Pharmacogenomics of Coronary Artery Response to Intravenous Gamma Globulin in Kawasaki Disease

Short Title: Pharmacogenomics of KD-related Coronary Artery Aneurysm

Sadeep Shrestha¹, Howard W Wiener¹, Sabrina Chowdhury¹, Hidemi Kajimoto², Vinodh Srinivasasainagendra³, Olga A Mamaeva¹, Ujval N Brahmbhatt¹, Dolen Ledee², Yung Lau⁴, Luz A Padilla¹, Jake Chen⁵, Nagib Dahdah⁵, Hemant K Tiwari², Michael A Portman²

¹Department of Epidemiology, School of Public Health, University of Alabama at Birmingham; ²Division of Cardiology, Seattle Children’s and University of Washington Department of Pediatrics; ³Department of Biostatistics, School of Public Health, University of Alabama at Birmingham; ⁴Division of Pediatric Cardiology, Children’s of Alabama; ⁵Department of Biomedical Informatics and Data Science, Heersink School of Medicine, University of Alabama at Birmingham; ⁶CHU Ste-Justine, Universite de Montreal

Corresponding Author
Sadeep Shrestha, PhD
Professor,
Department of Epidemiology, RPHB 217L
University of Alabama at Birmingham
sshrestha@uab.edu

Total Word Count:

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.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

AUC: Area Under the Curve
CAA: Coronary Artery Aneurysm
CAA/L: Large Coronary Artery Aneurysm
CADD: Combined Annotation-Dependent Depletion
DNA: Deoxyribonucleic Acid
eQTL: Expression Quantitative Trait Loci
FUMA: Functional Mapping and Annotation
GATK: Genome Analysis Tool Kit
GRS: Gene Risk Score
GTEx: Genotype Tissue Expression
HWE: Hardy-Weinberg equilibrium
IVIG: Intravenous Immunoglobulin
KD: Kawasaki Disease
LAD: Left Anterior Descending Artery
LD: Linkage Disequilibrium
LMCA: Left Main Coronary Artery
MAF: Minor Allele Frequency
PCA: Principal Components Analysis
QC: Quality Control
QQ: Quantile-quantile
RCA: Right Coronary Artery
ROC: Receiver Operating Characteristic
SNP: Single Nucleotide Polymorphism
WGS: Whole Genome Sequencing
ABSTRACT

Background: Kawasaki disease (KD) is a multisystem inflammatory illness of infants and young children that can result in acute vasculitis. The pathological walls of afflicted coronary arteries show propensity for forming thrombosis and aneurysms. The mechanism of coronary artery aneurysms (CAA) despite intravenous gamma globulin (IVIG) treatment is not known.

Methods: We performed a Whole Genome Sequencing (WGS) association analysis in a racially diverse cohort of KD patients treated with IVIG, both using AHA guidelines. We defined coronary aneurysm (CAA) (N = 234) as coronary z>2.5 and large coronary aneurysm (CAA/L) (N = 92) as z>5.0. We conducted logistic regression models to examine the association of genetic variants with CAA/L during acute KD and with persistence >6 weeks using an additive model between cases and 238 controls with no CAA. We adjusted for age, gender and three principal components of genetic ancestry. We performed functional mapping and annotation (FUMA) analysis and further assessed the predictive risk score of genomic risk loci using the area under the receiver operating characteristic curve (AUC).

Results: The top significant variants associated with CAA/L were in the intergenic regions (rs62154092 p<6.32E-08 most significant). Variants in SMAT4, LOC100127, PTPRD, TCAF2 and KLRC2 were the most significant non-intergenic SNPs. FUMA identified 12 genomic risk loci with eQTL or chromatin interactions mapped to 48 genes. Of these NDUFA5 has been implicated in KD CAA and MICU and ZMAT4 has potential functional implications. Genetic risk score using these 12 genomic risk loci yielded an AUC of 0.86.

Conclusions: This pharmacogenomics study provides insights into the pathogenesis of CAA/L in IVIG-treated KD patients. We have identified multiple novel SNPs associated with CAA/L and related genes with potential functional implications. The study shows that genomics can help define the cause of CAA/L to guide management and improve risk stratification of KD patients.

Key Words: Kawasaki Disease, Whole Genome Sequencing, IVIG refractory, Coronary Aneurysm
Background

