Whole Blood Targeted Bisulfite Sequencing and Differential Methylation in the *C6ORF10* Gene of Patients with Rheumatoid Arthritis

Vidyanand Anaparti[1] (ID), Prasoon Agarwal[2] (ID), Irene Smolik[3] (ID), Neeloffer Mookherjee[4] (ID), and Hani El-Gabalawy[1] (ID)

ABSTRACT. **Objective.** Polymorphisms in human major histocompatibility complex (MHC) are the strongest genetic associations with rheumatoid arthritis (RA). Epigenome-wide methylation studies suggest DNA methylation changes within MHC may contribute to disease susceptibility. We profiled MHC-specific methylated CpG (5′–C–phosphate–G–3′) in autoantibody-positive patients with RA and matched unaffected anticitrullinated protein antibodies–negative first-degree relatives (ACPA–/FDR) from an indigenous North American (INA) population that is known to have prevalent RA.
Methods. DNA was isolated from whole blood and targeted bisulfite sequencing was used to profile methylated CpG in patients with RA and ACPA–/FDR. Differentially methylated CpG loci (DML) were mapped and gene annotated. Ingenuity pathway analysis (IPA) was used for curating biomolecular networks of mapped genes. Transcript abundance was determined by quantitative (q)PCR.
Results. We identified 74 uniquely methylated CpG sites within the MHC region that were differentially methylated in patients with RA ($p < 0.05$), compared to ACPA–/FDR. Of these, 32 DML were located on 22 genes. IPA showed these genes are involved in regulating the nuclear factor–κB complex and processes involved in antigen presentation, and immune cell crosstalk in autoimmunity. Pearson correlation analysis demonstrated a negative association between differentially methylated CpG in the *C6ORF10* gene and risk factors associated with RA. Analysis by qPCR confirmed differential abundance of *C6ORF10*, *TNXB*, and *HCG18* mRNA in patients with RA compared to ACPA–/FDR.
Conclusion. Our results confirm the presence of differential methylation at specific gene loci within the MHC region of INA patients with RA. These epigenetic signatures may precede disease onset, or alternatively, may be a result of developing RA.

Key Indexing Terms: C6ORF10 gene, DNA methylation, Indigenous North American, MHC, rheumatoid arthritis, targeted bisulfite sequencing

Rheumatoid arthritis (RA) is a chronic, progressive systemic autoimmune inflammatory disease primarily affecting the peripheral joints. It is now well established that the risk of developing RA involves a complex interplay between genetic and environmental factors[1,2]. Although RA prevalence worldwide has been estimated between 0.5–1%, we and others have demonstrated

a disproportionately high RA prevalence in indigenous North American (INA) populations[3]. Risk factors for RA development in several INA populations include a high prevalence of predisposing HLA-DRB1 shared epitope (SE) encoding alleles, familial clustering of cases, high smoking rates, prevalent periodontal disease, and a high frequency of rheumatoid factor (RF) and anticitrullinated protein antibodies (ACPA) in the unaffected family members of patients with RA[4–10]. Further, studies of the Dene INA people suggest a unique clustering of biological features that may affect innate and adaptive immunity[11,12,13,14].

Located on chromosome 6 (6p21.3), the MHC encompasses 3 major gene clusters (class I, II, and III) and encodes ~252 genes, the majority of which are involved in antigen processing, presentation, and immune regulation[15]. Genetic variants in the MHC region are strongly associated with susceptibility to RA, both in seropositive and seronegative patients, although different variants are associated with these 2 major subsets[16,17]. In understanding the basis for this key association, a potential role of epigenetic changes in the MHC has been demonstrated in seminal studies that suggested that differential DNA methylation in the MHC

region may be involved in mediating the genotype-phenotype associations in RA[18,19]. Independent studies in another, predominantly white population have supported this finding[20]. The potential role of MHC epigenetic changes in predisposing INA populations to RA has not been studied despite the high prevalence of disease in these populations. We therefore studied the presence of differential methylation signatures in a cohort of INA seropositive patients with RA and unaffected seronegative first-degree relatives (FDR) in an INA population in central Canada where we have shown that RA is particularly prevalent and severe.

MATERIALS AND METHODS

Study design. INA study participants were recruited from Cree, Ojibway, and Ojicree communities in central Canada[6,11]. The Biomedical Research Ethics Board of the University of Manitoba approved the overall design of the study and consent forms (Ethics: 2005:093, Protocol: HS14453). The conduct of the study was guided by the principles of Community Based Participatory Research, a cornerstone of the Canadian Institutes of Health Research guidelines for Aboriginal health research (www.cihr-irsc. gc.ca/e/29134.html). The study participants provided informed consent after the study was explained to them in detail, with the help of an INA translator from their community, where necessary. The following 2 groups were included: (1) anticitrullinated protein antibody (ACPA)–positive patients with RA, and (2) unaffected FDR (ACPA–/FDR) who were seronegative for ACPA and RF. Patients with RA were unrelated to the ACPA–/FDR. The demographics of the study groups are summarized in Table 1. RA diagnosis was made based on fulfilling the 2010 American College of Rheumatology/European League Against Rheumatism classification criteria, as determined by a rheumatologist (HEG).

Sample collection and genomic DNA isolation. Venous blood was collected into Paxgene Blood DNA tubes (Qiagen Inc.), and processed using QIAamp DNA Blood Maxi Kit (Qiagen Inc.) per the manufacturer's instructions to isolate genomic DNA. Quality and concentration of isolated DNA was determined by Nanodrop and Qubit high sensitivity dsDNA assays. For total RNA isolation, venous blood collected into Paxgene Blood RNA tubes (Qiagen Inc.) was used.

Targeted bisulfite sequencing (TBSeq). We used CATCH-Seq target enrichment technology (Ubiquity Genomics) to perform bisulfite sequencing of coding and noncoding regions encompassing ~3.8 Mb around the MHC locus located within chromosome 6 (chr6:29530000-33500000). Library creation, bisulfite treatment, and Illumina sequencing for each DNA sample

and β value estimation for each CpG (5′–C–phosphate–G–3′) across the MHC region were performed by Ubiquity Genomics Inc. according to their standardized methodologies[21]. According to the protocol optimized by Ubiquity Genomics, all the samples had a bisulfite conversion rate > 97%. Briefly, DNA samples were sheared, end-repaired, hybridized to streptavidin-coated Dyna beads, and ligated to uniquely barcoded custom methylated/nonmethylated adapters. For target enrichment, pooled DNA samples bound to streptavidin-coated Dyna beads were collected with biotinylated probes generated from the human bacterial artificial chromosome library. Following collection, target enriched library was bisulfite-treated (EpiTect Bisulfite Kit, Qiagen Inc.), amplified by PCR, and sequenced using an Illumina HiSeq2500. Bowtie 2 was used for mapping and sequence alignment against hg19 version of reference genome and Bismark was used to calculate β scores indicating methylation level at each CpG locus across the target region after performing batch normalization. Additional information on the sequencing results is included as Supplementary Table 1 (available with the online version of this article).

