538	0.01058	0.3274	0.7105
	exm255453	198482574	G	T	0.007035	0.006971	1	1.009
	exm255456	198495767	G	A	0.005528	0.005048	0.8503	1.096
	exm255464	198498566	C	A	0.01206	0.0101	0.509	1.197

	exm255467	198498595	T	C	0.003015	0.002644	0.7984	1.141
	exm2265254	198511086	A	G	0.4894	0.4899	0.9783	0.9982
	exm-rs700651	198631714	G	A	0.3819	0.3492	0.01338	1.152
	exm255599	198646549	T	G	0.003015	0.002644	0.7984	1.141
5q31	exm477708	132387979	A	T	0.01759	0.01948	0.6888	0.9011
	exm477724	132408967	A	G	0.01809	0.0274	0.0271	0.6539
	exm477788	132437499	T	C	0.01005	0.01178	0.6062	0.8518
	exm477823	132535046	A	G	0.192	0.1873	0.676	1.031
	exm477871	132556523	T	C	0.00603	0.005529	0.8566	1.091
	exm477889	132561468	A	C	0.08543	0.08317	0.7682	1.03
	exm2266006	132573216	A	G	0.3427	0.3452	0.8634	0.9891
	exm2270247	132605601	C	T	0.03166	0.02163	0.02257	1.478
	exm477931	132652281	T	C	0.1211	0.1106	0.2299	1.108
	exm2256932	132754003	A	G	0.3663	0.3887	0.09237	0.9092
	rs113816216*	132846228	NA	NA	NA	NA	NA	NA
6q24	exm2270422	145677803	T	C	0.4417	0.4233	0.1773	1.078
	rs75018213*	146052178	NA	NA	NA	NA	NA	NA
	exm584253	146125793	T	A	0.4106	0.4248	0.2942	0.9433
	exm584278	146126419	T	C	0.4106	0.4248	0.2942	0.9433
	exm584288	146126580	A	G	0.01307	0.01322	1	0.9881
	exm584389	146243968	A	G	0.00603	0.007933	0.5226	0.7587
	exm584484	146273614	T	C	0.003518	0.001923	0.2711	1.832
	exm2266267	146394655	G	A	0.4387	0.4397	0.9562	0.9961
7p22	exm2266511	19104468	A	G	0.04221	0.03413	0.128	1.247
	exm2270585	19337064	C	T	0.4739	0.4865	0.3539	0.9505

rs10230207*	19577832	NA	NA	NA	NA	NA	NA
exm-rs4470914	19616522	T	C	0.1648	0.1675	0.7981	0.9805
exm-rs6977660	19805480	T	C	0.1337	0.144	0.2899	0.9173
exm2270730	19980396	C	T	0.1412	0.1127	0.001548	1.294
exm2270586	19999873	T	C	0.3492	0.337	0.3576	1.056
exm-rs38152	20047575	G	T	0.02412	0.02885	0.316	0.8321

SNP indicates single nucleotide polymorphism; BP, basepair position; A1, allele 1; A2, allele 2; F, frequency; P, p-value; OR, odds ratio;
NA, not available.

Common SNPs (MAF > 0.05) are displayed in grey, and low-frequent SNPs (MAF < 0.05) are displayed in black.

** Index SNPs from previous GWAS in which the risk locus was discovered*
*** Genomic coordinates are based on the reference assembly GRCh37 from the Genomic Reference Consortium*

Part II

Shared genetic background of intracranial aneurysms and related diseases

Chapter 5

Shared genetic risk factors of intracranial, abdominal and thoracic aneurysms

FNG van 't Hof, YM Ruigrok, CH Lee, S Ripke, G Anderson, M de Andrade, AF Baas, JD Blankensteijn, EP Böttinger, MJ Bown, J Broderick, P Bijlenga, DS Carrell, DC Crawford, DR Crosslin, C Ebeling, JG Eriksson, M Fornage, T Foroud, M von Und Zu Fraunberg, CM Friedrich, EI Gaál, O Gottesman, DC Guo, SC Harrison, J Hernesniemi, A Hofman, I Inoue, JE Jääskeläinen, GT Jones, LA Kiemeney, R Kivisaari, N Ko, S Koskinen, M Kubo, IJ Kullo, H Kuivaniemi, MI Kurki, A Laakso, D Lai, SM Leal, H Lehto, SA LeMaire, SK Low, J Malinowski, CA McCarty, DM Milewicz, TH Mosley, Y Nakamura, H Nakaoka, M Niemelä, J Pacheco, PL Peissig, J Pera, L Rasmussen-Torvik, MD Ritchie, F Rivadeneira, AM van Rij, RL Santos-Cortez, A Saratzis, A Slowik, A Takahashi, G Tromp, AG Uitterlinden, SS Verma, SH Vermeulen, GT Wang; Aneurysm Consortium; Vascular Research Consortium of New Zealand, B Han, GJE Rinkel, PIW de Bakker.

Journal of the American Heart Association. 2016 Jul 14;5(7).

Abstract

Background

Intracranial aneurysms (IA), abdominal aortic aneurysms (AAA) and thoracic aortic aneurysms (TAA) all have a familial predisposition. As aneurysm types are known to co-occur, we hypothesized that there may be shared genetic risk factors for IA, AAA and TAA.

Methods and results

We performed a mega-analysis of 1000 Genomes Project-imputed genome-wide association study (GWAS) data of four previously published aneurysm cohorts: two IA cohorts (in total 1,516 cases, 4,305 controls), one AAA cohort (818 cases, 3,004 controls), and one TAA cohort (760 cases, 2,212 controls), and observed associations of four known IA, AAA and/or TAA risk loci (9p21, 18q11, 15q21 and 2q33) with consistent effect directions in all four cohorts. We calculated polygenic scores based on IA-, AAA- and TAA-associated SNPs and tested these scores for association to case-control status in the other aneurysm cohorts; this revealed no shared polygenic effects. Similarly, LD-score regression analyses did not show significant correlations between any pair of aneurysm subtypes. Lastly, we evaluated the evidence for 14 previously published aneurysm risk SNPs through collaboration in extended aneurysm cohorts, with a total of 6,548 cases and 16,843 controls (IA), and 4,391 cases and 37,904 controls (AAA), and found nominally significant associations for IA risk locus 18q11 near *RBBP8* to AAA (OR=1.11, $p=4.1x10^{-5}$) and for TAA risk locus 15q21 near *FBN1* to AAA (OR=1.07, $p=1.1x10^{-3}$).

Conclusion

Although there was no evidence for polygenic overlap between IA, AAA and TAA, we found nominally significant effects of two established risk loci for IA and TAA in AAA. These two loci will require further replication.

Introduction

Intracranial aneurysms (IA), abdominal aortic aneurysms (AAA) and thoracic aortic aneurysms (TAA) are three different forms of arterial vessel wall dilatations, which can all lead to rupture with a high case fatality rate.[2, 96, 97] A co-occurrence of AAA and TAA is known,[16, 17] and a weak co-occurrence of IA and AAA has also been suggested.[18, 19] Furthermore, IA, TAA and AAA cluster within affected families,[20, 21] suggesting a shared genetic background of these three diseases.

Genome-wide association studies (GWAS) have revealed a handful of genetic risk factors for IA,[10, 11, 13] AAA[98-101] and TAA,[102] and of these identified loci, locus 9p21 near *CDKN2A*, *CDKN2B* and *CDKN2BAS* appears to be shared by IA and AAA.[11, 12] However, the heritability explained by the risk loci identified to date is low for all three diseases, suggesting that many aneurysm risk loci remain to be discovered. Among these unknown and known risk loci, there could be shared risk loci for multiple types of aneurysms.

In this study, we searched for shared genetic risk factors for aneurysms. We combined individual participant GWAS data from a Dutch and a Finnish IA cohort,[10] a Dutch AAA cohort[99] and a TAA cohort from the USA[102] for a GWAS mega-analysis across these traits. In a complementary approach, we performed a polygenic analysis to test groups of single nucleotide polymorphisms (SNPs) for a joint effect on risk across diseases. Finally, we tested the effect of *bona fide* risk SNPs from previously published IA, AAA and TAA GWAS on the other aneurysm types by meta-analyzing summary statistics in the GWAS cohorts of IA, AAA and TAA, extended by association results of additional IA and AAA GWAS cohorts.

Methods

Study populations

For the GWAS mega-analysis and polygenic analysis, we used data of subjects genotyped in previously published GWAS cohorts of aneurysm cases and controls: two IA cohorts, one from the Netherlands[10, 103, 104] and one from Finland,[10, 11, 105, 106] one AAA cohort from the Netherlands[12, 99] and one TAA cohort from the USA.[102] All studies were approved by the relevant medical ethical committees, and all participants provided written informed consent. All study populations were previously described in detail.[10, 102-104] Below is a brief description of each study population.

IA cases in the Dutch cohort (n=786) were admitted to the University Medical Center Utrecht, the Netherlands between 1997 and 2011. All cases were genotyped on Illumina CNV370 Duo BeadChips. Controls (n=2,089) were ascertained via the Rotterdam Study, a population-based cohort of subjects aged 45 years and older recruited from a district in Rotterdam (The Netherlands). These controls were genotyped on Illumina HumanHap550 BeadChips.[107]

The Finnish IA cohort consisted of 790 cases treated at the Helsinki and Kuopio University Hospitals, and 2,396 controls that were genetically matched to cases.[15] Of these, 1,666 controls were extracted from the Helsinki Birth Cohort Study (HBCS).[105] Additionally, 651 controls were extracted from anonymous donors from Kuopio University Hospital and Helsinki, and from the Health 2000 study (H2000).[106] All cases were genotyped on Illumina CNV370 Duo BeadChips, and controls were genotyped on Illumina HumanHap550 BeadChips (HBCS), and on Illumina CNV370 Duo BeadChips (anonymous donors and H2000).

The Dutch and the Finnish IA cohort both included cases with ruptured and unruptured IA. Ruptured IA cases were defined by symptoms suggestive of SAH combined with subarachnoid blood on a computed tomography (CT) scan and a proven IA at angiography (conventional angiogram, CT- or magnetic resonance (MR)-angiogram). Unruptured IA cases were identified by CT or MR angiography or conventional angiography in the absence of clinical or radiological signs of SAH. Patients with fusiform IA, possible traumatic SAH, and polycystic kidney disease were excluded.

The AAA cohort consisted of 859 cases, predominantly with unruptured AAA. These cases were recruited from eight medical centers in the Netherlands, mainly when individuals visited their vascular surgeon in the outpatient clinic or, in some cases, during hospital admission for elective or emergency AAA surgery. An AAA was defined as an infrarenal aorta diameter of ≥ 30 mm. The mean AAA diameter was 58.4 mm. Of these patients, 530 had undergone surgery, including 43 with rupture. Genotyping was performed on Illumina HumanHap610 chips.[99] As controls, we included 3,110 Dutch subjects who were recruited as part of the Nijmegen Biomedical Study (n=1,832) and the Nijmegen Bladder Cancer Study (n=1,278).[103, 104] These controls were genotyped on Illumina CNV370 Duo BeadChips.

The TAA cohort consisted of 765 cases with either an ascending thoracic aortic aneurysm without dissection (n=401) or with a type A and/or type B

aortic dissection (n=364). All cases were more than 30 years old, did not have a first-degree relative with TAA or dissection, and had no evidence of a syndromic form of TAA or dissection. The samples were genotyped with Illumina CNV370-Quad BeadChips. Controls (n=2,229) were included from the Wellcome Trust Case-Control Consortium (WTCCC) 1958 Birth Cohort (n=1,355), and from the US National Institute of Neurological Disorders and Stroke (NINDS) Repository's Neurologically Normal Control Collection (n=874).

Quality control

We performed quality control (QC) in each of the four cohorts separately, using PLINK version 1.07.[108] After removal of SNPs with A/T or C/G alleles and SNPs that were not called in any individual, we performed sample QC and SNP QC.

Sample QC was performed after merging cases and controls, using a subset of common, high-quality SNPs (as defined by SNPs without deviation from Hardy-Weinberg equilibrium (HWE) ($p>0.001$), with high minor allele frequency (MAF) (>20%) and with low rate of missing genotypes (<1%)). Linkage disequilibrium (LD) pruning ($r^2>0.5$) was performed. Subjects were removed based on the following three criteria: missing genotypes (subjects with a call rate below 95% were removed), heterozygosity (subjects were excluded if the inbreeding coefficient deviated more than 3 standard deviations from the mean) and cryptic relatedness (by calculating identity-by-descent (IBD) for each pair of individuals). In each pair with an IBD proportion of >20%, a subject was excluded, if it exhibited distant relatedness with more than one individual. For case-control pairs, we removed the control subject. In the case-case or control-control pairs, the subject with the lowest call rate was excluded.

Using these common, high-quality SNPs, we performed principal components (PC) analysis using EIGENSTRAT on the remaining study subjects and HapMap-CEU subjects. We excluded SNPs from three regions with known long-distance LD: the major histocompatibility (MHC) region (chr6: 25.8-36 Mbp), the chromosome 8 inversion (chr8: 6-16 Mbp) and a chromosome 17 region (chr17: 40-45 Mbp). We created PC plots with the first four PCs, using R version 2.11.[46] Based on visual inspection of these plots, we excluded subjects that appeared to be outliers with respect to the CEU or the study population. After outlier removal, we recomputed PCs for them to be included as covariates in the logistic regression models. PC plots after outlier removal are shown in Figure 1.

After sample QC, we excluded SNPs with more than 2% missing genotypes, MAF < 1%, missing genotype rate higher than MAF, and HWE deviation ($p<0.001$). Because cases and controls had been genotyped separately, we performed these QC steps in each study cohort separately and again after merging cases and controls. We also removed SNPs with a differential degree of missing genotypes between cases and controls ($p<1x10^{-5}$; chi-squared test).

Imputation

For each case-control dataset, we performed genotype imputation using the pre-phasing/imputation stepwise approach implemented in IMPUTE2 and SHAPEIT (chunk size of 3 Mb and default parameters).[109, 110] The imputation reference set consisted of 2,184 phased haplotypes from the full 1000 Genomes Project data set (February 2012; 40,318,253 variants). All genomic locations are given in NCBI Build 37/UCSC hg19 coordinates. After imputation, SNPs with an imputation accuracy score < 0.6 or MAF < 0.5% were excluded.

GWAS mega-analysis across IA, AAA and TAA

We performed a mega-analysis on all four GWAS cohorts. Association testing was carried out in PLINK[108] using imputed SNP dosages. We included as covariates the first four PCs, and an indicator variable to adjust for each case-control dataset. SNPs with $p<5x10^{-8}$ were considered as genome-wide significant. We also performed a GWAS on each cohort separately and a combined analysis of only the Dutch and Finnish IA cohort, which was needed for the polygenic analysis as described below. We calculated genomic inflation factors (λ_{GC}) for each GWAS and the mega-analysis, defined as the ratio of the median of the empirically observed distribution of the test statistic to the expected median.[80]

We calculated the statistical power for detecting a significant association ($p<5x10^{-8}$) in the mega-analysis using the genetic power calculator.[86] In case of a risk allele frequency of 10%, the resulting power is 0.03% at a relative risk of 1.1 per allele, and 68.7% at a relative risk of 1.3 per allele, assuming additive effects. In case of a higher risk allele frequency of 20%, the power is 0.3% at a relative risk of 1.1, and 98.9% at a relative risk of 1.3 per allele.

Polygenic analysis

We performed polygenic analysis as previously described.[111] We used the IA cohorts (Dutch and Finnish cohorts combined) as a discovery sample

and the AAA and TAA cohorts as two separate target samples, and vice versa, in all possible combinations (six in total).

We pruned the SNPs genotyped or imputed in the discovery sample, using a LD threshold of r^2>0.1. For each genomic region we chose SNPs with the lowest p-values in the GWAS of the discovery sample, in order to retain a set of independent, maximally associated SNPs. Next, we created sets of SNPs with disease association in the discovery sample at twelve different significance thresholds, increasing from $p<5x10^{-8}$ to p<0.5. For each SNP set, we calculated a polygenic risk score in each individual of target sample as follows:

Genetic score $= \beta_1 x_1 + \beta_2 x_2 + ... + \beta_n x_n$,

where x_i is the estimated allele dosage (between 0 and 2) in a given individual, and β_i is the effect size from the GWAS in the discovery sample for the i^{th} SNP. We tested the association between these polygenic risk scores and case-control status in the target sample using logistic regression, adjusting for the first four PCs in the target sample. For analyses involving the combined IA cohort as target sample, we also adjusted for population (Finnish or Dutch). For analyses involving the IA and AAA cohorts as target and discovery sample, we also adjusted for the genotype of the known shared IA/AAA risk SNPs on locus 9p21.[11, 12]

LD score regression: heritability estimation and genetic correlation analyses

We used LD score regression (LDSC) for heritability estimation and genetic correlation analysis.[112] These analyses were applied to the same four GWAS datasets as used in the GWAS mega-analysis and polygenic analysis described above (Dutch IA, Finnish IA, AAA and TAA), but we only included genotyped SNPs (after QC) and not imputed SNPs for this analysis.

LDSC is a new approach that implements LD-score weighted linear regression methods to estimate the variance (and covariance) explained by all SNPs on the whole genome for a complex trait. This method can distinguish true polygenicity from confounding effects due to population structure and cryptic relatedness. Under a polygenic model, the more genetic variants an index variant tags, the more the probability that the index variant will be significant.[112] Under this reasoning, the expected χ^2 statistic of variant j can be constructed, as follows:

$$E\left[\chi^2 \middle| \ell_j\right] = Nh^2\ell_j/M + Na + 1,$$

where N is the sample size; M is the number of SNPs, so that h^2/M is the average heritability per SNP; *a* measures the contribution of confounding biases, like cryptic relatedness and population stratification; and $\ell_j = \sum_k r_{jk}^2$ is the sum of the r^2 values to all variants that a variant j tags, which is called the LD score of variant j. The LD score can be calculated using reference panel containing whole genome information of the population. For the analysis of European population, we used the LD estimates that are from the European-ancestry samples in the 1000 Genomes Project.[38] Details were elaborated from the LDSC's GitHub repository.[112]

In addition to estimating heritability using LDSC, we also used LDSC to estimate genetic correlation (ρ) between each pair of three diseases, AAA, IA, and TAA.[113] We investigated the Dutch and Finnish IA cohort separately in this analysis, and also tested the correlation between these two IA datasets, to evaluate the genetic correlation within one disease.

Effects of previously established risk SNPs

We investigated the effect of established IA,[10, 11, 13] AAA[98-101] and TAA[102] risk SNPs from previous GWAS in each of the other aneurysm types. We looked up the effect sizes and significance of these SNPs in separate case-control data sets for IA, AAA and TAA beyond the individual-participant data sets already described above. We did not have access to all genotype data of these data sets, but we obtained association results for these candidate SNPs only. For IA, we used GWAS results of two Japanese cohorts,[11] a combined cohort of several previously IA studies that recruited subjects from mainly North America, but also Poland and Australia (further referred to as 'USA+'),[14] and the @neurIST study.[11] For AAA, we used GWAS results of a cohort from Iceland, the United Kingdom, the USA, and two cohorts from New Zealand.[100, 101, 114, 115] For TAA, no additional GWAS cohorts are currently available to our knowledge. The sample sizes and other details per cohort are listed in Table 1. In total, the target samples available for this look-up analysis contained 6,548 cases and 16,843 controls for IA, 4,391 cases and 37,904 controls for AAA, and 760 cases and 2,212 controls for TAA.

We combined the GWAS results for the selected SNPs in the extended IA and AAA cohorts using an inverse-variance fixed-effects meta-analysis. For each SNP, we first calculated z-scores from the provided p-values of each GWAS, and summed the z-scores across all studies using the effective sample size of each study as weights.[116] The resulting z-scores were

converted into chi-square values and two-sided p-values. We applied Bonferroni correction for performing 28 association tests (14 selected SNPs tested in two aneurysm types), and considered associations with $p<1.8x10^{-3}$ (0.05/28) as significant.

Figure 2 gives an overview of the methods described above.

Results

Study populations

Table 2 shows the numbers of cases, controls and SNPs of all four cohorts after quality control and imputation. Quantile-quantile plots for each GWAS per cohort are shown in Figure 3.

GWAS mega-analysis across IA, AAA and TAA

In total, 3,094 cases, 9,507 controls and 9,245,988 SNPs were available for the mega-analysis across all four aneurysm cohorts (λ_{GC} = 1.06). The results of this mega-analysis are shown in a Manhattan plot (Figure 4). We found four genome-wide significant loci, though all these loci were previously described as risk loci for IA, AAA and/or TAA. The direction of effect for these loci was consistent across all four aneurysm cohorts (see forest plots in Figure 5). First, SNPs at the known IA and AAA risk locus 9p21 near *CDKN2A*, *CDKN2B* and *CDKN2BAS* were associated.[11, 12] The strongest association at this locus was found for rs7866503, with $p=2.1x10^{-13}$. The second association was found for SNPs at the known IA risk locus 18q11 near *RBBP8*,[11] with the strongest association for rs8087799 ($p=1.6x10^{-9}$). The third association was found for SNPs at the known TAA risk locus 15q21 near *FBN1*,[102] with the strongest association for rs595222 ($p=1.0x10^{-8}$). The fourth association was found for rs919433 ($p=4.6x10^{-8}$), which is located at 2q33 near *ANKRD44*. The same SNP was previously found to be associated with IA in a Finnish and Dutch population.[15] This SNP is also in strong LD ($r^2=0.7$) with a nearby SNP (rs700651), which was previously found to be associated with IA in a Dutch, Finnish and Japanese population,[10] but did not reach genome-wide significance after adding other populations of IA patients.[11]

Polygenic analysis

Next, we investigated whether groups of SNPs associated with one type of aneurysm (e.g. IA) were also associated with the other types (e.g. AAA or TAA). The results of these polygenic analyses with the IA, AAA and TAA GWAS cohorts are shown in Tables 3-8, and in Figure 6. No SNP sets with a significant joint effect on another aneurysm type were observed, except

for a small group of seven SNPs associated with IA with $p<10^{-6}$. This SNP set was associated with AAA with $p=5x10^{-3}$ (Table 7). When taking a closer look at this SNP set, it appears that the polygenic association is driven by two SNPs (rs36071109 at 2q33, $p_{AAA}=2.3x10^{-3}$ and rs4330012 at 18q11, $p_{AAA}=1.3x10^{-2}$), both in very strong LD to genome-wide significant SNPs in our mega-analysis described above. (The pleiotropic 9p21 locus shared by IA and AAA had already been adjusted for.)

