UNIVERSITÀ DEGLI STUDI DI PADOVA

DIPARTIMENTO DI BIOLOGIA

Corso di Laurea magistrale in *Molecular Biology*

TESI DI LAUREA

GENETIC BASIS OF KAWASAKI DISEASE

Relatore: Prof. Luca Pagani
Dipartimento di Biologia

Laureanda: Cecilia Carmignoto

ANNO ACCADEMICO 2022/2023

TABLE OF CONTENTS

I. INTRODUCTION 1
1. HISTORY 1
2. EPIDEMIOLOGY 1
3. INFECTIOUS ETIOLOGY 2
4. GENETIC PREDISPOSITION 3
5. RISK FACTORS: AGE; SEX; SEASONALITY; EAST ASIAN ORIGIN 4
6. PHYSIOPATHOLOGY 5
7. DIAGNOSIS AND TREATMENT 5
II. AIM OF THE THESIS 7
III. MATERIALS AND METHODS 8
1. COHORT DESCRIPTION 8
2. SEQUENCING AND VARIANT CALLING 8
3. QUALITY CONTROL 9
4. POPULATION STRATIFICATION 11
5. VARIANT ANNOTATION 12
6. ASSOCIATION ANALYSIS 13
6.1 Hypothesis 13
6.2 Single variant based-analysis 13
6.3 Gene based-analysis 15
6.4 Variant Selection 16
6.5 P-value correction 16
IV. RESULTS 18
1. DATA QUALITY 18
1.1 Cohort description 18
2. SINGLE VARIANT ANALYSIS RESULTS 20
3. GENE BASED ANALYSIS RESULTS 24
3.1 Investigation of *GTF3C5* association signal 24
3.2 Investigation of the *OAS1*, *OAS2*, and *RNASEL.* MIS-C susceptibility genes 27
V. DISCUSSION AND CONCLUSIONS 28
VI. REFERENCES 32
VII. ANNEXES 29

ABSTRACT

Kawasaki disease (KD) is a rare acute systemic vasculitis syndrome that typically affects young children, the most frequent complication is the formation of a coronary artery aneurysm potentially followed by cardiovascular sequelae. KD is a worldwide illness and the leading cause of acquired heart disease among children in developed countries. While the exact etiology of Kawasaki disease remains unknown there is growing evidence to suggest the contribution of genetic factors to the disease. The leading theory to explain KD pathogenesis suggests the disease is triggered by an infection, which subsequently initiates an abnormal immune response in genetically susceptible individuals. To identify the genetic factors that may have a role in KD susceptibility we conducted whole-exome sequencing of 76 patients from all over the world. The association analysis identified the suggestive association of four common SNPs located in *CCHCR1* and five rare coding variants in the gene *GTF3C5*. These results provide genetic variants potentially involved in the development of KD, but they need to be validated in larger cohorts and new investigations appear necessary to gain further insights into this disorder.

ABSTRACT

La malattia di Kawasaki (KD) è una rara sindrome di vasculite sistemica acuta che colpisce tipicamente i bambini; la complicazione più frequente è la formazione di un aneurisma coronarico al quale può seguire una sequela cardiovascolare. La KD è una malattia diffusa in tutto il mondo ed è la principale causa di cardiopatia acquisita tra i bambini nei paesi sviluppati. Sebbene l'esatta eziologia della malattia di Kawasaki rimanga sconosciuta, vi sono sempre più prove che suggeriscono il coinvolgimento di fattori genetici nella malattia. La teoria principale per spiegare la patogenesi della KD suggerisce che la malattia sia innescata da un'infezione, che successivamente avvia una risposta immunitaria anomala in individui geneticamente suscettibili. Per identificare i fattori genetici che possono avere un ruolo nella suscettibilità alla KD, abbiamo condotto il sequenziamento dell'intero esoma di 76 pazienti provenienti da tutto il mondo. L'analisi di associazione ha identificato l'associazione suggestiva di quattro SNP comuni localizzati in CCHCR1 e cinque varianti codificanti rare nel gene GTF3C5. Questi risultati forniscono varianti genetiche potenzialmente coinvolte nello sviluppo della KD, ma devono essere convalidati in coorti più ampie e nuove indagini appaiono necessarie per ottenere ulteriori approfondimenti su questa patologia.

I. INTRODUCTION

1. HISTORY

In January 1961 the first case of Kawasaki disease was seen by Tomisaku Kawasaki. Over the next 5 years he reported 50 patients with the same clinical symptoms and published the first KD report in Japan in 1967. After the initial description of the syndrome there ensued a controversy among scientists whether KD was self-limited with no sequelae or rather was related to subsequent cardiac consequences as seen in a number of the cases reported. The discussion was resolved in 1970 when in Japan the first national survey reported 10 cases of KD with fatal coronary artery aneurysm. During the 60s and 70s KD cases started to be recognized all over the world (India, Hawaii, USA). The explanation for this simultaneous recognition is currently under investigation. Possibly, the disease emerged in Japan and then spread to the Western world through Hawaii. Alternatively, KD was existing before but was simply never recognized as a distinct clinical entity because of rash/fever illness high incidence during the pre-antibiotic. Since 1970 Japan has been conducting a biannual survey to improve the understanding of KD epidemiology and hence, help in the identification of KD pathogenesis.

2. EPIDEMIOLOGY

Kawasaki disease (KD) is a rare acute systemic vasculitis syndrome, generally self-limited, that typically affects children between the ages of 6 months and 5 years. Kawasaki disease is considered as a worldwide illness since cases are reported in more than 60 countries in the 5 continents. Based on national survey data and national health insurance reviews the highest prevalence is seen in Northeast Asian countries (Japan, South Korea, China, Taiwan) with rates 10 to 30 times higher than the United States or Europe (Rowley and Shulman, 2018). The incidence rate reported in Japan is 308 per 100 000 children under the age of five in 2014 and rising to 359.0/ 100 000 in 2017-2018 (Makino et al., 2018). Second

and third highest incidence are seen in South Korea where 194.7 cases per 100,000 children <5y.o were reported in 2014 (Kim et al., 2017) and Taiwan with 82.8 cases per 100,000 children <5y.o in 2010 (Lin et al., 2015). In Japan KD incidence has risen rapidly since 1990, while in the western world it appears to be stable with an incidence up to 25 per 100,000 children in the USA. Interestingly, after the outbreak of COVID-19 pandemic in 2020, the number of KD cases decreased significantly. One possible explanation for this change is mask wearing, that would be supporting the hypothesis that KD is triggered by inhalation of a ubiquitous respiratory agent (Ae et al., 2022).

3. INFECTIOUS ETIOLOGY

The etiology of Kawasaki disease is still unknown after more than 50 years from the first reported case. The leading hypothesis is that Kawasaki disease is caused by an abnormal immunological reaction triggered by the combination of an unknown infectious agent and a genetically predisposed individual. The theory of an infectious etiology is supported by several lines of evidence related to the clinical manifestations, the age distribution, the spatiotemporal and familiar clustering, and other factors. KD symptoms such as sudden high fever arising in previously healthy children, coupled with the lack of response to antibiotics, is suggestive of an acute infection possibly driven by non-bacterial pathogens (Rowley, 2018). Immunological features observed in KD patients, being neutrophils predominance in the peripheral blood during the first phase and later T lymphocytic infiltrates, are totally compatible with an innate immune response followed by an adaptive immune response to an acute infection. KD age distribution, where incidence rates are higher for children < 2 yr. and lower for < 6-month-old, is typical of childhood infection. Indeed, the presence of transplacental immunity in infants and the development of a functional immune system in older children would be protective against KD's pathogen infection (Rowley and Shulman, 2018). KD's seasonality and wave-like spreading with peaks during

January and June/July is possibly related to the specific times of the year during which the causal pathogen/pathogens is/are more present or easily transmitted (Burns et al., 2013). Comparison of epidemiological patterns between KD and other infectious diseases suggests that the etiological agent is transmitted by close contact and that it remains asymptomatic in most of the hosts. Multiple infectious organisms have been proposed as the causal agent for KD, including bacteria, fungal agents, and viruses (Chang et al., 2014). Multiple studies report enrichment of some viruses (i.e., EBV antibodies, coronavirus, retrovirus) in KD patients compared to controls but it was never possible to prove an etiological role (Nakamura et al., 2019).

In April 2020, a new multisystem inflammatory syndrome in children (MIS-C) that resembles KD emerged. Epidemiological data indicate that MIS-C typically occurs about 1 month after SARS- CoV-2 infection. These findings support the hypothesis of viral triggers for the various forms of classic KD (Sancho-Shimizu et al., 2021).

4. GENETIC PREDISPOSITION

The epidemiological findings strongly suggest that a genetic component plays a major role in KD etiology. KD occurrence, strikingly higher in East Asian populations, is suggestive of an ethnic-specific incidence rate that is explainable by either a genetic predisposition or by the presence of a particular environmental-lifestyle (Onouchi, 2018). This second option has been excluded thanks to a study conducted in Hawaii where children of both Japanese and American ancestry are living in the same environmental conditions (Holman et al., 2005). KD is also demonstrated to show familial aggregation: it is observed that siblings and first-degree relatives of children affected by KD have a tenfold higher risk of developing the disease compared to that of the general population (Uehara et al., 2003). Therefore, epidemiological data strongly suggest that interindividual variability in KD susceptibility and different KD prevalence among ethnicities are

related to a genetic component. These premises lead to a common effort towards the identification of genetic polymorphisms associated with KD susceptibility and disease severity. To date, human genetic studies of KD focused on common variants by means of candidate gene studies or genome-wide association studies. Four GWASs on KD performed in Europeans and East Asians population have identified several significant genome-wide hits within the human antigen leukocyte region (HAL), *ITPKC* (Onouchi et al., 2007) gene that is involved in T cell activation, *CD40* and *BLK* genes involved in B cell activity, *FCGR2A*, *CASP3* and others (Onouchi et al., 2012). However, effect size is modest (OR < 1.5) and cannot explain the full clinical variability.

5. RISK FACTORS: AGE; SEX; SEASONALITY; EAST ASIAN ORIGIN

Kawasaki disease age at onset is for 85-90% of the cases between 6 months and 5 years, with peaks in children of 18-24 months in Japan and 6-12 months in the USA. Very few cases are reported for patients of 18-30 years old in various countries. Higher cases are reported for males compared to females with a male: female ratio of 1.5:1 (Rowley and Shulman, 2018).

Kawasaki disease appears to be distributed in clusters, three epidemics have been reported in Japan during the years 1979, 1982 and 1986 (Uehara et al., 2012). In some geographical regions such as Japan, Hawaii and San Diego, KD shows a temporal cluster that follows a bimodal seasonality. The peaks are reported in January, the coldest month, and June/July, months with the highest precipitation (Burns et al., 2013).

Studies in the USA describe a clear variation in KD incidence based on self-reported ancestry. The highest incidence is seen among East Asians and Pacific Islanders followed by African American, European- Americans and Native American. East Asian ethnicity is a well-known risk factor for the occurrence of KD in Hawaii and USA. In Hawaii, East Asian children resident show incidence of

~2.5 times higher than in any other US state, and KD incidences for the various ancestries are concordant with the numbers reported in the mainland (Holman et al., 2005).

