

RESEARCH

Open Access



Genetic epidemiology of autoinflammatory disease variants in Indian population from 1029 whole genomes

Abhinav Jain^{1,2}, Rahul C. Bhojar¹, Kavita Pandhare¹, Anushree Mishra¹, Disha Sharma¹, Mohamed Imran^{1,2}, Vigneshwar Senthivel^{1,2}, Mohit Kumar Divakar^{1,2}, Mercy Rophina^{1,2}, Bani Jolly^{1,2}, Arushi Batra^{1,2}, Sumit Sharma¹, Sanjay Siwach¹, Arun G. Jadhao³, Nikhil V. Palande⁴, Ganga Nath Jha⁵, Nishat Ashrafi⁵, Prashant Kumar Mishra⁶, Vidhya A.K.⁷, Suman Jain⁸, Debasis Dash^{1,2}, Nachimuthu Senthil Kumar⁹, Andrew Vanlallawma⁹, Ranjan Jyoti Sarma⁹, Lalchhandama Chhakchhuak¹⁰, Shantaraman Kalyanaraman¹¹, Radha Mahadevan¹¹, Sunitha Kandasamy¹¹, Pabitha B. M.¹¹, Raskin Erusan Rajagopal¹¹, Ezhil Ramya J.¹¹, Nirmala Devi P.¹¹, Anjali Bajaj^{1,2}, Vishu Gupta^{1,2}, Samatha Mathew^{1,2}, Sangam Goswami^{1,2}, Mohit Mangla^{1,2}, Savinitha Prakash¹, Kandarp Joshi¹, Meyakumla¹, Sreedevi S.¹², Devarshi Gajjar¹³, Ronibala Soraisham¹⁴, Rohit Yadav^{1,2}, Yumnam Silla Devi¹⁵, Aayush Gupta¹⁶, Mitali Mukerji^{1,2}, Sivaprakash Ramalingam^{1,2}, Binukumar B. K.^{1,2}, Vinod Scaria^{1,2*}  and Sridhar Sivasubbu^{1,2*}

Abstract

Background: Autoinflammatory disorders are the group of inherited inflammatory disorders caused due to the genetic defect in the genes that regulates innate immune systems. These have been clinically characterized based on the duration and occurrence of unprovoked fever, skin rash, and patient's ancestry. There are several autoinflammatory disorders that are found to be prevalent in a specific population and whose disease genetic epidemiology within the population has been well understood. However, India has a limited number of genetic studies reported for autoinflammatory disorders till date. The whole genome sequencing and analysis of 1029 Indian individuals performed under the IndiGen project persuaded us to perform the genetic epidemiology of the autoinflammatory disorders in India.

Results: We have systematically annotated the genetic variants of 56 genes implicated in autoinflammatory disorder. These genetic variants were reclassified into five categories (i.e., pathogenic, likely pathogenic, benign, likely benign, and variant of uncertain significance (VUS)) according to the American College of Medical Genetics and Association of Molecular pathology (ACMG-AMP) guidelines. Our analysis revealed 20 pathogenic and likely pathogenic variants with significant differences in the allele frequency compared with the global population. We also found six causal founder variants in the IndiGen dataset belonging to different ancestry. We have performed haplotype prediction analysis for founder mutations haplotype that reveals the admixture of the South Asian population with other populations. The cumulative carrier frequency of the autoinflammatory disorder in India was found to be 3.5% which is much higher than reported.

*Correspondence: vinods@igib.in; sridhar@igib.in

² Academy of Scientific and Innovative Research (AcSIR), Ghaziabad, Uttar Pradesh 201002, India

Full list of author information is available at the end of the article

Conclusion: With such frequency in the Indian population, there is a great need for awareness among clinicians as well as the general public regarding the autoinflammatory disorder. To the best of our knowledge, this is the first and most comprehensive population scale genetic epidemiological study being reported from India.

Keywords: Autoinflammatory disorder, Genetic epidemiology, American College of Medical Genetics and Genomics, Allele frequency, Haplotype ancestry

Background

Autoinflammatory disorders are the growing group of Mendelian disorders caused by genetic defects in the genes which regulate the innate immune system. These disorders have been characterized by recurrent episodes of fever, abdominal pain, skin rashes, arthritis, serositis, conjunctivitis, or cutaneous signs that lack specificity and therefore make diagnosis difficult [1, 2]. These disorders are clinically diagnosed based on the age of onset, duration of fever and flares, type of rash, family history, and patient's ancestry [3]. Autoinflammatory disorders have a strong genetic background, multiple new genes or variants have been recently discovered [4–7]. The recent advancement in next-generation sequencing has resulted in the identification of more than 30 new genes associated with autoinflammatory disorders [8, 9]. Classically, most monogenic autoinflammatory disorders follow autosomal dominant and autosomal recessive modes of inheritance. Molecular diagnosis along with clinical criteria is essential for confirmation of the disease as well as it aids in discovering new disease [10]. Even autoinflammatory disorders manifest heterogeneity in phenotype–genotype correlation (i.e., mutation in the same gene can have varying severity and clinical manifestation and mutation in different genes can result in similar clinical characteristics) [11–13].

Identification of the causal variants in the genes responsible for autoinflammatory disorder mainly uses clinical or whole exome sequencing. However, 60% of patients suspected with autoinflammatory disorder remain molecularly undiagnosed using these sequencing technologies [14]. The undiagnosis could be due to the inability of the sequencing platforms in calling structural variants (SVs) [15], missing non-coding variants [16] as well as inadequate coverage of the coding region [17], and high-quality SNVs [18]. However the whole genome sequencing (WGS) has the ability to identify variants in difficult-to-diagnose cases. Recently, our group has identified ~ 5 Kb deletion in the primary immunodeficiency disorder (PID) patients that could not be identified using whole exome sequencing [19]. Also, Thaventhiran et al. has identified eight SVs by performing the WGS of 1318 patients affected with PID, that could be missed using the WES [20]. WGS has also

been implemented at the population scale for autoinflammatory disorders to comprehend the genetic epidemiology of the Qatari population [21]. Even though India is a hotspot for the genetic disorder due to the prevalent practice of endogamy and consanguineous marriages, migration, and large population size, the genetic epidemiology of autoinflammatory disorders in the country has not been studied much except for a few reports [22].

The whole-genome sequence data for cosmopolitan Indian populations encompassing 1029 individuals as part of the IndiGen programme [23] motivated us to estimate the genetic epidemiology of autoinflammatory disorders. In the present analysis, we have performed extensive data mining as well as integrative analysis to evaluate the pathogenicity of the variant according to ACMG-AMP guidelines. We have also further analyzed the variant prevalence in India compared to the global population. This is the first most comprehensive genetic epidemiology performed for autoinflammatory disorders in India.

Methods

IndiGen dataset and variant annotation

A total of 59,646,267 genetic variants including single-nucleotide variants and Indels from the IndiGen dataset were considered for the analysis. The data was derived from whole genome sequencing of 1029 cosmopolitan healthy Indians with a well-written informed consent obtained [23]. These are self-declared healthy individuals selected from different states of India without any provisional diagnosis of autoinflammatory disorder. It includes 495 males and 534 females with mean ages of 41.35 and 32.96 years, respectively [23]. These variants obtained were annotated using a tool called ANNOVAR (v. 2018-04-06) [24]. ANNOVAR provides annotations from multiple databases that include RefGene [25] and dbSNP (avsnp150) [26] that provide detailed information about the variant, dbNSFP35a that comprises data of multiple pathogenicity prediction tools [27]; global population databases (gnomAD V3 [28], 1000 Genome Project [29], Esp6500 [30], and Greater Middle East (GME)) provide allele frequency of different ancestry [31]. Finally, variant clinical

significance of variants was retrieved from ClinVar (ver 2020-01-13) database [32].

IndiGen variant filtering

Out of the total IndiGen variants, we have extracted variants from 56 genes associated with the 47 auto-inflammatory disorders. These auto-inflammatory disorder genes were selected by the experts of the International Union of Immunological Societies (IUIS) [33] and Infever as tabulated in Table 1. Further, those variants that are mapped to the exonic (except synonymous) and splicing region or those that were pathogenic and likely pathogenic variants at the other genomic positions as per ClinVar were prioritized. Also, variants whose allele frequency is less than 0.05 in the global population datasets (1000 genome project, gnomAD V3, and Esp6500) were considered for further analysis. As the variants with allele frequency greater than 0.05 were considered polymorphic and present in the large number of healthy individuals across different populations.

Datasets of disease-associated genetic variants

We downloaded genetic variants from two well-annotated databases (i.e., ClinVar (ver 2020-01-13) [32] and Infevers [8, 34]). ClinVar is an up-to-date database composed of genetic variants in multiple genes associated with multiple disorders classified as pathogenic, likely pathogenic, benign, likely benign, variant of uncertain significance (VUS), or of conflicting evidence from different sources. We filtered variants that were pathogenic, likely pathogenic, VUS, and conflicting evidence as per ClinVar and retrieved variants in 56 genes associated with 47 auto-inflammatory disorders. The Infevers is a publicly available database that [35] has a total of 2112 genetic variants in 38 genes involved in 34 auto-inflammatory disorders. The filtered genetic variants from ClinVar and Infever were merged and mapped to the IndiGen filtered variants.

Classification of genetic variants according to ACMG and AMP guidelines

The American College of Medical Genetics and Genomics and the Association of Molecular Pathology (ACMG-AMP) experts have provided comprehensive guidelines which include annotation of 28 features for variant classification into five broad categories (i.e., pathogenic, likely pathogenic, benign, likely benign, and VUS) [36]. The variant classification was calculated based on the weighted 28 attributes. For variants, pathogenicity was weighted as very strong (PVS1),

strong (PS1-4), moderate (PM1-6), or supporting (PP1-5). Similarly, for benign, it was weighted as stand-alone (BA1), strong (BS1-4), or supporting (BP1-7). The combination of these attributes put together in the algorithm of the genetic variant interpretation tool [37] classifies the variant. The guidelines associated with the weighted attributes have been detailed in Supplementary Data 1.

Statistical significance of pathogenic variants with global population

The statistical differences in the allele frequencies of pathogenic or likely pathogenic variants were estimated between the Indian population of IndiGen dataset with the global population dataset. The global population datasets include Genome Aggregation Database (gnomAD v3), which is the largest database of 71,702 whole-genome-sequenced individuals from eight populations (African, Amish, Ashkenazi Jews, East Asian, Finnish, Non-Finnish, Latino, and South Asian) [28]; the 1000 Genome Project (1000g2015aug_all) was composed of whole genome sequencing of 2504 individuals from five super populations (Africa, America, Europe, East Asia, and South Asia) [29], and ESP6500 (esp6500siv2_all) composed of 6503 whole exome-sequenced dataset of 2203 African-American and 4300 European-American healthy individuals [38]. The statistical significance was tested by Fisher's exact test; a *P* value of less than 0.05 was considered significant.

