JMolMed

ORIGINAL ARTICLE

Gene expression profiling and association studies implicate the neuregulin signaling pathway in Behçet's disease susceptibility

**Joana M. Xavier · Tiago Krug · Fereydoun Davatchi · Farhad Shahram ·
Benedita V. Fonseca · Gorete Jesus · Filipe Barcelos · Joana Vedes ·
Manuel Salgado · Bahar Sadeghi Abdollahi · Abdolhadi Nadji ·
Maria Francisca Moraes-Fontes · Niloofar Mojarad Shafiee ·
Fahmida Ghaderibarmi · José Vaz Patto · Jorge Crespo ·
Sofia A. Oliveira**

Abstract Behçet's disease (BD) is a complex disease with genetic and environmental risk factors implicated in its etiology; however, its pathophysiology is poorly understood. To decipher BD's genetic underpinnings, we combined gene expression profiling with pathway analysis and association studies. We compared the gene expression profiles in peripheral blood mononuclear cells (PBMCs) of 15 patients and 14 matched controls using Affymetrix microarrays and found that the neuregulin signaling pathway was over-represented among the differentially expressed genes. The Epiregulin (*EREG*), Amphiregulin (*AREG*), and Neuregulin-1 (*NRG1*) genes of this pathway stand out as they are also among the top differentially expressed genes. Twelve haplotype tagging SNPs at the *EREG-AREG* locus and 15 SNPs in *NRG1* found associated in at least one published BD genome-wide association study were tested for association with BD in a dataset of 976 Iranian patients and 839 controls. We found a novel association with BD for the rs6845297 SNP located downstream of *EREG*, and replicated three associations at *NRG1* (rs4489285, rs383632, and rs1462891). Multifactor dimensionality reduction analysis indicated the existence of epistatic interactions between *EREG* and *NRG1* variants. *EREG-AREG* and *NRG1*, which are members of the epidermal growth factor (EGF) family, seem to modulate BD susceptibility through main effects and gene–gene interactions. These association findings support a role for the EGF/ErbB signaling pathway in

J. M. Xavier · T. Krug · B. V. Fonseca · S. A. Oliveira (✉)
Instituto de Medicina Molecular, Faculdade de Medicina da
Universidade de Lisboa, Av. Prof. Egas Moniz,
Edifício Egas Moniz,
1649-028 Lisboa, Portugal
e-mail: aaoliveira@fm.ul.pt

J. M. Xavier · T. Krug · B. V. Fonseca · S. A. Oliveira
Instituto Gulbenkian de Ciência, Oeiras, Portugal

F. Davatchi · F. Shahram · B. S. Abdollahi · A. Nadji ·
N. M. Shafiee · F. Ghaderibarmi
Rheumatology Research Center, Tehran University
of Medical Sciences, Tehran, Iran

G. Jesus
Hospital Infante D. Pedro, Aveiro, Portugal

F. Barcelos · J. Vaz Patto
Instituto Português de Reumatologia,
Lisboa, Portugal

J. Vedes
Hospital de Sousa Martins, Guarda, Portugal

M. Salgado
Hospital Pediátrico de Coimbra,
Coimbra, Portugal

M. F. Moraes-Fontes
Hospital Curry Cabral, Lisboa, Portugal

J. Crespo
Hospitais da Universidade de Coimbra,
Coimbra, Portugal

Springer

BD pathogenesis that warrants further investigation and highlight the importance of combining genetic and genomic approaches to dissect the genetic architecture of complex diseases.

Keywords Genetics · Microarray · Association study · Behçet's disease · Genetic epidemiology

Introduction

Behçet's disease (BD) is a systemic immuno-inflammatory disease characterized by the peculiar form it affects the organism, especially in the mucocutaneous and ocular structures. Its most common clinical manifestations are recurrent orogenital ulcerations and inflammatory eye lesions. Although its etiopathogenesis remains to be elucidated, genetic (e.g., HLA-B51) and environmental factors as well as immune-mediated mechanisms have been pointed out as important factors [1].

Genome-wide association studies (GWAS) for BD recently revealed two novel susceptibility loci at a genome-wide significant level: *IL23R-IL12RB2* and *IL10* [2, 3]. However, the genetic susceptibility to BD is governed by additional genes that cannot be pinpointed using this approach. To identify additional genetic risk factors for BD, we hereby pursued another strategy which intersects data from a genome-wide expression study with an in silico pathway analysis and association studies.

Gene expression profiling allows the simultaneous monitoring of the transcriptional behavior of thousands of genes, and genes differentially expressed between patients and controls in a relevant tissue are likely to be involved at some stage of the pathogenic process. Gene profiling experiments have allowed the identification of novel and unsuspected pathways involved in disease etiology, including in rheumatic and immune diseases [4–6]. Pathway analysis of differentially expressed genes is a useful approach to highlight undetected patterns in the data and to determine which differentially expressed genes to follow-up in independent assays such as association studies. This convergence of genomic and genetic approaches enables data mining beyond the top findings. With this strategy, we were able to highlight the role of the neuregulin signaling pathway in BD pathogenesis and strengthen the previously observed association of variants in *NRG1* with BD.

Materials and methods

Gene profiling study subjects

Fourteen BD patients and 15 healthy controls were enrolled in the gene expression study through the Hospital Infante D. Pedro, Hospital de Sousa Martins, Hospital Pediátrico de Coimbra and Instituto Português de Reumatologia in Portugal. These participants are of Portuguese Caucasian origin. All patients satisfied the International Criteria for Behçet's disease (ICBD) [7]. Patients were considered immunosuppressed if medicated with azathioprine, cyclosporine, deflazacort, or thalidomide. BD patients with age-at-diagnosis after 60 years were excluded. Controls were evaluated using the same evaluation procedures as the cases and selected when negative for BD and any other rheumatologic or autoimmune disorder. This research was approved by the ethics committee at Hospital Infante D. Pedro and Instituto Português de Reumatologia where the samples were collected. All participants were informed of the study and provided informed written consent.