6. PHYSIOPATHOLOGY

KD onset is followed by abnormal activation of both innate and adaptive immunity and their infiltration in the artery walls. During the acute phase of the disease there is a prevalence of neutrophils, macrophages, and monocytes in the peripheral blood, accompanied by their infiltration in the artery wall. In the first two weeks, neutrophilic infiltrations gradually destroy the artery wall leading to necrosis. The progression of the inflammatory process is characterized by the presence of CD8+ T cells, IgA+, plasma cells and eosinophils that contribute to the secretion of pro-inflammatory cytokines such as IL-1beta (Noval Rivas and Arditi, 2020). Contribution to the overall damage is given by activated coronary endothelial cells and smooth muscle cells. The possible result of necrotizing arteritis is the development of coronary artery aneurysm (CAA) followed by subacute or chronic vasculitis and luminal myofibroblast proliferation that can be observed for months or years.

7. DIAGNOSIS AND TREATMENT

The first clinical manifestation of Kawasaki syndrome is persistent high fever for at least five days followed by 4 or 5 of the diagnostic criteria:

- changes in the oral cavity such as cracked lips or strawberry tongue
- polymorphous rash
- bilateral conjunctivitis
- changes in the extremities such as desquamation of hands and toes
- cervical lymphadenopathy

Some laboratory and echocardiographic tests can be used to help the diagnosis and exclude other diseases. Patients that do not satisfy all the criteria for complete Kawasaki syndrome diagnosis can be classified as atypical or incomplete Kawasaki.

Standard treatment consists of a high dose of intravenous immunoglobulin (IVIG), most effective when administered in the first 10 days from disease's onset. IVIG is shown to greatly reduce the incidence of coronary aneurysm from about 30% to 5-7%, possibly thanks to the inhibition of inflammatory cytokines (IL-1beta). However, the precise mechanism by which IVIG reduces the inflammatory response is still unknown; up to 20% of the patients do not respond to the treatment and are at higher risk to develop coronary artery aneurysm. For patients that are refractory to IVGI therapy possible treatment options include corticosteroids and TNF-alpha inhibitors.

The patient prognosis is mainly dependent upon the extent of coronary artery involvement. Case- fatality rate is less than 0.2% in Japan and United States, principal cause of death is myocardial infarction resulting from coronary artery occlusion (Rife and Gedalia, 2020)

II. AIM OF THE THESIS

The main aim of the thesis is to investigate the genetic predisposition to Kawasaki diseases focusing mainly on the role of rare genetic variants. The hypothesis is that rare monogenic inborn error of immunity (IEI) (Casanova, 2015) with strong effect could underly Kawasaki disease. Interestingly, rare autosomal recessive deficiencies of OAS1, OAS2, or RNASEL were recently reported in five unrelated children with MIS-C (Lee et al, 2023)), further supporting our hypothesis.

III. MATERIALS AND METHODS

1. COHORT DESCRIPTION

Over the last 10 years, the laboratory of Human Genetics of Infectious Diseases has enrolled 202 infants and young children under the age of 5 with Kawasaki disease from all over the world. Among them, 79 had whole exomes sequencing data available at the time of the internship. The individuals used as controls for the analysis were selected from the laboratory of Human Genetics of Infectious Diseases (HGID) inhouse database which includes children and adult patients from various ethnic origins and with various infectious diseases and for which whole exome or whole genome sequencing data is available. A total of 3577 samples not suffering of severe viral illness were used as controls in this study. Consent forms for clinical and genetic studies were signed by each participant or by their parents, and all research was conducted according to the ethical standards defined by the Helsinki declaration (General Assembly of the World Medical Association, 2001).

2. SEQUENCING AND VARIANT CALLING

The whole-exome (N= 79 cases and 3393 controls) or whole-genome (N=0 cases and 181 controls) was sequenced at several sequencing centers, including the Yale Center for Genome Analysis (USA), and the New-York Genome Center (NY, USA). Libraries for WES were generated with Agilent SureSelect (Human All Exon V4, V4+UTRs, and V6) panels. Raw reads were aligned and mapped to the human reference genome assembly hg19 – NCBI build 37 using the Burrows-Wheeler Aligner (BWA). Post alignment processing procedures from Genome Analysis Software Kit (GATK version 3.4-46) best-practice pipeline were applied to minimize eventual artifacts that may affect the quality of WES/WGS data. PCR duplicates were removed with Picard tools (broadinstitute.github.io/picard/). The

GATK base quality score recalibration (BQSR) was applied to correct sequencing artifacts. Individual genomic variant call files (gVCF) were generated with GATK HaplotypeCaller, and joint genotyping of all cases and controls was performed with GATK Genotypic in union interval of all the main WES capture kits ±200 bp.

3. QUALITY CONTROL

Data quality control is a fundamental step and consists of removing bad quality variants that are likely to be the result of a sequencing or genotyping error. The inclusion of those variants in the analysis would introduce a source of error and affect both the type I error and the power of the study (Lee et al., 2014). Data quality control is performed at three different levels: genotype, variant and individual level.

Genotype level: 3 filters.

- Genotype quality (GQ) - Genotypes showing GQ<20 were set to missing. GQ is computed by GATK as a measure of the genotype quality. It corresponds to a Phred-scaled confidence that the genotype called is correct. It is based on a second measure called PL that is the Phred-scaled likelihood of the possible genotypes, set to 0 for the most likely genotype. The GQ score is calculated by subtraction between the PLs of the second most likely genotype minus the one for the most likely genotype. The top value for the second most likely genotype is set to 99 (no information is added if the PL goes over this threshold). Therefore, the lower the GQ the less confident is the called genotype since its PL has little difference from the second most likely genotype. GQ score <20 were set to missing
- Depth of coverage (DP) - measure of coverage represented by the number of reads that are available per position for each sample. Genotypes with DP<8 were set to missing.
- Minor reads ratio (MRR) - ratio of the number of reads for the minor allele over the total number of reads in the same genomic position for

heterozygous calls. Genotypes with MRR<0.2 were set to missing.

Variant level: 4 filters

- Hardy-Weinberg equilibrium (HWE) test - In absence of particular events, such as migration or natural selection, the genotype frequency at any locus simply is a function of the allele frequencies. This typical situation is defined as Hardy-Weinberg equilibrium. Deviation from the HWE may indicate the presence of population stratification and genotyping errors. Since our samples come from various ethnic origin, we used a very liberal HWE p-value threshold of e-20 to only exclude highly suspect variants (e.g., only heterozygous genotype observed)
- Call rate (CR) - Variants with more than 5% of missing genotypes are exclude. The call rate represents, for each genotyped position, the percentage of variants presenting a genotype. Low call rate indicates low genotype quality and hence must be excluded.
- Differential missingness - Missing rates between cases and controls are compared by performing Fisher's exact test. Significant differences of missingness between cases and controls may suggest batch/sequencing platform biases. Variants were excluded if the p-value of the test was below 0.00001.
- Multi-allelic filter - Variants with more than four alternative alleles were excluded.
- Indel length – insertions or deletions with length below 15 were excluded.
- Variants falling in low complexity and decoy regions were excluded. Reads mapping in LCR have high probability of being misaligned because of the repetitive nature of those regions. Decoy regions are known human genomic sequences that are not present in the reference genome, such as the areas around the centromeres, in absence of the decoy strategy those reads would align elsewhere in the genome creating noise.

Quality evaluation of the variants remaining after the filtering is assessed by

computing the ratio of transitions (A↔G, T↔C) to transversion (G↔T, G↔C). Transition mutations are more likely to happen because of the biochemical configuration of DNA double helix, transversions indeed require greater distortion of the helix. The ratio Ts/Tv is expected to be around two for genome-wide and almost three exome-wide.

Individual level:

- Ancestry and sex information have been reported by clinicians for most of the subjects of the cohort. To fill the missing information and confirm the existing ones, both sex and ancestry are inferred from the exome of each individual with a script developed by the lab.
- Heterozygosity rates are computed with both an inhouse script and bcftools in order to exclude contaminated samples (samples with higher heterozygosity rate than expected).
- Kinship coefficient is calculated using KING software. The presence of related individuals in the cohort requires specific cares. Indeed, classical association tests assumes the independence of the samples.
- Call rate at sample level to exclude bad quality samples.

4. POPULATION STRATIFICATION

A major issue in genetic association studies is population stratification that, if not accounted for, leads to spurious association. Population stratification consists in the presence of differences in allele frequency between cases and controls due to systematic ancestry differences. Without the right correction the results of the association between genotype and studied trait could reflect the different ancestral components between cases and controls rather than a true association with the disease. Genetic association studies need to account for this population structure especially when multiple ethnicities are present in the cohort. Typically, the chosen method to identify population stratification in GWAS is the Principal Component Analysis (PCA). PCA is a statistical method that reduces the

complexity of the genetic data. Data complexity is reduced to a small number of components named Principal Components (PCs), where each PC describes a proportion of the genetic variation, PC1 being the one representing most of the variability. Depending on the data a certain number of PCs will be chosen and used in a regression model to account for the population stratification. It has been shown that principal components can be used to correct for population stratification also in rare variant association studies. Moreover, it has been shown that in the context of rare diseases where only few cases are available, the addition of controls, regardless of their ethnic origin, results in a gain of power in the study provided that population stratification is adequately controlled (Bouaziz et al., 2021).

For this analysis, the PCA was conducted with the PLINK software on a subset of 29900 SNPs with MAF > 0.01 and in linkage equilibrium. Linkage disequilibrium pruning was performed using a sliding window of 50 kb, step size 5 kb and maximum r^2 of linkage of 0.2 between SNP pairs.

5. VARIANT ANNOTATION

Variant annotation is the process of adding metadata to the DNA variants selected, it is fundamental for the interpretation of the analysis' results. The prediction of the functional impact for each variant was performed using Ensembl VEP (Variant Effect Predictor) software, using the GRChr37 MANE Select transcript dataset (https://tark.ensembl.org/web/mane_GRCh37_list/). The Matched Annotation from the NCBI and EMBL-EBI (MANE) is a collaborative project that aims to converge on human gene and transcript annotation and to define a genome wide set of representative transcripts and corresponding proteins for human protein-coding genes. Each MANE transcript represents an exact match in exonic regions between a Refseq transcript and its counterpart in the Ensembl/GENCODE annotation such that the two identifiers can be used synonymously (Morales et al., 2022). Allele frequencies for different populations from the Genome Aggregation Database (GnomAD) v2.1.1 were assigned to each variant using bcftools v1.9. GnomAD spans 125,748 exome sequences and 15,708 whole-genome sequences

from unrelated individuals sequenced as part of various disease-specific and population genetic studies. Finally, the Combined Annotation Dependent Depletion (CADD) score was used to predict the potential deleteriousness of the variants. CADD is a machine learning based tool that allows to score the deleteriousness of SNPs and INDELs in the human genome. A logistic regression model is trained on a set of observed and simulated variants and a combination of different metrics is used to assign a CADD score to each possible substitution in the human genome. The CADD score is then 'normalized' to a Phred-like score ranging from 1 to 99, where the higher the score the stronger the deleteriousness of the variant. However, CADD score for known pathogenic variants across different genes can vary significantly. To solve this issue, the lab designed a gene specific threshold called MSC (Mutation Significance Cutoff) that sets the lower limit of the confidence interval (90%, 95%, 99%) for the CADD score for all its pathogenic mutations.

