



## Association of an *IGHV3-66* gene variant with Kawasaki disease

Todd A. Johnson<sup>1</sup> · Yoichi Mashimo<sup>2</sup> · Jer-Yuarn Wu<sup>3</sup> · Dankyu Yoon<sup>4</sup> · Akira Hata<sup>2</sup> · Michiaki Kubo<sup>5</sup> · Atsushi Takahashi<sup>6,7</sup> · Tatsuhiko Tsunoda<sup>1,8,9</sup> · Kouichi Ozaki<sup>10,11</sup> · Toshihiro Tanaka<sup>10,12</sup> · Kaoru Ito<sup>10</sup> · Hiroyuki Suzuki<sup>13</sup> · Hiromichi Hamada<sup>14</sup> · Tohru Kobayashi<sup>15</sup> · Toshiro Hara<sup>16</sup> · Chien-Hsiun Chen<sup>3</sup> · Yi-Ching Lee<sup>17</sup> · Yi-Min Liu<sup>3</sup> · Li-Ching Chang<sup>3</sup> · Chun-Ping Chang<sup>3</sup> · Young-Mi Hong<sup>18</sup> · Gi-Young Jang<sup>19</sup> · Sin-Weon Yun<sup>20</sup> · Jeong-Jin Yu<sup>21</sup> · Kyung-Yil Lee<sup>22</sup> · Jae-Jung Kim<sup>23</sup> · Taesung Park<sup>24</sup> · Korean Kawasaki Disease Genetics Consortium, Taiwan Kawasaki Disease Genetics Consortium, Taiwan Pediatric ID Alliance, Japan Kawasaki Disease Genome Consortium · Jong-Keuk Lee<sup>23</sup> · Yuan-Tsong Chen<sup>3</sup> · Yoshihiro Onouchi<sup>2,10</sup>

Received: 17 July 2020 / Revised: 13 October 2020 / Accepted: 13 October 2020 / Published online: 26 October 2020  
© The Author(s), under exclusive licence to The Japan Society of Human Genetics 2020

### Abstract

In a meta-analysis of three GWAS for susceptibility to Kawasaki disease (KD) conducted in Japan, Korea, and Taiwan and follow-up studies with a total of 11,265 subjects (3428 cases and 7837 controls), a significantly associated SNV in the *immunoglobulin heavy variable* gene (*IGHV*) cluster in 14q33.32 was identified (rs4774175; OR = 1.20,  $P = 6.0 \times 10^{-9}$ ). Investigation of nonsynonymous SNVs of the *IGHV* cluster in 9335 Japanese subjects identified the C allele of rs6423677, located in *IGHV3-66*, as the most significant reproducible association (OR = 1.25,  $P = 6.8 \times 10^{-10}$  in 3603 cases and 5731 controls). We observed highly skewed allelic usage of *IGHV3-66*, wherein the rs6423677 A allele was nearly abolished in the transcripts in peripheral blood mononuclear cells of both KD patients and healthy adults. Association of the high-expression allele with KD strongly indicates some active roles of B-cells or endogenous immunoglobulins in the disease pathogenesis. Considering that significant association of SNVs in the *IGHV* region with disease susceptibility was previously known only for rheumatic heart disease (RHD), a complication of acute rheumatic fever (ARF), these observations suggest that common B-cell related mechanisms may mediate the symptomology of KD and ARF as well as RHD.

These authors contributed equally: Todd A. Johnson, Yoichi Mashimo, Jer-Yuarn Wu, Dankyu Yoon

Members of the Korean Kawasaki Disease Genetics Consortium, Taiwan Kawasaki Disease Genetics Consortium, Taiwan Pediatric ID Alliance, Japan Kawasaki Disease Genome Consortium are listed in the Supplementary information.

**Supplementary information** The online version of this article (<https://doi.org/10.1038/s10038-020-00864-z>) contains supplementary material, which is available to authorized users.

- ✉ Jer-Yuarn Wu  
jywu@ibms.sinica.edu.tw
- ✉ Jong-Keuk Lee  
cookie\_jklee@hotmail.com
- ✉ Yuan-Tsong Chen  
chen0010@ibms.sinica.edu.tw
- ✉ Yoshihiro Onouchi  
onouchy@chiba-u.jp

Extended author information available on the last page of the article

### Introduction

Kawasaki disease (KD) is an acute systemic vasculitis syndrome characterized by high fever, bilateral conjunctivitis, polymorphous skin rash, reddening of lips and oral cavity, changes in extremities and nonsuppurative cervical lymphadenopathy [1] and predominantly affects infants and children younger than 5 years [2]. In many cases, KD is self-limiting. However, it causes coronary artery complications such as dilatations and aneurysms (coronary artery lesions; CALs) in 20–25% of untreated patients [3]. Replacing acute rheumatic fever (ARF), KD is now the leading cause of acquired heart diseases of children in developed countries [4]. Most KD patients are treated with high dose intravenous immunoglobulin (IVIG) infusion combined with oral aspirin, which was established in the 1980s and 1990s and has become a standard treatment that is effective at resolving inflammation and reducing CALs [5–7]. However, the mechanism of IVIG action on KD has not been revealed, and

10–15% of KD patients do not respond to the treatment and have a higher risk for CAL. Recently a series of genome-wide association studies (GWAS) revealed several definitive KD susceptibility loci [8–10]. However, these genetic factors can explain only a part of the etiology. Also, the reason for the high incidence among East Asian children [11], which is up to 20-fold higher than those in Western countries and is a crucial epidemiologic feature of KD, has not yet been explained. It might be attributed in part to lack of statistical power in the previous GWAS analyses that were carried out using modest sample-sizes, and therefore, many common genetic factors with relatively low penetrance may have gone undetected. In this study, to identify novel susceptibility loci for KD, we conducted a meta-analysis of GWAS from Japan, Korea, and Taiwan, the three countries with the highest incidence of KD in the world.

## Materials and methods

### Subjects and design of association analyses in identifying novel susceptibility loci for KD

In this study, susceptibility loci for KD were screened and verified in a two-stage association analysis (Fig. 1). Stage 1 was a whole-genome meta-analysis of results from GWAS conducted in Japan [9], Taiwan [10], and Korea [12] involving 6362 individuals. We carried out follow-up association studies using three case-control panels comprised of 1418 KD patients and 1700 controls (Japanese), 473 KD patients and 484 controls (Korean) and 261 KD patients and 567 controls (Taiwanese) independent from the subjects in the GWAS and performed meta-analyses with the Stage 1 results (Stage 2). Loci for the follow-up studies were selected based on their predicted potential to achieve

**Fig. 1** Flow of the screening of the novel susceptibility loci for KD in this study

#### Stage 1

##### Whole genome imputation and a meta-analysis of three GWAS results

###### GWAS genotype data

**Japan**  
Illumina HumanHap 550k/610k  
KD ct. = 428, Control ct. = 3,379

**Taiwan**  
Affymetrix SNP 6.0  
KD ct. = 662, Control ct. = 1,107

**Korea**  
Affymetrix SNP 6.0  
KD ct. = 186, Control ct. = 600

**Quality control**  
 $HWE-P \geq 0.0001$ ,  
Call rate  $\geq 0.95$  &  
MAF  $\geq 0.01$

- 1) **Imputation**  
Impute2 & minimac softwares  
1000genomes East Asian 572 haplotypes
- 2) **Quality control**  
a)  $MAF_{cases} \geq 0.01$  &  $MAF_{controls} \geq 0.01$   
b)  $Info \geq 0.4$
- 3) **Genome-wide meta-analysis**  
6,265,045 autosomal SNVs
- 4) **Selection SNVs with  $P < 0.001$**   
**15,290 SNVs** → Stage 2

#### Stage 2

##### Follow-up studies of marginally associated SNVs in the Stage 1 analysis

**Selection SNVs with high probability of achieving  $P < 5.0 \times 10^{-8}$**

- 1) **Creation of hypothetical case control panels**
  1. Creation of virtual case and control cohorts (100,000 individuals each) for each population that have the minor allele frequencies of the candidate SNVs same with the GWAS samples.
  2. Random sampling of hypothetical case controls from the virtual cohorts making their sizes same with the follow-up panels prepared in each population.
- 2) **Association study of the candidate SNVs in the hypothetical case control panels**
- 3) **Evaluation of P values in meta-analyses with the Stage 1 association results**
- 4) **Iteration of the 1)-3) steps 100 times**
- 5) **Selection of SNVs showing  $P < 5.0 \times 10^{-8}$  more than 80 times → 49 SNVs**
- 6) **Defining the groups of SNVs linked with the 49 SNVs selected in 5) ( $r^2 > 0.1$ )**
- 7) **Pick-up representative SNVs for genotyping**  
**48 SNVs** (1 SNV failed in assay design)

###### 8) Follow-up association studies

###### Sample panels for follow-up studies

**Japan**  
Invader  
KD ct. = 1,418, Control ct. = 1,700

**Taiwan**  
MassARRAY or TaqMan  
KD ct. = 261, Control ct. = 567

**Korea**  
Golden Gate or TaqMan  
KD ct. = 473, Control ct. = 484

###### 9) Meta-analyses with the GWAS data

$P$  values less than  $5.0 \times 10^{-8}$  in the meta-analyses, with the prediction made by iterative simulations of follow-up studies with virtual case and control cohorts (detailed below). To further validate the associations of rs4774175 and rs6423677 in 14q32.33, an additional 1758 KD cases and 653 controls collected in Japan were used. The number of KD cases and controls, as well as platforms in the three previous GWAS and follow-up studies in Japan, Korea, and Taiwan are summarized in Supplementary Table 1.

