Title: Integrative genomics analysis implicates decreased *FGD6* expression underlying risk of intracranial aneurysm rupture

Authors: Andrew T. Hale MD, PhD, Jing He MS, and Jesse Jones MD

Affiliations: 1Department of Neurosurgery, University of Alabama at Birmingham, Birmingham, AL; 2Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN.

Corresponding Author:
Andrew T Hale, MD, PhD
Department of Neurosurgery
University of Alabama at Birmingham
FOT Suite 1060, 1720 2nd Ave S.
Birmingham, AL 35294
Andrewthale20@gmail.com

The authors have declared that no conflict of interest exists.

Word count: 2,348

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
Abstract

Importance: The genetic determinants and mechanisms underlying intracranial aneurysm rupture (rIA) are largely unknown. Given the ~50% mortality rate of rIA, approaches to identify patients at high-risk will inform screening, diagnostic, and preventative measures.

Objective: To identify and characterize the genetic basis of rIA.

Design: We perform a genome-wide association study (GWAS) use functional genomics approaches to identify and characterize rIA-associated loci and genes.

Setting: We perform a meta-analysis across 24 published GWAS of rIA.

Participants: Our cohort contains 84,353 individuals (7,843 rIA cases and 76,510 controls).

Main Outcome and Measure: Single nucleotide polymorphisms (SNP), gene-burden analysis, and functional genomics identify and characterize genetic risk factors for rIA.

Results: We identify 5 independent genetic loci reaching genome-wide significance (p<5.0x10^{-8}) for rIA including rs12310399 (FGD6, OR=1.16), which to our knowledge, has not been implicated in prior GWAS of (r)IA. We then quantified gene-level mutation-burden across ~20,000 genes, and only FGD6 (containing 21 rIA-associated SNPs) reached transcriptome-wide significance. Expression quantitative trait loci (eQTL) mapping indicates that rs12310399 causes decreased FGD6 gene expression in arterial tissue. Single-cell RNA sequencing of normal human cerebrovascular cells obtained during resection surgery identified high expression of FGD6 in 1 of 3 arterial lineages but absent in perivascular cells. Patients with a Mendelian disease caused by FGD6 mutations (Aarskog-Scott syndrome) have been diagnosed with rIA, among other features, providing direct causal evidence for FGD6 in rIA pathogenesis.

Conclusions and Relevance: We identify and characterize a previously unknown risk loci for rIA containing FGD6. Variable expressivity of FGD6 causes a range of clinical features from Mendelian disease to sporadic rIA. Elucidation of high-risk genetic loci may instruct population-genetic screening and clinical-genetic testing strategies to identify patients predisposed to rIA.

Funding: No funding sources were used for the material presented herein.
Key Points

Question: What are the genetic determinants of intracranial aneurysm rupture?

Findings: We identify FGD6 as a new risk locus for rIA using GWAS and gene-burden approaches. Characterization of the FGD6 locus and gene-level function using integrative genomics elucidates consequences of decreased FGD6 expression.

Meaning: Elucidating the genetic risk factors for rIA will instruct population-screening, genetic testing, and preventative measures aimed at reducing morbidity caused by rIA.
Introduction

Intracranial aneurysms (IA) are abnormal dilations of cerebral arteries that occur in response to vessel wall dysfunction. Spontaneous rupture of an IA (rIA) causes severe neurological injury and death.¹² IA is estimated to affect around 5% of the world’s population with lifetime rupture-risk of ≈25%.³ However, the mortality of rIA is ≈50%.⁴ While many rIA risk-factors including family history, female gender, aneurysm location/size, cigarette smoking, and hypertension have been elucidated in large epidemiological studies,⁵ these factors are largely non-specific. Thus, personalized genetic approaches to 1) identify patients with IA, and 2) predict which patients with IA are most likely to rupture are needed. Elucidating causal genetic factors associated with rIA has direct implications for disease prevention, diagnosis, and treatment. While rare inherited causes of IA have been described,⁶ these patients comprise the minority of total IA disease burden.⁷ Family history of rIA in first-degree relatives, even in the absence of a known IA-associated genetic syndrome, increases one’s risk of being diagnosed with rIA by nearly four-fold.⁸ Thus, a strong genetic component is thought to underlie IA,⁹ and unbiased, genome-wide approaches to understand the germline-genetic determinants of IA formation and rupture are critical to understanding IA pathophysiology. Since one’s genetic information in large part does not change over a lifetime (excluding somatic mutations), elucidating the genetic architecture of rIA can inform screening, detection, and treatment (for patients at the highest risk) strategies. In this study, we systematically delineate the genetic determinants of rIA using convergent human-genetic and functional genomics approaches.