Kawasaki disease (KD) is a life-threatening acute vasculitis that diffusely affects multiple organ systems in children. Coronary artery dilatations and aneurysms can occur and represent the most serious KD complications.\(^1,2\) For most patients, KD is self-limited and lacks the chronic nature of other autoimmune diseases; however, the pathological walls of afflicted vessels show a propensity for forming thrombosis and aneurysms.\(^3\) If untreated or treatment fails, the vasculitis can lead to coronary aneurysm or thrombosis in 20–25% of cases, potentially resulting in ischemic heart disease, myocardial infarction, or death.\(^4,5\) Clinical trials conducted in the 1980s and 1990s showed that IVIG treatment dramatically reduced occurrence of persistent CAA defined, primarily by Japanese Ministry of Health criteria.\(^6,7\) These criteria stated that coronary artery diameters $\geq 3$ mm in children $<5$ years and $\geq 4$ mm in children $\geq 5$ years were classified as abnormal. Echocardiographic detection and definition of significant CAA has dramatically improved over the past 2 to 3 decades. A 2007 study by the National Heart Lung and Blood Institute (NHLBI) showed that approximately 18 to 20% of patients had persistent CAA determined using coronary artery $z$ scores.\(^8,9\) The 2017 AHA guidelines adjusted the definition for CAA as $z$ score $\geq 2.5$, and $\geq 5$ defines a medium to large CAA.\(^1\) Although these coronary abnormalities show higher prevalence in IVIG refractory patients, they can still occur in patients seemingly responsive and showing fever resolution. Most of these children with large aneurysms require daily lifelong anti-coagulation, often with twice daily painful low molecular weight heparin injections, as warfarin is difficult to maintain within therapeutic range in children.

A study based on a large-scale Japanese cohort reported that coronary events did not occur in patients with small CAA; however, 5% of patients with medium CAA and 35% of patients with
large CAA had coronary events.\textsuperscript{10} North American studies support these data from Japan.\textsuperscript{11} The most severe form, giant coronary artery aneurysm (GCA), has been shown to be associated with complications such as luminal narrowing, thrombosis, and major cardiac events,\textsuperscript{11,12} and substantially alters quality of life for KD patients. Children with GCA require lifelong anticoagulation, exercise restrictions, and often treatment for ischemic heart disease such as coronary stenting and/or bypass grafting.\textsuperscript{1} The majority of patients with GCA develop clinically important stenosis from intimal hypertrophy during the late convalescent phase.\textsuperscript{13} Japanese males with coronary artery involvement have a mortality rate 2.4 times higher than expected in general population\textsuperscript{14}, but the overall impact on KD patients in the U.S. still requires definition. A recent systematic review showed that mid- to large-sized CAA provided the most significant risk factor for reducing survival of patients with KD.\textsuperscript{15} Mid- to large-sized CAA showed a slower recovery with worse prognosis than smaller CAA.\textsuperscript{16} Another study reported that the high persistence probability of mid- to large-sized CAA significantly increased the cardiovascular risk at 1 year after KD onset, when approximately two-thirds of the acute myocardial infarction cases occur.\textsuperscript{17}

Predicting increased risk for persistent large (medium to giant) size coronary aneurysms despite IVIG treatment is clinically important for intensification of treatment and disease management. Algorithms combining clinical and lab data in Japanese populations predict risk with respect to persistent CAA.\textsuperscript{18} However, the Japanese algorithms show poor predictive value for risk in North American and European cohorts,\textsuperscript{8,19} and fail even in some Asian populations.\textsuperscript{20,21} Thus, no universal biomarker or algorithm accurately predicts risk for persistent CAA in North America.\textsuperscript{22} Currently available data indicate that KD susceptibility and treatment response depend on an individual patient’s genetic background.\textsuperscript{23-31} Discrepancies among races or ethnicities also
suggest that pathogenesis of KD might vary.\textsuperscript{32,33} To date, very few studies have focused on genetic risk factors for CAA development in KD patients. We performed the first Whole Genome Sequencing (WGS) association analysis in a cohort of KD patients in a racially diverse North American population exhibiting differences in artery aneurysm formation. We identified multiple loci associated with CAA formation among a pediatric KD population receiving IVIG that can inform on risk stratification, potentially serve as treatment response predictors and guidance toward new therapeutic targets.

**Methods**

**Study populations and Primary Outcome**

We performed whole genome sequencing in 504 KD patients who were diagnosed and treated with IVIG (2g/Kg on a single infusion) and aspirin,\textsuperscript{34} both using the American Heart Association (AHA) criteria.\textsuperscript{1,35} All patients included in the study had echocardiography data that assessed for coronary artery aneurysm (CAA) and were included in this study. Coronary artery internal diameters in the left main coronary artery (LMCA), left anterior descending artery (LAD), and right coronary artery (RCA) obtained by echocardiography were converted to Z scores. Echocardiographic data were collected at baseline, at 2 weeks, and 5-6 weeks or after following fever onset. The > 5 -6 week echocardiograms were assessed for persistence of the CAAs. We used the Boston z -score model\textsuperscript{1} and as defined by AHA, coronary abnormality or aneurysm was considered if $z \geq 2.5$. We categorized positive or negative coronary artery aneurysm (CAA) occurring at any time point with $z$ score $\geq 2.5$ as “CAA”. Persistent CAA (P-CAA) was for any aneurysm $z > 2.5$ at $\geq 5$ – 6 weeks as P-CAA We also categorized large as medium to giant
coronary aneurysm for a z score ≥5.0 at any time as large CAA (CCA/L); and then persistent CCA/L (P-CCA/L) if at >5 – 6 weeks with z ≥ 5.0. Genomic comparisons were made for each of the 4 categories versus those without any aneurysm (z < 2.5).

