Genetic landscape of rare autoinflammatory disease variants in Qatar and Middle Eastern populations through the integration of genomic datasets

Parul Sharma[$, 1], Abhinav Jain[$,2,3], Vinod Scaria[1,2,3]

[1]Center for Computational Biology, Indraprastha Institute of Information Technology Delhi (IIIT-Delhi), Okhla Industrial Estate, Phase III, New Delhi 110020, India
[2]CSIR-Institute of Genomics and Integrative Biology, Mathura Road, Delhi 110025, India
[3]Academy of Scientific and Innovative Research (AcSIR), Mathura Road, Delhi 110025, India

$ Contributed equally and would like to be known as joint first authors.
** Address for correspondence: vinods@igib.in*

Abstract

Rare autoinflammatory diseases are a group of recurrent inflammatory genetic disorders caused due to genetic mutations in over 37 genes and typically have a recessive mode of inheritance. While a number of these disorders have been identified and reported from the Middle Eastern populations, the carrier frequency of these genetic mutations in the Middle Eastern populations is not known. The availability of whole-genome and exome datasets of over a thousand individuals from Qatar persuaded us to explore the genetic epidemiology of rare autoinflammatory genetic variants in the Middle East. We have systematically analyzed genetic variants in genome-scale datasets from Qatar with a compendium of variants associated with autoinflammatory diseases. The variants were further systematically reclassified according to the American College of Medical Genetics and Genomics (ACMG & AMP) guidelines for interpretation of pathogenicity of sequence variants. Our analysis identified 7 pathogenic and likely pathogenic variants with a significant difference in their allele frequencies compared to the global population. The cumulative carrier frequency of these variants was 2.58%. Furthermore, our analysis revealed that 5 genes implicated in rare autoinflammatory diseases were under natural selection. To our best knowledge, this is the first and comprehensive study on the population-scale analysis and genetic epidemiology for genetic variants causing rare autoinflammatory disease in any Middle Eastern population.

Keywords: *Autoinflammatory disease, Qatar, Middle East, genome, Epidemiology*

Introduction

Autoinflammatory diseases are a group of Mendelian genetic disorders characterized by recurrent inflammatory episodes due to an abnormal innate immune system. Until recently, these disorders were only defined by phenotypic traits including recurrent attacks of fever, abdominal pain, arthritis, skin rashes or cutaneous signs, which coincide with other diseases and therefore leads to the inaccurate or difficult diagnosis [1,2]. Recent advancements in the understanding of the molecular basis of these disorders have resulted in their accurate classification presently encompassing more than 31 disorders involving over 30 genes [3–54,6–8]. Several studies in recent years have established a molecular basis for the diagnosis of autoinflammatory diseases] [9,10].

Genetic testing is widely available for heritable monogenic disorders including Familial Mediterranean Fever (FMF), TNF receptor-associated periodic syndrome (TRAPS) and Pyogenic Sterile Arthritis, Pyoderma Gangrenosum, and Acne Syndrome (PAPA). Molecular testing, however, has poor diagnostic value for multifactorial conditions but can help in the prognosis of the disease. The most common among all autoinflammatory disease is Familial Mediterranean Fever (FMF). It is caused by the mutation in the *MEFV* gene and is highly prevalent in the Middle Eastern countries, where it is considered to be an important clinical and public health problem [11,12]. Due to the high level of consanguinity, which is accepted as a natural practice in these countries, the occurrence of rare and recessive alleles becomes prominent. FMF is found to be very common in Armenians, Israel, North Africa, Turkey, Lebanese and Iraq [13–16]. It is also present in high numbers in other countries like Syria, Egypt, Kuwait, and Saudi Arabia. Previously, we studied the population-specific frequencies of the variants found in the *MEFV* gene responsible for causing FMF. The study revealed a comprehensive report on the genetic epidemiology of the disease among the 2000 individual samples from the Mediterranean region [17]. Other autoinflammatory disorders including Majeed Syndrome are also prevalent in various Arab countries [18].

There are multiple reports of autoinflammatory disorders from Middle Eastern countries including reports from a single hospital in Riyadh where thirty-four patients were admitted due to autoinflammatory disorders in 10 years [19]. Also, there are multiple individuals from Arab countries affected by Majeed syndrome, the rare autoinflammatory condition characterized by chronic recurrent multifocal osteomyelitis (CRMO) and congenital dyserythropoietic anemia (CDA) [19–21]. A retrospective study by Aladbe et al. in Qatari children for 5 years using medical records found

around 70 symptomatic and carrier individuals and reported its expansion in Qatar [22]. Epidemiology survey of Behçet's Syndrome found that the highest prevalence is in Turkey and is as high as 421 in 100,000 adult population [23]. Also, in the international registry of autoinflammatory disorders, eurofever has the second largest number of patients (16% of all individuals) from eastern and southern Mediterranean region which includes Turkey, Israel, and North Africa [24]. Incidence reports from Sweden on autoinflammatory disorders and hereditary amyloidosis found 98% of individuals from the Eastern Mediterranean region mainly young Syrian descendants [25]. Another report, where Lebanese patients were affected by deficiency of IL-1 receptor antagonist (DIRA), could be possible founder *IL1RN* mutation in Lebanon population [16,25]. There is another report of Palestinian patient affected by deficiency of interleukin-36 (IL-36) receptor antagonist (DITRA) caused due to stop gain mutation in *IL36RN* [26]. Due to highly consanguineous marriages in Arab, it gives great opportunity to find novel risk loci like *LACC1* and *LRBA* which were found to be associated with Crohn's disease and inflammatory bowel disease with combined immunodeficiency respectively in Arab population [16,27,28].

The recent availability of genome-scale datasets of many populations not previously covered in global sequencing projects provides a unique opportunity to study genetic epidemiology of diseases which are prevalent in such populations. The recent availability of data from Middle East populations has made it possible to understand the genetic epidemiology of autoinflammatory diseases [29,30]. In the present study, we have used an integrative analysis of extensive data mining to formulate a comprehensive database of known variations prone to cause Autoinflammatory disorders. We have further attempted to understand the prevalence of these disorders by studying the allele and genotype frequencies of the variants in the Qatar population. Our analysis points to significant differences in the allele frequencies even for the small population studied.

Results

Genetic variants in the autoinflammatory genes from different database

We retrieved pathogenic or likely pathogenic genetic variants which encompassed a total of 72,331 variants from ClinVar (v. date 2018-09-30) database. These variants were further filtered to the list of 37 genes associated with autoinflammatory diseases as detailed in Table 1 which encompassed a total of 270 variants. These variants were further mapped on Qatar dataset and retrieved 5 variants. In a similar way, Infevers database had a total of 29 rare autoinflammatory

genes excluding MEFV which compose of 1,254 genetic variants. These genetic variants from the Infevers database were mapped on the Qatar dataset and this resulted in a total of 95 overlapping variants. Similarly, for HGMD, a total of 1,209 genetic variants exist in 37 rare autoinflammatory genes. These variants on manual overlapped with Qatar dataset retrieved 72 genetic variants.

ClinVar, infevers, and HGMD had various overlapping variants which include, one variant which was common to all the databases, three variants were common to ClinVar and HGMD, 22 variants were common to HGMD and infevers whereas there was no common variant between ClinVar and Infever. There were also variants which were unique to the database as one variant in ClinVar, 46 in HGMD and 72 in infevers. So, we had a total of 145 unique variants in Qatar which overlapped with the databases. Venn diagram for variant numbers and databases were represented in Fig. 2. The information for the total number of variants from each database and genes had been tabulated in Supplementary Table S1.

Variant Annotation based on ACMG and AMP classification

These 145 variants were annotated using 28 stringent criteria put forward by ACMG and AMP. Each of the variants was manually curated and categorized, based on the criteria detailed in Supplementary data S1. According to ACMG guidelines we have annotated and found single pathogenic and six likely pathogenic variants, while 92 variants were benign and likely benign and 42 were VUS and VUS with conflicting evidence. The pathogenic and likely pathogenic variants were tabulated in Table 2. The detailed variants annotation were in Supplementary Table S2.