Ancestry haplotyping of a founder mutation

Since several auto-inflammatory conditions have geographical associations, we were interested in the pathogenic or likely pathogenic variants which were prevalent and considered as a founder variant in the population. We further explored the haplotype similarity pattern of the founder variant of each individual from the IndiGen dataset to the global population dataset of the 1000 Genome Project. To evaluate the haplotype similarity, we used a tool called fineSTRUCTURE [39] (version 2.1.3) that uses linkage disequilibrium model-based STRUCTURE approach and principal component analysis for population haplotype prediction. Each individual specific chromosomal variant, where founder variants are located, were merged with 2548 individual variants from the 1000 Genome Project. Further, merged variants were pruned by applying a filter of allele frequency of > 1% and allele count > 1000 to get maximum genotype rate using a bespoke bash script. We phased the filtered merged VCF using SHAPEIT v2.r900 tool [40]. Further, the fineSTRUCTURE pipeline was used to analyze the phased VCF and plotted using R script. The overall workflow adopted in this study has been represented in Fig. 1.

Table 1 Genes associated with autoinflammatory disorder with their respective mode of inheritance

Gene	Disease	Mode of inheritance
<i>ACPS</i>	Spondyloenchondrodysplasia with immune dysregulation	AR
<i>ADA2</i>	Deficiency of adenosine deaminase 2 (DADA2)	AR
<i>ADAM17</i>	Inflammatory skin and bowel disease, neonatal	AR
<i>ADAR1; ADAR</i>	Aicardi-Goutières syndrome	AR
<i>ALPI</i>	Monogenic inflammatory bowel disease	AR
<i>AP1S3</i>	Psoriasis 15, pustular	AD
<i>CARD14</i>	Familial psoriasis/ CARD14-mediated pustular psoriasis (CAMPS/PSORS2)	AD
<i>CDC42</i>	Takenouchi-Kosaki syndrome	AD
<i>COPA</i>	Autoimmune interstitial lung, joint, and kidney disease	AD
<i>DNASE1L3</i>	Systemic lupus erythematosus	AR
<i>DNASE2</i>	Type I interferon-mediated autoinflammation	AD
<i>HAVCR2</i>	T cell lymphoma, subcutaneous panniculitis	AR
<i>IFIH1</i>	Aicardi-Goutières syndrome	AD
<i>IL10</i>	Interleukin 10 deficiency (IL10D)	AR
<i>IL10RA</i>	Inflammatory bowel disease 28 (IBD28)/ interleukin 10 receptor A deficiency (IL10R1D)	AR
<i>IL10RB</i>	Inflammatory bowel disease 25 (IBD25)/ interleukin 10 receptor B deficiency (IL10R2D)	AR
<i>IL1RN</i>	Deficiency of interleukin-1 β (IL-1 β) receptor antagonist/ osteomyelitis, sterile multifocal with periostitis pustulosis (DIRA/ OMPP)	AR
<i>IL36RN</i>	Deficiency of interleukin-36-receptor antagonist/ generalized pustular psoriasis (GPP)/ (DITRA/PSORP)	AR
<i>LACC1</i>	Juvenile arthritis	AR
<i>LPIN2</i>	MAJEED/ chronic recurrent multifocal osteomyelitis, congenital dyserythropoietic anemia, & neutrophilic dermatosis	AR
<i>MEFV</i>	Familial Mediterranean fever	AR; AD
<i>MVK</i>	Hyperimmunoglobulinemia D with periodic fever syndrome (HIDS); melvanoic aciduria (MA)	AR
<i>NCSTN</i>	Acne inversa, familial	AD
<i>NLRC4</i>	Autoinflammation with infantile enterocolitis (AIFEC)	AD
<i>NLRP12</i>	NLRP12-associated periodic fever syndrome/ familial cold autoinflammatory syndrome 2, or Guadeloupe periodic fever (NLRP12/FCAS2)	AD
<i>NLRP1</i>	Autoinflammation with arthritis and dyskeratosis	AR; AD
<i>NLRP3</i>	Familial cold autoinflammatory syndrome (FCAS); Muckle-Wells syndrome (MWS); neonatal onset multisystemic inflammatory disorder/ chronic infantile neurological cutaneous articular syndrome (NOMID/CINCA)	AD
<i>NLRP7</i>	Hydatidiform mole, recurrent, 1 (HYDM1)	AR
<i>NOD2</i>	Juvenile systemic granulomatosis–Blau syndrome, pediatric granulomatous arthritis (PGA), Crohn's disease early onset sarcoidosis, or Jabs syndrome (BLAU/PGA/EOS)	AD
<i>OAS1</i>	Pulmonary alveolar proteinosis with hypogammaglobulinemia	AR
<i>OTULIN</i>	Otulipenia	AR
<i>PLCG2</i>	PLCG2-associated antibody deficiency and immune dysregulation/ familial atypical cold urticaria (FACU) (PLAID/FCAS3); autoinflammation and PLCG2-associated antibody deficiency and immune dysregulation (APLAID)	AD
<i>POLA1</i>	Type I interferon-mediated autoinflammation	XLR
<i>POMP</i>	Chronic atypical neutrophilic dermatosis with lipodystrophy and elevated temperature—Nakajo Nishimura syndrome (CANDLE/PRAAS)	AR
<i>PSMA3</i>		
<i>PSMB4</i>		
<i>PSMB8</i>		
<i>PSMB9</i>		
<i>PSMG2</i>	Chronic atypical neutrophilic dermatosis with lipodystrophy and elevated temperature (CANDLE)	AR;AD
<i>PSTPIP1</i>	Pyogenic sterile arthritis, pyoderma gangrenosum and acne syndrome (PAPA)	AD
<i>RBCK1</i>	Polyglucosan body myopathy, early-onset, with or without immunodeficiency (PBMEI)	AR
<i>RNASEH2A</i>	Aicardi-Goutières syndrome	AR
<i>RNASEH2B</i>		
<i>RNASEH2C</i>		
<i>SAMHD1</i>	Chilblain lupus	AD

Table 1 (continued)

Gene	Disease	Mode of inheritance
<i>SH3BP2</i>	Cherubism	AD
<i>SLC29A3</i>	SLC29A3 spectrum disorder (SLC29A3)	AR
<i>TMEM173</i>	Sting-associated vasculopathy, infantile-onset (SAVI)	AD
<i>TNFAIP3</i>	Autoinflammatory syndrome, familial, Behcet-like (AISBL)	AD
<i>TNFRSF11A</i>	TNFRSF11A-associated hereditary fever disease (TRAPS11)	AD
<i>TNFRSF1A</i>	TNF receptor-associated periodic syndrome (TRAPS)	AD
<i>TREX1</i>	Systemic lupus erythematosus; Aicardi-Goutières syndrome; Chilblain lupus	AR; AD
<i>TRIM22</i>	Inflammatory bowel disease	AR
<i>TRNT1</i>	Sideroblastic anemia with B cell immunodeficiency, periodic fevers, and developmental delay	AR
<i>USP18</i>	Pseudo-TORCH syndrome	AR
<i>WDR1</i>	Autoinflammatory periodic fever, immunodeficiency, and thrombocytopenia (PFIT)	AR

