

Research Article

Investigating the GWAS-Implicated Loci for Rheumatoid Arthritis in the Pakistani Population

Muhammad Muaaz Aslam ^{1,2}, Peter John,¹ Kang-Hsien Fan,² Attya Bhatti,¹ Wajahat Aziz,³ Bashir Ahmed,³ Eleanor Feingold,² F. Yesim Demirci,² and M. Ilyas Kamboh ²

¹Atta-ur-Rahman School of Applied Biosciences (ASAB), National University of Sciences and Technology (NUST), Islamabad, Pakistan

²Department of Human Genetics, Graduate School of Public Health (GSPH), University of Pittsburgh, Pittsburgh, PA, USA

³Department of Rheumatology, Pakistan Institute of Medical Sciences (PIMS), Islamabad, Pakistan

Correspondence should be addressed to M. Ilyas Kamboh; kamboh@pitt.edu

Received 26 February 2020; Revised 17 July 2020; Accepted 23 July 2020; Published 3 August 2020

Academic Editor: Wen-Jun Tu

Copyright © 2020 Muhammad Muaaz Aslam et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Rheumatoid arthritis (RA) is a complex and multifactorial autoimmune disorder with the involvement of multiple genetic and environmental factors. Genome-wide association studies (GWAS) have identified more than 50 RA genetic loci in European populations. Given the anticipated overlap of RA-relevant genes and pathways across different ethnic groups, we sought to replicate 58 GWAS-implicated SNPs reported in Europeans in Pakistani subjects. 1,959 unrelated subjects comprising 1,222 RA cases and 737 controls were collected from three rheumatology facilities in Pakistan. Genotyping was performed using iPLEX or TaqMan® methods. A total of 50 SNPs were included in the final association analysis after excluding those that failed assay design/run or post-run QC analysis. Fourteen SNPs (*LINC00824/rs1516971*, *PADI4/rs2240336*, *CEP57/rs4409785*, *CTLA4/rs3087243*, *STAT4/rs13426947*, *HLA-B/MICA/rs2596565*, *C5orf30/rs26232*, *CCL21/rs951005*, *GATA3/rs2275806*, *VPS37C/rs595158*, *HLA-DRB1/rs660895*, *EOMES/rs3806624*, *SPRED2/rs934734*, and *RUNX1/rs9979383*) were replicated in our Pakistani sample at false discovery rate (FDR) of <0.20 with nominal *p* values ranging from 4.73E-06 to 3.48E-02. Our results indicate that several RA susceptibility loci are shared between Pakistani and European populations, supporting the role of common genes/pathways.

1. Introduction

Rheumatoid arthritis (RA) is a complex autoimmune disorder due to the involvement of many genetic, environmental, and stochastic elements in its etiology [1, 2]. It is characterized by continuous inflammation associated with altered expression of different proinflammatory (TNF α , IL-1, IL-17) and anti-inflammatory (IL-4, IL-10, IL-13, IL-35) cytokines and activation of B and T cells, which leads to the destruction of synovial joints and ultimately physical disability [3, 4]. Worldwide, RA affects 0.5 to 1% of the population [5]. RA prevalence does not differ significantly between rural and urban areas [6]. It can affect both sexes at any age, but data suggest that

women are three times more prone than men [7]. There is a significant genetic susceptibility to RA with 50-60% heritability estimates [8]. Twin studies also revealed high concordance rates in monozygotic twins (12.3-15.4%) relative to dizygotic twins (3.5%), strongly indicating the presence of a genetic component in RA [9].

During the last two decades, many genetic studies, including genome-wide association studies (GWAS), have been conducted to understand the genetic basis of RA and this has led to the identification of more than 50 risk loci [10]. Besides *HLA-DRB1*, many non-*HLA* loci have been implicated that are involved in multiple functions/pathways including immune activation (NF- κ B signaling, JAK-STAT pathway, regulation, and activation of CD4⁺T-cells,

TABLE 1: List of selected GWAS-implicated RA SNPs examined in this study.