Comparison of Allele Frequencies with the world population

Allele frequencies of the seven pathogenic or likely pathogenic variants were compared with Global datasets of overall control population i.e. 1000 genome (1000g2015aug_all), ExAC (ExAC_all) and gnomAD (gnomAD_all). We implemented Fisher's exact test with a *p*-value less than 0.05 (Table 3) and found four out of seven variants were significantly different in genotype frequencies . A likely pathogenic variant p.A125V rs5029941 in *TNFAIP3* hada significantly lower frequency (p-value < 0.05, Fisher's Exact test) in comparison to the global frequencies from the 1000 Genome . Its allele frequency in Qatar population was 0.16% and in GME, it was 0.05%, whereas allele frequency is quite high in i.e. 0.6% in 1000 Genome Project. Another likely pathogenic variant p.V377I rs28934897 in *MVK* was also significantly higher in Qatar Arab subpopulation compared to ExAC The allele frequency was 0.8% in Qatar Arab population

whereas it was absent in 1000 Genome, 0.14% in ExACI and 0.16% in gnomAD. Interestingly, this variant was also significantly higher in GME subpopulation from Arabian Peninsula (AP) with an allele frequency of 0.87% in comparison to ExAC. A pathogenic variant p.R82C rs753966933 in *RAB27A* was significantly higher in Qatar Arab subpopulation in comparison to the ExAC. The allele frequency in Qatar Arab subpopulation was 2.7%, whereas in global population i.e. in 1000 Genomewas absent and in ExAC and gnomAD it has very low allele frequency i.e. 8×10^{-4} % and 1.6×10^{-3} % respectively and was absent in GME dataset. Another likely pathogenic variant p.R352C in NLRP12 rs199881207 was significantly higher in two Qatar subpopulations i.e. European and South Asian whose allele frequency was 12.5% and 0.8% respectively in comparison to ExAC. This variant was absent in the 1000 genome and had very low allele frequencies in GME, ExAC and gnomAD i.e. 0.05%, 0.04% and 0.038% respectively. Another likely pathogenic variant p.N852S was significantly higher in GME population especially in Northeast African subpopulation in comparison to the 1000 genome. Its allele frequency in GME and its subpopulation Northeast Africa was 0.3% and 0.5% respectively whereas in 1000 genome, ExAC, and gnomADit was 0.04%, 0.12%, and 0.10% respectively. Comparison of allele frequency of Qatar dataset and its subpopulations with GME and global population datasets are well represented in Fig. 3. The comparison of allele frequency of Qatar population among the global population (1000g2015aug_all, ExAC_all, and gnomAD_al) and GME has been detailed in Supplementary Table S3.

Genes with signals of natural selection

We explored signals for natural selection in the rare autoinflammatory genes. We calculated iHS and Fst score for natural selection prediction for Q1005. iHS predicted selection based on variant on the long haplotype and Fst predicted it by comparing allele frequency of Qatar and its subpopulation with 1000 Genome Project and its subpopulation (African, European and South Asian). We took 1% of iHS scored variants and found that 5 rare autoinflammatory genes fall into top 1% naturally selected variants. These 5 genes were IL1RN, IL36RN, NLRP3, PSMB9, and RAB27A. Fig. 4 represented iHS and Fst plot for a natural selection of the RAB27A gene, we chose RAB27A gene as one variant in RAB27A had been annotated as likely pathogenic and had very high carrier allele frequency especially in Arab subpopulation in comparison to global control population.

Discussion

Autoinflammatory diseases are a group of Mendelian disorders which are caused due to abnormal development and maturation of innate immune cells which leads to unprovoked recurrent fever with rashes, systemic inflammation, pain in the chest, arthritis, and serositis [13,44]. The spectrum of diseases encompasses over 31 distinct autoinflammatory diseases caused by mutations in 30 autoinflammatory genes as characterized by an expert panel of Inborn Errors of Immunity Committee and Infevers [4,324,6–84,32]. A large number of cases of clinically and molecularly characterized autoinflammatory disorders has been reported in the literature, with the positivity of less than 10% that includes variants of uncertain significance (VUS) [45]. These significant numbers of autoinflammatory conditions have also been reported in the Middle Eastern population but genetic epidemiology of autoinflammatory disorders is not yet known [45–48]. The available whole genome and whole exome sequence datasets of Middle Eastern populations were used to understand the genetic epidemiology of autoinflammatory disorders in the population [29,30].

In the present study, we compiled a comprehensive list of genes and genetic variants from a number of resources. A total of 37 genes associated with 35 autoinflammatory conditions comprise data from Infever database, expert panel of Inborn Errors of Immunity Committee and manually curated literature which formed the basis of the analysis [4,32,334,6–84,32,33]. We did not consider Familial Mediterranean Fever and *MEFV* genes for the analysis since earlier work from our group comprehensively analyzed the genetic epidemiology of the genetic variants for this gene [17]. The allele frequencies of a compendium of variants were checked in the whole genome/exome dataset from Qatar, encompassing 917 exomes and 88 genomes to reveal a total of 145 unique variants. The annotation of pathogenicity of variants was systematically performed as per the ACMG/AMP guidelines for the interpretation of sequence variants. Systematic filtering and annotation by ACMG/AMP guidelines resulted in 7 pathogenic and likely pathogenic variants. The final list was composed of 6 nonsynonymous mutations and 1 stop gain mutation in 5 genes(one in *MVK*, one in *RAB27A*, one in *NOD2*, one in *AP1S3*, one in *TNFAIP3*, and two in *NLRP12*).

The nonsynonymous mutation in *AP1S3* i.e. Phe4Cys (p.F4C) rs116107386 annotated as likely pathogenic by ACMG & AMP guidelines had been functionally validated by Setta-Kaffetzi et al. using transfection studies in HEK293 cells and HaCaT keratinocytes which showed the decreased

activity of the mutant protein [49]. The F4C variant was located in the functionally important domain of AP1S3 and it had been predicted as deleterious by all three computational software (SIFT, PolyPhen-2, and CADD). Another nonsynonymous likely pathogenic variants in *TNFAIP3* i.e. Ala125Val (p.A125V) leads to inflammatory bowel disease (IBD). Lodolce and group transfected HEK293 cell lines with wild type and mutant allele and found a decrease in enzymatic activity of *TNFAIP3* to deubiquitinase the target protein TRAF2 in mutant allele [50]. This mutation also falls in a functionally important domain, which affects the functionality of the protein. Third nonsynonymous likely pathogenic mutation in *MVK* i.e. V377I, V325I had been reported in various studies as a cause of Hyperimmunoglobulinemia D with Periodic Fever Syndrome (HIDS) and Mevalonic aciduria (MA). Both of these diseases were rare and had an autosomal recessive mode of inheritance. This variant was popularly known as the Dutch mutation with a carrier frequency of 1:65 individuals [51]. A number of studies have shown this variant was present in homozygous or compound heterozygous states in multiple patients affected by HIDS with different ethnic backgrounds. Various in-vitro and in-vivo functional studies have been proved p.V377I as disease-causing with very low allele frequency in control databases [52–56]. Fourth nonsynonymous likely pathogenic mutation was in *RAB27A* i.e. Arg82Cys (R82C) had been predicted as a causal variant for hemophagocytic lymphohistiocytosis (HLH). This variant was absent in the healthy population i.e. in 1000 genome and esp6500 and present in very low frequency in ExAC and gnomAD i.e. 8.3×10^{-6} and 1.63×10^{-5}. It had been located in a well established functional domain of the *RAB27A*. In 2016 Netter et al. also reported family members were severely affected by homozygous p.R82C variant and less affected by heterozygous p.R82C [57]. The functional assay also showed the significant effect of p.R82C on RAB27A binding activity to melanophilin and Munc13-4 [58]. Fifth nonsynonymous likely pathogenic mutation was in *NOD2* i.e. Asn852Ser (p.N852S) which were associated with Blau syndrome, Crohn's disease, and Yao syndrome. These syndromes were associated with abnormal inflammation of the body. Horowitz *et al.* in 2017 identified compound heterozygous variants in two affected individuals. The first individual was compound heterozygous for p.G908R and p.N852S and in second was for novel mutation p.S506Vfs*73 and p.G908R [59]. The functional assessment had shown that p.N852S impaired NOD2 ligand muramyl dipeptide(MDP)-induced NF-κB activation could lead to Crohn's disease [60]. p.N852S had been co-segregated within the Ashkenazi Jewish family with Crohn's disease. It had been reported that 15% of Crohn patients of Ashkenazi Jewish ancestry had the p.N852S variant [61]. Sixth nonsynonymous likely pathogenic variant from *NLRP12* i.e. Arg352Cys (p.R352C) found to be associated with NLRP12-Associated Periodic Fever Syndrome. Jeru and group in 2011

performed an in-vivo experiment in which they transfected the HEK293 cell line with wild type plasmid and mutant R352C plasmid. They showed that mutation enhanced the signaling activity of procaspase 1 which leads to abnormal inflammation and causes NLRP12-Associated Periodic Fever Syndrome [62]. The R352C variant also falls in the functionally important domain of NLRP12. The seventh stop-gain pathogenic mutation which falls in NLRP12 i.e. p.R248X which causes NLRP12-Associated Periodic Fever Syndrome. This stop gain mutation leads to aberrant termination of protein synthesis which disrupts the integrity of protein structure and ultimately its function. NLRP12 is responsible to inhibit the inflammatory response but due to stop gain mutation, it was not regulated which caused abnormal inflammation and periodic fever.