Total RNA isolation and microarray hybridization

Whole blood samples were obtained by venipuncture and collected in BD Vacutainer CPT tubes (BD, Franklin Lakes, NJ, USA). These samples were centrifuged to isolate PBMCs, which were then washed twice. Total RNA was extracted from PBMCs using RNeasy Mini kit (Qiagen, Hilden, Germany) according to the protocol recommended by the manufacturer. Total RNA (3.5 μg) from each individual was hybridized to GeneChip Human Genome U133 Plus 2.0 microarrays (Affymetrix, Santa Clara, CA, USA) at the Instituto Gulbenkian de Ciência's Affymetrix Core Facility following manufacturer's protocols. Extensive quality control checks were performed in all steps of the process.

Gene expression data normalization and statistical analysis

The intensity array data were analyzed with their respective CDF file from Affymetrix on the Partek software (Partek Incorporated, St Louis, MO, USA). Background correction, normalization, and summarization of the CEL files were performed using the robust multichip average algorithm. Differentially expressed genes among cases and controls were identified with analysis of variance (ANOVA), adjusting for known experimental (immunosuppression status) and study design (geographic origin and scan date) covariates. The genes with more than 1.20-fold-change and $P \leq 0.05$ were considered differentially expressed. To account for multiple testing, we calculated false discovery rates (FDR) using the Partek Q value method [8].

The gene expression profiling was conducted and reported in accordance with the minimum information about microarray experiment (MIAME) guidelines [9]. The microarray data were deposited on Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/projects/geo/) with the accession number GSE17114.

Hierarchical clustering (HC) analysis was performed using the Partek software. The effects of the geographic origin of the participants, the scan date of the microarrays

and the immunosuppression status of the participants were removed using the batch-remove tool implemented in Partek prior to visualization, since these visualization tools cannot correct for study design batch effects.

Quantitative real-time polymerase chain reaction confirmation of some of the microarray results was not performed because of an insufficient amount of RNA available for several of the controls and patients used in this study.

Pathway analyses

Ingenuity Pathway Analysis 9.0 (IPA; Ingenuity Systems, http://www.ingenuity.com) was used to functionally annotate genes according to molecular networks, biological function, and canonical pathways. An IPA list is built upon a very large manually curated and up-to-date database of genes, proteins, functions, interactions/networks, and pathways. Functional analysis was performed on the 621 probe IDs differentially expressed ($P \leq 0.05$ and absolute fold-change>1.20). The IPA "Core Analysis" was run using the Human Genome U133 Plus 2.0 array as the reference set, using direct and indirect relationships for network analysis, including endogenous chemicals, data from Human species only, experimental observed confidence, and all tissues, cell lines, and data sources. The significance (P values) of the association between a dataset and a canonical pathway was determined by comparing the number of genes in a dataset that participate in a given pathway to the total number of occurrences of these genes in all pathway annotations. A Fisher's exact test was used to calculate the P value to determine the probability that the association between the genes in the dataset and the canonical pathway is explained only by chance. The level of statistical significance was set to $P \leq 0.05$.

Association study subjects

Patients (976) and 839 controls, ascertained and collected as described previously [10], were included in the association study. Diagnosis of BD was made according to the ICBD criteria [7]. This research was approved by the ethics committee at the Tehran University for Medical Sciences, Iran. All participants were informed of the study and provided informed written consent.

Genotyping

Thirteen haplotype tagging SNPs (htSNP) at the *EREG-AREG* locus (chr. 4: 75449724–75545341 bp) were identified in Haploview 4.1 [11] using genotypes of 30 European (CEU) trios downloaded from the HapMap Release 24/phase II Nov08 (NCBI B36 assembly, dbSNP b126), and with the following options: pairwise mode, $r^2>0.80$, and minor allele frequency (MAF)>0.1. Fifteen SNPs with

a $r^2<0.8$ in *NRG1* (chr. 8: 31517663–32066776 bp) found associated with BD in either one of two published GWAS (personal communications by Dr. Elaine Remmers and Dr. Nobuhisa Mizuki) were genotyped.

Genomic DNA was extracted from whole blood samples using a salting out procedure. SNPs were genotyped using Sequenom's (San Diego, USA) iPlex assay (primer extension of multiplex products with detection by matrix assisted laser desorption/ionization time-of-flight mass spectrometry) following the manufacturer's protocol and detected in a Sequenom MassArray K2 platform. The primer sequences are available in Supplementary Table 1 and were designed using Sequenom's MassARRAY ® Assay Design 3.0 software. The genotyping was performed at the Instituto Gulbenkian de Ciência's Genomics Unit.

Extensive quality control was performed using eight HapMap (http://hapmap.ncbi.nlm.nih.gov/) controls of diverse ethnic affiliation, sample duplication within and across plates, Mendelian inheritance check in three large pedigrees, Hardy–Weinberg equilibrium (HWE) in the control group ($P>0.01$), and a minimum of 95 % call rate for each SNP. Genotype determinations were performed blinded to affection status.

Association analyses

Student's unpaired t tests and chi-square tests were used to compare quantitative and qualitative clinical and demographic data, respectively, between BD cases and controls. χ^2 tests for HWE in controls and crude allelic association of SNPs with BD risk were performed using Haploview 4.1. The SNPassoc v.1.4-9 package [12] implemented in the R freeware (http://cran.r-project.org/) was used to assess the associations with BD adjusted for gender. Odds ratios (ORs) and their associated 95 % confidence intervals (CIs) were calculated for SNPs with significant allelic associations with BD. Results were considered significant below the conventional level of 0.05.

Meta-analyses

GWAS data for *EREG-AREG* and *NRG1* SNPs (e.g., allele counts, OR, CI) was obtained by personal communications from Drs. Ahmet Gul and Elaine F. Remmers for the Turkish dataset [2], and from Drs. Shigeaki Ohno, Nobuhisa Mizuki and Akira Meguro for the Japanese dataset [3]. Fixed effects (Mantel-Haenszel) meta-analyses were performed using the rmeta package in R.2.7.2 and PLINK v1.047.