AR autosomal recessive, AD autosomal dominant, XLR X-linked recessive

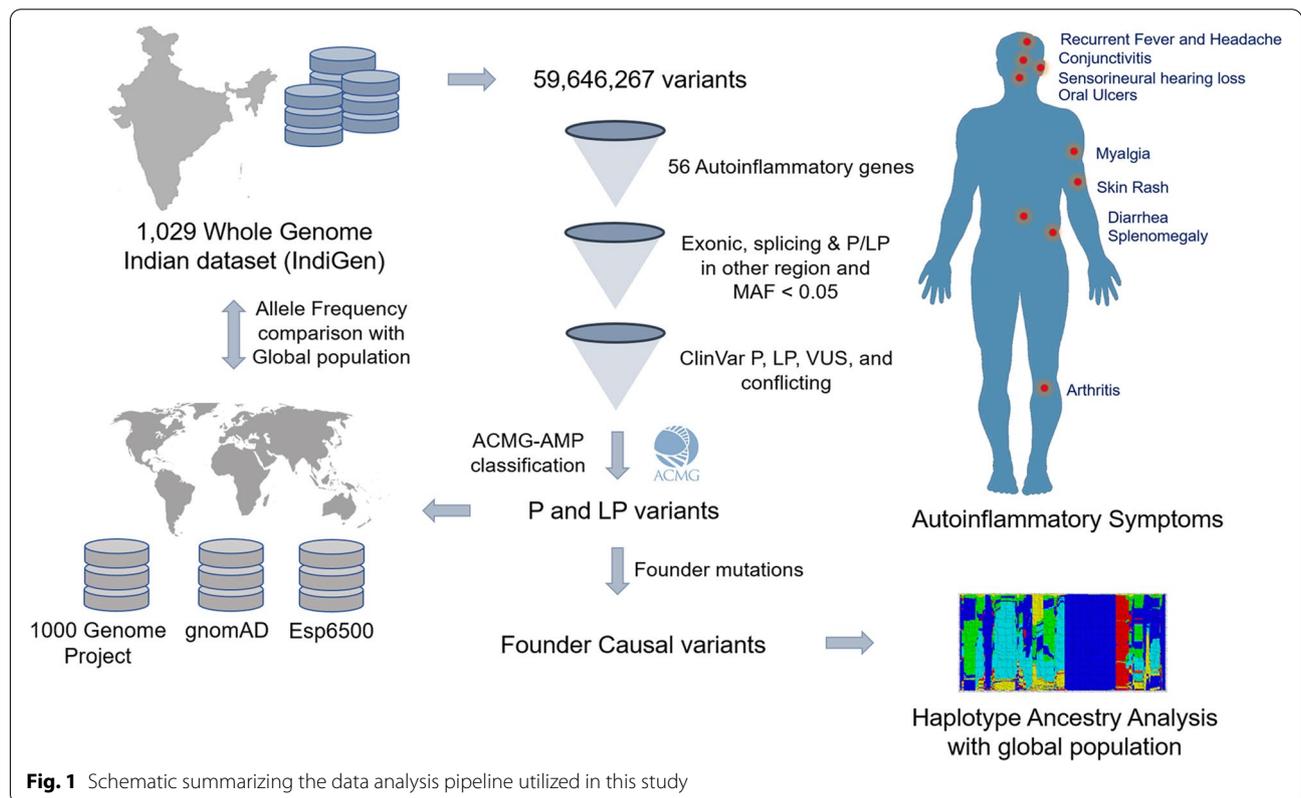


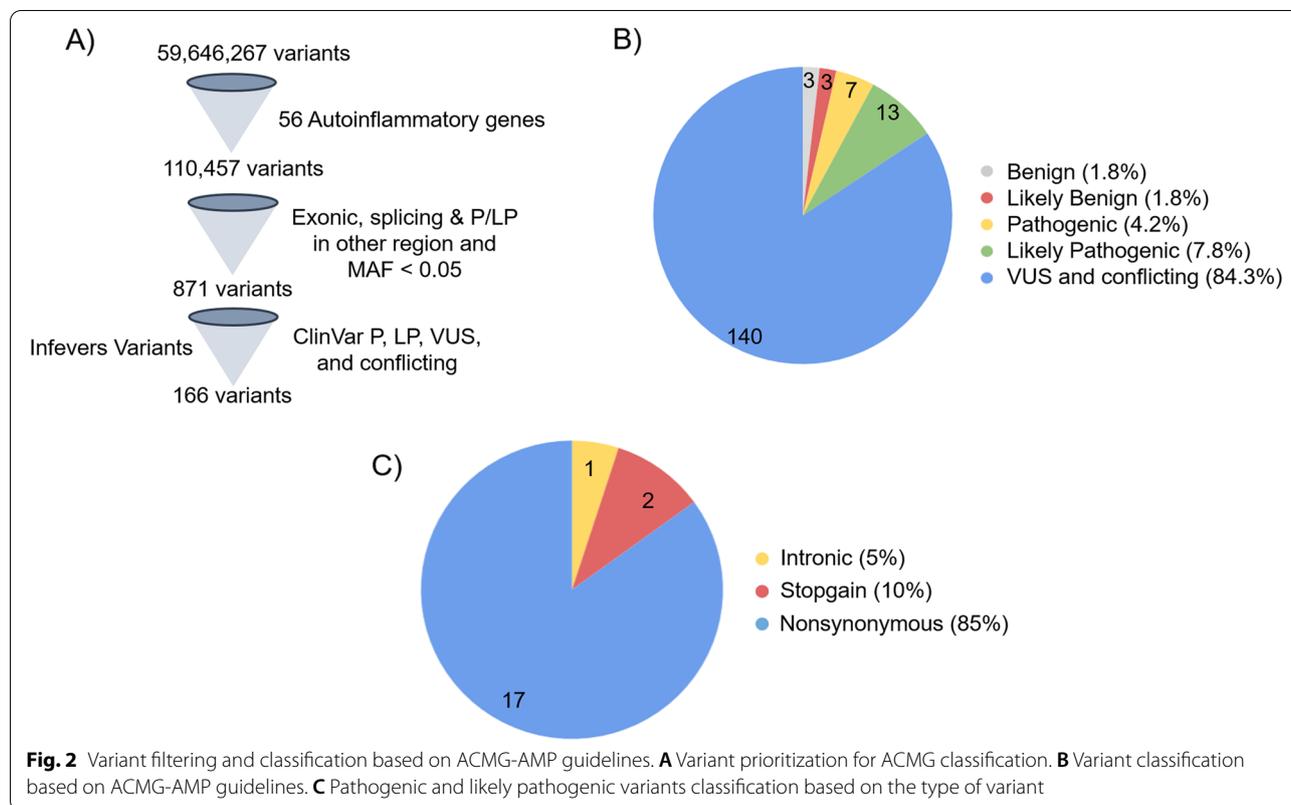
Fig. 1 Schematic summarizing the data analysis pipeline utilized in this study

Results

Genetic variants in the autoinflammatory genes from different datasets

The IndiGen dataset consisted of a total of 50,517,048 single-nucleotide variants and 5,381,074 InDels, out of which, 110,457 variants were retrieved from 56 genes associated with the 47 autoinflammatory disorders. We

further prioritized 871 exonic, splicing and ClinVar pathogenic and likely variants in other regions whose minor allele frequency is less than 0.05. Further, these variants were mapped on the merged ClinVar pathogenic, likely pathogenic, VUS, and conflicting variants and on Infevers variants that retrieved 166 variants for further analysis as shown in Fig. 2A.



Variant classification based on ACMG and AMP guidelines

These 166 variants were further scrutinized according to the ACMG-AMP guidelines and classified as pathogenic, likely pathogenic, benign, likely benign, and VUS. In our analysis, we found 7 variants as pathogenic and 13 variants as likely pathogenic (Table 2), while 6 variants were benign and likely benign, and 140 variants were VUS or variants with conflicting evidence (Fig. 2B). Out of 20 pathogenic and likely pathogenic variants, there are 1 intronic, 2 stopgain, and 17 nonsynonymous variants (Fig. 2C). The detailed variant annotation and classification are tabulated in Supplementary Table 1.

Comparison of variant frequency with the global population

Allele frequencies of 20 pathogenic or likely pathogenic variants were compared with the global populations that include gnomAD V3, 1000 Genome Project (1000g2015aug_all), and ESP6500 (Esp6500siv2_all). We found 17 out of 20 variants were significantly different from the Indian population compared with the global populations (Fisher’s exact $P < 0.05$). All the 17 variants were significantly different in comparison with the gnomAD database or its subpopulations, while 3 variants

had significantly different allele frequencies compared with the 1000 Genomes dataset and 1 variant compared with the Esp6500 European exome dataset. We could not perform Fisher’s exact test for 12 variants of the 1000 Genome Project and Esp6500 each as the allele frequencies were unavailable in the respective datasets.

In this analysis, two pathogenic variants rs144478519 and rs148755083 in the *IL36RN* gene each had 0.1% allele frequency in the Indian population (IndiGen) that were significantly less in comparison with the European population and East Asian population respectively in both 1000 Genome Project and gnomAD. Interestingly, the latter variant, rs148755083, is private to the East Asian population. Similarly, a pathogenic variant rs78635798 in the *RNASEH2C* gene had 0.05% allele frequency in the IndiGen dataset and was present only in the South Asian population. A likely pathogenic variant rs104895492 in the *NOD2* gene with high allele frequency in Indian and South Asian population dataset of 0.15 and 0.16%, while absent in other populations. Another likely pathogenic variant rs116107386 in the *APIS3* gene whose allele frequency is significantly low in the Indian population (i.e., 0.05% in comparison with the four populations of the gnomAD that includes Amish, European (Finnish and Non-Finnish), and Latino as well as European population

Table 2 Pathogenic and likely pathogenic variants classified by ACMG-AMP guidelines

Gene	Variant	SNPID	Protein change	ACMG criteria	Classification	Inheritance	Disease
ADAR	chr1:154602065:G>C	rs145588689	c.C577G: p.P193A	PM1, PP3, PM3, PP4	Likely pathogenic 5	AR	AGS
IL36RN	chr2:113060943:T>C	rs148755083	c.115+6T>C: p.Arg10ArgfsX1	PS2, PS3, PP5	Pathogenic 2	AR	DITRA
IL36RN	chr2:113062547:C>T	rs144478519	c.C338T: p.S113L	PM1, PP3, PS3, PM3, PP1, PP4	Pathogenic 3b	AR	DITRA
AP1S3	chr2:223777862:A>C	rs116107386	c.T11G: p.F4C	PM1, PP3, PS3, BS4	Likely pathogenic 2	AD	Pustular psoriasis
RNASEH2C	chr11:65720385:G>A	rs78635798	c.C205T: p.R69W	PM1, PM2, PP3, PP5, PS3, PP1-M	Pathogenic 3a	AR	AGS
MVK	chr12:109586107:A>G	rs104895364	c.A613G: p.N205D	PM1, PM2, PP3, PP4	Likely pathogenic 5	AR	HIDS
MVK	chr12:109596515:G>A	rs28934897	c.G1129A: p.V377I	BP4, PS4, PM3, PP1-M	Likely pathogenic 2	AR	HIDS
RNASEH2B	chr13:50945445:G>A	rs75184679	c.G529A: p.A177T	PM1, PP3, PP5, PS3, BS3, PP4	Likely pathogenic 2	AR	AGS
PSTPIP1	chr15:77031192:C>T	rs751668240	c.C850T: p.Q284X	PVS1, PM2, PP3	Pathogenic 1c	AD	PAPA
NOD2	chr16:50711206:C>T	rs2076754	c.C1295T: p.A432V	PM1, PM2, PP3, BS3, PS4	Likely pathogenic 2	AD	Blau Syndrome
NOD2	chr16:50711301:G>T	rs104895492	c.G1390T: p.G464W	PM2, PP3, PS3, PP1	Likely pathogenic 2	AD	Blau Syndrome
NOD2	chr16:50712048:C>T	rs104895440	c.C2137T: p.R713C	PM2, PP3, PS3	Likely pathogenic 2	AD	Blau Syndrome
CARD14	chr17:80184015:G>A	rs200731780	c.G452A: p.R151Q	PM2, BP4, PS3	Likely pathogenic 2	AD	CAMP5
CARD14	chr17:80198685:C>T	rs200379060	c.C1234T: p.R412X	PVS1, PM2	Likely pathogenic 1	AD	CAMP5
NLRP7	chr19:54936400:G>A	rs104895525	c.C2161T: p.R721W	PM2, PP3, PS1	Likely pathogenic 2	AR	HYDM1
NLRP12	chr19:53810605:G>A	rs199881207	c.C1054T: p.R352C	PM1, PM2, PS3, PP3	Likely pathogenic 2	AD	FACS
ADA2	chr22:17203564:G>A	rs148936893	c.C626T: p.P209L	PM1, PM2, BP4, PP5, PS3, PM3, PP1	Pathogenic 3a	AR	DADA2
ADA2	chr22:17209538:C>A	rs200930463	c.G140T: p.G47V	PM1, PM2, PP3, PP5, PM5	Likely pathogenic 4	AR	DADA2
ADA2	chr22:17209539:C>G	rs202134424-G	c.G139C: p.G47R	PM1, PM2, PP3, PS1, PS3, PS4	Pathogenic 2	AR	DADA2
ADA2	chr22:17209539:C>T	rs202134424-T	c.G139A: p.G47R	PM1, PM2, PP3, PP5, PS3, PS4	Pathogenic 2	AR	DADA2

AGS Aicardi-Goutières syndrome, DITRA deficiency of interleukin-36-receptor antagonist, HIDS hyper IgD syndrome, CAMP5 CARD14-mediated pustular psoriasis, HYDM1 hydantidiform mole, recurrent, 1, DADA2 deficiency of adenosine deaminase 2, AR autosomal recessive, AD autosomal dominant

of the 1000 Genome Project and ESP6500). An interesting likely pathogenic variant rs28934897 in the *MVK* gene also popularly known as Dutch mutation has high allele frequency in the Indian population (i.e., 0.34%, while absent in the South Asian populations of the control population database). The allele frequency, allele count, and allele number of all the pathogenic variants with a comparison with the gnomAD database and its population have been summarized in Table 3. The comparison with all three global populations 1000 Genome Project, gnomAD, and ESP6500 and also with Greater Middle East and Qatar populations 27,408,750 has been tabulated in Supplementary Table 2 and well represented in Fig. 3.