The allele frequencies of 5 out of 7 variants classified as or likely pathogenic were significantly different in Qatari and GME population datasets in comparison with global populations. The allele frequencies of pathogenic or likely pathogenic variants showed concordance with the previous study in this region. In 2005, Mohammed Hammoudeh reported for the first time an Arab child been affected by HIDS, with the homozygous variant p.V377I in MVK gene [63]. Moussa et al. in 2015 also reported p.V377I segregation with HIDS in an Arabic family where 2 siblings were homozygous p.V377I and affected [64]. Also, there were various reports for a patient suffering from HIDS from the Middle East and nearby region which include Turkey, Armenia, Kuwait, Israel, Palestine [65–68]. In the Qatar population dataset, overall carrier allele frequency was 0.32% while the Arab subpopulation had a very high carrier allele frequency of 0.82%. Similarly, in the GME population dataset, carrier allele frequency was0.35% while in the African Pygmy subpopulation, a high carrier allele frequency of 0.87% was observed. Another pathogenic mutation p.R82C in RAB27A had very high carrier allele frequency in Qatari Arab subpopulation i.e. 2.7% while the allele frequency was 0.5% in the Qatari population dataset. A large study conducted by Tukel et al. in 2004 to find the frequency of Crohn disease-associated Ashkenazi, Sephardi, and Oriental Jewish Families in middle eastern countries which include Ashkenazi, Sephardi and Oriental Jewish from Israel, Morocco, Turkey, Tunisia, Iraq, Kurds, Iran, Yemen, and Syria. They found 15% of Ashkenazi Jews suffering from Crohn's disease had p.N852S mutation [61]. Another study which involves p.N852S frequency in Turkish population, but could not find N852S association with Crohn's disease [69]. In our study we found p.N852S had high carrier allele frequency in GME subpopulations as 0.8% in Syrian Desert, 0.5% in North-East Africa and 0.3% in Turkish Peninsula and Qatar allele frequency was 0.1% and its subpopulation Bedouin was 0.2%.

The distinct frequencies of many variants prompted us to explore whether any of these genes showed signals of natural selection. Our analysis revealed a significant signal as revealed by high iHS scores in the *RAB27A* and *IL1RN* genes RAB27A protein mainly involved in the transportation of melanin for skin pigmentation. It also plays an important role in cytotoxic T lymphocytes for killing or neutralizing of foreign invaders. *RAB27A* had been evolutionarily selected among different species, as editing levels of *RAB27A* showed the highest difference among humans and rhesus monkey [70]. Similarly, *IL1RN* identified to be under selection in the European-American population [71,72]. Whereas *IL36RN* was found to be evolutionary conserved among different species [73]. Other genes like NLRP3 and PSMB9 did not show any strong signals of natural selection or conservation.

In summary, our analysis integrating population-scale genomic datasets provide the first comprehensive insights into the genetic epidemiology of autoinflammatory diseases in the Middle Eastern populations. Our analysis suggests distinct allele frequencies in subpopulations and also suggests at least two genes show strong signals of natural selection in the middle eastern population. As more population-scale genomic datasets from the region become available, including from national initiatives like the Bahrain genome program https://www.moh.gov.bh/GenomeProject/ and the Emirati genome initiative apart from the ongoing Qatar Genome project, would provide a higher resolution towards understanding frequencies of rarer genetic variants [74].

Methods

Population-scale datasets of genetic variants:

The analysis was performed on population-scale datasets which involves 1005 Qatar Genomes and Exomes [29] and 1,111 Greater Middle East (GME) exomes [30]. The Qatar dataset encompassed 1005 unrelated individuals data, which comprised a total of 88 whole genomes and 917 whole exomes sequences and a total of 20,937,965 genetic variants were mapped to hg19/GRCh37 reference human genome. The individuals sequenced belonged to seven different sub-populations within Qatar namely: European (EUR, n=5), South Asian (SA, n=82), Bedouin (BED, n=566), African Pygmy (AP, n=1), Arab (ARB, n=236), Persian (PER, n=194), and Sub Saharan African (SAF, n=77) [29]. The GME Variome Project comprised of high quality exome data of 1,111 unrelated individual from six different regions of GME which includes Northwest Africa (NWA, n=85),

Northeast Africa (NEA, n=423), Turkish Peninsula (TP, n=140), Syrian Desert (SD, n=81), Arabian Peninsula (AP, n=214), and Persia and Pakistan (PP, n=168). It consists of total 689,297 genetic variants mapped to the hg19/GRCh37 reference genome [30].

Datasets of disease-associated genetic variants

Three independent datasets of genetic variants were compiled for the analysis. This included ClinVar, Infevers and the Human Gene Mutation Database (HGMD).

Infevers is a publicly available database of genetic mutations in autoinflammatory disorders. The database lists a total of 1,586 genetic variants for 30 genes associated with 31 autoinflammatory disorders [4,6–8]. The complete list along with the disease classification is provided on the website https://infevers.umai-montpellier.fr/. The rare autoinflammatory variants for 29 genes excluding *MEFV* as it has already been analyzed by our group [6,17] were downloaded and compiled. These variants were further mapped on Qatar variants.

The ClinVar database [ClinVar version: 2018-02-25] variants which were annotated as pathogenic or likely pathogenic were retrieved [31]. These variants were further filtered for a list of 37 rare autoinflammatory genes which compose 29 genes from Infever database [4,6–8] and 8 genes were curated from the literature and other sources [32,33]. Further, the remaining variants were mapped with the Qatari variants.

The Human Gene Mutation Database (HGMD) is a comprehensive resource and a collection of genetic variants in human genes [6,17,34]. The public version of the database lists over 141,000 mutations in over 5,700 genes. The data for all 37 autoinflammatory genes were manually compiled from the public version of the HGMD (http://www.hgmd.cf.ac.uk/ac/index.php) and mapped it over the Qatar variants.

Computational Annotation protocol for variants

All the autoinflammatory variants from ClinVar, HGMD, and Infevers mapped to Qatar dataset were further annotated by using software ANNOVAR for their functional annotation which consisted of annotation from various databases [35]. These databases include RefGene, dbnsfp33a, dbscsnv11, avsnp147, intervar_20170202, kaviar_20150923 and GWAS catalog for prediction of variants pathogenicity and other characteristics. There are databases for allele frequency across

the global population which consist of esp6500si_all, exac03, and 1000g2015aug_all. We also annotated our variants by using ClinVar (v. date 2018-02-25) which comprises human variation associated with a phenotype [31].

Annotation of genetic variants according to ACMG and AMP guidelines

The expert panel from ACMG and Association for Molecular Pathology (AMP) put forward 28 criteria for variant classification into 5 broad categories i.e. pathogenic, likely pathogenic, benign, likely benign and variant of uncertain significance (VUS). Each criterion depending on the pathogenicity of the variant was weighted as very strong (PVS1), strong (PS1-4), moderate (PM1-6) or supporting (PP1-5). Similar cases for benign characteristics were weighted as stand-alone (BA1), strong (BS1-4) or supporting (BP1-7). These criteria in combination classify the variant into these 5 broad categories [36]. Method for annotation and criteria is detailed in Supplementary data S1.

Each attribute has been assigned a weight by ACMG experts based on their pathogenicity. These weight has been aggregated in the Genetic variant interpretation tool https://www.medschool.umaryland.edu/Genetic_Variant_Interpretation_Tool1.html/ which classify them into five categories viz pathogenic, likely pathogenic, benign, likely benign and VUS.

Global Population Datasets

We used four popular and large datasets for global populations to compare allele frequencies.

The 1000 Genomes project (1000g2015aug_all) dataset encompasses genomes of 2504 individuals from 26 populations in Europe (EUR), South Asia (SAS), Africa (AFR), East Asia (EAS), and the Americas (AMR) which comprises of 84.7 million SNPs, 3.6 million small indels and 60,000 structural variants [37].

The National Heart Lung and Blood Institute (NHLBI) Exome Sequencing Project (ESP) comprised of unrelated 2203 African-American and 4300 European-American totaling to 6503 individual exome sequenced database. These individuals were composed of healthy controls, certain traits related to disease [38].

The Exome Aggregation Consortium (ExAC) dataset is a compilation of exome sequencing data from a variety of sequencing projects all across the world, summarized together to provide a single platform for scientific studies. The data set comprised of exome sequences obtained from 60,706 unrelated individuals involved in disease-specific and populations genetics studies [38,39].

The Genome Aggregation Database (gnomAD) comprises date from 123,136 exomes and 15,496 whole genomes of unrelated individuals. These individuals belong to 7 ancestries i.e. African/African American, Latino, Ashkenazi Jewish, East Asian, Finnish, Non-Finish European, South Asian and other [40].

