European Journal of Human Genetics (2015) 23, 678–687
© 2015 Macmillan Publishers Limited All rights reserved 1018-4813/15
www.nature.com/ejhg

ARTICLE

Identification of possible pathogenic pathways in Behçet's disease using genome-wide association study data from two different populations

Burcu Bakir-Gungor*[,1,2], Elaine F Remmers[3], Akira Meguro[4], Nobuhisa Mizuki[4], Daniel L Kastner[3], Ahmet Gul[5] and Osman U Sezerman[6]

Behçet's disease (BD) is a multi-system inflammatory disorder of unknown etiology. Two recent genome-wide association studies (GWASs) of BD confirmed a strong association with the MHC class I region and identified two non-HLA common genetic variations. In complex diseases, multiple factors may target different sets of genes in the same pathway and thus may cause the same disease phenotype. We therefore hypothesized that identification of disease-associated pathways is critical to elucidate mechanisms underlying BD, and those pathways may be conserved within and across populations. To identify the disease-associated pathways, we developed a novel methodology that combines nominally significant evidence of genetic association with current knowledge of biochemical pathways, protein–protein interaction networks, and functional information of selected SNPs. Using this methodology, we searched for the disease-related pathways in two BD GWASs in Turkish and Japanese case–control groups. We found that 6 of the top 10 identified pathways in both populations were overlapping, even though there were few significantly conserved SNPs/genes within and between populations. The probability of random occurrence of such an event was 2.24E −39. These shared pathways were *focal adhesion*, *MAPK signaling*, *TGF-β signaling*, *ECM–receptor interaction*, *complement and coagulation cascades*, *and proteasome* pathways. Even though each individual has a unique combination of factors involved in their disease development, the targeted pathways are expected to be mostly the same. Hence, the identification of shared pathways between the Turkish and the Japanese patients using GWAS data may help further elucidate the inflammatory mechanisms in BD pathogenesis.
European Journal of Human Genetics (2015) **23,** 678–687; doi:10.1038/ejhg.2014.158; published online 17 September 2014

INTRODUCTION

Behçet's disease (BD) is a multisystem inflammatory disorder characterized by recurrent exacerbations affecting several organs including orogenital mucosa, eyes, and skin. Although BD exists worldwide, it is more prevalent in countries along the ancient Silk Road spanning from the Far East to the Middle East and the Mediterranean basin, with the highest prevalence of 4.2/1000 in Turkey.[1] Its etiology remains poorly characterized, but a genetic tendency to uncontrolled inflammatory reactions induced by various environmental triggers is considered to play a critical role in its development. Recently, two genome-wide association studies (GWASs) in BD were conducted on Turkish[1] and Japanese[2] populations. In these studies, HLA-B51 and other variations around the *HLA-B* gene were found as the most strongly associated genetic factor to BD. In addition, two other non-HLA associations with the *IL10* and *IL23R/IL12RB2* genes were confirmed using these sets of samples. A family-based linkage study suggested that HLA-B51 accounts for <20% of the genetic risk,[3] suggesting that other genetic factors including rare variants and copy number variations are waiting to be discovered.

Although GWASs have been advocated as the most powerful approach to explore the contribution of common polymorphisms to polygenic traits in many complex diseases,[4] numerous challenges persist in the identification of weaker associations with qualitative and quantitative phenotypes at the genome-wide level.[5] Standard GWAS approaches focus on the analysis of single SNPs; however, for multifactorial diseases, no particular variant on a particular gene may have a strong effect, but the combination of multiple variants with small effects explains the overall susceptibility to the disease.[6,7] The disruption of different biological pathways is thought to determine the intrinsic biological processes of multifactorial diseases. In this regard, pathway-based approaches to GWAS may be more helpful in a search for multiple genes involved in the same biological pathway, where the common variations in each of these genes have little correlation with disease risk.[8–14]

It has been observed that genes that have aberrations associated with a given complex disease tend to be part of the same subnetwork of the overall protein–protein interaction (PPI) network.[15,16] Hence, to be able to explain the connections between genotypic and phenotypic data, perturbed network modules need to be detected.

[1]Department of Genetics and Bioinformatics, Faculty of Arts and Sciences, Bahcesehir University, Istanbul, Turkey; [2]Department of Computer Engineering, Faculty of Engineering and Natural Sciences, Abdullah Gul University, Kayseri, Turkey; [3]Inflammatory Disease Section, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA; [4]Department of Ophthalmology, Yokohama City University School of Medicine, Yokohama, Kanagawa, Japan; [5]Division of Rheumatology, Department of Internal Medicine, Istanbul University, Istanbul, Turkey; [6]Biological Sciences and Bioengineering, Faculty of Engineering and Natural Sciences, Sabanci University, Istanbul, Turkey
*Correspondence: Professor B Bakir-Gungor, Department of Computer Engineering, Faculty of Engineering and Natural Sciences, Abdullah Gul University, Melikgazi, Kayseri, Turkey. Tel: +90 352 2248800; Fax: +90 352 3388828; E-mail: burcub@gatech.edu
Received 27 December 2013; revised 8 July 2014; accepted 10 July 2014; published online 17 September 2014

Genotype data (eg, SNP, copy number variation) have been used to aid the identification of these modules.[15,16] In this respect, the use of PPI networks has demonstrated great success in enhancing the outcome of GWASs.[8–10,13,17–20]

Another important piece of information that could improve the analysis of GWAS data sets is the functional effect of a SNP.[21,22] Although DNA variations that alter protein function can have significant effects, such as *NOD2* mutations in inflammatory bowel disease[23] and *FLG* mutations in eczema,[24] other types of SNPs do not have such serious consequences in disease development mechanisms. In addition, to uncover the links between the genetics and pathogenesis of human complex diseases, the potential of conducting research on disparate populations have been discussed in terms of GWAS.[25] As summarized here, many different kinds of knowledge must be combined in order to mine GWAS results further. Yet, to the best of our knowledge, none of the existing platforms can successfully integrate functional information of typed SNPs in a GWAS with PPI networks to identify SNP-targeted pathways, and make a comparative evaluation between different populations.