Gene–gene interactions

Epistasis between *EREG-AREG* and *NRG1* variants was detected and characterized using the multifactor dimensionality reduction (MDR) method (v2.0 beta 8.3) (www.epistasis.org) [13]. Missing genotypes were first imputed for each SNP using

the PLINK database (http://pngu.mgh.harvard.edu/purcell/plink/) [14] with genotypes from individuals in the CEU HapMap population as a reference. MDR reduces high-dimensional data into a single dimension data by pooling multilocus genotypes into "high- risk" and "low-risk" groups according to ratio of affected and unaffected individuals within each genotype combination. This new multilocus variable is then tested for its ability to predict disease status. Cross validation and testing balance accuracy (TBA) are used to select the best models [13, 15]. The statistical significances of the best models were calculated after 1,000 permutations using the MDR Permutation Tool (v1.0, beta 2). Entropy-based interaction dendrograms were drawn to interpret epistasis models.

Results

Gene expression study

The principal clinical and demographic characteristics of the 29 individuals used in the gene profiling study are shown in Table 1. Cases and controls were matched for age-at-examination (AEE) (mean AAE±SD of 37.1±11.0 years in cases and 36.7±13.5 years in controls, $P=9.39E-01$). All of the microarrays performed were of good quality (average±SD of present calls and of background were 45.0±3.0 % and 42.0±6.1 %, respectively).

Using analysis of variance on the normalized expression data, 621 probe sets representing 508 genes (Supplementary Table 2) were found differentially expressed among BD cases and controls with a threshold of 1.20-fold-change, and $P \leq 0.05$. 373 of these probe sets (representing 314 genes) were down-regulated. All the probe sets with $P \leq 0.05$ also had a Q value ≤ 0.05. The false discovery rate was determined based on Q values, as these have a higher apparent power when compared to other standard methods [8]. The hierarchical cluster diagram shows a distinct gene expression profile between patients and controls that clearly separates the two groups of samples (Fig. 1). The top 30 genes with larger fold-change differences between cases and controls are shown in Table 2. EREG, S100B (S100 calcium binding protein B), NAMPT (nicotinamide phosphoribosyltransferase), AREG, and NRG1 were the top five genes under-expressed in patients, and HBG1/HBG2 (hemoglobin subunit gamma-1/gamma-2), SRSF6 (serine/arginine-rich splicing factor 6), LGALS2 (lectin, galactoside-binding, soluble, 2), NEBL (nebulette), and TRD@ (T cell receptor delta locus) the top five over-expressed genes in BD patients.

Pathway analysis

Pathway analysis of the differentially expressed genes was performed using the Ingenuity Pathway Analysis (IPA).

This software enables the discovery and analyses of functional relationships between differentially expressed genes and gives information about the molecular networks, biological functions, and canonical pathways over-represented among those genes.

Fourteen canonical pathways (Supplementary Table 3) were found to be significantly over-represented among the differentially expressed genes, including the lipid antigen presentation by CD1 (4/22 genes, $P=4.34E-04$) and antigen presentation (4/40 genes, $P=9.12E-03$) pathways which have functions related to cell-mediated immunity. IPA analysis also highlighted several pathways related to cell signaling such as the RhoA signaling (9/22 genes, $P=1.72E-03$), Cdc42 signaling (9/143 genes, $P=3.51E-03$), JAK2 in hormone-like cytokine signaling (4/34 genes, $P=8.17E-03$), caveolar-mediated endocytosis signaling (6/81 genes, $P=8.88E-03$), and neuregulin signaling (6/95 genes, $P=2.04E-02$). Interestingly, three genes of the neuregulin signaling pathway—EREG (-2.39 fold-change, $P=2.99E-02$), AREG (-1.80-fold-change, $P=4.90E-02$), and NRG1 (-1.76-fold-change, $P=3.12E-02$)—were among the most differentially expressed genes between cases and controls, and therefore these three genes were selected for follow up in an association study.

Association study

To evaluate the association between genes of the neuregulin signaling pathway (EREG-AREG and NRG1) and susceptibility to BD, we performed a case–control association study. The general characteristics of the association study dataset are shown in Table 3. The distribution of clinical symptoms in the Iranian BD patients (e.g., 98.7 % with oral aphthosis, 62.3 % with genital aphthosis, 55.2 % with skin lesions, 59.7 % with ocular lesions) is in line with what has been observed for larger datasets [16], suggesting that this is a representative group. Cases and controls were matched for age-at-examination (mean AAE±SD of 39.1±11.0 years in BD and 40.0±12.3 years in controls, $P=1.93E-01$).

Since no SNPs in the EREG-AREG locus showed evidence of association in the two published BD GWAS (personal communications by Dr. Elaine Remmers and Dr. Nobuhisa Mizuki), we followed a haplotype tagging SNP approach in order to explore the association of this locus with BD. In an extended region of chromosome 8 (chr. 8: 31517663–32713793 bp), including NRG1, 17 and 22 SNPs have been associated with BD in the Japanese and Turkish populations ($P<0.05$), respectively (personal communications by Dr. Elaine Remmers and Dr. Nobuhisa Mizuki). Since NRG1 is a very large gene spanning over 1.1 Mb, we focused on 15 independent SNPs located within the smaller region showing stronger evidence of association (chr. 8: 31517663–32066776 bp). In the EREG-AREG region, one