SNP	Locus	Chromosomal location (GRCh38)	Variant type	Gene	Reference
rs2228145	1q21	154454494	Missense variant	<i>IL6R</i>	[15]
rs2105325	1q25.1	173380586	Intron variant	<i>LOC100506023</i>	[16]
rs2843401	1p36	2596694	Intron variant	<i>MMEL1</i>	[15]
rs28411352	1p34.3	37812907	3 prime UTR variant	<i>MTF1</i>	[16]
rs883220	1p34	38151199	Intron variant	<i>LOC105378654</i>	[15]
rs2476601	1p13	113834946	Missense variant	<i>PTPN22</i>	[15]
rs6732565	2q13	110850255	Intron variant	<i>ACOXL</i>	[16]
rs11676922	2q11	100190478	Intergenic	<i>AFF3</i>	[17]
rs6715284	2q33.1	201289674	Intron variant	<i>ALS2CR12</i>	[16]
rs3087243	2q33	203874196	Downstream variant	<i>CTLA4</i>	[17]
rs34695944	2p16	60897715	Intron variant	<i>REL</i>	[15]
rs934734	2p14	65368452	Intron variant	<i>SPRED2</i>	[17]
rs13426947	2q32	191068528	Intron variant	<i>STAT4</i>	[15]
rs3806624	3p24.1	27723132	Upstream variant	<i>EOMES</i>	[16]
rs9826828	3q22.3	136683218	Intron variant	<i>STAG1</i>	[16]
rs4452313	3p24.3	17005540	Intron variant	<i>PLCL2</i>	[16]
rs13315591	3p14	58571114	Intron variant	<i>FAM107A</i>	[17]
rs874040	4p15	26106575	Upstream variant	<i>RBPJ</i>	[17]
rs2664035	4p11	48218822	Intron variant	<i>TEC</i>	[16]
rs71624119	5q11	56144903	Intron variant	<i>ANKRD55</i>	[15]
rs26232	5q21	103261019	Intron variant	<i>C5orf30</i>	[17]
rs3093023	6q27	167120802	Intron variant	<i>CCR6</i>	[16]
rs2234067	6p21.31	36387877	Upstream variant	<i>ETV7</i>	[16]
rs6920220	6q23	137685367	Upstream variant	<i>TNFAIP3</i>	[15]
rs2596565	6p21.33	31385552	Upstream variant	<i>HLA-B/MICA</i>	[18]
rs6910071	6p21	32315077	Intron variant	<i>TSBP1</i>	[17]
rs660895	6p21	32609603	Intron variant	<i>HLA-DRB1</i>	[14]
rs4272	7q21.2	92607515	3 prime UTR variant	<i>CDK6</i>	[16]
rs10488631	7q32	128954129	Downstream variant	<i>IRF5</i>	[17]
rs67250450	7p15.1	28135367	Intron variant	<i>JAZF1</i>	[16]
rs678347	8q22.3	101451374	Upstream variant	<i>GRHL2</i>	[16]
rs1516971	8q24.21	128529854	Intron variant	<i>LINC00824</i>	[16]
rs998731	8q21.13	80183160	Intron variant	<i>TPD52</i>	[16]
rs951005	9p13	34743684	Intergenic	<i>CCL21</i>	[17]
rs12764378	10q21	62040245	Intron variant	<i>ARID5B</i>	[15]
rs2275806	10p14	8053377	Intron variant	<i>GATA3</i>	[15]
rs706778	10p15	6056986	Intron variant	<i>IL2RA</i>	[17]
rs595158	11q12	61142109	Intron variant	<i>VPS37C</i>	[15]
rs4409785	11q21	95578258	Intergenic/unknown	<i>CEP57</i>	[16]
rs773125	12q13.2	56001170	Intron variant	<i>SUOX</i>	[16]
rs1950897	14q24.1	68293424	Intron variant	<i>RAD51B</i>	[16]
rs8043085	15q14	38535939	Intron variant	<i>RASGRP1</i>	[15]
rs8026898	15q23	69699078	Intergenic/unknown	<i>TLE3</i>	[15]
rs13330176	16q24	85985481	Intergenic/unknown	<i>IRF8</i>	[15]
rs4780401	16p13.13	11745470	Upstream variant	<i>TXNDC11</i>	[16]
rs12936409	17q12	39887396	Intergenic/unknown	<i>IKZF3</i>	[15]

TABLE 1: Continued.

SNP	Locus	Chromosomal location (GRCh38)	Variant type	Gene	Reference
rs34536443	19p13	10352442	Missense variant	<i>TYK2</i>	[15]
rs4810485	20q13	46119308	Intron variant	<i>CD40</i>	[17]
rs9979383	21q22	35343463	Intron variant	<i>RUNX1</i>	[15]
rs1893592	21q22.3	42434957	Intron variant	<i>UBASH3A</i>	[16]
rs909685	22q13.1	39351666	Intron variant	<i>SYNGR1</i>	[16]
rs2240336	1p36	17347907	Intron variant	<i>PADI4</i>	[16]
rs10175798	2p23.1	30226728	Upstream variant	<i>LBH</i>	[16]
rs968567	11q12.2	61828092	Intron variant	<i>FADS2</i>	[16]
rs10774624	12q24.12	111395984	Upstream variant	<i>LINC02356</i>	[16]
rs9603616	13q14.11	39793932	Downstream variant	<i>COG6</i>	[16]
rs73194058	21q22.11	33391982	Intergenic/unknown	<i>IFNGR2</i>	[16]
rs2834512	21q22	34539301	Intron variant	<i>RCAN1</i>	[15]

complement activation), fibroblast differentiation and dedifferentiation, and bone modeling and repair [11, 12].

Thus far, the major focus of genetic studies on RA has largely been on Europeans or European-derived populations, with only limited data available on other populations [13], especially in South Asians [14]. In an effort to expand genetic studies on RA in South Asians and to further delineate its genetic basis, we sought to replicate 58 genome-wide significant single nucleotide polymorphisms (SNPs) from 58 loci previously implicated in Europeans or European-derived populations [15–18] in Pakistani population where there is the paucity of such data.

2. Material and Methods

2.1. Study Subjects. A total of 1,959 subjects (1,222 cases, 737 controls) were included in our study. Blood samples and relevant clinical information were collected from three major public or private rheumatology clinics in Pakistan: Pakistan Institute of medical sciences (PIMS), Military hospital, and Rehmat Noor Clinic. All cases included in this study (mean age \pm SD = 43.1 \pm 12.33, 78.6% women) were clinically diagnosed by rheumatologists and met the American College of Rheumatology (ACR) 1987 classification criteria for RA [19]. All controls included in this study (mean age \pm SD = 40.7 \pm 12.49, 39.5% women) were recruited from the general population and were free of any autoimmune disease at the time of recruitment. A screening questionnaire was filled out and a written informed consent was obtained from each subject at the time of the recruitment. All blood samples were collected in EDTA tubes to avoid coagulation and processed shortly after the collection. The study was approved by the Institutional review board (IRB) of the University of Pittsburgh, USA (IRB no. PRO12110472).

2.2. Genomic DNA Extraction. Genomic DNA was extracted from whole blood using either the standard phenol-chloroform extraction method or GeneJET Whole Blood

Genomic DNA Purification (Thermo Scientific USA) and quantified using the NanoDrop™ 2000 spectrophotometer (Thermo Scientific USA).

2.3. Genotyping. A total of 58 RA-associated genome-wide significant SNPs ($p < 5E - 08$) was selected from the previously published GWAS in subjects of European ancestry (Table 1). SNP genotyping was performed using either the TaqMan® (Applied Biosystems, Thermo Fisher Scientific) or iPLEX® Gold (Agena Bioscience) methods and following the manufacturer’s design/order instructions and protocols. After the thermal cycling of the TaqMan® assays and DNAs on 384-well plates, the endpoint fluorescence reading was performed on a QuantStudio™ 12K Flex system (Applied Biosystems, Thermo Fisher Scientific). The iPLEX® Gold genotyping was performed in the Core laboratories of the University of Pittsburgh. 18% replicates were used to test genotyping consistency.