Ancestry haplotyping

We found six causal founder variants (i.e., rs28934897, rs148755083, rs202134424-T, rs202134424-G, rs200930463,

and rs78635798 in the *MVK*, *IL36RN*, *ADA2*, *ADA2*, *ADA2*, and *RNASEH2C* genes, respectively). The first three variants rs28934897, rs202134424-T, and rs148755083 were present in seven, four, and two individuals, respectively while other three variants (rs202134424-G, rs200930463, and rs78635798) were found in a single individual of the IndiGen dataset in heterozygous state. To identify the haplotype ancestry, we have applied filters of allele frequency > 1% on the merged IndiGen individual variant with the 1000 Genome Project that results on an average of 23,843, 784,354, 137,671, 137,701, 137,783, and 473,514 variants with approximately 99% genotype rate for variants rs28934897, rs148755083, rs202134424-T, rs202134424-G, rs200930463, and rs78635798, respectively. After performing chromosomal painting using fineSTRUCTURE, we identified six out of seven individuals had admixed European haplotype around the rs28934897 variant along with South Asian in four individuals and American haplotype in

Table 3 IndiGen allele frequency comparison *P* value (< 0.05) of autoimmune disorder pathogenic and likely pathogenic variants with the gnomAD (global) database

Pathogenic variant	Allele	IndiGen	gnomA D All	gnomAD								
				AMI	EUR (Fin)	EUR (Non-Fin)	Latino	AFR	ASJ	EAS	SAS	Other
chr1-154602065	Ref: G	0.9980	0.998	1.000	1.000	0.997	0.996	0.999	0.999	1.000	0.999	0.997
rs145588689	Alt: C	0.0020	0.002	0.000	0.001	0.004	0.004	0.001	0.001	0.000	0.001	0.003
ADAR	AC / AN	4 / 2050	327 / 143292	0 / 900	5 / 10484	224 / 64564	54 / 13656	34 / 42038	2 / 3324	0 / 3134	2 / 3042	6 / 2150
	P-value:		1.000	0.321	0.045*	0.334	0.238	0.099	0.211	0.02*	0.228	0.754
chr2-113062547	Ref: C	0.9990	0.997	1.000	0.994	0.996	0.997	0.999	1.000	1.000	0.999	0.997
rs144478519	Alt: T	0.0010	0.003	0.000	0.006	0.004	0.003	0.001	0.000	0.000	0.001	0.003
IL36RN	AC / AN	2 / 2048	372 / 143266	0 / 900	64 / 10476	233 / 64558	42 / 13660	25 / 42026	0 / 3320	0 / 3134	2 / 3044	6 / 2148
	P-value:		0.187	1.000	0.001*	0.054	0.115	0.359	0.146	0.156	1.000	0.290
chr2-113060943	Ref: T	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.991	1.000	1.000
rs148755083	Alt: C	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.009	0.000	0.000
IL36RN	AC / AN	2/2052	29/143252	0/900	0/10464	0/64562	0/13656	0/3324	0/3324	29/3130	0/3046	0/2152
	P-value:		0.071	1.000	0.026*	0.001*	0.017*	0.146	0.146	0.000055*	0.162	0.238
chr2-223777862	Ref: A	0.99951	0.991	0.978	0.993	0.986	0.993	0.997	0.996	1.000	0.999	0.986
rs116107386	Alt: C	0.0005	0.009	0.022	0.007	0.014	0.007	0.003	0.004	0.000	0.001	0.014
AP1S3	AC / AN	1 / 2048	1264 / 143304	20 / 900	907 / 64574	71 / 10482	94 / 13666	125 / 42030	13 / 3322	0 / 3132	4 / 3048	30 / 2150
	P-value:		7.52E-07*	8.59E-10*	3.33E-11*	0.00007*	0.00004*	0.032	0.023	0.396	0.654	3.84E-08*
chr11-65720385	Ref: G	0.99951	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	1.000
rs78635798	Alt: A	0.0005	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000
RNASEH2C	AC / AN	1 / 2038	2 / 143340	0 / 898	0 / 10482	0 / 64574	0 / 13666	0 / 42068	0 / 3322	0 / 3130	2 / 3048	0 / 2152
	P-value:		0.0415*	1.000	0.163	0.03*	0.130	0.046*	0.380	0.395	1.000	0.487
chr12-109586107	Ref: A	0.99951	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
rs104895364	Alt: G	0.0005	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
MVK	AC / AN	1 / 2048	2 / 143326	0 / 900	0 / 10478	0 / 64574	1 / 13664	1 / 42054	0 / 3322	0 / 3130	0 / 3050	0 / 2154
	P-value:		0.04169*	1.000	0.164	0.031*	0.244	0.091	0.382	0.396	0.402	0.488
chr12-109596515	Ref: G	0.99658	0.999	1.000	0.999	0.998	0.999	1.000	0.999	1.000	1.000	0.999
rs28934897	Alt: A	0.0034	0.002	0.000	0.001	0.003	0.001	0.000	0.001	0.000	0.000	0.001
MVK	AC / AN	7 / 2046	219 / 143294	0 / 898	15 / 10474	164 / 64572	18 / 13654	17 / 42046	3 / 3322	0 / 3132	0 / 3044	2 / 2152
	P-value:		0.0425*	0.109	0.075	0.375	0.037*	0.00008*	0.050	0.002*	0.002*	0.101
chr13-50945445	Ref: G	0.99951	0.999	1.000	1.000	0.998	0.997	1.000	1.000	1.000	0.999	0.998
rs75184679	Alt: A	0.0005	0.001	0.000	0.000	0.002	0.003	0.000	0.000	0.000	0.001	0.002
RNASEH2B	AC / AN	1 / 2046	200 / 143276	0 / 900	2 / 10478	141 / 64562	38 / 13654	12 / 42036	1 / 3324	0 / 3134	2 / 3034	4 / 2154
	P-value:		0.538	1.000	0.415	0.138	0.054	0.461	1.000	0.395	1.000	0.376
chr15-77031192	Ref: C	0.99951	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
rs751668240	Alt: T	0.0005	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
PSTPIP1	AC / AN	1/2046	4/143332	0/898	0/10478	1/64578	3/13664	0/42060	0/3318	0/3134	0/3050	0/2152
	P-value:		0.068	1.000	0.163	0.061	0.428	0.046*	0.382	0.395	0.402	0.487
chr16-50711206	Ref: C	0.99951	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
rs2076754	Alt: T	0.0005	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.001
NOD2	AC / AN	1 / 2048	33 / 143338	0 / 900	0 / 10480	32 / 64582	0 / 13666	0 / 42050	0 / 3324	0 / 3134	0 / 3048	1 / 2154
	P-value:		0.383	1.000	0.164	1.000	0.130	0.046*	0.381	0.395	0.402	1.000
chr16-50711301	Ref: G	0.99854	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.998	1.000
rs104895492	Alt: T	0.0015	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.000
NOD2	AC / AN	3 / 2048	5 / 143312	0 / 900	0 / 10482	0 / 64570	0 / 13662	0 / 42044	0 / 3324	0 / 3128	5 / 3050	0 / 2152

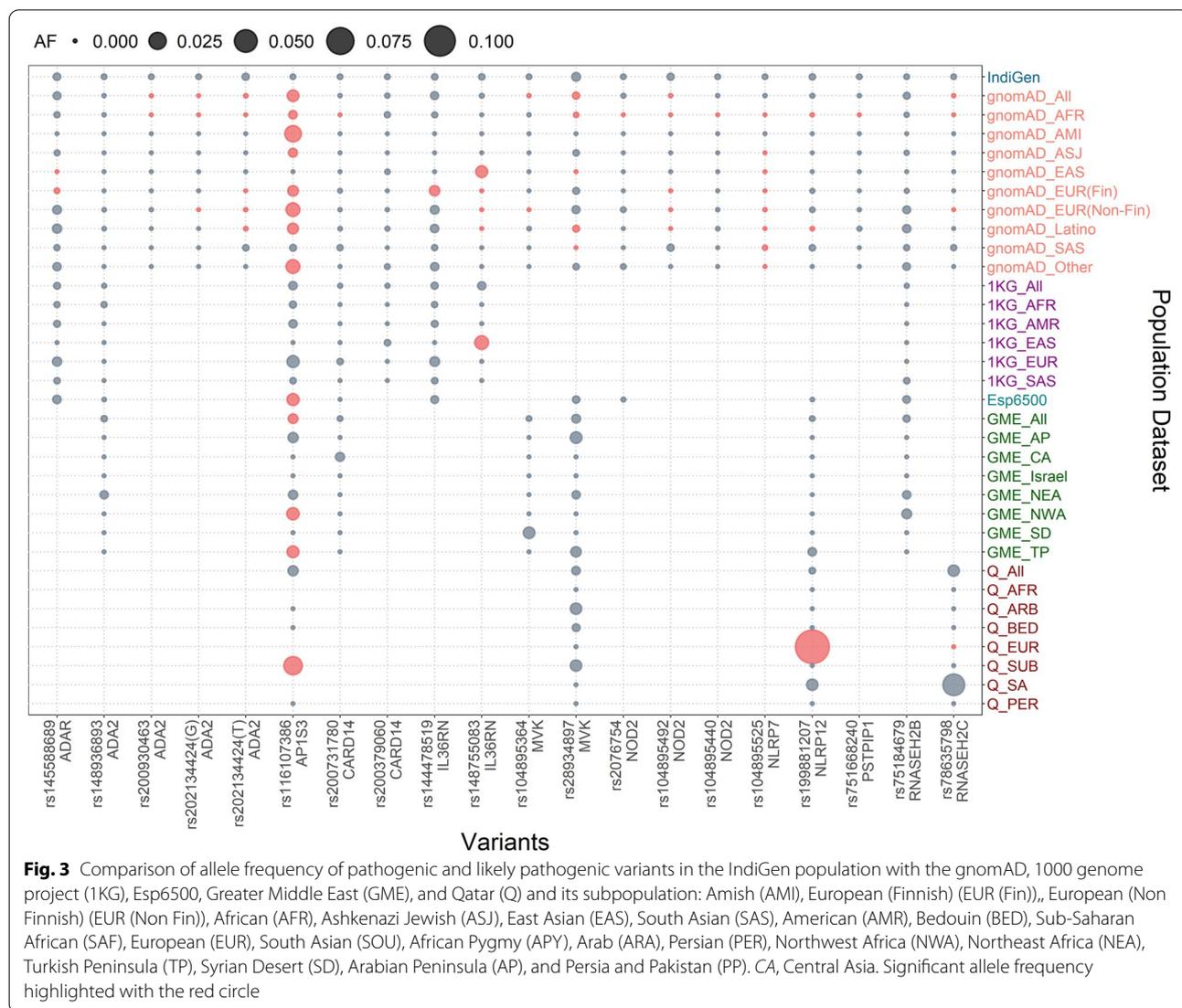
Table 3 (continued)

