Genome-wide linkage and association mapping identify susceptibility alleles in ABCC4 for Kawasaki disease


Published in: Journal of Medical Genetics

DOI: 10.1136/jmg.2010.086611

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
ORIGINAL ARTICLE

Genome-wide linkage and association mapping identify susceptibility alleles in \textit{ABCC4} for Kawasaki disease

Chia Chuen Khor,1,2 Sonia Davila,1,3 Chisato Shimizu,4 Stephanie Sheng,4 Tomoyo Matsubara,5 Yasuo Suzuki,6 Jane W Newburger,7 Annette Baker,7 David Burgner,8 Willeijnj Breunis,9,10 Taco Kuijpers,9,10 Victoria J Wright,11 Michael Levin,11 Martin L Hibberd,1,2 Jane C Burns,4 on behalf of the US and International Kawasaki Disease Genetics Consortia*

\textbf{ABSTRACT}

\textbf{Background} Kawasaki disease (KD) is a self limited vasculitis in which host genetics plays a prominent role. To further the understanding of the role of host genetics in KD, a three-stage genetic study was conducted that began with a family linkage study and ultimately involved more than 3000 individuals to identify new genetic contributions to KD susceptibility.

\textbf{Methods and results} A 26-family linkage study followed by fine mapping was performed in a cohort of 1284 KD subjects and their family members (total 3248 individuals). Suggestive evidence of disease linkage (logarithm of odds (LOD) $\geq 3.0$, $p<1.00 \times 10^{-7}$) was found for five genomic locations (Chr 3q, 4q, 10p, 13q, 21q). Two of these loci (Chr 4q and Chr 13q) overlapped with validated findings from a recent KD genome-wide association study. Fine mapping analysis revealed three single nucleotide polymorphisms (SNPs) in \textit{ABCC4} with validated findings from a recent KD genome-wide association study. The Chr 13q linkage peak showing evidence of association to KD (lowest $p=8.82 \times 10^{-3}$; combined OR 2.00, 95% CI 1.41 to 2.83). \textit{ABCC4} is a multifunctional cyclic nucleotide transporter that stimulates the migratory capacity of dendritic cells. It is also a mediator of prostanoid efflux from human cells and is inhibited by non-steroidal anti-inflammatory medications such as aspirin.

\textbf{Conclusion} These genetic data suggest that \textit{ABCC4} could play a fundamental role in KD pathogenesis with effects on immune activation and vascular response to injury.

\textbf{INTRODUCTION}

Kawasaki disease (KD) is a systemic vasculitis of medium sized arteries primarily affecting children. The cause of the disease is unknown, and although many patients experience an uneventful recovery, up to 25% of untreated children will develop potentially fatal coronary arterial aneurysms. An infectious aetiology for KD has been suspected but the agent remains elusive.1 Clues suggesting a possible genetic link to KD include strikingly different rates of occurrence in different ethnic groups (individuals of Asian descent are notably more susceptible), and higher-than-expected concordance between sibling pairs.2 Accordingly, a genome-wide association study (GWAS) of KD was performed and reported suggestive evidence of association ($p \sim 10^{-6}$) of two genes (\textit{NAALADL2} and \textit{ZFHIX}) with KD.3 Due to limited statistical power, the modest GWAS sample size (107 cases and 154 controls) precluded unequivocal detection of other KD susceptibility genes of smaller effect size and/or rarer allele frequencies.

We now utilise a combined linkage and association approach to identify other potential KD susceptibility genes which were not identified by the GWAS.

\textbf{METHODS}

KD was diagnosed according to the criteria of the American Heart Association,4 which include fever, together with at least four of the five classical diagnostic criteria:

1. Polymorphous erythematous rash
2. Non-purulent bilateral conjunctival injection
3. Oropharyngeal changes, including diffuse hyperaemia, strawberry tongue, and lip changes (eg, swelling, fissuring, erythema, bleeding)
4. Peripheral extremity changes, including erythema, oedema, induration, and desquamation
5. Non-purulent cervical lymphadenopathy

Children with at least 5 days of fever and at least two diagnostic criteria with echocardiographic findings of coronary artery damage during the acute and/or convalescent phases of KD were also included, as these coronary artery manifestations are pathognomonic for KD. Cases of incomplete KD, who have fever, <4 diagnostic criteria, and no coronary artery manifestations (who constitute approximately 15% of KD cases receiving clinical treatment), were excluded to maximise the homogeneity of the clinical phenotype.

