



Pharmacogenomics of coronary artery response to intravenous gamma globulin in kawasaki disease



Sadeep Shrestha ¹✉, Howard W. Wiener¹, Sabrina Chowdhury¹, Hidemi Kajimoto², Vinodh Srinivasasainagendra³, Olga A. Mamaeva¹, Ujval N. Brahmbhatt¹, Dolena Ledee², Yung R. Lau⁴, Luz A. Padilla¹, Jake Y. Chen⁵, Nagib Dahdah⁶, Hemant K. Tiwari³ & Michael A. Portman²

Kawasaki disease (KD) is a multisystem inflammatory illness of infants and young children that can result in acute vasculitis. The mechanism of coronary artery aneurysms (CAA) in KD despite intravenous gamma globulin (IVIG) treatment is not known. 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 $z \geq 2.5$ and large coronary aneurysm (CAA/L) ($N = 92$) as $z \geq 5.0$. 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. 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. Functional mapping and annotation (FUMA) analysis 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 area under the receiver operating characteristic curve (AUC) of 0.86. This pharmacogenomics study provides insights into the pathogenesis of CAA/L in IVIG-treated KD and shows that genomics can help define the cause of CAA/L to guide management and improve risk stratification of KD patients.

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 aneurysms³. 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 ≥ 3 mm in children <5 years and >4 mm in children ≥ 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 ≥ 2.5 and ≥ 5 defines a medium to large CAA¹. Although these coronary abnormalities show higher prevalence in IVIG refractory patients, they can

¹Department of Epidemiology, School of Public Health, University of Alabama at Birmingham, Birmingham, AL, USA. ²Division of Cardiology, Seattle Children's and University of Washington Department of Pediatrics, Seattle, WA, USA. ³Department of Biostatistics, School of Public Health, University of Alabama at Birmingham, Birmingham, AL, USA. ⁴Division of Pediatric Cardiology, Heersink School of Medicine, University of Alabama at Birmingham, Birmingham, AL, USA. ⁵Department of Biomedical Informatics and Data Science, Heersink School of Medicine, University of Alabama at Birmingham, Birmingham, AL, USA. ⁶Division of Pediatric Cardiology, CHU Ste-Justine, Universite de Montreal, Montreal, QC, Canada. ✉e-mail: sshrestha@uab.edu

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 events¹⁰. North American studies support these data from Japan¹¹. 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^{11,12}, and substantially alters quality of life for KD patients. Children with GCA require lifelong anti-coagulation, exercise restrictions, and often treatment for ischemic heart disease such as coronary stenting and/or bypass grafting¹. The majority of patients with GCA develop clinically important stenosis from intimal hypertrophy during the late convalescent phase¹³. Japanese males with coronary artery involvement have a mortality rate 2.4 times higher than expected in general population¹⁴, 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 KD¹⁵. Mid- to large-sized CAA showed a slower recovery with worse prognosis than smaller CAA¹⁶. 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 occur¹⁷.

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 CAA¹⁸. However, the Japanese algorithms show poor predictive value for risk in North American and European cohorts^{8,19}, and fail even in some Asian populations^{20,21}. Thus, no universal biomarker or algorithm accurately predicts risk for persistent CAA in North America²².

A single U.S center retrospective study showed that presence of early coronary artery dilation is moderately useful in predicting persistent dilation. However, that study did not specifically evaluate for larger higher risk aneurysms²³. Currently available data indicate that KD susceptibility and treatment response depend on an individual patient's genetic background²⁴⁻³³. Discrepancies among races or ethnicities also suggest that pathogenesis of KD might vary^{34,35}. 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.

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 described³⁶. Basic demographics of the study population with CCA/L ($N = 91$) and no aneurysm ($N = 278$) 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 study³⁶. 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 (Fig. 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 rs1424006606, rs1396081550, rs1258107032, rs1379390981, rs1424309393) 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 (2.20E-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* (intronic rs600075, rs5896385) and *TCAF2* (intronic rs1218424730) 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 (Fig. 1b-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 Fig. 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 (Fig. 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 ($P = 4.84E-06$). An eQTL analysis showed that with the increasing number of risk alleles of rs34163760, there was a higher mRNA level of *NDUFA5* in the Esophagus. The CADD score of rs34163760 is 14.39 indicating a deleterious mutation. *ZMAT4* is located in chromosome 8, and its lead SNP rs9643846 is located in the intron of the gene ($P = 6.57E-07$). An eQTL analysis showed that with the decreasing number of risk alleles of rs9643846, there was a higher mRNA level of *ZMAT4* in thyroid. CADD score of rs9643846 is 15 indicating a deleterious mutation. The third gene *MICU2* is located in chromosome 13, and its lead SNP rs12585631 is located in the intron of the gene ($P = 4.11E-06$). An eQTL analysis showed that with the decreasing number of risk alleles of rs12585631, there was a higher mRNA level of *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 (Fig. 3, Supplementary Table S3).

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

Genetic risk score (GRS)

Twelve genomic risk loci identified from FUMA yielded an AUC of 0.86. As shown in Fig. 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.