### Whole-genome imputation and meta-analyses

For the Stage 1 analysis, each study center's genotype data for the Illumina Human Hap550/610 or Affymetrix SNP 6.0 arrays (Supplementary Table 1) was oriented to the forward strand of the hg19 human reference genome. Genotype data were filtered for minimum quality-control parameter cutoffs such as HWE- $P \geq 0.0001$ , Call-rate  $\geq 0.95$ , and minor allele frequency (MAF)  $\geq 0.01$ . Each study center performed whole-genome imputation of autosomal variants with 572 East Asian haplotypes from the 1000 Genomes Project Phase 1 Version 3 (ref. [13]) as reference using pre-phasing with SHAPEIT v1 (ref. [14]) and imputation with IMPUTE2 and minimac softwares [15, 16].

Imputed genotype data were analyzed by each study center in a case-control logistic regression analysis, and the output was merged in the R statistics environment (URL: <https://www.R-project.org/>) and filtered for variants that were polymorphic in both cases and controls (MAF<sub>cases</sub>  $\geq 0.01$  and MAF<sub>controls</sub>  $\geq 0.01$ ) and had info  $\geq 0.4$  in all three studies' dataset; there were 6,265,045 SNVs in the final filtered dataset. A fixed-effects meta-analysis of the three data sets' beta-coefficients and standard errors was performed using the R package metafor (<https://cran.r-project.org/package=metafor>).

### Linkage disequilibrium-based region definition

Each chromosome's SNVs were filtered for those with a meta-analysis  $P < 0.01$  and then labelled based on linkage disequilibrium to regional "top SNVs". Briefly, SNVs with  $P < 1 \times 10^{-4}$  were sorted by  $P$  value, the top SNV identified, and any SNVs within  $\pm 5$  Mb that had  $r^2 > 0.1$  to that top SNV assigned to its region; data on a chromosome was processed as such until no SNVs with  $P < 1 \times 10^{-4}$  were remaining. Each top SNV region was labelled based on chromosome and minimum and maximum positions of the linked SNVs. Based on that process, each region label would be unique, but regions could overlap with each other. Calculation of  $r^2$  was performed using the `ld` function in the R Bioconductor `snpStats` package using East Asian genotype data from the 1000 Genomes Phase 1 Version 3 release.

### P value simulation

For simplicity, we will refer to the regions of nominally linked SNVs described above as "loci". To perform the Stage 2 follow-up study efficiently, loci that had a high potential of achieving  $P$  values less than  $5.0 \times 10^{-8}$  in a meta-analysis of Stage 1 and 2 results were selected by  $P$  value simulation as follows. For each locus, we first selected any nominally associated SNVs ( $P < 0.001$ ) identified in the Stage 1 analysis, and for each SNV created a virtual set of 100,000 case genotypes and a virtual set of 100,000 control genotypes for each of the three Stage 1 case-control panels. Each virtual set was generated based on the genotype frequencies observed in that panel's Stage 1 cases or controls at a particular SNV, and each of those virtual sets was then re-sampled in R to randomize their order. Each virtual case and control set was then sampled 100 times to create 100 virtual case-control cohorts; the numbers of cases and controls sampled from the virtual genotype sets were the case-control counts that were expected in the three collaborators' follow-up studies (RIKEN in Japan:  $n_{cases} = 1300$ ,  $n_{controls} = 1300$ ; Academia Sinica in Taiwan:  $n_{cases} = 500$ ,  $n_{controls} = 2000$ ; Asan Medical Center, University of Ulsan in Korea:  $n_{cases} = 412$ ,  $n_{controls} = 600$ ). For each iteration, associations of the candidate SNVs were evaluated in a meta-analysis of the virtual cohorts and the three GWAS data sets. For each candidate SNV, the frequency of observing  $P$  values of  $5.0 \times 10^{-8}$  or smaller in 100 iterations of the simulated meta-analysis was scored as the simulation score. Loci with at least one SNV having simulation scores of 0.8 or higher were considered to be promising, and for de novo genotyping, a representative SNV was selected for which assays were designable across the different platforms employed in each research center (Invader in Japan, Sequenom MassARRAY or TaqMan in Taiwan and VeraCode GoldenGate Genotyping kit or TaqMan in Korea, respectively) (Supplementary Table 1).

### Genotyping of SNVs in *IGHV* genes

Nonsynonymous SNVs in *IGHV* genes were genotyped basically by the Invader Assay. Primers and the probes were carefully designed in order to ensure specificity of the assay. We refrained from using multiplex PCR to avoid both expected and unexpected nonspecific amplification of DNA fragments of high sequence homology which will allow cross reaction between amplicons and probes for different loci. Sequences of the primers and probes for 18 nonsynonymous SNVs in *IGHV* genes and rs4774175, as well as representative genotyping results of rs6423677, are provided in Supplementary Tables 2 and 3 and Supplementary Fig. 1, respectively.

**Table 1** Loci achieving  $P$  values  $< 5.0 \times 10^{-8}$  in the meta-analyses of Stage 1 and 2 data

dbSNP ID	Chr: location <sup>b</sup> Closet genes Position within gene	Alleles Ref/Alt	Population	Association statistics					
				Stage 1		Stage 2		Meta-analysis <sup>a</sup>	
				OR	$P$	OR	$P$	OR (95% CI)	$P$ value
rs2720378	4: 185568113 <i>CASP3</i> intron	C/G	Japan	0.64	$2.5 \times 10^{-8}$	0.85	$1.7 \times 10^{-3}$	0.83 (0.78–0.89)	$1.6 \times 10^{-8}$
			Korea	0.83	0.15	1.00	0.96		
			Taiwan	0.81	$6.6 \times 10^{-3}$	1.03	0.78		
rs1873212	6: 27869631 <i>TRX-CAT1-7</i> 3' flanking	C/T	Japan	1.35	$1.7 \times 10^{-4}$	1.20	$1.8 \times 10^{-3}$	1.27 (1.18–1.36)	$2.0 \times 10^{-10}$
			Korea	1.43	0.017	1.34	0.013		
			Taiwan	1.29	0.012	1.17	0.27		
rs1778477	6: 28248594 <i>PGBD1</i> 5' flanking	A/T	Japan	0.79	0.28	0.76	$7.1 \times 10^{-5}$	0.77 (0.70–0.83)	$7.2 \times 10^{-10}$
			Korea	0.66	$7.2 \times 10^{-3}$	0.83	0.094		
			Taiwan	0.72	$2.9 \times 10^{-3}$	0.84	0.31		
rs1264516	6: 30411903 <i>LOC105375012</i> 3' flanking	C/A	Japan	1.31	$5.2 \times 10^{-4}$	1.21	$2.8 \times 10^{-4}$	1.24 (1.16–1.32)	$3.9 \times 10^{-11}$
			Korea	1.38	$8.0 \times 10^{-3}$	1.29	$3.5 \times 10^{-3}$		
			Taiwan	1.20	0.27	1.11	0.35		
rs2857602	6: 31533378 <i>LTA</i> intron	G/A	Japan	0.73	$8.9 \times 10^{-6}$	0.77	$2.5 \times 10^{-7}$	0.78 (0.73–0.83)	$2.8 \times 10^{-15}$
			Korea	0.67	$1.2 \times 10^{-3}$	0.81	0.027		
			Taiwan	0.86	0.059	0.84	0.13		
rs3129960	6: 32300809 <i>C6orf10/ LOC101929163</i> intron/intron	A/G	Japan	0.72	$3.5 \times 10^{-4}$	0.66	$7.1 \times 10^{-8}$	0.73 (0.67–0.80)	$6.7 \times 10^{-13}$
			Korea	0.84	0.37	0.83	0.22		
			Taiwan	0.75	$1.7 \times 10^{-3}$	0.83	0.17		
rs7775228	6: 32658079 <i>HLA-DQB1</i> 5' flanking	T/C	Japan	1.29	$5.4 \times 10^{-4}$	1.36	$3.1 \times 10^{-9}$	1.26 (1.19–1.35)	$7.0 \times 10^{-13}$
			Korea	1.07	0.60	1.05	0.65		
			Taiwan	1.35	$2.0 \times 10^{-4}$	1.08	0.54		
rs2071473	6: 32782605 <i>HLA-DOB</i> intron	C/T	Japan	1.39	$8.4 \times 10^{-6}$	1.29	$1.2 \times 10^{-6}$	1.26 (1.18–1.34)	$6.7 \times 10^{-13}$
			Korea	1.32	0.23	1.30	$6.1 \times 10^{-3}$		
			Taiwan	1.09	0.26	1.15	0.22		
rs2736340	8: 11343973 <i>BLK</i> 5' flanking	C/T	Japan	1.71	$1.3 \times 10^{-9}$	1.51	$6.2 \times 10^{-13}$	1.55 (1.44–1.66)	$6.6 \times 10^{-33}$
			Korea	1.48	$6.1 \times 10^{-3}$	1.48	$3.8 \times 10^{-4}$		
			Taiwan	1.56	$2.0 \times 10^{-7}$	1.55	$7.6 \times 10^{-4}$		
rs4774175	14: 107152027 <i>IGHV1-69</i> 3' flanking	G/A	Japan	1.22	$4.9 \times 10^{-3}$	1.22	$1.7 \times 10^{-4}$	1.20 (1.13–1.28)	$6.0 \times 10^{-9}$
			Korea	1.40	$7.6 \times 10^{-3}$	1.20	0.059		
			Taiwan	1.16	0.042	1.06	0.63		
rs28493229	19: 41224204 <i>ITPKC</i> intron	G/C	Japan	1.50	$3.4 \times 10^{-6}$	1.38	$1.3 \times 10^{-6}$	1.41 (1.29–1.53)	$1.1 \times 10^{-14}$
			Korea	0.98	0.91	1.25	0.14		
			Taiwan	1.97	$1.9 \times 10^{-6}$	1.24	0.27		
rs1883832	20: 44746982 <i>CD40</i> 5' UTR	T/C	Japan	1.34	$1.4 \times 10^{-4}$	1.18	$2.1 \times 10^{-3}$	1.28 (1.20–1.36)	$1.5 \times 10^{-13}$
			Korea	1.07	0.61	1.32	$5.1 \times 10^{-3}$		
			Taiwan	1.41	$4.9 \times 10^{-6}$	1.42	$2.3 \times 10^{-3}$		