LD score regression: heritability estimation and genetic correlation analyses

We attempted to analyze heritability of the three aneurysm subtypes using LDSC. The estimated heritability was 0.15 for AAA, 0.31 for the Finnish IA cohort, 0.34 for the Dutch IA cohort, and 0.40 for TAA (Table 9). These estimates are smaller than the twin-based estimates as reported in the literature (0.41 for IA,[6] 0.70 for AAA[101] and unknown for TAA) but the standard errors (SE) are large, so the differences are not statistically significant. Note that methods estimating heritability using SNP data can often underestimate the heritability if the SNP set does not tag all underlying causal variants.

We then performed genetic correlation analyses (Table 10). We found that all pairs did not show statistically significant genetic correlations ($p>0.05$). The direction of correlation was either close to zero or positive, but the SE values were large. As expected, we observed the largest genetic correlation between the two cohorts of the same IA disease ($\rho=1.59$). However, this correlation was also not significant ($p=0.09$, SE=0.93).

Effects of previously established risk SNPs

Table 11 shows the results of the look-up of previously published aneurysm risk SNPs in IA, AAA and TAA GWAS results. Besides the two IA and AAA risk SNPs at 9p21, which are associated with IA and AAA but not TAA, we observed two SNPs with significant associations to another aneurysm type after multiple testing correction. First, the IA risk SNP rs11661542 at 18q11 near *RBBP8* was associated with AAA (OR=1.11, $p=4.1x10^{-5}$). Second, the TAA risk SNP rs2118181 at 15q21 near *FBN1* was associated with AAA (OR=1.07, $p=1.1x10^{-3}$). The other TAA risk SNP rs10519177, which lies at the same locus but is independent from rs2118181, showed a suggestive but not statistically significant association to AAA (OR=1.01, $p=0.016$). For both significant associations, the direction of effect was concordant in all but one of the six AAA cohorts in the analysis, with no significant heterogeneity in the meta-analysis of the six cohorts ($p_{Cochran's\ Q}$ = 0.11 for rs11661542 at 18q11 and 0.45 for rs2118181 at 15q21).

Discussion

In this study we have applied multiple analytic approaches to detect a possible genetic overlap between IA, TAA and AAA. By performing a GWAS mega-analysis and polygenic analysis, we considered both sharing of significant risk loci with individually large effects on disease risk, as well as a cumulative effect of many loci with individually weak effects. Although we did not find novel shared aneurysm risk loci that were previously not described as risk loci for IA, AAA and/or TAA, we did find some evidence for a shared genetic background of IA, AAA and TAA. In the GWAS mega-analysis across IA, AAA and TAA, we detected genome-wide significant associations for SNPs at four loci: 9p21, 18q11, 15q21 and 2q33. These were all previously described to be associated with IA, AAA or TAA, but in this study they showed globally consistent effects across all three aneurysm types. Polygenic analysis did not reveal any groups of weakly IA-, AAA- or TAA-associated loci with a joint effect on other aneurysm types. Similarly, LD-score regression analyses did not show significant correlations between any pair of aneurysm subtypes. The correlation between two cohorts of the same subtype (IA) was relatively high but also not significant. The absence of novel shared loci, a polygenic association or a genetic correlation can be due to the modest sample size and power, certainly in comparison to community-wide efforts for coronary artery disease.[117] It may therefore be premature to claim that there are genuinely no novel shared loci or shared polygenic effects for IA, AAA and TAA.

After testing the association of *bona fide* aneurysm risk loci in other aneurysm types in much larger aneurysm GWAS cohorts (consisting of the mega-analysis data plus additional IA and AAA GWAS cohorts), we found nominally significant associations of the IA risk locus 18q11 and the TAA risk locus 15q21 to AAA, both of which were previously unknown to be associated with AAA. The 15q21 locus has reported biological functions that could plausibly be related to aneurysm development, because it encompasses the *FBN1* gene, which encodes fibrillin-1, an extracellular matrix protein in the elastic fibers of the aortic wall. Mutations in *FBN1* cause Marfan syndrome, which is often associated with (mainly thoracic) aortic aneurysms.[102]

There are extensive differences in pathophysiology and epidemiological risk factors between the three diseases. IA are mostly saccular shaped dilatations, while AAA and TAA are more often fusiform; the vessel wall structure differs between the locations where IA, AAA and TAA occur; and atherosclerosis has a clear role in AAA, in contrast to IA and TAA.[118-121]

However, there have been previous reports of a co-occurrence of AAA and TAA,[16, 17] and to a lesser extent of IA and aortic aneurysms.[18, 19] Also, IA mainly co-occurs with AAA and TAA within families,[20, 21] so genetic sharing between IA, AAA and TAA could theoretically be present in the familial form, but not in sporadic cases (which were primarily studied here). For example, mutations in *TGFBR1* and *TGFBR2* were found in families in which all three aneurysm types occur.[20] Furthermore, a genetic linkage study in 26 families with both IA and AAA or TAA found linkage peaks at locus 6p23 and 11q24.[122] Independent linkage studies on IA, AAA and TAA also found linkage peaks at 11q24.[123] We did not find associations of these loci in our study, which can be explained by our focus on the sporadic aneurysm cases, in contrast to the familial cases in the linkage studies.

In conclusion, our study results do not reveal new risk loci shared between IA, AAA and TAA, but the effects of established IA, AAA and TAA risk loci in other aneurysm types do suggest a shared genetic background. Future studies with larger sample sizes should increase the statistical power to detect common genetic variants with a smaller effect on disease risk, and to draw definitive conclusions on genetic correlation between aneurysm subtypes.

Figures

Figure 1 is available with the online version of this paper.

Figure 2. Work-flow figure. This figure gives an overview of the study methods.

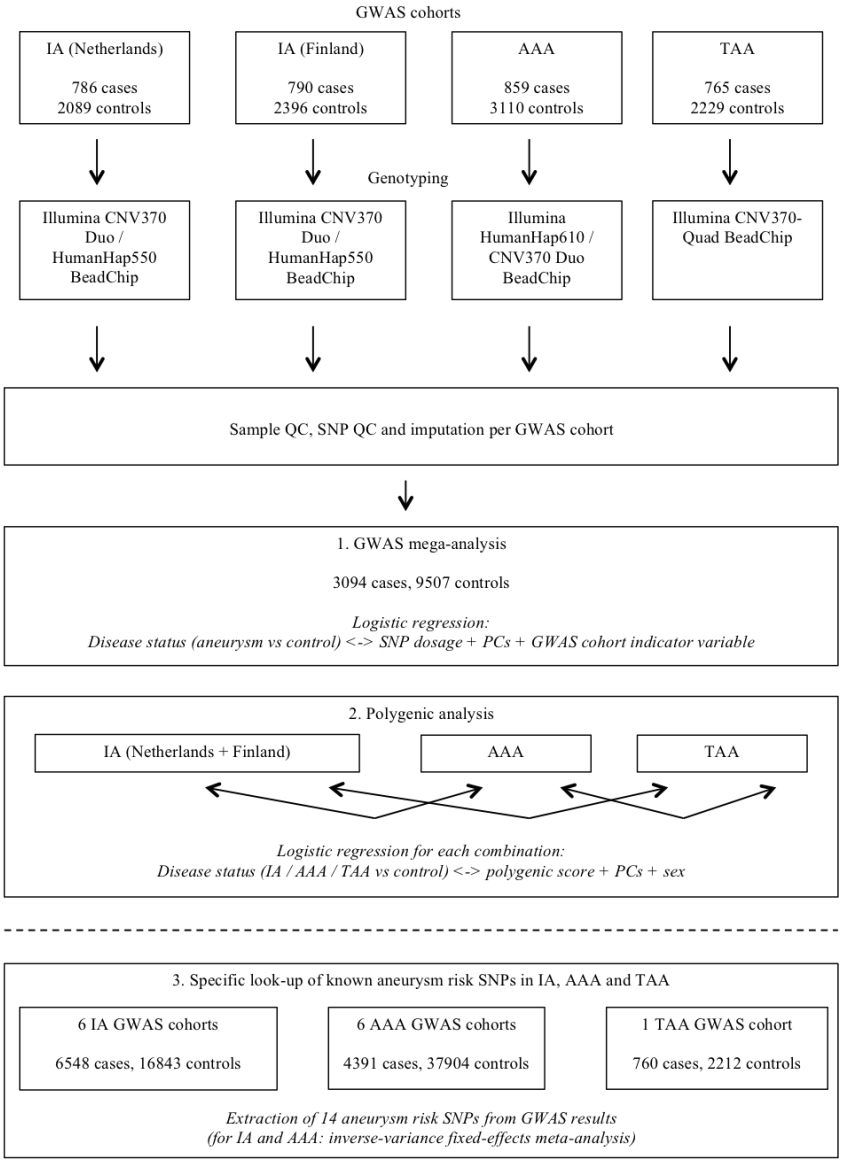
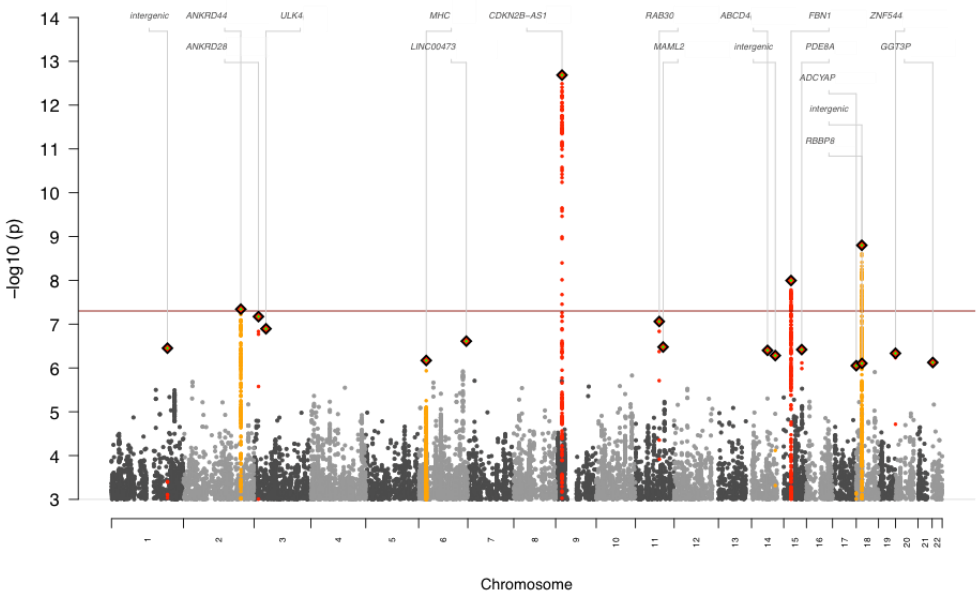


Figure 3 is available with the online version of this paper.

Figure 4. Manhattan plot of the aneurysm GWAS mega-analysis.

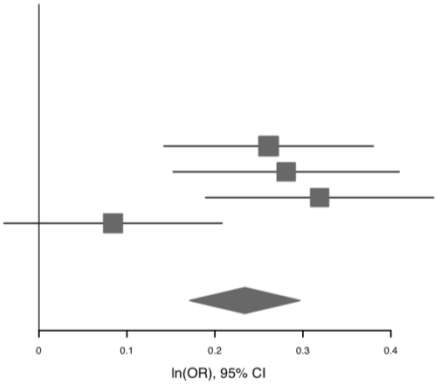


This Manhattan plot shows the p-values of all SNPs with an association with $p<10^{-4}$ to disease (IA, AAA or TAA). P-values on the y-axis are presented on an inverse log scale. The x-axis represents the genomic position of each SNP. The red horizontal line represents $p=5x10^{-8}$, the cut-off value for genome-wide association. Index SNPs with $p<1x10^{-6}$ are depicted as diamonds, while SNPs in the same LD block as these SNPs are depicted as yellow and red dots.

Figure 5. Forest plots of significant SNPs from the aneurysm mega-analysis.

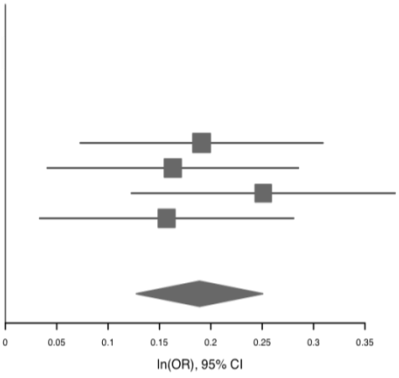
A. rs7866503 (risk allele T) at locus 9p21

			het_P:	0.05			
Cohort	**info**	**p_value**	**f_ca(n)**	**f_co(n)**	**OR**	**ln(OR)**	**STDerr**
AAA	0.88	1.72E-05	0.480(818)	0.427(2991)	1.30	0.261	0.0607
IA (Finland)	0.86	1.79E-05	0.426(799)	0.368(2317)	1.32	0.281	0.0655
IA (Netherlands)	0.90	1.44E-06	0.489(717)	0.421(1987)	1.38	0.319	0.0661
TAA	0.91	1.82E-01	0.465(760)	0.445(2212)	1.09	0.0842	0.0632
Mega-analysis	**0.89**	**2.06E-13**	**0.465(3094)**	**0.416(9507)**	**1.26**	**0.234**	**0.0319**

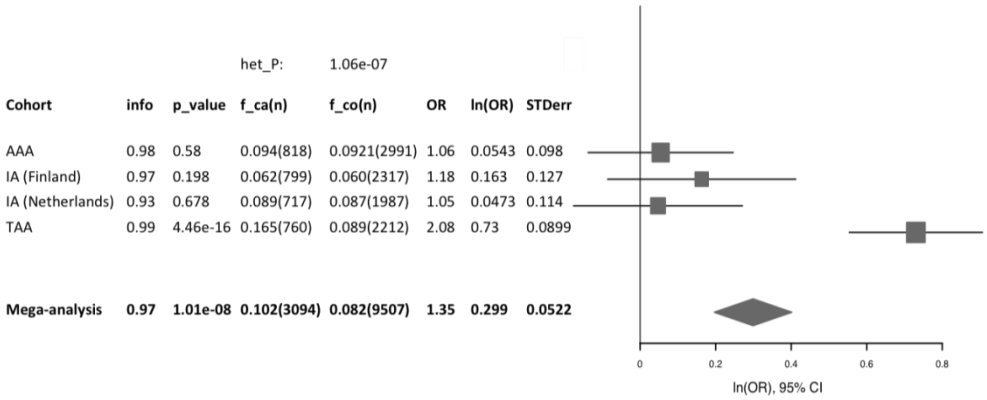


B. rs8087799 (risk allele A) at locus 18q11

			het_P:	0.73			
Cohort	**info**	**p_value**	**f_ca(n)**	**f_co(n)**	**OR**	**ln(OR)**	**STDerr**
AAA	0.96	0.00153	0.364(818)	0.321(2991)	1.21	0.191	0.0603
IA (Finland)	1.02	0.00898	0.353(799)	0.319(2317)	1.18	0.163	0.0624
IA (Netherlands)	0.98	0.00013	0.387(717)	0.332(1987)	1.29	0.251	0.0655
TAA	0.99	0.0126	0.367(760)	0.327(2212)	1.17	0.157	0.0631
Mega-analysis	**0.99**	**1.58e-09**	**0.367(3094)**	**0.324(9507)**	**1.21**	**0.189**	**0.0314**

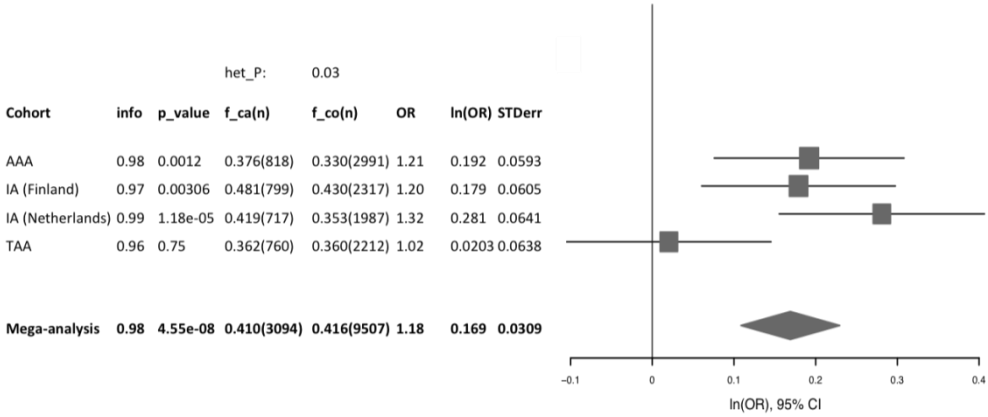


C. rs595244 (risk allele T) at locus 15q21



			het_P:	1.06e-07			
Cohort	**info**	**p_value**	**f_ca(n)**	**f_co(n)**	**OR**	**ln(OR)**	**STDerr**
AAA	0.98	0.58	0.094(818)	0.0921(2991)	1.06	0.0543	0.098
IA (Finland)	0.97	0.198	0.062(799)	0.060(2317)	1.18	0.163	0.127
IA (Netherlands)	0.93	0.678	0.089(717)	0.087(1987)	1.05	0.0473	0.114
TAA	0.99	4.46e-16	0.165(760)	0.089(2212)	2.08	0.73	0.0899
Mega-analysis	**0.97**	**1.01e-08**	**0.102(3094)**	**0.082(9507)**	**1.35**	**0.299**	**0.0522**

D. rs919433 (risk allele A) at locus 2q33



			het_P:	0.03			
Cohort	**info**	**p_value**	**f_ca(n)**	**f_co(n)**	**OR**	**ln(OR)**	**STDerr**
AAA	0.98	0.0012	0.376(818)	0.330(2991)	1.21	0.192	0.0593
IA (Finland)	0.97	0.00306	0.481(799)	0.430(2317)	1.20	0.179	0.0605
IA (Netherlands)	0.99	1.18e-05	0.419(717)	0.353(1987)	1.32	0.281	0.0641
TAA	0.96	0.75	0.362(760)	0.360(2212)	1.02	0.0203	0.0638
Mega-analysis	**0.98**	**4.55e-08**	**0.410(3094)**	**0.416(9507)**	**1.18**	**0.169**	**0.0309**

These figures show forest plots for each of the four genome-wide significant SNPs from the mega-analysis of IA, AAA and TAA GWAS cohorts. AAA indicates abdominal aortic aneurysm; IA, intracranial aneurysm; TAA, thoracic aortic aneurysm; info, imputation accuracy score; p_value, p-value for association of risk allele in aneurysm cohort; f_ca(n), risk allele frequency in cases (number of cases); f_co(n), risk allele frequency in controls (number of controls); OR, odds ratio; ln(OR), natural log of odds ratio; STDerr, standard error; het_P, p-value for heterogeneity.

Figure 6 is available with the online version of this paper.

Tables

Table 1. Overview of study populations used for the specific look-up of known IA, AAA and TAA risk SNPs

Cohort	Cases (n)	Controls (n)
IA		
Netherlands	717	1,987
Finland	799	2,317
Japan 1	288	194
Japan 2	1,383	5,484
USA+	2,617	2,548
@neurIST	717	3,296
Total	*6,548*	*16,843*
AAA		
Netherlands	812	2,998
Iceland	430	27,712
USA	724	1,604
Aneurysm Consortium	1,846	5,605
New Zealand 1	608	612
New Zealand 2	400	384
Total	*4,391*	*37,904*
TAA		
USA	760	2,212

IA indicates intracranial aneurysm; AAA, abdominal aortic aneurysm; TAA, thoracic aortic aneurysm.

Table 2. Baseline characteristics after quality control and imputation of the study populations used for the mega-analysis and polygenic analysis of IA, AAA and TAA GWAS data

	Study cohort			
Characteristics	IA (Netherlands)	IA (Finland)	AAA	TAA
Cases (n)	717	799	818	760
Women (%)	64.3	57.8	10.5	34.3
Controls (n)	1988	2317	3004	2212
Women (%)	56.2	57.2	37.7	53.0
SNPs (n)	10,683,725	10,524,028	10,684,772	10,750,239
Genomic inflation factor	1.10	1.06	1.04	1.05

IA indicates intracranial aneurysm; AAA, abdominal aortic aneurysm; TAA, thoracic aortic aneurysm.

Tables 3-8 are available with the online version of this paper.

Table 9. Heritability analysis results from LDSC (linkage disequilibrium score regression).

Cohort	h^2	SE (h^2)
AAA	0.160	0.179
IA (Finland)	0.314	0.258
IA (The Netherlands)	0.341	0.231
TAA	0.396	0.280

h^2 indicates proportion of phenotypic variance explained by genetic effects; SE, standard error; IA, intracranial aneurysm; AAA, abdominal aortic aneurysm; TAA, thoracic aortic aneurysm.

Table 10. Genetic correlation analysis results from LDSC (linkage disequilibrium score regression). The test evaluates the genetic correlation between diseases based on their common SNPs set corresponding to the structured LD from the reference panel.