2.4. Statistical Analysis. Concordance to Hardy-Weinberg Equilibrium (HWE) was tested using the Chi-square goodness of fit test. Departure from HWE was considered at arbitrarily $p < 1E - 05$. Logistic regression using an additive model and minor allele as the effect allele was employed for case-control association analysis using sex and age as covariates. The Benjamin Hochberg false discovery rate (FDR) was applied to correct for multiple testing [20]. $p < 0.05$ was considered as suggestive evidence of association and FDR (q value) of < 0.20 as statistically significant as used in previous reports [21, 22]. All analyses were implemented in R, version 3.4.4.

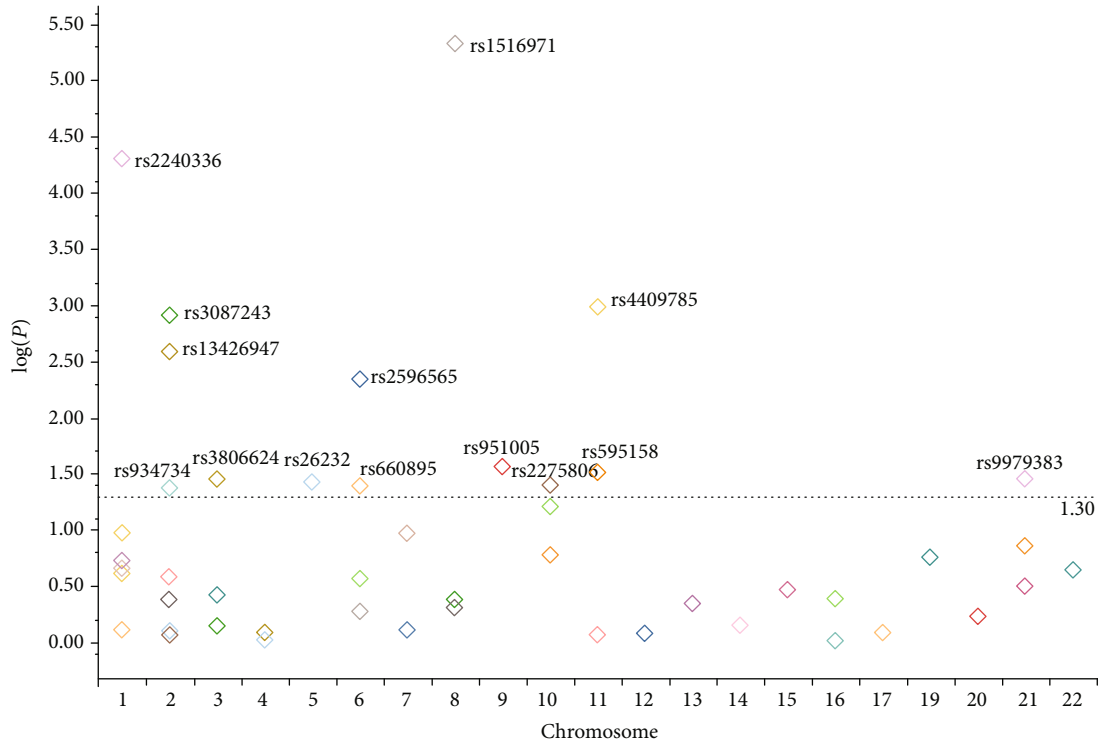
2.5. Functional Annotations. To evaluate the potential biological significance of reported genome-wide significant SNPs, we used the Genotype-Tissue Expression (GTEx) database (<https://gtexportal.org/home/>) to search for expression quantitative trait loci (eQTL) in RA-relevant tissues and whole blood. We also used the RegulomeDB online database

TABLE 2: Association analysis results for GWAS-implicated RA SNPs in the Pakistani population.