In all cohorts, clinical and laboratory data were obtained directly from patient medical files and supplemented with parental questionnaires. Phenotypic data were reviewed in all cases by experienced pediatricians. Ethical approval was obtained from the appropriate national and regional institutional review boards for each study population.3 Informed consent and assent as appropriate were obtained from participating families.
Complex traits

Figure 1  Genome-wide linkage results. Genome-wide linkage results are shown, with the dotted line indicating maximum logarithm of odds (LOD) score (MLS)=3.

Linkage cohort
All families enrolled in the KD Genetics Study in the USA were asked regarding other affected family members and attempts were made to collect DNA samples on all individuals in the pedigree. Collaborators in Japan identified four families with more than one affected member. For the linkage cohort, 26 out of 30 pedigrees recruited were analysed (supplementary table 1, supplementary figure 1), as samples from both parents were available for the affected individuals.

Case–control cohort
This cohort was enrolled from the Netherlands and comprises 191 KD cases and 225 controls (416 individuals of European descent). All KD cases were identified by collaborating paediatricians and sent for cardiac evaluations and subsequent follow-up by one of the co-authors (TWK; more information on the clinical characteristics of subjects can be found in supplementary table 2). The controls were unrelated, adult blood donors living in the same geographical area. Ethnicity was determined by self report or parental report on behalf of the child.

Family based cohorts
The cohorts used for family based association testing included 715 complete parent affected offspring trios (2145 individuals) plus an additional 578 cases with either one (n=191) or no (n=187) parents providing a DNA sample. An unaffected sibling was also included in 198 of the families. Subjects were recruited from Australia, the Netherlands, USA, and UK. These cases were identified from paediatric hospital databases, from parental support groups, and through media releases. More detailed information can be found in supplementary table 2.

Study design
We used a three-stage study design to evaluate the contribution of genetic factors to KD susceptibility. Stage 1 was a genetic linkage study analysing 26 KD pedigrees (supplementary figure 1). Stage 2 involved searching for regions of overlap between linkage peaks (logarithm of odds (LOD) >3) observed in KD pedigrees and replicated associations reported in the previous GWAS. Stage 3 involved fine-scale mapping of the strongest associated region (stage 2 p<1.00x10^-5) identified in stage 2. This fine-scale mapping was carried out in 191 KD cases, 225 controls, and 715 complete case–parent trios.

Genotyping methods
DNA samples from the linkage study were genotyped using the Illumina Linkage 12 panel (http://www.illumina.com; 6089 single nucleotide polymorphisms (SNPs) spaced evenly throughout the human genome). Follow-up genotyping was performed using the Sequenom Mass-Array iPLEX platform (http://www.sequenom.com).

Statistical analysis
Linkage analysis was performed using the MERLIN statistical package, as previously described (http://www.sph.umich.edu/csg/abecasis/merlin/download/). The Kong and Cox non-parametric, exponential model was chosen to allow sharing of a small number of large effect alleles among a small number of families to be more reliably detected. The non-parametric model is unguided (model-free) and does not assume any given mode of inheritance. This is appropriate since the mode of inheritance of the susceptibility loci for KD is unknown and not necessarily the same for each locus. The Kong and Cox algorithm performs multipoint calculations and hence allows the full utilisation of the information available on descent and the inheritance pattern at every locus. An LOD score of 3 accompanied by p<1.00x10^-4 was considered suggestive evidence of linkage.