For all single nucleotide variants (SNVs), association in the logistic regression model was evaluated and odds ratios were calculated for the effects of the alternative alleles

Chr chromosomes, Ref reference allele, Alt alternative allele, OR odds ratio, CI confidence interval, UTR untranslated region

<sup>a</sup>Combined odds ratios and  $P$  values were calculated using fixed-effects meta-analytic model

<sup>b</sup>Location of the SNVs are derived from NCBI human genome reference sequence Build 37.3

## Next-generation sequencing (NGS) of *IGHV* repertoires

Two milliliters of venous blood was drawn from patients who were admitted to the hospitals for KD at four time points including (1) acute phase before receiving IVIG (3–8 illness days), (2) 48 h after the patients became afebrile (8–18 illness days), (3) the first follow-up visit to the pediatric clinic after discharge (17–50 days after the disease onset), and (4) the second follow-up visit to the pediatric clinic after discharge (3–4 months after the disease onset). Blood samples were collected into Vacutainer CPT Cell Preparation Tube (BD) and mononuclear cells were separated according to the manufacturer's instruction. Total RNA from the mononuclear cells was extracted by using the RNeasy Mini Kit (QIAGEN). 1.0 µg of RNA was reverse transcribed with PrimeScript (TAKARA) and the mixed oligonucleotides of random hexamer and oligo-dT primers. Isotype-specific libraries for NGS were prepared as follows. Mixed forward primers covering the framework region 1 of 7 subgroups of *IGHV* genes (V1–V7) [17] and reverse primers specific to each *IGHC* gene for IgM, IgD, IgG, and IgA (including a partial Illumina adapter sequence in the 5' ends of both primers) were designed for the 1st round PCR. Sequences of the primers are provided in Supplementary Table 4. 6-base barcode sequence and the full Illumina adapter sequence were added at 5' and 3' ends of the immunoglobulin amplicons in the 2nd round PCR. The barcode sequences were used to distinguish the patients and the sampling time points. The libraries were sequenced with MiSeq Reagent Kit v3 (600-cycle) (Illumina).

Forward and reverse sequence reads were merged by using FLASH software [18]. Sequences unmerged due to insufficient overlapping length were excluded from subsequent analyses. Merged sequence reads were classified into isotypes and subclasses based on primer sequences by using blast software (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>), and then quality filtering and removing the primer sequence were performed by using the FASTX-Toolkit ([http://hannonlab.cshl.edu/fastx\\_toolkit/index.html](http://hannonlab.cshl.edu/fastx_toolkit/index.html)).

## Repertoire analyses of immunoglobulin heavy chains

Immunoglobulin repertoires were determined by migmap-1.0.2 software (<https://github.com/mikessh/migmap>). Sequence reads that terminated prematurely or had a complementarity determining region (CDR) 3 that was non-canonical (i.e., lacking consensus amino acids at both ends or not fully mapped) were excluded from subsequent analyses. Clonotypes were defined by combinations of *IGHV*, *IGH diversity (IGHD)*, and *IGH joining (IGHJ)* gene alleles. Correlation trend between the proportions of

the *IGH* clonotypes using *IGHV3-66* and the C allele counts at rs6423677 was evaluated by the Jonckheere-Terpstra test using R package PMCMR (<https://cran.r-project.org/package=PMCMR>). Change of each clonotype proportion from baseline (the 1st sampling point) was calculated by subtracting the baseline proportion from those at 2nd, 3rd, and 4th points. The diversity of CDR3 clonotypes was evaluated by the inverse Simpson's index using the R package vegan (<https://cran.r-project.org/package=vegan>).

## Results

### Significant associations in the meta-analyses of three GWAS data sets

In the Stage 1 analysis, a meta-analysis of three GWAS data sets, genome-wide level associations were seen in known susceptibility loci such as *FAM167A-BLK* (rs2736340;  $P = 1.23 \times 10^{-16}$ ), *ITPKC* (rs28493229;  $P = 3.07 \times 10^{-9}$ ), *CASP3* (rs2720377;  $P = 2.66 \times 10^{-9}$ ) and *CD40* (rs1883832;  $P = 1.76 \times 10^{-8}$ ). Four top SNVs in *HLA class III~ class II* regions (rs2844485, rs73729123, rs7739458, and rs146650659) and rs13330932 in the 16q24.1 region also showed significant associations (Supplementary Table 5, Supplementary Fig. 2). From a standard test for inflation, the genomic control inflation factor  $\lambda_{GC} = 1.0216$  (Supplementary Fig. 2).

### Follow-up of analyses of 49 most promising loci

In a simulation of meta-analyses, a “simulation score” was calculated for each nominally associated SNV ( $P \leq 0.001$ ; see Methods). Forty-nine loci were identified with one or more SNVs expected to satisfy a  $P$  value threshold of  $5.0 \times 10^{-8}$  with simulation score of 0.8 or higher (Supplementary Table 6). From every locus, we selected representative SNVs for genotyping in replication sample panels and then performed a meta-analysis with data from the three GWAS. For one particular SNV locus (rs181511609) that satisfied the criteria, a genotyping assay could not be designed due to the complexity of the surrounding sequence. As a result, 12 out of 48 examined loci showed significant associations (Table 1, Supplementary Table 7). These included four previously known susceptibility gene loci such as *CASP3*, *FAM167A-BLK*, *ITPKC*, and *CD40*, seven on chromosome 6p21 (NC\_000006.11: 25.5–32.78 Mb) and one in the *immunoglobulin heavy variable* gene (*IGHV*) region on chromosome 14q32.33. The trend of association for rs13330932 on chromosome 16q24.1 was not replicated in the follow-up sample panels from the three populations (Supplementary Table 7).

**Table 2** Association of the 18 nonsynonymous SNVs of *IGHV* genes as candidates and rs4774175

SNVs	Position	Alleles Ref/Alt <sup>a</sup>	Gene	$r^2$ with rs6423677	Distance from rs6423677	RAF KD <sup>b</sup> /Ctrl <sup>b</sup>	$P_{HWE}$	Association statistics OR (95% CI)	
								Single-SNP	Conditional <sup>c</sup>
rs61999676	106926328	A/C	V3-43	0.15	204836	0.33/0.32	0.60	1.07 (0.97–1.20)	0.96 (0.85–1.08)
rs79008247	106926456	A/C	V3-43	0.16	204708	0.33/0.31	0.21	1.11 (0.99–1.86)	1.00 (0.88–1.12)
rs7141669	106963101	C/T	V1-45	0.17	168063	0.34/0.33	0.98	0.063	0.95
rs7148607	106993945	T/C	V3-48	0.17	137219	0.36/0.35	0.78	1.08 (0.97–1.20)	0.96 (0.86–1.09)
rs2073674	107013129	A/C	V3-49	0.24	118035	0.41/0.39	0.16	1.09 (0.98–1.20)	0.97 (0.86–1.09)
rs2073673	107013201	G/T	V3-49	0.23	117963	0.59/0.58	0.18	0.12	0.60
rs2516904	107078567	T/C	V1-58	0.68	52597	0.54/0.49	0.03	1.07 (0.97–1.19)	0.93 (0.83–1.05)
rs72483334	107099262	T/C	V3-62	0.74	31902	0.54/0.49	0.24	0.17	0.26
rs113692972	107099328	A/T	V3-62	0.74	31836	0.53/0.49	0.21	1.03 (0.93–1.14)	0.89 (0.79–1.00)
rs2072045	107113851	C/T	V3-64	0.76	17313	0.54/0.49	0.19	0.59	0.044
rs11846079	107114009	T/C	V3-64	0.75	17155	0.54/0.49	0.49	1.20 (1.09–1.32)	0.96 (0.80–1.15)
rs2073670	107114193	A/G	V3-64	0.76	16971	0.54/0.49	0.31	$3.0 \times 10^{-4}$	0.65
rs6423677	107131164	A/C	V3-66	NA	0	0.49/0.43	0.09	1.22 (1.10–1.34)	0.97 (0.79–1.00)
rs149638514	107131290	T/C	V3-66	0.90	126	0.50/0.45	0.23	$1.1 \times 10^{-4}$	0.79
rs4774175	107152027	G/A	Intergene	0.85	20863	0.50/0.45	0.56	1.20 (1.08–1.32)	0.92 (0.76–1.13)
rs8009570	107170055	G/A	V1-69	0.72	38891	0.54/0.49	0.32	$3.0 \times 10^{-4}$	0.44
rs55891010	107170062	A/G	V1-69	0.67	38898	0.52/0.47	0.11	1.20 (1.08–1.32)	0.90 (0.73–1.10)
rs11845244	107170077	C/T	V1-69	0.81	38913	0.52/0.47	0.48	$3.6 \times 10^{-4}$	0.30
								$4.3 \times 10^{-4}$	0.89 (0.72–1.10)
								$3.2 \times 10^{-5}$	0.28
								$4.5 \times 10^{-6}$	0.36
								1.26 (1.14–1.39)	NA
								$1.0 \times 10^{-4}$	0.74 (0.52–1.04)
								1.22 (1.10–1.35)	0.09
								$1.7 \times 10^{-4}$	0.89 (0.68–1.17)
								$5.7 \times 10^{-5}$	0.40
								$1.23 (1.11–1.35)$	1.01 (0.82–1.23)
								$4.4 \times 10^{-5}$	0.94
								$1.24 (1.12–1.37)$	1.05 (0.87–1.26)
								$1.23 (1.12–1.36)$	0.63
								$3.4 \times 10^{-5}$	0.99 (0.77–1.26)