Gene/SNP	Major allele	Minor allele	MAF	Reported GWAS <i>p</i> value	Nominal <i>p</i> value in Pakistanis	OR (95% CI)	FDR (<i>q</i>)
<i>LINC00824</i> /rs1516971	T	C	0.099	3.20E-11	4.73E-06	0.57 (0.45, 0.73)	2.37E-04
<i>PADI4</i> /rs2240336	C	T	0.452	5.9E-9	5.00E-05	0.74 (0.64, 0.86)	1.25E-03
<i>CEP57</i> /rs4409785	T	C	0.262	3.60E-09	1.03E-03	1.33 (1.12, 1.58)	1.54E-02
<i>CTLA4</i> /rs3087243	A	G	0.422	1.2E-8	1.23E-03	1.28 (1.1, 1.49)	1.54E-02
<i>STAT4</i> /rs13426947	G	A	0.238	7.2E-10	2.59E-03	1.31 (1.1, 1.56)	2.59E-02
<i>HLA-B/MICA</i> /rs2596565	G	A	0.119	9.26E-09	4.53E-03	1.4 (1.11, 1.77)	3.77E-02
<i>C5orf30</i> /rs26232	C	T	0.181	4.10E-08	3.73E-02	0.83 (0.69, 0.99)	1.52E-01
<i>CCL21</i> /rs951005	A	G	0.248	3.90E-10	2.75E-02	0.82 (0.69, 0.98)	1.52E-01
<i>GATA3</i> /rs2275806	A	G	0.397	4.6E-8	4.02E-02	1.17 (1.01, 1.36)	1.52E-01
<i>VPS37C</i> /rs595158	A	C	0.320	3.4E-8	3.10E-02	1.19 (1.02, 1.38)	1.52E-01
<i>HLA-DRB1</i> /rs660895	A	G	0.111	<1E-300	4.02E-02	1.27 (1.01, 1.59)	1.52E-01
<i>EOMES</i> /rs3806624	G	A	0.262	2.80E-08	3.54E-02	0.84 (0.71, 0.99)	1.52E-01
<i>SPRED2</i> /rs934734	A	G	0.406	5.30E-10	4.24E-02	1.17 (1.01, 1.35)	1.52E-01
<i>RUNX1</i> /rs9979383	T	C	0.300	5.0E-10	3.48E-02	0.84 (0.72, 0.99)	1.52E-01
<i>IL2RA</i> /rs706778	T	C	0.458	1.40E-11	6.21E-02	0.87 (0.75, 1.01)	2.07E-01
<i>LOC105378654</i> /rs883220	C	A	0.212	2.1E-8	1.08E-01	0.87 (0.73, 1.03)	3.17E-01
<i>IRF5</i> /rs10488631	T	C	0.196	4.20E-11	1.07E-01	1.16 (0.97, 1.39)	3.17E-01
<i>RCAN1</i> /rs2834512	G	A	0.160	2.1E-8	1.39E-01	0.86 (0.71, 1.05)	3.86E-01
<i>ARID5B</i> /rs12764378	G	A	0.241	4.5E-10	1.64E-01	1.13 (0.95, 1.34)	4.33E-01
<i>TYK2</i> /rs34536443	G	C	0.010	2.3E-14	1.73E-01	0.62 (0.31, 1.23)	4.33E-01
<i>IL6R</i> /rs2228145	A	C	0.323	1.3E-8	1.89E-01	0.9 (0.77, 1.05)	4.50E-01
<i>PTPN22</i> /rs2476601	G	A	0.016	7.5E-77	2.21E-01	1.47 (0.79, 2.73)	4.93E-01
<i>SYNGR1</i> /rs909685	T	A	0.480	6.40E-12	2.27E-01	1.09 (0.95, 1.26)	4.93E-01
<i>MTF1</i> /rs28411352	C	T	0.222	5.90E-09	2.42E-01	1.11 (0.93, 1.32)	5.05E-01
<i>CCR6</i> /rs3093023	G	A	0.425	1.50E-11	2.70E-01	1.09 (0.94, 1.26)	5.19E-01
<i>ALS2CR12</i> /rs6715284	C	G	0.185	2.50E-09	2.61E-01	1.11 (0.92, 1.35)	5.19E-01
<i>UBASH3A</i> /rs1893592	A	C	0.241	9.80E-09	3.11E-01	0.92 (0.77, 1.08)	5.75E-01
<i>IRF8</i> /rs13330176	T	A	0.243	4.0E-8	4.04E-01	0.93 (0.78, 1.11)	6.13E-01
<i>PLCL2</i> /rs4452313	A	T	0.492	5.2E-11	3.76E-01	1.07 (0.92, 1.24)	6.13E-01
<i>ACOXL</i> /rs6732565	G	A	0.444	9.40E-09	4.15E-01	0.94 (0.81, 1.09)	6.13E-01
<i>GRHL2</i> /rs678347	A	G	0.425	7.30E-10	4.17E-01	1.06 (0.92, 1.23)	6.13E-01
<i>TLE3</i> /rs8026898	G	A	0.303	9.16E-14	3.50E-01	0.93 (0.79, 1.09)	6.13E-01
<i>RASGRP1</i> /rs8043085	G	T	0.280	1.4E-10	3.64E-01	1.08 (0.92, 1.27)	6.13E-01
<i>REL</i> /rs34695944	T	C	0.112	1.4E-10	4.05E-01	1.1 (0.88, 1.39)	6.13E-01
<i>COG6</i> /rs9603616	C	T	0.260	2.80E-11	4.47E-01	0.94 (0.8, 1.11)	6.38E-01
<i>TPD52</i> /rs998731	T	C	0.434	6.60E-09	5.04E-01	0.95 (0.82, 1.1)	7.00E-01
<i>TNFAIP3</i> /rs6920220	G	A	0.132	2.3E-13	5.33E-01	1.07 (0.86, 1.34)	7.20E-01
<i>CD40</i> /rs4810485	G	T	0.250	2.8E-9	6.03E-01	0.96 (0.81, 1.13)	7.94E-01
<i>RAD51B</i> /rs1950897	T	C	0.208	5.00E-08	7.02E-01	1.04 (0.86, 1.24)	8.98E-01
<i>FAM107A</i> /rs13315591	T	C	0.043	4.60E-08	7.25E-01	1.06 (0.75, 1.51)	8.98E-01
<i>MMEL1</i> /rs2843401	C	T	0.461	6.6E-9	7.80E-01	1.02 (0.88, 1.18)	8.98E-01
<i>RBPJ</i> /rs874040	G	C	0.159	1.00E-16	8.20E-01	0.98 (0.81, 1.18)	8.98E-01
<i>IKZF3</i> /rs12936409	C	T	0.392	2.8E-9	8.19E-01	0.98 (0.85, 1.14)	8.98E-01
<i>LBH</i> /rs10175798	A	G	0.474	4.20E-08	8.46E-01	1.01 (0.88, 1.17)	8.98E-01
<i>LINC02356</i> /rs10774624	A	G	0.138	6.90E-09	8.38E-01	0.98 (0.79, 1.2)	8.98E-01
<i>FADS2</i> /rs968567	G	A	0.084	1.80E-08	8.62E-01	0.98 (0.75, 1.27)	8.98E-01

TABLE 2: Continued.