Statistical Significance of pathogenic variants

The differences in allele frequencies of variants annotated as pathogenic or likely pathogenic by ACMG guidelines were tested for statistical significance. The allele counts of the pathogenic variants in Qatar and GME were compared with 1000 Genome project (Ensembl 87 browser), ExAC and in some cases gnomAD. For statistical significance, we implemented Fisher's exact test with a *p*-value less than 0.05 were considered significant.

Signals of Natural Selection in genes

Natural Selection of genes were predicted based on the extended haplotype homozygosity (EHH) statistics which predicts the positive selection of the derived or ancestral allele depends on their localization of the long haplotype. The statistical analysis provides the integrated Haplotype Score for the variant which should be >2 or <-2 to be predicted as positively selected. Another methodology which is based on the allele frequency difference between population can deduce the selection pressure i.e. Wright's fixation index (Fst) [40,41]. So we have calculated iHS for all variants of the Qatar population using selection tools pipeline [42] freely available on GitHub http://www.github.com/smilefreak/selectionTools/. Out of which the top 1% of variants were filtered. Then we filtered out variants of the autoinflammatory genes to know the selection pressure in these genes. Those autoinflammatory genes which fall in the top 1% criteria of iHS have taken further and calculated the Fst using PLINK algorithm to infer selection pressure among populations [42,43]. We calculated the Fst score among populations i.e. Qatar population and its subpopulations with 1000 genome subpopulations (African, European and South Asian). The overall methodology has been well depicted in Fig 1.

Data Availability

The data that support the findings of this study are available from the NCBI Sequence Read Archive (SRA accessions SRP060765, SRP061943 and SRP061463, accessible online at http://www.ncbi.nlm.nih.gov/Traces/study/?acc=SRP060765%2CSRP061943%2CSRP061463&go=go. (SRA accession SRP061943).

Reference

1. Hausmann, J. S., Lomax, K. G., Shapiro, A. & Durrant, K. The patient journey to diagnosis and treatment of autoinflammatory diseases. *Orphanet J. Rare Dis.* 13, 156 (2018).
2. Ben-Chetrit, E. *et al.* Consensus proposal for taxonomy and definition of the autoinflammatory diseases (AIDs): a Delphi study. *Ann. Rheum. Dis.* 77, 1558–1565 (2018).
3. Ciccarelli, F., De Martinis, M. & Ginaldi, L. An update on autoinflammatory diseases. *Curr. Med. Chem.* 21, 261–269 (2014).
4. Touitou, I. *et al.* Infevers: an evolving mutation database for auto-inflammatory syndromes. *Hum. Mutat.* 24, 194–198 (2004).
5. Martinez-Quiles, N. & Goldbach-Mansky, R. Updates on autoinflammatory diseases. *Curr. Opin. Immunol.* 55, 97–105 (2018).
6. Milhavet, F. *et al.* The infevers autoinflammatory mutation online registry: update with new genes and functions. *Hum. Mutat.* 29, 803–808 (2008).
7. Van Gijn, M. E. *et al.* New workflow for classification of genetic variants' pathogenicity applied to hereditary recurrent fevers by the International Study Group for Systemic Autoinflammatory Diseases (INSAID). *J. Med. Genet.* 55, 530–537 (2018).
8. Sarrauste de Menthière, C. *et al.* INFEVERS: the Registry for FMF and hereditary inflammatory disorders mutations. *Nucleic Acids Res.* 31, 282–285 (2003).
9. Yao, Q. & Furst, D. E. Autoinflammatory diseases: an update of clinical and genetic aspects. *Rheumatology* 47, 946–951 (2008).
10. Autoinflammatory Syndromes. in *Pediatric Rheumatology in Clinical Practice* 123–136 (Springer, London, 2007). doi:10.1007/978-1-84628-421-2_10.
11. The International FMF Consortium. Ancient Missense Mutations in a New Member of the RoRet Gene Family Are Likely to Cause Familial Mediterranean Fever. *Cell* 90, 797–807 (1997).
12. Salehzadeh, F. Familial Mediterranean Fever in Iran: A Report from FMF Registration Center.

Int. J. Rheumatol. 2015, 912137 (2015).

13. Cush, J. J. Autoinflammatory syndromes. *Dermatol. Clin.* 31, 471–480 (2013).
14. Sarkisian, T., Ajrapetian, H., Beglarian, A., Shahsuvarian, G. & Egiazarian, A. Familial Mediterranean Fever in Armenian population. *Georgian Med. News* 105–111 (2008).
15. Özen, S., Batu, E. D. & Demir, S. Familial Mediterranean Fever: Recent Developments in Pathogenesis and New Recommendations for Management. *Front. Immunol.* 8, 253 (2017).
16. Martorana, D., Bonatti, F., Mozzoni, P., Vaglio, A. & Percesepe, A. Monogenic Autoinflammatory Diseases with Mendelian Inheritance: Genes, Mutations, and Genotype/Phenotype Correlations. *Front. Immunol.* 8, 344 (2017).
17. Koshy, R., Sivadas, A. & Scaria, V. Genetic epidemiology of familial Mediterranean fever through integrative analysis of whole genome and exome sequences from Middle East and North Africa. *Clin. Genet.* 93, 92–102 (2018).
18. Teebi, A. S. & Farag, T. I. *Genetic Disorders Among Arab Populations.* (Oxford University Press, USA, 1997).
19. Alenazi, A., Al Sonbul, A., Al Jumaah, S., Al Mehaidib, A. & Al-Mayouf, S. M. A retrospective review of autoinflammatory diseases in Saudi children at a rheumatology clinic. *Ann. Saudi Med.* 32, 43–48 (2012).
20. Al-Mosawi, Z. S., Al-Saad, K. K., Ijadi-Maghsoodi, R., El-Shanti, H. I. & Ferguson, P. J. A splice site mutation confirms the role of LPIN2 in Majeed syndrome. *Arthritis Rheum.* 56, 960–964 (2007).
21. Majeed, H. A. *et al.* Congenital dyserythropoietic anemia and chronic recurrent multifocal osteomyelitis in three related children and the association with Sweet syndrome in two siblings. *J. Pediatr.* 115, 730–734 (1989).
22. Aladbe, B. *et al.* P02-007 - Childhood autoinflammatory disorders in Qatar. *Pediatric Rheumatology* vol. 11 A114 (2013).
23. Yurdakul, S. & Yazıcı, Y. Epidemiology of Behçet's Syndrome and Regional Differences in Disease Expression. in *Behçet's Syndrome* 35–52 (Springer, New York, NY, 2010). doi:10.1007/978-1-4419-5641-5_3.
24. Toplak, N. *et al.* An international registry on autoinflammatory diseases: the Eurofever experience. *Ann. Rheum. Dis.* 71, 1177–1182 (2012).
25. Hemminki, K., Li, X., Försti, A., Sundquist, J. & Sundquist, K. Incidence of hereditary amyloidosis and autoinflammatory diseases in Sweden: endemic and imported diseases. *BMC Med. Genet.* 14, 88 (2013).

26. Renert-Yuval, Y. *et al.* IL36RN mutation causing generalized pustular psoriasis in a Palestinian patient. *Int. J. Dermatol.* 53, 866–868 (2014).
27. Patel, N. *et al.* Study of Mendelian forms of Crohn's disease in Saudi Arabia reveals novel risk loci and alleles. *Gut* vol. 63 1831–1832 (2014).
28. Alangari, A. *et al.* LPS-responsive beige-like anchor (LRBA) gene mutation in a family with inflammatory bowel disease and combined immunodeficiency. *J. Allergy Clin. Immunol.* 130, 481–8.e2 (2012).
29. Fakhro, K. A. *et al.* The Qatar genome: a population-specific tool for precision medicine in the Middle East. *Hum Genome Var* 3, 16016 (2016).
30. Scott, E. M. *et al.* Characterization of Greater Middle Eastern genetic variation for enhanced disease gene discovery. *Nat. Genet.* 48, 1071–1076 (2016).
31. Landrum, M. J. *et al.* ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res.* 46, D1062–D1067 (2018).
32. Bousfiha, A. *et al.* The 2017 IUIS Phenotypic Classification for Primary Immunodeficiencies. *J. Clin. Immunol.* 38, 129–143 (2018).
33. Cimaz, R. *Periodic and Non-Periodic Fevers*. (Springer International Publishing, 2019). doi:10.1007/978-3-030-19055-2.
34. Stenson, P. D. *et al.* The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. *Hum. Genet.* 133, 1–9 (2014).
35. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 38, e164 (2010).
36. Richards, S. *et al.* Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.* 17, 405–424 (2015).
37. The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* 526, 68–74 (2015).
38. Fu, W. *et al.* Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. *Nature* 493, 216–220 (2013).
39. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 536, 285–291 (2016).
40. Karczewski, K. J. *et al.* Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. *bioRxiv* 531210

(2019) doi:10.1101/531210.