6. ASSOCIATION ANALYSIS

6.1 Hypothesis

The hypothesis behind variant association analysis is the existence of rare genetic variants predisposing to Kawasaki disease. Predisposing variants should be found enriched in the group of cases of the cohort analyzed. This hypothesis is tested with the assumption of genic homogeneity for at least a number of the cases tested, meaning at least a number of the cases carry variants in the same gene.

6.2 Single variant based-analysis

As strongly advised by Do et al., we performed a single variant association test previous to the gene-level analyses to verify the quality of the data and to inspect the QQ plots (quantile-quantile graph) to ensure that the statistical model is adequate and that the systematic bias due to population stratification was corrected

(Do et al., 2012). A clean QQ plot should show a solid line matching x = y until it sharply curves at the end, representing the small number of true associations. Any other deviation from the x = y line suggests a bias due to confounders. Single variant analysis also allows for the search of rather common variants that may be associated with KD. The software chosen to perform the association analysis for our binary trait is REGENIE (https://rgcgithub.github.io/regenie/). A software using a whole genome logistic regression approach and capable of accounting for relatedness and population stratification. REGENIE's workflow is divided into two main steps. The first step aims to capture most of the phenotypic variation attributable to genetic effects. This operation is performed by fitting a subset of the SNPs (MAF> 0.05) in a whole genome regression model to obtain the set of predictors. The genetic predictions are built with the LOCO (Leave One Chromosome Out) approach and stored in a matrix to be used in step 2 when testing for association. In step 2 the association between each genetic marker and phenotype is tested by means of Firth's bias corrected logistic regression including LOCO predictions as an offset, and covariates (Mbatchou et al., 2021). To account for population stratification, we used as covariates the first 10 components of the PCA. Firth logistic regression has been used to reduce the inherent bias present in the maximum likelihood estimates of the coefficients. This approach has demonstrated the ability to effectively control the type I error rate, even in scenarios characterized by small sample sizes and an unbalanced case-control ratio (Chen et al., 2021). Let Yi be the phenotype of the individual i, X_{ik} the genetic score of the individual i at SNP k, M_i the covariate matrix and β0, β1 and β2 the fixed regression coefficients. The logistic regression model can be written as follows:

$$logit(P(Y_i=1)) = \beta_0 + \beta_1 X_{ik} + \beta_2 M_i$$

The p-values resulting from the analysis are based on a likelihood ratio test (LRT). We considered three genetic models: additive, recessive, and

dominant. Under the additive model, the genetic score is coded 0, 1, or 2, according to the number of alternative alleles present for the SNP. In the recessive model, the genetic score is set to 1 only for the minor homozygous genotype, while it remains 0 for all other genotypes. In the dominant model, the genetic score is assigned a value of 1 for both the minor homozygous and heterozygous genotypes, while it is 0 for all other genotypes.

6.3 Gene based-analysis

Single-variant tests are underpowered to detect rare variant associations (Lee et al., 2014). To increase the statistical power for association analysis of rare genetic variants, alternative strategies based on the aggregation or collapsing of variants within a genetic unit have been proposed. In our study, we considered genes as the genetic unit and we evaluated the association between each gene and KD using a Cohort Allelic Sums Test (CAST) (Morgenthaler and Thilly, 2007) derived approach. CAST assumes that the presence of any rare variant in a test unit (i.e. a gene in our study) increases the disease risk. For a given gene, a genetic score is assigned to each individual depending on the presence or absence of at least one rare candidate variant and the zygosity status as reported in Table 1. The analysis was performed with REGENIE, similarly to the single variant analysis.

Table 1. *Coding of the genetic score for the gene-based rare variant association analysis*

Genetic score	At least one homozygous candidate varant	At least one heterozygous candidate variant	No candidate variant
Additive	2	1	0
Dominant	1	1	0
Recessive (homozygous)	1	0	0

6.4 Variant Selection

All variants with MAF >0.05 in the cohort were kept for the single variant analysis. Gene-based association analysis requires the definition of eligible variant sets for the analysis and consideration of which variants to be studied together: ideally, one would aggregate only harmful alleles and ignore neutral variation (Lee et al., 2014). To enrich for potentially harmful alleles, we considered several sets of variants on which to perform association testing, based on the MAF and the variant annotation:

- three subsets are created based on the MAF retrieved from Gnomad. The thresholds used are: 0.01, 0.001, 0.0001.
- two subsets are created based on the CADD score: all variants independently from the CADD score and only variants whose CADD score is above the MSC.
- two subsets are created depending on the predicted consequence in the annotation: predicted loss of function (LoF) variants only (i.e. stop gain, start lost, frameshift and essential splicing); predicted LoF, missense and inframe variants (missLoF).

The total number of variant groups obtained is 12. Each group is tested under the three different models: additive, dominant, recessive. The total number of tests performed is 36.

6.5 P-value correction

To control the false positive rate during multiple testing analysis, a threshold for p-values must be set to identify the truly significant associations. Bonferroni correction is chosen for both single variant analysis and gene-based analysis to address the issue of multiple testing. Bonferroni exome-wide significance threshold for the p-value is given by dividing the chosen alpha by the number of tests performed. For single variants, the number of tests considered was the number of variants with MAF > 5% multiplied by the three genetic models. For gene-based analysis, the number of tests considered was the sum of the number of

informative genes (i.e. with at least 3 carriers of candidate variant) for each of the variant sets and genetic hypothesis. It is important to remember that Bonferroni threshold is very conservative since the different sets tested are not independent.

IV. RESULTS

1. DATA QUALITY

A total of 3653 samples have been whole exome (N=3472) or whole genome (N=181) sequenced. At an individual level, three cases were excluded because of sex discrepancies between the inferred sex and the one reported by the clinicians. All the samples passed the contamination control. The study of the kinship coefficient showed no first-degree relatives in the cohort, and 134 individuals having one distant relative (5 cases and 129 controls having 0.0625<kinship coefficient<0.125). The total number of variants identified was 6 190 717 before QC filtering. After the QC filtering steps, 2 115 572 (~34% of the original variants) remained. Of note, most of the filtered variants did not belong to the captured exonic regions but to the 200 base pairs flanking regions. As shown in Table 2, the Ts/Tv ratio as well as the average call rate and depth per sample increased after the quality control. In particular, the Ts/Tv ratio increases from 2.29 to almost 3, as expected.

Table 2. *Quality metrics before quality control (Pre-QC) and after quality control (post-QC).*

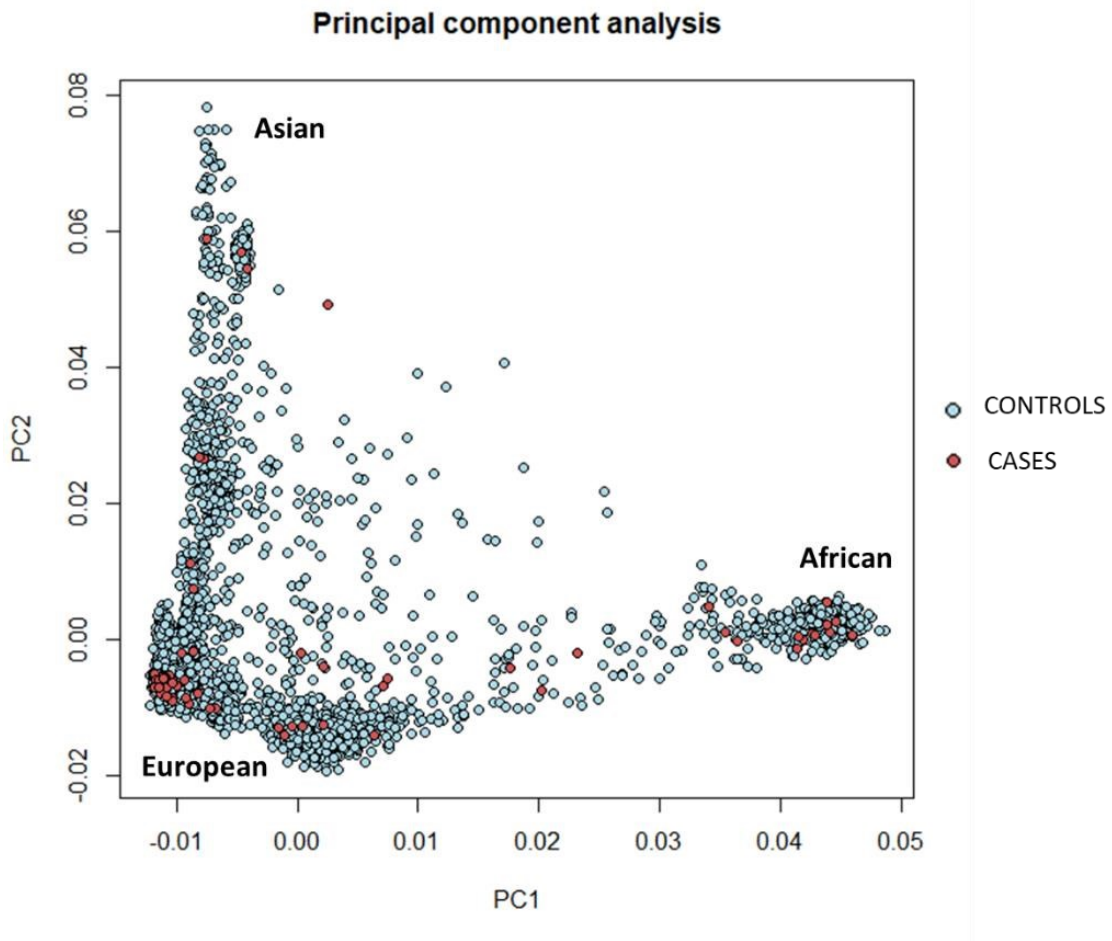
	N° samples	N° variants	Ts/Tv	N° singletons	N° heterozygous	CR per sample	Depth per sample
Pre-QC	3656	6190717	2.29	628	67713	0.96	26.40
Post-QC	3653	2115572	2.90	220	13017	0.99	36.05

1.1 Cohort description

The final cohort used for this study is composed of 3653 individuals, 46% being female and 54% being male. The male:female ratio in the cases is 3 and in the controls is 1.17. As expected, the cohort is very heterogeneous in self reported ancestry (Fig 2) and is composed mainly of Europeans (49%) and North Africans (24%) but also individuals having East Asian, Native American, and Sub-Saharan

African ancestry. The ethnic composition between the two groups of cases and controls sees a slight imbalance with an enrichment of Europeans and East Asians in the cases, this difference is acceptable and accountable for during the analysis using the data from Principal Component Analysis.

Figure 2. ***PCA -Principal Components Analysis****: Graphic representation of projection of samples, cases in red and controls in blue, to the first two principal components*



2. SINGLE VARIANT ANALYSIS RESULTS

In the single variant analysis, a total of 72 865 variants with a minor allele frequency (MAF) greater than 0.05 were examined under the three transmission models. The quantile-quantile (QQ) plots (Fig 3) showed no systematic deviation from the null hypothesis indicating that the ethnic heterogeneity of our cohort is well accounted for by the inclusion of the first 10 PCs in the logistic regression model.