Gene/SNP	Major allele	Minor allele	MAF	Reported GWAS p value	Nominal p value in Pakistanis	OR (95% CI)	FDR (q)
<i>AFF3</i> /rs11676922	T	A	0.485	1.00E-14	7.91E-01	0.98 (0.85, 1.14)	8.98E-01
<i>CDK6</i> /rs4272	A	G	0.083	1.2E-8	7.64E-01	0.96 (0.74, 1.24)	8.98E-01
<i>TEC</i> /rs2664035	G	A	0.256	3.30E-08	9.54E-01	1 (0.84, 1.17)	9.70E-01
<i>TXNDC11</i> /rs4780401	T	G	0.430	8.70E-09	9.70E-01	1 (0.86, 1.16)	9.70E-01

FIGURE 1: Annotated 50 tested SNPs. SNPs with p value < 0.05 are shown above the dotted line.

(<http://regulome.stanford.edu/>) to determine possible regulatory functions of the SNPs located in noncoding regions.

3. Results

A total of 1,222 unrelated RA cases and 737 controls were recruited for this research study. The prevalence of RA was higher in females (78%) than males (22%), supporting the earlier data that females are more prone to RA [14].

Eight of the 58 genotyped SNPs failed the QC (quality control) during assay design (either iPLEX® Gold/TaqMan® or both). The genotype distribution of all QC-passed 50 SNPs adhered to the HWE. The association analyses results in our Pakistani sample are presented in Table 2. Fourteen SNPs showed nominal significance at $p < 0.05$ and FDR of ≤ 0.20 .

We found the same direction of association where the minor allele showed the same risk or protective effect as compared to the previous findings on the tested allele (major or minor). In our data, *LINC00824*/rs1516971 showed the most significant association ($p = 4.73E - 06$). The second most significant SNP was *PADI4*/rs2240336 ($p = 5.00E - 05$)

followed by *CEP57*/rs4409785 ($p = 1.03E - 03$), *CTLA4*/rs3087243 ($p = 1.23E - 03$), *STAT4*/rs13426947 ($p = 2.59E - 03$), and *HLA-B-MICA*/rs2596565 ($p = 4.53E - 03$). Eight other SNPs showed marginal significance: *C5orf30*/rs26232 ($p = 3.73E - 02$), *CCL21*/rs951005 ($p = 2.75E - 02$), *GATA3*/rs2275806 ($p = 4.02E - 02$), *VPS37C*/rs595158 ($p = 3.10E - 02$), *HLA-DRB1*/rs660895 ($p = 4.02E - 02$), *EOMES*/rs3806624 ($p = 3.54E - 02$), *SPRED2*/rs934734 ($p = 4.24E - 02$), and *RUNX1*/rs9979383 ($p = 3.48E - 02$). Figure 1 shows the distribution of tested SNPs across the genome where the SNPs with p value < 0.05 are labeled.

Next, we examined the functional significance of all 50 SNPs using the GTEx and RegulomeDB databases. Table 3 shows the category summaries of RegulomeDB scores, and Table 4 shows 14 SNPs that had a RegulomeDB score of ≤ 3 , indicating strong evidence of potential regulatory role. SNPs falling in this category have more likelihood to affect the binding of transcriptional factors. Out of these fourteen SNPs, only five (*HLA-B/MICA*/rs2596565, *HLA-DRB1*/rs660895, *C5orf30*/rs26232, *CTLA4*/rs3087243, and *RUNX1*/rs9979383) had $p < 0.05$ in our association results.

TABLE 3: RegulomeDB score description.

Score	Description
Likely to affect binding and linked to the expression of a gene target	
1a	eQTL + TF binding + matched TF motif + matched DNase footprint + DNase peak
1b	eQTL + TF binding + any motif + DNase footprint + DNase peak
1c	eQTL + TF binding + matched TF motif + DNase peak
1d	eQTL + TF binding + any motif + DNase peak
1e	eQTL + TF binding + matched TF motif
1f	eQTL + TF binding/DNase peak
Likely to affect binding	
2a	TF binding + matched TF motif + matched DNase footprint + DNase peak
2b	TF binding + any motif + DNase footprint + DNase peak
2c	TF binding + matched TF motif + DNase peak
Less likely to affect binding	
3a	TF binding + any motif + DNase peak
3b	TF binding + matched TF motif
Minimal binding evidence	
4	TF binding + DNase peak
5	TF binding or DNase peak
6	Motif hit
7	No data available

SYNGR1/rs909685 had the top RegulomeDB score of 1b followed by *FADS2/rs968567*, *CD40/rs4810485*, *CCR6/rs3093023*, *HLA-B/MICA/rs2596565*, and *HLA-DRB1/rs660895* with a score of 1f. The latter two SNPs (*HLA-B/MICA/rs2596565* and *HLA-DRB1/rs660895*) were significant in our sample ($p = 0.0045$ and $p = 0.0402$) and based on RegulomeDB, both SNPs are eQTL for *HLA-DQA1*. While *HLA-DRB1/rs660895* affects the binding of two proteins (BCLAF1 and POLR2A), no such evidence was found for *HLA-B/MICA/rs2596565*. *SYNGR1/rs909685* is an intronic variant, which falls in the p53decamer binding motif and affects the binding of five different proteins (MAX, MYC, PAX5, TRIM28, and EBF1). *FADS2/rs968567* is also an intronic variant and it affects the binding of twenty-eight proteins, including MAX, POLR2A, NFKB1, and NFIC. The binding of these proteins is also affected by the *CD40/rs4810485* variant. An intronic *CCR6/rs3093023* variant falls in the Oct-1 binding motif and is eQTL for *CCR6*.

GTEx data eQTL showed the lowest p-eQTL (2.59E-34) for *FADS2/rs968567* that affects *FADS2* gene expression in transformed fibroblast cells; the same variant also affects *TMEM258* ($p - eQTL = 3.43E - 08$), *ZBTB3* ($p - eQTL = 0.0168$), *MYRF* ($p - eQTL = 0.0207$), *DAGLA* ($p - eQTL = 0.0246$), and *RAB31L1* ($p - eQTL = 0.0247$). In transformed fibroblast cells, *LINC00824/rs1516971* was a weak eQTL ($p = 0.0157$) for *RP11-89M16.1*. No data was available for *CTLA4/rs3087243* in GTEx.