Pathogenic variant	Allele	IndiGen	gnomAD All	gnomAD								
				AMI	EUR (Fin)	EUR (Non-Fin)	Latino	AFR	ASJ	EAS	SAS	Other
	P-value:		0.00015*	0.558	0.004*	0.00003*	0.00222*	0.0001*	0.056	0.062	1.000	0.116
chr16-50712048	Ref: C	0.99951	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
rs104895440	Alt: T	0.0005	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
NOD2	AC / AN	1 / 2044	3 / 143366	0 / 900	0 / 10488	2 / 64592	1 / 13658	0 / 42070	0 / 3322	0 / 3134	0 / 3048	0 / 2154
	P-value:		0.055	1.000	0.163	0.089	0.244	0.046*	0.381	0.395	0.402	0.487
chr17-80184015	Ref: G	0.99951	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	1.000
rs200731780	Alt: A	0.0005	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000
CARD14	AC / AN	1 / 2046	5 / 143066	0 / 894	1 / 10436	1 / 64494	1 / 13646	0 / 41946	0 / 3324	0 / 3132	2 / 3044	0 / 2150
	P-value:		0.082	1.000	0.301	0.061	0.244	0.046*	0.381	0.395	1.000	0.488
chr17-80198685	Ref: C	0.99951	1.000	1.000	1.000	1.000	1.000	0.999	1.000	1.000	1.000	1.000
rs200379060	Alt: T	0.0005	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000
CARD14	AC / AN	1/2046	38/143348	0/900	0/10484	1/64582	2/13662	33/42056	0/3324	450,711.000	0/3052	92,774.000
	P-value:		0.425	1.000	0.163	0.061	0.342	1.000	0.381	1.000	0.401	1.000
chr19-54936400	Ref: G	0.99951	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
rs104895525	Alt: A	0.0005	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
NLRP7	AC / AN	1 / 2048	3 / 143298	0 / 900	0 / 10474	2 / 64580	0 / 13632	0 / 42058	0 / 3324	0 / 3132	1 / 3046	0 / 2152
	P-value:		0.055	1.000	0.0044*	0.0000128*	0.00004*	1.01E-08*	0.0012*	0.0006*	0.0018*	0.0008*
chr19-53810605	Ref: G	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	1.000
rs199881207	Alt: A	0.0010	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000
NLRP12	AC / AN	2 / 2044	29 / 143174	0 / 898	2 / 10480	22 / 64530	1 / 13626	1 / 42000	0 / 3320	0 / 3130	3 / 3036	0 / 2154
	P-value:		0.070	1.000	0.127	0.167	0.0466*	0.006*	0.145	0.156	1.000	0.237
chr22-17203564	Ref: G	0.99951	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
rs148936893	Alt: A	0.0005	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
ADA2	AC / AN	1 / 2050	5 / 143304	0 / 900	0 / 10478	3 / 64576	1 / 13648	1 / 42046	0 / 3324	0 / 3134	0 / 3048	0 / 2150
	P-value:		0.082	1.000	0.164	0.118	0.244	0.091	0.382	0.396	0.402	0.488
chr22-17209538	Ref: C	0.99951	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
rs200930463	Alt: A	0.0005	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
ADA2	AC / AN	1 / 2052	1 / 143248	0 / 900	0 / 10458	1 / 64574	0 / 13634	0 / 42032	0 / 3320	0 / 3130	0 / 3048	0 / 2152
	P-value:		0.0281*	1.000	0.164	0.061	0.131	0.046*	0.382	0.396	0.403	0.488
chr22-17209539	Ref: C	0.99951	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
rs202134424	Alt: G	0.0005	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
ADA2	AC / AN	1 / 2052	0 / 143248	0 / 900	0 / 10458	0 / 64574	0 / 13634	0 / 42032	0 / 3320	0 / 3130	0 / 3048	0 / 2152
	P-value:		0.0141*	1.000	0.164	0.0308*	0.131	0.046*	0.382	0.396	0.403	0.488
chr22-17209539	Ref: C	0.99854	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	1.000
rs202134424	Alt: A	0.00146	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000
ADA2	AC / AN	3 / 2052	6 / 143216	0 / 900	0 / 10458	2 / 64574	1 / 13622	0 / 42026	0 / 3318	0 / 3130	3 / 3038	0 / 2150
	P-value:		0.00022*	0.558	0.0044*	0.00027*	0.0081*	0.0001*	0.056	0.062	0.691	0.117

AMI Amish, EUR (Fin) European (Finnish), EUR (Non-Fin) European (Non-Finnish), AFR African, ASJ Ashkenazi Jewish, EAS East Asian, SAS South Asian
 Significant values are marked with asterisk (*) and cells colored in red

two individuals, while one individual had a complete South Asian haplotype. Similarly, those with variant rs202134424-G/T had admixed European haplotype in three individuals, out of which two had admixed American and one had admixed South Asian haplotype, while one individual had complete South Asian haplotype. Interestingly, a variant rs148755083 identified in two individuals in IndiGen

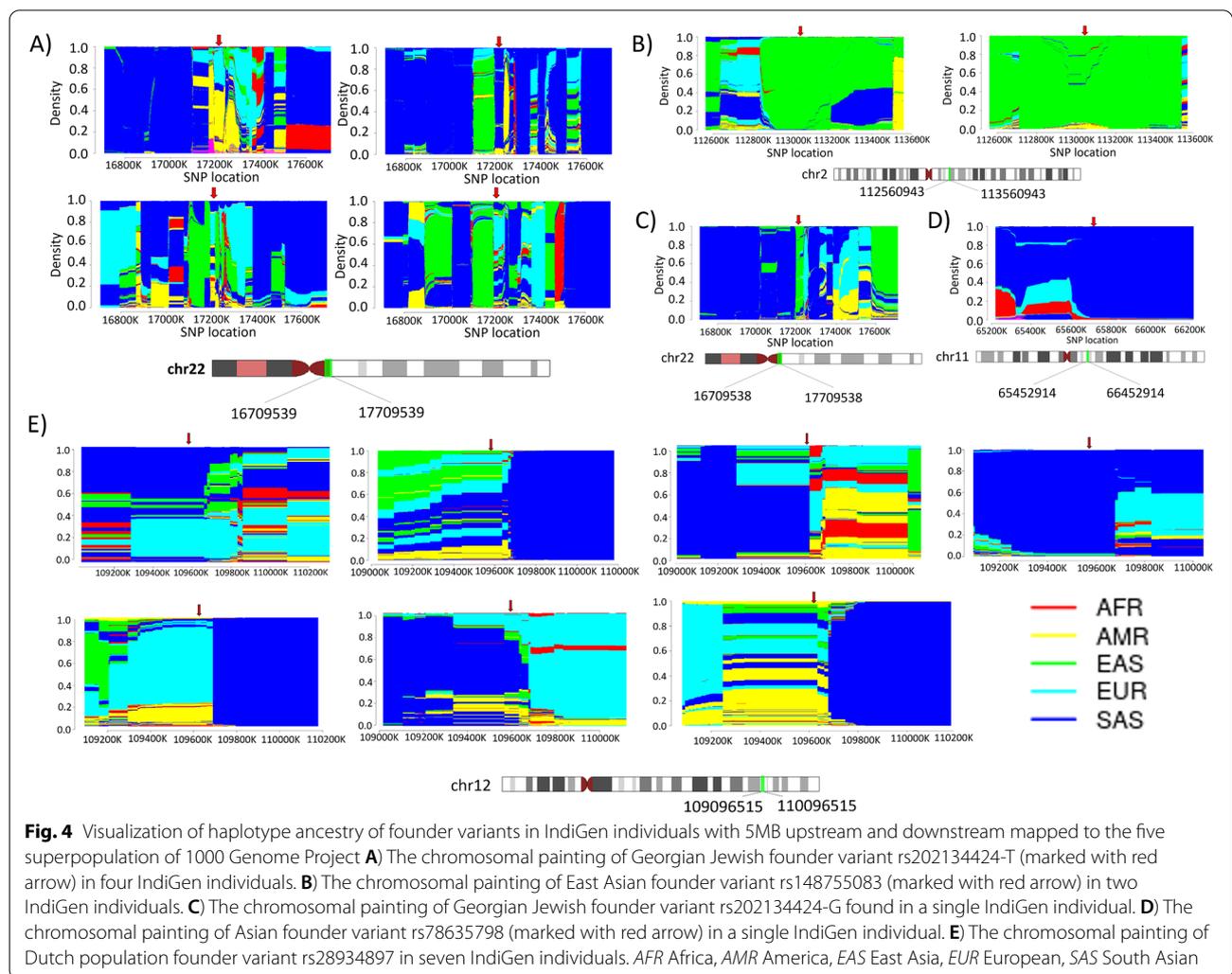
was found to have East Asian haplotype. Also, the variant rs200930463 harbored by an IndiGen individual was found to have East Asian haplotype. While an individual harboring a variant rs78635798 had a South Asian haplotype around the variant. The painted chromosomal region 500 KB upstream and downstream of the founder variant has been represented in Fig. 4.



Discussion

Autoinflammatory disorders are a group of Mendelian disorders caused by genetic defects in the genes involved in the regulation of innate immune systems. There are more than 50 genes that are associated with the autoinflammatory disorders as curated by the experts of the International Union of Immunological Societies (IUIS) [33] and Infervers [8, 34]. There are a number of genetic epidemiological studies that have been carried out around the world which suggest high prevalence of the distinct autoinflammatory disorders in different regions of the world [41–46]. The most common autoinflammatory disorder is familial Mediterranean fever (FMF) found to be very common in Middle Eastern countries [42, 44, 45]. A study performed on 1299 Armenian patients affected with FMF found a high likelihood

of carrier individuals with disease manifestation [45]. Another study revealed the carrier with high frequency of *MEFV* mutation (i.e., 1 in 5 in healthy Armenian individuals) [44]. A comprehensive study performed by our group on more than 2000 whole exome sequence dataset of the Mediterranean region revealed a carrier frequency of 8% in the population [41]. The 8-year epidemiological retrospective study for the cryopyrin-associated periodic syndrome (CAPS) caused due to the *NLRP3* gene estimated 1/360,000 prevalence in France [43]. Another study performed by Houten et al. estimated 1:65 carrier frequency of *MVK* mutation in the Dutch population associated with hyper IgD syndrome [46]. While only a handful of genetically characterized autoinflammatory diseases have been reported in India [22, 47] including from our group [48], the genetic frequencies of variants



in the population remain an enigma. The recent availability of genome-wide data for the cosmopolitan Indian population has motivated us to understand the genetic epidemiology of autoinflammatory diseases in India. Our systematic analysis of variants and reclassification according to the ACMG & AMP guidelines revealed a total of 20 genetic variants which could be classified as pathogenic or likely pathogenic.