No SNP show significant association with KD at the Bonferroni corrected p- value threshold of 2.3e-07 (=0.05/ [72 865x3]; Fig 4). At a less stringent e-05 suggestive threshold, we identified a cluster of 4 SNPs in high linkage disequilibrium within the HLA region on chromosome 6 associated with KD (Table 3 and Fig 5). The minor T allele of the top associated SNP rs3094226 increased the risk of KD under the dominant transmission model (OR [95% CI] = 2.89, p=3.81e-06). The four SNPs are located within a single gene, *CCHCR1*, which encodes for a coiled-coil alpha- helical rod protein. Among the four variants, rs3094226, rs2073719 are intronic variants while rs130078, rs309425 are two synonymous variants.

Using the NIH tool LDproxy (https://analysistools.cancer.gov/LDlink/?tab=ldproxy) and querying rs3094226 to explore for proxy and putatively functional variants, we identified SNPs in high LD (r^2>0.8) in all populations (Fig 5). The two sequenced SNPs rs130078 and rs309425 had a RegulomeDB score of 1f, which strongly suggest they may have regulatory consequences. Scores assigned by RegulomeDB span from 1 to 7 with 1 indicating the highest regulatory potential score. Indeed, the variants were shown to be eQTL in various tissues for several genes in the region such as *CCHCR1, HLA-C, HLA-DRB5* and *DRB1*.

***Table 3**. Best hits summary for single variant analysis results. Are reported variants having p-value<e-04 and at least 5 case carriers. Per each variant are reported position and IDs, Odd Ratio, Alternative Allele Frequency, number of carriers in the cases and controls with the state of the variant (homozygous or heterozygous), the transmission model.*

CHR	POS	REF	ALT	ID	OR	*GENE*	AAF	P-VALUE	N°CASES	N° CONTROLS	MODEL
6	31112899	C	T	rs3094226	2.89E+00	***CCHCR1***	0.199202	3.81E-06	51 (49het,2hom)	1241 (1088het, 153hom)	dominant
6	31112925	C	T	rs2073719	2.89E+00	***CCHCR1***	0.199616	3.90E-06	51 (49het,2hom)	1248 (1095het, 153hom)	dominant
6	31118565	C	G	rs130078	2.83E+00	***CCHCR1***	0.218159	6.31E-06	52 (48het,4hom)	1355 (1173het, 182hom)	dominant
6	31113052	G	A	rs3094225	2.91E+00	***CCHCR1***	0.255635	8.13E-06	57 (49het,8hom)	1535 (1275het, 260hom)	dominant
6	31129642	A	C	rs2073722	2.78E+00	*TCF19*	0.194429	1.18E-05	48 (44het,4hom)	1213 (1061het, 152hom)	dominant
6	31129616	A	G	rs2073721	2.77E+00	*TCF19*	0.19449	1.22E-05	48 (44het,4hom)	1215 (1063het, 152hom)	dominant
16	84034434	GAGGGAGACA GAGGGAAGT	G	rs141446650	3.67E+00	*NECAB2*	0.0528603	1.30E-05	22 (21het,1hom)	336 (318het, 18hom)	dominant
6	31130502	T	C	rs1065461	2.74E+00	*TCF19*	0.195051	1.52E-05	48 (44het,4hom)	1208 (1057het, 151hom)	dominant
X	100387337	C	T	rs7883144	1.64E+00	*CENPI*	0.202315	1.59E-05	29 (6het,23hom)	957 (515het, 442hom)	recessive
12	32487644	T	A	rs10844188	3.61E-01	*BICD1*	0.2055	1.89E-05	16 (16het,0hom)	1313 (1155het, 158hom)	additive
11	60059810	A	G	rs10750931	2.57E+00	*MS4A4A*	0.159271	1.95E-05	38 (31het,7hom)	1017 (916het, 101hom)	additive
11	94126620	C	T	rs78029801	9.36E+00	*GPR83*	0.0111447	2.03E-05	7 (7het,0hom)	74 (74het, 0hom)	dominant
6	138644775	A	C	rs2010579	9.02E-01	*KIAA1244*	0.0626383	2.06E-05	10 (5het,5hom)	416 (394het, 22hom)	recessive
11	60070176	A	G	rs6591561	2.75E+00	*MS4A4A*	0.319566	3.36E-05	57 (46het,11hom)	1878 (1495het, 383hom)	dominant
6	31111180	T	C	rs130073	2.61E+00	*CCHCR1*	0.261536	4.48E-05	54 (49het,5hom)	1567 (1300het, 267hom)	dominant

Figure 3. ***Quantile-Quantile plot for single variant analysis****: allelic association analysis of expected versus observed p-values of 72 865 SNPs. The black circles are showing the deviation from the line of expected p-values x=y. Additive, dominant and recessive transmission models are reported.*

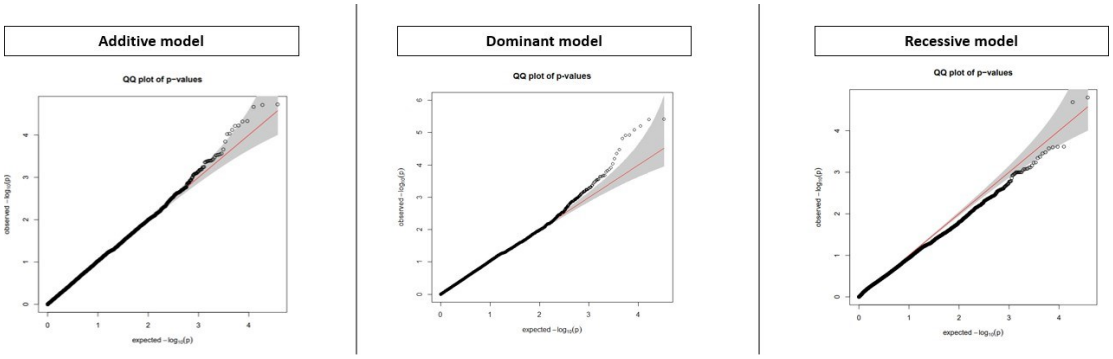


Figure 4. ***Manhattan plot for single variant analysis****: the best hits among the three transmission models are reported. The 72 865 SNPs with MAF>0.05 are reported. The x axis displays the chromosomal position and the y axis the p-values per variant in logarithmic scale. The two red lines represent the Bonferroni corrected significance threshold (higher line) and the less stringent threshold of e-05 (lower line).*

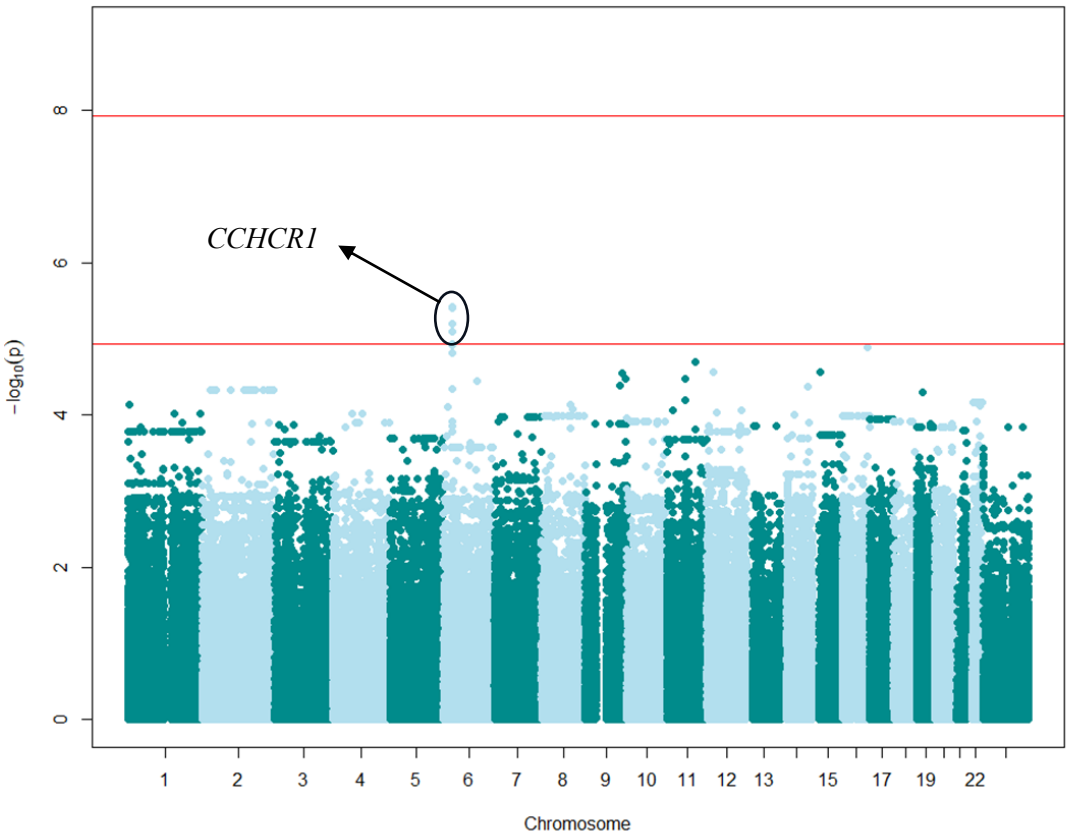
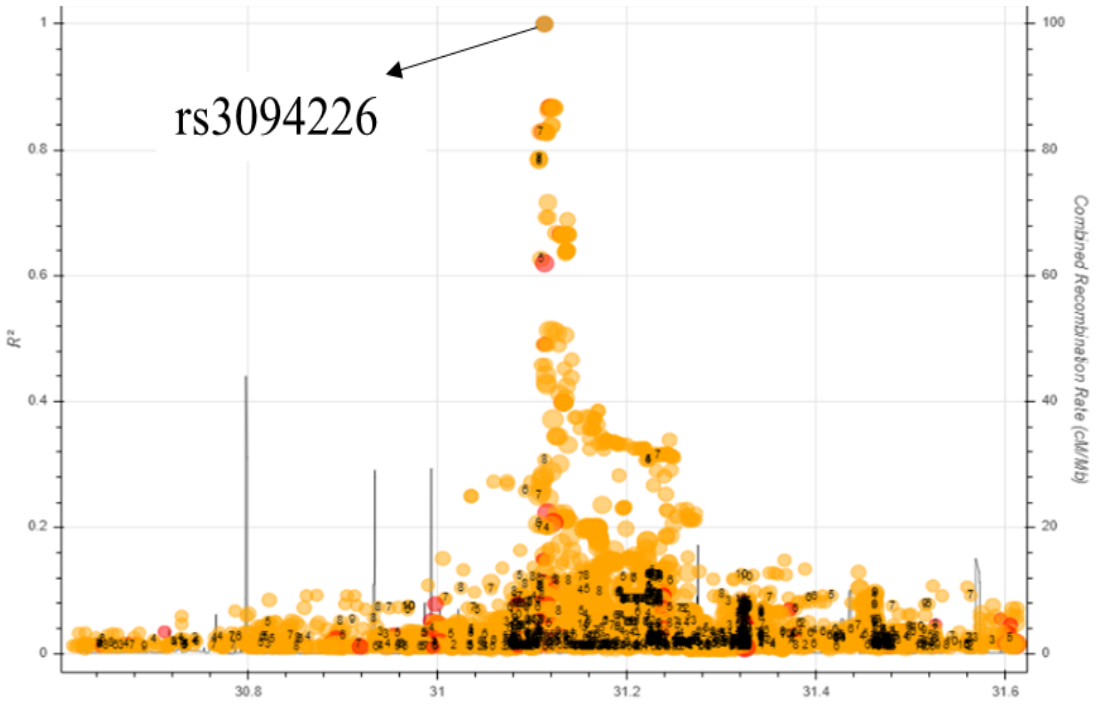


Figure 5. The LD proxy plot for rs3094226*: rs3094226 is displayed together with its proxy variants. The x axis represents the chromosomal coordinates in the genome and the y axis the pairwise R^2 with rs3094226 as well as the Combined Recombination Rate.*



3. GENE BASED ANALYSIS RESULTS

Association analysis was performed at gene level for the 12 different variant sets and 3 transmission models, leading to a total number of 36 tests and a Bonferroni corrected threshold of 8e-08. Figure 6 shows the QQ plots for missLoF variant set with Gnomad AAF > 0.05 and no CADD filtering, according to the genetic model considered. As for the single variant analysis, the QQ plots didn't show systematic deviation from the null hypothesis indicating that population stratification was well accounted for in gene-based analysis. No gene displayed a significant association with Kawasaki disease under any of the three transmission models, as evidenced by the Manhattan plot (Fig 7). At a less stringent e-05 suggestive threshold, one gene showed enrichment of rare candidate variants in KD patients compared to controls: *GTF3C5* (Table 4).