Results

Our cohort contained 84,353 individuals (7,843 rIA cases and 76,510 controls) across 24 published GWAS.¹⁰ Our GWAS identified 5 independent genetic loci reaching genome-wide significance for rIA (Figure 1A): rs73392700 (CDKN2B, p=4.84x10⁻¹⁷, OR=1.12), rs6841581 (EDNRA, p=3.52x10⁻¹⁴, OR=0.77), rs11661542 (RBBP8, p=3.18x10⁻¹³, OR=0.84), rs62516550 (RP1, p=2.94x10⁻¹¹, OR=1.20), and rs12310399 (FGD6, p=3.19x10⁻¹⁰, OR=1.16). The FGD6 loci, to our knowledge, has not been implicated in prior GWAS of (r)IA. We confirmed that population stratification and genomic control was
adequately achieved (i.e., no early deviation of p-values from expected) by generating a quantile-quantile plot (Figure 1B).

Consistent with GWAS studies of complex and polygenic diseases, the vast majority of identified genetic risk factors are located in non-coding regions of the genome, making causal inference of disease mechanisms challenging. To overcome these challenges and delineate the functional consequences of rIA-associated single-nucleotide polymorphisms (SNPs), we analyzed the summative contribution of rIA associated SNPs across ~20,000 individual genes using gene-burden analysis. Determining gene-based, rather than variant-based, associations with rIA will facilitate ease of biologic interpretation and design/interpretation of mechanistic studies. Across all rIA-associated SNPs and genes tested, 958 genes were associated with rIA (p< 0.05). However, after multiple-testing correction, only FGD6 (harboring 21 rIA-associated SNPs) reached transcriptome-wide significance.

To visualize the FGD6 locus for rIA-associated SNPs, we perform fine-mapping and present a LocusZoom plot (Figure 2). These data demonstrate the extent to which multiple variants in this locus, alone and in concert through linkage disequilibrium (LD), contribute to rIA risk. FGD6 (FYVE RhoGef and Ph domain containing protein 6) is a RhoGTPase that activates CDC42, a cell cycle-associated gene, and FGD6 has been shown to play a role in regulating angiogenesis and actin cytoskeleton rearrangement, functions consistent with (r)IA. Finally, mutations in FGD6 are associated with Aarskog-Scott Syndrome, a Mendelian disease characterized by faciogential dysplasia, where some patients have also been diagnosed with to (r)IA and other cerebrovascular anomalies. These data provide direct causal evidence for FGD6 dysfunction in rIA.

To delineate the effect of rIA-associated SNPs on gene expression, we utilize eQTL mapping across 49 tissue types using the Genotype-Tissue Expression Project (GTEx) portal and demonstrate that the most significant rIA-associated SNP in the FGD6 loci (rs12310399) causes decreased FGD6 gene expression most-significantly in arterial tissue (Figure 3). In addition, rs12310399 overlaps splicing QTLs for vezatin (VEZT, p< 8.20x10^{-9}, skeletal muscle, gtexportal.org), a transmembrane protein in the adherens junction that binds to myosin VI (Myo6) and NADH ubiquinone oxidoreductase subunit A12 (NDUFA12, p< 8.20x10^{-9}, skeletal muscle, gtexportal.org), a gene previously implicated in IA biology.
These data provide additional support for this locus and suggests that multiple mechanisms are likely contributing to rIA risk. While identification of new rIA-associated genes is important, it is the function of that gene in relevant tissues/cells that will expand our pathophysiologic understanding of IA formation and rupture. To further delineate the cellular origin and functional consequences of \textit{FGD6} expression in the cerebrovascular system, we utilize scRNA-seq data derived from normal adult human brain endothelial and perivascular cells as previously described.\textsuperscript{19} We find that \textit{FGD6} is highly expressed in 1 of 3 arterial lineages, venules, and veins but \textit{FGD6} expression is largely undetectable across 12 perivascular cell lineages (Figure 5A-B). These findings will facilitate additional mechanistic studies aimed at understanding the molecular basis of \textit{FGD6} dysfunction underlying rIA.

\textbf{Discussion}

We identify 5 independent genetic loci reaching genome-wide significance for rIA, including a novel locus containing \textit{FGD6} not previously implicated in (r)IA biology. Gene-burden analysis provides convergent evidence supporting the \textit{FGD6} gene in rIA risk. Functional characterization indicates that genetic variation in \textit{FGD6} loci contributes to rIA risk by decreasing \textit{FGD6} gene expression in arterial tissue. Utilization of single-cell RNA sequencing analysis of human brain endothelial and perivascular cells\textsuperscript{19} characterizes \textit{FGD6} expression across rIA-relevant neurovascular cell types where alterations in \textit{FGD6} expression may alter cell identity and function. These data will inform detailed mechanistic analysis of \textit{FGD6} in rIA biology and perhaps even treatment strategies. Finally, analysis of the Online Mendelian Inheritance in Man database (omim.org) identifies a Mendelian disease (Aarskog-Scott syndrome) caused by \textit{FGD6} loss-of-function, in whom rIA has been described, providing causal evidence for \textit{FGD6} dysfunction in rIA pathogenesis. These data suggest that variable expressivity of \textit{FDG6} causes a range of clinical features from Mendelian disease to sporadic rIA.