Discussion

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

Table 1 | Kawasaki Disease Patients With (medium/large, any) and Without Coronary Aneurysm included in the whole genome sequencing analyses

	Medium/Giant Coronary Aneurysm $Z \geq 5$	Persistent Medium/Giant Coronary Aneurysm	Any Coronary Aneurysm $Z \geq 2.5$	Persistent Coronary Aneurysm	No Coronary Aneurysm $Z < 2.5$
N	91	79	233	145	278
Median Age (months)	20	21.5	27	26	36
Gender	58	48	154	95	157
Male	34	31	79	50	119
Female					

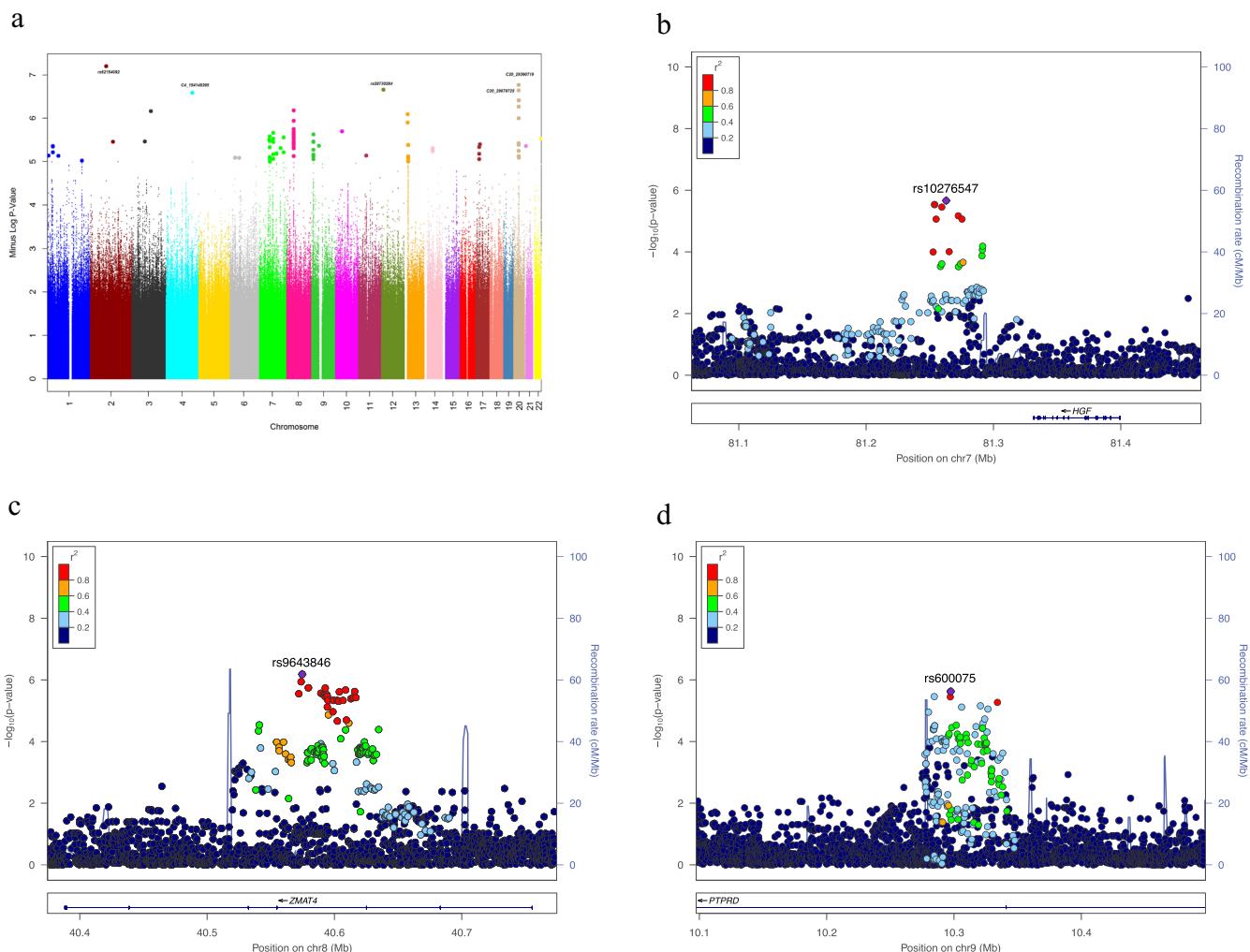


Fig. 1 | Overall and regional association results. **a** Manhattan Plot displaying Whole Genome Sequence Association results with—Large (medium/giant) Coronary Aneurysm ($N = 92$) vs no Coronary Artery Aneurysm ($N = 276$). Negative \log_{10} -transformed P values from the logistic regression model (additive 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 aneurysms. 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.

refractoriness as defined by AHA guidelines¹. 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 ($Z \geq 5$) as these show a lower chance for early regression than do smaller aneurysms ($Z \geq 2.5$, but < 5). However, we also found that statistical genetic associations were consistent

among all coronary phenotype groups as shown in Table 2 and Fig. 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