Table 4. *Top results for gene-based analysis among the three models*

CHR	POS	TRANSCRIPT	*GENE*	MAF	VARIANT	CADD> MSC	OR CI[95%]	PVAL	AAF	MODEL
9	1.36E+08	ENSG00000148308	***GTF3C5***	0.0001	LoF+ missense	True	20.5159	9.76e-06	0.003011	dominant
12	96370259	ENSG00000084110	*HAL*	0.0001	LoF+ missense	Tue	23.5162	2.53e-05	0.002464	dominant
12	53818090	ENSG00000135409	*AMHR2*	0.0001	LoF	True	68.2497	4.88e-05	0.000684	dominant
11	64064976	ENSG00000182450	*KCNK4*	0.0001	LoF+ missense	True	54.3389	5.58e-05	0.000958	dominant
17	79898714	ENSG00000185105	*MYADML2*	0.01	LoF+ missense	True	11.0745	9.98e-05	0.005325	dominant

3.1 Investigation of *GTF3C5* association signal

Five KD patients versus 13 controls carried a rare (Gnomad AF < 0.0001) missense variant with CADD > MSC in heterozygous state ($OR_{dominant}$ [95%CI] = 20.5; p = 9.76085e-06). Looking at the ethnic origin of the carriers we find 13 with European and 5 with North African ancestry (Table 5). Among the five variants identified in Kawasaki patients, variant rs181363313 has been classified as "probably damaging" and "deleterious" by the prediction tools Polyphen and SIFT, respectively. *GTF3C5* is located on chromosome 9 and encodes the General Transcription Factor IIIC Subunit 5, a subunit of the DNA-binding subcomplex.

Figure 6. Quantile-Quantile plot for gene-based analysis*: allelic association analysis of expected versus observed p-values. The black circles are showing the deviation from the line of expected p-values x=y. The set of variants analyzed in the QQ-plots is of SNPs with MAF>0.05, predicted variant consequence either missense or loss of function and no filtering on the CADD score.*

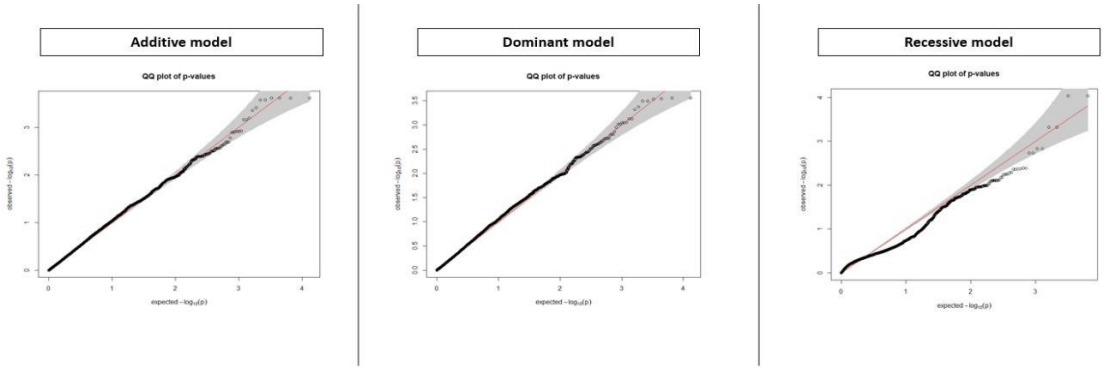
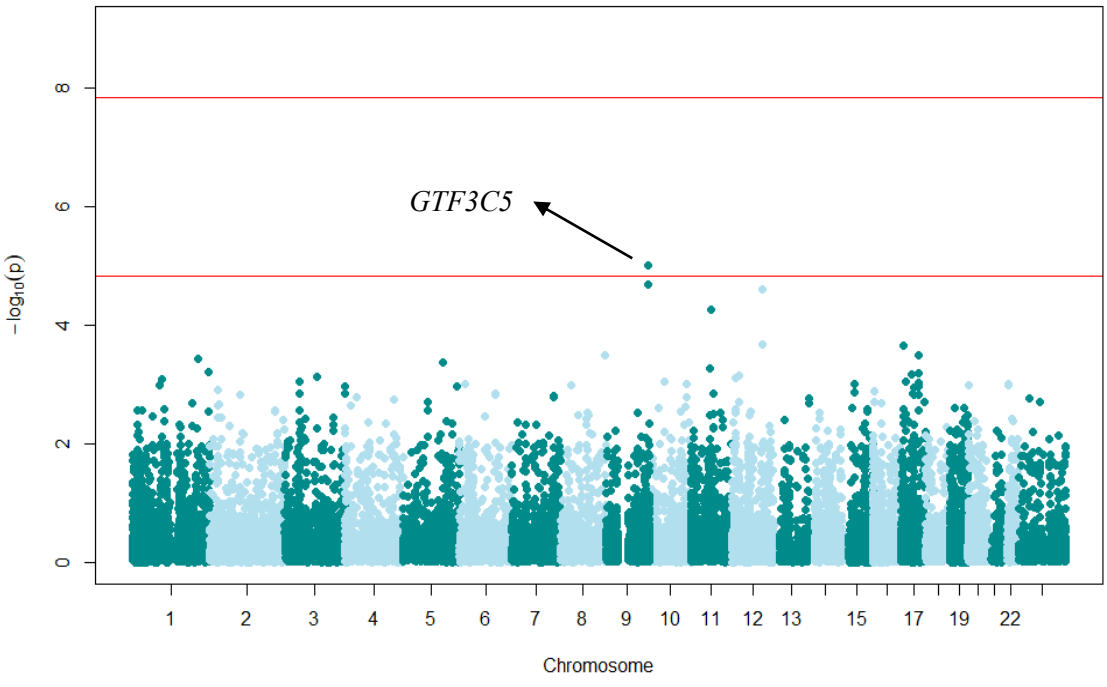


Figure 7. Manhattan plot. *Best hits for the three transmission models are reported. Along the x axis is reported the SNPs' chromosomal position, along the y axis the p-values in logarithmic scale. The top red line indicates the Bonferroni threshold of significance, and the bottom red line indicates a less stringent threshold of e-05.*



(TFIIIC2) of transcription factor IIIC (TFIIIC) and is expressed in the nucleus of all cell types including immune cells' (https://www.proteinatlas.org/ENSG00000148308-GTF3C5/immune+cell).

GTF3C5 constraint evaluation is performed using two scores assigned by Gnomad: the observed/expected (o/e= 0.88 (0.8-0.97)) and Z (Z=0.77) scores. The o/e score aims to identify genes intolerance to variation by predicting the number of variants expected to be seen and comparing those expectation to the observed number of variations (Lek et al., 2016). Positive Z score and low o/e value indicate that the gene has fewer variant than expected. *GTF3C5* scores are not indicating any particular constraint to missense mutations.

Table 5. *Variants included for the analysis of GTF3C5 genes.*

GTF35C variants	ID	Variant type	CADD (MSC=13.97)	Gnomad frequency	Transcript	Phenotype	Ethnicity
9:135919148 C/T	.	missense	22.5	2.12E-05	ENSP00000361169.5: p.Thr136Met	control	European
9:135919198 C/T	rs181363313	missense	26.9	2.39E-05	ENSP00000361169.5: p.Arg153Trp	case	European
9:135919250 C/T	.	missense	25.2	1.19E-05	ENSP00000361169.5: p.Pro170Leu	control	European
9:135919252 C/A	.	missense	23.8	7.95E-06	ENSP00000361169.5: p.Pro171Thr	control	European
9:135919252 C/A	.	missense	23.8	7.95E-06	ENSP00000361169.5: p.Pro171Thr	control	European
9:135927501 G/A	.	missense	16.15	1.21E-05	ENSP00000361169.5: p.Val275Ile	control	North_African
9:135929249 T/A	.	missense	14.79	0	ENSP00000361169.5: p.Phe303Tyr	control	European
9:135929329 G/A	.	missense	38	8.06E-06	ENSP00000361169.5: p.Gly330Ser	control	European
9:135929851 C/T	rs199931911	missense	23.4	9.90E-05	ENSP00000361169.5: p.Leu349Phe	case	European
9:135929851 C/T	rs199931911	missense	23.4	9.90E-05	ENSP00000361169.5: p.Leu349Phe	case	European
9:135929851 C/T	rs199931911	missense	23.4	9.90E-05	ENSP00000361169.5: p.Leu349Phe	control	European
9:135930164 C/T	.	missense	22.7	2.01E-05	ENSP00000361169.5: p.Arg379Trp	case	European
9:135931411 A/C	rs140832971	missense	21.4	3.89E-05	ENSP00000361169.5: p.Ile394Leu	control	North_African
9:135931413 C/G	.	missense	24.9	3.98E-06	ENSP00000361169.5: p.Ile394Met	control	European
9:135931417 C/T	.	missense	32	0	ENSP00000361169.5: p.Arg396Trp	control	North_African
9:135932176 A/G	rs373718885	missense	23.4	4.07E-06	ENSP00000361169.5: p.Asn425Ser	control	North_African
9:135932259 C/T	.	missense	19.52	0	ENSP00000361169.5: p.Leu453Phe	case	European
9:135933335 G/A	.	missense	31	3.98E-06	ENSP00000361169.5: p.Glu510Lys	control	North_African

3.2 Investigation of the *OAS1*, *OAS2*, and *RNASEL.* MIS-C susceptibility genes

Finally, we took a closer look at the results obtained for *OAS1*, *OAS2* and *RNASEL* genes. Autosomal recessive deficiencies of these three genes have been recognized as causing MIS-C (Lee et al., 2022). Our analysis of the recessive model revealed that none of the Kawasaki cases in our cohort exhibited any rare homozygous variants, whether they were loss-of-function (pLOF) or missense variants. While when examining heterozygous pLOF or missense variants we observed a trend of enrichment towards the cases, even if not significant.