Cohorts	Correlation (ρ)	Standard error (ρ)	p-value
AAA – IA (Finland)	-0.032	0.557	0.955
AAA – IA (Netherlands)	-0.384	0.658	0.560
AAA - TAA	0.243	0.777	0.754
IA (Finland) - TAA	0.685	0.509	0.179
IA (Netherlands) - TAA	0.847	0.740	0.253
IA (Finland) - IA (Netherlands)	1.591	0.928	0.086

IA indicates intracranial aneurysm; AAA, abdominal aortic aneurysm; TAA, thoracic aortic aneurysm; ρ, genetic correlation.

Table 11. Specific look-up of previously published aneurysm risk SNPs in six case-control datasets of IA, six case-control datasets of AAA and one case-control dataset of TAA. Look-up results in different cohorts were combined per subtype, using an inverse-variance fixed-effects meta-analysis.

				AAA				TAA			
SNP - IA	*Gene*	*Chr*	*RA*	*p*	*OR*	*Dir**	p_{CQ}	*p*	*OR*	*Dir**	p_{CQ}
rs6841581[13]	*EDNRA*	4	G	0.62	0.98	+----+	0.45	0.43	1.07	NA	NA
rs10958409[12 †]	*SOX17*	8	A	0.81	0.99	--++--	0.14	0.37	1.07	NA	NA
rs9298506[12 †]	*SOX17*	8	A	0.81	0.99	-+-+--	0.09	0.67	0.97	NA	NA
rs1333040[12]	*CDKN2BAS*	9	T	$1.5 \cdot 10^{-8}$	1.15	+++++-	0.07	0.52	1.04	NA	NA
rs12413409[11]	*CNNM2*	10	G	0.69	1.02	+---++	0.70	0.95	1.01	NA	NA
rs9315204[11]	*STARD13-KL*	13	T	0.73	1.01	--++++	0.21	0.79	1.02	NA	NA

rs11661542[11]	*RBBP8*	18	C	$4.1 \cdot 10^{-5}$	1.11	++++-+	0.11	0.08	1.11	NA	NA

				IA				TAA			
SNP - AAA	*Gene*	*Chr*	*RA*	*p*	*OR*	*Dir**	p_{CQ}	*p*	*OR*	*Dir**	p_{CQ}
rs10757278[12]	*CDKN2BAS*	9	G	$5.9 \cdot 10^{-17}$	1.30	+++…	0.81	0.42	1.05	NA	NA
rs1466535[98]	*LRP1*	12	C	0.95	1.00	--++-+	0.66	0.07	1.12	NA	NA
rs7025486[99]	*DAB2IP*	9	A	0.25	1.03	+-++-+	0.36	0.70	0.97	NA	NA
rs599839[101]	*SORT1*	1	G	0.41	0.97	-+-…	0.76	0.42	0.94	NA	NA
rs6511720[100]	*LDLR*	19	G	0.15	1.07	+-+-..	0.09	0.05	1.21	NA	NA

				IA				AAA			
SNP - TAA		*Chr*	*RA*	*p*	*OR*	*Dir**	p_{CQ}	*p*	*OR*	*Dir**	p_{CQ}
rs10519177[102] ‡	*FBN1*	15	G	0.99	1.00	++--++	0.25	0.016	1.01	-+++-+	0.32
rs2118181[102] ‡	*FBN1*	15	G	0.20	0.96	++---+	0.08	$1.1 \cdot 10^{-3}$	1.07	+-++++	0.45

SNP indicates single nucleotide polymorphism; IA, intracranial aneurysm; AAA, abdominal aortic aneurysm; TAA, thoracic aortic aneurysm; Chr, chromosome; RA, risk allele; OR, odds ratio; Dir, direction of effect; p, p-value; p_{CQ}, p-value from Cochran's Q test.

** Direction of effect per cohort. For AAA, directions correspond to the cohorts 'Iceland', 'Netherlands', 'USA', 'Aneurysm Consortium', 'New Zealand 1', 'New Zealand 2' respectively. For IA, directions correspond to the cohorts 'Netherlands', 'Finland', 'USA',' @neurIST', 'Japan 1', 'Japan 2' respectively. '+' indicates same direction as previously published aneurysm risk SNP, '-' indicates opposite direction, '.' indicates absence of the SNP in target cohort.*

† These two SNPs are independent (r^2=0.06)

‡ These two SNPs are independent (r^2=0.24)

Chapter 6

The impact of inherited genetic variants associated with lipid profile, hypertension, and coronary artery disease on the risk of intracranial and abdominal aortic aneurysms.

FNG van 't Hof, YM Ruigrok, AF Baas, LA Kiemeney, SH Vermeulen, AG Uitterlinden, A Hofman, F Rivadeneira, GJE Rinkel, PIW de Bakker.

Circulation: Cardiovascular Genetics. 2013 Jun;6(3):264-70.

Abstract

Background

Epidemiological studies show that an unfavourable lipid profile and coronary artery disease (CAD) are risk traits for abdominal aortic aneurysms (AAA) but not for intracranial aneurysms (IA), and that hypertension is a main risk trait for IA but not for AAA. To evaluate these observations, we investigated single nucleotide polymorphisms (SNPs) associated with serum lipid levels, hypertension, and CAD, and tested their contribution to AAA and IA risk.

Methods and results

We defined sets of SNPs previously reported to be associated with serum lipid levels, CAD and blood pressure (BP). From previously collected genome-wide data, we extracted genotypes for these SNP sets in 709 IA cases and 2692 controls, and 807 AAA cases and 1905 controls (all of Dutch origin). We computed genetic scores for each individual by summing the observed number of risk alleles weighted by their previously published effect size. Using logistic regression we tested the genetic scores for association to IA and AAA, and found significant associations for genetic scores of total cholesterol ($p=3.6x10^{-6}$), low-density lipoprotein cholesterol ($p=5.7x10^{-7}$) and CAD ($p=0.0014$) with AAA, and for the BP score with IA ($p=0.0030$).

Conclusion

We demonstrate that genetic risk profiles of lipid factors and CAD are associated with AAA but not with IA, and the genetic risk profile of BP is associated with IA but not with AAA. These findings are consistent with epidemiological observations.

Introduction

Intracranial aneurysms (IA) and abdominal aortic aneurysms (AAA) share a pathophysiological background (including upregulation of proteolytic pathways, inflammation, and loss of arterial wall matrix)[120] as well as epidemiological risk factors, such as age and smoking.[30, 124] For both diseases, family history is a main risk factor, indicating that inherited genetic variation influences disease risk.[30, 124] The disease risk for first-degree relatives of patients is up to eight times increased in AAA,[125, 126] and up to seven times in IA.[5] Co-occurrence of the two diseases has been described, mainly within affected families, suggesting a shared genetic basis for these diseases.[21]

Recent genome-wide association studies (GWAS) have highlighted a role for a risk allele at locus 9p21.3 near *CDKN2A* and *CDKN2B* in both IA and AAA.[10-12] In addition, risk alleles five risk loci near the genes *STARD13-KL, RBBP8, SOX17, CNNM2* and *EDNRA* are associated with IA,[10, 11, 13] while risk alleles at two loci near *DAB2IP* and *LRP1* are associated with AAA.[98, 99]

Epidemiological studies showed that high total cholesterol (TC), low-density lipoprotein cholesterol (LDL-C), triglyceride (TG), low high-density lipoprotein cholesterol (HDL-C) levels and coronary artery disease (CAD) are risk factors for AAA.[127-132] In contrast, hypercholesterolemia is not associated with increased risk on IA,[30, 133] and no significant comorbidity of CAD and IA has been observed, according to one study.[20] Similarly, hypertension is a major risk factor for IA,[30] but is only weakly associated with AAA.[131]

In this study, we test the effect of genetic variants associated with lipids, CAD and blood pressure (BP) on IA and AAA disease risk. This genetic design has some advantages compared to epidemiological studies, which may be more susceptible to confounding and reverse causation.[134] We hypothesize that the genetic risk profiles of lipid factors and CAD are associated with AAA but not with IA, and the genetic risk profile of BP is associated with IA but not with AAA. In the future, this may help to identify additional genes or pathways in aneurysm etiology.

Methods

Study populations

We used data from Dutch subjects genotyped in previous GWAS.[10-12, 99, 103, 104, 135] All studies were approved by the relevant medical ethical committees and all participants provided written informed consent. All study populations were previously described in detail.[10-12, 99, 103, 104, 135] Here we give a brief description of each study population.

IA patients (n=786) were admitted to the Utrecht University Medical Center (the Netherlands) between 1997 and 2007. The population consisted of 247 men and 539 women and included both patients with ruptured (727) and patients with only unruptured (59) IA. Ruptured IA were defined by symptoms suggestive of subarachnoid hemorrhage (SAH) combined with subarachnoid blood on a computed tomography (CT) scan and a proven aneurysm at angiography (conventional angiogram, CT- or magnetic resonance (MR)-angiogram). Unruptured IA were identified by CT or MR angiography or conventional angiography in the absence of clinical or radiological signs of SAH.[10-12] As controls, we included 3110 Dutch subjects, who were recruited as part of the Nijmegen Biomedical Study (n=1832) and the Nijmegen Bladder Cancer Study (n=1278).[103, 104] All case and control subjects were genotyped on Illumina CNV370 Duo BeadChips.

AAA patients (n=859) were recruited from eight medical centers in the Netherlands, mainly when individuals visited their vascular surgeon in the outpatient clinic or, in some cases, during hospital admission for elective or emergency AAA surgery. An AAA was defined as an infrarenal aorta diameter of ≥ 30 mm. The cohort consisted of 772 men and 87 women. The mean AAA diameter was 58.4 mm. Of these patients 530 had received surgery, of which 43 after rupture. Genotyping was performed on Illumina HumanHap610 chips.[99] Control subjects (n=2089) were ascertained via the Rotterdam Study, a population based cohort of subjects aged 45 years and older recruited from a district in Rotterdam (The Netherlands). These subjects were genotyped on Illumina HumanHap550 chips.[135]

Quality control (QC)

For the IA and AAA GWAS, PLINK version 1.07[108] was used for quality control of both subjects and single nucleotide polymorphisms (SNPs). After removal of SNPs with A/T or C/G alleles and SNPs that were not called in any individual, we performed sample QC and SNP QC as described below.

We performed sample QC after merging cases and controls, using a subset of common, high-quality SNPs (as defined by SNPs without deviation from Hardy-Weinberg equilibrium (HWE) ($p > 0.001$), with high minor allele frequency (> 20%) and with low missingness (<1%)), and performed pruning based on linkage disequilibrium ($r^2 > 0.5$). Subjects were removed based on the following three criteria: genotype missingness (subjects with a call rate below 95% were removed), heterozygosity (subjects were excluded if the inbreeding coefficient deviated more than 3 standard deviations from the mean) and cryptic relatedness (by calculating identity-by-descent (IBD) for each pair of individuals). In each pair with an IBD proportion of at least 5%, a subject was excluded, if it exhibited distant relatedness with multiple individuals. For case-control pairs, we removed the control subject. In the remaining pairs, the subject with the lowest call rate was excluded.

We performed principal component analysis (PCA) using EIGENSTRAT[79] on the study subjects and HapMap-CEU subjects.[136] We excluded SNPs from three regions with known long-distance linkage disequilibrium (LD): the major histocompatibility (MHC) region (chr6: 25.8-36 Mbp), the chromosome 8 inversion (chr8: 6-16 Mbp) and a chromosome 17 region (chr17: 40-45 Mbp). We created multi-dimensional scaling plots with the first 4 principal components (PCs), using R version 2.11.[46] Based on visual inspection of these plots, we excluded subjects that appeared to be outliers with respect to the CEU or the study population. After outlier removal, we recomputed principal components to include as covariates for logistic regression.

After sample QC, we excluded SNPs with more than 5% missing genotypes, a minor allele frequency < 1%, genotype missingness higher than the minor allele frequency and HWE deviation ($p < 1x10^{-6}$). We performed these QC steps in each study cohort separately and again after merging cases and controls. We also removed SNPs with a differential degree of missingness between cases and controls ($p < 1x10^{-5}$; chi-squared test).

Imputation

For imputation of ungenotyped SNPs, we used BEAGLE software version 3.0.4,137 with the HapMap phase II CEU population136 as the reference panel. Genotype probability scores were converted to allele dosages, ranging from 0 to 2. After imputation, we removed SNPs with imputation quality scores (i.e. ratio of the observed variance of the allele dosage to the expected variance) below 0.1 or above 1.1.

Association analysis

We performed logistic regression analyses of each SNP versus disease state in both the IA and AAA study, using PLINK. From the 10 PCs calculated in each study cohort, we included the PCs associated with case/control status ($p<0.05$; logistic regression) as covariates in this analysis.

Genetic scores

We investigated the genetic overlap between IA and AAA and aneurysm risk traits (plasma lipid levels, blood pressure and CAD) by creating genetic scores for each trait in our study subjects, based on established risk SNPs for these traits. For plasma lipid levels, we listed 52 SNPs known to be associated with TC, 37 SNPs with LDL-C, 47 SNPs with HDL-C and 31 with TG.[138] A total number of 26 SNPs are associated with systolic blood pressure (SBP), 26 with diastolic blood pressure (DBP),[139, 140] 22 with mean arterial pressure (MAP) and 10 with pulse pressure (PP).[141] Another 31 risk SNPs are associated with CAD.[142-144] As a negative control, we also used 180 SNPs associated with height.[145] For each trait, we listed the associated SNPs with their corresponding risk alleles and published effect sizes. When the association was reported in multiple studies, we took the effect size from the largest study (i.e. CARDIoGRAM consortium[142] for previously identified CAD SNPs and International Consortium for Blood Pressure Genome-Wide Association Studies[139] for previously identified BP SNPs). We only included SNPs that were independent (in low LD; $r^2<0.1$). We used SNAP[146] to calculate LD between SNPs (based on genotype data from HapMap release 22) and to search for proxy SNPs in case risk SNPs were absent from our data.

For CAD, a total number of 30 published risk SNPs were independent. From these, two SNPs were not genotyped or imputed in our data, because they were absent from HapMap. For one SNP (rs17465637 near *MIA3*[142]) we used a perfect proxy (rs17011681, $r^2=1$ in 1000 Genomes pilot 1). The other SNP (rs3798220 near *LPA*[142]) was excluded because no proxies were available. This resulted in a total number of 29 SNPs used for genetic score calculation.

To improve power for the BP scores, we constructed a composite BP-score incorporating all SBP, DBP, MAP and PP SNPs into a single score, by including only independent SNPs with concordant effect directions across these four traits. This resulted in a genetic score of 35 SNPs, after excluding three PP associated SNPs with a discordant effect on DBP.[141]

For this genetic score calculation we ascribed each SNP an equal weight, because the effect sizes differed across these four BP traits.

For all SNPs associated with lipid factors, CAD and BP, we listed the p-values, OR and direction of effect in our IA and AAA GWAS. In accordance with our hypothesis, we expect that the direction of effect of these SNPs is the same between each trait and IA / AAA (e.g. BP increasing alleles increase AAA risk), with exception of HDL-C increasing SNPs, which are expected to decrease AAA risk.

Using these SNPs, we calculated genetic scores for each trait in each individual of the IA and AAA cohort as follows:

$$\textit{Genetic score} = \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n,$$

where x_i is the estimated allele dosage (between 0 and 2) in a given individual, and β_i is the reported effect size (lipids, BP and height) or the natural log of the reported OR (CAD) for the i^{th} SNP.

We tested the resulting genetic scores for association with IA and AAA using logistic regression. Sex and PCs significantly associated with disease in the study cohorts were added as covariates in the analysis. We also adjusted for the 7 known IA risk SNPs (rs6841581 at locus 4q31, rs1333040 at 9p21, rs10958409 and rs9298506 at 8q11-12 (r^2=0.058), rs12413409 at 10q24, rs9315204 at 13q13 and rs11661542 at 18q11)[11, 13] and for the 3 AAA risk SNPs (rs1466535 at 12q13, rs10757278 at 9p21 and rs7025486 at 9q33).[12, 13]

To evaluate the effect of each genetic score, we divided the IA and AAA cohorts in quartiles based on the genetic score derived from each subject, and calculated odds ratios (OR) and corresponding 95% confidence intervals (CI) for IA/AAA disease risk in the highest quartile compared to that in the lowest quartile, using logistic regression. We adjusted for significantly associated PCs, known IA and AAA risk SNPs and sex.

Results

IA and AAA cohorts

The IA cohort consisted of 709 cases and 2692 controls, and the AAA cohort contained 807 cases and 1905 controls. After imputation, both cohorts comprised 2,393,271 autosomal SNPs. Table 1 shows the baseline characteristics of the IA and AAA cohorts after quality control.

Genetic scores

For all SNPs associated with lipid factors, CAD and BP, our IA and AAA GWAS results are listed in Supplemental Tables 1-6. Using these lists of SNPs, we constructed genetic scores for TC, LDL-C, HDL-C, TG, CAD, SBP, DBP, MAP, PP, composite BP and height, and tested these for association to IA and AAA. As an example, the distribution of LDL-C-scores in the AAA cohort is shown in the Figure. The distribution of all genetic scores in the IA and AAA cohort are shown in the Supplemental Figure. We estimated the effect size of these genetic scores by comparing the disease risk in the highest genetic score quartile with that in the lowest quartile (Table 2).

For IA, we found associations for the genetic score of SBP (p=0.031), DBP (p=0.020), MAP (p=0.0065) and the composite BP-score (p=0.0030). For the other genetic scores (TC, LDL-C, HDL-C, TG and CAD), none reached nominal significance. For the significantly associated genetic scores, the OR for IA risk in the highest versus the lowest genetic score quartile was 1.09 for the SBP-score (95% CI 1.00 - 1.19), 1.10 for the DBP-score (95% CI 1.01 – 1.20), 1.10 for the MAP-score (95% CI 1.01 – 1.19) and 1.11 for the composite BP-score (95% CI 1.02 - 1.21).

In AAA, we found significant associations for genetic scores of TC ($p=3.6x10^{-6}$), LDL-C ($p=5.7x10^{-7}$), and CAD ($p=1.4x10^{-3}$) and a weaker association with the HDL-C-score (p=0.020). We did not observe an association with the TG- and the BP-scores. The OR for AAA risk in the highest versus the lowest genetic score quartile was 1.24 for the TC-score (95% CI 1.13 – 1.35), 1.21 for the LDL-C-score (95% CI 1.10 – 1.32), 0.94 for the HDL-C-score (95% CI 0.86 – 1.03) and 1.18 for the CAD-score (95% CI 1.06 – 1.32).

To test the impact of winner's curse on the published effect estimates of the SNP associations, we repeated the analyses by re-calculating the genetic risk scores without weighting each risk allele by the published effect size, computing effectively the total number of risk alleles in each individual. The results were essentially unchanged (Supplemental Table 7).

We explored the individual contributions of SNPs making up the scores that were significantly associated with IA or AAA. We observed that one SNP (rs6511720) associated with TC and LDL-C is also strongly associated with AAA ($p=3.9x10^{-5}$). This SNP is located at the *LDLR* gene, which is known to be associated with CAD[142] and with familial hypercholesterolemia.[138] For

the other SNPs in the genetic scores, the individual associations with IA or AAA were much weaker (Supplemental Tables 1-6).

We did not find evidence for a relation between the height-score with either IA (p=0.67) or AAA (p=0.20).

Discussion

In this study, we used a genetic approach to study the role of epidemiological risk traits in IA and AAA. We found that genetic scores based on validated SNPs associated with CAD and lipid factors influence disease risk of AAA but not of IA. In contrast, genetic scores based on BP associated SNPs are associated with IA but not with AAA.

The absence of association between genetic scores based on lipid risk alleles and IA is consistent with previous epidemiological studies reporting that hypercholesterolemia is not associated with higher risk on IA, but possibly a risk-reducing factor.[30] In contrast, we did find significant associations of TC- and LDL-C-scores to AAA. Previous epidemiological studies considering lipid factors in AAA report conflicting results, and suffer from several limitations. For example, most studies did not adjust for lipid-lowering medication as a potential confounder,[124, 127, 130] with exception of some large studies that found an association for TC [127] and HDL-C[127, 130] but not for TG[127, 130] and disagreed about an association for LDL-C.[128, 130] Overall, our results are largely consistent with large epidemiological studies which adjusted for lipid-lowering medication.

To our knowledge, a shared genetic background of lipid factors and AAA has not been described before. A recent GWAS reported an association between a common variant in *LRP1* (rs1466535) and AAA.[98] Another variant mapping to the same gene (rs11613352) was found to be associated with HDL and TG.[138] However, close examination of the *LRP1* locus reveals that these two SNPs are not in LD (r^2=0.04). Furthermore, the AAA risk SNP at *LRP1* was not associated with lipid levels in a previous meta-analysis.[98] In this study, we found that one SNP (rs6511720) driving the association between the TC and LDL-C score and AAA maps to *LDLR*. This gene is of particular interest, because mutations in *LDLR* are a well-known cause of familial hypercholesterolemia.[147] Taken together, our findings suggest that a part of the (yet unknown) genetic risk variants for AAA may act through changes in lipid factors.

The association of the genetic CAD-score with AAA (and lack of association with IA) may suggest a shared etiology between CAD and AAA. Epidemiologically, the risk for CAD is increased in the years after discovery of AAA and after rupture of IA.[148, 149] However, a history of myocardial infarction is a risk trait for AAA,[131] and CAD is frequently present at the time of diagnosis of AAA.[131, 132] No significant comorbidity between CAD and IA has been reported, although to our knowledge there has been only one study that investigated this relation.[133] This study suggests a stronger association between CAD and AAA than between CAD and IA, which is consistent with our study findings.

Although the 9p21 locus is shared between CAD, IA and AAA,[12] we did not find evidence for additional genetic sharing between CAD and IA. Considering the pleiotropic effects of 9p21 across a wide range of human conditions (including ones not associated with vascular diseases), sharing of this locus may not necessarily imply an overlap in disease etiology.