Immunoassays. Serum C-reactive protein (CRP) levels were tested in serum samples using a human high-sensitivity CRP (hsCRP) ELISA kit (Biomatik) per the manufacturer's instructions. ACPA status was determined using the BioPlex 2200 System Anti-CCP reagent kit (Bio-Rad).

Genotyping of HLA-DRB1. High-resolution HLA-DRB1 genotyping was performed at the Dalhousie University HLA Typing Laboratory (Halifax, Nova Scotia, Canada) using a sequence-specific PCR method.

Total RNA extraction and quantitative real-time PCR (qRT-PCR). Total RNA was isolated from whole blood using the miRVANA isolation kit (Ambion, AM1561) per manufacturer's instructions. RNA quality was determined on Agilent Bioanalyzer using the Agilent RNA 6000 Nano kit (Agilent Technologies). Total RNA with an A260/280 > 2.0 and an RNA integrity number > 7.0 was used for downstream processing. For mRNA amplification, first-strand cDNA was synthesized from total RNA (1 μg) using Superscript VILO Mastermix (Thermo Scientific) per manufacturer's instructions. Target mRNA was amplified using Applied Biosystems Power SYBR Green Mastermix (Thermo Scientific) per manufacturer's instructions. Primers used for mRNA amplification are listed in Supplementary Table 2 (available with the online version of this article).

Data analysis and statistics. GraphPad Prism version 7.0 and SPSS v25.0 (IBM) were used for data analysis and generating volcano plots, scatter plots, and bar graphs. We performed data control as follows: first, only CpG with a minimum depth coverage > 15× were included and sites containing missing values were excluded. Upon filtering based on depth, CpG included for analysis comprised 26% of total reads. Filtered β values were $\log_2$

Table 1. Clinical characteristics of the study population.

| | TBSeq Cohort | | qPCR Cohort | |
	ACPA–/FDR, n = 17	ACPA+/RA Patients, n = 17	ACPA–/FDR, n = 12	ACPA+/RA Patients, n = 15
Age, yrs	43 (28–55)	48.5 (20–70)	40 (23–66)	46.5 (29–70)
Sex (female/male)	12/5	15/2	10/2	12/3
Disease duration, yrs	NA	6.7 (0.3–35)	NA	11.1 (0.1–35.6)
CRP titer	5.6 (1–16)	7.44 (2.25–302)	3.35 (1.07–9.25)	6.91 (2–42.6)
RF titer, IU/ml	< 20	482 (50–4110)	< 20	289 (20–1540)
Anti-CCP titer	4 (2–13)	201 (38–248)	1 (0.4–2.0)	201 (19–289)
BMI, kg/m²	32.0 (19–52)	28 (21–40)	29.54 (20–34)	27.4 (20–40)
DAS28	NA	3.55 (2.2–6.6)	NA	3.28 (1.4–6.8)

Except where indicated, all values are reported as median (range). TBSeq: targeted bisulfite sequencing; ACPA–/FDR: anticitrullinated protein antibodies–negative first-degree relatives; RA: rheumatoid arthritis; RF: rheumatoid factor; anti-CCP: anticyclic citrullinated protein antibody; CRP: C-reactive protein; NA: not applicable; BMI: body mass index; DAS28: 28-joint count Disease Activity Score.

transformed and differential methylation at each CpG site was calculated (RA-ACPA−/FDR). Benjamini-Hochberg correction was used to perform multiple correction (FDR > 5%) and differentially methylated loci (DML) with FDR-corrected p values < 0.05 were considered significant. To annotate and determine the genomic distribution of identified DML, we used diffReps script region_analysis.pl[22]. Ingenuity pathway analysis (IPA) was used to perform gene network and pathway analyses. Gene expression by qPCR was determined in samples after normalization using 18S ribosomal RNA as endogenous control, and relative fold changes were calculated using the ΔΔCt method[23].

RESULTS

Demographics of the study population. Participants were age-matched, ethnically homogeneous individuals, and about 80% of the study participants were women (Table 1). As expected, patients with RA in the TBSeq cohort (n = 17) demonstrated higher hsCRP (mean ± SD 32.23 ± 72.73 µg/ml) and ACPA levels (mean ± SD 173.56 ± 59.32 µg/ml) compared to ACPA−/FDR (n = 17). The demographics of ACPA−/FDR (n = 12) and patients with RA (n = 15) in the qPCR cohort were identical to those in the TBSeq cohort and have been described previously[4]. All the patients with RA were taking stable disease-modifying antirheumatic drugs, although the disease activity levels as determined by 28-joint count Disease Activity Score varied among the study subjects (Table 1, and Supplementary Table 3, available with the online version of this article).

Because of the known association with RA, we performed high-resolution HLA-DRB1 typing on samples from ACPA−/FDR and patients with RA included in the TBSeq cohort. Of the HLA-DRB1 alleles, there was a predominance of the SE encoding *04:04 and *14:02 alleles in both the study groups compared to other alleles. Even in this modestly sized cohort of patients and controls, bearing a single copy of SE allele was significantly associated with being an ACPA−/FDR, while a higher frequency of patients with RA had 2 copies of SE alleles (Table 2A).

Binary logistic regression analysis showed a significant association between presence of any SE alleles with body mass index (p = 0.017), while presence of 2 copies of SE associated significantly with CRP levels in this population (p = 0.021). Further, presence of *14:02 alleles associated significantly with autoantibodies [anticyclic citrullinated protein antibodies (anti-CCP) and RF] and CRP levels, suggesting that *14:02 positivity may be an important factor underlying autoantibody production and inflammation observed in patients with RA (Table 2B). The *04:04 positivity did not show association with any of the RA risk factors in our analysis.

Differences in MHC-region specific methylation between RA patients and ACPA−/FDR. Using CATCH-Seq-based complete MHC collection kits and targeted bisulfite sequencing, we examined the single-CpG site methylation levels specifically within the 3.8 Mb human MHC region (from 29.6 Mb to 33.2 Mb) of ACPA−/FDR and patients with RA. First, we identified the overall genomic distribution of all methylated CpG loci identified within the region. While 50% of the identified CpG

Table 2A. Autoantibody status and frequency of HLA-DRβ1 alleles in controls and patients with RA.