**Whole Genome Sequencing and Variant Calling**

With consent from the parents or legal guardian, whole blood or saliva was collected to extract genomic DNA, as previously described.\(^{34}\) PCR-free libraries were generated using the BGI DNBSEQ True PCR-Free platform (Beijing Genomics Institute; Guangdong, Shenzhen, China) and whole genome sequencing was performed on the MGISEQ-2000 instrument (Beijing Genomics Institute; Guangdong, Shenzhen, China) to generate 100 bp paired-end reads, as previously described.\(^{34}\) The average sequencing depth was 30x per individual. Broad Institute’s Genome Analysis Tool Kit (GATK) best practices workflow was used for quality control and informatics pre-processing of the data.

**Whole Genome Sequencing (WGS) Association - Single-Variant Analysis**

Intensive quality control of the genetic data including minor allele frequency (MAF), call rate (CR), and p-values of Hardy-Weinberg equilibrium (HWE), were applied to filter uncertain SNPs as described previously,\(^{34}\) resulting in 46,718,826 variants (21,675,492 singletons) and 25,043,334 polymorphic SNPs were included in the analysis. Logistic regressions were conducted to examine the association of individual autosomal SNPs using an additive model in a case/control design, for the main outcome (medium/giant aneurysm) and the three secondary outcomes. Age, gender and three principal components (PCAs) of genetic ancestry were adjusted in the models. Quantile-quantile (QQ) plots and Manhattan plots were produced with the qqman
package in R. The crude and adjusted odds ratios (ORs) and 95% confidence intervals (CIs) were also calculated for the top hits using unconditional univariate logistic regression analysis to evaluate the associations between genotypes and medium/giant aneurysm.

**Identification of Genes and Their Roles Using FUMA**

Functional annotation was conducted in Functional Mapping and Annotation (FUMA) v1.3.0,\(^\text{36}\) using variants of interest from the WGS association analysis (\(p < 1.0 \times 10^{-5}\) and all variants in \(r^2 < 0.6\) with them). Lead SNPs were defined from these independent statistically significant SNPs if pairwise SNPs had \(r^2 < 0.1\). The maximum distance between LD blocks to merge into a genomic locus was 250 kb. The genetic data of mixed population in 1000G phase3 were used as reference to estimate LD. Three methods were used to map SNPs to genes: a) physical distance (within a 10-kb window) from known protein-coding genes in the human reference assembly, b) expression quantitative trait loci (eQTL) variant mapping using,\(^\text{37}\) and c) 3D chromatin interaction mapping (Hi-C).\(^\text{38}\) Combined Annotation-Dependent Depletion (CADD) analysis\(^\text{39}\) with a minimum score of \(> 12.37\) (considered to be suggestive deleterious) was used to filter the variants. Annotation of enhancers,\(^\text{40}\) tissue-specific expression of genes identified via Hi-C and eQTL mapping\(^\text{37}\) were queried in FUMA tool and Genotype Tissue Expression (GTEx) database (https://gtexportal.org/home/).

**Genetic Risk Score Computation**

All genomic risk loci from FUMA were used to estimate gene risk score (GRS) by a simple risk alleles count method. Discriminative power attributable to the GRS was estimated and compared by plotting receiver operating characteristic (ROC) curves and calculating the area under the
curve (AUC) for the case-control samples. The AUC compares the rates of true positives (sensitivity) and false positives (1 – specificity) and assesses the overall performance of genetic risk score models. Next, case and control status was randomly permuted 10,000 times and AUC was estimated with each pseudo case-control status. An empirical p-value (eP), which is the proportion of AUC based on randomization distribution of cases and controls that are more extreme than our observed AUC from the actual case (medium/giant aneurysm) and control (no aneurysm) status, was then calculated.

Results

Genetic Association Analysis Between Individual SNPs and the Risk of Large (medium/giant) Aneurysm

To identify SNPs associated with KD-associated large (medium/giant) coronary aneurysm (CCA/L), clinical data were linked in KD patients with whole genome sequencing data, nested in a clinical cohort as previously described. Basic demographics of the study population with CCA/L (N = 92) and no aneurysm (N = 276) are described in Table 1. Principal component analysis confirmed a good match between KD patients with CCA/L and those without any aneurysm. However, 3 PC were adjusted for in the analyses, as estimated in previous study. A quantile-quantile plot indicated that population stratification had negligible effects on the statistical results (λ genomic control = 0.955). There were several SNPs that exhibited suggestive statistical significance (p < 10^{-5}) in the additive genetic model, as shown in the Manhattan plot (Figure 1). Of all the SNPs examined, rs62154092 in the intragenic region (nearest gene ACTR3BP2) was the most statistically significant (6.32E-08). Among the overall top 10 most
significant SNPs, 5 SNPs (all intergenic C20_29360719, C20_29679725, C20_28616821, C20_29821717, C20_28616815) were in chromosome 20, although in different regions. All SNPs statistically significant at p < 10^{-4} are listed in Supplementary Table S1. Among the non-intergenic SNPs, rs28730284 upstream of KLRC2 gene was the most statistically significant (2.20 E-07) and among the top 15 non-intergenic, they were mostly intronic (rs9643846, rs9643847, rs57504215, rs60545202, rs59556769, rs73677451, rs12676292, rs4332118, rs6988966) located in SMAT4 and others in LOC100127 (non-coding RNA rs10276547, rs10280266), PTPRD (intrinsic rs600075, rs5896385) and TCAF2 (intrinsic C7_142457900) genes. The most significant exonic SNPs (rs11259953 and rs11259954) were in WHAMM gene (Supplementary Table S1). Regional association plots with cluster of SNPs in LD in chromosomes 7, 8 and 9 are shown in Figure 1 b-d. Results of corresponding single SNP association with developing any CAA (N = 233), any persistent CAA (N = 145) for 2 years, and P-CCA/L for 2 years (N = 79) vs no aneurysm (N = 276) are shown in Figure 2 and Supplementary Table S1.