We observed a strong association of the composite BP-score with IA, which was also supported by associations of the individual SBP, DBP and MAP scores. The BP-scores were not associated with AAA, but the effect directions of these scores were similar for BP increase and AAA disease risk. The difference in effect of the BP-scores on IA and AAA disease risk is in line with epidemiological findings: hypertension is one of the most prominent risk factors of IA (OR=2.6, 95% CI 2.0 - 3.1)[30] while for AAA a more modest effect on disease risk has been reported (OR=1.3, 95% CI 1.14 -1.55).[131] Of note, it has been suggested that hypertension is a risk factor for AAA rupture, but not for development of AAA.[131] This can explain our study results, because our AAA study population largely exists of patients with unruptured aneurysms.

Our study findings are consistent with previous genetic studies on IA and BP. One recent study discovered that a suggestive IA risk SNP at locus 5q23.2 in *PRDM6* is also associated with high SBP.[150] In addition, previous GWAS independently discovered that locus 10q24 is associated with IA,[11] and that the same SNP at this locus is also associated with SBP and DBP.[139, 140] From the genes in this region, *CYP17A1* is a functional candidate gene for BP regulation, as it is involved in biosynthesis of mineralocorticoids and glucocorticoids.[140] In contrast to IA, no sharing of risk alleles has been reported for BP and AAA.

In this study, we investigated the effect of SNPs associated with lipid factors, CAD and BP on disease risk of IA and AAA. The genetic approach

of this study has advantages compared to traditional epidemiological studies, which may suffer from potential confounding. In addition, reverse causation, where the disease influences the risk factor under study, may lead to false interpretations of causality. These limitations are less likely to play a role in genetic studies, because alleles are randomly transmitted from both parents to offspring, are assumed to remain stable over a lifetime, and are not changed by disease or by trait-modifying factors like medication.[134]

One limitation of our study is the modest sample size of our IA and AAA case-control collection, which puts a limit to what we can detect reliably. To illustrate this, we calculated the power for detecting a nominally significant association ($p<0.05$) in the IA analysis for a risk allele with a frequency of 10% using the genetic power calculator.[86] The resulting power is 18% at a relative risk of 1.1, and 49% at a relative risk of 1.2 per allele, assuming additive effects. Power of the AAA analysis is virtually identical (because the sample sizes are very similar). Under a model that multiple risk alleles have a cumulative effect on aneurysm risk, the power of detecting an association between the composite score and outcome is much increased. The effective power increase will ultimately be dependent on the number of truly associated variants relative to the number of "null" variants (that may introduce noise), and on the validity of the assumption of additivity and independence of the different variants. Needless to say, lack of an association with one of our genetic risk scores tested does not necessarily imply that none of the SNPs are truly associated. Lastly, we note that the sample sizes of the IA and AAA cohorts are comparable, so the tests performed should represent a fair comparison between IA and AAA. That is to say, if we observe a significant effect in the IA case-control analysis, we should expect to have sufficient power for a *similar* effect to be detected in the AAA case-control analysis. The significant lipid-AAA association ($p=3.6 \times 10\text{-}06$ for TC and $p=5.7 \times 10\text{-}07$ for LDL-C) and the insignificant lipid-IA association ($p=0.2$ and $p=0.6$, respectively) seem therefore surprising and unexpected under the null. The association between the blood pressure score and IA ($p=0.003$) and the insignificant association for AAA ($p=0.2$) is perhaps less striking, but the overall consistency between the (known) epidemiological relations and the observed genetic associations (namely, that hypertension is related to IA, and that lipids and CAD are related to AAA) is compelling.

We did not observe a significant association of the genetic score of human height, a highly polygenic trait, with IA and AAA. This reduces the likelihood that our results are due to population stratification, which is a possible

confounder in genetic analyses.[79] We also confirmed that the observed effects were not due to inflated effect sizes (winner's curse) in the initial association studies.

In conclusion, this study demonstrates that the genetic scores for TC, LDL-C, HDL-C and CAD are associated with the risk of AAA but not with the risk of IA, and genetic BP-scores increase risk of IA but not of AAA. With this genetic approach, we confirm relationships observed in epidemiological studies. This study illustrates how genetic studies can help to elucidate the role of risk traits in aneurysm pathology. This study does not allow us to confirm whether the observed effects are actually mediated through these risk traits. Future research can include Mendelian randomization studies[151] to determine whether these risk traits play a causal role in the disease mechanism of aneurysms.

Tables

Table 1. Overview of the baseline characteristics of the IA and AAA study cohorts (numbers after quality control)

	IA cohort	AAA cohort
Number of cases	709	807
Men (%)	31.2	89.7
Mean age (yrs)	54.4	68.3
Ruptured aneurysms (%)	92.5	5.5
Number of controls	2692	1905
Men (%)	62.8	43.9
Mean age (yrs)	61.7	56.0
Number of genotyped SNPs	311,037	539,589

IA indicates intracranial aneurysm; AAA, abdominal aortic aneurysm; SNP, single nucleotide polymorphism; yrs, years.

Table 2. Genetic scores based on lipid factor, CAD, BP and height-associated SNPs tested for association to IA and AAA disease risk

		IA			AAA		
		G-score (cont)*	G-score (quart)†		G-score (cont)*	G-score (quart)†	
Trait	No. SNPs	*p*	OR	95% CI	*p*	OR	95% CI
Lipid factors							
TC	52	0.20	0.95	0.87 - 1.03	$3.6 \cdot 10^{-6}$	1.24	1.13 - 1.35
LDL-C	37	0.57	0.97	0.89 - 1.06	$5.7 \cdot 10^{-7}$	1.21	1.10 - 1.32
HDL-C	47	0.38	0.96	0.88 - 1.04	0.020	0.94	0.86 - 1.03
TG	31	0.66	0.99	0.91 - 1.07	0.066	1.12	1.03 - 1.23

CAD	29	0.75	0.99	0.89 - 1.09	0.0014	1.18	1.06 - 1.32
BP traits							
SBP	26	0.031	1.09	1.00 - 1.19	0.16	1.06	0.97 - 1.15
DBP	26	0.020	1.10	1.01 - 1.20	0.19	1.05	0.96 - 1.14
MAP	22	0.0065	1.10	1.01 - 1.19	0.12	1.09	1.00 - 1.19
PP	10	0.59	1.00	0.92 - 1.09	0.10	1.07	0.98 - 1.16
Comp BP	33	0.0030	1.11	1.02 - 1.21	0.16	1.04	0.95 - 1.13
Height	180	0.67	1.03	0.94 - 1.12	0.20	1.05	0.97 - 1.15

IA indicates intracranial aneurysm; AAA, abdominal aortic aneurysm; OR, odds ratio; CI, confidence interval; TC, total cholesterol; LDL-C, low-density lipoprotein cholesterol; HDL-C, high-density lipoprotein cholesterol; TG, triglycerides; CAD, coronary artery disease; BP, blood pressure; SBP, systolic blood pressure; DBP, diastolic blood pressure; MAP, mean arterial pressure; PP, pulse pressure; p, p-value; G-score, genetic score; cont, continuous; quart, quartiles; Comp BP, Composite BP

**Association between genetic score and disease risk, calculated with logistic regression adjusted for sex, risk SNPs and significantly associated principal components.*
†Disease risk in the highest versus the lowest genetic score quartile, calculated with logistic regression adjusted for sex, risk SNPs and significantly associated principal components

Figures

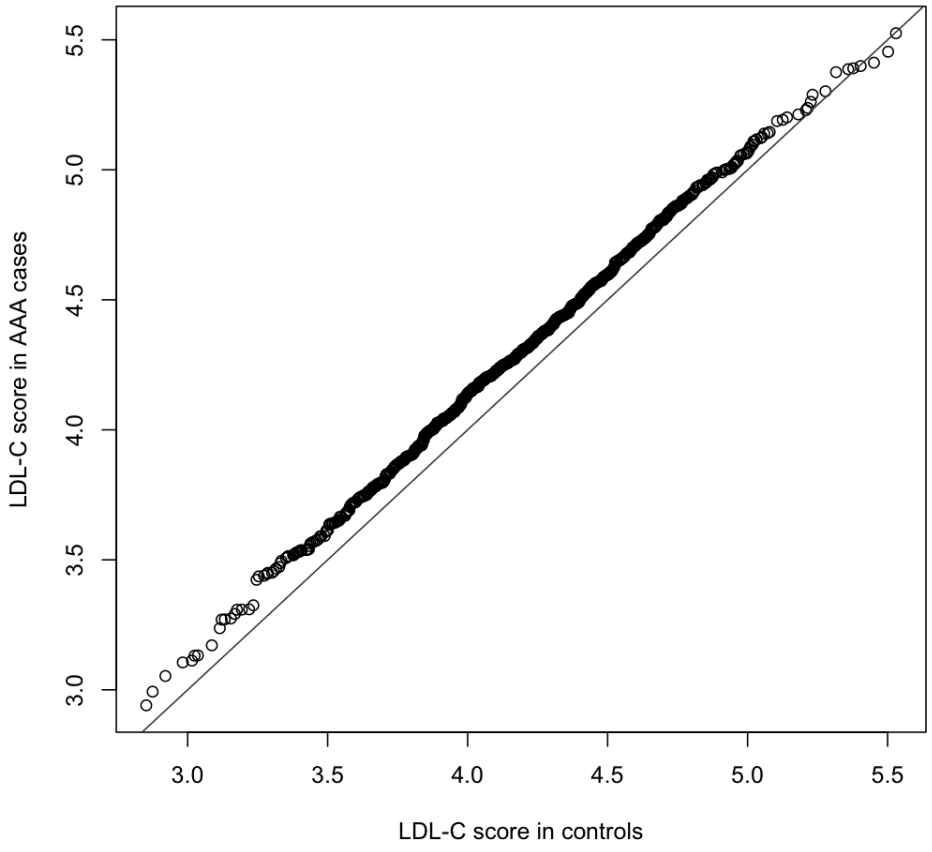


Figure: Quantile-quantile plot of genetic scores of LDL-C in the AAA cohort.
This figure shows the distribution of genetic scores based on LDL-C associated SNPs in AAA cases versus controls. We calculated this score for each AAA case and control subject individually, and tested the genetic score for association with AAA using logistic regression, adjusting for associated PCs, known AAA risk SNPs and sex ($p=5.7 \times 10^{-7}$).

Supplemental information
Supplemental Tables 1-6 and Supplemental Figure are available with the online version of this paper

Chapter 7

Genetic variants associated with type 2 diabetes and adiposity and risk of intracranial and abdominal aortic aneurysms

FNG van 't Hof, J Vaucher, MV Holmes, A de Wilde, AF Baas, JD Blankensteijn, A Hofman, LA Kiemeney, F Rivadeneira, AG Uitterlinden, SH Vermeulen, GJE Rinkel, PIW de Bakker, YM Ruigrok.

European Journal of Human Genetics 2017 Apr 5. (epub ahead of print)

Abstract

Epidemiological studies show that type 2 diabetes (T2D) is inversely associated with intracranial aneurysms (IA) and abdominal aortic aneurysms (AAA). Although adiposity has not been considered a risk factor for IA, there have been inconsistent reports relating adiposity to AAA risk. We assessed whether these observations have a genetic, causal basis. To this end, we extracted genotypes of validated single nucleotide polymorphisms (SNPs) associated with T2D (n=65), body mass index (BMI) (n=97) and waist-hip ratio adjusted for BMI (WHRadjBMI) (n=47) from genotype data collected in 717 IA cases and 1988 controls, and in 818 AAA cases and 3004 controls, all of Dutch descent. For each of these three traits, we computed genetic risk scores (GRS) for each individual in these case-control data sets by summing the number of risk alleles weighted by their published effect size, and tested whether these GRS were associated to risk of aneurysm. We divided the cohorts into GRS quartiles, and compared IA and AAA risk in the highest to the lowest GRS quartile using logistic regression. We found no significant difference in IA or AAA risk between top and bottom quartiles for the genetic risk scores for T2D, BMI, and WHRadjBMI. However, additional Mendelian randomization analyses suggested a trend to potentially causal associations between BMI and WHRadjBMI and risk of AAA. Overall, our results do not support epidemiological observations relating T2D to aneurysm risk, but may indicate a potential role of adiposity in AAA that requires further investigation.

Introduction

Thanks to the success of genome-wide association studies (GWAS), we are rapidly gaining novel insights into the genetic basis of traits and diseases as well as the overlap between them. An important advantage compared to observational studies, genetic analysis can rigorously test whether one trait (e.g. circulating levels of a biomarker) affects risk of disease in a *causal* manner. In a previous study, for example, we showed that genetic risk profiles for lipid factors and coronary artery disease (CAD) susceptibility were associated with abdominal aortic aneurysms (AAA) but not intracranial aneurysms (IA), whereas the genetic risk profile for blood pressure (BP) was instead associated with IA but not AAA,[152] consistent with epidemiological observations.[30, 127, 131] This illustrated that a genetic approach can help assess the causal role of traditional risk factors in disease.

Here, we apply the same methodology to study the effect of other metabolic risk factors which do not yet have an established role in aneurysm pathology. Epidemiological studies have reported an inverse association between type 2 diabetes (T2D) and both IA[30, 153, 154] and AAA,[154-156] but it remains unclear whether these relations are truly causal or may have arisen due to confounding. Adiposity is not considered a *bona fide* risk factor for IA,[30, 157] and for AAA the impact of adiposity is still unclear due to conflicting results from observational studies.[158, 159] (e.g. one study suggested that abdominal adiposity rather than total obesity may play a role in AAA).[159] In this study, we aim to test whether there is a causal association between T2D or adiposity and risk of IA and AAA, by testing the combined effect of established genetic risk variants for these traits on the risk of IA and AAA. For adiposity, we tested genetic risk variants for both BMI as a measure of total adiposity, and for waist-hip ratio adjusted for BMI (WHRadjBMI) as a measure of central adiposity. In addition, we perform Mendelian randomization analyses to test whether any of these risk factors have causal effect on aneurysm risk.

Methods

Study populations

We used genotype data from Dutch IA and AAA patients and controls. These cohorts have been described in detail elsewhere.[10, 12, 99, 103, 104] All studies have been approved by the relevant medical ethical committees and all participants provided written informed consent. Below is a brief description of each study population.

IA cases in the Dutch cohort (n=1,080) were admitted to the University Medical Center Utrecht, the Netherlands between 1997 and 2011. The cohort both included cases with ruptured and unruptured IA. Ruptured IA cases were defined by symptoms suggestive of subarachnoid hemorrhage (SAH) combined with subarachnoid blood on a computed tomography (CT) scan and a proven IA at angiography (conventional angiogram, CT- or magnetic resonance (MR)-angiogram). Unruptured IA cases were identified by CT or MR angiography or conventional angiography in the absence of clinical or radiological signs of SAH. Patients with fusiform IA, possible traumatic SAH, and polycystic kidney disease were excluded. All cases were genotyped on Illumina CNV370 Duo BeadChips.[10] Controls (n=2,089) were ascertained via the Rotterdam Study, a population-based cohort of subjects aged 45 years and older recruited from a district in Rotterdam (The Netherlands). These controls were genotyped on Illumina HumanHap550 BeadChips.[135]

AAA patients (n=859) were recruited from eight medical centers in the Netherlands, mainly when individuals visited their vascular surgeon in the outpatient clinic or, in some cases, during hospital admission for elective or emergency AAA surgery. An AAA was defined as an infrarenal aorta diameter of ≥ 30 mm. The mean AAA diameter was 58.4 mm. Of these patients 530 had received surgery, of which 43 after rupture. Genotyping was performed on Illumina HumanHap610 chips.[99] As controls, we included 3,110 Dutch subjects who were recruited as part of the Nijmegen Biomedical Study (n=1,832) and the Nijmegen Bladder Cancer Study (n=1,278).[103, 104] These controls were genotyped on Illumina CNV370 Duo BeadChips.

Quality control, imputation and association analysis

We performed quality control of samples and single nucleotide polymorphisms (SNPs) as described previously.[152] For each case-control dataset, genotype imputation was performed in Ricopili (Rapid Imputation Consortium Pipeline, https://sites.google.com/a/broadinstitute.org/ricopili). For imputation, the prephasing/imputation stepwise approach implemented in IMPUTE2 and SHAPEIT was used (chunk size of 3 Mb and default parameters).[109, 110] The imputation reference set consisted of 2,184 phased haplotypes from the full 1000 Genomes Project data set (February 2012; 40,318,253 variants).[160] All genomic locations are given in NCBI Build 37/UCSC hg19 coordinates. After imputation, SNPs with an imputation accuracy score < 0.6 or minor allele frequency (MAF) < 0.5% were excluded. Next, we performed an association analysis in both the IA and

AAA cohorts using PLINK.[108] We used logistic regression models where we included the allelic dosage (0-2) of each SNP as an independent variable, and case-control status as the dependent variable (outcome). The first four principal components (PCs) in each study cohort were included as covariates to correct for population stratification.

Genetic risk scores

From the literature, we found 65 established risk SNPs for T2D,[161] 97 risk SNPs for BMI,[162] and 49 SNPs for WHRadjBMI.[163] One BMI risk SNP (rs12016871:C>T) was not present in our association study, and was replaced by a proxy SNP (rs9581854:C>T; r^2=1). One WHRadjBMI risk SNP (rs7759742:A>T) was excluded from further analyses, because it was not present in our association study, and no proxy SNPs (r^2>0.8) were available. For each of these SNPs, we looked up the corresponding risk allele and its published effect size. We extracted the results of these SNPs from our IA and AAA association analysis, and computed genetic risk scores for T2D, BMI and WHRadjBMI in each individual of the IA and AAA cohorts as follows:

$$Genetic\ risk\ score = \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n$$

where x_i is the estimated allele dosage (between 0 and 2) in a given individual, and β_i is the published effect size (BMI and WHRadjBMI) or the natural log of the reported odds ratio (T2D) for the i^{th} SNP. We tested the resulting genetic risk scores for association with IA and AAA using logistic regression, adjusting for sex and the first four PCs in the study cohorts. To evaluate the magnitude of effect of the genetic risk scores on IA or AAA disease risk, we divided the IA and AAA cohorts into quartiles on the basis of the genetic risk score distributions in these two cohorts, and calculated odds ratios and corresponding 95% confidence intervals (CI) for IA or AAA disease risk in the highest quartile compared to that in the lowest quartile, using logistic regression, adjusting for sex and the first four PCs in the study cohorts.

Mendelian randomization analyses

To assess any causal association between T2D, adiposity (indexed by BMI and WHRadjBMI) on risk of aneurysms, we performed a series of complimentary Mendelian randomization (MR) analyses.[164] Conventional MR findings (based on an inverse variance weighted (IVW) approach) were contrasted with results from sensitivity analyses, including MR-Egger, MR-Egger adjusted for simulation extrapolation, and weighted median MR.[165] [166] [167] MR-Egger tests for the presence of unbalanced pleiotropic effects of

the SNPs under analysis that may distort the causal effect estimate. For all MR analyses, SNPs were oriented towards an increase in the exposure. The standard error was obtained by bootstrap resampling 10,000 times. IVW was performed using linear regression of the SNPs-exposure (T2D, BMI or WHRadjBMI) estimates against SNPs-outcome (IA or AAA) estimates, weighted by the minor allele frequencies of each SNP and forced to pass through the origin.[165] MR-Egger is similar to IVW but the constant was not constrained to pass through the origin.[165] A significant departure from the origin provides evidence for the presence of unbalanced pleiotropy (tested using the Egger test; i.e. a p-value significant at the 5% level indicates presence of pleiotropic effects). For MR-Egger adjusted for simulation extrapolation using I^2-statistics, we first quantified any violation of the NOME (NO Measurement Error [NOME]) assumption (i.e., testing the fact that the SNP-exposure association is true).[167] We then applied simulation extrapolation (SIMEX; implemented in R using the simex package) to adjust the MR-Egger causal estimates to account for a potential NOME violation (results based on 1,000 simulations).[167] Weighted median MR yields robust and precise results even when up to 50% of the weight in the analysis comes from invalid genetic variants.[166] We specifically performed a penalized weighted median MR analysis (implemented in Stata using the mrrobust package; available at: https://github.com/remlapmot/mrrobust), which favours SNPs with causal estimates close to the median causal estimate.

Data availability

The study data (summary-level data of primary study results) have been made publicly available. They can be obtained from GWAS Central: http://www.gwascentral.org/study/HGVST1837.

Results

IA and AAA cohorts

After quality control and imputation, the IA cohort consisted of 717 cases and 1988 controls, and the AAA cohort comprised 818 cases and 3004 controls. Table 1 shows the baseline characteristics of the IA and AAA cohorts.

Genetic risk scores

The association results of the individual T2D, BMI and WHRadjBMI associated SNPs in the IA and AAA cohorts are listed in Supplemental Tables 1, 2 and 3. None of the T2D, BMI and WHRadjBMI SNPs were significantly associated with IA or AAA (p>0.002).

The association results of the T2D, BMI and WHRadjBMI genetic risk scores with IA risk are shown in Table 2 and with AAA in Table 3. We did not detect any evidence for associations between the weighted GRS and IA/AAA case status (p>=0.06). There was no evidence for association in IA risk between top and bottom quartiles for the T2D-score (odds ratio (OR)=1.03, 95% confidence interval (CI)=0.95-1.11), for the BMI-score (OR=1.04, 95% CI=0.96-1.13), or for the WHRadjBMI-score (OR=0.96, 95% CI=0.88-1.04). Also for AAA risk, no evidence for association was found between top and bottom quartiles for the T2D-score (OR=1.03, 95% CI=0.96-1.11), for the BMI-score (OR=1.05, 95% CI=0.97-1.13), or for the WHRadjBMI-score (OR=1.04, 95% CI=0.96-1.13).