	ACPA−/FDR, n = 17	RA, n = 17	p
Any SE	15 (88.2)	13 (76.5)	0.368
SE/SE	6 (35.3)	11 (64.7)	0.086
SE/X	9 (53)	3 (17.64)	0.031
*14:02	7 (41.2)	8 (47.05)	0.730
*04:04	6 (35.3)	6 (35.3)	1.000
*09:01	1 (6.67)	1 (7.69)	1.000

Data are n (%). These HLA-DRB1 alleles were tested: 0101, 0102, 0103, 0301, 0401, 0402, 0403, 0404, 0407, 0408, 0410, 0701, 0801, 0802, 0811, 0901, 1001, 1101, 1103, 1301, 1302, 1401, 1402, 1454, 1501, 1602. Shared-epitope alleles: 0101, 0102, 0401, 0404, 0405, 0408, 0410, 0901, 1001, and 1402. Chi-square test (2-sided) was used to determine the statistical difference between the groups. RA: rheumatoid arthritis; ACPA: anticitrullinated protein antibodies; FDR: first-degree relatives; SE: shared epitope.

sites were localized to intergenic regions, 31% CpG were within the gene body, and the remaining were localized to proximal and distal promoter regions (Figure 1A). No significant difference was observed in mean methylation levels within the MHC cluster between patients with RA and ACPA−/FDR (Figure 1B; p = 0.9798; unpaired t test with Welch's correction). However, volcano plot analyses revealed the presence of 74 FDR-corrected DML out of 9782 filtered CpG loci (0.76%; p < 0.05, q < 0.05; Benjamini-Hochberg correction; Supplementary Table 4, available with the online version of this article), among which 74% of regions were hypomethylated and 26% were hypermethylated CpG sites (Figures 1C and 1D). Further, unsupervised hierarchical clustering based on these 74 DML identified within the MHC region demonstrated a distinctly independent clustering of patients and controls with minor overlap between the groups (Figure 1E). It is interesting to note that a majority of these DML could not be annotated to any specific gene in the MHC region and are located within the intronic regions. Overall, our analyses revealed the presence of intrinsic methylation differences at unique CpG loci within the MHC cluster in INA patients with RA compared to ACPA−/FDR from the same population.

IPA analysis of hypermethylated and hypomethylated genes. We performed IPA analyses on the annotated genes mapped to differentially methylated CpG to characterize the biological mechanisms predictably influenced by aberrant DNA methylation in patients with RA. The IPA biomolecular network identified nuclear factor (NF)-κB complex as a common hub regulated by both hypermethylated and hypomethylated DML-harboring genes (Figure 2A and 2B). Diseases predicted to be regulated included metabolic and immunological disorders, particularly insulin-dependent diabetes mellitus and RA (Fisher's exact test; log p value > 2.0; threshold value = 0.05). As expected, IPA also revealed that antigen presentation pathway was the most represented canonical pathway, while interferon (IFN)-α receptor was the most common upstream regulator. This suggests that differential

Table 2B. Logistic regression analysis based on *14:02 positivity in patients with RA and ACPA–/FDR.

Variables Entered on Step 1	B	SE	Wald	df	p	Exp(B)	95% CI for Exp(B)	
							Lower	Upper
Smoking	0.562	1.126	0.249	1	0.618	1.754	0.193	15.938
log2_anti-CCP	−1.492	0.635	5.515	1	0.019	0.225	0.065	0.781
log2_RF	1.529	0.668	5.241	1	0.022	4.612	1.246	17.074
log2_BMI	2.157	1.742	1.533	1	0.216	8.642	0.285	262.440
log2_CRP	1.548	0.700	4.892	1	0.027	4.702	1.193	18.536
Constant	−18.648	10.263	3.301	1	0.069	0.000		

Analysis was performed using *14:02 positivity as a dependent variable and $\log_2$-transformed values of anti-CCP, RF, BMI, and CRP. $P < 0.05$ was considered significant. RA: rheumatoid arthritis; ACPA: anticitrullinated protein antibodies; FDR: first-degree relatives; df: degrees of freedom; SE: shared epitope; Exp(B): exponentiated coefficient; anti-CCP: anticyclic citrullinated protein antibodies; RF: rheumatoid factor; BMI: body mass index; CRP: C-reactive protein.
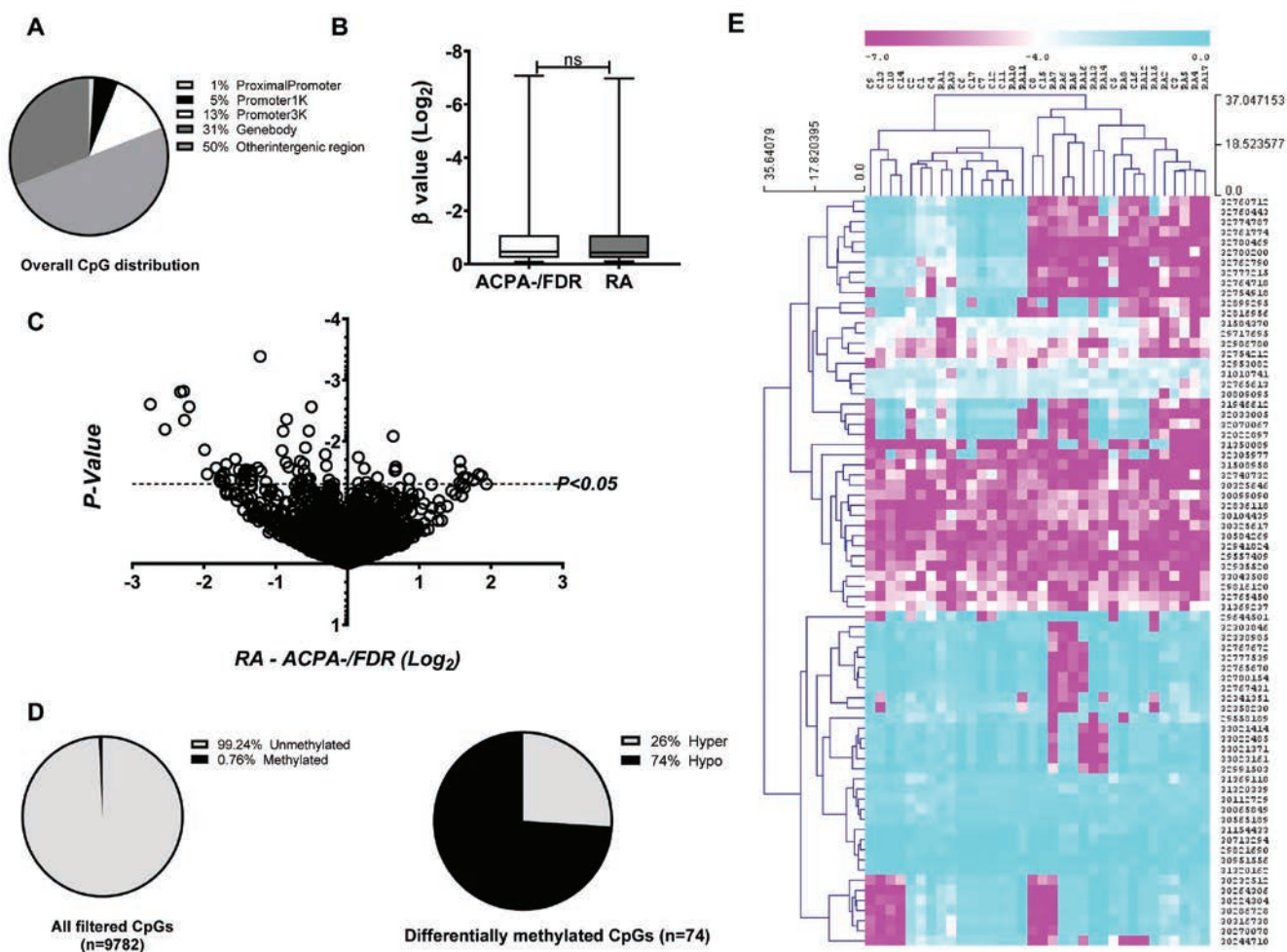