In total, 36 genomes in the IndiGen dataset had at least one pathogenic/likely pathogenic variant. Of the individuals with carriers for any of the pathogenic/likely pathogenic variants, in the *ADA2* genes, causing deficiency of adenosine deaminase 2 (DADA2) had the maximum number of 4 unique variants in 6 individuals. The second maximum number of variants that was present in the *NOD2* gene that causes Blau syndrome and had 3 unique variants in 5 individuals. This was followed by the *MVK*, *NLRP12*, and *IL36RN* genes causing hyper IgD syndrome, (HIDS) and generalized pustular psoriasis (GPP) with

2 unique variants each, in 8 and 4 individuals, respectively. A total of 4 and 2 individuals were carriers for a causal variant in the *ADAR* and *NLRP12* gene causative for Aicardi-Goutières syndrome (AGS) and familial cold autoinflammatory syndrome (FACS). While the remaining 7 genes (*APIS3*, *RNASEH2C*, *RNASEH2B*, *CARD14*, *NLRP7*, *PSTPIP1*, and *CARD14*) have only a single individual carrier for the causal variant.

Out of 20 causal variants, 7 were pathogenic (1 intronic, 1 stopgain, and 5 nonsynonymous), and 13 are likely pathogenic variants (1 stopgain and 12 nonsynonymous) by ACMG-AMP guidelines. We identified two pathogenic variants in the *IL36RN* gene that includes a nonsynonymous variant (c.338C>T: p.S113L) rs144478519 and intronic variant rs148755083 (c.115+6T>C: p.Arg10ArgfsX1). A nonsynonymous variant rs144478519 was found in multiple unrelated patients of different ancestries affected with GPP in the homozygous or compound heterozygous state. This

variant falls in a region that is evolutionarily conserved with proximity to the binding site involved in receptor interaction. This interaction is responsible for the IL-36 signaling system for the inhibition of the activity of interleukin-36 that further inhibits the local autoinflammatory response [23]. Functional studies also revealed a significant increase in proinflammatory cytokines [49–51]. The latter intronic variant rs148755083 is predominantly found in the East Asian populations with allele frequency of ~ 1% in the global dataset. This variant has been found to be highly prevalent in patients of Japanese and Chinese ancestry and was identified as founder mutation in both the population [50, 52–54]. A nonsynonymous pathogenic variant rs78635798 (c.205C>T: p.R69W) in the *RNASEH2C* gene causative of AGS had very low allele frequency in the global populations (i.e., < 0.01%) and falls in the functionally important domain. Functional studies revealed *RNASEH2C* p.R69W had a significant reduction in the thermal stability of RNase H2 complex [55]. This variant has been recurrent and considered as a founder mutation in the Asian population as well; it segregates with the disease in the family [56]. Other founder pathogenic variants from the Georgian Jewish population (p.G47R) rs200930463-T and rs200930463-G in the *ADA2* gene cause DADA2 with the carrier frequency in this population as 10.2% with the high prevalence of the disease [57]. However, in the global population datasets, it has a very low allele frequency of < 0.0002%. The functional analysis showed a marked reduction of ADA2 activity in comparison with the wild type in the homozygous state [57, 58]. Recently, a case series of 33 DADA2-affected patients from India has been reported and found this variant p.G47R to be prevalent in the Jain/Aggarwal community [59]. Another *ADA2* variant at the same amino acid position of the ADA2 protein (p.G47V) rs200930463 was in a trans compound heterozygous state with p.W246S in the patient affected with DADA2 [57]. Also, functional studies revealed the complete absence of ADA2 protein in cells transfected with p.G47V and p.W246S as well as lower amounts in drosophila S2 cells [57]. The pathogenic nonsynonymous variant rs148936893 (c.C752T: p.P251L) in the *ADA2* gene associated with the DADA2 has a low frequency of < 0.02% in the global population datasets. This variant was found to be segregated with the disease in a German family in a trans compound heterozygous state. Functional studies revealed ADA2 activity has been severely compromised and also indicated intracellular elevated levels of ADA2 protein [57].

Another founder mutation of Dutch ancestry (c.G1129A: p.V377I) rs28934897 in the *MVK* gene causes HIDS with very high carrier frequency (i.e.,

1:65) [46]. This variant has been reported in multiple patients with HIDS of different ancestry and found variant segregation with the disease in the family either in compound heterozygous or homozygous state [60–62]. In vivo and in vitro functional studies have revealed a significant decrease in the enzymatic activity of MVK protein [63]. Recently, we have found a p.V377I variant in our six patients of South India ancestry in a trans compound heterozygous state residing in the same region and identified as founder in the South Indian population [48]. Another likely pathogenic *MVK* variant rs104895364 (c.613A>G: p.N205D) was found in a trans compound heterozygous state in two patients affected with HIDS of the same family [48]. It has a very low allele frequency of < 0.005% in the global population datasets (i.e., 1000 Genome Project, gnomAD, and Esp6500). The variant falls in the functionally important domain and was found in multiple patients affected with HIDS of different ancestry [64–67]. A likely pathogenic nonsynonymous variant rs145588689 (c.577C>G: p.P193A) in the *ADAR* gene associated with Aicardi-Goutières syndrome (AGS) was found to be segregated with disease in compound heterozygous state in 22 out of 23 unrelated families as well as in multiple unrelated AGS patients [68, 69]. This variant falls in the functional domain that disrupts the interaction between Z-DNA/Z-RNA binding thus upregulating IFN-stimulated genes [70]. Another likely pathogenic variant rs2076754 (c.C1295T: p.A432V) in the *NOD2* gene associated with Blau syndrome/ Crohn's disease does not have any significant difference in comparison with the wild type [71]. However, it has a low allele frequency of \leq 0.02% in the global population dataset, and also the odds of occurring in Crohn's disease patients is significantly higher than the controls [72]. Another nonsynonymous likely pathogenic mutation (c.G1390T: p.G464W) rs104895492 in the *NOD2* gene has a low allele frequency of 0.003% in the global population. A reporter assay revealed this mutation to cause hyperactivity of NOD2-mediated NF-KB signaling in the absence of ligands [73]. The variant segregates with the disease in the mother and daughter, both were affected with Blau syndrome [74]. Another likely pathogenic nonsynonymous variant rs104895440 (c.C2137T: p.R713C) in the *NOD2* gene has very low allele frequency of < 0.003% in the global population datasets. In vitro functional studies of the variant revealed major impairment of the peptidoglycan-induced response [71]. A nonsynonymous mutation (c.C1054T: p.R352C) in the *NLRP12* gene found in two patients suffering from familial cold autoinflammatory syndrome (FCAS) of different ancestries. Functional study of the mutation revealed the increase in the speck formation as well as activation

of the caspase 1 signaling in comparison with the wild type [11]. A likely pathogenic nonsynonymous variant rs200731780 (c.452G>A: p.R151Q) in the *CARD14* gene associated with *CARD14*-mediated pustular psoriasis (CAMPS) that has been predicted as benign by computational prediction tools. However, in vitro functional studies of p.R151Q have shown a significant increase (i.e., 18 folds in the NF-KB activation that leads to CAMPS) [75, 76]. It has a very low allele frequency of < 0.03% in the global population datasets (i.e., 1000 Genome Project, gnomAD, and Esp6500). Another likely pathogenic variant rs116107386 (c.11T>G: p.F4C) in the *APIS3* gene associated with pustular psoriasis has a frequency of ~ 1% in the global population. However, in transfection studies in HEK293 cells, it showed a significant decrease of the mutant protein in comparison with the wild type that leads to marked inhibition of downstream signaling. The allele frequency was also found to be significantly higher in the patient than controls and falls in functionally important domains [77].

In this study, we have prioritized six causal founder variants. This includes rs28934897, rs148755083, rs202134424-T, rs202134424-G, rs200930463, and rs78635798 in the *MVK*, *IL36RN*, *ADA2*, *ADA2*, *ADA2*, and *RNASEH2C* genes, respectively. The variant (rs28934897) in the *MVK* gene is harbored in seven individuals of the IndiGen dataset in heterozygous state. On performing haplotype ancestry analysis, we found six out of seven individuals had admixed European ancestry. The occurrence of the European haplotype at the founder variant in the Indian population could be due to the invasion or migration of the Europeans in India [78]. The founder Dutch mutation p.V377I (rs28934897) along with the splicing mutation c.226+2delT in the *MVK* gene in trans compound heterozygous state was found to be more common in the Indian population than reported [48]. An intronic variant c.115+6T>C (rs148755083) in the gene *IL36RN* implicated in GPP was predominantly identified and considered as a founder variant in the East Asian population [50, 52–54]. We have identified this variant in two IndiGen individuals belonging to the Eastern part of India and was found to have East Asian ancestry. Since it has high frequency in the Eastern parts of India, variants could be screened, and the government could take proactive measures. We have also identified three variants p.G47V (rs202134424-T), p.G47V (rs202134424-G), and G47R (rs200930463) in the *ADA2* gene implicated in DADA2 known to be founder with high frequency in the Georgian Jewish population [57]. In IndiGen, out of the five individuals harboring these variants, three of them had European admixture, one East Asian, and one South Asian ancestry. A recent study performed in Indian DADA2 patients identified p.G47R as a founder

variant in the Aggarwal/Jain community [59]. The Aggarwal community mainly resides in North India and is a descendant of the Indo-European migrants that had high frequency of the *ADA2* causal variant [59, 79]. The *RNASEH2C* variant p.R69W (rs78635798) implicated in AGS found in a single individual of IndiGen had South Asian haplotype ancestry. This variant was considered as a founder variant in the Asian populations [56].

In contrast with similar approaches towards understanding the genetic epidemiology of autoinflammatory diseases in other populations, three of the causal variants in the present analysis overlapped with our previous analysis of the Middle Eastern population [21]. Compared to the variant frequencies in the global populations, a number of disease alleles have frequencies in Indian population higher than global datasets including rs116107386, rs78635798, rs104895364, rs28934897, rs104895492, rs200930463, and rs202134424. Similarly, a number of variants have allele frequencies less than global populations like rs148755083 in *IL36RN* for pustular psoriasis. We surmise that the allele frequencies also correlate with the frequency of variants in clinical settings. For example, a recent case series of HIDS from our group suggests the clinical and genetic characteristics of patients. Incidentally, the prevalent variant rs148755083 with a founder effect was also the frequency variant identified in the case series [48].

The study has many caveats, the major being the dataset encompasses only a limited sample of cosmopolitan Indians, and therefore might not adequately cover smaller endogamous and ethnic groups. Secondly, the annotation of variants based on evidence from already proven cases, and therefore preclude novel and potentially pathogenic genetic variants.

Conclusions

The present analysis of genomes suggests that a number of autoinflammatory disease variants are prevalent in India. A subset of the variants were founders and were mainly descendants of different ancestry (i.e., European and East Asian due to migration or invasion in India). The causal founder autoinflammatory variant had high frequency with respect to their geographical regions or community. That could be considered a hotspot variant for the distinct population. The respective government could undertake the initiative and could perform the low-cost population screening so that it could provide better health care facilities to the population.