Mendelian randomization analyses

MR analyses showed consistent directions of effect of BMI on AAA and WHRadjBMI on AAA, but the 95% CI of all effect sizes included the value of 1 (Table 4). The MR analysis of BMI was suggestive of a causal relationship with risk of AAA on conventional (IVW) MR analysis (OR for 1-SD increase in BMI on risk of AAA=1.63, 95% CI 0.99-2.61). The causal effect estimates of the associations between T2D and aneurysms were close to the null. In the Egger tests we did not find evidence against the null hypothesis of no pleiotropy of the SNPs under analysis, except for WHRadjBMI on IA analysis. Supplemental Figures 1, 2 and 3 present the effect estimates from conventional and Egger Mendelian randomization as regression slopes, for each combination of a risk trait and IA or AAA. These slopes are all parallel, except for the analysis of WHRadjBMI on IA, indicating the presence of pleiotropic effects in this analysis.

Discussion

In this study, we found no evidence for a genetic overlap between genetic variants associated with T2D, BMI and WHRadjBMI and risk of IA and AAA. Nonetheless, Mendelian randomization analyses suggest a potentially causal effect of BMI and WHRadjBMI on AAA risk. The lack of a genetic association between T2D and IA or AAA, further supported by Mendelian randomization analyses, suggests that the observed inverse relation between T2D and aneurysms is likely confounded and not real. Several explanations for this inverse relation have been proposed. For example, hyperglycemia and hyperinsulinemia could lead to biological changes protecting the aortic wall from AAA development.[124, 156] Confounding factors have also been suggested, like competing morbidity from other (cardiovascular) diseases induced by T2D.[30, 168] Furthermore, effects of

medication used in the management of T2D could protect against aneurysm development.[124, 156]

The absence of a genetic association between adiposity and IA in this study is consistent with epidemiological studies, which have ruled BMI out as a risk factor for IA.[30, 157] The association between adiposity and AAA is not clear, however, and our data suggest that there may well be a causal relationship.

In our study we chose to index adiposity by BMI and WHRadjBMI. In particular, we favoured WHRadjBMI over WHR alone as the former represents a better surrogate of body fat distribution (i.e. central obesity)[169] and previous studies suggest that the risk of AAA is influenced by central obesity rather than total obesity.[159] Nevertheless, it is important to realize that the adjustment for the heritable BMI trait in WHRadjBMI may have created spurious associations, a phenomenon referred to as collider bias.[170] To investigate this further, future studies should investigate WHR and BMI separately using multivariable MR approaches.[171]

MR analysis has the advantage of reducing bias from confounding and reverse causation. Inherited alleles from parents remain stable throughout life and are not influenced by environmental factors. Whereas standard genetic association studies (using GRS) help identify common genetic determinants between traits and disease outcomes, MR more formally investigates this and synthesizes a causal estimate for a given difference in genetically-instrumented exposure. Thus, a causal effect discovered by MR analysis is likely to reflect a true association.[134] Here, we did not observe strong evidence for causal relationships between T2D, BMI or WHRadjBMI and risk of aneurysms, although our MR analyses consistently suggest a potential association between adiposity (particularly BMI) and risk of AAA. We emphasize that these results are hampered by limited statistical power and studies involving larger sample sizes would be required to provide a definitive answer. Even so, we were previously able to show significant associations between genetic risk scores for well-established risk factors (hypertension and serum lipid levels) and IA or AAA risk, respectively, using the same data.[152] This suggests that the power of the present study should have been sufficient under the assumption that the effect is real and has a comparable magnitude of effect.

In summary, we found no evidence for a causal role of T2D in aneurysms, and only suggestive evidence for a potentially causal role of adiposity traits

in AAA. Future studies with larger datasets would be needed to further investigate the nature of this relationship.

Tables

Table 1: Overview of the baseline characteristics of the IA and AAA study cohorts (numbers after quality control).

	IA cohort	AAA cohort
Number of cases	717	818
Women (%)	64.3	10.5
Mean age (yrs)	61	68
Ruptured aneurysms (%)	92.5	5.5
Number of controls	1988	3004
Women (%)	57.2	37.7
Mean age (yrs)	56	62

IA indicates intracranial aneurysm; AAA, abdominal aortic aneurysm; SNP, single nucleotide polymorphism; yrs, years.

Table 2. Genetic risk scores based on T2D, BMI and WHRadjBMI associated SNPs tested for association with IA disease risk.

		GRS (continuous *)	GRS (quartiles **)	
Trait	*Number of SNPs in GRS*	*p-value*	*Odds ratio*	*95% CI*
T2D	65	0.56	1.03	0.95-1.11
BMI	97	0.66	1.04	0.96-1.13
WHRadjBMI	48	0.12	0.96	0.88-1.04

GRS indicates genetic risk score; IA, intracranial aneurysm; SNPs, single nucleotide polymorphism; T2D, type 2 diabetes; BMI, body mass index; WHRadjBMI, waist-hip ratio adjusted for BMI; CI, confidence interval.
** Association between IA risk and genetic score as a continuous variable*
*** IA risk in the highest versus the lowest genetic score quartile*

Table 3. Genetic risk scores based on T2D, BMI and WHRadjBMI associated SNPs tested for association with AAA disease risk.

		GRS (continuous *)	GRS (quartiles **)	
Trait	*Number of SNPs in GRS*	*p-value*	*Odds ratio*	*95% CI*
T2D	65	0.52	1.03	0.96-1.11
BMI	97	0.15	1.05	0.97-1.13
WHRadjBMI	48	0.06	1.04	0.96-1.13

GRS indicates genetic risk score; AAA, abdominal aortic aneurysm; SNPs, single nucleotide polymorphism; T2D, type 2 diabetes; BMI, body mass index; WHRadjBMI, waist-hip ratio adjusted for BMI; CI, confidence interval.

** Association between AAA risk and genetic score as a continuous variable*
*** AAA risk in the highest versus the lowest genetic score quartile*

Table 4. Mendelian randomization (MR) analysis results for T2D, BMI and WHRadjBMI risk SNPs tested in IA and AAA.

Analysis	MR analysis	Effect estimate	95% CI		I^2	Egger test*
T2D vs IA	IVW	1.04	0.86	1.25	0.91	0.66
	MR-Egger	1.14	0.75	1.71		
	MR-Egger+SIMEX	1.17	0.69	1.96		
	weighted median MR	1.11	0.79	1.56		
T2D vs AAA	IVW	1.04	0.87	1.24	0.91	0.78
	MR-Egger	0.99	0.67	1.48		
	MR-Egger+SIMEX	0.97	0.62	1.54		
	weighted median MR	0.94	0.72	1.24		

BMI vs IA	IVW	1.15	0.67	1.97		
	MR-Egger	0.91	0.24	3.58	0.88	0.76
	MR-Egger+SIMEX	0.89	0.14	5.65		
	weighted median MR	1.17	0.44	3.13		
BMI vs AAA	IVW	1.63	0.99	2.61		
	MR-Egger	1.34	0.39	4.33	0.88	0.72
	MR-Egger+SIMEX	1.42	0.38	5.36		
	weighted median MR	1.72	0.73	4.03		
WHRadjBMI vs IA	IVW	0.49	0.23	1.09		
	MR-Egger	15.33	0.35	162.35	0.70	0.03
	MR-Egger+SIMEX	46.72	0.81	2684.53		
	weighted median MR	0.62	0.20	1.97		
WHRadjBMI vs AAA	IVW	1.84	0.92	3.57		
	MR-Egger	1.61	0.11	18.88	0.70	0.91
	MR-Egger+SIMEX	1.96	0.09	43.99		
	weighted median MR	2.10	0.80	5.47		

T2D indicates type 2 diabetes; BMI, body mass index; WHRadjBMI, waist-hip ratio adjusted for BMI; IA, intracranial aneurysms; AAA, abdominal aortic aneurysms; IVW, inverse variance weighted MR; SIMEX, simulation extrapolation; CI, confidence interval; I^2, expected relative bias of the MR-Egger causal estimate (i.e., dilution due to potential spurious measurements of the SNP-exposure association).

** p-value related to the constant derived from MR-Egger regression, indicating presence of pleiotropic effects.*

Supplemental information

Supplemental Tables 1-3 and Supplemental Figures 1-3 are available with the online version of this paper

Chapter 8

Summary and conclusions

This thesis, 'The genetics of intracranial aneurysms and related diseases,' describes the search for new genetic risk factors involved in intracranial aneurysms (IA), genetic differences between IA subtypes and similarities and differences with the genetic background of aortic aneurysms and traits related to aneurysms. Below, I summarize the main findings from this thesis and discuss the most important lessons that can be learned from these studies. Then, I describe in more detail the different challenges in studying the genetic background of IA, including suggestions for future directions.

Summary

1. Intracranial aneurysms are genetically heterogeneous (Chapter 2)

In genetic studies, IA are typically considered as one and the same trait. However, from a clinical perspective, it is clear that IA patients present with different phenotypic subtypes, the most important difference being the location of the aneurysm at the circle of Willis. Some patient characteristics, like sex, age and presence of familial IA predispose to IA at different locations,[26, 33-35] and the risk of rupture differs between IA locations.[35, 172]

These differences led to the hypothesis that the genetic risk underlying IA may also play a different role depending on the phenotypic subtype. Indeed, we observed an enrichment of risk alleles at seven IA risk loci, which were previously established in genome-wide association studies (GWAS), in patients with IA located at the middle cerebral artery (MCA). This observation was replicated in an independent population of Finnish IA patients. The difference in genetic load according to IA location implies that IA is not only clinically, but also genetically a heterogeneous disease. This may be important for future genetic studies.

2. No altered gene expression can be detected in blood of previous SAH patients (Chapter 3)

We hypothesized that gene expression differences would be measurable years after the time of rupture, because IA seems to be a chronic vascular disease rather than a single event; this hypothesis is based on the high recurrence rate of IA,[45] and on pathologic studies of the vessel wall.[173] However, we did not find convincing evidence for differences in gene expression between patients with a history of subarachnoid haemorrhage (SAH) from ruptured IA and controls. Only one group of genes with a case-control difference in co-expression was observed, but we could not ascribe biological meaning to either this group of genes, or individual genes within this group. Overall, the results may imply that disease-relevant gene expression differences relevant to IA formation do not occur in blood.

3. Exome chip analysis of IA suggests a possible role of the *FBLN2* gene in IA (Chapter 4)

In addition to common genetic variants that are detected by GWAS, genetic variants with lower frequencies in the general population (<5%) also play a role in complex diseases.[69] The exome chip captures a large number of these variants with lower frequencies that are located in coding regions (the exome) of the genome. Using this exome chip in our study, comprising 995 IA patients and 2,080 controls, we did not identify any variants that were individually associated with IA. This is likely due to lack of power, as we know that large samples are needed to detect low-frequency (minor allele frequency <5%) modest effect variants.[174] However, when testing multiple variants in aggregate based on their locations in the exome of specific genes, we found that variants located in *FBLN2* together were associated with IA at nominal significance ($p=8x10^{-6}$). The function of *FBLN2* could hypothetically be related to the pathophysiology of IA, because it encodes an extracellular matrix protein that belongs to the fibulin family.[88] *FBLN2* plays a role in the formation of the elastic lamina of blood vessels.[89] The

suggestive association of this gene found in this study remains to be confirmed in separate cohorts.

4. The genetic backgrounds of intracranial and aortic aneurysms mainly consist of differences, but also some similarities (Chapter 5)

Because a co-occurrence of IA, abdominal aortic aneurysms (AAA) and thoracic aortic aneurysms (TAA) has been described,[16, 17,18, 19] mainly within families,[20, 21] we hypothesized the presence of an overlap in genetic background between these diseases. We mainly found arguments against polygenic overlap: we discovered no new shared genetic factors after combining GWAS data of these three diseases, we observed no genetic c orrelation between these diseases, and we found no evidence of polygenic effects (joint effect of groups of SNPs) across the three diseases. These findings may be in line with the differences between IA, AAA and TAA with respect to pathophysiology and epidemiological risk factors. However, we also observed some evidence of genetic overlap of individual risk loci: we replicated associations of SNPs at four loci (9p21, 18q11, 15q21 and 2q33) that were previously described to be associated with IA, AAA or TAA, and observed consistent effects of these SNPs across all three aneurysm types. In addition, the IA risk locus 18q11 and the TAA risk locus 15q21 were found to be associated with AAA with nominal significance. The robustness of these findings remains to be confirmed in additional datasets; all data from all AAA GWAS that are currently available, were present in this study.

5. Genetic data can be used to study the interaction between IA and related traits (Chapter 6 and 7)

We evaluated whether the role of known epidemiological risk traits for IA (high blood pressure) and AAA (hypercholesterolemia and coronary artery disease (CAD)) could be confirmed by testing their genetic overlap. We created genetic risk scores for aneurysm patients and controls, based on alleles known to be associated with these aneurysm risk traits. The connections that were expected from epidemiological knowledge were indeed confirmed. This finding illustrated that a genetic approach can help assess the causal role of traditional risk factors in complex diseases like aneurysms; using genetic data has advantages compared to epidemiological studies, which may be more susceptible to confounding and reverse causation.[134]

Based on these observations, we next performed a comparable study with traits for which the association with aneurysms is yet unclear. Epidemiological studies show that type 2 diabetes (T2D) is inversely associated with IA and AAA. Although adiposity has not been considered a

risk factor for IA, there have been inconsistent reports relating adiposity to AAA risk. We assessed whether these observations have a genetic basis, using the same method of genetic risk scores based on the known risk alleles for these traits. These genetic risk scores were not associated with IA or AAA. Nonetheless, in this second study we also applied a recently developed type of a Mendelian randomization analysis (MR Egger),[165] which can be used to assess causal effects that would remain undetected in a conventional approach of risk score analyses, due to pleiotropic effects. The results of these analyses suggested a potentially causal effect of traits related to adiposity on AAA risk, but they did not reach significance. Replication in other cohorts should further clarify this suggestive association. The absence of a genetic association between T2D and IA or AAA, further supported by Mendelian randomization analyses, suggests that the observed inverse relation between T2D and aneurysms is likely confounded and not real.

In conclusion, these two studies showed that the role of known epidemiological risk traits for IA and AAA can be confirmed by testing their genetic overlap, and that genetic study approaches can also help to clarify possible causality of traits for which the role in IA and AAA is yet unclear from epidemiological observations.

Genetic studies of IA: challenges, changes, and next steps

Though the studies described here have helped in elucidating the genetic contribution to risk of IA, much of the disease etiology remains to be revealed. Genetic studies of IA face a number of challenges, including (1) the potential for bias, as well as (2) limited statistical power due to small sample sizes, (3) the potential role for rare variation, (4) phenotypic heterogeneity, and (5) choices to be made in design of gene expression studies. Below, I discuss these challenges, potential study design changes that could be made to further elucidate the genetic underpinnings of IA, and ways in which future research can move our understanding towards clinical application, thus directly impacting patient treatment.

1. Bias and false-positive associations

GWAS and genomic studies can be susceptible to false-positive findings because of their study design. Up to hundreds of thousands genetic variants or gene products are tested for association to a trait simultaneously in a hypothesis-free way. This results in a big pool of test statistics, all of which may be affected by random statistical fluctuations and hidden confounders in the data. This means that typically, a subgroup of

variants with lower p-values will always be found by chance, although these signals do not represent a true disease association but rather SNP frequency differences between cases and controls observed by chance. Multiple testing adjustments for claiming genome-wide significance (Bonferroni for GWAS, or False Discovery Rate (FDR) for genomic studies) are generally applied to ensure that studies are corrected for multiple testing, and that claims of association are made only for those SNPs or genes likely to be truly associated to disease. Bonferroni correction or FDR minimizes, but does not eliminate, the possibility of false-positive associations.

Additionally, these kinds of studies are designed to detect genetic factors that are common in the general population, and thus mostly confer relatively small effects. This means that large study groups are needed to achieve sufficient statistical power to detect genetic factors with small effects at a genome-wide significant level. Also, replication of findings in an independent cohort is important to confirm the association of genetic factors.

Secondly, technical artifacts and population stratification can cause confounding. For example, if cases and controls come from different populations, are related to each other (either through families or cryptic relatedness), or if they are genotyped on different platforms, subtle differences between cases and controls across all genotyped variants can arise. This can result in the overall inflation of genome-wide test statistics (measured by the so-called genomic inflation factor, or lambda), and subsequently lead to false-positive associations. For these reasons, it is widely known that thorough quality control of genetic and genomic data is extremely important. Even then, confounding by these factors cannot be ruled out, as was demonstrated by experiences from analyses in this thesis. We applied extensive quality control on the GWAS data of IA and AAA cases and control groups, but still observed a high genomic inflation factor. This affected the polygenic analyses for testing the genetic overlap between aneurysm subtypes (Chapter 5). At first, we observed highly significant results for these analyses, but after all control samples were exchanged between the cohorts, this effect completely disappeared. Cases and controls appeared to be genotyped in two different laboratories, and after exchanging all controls, laboratories were equal within each GWAS cohort. This example emphasizes the importance of always accounting for confounding factors when possible, especially when associations are more significant than would be expected based on the statistical power (and thus sample size and SNP frequency) of a study.

2. Where is the missing heritability?

To date, several GWAS studies have been performed in IA, including a "global" GWAS of all IA genotyping data currently available. These efforts resulted in the discovery of up to ten risk loci, which only explain a few percent of the heritability of IA.[13, 15] This gap between explained heritability and total heritability (the "missing heritability") could be explained by several phenomena, discussed here.

Overestimation of heritability

Based on twin studies, the heritability of IA is estimated to be around 40%.[6] However, it is possible that family studies have yielded overestimated heritability estimates, because risk factors related to lifestyle, like smoking, will also cluster in families.

Lack of power

The stringent significant threshold for GWAS, applied to detect only those variants involved in disease and prevent false-positive associations, will inevitably also result in truly associated variants (with small effects) that do not exceed genome-wide significance. An example of this is demonstrated by estimates of explained heritability that are typically modest when considering only genome-wide significant SNPs, but increase substantially when considering all SNPs (regardless of their p-value).[175, 176] Increasing statistical power by increasing the sample size of GWAS can improve the detection of these variants, but for a relatively uncommon disease like IA, this remains a challenge: the prevalence of IA is about 3%,[1] but as the largest part will never rupture, many of these remain undetected. The largest GWAS of IA to date included 2,780 IA cases from different study cohorts,[11] but combining all GWAS data available could increase this number.

Genetic heterogeneity

Genetic heterogeneity can also decrease the power to find associated variants. Firstly, genetic associations can be population-specific: variants near *HDAC9* were only found in the European IA population[14] and low-frequency variants near *FSTL1* and *EPM2A* were only associated with IA in the Dutch and Finnish population.[15] By analyzing all worldwide GWAS data together, population-specific variants could (partly) remain undetected.

Secondly, the phenotypic heterogeneity of a disease can imply presence of genetic heterogeneity. In Chapter 2, we described that established genetic risk variants of IA have a larger effect in IA located at the MCA. The exact meaning of this finding is unknown, but it may suggest that IA risk alleles identified to date play a more important role in the development of IA at the

MCA than at other sites. As the exact pathophysiological mechanisms behind the effect of both environmental and genetic IA risk factors are not completely understood, it is also not known how an increased genetic burden of IA risk alleles predisposes to IA located at MCA. However, it does imply that different phenotypic subtypes may be underpinned by different genetic variants, and pooling subtypes together may not improve power to detect additional risk loci. Instead, studies that increase sample size within a specific subtype (like IA at the MCA) could help to improve power.

Rare variants

A part of the missing heritability may be found in rare variants as well. GWAS are designed to test and detect genetic variants that are relatively common in the population, with minor allele frequencies of at least 5%. Although common variants are indeed likely to explain a large part of the genetic background of complex diseases,[177] whole-genome and whole-exome sequencing studies have showed that rare genetic variants are also likely to play a role.[69] Exome chips have been designed as an affordable way to detect rare and low-frequency genetic variants in coding regions of the genome on large scale.[70, 71] This has resulted in discovery of rare variants involved various traits and diseases, like insulin secretion,[72] myocardial infarction and cholesterol levels,[74] and blood pressure.[93] Even more genetic information can be gained by sequencing (coding regions of) the genome. As whole-exome and genome sequencing at large scale becomes more and more affordable, sequencing promises to help in identification of rare, larger-effect variants that may be useful in clinical care (for example, in patient diagnosis).[178] In the future, such studies could also become a possibility for studying the genetics of IA.

At this moment, discovery of rare variants involved in IA appears to be restricted by the lack of statistical power due to the relatively small sample size of available cohorts with genotypes of rare variants, as shown in Chapter 3. However, studying the role of rare variants does not always require extremely large cohorts. As an example, whole-exome sequencing in families affected with IA can be used to explore the role of rare variants involved in IA.[179] This has recently resulted in discovery of associations between the genes *ADAMTS15*,[180] *RNF213*[181] and *THSD1*,[182] and familial IA. Familial studies can also help to identify genes that are relevant for risk on sporadic IA. This has already been shown for other diseases such as osteoporosis and amyotrophic lateral sclerosis, for which genes identified in familial studies also appeared to play a role in disease in the general population.[176, 183]

Further, genetic studies (including rare variants) in population isolates can help to discover new risk loci, as risk loci may rise to much higher frequency in population isolates compared to randomly-mating populations, typically studied in GWAS. In a GWAS of IA in Finland, a population isolate with an increased incidence of IA, four new risk loci were discovered, of which two were replicated in our Dutch GWAS cohort.[15]

3. Clarifying the role of pleiotropy

Pleiotropy, the phenomenon that one genetic variant is associated with multiple traits, also appears to play a role in IA (and AAA) pathology. In this thesis, we described that the genetic background of aneurysms harbours a substantial overlap with that of related traits (Chapter 6 and 7). Similarly, pleiotropic effects have been described for the 5q23.2[150] and 10q24[11] loci: alleles at these loci were found to be (suggestively) associated with IA[150, 11] as well as hypertension,[139, 140] the most important risk factor for IA.[30] These findings imply that it is not always clear whether a risk locus is directly associated with aneurysms, or that the risk effect acts through the risk of associated traits. Mendelian randomization studies can further help to make this distinction, disentangling those alleles that are associated from those that are specifically causal for disease.[151] Therefore, it is important that detailed phenotype information, which is necessary for successful Mendelian randomization studies, will be available for samples in genetic studies.