Table 1 Characterization of the sample used in the gene profiling study

ID	Affection status	Sex	AAE (years)	AAD (years)	Major clinical symptoms	IS
1	Case	F	29	21	Oral and genital aphtosis, pseudofolliculitis	Yes
2	Case	F	40	35	Oral and genital aphtosis, pseudofolliculitis, erythema nodosum, large vein thrombosis	Yes
3	Case	F	36	35	Oral and genital aphtosis, pseudofolliculitis, positive pathergy	No
4	Case	F	29	21	Oral and genital aphtosis	No
5	Case	F	55	32	Oral and genital aphtosis, erythema nodosum, positive pathergy	No
6	Case	F	30	29	Oral and genital aphtosis, erythema nodosum, positive pathergy	Yes
7	Case	F	44	30	Oral and genital aphtosis, pseudofolliculitis and posterior uveitis	Yes
8	Case	F	46	44	Oral and genital aphtosis, pseudofolliculitis, erythema nodosum, large vein thrombosis	No
9	Case	M	20	17	Oral and genital aphtosis, pseudofolliculitis	Yes
10	Case	M	57	36	Oral aphtosis, pseudofolliculitis, erythema nodosum, posterior uveitis	No
11	Case	M	50	30	Oral and genital aphtosis, erythema nodosum, large vein thrombosis	Yes
12	Case	M	30	30	Oral aphtosis, pseudofolliculitis, erythema nodosum, large vein thrombosis, positive pathergy	Yes
13	Case	M	33	21	Oral, genital and skin aphtosis, pseudofolliculitis, erythema nodosum, anterior and posterior uveitis	Yes
14	Case	M	29	27	Oral and genital aphtosis, pseudofolliculitis	No
15	Case	M	28	24	Oral and genital aphtosis, erythema nodosum, anterior and posterior uveitis, retinal vasculitis	Yes
		Mean±SD	37.1±11.0	28.8±7.2		
16	Control	F	27	–	–	No
17	Control	F	32	–	–	No
18	Control	F	62	–	–	No
19	Control	F	51	–	–	No
20	Control	F	26	–	–	No
21	Control	F	26	–	–	No
22	Control	F	46	–	–	No
23	Control	M	35	–	–	No
24	Control	M	42	–	–	No
25	Control	M	31	–	–	No
26	Control	M	28	–	–	No
27	Control	M	61	–	–	No
28	Control	M	26	–	–	No
29	Control	M	21	–	–	No
		Mean±SD	36.7±13.5	–		

The principal demographic and clinical characteristics of the study subjects are indicated, including the age-at-examination (AAE) for the patients and controls and the age-at-diagnosis (AAD) for the patients. Patients were considered immunosuppressed (IS) if medicated with azathioprine, cyclosporine, deflazacort, or thalidomide at the time of blood sample collection

SNP failed genotyping (rs2132065) and was not included in the analysis. All the remaining SNPs were successfully genotyped (Supplementary Table 4), and the association results are shown in Supplementary Table 5.

In the *EREG-AREG* locus, we found an allelic association at rs6845297 (P=2.51E-02, OR_A [95%CI]=0.86 [0.75–0.98]), located downstream of *EREG*. In *NRG1*, three markers demonstrated an association with BD risk in the allelic test: rs4489285 (P=2.01E-02, OR_G [95%CI]=0.85 [0.73–0.97]) located upstream *NRG1*, and rs383632 (P=2.78E-02, OR_T [95%CI]=1.22 [1.02–1.46]) and rs1462891 (P=3.92E-02, OR_T [95 % CI]=1.18 [1.01–1.37]) located in *NRG1*. rs4489285 was also associated with BD in a Japanese sample, and rs383632 and rs1462891 were associated with BD in a Turkish dataset (personal communications by Dr. Elaine Remmers and Dr. Nobuhisa Mizuki). Since male to female
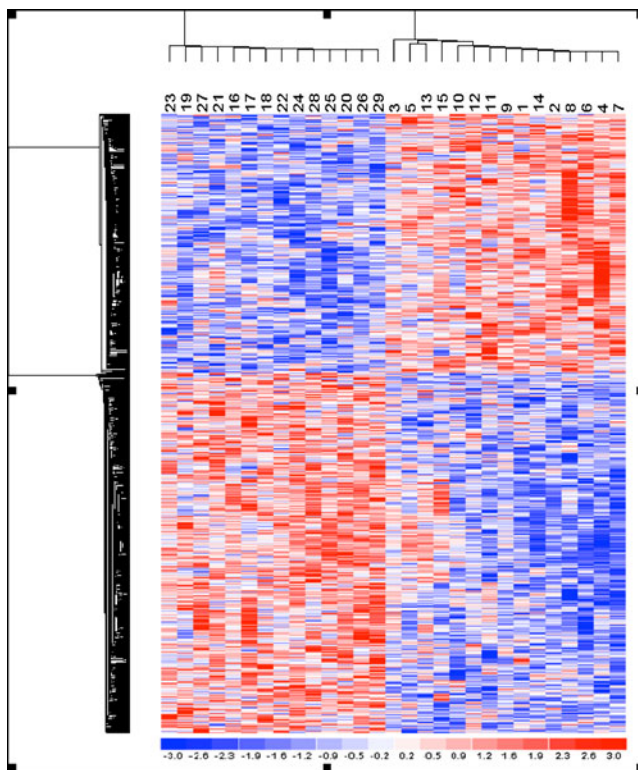
Fig. 1 Illustration of the expression pattern differences among BD cases and controls. Hierarchical clustering analysis of the 29 samples using the 621 probe sets differentially expressed between BD cases and controls, with a threshold of a 1.2-fold-change and $P \leq 0.05$. Each column represents an individual, and each row a probe set. Higher expression levels are *dark red* and lower levels are *dark blue*. Refer to Table 1 for sample identification. This figure is produced with the Partek software

ratio was significantly higher in the BD group than in the control group (52.5 % and 41.5 %, respectively, $P=3.76E-06$), we tested the association of these four SNPs adjusted for gender. They remained significantly associated ($P \leq 0.05$), albeit less so due to the concurrent loss of power associated with co-variate adjustment. None of these SNPs survive Bonferroni multiple testing correction. Haplotype associations did not provide additional insights over single marker associations (data not shown).