Table 2 (continued)

SNVs	Position	Alleles Ref/Alt <sup>a</sup>	Gene	<i>r</i> <sup>2</sup> with rs6423677	Distance from rs6423677	RAF KD <sup>b</sup> /Ctrl <sup>b</sup>	<i>P</i> <sub>HWE</sub>	Association statistics OR (95% CI)	
								Single-SNP	Conditional <sup>c</sup>
rs2073669	107178965	A/C	V2-70	0.84	47801	0.51/0.46	0.80	1.23 (1.11–1.36)	0.91 (0.69–1.20)
								6.1 × 10 <sup>-5</sup>	0.51

SNV single nucleotide variant, Ref reference allele, Alt alternative allele, RAF risk allele frequency, Ctrl control, OR odds ratio, CI confidence interval, P<sub>HWE</sub> *p* value for deviation from Hardy–Weinberg equilibrium in the control population

<sup>a</sup>Alleles associated with KD are underlined

<sup>b</sup>Japanese KD cases (*n* = 1418) and controls (*n* = 1700) prepared as a follow-up sample panel

<sup>c</sup>Conditional logistic regression analyses

### Association signals that newly achieved genome-wide significance after the meta-analyses of three GWAS and follow-up association studies

A significant association of rs2857151 located near *HLA-DOB* and *HLA-DQB2* genes with KD susceptibility was reported in the earlier Japanese GWAS [9]. However, the present study revealed that the association statistics for rs2857151 were not consistent across the other East Asian populations (Supplementary Table 8). Instead, 7 out of 12 groups of SNVs in the 6p21 region examined in the Stage 2 analyses showed significant association in the meta-analyses of the data sets in the three GWAS as well as in the follow-up studies (Table 1 and Supplementary Fig. 3A). To verify that those seven signals were associated independently from rs2857151, we calculated LD between these SNVs and rs2857151, and we also performed logistic regression analyses conditioning on rs2857151 using the Japanese Stage 2 sample set. Consistent with information from 1000 Genomes, the seven candidates were not in high LD with each other (Supplementary Fig. 3B). The association was most significant for rs2857151, and in the conditioned logistic regression analyses, the *P* value for rs2071473, which was located just 19 kb distal to and in marginal LD with rs2857151 (*D'* = 0.99, *r*<sup>2</sup> = 0.36), became nonsignificant (*P* > 0.05). After conditioning, *P* values for the other SNVs including rs2857151 itself increased but were still significant (*P* ≤ 0.05). (Supplementary Table 9). These included SNVs with previous information on their functional significance or association with diseases such as rs2857602, where the associated allele is linked to the A allele at rs2239704 (*r*<sup>2</sup> = 1.0 in the 1000 Genomes JPT population), which has been associated with reduced expression of lymphotoxin A [19], and rs7775228, which was previously associated with asthma in a Japanese population [20]. Thus, further investigation including that in each ethnic group is needed to unravel the involvement of the variants in this region in KD susceptibility.

In the meta-analysis of the Stage 1 and 2 studies, a significant association was also obtained for an SNV within the chromosome 14q32.33 region represented by rs4774175 (*P* = 6.0 × 10<sup>-9</sup>) (Table 1). Five hundred and nineteen SNVs, linked with rs4774175 (*r*<sup>2</sup> > 0.1), and having simulation score of 0.5 or higher were distributed across a 340 kb region (106.88–107.22 Mb) of this locus where many *IGHV* genes as well as their pseudogenes are clustered in tandem. Similar to the *HLA* region, most *IGHV* genes harbor common nonsynonymous SNVs within them contributing to the high level of diversity observed within the immunoglobulin heavy chain repertoire. Therefore, we proceeded to analyze the nonsynonymous SNVs in that locus. Among the 519 SNVs, 18 were nonsynonymous and the others were intergenic, intronic, in the untranslated region, or synonymous

**Table 3** Association of rs6423677 and KD in the Japanese case and control panels

SNV	Case-control sets	KD/Control	Association statistics				
			Per cohort		Meta-analysis		
			OR (95% CI)	$P_{logistic}$	OR (95% CI)	$P_{meta}$	QE $P$
rs4774175	GWAS	428/3379	1.22 (1.06–1.41)	$4.9 \times 10^{-2}$	1.19 (1.11–1.28)	$3.4 \times 10^{-6}$	0.70
	Follow-up panel	1402/1652	1.22 (1.10–1.35)	$1.7 \times 10^{-4}$			
	Replication panel	1729/651	1.14 (1.00–1.29)	$5.0 \times 10^{-2}$			
rs6423677	GWAS	428/3379	1.28 (1.09–1.52)	$3.3 \times 10^{-2}$	1.25 (1.16–1.34)	$6.8 \times 10^{-10}$	0.84
	Follow-up panel	1417/1699	1.27 (1.15–1.40)	$4.5 \times 10^{-6}$			
	Replication panel	1758/653	1.22 (1.07–1.38)	$4.2 \times 10^{-3}$			

KD Kawasaki disease, SNV single nucleotide variant, OR odds ratio, CI confidence interval, QE  $P$  values for QE statistics in tests for residual heterogeneity

SNVs. When the Japanese sample set used in the Stage 2 analysis was genotyped for these 18 nonsynonymous SNVs, rs6423677 in the *IGHV3-66* gene showed the most significant association (Table 2). As expected, some other nonsynonymous SNVs also showed a similar trend of association. However, those associations, including that of rs4774175, were considered to be explained by LD with rs6423677 because they were no longer significant after conditioning the logistic regression analyses on rs6423677 (Table 2). In contrast, rs6423677 remained significant ( $P < 0.05$ ) after adjusting for any of the other SNVs (data not shown). In a Stage 3 analysis, rs4774175 and rs6423677 were further examined in an additional Japanese sample set, and rs6423677 achieved genome-wide significance in a meta-analysis of the Japanese sample panels (Table 3). Meanwhile, the association trend of rs6423677 was weaker in Korean and unreproducible in Taiwanese subjects (Supplementary Table 10).