4. Design of gene expression studies

As we described in Chapter 3, no structural gene expression differences were found in blood of previous SAH patients compared to controls. Recently, two other gene-expression studies using blood of IA/SAH patients have been performed, but these studies focused on gene-expression differences after IA rupture,[62, 63] so their results rather reflect the consequences than the causes of IA and SAH. The absence of gene expression differences in our study may imply that gene expression differences relevant to IA development or rupture do not occur in blood. In contrast, multiple gene expression studies on IA tissue did show differential expression of genes, which were confirmed by other studies.[22] Also, a recent large gene expression study in IA tissue showed expression differences for genes involved in extracellular matrix and immune response,[23] processes that are known to play a role in IA pathology.[184] The contrast between these findings and our study in blood may be explained by the fact that (regulation of) gene expression varies across different tissues.[185, 186] Although blood is much easier to obtain and could therefore also be used for clinical practice in case of relevant findings, it seems less

effective to detect relevant gene expression differences. Alternatively, future studies could aim at detecting changes in microRNA, which plays an important role in the post-transcriptional regulation of gene expression, and is more stable in blood than RNA.[187] Also, the choice of patients in a (micro)RNA study is important: we hypothesized that IA is a chronic vessel disease and that gene expression differences could be measured long after IA rupture and therapy, but this might not be true for all patients. Including patients with yet unruptured IA that grow over time could therefore be more successful to detect gene expression differences secondary to IA growth, and such findings could also have more clinical relevance, for example for treatment decision-making in patients with unruptured IA.

5. Clinical applicability of genetic studies

Although most of the GWAS findings in complex diseases have not led to a change in clinical care, there are numerous examples demonstrating that GWAS results have provided insights in pathophysiological processes behind many complex diseases, sometimes with direct clinical relevance. For example, genes discovered by GWAS for autoimmune diseases appear to be new therapeutic targets,[176] and discovery of the association between lipids and *PCSK9* (first discovered in hyperlipidemia families and then through GWAS of myocardial infarction)[188, 189] have led to the development of *PCSK9* inhibitors, already on the market in the United States.[190]

Genetic studies of IA have not yet resulted in new insights in IA pathology, but they do provide additional arguments for existing hypotheses about the mechanisms underlying IA formation, and they do so through the study of a human model (i.e., DNA variation in humans). For example, the function of genes located in IA risk loci detected by GWAS is not always known, but in general they point to a possible role of cell cycle dysfunction (*SOX17, CDKN2A, CDKN2B, STARD13*) and endothelial dysfunction (*EDNRA*) in the pathogenesis of IA.[11, 13] This is in line with molecular biology studies, indicating that IA formation involves endothelial dysfunction, followed by pathological remodeling with degenerative changes of vascular walls, with an additional role for immunological processes.[184, 191] Results from gene expression studies also further confirm involvement of genes with functions related to these mechanisms.[22, 23] Future genetic studies could lead to new hypotheses about biological processes behind IA development. Besides discovery of new genetic factors associated with IA, fine-mapping of the known IA risk loci by deep sequencing could uncover the causal variant driving the SNP association picked up by GWAS. Together with identification of the functional consequences of these variants, this information can result in more information about disease mechanisms.[192]

Clinical risk prediction of IA based on genetic factors is unlikely to be realized in the near future: the low odds ratios of established IA risk SNPs from GWAS (up to 1.3) suggest that these SNPs are unlikely to add substantially to the risk factors known from clinical observations. Indeed, a recent study showed that the presence of IA in the general (Dutch) population cannot be predicted by a risk score based on ten known risk SNPs of IA, nor by risk scores based on SNPs associated with high blood pressure.[193] However, genetic risk scores have shown to be useful for clinical practice in, for example, CAD: a recent large community-based study showed that people with a high genetic risk for CAD had the largest clinical benefit from statin therapy.[194] When the genetic basis of IA will be further revealed (including the detection of the causal variants within IA risk loci detected by GWAS), genetic risk prediction could similarly help to identify people at high risk of IA, and especially IA with high risk of rupture, who should undergo radiological screening.

Conclusion

IA is a typical example of a complex trait: environmental factors play an important role in disease risk, but there is also a strong heritable component. Multiple genetic risk factors for IA have been discovered, but the genetic background of IA is still largely unknown. Until now, the successes of studies uncovering the genetic background of complex traits have mainly been determined by the sample size of the study cohorts, which is a limiting factor for a relatively uncommon trait like IA. This limitation also played a role in studies described in this thesis, and underlines the importance of collaboration and sharing of genetic data in the search for new genetic risk factors. This thesis also illustrates how a genetic study approach can give insight in the interaction between IA and related traits, including other types of aneurysms. Furthermore, we show that genetic heterogeneity exists for IA. These findings emphasize the need for ‘deep phenotyping’ in genetic studies, in addition to the collection and analysis of more comprehensive genotype information.

Future studies on genetics of IA should focus on detection of rare variants by whole-genome sequencing of (familial) IA patients, proving causality of associated variants by Mendelian randomization studies, and identifying the functional consequences of associated variants. This can give new insights in the pathophysiology of IA formation, and also lead to discoveries that are relevant for clinical practice, including risk stratification, disease prevention, and therapeutic developments.

Addendum

Samenvatting in het Nederlands

Achtergrond

Intracraniële aneurysmata (IA) zijn uitstulpingen in de vaatwand van slagaders in het hoofd. Ongeveer 3% van alle mensen heeft een IA, maar het grootste deel zal dit nooit merken. Bij een klein deel kan een IA echter scheuren, en daarmee leiden tot een subarachnoïdale bloeding (SAB): een bloeding tussen de hersenvliezen. Een derde van de patiënten overlijdt aan deze aandoening. Hoewel de oorzaak van IA niet in alle gevallen bekend is, weten we dat enkele beïnvloedbare factoren (roken, hoge bloeddruk, overmatig alcoholgebruik) een verhoogd risico geven op IA. Ook genetische factoren spelen waarschijnlijk een rol, want eerstegraads familieleden van SAB-patiënten hebben een zeven maal verhoogd risico om zelf ook een SAB te krijgen.

Bij de meeste mensen met IA komt dit echter niet in de familie voor; in deze gevallen leidt de combinatie van beïnvloedbare en meerdere genetische factoren waarschijnlijk tot de ziekte. De afgelopen jaren zijn verschillende studies gedaan met als doel deze genetische factoren op te helderen. Door middel van genoomwijde associatiestudies, waarbij de associatie tussen veelvoorkomende varianten in het DNA en de aanwezigheid van een ziekte wordt getest, zijn er inmiddels tien gebieden in het genoom gevonden die een verband hebben met IA. Toch is de kans op IA voor een individuele patiënt niet te voorspellen aan de hand van alleen deze gevonden gebieden. Een hypothese is dat meer ziekte-geassocieerde gebieden gevonden kunnen worden door ook zeldzamere genetische varianten te onderzoeken. Dit is door technologische ontwikkelingen pas recent praktisch mogelijk geworden. Naast het in kaart brengen van meer genetische risicofactoren, kan ook het mechanisme waarmee genetische risicofactoren tot ziekte leiden nog verder worden opgehelderd. Meer

kennis over de rol van genproducten (zoals RNA, het afleesproduct van DNA) kan hier onder andere aan bijdragen.

Dit proefschrift

In dit proefschrift beschrijf ik de zoektocht naar nieuwe genetische factoren voor IA, naar genetische verschillen binnen subtypen van IA, en naar een relatie tussen de genetische achtergrond van IA en gerelateerde aandoeningen. Hieronder vat ik de belangrijkste bevindingen samen.

Het genetisch risicoprofiel van IA verschilt per plek in het vaatstelsel in het hoofd (Hoofdstuk 2)

In genetische studies worden mensen met IA in het algemeen beschouwd als een homogene groep, maar vanuit klinische ervaring weten we dat er verschillende uitingsvormen zijn. Zo kunnen IA op verschillende plaatsen in het vaatstelsel van het hoofd voorkomen, en zijn er kleine verschillen in de niet-genetische risicoprofielen van patiënten (leeftijd, geslacht, familiaire belasting) per plaats. Onze hypothese was dat de rol van genetische factoren ook zou kunnen verschillen tussen subgroepen van patiënten. Daarop hebben we het genetische risicoprofiel voor IA per patiënten-subgroep samengesteld en vergeleken. Hieruit bleek dat patiënten met IA op de middelste hersenslagader een duidelijk sterker genetisch risicoprofiel hebben. Daarbij is het opvallend dat IA vaker op deze locatie voorkomen bij patiënten met een familiaire belasting voor SAB. De mate waarin genetische factoren een rol spelen, lijkt dus afhankelijk van de specifieke locatie van IA. Deze nieuwe kennis kan gebruikt worden bij toekomstige genetische studies.

Er is geen veranderd profiel van genexpressie (RNA) in bloed van voormalig SAB-patiënten (Hoofdstuk 3)

We vergeleken het profiel van genexpressie in bloed van mensen met een jaren geleden doorgemaakte SAB en een controlegroep. Er waren geen noemenswaardige verschillen, zodat het niet mogelijk bleek om mensen met en zonder doorgemaakte SAB van elkaar te onderscheiden op basis van hun genexpressieprofiel. Opmerkelijk is dat eerdere studies wel genexpressieveranderingen vonden in IA-weefsel dat werd afgenomen bij operaties. Blijkbaar zijn veranderingen in genexpressie bij IA niet duidelijk in bloed meetbaar, in ieder geval niet meer jaren na de doorgemaakte SAB.

Genetische varianten in het gen FBLN2 zijn mogelijk geassocieerd met IA (Hoofdstuk 4)

We onderzochten de associatie van weinig voorkomende genetische varianten, verspreid over het exoom (het voor eiwitten coderende deel van het genoom), met IA. We konden geen associatie van één van deze

varianten aantonen; waarschijnlijk was onze studiegroep te klein voor het betrouwbaar statistisch aantonen van deze individuele associaties. We onderzochten vervolgens ook de aanwezigheid van meerdere varianten in hetzelfde gen. Daarbij vonden we wel enig verband tussen IA en varianten in het gen *FBLN2*, een gen met een veronderstelde rol in de vorming van de vaatwand. Om meer bewijs te verzamelen voor dit verband wordt onze bevinding momenteel verder onderzocht in een andere studiegroep.

Aneurysmata van hoofd, borst en buik zijn genetisch vooral verschillend, maar er zijn ook overeenkomsten (Hoofdstuk 5)

Tussen de drie typen aneurysmata in het hoofd (IA), de aorta in de buik (AAA) en borst (TAA) bestaan belangrijke verschillen, zoals de vaatopbouw op de plaats van het aneurysma en het niet-genetische risicoprofiel van de patiënten. Ook zijn er juist overeenkomsten: ze komen relatief vaak tegelijk voor binnen families en ook is er één genetische variant bekend die sterk geassocieerd is met zowel IA als AAA. We wilden in onze grote groep patiënten met verschillende typen aneurysmata onderzoeken of er meer overlap in de genetische achtergrond zou kunnen zijn. Allereerst vonden we geen nieuwe genetische risicovarianten voor aneurysmata in het algemeen. Ook vertoonden de genetische profielen van IA, AAA en TAA over het geheel vooral verschillen. Toch bleken enkele bekende genetische risicovarianten voor IA en TAA individueel vrij sterk geassocieerd te zijn met AAA. Dit is een aanwijzing dat er toch meer gedeelde risicofactoren bestaan voor verschillende typen aneurysmata dan tot nu toe bekend was.

Genetische informatie is bruikbaar voor onderzoek naar de relatie tussen aneurysmata en andere aandoeningen (Hoofdstuk 6 en 7)

Er zijn verschillende aandoeningen die vaak samen gezien worden met aneurysmata. Zo komt IA vaak voor met een hoge bloeddruk, en AAA met een ongunstig cholesterolprofiel en verstopte kransslagaders van het hart. Voor elk van deze aandoeningen zijn tientallen genetische risicovarianten bekend. Op basis van deze informatie hebben we een aandoening-specifiek genetisch profiel bepaald bij IA- en AAA-patiënten, en dat vergeleken met een controlegroep. Hieruit bleek dat de specifieke profielen precies zo gerelateerd waren aan IA en AAA als we hadden verwacht op basis van de geobserveerde relaties in de praktijk. Deze bevindingen brachten ons op het idee om op dezelfde manier ook omstreden relaties van aandoeningen met IA en AAA te testen. Ten eerste wordt in de literatuur het oorzakelijke verband tussen obesitas en AAA niet consistent aangetoond. Op onze beurt vonden wij zwakke aanwijzingen voor een genetische relatie. Ten tweede lijkt diabetes mellitus type 2 minder vaak voor te komen bij aneurysmapatiënten, zodat soms wordt gedacht dat

diabetes beschermt tegen het krijgen van een aneurysma. Wij vonden echter geen enkel genetisch verband tussen diabetes en IA of AAA. Deze genetisch onderbouwde bevinding maakt het onwaarschijnlijk dat er een oorzakelijk verband tussen deze aandoeningen is.

Conclusie

1. Patiënten met IA op de middelste hersenslagader hebben een hogere genetische risico-belasting dan andere IA-patiënten.
2. Er kunnen geen noemenswaardige RNA-veranderingen worden gemeten in bloed van mensen met een doorgemaakte SAB.
3. Weinig voorkomende genetische varianten in het *FBLN2*-gen lijken gezamenlijk geassocieerd te zijn met IA.
4. IA, AAA en TAA vertonen zowel klinisch als genetisch vooral verschillen, maar enkele genetische risicovarianten voor IA en TAA hebben ook een relatie met AAA.
5. Aandoeningen die het risico op IA of AAA verhogen, delen ook een genetische achtergrond met IA of AAA.

Dankwoord

Na jaren is dit proefschrift dan eindelijk af. Dit zou nooit gebeurd zijn zonder hulp van anderen. Een aantal mensen wil ik hiervoor in het bijzonder bedanken.

Allereerst mijn (co-)promotoren, dr. Y.M. Ruigrok, prof. dr. P.I.W. de Bakker en prof. dr. G.J.E. Rinkel. Jullie verschillende invalshoeken en stijl van begeleiding waren ieder op hun eigen manier heel waardevol en vulden elkaar goed aan.

Beste Ynte, als dagelijks begeleider was je altijd enorm betrokken, behulpzaam en goed op de hoogte. Je pragmatische aanpak en ontspannen, positieve manier waarmee je een wetenschappelijke carrière, klinisch werk en een druk gezinsleven met elkaar combineert, vormen een belangrijk voorbeeld voor me. Als mijn gedachten of teksten even vastliepen, wist jij er altijd een schijnbaar simpele draai aan te geven en me weer op het goede spoor te brengen. Ik ben je heel dankbaar voor je onvoorwaardelijke steun in de afgelopen jaren.

Beste Paul, jij hebt me wegwijs gemaakt in het genetisch onderzoek. Eigenlijk moet ik in de eerste plaats Marike bedanken, die al voor aanvang van mijn promotieonderzoek een stage bij jou in Boston voorstelde. Je nam me daar als vanzelfsprekend op in de groep, en na je overstap naar Utrecht deed je dat nog eens. Je geduldige en heldere uitleg, kritische blik, vernieuwende ideeën, en ook je gevoel voor humor zijn ontzettend belangrijk voor me geweest. Ik zal altijd met plezier blijven terugdenken aan onze discussies en lab-meetings.

Beste Gabriël, als klinisch onderzoeker zei je altijd weinig van genetica te weten, maar je wist toch met verdacht veel goede suggesties te komen tijdens onze overlegmomenten. Juist je ervaring vanuit de kliniek leidde tot goede praktische ideeën voor bijvoorbeeld de opzet van het genexpressie-onderzoek, en voorkwam een genetische tunnelvisie. Bedankt voor je betrokkenheid.

Dear members of the reading committee, prof. dr. F.W. Asselbergs, prof. dr. V.V.A.M. Knoers, prof. dr. M.L. Bots, prof. dr. D.P.V. de Kleijn and prof. dr. D. Woo, thank you for your efforts to critically read this thesis.

Prof. dr. J.H. Veldink, beste Jan, dank je wel voor je begeleiding bij het genexpressie-project. Hoewel er zogezegd 'niks uit kwam', had je wel iedere mogelijke analysetechniek in de vingers om dat duidelijk te krijgen. Bedankt voor al je uitleg over dit boeiende deel van de genetica.

Prof. dr. Wijmenga, beste Cisca, je was helaas al naar Groningen vertrokken toen ik naar Utrecht kwam, maar ik kon mijn onderzoek doen dankzij de basis die jij daarvoor in Utrecht hebt gelegd, inclusief de onderzoekssubsidie die er dankzij jou is gekomen. Bedankt daarvoor.

Dear former colleagues in Boston (Sara, Paul McLaren, Sherman, Nikos, Lori, Eli, Fina): thank you for teaching me the basics of bioinformatics – almost from the level of turning on a computer, to running a GWAS - and also for the great time together!

Dear members of the de Bakker group in Utrecht (Jessica, Sara, Laurent, Vinicius, Androniki, Mircea, Balder, Sander, Maarten, Jytte en Clara): thank you for surrounding me for years with cleverness, good company, computer scripts, useful advice and discussions. It was an honour to be a member of this team for a few years.

Jessica van Setten, aan jou heb ik in het bijzonder veel te danken wat betreft hulp bij de computeranalyses. Maar je bleek ook nog een goede reisgenoot; dank je wel voor de mooie tijd in Seattle en omstreken!

De leden van het ALS-lab wil ik bedanken voor het adopteren van een vasculair onderzoeker in de periode van de genexpressie-analyses, voor het uitlenen van ruimte op de harde schijf (het mag er nu af), de koffie, gezelligheid, en natuurlijk Frank Diekstra en Wouter van Rheenen voor hulp bij het doorgronden van de genexpressie-data.

Jelena Medic, bedankt voor je ondersteuning in het lab. Het was leerzaam om eens met je mee te kijken – dat onderdeel in het hoofdstuk 'Methods' is toch een stuk makkelijker opgeschreven dan gedaan!

Studenten Duncan Lee, Anar Memmedov en Arno de Wilde, bedankt voor jullie interesse en hulp. Anar, bedankt voor het vele prikwerk in de vroege ochtend.

Prof. dr. J.H.J. Wokke en dr. T. Seute, beste John en Tatjana, wat ben ik trots dat ik onder jullie begeleiding tot neuroloog word opgeleid, in een kliniek waar zoveel ruimte is voor het doen van onderzoek naast klinisch werk, en tegelijkertijd ook begrip is voor de drukte die dit met zich meebrengt.

Arts-assistenten neurologie, wat fijn om dit boeiende werk samen te kunnen doen met zo'n leuke en hechte groep collega's! In het bijzonder mijn trouwe jaargenootjes Celine, Merel, Marjolein en Oliver, maar ook Renske, roostermaatjes Kim en Nora, en alle anderen voor de momenten samen op het werk en daarbuiten.

Mijn paranimfen, Hanneke Miedema en Sara Pulit:

Hanneke, twintig jaar geleden worstelden we ons in het NHJO al samen door grote symfonieën, maar we hebben op de fiets ook veel tegenwind en steile hellingen getrotseerd, en zelfs op de bodem van een troebele oceaan liet je me niet stikken. Ik ben heel blij dat je ook bij deze promotie aan mijn zijde wilt staan!

Sara, vanaf de eerste dag in Boston tot aan de laatste zin van dit proefschrift ben je een belangrijke factor geweest in mijn promotietijd. Je stond altijd voor me klaar met hulp, uitleg en motiverende woorden, maar wat hebben we ook veel leuke momenten beleefd buiten het werk in al die tijd! Ik heb bewondering voor je onderzoekstalent en je positiviteit, en voel me enorm gesteund met jou als paranimf.

Lieve papa en mama, zonder jullie was dit proefschrift letterlijk nooit af geweest. Wat ben ik blij dat ik onbeperkt bij jullie kon komen werken als dat nodig was, terwijl meteen ook op de kinderen werd gepast, eten werd klaargemaakt, mijn werk direct op correct Engels taalgebruik werd gecontroleerd en ik ook nog de nodige slaap kon inhalen. Bovendien was het zo gezellig dat ik het bijna jammer vind dat dit laatste deel promotietijd voorbij is! Ik ben ontzettend verwend met zulke ouders. Dit geldt natuurlijk

ook voor mijn schoonouders Cor en Emmy, die altijd klaarstonden om te helpen waar nodig.

Tot slot: lieve Frank, dank je wel voor alle afleiding naast, maar vooral ook voor je betrokkenheid bij mijn werk. Met je kritische vragen vanuit een heel andere invalshoek bleef je me aan het denken zetten, en ben je dus niet alleen verantwoordelijk voor de buitenkant, maar indirect ook voor de inhoud van dit proefschrift. Ik kijk uit naar meer van dit soort momenten die we ongetwijfeld samen nog zullen beleven.