STAT4/rs13426947 was eQTL for *STAT4*, *AC005540.3*, and *RP11-647K16.1* in EBV-transformed lymphocyte cells. *HLA-B/MICA/rs2596565* was eQTL for multiple immune-related genes including *C4A*, *C4B*, *HLA-S*, *HLA-B*, *HLA-C*, and *MICA* in all three tested tissues. *C5orf30/rs26232* showed

the strongest eQTL ($p - eQTL = 1.72E - 19$) with PPIP5K2 only in whole blood. Additional details of GTEx data are given in supplementary Table S1.

4. Discussion

There have been a number of genome-wide association studies on RA over the past decade that have resulted in the identification of many RA susceptibility loci, thus improving our understanding of the complex genetic underpinning of RA [15–17]. Most of the recently published GWAS have been conducted on Europeans and North Americans, which have identified many new RA risk loci and replicated and confirmed the previously reported putative risk loci. Many of these reported risk loci are shared among different ethnic groups, while some are specific to certain populations [23]. Replication studies across different ethnic groups are necessary to better understand the globe-wide population specificity of these established RA risk loci and guide future directions. Since South Asians, including the Pakistani population, have significant European ancestry [24, 25], we chose European originated GWAS-implicated SNPs for replication in Pakistanis. For this purpose, we enrolled 1,959 unrelated RA cases and controls from two public and private Rheumatology facilities, where people from different demographic regions of Pakistan visit for treatment, and we successfully genotyped 50 GWAS—implicated SNPs in those subjects.

In our data set, *rs1516971* was the most significant SNP ($p = 4.73E - 06$), which is an intronic variant in the *LINC00824* gene at chromosome 8q24.21. Previously, this variant has also been reported in a trans-ethnic association study of RA ($p = 1.3E - 10$) [16]. Another SNP (*rs6651252*)

TABLE 4: Details of study SNPs (RegulomeDB Score ≤ 3) with putative regulatory functions.

SNP	Chr.	Score	eQTL	Bound protein	Motifs
<i>SYNGR1</i> /rs909685	chr22	1b	<i>SYNGR1</i>	MAX, MYC, PAX5, TRIM28, EBF1 CTCF, E2F4, EP300, ETS1, FOSL2, FOXP2, GABPA, GABPB1, GATA1, HNF4A, HNF4G, MAX, MYBL2, NFIC, NFKB1, NFYB, PML, POLR2A, RAD21, REST, SIN3A, SPI1, SREBF1, SREBF2, TAF1, TCF12, YY1, ZBTB7A	p53decamer
<i>FADS2</i> /rs968567	chr11	1f	<i>NXF1</i>	POLR2A, NFKB1, NFIC, MEF2C, MEF2A, IRF1, IKZF1, FOXM1, BCL3, BATF	
<i>CD40</i> /rs4810485	chr20	1f	<i>CD40</i>		
<i>CCR6</i> /rs3093023	chr6	1f	<i>CCR6</i>		Oct_1
<i>HLAB/MICA</i> /rs2596565	chr6	1f	BTN3A2, HLA-A, HLA-C, HLA-DPB1, HLA-DQA1, HLA- DQB1, HLA-DRB1, HLA-G, HLA-H, VARSL		
<i>HLA-DRB1</i> /rs660895	chr6	1f	<i>HLA-DQA2</i> , <i>HLA-DQA1</i>	BCLAF1, POLR2A	
<i>GRHL2</i> /rs678347	chr8	2a		YY1, CEBPB, EP300, FOXA2, JUND, STAT3, TCF12, USF2, SIN3A	HNF3, Foxa2, FOXP2, Freac-2, Freac-4, FOXA1, HNF3, FOXC1
<i>PTPN22</i> /rs2476601	chr1	2b		FOS, STAT3	Gabpa, Etv1, Elk1, Erg, Ehf, PU.1
<i>LBH</i> /rs10175798	chr2	2b		ATF2, FOXM1, RUNX3, NFIC, SPI1, MEF2A, BATF	STAT1:STAT1
<i>RBPJ</i> /rs874040	chr4	2b		GATA2, PML, TAL1	MEF-2
<i>C5orf30</i> /rs26232	chr5	2b		BHLHE40, CHD1, EP300, FOXM1, MAZ, NFATC1, NFIC, RUNX3, TBL1XR1, TBP, USF2, ATF2, MEF2A, MEF2C, NFKB1	ICSBP
<i>RUNX1</i> /rs9979383	chr21	3a		MYC, NFKB1	E47, FIGLA, ID4, MESP1, SNAI2
<i>CTLA4</i> /rs3087243	chr2	3a		MAX, FOS, STAT3	Bbx
<i>IRF5</i> /rs10488631	chr7	3a		GATA2	Nanog

from the same region, which is in complete LD ($r^2 = 1.00$) with rs1516971, has also been reported to be associated with RA and Crohn's disease [15, 26, 27]. GTEx expression data suggests that rs1516971 affects the expression of *RP11-89M16.1* at chromosome 8 in transformed fibroblast cells ($p - eQTL = 0.0157$).