For the case–control study, tests of disease marker association were performed using the model-free allelic test which compares the difference in SNP minor allele frequency between KD

Figure 2  Association scan of the 1-LOD-drop region underlying Chr 13q. Top panel: Linkage scan results. Middle panel: Fine mapping association results. Boxes represent genes. The genes, from left to right are: ABCC4, CLDN10, DZIP1, DNAJC3, UGCG2, HS6ST3, MRNL2, FARP1, DCK9, and VCNL1. Bottom panel: Fine mapping association results in the context of the 395 single nucleotide polymorphisms (SNPs) typed in the genome-wide association study (GWAS). ◆ denotes meta-p value of rs7320375, ▲ for rs7986087, and # for rs7329490. * denotes the original GWAS p values for rs7320375 and rs7986087.
patients and controls. The p value was computed from an accompanying χ² statistic (with 1 degree of freedom) which measures the significance of the difference in allele frequencies observed between cases and controls. The calculated OR is modelled from the allelic test, which represents the increase in risk of KD per copy of the risk allele. For family based cohorts, tests of disease marker association were performed using the transmission–disequilibrium test (TDT) as implemented in the PLINK software (http://pngu.mgh.harvard.edu/purcell/plink/). The TDT statistic has a χ² distribution and is expressed as χ² = (T − U)²/(T + U) whereby T is the total number of times the risk allele was transmitted and U is the number of times the risk allele was not transmitted. Transmissions are only scored from heterozygous parents to affected offspring. Only complete parent–offspring trios and unambiguous transmissions were scored with the test.

For meta-analysis, data from all three cohorts were pooled and common odds ratios (ORs) across all individuals were calculated.
using the inverse variance method. The meta-p value was calculated using Fisher’s combined probability test, as previously described. This method of meta-analysis combines p values from several independent cohorts into one test statistic $X^2$ using the formula defined by

$$X^2 = -2 \sum_{i=1}^k \log (p_i),$$

whereby $p_i$ is the p value for the $i^{th}$ independent cohort. The number of degrees of freedom is 2k.

The significance of the meta-p values was evaluated in the context of 429 independent tests (395 from the GWAS and 34 from fine mapping; see Results), and meta-p<0.00012 was considered to be significant experiment-wide. All p values reported here are two-tailed.

RESULTS

The results of the genome-wide linkage analysis are displayed in figure 1 (regions showing a nominal LOD $>2.0$ are included in supplementary table 3). Five chromosomal segments (Chr 3q, 4q, 10p, 13q, 21q) showed suggestive evidence of linkage to KD (figure 1 and supplementary figure 2; LOD score $>5.0$, corresponding to $p=1.0 \times 10^{-4}$). We then examined data from our previous GWAS$^3$ for validated SNPs (SNPs which were associated in both the GWAS and replication study) underlying the five linkage peaks. Two such SNPs were found underlying two of the five peaks (rs16869859 with $p=0.00012$ underlying Chr 4q, and rs285052 with $p=1.7 \times 10^{-5}$ underlying Chr 13q) (supplementary table 4).

We chose Chr 13q for in-depth assessment because the statistical evidence of association exceeded the pre-set threshold of $p<10^{-4}$ and was much stronger than the association to Chr 4q.

By first systematically examining genotype data from the GWAS$^3$ for SNP coverage within the 1-LOD drop interval defining the Chr 13q linkage region (between genomic positions 94.46 Mb and 101 Mb, figure 2), we identified 395 genotyped SNPs that passed quality control (see supplementary methods). From these 395 SNPs, we observed an excess of SNPs (n=8) showing association at p<0.01 in the GWAS (N=107 cases and 134 controls, table 1) compared to chance alone (n=4). When these eight SNPs were genotyped in all remaining samples excluding the GWAS samples, (N=1177 KD cases within a total of 5010 subjects), two SNPs (rs7986087 and rs7520575) within ABCC4 retained evidence of significant association with KD (table 1).