List of publications

In this thesis

van 't Hof FN, Ruigrok YM, Baas AF, Kiemeney LA, Vermeulen SH, Uitterlinden AG, Hofman A, Rivadeneira F, Rinkel GJ, de Bakker PI. **Impact of inherited genetic variants associated with lipid profile, hypertension, and coronary artery disease on the risk of intracranial and abdominal aortic aneurysms.** *Circ Cardiovasc Genet*. 2013 Jun;6(3):264-70.

van 't Hof FN, Kurki MI, Kleinloog R, de Bakker PI, von und zu Fraunberg M, Jääskeläinen JE, Gaál EI, Lehto H, Kivisaari R, Laakso A, Niemelä M, Hernesniemi J, Brouwer MC, van de Beek D, Rinkel GJ, Ruigrok YM. **Genetic risk load according to the site of intracranial aneurysms.** *Neurology.* 2014 Jul 1;83(1):34-9.

van 't Hof FN, Ruigrok YM, Medic J, Sanjabi B, van der Vlies P, Rinkel GJ, Veldink JH. **Whole Blood Gene Expression Profiles of Patients with a Past Aneurysmal Subarachnoid Hemorrhage.** *PLoS One*. 2015 Oct 6;10(10):e0139352.

van 't Hof FN, Ruigrok YM, Lee CH, Ripke S, Anderson G, de Andrade M, Baas AF, Blankensteijn JD, Böttinger EP, Bown MJ, Broderick J, Bijlenga P, Carrell DS, Crawford DC, Crosslin DR, Ebeling C, Eriksson JG, Fornage M, Foroud T, von Und Zu Fraunberg M, Friedrich CM, Gaál EI, Gottesman O, Guo DC, Harrison SC, Hernesniemi J, Hofman A, Inoue I, Jääskeläinen JE, Jones GT, Kiemeney LA, Kivisaari R, Ko N, Koskinen S, Kubo M, Kullo IJ, Kuivaniemi H, Kurki MI, Laakso A, Lai D, Leal SM, Lehto H, LeMaire SA, Low SK, Malinowski J, McCarty CA, Milewicz DM, Mosley TH, Nakamura Y,

Nakaoka H, Niemelä M, Pacheco J, Peissig PL, Pera J, Rasmussen-Torvik L, Ritchie MD, Rivadeneira F, van Rij AM, Santos-Cortez RL, Saratzis A, Slowik A, Takahashi A, Tromp G, Uitterlinden AG, Verma SS, Vermeulen SH, Wang GT; Aneurysm Consortium; Vascular Research Consortium of New Zealand, Han B, Rinkel GJ, de Bakker PI. **Shared Genetic Risk Factors of Intracranial, Abdominal, and Thoracic Aneurysms.** *J Am Heart Assoc*. 2016 Jul 14;5(7).

van 't Hof FN, Vaucher J, Holmes MV, de Wilde A, Baas AF, Blankensteijn JD, Hofman A, Kiemeney LA, Rivadeneira F, Uitterlinden AG, Vermeulen SH, Rinkel GJ, de Bakker PI, Ruigrok YM**. Genetic variants associated with type 2 diabetes and adiposity and risk of intracranial and abdominal aortic aneurysms.** *Eur J Hum Genet.* 2017 Apr 5. (epub ahead of print)

Not included in this thesis

van 't Hof FNG, Carpay JA. **Discontinuation of anti-epileptic medication: what is the ratio and the evidence?** *Tijdschr Neurol Neurochir* 2008 Dec;109:345-51 (in Dutch)

Kleinloog R, **van 't Hof FN**, Wolters FJ, Rasing I, van der Schaaf IC, Rinkel GJ, Ruigrok YM. **The association between genetic risk factors and the size of intracranial aneurysms at time of rupture.** *Neurosurgery.* 2013 Oct;73(4):705-8.

Harrison SC, Smith AJ, Jones GT, Swerdlow DI, Rampuri R, Bown MJ; Aneurysm Consortium, Folkersen L, Baas AF, de Borst GJ, Blankensteijn JD, Price JF, van der Graaf Y, McLachlan S, Agu O, Hofman A, Uitterlinden AG, Franco-Cereceda A, Ruigrok YM, **van't Hof FN**, Powell JT, van Rij AM, Casas JP, Eriksson P, Holmes MV, Asselbergs FW, Hingorani AD, Humphries SE. **Interleukin-6 receptor pathways in abdominal aortic aneurysm.** *Eur Heart J.* 2013 Dec;34(48):3707-16.

Kurki MI, Gaál EI, Kettunen J, Lappalainen T, Menelaou A, Anttila V, **van 't Hof FN**, von Und Zu Fraunberg M, Helisalmi S, Hiltunen M, Lehto H, Laakso A, Kivisaari R, Koivisto T, Ronkainen A, Rinne J, Kiemeney LA, Vermeulen SH, Kaunisto MA, Eriksson JG, Aromaa A, Perola M, Lehtimäki T, Raitakari OT, Salomaa V, Gunel M, Dermitzakis ET, Ruigrok YM, Rinkel

GJ, Niemelä M, Hernesniemi J, Ripatti S, de Bakker PI, Palotie A, Jääskeläinen JE. **High risk population isolate reveals low frequency variants predisposing to intracranial aneurysms.** *PLoS Genet.* 2014 Jan 30;10(1):e1004134

Jones GT, Tromp G, Kuivaniemi H, Gretarsdottir S, Baas AF, Giusti B, Strauss E, **Van't Hof FN**, Webb TR, Erdman R, Ritchie MD, Elmore JR, Verma A, Pendergrass S, Kullo IJ, Ye Z, Peissig PL, Gottesman O, Verma SS, Malinowski J, Rasmussen-Torvik LJ, Borthwick KM, Smelser DT, Crosslin DR, de Andrade M, Ryer EJ, McCarty CA, Böttinger EP, Pacheco JA, Crawford DC, Carrell DS, Gerhard GS, Franklin DP, Carey DJ, Phillips VL, Williams MJ, Wei W, Blair R, Hill AA, Vasudevan TM, Lewis DR, Thomson IA, Krysa J, Hill GB, Roake J, Merriman TR, Oszkinis G, Galora S, Saracini C, Abbate R, Pulli R, Pratesi C, Saratzis A, Verissimo AR, Bumpstead S, Badger SA, Clough RE, Cockerill G, Hafez H, Scott DJ, Futers TS, Romaine SP, Bridge K, Griffin KJ, Bailey MA, Smith A, Thompson MM, van Bockxmeer FM, Matthiasson SE, Thorleifsson G, Thorsteinsdottir U, Blankensteijn JD, Teijink JA, Wijmenga C, de Graaf J, Kiemeney LA, Lindholt JS, Hughes A, Bradley DT, Stirrups K, Golledge J, Norman PE, Powell JT, Humphries SE, Hamby SE, Goodall AH, Nelson CP, Sakalihasan N, Courtois A, Ferrell RE, Eriksson P, Folkersen L, Franco-Cereceda A, Eicher JD, Johnson AD, Betsholtz C, Ruusalepp A, Franzén O, Schadt EE, Björkegren JL, Lipovich L, Drolet AM, Verhoeven EL, Zeebregts CJ, Geelkerken RH, van Sambeek MR, van Sterkenburg SM, de Vries JP, Stefansson K, Thompson JR, de Bakker PI, Deloukas P, Sayers RD, Harrison SC, van Rij AM, Samani NJ, Bown MJ. **Meta-Analysis of Genome-Wide Association Studies for Abdominal Aortic Aneurysm Identifies Four New Disease-Specific Risk Loci**. Circ Res. 2017 Jan 20;120(2):341-353.

Curriculum Vitae

Femke van ‘t Hof werd geboren op 20 januari 1981 te Beverwijk. In 1999 behaalde ze haar VWO-diploma aan het Jac. P. Thijsse College te Castricum, waarna ze begon aan haar studie Geneeskunde aan de Vrije Universiteit te Amsterdam. Voor ze met haar co-schappen begon, deed ze een onderszoeksstage bij de afdeling cardiofysiologie aan de VU. Na haar artsexamen in 2006 had ze verschillende aanstellingen als ANIOS, waarbij ze uiteindelijk uitkwam bij de afdeling neurologie van het UMC Utrecht. In 2009 begon ze daar met haar promotieonderzoek onder begeleiding van dr. Ynte Ruigrok, prof. dr. Gabriel Rinkel en prof. dr. Paul de Bakker. De resultaten hiervan staan in dit proefschrift beschreven. Ze verrichtte in het kader van haar promotieonderzoek een onderzoeksstage van 4 maanden op de afdeling genetica van Harvard Medical School en Brigham and Women’s Hospital in Boston, waar haar promotor Paul de Bakker destijds werkte als assistent-professor. In 2011 begon ze aan haar opleiding tot neuroloog in het UMC Utrecht. In het kader van deze opleiding zal ze kort na haar promotie beginnen aan een keuzestage in het Epilepsy Research Centre in Melbourne. Ze verwacht haar specialisatie begin 2019 af te ronden.

Bibliography

1. Vlak MH, Algra A, Brandenburg R, Rinkel GJ. Prevalence of unruptured intracranial aneurysms, with emphasis on sex, age, comorbidity, country, and time period: a systematic review and meta-analysis. Lancet Neurol 2011;10:626-636.
2. Nieuwkamp DJ, Setz LE, Algra A, Linn FH, de Rooij NK, Rinkel GJ. Changes in case fatality of aneurysmal subarachnoid haemorrhage over time, according to age, sex, and region: a meta-analysis. Lancet Neurol 2009;8:635-642.
3. Frosen J, Tulamo R, Paetau A, et al. Saccular intracranial aneurysm: pathology and mechanisms. Acta neuropathologica 2012;123:773-786.
4. Aoki T, Nishimura M. The development and the use of experimental animal models to study the underlying mechanisms of CA formation. Journal of biomedicine & biotechnology 2011;2011:535921.
5. Ruigrok YM, Rinkel GJ, Wijmenga C. Genetics of intracranial aneurysms. Lancet Neurol 2005;4:179-189.
6. Korja M, Silventoinen K, McCarron P, et al. Genetic epidemiology of spontaneous subarachnoid hemorrhage: Nordic Twin Study. Stroke 2010;41:2458-2462.
7. Cardon LR, Bell JI. Association study designs for complex diseases. Nature reviews Genetics 2001;2:91-99.
8. Ruigrok YM, Rinkel GJ. Genetics of intracranial aneurysms. Stroke 2008;39:1049-1055.
9. Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of complex diseases. Nature 2009;461:747-753.
10. Bilguvar K, Yasuno K, Niemela M, et al. Susceptibility loci for intracranial aneurysm in European and Japanese populations. Nat Genet 2008;40:1472-1477.
11. Yasuno K, Bilguvar K, Bijlenga P, et al. Genome-wide association study of intracranial aneurysm identifies three new risk loci. Nature genetics 2010;42:420-425.
12. Helgadottir A, Thorleifsson G, Magnusson KP, et al. The same sequence variant on 9p21 associates with myocardial infarction, abdominal aortic aneurysm and intracranial aneurysm. Nat Genet 2008;40:217-224.

13. Yasuno K, Bakircioglu M, Low SK, et al. Common variant near the endothelin receptor type A (EDNRA) gene is associated with intracranial aneurysm risk. Proc Natl Acad Sci U S A 2011;108:19707-19712.
14. Foroud T, Lai D, Koller D, et al. Genome-wide association study of intracranial aneurysm identifies a new association on chromosome 7. Stroke 2014;45:3194-3199.
15. Kurki MI, Gaal EI, Kettunen J, et al. High risk population isolate reveals low frequency variants predisposing to intracranial aneurysms. PLoS Genet 2014;10:e1004134.
16. Larsson E, Vishnevskaya L, Kalin B, Granath F, Swedenborg J, Hultgren R. High frequency of thoracic aneurysms in patients with abdominal aortic aneurysms. Annals of surgery 2011;253:180-184.
17. Svensjo S, Bengtsson H, Bergqvist D. Thoracic and thoracoabdominal aortic aneurysm and dissection: an investigation based on autopsy. Br J Surg 1996;83:68-71.
18. Miyazawa N, Akiyama I, Yamagata Z. Risk factors for the association of intracranial and aortic aneurysms. Acta Neurochir (Wien) 2007;149:221-229; discussion 229.
19. Norrgard O, Angqvist KA, Fodstad H, Forssell A, Lindberg M. Co-existence of abdominal aortic aneurysms and intracranial aneurysms. Acta Neurochir (Wien) 1987;87:34-39.
20. Regalado E, Medrek S, Tran-Fadulu V, et al. Autosomal dominant inheritance of a predisposition to thoracic aortic aneurysms and dissections and intracranial saccular aneurysms. American journal of medical genetics Part A 2011;155A:2125-2130.
21. Kim DH, Van Ginhoven G, Milewicz DM. Familial aggregation of both aortic and cerebral aneurysms: evidence for a common genetic basis in a subset of families. Neurosurgery 2005;56:655-661; discussion 655-661.
22. Roder C, Kasuya H, Harati A, Tatagiba M, Inoue I, Krischek B. Meta-analysis of microarray gene expression studies on intracranial aneurysms. Neuroscience 2012;201:105-113.
23. Kleinloog R, Verweij BH, van der Vlies P, et al. RNA Sequencing Analysis of Intracranial Aneurysm Walls Reveals Involvement of Lysosomes and Immunoglobulins in Rupture. Stroke 2016;47:1286-1293.
24. Zhan X, Jickling GC, Tian Y, et al. Transient ischemic attacks characterized by RNA profiles in blood. Neurology 2011;77:1718-1724.
25. Stamova B, Xu H, Jickling G, et al. Gene expression profiling of blood for the prediction of ischemic stroke. Stroke; a journal of cerebral circulation 2010;41:2171-2177.
26. Lindner SH, Bor AS, Rinkel GJ. Differences in risk factors according to the site of intracranial aneurysms. Journal of neurology, neurosurgery, and psychiatry 2010;81:116-118.
27. Bromberg JE, Rinkel GJ, Algra A, et al. Familial subarachnoid hemorrhage: distinctive features and patterns of inheritance. Annals of neurology 1995;38:929-934.

28. Ronkainen A, Hernesniemi J, Tromp G. Special features of familial intracranial aneurysms: report of 215 familial aneurysms. Neurosurgery 1995;37:43-46; discussion 46-47.
29. Mackey J, Brown RD, Jr., Moomaw CJ, et al. Unruptured intracranial aneurysms in the Familial Intracranial Aneurysm and International Study of Unruptured Intracranial Aneurysms cohorts: differences in multiplicity and location. Journal of neurosurgery 2012;117:60-64.
30. Feigin VL, Rinkel GJ, Lawes CM, et al. Risk factors for subarachnoid hemorrhage: an updated systematic review of epidemiological studies. Stroke; a journal of cerebral circulation 2005;36:2773-2780.
31. Greebe P, Bromberg JE, Rinkel GJ, Algra A, van Gijn J. Family history of subarachnoid haemorrhage: supplemental value of scrutinizing all relatives. Journal of neurology, neurosurgery, and psychiatry 1997;62:273-275.
32. Mackey J, Brown RD, Jr., Moomaw CJ, et al. Familial intracranial aneurysms: is anatomic vulnerability heritable? Stroke; a journal of cerebral circulation 2013;44:38-42.
33. Park SK, Kim JM, Kim JH, Cheong JH, Bak KH, Kim CH. Aneurysmal subarachnoid hemorrhage in young adults: a gender comparison study. Journal of clinical neuroscience : official journal of the Neurosurgical Society of Australasia 2008;15:389-392.
34. Horiuchi T, Tanaka Y, Hongo K. Sex-related differences in patients treated surgically for aneurysmal subarachnoid hemorrhage. Neurologia medico-chirurgica 2006;46:328-332; discussion 332.
35. Inagawa T. Site of ruptured intracranial saccular aneurysms in patients in Izumo City, Japan. Cerebrovasc Dis 2010;30:72-84.
36. Nakaoka H, Takahashi T, Akiyama K, et al. Differential effects of chromosome 9p21 variation on subphenotypes of intracranial aneurysm: site distribution. Stroke; a journal of cerebral circulation 2010;41:1593-1598.
37. van Munster CE, von und zu Fraunberg M, Rinkel GJ, Rinne J, Koivisto T, Ronkainen A. Differences in aneurysm and patient characteristics between cohorts of Finnish and Dutch patients with subarachnoid hemorrhage: time trends between 1986 and 2005. Stroke 2008;39:3166-3171.
38. Abecasis GR, Auton A, Brooks LD, et al. An integrated map of genetic variation from 1,092 human genomes. Nature 2012;491:56-65.
39. Chibnik LB, Keenan BT, Cui J, et al. Genetic risk score predicting risk of rheumatoid arthritis phenotypes and age of symptom onset. PloS one 2011;6:e24380.
40. Taylor KE, Chung SA, Graham RR, et al. Risk alleles for systemic lupus erythematosus in a large case-control collection and associations with clinical subphenotypes. PLoS genetics 2011;7:e1001311.
41. Raaymakers TW. Aneurysms in relatives of patients with subarachnoid hemorrhage: frequency and risk factors. MARS Study Group. Magnetic Resonance Angiography in Relatives of patients with Subarachnoid hemorrhage. Neurology 1999;53:982-988.
42. Bor AS, Rinkel GJ, van Norden J, Wermer MJ. Long-term, serial screening for intracranial aneurysms in individuals with a family history of

aneurysmal subarachnoid haemorrhage: a cohort study. Lancet neurology 2014;13:385-392.
43. Bor AS, Koffijberg H, Wermer MJ, Rinkel GJ. Optimal screening strategy for familial intracranial aneurysms: a cost-effectiveness analysis. Neurology 2010;74:1671-1679.
44. Bromberg JE, Rinkel GJ, Algra A, et al. Subarachnoid haemorrhage in first and second degree relatives of patients with subarachnoid haemorrhage. BMJ 1995;311:288-289.
45. Wermer MJ, van der Schaaf IC, Velthuis BK, Algra A, Buskens E, Rinkel GJ. Follow-up screening after subarachnoid haemorrhage: frequency and determinants of new aneurysms and enlargement of existing aneurysms. Brain : a journal of neurology 2005;128:2421-2429.
46. R Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.
47. Kent WJ. BLAT--the BLAST-like alignment tool. Genome research 2002;12:656-664.
48. Leek JT, Storey JD. Capturing heterogeneity in gene expression studies by surrogate variable analysis. PLoS genetics 2007;3:1724-1735.
49. Hudson NJ, Dalrymple BP, Reverter A. Beyond differential expression: the quest for causal mutations and effector molecules. BMC genomics 2012;13:356.
50. Hudson NJ, Reverter A, Dalrymple BP. A differential wiring analysis of expression data correctly identifies the gene containing the causal mutation. PLoS computational biology 2009;5:e1000382.
51. Langfelder P, Zhang B, Horvath S. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. Bioinformatics 2008;24:719-720.
52. Tesson BM, Breitling R, Jansen RC. DiffCoEx: a simple and sensitive method to find differentially coexpressed gene modules. BMC bioinformatics 2010;11:497.
53. Langfelder P, Luo R, Oldham MC, Horvath S. Is my network module preserved and reproducible? PLoS computational biology 2011;7:e1001057.
54. Choi Y, Kendziorski C. Statistical methods for gene set co-expression analysis. Bioinformatics 2009;25:2780-2786.
55. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature protocols 2009;4:44-57.
56. Sporns O, Honey CJ, Kotter R. Identification and classification of hubs in brain networks. PloS one 2007;2:e1049.
57. Bullmore E, Sporns O. Complex brain networks: graph theoretical analysis of structural and functional systems. Nature reviews Neuroscience 2009;10:186-198.
58. van den Heuvel MP, Mandl RC, Stam CJ, Kahn RS, Hulshoff Pol HE. Aberrant frontal and temporal complex network structure in schizophrenia: a graph theoretical analysis. The Journal of neuroscience : the official journal of the Society for Neuroscience 2010;30:15915-15926.

59. Tibshirani R, Hastie T, Narasimhan B, Chu G. Diagnosis of multiple cancer types by shrunken centroids of gene expression. Proceedings of the National Academy of Sciences of the United States of America 2002;99:6567-6572.
60. Tomaszewski M, Debiec R, Braund PS, et al. Genetic architecture of ambulatory blood pressure in the general population: insights from cardiovascular gene-centric array. Hypertension 2010;56:1069-1076.
61. Kurki MI, Hakkinen SK, Frosen J, et al. Upregulated signaling pathways in ruptured human saccular intracranial aneurysm wall: an emerging regulative role of Toll-like receptor signaling and nuclear factor-kappaB, hypoxia-inducible factor-1A, and ETS transcription factors. Neurosurgery 2011;68:1667-1675; discussion 1675-1666.
62. Pera J, Korostynski M, Golda S, et al. Gene expression profiling of blood in ruptured intracranial aneurysms: in search of biomarkers. Journal of cerebral blood flow and metabolism : official journal of the International Society of Cerebral Blood Flow and Metabolism 2013;33:1025-1031.
63. Sabatino G, Rigante L, Minella D, et al. Transcriptional profile characterization for the identification of peripheral blood biomarkers in patients with cerebral aneurysms. Journal of biological regulators and homeostatic agents 2013;27:729-738.
64. Giusti B, Rossi L, Lapini I, et al. Gene expression profiling of peripheral blood in patients with abdominal aortic aneurysm. European journal of vascular and endovascular surgery : the official journal of the European Society for Vascular Surgery 2009;38:104-112.
65. Wang Y, Barbacioru CC, Shiffman D, et al. Gene expression signature in peripheral blood detects thoracic aortic aneurysm. PloS one 2007;2:e1050.
66. Aziz H, Zaas A, Ginsburg GS. Peripheral blood gene expression profiling for cardiovascular disease assessment. Genomic medicine 2007;1:105-112.
67. Juvela S, Poussa K, Porras M. Factors affecting formation and growth of intracranial aneurysms: a long-term follow-up study. Stroke; a journal of cerebral circulation 2001;32:485-491.
68. Bor AS, Rinkel GJ, Adami J, et al. Risk of subarachnoid haemorrhage according to number of affected relatives: a population based case-control study. Brain 2008;131:2662-2665.
69. Panoutsopoulou K, Tachmazidou I, Zeggini E. In search of low-frequency and rare variants affecting complex traits. Hum Mol Genet 2013;22:R16-21.
70. Abecasis Lab. Exome Chip Design Wiki Site (http://genome.sph.umich.edu/wiki/Exome_Chip_Design).
71. Guo Y, He J, Zhao S, et al. Illumina human exome genotyping array clustering and quality control. Nature protocols 2014;9:2643-2662.
72. Huyghe JR, Jackson AU, Fogarty MP, et al. Exome array analysis identifies new loci and low-frequency variants influencing insulin processing and secretion. Nat Genet 2013;45:197-201.
73. Seddon JM, Yu Y, Miller EC, et al. Rare variants in CFI, C3 and C9 are associated with high risk of advanced age-related macular degeneration. Nat Genet 2013;45:1366-1370.