### rs6423677 genotypes and *IGHV3-66* gene usage in the immunoglobulin heavy chain repertoire

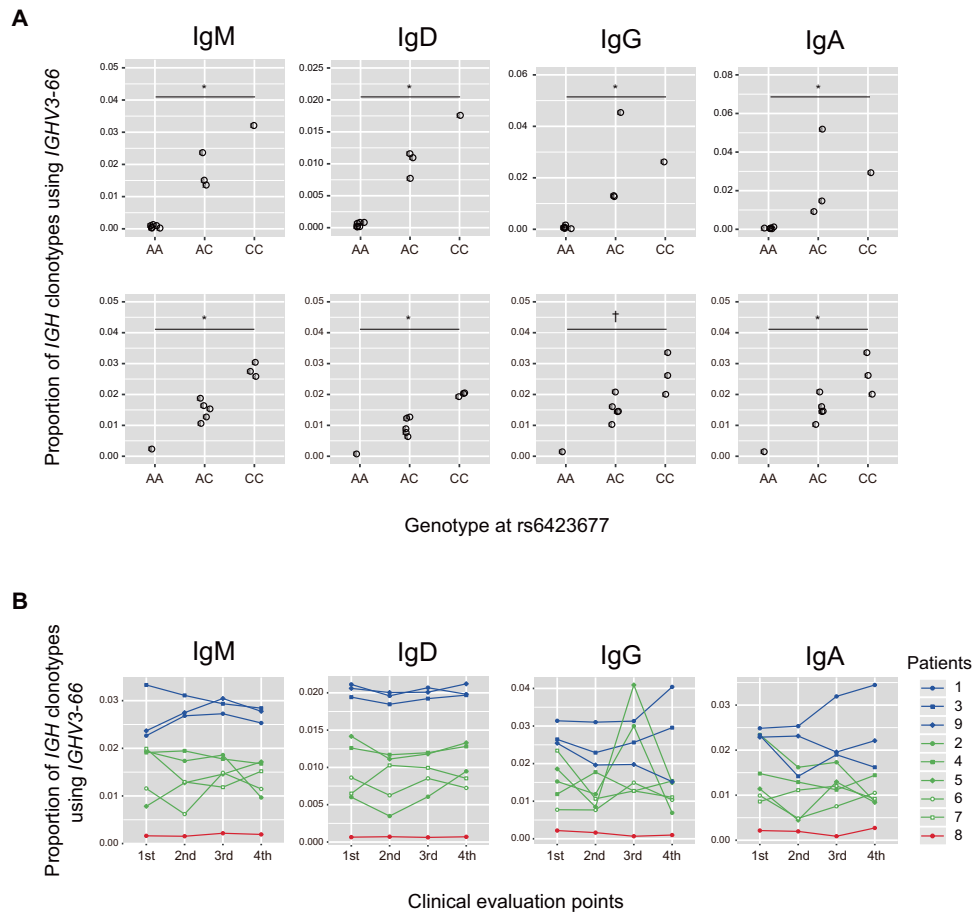
The significant association of an SNV within *IGHV3-66* with KD prompted us to investigate the impact of rs6423677 on the function of immunoglobulin molecules or B cell receptors (BCRs) harboring *IGHV3-66* in their heavy chain variable regions. Firstly, we examined *IGHV* usage in different immunoglobulin isotypes (IgM, IgD, IgG, and IgA) by analyzing the immunoglobulin heavy chain repertoire in peripheral B lymphocytes from healthy Japanese individuals ( $n = 10$ ) and patients with acute KD ( $n = 9$ ) by NGS. For every Ig isotype, we could identify *IGHV3-66* clonotypes in each individual and found that the *IGHV3-66* usage tended to be correlated with the number of C alleles each individual had (Fig. 2A and Supplementary Fig. 4). Consistent with the correlation trend, we observed significantly skewed usage of the alleles in the analyses of the

five patients who were heterozygous for rs6423677, with the protective allele (A) almost completely suppressed in all immunoglobulin classes and time points examined (Supplementary Fig. 5). Similar correlation trend was also observed between rs6423677 genotypes and relative levels of *IGH* transcript for *IGHV3-64*, the closest neighboring gene to *IGHV3-66*. However, the correlations were more modest than that seen for *IGHV3-66* (Supplementary Fig. 6). Paralogues or pseudogenes of *IGHV3-66* which have the ‘A’ nucleotide at the corresponding position of rs6423677 as well as PCR amplification bias in library preparation caused by additional variants interfering with the primer annealing could lead to such results. However, the distinct separation pattern of the allele discrimination plot in the Invader Assay and the clear electropherogram data in the direct sequencing of PCR-amplified genomic DNA from heterozygous patients does not support that possibility (Supplementary Fig. 1). Thus, we consider these observations not to be artificial in origin. We also negated the chance of misassignment of *IGHV3-66* with the A allele at rs6423677 (*IGHV3-66\*03*) to other alleles (*\*01*, *\*02*, or *\*04*) or paralogues such as *IGHV3-53* by confirming that the mipmap software could correctly assign an artificially created sequence with the rs6423677 A allele as *IGHV3-66\*03* (data not shown).

### A feature of the immunoglobulin repertoire of peripheral B lymphocytes from KD patients

To obtain an insight into the role of immunoglobulins using *IGHV3-66* in their heavy chain variable region in the pathogenesis of KD, we investigated how *IGHV3-66* usage in the immunoglobulin heavy chain gene repertoire changed over time during the clinical courses of nine KD patients (Supplementary Table 11). Proportions of *IGH* clones harboring *IGHV3-66* in their variable region seemed to be stable or not to have common varying patterns for IgM,





**Fig. 2** Effect of rs6423677 genotypes on the relative expression level of immunoglobulin heavy chain transcripts with *IGHV3-66*. **A** The proportion of clonotypes with *IGHV3-66* among all *IGH* transcripts displayed by rs6423677 genotype for each immunoglobulin isotype. Ten healthy adult volunteers (upper) and nine KD patients (lower) were analyzed. For KD patients, mean data of proportions at the four evaluation points were plotted. The increasing trend of the proportions according to the number of the C allele at rs6423677 was statistically evaluated with Jonckheere-Terpstra trend test. \* $P < 0.005$ , † $P < 0.01$ . **B** Time-course change of the sequence reads proportions of immunoglobulin heavy chain transcripts with *IGHV3-66* of acute KD

patients. Evaluation was done on four time points including (1) acute phase before receiving IVIG (3–8 illness days), (2) 48 h after the patients became afebrile (8–18 illness days), (3) the first follow-up visit to the pediatric clinic after discharge (17–50 days after the disease onset) and (4) the second follow-up visit to the pediatric clinic after discharge (3–4 months after the disease onset). In each panel, symbols and lines of red, green, and blue colors are representing KD patients with genotypes of AA, AC, and CC at rs6423677, respectively. The analyses were carried out for four immunoglobulin classes (IgM, IgD, IgG, and IgA)

IgD, and IgA, while in two patients (patients 2 and 5), transient increases were observed for IgG at the third evaluation point (Fig. 2B). Then, we investigated the increased clonotypes in the IgG heavy chain repertoire more precisely by defining them by the combinations of *IGHV*, *IGHD*, and *IGHJ* genes. We found that both patients 2 and 5 had single clonotypes with *IGHV3-66* that increased 1.0% or more as a proportion of the total from the first to the third evaluation points. However, the gene combinations of them were not the same (Table 4), and CDR3 amino acid sequences corresponding to the *V-D-J* combinations also differed between the patients (data not shown). Five of the remaining seven patients also had one or more IgG heavy chain clonotypes

with *IGHV3-66* that increased 0.1% or more at the same follow-up time point. However, cooccurring *V-D-J* combinations in the increased clonotypes were only seen between patients 2 and 6.

On the other hand, some of the CDR3 clonotypes which have recently been reported to dominate in the IgM heavy chain repertoire (>0.001%) in the acute pretreatment phase of Taiwanese KD patients [21] were also prevalent at a similar time point in the Japanese KD patients in this study (Supplementary Table 12). A distinct increase in diversity of the CDR3 clonotypes of IgM heavy chain during the convalescent phase of KD, which has also been reported in previous research of Taiwanese KD patients [21], was

**Table 4** IgG heavy chain clonotypes using IGHV3-66 that increased more than 0.1% as a proportion at the third evaluation point

Patient	Genotype at rs6423677	Clonotypes			Proportion in the repertoire at each clinical evaluation point				Increase during 1st and 3rd points
		IGHV-gene	IGHD-gene	IGHJ-gene	1st	2nd	3rd	4th	
1	CC	3-66*02	5-24*01	4*02	0.00016	0.00026	0.0099	0.002	0.0098
		3-66*02	3-16*01	2*01	$4.6 \times 10^{-6}$	$4.3 \times 10^{-6}$	0.0041	0	0.0041
2	AC	<b>3-66*01</b>	<b>6-19*01</b>	<b>4*02</b>	<b><math>4.0 \times 10^{-5}</math></b>	<b>0.00045</b>	<b>0.011</b>	<b>0.00065</b>	<b>0.011</b>
		3-66*01	2-21*01	4*02	0.0012	0.00032	0.0068	0.00014	0.0056
		3-66*01	3-10*01	4*02	0.00097	0.00036	0.0042	$8.1 \times 10^{-5}$	0.0032
3	CC	3-66*02	1-1*01	6*02	$5.1 \times 10^{-6}$	0.00016	0.0053	0.00015	0.0053
		3-66*02	3-22*01	6*02	0.00029	0.0012	0.0019	0.0024	0.0016
		3-66*02	6-19*01	2*01	0.00012	0.00094	0.0017	0.0009	0.0015
		3-66*02	4-17*01	4*02	$1.8 \times 10^{-5}$	0.00015	0.0014	$2.2 \times 10^{-5}$	0.0014
4	AC	3-66*01	5-24*01	4*02	0.00051	0.00015	0.0017	0.0009	0.0012
5	AC	<b>3-66*01</b>	<b>3-3*01</b>	<b>4*02</b>	<b><math>3.3 \times 10^{-5}</math></b>	<b>0.0012</b>	<b>0.028</b>	<b><math>1.8 \times 10^{-5}</math></b>	<b>0.028</b>
		3-66*01	3-9*01	3*02	0	$6.2 \times 10^{-5}$	0.0018	0	0.0018
		3-66*01	4-17*01	3*02	$4.7 \times 10^{-6}$	$4.9 \times 10^{-5}$	0.0014	0	0.0014
6	AC	3-66*01	3-22*01	4*02	0.00028	0.00073	0.0025	0.0018	0.0022
		3-66*01	3-10*01	4*02	0.0007	0.00027	0.0018	0.00058	0.0011
		3-66*01	6-19*01	4*02	$8.6 \times 10^{-5}$	0.00036	0.0012	$5.2 \times 10^{-5}$	0.0011
7	AC	3-66*01	1-1*01	6*02	0.00013	0.00034	0.0014	$3.4 \times 10^{-5}$	0.0012
8	AA					none			
9	CC					none			

Clonotypes increased more than 1% during 1st and 3rd evaluation points are bold faced

observed for four of the nine KD patients (Supplementary Fig. 7).