We then proceeded to fine map the region delineated by the two SNPs within ABCC4 in all available KD subjects and controls samples (191 KD subjects, 225 controls, and 715 KD case-parent trios), as well as to exclude the involvement of the region around rs285052 (near FARP1) more definitively. This was done by genotyping an additional 34 SNPs within the ABCC4 and FARP1 regions (table 2). Fine mapping revealed a third SNP within ABCC4 (rs7529490), located <5000 base pairs away from the originally associated SNPs (rs7986087 and rs7520575) which also showed suggestive evidence of association with KD ($p=5.8 \times 10^{-5}$). In contrast, no association with KD was observed for all 11 SNPs genotyped in the vicinity of FARP1 (table 2, figure 2). The overall/gene-wide extent of linkage disequilibrium was moderately low within ABCC4.

However, the three associated SNPs were in very tight LD ($r^2 >0.95$), confined within a block 7000 bp in size (supplementary figure 3), thus suggesting this 7 Kb block to be the region of association where the causative variant(s) may lie.

Combining the data from all available samples, we found evidence of association with KD for three ABCC4 SNPs: meta-$p=5.8 \times 10^{-4}$ to $8.8 \times 10^{-5}$, and per-allele ORs=1.87 to 2.00, table 2, figure 2). Detailed genotype analysis of all KD cases and the population based controls for all three SNPs revealed that individuals carrying a single risk allele were significantly over-represented among the KD cases, and almost all individuals homozygous for the risk allele were KD cases (table 3). Homozygosity for any one SNP was linked to homozygosity at the other two loci. Haplotype analysis between the three SNPs did not reveal levels of significance exceeding that observed with single marker analysis, likely due to their very strong inter-marker LD.

DISCUSSION

We utilised three distinct but complementary study approaches to search for disease genes. In the first stage, linkage analysis tracked sharing of large chromosomal segments between family members affected with KD (see supplementary figure 4 for informative families significantly contributing to the LOD score), assisting in identifying ‘regions of interest’ for further detailed enquiry. The finer resolution of the case-control approach allowed comparison of allele frequencies of genetic variants between KD cases and controls. While having greater statistical power to detect disease associations, case-control

---

**Table 3** Genotype distributions of the three ABCC4 single nucleotide polymorphisms (SNPs) in cases and controls

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Dutch</th>
<th>UK + Australia</th>
<th>All cases</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>%</td>
<td>n</td>
<td>%</td>
</tr>
<tr>
<td>a) rs7986087</td>
<td>11</td>
<td>173</td>
<td>240</td>
<td>87.59</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>16</td>
<td>8.38</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>22</td>
<td>2</td>
<td>1.05</td>
<td>3</td>
</tr>
<tr>
<td>Total</td>
<td>191</td>
<td>274</td>
<td>465</td>
<td>225</td>
</tr>
<tr>
<td>b) rs7320375</td>
<td>11</td>
<td>171</td>
<td>89.53</td>
<td>360</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>18</td>
<td>9.42</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>22</td>
<td>2</td>
<td>1.05</td>
<td>3</td>
</tr>
<tr>
<td>Total</td>
<td>191</td>
<td>274</td>
<td>465</td>
<td>225</td>
</tr>
<tr>
<td>c) rs7329490</td>
<td>11</td>
<td>175</td>
<td>89.29</td>
<td>360</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>16</td>
<td>8.35</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>22</td>
<td>2</td>
<td>1.06</td>
<td>3</td>
</tr>
<tr>
<td>Total</td>
<td>191</td>
<td>275</td>
<td>474</td>
<td>225</td>
</tr>
</tbody>
</table>

*Cases from the US cohorts were not included for this analysis due to the low proportion of individuals of European descent in these cohorts.*

Table 4 Independent statistical evidence summarising support for the Chr 13q locus at ABCC4

<table>
<thead>
<tr>
<th>N</th>
<th>Analysis</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linkage results</td>
<td>26 pedigrees</td>
<td>Linkage</td>
</tr>
<tr>
<td>LD mapping in case—controls</td>
<td>191 cases, 225 controls</td>
<td>Case—control</td>
</tr>
<tr>
<td>LD mapping in trios</td>
<td>715 complete trios</td>
<td>TDT</td>
</tr>
<tr>
<td>All LD mapping results</td>
<td>Fisher’s test</td>
<td>$8.8 \times 10^{-5}$</td>
</tr>
<tr>
<td>Addition of WTCCC controls</td>
<td>4674</td>
<td>Case—control</td>
</tr>
</tbody>
</table>