Meta-analysis of associated SNPs in *NRG1*

A combined analysis of the three associated SNPs at *NRG1* with data from two published GWAS [2, 3] reveals that the G allele of rs4489285 confers protection to BD ($P=2.10E-03$, OR_G [95 % CI]=0.83 [0.74–0.93]) and that the T alleles of rs383632 ($P=6.61E-04$, OR_T [95 % CI]=1.23 [1.09–1.38]) and rs1462891 ($P=7.88E-04$, OR_T[95 % CI]=1.19[1.07–1.32]) confer risk to BD (Table 4). The association consistency of these *NRG1* polymorphisms is reinforced by this meta-analysis and further supports their role in BD susceptibility.

Epistatic interaction

Given that BD has clearly a multigenic inheritance and that *EREG*, *AREG*, and *NRG1* belong to the same cellular pathway, we tested for the existence of non-additive gene–gene interactions using the multifactor-dimensionality reduction method. Supplementary Table 6 summarizes the best interaction models among all possible 2-, 3-, and 4-marker models tested. The best 4-marker interaction model was a significant predictor of Behçet's disease. The interaction between rs6845297 and rs9992496 (both located downstream of *EREG*) and rs2345991 and rs956203 (located in *NRG1*) had a statistically significant testing balanced accuracy of 0.566 (thus correctly classifying 56.6 % of the individuals tested), and a cross-validation consistency of 5/10, indicating that the model was selected five times out of ten cross-validation subsets ($P=8.00E-03$ after 1,000 permutations). The interaction dendogram depicted in Fig. 2 shows a strong non-linear (epistatic) synergistic interaction between rs2345991 and rs956203 (located in *NRG1*) with rs9992496 (located downstream of *EREG*). The dot-dash line indicates an independent effect of rs6845297 (located downstream of *EREG*) over the three other SNPs. These results suggest that BD susceptibility can be modulated by epistatic interactions between epidermal growth factor receptor genes.

Discussion

Using a genomic and genetic convergence strategy for the first time in the BD genetics field, the neuregulin signaling pathway emerged as a novel player. Several genes in this pathway (*EREG*, *AREG*, *NRG1*) were under-expressed in BD cases when compared to controls, and genetic markers in these loci were associated with BD susceptibility.

To the best of our knowledge, we performed the first genome-wide study comparing the expression profiles of BD patients and controls. PBMCs were considered a relevant tissue for expression profiling since BD is a generalized vasculitis characterized by a profound inflammatory and immune dysregulation [1]. Furthermore, given that BD is a complex disease with multiple genes and environmental factors implicated in its etiology, we did not expect to find genes with major expression differences. Instead, the cumulative effect of small changes in expression levels of a large number of genes, over time, will result in the phenotype. Supporting this notion is the finding that *HLA-DQA1*, a well-established locus for BD, demonstrated a moderate differential expression in our study (1.56-fold-change, $P=3.19E-02$). Therefore, we selected genes to follow up based on a conservative threshold of 1.20-fold-change.

Table 2 Top genes differentially expressed between BD cases and controls

Probe set ID	Gene Symbol	Gene name	P	Fold-change
204419_x_at	HBG /HBG2	Hemoglobin, gamma A/ hemoglobin, gamma G	1.35E-02	2.53
206108_s_at	SRSF6	Serine/arginine-rich splicing factor 6	3.96E-02	2.45
208450_at	LGALS2	Lectin, galactoside-binding, soluble, 2	1.11E-02	2.20
203961_at	NEBL	Nebulette	2.59E-02	2.10
234964_at	TRD@	T cell receptor delta locus	1.91E-02	1.81
222027_at	NUCKS1	Nuclear casein kinase and cyclin-dependent kinase substrate 1	3.06E-02	1.68
224009_x_at	DHRS9	Dehydrogenase/reductase (SDR family) member 9	4.92E-03	1.60
209642_at	BUB1	Budding uninhibited by benzimidazoles 1 homolog	4.54E-02	1.59
212671_s_at	HLA-DQA1/2	Major histocompatibility complex, class II, DQ alpha 1/alpha 2	3.19E-02	1.57
216733_s_at	GATM	Glycine amidinotransferase	8.85E-03	1.55
221530_s_at	BHLHE41	Basic helix-loop-helix family, member e41	4.08E-03	1.53
223751_x_at	TLR10	Toll-like receptor 10	2.58E-03	1.51
228285_at	TDRD9	Tudor domain containing 9	4.74E-02	−1.52
207314_x_at	KIR3DL1/2 /LOC727787	Killer cell immunoglobulin-like receptor, three domains, long cytoplasmic tail 1/2 /tail, 2-like	2.45E-02	−1.53
219359_at	ATHL1	ATH1, acid trehalase-like 1	2.40E-02	−1.53
204908_s_at	BCL3	B-cell CLL/lymphoma 3	1.24E-02	−1.53
216905_s_at	ST14	Suppression of tumorigenicity 14	3.77E-03	−1.55
227510_x_at	MALAT1	Metastasis associated lung adenocarcinoma transcript 1	1.12E-02	−1.57
204036_at	LPAR1	Lysophosphatidic acid receptor 1	3.45E-02	−1.57
202464_s_at	PFKFB3	6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 3	2.35E-03	−1.58
213986_s_at	C19orf6	Chromosome 19 open reading frame 6	5.46E-03	−1.62
1569599_at	SAMSN1	SAM domain, SH3 domain and nuclear localization signals 1	3.84E-02	−1.63
209264_s_at	TSPAN4	Tetraspanin 4	1.30E-02	−1.65
236244_at	HNRNPU	Heterogeneous nuclear ribonucleoprotein U	2.42E-02	−1.67
232530_at	PLD1	Phospholipase D1, phosphatidylcholine-specific	1.06E-02	−1.72
206343_s_at	NRG1	Neuregulin 1	3.12E-02	−1.76
205239_at	AREG	Amphiregulin	4.90E-02	−1.80
243296_at	NAMPT	Nicotinamide phosphoribosyltransferase	5.24E-04	−1.86
209686_at	S100B	S100 calcium binding protein B	7.92E-03	−2.25
205767_at	EREG	Epiregulin	2.99E-02	−2.39