Abbreviations

VUS: Variant of uncertain significance; ACMG-AMP: American College of Medical Genetics and Genomics and Association of Molecular Pathology; FMF: Familial Mediterranean fever; CAPS: Cryopyrin-associated periodic syndrome;

GME: Greater Middle East; IUIS: International Union of Immunological Societies; DADA2: Deficiency of adenosine deaminase 2; HIDS: Hyper IgD syndrome; GPP: Generalized pustular psoriasis; AGS: Aicardi-Goutières syndrome; FCAS: Familial cold autoinflammatory syndrome; CAMPS: CARD14-mediated pustular psoriasis.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s43141-021-00268-2>.

Additional file 1: Supplementary Data 1. Detailed description of 28 ACMG-AMP guidelines for variant classification.

Additional file 2: Supplementary Table 1. Autoinflammatory variants annotation and their classification according to the ACMG-AMP guidelines.

Additional file 3: Supplementary Table 2. IndiGen allele frequency comparison p -value (<0.05) of autoinflammatory disorder pathogenic and likely pathogenic variants with the global databases included gnomAD V3, 1000 Genome Project, Esp6500, GME, and Qatar with their subpopulation. AML: Amish, EUR (Fin): European (Finnish), EUR (Non Fin): European (Non Finnish), AFR: African, ASJ: Ashkenazi Jewish, EAS: East Asian, SAS: South Asian, AMR: American, BED Bedouin, SAF Sub-Saharan African, EUR European, SOU South Asian, APY African Pygmy, ARA ARAB, PER Persian, NWA: Northwest Africa, NEA: Northeast Africa, TP: Turkish Peninsula, SD: Syrian Desert, AP: Arabian Peninsula, and PP: Persia and Pakistan, NA not applicable. Significant values are marked with * and cells colored in red.

Acknowledgments

The authors acknowledge the volunteers who participated in this study. Authors acknowledge Mukta Poojary and Vishu Gupta for constructive suggestions and Sr. B. Velangini, Principal, St. Pious X Degree & PG College for Women, Hyderabad for encouragement.

Authors' contributions

AJ was involved in the ACMG analysis, manuscript writing, and editing. RCB, AM, MI, VSen, MKD, SSharma, and SSiwach were involved in the whole genome sequencing. AJ, KV, DS, MR, and BJ were involved in the analysis of the whole genome sequencing. VScarria and SSB designed the study and were involved in the manuscript editing. The other authors were involved in the collection of the samples and formal analysis. The authors read and approved the final manuscript.

Funding

The study was funded by the Council of Scientific and Industrial Research (CSIR), India (MLP1809, MLP2001), CSIR fellowship (to A.J., M.K.D., B.J., A.B., S.M., R.Y.), Intel Research Fellowship (to D.S.), and UGC fellowship (to V.G., S.G.). Funding for open access charge: Council of Scientific and Industrial Research, India (MLP1809, MLP2001).

Availability of data and materials

All data generated or analyzed during this study are included in this published article (and its Supplementary information files).

Declarations

Ethics approval and consent to participate

This study was approved by the Institutional Human Ethics Committee (IHEC) (ethics approval no. CSIR-IgIB/IHEC/2018-19 Dt 21/02/2019) of CSIR-Institute of Genomics and Integrative Biology. The participants were explained about the informed consent process as per the approved IHEC guidelines and obtained written consent.

Consent for publication

A well informed consent has been obtained regarding publication from the participants.

Competing interests

The authors declare that they have no competing interests.

Author details

¹CSIR-Institute of Genomics and Integrative Biology, New Delhi 110025, India. ²Academy of Scientific and Innovative Research (AcSIR), Ghaziabad, Uttar Pradesh 201002, India. ³Department of Zoology, RTM Nagpur University, Nagpur, Maharashtra 440033, India. ⁴Department of Zoology, Shri Mathuradas Mohota College of Science, Nagpur, Maharashtra 440009, India. ⁵Department of Anthropology, Vinoba Bhave University, Hazaribag, Jharkhand 825301, India. ⁶Department of Biotechnology, Vinoba Bhave University, Hazaribag, Jharkhand 825301, India. ⁷Department of Biochemistry, Dr. Kongu Science and Art College, Erode, Tamil Nadu 638107, India. ⁸Thalassemia and Sickle Cell Society, Hyderabad, Telangana 500052, India. ⁹Department of Biotechnology, Mizoram University, Aizawl, Mizoram 796004, India. ¹⁰Department of Pathology, Civil Hospital Aizawl, Mizoram 796001, India. ¹¹TVMC, Tirunelveli Medical College, Tirunelveli, Tamil Nadu 627011, India. ¹²Department of Microbiology, St.Pious X Degree & PG College for Women, Hyderabad, Telangana 500076, India. ¹³Department of Microbiology, The Maharaja Sayajirao University of Baroda, Vadodara, Gujarat 390002, India. ¹⁴Department of Dermatology, Venereology and Leprology, Regional Institute of Medical Sciences, Imphal, Manipur 795004, India. ¹⁵CSIR- North East Institute of Science and Technology, Jorhat, Assam 785006, India. ¹⁶Department of Dermatology, Dr. D.Y. Patil Medical College, Pune, Maharashtra 411018, India.

Received: 26 July 2021 Accepted: 13 October 2021

Published online: 14 December 2021

References

- Rowczenio D, Shinar Y, Ceccherini I, Sheils K, Van Gijn M, Patton SJ et al (2019) Current practices for the genetic diagnosis of autoinflammatory diseases: results of a European Molecular Genetics Quality Network Survey. *Eur J Hum Genet* 27:1502–1508
- Hausmann JS, Lomax KG, Shapiro A, Durrant K (2018) The patient journey to diagnosis and treatment of autoinflammatory diseases. *Orphanet J Rare Dis* 13:156
- Federici S, Sormani MP, Ozen S, Lachmann HJ, Amaryan G, Woo P et al (2015) Evidence-based provisional clinical classification criteria for autoinflammatory periodic fevers. *Ann Rheum Dis* 74:799–805
- Jia T, Zheng Y, Feng C, Yang T, Geng S (2020) A Chinese case of Nakajima-Nishimura syndrome with novel compound heterozygous mutations of the PSM88 gene. *BMC Med Genet* 21:126
- Jia X, Shi N, Feng Y, Li Y, Tan J, Xu F et al (2020) Identification of 67 pleiotropic genes associated with seven autoimmune/autoinflammatory diseases using multivariate statistical analysis. *Front Immunol* 11:30
- Chear CT, Nallusamy R, Canna SW, Chan KC, Baharin MF, Hishamshah M et al (2020) A novel de novo NLR4 mutation reinforces the likely pathogenicity of specific LRR domain mutation. *Clin Immunol* 211:108328
- Tao P, Sun J, Wu Z, Wang S, Wang J, Li W et al (2020) A dominant autoinflammatory disease caused by non-cleavable variants of RIPK1. *Nature* 577:109–114
- Touitou I, Lesage S, McDermott M, Cuisset L, Hoffman H, Dode C et al (2004) Infevers: an evolving mutation database for auto-inflammatory syndromes. *Hum Mutat* 24:194–198
- Hashkes PJ, Laxer RM, Simon A (2019) Textbook of autoinflammation. Springer, Cham. <https://doi.org/10.1007/978-3-319-98605-0>
- Rusmini M, Federici S, Caroli F, Grossi A, Baldi M, Obici L et al (2016) Next-generation sequencing and its initial applications for molecular diagnosis of systemic auto-inflammatory diseases. *Ann Rheum Dis* 75:1550–1557
- Jéru I, Le Borgne G, Cochet E, Hayrapetyan H, Duquesnoy P, Grateau G et al (2011) Identification and functional consequences of a recurrent NLRP12 missense mutation in periodic fever syndromes. *Arthritis Rheum* 63:1459–1464
- Hoffman HM, Mueller JL, Broide DH, Wanderer AA, Kolodner RD (2001) Mutation of a new gene encoding a putative pyrin-like protein causes familial cold autoinflammatory syndrome and Muckle-Wells syndrome. *Nat Genet* 29:301–305