**Gene mapping**

Using three gene mapping strategies (position mapping, eQTL mapping and chromatin interaction mapping) in FUMA, we mapped the significant association variants (P < 10^{-5}) to genes and identified 12 genomic risk loci (Supplementary Table S2) and 48 mapped genes associated with CCA/L (Figure 3, Supplementary Table S3). None of the genes were mapped by all three strategies. Three genes NDUFA5, ZMAT4 and MICU2 were mapped by physical and eQTL - NDUFA5 is located at the chromosome 7, and its lead SNP rs34163760 is located in the intron of the gene (P = 4.84 × 10^{-6}). An eQTL analysis showed that with the increasing number
of risk alleles of rs34163760, there was a higher mRNA level of \textit{NDUFA5} in the Esophagus. The CADD score of rs34163760 is 14.39 indicating a deleterious mutation. \textit{ZMAT4} is located in chromosome 8, and its lead SNP rs9643846 is located in the intron of the gene ($P = 6.57 \times 10^{-7}$).

An eQTL analysis showed that with the decreasing number of risk alleles of rs9643846, there was a higher mRNA level of \textit{ZMAT4} in thyroid. CADD score of rs9643846 is 15 indicating a deleterious mutation. The third gene \textit{MICU2} is located in chromosome 13, and its lead SNP rs12585631 is located in the intron of the gene ($P = 4.11 \times 10^{-6}$). An eQTL analysis showed that with the decreasing number of risk alleles of rs12585631, there was a higher mRNA level of \textit{MICU2} in thyroid. The CADD score of rs12585631 is 15.22 indicating a deleterious mutation.

Several genes in chromosomes 4, 7 and 13 were also identified to interact with the chromatin at those sites (\textbf{Figure 3, Supplementary Table S3}).

Expression patterns of the 48 prioritized genes were estimated in 54 different tissues (\textbf{Supplementary Table S4 Supplementary Figure S1}). Several of these genes show high expression in aorta and coronary artery tissues.

\textbf{Genetic Risk Score (GRS)}

Twelve genomic risk loci identified from FUMA yielded an AUC of 0.86. As shown in \textbf{Figure 4}, based on the empirical distribution of the AUC from the permutation test, eP was <0.0001, suggesting highly significant genetic risk score from the 12 genomic risk loci. For sensitivity analyses, when GRS for CCA/L was conducted separately in four specific races, AUC of 0.83, 0.78, 0.81 and 0.97 were obtained among White, Asian, Hispanic and African American KD patients.
DISCUSSIONS

We analyzed a relatively large North American KD cohort using whole genome sequencing. Treatment with IVIG during the acute phase was an inclusion criterion, so lack of treatment was not a confounding factor. Additionally, persistent coronary artery aneurysm (P-CAA or P-CCA/L) should be considered a failure of IVIG therapy. As noted in Table 1, the vast majority of medium to giant coronary aneurysms persisted. We identified for the most part novel gene loci that appear to have a relationship with coronary artery aneurysm formation and persistence in KD patients. We have used this same WGS strategy to identify genes related to IVIG refractoriness as defined by AHA guidelines. Prior studies searching for CAA genetic risk variants have used either hypothesis driven strategies or genome-wide association strategy with their inherent limitations. In this study, we specifically used the newest AHA classifications, and sought to determine genetic associations with CCA/L (Z ≥ 5) as these show a lower chance for early regression than do smaller aneurysms (Z > 2.5, but < 5). However, we also found that statistical genetic associations were consistent among all coronary phenotype groups as shown in Table 2 and Figure 2. This suggests that pathobiology or at least genetic risk is consistent regardless of size of the aneurysm or propensity for regression.

The most significant SNP related to moderate to giant aneurysm was located in the intergenic region with the closest gene being ACTR3BP2, a pseudogene with unknown function. However, among the top SNPs in the gene region, rs28730284 was located just upstream of the KLRC2 gene and has intriguing potential biological relevance to KD. The KLRC2 gene encodes the C-type lectin NKG2C (Killer cell lectin receptor-2). Natural Killer (NK) cells mediate innate immune responses against virally infected and malignant cells. NK cell function such as
production of proinflammatory cytokines, depends on a balance between activating and
inhibiting signals triggered by multiple surface receptors, including NKG2C. Polymorphisms in
*KLR2* have been shown to influence both function and expression of NK cells. Furthermore,
SNPs in *KLRC2* are associated with microvascular inflammation during renal graft transplant
rejection. Importantly, all type NK cell (CD56⁺CD16⁻, CD56⁺CD16⁺, CD56⁻CD16⁺) expression is reduced in KD patients compared to febrile or non-febrile controls, while
CD56⁻CD16⁺ NK cell expression was significantly lower in IVIG-resistant patients than in the
IVIG-responsive.