Functional analysis indicates that \textit{FGD6} regulates angiogenesis and actin cytoskeleton rearrangement.\textsuperscript{13} Thus, it is reasonable to posit that \textit{FGD6}-dependent alterations in blood vessel integrity and/or formation may, at least in part, contribute to rIA. \textit{FGD6} mediated genetic susceptibility coupled with clinical risk factors (hypertension, smoking, etc.) converge to contribute summative risk for rIA.
Whole-gene deletion or introduction of FGD6 disease-causing mutations disrupts endothelial cell network formation which could be rescued by introduction of the wild-type gene product. In addition, mutations in FGD6 have been previously identified in patients with a Mendelian disease Aarskog-Scott syndrome, of which rIA is a component, as well as in patients with polypoidal choroidal vasculopathy.13-15 These data provide causal evidence for FGD6 dysfunction in rIA pathology and regulation of the human cerebrovascular system. These data add to the growing body of evidence that Mendelian genes influence common diseases20 likely through a variable expressivity mechanism.

While the focus of this manuscript is on identification and characterization of the FGD6 loci, multiple genetic loci are contributing to rIA (Figure 1). Although the magnitude of effect is relatively small for the lead SNP in the FGD6 loci (OR= 1.16), the clinical consequences of genetic-testing underlying rIA are more likely to be representing through assessment of polygenic risk score (PRS).21 The weighted contribution of multiple loci through creation of a PRS will improve our understanding of rIA genetic architecture.22 In addition, combining PRS with important clinical and radiologic features of patients with (r)IA will improve translational of these genetic findings to clinical care. As our ability to acquire, analyze, and interpret genetic information improves, we surmise that translating genetic information into clinical practice will become streamlined.

Our cohort is the largest studies of rIA to date, but additional genetic data from diverse populations will increase power to detect additional risk loci and improve applicability of the genetic information across diverse populations. One limitation of array based GWAS is our relative inability to include rare variants, although detailed analysis of IA rare variants has been performed elsewhere.23,24 Multiple studies of (r)IA have implicated common variants underling IA risk.10,18,25-30 Thus, it is likely that both common and rare variants in both coding and non-coding regions confer risk to IA, alone and in concert. In addition, recent studies suggest that somatic mutations may play a role in IA pathology,31,32 which we did not assess herein, but may play a significant role in (r)IA pathogenesis. Analysis of lesional tissue acquired through advances in endoluminal biopsy will promote our understanding of somatic contribution to (r)IA biology.33 Integration of clinical and aneurysm-specific factors (size, location, etc.) will enable personalized medicine approaches to IA treatment and expand the clinician’s armamentarium.
Here we perform a large genetic study of rIA and identify and characterize a previously unknown risk locus for rIA containing FGD6. Since the genetic contribution to rIA risk does not change (excluding somatic mutations), approaches to quantify genetic risk of rIA may be extremely valuable for clinical treatment. Integration of clinical and radiologic IA factors with genetic data will further improve our ability to comprehensively evaluate rIA risk. Understanding the genetic determinants of rIA may inform gene-based therapies aimed at IA remodeling or obliteration. Strategies to incorporate genetic information into clinical care will inform personalized medicine approaches to manage IA.

**Methods**

*Genome wide association study analysis*

Genetic data were obtained using the Cerebrovascular Disease Knowledge Portal (cerebrovascularportal.org). GWAS was performed as previously described. Whole-genome genotyping was performed using Illumina HumanOmniExpressExome BeadChip. Imputation was performed using the 1000 genome reference panel, Haplotype Reference Consortium, or the UK10K using IMPUTE4 software. Meta-analysis was performed using inverse-variance method and Cochran’s Q test (to estimate heterogeneity) with METAL. Age, sex, and the first twenty principal components were included as covariates in logistic regression models to identify IA-associated SNPs as previously described. Threshold for genome-wide significance was defined after Bonferroni correction for the total number of SNPs tested \( (p \leq 5.0 \times 10^{-8}) \). Our cohort contained patients of both European and East-Asian ancestry. Assuming a disease prevalence for IA of 100 in 1,000, \( rs73392700 \) allele frequency of 0.12 (the most statistically-significant SNP identified in our GWAS), and genotype relative risk of 1.5, our study is appropriately powered to 1.0 (GAS power calculator from the University of Michigan, csg.sph.umich.edu). STREGA reporting guidelines for genetic association studies were followed.