The second most significant SNP in our data was rs2240336 ($p = 5.00E - 05$), which is present in the 9th intron of *PADI4* at chromosome 1p36. *PADI4* is one of the most important RA susceptibility loci in multiple ethnic groups, including Europeans, Asians, and Latin Americans [28, 29]. Another SNP in this region, rs2301888 that is in strong LD ($r^2 = 0.8$) with rs2240336, has also been reported to be associated with RA in Europeans and Koreans [30]. However, a study on Iranian population found no significant association of two other SNPs (rs11203367 and rs874881) in the *PADI4* region with RA, which are not in strong LD ($r^2 < 0.8$) with rs2240336 [31]. GTEx expression data showed that rs2240336 is eQTL for *RP4-798A10.2* ($p - eQTL = 0.0191$) and *PADI3* ($p - eQTL = 0.0203$) in EBV transformed lym-

phocytes. The rs2240336 SNP also controls the expression of *FBXO42* ($p - eQTL = 0.00526$) in transformed fibroblasts. The third most important SNP in our data is *CEP57*/rs4409785 ($p = 1.03E - 03$), which is an intergenic variant on chromosome 11q21. This SNP has also been reported to be associated with other immune-related diseases such as multiple sclerosis, vitiligo, and Graves' disease [32–34]. In GTEx expression data, rs4409785 was eQTL for *JRKL* ($p - eQTL = 0.037$) in EBV-transformed lymphocytes, and for *AP001877.1* ($p - eQTL = 0.0188$), and *MAML2* ($p - eQTL = 0.034$) in transformed fibroblast cells.

Among the reported genome-wide significant SNPs that did not show significant association in our sample, the top functional ones based on the GTEx and RegulomeDB database included *SYNGR1*/rs909685, *FADS2*/rs968567, *CD40*/rs4810485, *CCR6*/rs3093023, and *IKZF3*/rs12936409. *SYNGR1*/rs909685 has a RegulomeDB score of 1b and is a strong eQTL for *SYNGR1* in EBV-transformed lymphocytes ($p - eQTL = 3.74E - 14$) and whole blood ($p - eQTL = 5.68E - 21$). *FADS2*/rs968567

has a RegulomeDB score of 1f and is eQTL for *FADS2* in EBV-transformed lymphocytes ($p - eQTL = 2.05E - 05$), whole blood and ($p - eQTL = 1.86E - 71$) transformed fibroblasts ($p - eQTL = 2.59E - 34$). *CD40/rs4810485* also has a high RegulomeDB score of 1f and is strong eQTL for *CD40* in whole blood ($p - eQTL = 1.33E - 09$) and transformed fibroblasts ($p - eQTL = 8.58E - 11$) but weak eQTL for *CD40* in EBV-transformed lymphocytes ($p - eQTL = 0.0272$). *CCR6/rs3093023* with RegulomeDB score of 1f was the strongest eQTL for *RNASET2* in transformed fibroblasts ($p - eQTL = 1.08E - 16$) and whole blood ($p - eQTL = 1.31E - 15$) with no data reported for EBV-transformed lymphocytes in GTEx. *IKZF3/rs12936409* was a strong eQTL for *ORMDL3* in all three tested tissues. In GTEx expression data, rs12936409 was found to control the expression of *GSDMA*, *GSDMB*, and *ORMDL3* in all tested cells and whole blood.

5. Conclusions

We were able to successfully replicate 14 of 50 SNPs selected from previously published GWAS results in European populations in our unique Pakistani population. These findings suggest that there is a sharing of RA risk loci among different population groups. A weakness of our study is the availability of relatively small-sized sample, which may have prevented 36 SNPs from achieving the significance threshold, although they showed a trend for the same directional effects as the reported ones. Further studies using larger samples may help to identify and replicate more RA risk loci in the Pakistani population.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

All authors declared no competing interests.

Acknowledgments

We are very thankful to all Rheumatologists and their staff for providing us blood samples of RA patients and the Higher Education Commission (HEC) and the University of Pittsburgh for providing funding. We thank all the study subjects (RA patients and control subjects) for their participation in this study.

Supplementary Materials

S1: additional details of GTEx data. (*Supplementary Materials*)

References

- [1] P. Szodoray, Z. Szabó, A. Kapitány et al., "Anti-citrullinated protein/peptide autoantibodies in association with genetic and environmental factors as indicators of disease outcome in rheumatoid arthritis," *Autoimmunity Reviews*, vol. 9, no. 3, pp. 140–143, 2010.
- [2] M. M. Aslam, P. John, A. Bhatti, S. Jahangir, and M. I. Kambh, "Vitamin D as a principal factor in mediating rheumatoid arthritis-derived immune response," *BioMed Research International*, vol. 2019, Article ID 3494937, 12 pages, 2019.
- [3] F. A. S. Kurreeman, L. Padyukov, R. B. Marques et al., "A candidate gene approach identifies the TRAF1/C5 region as a risk factor for rheumatoid arthritis," *PLoS Medicine*, vol. 4, no. 9, article e278, 2007.
- [4] G. S. Firestein, "Evolving concepts of rheumatoid arthritis," *Nature*, vol. 423, no. 6937, pp. 356–361, 2003.
- [5] M. Bax, J. van Heemst, T. W. J. Huizinga, and R. E. M. Toes, "Genetics of rheumatoid arthritis: what have we learned?," *Immunogenetics*, vol. 63, no. 8, pp. 459–466, 2011.
- [6] I. Rudan, S. Sidhu, A. Papana et al., "Prevalence of rheumatoid arthritis in low- and middle-income countries: a systematic review and analysis," *Journal of Global Health*, vol. 5, no. 1, article 010409, 2015.
- [7] S. S. Dhawan and A. A. Quyyumi, "Rheumatoid arthritis and cardiovascular disease," *Current Atherosclerosis Reports*, vol. 10, no. 2, pp. 128–133, 2008.
- [8] E. J. M. Toonen, P. Barrera, T. R. D. J. Radstake et al., "Gene expression profiling in rheumatoid arthritis: current concepts and future directions," *Annals of the Rheumatic Diseases*, vol. 67, no. 12, pp. 1663–1669, 2008.
- [9] A. J. MacGregor, H. Snieder, A. S. Rigby et al., "Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins," *Arthritis and Rheumatism*, vol. 43, no. 1, pp. 30–37, 2000.
- [10] Y. Ito and S. Sakaguchi, "Rheumatoid arthritis, genetics," in *Encyclopedia of Medical Immunology*, I. R. Mackay, N. R. Rose, B. Diamond, and A. Davidson, Eds., pp. 996–1001, Springer, New York, NY, USA, 2014.
- [11] The Wellcome Trust Case Control Consortium, "Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls," *Nature*, vol. 447, no. 7145, pp. 661–678, 2007.
- [12] D. Diogo, Y. Okada, and R. M. Plenge, "Genome-wide association studies to advance our understanding of critical cell types and pathways in rheumatoid arthritis: recent findings and challenges," *Current Opinion in Rheumatology*, vol. 26, no. 1, pp. 85–92, 2014.
- [13] J. Freudenberg, H. S. Lee, B. G. Han et al., "Genome-wide association study of rheumatoid arthritis in Koreans: population-specific loci as well as overlap with European susceptibility loci," *Arthritis & Rheumatism*, vol. 63, no. 4, pp. 884–893, 2011.
- [14] S. F. Jalil, A. Bhatti, F. Y. Demirci et al., "Replication of European rheumatoid arthritis loci in a Pakistani population," *The Journal of Rheumatology*, vol. 40, no. 4, pp. 401–407, 2013.
- [15] S. Eyre, J. Bowes, D. Diogo et al., "High-density genetic mapping identifies new susceptibility loci for rheumatoid arthritis," *Nature Genetics*, vol. 44, no. 12, pp. 1336–1340, 2012.
- [16] Y. Okada, D. Wu, G. Trynka et al., "Genetics of rheumatoid arthritis contributes to biology and drug discovery," *Nature*, vol. 506, no. 7488, pp. 376–381, 2014.
- [17] E. A. Stahl, S. Raychaudhuri, E. F. Remmers et al., "Genome-wide association study meta-analysis identifies seven new