Figure 1. Analysis and distribution of MHC-specific CpG in patients with RA and ACPA–/FDR. A. Location and overall distribution of all CpG within the MHC cluster. B. Histogram of $\log_2$ normalized β values of MHC-specific CpG in whole blood of ACPA–/FDR and RA patients. C. Distribution of all DML in patients with RA. Methylated CpG that are significantly altered in both the groups are located above the horizontal dashed line corresponding to $p < 0.05$. D. Overall distribution of methylated vs unmethylated CpG within all the filtered CpG and distribution of hyper- and hypomethylated CpG within all the DML. E. Unsupervised hierarchical clustering generated using normalized β values of DML in patients with RA compared to ACPA–/FDR. Color scheme: violet (hypermethylated), cyan (hypomethylated), and white (unchanged). CpG: 5′–C–phosphate–G–3′; RA: rheumatoid arthritis; ACPA: anticitrullinated protein antibodies; FDR: first-degree relatives; DML: differentially methylated loci; ns: not significant.

epigenetic methylation in patients with RA can potentially target antigen presentation and type 1 IFN pathways (Supplementary Table 5, available with the online version of this article), which can together influence NF-κB complex and modulate RA pathogenesis. We also observed a representation of pathways involved in immune cell crosstalk and hematological system development, which are probably overrepresented in chronic diseases.

Verification of DML-annotated genes. Using UCSC genome browser, we mapped 32 out of 74 DML to 21 annotated genes (hg19). For this, we created a bed file using the genomic location of DML and obtained coordinates following addition and subtraction of 1000 bases to the genomic location. It is interesting to note that the majority of DML were located within the gene body (17 CpG loci; Table 3). Among the 32 gene-annotated DML, hypermethylated DML were harbored in 6 genes and hypomethylated DML were harbored in 15 genes (Table 3).

We downloaded a public GEO dataset (GSE42861) to verify the findings from our TBSeq data. GSE42861 is an epigenome-wide association study (EWAS) measuring the DNA methylome in the peripheral blood of 354 ACPA+ patients with RA and 337 healthy controls using Illumina HumanMethylation450 arrays. We specifically downloaded DNA methylome data of chromosome 6 and filtered it by using coordinates of the MHC region (chr6:29530000-33500000). The verification was applied to only 74 DML identified in our primary analysis. Of these, the locations of 4 DML matched exactly with the positions identified by Liu, *et al* to be differentially methylated in patients with RA[19]. However, the CpG positions of the remaining 70 DML were in close proximity (within ~20 Mb−1 kb) to the differentially methylated regions mapped in the public dataset, but in the same gene (Supplementary Table 6, available with the online version of this article). The discrepancies in methylated CpG locations can be attributed to the use of different methodologies and the sample size, which makes it difficult to demonstrate a 100% alignment between the CpG. However, identification of differentially methylated regions within the same gene in the GEO dataset indirectly confirms our findings.

Methylation differences in CpG located within the *C6ORF10* gene were previously shown to be strongly associated with risk of developing RA. Therefore, we performed a Pearson correlation analysis between RA risk factors and methylation values of *C6ORF10* DML identified in our study. Of the 4 DML, methylation at only 1 location (CpG32341351) demonstrated a significantly strong negative correlation with the anti-CCP levels (r = −0.402; p = 0.018), RF antibodies (r = −0.376; p = 0.029), the presence of 2 copies of HLA-DRB1 SE alleles (r = −0.407; p = 0.017), and the presence of *0404 SE allele (r = −0.360; p = 0.037).