The 30 genes with higher absolute value of fold-change difference between cases and controls are shown in this table. Genes with more than one probe differentially expressed are represented by the probe with the higher fold-change difference between cases and controls

Pathway analysis of differentially expressed genes drew our attention to the neuregulin signaling pathway and to some of its constituents, in particular to *EREG*, *AREG*, and *NRG1*, which were among the most under-expressed genes in BD patients. Association studies of these genes revealed one polymorphism in the *EREG-AREG* locus (rs6845297) and three SNPs in *NRG1* (rs4489285, rs383632, and rs1462891) associated with BD. While the association of rs6845297 with BD has not been previously tested, meta-analyses of our associations in *NRG1* with those from previous reports in Turkish and Japanese [2, 3] indicate that the susceptibility conferred by these variants is not restricted to the Iranian population and strengthen the role of *NRG1* in BD susceptibility. The effect size of these *NRG1* variants in BD susceptibility (ORs for the risk alleles of approximately 1.20) is much smaller than the effect of the well-established *HLA-B51* allele (OR of approximately 3.50) [2, 3] but is of the same magnitude as reported non-HLA allelic associations (ORs typically between 1.20 and 1.60) in genome-wide associations studies for BD [2, 3, 17, 18]. Moreover, these three genetic markers have minor allele frequencies ranging from 0.122 to 0.473 in HapMap Caucasian (CEU) and Asian (CHB and JPT) populations, suggesting that these common variants may influence risk for BD in a substantial proportion of the population.

Concerns with multiple testing arise when screening a large number of genes or genetic variants. The results of the gene expression profiling were assessed for false discovery rate using the Q value, but no correction for the number of SNPs tested was performed. However, the *NRG1* meta-analyses and the observed epistatic

Table 3 Principal demographic and clinical characteristics of the Iranian case–control sample used in the association study

Characteristic	Controls	BD cases
N	839	976
Gender (*n/N*, % males)	348/839 (41.5)	512/976 (52.5)
Mean age-at-examination (years)±SD	40.0±12.3	39.1±11.0
Mean age-at-diagnosis (years)±SD	–	32.2±9.2
Oral aphthosis (*n/N*, %)	0/839 (0)	963/976 (98.7)
Genital aphthosis (*n/N*, %)	0/839 (0)	608/976 (62.3)
Skin lesions (*n/N*, %)	–	539/976 (55.2)
Pseudofolliculitis (*n/N*, %)	–	409/539 (75.9)
Erythema nodosum (*n/N*, %)	–	211/539 (39.1)
Skin aphthosis (*n/N*, %)	–	30/539 (5.6)
Ophthalmologic manifestations (*n/N*, %)	–	583/976 (59.7)
Anterior uveitis (*n/N*, %)	–	420/583 (72.0)
Posterior uveitis (*n/N*, %)	–	495/583 (84.9)
Retinal vasculitis (*n/N*, %)	–	343/583 (58.8)
Joint manifestations (*n/N*, %)	–	301/976 (30.8)
Arthralgia (*n/N*, %)	–	139/301 (46.2)
Arthritis (*n/N*, %)	–	188/301 (62.5)
Ankylosing spondylitis (*n/N*, %)	–	18/301 (6.0)
Neurological manifestations (*n/N*, %)	–	62/976 (6.4)
Vascular involvement (*n/N*, %)	–	51/976 (5.2)
Gastro-intestinal manifestations (*n/N*, %)	–	39/976 (4.0)
Epididymitis (*n/N*, %)	–	22/512 (4.3)
Cardiac involvement (*n/N*, %)	–	6/976 (0.6)
Pleuro-pulmonary involvement (*n/N*, %)	–	6/976 (0.6)
Pathergy phenomenon (*n/N*, %)	–	441/960 (45.9)
Family history of BD (*n/N*, %)	–	86/936 (9.2)

interactions between genetic markers in *NRG1* and *EREG-AREG* reinforce the role of this pathway in BD pathogenesis and warrant further validation in independent population samples.

To date, these four genetic markers have not been associated with other diseases (NHGRI catalogue of GWAS), and no functions have been linked to these variants. However, *NRG1* is encoded by 21 alternatively spliced exons spanning more than 1.1 Mb which, through alternative promoter usage and splicing, produce a variety of isoforms with different affinities to their receptors [19]. Interestingly, rs4489285 is in complete LD with rs1476540 (r^2=1, CEU population) that is located in a conserved transcription factor binding site for GATA-1, GATA-2, and GATA-3 in humans, and rs383632 is in complete LD with rs967205 (r^2=1, CEU population) which localizes to a conserved transcription factor binding site for FOXO3a, FOXO3b, and FOXD1 in humans (http://snp-nexus.org). The associated SNPs may therefore work as proxies for variants in intronic enhancers.

Epiregulin, the most under-expressed gene in BD patients, plays an essential role in immune/inflammatory-related responses in keratinocytes and macrophages in the epidermal layer [20] and has a pivotal role in peptidoglycan-induced proinflammatory cytokine production by antigen presenting cells [21]. Indeed, deficiency of epiregulin in mice results in chronic dermatitis and is correlated with an enhanced expression of the pro-inflammatory cytokine IL-18 by keratinocytes, supporting the role of this molecule in inflammatory diseases. An in silico analysis of the functional relevance of rs6845297 (located downstream of *EREG*) using SNPnexus (http://snp-nexus.org/) did not reveal previous associations of this polymorphism with other phenotypes or potential effects on regulatory elements or conserved sequences.

Amphiregulin binds exclusively to ErbB1 [22], and its overexpression has been associated with several autoimmune disorders such as systemic lupus erythematosus, psoriasis, and rheumatoid arthritis, and to synovial membrane inflammation [23–25]. Studies in transgenic mice demonstrate that amphiregulin overexpression in both the epidermis basal and suprabasal layer leads to a severe psoriasis-like phenotype and skin inflammation with a rich dermal and epidermal infiltration of neutrophils and lymphocytes [26]. Epidermal *AREG* expression is a possible mediator of innate cutaneous immunity and epidermal proliferation and a potential trigger of both cutaneous psoriasis and psoriatic arthritis [26].