## Discussion

It has generally been considered that the immune-mediated vasculitis of KD is triggered in response to infection with some type of microorganism. This assumption was made based on factors indicative of a primary infection, such as its clinical symptoms, which included fever and skin rash, epidemiologic findings such as its peak age of onset (6 months to 1 year), along with the seasonality of the occurrence of regional outbreaks and nation-wide epidemics [2, 11]. Despite tremendous efforts, no single microorganism has been conclusively proven as the pathogen of KD and lack of information about the pathogen has been a significant obstacle to causal treatment and disease prevention.

Histopathological and immunological studies have revealed activation of neutrophils, macrophages, and monocytes in the acute phase of KD [22, 23], and now the leading hypothesis for the pathophysiology of KD inflammation is an attack of the dysregulated or hyperactive innate

immune system against vascular walls [24]. In addition, genetic studies using a genome-wide approach have identified several robust susceptibility loci/genes for KD [8–10, 12, 25–32], and an in silico prediction of responsible variants and the types of cells where they have biological significance has highlighted the importance of B cells in the pathogenesis of KD [33]. In previous literature, polyclonal activation and increase of B cells [34, 35] and detection of auto-antibodies against components of vascular endothelial cells or neutrophils in peripheral blood of acute KD patients have been documented [36, 37]. Infiltration of oligoclonal plasma cells producing IgA into the vascular wall of KD patients was also reported [38]. These previous findings have suggested the involvement of B cells in KD inflammation. However, a recent transcriptomic study revealed downregulation of genes related to BCR signaling in acute KD patients [39]. Thus, there have been both supportive and unsupportive pieces of evidence, and there is no consensus view on the role of B cells, as well as of the adaptive immune system in KD pathogenesis. In this study, we identified genetic variants within the IGHV region significantly associated with KD in the Japanese population. The immunoglobulin heavy (IGH) locus spans 1.3 Mb (from 106.0 to 107.3 Mb on NC\_000014.8) of the chromosome

14q32.33 region and encompasses serially arranged *IGH* constant (*IGHC*), *IGHJ*, *IGHD*, and *IGHV* clusters. Because of high sequence redundancy which obstructs designing of specific genotyping assays, there is a large blank area (~0.8 Mb) which was not covered by the SNP arrays and therefore we lack association information about the SNVs in the area (Supplementary Fig. 8). However, because SNVs within the gap are not in high LD ( $r^2 < 0.3$ ) with rs6423677 or rs4774175 in the 1000 genomes JPT population (Supplementary Fig. 8) and we did not observe association signals in the chromosomal region on the opposite side of the blank area, the association signals represented by rs6423677 or rs4774175 might be localized within the region where we could investigate in this study.

In contrast to the robust association of rs6423677 in the Japanese subjects, it was not consistent in the Korean and Taiwanese subjects. Although neither significant nor consistent with the results in the current study, a marginal trend of association between SNVs within the *IGHV* region and linked to rs4774175 ( $r^2 = 0.45$ – $0.66$  in the 1000genomes CHB population) and KD in the Taiwanese population had already been reported [28]. Findings in the previous and the current studies are strongly indicating that variants in the *IGHV* region are commonly involved in KD pathogenesis at least in the Japanese and Taiwanese populations. However, the robustness of association might not be uniform among the populations. Given that multiple pathogens with regionally or seasonally differing epidemic patterns could be triggering KD, it might be possible that various susceptibility genes or alleles corresponding to different antigenicity for the pathogens exist in this locus and are related to the mixed robustness of the association.

Highly skewed expression of *IGH* transcripts with the risk-associated allele adds support to consider that *IGHV3-66* is the functional target of the associated variants. The A to C nucleotide change at rs6423677 results in an amino acid alteration from cysteine to glycine within the CDR2 region of *IGHV3-66*. Thus, the modification might affect the affinity to some antigens of immunoglobulins carrying *IGHV3-66*. However, we consider that the mechanism of increased susceptibility to KD associated with the C allele might not be due to reduced host defense to particular agents because it is inconsistent with the observation that the A allele, which is protective against KD and expected to have a higher neutralizing ability in this scenario, seemed to be nearly silenced. *IGHV3-66* has been recognized as one of the functional *IGHV* genes and purported to be utilized at frequencies of around 1% [40]. The average proportions of *IGHV3-66* clonotypes in the 10 healthy adults in our study (0.88% for IgM, 0.50% for IgD, 1.0% for IgG, and 1.1% for IgA), were consistent with that previous information. It is currently unknown in what stage, i.e., somatic

recombination, RNA transcription, and processing of the pre-mRNA, the A allele was excluded and resulted in the significantly skewed allelic usage to the C allele of rs6423677. It is suggestive that usage of *IGHV3-64*, the nearest functional gene to *IGHV3-66*, showed difference among genotypes at rs6423677 which was similar to that of *IGHV3-66* (Supplementary Fig. 6). One potential reason is suggested by the predicted binding of CTCF transcription repressor in the 0.4 kb region encompassing rs6423677 (Supplementary Fig. 9). CTCF has been reported to interact with multiple sites in the *IGH* region and plays essential roles in somatic recombination [41] of the distal area of the *IGHV* gene region. The association data of rs6423677 and the increased usage of *IGH* transcripts with the risk-associated allele are indicative of some active role of the immunoglobulin molecules as antibodies or as components of BCRs in the development of KD. One possible role of such immunoglobulins might be activation of B cells. Established KD susceptibility gene products such as B lymphoid tyrosine kinase (BLK) [8–10] and Inositol 1,4,5 trisphosphate 3-kinase C (ITPKC) [8, 9] are involved in BCR signaling. If IVIG acts through competing with such immunoglobulins for agents or antigens relevant to KD pathogenesis, the requirement of a high dose administration (1–2 g/kg) of IVIG to treat KD might be reasonable because only a fraction of the IgG would contribute to the therapeutic effect, with *IGHV3-66* expected to only account for up to several percent of the IVIG preparations. B cells can be activated by nonspecific binding of microbial products such as superantigens (SAGs) to BCRs. Intriguingly, B cell SAGs restricted to *IGHV3* segment of immunoglobulins have been known [42]. However, considering that the innate immune system has been thought to play a central role in the KD vasculitis and that KD can develop in patients with X-linked agammaglobulinemia, who lack or have small numbers of B cells [43], the activity of B cells or immunoglobulins in KD might be relevant to initiation or enhancement of the innate immune activation but may be substitutable.

Recently, a significant association of an allele of the *IGHV4-61* gene (*IGHV4-61\*02*) with susceptibility to rheumatic heart disease (RHD), which is a long-term complication of ARF, was reported [44]. *IGHV4-61* is located only 36 kb downstream of *IGHV3-66* (Supplementary Fig. 6). ARF develops as a sequela of *Streptococcus pyogenes* (*S. pyogenes*) infection and, similar to KD, has been recognized to affect genetically susceptible individuals [45]. Among reports of GWAS for human diseases, a genome-wide significant association of variants in the *IGHV* gene region has been identified only for RHD. Although *S. pyogenes* is not recognized as the cause of KD, considering that KD shares some characteristic symptoms such as skin rash and strawberry tongue with *S. pyogenes* infection, it is

suggestive that the previously discussed role of B cells might be related to some underlying mechanisms of the common symptoms.

In 2020, multiple series of patients with Pediatric Inflammatory Multisystem Syndrome Temporally Associated with SARS-CoV2 infection (PIMS-TS) or (MIS-C) having KD-like symptoms or increase of severe KD patients after the SARS-CoV2 epidemic were reported from the US and European countries [46, 47]. In the latest studies, overrepresentation of IGHV3-53 and IGHV3-66 in neutralizing monoclonal antibodies against the receptor binding domain of SARS-CoV2 spike were also reported [48, 49]. Given immunoglobulin with IGHV3-66 play a role in KD, it might be possible that KD-like symptoms seen in such patients are mediated by interaction between SARS-CoV2 and B cells expressing IGHV3-66.

We also found some commonality between the CDR3 clonotypes that were increased in the IgM heavy chains of the Taiwanese and the Japanese KD patients (Supplementary Table 12). IGHV3-66 was used only in one commonly increased CDR3 clonotype (Supplementary Table 13). However, as far as can be understood from the limited number of observations, *IGHV3-66* seemed not to be important in the IgM response in the acute pretreatment phase.

Future characterization of the endogenous immunoglobulin molecules that are increased in KD patients utilizing information of the light chains that can be obtained simultaneously in single-cell analyses [50] will facilitate identification of the agent triggering KD as well as understanding the mechanism of action of IVIG treatment.

There are limitations in this study. First, in addition to the gap above where we could not examine the association of the variants, our strategy to focus only on nonsynonymous SNVs on *IGHV* genes left the possibility that rs6423677 is just a proxy of the genuinely responsible variant located outside *IGHV3-66*. Second, we did not analyze the time-course change of the immunoglobulin heavy chain repertoire in other infectious diseases. So it is uncertain whether the upregulation of IgG heavy chain transcripts with *IGHV3-66* is a specific observation for KD or not. Third, our results might not directly reflect changes of the immunoglobulins at the protein level because we lack information about the correlation between the proportions of particular *IGH* clones in the transcripts from B cells and in the proteins expressed on the cell surface or circulating in the serum.