Considering that CpG methylation in DNA is strongly implicated in regulating gene transcription[24], it is important to understand the functional relationship between altered
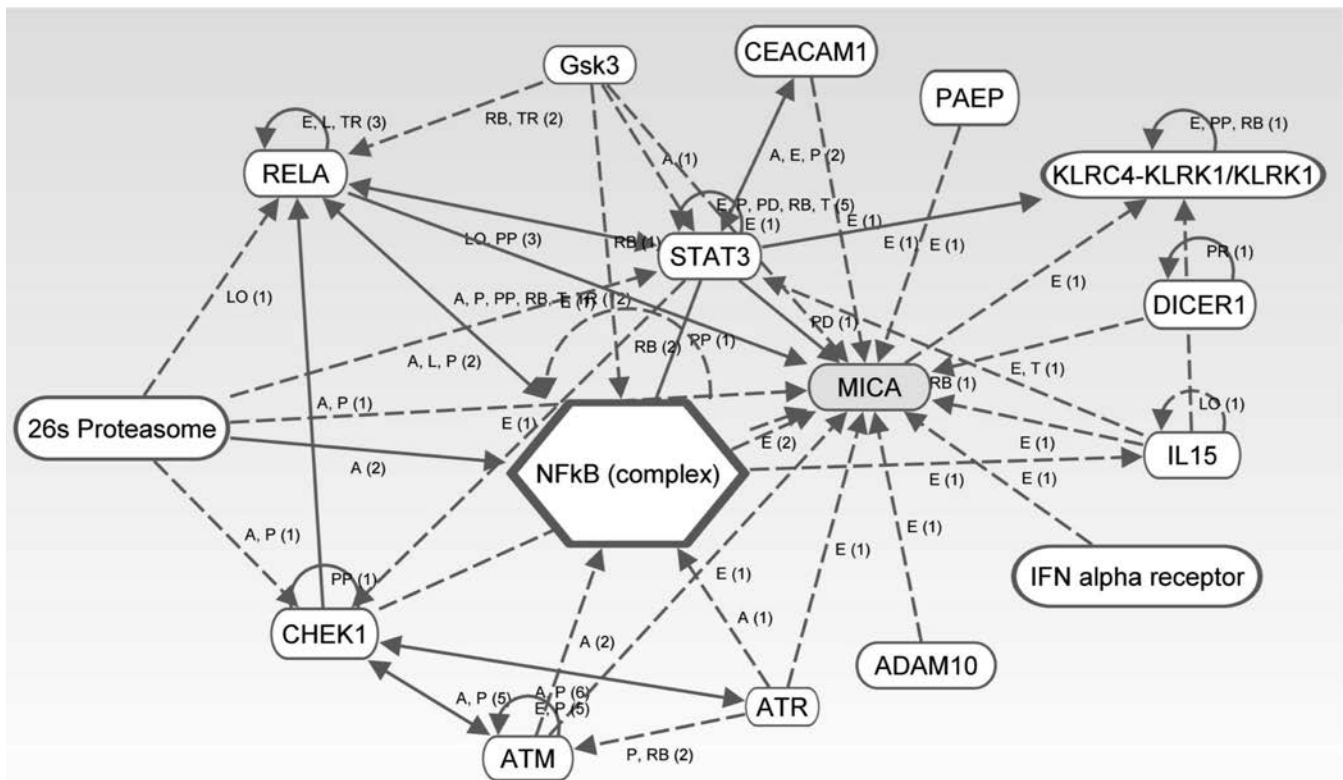

Figure 2A. Curated molecular interactions between genes associated with hypermethylated DML, in patients with RA. RA: rheumatoid arthritis; DML: differentially methylated loci; NF-κB: nuclear factor-κB; MICA: MHC class I–related chain A; IL: interleukin; IFN: interferon.
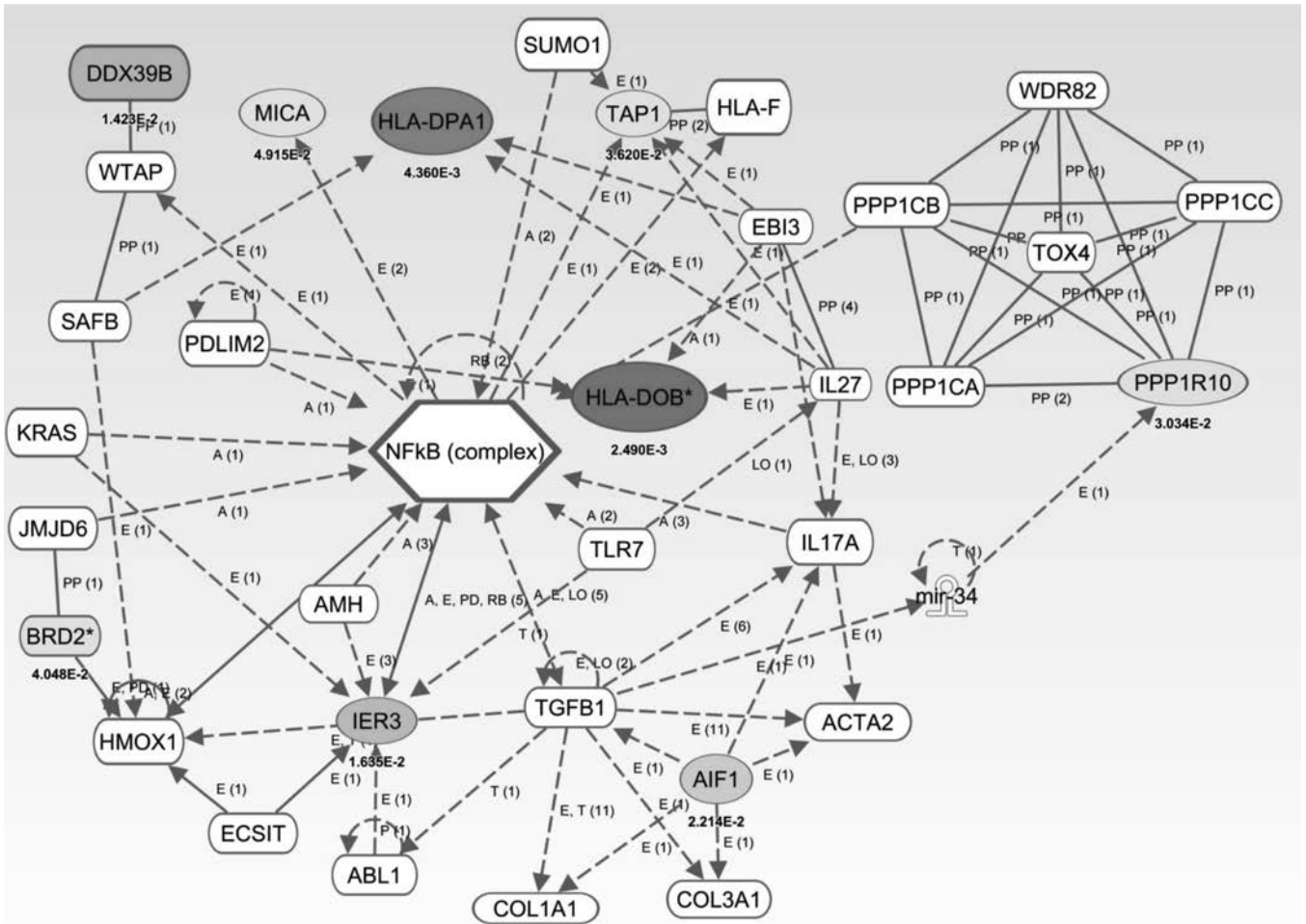
Figure 2B. Curated molecular interactions between genes associated with hypomethylated DML, in patients with RA. RA: rheumatoid arthritis; DML: differentially methylated loci; NF-κB: nuclear factor-κB; MICA: MHC class I–related chain A; IL: interleukin.

methylation and the relative abundance of genes associated with DML. Therefore, we validated bisulfite sequencing data by performing gene expression analyses by qPCR in mRNA isolated from whole blood samples collected from an independent subset of INA people (qPCR cohort). These individuals are nested within the same INA cohort and are independent from those on whom pyrosequencing was performed. Specifically, we focused on determining mRNA abundance of genes associated with multiple DML located either within the genebody or promoter1k regions (Figure 3). Based on this, we selected 5 genes for the qPCR analysis: *C6ORF10* (4 DML), *HLA-DOB* (3 DML), *HCG18* (3 DML), *TNXB* (3 DML), and *TRIM40* (2 DML). Among the genes harboring multiple hypomethylated DML, mRNA abundance of *C6ORF10* was significantly higher in patients with RA compared to ACPA–/FDR. Even though an increasing trend and a dichotomous pattern was observed in *TNXB* abundance, the effect was statistically nonsignificant. *HLA-DOB* transcript levels remained unaltered between patients with RA and ACPA–/FDR. A significantly lesser abundance of *HCG18* transcript was observed in patients with RA relative to ACPA–/FDR, while *TRIM40* abundance was

unchanged. This suggests that differential methylation at specific loci on some of these genes can influence their gene expression.