13. Tsimaratos M, Kone-Paut I, Divry P, Philip N, Chabrol B (2001) Mevalonic aciduria and hyper-IgD syndrome: two sides of the same coin? *J Inher Metab Dis* 24:413–414
14. Schnappauf O, Aksentijevich I (2019) Current and future advances in genetic testing in systemic autoinflammatory diseases. *Rheumatology* 58:vi44–vi55
15. Biesecker LG, Shianna KV, Mullikin JC (2011) Exome sequencing: the expert view. *Genome Biol* 12:128
16. Chou J, Ohsumi TK, Geha RS (2012) Use of whole exome and genome sequencing in the identification of genetic causes of primary immunodeficiencies. *Curr Opin Allergy Clin Immunol* 12:623–628
17. Sheppard S, Biswas S, Li MH, Jayaraman V, Slack I, Romasko EJ et al (2018) Utility and limitations of exome sequencing as a genetic diagnostic tool for children with hearing loss. *Genet Med* 20:1663–1676
18. Belkadi A, Bolze A, Itan Y, Cobat A, Vincent QB, Antipenko A et al (2015) Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. *Proc Natl Acad Sci U S A* 112:5473–5478
19. Jain A, Govindaraj GM, Edavazhippurath A, Faisal N, Bhojar RC, Gupta V et al (2021) Whole genome sequencing identifies novel structural variant in a large Indian family affected with X-linked agammaglobulinemia. *PLoS One* 16:e0254407
20. Thaventhiran JED, Lango Allen H, Burren OS, Rae W, Greene D, Staples E et al (2020) Whole-genome sequencing of a sporadic primary immunodeficiency cohort. *Nature* 583:90–95
21. Sharma P, Jain A, Scaria V (2021) Genetic landscape of rare autoinflammatory disease variants in Qatar and Middle Eastern populations through the integration of whole-genome and exome datasets. *Front Genet* 12:631340
22. Kumar BS, Kumar PS, Sowgandhi N, Prajwal BM, Mohan A, Sarma KVS et al (2016) Identification of novel mutations in CD2BP1 gene in clinically proven rheumatoid arthritis patients of South India. *Eur J Med Genet* 59:404–412
23. Jain A, Bhojar RC, Pandhare K, Mishra A, Sharma D, Imran M et al (2021) IndiGenomes: a comprehensive resource of genetic variants from over 1000 Indian genomes. *Nucleic Acids Res* 49:D1225–D1232
24. Wang K, Li M, Hakonarson H (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 38:e164
25. Pruitt KD, Tatusova T, Maglott DR (2007) NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res* 35:D61–D65
26. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM et al (2001) dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* 29:308–311
27. Liu X, Wu C, Li C, Boerwinkle E (2016) dbNSFP v3.0: a one-stop database of functional predictions and annotations for human nonsynonymous and splice-site SNVs. *Hum Mutat* 37:235–241
28. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q et al (2020) The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* 581:434–443
29. 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM et al (2015) A global reference for human genetic variation. *Nature* 526:68–74
30. Exome Variant Server n.d. <http://evs.gs.washington.edu/EVS/> (Accessed 28 Mar 2021).
31. Scott EM, Halees A, Itan Y, Spencer EG, He Y, Azab MA et al (2016) Characterization of Greater Middle Eastern genetic variation for enhanced disease gene discovery. *Nat Genet* 48:1071–1076
32. Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S et al (2018) ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res* 46:D1062–D1067
33. Tangye SG, Al-Herz W, Bousfiha A, Chatila T, Cunningham-Rundles C, Etzioni A et al (2020) Human inborn errors of immunity: 2019 update on the classification from the International Union of Immunological Societies Expert Committee. *J Clin Immunol* 40:24–64
34. Milhavel F, Cuisset L, Hoffman HM, Slim R, El-Shanti H, Aksentijevich I et al (2008) The Infevers autoinflammatory mutation online registry: update with new genes and functions. *Hum Mutat* 29:803–808
35. Infevers <http://fmf.igh.cnrs.fr/ISSAID/infevers/index.php> (Accessed 28 Mar 2021).
36. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J et al (2015) 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
37. Genetic Variant Interpretation Tool https://www.medschool.umaryland.edu/Genetic_Variant_Interpretation_Tool1.html/ (Accessed 28 Mar 2021).
38. Exome Variant Server <http://evs.gs.washington.edu/EVS/> (Accessed 28 Mar 2021).
39. Lawson DJ, Hellenthal G, Myers S, Falush D (2012) Inference of population structure using dense haplotype data. *PLoS Genet* 8:e1002453
40. Delaneau O, Coulonges C, Zagury J-F (2008) Shape-IT: new rapid and accurate algorithm for haplotype inference. *BMC Bioinformatics* 9:540
41. Koshy R, Sivadas A, Scaria V (2018) 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
42. Cush JJ (2013) Autoinflammatory syndromes. *Dermatol Clin* 31:471–480
43. Cuisset L, Jeru I, Dumont B, Fabre A, Cochet E, Le Bozec J et al (2011) Mutations in the autoinflammatory cryopyrin-associated periodic syndrome gene: epidemiological study and lessons from eight years of genetic analysis in France. *Ann Rheum Dis* 70:495–499
44. Sarkisian T, Ajrapetian H, Beglarian A, Shahsuvarian G, Egiazarian A (2008). Familial Mediterranean fever in Armenian population. *Georgian Med News* 105–11.
45. Moradian MM, Sarkisian T, Ajrapetyan H, Avanesian N (2010) Genotype-phenotype studies in a large cohort of Armenian patients with familial Mediterranean fever suggest clinical disease with heterozygous MEFV mutations. *J Hum Genet* 55:389–393
46. Houten SM, van Woerden CS, Wijburg FA, Wanders RJA, Waterham HR (2003) 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
47. Sandhya P, Vellarikkal SK, Nair A, Ravi R, Mathew J, Jayarajan R et al (2017) Egyptian tale from India: application of whole-exome sequencing in diagnosis of atypical familial Mediterranean fever. *Int J Rheum Dis* 20:1770–1775
48. Govindaraj GM, Jain A, Peethambaran G, Bhojar RC, Vellarikkal SK, Ganapati A et al (2020) Spectrum of clinical features and genetic variants in mevalonate kinase (MVK) gene of South Indian families suffering from hyperimmunoglobulin D syndrome. *PLoS One* 15:e0237999
49. Körber A, Mössner R, Renner R, Sticht H, Wilsmann-Theis D, Schulz P et al (2013) Mutations in IL36RN in patients with generalized pustular psoriasis. *J Invest Dermatol* 133:2634–2637
50. Setta-Kaffetzi N, Navarini AA, Patel VM, Pullabhatla V, Pink AE, Choon S-E et al (2013) Rare pathogenic variants in IL36RN underlie a spectrum of psoriasis-associated pustular phenotypes. *J Invest Dermatol* 133:1366–1369
51. Onoufriadis A, Simpson MA, Pink AE, Di Meglio P, Smith CH, Pullabhatla V et al (2011) Mutations in IL36RN/IL1F5 are associated with the severe episodic inflammatory skin disease known as generalized pustular psoriasis. *Am J Hum Genet* 89:432–437
52. Li M, Han J, Lu Z, Li H, Zhu K, Cheng R et al (2013) Prevalent and rare mutations in IL-36RN gene in Chinese patients with generalized pustular psoriasis and psoriasis vulgaris. *J Invest Dermatol* 133:2637–2639
53. Farooq M, Nakai H, Fujimoto A, Fujikawa H, Matsuyama A, Kariya N et al (2013) Mutation analysis of the IL36RN gene in 14 Japanese patients with generalized pustular psoriasis. *Hum Mutat* 34:176–183
54. Sugiura K, Takemoto A, Yamaguchi M, Takahashi H, Shoda Y, Mitsuma T et al (2013) The majority of generalized pustular psoriasis without psoriasis vulgaris is caused by deficiency of interleukin-36 receptor antagonist. *J Invest Dermatol* 133:2514–2521
55. Reijns MAM, Bubeck D, Gibson LCD, Graham SC, Baillie GS, Jones EY et al (2011) The structure of the human RNase H2 complex defines key interaction interfaces relevant to enzyme function and human disease. *J Biol Chem* 286:10530–10539
56. Vogt J, Agrawal S, Ibrahim Z, Southwood TR, Philip S, Macpherson L et al (2013) Striking intrafamilial phenotypic variability in Aicardi-Goutières syndrome associated with the recurrent Asian founder mutation in RNASEH2C. *Am J Med Genet A* 161A:338–342
57. Navon Elkan P, Pierce SB, Segel R, Walsh T, Barash J, Padeh S et al (2014) Mutant adenosine deaminase 2 in a polyarteritis nodosa vasculopathy. *N Engl J Med* 370:921–931

58. Zhou Q, Yang D, Ombrello AK, Zavialov AV, Toro C, Zavialov AV et al (2014) Early-onset stroke and vasculopathy associated with mutations in ADA2. *N Engl J Med* 370:911–920
59. Sharma A, Naidu G, Sharma V, Jha S, Dhooria A, Dhir V et al (2021) Deficiency of adenosine deaminase 2 in adults and children: experience from India. *Arthritis Rheumatol* 73:276–285
60. Houten SM, Koster J, Romeijn GJ, Frenkel J, Di Rocco M, Caruso U et al (2001) Organization of the mevalonate kinase (MVK) gene and identification of novel mutations causing mevalonic aciduria and hyperimmunoglobulinemia D and periodic fever syndrome. *Eur J Hum Genet* 9:253–259
61. Drenth JP, Cuisset L, Grateau G, Vasseur C, van de Velde-Visser SD, de Jong JG et al (1999) Mutations in the gene encoding mevalonate kinase cause hyper-IgD and periodic fever syndrome. *International Hyper-IgD Study Group. Nat Genet* 22:178–181
62. Houten SM, Kuis W, Duran M, de Koning TJ, van Royen-Kerkhof A, Romeijn GJ et al (1999) Mutations in MVK, encoding mevalonate kinase, cause hyperimmunoglobulinemia D and periodic fever syndrome. *Nat Genet* 22:175–177
63. Cuisset L, Drenth JP, Simon A, Vincent MF, van der Velde VS, van der Meer JW et al (2001) Molecular analysis of MVK mutations and enzymatic activity in hyper-IgD and periodic fever syndrome. *Eur J Hum Genet* 9:260–266
64. Tanaka T, Yoshioka K, Nishikomori R, Sakai H, Abe J, Yamashita Y et al (2019) National survey of Japanese patients with mevalonate kinase deficiency reveals distinctive genetic and clinical characteristics. *Mod Rheumatol* 29:181–187
65. Galeotti C, Meinzer U, Quartier P, Rossi-Semerano L, Bader-Meunier B, Pillet P et al (2012) Efficacy of interleukin-1-targeting drugs in mevalonate kinase deficiency. *Rheumatology* 51:1855–1859
66. Mandey SHL, Schneiders MS, Koster J, Waterham HR (2006) Mutational spectrum and genotype-phenotype correlations in mevalonate kinase deficiency. *Hum Mutat* 27:796–802
67. Çakan M, Aktay-Ayaz N, Keskindemirci G, Karadağ ŞG (2017) Two cases of periodic fever syndrome with coexistent mevalonate kinase and Mediterranean fever gene mutations. *Turk J Pediatr* 59:467–470
68. Rice GI, Kitabayashi N, Barth M, Briggs TA, Burton ACE, Carpanelli ML et al (2017) Genetic, phenotypic, and interferon biomarker status in ADAR1-related neurological disease. *Neuropediatrics* 48:166–184
69. Livingston JH, Lin J-P, Dale RC, Gill D, Brogan P, Munnich A et al (2014) A type I interferon signature identifies bilateral striatal necrosis due to mutations in ADAR1. *J Med Genet* 51:76–82
70. Rice GI, Kashner PR, Forte GMA, Mannion NM, Greenwood SM, Szykiewicz M et al (2012) Mutations in ADAR1 cause Aicardi-Goutières syndrome associated with a type I interferon signature. *Nat Genet* 44:1243–1248
71. Chamaillard M, Philpott D, Girardin SE, Zouali H, Lesage S, Chareyre F et al (2003) Gene-environment interaction modulated by allelic heterogeneity in inflammatory diseases. *Proc Natl Acad Sci U S A* 100:3455–3460
72. Naderi N, Farnood A, Habibi M, Zojaji H, Balahi H, Firouzi F et al (2011) NOD2 exonic variations in Iranian Crohn's disease patients. *Int J Colorectal Dis* 26:775–781
73. Parkhouse R, Boyle JP, Monie TP (2014) Blau syndrome polymorphisms in NOD2 identify nucleotide hydrolysis and helical domain 1 as signalling regulators. *FEBS Lett* 588:3382–3389
74. Khubchandani RP, Hasija R, Touitou I, Khemani C, Wouters CH, Rose CD (2012) Blau arteritis resembling Takayasu disease with a novel NOD2 mutation. *J Rheumatol* 39:1888–1892
75. Israel L, Mellett M (2018) Clinical and genetic heterogeneity of mutations in psoriatic skin disease. *Front Immunol* 9:2239
76. Ammar M, Jordan CT, Cao L, Lim E, Bouchlaka Souissi C, Jrad A et al (2016) CARD14 alterations in Tunisian patients with psoriasis and further characterization in European cohorts. *Br J Dermatol* 174:330–337
77. Setta-Kaffetzi N, Simpson MA, Navarini AA, Patel VM, Lu H-C, Allen MH et al (2014) AP1S3 mutations are associated with pustular psoriasis and impaired Toll-like receptor 3 trafficking. *Am J Hum Genet* 94:790–797
78. Bamshad M, Kivisild T, Watkins WS, Dixon ME, Ricker CE, Rao BB et al (2001) Genetic evidence on the origins of Indian caste populations. *Genome Res* 11:994–1004
79. Zhao Z, Khan F, Borkar M, Herrera R, Agrawal S (2009) Presence of three different paternal lineages among North Indians: a study of 560 Y chromosomes. *Ann Hum Biol* 36:46–59

Publisher's Note

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

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)