Multiple intronic SNPs were found in *ZMAT4* gene, which encodes the Zinc Finger Matrin-Type 4 protein. This gene was also identified by FUMA. SNPs within *ZMAT4* are associated with
diseases such as Spinocerebellar Ataxia and Myopia, and copy number variations are
associated with hematological malignancies. Function of this particular Zinc-Finger protein
remains undefined, so a potential biological role in KD would be unclear. The top two exonic
SNPs were in the *WHAMM* gene. This gene encodes a protein nucleation-promoting factor that
regulates the Actin-related protein 2/3 complex, but any biological relevance to KD would be highly speculative. Additionally, we found SNP (rs1052373) within the *MYBPC3* exon region as
marginally significant. *MYBPC3* (myosin binding protein c3) function is well established and
mutations are involved in the pathology of hypertrophic cardiomyopathy. This particular
SNP has also been associated with athletic endurance. However, any suggestion of biological
relevance for these exonic SNPs to KD would be highly speculative.

Multiple SNPs were also found in regions near Facioscapulohumeral muscular dystrophy
(FSHD) region-1(*FRGIDP*). *FRG1* acts on upstream of FGF2, which signals activation of the
AKT/ERK signaling axis in endothelial cells. Interestingly, we previously reported that this gene is associated with IVIG response in KD patients. *FRG1DP* has been linked to angiogenesis and retinal vasculopathy in FSHD patients including development of micro aneurysms. Additionally, altered expression for FRG-1 protein leads altered angiogenesis in human umbilical vein endothelial cells (HUVECs).

Using FUMA we found several genes potentially related to CAA/L in IVIG-treated KD patients. *MICU2* is a calcium sensitive regulatory subunit of the mitochondrial calcium uniporter and is important for reducing oxidative stress particularly in endothelial cells. *MICU2*−/− mice exhibit abnormal cardiac diastolic relaxation but also develop abdominal aortic aneurysms, which spontaneously rupture with only modest increases in blood pressure. *NDUFA5* encodes NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 5, a critical component of mitochondrial respiratory complex 1, which facilitates the translocation of protons across to the mitochondrial inner membrane. A SNP in *NDUFA5* was also a top hit in a small Taiwanese KD study evaluating genetic risk for CAA formation.\(^5\) Thus, our findings for these two genes regulating oxidative stress response suggest that mitochondria play a role in development of CAA and should be a target for future research.\(^5\)

We previously published a WGS pharmacogenomics analyses of IVIG response in KD using “persistent or recurrent fever” as the benchmark for IVIG resistance.\(^3\) Data suggests that IVIG resistance or refractoriness is a risk factor for persistent CAA. However, we did not find numbers of variants that were associated with both coronary aneurysms and IVIG resistance. This lack of genetic association uniformity between these two different outcomes suggests that their
biological pathways and mechanisms may be different. However, the list of novel genes provides new insights into the pathogenesis of KD.

Genetic analyses of coronary aneurysms in KD are complicated by multiple potential confounding factors. Prior studies, which have identified numerous significant variants associated with coronary aneurysms, did not account for the IVIG therapeutic effect.\textsuperscript{53, 54, 55} Clinical trials clearly show that IVIG reduces the risk of CAA.\textsuperscript{56} However, those previous genetic studies for the most part do not clarify whether study participants received appropriate IVIG treatment. Thus those prior cohorts could and probably do include patients whom did not receive timely IVIG. We used strict criteria for IVIG treatment in our study subjects in accordance with pharmacogenomics design. Accordingly, our results using different design and methodology did not replicate findings from prior studies such as associations with coronary aneurysms for variants in \textit{ITPKC}\textsuperscript{55}, \textit{KCNN2}\textsuperscript{54}, \textit{NEBL} and \textit{TUBA3C}\textsuperscript{51}, \textit{SLC8A1}\textsuperscript{57}, and the matrix metalloproteinase (\textit{MMP}) gene family\textsuperscript{58}.

Two other studies indicated association of TIFAB\textsuperscript{59} and PLCB1\textsuperscript{60} genes. Although the same SNPs were not replicated in our study, we found several SNPs in these gene regions that were associated with CAA in our study (Table S1). Unlike studies predominantly using Asian populations, our cohort included 4 races (Whites, Asians, African-Americans, and Hispanics). We had fewer cases of African-Americans and Hispanics and we were underpowered to conduct race-specific analyses (Supplementary Table S5). However, in the main combined cohort, we adjusted for three principal components which resulted in $\lambda$ genomic control $= 0.955$, suggesting no major confounding factors (Supplementary Figure S2).
We also used 12 genomic loci from FUMA to test for overall prediction of risk for developing CAA/L, a genetic predictive risk score. The AUC based on these markers is promising and the empirical risk models based on 10,000 simulated cases and controls had considerably higher AUC than theoretically achievable. Although sample size is small, all race-specific analyses also showed similar trends with high AUC. These specific markers need to be validated in different IVIG-treated KD populations; however, there is potential clinical utility in developing a point-of-care assay based on panels of genetic risk markers to predict who will develop CAA/L and/or other sequelae. Accurate prediction can assist in developing a treatment plan during the acute KD phase.\textsuperscript{61}

The main limitation for this study is the lack of a validation cohort. This is a common limitation of clinical trials and pharmacogenomics studies of rare diseases, although our cohort is the largest reported to date with a clear IVIG treatment phenotype. Further validations as well as functional studies of these variants will be needed in the future.