DISCUSSION
In our present study, we compared the MHC-specific DNA methylation patterns in whole blood samples from a cohort of autoantibody-positive INA patients with RA and unrelated seronegative FDR with no clinically detectable autoimmunity. Using bisulfite sequencing, we identified distinct and differentially methylated CpG loci within MHC region of patients with RA, several of which were annotated to specific genes that regulate key pathways in RA pathogenesis, such as antigen presentation and immune cell crosstalk through the NF-κB complex.

DNA methylation, particularly in CpG islands embedded with gene promoter regions, is significantly associated with linking genetic susceptibility to phenotypic traits[25]. Microarray-based EWAS on European whites demonstrated directionally similar differential methylation patterns in the MHC region of patients with RA compared to healthy controls[19,20]. We integrated these observations and applied a high-throughput, cost-effective, and accurate pyrosequencing methodology (TBSeq) to profile

Table 3. List of gene-annotated hypermethylated and hypomethylated DML.

CpG	p	Location	Gene	β Difference
Hypomethylated DML				
32341351	0.020	Promoter3k	*C6orf10*	−1.564
32305977	0.024	Genebody	*C6orf10*	−1.689
32303846	0.029	Genebody	*C6orf10*	−1.509
32338985	0.031	Promoter1k	*C6orf10*	−1.208
32780469	0.002	Genebody	*HLA-DOB*	−2.749
32780200	0.006	Genebody	*HLA-DOB*	−2.544
32780154	0.036	Genebody	*HLA-DOB*	−1.452
32070067	0.035	Genebody	*TNXB*	−1.954
32033005	0.043	Genebody	*TNXB*	−1.623
32022897	0.047	Genebody	*TNXB*	−1.764
32941824	0.040	Promoter1k	*BRD2*	−0.524
32935520	0.043	Promoter1k	*BRD2*	−0.572
31369237	0.049	Promoter1k	*MICA*	−0.650
32816956	0.036	Genebody	*TAP1*	−1.754
31946612	0.038	Genebody	*STK19*	−1.768
32358230	0.044	ProximalPromoter	*HCG23*	−1.349
33043508	0.004	Genebody	*HLA-DPA1*	−0.847
31508958	0.014	Promoter3k	*DDX39B*	−0.912
30713294	0.016	Promoter1k	*IER3*	−0.309
31584370	0.022	Promoter1k	*AIF1*	−0.856
29557409	0.023	Promoter3k	*OR2H2*	−0.609
29717695	0.026	Promoter1k	*HLA-F-AS1*	−0.820
30584269	0.030	Promoter1k	*PPP1R10*	−0.603
Hypermethylated DML				
30286728	0.040	Genebody	*HCG18*	1.672
30270078	0.048	Genebody	*HCG18*	1.564
30264306	0.037	Genebody	*HCG18*	1.870
30112729	0.028	Genebody	*TRIM40*	0.233
30104439	0.030	ProximalPromoter	*TRIM40*	0.661
30232512	0.021	Genebody	*HLA-L*	1.568
29644501	0.039	Promoter1k	*ZFP57*	1.620
31369118	0.039	Promoter1k	*MICA*	0.417
30951556	0.018	Genebody	*MUC21*	0.153

β difference refers to the difference in mean methylation level at the specific CpG site between patients with RA and ACPA−/FDR (mean methylation level in low-fatigue subtracted from mean methylation level in high-fatigue patients). DML: differentially methylated CpG loci; Chr: chromosome 6; RA: rheumatoid arthritis; ACPA−/FDR: anticitrullinated protein antibodies–negative first-degree relatives.

methylated CpG loci specifically in and around the MHC region in a predisposed non-white population, aiming to replicate previously reported findings.

EWAS studies by Liu, *et al* and others underscore the importance of identifying altered DNA methylation loci within the MHC region, and describe their effect on gene expression and delineating molecular mechanisms that can eventually influence RA phenotype[19,20]. In our study, we observed a representation of multiple DML within the genebody of certain annotated genes (*C6ORF10, TNXB, HLA-DOB,* and *HCG18*), which indicates the presence of a higher methylation frequency and can possibly regulate their transcript expression. *C6ORF10, TNXB,* and *HCG18* are novel RA-associated pleiotropic genes also identified in various genome-wide association studies and EWAS, whose physiological functions are yet to be deciphered[19,26,27,28,29]. A strong negative correlation between DML in *C6ORF10* and RA-associated autoantibodies (ACPA and RF) suggests that hypomethylated *C6ORF10* might be involved in amplifying the

levels of circulating RA autoantibodies in genetically susceptible INA individuals with SE alleles. In fact, disease-related functional gene networks found using the mapped SNP database identify *C6ORF10* as a functional hotspot interacting with at least 14 different genes that are implicated in autoimmune conditions such as systemic lupus erythematosus, type 1 diabetes, psoriasis, and RA[26]. Therefore, additional investigations are warranted to further unravel the role of *C6ORF10* in mediating the onset of inflammatory arthritis in INA and other populations.