Under the dominant model *RNASEL* displayed 37 predicted missense variants, one frameshift and stop gain with Gnomad AF< 0.0001, in the heterozygous state ($OR_{dominant}$= 3.6, p=0.159432). Two out of 47 carriers are Kawasaki patients (Annex 1). *OAS1* showed 19 predicted missense and two stop gained variants (Gnomad AF < 0.0001). The individuals carrying the variants were 25 controls and one Kawasaki patients ($OR_{dominant}$=2.5, p=0.518341) (Annex 2). The last investigated gene *OAS2* displayed 33 predicted missense, four frameshift and one stop gain variants in the heterozygous state (Gnomad AF < 0.0001) with one Kawasaki patient carrier and 43 controls ($OR_{dominant}$ =1.2, p= 0.856058) (Annex 3). Even if no gene displayed significant association we investigated the variants taken into consideration for the analysis (Annex 1, Annex 2, Annex3).

Table 6. *Results for genes RNASEL, OAS1, OAS2*

CHR	*POS*	*TRANSCRIPT*	*GENE*	MAF	variant	CADD> MSC	OR; CI[95%]	Pval	AAF	Model
1	*182545389*	*ENSG00000135828*	*RNASEL*	1.00E-04	LoF+ missense	False	3.6010	0.159	6.7e-03	dominant
12	*113344924*	*ENSG00000089127*	*OAS1*	1.00E-04	LoF+ missense	False	2.478890	0.518	3.8e-03	dominant
12	*113416426*	*ENSG00000111335*	*OAS2*	1.00E-04	LoF+ missense	False	1.225330	0.856	6.6e-03	dominant

V. DISCUSSION AND CONCLUSIONS

The technological advancement of high-throughput sequencing has led to the identification of several genes related to monogenic inborn errors of immunity. Genome-wide analysis and family linkage studies succeeded in identifying multiple genes that contribute to Kawasaki diseases susceptibility and outcome. However, these findings have not significantly improved the understanding of KD pathogenesis, and the majority of the genetic factors of KD remain unidentified (Lee et al., 2014). In this study, we analyzed whole-exome sequencing (WES) data to identify rare variant with strong effect size that contribute to KD susceptibility. In our association study, involving a cohort of 3653 individuals, no variant or gene was found significantly associated with KD. However, some suggestive results have been obtained for both common and rare variants. A cluster of common variants located in the gene *CCHCR1* and rare missense variants in one gene, *GTF3R5*, showed suggestive association signal that need to be replicated and further investigated.

Single variant analysis, using a threshold of e-05, which is less stringent than Bonferroni's, reveled a suggestive association between a cluster of four SNPs with KD. This association was observed under the dominant model, with OR= 2.89 [95%CI]. The four SNPs, with a MAF ~0.2, are in strong linkage disequilibrium and located within the *CCHCR1* gene, that is flanking the human leukocyte antigen (HLA-C). Further analysis are needed to overcome the challenges presented by the high LD and identify the specific contribution of individual variants within this cluster. Our hypothesis suggests that these SNPs are likely expression quantitative trait loci (eQTLs) that regulate multiple genes, such as *HLA-C* and *PSORS1C2*. The top hit SNP, rs3094226, and rs130078 are assigned a RegulomeDB score of 1f, indicating that they are located in a region of chromatin accessibility peak and transcription factor binging site. The protein product of *CCHCR1*, called coiled-coil alpha helical rod protein 1, is believed to play a role in mRNA metabolism regulation, as well as in the proliferation and differentiation

of keratinocytes and steroidogenesis (Tiala et al., 2007). *CCHCR1* natural variant are associated with psoriasis, an immune-mediated disease that causes inflammation in the body (Tervaniemi et al., 2018). Additionally, GWAS meta-analysis identified an association between *CCHCR1* and COVID-19 severity (Pairo-Castineira et al., 2021). Data also suggest that *CCHCR1* protein interacts with the E2 protein Human Papillomavirus type 16 potentially interfering with HPV activation of the keratinocytes (Muller et al., 2014). The findings suggest that *CCHCR1* is involved in viral defense and skin proliferation, both of which are relevant to the onset and progression of Kawasaki disease. Hence, it is reasonable to hypothesize that the identified variants exert a regulatory influence on genes associated with Kawasaki disease. These results should be validated in independent and larger cohorts.

Gene-based CAST analysis performed under the hypothesis of a dominant model and investigated with a less stringent threshold of e-05, resulted in a suggestively associated gene with KD: *GTF3C5*. Enrichment in rare missense variants with CADD >MSC is found in the cases (6.58%) compared to the controls (0.47%). *GTF3C5* encodes for a subunit of the DNA-binding subcomplex (TFIIIC2) of transcription factor IIIC (TFIIIC); TFIIIC2 subcomplex directly binds tRNA and virus-associated RNA promoters (Sinn et al., 1995). *GTF3C5* is crucial for general transcription and is expressed in all tissue and its functions remains largely unknown. Recent studies have suggested the association of *GTF3C5* with Hypomelanosis of Ito (HMI) (Saida et al., 2022), a rare neurocutaneous syndrome classified as a mosaic cutaneous disorder, typically caused by de novo postzygotic mutations (Arora et al., 2022). Additionally, *GTF3C5* is found associated with Herpetic Whitlow, a skin infectious disease caused by HSV-1 or HSV-2, known to be associated to *SIGLEC5* and RNA Polymerase III Transcription Initiation (Dremel et al., 2022). *GTF3C5* roles have not been deeply investigated but from the little that is known, its involvement in KD disease is plausible.

Recessive deficiency of *RNASEL, OAS1, OAS2* were found to cause MIS-C, a SARS-CoV-2 related disease of children resembling KD. Single genes recessive inborn errors of the OAS- RNASEL pathway are proved unleash the production of

SARS-CoV-2–triggered inflammatory cytokines by mononuclear phagocytes, and are hence believed to underly at least some MIS-C cases. The gene based analysis of *RNASEL, OAS1, OAS2* were analyzed but none rare homozygous pLoF or missense variants were identified. However, under the dominant model, we found a slight, non-significant, enrichment of rare missense variants in KD as compared to controls. The absence of a notable signal in our analysis can be partially attributed to the comparatively lower level of genetic homogeneity associated with KD compared to MIS-C. Unlike MIS-C, which is believed to be caused by a single agent (SARS-CoV-2), KD is potentially triggered by multiple diverse and unidentified agents. Hence we could speculate that the susceptibility to KD may involve various potential susceptibility variants that are more challenging to identify compared to MIS-C.

Among the various statistical method developed to carry out association studies, the CAST burden test was chosen in this study, mainly because of its ease of interpretation. This test is the most powerful when the variants have the same magnitude and effect direction, while it loses power in the presence of both trait-increasing and trait-decreasing variants. Only variants predisposing to Kawasaki disease were possibly identified, neglecting the existence of potential protective variants co-existing in the unit with the deleterious ones. To address this issue alternative tests could be used that take into account the variant effect direction and magnitude of effect, namely variance component-based tests as SKAT and combined tests as SKAT-O (Lee et al., 2014).

WES analysis have the advantage of being cost-effective while allowing to screen for also rare potential causal variant in a high-value portion of the genome. The main and most obvious limitation of this approach is that only the protein coding and splice regions are sequenced (1%- 2% of the genome); most of variants located in the regulatory regions are lost even if they might have significant biological function. Additional loss of candidate variants is seen during the data quality control, where strict filtering may be needed to remove the artifacts due to technological issue of the sequencing such as coverage. This technology doesn’t

allow the detection of structural variant such as inversions or copy number variations (CNVs). Therefore, in the analysis only a small proportion of the potential causal variants is taken into account. It is expected that the focus for rare variant studies will extend to genome-wide studies as the sequencing cost gradually decreases and the annotation of non-coding variants improves (Lee et al., 2014).

This WES study of KD provides some interesting potential coding variants both rare and common associated with KD. To be considered is that the results obtain are preliminary, especially because they lack significance when Bonferroni correction is applied, and need to be replicated in an independent cohort. The replication cohort will be composed by 123 additional Kawasaki cases collected by the laboratory in the last two years and recently whole-exome sequenced, and patients with non-viral illnesses as controls. The focus of the validation study will be on *CCHCR1* variants and *GTF3C5*. Additionally, we plan to perform a combined analysis including all the cases and controls. The ultimately significant genes and variants will be investigated by the genetic immunology team of the lab. It is hoped that the discovery of KD genetic predisposing will help the understanding of KD etiology, and possibly allow the development of new therapies and preventive strategies.

VI. REFERENCES

2765(03)00433-7.

1. Ae R, Makino N, Kuwabara M, et al. Incidence of Kawasaki Disease Before and After the COVID-19 Pandemic in Japan: Results of the 26th Nationwide Survey, 2019 to 2020. JAMA Pediatr. 2022;176(12):1217–1224. doi:10.1001/jamapediatrics.2022.3756
2. Arora, V., Tandon, R., Puri, R.D. et al. Hypomelanosis of Ito. Indian J Pediatr 89, 1117– 1119 (2022). https://doi.org/10.1007/s12098-022-04208-x
3. Bouaziz, M., Mullaert, J., Bigio, B. et al. Controlling for human population stratification in rare variant association studies. Sci Rep 11, 19015 (2021). https://doi.org/10.1038/s41598-021-98370-5
4. Burns JC, Herzog L, Fabri O, Tremoulet AH, Rodó X, Uehara R, Burgner D, Bainto E, Pierce D, Tyree M, Cayan D; Kawasaki Disease Global Climate Consortium. Seasonality of Kawasaki disease: a global perspective. PLoS One. 2013 Sep 18;8(9):e74529. doi: 10.1371/journal.pone.0074529.
5. Casanova JL. Severe infectious diseases of childhood as monogenic inborn errors of immunity. Proc Natl Acad Sci U S A. 2015 Dec 22;112(51):E7128-37. doi: 10.1073/pnas.1521651112
6. Chang LY, Lu CY, Shao PL, Lee PI, Lin MT, Fan TY, Cheng AL, Lee WL, Hu JJ, Yeh SJ, Chang CC, Chiang BL, Wu MH, Huang LM. Viral infections associated with Kawasaki disease. J Formos Med Assoc. 2014 Mar;113(3):148-54. doi: 10.1016/j.jfma.2013.12.008.
7. Chen, MH., Pitsillides, A. & Yang, Q. An evaluation of approaches for rare variant association analyses of binary traits in related samples. Sci Rep 11, 3145 (2021). https://doi.org/10.1038/s41598-021-82547-z

8. Danyel Lee et al. ,Inborn errors of OAS–RNase L in SARS-CoV-2–related multisystem inflammatory syndrome in children.Science379,eabo3627(2023).DOI:10.1126/science.abo3627
9. Dremel SE, Sivrich FL, Tucker JM, Glaunsinger BA, DeLuca NA. Manipulation of RNA polymerase III by Herpes Simplex Virus-1. Nat Commun. 2022 Feb 2;13(1):623. doi: 10.1038/s41467-022-28144-8
10. Holman RC, Curns AT, Belay ED, Steiner CA, Effler PV, Yorita KL, Miyamura J, Forbes S, Schonberger LB, Melish M. Kawasaki syndrome in Hawaii. Pediatr Infect Dis J. 2005 May;24(5):429-33. doi: 10.1097/01.inf.0000160946.05295.91.