In summary, using WGS we have identified several novel genes and loci, which could have a functional impact on coronary artery response to IVIG in KD. Additionally, these loci could be used in developing an important predictive risk score for persistence of coronary artery aneurysms despite IVIG treatment.
Acknowledgments

The parent cohort/study and this genomic study conformed to the procedures for informed consent (parental permission) approved by institutional review boards at all sponsoring organizations and to human-experimentation guidelines set forth by the United States Department of Health and Human Services. We thank the participating patients and their parents.

This study was supported by the National Institutes of Health grant NHLBI-1R01HL146130.
<table>
<thead>
<tr>
<th></th>
<th>Medium/Large Coronary Aneurysm $Z \geq 5$</th>
<th>Persistent Medium/Large Coronary Aneurysm</th>
<th>Any Coronary Aneurysm $Z \geq 2.5$</th>
<th>Persistent Coronary Aneurysm</th>
<th>No Coronary Aneurysm $Z &lt; 2.5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>92</td>
<td>79</td>
<td>233</td>
<td>145</td>
<td>276</td>
</tr>
<tr>
<td>Median Age (months)</td>
<td>20</td>
<td>21.5</td>
<td>27</td>
<td>26</td>
<td>36</td>
</tr>
<tr>
<td>Gender</td>
<td>Male</td>
<td>58</td>
<td>48</td>
<td>154</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>34</td>
<td>31</td>
<td>79</td>
<td>50</td>
</tr>
</tbody>
</table>
Table 2. Top Genes Associated with Large (Medium/Giant) Coronary Aneurysm (CCA/L) (z ≥5), Persistent CCA/L (P-CCA/L), any coronary artery aneurysm (CAA) (z≥2.5), and persistent CAA (P-CAA)

<table>
<thead>
<tr>
<th>SNP</th>
<th>Genes and Gene Regions</th>
<th>Large Aneurysm</th>
<th>Persistent Large Aneurysm</th>
<th>Any Aneurysm</th>
<th>Persistent Any Aneurysm</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>OR (95% CI)</td>
<td>P-value</td>
<td>OR (95% CI)</td>
<td>P-value</td>
</tr>
<tr>
<td>rs62154092</td>
<td>ACTR3BP2;NONE</td>
<td>0.17 [0.09, 0.33]</td>
<td>6.32E-08</td>
<td>0.16, [0.08, 0.30]</td>
<td>3.21E-08</td>
</tr>
<tr>
<td>C20_29360719</td>
<td>FGR1DP;FGR2EP</td>
<td>0.15 [0.08, 0.31]</td>
<td>1.71E-07</td>
<td>0.19, [0.09, 0.38]</td>
<td>0.000003368</td>
</tr>
<tr>
<td>rs28730284</td>
<td>KLR2</td>
<td>0.18 [0.10, 0.35]</td>
<td>2.20E-07</td>
<td>0.20, [0.10, 0.39]</td>
<td>0.000001955</td>
</tr>
<tr>
<td>C20_29679725</td>
<td>FAM242B;MIR4477A</td>
<td>0.24 [0.14, 0.41]</td>
<td>2.28E-07</td>
<td>0.22, [0.13, 0.39]</td>
<td>1.971E-07</td>
</tr>
<tr>
<td>C4_154148395</td>
<td>SFRP2;DCHS2</td>
<td>0.19 [0.10, 0.36]</td>
<td>2.59E-07</td>
<td>0.20, [0.10, 0.38]</td>
<td>0.000001086</td>
</tr>
<tr>
<td>C20_28616821</td>
<td>FGR1CP;FGR1DP</td>
<td>5.76 [2.93, 11.32]</td>
<td>3.83E-07</td>
<td>6.11, [2.96, 12.63]</td>
<td>0.000001039</td>
</tr>
<tr>
<td>C20_298221717</td>
<td>FAM242B;MIR4477A</td>
<td>0.21 [0.11, 0.38]</td>
<td>3.86E-07</td>
<td>0.20, [0.11, 0.38]</td>
<td>8.742E-07</td>
</tr>
<tr>
<td>C20_28616815</td>
<td>FGR1CP;FGR1DP</td>
<td>5.64 [2.87, 11.08]</td>
<td>5.37E-07</td>
<td>5.97, [2.88, 12.34]</td>
<td>0.000001452</td>
</tr>
<tr>
<td>C3_110637278</td>
<td>MIR4445;NECTIN3-AS1</td>
<td>4.60 [2.52, 8.40]</td>
<td>6.86E-07</td>
<td>5.24, [2.80, 9.80]</td>
<td>2.227E-07</td>
</tr>
</tbody>
</table>