*WTCCC controls were added to the case—control sample and re-analysed. LD, linkage disequilibrium; TDT, transmission disequilibrium test; WTCCC, Wellcome Trust Case Control Consortium.*
studies potentially suffer from biases such as population stratification and phenotype misclassification.

Fortunately, KD is rare and the proportion of controls with unrecognised KD will be very low. The transmission disequilibrium test is substantially more robust compared to the case-control approach in that it protects against population stratification by scoring transmissions only from heterozygous parents to affected offspring. Consistency of results observed across all three study methodologies confirmed disease association independent of linkage signals.

Nonetheless, our study suffers from limitations. The sample size for the linkage screen was very modest (N = 26 pedigrees analysed); in light of this, statistical power was limited and other significant disease loci with smaller effect sizes compared to that observed at ABCC4 (per-allele OR = 2.0) could have been missed. This was likely the case for NAALADL2 and ZFHX3, two significantly associated genes in the GWAS (per-allele OR = 1.45 and 1.56, respectively) where no major evidence of linkage in their chromosomal segments was detected in these 26 families. Thus, our linkage study was powered only to detect moderately large genetic effects reliably exceeding experiment-wide thresholds of significance, and a larger sample of pedigrees will be required for detection of more modestly contributing alleles (eg, OR ~1.5).

Turning to the experimental findings, we noted that in the case-control association study, individuals homozygous for the risk allele of the three ABCC4 SNPs were only found among the KD cases, with none observed in our initial Dutch (n = 225) and HapMap CEU controls (n = 60). Only by adding controls from the WTCCC2 (n = 4674), which had data for SNP rs7986087 and thus allowed a total of 4959 controls to be examined, were we able to observe homozygosity (n=4, frequency of homozygous mutants=0.08%). Thus, the estimate of the odds of disease conferred by homozygosity is 8.68 (95% CI 1.94 to 36.9; p < 10^{-7}, table 4, supplementary table S). We noted that the minor allele frequency of ABCC4 rs7986087 was not significantly different across all three Caucasian control groups (Dutch MAF = 1.8%, n = 225, HapMap CEU = 5.8%, n = 60, and WTCCC2 = 2.3%, n = 4674). Strikingly, homozygosity for the risk allele at the three SNPs was shared across the same individuals, in direct keeping with observations in the linkage study, lending credence that convergence of linkage and association signals could be used to identify disease genes. One other phenotype which showed such consistency between linkage, family based, and association studies is type 2 diabetes, where the gene TCF7L2 (RR = 1.45−1.5, current meta-p on all known cohorts <10−100) was identified from an original linkage signal (LOD = 2.79). Future studies will focus on re-sequencing within the associated region of ABCC4 to identify the causative variant(s).

ABCC4 (also known as MRP4; multi-drug resistance protein 4) belongs to a family of ATP binding cassette transporters. It has been shown to be crucial in the release of prostaglandins from cells; in particular, ABCC4 mediates the efflux of PGE1 and PGE2, 13 a function that is inhibited by non-steroidal anti-inflammatory agents in vitro. 14−16 Our genetic findings thus suggest a role for ABCC4, a protein responsible for prostaglandin transport and augmentation of inflammation, in the pathogenesis of KD. Re-sequencing of ABCC4 is warranted to better understand the risk alleles associated with KD susceptibility. ABCC4 is a large gene (282 Kb) and thus re-sequencing would need to be targeted to regions of highest disease association. As carriers of these ABCC4 genetic variants have on average, a twofold per-allele increased risk of KD, knowledge of this gene’s function in KD may provide valuable insight into disease pathogenesis.