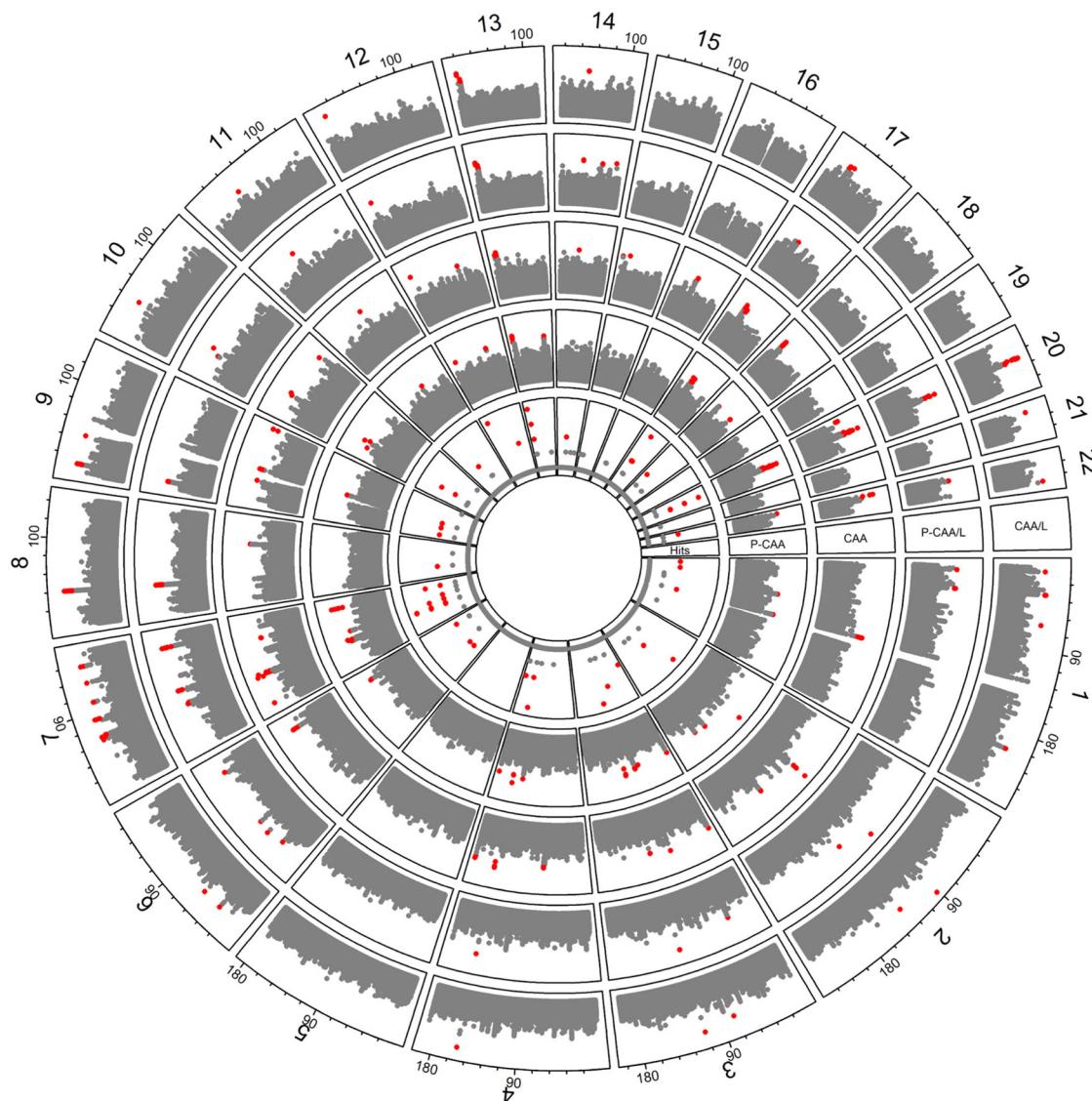


Fig. 2 | Circo plot summarizing whole genome sequence associations. The outer plot is the association of large aneurysm ($N = 92$), the next inner plot is the association of persistent large aneurysm ($N = 79$), the next inner plot is the association of

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 $p < 1.0E-05$.

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^{37–39}. NK cell function such as production of proinflammatory cytokines, depends on a balance between activating and inhibiting signals triggered by multiple surface receptors, including NKG2C⁴⁰. Polymorphisms in *KLRC2* 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-responsive⁴¹.

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 Myopia⁴², and copy number variations are associated with hematological malignancies⁴³. 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^{44,45}. This particular SNP has also been associated with athletic endurance⁴⁶. 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(*FRG1DP*). *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

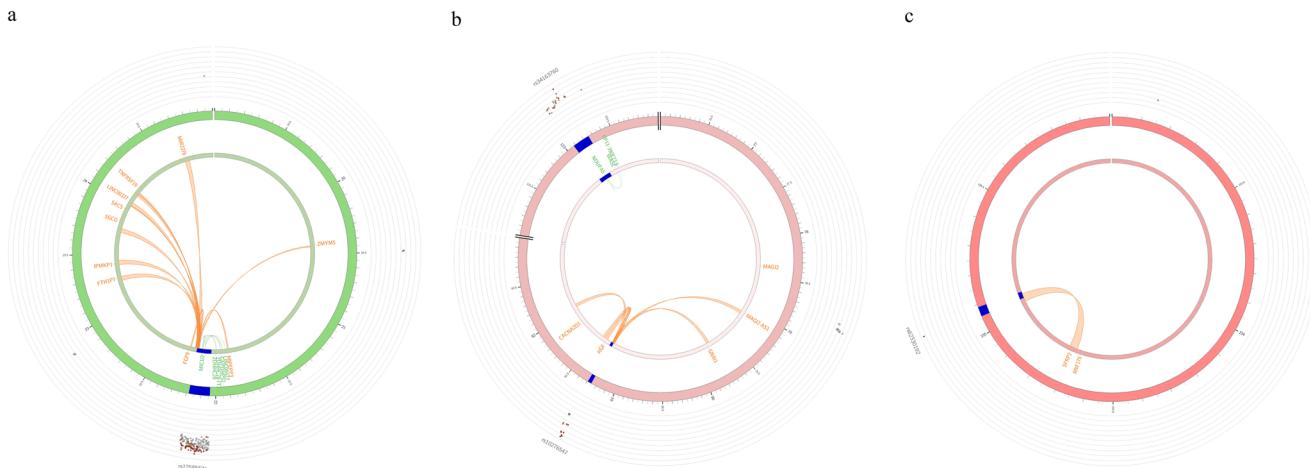


Fig. 3 | FUMA circos plots of mapped genes in genomic risk loci. The most outer layer is the Manhattan plot (only SNPs with $P < 0.05$ are displayed). Genomic risk loci are highlighted in blue and the strength of linkage disequilibrium r^2 between each SNP to the lead SNP is given by the following color code: red ($r^2 > 0.8$), orange ($r^2 > 0.6$),

green ($r^2 > 0.4$), blue ($r^2 > 0.2$) and gray ($r^2 \leq 0.2$). 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.

oxidative stress particularly in endothelial cells. MICU 2 $-/-$ 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 formation⁴⁷. 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 research⁴⁸.