    Induced Antiviral Protein 2′-5′-Oligoadenylate Synthetase." Molecular Cell, Volume 12, Issue 5, 2003, Pages 1173-1185, ISSN 1097-2765, https://doi.org/10.1016/S1097-
11. Kim GB, Park S, Eun LY, Han JW, Lee SY, Yoon KL, Yu JJ, Choi JW, Lee KY. Epidemiology and Clinical Features of Kawasaki Disease in South Korea, 2012-2014. Pediatr Infect Dis J. 2017 May;36(5):482-485. doi: 10.1097/INF.0000000000001474.
12. Kwon, Y. S., Park, S. H., Lee, H. R., Kim, J. H., Sohn, S., & Han, J. W. (2019). Reality of Kawasaki disease epidemiology. Korean Journal of Pediatrics, 62(8), 292–296. doi:10.3345/kjp.2019.00157
13. Lee S, Abecasis GR, Boehnke M, Lin X. Rare-variant association analysis: study designs and statistical tests. Am J Hum Genet. 2014 Jul 3;95(1):5-23. doi: 10.1016/j.ajhg.2014.06.009.
14. Lin MC, Lai MS, Jan SL, Fu YC. Epidemiologic features of Kawasaki disease in acute stages in Taiwan, 1997-2010: effect of different case definitions in claims data analysis. J Chin Med Assoc. 2015 Feb;78(2):121-6. doi: 10.1016/j.jcma.2014.03.009.
15. 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. doi:10.1111/ped.13544

16. Mbatchou, J., Barnard, L., Backman, J. et al. Computationally efficient whole-genome regression for quantitative and binary traits. Nat Genet 53, 1097–1103 (2021). https://doi.org/10.1038/s41588-021-00870-7
17. Morales, J., Pujar, S., Loveland, J.E. et al. A joint NCBI and EMBL-EBI transcript set for clinical genomics and research. Nature 604, 310–315 (2022). https://doi.org/10.1038/s41586-022-04558-8
18. Muller M, Demeret C. CCHCR1 interacts specifically with the E2 protein of human papillomavirus type 16 on a surface overlapping BRD4 binding. PLoS One. 2014 Mar 24;9(3):e92581. doi: 10.1371/journal.pone.0092581
19. Nakamura, A., Ikeda, K., & Hamaoka, K. (2019). Aetiological Significance of Infectious Stimuli in Kawasaki Disease. Frontiers in Pediatrics, 7, Article number 244. https://doi.org/10.3389/fped.2019.00244
20. Noval Rivas, M., & Arditi, M. (2020). Kawasaki disease: pathophysiology and insights from mouse models. Nature Reviews Rheumatology, 16, 391–405. https://doi.org/10.1038/s41584-020-0426-0
21. Onouchi Y. The genetics of Kawasaki disease. Int J Rheum Dis. 2018 Jan;21(1):26-30. doi: 10.1111/1756-185X.13218.
22. Onouchi, Y., Gunji, T., Burns, J. et al. ITPKC functional polymorphism associated with Kawasaki disease susceptibility and formation of coronary artery aneurysms. Nat Genet 40, 35–42 (2008). https://doi.org/10.1038/ng.2007.59
23. Pairo-Castineira, E., Clohisey, S., Klaric, L. et al. Genetic mechanisms of critical illness in COVID-19. Nature 591, 92–98 (2021). https://doi.org/10.1038/s41586-020-03065-y
24. Rife E, Gedalia A. Kawasaki Disease: an Update. Curr Rheumatol Rep. 2020 Sep 13;22(10):75. doi: 10.1007/s11926-020-00941-4
25. Rowley, A. H., & Shulman, S. T. (2018). The Epidemiology and

Pathogenesis of Kawasaki Disease. Frontiers in Pediatrics, 6, Article number 374. https://doi.org/10.3389/fped.2018.00374

26. Rowley, A.H. (2018), Is Kawasaki disease an infectious disorder?. Int J Rheum Dis, 21: 20-25. https://doi.org/10.1111/1756-185X.13213.
27. Rune Hartmann, Just Justesen, Saumendra N Sarkar, Ganes C Sen, Vivien C Yee. "Crystal Structure of the 2′-Specific and Double-Stranded RNA-Activated Interferon-
28. Saida, K., Chong, P.F., Yamaguchi, A. et al. Monogenic causes of pigmentary mosaicism. Hum Genet 141, 1771–1784 (2022). https://doi.org/10.1007/s00439-022- 02437-w
29. Tervaniemi, M.H., Katayama, S., Skoog, T. et al. Intracellular signalling pathways and cytoskeletal functions converge on the psoriasis candidate gene CCHCR1 expressed at P-bodies and centrosomes. BMC Genomics 19, 432 (2018). doi: https://doi.org/10.1186/s12864-018-4810-y
30. Tiala I, Suomela S, Huuhtanen J, Wakkinen J, Hölttä-Vuori M, Kainu K, Ranta S, Turpeinen U, Hämäläinen E, Jiao H, Karvonen SL, Ikonen E, Kere J, Saarialho-Kere U, Elomaa O. The CCHCR1 (HCR) gene is relevant for skin steroidogenesis and downregulated in cultured psoriatic keratinocytes. J Mol Med (Berl). 2007 Jun;85(6):589-601. doi: 10.1007/s00109-006-0155-0
31. Timothy J. Brown and others, CD8 T Lymphocytes and Macrophages Infiltrate Coronary Artery Aneurysms in Acute Kawasaki Disease, The Journal of Infectious Diseases, Volume 184, Issue 7, 1 October 2001, Pages 940–943, https://doi.org/10.1086/323155
32. Uehara R, Belay ED. Epidemiology of Kawasaki disease in Asia, Europe, and the United States. J Epidemiol. 2012;22(2):79-85. doi: 10.2188/jea.je20110131
33. Uehara R, Yashiro M, Nakamura Y, Yanagawa H. Kawasaki disease in parents and children. Acta Paediatr. 2003 Jun;92(6):694-7. doi: 10.1080/08035320310002768.

34. Vanessa Sancho-Shimizu, Petter Brodin, Aurélie Cobat, Catherine M. Biggs, Julie Toubiana, Carrie L. Lucas, Sarah E. Henrickson, Alexandre Belot, MIS-C@CHGE, Stuart G. Tangye, Joshua D. Milner, Michael Levin, Laurent Abel, Dusan Bogunovic, Jean-Laurent Casanova, Shen-Ying Zhang; SARS-CoV-2–related MIS-C: A key to the viral and genetic causes of Kawasaki disease?. J Exp Med 7 June 2021; 218 (6): e20210446. doi: https://doi.org/10.1084/jem.20210446
35. Zhang Y, Shen X, Pan W. Adjusting for population stratification in a fine scale with principal components and sequencing data. Genet Epidemiol. 2013 Dec;37(8):787-801. doi: 10.1002/gepi.21764

VII. ANNEXES

Annex 1. *Full list of missense and pLoF variants of RNASEL gene identified in cases and controls.*

RNASEL variants	ID	Variant type	CADD MSC=20.17	Gnomad frequency	Transcript	phenotype	Ethnicity
1:182545412_A/G	rs150721457	missense	22.7	2.48E-05	ENSP00000356530.3:p.Ile673Thr	control	European
1:182545431_G/A	rs374614472	missense	24.2	7.99E-06	ENSP00000356530.3:p.Arg667Trp	control	African
1:182545431_G/A	rs374614472	missense	24.2	7.99E-06	ENSP00000356530.3:p.Arg667Trp	control	African
1:182545452_C/T	.	missense	15.75	0	ENSP00000356530.3:p.Gly660Ser	control	Middle Eastern
1:182550367_G/A	.	missense	22.4	1.41E-05	ENSP00000356530.3:p.Thr633Met	control	North African
1:182550391_T/C	.	missense	0.001	0	ENSP00000356530.3:p.His625Arg	control	European
1:182550432_T/G	rs376567380	missense	0.004	2.48E-05	ENSP00000356530.3:p.Glu611Asp	control	African
1:182550432_T/G	rs376567380	missense	0.004	2.48E-05	ENSP00000356530.3:p.Glu611Asp	control	African
1:182550432_T/G	rs376567380	missense	0.004	2.48E-05	ENSP00000356530.3:p.Glu611Asp	control	African
1:182551267_G/A	rs200397730	missense	15.22	1.07E-05	ENSP00000356530.3:p.Arg565Cys	control	North African
1:182551267_G/A	rs200397730	missense	15.22	1.07E-05	ENSP00000356530.3:p.Arg565Cys	control	North African
1:182551270_G/A	.	missense	3.142	0	ENSP00000356530.3:p.His564Tyr	control	European
1:182551329_G/T	rs141087868	missense	9.913	8.56E-05	ENSP00000356530.3:p.Ala544Asp	control	African
1:182551366_C/G	rs193195484	missense	22.2	4.02E-06	ENSP00000356530.3:p.Val532Leu	case	Asian
1:182551366_C/G	rs193195484	missense	22.2	4.02E-06	ENSP00000356530.3:p.Val532Leu	case	Asian
1:182554473_T/C	.	missense	25.4	9.16E-05	ENSP00000356530.3:p.Asn490Ser	control	Asian
1:182554473_T/C	.	missense	25.4	9.16E-05	ENSP00000356530.3:p.Asn490Ser	control	Asian
1:182554492_G/A	.	stop_gained	39	0	ENSP00000356530.3:p.Gln484Ter	control	American
1:182554500_T/TTGCAG	.	frameshift	26.1	2.39E-05	ENSP00000356530.3:p.Tyr481SerfsTer14	control	European
1:182554504_C/T	.	missense	22.8	0	ENSP00000356530.3:p.Gly480Arg	control	American
1:182554650_A/G	.	missense	20.5	0	ENSP00000356530.3:p.Phe431Ser	control	European
1:182554668_C/T	.	missense	11.8	0	ENSP00000356530.3:p.Ser425Asn	control	North African
1:182554744_G/A	rs377048179	missense	13.39	2.39E-05	ENSP00000356530.3:p.Arg400Cys	control	European
1:182554744_G/A	rs377048179	missense	13.39	2.39E-05	ENSP00000356530.3:p.Arg400Cys	control	Middle Eastern
1:182554753_C/T	.	missense	15.95	0	ENSP00000356530.3:p.Gly397Ser	control	European
1:182554764_G/A	.	missense	10.7	1.20E-05	ENSP00000356530.3:p.Thr393Met	control	European
1:182554778_T/G	.	missense	20.3	1.20E-05	ENSP00000356530.3:p.Glu388Asp	control	American
1:182554871_C/G	.	missense	19.03	3.99E-06	ENSP00000356530.3:p.Met357Ile	control	European
1:182554879_G/A	rs148464936	missense	25.8	3.55E-05	ENSP00000356530.3:p.Arg355Cys	control	European
1:182554932_T/A	.	missense	9.488	3.98E-06	ENSP00000356530.3:p.Lys337Met	control	European
1:182554995_A/G	.	missense	22.8	0	ENSP00000356530.3:p.Leu316Pro	control	American
1:182554998_G/C	.	missense	0.006	0	ENSP00000356530.3:p.Ser315Cys	control	African
1:182554998_G/A	.	missense	0.004	0	ENSP00000356530.3:p.Ser315Phe	control	North African
1:182555337_C/T	.	missense	22.7	6.77E-05	ENSP00000356530.3:p.Gly202Asp	control	Middle Eastern
1:182555340_A/G	.	missense	7.8	3.98E-06	ENSP00000356530.3:p.Met201Thr	control	European
1:182555353_C/G	.	missense	18.87	3.98E-06	ENSP00000356530.3:p.Ala197Pro	control	European
1:182555508_T/C	.	missense	16.61	3.18E-05	ENSP00000356530.3:p.Tyr145Cys	control	European
1:182555508_T/C	.	missense	16.61	3.18E-05	ENSP00000356530.3:p.Tyr145Cys	control	Middle Eastern