- rheumatoid arthritis risk loci," *Nature Genetics*, vol. 42, no. 6, pp. 508–514, 2010.
- [18] L. Bossini-Castillo, C. de Kovel, H. Kallberg et al., "A genome-wide association study of rheumatoid arthritis without antibodies against citrullinated peptides," *Annals of the Rheumatic Diseases*, vol. 74, no. 3, p. e15, 2015.
- [19] F. C. Arnett, S. M. Edworthy, D. A. Bloch et al., "The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis," *Arthritis & Rheumatism*, vol. 31, no. 3, pp. 315–324, 1988.
- [20] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate - a practical and powerful approach to multiple testing," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 57, no. 1, pp. 289–300, 1995.
- [21] Z. H. Radwan, X. Wang, F. Waqar et al., "Comprehensive evaluation of the association of APOE genetic variation with plasma lipoprotein traits in U.S. whites and African blacks," *PLoS One*, vol. 9, no. 12, article e114618, 2014.
- [22] V. Niemsiri, X. Wang, D. Pirim et al., "Genetic contribution of SCARB1 variants to lipid traits in African blacks: a candidate gene association study," *BMC Medical Genetics*, vol. 16, no. 1, p. 106, 2015.
- [23] A. Suzuki and K. Yamamoto, "From genetics to functional insights into rheumatoid arthritis," *Clinical and Experimental Rheumatology*, vol. 33, no. 4, pp. S40–S53, 2015.
- [24] L. Quintana-Murci, R. Chaix, R. S. Wells et al., "Where west meets east: the complex mtDNA landscape of the southwest and central Asian corridor," *American Journal of Human Genetics*, vol. 74, no. 5, pp. 827–845, 2004.
- [25] N. A. Rosenberg, S. Mahajan, S. Ramachandran, C. Zhao, J. K. Pritchard, and M. W. Feldman, "Clines, clusters, and the effect of study design on the inference of human population structure," *PLoS Genetics*, vol. 1, no. 6, article e70, 2005.
- [26] A. Franke, D. P. B. McGovern, J. C. Barrett et al., "Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci," *Nature Genetics*, vol. 42, no. 12, pp. 1118–1125, 2010.
- [27] S. Matthews, M. Eshelman, K. Schieffer, W. Koltun, and G. S. Yochum, "Tu1804 - Crohn's Disease-Associated Snp, Rs6651252, Demarcates a Wntresponsive DNA element that correlates with increased *Myc* expression in Cd patients," *Gastroenterology*, vol. 154, no. 6, pp. S-1024, 2018.
- [28] X. K. Yang, J. Liu, J. Liu et al., "Associations between PADI4 gene polymorphisms and rheumatoid arthritis: an updated meta-analysis," *Archives of Medical Research*, vol. 46, no. 4, pp. 317–325, 2015.
- [29] C. L. Too, S. Murad, J. S. Dhaliwal et al., "Polymorphisms in peptidylarginine deiminase (PADI) associate with rheumatoid arthritis in diverse Asian populations: evidence from MyEIRA study and meta-analysis," *Arthritis Research & Therapy*, vol. 14, no. 6, article R250, 2012.
- [30] K. Kim, S. Y. Bang, H. S. Lee et al., "High-density genotyping of immune loci in Koreans and Europeans identifies eight new rheumatoid arthritis risk loci," *Annals of the Rheumatic Diseases*, vol. 74, no. 3, article e13, 2015.
- [31] E. Shamsian, M. Azarian, M. Akhlaghi et al., "PADI4 polymorphisms in Iranian patients with rheumatoid arthritis," *Acta Reumatológica Portuguesa*, vol. 41, no. 4, pp. 338–343, 2016.
- [32] The International Multiple Sclerosis Genetics Consortium & The Wellcome Trust Case Control Consortium 2, "Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis," *Nature*, vol. 476, no. 7359, pp. 214–219, 2011.
- [33] Y. Jin, S. A. Birlea, P. R. Fain et al., "Genome-wide association analyses identify 13 new susceptibility loci for generalized vitiligo," *Nature Genetics*, vol. 44, no. 6, pp. 676–680, 2012.
- [34] J. J. Khong, K. P. Burdon, Y. Lu et al., "Pooled genome wide association detects association upstream of FCRL3 with Graves' disease," *BMC Genomics*, vol. 17, no. 1, p. 939, 2016.