We previously published a WGS pharmacogenomics analyses of IVIG response in KD using “persistent or recurrent fever” as the benchmark for IVIG resistance³⁶. 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^{49–51}. Clinical trials clearly show that IVIG reduces the risk of CAA⁵². 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 *ITPKC*⁵¹, *KCNN2*⁵⁰, *NEBL* and *TUBA3C*⁴⁷, *SLC8A1*⁵³, and the matrix metalloproteinase (MMP) gene family⁵⁴. Likewise, Huang et al⁵⁵ reported TET mRNA levels associated with IVIG and DMNT1 mRNA levels with CAA; however, there were no SNPs in these gene regions that were statistically significant in our study.

Two other studies indicated association of *TIFAB*⁵⁶ and *PLCB1*⁵⁷ 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 λ 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 phase⁵⁸.

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.

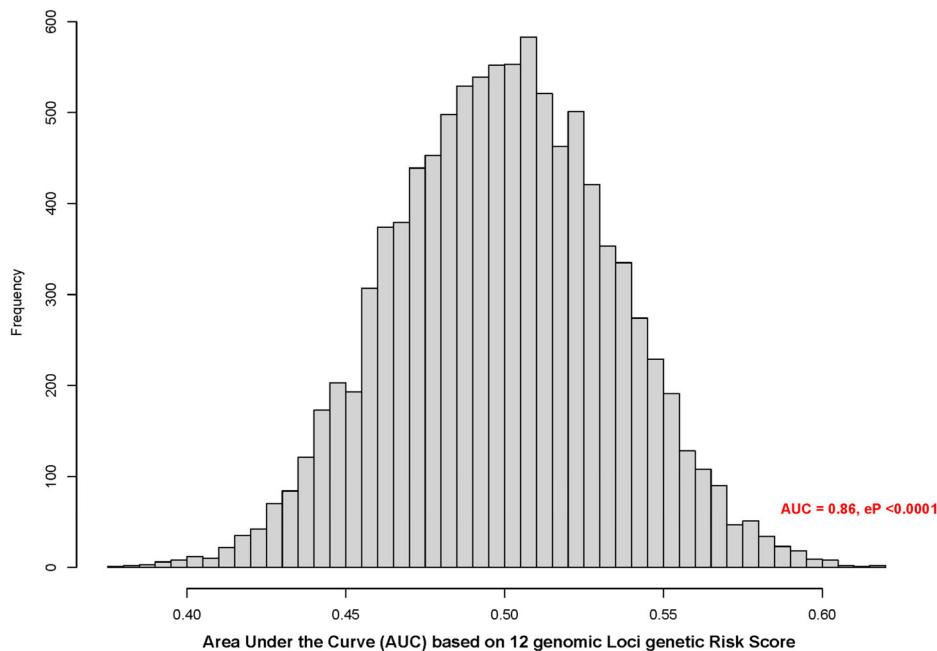
In summary, using WGS we have identified several novel genes and loci, which could have a functional impact on coronary artery response to IVIG in KD. Additionally, these loci could be used in identifying new personalized therapeutic avenues as well as developing an important predictive risk score for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

Methods

Study populations and primary outcome

We performed whole genome sequencing in 504 KD patients who were diagnosed and treated with IVIG (2 g/Kg on a single infusion) and aspirin³⁶, both using the American Heart Association (AHA) criteria^{1,59}. All patients included in the study had echocardiography data that assessed for coronary artery aneurysm (CAA) (212 Whites, 75 Asians, 50 Hispanics and 32 Blacks). Coronary artery internal diameters in the left main coronary artery (LMCA), left anterior descending artery (LAD), and right coronary artery (RCA) obtained by echocardiography. Coronary artery dimensions were normalized for body surface area and converted to z-scores (SDs from a predicted normal mean) based on nonlinear regression equations derived from a normal nonfebrile population⁶⁰. 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.

Fig. 4 | Empirical curve based on the area under the curve (AUC) from 10,000 permutation. X-axis is the AUC value after a random permutation of the outcome variable (medium to large coronary aneurysm, $z > 5.0$) based on the 12 genomic loci from FUMA in predicting the risk score of having a large aneurysm and Y-axis is the frequency. Empirical P-value (eP) is proportion of permutations resulting in a larger AUC than original data.



We used the Boston z -score model¹ and as defined by AHA, coronary abnormality or aneurysm was considered if $z \geq 2.5$. According to AHA guidelines, we categorized positive or negative coronary artery aneurysm (CAA) occurring at any time point with z score ≥ 2.5 as “CAA”. Persistent CAA (P-CAA) was defined as any aneurysm $z \geq 2.5$ for up to 5–6 weeks. We also categorized large CAA (CAA/L) as medium to giant coronary aneurysm for a z score ≥ 5.0 at any time; and then persistent CCA/L (P-CCA/L) if aneurysm $z \geq 5.0$ remained for up to 5–6 weeks. Genomic comparisons were made for each of the 4 categories versus those without any aneurysm ($z < 2.5$).

The parent cohort/study and this pharmacogenomic study conformed to the procedures for informed consent (parental permission) approved by institutional review boards at all sponsoring organizations. The pharmacogenomic data management and analysis procedure was approved by the University of Alabama at Birmingham Institutional Review Board (IRB). The study was conducted in accordance with the local legislation and institutional requirements following the ethical guidelines of the Declaration of Helsinki. Written informed consent for minor participation (children) in this study was provided by the participants’ legal guardians/next of kin.