41. Vitti, J. J., Grossman, S. R. & Sabeti, P. C. Detecting natural selection in genomic data. *Annu. Rev. Genet.* 47, 97–120 (2013).
42. Cadzow, M. *et al.* A bioinformatics workflow for detecting signatures of selection in genomic data. *Front. Genet.* 5, 293 (2014).
43. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575 (2007).
44. McDermott, M. F. *et al.* Germline Mutations in the Extracellular Domains of the 55 kDa TNF Receptor, TNFR1, Define a Family of Dominantly Inherited Autoinflammatory Syndromes. *Cell* 97, 133–144 (1999).
45. Ter Haar, N. M. *et al.* Clinical characteristics and genetic analyses of 187 patients with undefined autoinflammatory diseases. *Ann. Rheum. Dis.* (2019) doi:10.1136/annrheumdis-2018-214472.
46. Tunca, M. & Ozdogan, H. Molecular and genetic characteristics of hereditary autoinflammatory diseases. *Curr. Drug Targets Inflamm. Allergy* 4, 77–80 (2005).
47. de Jesus, A. A., Canna, S. W., Liu, Y. & Goldbach-Mansky, R. Molecular mechanisms in genetically defined autoinflammatory diseases: disorders of amplified danger signaling. *Annu. Rev. Immunol.* 33, 823–874 (2015).
48. Manthiram, K., Zhou, Q., Aksentijevich, I. & Kastner, D. L. Erratum: Corrigendum: The monogenic autoinflammatory diseases define new pathways in human innate immunity and inflammation. *Nature Immunology* vol. 18 1271–1271 (2017).
49. Setta-Kaffetzi, N. *et al.* AP1S3 mutations are associated with pustular psoriasis and impaired Toll-like receptor 3 trafficking. *Am. J. Hum. Genet.* 94, 790–797 (2014).
50. Lodolce, J. P. *et al.* African-derived genetic polymorphisms in TNFAIP3 mediate risk for autoimmunity. *J. Immunol.* 184, 7001–7009 (2010).
51. Houten, S. M., van Woerden, C. S., Wijburg, F. A., Wanders, R. J. A. & Waterham, H. R. Carrier frequency of the V377I (1129G>A) MVK mutation, associated with Hyper-IgD and periodic fever syndrome, in the Netherlands. *Eur. J. Hum. Genet.* 11, 196–200 (2003).
52. Houten, S. M. *et al.* Mutations in MVK, encoding mevalonate kinase, cause hyperimmunoglobulinaemia D and periodic fever syndrome. *Nat. Genet.* 22, 175–177 (1999).
53. Drenth, J. P. *et al.* Mutations in the gene encoding mevalonate kinase cause hyper-IgD and periodic fever syndrome. International Hyper-IgD Study Group. *Nat. Genet.* 22, 178–181 (1999).

54. Houten, S. M. *et al.* Organization of the mevalonate kinase (MVK) gene and identification of novel mutations causing mevalonic aciduria and hyperimmunoglobulinaemia D and periodic fever syndrome. *Eur. J. Hum. Genet.* 9, 253–259 (2001).
55. Cuisset, L. *et al.* Molecular analysis of MVK mutations and enzymatic activity in hyper-IgD and periodic fever syndrome. *Eur. J. Hum. Genet.* 9, 260–266 (2001).
56. Levy, M. *et al.* Severe early-onset colitis revealing mevalonate kinase deficiency. *Pediatrics* 132, e779–83 (2013).
57. Netter, P. *et al.* A novel Rab27a mutation binds melanophilin, but not Munc13-4, causing immunodeficiency without albinism. *J. Allergy Clin. Immunol.* 138, 599–601.e3 (2016).
58. Cetica, V. *et al.* Patients with Griscelli syndrome and normal pigmentation identify RAB27A mutations that selectively disrupt MUNC13-4 binding. *J. Allergy Clin. Immunol.* 135, 1310–8.e1 (2015).
59. Horowitz, J. E. *et al.* Mutation spectrum of NOD2 reveals recessive inheritance as a main driver of Early Onset Crohn's Disease. doi:10.1101/098574.
60. Rivas, M. A. *et al.* Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. *Nat. Genet.* 43, 1066–1073 (2011).
61. Tukel, T. *et al.* Crohn disease: frequency and nature of CARD15 mutations in Ashkenazi and Sephardi/Oriental Jewish families. *Am. J. Hum. Genet.* 74, 623–636 (2004).
62. Jéru, I. *et al.* Identification and functional consequences of a recurrent NLRP12 missense mutation in periodic fever syndromes. *Arthritis Rheum.* 63, 1459–1464 (2011).
63. Hammoudeh, M. Hyperimmunoglobulinemia D syndrome in an Arab child. *Clin. Rheumatol.* 24, 92–94 (2005).
64. Moussa, T. *et al.* Overlap of familial Mediterranean fever and hyper-IgD syndrome in an Arabic kindred. *J. Clin. Immunol.* 35, 249–253 (2015).
65. Tas, D. A., Dınkcı, S. & Erken, E. Different clinical presentation of the hyperimmunoglobulin D syndrome (HIDS) (four cases from Turkey). *Clin. Rheumatol.* 31, 889–893 (2012).
66. van der Hilst, J. C. H. *et al.* Long-term follow-up, clinical features, and quality of life in a series of 103 patients with hyperimmunoglobulinemia D syndrome. *Medicine* 87, 301–310 (2008).
67. Megged, O., Navon-Elkan, P., Anikster, Y. & Kleid, D. Recurrent Fever, arthritis, lymphadenopathy, and hepatosplenomegaly. *Pediatr. Infect. Dis. J.* 30, 182–185 (2011).
68. Harel-Meir, M., Bujanover, Y., Berkun, Y., Anikster, Y. & Anikster, Y. Mevalonic Aciduria in a Child Featuring Hepatic Fibrosis and Novel Mevalonate Kinase Mutations. *The Open Pediatric Medicine Journal* vol. 3 45–47 (2009).

69. Budak Diler, S. & Yaraş, S. CTLA-4 (+49A/G) and NOD2/CARD15 (N852S) polymorphisms with inflammatory bowel disease in Turkish patients. *Cell. Mol. Biol.* 64, 97–101 (2018).
70. Paz-Yaacov, N. *et al.* Adenosine-to-inosine RNA editing shapes transcriptome diversity in primates. *Proc. Natl. Acad. Sci. U. S. A.* 107, 12174–12179 (2010).
71. Jaffe, S. *et al.* Unique variation in genetic selection among Black North American women and its potential influence on pregnancy outcome. *Med. Hypotheses* 81, 919–922 (2013).
72. Akey, J. M. *et al.* Population history and natural selection shape patterns of genetic variation in 132 genes. *PLoS Biol.* 2, e286 (2004).
73. Lv, Z. *et al.* Integrative genomic analysis of interleukin-36RN and its prognostic value in cancer. *Mol. Med. Rep.* 13, 1404–1412 (2016).
74. Al-Ali, M., Osman, W., Tay, G. K. & AlSafar, H. S. A 1000 Arab genome project to study the Emirati population. *J. Hum. Genet.* 63, 533–536 (2018).

Acknowledgments

Authors thank Disha Sharma, Mukta Poojary, and Bani Jolly for suggestions which enriched the manuscript. Authors acknowledge funding from the Council of Scientific and Industrial Research (CSIR, India) through Grant MLP1809 (IndiGen). Abhinav Jain is a recipient of Senior Research Fellowship from Council of Scientific and Industrial Research (CSIR, India). The funders had no role in the preparation of the manuscript or decision to submit. We acknowledge the researchers at Weill Cornell Medicine for sharing the genome and exome data sets without which this analysis was not possible.

Author contributions

P.S. collected the data of the variants and genes. A.J. performed the annotations and classification as per the ACMG & AMP guidelines along with the data analysis of allele frequencies. V.S. conceptualized the project and supervised the analysis. All authors contributed to writing the manuscript.

Conflict of Interest

The author(s) declares no competing interests.

Figure legends

Figure 1: Schematic summarising the data analysis pipeline utilised for this study.

Figure 2: Venn Diagram for the variants present in HGMD, ClinVar and Infevers.

Figure 3: Comparison of allele frequency of pathogenic and likely pathogenic variants in Qatari population with its subpopulation and GME and its subpopulation with 1000 genome and its subpopulation and ExAC and its subpopulation

Significant allele frequency highlighted with red circle.