Top 5 exonic SNPs

<table>
<thead>
<tr>
<th>SNP</th>
<th>Gene</th>
<th>OR (95% CI)</th>
<th>P-value</th>
<th>OR (95% CI)</th>
<th>P-value</th>
<th>OR (95% CI)</th>
<th>P-value</th>
<th>OR (95% CI)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs11259953</td>
<td>WHAMM</td>
<td>3.30 [1.90, 5.74]</td>
<td>2.24E-05</td>
<td>3.34, [1.88, 5.93]</td>
<td>3.743E-05</td>
<td>1.98[1.29, 3.02]</td>
<td>0.001652</td>
<td>2.01[1.24, 3.24]</td>
<td>0.004333</td>
</tr>
<tr>
<td>rs11259954</td>
<td>WHAMM</td>
<td>3.30 [1.90, 5.74]</td>
<td>2.24E-05</td>
<td>3.34, [1.88, 5.93]</td>
<td>3.743E-05</td>
<td>1.98[1.29, 3.02]</td>
<td>0.001652</td>
<td>2.01[1.24, 3.24]</td>
<td>0.004333</td>
</tr>
<tr>
<td>rs1055110*</td>
<td>MADD-AS1</td>
<td>0.42 [0.27, 0.65]</td>
<td>8.76E-05</td>
<td>0.36, [0.23, 0.58]</td>
<td>2.161E-05</td>
<td>0.59, [0.36, 0.79]</td>
<td>0.0004243</td>
<td>0.51[0.36, 0.72]</td>
<td>0.001133</td>
</tr>
<tr>
<td>rs1052373</td>
<td>MYBPC3</td>
<td>0.44 [0.29, 0.68]</td>
<td>0.0001853</td>
<td>0.39, [0.25, 0.61]</td>
<td>5.28E-05</td>
<td>0.60, [0.45, 0.81]</td>
<td>0.0007693</td>
<td>0.53, [0.37, 0.74]</td>
<td>0.0002178</td>
</tr>
<tr>
<td>rs77868901</td>
<td>TMEM259</td>
<td>9.00 [2.76, 29.31]</td>
<td>0.000266</td>
<td>8.57, [2.62, 28.05]</td>
<td>0.0003862</td>
<td>3.60, [1.21, 10.71]</td>
<td>0.02129</td>
<td>4.69, [1.51, 14.54]</td>
<td>0.007461</td>
</tr>
</tbody>
</table>

Top 10 non-exonic SNPs in the gene

<table>
<thead>
<tr>
<th>SNP</th>
<th>Gene</th>
<th>OR (95% CI)</th>
<th>P-value</th>
<th>OR (95% CI)</th>
<th>P-value</th>
<th>OR (95% CI)</th>
<th>P-value</th>
<th>OR (95% CI)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs28730284</td>
<td>Upstream KLR2</td>
<td>0.18 [0.10, 0.35]</td>
<td>2.202E-07</td>
<td>0.20, [0.1, 0.39]</td>
<td>1.955E-06</td>
<td>0.37, [0.25, 0.55]</td>
<td>6.343E-07</td>
<td>0.35, [0.22, 0.55]</td>
<td>6.858E-06</td>
</tr>
<tr>
<td>rs57504215</td>
<td>Intronic ZMAT4</td>
<td>3.88(2.23, 6.78)</td>
<td>1.783E-06</td>
<td>3.90(2.2, 6.92)</td>
<td>3.042E-06</td>
<td>2.18(1.45, 3.27)</td>
<td>0.0001609</td>
<td>1.9(1.21, 2.98)</td>
<td>0.005597</td>
</tr>
<tr>
<td>rs60545202</td>
<td>Intronic ZMAT4</td>
<td>3.91(2.23, 6.84)</td>
<td>1.821E-06</td>
<td>3.96(2.22, 7.06)</td>
<td>2.973E-06</td>
<td>2.13(1.41, 3.23)</td>
<td>0.000347</td>
<td>1.98(1.26, 3.14)</td>
<td>0.003374</td>
</tr>
<tr>
<td>rs59556769</td>
<td>Intronic ZMAT4</td>
<td>3.73(2.16, 6.42)</td>
<td>2.109E-06</td>
<td>3.87(2.21, 6.77)</td>
<td>2.182E-06</td>
<td>2.01(1.35, 2.99)</td>
<td>0.0006387</td>
<td>1.98(1.27, 3.08)</td>
<td>0.002665</td>
</tr>
<tr>
<td>rs10276547</td>
<td>ncRNA intronic LOC100128317</td>
<td>5.55(2.73, 11.27)</td>
<td>2.171E-06</td>
<td>5.93(2.88, 12.2)</td>
<td>1.343E-06</td>
<td>2.74(1.56, 4.81)</td>
<td>0.0004307</td>
<td>3.88(2.09, 7.22)</td>
<td>1.781E-05</td>
</tr>
<tr>
<td>rs73677451</td>
<td>Intronic ZMAT4</td>
<td>3.70(2.15, 6.38)</td>
<td>2.355E-06</td>
<td>3.72(2.13, 6.52)</td>
<td>4.143E-06</td>
<td>2.08(1.39, 3.09)</td>
<td>0.0003341</td>
<td>1.98(1.27, 3.09)</td>
<td>0.002503</td>
</tr>
<tr>
<td>rs600075</td>
<td>Intronic PTPRD</td>
<td>3.59(2.11, 6.09)</td>
<td>2.358E-06</td>
<td>3.55(2.05, 6.16)</td>
<td>6.451E-06</td>
<td>2.30(1.57, 3.37)</td>
<td>1.79E-05</td>
<td>2.41(1.56, 3.74)</td>
<td>8.241E-05</td>
</tr>
<tr>
<td>rs5896385</td>
<td>Intronic PTPRD</td>
<td>3.59(2.11, 6.09)</td>
<td>2.358E-06</td>
<td>3.55(2.05, 6.16)</td>
<td>6.451E-06</td>
<td>2.30(1.57, 3.37)</td>
<td>1.79E-05</td>
<td>2.41(1.56, 3.74)</td>
<td>8.241E-05</td>
</tr>
</tbody>
</table>
Figure Legends