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 described³⁶. 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 described³⁶. All reads that passed were aligned to the human reference genome (hg38) using Burrows-Wheeler aligner (BWA) v 0.7.17. 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. Variant-level QC was performed using the Variant Quality Score Recalibration tool (VQSR) from the Genome Analysis Toolkit (GATK), using the recommended threshold of 99% sensitivity for the “true” variant. As we previously reported, we included 5 duplicate samples, which showed overall high SNP genotype concordance, with a kinship coefficient estimate, $\Phi > 0.497$ between duplicates. For the SNPs we report association, the concordant genotypes were confirmed between all duplicate samples.

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 previously³⁶, resulting in 46,718,826 variants (21,675,492 singletons) and 25,043,334 polymorphic SNPs were included in the analysis. Logistic regression models were conducted using PLINK 1.90 to examine the association of individual autosomal SNPs using an additive model in a case/control design, for the main outcome (medium/giant aneurysm, 91 cases (52 W, 3AA, 20His, 16As) and 278 controls (160 W, 29AA, 30His, 59As) and the three secondary outcomes (Table 1). 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) v1.3.0⁶¹, 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 using⁶², and (c) 3D chromatin interaction mapping (Hi-C)⁶³. Combined Annotation-Dependent Depletion (CADD) analysis⁶⁴ with a minimum score of >12.37 (considered to be suggestive deleterious) was used to filter the variants. Annotation of enhancers⁶⁵, tissue-specific expression of genes identified via Hi-C and eQTL mapping⁶² 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

Table 2 | Top genes associated with large (medium/giant) coronary aneurysm (CCA/L) ($z \geq 5$), persistent CCA/L (P-CCA/L), any coronary artery aneurysm (CAA) ($z \geq 2.5$), and persistent CAA (P-CAA)

Overall Top 10 Hits						
SNP*	Genes and gene regions	Large Aneurysm		Persistent Large Aneurysm		Persistent Any Aneurysm
		Closest gene	OR (95% CI)	P value	OR (95% CI)	
rs62154092 (T/c)	ACTR3BP2;NONE	0.17 [0.09, 0.33]	6.32E-08	0.16 [0.08, 0.30]	3.22E-08	0.23 [0.14, 0.39]
rs1424006606 (C/t)	FRG1DP;FRG2EP	0.15 [0.08, 0.31]	1.71E-07	0.19 [0.09, 0.38]	3.37E-06	0.40 [0.26, 0.59]
rs28730284 (G/a)	KLRC2	0.18 [0.10, 0.35]	2.20E-07	0.20 [0.10, 0.39]	1.96E-06	0.37 [0.25, 0.55]
rs1396081550 (C/t)	FAM242B;MIR4477A	0.24 [0.14, 0.41]	2.28E-07	0.22 [0.13, 0.39]	1.97E-07	0.35 [0.24, 0.52]
rs6230192 (A/g)	SFRP2;DCHS2	0.19 [0.10, 0.36]	2.59E-07	0.20 [0.10, 0.38]	1.09E-06	0.38 [0.25, 0.56]
rs1258107032 (T/a)	FRG1CP;FRG1DP	5.76 [2.93, 11.32]	3.83E-07	6.11 [2.96, 12.63]	1.04E-06	2.70 [1.82, 4.00]
rs1379390981 (T/g)	FAM242B;MIR4477A	0.21 [0.11, 0.38]	3.86E-07	0.20 [0.11, 0.38]	8.74E-07	0.27 [0.17, 0.44]
rs1424308393 (A/t)	FRG1CP;FRG1DP	5.64 [2.87, 11.08]	5.37E-07	5.97 [2.88, 12.34]	1.45E-06	2.74 [1.84, 4.07]
rs9643846 (G/t)	ZMAT4	4.12 [2.36, 7.20]	6.57E-07	4.16 [2.35, 7.39]	1.10E-06	2.30 [1.54, 3.44]
rs377109159 (T/c)	MIR4445;NECTIN3-AS1	4.60 [2.52, 8.40]	6.86E-07	5.24 [2.80, 9.80]	2.23E-07	2.83 [1.79, 4.47]
Top 5 exonic SNPs						
SNP	Gene	OR (95% CI)	P value	OR (95% CI)	P value	OR (95% CI)
rs11259953 (A/c)	WHAMM	3.30 [1.90, 5.74]	2.25E-05	3.34 [1.88, 5.93]	3.74E-05	1.98 [1.29, 3.02]
rs11259954 (C/g)	WHAMM	3.30 [1.9, 5.74]	2.25E-05	3.34 [1.88, 5.93]	3.74E-05	1.98 [1.29, 3.02]
rs105510 (A/g)	MADD-AS1	0.42 [0.27, 0.65]	8.76E-05	0.36 [0.23, 0.58]	2.16E-05	0.59 [0.43, 0.79]
rs1052373 (C/t)	NYBPC3	0.44 [0.29, 0.68]	1.85E-04	0.39 [0.25, 0.61]	5.28E-05	0.60 [0.45, 0.81]
rs77888901 (G/c)	TMEM259	9.00 [2.76, 29.31]	2.66E-04	8.57 [2.62, 28.05]	3.86E-04	3.60 [1.21, 10.71]
Top 10 non-exonic SNPs in the gene						
SNP	Gene	OR (95% CI)	P value	OR (95% CI)	P value	OR (95% CI)
rs28730284 (A/g)	Upstream KLRC2	0.18 [0.10, 0.35]	2.20E-07	0.20 [0.1, 0.39]	1.96E-06	0.37 [0.25, 0.55]
rs9643846 (T/g)	Intronic ZMAT4	4.12 [2.36, 7.2]	6.57E-07	4.16 [2.35, 7.39]	1.10E-06	2.31 [1.54, 3.44]
rs9643847 (G/a)	Intronic ZMAT4	3.95 [2.27, 6.86]	1.14E-06	3.86 [2.19, 6.82]	3.19E-06	2.14 [1.44, 3.2]
rs57504215 (G/a)	Intronic ZMAT4	3.88 [2.3, 6.78]	1.78E-06	3.90 [2.12, 6.92]	3.04E-06	2.18 [1.45, 3.27]
rs60545202 (A/g)	Intronic ZMAT4	3.91 [2.23, 6.84]	1.82E-06	3.96 [2.22, 7.06]	2.97E-06	2.13 [1.41, 3.23]
rs59566769 (G/a)	Intronic ZMAT4	3.73 [2.16, 6.42]	2.11E-06	3.87 [2.21, 6.77]	2.18E-06	2.01 [1.35, 2.99]
rs1026547 (C/t)	ncRNA_intronic LOC100128317	5.55 [2.73, 11.27]	2.17E-06	5.93 [2.38, 12.2]	1.34E-06	2.74 [1.56, 4.81]
rs73677451 (G/t)	Intronic ZMAT4	3.70 [2.15, 6.38]	2.36E-06	3.72 [2.13, 6.52]	4.14E-06	2.08 [1.39, 3.09]
rs600075 (G/a)	Intronic PTPRD	3.59 [2.11, 6.09]	2.36E-06	3.55 [2.05, 6.16]	6.45E-06	2.30 [1.57, 3.37]
rs5896385 (-a)	Intronic PTPRD	3.59 [2.11, 6.09]	2.36E-06	3.55 [2.05, 6.16]	6.45E-06	2.30 [1.57, 3.37]