74. Holmen OL, Zhang H, Fan Y, et al. Systematic evaluation of coding variation identifies a candidate causal variant in TM6SF2 influencing total cholesterol and myocardial infarction risk. Nat Genet 2014;46:345-351.
75. Service SK, Teslovich TM, Fuchsberger C, et al. Re-sequencing expands our understanding of the phenotypic impact of variants at GWAS loci. PLoS Genet 2014;10:e1004147.
76. Grobbee DE, Hoes AW, Verheij TJ, Schrijvers AJ, van Ameijden EJ, Numans ME. The Utrecht Health Project: optimization of routine healthcare data for research. Eur J Epidemiol 2005;20:285-287.
77. Goldstein JI, Crenshaw A, Carey J, et al. zCall: a rare variant caller for array-based genotyping: genetics and population analysis. Bioinformatics 2012;28:2543-2545.
78. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007;81:559-575.
79. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 2006;38:904-909.
80. Devlin B, Roeder K. Genomic control for association studies. Biometrics 1999;55:997-1004.
81. Exome Aggregation Consortium (ExAC), Cambridge, MA (URL: http://exac.broadinstitute.org) [accessed in September 2015].
82. Lin DY, Tang ZZ. A general framework for detecting disease associations with rare variants in sequencing studies. Am J Hum Genet 2011;89:354-367.
83. Johansen CT, Wang J, Lanktree MB, et al. Excess of rare variants in genes identified by genome-wide association study of hypertriglyceridemia. Nat Genet 2010;42:684-687.
84. Price AL, Kryukov GV, de Bakker PI, et al. Pooled association tests for rare variants in exon-resequencing studies. Am J Hum Genet 2010;86:832-838.
85. Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X. Rare-variant association testing for sequencing data with the sequence kernel association test. Am J Hum Genet 2011;89:82-93.
86. Purcell S, Cherny SS, Sham PC. Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. Bioinformatics 2003;19:149-150.
87. Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature 2016;536:285-291.
88. Pruitt KD, Brown GR, Hiatt SM, et al. RefSeq: an update on mammalian reference sequences. Nucleic acids research 2014;42:D756-763.
89. Chapman SL, Sicot FX, Davis EC, et al. Fibulin-2 and fibulin-5 cooperatively function to form the internal elastic lamina and protect from vascular injury. Arterioscler Thromb Vasc Biol 2010;30:68-74.
90. Kuivaniemi H, Marshall A, Ganguly A, Chu ML, Abbott WM, Tromp G. Fibulin-2 exhibits high degree of variability, but no structural changes concordant with abdominal aortic aneurysms. Eur J Hum Genet 1998;6:642-646.

91. Hasham SN, Willing MC, Guo DC, et al. Mapping a locus for familial thoracic aortic aneurysms and dissections (TAAD2) to 3p24-25. Circulation 2003;107:3184-3190.
92. Radmanesh F, Falcone GJ, Anderson CD, et al. Rare Coding Variation and Risk of Intracerebral Hemorrhage. Stroke 2015;46:2299-2301.
93. Liu C, Kraja AT, Smith JA, et al. Meta-analysis identifies common and rare variants influencing blood pressure and overlapping with metabolic trait loci. Nat Genet 2016;48:1162-1170.
94. Kloek AT, van Setten J, van der Ende A, et al. Exome Array Analysis of Susceptibility to Pneumococcal Meningitis. Scientific reports 2016;6:29351.
95. Zheng HF, Forgetta V, Hsu YH, et al. Whole-genome sequencing identifies EN1 as a determinant of bone density and fracture. Nature 2015.
96. Sakalihasan N, Limet R, Defawe OD. Abdominal aortic aneurysm. Lancet 2005;365:1577-1589.
97. Chau KH, Elefteriades JA. Natural history of thoracic aortic aneurysms: size matters, plus moving beyond size. Progress in cardiovascular diseases 2013;56:74-80.
98. Bown MJ, Jones GT, Harrison SC, et al. Abdominal aortic aneurysm is associated with a variant in low-density lipoprotein receptor-related protein 1. Am J Hum Genet 2011;89:619-627.
99. Gretarsdottir S, Baas AF, Thorleifsson G, et al. Genome-wide association study identifies a sequence variant within the DAB2IP gene conferring susceptibility to abdominal aortic aneurysm. Nat Genet 2010;42:692-697.
100. Bradley DT, Hughes AE, Badger SA, et al. A variant in LDLR is associated with abdominal aortic aneurysm. Circulation Cardiovascular genetics 2013;6:498-504.
101. Jones GT, Bown MJ, Gretarsdottir S, et al. A sequence variant associated with sortilin-1 (SORT1) on 1p13.3 is independently associated with abdominal aortic aneurysm. Hum Mol Genet 2013;22:2941-2947.
102. Lemaire SA, McDonald ML, Guo DC, et al. Genome-wide association study identifies a susceptibility locus for thoracic aortic aneurysms and aortic dissections spanning FBN1 at 15q21.1. Nat Genet 2011;43:996-1000.
103. Kiemeney LA, Thorlacius S, Sulem P, et al. Sequence variant on 8q24 confers susceptibility to urinary bladder cancer. Nat Genet 2008;40:1307-1312.
104. Wetzels JF, Kiemeney LA, Swinkels DW, Willems HL, den Heijer M. Age- and gender-specific reference values of estimated GFR in Caucasians: the Nijmegen Biomedical Study. Kidney Int 2007;72:632-637.
105. Barker DJ, Osmond C, Forsen TJ, Kajantie E, Eriksson JG. Trajectories of growth among children who have coronary events as adults. N Engl J Med 2005;353:1802-1809.
106. Aromaa A, Koskinen S. HEALTH AND FUNCTIONAL CAPACITY IN FINLAND. Baseline Results of the Health 2000 Health Examination Survey. Publications of the National Public Health Institute 2004.
107. Hofman A, Darwish Murad S, van Duijn CM, et al. The Rotterdam Study: 2014 objectives and design update. Eur J Epidemiol 2013;28:889-926.

108. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007;81:559-575.
109. Howie B, Marchini J, Stephens M. Genotype imputation with thousands of genomes. G3 (Bethesda) 2011;1:457-470.
110. Delaneau O, Marchini J, Zagury JF. A linear complexity phasing method for thousands of genomes. Nature methods 2012;9:179-181.
111. Purcell SM, Wray NR, Stone JL, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature 2009;460:748-752.
112. Bulik-Sullivan BK, Loh PR, Finucane HK, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet 2015;47:291-295.
113. Bulik-Sullivan B, Finucane HK, Anttila V, et al. An atlas of genetic correlations across human diseases and traits. Nat Genet 2015;47:1236-1241.
114. Smelser DT, Tromp G, Elmore JR, et al. Population risk factor estimates for abdominal aortic aneurysm from electronic medical records: a case control study. BMC cardiovascular disorders 2014;14:174.
115. Verma SS, de Andrade M, Tromp G, et al. Imputation and quality control steps for combining multiple genome-wide datasets. Frontiers in genetics 2014;5:370.
116. de Bakker PI, Ferreira MA, Jia X, Neale BM, Raychaudhuri S, Voight BF. Practical aspects of imputation-driven meta-analysis of genome-wide association studies. Hum Mol Genet 2008;17:R122-128.
117. Deloukas P, Kanoni S, Willenborg C, et al. Large-scale association analysis identifies new risk loci for coronary artery disease. Nat Genet 2013;45:25-33.
118. Penn DL, Witte SR, Komotar RJ, Sander Connolly E, Jr. The role of vascular remodeling and inflammation in the pathogenesis of intracranial aneurysms. Journal of clinical neuroscience : official journal of the Neurosurgical Society of Australasia 2014;21:28-32.
119. Guo DC, Papke CL, He R, Milewicz DM. Pathogenesis of thoracic and abdominal aortic aneurysms. Ann N Y Acad Sci 2006;1085:339-352.
120. Norman PE, Powell JT. Site specificity of aneurysmal disease. Circulation 2010;121:560-568.
121. Tromp G, Kuivaniemi H, Hinterseher I, Carey DJ. Novel genetic mechanisms for aortic aneurysms. Current atherosclerosis reports 2010;12:259-266.
122. Worrall BB, Foroud T, Brown RD, Jr., et al. Genome screen to detect linkage to common susceptibility genes for intracranial and aortic aneurysms. Stroke 2009;40:71-76.
123. Ruigrok YM, Elias R, Wijmenga C, Rinkel GJ. A comparison of genetic chromosomal loci for intracranial, thoracic aortic, and abdominal aortic aneurysms in search of common genetic risk factors. Cardiovascular pathology : the official journal of the Society for Cardiovascular Pathology 2008;17:40-47.
124. Nordon IM, Hinchliffe RJ, Loftus IM, Thompson MM. Pathophysiology and epidemiology of abdominal aortic aneurysms. Nat Rev Cardiol 2011;8:92-102.

125. Larsson E, Granath F, Swedenborg J, Hultgren R. A population-based case-control study of the familial risk of abdominal aortic aneurysm. J Vasc Surg 2009;49:47-50; discussion 51.
126. Ogata T, MacKean GL, Cole CW, et al. The lifetime prevalence of abdominal aortic aneurysms among siblings of aneurysm patients is eightfold higher than among siblings of spouses: an analysis of 187 aneurysm families in Nova Scotia, Canada. J Vasc Surg 2005;42:891-897.
127. Forsdahl SH, Singh K, Solberg S, Jacobsen BK. Risk factors for abdominal aortic aneurysms: a 7-year prospective study: the Tromso Study, 1994-2001. Circulation 2009;119:2202-2208.
128. Hobbs SD, Claridge MW, Quick CR, Day NE, Bradbury AW, Wilmink AB. LDL cholesterol is associated with small abdominal aortic aneurysms. Eur J Vasc Endovasc Surg 2003;26:618-622.
129. Reed D, Reed C, Stemmermann G, Hayashi T. Are aortic aneurysms caused by atherosclerosis? Circulation 1992;85:205-211.
130. Golledge J, van Bockxmeer F, Jamrozik K, McCann M, Norman PE. Association between serum lipoproteins and abdominal aortic aneurysm. Am J Cardiol 2010;105:1480-1484.
131. Cornuz J, Sidoti Pinto C, Tevaearai H, Egger M. Risk factors for asymptomatic abdominal aortic aneurysm: systematic review and meta-analysis of population-based screening studies. Eur J Public Health 2004;14:343-349.
132. Lederle FA, Johnson GR, Wilson SE, et al. Prevalence and associations of abdominal aortic aneurysm detected through screening. Aneurysm Detection and Management (ADAM) Veterans Affairs Cooperative Study Group. Ann Intern Med 1997;126:441-449.
133. Nguyen TV, Chandrashekar K, Qin Z, Parent AD, Zhang J. Epidemiology of intracranial aneurysms of Mississippi: a 10-year (1997-2007) retrospective study. Journal of stroke and cerebrovascular diseases : the official journal of National Stroke Association 2009;18:374-380.
134. Smith GD, Lawlor DA, Harbord R, Timpson N, Day I, Ebrahim S. Clustered environments and randomized genes: a fundamental distinction between conventional and genetic epidemiology. PLoS medicine 2007;4:e352.
135. Hofman A, van Duijn CM, Franco OH, et al. The Rotterdam Study: 2012 objectives and design update. Eur J Epidemiol 2011;26:657-686.
136. International HapMap Consortium. A second generation human haplotype map of over 3.1 million SNPs. Nature 2007;449:851-861.
137. Browning BL, Browning SR. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. Am J Hum Genet 2009;84:210-223.
138. Teslovich TM, Musunuru K, Smith AV, et al. Biological, clinical and population relevance of 95 loci for blood lipids. Nature 2010;466:707-713.
139. International Consortium for Blood Pressure Genome-Wide Association Studies. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. Nature 2011;478:103-109.
140. Global BPgen consortium. Genome-wide association study identifies eight loci associated with blood pressure. Nat Genet 2009;41:666-676.

141. International Consortium for Blood Pressure Genome-Wide Association Studies. Genome-wide association study identifies six new loci influencing pulse pressure and mean arterial pressure. Nat Genet 2011;43:1005-1011.
142. CARDIoGRAM Consortium. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. Nat Genet 2011;43:333-338.
143. Coronary Artery Disease (C4D) Genetics Consortium. A genome-wide association study in Europeans and South Asians identifies five new loci for coronary artery disease. Nat Genet 2011;43:339-344.
144. IL6R Genetics Consortium Emerging Risk Factors Collaboration. Interleukin-6 receptor pathways in coronary heart disease: a collaborative meta-analysis of 82 studies. Lancet 2012;379:1205-1213.
145. GIANT Consortium. Hundreds of variants clustered in genomic loci and biological pathways affect human height. Nature 2010;467:832-838.
146. Johnson AD, Handsaker RE, Pulit SL, Nizzari MM, O'Donnell CJ, de Bakker PI. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. Bioinformatics 2008;24:2938-2939.
147. Hobbs HH, Brown MS, Goldstein JL. Molecular genetics of the LDL receptor gene in familial hypercholesterolemia. Hum Mutat 1992;1:445-466.
148. Freiberg MS, Arnold AM, Newman AB, Edwards MS, Kraemer KL, Kuller LH. Abdominal aortic aneurysms, increasing infrarenal aortic diameter, and risk of total mortality and incident cardiovascular disease events: 10-year follow-up data from the Cardiovascular Health Study. Circulation 2008;117:1010-1017.
149. Nieuwkamp DJ, Algra A, Blomqvist P, et al. Excess mortality and cardiovascular events in patients surviving subarachnoid hemorrhage: a nationwide study in Sweden. Stroke 2011;42:902-907.
150. Gaal EI, Salo P, Kristiansson K, et al. Intracranial Aneurysm Risk Locus 5q23.2 Is Associated with Elevated Systolic Blood Pressure. PLoS Genet 2012;8:e1002563.
151. Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? Int J Epidemiol 2003;32:1-22.
152. van 't Hof FN, Ruigrok YM, Baas AF, et al. Impact of inherited genetic variants associated with lipid profile, hypertension, and coronary artery disease on the risk of intracranial and abdominal aortic aneurysms. Circulation Cardiovascular genetics 2013;6:264-270.
153. Inagawa T. Risk factors for aneurysmal subarachnoid hemorrhage in patients in Izumo City, Japan. J Neurosurg 2005;102:60-67.
154. Shah AD, Langenberg C, Rapsomaniki E, et al. Type 2 diabetes and incidence of cardiovascular diseases: a cohort study in 1.9 million people. The lancet Diabetes & endocrinology 2015;3:105-113.
155. Ohrlander T, Merlo J, Ohlsson H, Sonesson B, Acosta S. Socioeconomic position, comorbidity, and mortality in aortic aneurysms: a 13-year prospective cohort study. Annals of vascular surgery 2012;26:312-321.
156. Shantikumar S, Ajjan R, Porter KE, Scott DJ. Diabetes and the abdominal aortic aneurysm. Eur J Vasc Endovasc Surg 2010;39:200-207.

157. Vlak MH, Rinkel GJ, Greebe P, Algra A. Independent risk factors for intracranial aneurysms and their joint effect: a case-control study. Stroke 2013;44:984-987.
158. Kent KC, Zwolak RM, Egorova NN, et al. Analysis of risk factors for abdominal aortic aneurysm in a cohort of more than 3 million individuals. J Vasc Surg 2010;52:539-548.
159. Stackelberg O, Bjorck M, Sadr-Azodi O, Larsson SC, Orsini N, Wolk A. Obesity and abdominal aortic aneurysm. Br J Surg 2013;100:360-366.
160. Abecasis GR, Altshuler D, Auton A, et al. A map of human genome variation from population-scale sequencing. Nature 2010;467:1061-1073.
161. Morris AP, Voight BF, Teslovich TM, et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. Nat Genet 2012;44:981-990.
162. Locke AE, Kahali B, Berndt SI, et al. Genetic studies of body mass index yield new insights for obesity biology. Nature 2015;518:197-206.
163. Shungin D, Winkler TW, Croteau-Chonka DC, et al. New genetic loci link adipose and insulin biology to body fat distribution. Nature 2015;518:187-196.
164. Smith GD, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? Int J Epidemiol 2003;32:1-22.
165. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. Int J Epidemiol 2015;44:512-525.
166. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. Genetic epidemiology 2016;40:304-314.
167. Bowden J, Del Greco MF, Minelli C, Davey Smith G, Sheehan NA, Thompson JR. Assessing the suitability of summary data for two-sample Mendelian randomization analyses using MR-Egger regression: the role of the I2 statistic. Int J Epidemiol 2016.
168. Rodin MB, Daviglus ML, Wong GC, et al. Middle age cardiovascular risk factors and abdominal aortic aneurysm in older age. Hypertension 2003;42:61-68.
169. Heid IM, Jackson AU, Randall JC, et al. Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. Nat Genet 2010;42:949-960.
170. Aschard H, Vilhjalmsson BJ, Joshi AD, Price AL, Kraft P. Adjusting for heritable covariates can bias effect estimates in genome-wide association studies. Am J Hum Genet 2015;96:329-339.
171. Burgess S, Thompson SG. Multivariable Mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects. Am J Epidemiol 2015;181:251-260.
172. Rinkel GJ, Djibuti M, Algra A, van Gijn J. Prevalence and risk of rupture of intracranial aneurysms: a systematic review. Stroke 1998;29:251-256.
173. Krings T, Mandell DM, Kiehl TR, et al. Intracranial aneurysms: from vessel wall pathology to therapeutic approach. Nature reviews Neurology 2011;7:547-559.

174. Kosmicki JA, Churchhouse CL, Rivas MA, Neale BM. Discovery of rare variants for complex phenotypes. Hum Genet 2016;135:625-634.
175. Yang J, Manolio TA, Pasquale LR, et al. Genome partitioning of genetic variation for complex traits using common SNPs. Nat Genet 2011;43:519-525.
176. Visscher PM, Brown MA, McCarthy MI, Yang J. Five years of GWAS discovery. Am J Hum Genet 2012;90:7-24.
177. Fuchsberger C, Flannick J, Teslovich TM, et al. The genetic architecture of type 2 diabetes. Nature 2016;536:41-47.
178. Dewey FE, Murray MF, Overton JD, et al. Distribution and clinical impact of functional variants in 50,726 whole-exome sequences from the DiscovEHR study. Science 2016;354.
179. Farlow JL, Lin H, Sauerbeck L, et al. Lessons learned from whole exome sequencing in multiplex families affected by a complex genetic disorder, intracranial aneurysm. PLoS One 2015;10:e0121104.
180. Yan J, Hitomi T, Takenaka K, et al. Genetic study of intracranial aneurysms. Stroke 2015;46:620-626.
181. Zhou S, Ambalavanan A, Rochefort D, et al. RNF213 Is Associated with Intracranial Aneurysms in the French-Canadian Population. Am J Hum Genet 2016;99:1072-1085.
182. Santiago-Sim T, Fang X, Hennessy ML, et al. THSD1 (Thrombospondin Type 1 Domain Containing Protein 1) Mutation in the Pathogenesis of Intracranial Aneurysm and Subarachnoid Hemorrhage. Stroke 2016;47:3005-3013.
183. Renton AE, Majounie E, Waite A, et al. A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. Neuron 2011;72:257-268.
184. Brisman JL, Song JK, Newell DW. Cerebral aneurysms. N Engl J Med 2006;355:928-939.
185. Mele M, Ferreira PG, Reverter F, et al. Human genomics. The human transcriptome across tissues and individuals. Science 2015;348:660-665.
186. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science 2015;348:648-660.
187. Bartel DP. MicroRNAs: target recognition and regulatory functions. Cell 2009;136:215-233.
188. Abifadel M, Varret M, Rabes JP, et al. Mutations in PCSK9 cause autosomal dominant hypercholesterolemia. Nat Genet 2003;34:154-156.
189. Cohen JC, Boerwinkle E, Mosley TH, Jr., Hobbs HH. Sequence variations in PCSK9, low LDL, and protection against coronary heart disease. N Engl J Med 2006;354:1264-1272.
190. Roberts R. PCSK9 inhibition--a new thrust in the prevention of heart disease: genetics does it again. The Canadian journal of cardiology 2013;29:899-901.
191. Kataoka H. Molecular mechanisms of the formation and progression of intracranial aneurysms. Neurol Med Chir (Tokyo) 2015;55:214-229.
192. Gaulton KJ, Ferreira T, Lee Y, et al. Genetic fine mapping and genomic annotation defines causal mechanisms at type 2 diabetes susceptibility loci. Nat Genet 2015;47:1415-1425.

193. Peymani A, Adams HH, Cremers LG, et al. Genetic Determinants of Unruptured Intracranial Aneurysms in the General Population. Stroke 2015;46:2961-2964.
194. Mega JL, Stitziel NO, Smith JG, et al. Genetic risk, coronary heart disease events, and the clinical benefit of statin therapy: an analysis of primary and secondary prevention trials. Lancet 2015;385:2264-2271.