*Functional genomics analysis*

The latest data release of GTEx (version 8) was used. The GTEx portal (gtexportal.org) was used for data visualization. Expression quantitative trait loci (eQTL) analysis, which estimates the functional effect of a SNP/loci on neighboring gene expression, was performed as previously
Splicing quantitative trait loci (sQTL) was used to determine the potential impact of a SNP on alternative splicing.\footnote{34,43} In addition, we used Haploreg,\footnote{44} a functional genomics tool to explore the function of variants in non-coding regions of the genome. Haploreg combines data from the 1000 genomes project, Roadmap Epigenomics, and ENCODE projects to estimate the effect of SNPs on functions such as chromatin state, protein-binding, and gene-expression. Data visualization was performed using gtexportal.org. Finally, we utilized single-cell RNA sequencing analysis of adult human brain endothelial and perivascular cells obtained from normal brain cortex biopsies obtained as previously described.\footnote{19} 

**Statistics**

Experiment-specific statistical methods are discussed above. Thresholds, controls, and statistical tests are indicated in each Figure legend.

**Study approval**

All genetic and functional genomics data were acquired through publicly available sources, where individual data-sharing policies and waivers of informed consent were obtained.

**Author contributions**

ATH, JH, and JJ were responsible for project design and analysis. ATH drafted the manuscript. ATH and JH were responsible for statistical methods and analysis of the study. ATH, JH, and JJ interpreted results. All authors engaged in critical review of the manuscript.

**Acknowledgements**

The authors would like to thank the creators of the Cerebrovascular Disease Knowledge Portal. In addition, we would like to thank GTEx which is supported by the Common Fund of the Office of the Director of the NIH, the National Cancer Institute (NCI), the National Human Genome Research Institute (NHGRI), the National Heart, Lung, and Blood Institute (NHLBI), the National Institute on Drug Abuse (NIDA), the National Institute of Mental Health (NIMH), and the National Institute of Neurological
Disorders and Stroke (NINDS). We would also like to thank Haploreg (Broad Institute) - which also utilizes Roadmap Epigenomics, 1000 Genomes Project, and ENCODE data - for use of their resource.
References


Figure 1. Genome-wide association study (GWAS) of ruptured intracranial aneurysms (rIA). (A) Manhattan plot of our rIA GWAS containing 84,353 individuals (7,843 rIA cases and 76,510 controls) and identify 5 independent genetic loci reaching genome-wide significance (p<5.0x10^-8) for rIA: rs73392700 (CDKN2B, OR=1.12), rs6841581 (EDNRA, OR=0.77), rs11661542 (RBBP8, OR=0.84), rs62516550 (RP1, OR=1.20), and rs12310399 (FGD6, OR=1.16). The horizontal red line delineates the threshold for genome-wide significance (p = 5.0x10^-8). (B) Quantile-Quantile (Q-Q) plot demonstrating adequate genomic control for population stratification. These data are visualized using the cd.hugemapt.org.
Figure 2. LocusZoom plot after fine-mapping of rIA-associated SNPs in the FGD6 locus.
Figure 3. Expression quantitative trait loci (eQTL) analysis of the *FGD6* locus sentinel variant (rs12310399) across 49 tissues in GTEx. NES = normalized enrichment score. The m-value is the posterior probability from MetaSoft which is the probability than an eQTL effect exists in the tested cross-tissue analysis 46,47. A m-value < 0.01 is highly suggestive that an eQTL effect does not exist, whereas a m-value > 0.9 is highly predictive of an eQTL effect.
Figure 4. Functional mapping of the *FGD6* locus in arterial tissue. Expression quantitative trait loci (eQTL), splicing quantitative trait loci (sQTL), and linkage disequilibrium (LD) mapping of the FGD6 locus. Abbreviations are as follows: Transcription end site (TES), Transcription start site (TSS), Aortic artery (ARTORT), Tibial artery (ARTTBL). Data were visualized using gtexportal.org.
Figure 5. Single-cell RNA sequencing (scRNA-seq) analysis of FGD6 across adult human brain (A) endothelial and (B) perivascular cells obtained from normal resected brain tissue. Endothelial cells were characterized into 3 arterial lineages (Art1-3), capillary (cap), venous, and venule cell-lineages. Blue dots indicate undetectable expression of FGD6, and increasingly dark shades of orange indicate higher expression of FGD6. Perivascular cells were partitioned into 7 smooth muscle (SMC1-7), 2 fibroblast (FB), 2 fibromyocyte (FbM1-2), and 1 pericyte (PC) lineage. Green dots indicate undetectable expression of FGD6 and red dots indicate higher expression of FGD6. UMAP data visualization and gene-expression markers for cell populations was performed using the UCSC Cell Browser.