Figure 4: The Fst and iHS score depicted for RAB27A gene from Qatari population (QALL: Qatar 1005 dataset; 1AFR,1000 genome AFRICAN; 1EUR, 1000genome EUROPEAN; 1SAS, 1000 genomeSouth Asian; QPER:Qatar-Persian subset;QBED,:Qatar-Bedouin subset; QARB:Qatar-Arab-subset; QAFR: Qatar-African-subset; QSAS:Qatar-South-Asian-subset).

Tables

Disease Name	Gene	Inheritance
Familial Cold Autoinflammatory Syndrome (FCAS)	*NLRP3*	Autosomal Dominant
Muckle-Wells syndrome (MWS)	*NLRP3*	Autosomal Dominant
Neonatal onset multi-systemic inflammatory disorder/ Chronic Infantile Neurological Cutaneous Articular Syndrome (NOMID/CINCA)	*NLRP3*	Autosomal Dominant
TNF receptor-associated periodic syndrome (TRAPS)	*TNFRSF1A*	Autosomal Dominant
Hyperimmunoglobulinemia D with Periodic Fever Syndrome (HIDS)	*MVK*	Autosomal Recessive
Mevalonate Aciduria (MA)	*MVK*	Autosomal Recessive
Porokeratosis 3, multiple types	*MVK*	Autosomal Dominant
Deficiency of Interleukin-1ß (IL-1ß) Receptor Antagonist/ Osteomyelitis, Sterile Multifocal w/Periostitis Pustulosis (DIRA/OMPP)	*IL1RN*	Autosomal Recessive
MAJEED/ Chronic Recurrent Multifocal Osteomyelitis, Congenital Dyserythropoietic Anemia, & Neutrophilic Dermatosis	*LPIN2*	Autosomal Recessive
Deficiency of Interleukin-36-Receptor Antagonist/ Generalized	*IL36RN*	Autosomal

Pustular Psoriasis (GPP)/ (DITRA/PSORP)		Recessive
Familial Psoriasis/ CARD14-Mediated Pustular Psoriasis (CAMPS/PSORS2)	*CARD14*	Autosomal Dominant
Pityriasis rubra pilaris	*CARD14*	Autosomal Dominant
Pyogenic Sterile Arthritis, Pyoderma Gangrenosum and Acne Syndrome (PAPA)	*PSTPIP1*	Autosomal Dominant
Juvenile Systemic Granulomatosis– Blau syndrome, Pediatric Granulomatous Arthritis (PGA), Early Onset Sarcoidosis, or Jabs Syndrome (BLAU/PGA/EOS)	*NOD2*	Autosomal Dominant
NLRP12-Associated Periodic Fever Syndrome/ Familial Cold Autoinflammatory Syndrome 2, or Guadaloupe Periodic Fever (NLRP12/FCAS2)	*NLRP12*	Autosomal Dominant
Chronic Atypical Neutrophilic Dermatosis w/ Lipodystrophy and Elevated Temperature- Nakajo Nishimura Syndrome (CANDLE/PRAAS)	*PSMB8; also PSMB4, PSMB9,PSMA3, POMP*	Autosomal Recessive
(Primary) Familial Hemophagocytic Lymphohistiocytosis/ a Familial Erythrophagocytic Lymphohistiocytosis (1°HLH/FHL)	*PRF1, STX11, STXBP2, MUNC13-4/UNC 13D, RAB27A*	Autosomal recessive
(Primary) Familial Hemophagocytic Lymphohistiocytosis/ a Familial Erythrophagocytic Lymphohistiocytosis (1°HLH/FHL)	*SH2D1A,, BIRC4/XIAP*	X-linked recessive
PLCG2-associated Antibody Deficiency and Immune Dysregulation/ Familial Atypical Cold Urticaria (FACU) (PLAID/FCAS3)	*PLCG2*	Autosomal Dominant
Autoinflammation and PLCG2-associated Antibody Deficiency and Immune Dysregulation (APLAID)	*PLCG2*	Autosomal Dominant
SLC29A3 Spectrum Disorder (SLC29A3)	*SLC29A3*	Autosomal Recessive
Psoriasis 15, Pustular	*AP1S3*	Autosomal Dominant

Deficiency of Adenosine Deaminase 2 (DADA2)	*CECR1/ADA2*	Autosomal Recessive
Sneddon syndrome	*CECR1/ADA2*	Autosomal Recessive
Interleukin 10 deficiency (IL10D)	*IL10*	Autosomal Recessive
Inflammatory bowel disease 28 (IBD28)/ Interleukin 10 receptor A deficiency (IL10R1D)	*IL10RA*	Autosomal Recessive
Inflammatory bowel disease 25 (IBD25)/ Interleukin 10 receptor B deficiency (IL10R2D)	*IL10RB*	Autosomal Recessive
Autoinflammation with infantile enterocolitis (AIFEC)	*NLRC4*	Autosomal Dominant
Hydatidiform Mole, Recurrent, 1 (HYDM1)	*NLRP7*	Autosomal Recessive
Otulipenia	*OTULIN*	Autosomal Recessive
Polyglucosan Body Myopathy, Early-onset, with Or without Immunodeficiency (PBMEI)	*RBCK1*	Autosomal Recessive
Cherubism	*SH3BP2*	Autosomal Dominant
Sting-Associated Vasculopathy, Infantile-Onset (SAVI)	*TMEM173*	Autosomal Dominant
Autoinflammatory syndrome, familial, Behcet-like (AISBL)	*TNFAIP3*	Autosomal Dominant
TNFRSF11A-associated hereditary fever disease (TRAPS11)	TNFRSF11A	Autosomal Dominant

Table 1. Genes and associated autoinflammatory diseases considered in the present analysis and their respective inheritance patterns.

Gene	Chr-pos	Variant	SNPID	Protein Change	ACMG criteria	Annotation	Inheritance	Disease
AP1S3	2-224642579	A>C	rs116107386	F4C	PS3,PM1,PP3,BS4	Likely pathogenic (II)	AD	Psoriasis 15
TNFAIP3	6-138196060	C>T	rs5029941	A125V	PS3,PM1,PP3	Likely pathogenic (II)	AD	Autoinflammatory syndrome, familial, IBD
MVK	12-110034320	G>A	rs28934897	V377I, V325I	PS4,PM3,PP1,PP5,BP4	Likely pathogenic (II)	AR	Hyperimmunoglobulinemia D with Periodic Fever Syndrome (HIDS) and Mevalonic aciduria (MA)
RAB27A	15-55520906	G>A	rs753966933	R82C	PM1,PM2, PP3,PP5	Pathogenic III (b)	AR	Hemophagocytic Lymphohistiocytosis
NOD2	16-50750810	A>G	rs104895467	N852S, N825S	PS3,PM3,PP3	Likely pathogenic (II)	AD	Blau syndrome, Pediatric
NLRP12	19-54313859	G>A	rs199881207	R352C	PS3,PM1,PP3,PP5	Likely pathogenic (II)	AD	NLRP12-Associated Periodic Fever Syndrome
NLRP12	19-54314063	G>A	rs104895564	R284X	PVS1,PM1, PP5	Pathogenic 1(c)	AD	NLRP12-Associated Periodic Fever Syndrome

Table 2. Pathogenic and Likely Pathogenic variants annotated as per ACMG guidelines for interpretation of variants

Pathogenic variant	Gene	Annotation	Allele	Qatar_All	Qatar Sub populations						
					EUR	SOU	BED	AYG	ARA	PER	SAF
2-224642579	AP1S3	Likely pathogenic (II)	Ref-A	0.994318	Not found	1	1	Not found	1	1	0.97059
rs116107386			Alt-C	0.005682	Not found	0	0	Not found	0	0	0.02941
			p-value	0.425	NA	NA	NA	NA	NA	NA	0.1028
6-138196060	TNFAIP3	Likely pathogenic (II)	Ref: C	0.998406	1	0.99231	1	1	1	1	0.98413
rs50299			Alt: T	0.00159	0	0.00769	0	0	0	0	0.01587