We applied computational predictive analyses using IPA to identify functional biological relevance and potentially common downstream signaling networks regulated by DML-harboring genes found in patients with RA. Emergence of NF-κB complex as a central node and autoimmune disorders as the top-related diseases associated with the differentially methylated MHC-specific genes is not surprising. Of the 98 gene loci established to have significant risk to RA
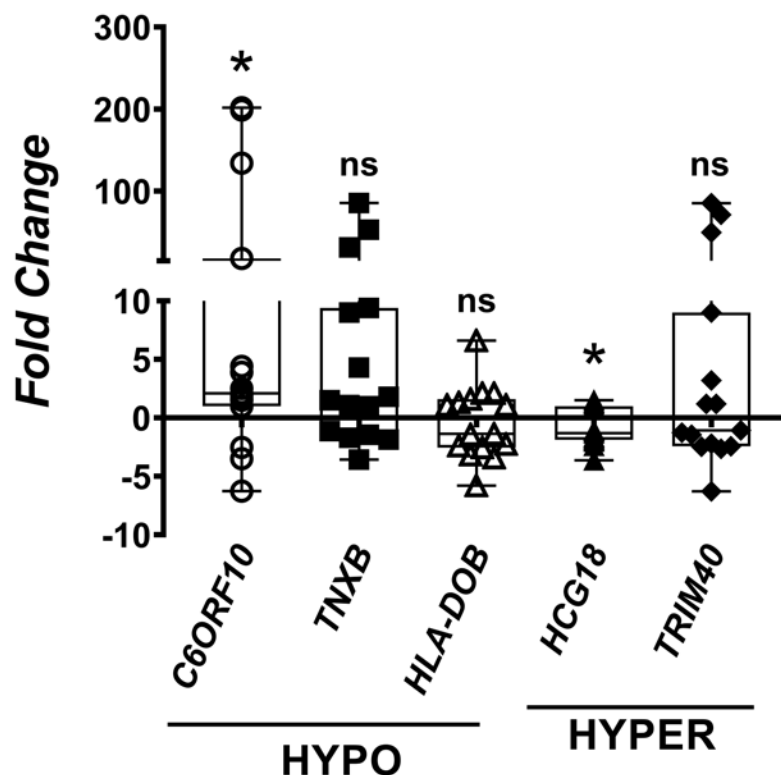
Figure 3. mRNA abundance by qPCR. Transcript abundance in patients with RA. mRNA abundance of target genes with multiple DML (*C6ORF10, TNXB, HLA-DOB, BRD2, HCG18,* and *TRIM40*) was determined by qPCR using gene-specific primers in RNA isolated from WB of ACPA–/FDR (controls) and patients with RA (cases) from the qPCR cohort. Data are represented as fold change relative to ACPA–/FDR. Error bars represent median value and are analyzed by Mann-Whitney U test (* p < 0.05). NS: nonsignificant; qPCR: quantitative real-time polymerase chain reaction; RA: rheumatoid arthritis; DML: differentially methylated loci; WB: whole blood; ACPA: anticitrullinated protein antibodies; FDR: first-degree relatives.

susceptibility, a vast majority are involved in NF-κB signaling, thus establishing the role of NF-κB complex in RA pathogenesis[30]. Activation of NF-κB complex in the rheumatoid synovium, either canonically or noncanonically, upregulates proinflammatory cytokine expression, augments immune cell recruitment, and promotes proliferation of synovial fibroblasts leading to joint destruction[31]. Network analysis identified an overrepresentation of pathways, presumably regulated upstream by IFN-α receptor, and involved in immune cell crosstalk and hematological cell development in the interactome of both hypo- and hypermethylated genes (Supplementary Table 4, available with the online version of this article). While MHC class I–related chain A (MICA) is key to crosstalk between dendritic and NK cells and might play a key role in regulation of autoreactivity[32], *HLA-DPA1, HLA-DOB,* and *TAP1* are involved in multiple immune cell–associated pathways such as interleukin 4 signaling and antigen presentation[33,34,35,36]. Molecules such as AIF1, HLA-DPA1, MICA, and TNXB are involved in vasculogenesis, erythropoietic cell reprogramming,

phosphoinositide-3-kinase–protein kinase B/Akt signaling, and T cell proliferation[37,38,39].

Key limitations of our study include small sample size and use of whole blood DNA as a starting material for methylation profiling. Considering that the primary objective of our study was to validate the observations made in previously published EWAS, we partially addressed the issue of small sample size by using an MHC-specific TBSeq technique and validating our findings in an independent qPCR cohort using gene-specific primers. Also, the qPCR cohort was different from the TBSeq cohort because of logistical limitations. However, we attempted to minimize variability by selecting a comparable group of individuals belonging to the same population. We acknowledge our inability to address the association between cell-type heterogeneity to methylation levels at specific DML identified in INA patients with RA because of the unavailability of reference DNA methylation database tailored to chromosome 6 and our region of interest in different blood cell types. Our study findings may also be affected by exposure to

RA medications and require future investigations to delineate their effect.

We present evidence that the presence of differential MHC-associated methylation profile is detectable in the whole blood of seropositive patients with RA from an INA population, with DML mapped to unique genes whose role in autoimmune RA is yet unknown. Multiple methylation studies, including ours, undeniably identify *C6ORF10* as a novel gene that is significantly dysregulated and can potentially modulate inflammatory processes involved in RA pathogenesis and possibly other autoimmune conditions across populations. Additional research to identify the interaction network of *C6ORF10* in FDR of patients with RA will delineate its functional contribution to RA onset in a high-risk INA population.

ONLINE SUPPLEMENT
Supplementary material accompanies the online version of this article.