Author affiliations
1Division of Infectious Diseases, Genome Institute of Singapore, Singapore
2Department of Epidemiology and Public Health, Centre for Molecular Epidemiology, National University of Singapore, Singapore
3Division of Human Genetics, Genome Institute of Singapore, Singapore
4Department of Pediatrics, University of California San Diego School of Medicine and Rady Children’s Hospital, San Diego, California, USA
5Department of Pediatrics, Juntendo University Graduate School of Medicine, Ube, Japan
6Department of Cardiology, Children’s Hospital Boston, Boston, Massachusetts, USA
7Murdoch Children’s Research Institute, The Royal Children’s Hospital, Parkville Victoria, Australia
8Department of Paediatric Haematology, Emma Children’s Hospital, Academic Medical Center (AMC), Immunology and Infectious Diseases, Amsterdam, The Netherlands
9Department of Blood Cell Research, Sanquin Research and Landsteiner Laboratory, University of Amsterdam, Amsterdam, The Netherlands
10Imperial College London, Department of Pediatrics, Division of Medicine, London, UK

Acknowledgements We thank all individuals with Kawasaki disease and their families for participating in this study. We thank Wei-Yang Meah, Hong-Boon Toh, Xiaoyin Chen, Khaï-Koon Heng, Chang-Hua Wong, Dennis Ek Tan, Nicholas Jw Tay, Poh-Ling Ng, Sherrina HY Chen, and Sigeeta Rajaram for technical support and expertise. This work supports activities by a grant (HL093133) from the National Institutes of Health, National Heart, Lung, Blood Institute (to JCB), the Raine Medical Research Foundation, the Ada Bartholomew Medical Research Trust, the University of Western Australia, the National Heart Foundation of Australia, and the Agency for Science, Technology and Research, Singapore.

Competing interests None.

Ethics approval Ethical approval was obtained from the appropriate national and regional institutional review boards for each study population.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES
9. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. PLINK: a tool for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007; 81:559−75.


APPENDIX 1

US Kawasaki Genetics Consortium

Jane W Newburger, Annette L Baker, Cardiology; Boston Children’s Hospital, Boston, MA; Anne H Rowley, Stanford T Shulman, Pediatrics, Northwestern University Feinberg School of Medicine, Children’s Memorial Hospital, Chicago, IL; Wilbert Mason, Pediatrics, Masato Takahashi, Cardiology, Los Angeles Children’s Hospital, Los Angeles, CA; Marian E Melish, Pediatrics, Kaplani Settings Children’s Hospital, Honolulu, Hawaii; Adriana H Tremoulet, University of California San Diego.

Members of the International Kawasaki Disease Genetics Consortium

Australia: Miranda Odam, Frank Christiansen, University of Western Australia, Campbell Witt, Royal Perth Hospital, Perth, Western Australia; Paul Goldwater, The Women’s & Children’s Hospital, South Australia; Nigel Curtis, Royal Children’s Hospital, Victoria; Pamela Palasanthiran, John Ziegler, Sydney Children’s Hospital, Randwick, New South Wales; Michael Nissan, Royal Children’s Hospital, Clare Nourse, Mater Hospital, Queensland.

The Netherlands: Irene M Kupers, Jaap J Ottenkamp, Judy Geissler, Maarten Biezeveld, Emma Children’s Hospital, Academic Medical Center; Luc Filippini, Juliana Children’s Hospital, The Hague.

Singapore: Ling Ling.

UK: Paul Brogan, Nigel Klein, Vanita Shah, Michael Dillon, The Institute of Child Health, London; Robert Boyo, Belane Shingadia, Anu Bose, Thomas Mukasa, Royal London Hospital, London; Robert Tulloh, Bristol Royal Hospital for Children, Bristol; Colin Michie, Ealing Hospital, London.
Genome-wide linkage and association mapping identify susceptibility alleles in ABCC4 for Kawasaki disease

Chiea Chuen Khor, Sonia Davila, Chisato Shimizu, et al.

doi: 10.1136/jmg.2010.086611