41				4							
			p-value	0.0175*	0	0.5492	0	0	0	0	0.1847
12-110034320	MVK	Pathogenic (IIIb)	Ref: G	0.996788	1	1	0.997845	1	0.991713	1	0.992063
rs28934897			Alt: A	0.003212	0	0	0.002155	0	0.008287	0	0.007937
			p-value	0.054	NA	NA	0.3785	NA	0.0155*	NA	0.1637
15-55520906	RAB27A	Likely pathogenic (V)	Ref: G	0.994155	1	1	0.998929	1	0.97268	1	1
rs753966933			Alt: A	0.005845	0	0	0.001071	0	0.02732	0	0
			p-value	< 0.00001 ***	NA	NA	0.0188	NA	< 0.00001 ***	NA	NA
16-50750810	NOD2	Likely pathogenic (V)	Ref: A	0.998828	1	1	0.99768	1	1	1	1
rs104895467			Alt: G	0.001172	0	0	0.00232	0	0	0	0
			p-value	0.2686	NA	NA	0.1054	NA	NA	NA	NA
19-54313859	NLRP12	Likely pathogenic (II)	Ref: G	0.998828	0.875	0.992063	1	1	1	1	1
rs199881207			Alt: A	0.001172	0.125	0.007937	0	0	0	0	0
			p-value	0.1292	0.0033*	0.0447*	NA	NA	NA	NA	NA
19-54314063	NLRP12	Pathogenic (Ib)	Ref: G	0.994318	NA	1	1	NA	1	1	0.97059
rs104895564			Alt: A	0.005682	NA	0	0	NA	0	0	0.02941
			p-value	0.3621	NA	NA	NA	NA	NA	NA	0.0843

Table 3. Allele frequency of pathogenic and likely pathogenic variants compared with control global population (1000 genome, ExAC or gnomAD) using Fisher exact test ($p<0.05$). BED Bedouin, SAF Sub-Saharan African, EUR European, SOU South Asian, APY African Pygmy, ARA ARAB, PER Persian, NA not applicable]. Significant values are marked with *

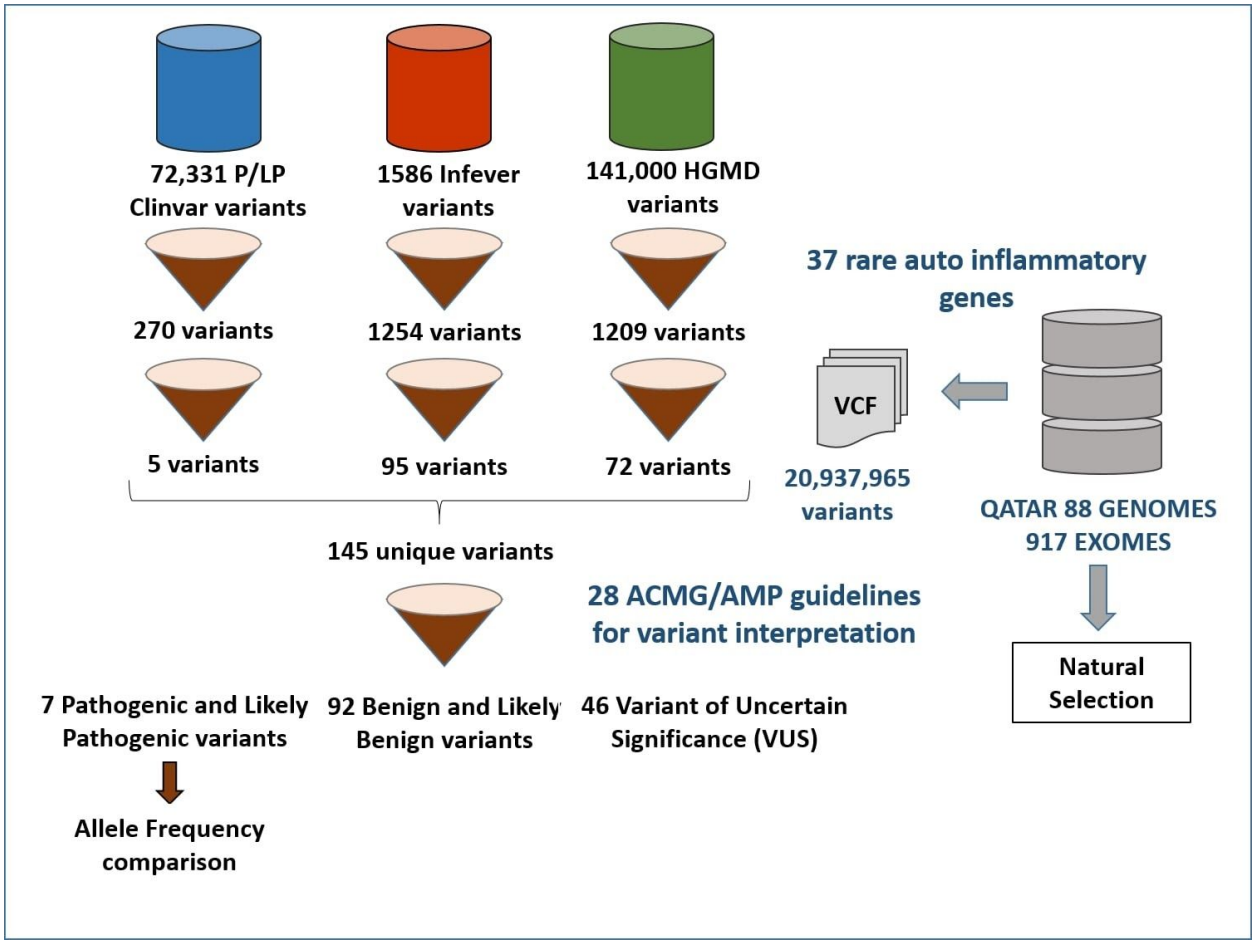

Figure 1: Schematic summarising the data analysis pipeline utilised for this study.

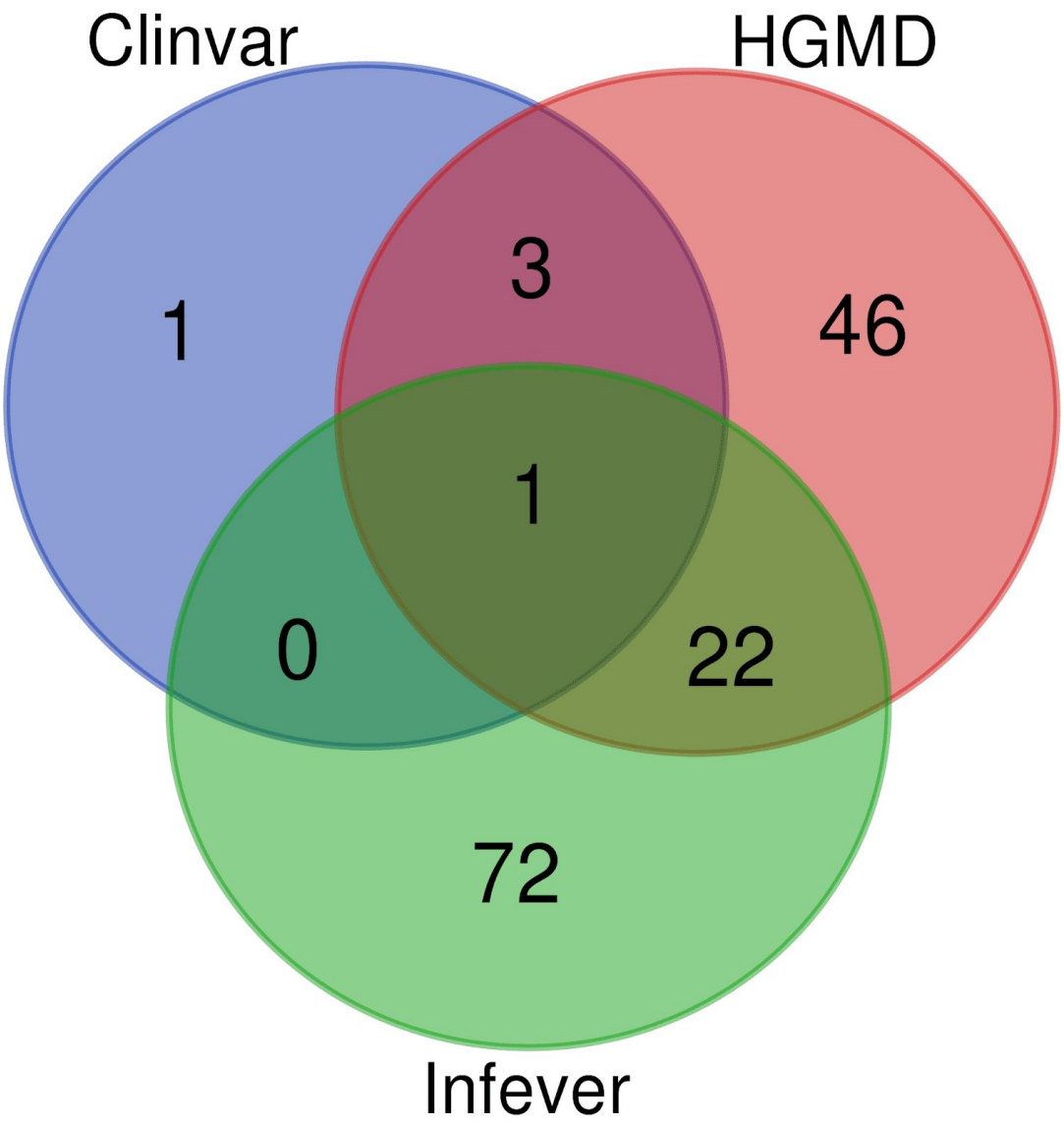

Figure 2: Venn Diagram for the variants present in HGMD, ClinVar and Infevers.