*First allele is the major allele and the second is the minor allele. Associations statistics (OR and p values) are based on the minor allele.

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.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The data have been deposited with links to BioProject accession number PRJNA1055092 in the NCBI BioProject database (<https://www.ncbi.nlm.nih.gov/bioproject/>). The datasets used and/or analyzed in this study are also available from the corresponding author on reasonable request.

Code availability

Analytical codes will be available upon request from the Corresponding author.

Received: 31 January 2024; Accepted: 9 May 2024;

Published online: 30 May 2024

References

1. McCrindle, B. W. et al. Diagnosis, treatment, and long-term management of Kawasaki disease: A scientific statement for health professionals From the American Heart Association. *Circulation* **135**, e927–e999 (2017).
2. Kawasaki, T. [Acute febrile mucocutaneous syndrome with lymphoid involvement with specific desquamation of the fingers and toes in children]. *Arerugi* **16**, 178–222 (1967).
3. Orenstein, J. M. et al. Three linked vasculopathic processes characterize Kawasaki disease: a light and transmission electron microscopic study. *PLoS One* **7**, e38998 (2012).
4. Kato, H. et al. Long-term consequences of Kawasaki disease. A 10- to 21-year follow-up study of 594 patients. *Circulation* **94**, 1379–1385 (1996).
5. Burns, J. C. & Glode, M. P. Kawasaki syndrome. *Lancet* **364**, 533–544 (2004).
6. Newburger, J. W. et al. A single intravenous infusion of gamma globulin as compared with four infusions in the treatment of acute Kawasaki syndrome. *N. Engl. J. Med.* **324**, 1633–1639 (1991).
7. Newburger, J. W. et al. The treatment of Kawasaki syndrome with intravenous gamma globulin. *N. Engl. J. Med.* **315**, 341–347 (1986).
8. Sleeper, L. A. et al. Evaluation of Kawasaki disease risk-scoring systems for intravenous immunoglobulin resistance. *J. Pediatr.* **158**, 831–835.e3 (2011).
9. Newburger, J. W. et al. Randomized trial of pulsed corticosteroid therapy for primary treatment of Kawasaki disease. *N. Engl. J. Med.* **356**, 663–675 (2007).
10. Miura, M. et al. Association of severity of coronary artery aneurysms in patients with kawasaki disease and risk of later coronary events. *JAMA Pediatr.* **172**, e180030 (2018).
11. McCrindle, B. W. et al. Medium-term complications associated with coronary artery aneurysms after Kawasaki disease: A study from the international kawasaki disease registry. *J. Am. Heart Assoc.* **9**, e016440 (2020).
12. Friedman, K. G. et al. Coronary artery aneurysms in Kawasaki disease: Risk factors for progressive disease and adverse cardiac events in the US population. *J. Am. Heart Assoc.* **5**, e003289 (2016).
13. Patel, A. S., Bruce, M., Harrington, W. & Portman, M. A. Coronary artery stenosis risk and time course in Kawasaki disease patients: Experience at a US tertiary pediatric centre. *Open Heart* **2**, e000206 (2015).
14. Newburger, J. W. et al. Summary and abstracts of the Seventh International Kawasaki Disease Symposium: December 4–7, 2001, Hakone, Japan. *Pediatr. Res.* **53**, 153–157 (2003).
15. Lee, J. J. Y. et al. The long-term cardiac and noncardiac prognosis of Kawasaki disease: A systematic review. *Pediatrics* **149**, e2021052567 (2022).
16. Hu, J. & Ren, W. Analysis of the risk factors in prognosis of Kawasaki disease with coronary artery lesions. *Front Pediatr.* **9**, 798148 (2021).
17. Lin, M. T. et al. Acute and late coronary outcomes in 1073 patients with Kawasaki disease with and without intravenous gamma-immunoglobulin therapy. *Arch. Dis. Child* **100**, 542–547 (2015).
18. Kobayashi, T. et al. Risk stratification in the decision to include prednisolone with intravenous immunoglobulin in primary therapy of Kawasaki disease. *Pediatr. Infect. Dis. J.* **28**, 498–502 (2009).
19. Davies, S. et al. Predicting IVIG resistance in UK Kawasaki disease. *Arch. Dis. Child* **100**, 366–368 (2015).
20. Fu, P. P., Du, Z. D. & Pan, Y. S. Novel predictors of intravenous immunoglobulin resistance in Chinese children with Kawasaki disease. *Pediatr. Infect. Dis. J.* **32**, e319–e323 (2013).