1:182555639_A/C	.	missense	6.285	0	ENSP00000356530.3:p.Ile101Met	control	European
1:182555658_G/C	.	missense	23.3	0	ENSP00000356530.3:p.Pro95Arg	control	European
1:182555688_G/A	.	missense	23.1	4.65E-05	ENSP00000356530.3:p.Pro85Leu	control	North African
1:182555688_G/A	.	missense	23.1	4.65E-05	ENSP00000356530.3:p.Pro85Leu	control	North African
1:182555721_A/G	.	missense	22	1.79E-05	ENSP00000356530.3:p.Ile74Thr	control	African
1:182555836_C/T	.	missense	11.93	3.98E-06	ENSP00000356530.3:p.Glu36Lys	control	North African
1:182555881_C/T	.	missense	0.905	1.99E-05	ENSP00000356530.3:p.Ala21Thr	control	American
1:182555917_G/A	.	missense	2.026	0	ENSP00000356530.3:p.Pro9Ser	control	European
1:182555918_G/T	.	missense	4.613	0	ENSP00000356530.3:p.Asn8Lys	control	European

Annex 2. *Full list of missense and pLoF variants of OAS1 gene identified in cases and controls.*

OAS1 variants	ID	Variant type	CADD MSC=18.41	Gnomad frequency	Transcript	phenotype	Ethnicity
12:113344940_T/A	rs371488150	missense_variant	0.255	2.12E-05	ENSP00000202917.5:p.His32Gln	control	European
12:113344965_C/G	.	missense_variant	22.7	1.77E-05	ENSP00000202917.5:p.Leu41Val	control	European
12:113344984_G/T	.	missense_variant	0.608	0	ENSP00000202917.5:p.Arg47Leu	control	North African
12:113345007_G/A	rs138921278	missense_variant	23.5	7.43E-05	ENSP00000202917.5:p.Val55Met	control	North African
12:113345007_G/A	rs138921278	missense_variant	23.5	7.43E-05	ENSP00000202917.5:p.Val55Met	control	North African
12:113346347_T/C	.	missense_variant	24	1.63E-05	ENSP00000202917.5:p.Ser63Pro	control	North African
12:113346347_T/C	.	missense_variant	24	1.63E-05	ENSP00000202917.5:p.Ser63Pro	control	North African
12:113346366_C/T	rs142847241	missense_variant	18.78	8.14E-05	ENSP00000202917.5:p.Thr69Ile	control	African
12:113346392_C/G	.	missense_variant	23.2	0	ENSP00000202917.5:p.Leu78Val	control	European
12:113346483_C/T	.	missense_variant	17.87	1.99E-05	ENSP00000202917.5:p.Ala108Val	control	European
12:113346516_T/C	.	missense_variant	23.5	0	ENSP00000202917.5:p.Phe119Ser	control	European
12:113346539_G/A	.	missense_variant	2.592	0	ENSP00000202917.5:p.Gly127Ser	control	European
12:113346552_C/T	.	missense_variant	0.047	3.99E-05	ENSP00000202917.5:p.Ala131Val	control	European
12:113346563_G/A	rs111902215	missense_variant	11.45	6.76E-05	ENSP00000202917.5:p.Val135Ile	case	Asian
12:113346563_G/A	rs111902215	missense_variant	11.45	6.76E-05	ENSP00000202917.5:p.Val135Ile	control	Asian
12:113346563_G/T	rs111902215	missense_variant	7.854	3.18E-05	ENSP00000202917.5:p.Val135Leu	control	African
12:113346563_G/T	rs111902215	missense_variant	7.854	3.18E-05	ENSP00000202917.5:p.Val135Leu	control	African
12:113348859_A/G	.	missense_variant	19.63	0	ENSP00000202917.5:p.Gln158Arg	control	European
12:113348882_A/G	.	missense_variant	0.001	1.59E-05	ENSP00000202917.5:p.Asn166Asp	control	European
12:113354444_T/A	.	missense_variant	23	0	ENSP00000202917.5:p.Val262Asp	control	American
12:113354465_G/A	rs138105298	missense_variant	22.7	4.60E-05	ENSP00000202917.5:p.Cys269Tyr	control	European
12:113354491_T/C	.	missense_variant	16.89	0	ENSP00000202917.5:p.Phe278Leu	control	European

12:113355367_C/A	.	missense_variant	23.5	0	ENSP00000202917.5:p.Asp300Glu	control	American
12:113357197_G/T	.	stop_gained	0.542	1.07E-05	ENSP00000202917.5:p.Glu348Ter	control	North African
12:113357197_G/T	.	stop_gained	0.542	1.07E-05	ENSP00000202917.5:p.Glu348Ter	control	North African
12:113357311_G/A	.	missense_variant	9.819	3.18E-05	ENSP00000202917.5:p.Ala386Thr	control	European

Annex 3. *Full list of missense and pLoF variants of OAS2 gene identified in cases and controls.*

OAS2 variants	ID	Variant type	Transcript	CADD	Gnomad frequency	phenotype	ethnicity
12:113416426_G/A	.	missense_variant	ENSP00000376362.2:p.Glu5Lys	2.085	0	control	North African
12:113416480_T/C	.	missense_variant	ENSP00000376362.2:p.Tyr23His	0.001	0	control	European
12:113416550_A/G	rs200545644	missense_variant	ENSP00000376362.2:p.Gln46Arg	7.599	4.38E-05	control	European
12:113424892_C/T	.	missense_variant	ENSP00000376362.2:p.Thr76Ile	8.822	0	control	European
12:113424935_G/T	.	missense_variant	ENSP00000376362.2:p.Gln90His	23.2	0	control	Asian
12:113424939_A/G	.	missense_variant	ENSP00000376362.2:p.Arg92Gly	0.241	2.12E-05	control	European
12:113424946_A/C	.	missense_variant	ENSP00000376362.2:p.Gln94Pro	17.18	1.59E-05	control	European
12:113424949_G/A	.	missense_variant	ENSP00000376362.2:p.Arg95His	0.018	1.06E-05	control	European
12:113424960_G/A	.	missense_variant	ENSP00000376362.2:p.Asp99Asn	0.237	1.99E-05	control	North African
12:113424961_A/G	.	missense_variant	ENSP00000376362.2:p.Asp99Gly	0.646	0	control	European
12:113424967_CTG/C	.	frameshift_variant	ENSP00000376362.2:p.Asp103Ter	16.38	3.19E-05	control	African
12:113435383_C/T	rs200589437	missense_variant	ENSP00000376362.2:p.Thr229Met	21.9	3.18E-05	control	European
12:113435383_C/T	rs200589437	missense_variant	ENSP00000376362.2:p.Thr229Met	21.9	3.18E-05	control	European
12:113436078_A/G	.	missense_variant	ENSP00000376362.2:p.Ile291Val	20.2	4.00E-06	control	European
12:113436093_G/A	.	missense_variant	ENSP00000376362.2:p.Asp296Asn	23.4	0	control	African
12:113436165_A/T	.	missense_variant	ENSP00000376362.2:p.Thr320Ser	0.047	0	control	North African
12:113436165_A/T	.	missense_variant	ENSP00000376362.2:p.Thr320Ser	0.047	0	control	North African
12:113440843_T/C	.	missense_variant	ENSP00000376362.2:p.Ile372Thr	8.225	0	control	African
12:113440848_C/T	rs142826885	missense_variant	ENSP00000376362.2:p.Arg374Cys	2.518	5.30E-05	control	American
12:113440848_C/T	rs142826885	missense_variant	ENSP00000376362.2:p.Arg374Cys	2.518	5.30E-05	control	American
12:113442773_CTG/C	.	frameshift_variant	ENSP00000376362.2:p.Gly406LeufsTer2	22.9	0.000244	control	African
12:113442773_CTG/C	.	frameshift_variant	ENSP00000376362.2:p.Gly406LeufsTer2	22.9	0.000244	control	African
12:113442773_CTG/C	.	frameshift_variant	ENSP00000376362.2:p.Gly406LeufsTer2	22.9	0.000244	control	African
12:113442796_G/T	.	missense_variant	ENSP00000376362.2:p.Val413Leu	13.94	7.96E-05	control	American
12:113442877_A/G	rs144830420	missense_variant	ENSP00000376362.2:p.Lys440Glu	0.356	1.06E-05	control	African
12:113442881_C/T	.	missense_variant	ENSP00000376362.2:p.Ala441Val	6.414	2.39E-05	control	European
12:113442890_G/A	rs371992121	missense_variant	ENSP00000376362.2:p.Arg444Lys	0.006	3.90E-05	control	African

12:113442890_G/A	rs371992121	missense_variant	ENSP00000376362.2:p.Arg444Lys	0.006	3.90E-05	control	African
12:113444334_C/T	rs142538466	stop_gained	ENSP00000376362.2:p.Arg529Ter	34	1.99E-05	control	North African
12:113444347_G/A	.	missense_variant	ENSP00000376362.2:p.Arg533His	0.056	1.99E-05	control	North African
12:113444347_G/A	.	missense_variant	ENSP00000376362.2:p.Arg533His	0.056	1.99E-05	control	North African
12:113444402_A/T	.	missense_variant	ENSP00000376362.2:p.Lys551Asn	20.7	4.05E-06	control	Middle Eastern
12:113445525_A/G	rs201568741	missense_variant	ENSP00000376362.2:p.Lys558Glu	22.3	1.59E-05	case	North African
12:113445541_T/G	.	missense_variant	ENSP00000376362.2:p.Leu563Trp	24.3	7.96E-06	control	American
12:113445558_T/G	.	missense_variant	ENSP00000376362.2:p.Leu569Val	22.7	3.98E-06	control	European
12:113445565_T/C	.	missense_variant	ENSP00000376362.2:p.Leu571Pro	24.8	1.19E-05	control	American
12:113445580_C/T	rs146756036	missense_variant	ENSP00000376362.2:p.Ala576Val	25	7.78E-05	control	African
12:113445592_G/A	rs373092983	missense_variant	ENSP00000376362.2:p.Gly580Glu	22.4	3.98E-06	control	North African
12:113445631_G/A	.	missense_variant	ENSP00000376362.2:p.Arg593Gln	22.5	2.39E-05	control	North African
12:113446966_C/G	.	missense_variant	ENSP00000376362.2:p.Ala657Gly	23.7	1.41E-05	control	European
12:113446983_T/C	.	missense_variant	ENSP00000376362.2:p.Trp663Arg	23.3	3.98E-06	control	European
12:113446983_T/C	.	missense_variant	ENSP00000376362.2:p.Trp663Arg	23.3	3.98E-06	control	European
12:113447023_C/A	.	missense_variant	ENSP00000376362.2:p.Pro676Gln	19.26	0	control	European
12:113447044_C/T	rs147522268	missense_variant	ENSP00000376362.2:p.Pro683Leu	23.5	4.61E-05	control	European