In conclusion, a significant association of a nonsynonymous SNV in the *IGHV3-66* gene with KD was observed. Further intensified study of the association in this region and repertoire analyses of immunoglobulins in different ethnicities and subpopulations of the patients with different demographic features would give insights into both the role

of B cells in the KD pathogenesis and the causal agent of the disease.

**Acknowledgements** This study was supported by grants from the Millennium Project, from the Japan Kawasaki Disease Research Center (2015 to YM, and 2018 and 2019 to YO), and from the Japan Agency for Medical Research and Development (JP18ek0410039 to TH). This study was also supported by a grant from the Ministry of Health & Welfare of the Republic of Korea (HI15C1575 to JKL). We are grateful to the KD patients and their family members as well as the medical staff taking care of the patients. We also thank Ms. Yoshie Kikuchi for her technical assistance.

**Author contributions** JKL, JYW, YTC, and YO supervised the study. JKL, JYW, and YO conceived the study. JYW, TAJ, TT, JKL, and YO designed the study. TAJ, YM, JKL, JYW, and YO wrote the manuscript. AH, HS, HH, TH, and Japan Kawasaki Disease Genome Consortium collected Japanese samples. YMH, GYJ, SWY, JYJ, KYL, and Korean Kawasaki Disease Genetics Consortium collected Korean samples. Taiwan Kawasaki Disease Genetics Consortium and Taiwan Pediatric ID Alliance collected Taiwanese samples. YML coordinated the multi-center collaboration in Taiwan as the project manager and collected samples and clinical information. MK performed GWAS assays for the Japanese samples. AT performed statistical analyses for the Japanese GWAS data. DY and TP performed statistical analyses for the Korean GWAS data. JJK conducted a follow-up study (Stage 2) for the Korean samples. CHC performed statistical analyses for the Taiwanese GWAS data and followed-up meta-analyses for the Taiwanese data. YCL supervised the GWAS and replication genotyping pipeline, performed the data analyses. LCC performed statistical analyses for the Taiwanese GWAS data and followed-up meta-analyses for the Taiwanese data. CPC performed genotyping and direct sequencing of Taiwanese samples. TAJ, DY, and CHC conducted the whole-genome imputation. TAJ performed *P* value simulation and meta-analyses. KO, TT, and KI performed genotyping and direct sequencing of the Japanese samples. YM and YO performed the NGS data analyses for the IGH repertoires.

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

**Ethical approval** This study was approved by the Institutional Review Board at all involved institutes.

**Informed consent** Written informed consent was obtained from all subjects.

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

## References

1. Kawasaki T. Acute febrile mucocutaneous syndrome with lymphoid involvement with specific desquamation of the fingers and toes in children. (Japanese) *Arerugi* 1967; 16: 178–222. English translation by Shike H, Burns JC, Shimizu C. *Pediatr Infect Dis J*. 2000;21:993–5.
2. Makino N, Nakamura Y, Yashiro M, Sano T, Ae R, Kosami K, et al. Epidemiological observations of Kawasaki disease in Japan, 2013–2014. *Pediatr Int*. 2018;60:581–7.

3. Kato H, Koike S, Yamamoto M, Ito Y, Yano E. Coronary aneurysms in infants and young children with acute febrile mucocutaneous lymph node syndrome. *J Pediatr.* 1975;86:892–8.
4. Taubert KA, Rowley AH, Shulman ST. Nationwide survey of Kawasaki disease and acute rheumatic fever. *J Pediatr.* 1991;119:279–82.
5. Furusho K, Sato K, Soeda T, Matsumoto H, Okabe T, Hirota T, et al. High-dose intravenous gammaglobulin for Kawasaki disease. *Lancet.* 1984;2:1055–8.
6. Newburger JW, Takahashi M, Burns JC, Beiser AS, Chung KJ, Duffy CE, et al. The treatment of Kawasaki syndrome with intravenous gamma globulin. *N. Engl J Med.* 1986;315:341–7.
7. Newburger JW, Takahashi M, Beiser AS, Burns JC, Bastian J, Chung KJ, 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.* 1991;324:1633–9.
8. Khor CC, Davila S, Breunis WB, Lee YC, Shimizu C, Wright VJ, et al. Genome-wide association study identifies *FCGR2A* as a susceptibility locus for Kawasaki disease. *Nat Genet.* 2011;43:1241–6.
9. Onouchi Y, Ozaki K, Burns JC, Shimizu C, Terai M, Hamada H, et al. A genome-wide association study identifies three new risk loci for Kawasaki disease. *Nat Genet.* 2012;44:517–21.
10. Lee YC, Kuo HC, Chang JS, Chang LY, Huang LM, Chen MR, et al. Two new susceptibility loci for Kawasaki disease identified through genome-wide association analysis. *Nat Genet.* 2012;44:522–5.
11. Burns JC, Kushner HI, Bastian JF, Shike H, Shimizu C, Matsubara T, et al. Kawasaki disease: a brief history. *Pediatrics.* 2000;106:E27.
12. Kim JJ, Hong YM, Sohn S, Jang GY, Ha KS, Yun SW, et al. A genome-wide association analysis reveals 1p31 and 2p13.3 as susceptibility loci for Kawasaki disease. *Hum Genet.* 2011;129:487–95.
13. 1000 Genomes Project Consortium, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature.* 2012;491:56–65.
14. Delaneau O, Marchini J, Zagury JF. A linear complexity phasing method for thousands of genomes. *Nat Methods.* 2012;9:179–81.
15. Howie B, Marchini J, Stephens M. Genotype imputation with thousands of genomes. *G3.* 2011;1:457–70.
16. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat Genet.* 2012;44:955–9.
17. Gao F, Lin E, Feng Y, Mack WJ, Shen Y, Wang K. Characterizing immunoglobulin repertoire from whole blood by a personal genome sequencer. *PLoS One.* 2013;8:e75294.
18. Magoč T, Salzberg SL. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics.* 2011;27:2957–63.
19. Knight JC, Keating BJ, Kwiatkowski DP. Allele-specific repression of lymphotoxin- $\alpha$  by activated B cell factor-1. *Nat Genet.* 2004;36:394–9.
20. Hirota T, Takahashi A, Kubo M, Tsunoda T, Tomita K, Doi S, et al. Genome-wide association study identifies three new susceptibility loci for adult asthma in the Japanese population. *Nat Genet.* 2011;43:893–6.
21. Ko TM, Kiyotani K, Chang JS, Park JH, Yin YP, Chen YT, et al. Immunoglobulin profiling identifies unique signatures in patients with Kawasaki disease during intravenous immunoglobulin treatment. *Hum Mol Genet.* 2018;27:2671–7.
22. Takahashi K, Oharaseki T, Naoe S, Wakayama M, Yokouchi Y. Neutrophilic involvement in the damage to coronary arteries in acute stage of Kawasaki disease. *Pediatr Int.* 2005;47:305–10.
23. Koga M, Ishihara T, Takahashi M, Umezawa Y, Furukawa S. Activation of peripheral blood monocytes and macrophages in Kawasaki disease: ultrastructural and immunocytochemical investigation. *Pathol Int.* 1998;48:512–7.
24. Hara T, Nakashima Y, Sakai Y, Nishio H, Motomura Y, Yamasaki S. Kawasaki disease: a matter of innate immunity. *Clin Exp Immunol.* 2016;186:134–43.
25. Onouchi Y, Gunji T, Burns JC, Shimizu C, Newburger JW, Yashiro M, et al. *ITPKC* functional polymorphism associated with Kawasaki disease susceptibility and formation of coronary artery aneurysms. *Nat Genet.* 2008;40:35–42.
26. Onouchi Y, Ozaki K, Burns JC, Shimizu C, Hamada H, Honda T, et al. Common variants in *CASP3* confer susceptibility to Kawasaki disease. *Hum Mol Genet.* 2010;19:2898–906.
27. Burgner D, Davila S, Breunis WB, Ng SB, Li Y, Bonnard C, et al. A genome-wide association study identifies novel and functionally related susceptibility Loci for Kawasaki disease. *PLoS Genet.* 2009;5:e1000319.
28. Tsai FJ, Lee YC, Chang JS, Huang LM, Huang FY, Chiu NC, et al. Identification of novel susceptibility Loci for Kawasaki disease in a Han chinese population by a genome-wide association study. *PLoS One.* 2011;6:e16853.
29. Khor CC, Davila S, Shimizu C, Sheng S, Matsubara T, Suzuki Y, et al. Genome-wide linkage and association mapping identify susceptibility alleles in *ABCC4* for Kawasaki disease. *J Med Genet.* 2011;48:467–72.
30. Shimizu C, Eleftherohorinou H, Wright VJ, Kim J, Alphonse MP, Perry JC, et al. Genetic variation in the *SLC8A1* calcium signaling pathway is associated with susceptibility to Kawasaki disease and coronary artery abnormalities. *Circ Cardiovasc Genet.* 2016;9:559–68.
31. Kim J, Shimizu C, Kingsmore SF, Veeraraghavan N, Levy E, Dos Santos AMR, et al. Whole genome sequencing of an African American family highlights *toll like receptor 6* variants in Kawasaki disease susceptibility. *PLoS One.* 2017;12:e0170977.
32. Kim JJ, Yun SW, Yu JJ, Yoon KL, Lee KY, Kil HR, et al. A genome-wide association analysis identifies *NMNAT2* and *HCP5* as susceptibility loci for Kawasaki disease. *J Hum Genet.* 2017;62:1023–9.
33. Farh KK, Marson A, Zhu J, Kleinewietfeld M, Housley WJ, Beik S, et al. Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature.* 2015;518:337–43.
34. Leung DY, Siegel RL, Grady S, Krensky A, Meade R, Reinherz EL, et al. Immunoregulatory abnormalities in mucocutaneous lymph node syndrome. *Clin Immunol Immunopathol.* 1982;23:100–12.
35. Furukawa S, Matsubara T, Yabuta K. Mononuclear cell subsets and coronary artery lesions in Kawasaki disease. *Arch Dis Child.* 1982;67:706–8.
36. Leung DY, Collins T, Lapierre LA, Geha RS, Pober JS. Immunoglobulin M antibodies present in the acute phase of Kawasaki syndrome lyse cultured vascular endothelial cells stimulated by gamma interferon. *J Clin Investig.* 1986;77:1428–35.
37. Savage CO, Tizard J, Jayne D, Lockwood CM, Dillon MJ. Antineutrophil cytoplasm antibodies in Kawasaki disease. *Arch Dis Child.* 1989;64:360–3.
38. Rowley AH, Shulman ST, Spike BT, Mask CA, Baker SC. Oligoclonal IgA response in the vascular wall in acute Kawasaki disease. *J Immunol.* 2001;166:1334–43.
39. Ikeda K, Yamaguchi K, Tanaka T, Mizuno Y, Hijikata A, Ohara O, et al. Unique activation status of peripheral blood mononuclear cells at acute phase of Kawasaki disease. *Clin Exp Immunol.* 2010;160:246–55.
40. Lange MD, Huang L, Yu Y, Li S, Liao H, Zemlin M, et al. Accumulation of VH replacement products in IgH genes derived