21. Shin, J., Lee, H. & Eun, L. Verification of current risk scores for Kawasaki disease in Korean children. *J. Korean Med. Sci.* **32**, 1991–1996 (2017).
22. Portman, M. A. & Shrestha, S. One size does not fit all: Genetic prediction of kawasaki disease treatment response in diverse populations. *Circ. Cardiovasc Genet* **10**, e001917 (2017).
23. Son, M. B. F. et al. Predicting coronary artery aneurysms in Kawasaki Disease at a North American Center: An assessment of Baseline z Scores. *J. Am. Heart Assoc.* **6**, e005378 (2017).
24. Makowsky, R. et al. FcgammaR gene copy number in Kawasaki disease and intravenous immunoglobulin treatment response. *Pharmacogenet Genomics* **23**, 455–462 (2013).
25. Shrestha, S. et al. Role of activating FcgammaR gene polymorphisms in Kawasaki disease susceptibility and intravenous immunoglobulin response. *Circ. Cardiovasc Genet* **5**, 309–316 (2012).
26. Shrestha, S. et al. Functional FCGR2B gene variants influence intravenous immunoglobulin response in patients with Kawasaki disease. *J. Allergy Clin. Immunol.* **128**, 677–680 (2011).
27. Wu, S. F., Chang, J. S., Wan, L., Tsai, C. H. & Tsai, F. J. Association of IL-1Ra gene polymorphism, but no association of IL-1beta and IL-4 gene polymorphisms, with Kawasaki disease. *J. Clin. Lab Anal.* **19**, 99–102 (2005).
28. Minami, T. et al. A polymorphism in plasma platelet-activating factor acetylhydrolase is involved in resistance to immunoglobulin treatment in Kawasaki disease. *J. Pediatr.* **147**, 78–83 (2005).
29. Burns, J. C. et al. Family-based association analysis implicates IL-4 in susceptibility to Kawasaki disease. *Genes Immun.* **6**, 438–444 (2005).
30. Burns, J. C. et al. Genetic variations in the receptor-ligand pair CCR5 and CCL3L1 are important determinants of susceptibility to Kawasaki disease. *J. Infect. Dis.* **192**, 344–349 (2005).
31. Khor, C. C. et al. Genome-wide association study identifies FCGR2A as a susceptibility locus for Kawasaki disease. *Nat. Genet.* **43**, 1241–1246 (2011).
32. Onouchi, Y. et al. A genome-wide association study identifies three new risk loci for Kawasaki disease. *Nat. Genet.* **44**, 517–521 (2012).
33. Elakabawi, K., Lin, J., Jiao, F., Guo, N. & Yuan, Z. Kawasaki disease: Global burden and genetic background. *Cardiol. Res* **11**, 9–14 (2020).
34. Holman, R. C. et al. Racial/ethnic differences in the incidence of Kawasaki syndrome among children in Hawaii. *Hawaii Med J.* **69**, 194–197 (2010).
35. Padilla, L. A. et al. Kawasaki disease and clinical outcome disparities among black children. *J. Pediatr.* **229**, 54–60 e2 (2021).

36. Shrestha, S. et al. Pharmacogenomics of intravenous immunoglobulin response in Kawasaki disease. *Front. Immunol.* **14**, 1287094 (2024).

37. Herberman, R. B., Nunn, M. E., Holden, H. T. & Lavrin, D. H. Natural cytotoxic reactivity of mouse lymphoid cells against syngeneic and allogeneic tumors. II. Characterization of effector cells. *Int J. Cancer* **16**, 230–239 (1975).

38. Kiessling, R., Klein, E. & Wigzell, H. “Natural” killer cells in the mouse. I. Cytotoxic cells with specificity for mouse Moloney leukemia cells. Specificity and distribution according to genotype. *Eur. J. Immunol.* **5**, 112–117 (1975).

39. Trinchieri, G. & Santoli, D. Anti-viral activity induced by culturing lymphocytes with tumor-derived or virus-transformed cells. Enhancement of human natural killer cell activity by interferon and antagonistic inhibition of susceptibility of target cells to lysis. *J. Exp. Med.* **147**, 1314–1333 (1978).

40. Long, E. O., Kim, H. S., Liu, D., Peterson, M. E. & Rajagopalan, S. Controlling natural killer cell responses: integration of signals for activation and inhibition. *Annu. Rev. Immunol.* **31**, 227–258 (2013).

41. Choi, I. S. et al. Circulating immune cell profile and changes in intravenous immunoglobulin responsiveness over the disease course in children With Kawasaki Disease. *Front Pediatr.* **9**, 792870 (2021).

42. Cheong, K. X., Yong, R. Y. Y., Tan, M. M. H., Tey, F. L. K. & Ang, B. C. H. Association of VIPR2 and ZMAT4 with high myopia. *Ophthalmic Genet* **41**, 41–48 (2020).