Table 4 Association results of three SNPs in *NRG1* in the Iranian dataset, in the discovery GWAS samples, and in the overall meta-analyses

SNP	Gene	Allele	Dataset	N		Frequency		P	OR [95 % CI]
				Cases	Controls	Cases	Controls		
rs4489285	*NRG1*	G	Iranian	976	839	0.302	0.339	**2.01E-02**	0.85 [0.73–0.97]
			Japanese [Ref. 3]	611	737	0.123	0.151	**4.32E-02**	0.79 [0.63–0.99]
			Overall					**2.10E-03**	0.83 [0.74–0.93]
rs383632	*NRG1*	T	Iranian	976	839	0.180	0.153	**2.78E-02**	1.22 [1.02–1.46]
			Turkish [Ref. 2]	1211	1270	0.166	0.140	**9.16E-03**	1.23 [1.05–1.43]
			Overall					**6.61E-04**	1.23 [1.09–1.38]
rs1462891	*NRG1*	T	Iranian	976	839	0.261	0.231	**3.92E-02**	1.18 [1.01–1.37]
			Turkish [Ref. 2]	1215	1278	0.236	0.205	**7.86E-03**	1.20 [1.05–1.37]
			Overall					**7.88E-04**	1.19 [1.07–1.32]

At each marker, the OR and CI refer to the allele on the forward strand of the human genome reference sequence. Significant *P* values are bolded

Because of its multiple isoforms and various functional roles, *NRG1* has been implicated in the pathophysiology of numerous disorders, including breast cancer, schizophrenia, and bipolar disorder in human studies, atherosclerosis, myocardial dysfunction, and multiple sclerosis in animal models [27]. Interestingly, a schizophrenia-associated missense mutation in *NRG1* has been associated with immune system deregulation [27], establishing a link between this gene and immunological disorders. NRG1 can bind either to ErbB3 (that has no active kinase domain) and/or ErbB4. *ErbB4*, as well as two other genes of further downstream in the pathway (*DCN*, *SOS*) (Supplementary Table 1), were differentially expressed but not among the top genes, and thus were not selected for follow-up in the association study. On the other hand, *ErbB1* was not differentially expressed, suggesting that the intracellular signaling pathway affected in BD may be via ErbB4 and not ErbB1. To further establish the role of this pathway in BD pathogenesis, comprehensive immunogenetic/cellular studies may be conducted.

We cannot formally exclude the possibility that the relatively small number of samples and the different immunossupressants (IS) taken by some patients had an impact on our findings in the expression study, and may prevented the detection of some important BD genes. Nevertheless, to minimize the effect of this potential confounding factor, we included the immunosuppression status as a co-variate in the analysis. The rationale behind including immunosuppressed patients in the expression profiling was that they are expected to be in a physiological state that more closely resembles the inactive phase of BD patients not taking immunosuppressants drugs. Our group of cases is therefore thought to be more homogeneous than a group of patients with varying degrees of disease activity. Furthermore, current treatments for BD target mostly environmental triggers (e.g., antibiotics to treat bacterial infections) or self-amplifying disease mechanisms (e.g., immunosuppressants to block auto-immunity, NSAIDs and corticosteroids to reduce inflammation), but are not specific for this disorder and are unlikely to affect the unique initial etiopathogenic mechanism(s) that we are searching for. Also, unlike canonical pathways related to immunity and inflammation, the neuroregulin signaling pathway is not as likely to be influenced by immunomodulatory drugs.

In summary, we have shown the power of combining genome-wide expression studies with in silico pathways

Fig. 2 Interaction dendrogram for the *EREG-AREG* and *NRG1* polymorphisms in BD susceptibility. The length of the dendrogram branch that connects two polymorphisms indicates the strength of interaction (the shorter the branch, the stronger is the interaction)
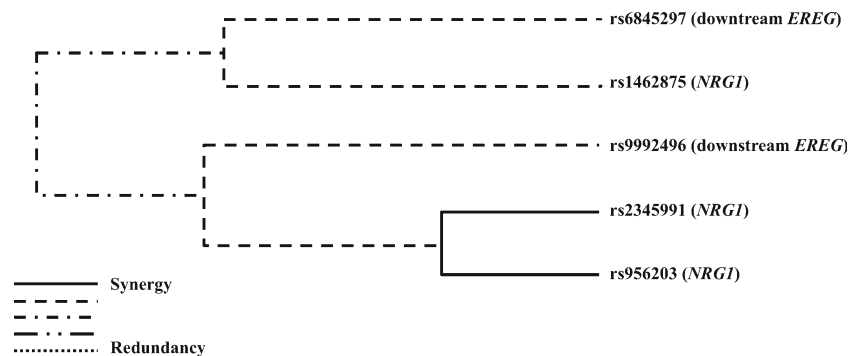
analysis and association studies in the identification of novel genetic risk factors for BD. We have also yielded a series of candidate genes, blood biomarkers, pathways, and mechanisms that are prime targets for follow-up in hypothesis driven studies. Future research towards understanding the role of these genes and pathways in BD pathogenesis include candidate gene association studies, deep re-sequencing of selected genes, biochemical, cellular, and animal work.

Disclosure of potential conflict of interest There is no conflict of interest to disclose.