**Figure 1 Overall and Regional Association Results** A) Manhattan Plot displaying Whole Genome Sequence Association results with - Large (medium/giant) Coronary Aneurysm (N = 92) vs no Coronary Artery Aneurysm (N = 276). Negative log_{10}-transformed $P$ values from the logistic regression model (additive model) are plotted on the $y$-axis and the SNP genomic locations on the $x$-axis (colors representing different autosomal chromosomes). Locus Zoom plots for selected gene regions in B) Chromosome 7 with rs10276547 the most significant SNP in the region, C) Chromosome 8 with rs9643846 the most significant SNP in the region, D) Chromosome 9 with rs600075 the most significant SNP in the region. Vertical axis (on the left) is the $-\log_{10}$ of the $p$-value, the horizontal axis is the chromosomal position. Each dot represents a SNP tested for association with large coronary aneurysm. Linkage disequilibrium between the most significant SNP, listed at the top of each plot, and the other SNPs in the plot is shown by the $r^2$ legend in each plot. Vertical axis (on the right) is the recombination – the site and rate are represented by red curves.

**Figure 2 Circo plot summarizing whole genome sequence associations** – the outer plot is the association of large aneurysm (N = 92), the next inner plot is the association of persistent large aneurysm (N = 79), the next inner plot is the association of any coronary aneurysm (N = 233) and the fourth inner plot is the association of persistent any coronary aneurysm vs no coronary aneurysm (for each outcome). The innermost plot indicates if SNPs were associated in 1-4 outcomes with $p < 10^{-5}$.

**Figure 3 FUMA circos plots of mapped genes in genomic risk loci** - The most outer layer is the Manhattan plot (only SNPs with $P < 0.05$ are displayed). Genomic risk loci are highlighted in blue and the strength of linkage disequilibrium $r^2$ between each SNP to the lead SNP is given by the following color code: red ($r^2 > 0.8$), orange ($r^2 > 0.6$), green ($r^2 > 0.4$), blue ($r^2 > 0.2$) and grey ($r^2 \leq 0.2$). Genes are mapped by 3-D chromatin interaction (orange) or eQTLs (green), or both (red). A) Circos plot for Chromosome 13 with lead SNP rs12585631, B) Circos plot for Chromosome 7 with lead SNP rs10276547, and C) Circos plot for Chromosome 4 with lead SNP rs62330192

**Figure 4 Empirical curve based on the AUC from 10,000 permutation.** X-axis is the AUC value based on the 12 genomic Loci from FUMA in predicting the risk score of having a large aneurysm and Y-axis is the frequency. True AUC of 0.
Supplementary Figures

Supplementary Figure S1. Heat map of 48 prioritized genes based on FUMA related to 12 genomic risk loci in 54 different tissues. The tissue-specific expression patterns are based on GTEx v6 expression data. Cells in red represent a higher expression, compared to the cells in blue. Gene expression comparisons between tissues (horizontal comparison) within a gene (y axis) are comparable but not those of different genes within a tissue (vertical comparison).

Supplementary Figure S2 QQ plot of whole genome sequence association analysis of large aneurysm in IVIG-treated Kawasaki Disease. Q-Q plot shows observed distributions of association statistics (-log(p)) in y-axis against those expected (-log(p)) in x-axis under the fitted model.
Supplementary Tables

Supplementary Table S1. Independent SNPs (r2 < 0.6) associated with large (medium/giant) aneurysm

Supplementary Table S2. Genomic risk loci of interest from large (medium/giant) aneurysm WGS Association

Supplementary Table S3. All genes mapped in SNP-based (FUMA) for large (medium/giant) aneurysm

Supplementary Table S4. Expression values of the genes prioritized by FUMA for large (medium/giant) aneurysm in 54 tissues

Supplementary Table S5. Minor allele frequencies of cases and control design for 4 CAA phenotype
REFERENCES


chr9:10297519
AUC = 0.86, eP < 0.0001