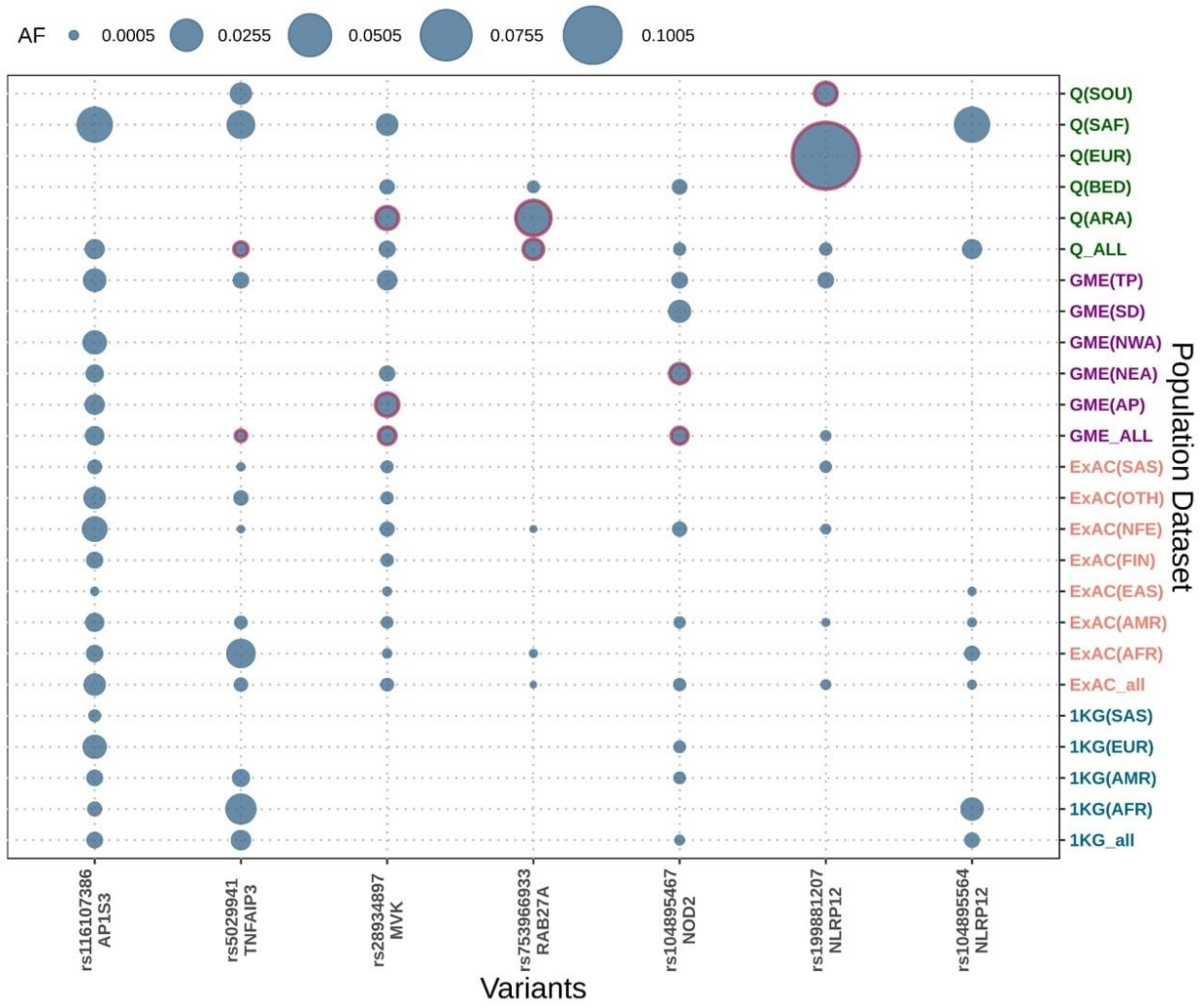

Figure 3: Comparison of allele frequency of pathogenic and likely pathogenic variants in Qatari population with its subpopulation and GME and its subpopulation with 1000 genome and its subpopulation and ExAC and its subpopulation

Significant allele frequency highlighted with red circle.

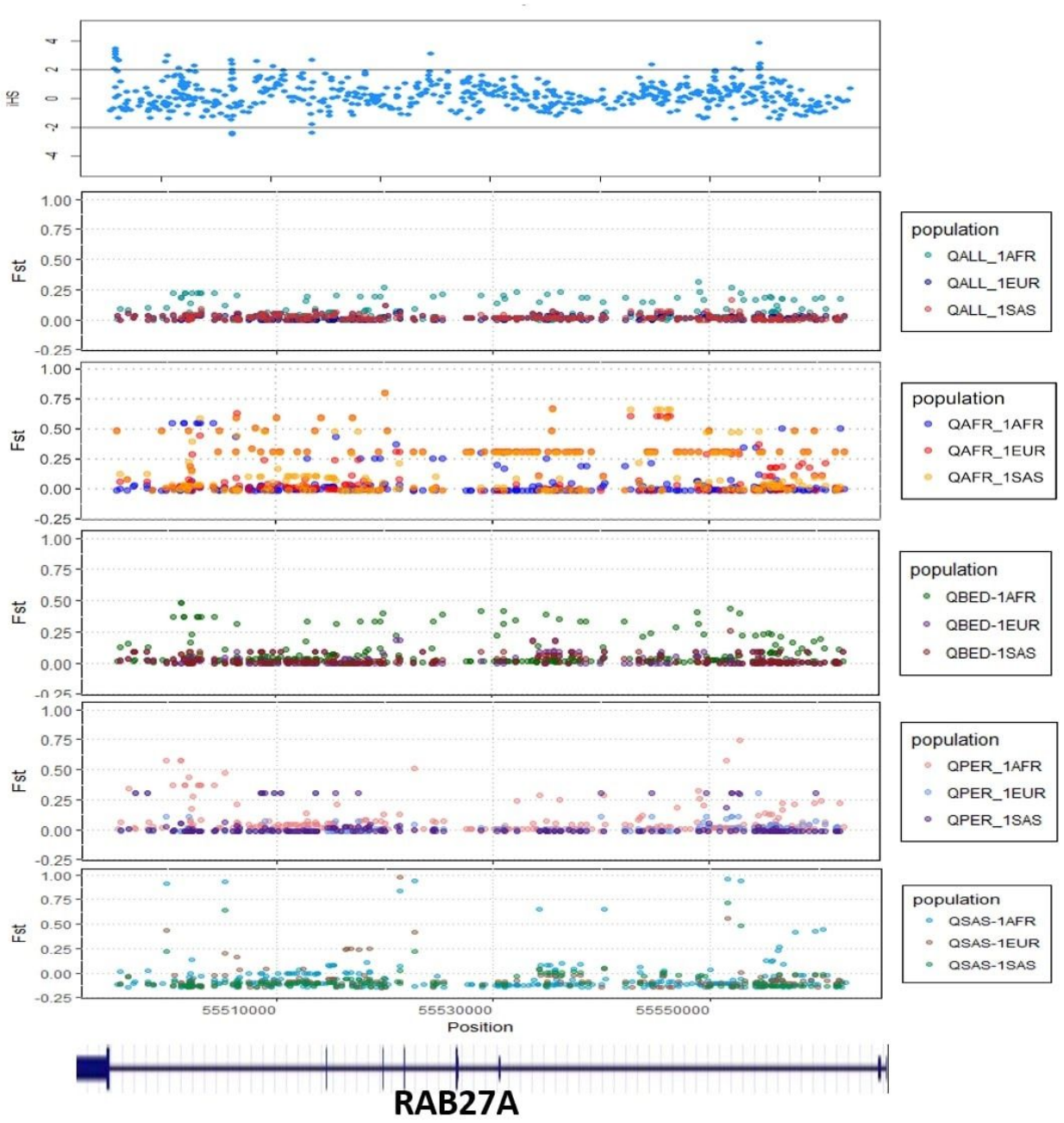

Figure 4: The Fst and iHS score depicted for RAB27A gene from Qatari population (QALL: Qatar 1005 dataset; 1AFR,1000 genome AFRICAN; 1EUR, 1000genome EUROPEAN; 1SAS, 1000 genomeSouth Asian; QPER:Qatar-Persian subset;QBED,:Qatar-Bedouin subset; QARB:Qatar-Arab-subset; QAFR: Qatar-African-subset; QSAS:Qatar-South-Asian-subset).