43. Wan, J. et al. The association between the copy-number variations of ZMAT4 and hematological malignancy. *Hematology* **16**, 20–23 (2011).

44. Helms, A. S. et al. Effects of MYBPC3 loss-of-function mutations preceding hypertrophic cardiomyopathy. *JCI Insight* **5**, e133782 (2020).

45. van Velzen, H. G. et al. Clinical characteristics and long-term outcome of hypertrophic cardiomyopathy in individuals with a MYBPC3 (myosin-binding protein C) founder mutation. *Circ. Cardiovasc. Genet* **10**, e001660 (2017).

46. Al-Khelaifi, F. et al. Genome-wide association study reveals a novel association between MYBPC3 gene polymorphism, endurance athlete status, aerobic capacity and steroid metabolism. *Front Genet* **11**, 595 (2020).

47. Kuo, H. C. et al. Genome-wide association study identifies novel susceptibility genes associated with coronary artery aneurysm formation in Kawasaki disease. *PLoS One* **11**, e0154943 (2016).

48. Beckley, M. A., Shrestha, S., Singh, K. K. & Portman, M. A. The role of mitochondria in the pathogenesis of Kawasaki disease. *Front Immunol.* **13**, 1017401 (2022).

49. Hoggart, C. et al. Identification of novel locus associated with coronary artery aneurysms and validation of loci for susceptibility to Kawasaki disease. *Eur. J. Hum. Genet* **29**, 1734–1744 (2021).

50. Kim, J. J. et al. Identification of KCNN2 as a susceptibility locus for coronary artery aneurysms in Kawasaki disease using genome-wide association analysis. *J. Hum. Genet* **58**, 521–525 (2013).

51. Liu, J., Yuan, P., Pang, Y. & Su, D. ITPKC polymorphism (rs7251246 T > C), coronary artery aneurysms, and thrombosis in patients with Kawasaki disease in a Southern Han Chinese population. *Front Immunol.* **14**, 1184162 (2023).

52. Oates-Whitehead, R. M. et al. Intravenous immunoglobulin for the treatment of Kawasaki disease in children. *Cochrane Database Syst. Rev.* **2003**, CD004000 (2003).

53. Shimizu, C. et al. Genetic variation in the SLC8A1 calcium signaling pathway is associated with susceptibility to Kawasaki disease and coronary artery abnormalities. *Circ. Cardiovasc Genet* **9**, 559–568 (2016).

54. Shimizu, C. et al. Matrix metalloproteinase haplotypes associated with coronary artery aneurysm formation in patients with Kawasaki disease. *J. Hum. Genet* **55**, 779–784 (2010).

55. Huang, Y. H. et al. Decreased DNA methyltransferases expression is associated with coronary artery lesion formation in Kawasaki disease. *Int J. Med. Sci.* **16**, 576–582 (2019).

56. Kwon, Y. C. et al. Identification of the TIFAB gene as a susceptibility locus for coronary artery aneurysm in patients with Kawasaki disease. *Pediatr. Cardiol.* **40**, 483–488 (2019).

57. Lin, Y. J. et al. Genetic variants in PLCB4/PLCB1 as susceptibility loci for coronary artery aneurysm formation in Kawasaki disease in Han Chinese in Taiwan. *Sci. Rep.* **5**, 14762 (2015).

58. Wu, J., Pfeiffer, R. M. & Gail, M. H. Strategies for developing prediction models from genome-wide association studies. *Genet Epidemiol.* **37**, 768–777 (2013).

59. Newburger, J. W. et al. Diagnosis, treatment, and long-term management of Kawasaki disease: A statement for health professionals from the Committee on Rheumatic Fever, Endocarditis, and Kawasaki Disease, Council on Cardiovascular Disease in the Young, American Heart Association. *Pediatrics* **114**, 1708–1733 (2004).

60. McCrindle, B. W. et al. Coronary artery involvement in children with Kawasaki disease: Risk factors from analysis of serial normalized measurements. *Circulation* **116**, 174–179 (2007).

61. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).

62. Consortium, G. T. Human genomics. The genotype-tissue expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).

63. van Berkum, N. L. et al. Hi-C: a method to study the three-dimensional architecture of genomes. *J. Vis. Exp.* **6**, 1869 (2010).

64. Kircher, M. et al. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet* **46**, 310–315 (2014).

65. Roadmap Epigenomics, C. et al. Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).

Acknowledgements

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.

Author contributions

SS: Conceptualization, Study design, Funding acquisition, Methodology, Analysis, Overall supervision, Results validation & visualization, Manuscript writing – original draft. HW: Data management & pre-processing, WGS analysis, Writing – review & editing. SC: Results tabulation and visualization, Writing – review & editing. HK: Clinical data management, Biospecimen handling, Writing – review & editing. VS: Data pre-processing, Bioinformatics, Results tabulation and visualization, Writing – review & editing. OAM: Sample processing, DNA extraction & quantification, Writing – review & editing. UNB: Results tabulation and visualization, Writing – review & editing. DL: Clinical data management, Project administration, Writing – review & editing. YL: Phenotyping and clinical validation, Writing – review & editing, LAP: Data management, Methodology, Writing – review & editing. JC: Supervision of genomic data pre-processing, Writing – review & editing. ND: Phenotyping and clinical validation, Methodology, Writing – review & editing. HT: Supervision of genomic data pre-processing & analysis, Methodology, Writing – review & editing. MP: Conceptualization, Phenotyping and clinical validation, Funding acquisition, Investigation, Methodology, Clinical Supervision, Results validation & visualization, Writing – review & editing

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41525-024-00419-7>.

Correspondence and requests for materials should be addressed to Sadeep Shrestha.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024