- from autoimmune diseases and anti-viral responses in human. *Front Immunol.* 2014;5:345.
41. Guo C, Yoon HS, Franklin A, Jain S, Ebert A, Cheng HL, et al. CTCF-binding elements mediate control of V(D)J recombination. *Nature.* 2011;477:424–30.
  42. Silverman GJ, Sasano M, Wormsley SB. Age-associated changes in binding of human B lymphocytes to a VH3-restricted unconventional bacterial antigen. *J Immunol.* 1993;151:5840–55.
  43. Behniafard N, Aghamohammadi A, Abolhassani H, Pourjabbar S, Sabouni F, Rezaei N. Autoimmunity in X-linked agammaglobulinemia: Kawasaki disease and review in the literature. *Expert Rev Clin Immunol.* 2012;8:155–9.
  44. Parks T, Mirabel MM, Kado J, Auckland K, Nowak J, Rautanen A, et al. Pacific Islands rheumatic heart disease genetics network. Association between a common immunoglobulin heavy chain allele and rheumatic heart disease risk in Oceania. *Nat Commun.* 2017;8:14946.
  45. Carapetis JR, Currie BJ, Mathews JD. Cumulative incidence of rheumatic fever in an endemic region: a guide to the susceptibility of the population? *Epidemiol Infect.* 2000;124:239–44.
  46. Riphagen S, Gomez X, Gonzalez-Martinez C, Wilkinson N, Theocharis P. Hyperinflammatory shock in children during COVID-19 pandemic. *Lancet.* 2020;395:1607–8.
  47. Whittaker E, Bamford A, Kenny J, Kaforou M, Jones CE, Shah P, et al. Clinical characteristics of 58 children with a pediatric inflammatory multisystem syndrome temporally associated With SARS-CoV-2. *JAMA.* 2020;324:259–69.
  48. Cao Y, Su B, Guo X, Sun W, Deng Y, Bao L, et al. Potent neutralizing antibodies against SARS-CoV-2 identified by high-throughput single-cell sequencing of convalescent patients' B cells. *Cell.* 2020;182:73–84.
  49. Barnes CO, West AP Jr, Huey-Tubman KE, Hoffmann MAG, Sharaf NG, Hoffman PR, et al. Structures of human antibodies bound to SARS-CoV-2 spike reveal common epitopes and recurrent features of antibodies. *Cell.* 2020;182:828–42.
  50. DeKosky BJ, Ippolito GC, Deschner RP, Lavinder JJ, Wine Y, Rawlings BM, et al. High-throughput sequencing of the paired human immunoglobulin heavy and light chain repertoire. *Nat Biotechnol.* 2013;31:166–9.

## Affiliations

Todd A. Johnson<sup>1</sup> · Yoichi Mashimo<sup>2</sup> · Jer-Yuarn Wu<sup>3</sup> · Dankyu Yoon<sup>4</sup> · Akira Hata<sup>2</sup> · Michiaki Kubo<sup>5</sup> · Atsushi Takahashi<sup>6,7</sup> · Tatsuhiro Tsunoda<sup>8,9</sup> · Kouichi Ozaki<sup>10,11</sup> · Toshihiro Tanaka<sup>10,12</sup> · Kaoru Ito<sup>10</sup> · Hiroyuki Suzuki<sup>13</sup> · Hiromichi Hamada<sup>14</sup> · Tohru Kobayashi<sup>15</sup> · Toshiro Hara<sup>16</sup> · Chien-Hsiun Chen<sup>3</sup> · Yi-Ching Lee<sup>17</sup> · Yi-Min Liu<sup>3</sup> · Li-Ching Chang<sup>3</sup> · Chun-Ping Chang<sup>3</sup> · Young-Mi Hong<sup>18</sup> · Gi-Young Jang<sup>19</sup> · Sin-Weon Yun<sup>20</sup> · Jeong-Jin Yu<sup>21</sup> · Kyung-Yil Lee<sup>22</sup> · Jae-Jung Kim<sup>23</sup> · Taesung Park<sup>24</sup> · Korean Kawasaki Disease Genetics Consortium, Taiwan Kawasaki Disease Genetics Consortium, Taiwan Pediatric ID Alliance, Japan Kawasaki Disease Genome Consortium · Jong-Keuk Lee<sup>23</sup> · Yuan-Tsong Chen<sup>3</sup> · Yoshihiro Onouchi<sup>2,10</sup>

<sup>1</sup> Laboratory for Medical Science Mathematics, RIKEN Center for Integrative Medical Sciences, Yokohama, Kanagawa 230-0045, Japan

<sup>2</sup> Department of Public Health, Chiba University Graduate School of Medicine, Chiba, Chiba 260-8670, Japan

<sup>3</sup> Institute of Biomedical Sciences, Academia Sinica, Taipei 11529, Taiwan

<sup>4</sup> Division of Allergy and Chronic Respiratory Diseases, Center for Biomedical Sciences, Korea National Institute of Health, Osong 28160, Korea

<sup>5</sup> RIKEN Center for Integrative Medical Sciences, Yokohama, Kanagawa 230-0045, Japan

<sup>6</sup> Laboratory for Statistical and Translational Genetics, RIKEN Center for Integrative Medical Sciences, Yokohama, Kanagawa 230-0045, Japan

<sup>7</sup> Department of Genomic Medicine, Research Institute, National Cerebral and Cardiovascular Center, Suita, Osaka 564-8565, Japan

<sup>8</sup> Department of Biological Sciences, Graduate School of Science, the University of Tokyo, Tokyo 113-0033, Japan

<sup>9</sup> Department of Medical Science Mathematics, Medical Research Institute, Tokyo Medical and Dental University, Tokyo 113-8510, Japan

<sup>10</sup> Laboratory for Cardiovascular Genomics and Informatics, RIKEN Center for Integrative Medical Sciences, Yokohama, Kanagawa 230-0045, Japan

<sup>11</sup> Division for Genomic Medicine, Medical Genome Center, National Center for Geriatrics and Gerontology, Obu, Aichi 474-8511, Japan

<sup>12</sup> Department of Human Genetics and Disease Diversity, Graduate School of Medical and Dental Sciences, Tokyo Medical and Dental University, Tokyo 113-8510, Japan

<sup>13</sup> Department of Pediatrics, Wakayama Medical University, Wakayama, Wakayama 641-8509, Japan

<sup>14</sup> Department of Pediatrics, Tokyo Women's Medical University Yachiyo Medical Center, Yachiyo, Chiba 276-8524, Japan

<sup>15</sup> Department of Management and Strategy, Clinical Research Center, National Center for Child Health and Development, Tokyo 157-8535, Japan

<sup>16</sup> Fukuoka Children's Hospital, Fukuoka, Fukuoka 813-0017, Japan

<sup>17</sup> Institute of Cellular and Organismic Biology, Academia Sinica, Taipei 11529, Taiwan

<sup>18</sup> Department of Pediatrics, Ewha Womans University College of Medicine, Seoul 07985, Korea

<sup>19</sup> Department of Pediatrics, Korea University Ansan Hospital, Ansan 15355, Korea

<sup>20</sup> Department of Pediatrics, Chung-Ang University Hospital, Seoul 06973, Korea

<sup>21</sup> Department of Pediatrics, Asan Medical Center, Seoul 05505, Korea

<sup>22</sup> Department of Pediatrics, Daejeon St. Mary's Hospital, Daejeon 34943, Korea

<sup>23</sup> Asan Institute for Life Sciences, University of Ulsan College of Medicine, Seoul 05505, Korea

<sup>24</sup> Department of Statistics, Seoul National University, Seoul 08826, Korea