REFERENCES
1. Liao KP, Alfredsson L, Karlson EW. Environmental influences on risk for rheumatoid arthritis. Curr Opin Rheumatol 2009; 21:279-83.
2. Ospelt C, Gay S, Klein K. Epige netics in the pathogenesis of RA. Semin Immunopathol 2017;39:409-19.
3. McDougall C, Hurd K, Barnabe C. Systematic review of rheumatic disease epidemiology in the indigenous populations of Canada, the United States, Australia, and New Zealand. Semin Arthritis Rheum 2017;46:675-86.
4. Anaparti V, Smolik I, Meng X, Spicer V, Mookherjee N, El-Gabalawy H. Whole blood microRNA expression pattern differentiates patients with rheumatoid arthritis, their seropositive first-degree relatives, and healthy unrelated control subjects. Arthritis Res Ther 2017;19:249.
5. Deane KD, El-Gabalawy H. Pathogenesis and prevention of rheumatic disease: focus on preclinical RA and SLE. Nat Rev Rheumatol 2014;10:212-28.
6. El-Gabalawy HS, Robinson DB, Hart D, Elias B, Markland J, Peschken CA, et al. Immunogenetic risks of anti-cyclical citrullinated peptide antibodies in a North American native population with rheumatoid arthritis and their first-degree relatives. J Rheumatol 2009;36:1130-5.
7. Mookherjee N, El-Gabalawy HS. High degree of correlation between whole blood and PBMC expression levels of miR-155 and miR-146a in healthy controls and rheumatoid arthritis patients. J Immunol Methods 2013;400-1:106-10.
8. Oen K, Robinson DB, Nickerson P, Katz SJ, Cheang M, Peschken CA, et al. Familial seropositive rheumatoid arthritis in North American native families: effects of shared epitope and cytokine genotypes. J Rheumatol 2005;32:983-91.
9. Peschken CA, Hitchon CA, Robinson DB, Smolik I, Barnabe CR, Prematilake S, et al. Rheumatoid arthritis in a North American native population: longitudinal followup and comparison with a white population. J Rheumatol 2010;37:1589-95.
10. Scally SW, Law SC, Ting YT, Heemst JV, Sokolove J, Deutsch AJ, et al. Molecular basis for increased susceptibility of indigenous North Americans to seropositive rheumatoid arthritis. Ann Rheum Dis 2017;76:1915-23.
11. El-Gabalawy HS, Robinson DB, Smolik I, Hart D, Elias B, Wong K, et al. Familial clustering of the serum cytokine profile in the relatives of rheumatoid arthritis patients. Arthritis Rheum 2012;64:1720-9.
12. Hitchon CA, Sun Y, Robinson DB, Peschken CA, Bernstein CN, Siminovitch KA, et al. Vitamin D receptor polymorphism rs2228570 (Fok1) is associated with rheumatoid arthritis in North American natives. J Rheumatol 2012;39:1792-7.
13. Larcombe L, Mookherjee N, Slater J, Slivinski C, Dantouze J, Singer M, et al. Vitamin D, serum 25(OH)D, ll-37 and polymorphisms in a Canadian first nation population with endemic tuberculosis. Int J Circumpolar Health 2015;74:28952.
14. Larcombe L, Mookherjee N, Slater J, Slivinski C, Singer M, Whaley C, et al. Vitamin D in a northern Canadian first nation population: dietary intake, serum concentrations and functional gene polymorphisms. PLoS One 2012;7:e49872.
15. Kilding R, Wilson AG. Mapping of a novel susceptibility gene for rheumatoid arthritis in the telomeric MHC region. Cytokine 2005;32:71-5.
16. Newton JL, Harney SM, Wordsworth BP, Brown MA. A review of the MHC genetics of rheumatoid arthritis. Genes Immun 2004;5:151-7.
17. Kim K, Bang SY, Lee HS, Bae SC. Update on the genetic architecture of rheumatoid arthritis. Nat Rev Rheumatol 2017;13:13-24.
18. Cribbs A, Feldmann M, Oppermann U. Towards an understanding of the role of DNA methylation in rheumatoid arthritis: therapeutic and diagnostic implications. Ther Adv Musculoskelet Dis 2015;7:206-19.
19. Liu Y, Aryee MJ, Padyukov L, Fallin MD, Hesselberg E, Runarsson A, et al. Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. Nat Biotechnol 2013;31:142-7.
20. van Steenbergen HW, Luijk R, Shoemaker R, Heijmans BT, Huizinga TW, van der Helm-van Mil AH. Differential methylation within the major histocompatibility complex region in rheumatoid arthritis: a replication study. Rheumatology 2014;53:2317-8.
21. Day K, Song J, Absher D. Targeted sequencing of large genomic regions with CATCH-Seq. PLoS One 2014;9:e111756.
22. Shen L, Shao NY, Liu X, Maze I, Feng J, Nestler EJ. diffReps: detecting differential chromatin modification sites from ChIP-seq data with biological replicates. PLoS One 2013;8:e65598.
23. Pfaffl MW. A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res 2001;29:e45.
24. Siegfried Z, Simon I. DNA methylation and gene expression. Wiley Interdiscip Rev Syst Biol Med 2010;2:362-71.
25. Szyf M, Bick J. DNA methylation: a mechanism for embedding early life experiences in the genome. Child Dev 2013;84:49-57.
26. Zheng W, Rao S. Knowledge-based analysis of genetic associations of rheumatoid arthritis to inform studies searching for pleiotropic genes: a literature review and network analysis. Arthritis Res Ther 2015;17:202.
27. Kaufman CS, Butler MG. Mutation in TNXB gene causes moderate to severe Ehlers-Danlos syndrome. World J Med Genet 2016; 6:17-21.

28. Frangogiannis NG. Matricellular proteins in cardiac adaptation and disease. Physiol Rev 2012;92:635-88.

29. Xi Y, Jiang T, Wang W, Yu J, Wang Y, Wu X, et al. Long non-coding HCG18 promotes intervertebral disc degeneration by sponging miR-146a-5p and regulating TRAF6 expression. Sci Rep 2017;7:13234.

30. Castro-Santos P, Diaz-Pena R. Genetics of rheumatoid arthritis: a new boost is needed in Latin American populations. Rev Bras Reumatol Engl Ed 2016;56:171-7.

31. Sun SC. The non-canonical NF-κB pathway in immunity and inflammation. Nat Rev Immunol 2017;17:545-58.

32. Poggi A, Zocchi MR. Antigen presenting cells and stromal cells trigger human natural killer lymphocytes to autoreactivity: evidence for the involvement of natural cytotoxicity receptors (NCR) and NKG2D. Clin Dev Immunol 2006;13:325-36.

33. Afroz S, Giddaluru J, Vishwakarma S, Naz S, Khan AA, Khan N. A comprehensive gene expression meta-analysis identifies novel immune signatures in rheumatoid arthritis patients. Front Immunol 2017;8:74.

34. Kang YJ, Zeng W, Song W, Reinhold B, Choi J, Brusic V, et al. Identification of human leucocyte antigen (HLA)-A*0201-restricted cytotoxic T lymphocyte epitopes derived from HLA-DOβ as a novel target for multiple myeloma. Br J Haematol 2013;163:343-51.

35. Peaper DR, Cresswell P. Regulation of MHC class I assembly and peptide binding. Annu Rev Cell Dev Biol 2008;24:343-68.

36. Tosello V, Zamarchi R, Merlo A, Gorza M, Piovan E, Mandruzzato S, et al. Differential expression of constitutive and inducible proteasome subunits in human monocyte-derived DC differentiated in the presence of IFN-alpha or IL-4. Eur J Immunol 2009; 39:56-66.

37. Tian Y, Jain S, Kelemen SE, Autieri MV. AIF-1 expression regulates endothelial cell activation, signal transduction, and vasculogenesis. Am J Physiol Cell Physiol 2009;296:C256-66.

38. Tangprasittipap A, Kaewprommal P, Sripichai O, Sathirapongsasuti N, Satirapod C, Shaw PJ, et al. Comparison of gene expression profiles between human erythroid cells derived from fetal liver and adult peripheral blood. PeerJ 2018;6:e5527.

39. Yan SP, Chu DX, Qiu HF, Xie Y, Wang CF, Zhang JY, et al. LncRNA LINC01305 silencing inhibits cell epithelial-mesenchymal transition in cervical cancer by inhibiting TNXB-mediated PI3K/ Akt signalling pathway. J Cell Mol Med 2019;23:2656-66.