References

1. Kapsimali VD, Kanakis MA, Vaiopoulos GA, Kaklamanis PG (2010) Etiopathogenesis of Behçet's disease with emphasis on the role of immunological aberrations. Clin Rheumatol 29:1211–1216

2. Remmers EF, Cosan F, Kirino Y, Ombrello MJ, Abaci N, Satorius C, Le JM, Yang B, Korman BD, Cakiris A et al (2010) Genome-wide association study identifies variants in the MHC class I, *IL10*, and *IL23R-IL12RB2* regions associated with Behçet's disease. Nat Genet 42:698–702

3. Mizuki N, Meguro A, Ota M, Ohno S, Shiota T, Kawagoe T, Ito N, Kera J, Okada E, Yatsu K et al (2010) Genome-wide association studies identify *IL23R-IL12RB2* and *IL10* as Behçet's disease susceptibility loci. Nat Genet 42:703–706

4. Olsen N, Sokka T, Seehorn CL, Kraft B, Maas K, Moore J, Aune TM (2004) A gene expression signature for recent onset rheumatoid arthritis in peripheral blood mononuclear cells. Ann Rheum Dis 63:1387–1392

5. Achiron A, Gurevich M, Friedman N, Kaminski N, Mandel M (2004) Blood transcriptional signatures of multiple sclerosis: unique gene expression of disease activity. Ann Neurol 55:410–417

6. Mandel M, Gurevich M, Pauzner R, Kaminski N, Achiron A (2004) Autoimmunity gene expression portrait: specific signature that intersects or differentiates between multiple sclerosis and systemic lupus erythematosus. Clin Exp Immunol 138:164–170

7. International Team for the Revision of the International Criteria for Behçet's Disease (ITR-ICBD) (2006) Revision of the International Criteria for Behçet's Disease (ICBD). Clin Exp Rheumatol 24(Suppl 42):S14–S15

8. Quian HR, Huang S (2005) Comparison of false discovery rate methods in identifying genes with differential expression. Genomics 86:495–503

9. Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C, Aach J, Ansorge W, Ball CA, Causton HC et al (2001) Minimum information about a microarray experiment (MIAME): toward standards for microarray data. Nat Genet 29:365–371

10. Xavier JM, Shahram F, Davatchi F, Rosa A, Crespo J, Abdollahi BS, Nadji A, Jesus G, Barcelos F, Vaz Patto J et al (2012) Association study of *IL10* and *IL23R-IL12RB2* in Iranian Behçet's disease patients. Arthritis Rheum 64:2761–2772

11. Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21:263–265

12. González JR, Armengol L, Solé X, Guinó E, Mercader JM, Estivill X, Moreno V (2007) SNPassoc: an R package to perform whole genome association studies. Bioinformatics 23:644–645

13. Ritchie MD, Hahn LW, Roodi N, Bailey LR, Dupont WD, Parl FF, Moore JH (2001) Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. Am J Hum Genet 69:138–147

14. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ et al (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81:559–575

15. Moore JH (2004) Computational analysis of gene–gene interactions using multifactor dimensionality reduction. Expert Rev Mol Diagn 4:795–803

16. Davatchi F, Shahram F, Chams-Davatchi C, Shams H, Nadji A, Akhlaghi M, Faezi T, Ghodsi Z, Faridar A, Ashofteh F et al (2010) Behçet's disease: from east to west. Clin Rheumatol 29:823–833

17. Hou S, Yang Z, Du L, Jiang Z, Shu Q, Chen Y, Li F, Zhou Q, Ohno S, Chen R et al (2012) Identification of a susceptibility locus in *STAT4* for Behçet's disease in Han Chinese in a genome-wide association study. Arthritis Rheum 64:4104–4013

18. Lee YJ, Horie Y, Wallace GR, Choi YS, Park JA, Song R, Kang YM, Kang SW, Baek HJ, Kitaichi N et al (2013) Genome-wide association study identifies *GIMAP* as a novel susceptibility locus for Behçet's disease. Ann Rheum Dis. doi:10.1136/annrheumdis-2011-200288

19. Falls DL (2003) Neuregulins: functions, forms, and signaling strategies. Exp Cell Res 284:14–30

20. Shirasawa S, Sugiyama S, Baba I, Inokuchi J, Sekine S, Ogino K, Kawamura Y, Dohi T, Fujimoto M, Sasazuki T (2004) Dermatitis due to epiregulin deficiency and a critical role of epiregulin in immune-related responses of keratinocyte and macrophage. Proc Natl Acad Sci U S A 101:13921–13926

21. Sugiyama S, Nakabayashi K, Baba I, Sasazuki T, Shirasawa S (2005) Role of epiregulin in peptidoglycan-induced proinflammatory cytokine production by antigen presenting cells. Biochem Biophys Res Commun 337:271–274

22. Shoyab M, Plowman GD, McDonald VL, Bradley JG, Todaro GJ (1989) Structure and function of human amphiregulin: a member of the epidermal growth factor family. Science 243:1074–1076

23. Ishii T, Onda H, Tanigawa A, Ohshima S, Fujiwara H, Mima T, Katada Y, Deguchi H, Suemura M, Miyake T et al (2005) Isolation and expression profiling of genes upregulated in the peripheral blood cells of systemic lupus erythematosus patients. DNA Res 12:429–439

24. Cook PW, Pittelkow MR, Keeble WW, Graves-Deal R, Coffey RJ Jr, Shipley GD (1992) Amphiregulin messenger RNA is elevated in psoriatic epidermis and gastrointestinal carcinomas. Cancer Res 52:3224–3227

25. Davies MR, Harding CJ, Raines S, Tolley K, Parker AE, Downey-Jones M, Needham MR (2005) Nurr1 dependent regulation of pro-inflammatory mediators in immortalised synovial fibroblasts. J Inflamm 2:15

26. Cook PW, Brown JR, Cornell KA, Pittelkow MR (2004) Suprabasal expression of human amphiregulin in the epidermis of transgenic mice induces a severe, earlyonset, psoriasis-like skin pathology: expression of amphiregulin in the basal epidermis is also associated with synovitis. Exp Dermatol 13:347–356

27. Marballi K, Quinones MP, Jimenez F, Escamilla MA, Raventós H, Soto-Bernardini MC, Ahuja SS, Walss-Bass C (2010) In vivo and in vitro genetic evidence of involvement of neuregulin 1 in immune system dysregulation. J Mol Med 88:1133–1141