In this study, we hypothesized that the few SNPs that are identified in GWASs and their associated genes may be targeting the same combinations of pathway, and these biological pathways show higher conservation across populations, making them potential markers for BD. In this respect, we analyzed the GWAS data from two different populations using the methodology that we had developed to identify disease-associated pathways by combining nominally significant evidence of genetic association with the known biochemical pathways, PPI networks, and the functional information of selected SNPs.[8,26–28] In the following sections, we will discuss our findings.

MATERIALS AND METHODS

Turkish population data set

Remmers *et al*[1] conducted a GWAS on 1215 Turkish BD cases and 1278 unaffected controls. The samples were typed using Human CNV370-Duo v1.0 and Human CNV370-Quad v3.0 chips (Illumina, San Diego, CA, USA). Initially, there were 335 887 autosomal SNPs on the Duo and 334 556 SNPs on the Quad chips. The SNP genotypes from these 2493 individuals were filtered using the following criteria, >95% call rate, >1% minor allele frequency, and Hardy–Weinberg equilibrium >0.00001, resulting in 311 459 SNPs in the final analysis.[1] The details of whole genome SNP typing, quality control, and association analysis can be found in Remmers *et al*[1] and in Supplementary Table S1.

Japanese population data set

Mizuki *et al*[2] tested 500 568 SNPs on 612 Japanese individuals with BD and 740 healthy controls. DNA samples were genotyped using the Affymetrix GeneChip Human Mapping 500K Array Set (Affymetrix, Santa Clara, CA, USA). Quality control was conducted on 500 568 SNPs, and they removed 28 702 SNPs with a call rate <95%, 14 044 SNPs with deviation from Hardy–Weinberg equilibrium in controls ($P<0.001$), and 137 384 SNPs with a MAF <5% overall. Finally, 320 438 SNPs were left for further analyses.[2] The details of whole genome SNP typing, quality control, and association analysis can be found in Mizuki *et al*[2] and in Supplementary Table S1.

For both BD sets, SNP data and the genotypic *P*-values of association for each tested SNP, which were calculated via allelic χ^2 test, were obtained using the original data by EFR and AM.

Clinical features of the patients

In the Turkish GWAS set, BD patients fulfilled the International Study Group (ISG) criteria for classification; the Japanese GWAS set comprised those fulfilling the Japanese diagnostic criteria that overlap to a great extent with the ISG criteria.[1,2] Turkish patients were recruited through a multi-disciplinary BD clinic, and Japanese patients were mainly recruited from ophthalmology clinics. Therefore, some differences were noted between the frequencies of the clinical manifestations in both groups, such as a higher frequency of uveitis among Japanese patients (83.3% *vs* 35.4%) and a more frequent vascular involvement in Turkish patients (25.2% *vs* 4.4%).[1,2]

PPI data set

A human PPI data set was obtained from Protein Interaction Network Analysis platform (PINA) that integrated PPI data from six manually curated databases (MINT, IntAct, DIP, BioGRID, HPRD, and MIPS/MPact).[29] The integrated set of PPIs contains 68 420 interactions between 11 411 genes.

Network and pathway oriented GWAS analysis

We had previously developed a methodology (Pathway and Network Oriented GWAS Analysis (PANOGA) to devise disease-related KEGG pathways through the identification of SNP targeted genes within these pathways.[8,19,20,26–28] PANOGA is publicly available at http://panoga.sabanciuniv.edu. In this study, we further improved PANOGA to identify SNP-targeted pathways with the genes responsible for BD susceptibility. A diagram of our multistep strategy is given in Figure 1.

PANOGA starts with considering the functional effects of SNPs, predicted at the splicing, transcriptional, translational, and post-translational levels. Next, following the scoring scheme of Saccone *et al*,[30] PANOGA combines SNP functional scores (FS) with GWAS *P*-values (*P*) and calculates a weighted *P*-value ($Pw = P/10^{FS}$) for each SNP.

To be able to move into pathways, PANOGA maps the SNPs to genes. However, while assigning SNPs to genes, SPOT program[30] checks 2000 bp upstream region of genes following dbSNP strategy. In this approach, if a SNP is located further than 2000 bp away, it cannot be assigned to any gene, and the information from this SNP is lost. Especially the HLA region, which is known to be important for BD, is rich in such SNPs. To be able to incorporate such SNPs to our system, in this study, we improved PANOGA such that if no gene is assigned to a SNP by SPOT program, we assign the closest gene to this SNP. Next, using these SNPs to gene mapping results, we transfer SNP weighted *P*-values to genes.

An active subnetwork is the connected subgraph of the interactome that has high total significance of genotypic *P*-values of the disease-predisposing SNPs with respect to the controls.[31] PANOGA searches out active subnetworks by using the above-mentioned *Pw* values of the genes. Differently from the original version of PANOGA, in this study, we adapted PANOGA to use PPI network by PINA that combines information from six manually curated databases. In addition, differently from the original version of PANOGA, in this study we used Entrez GeneIDs to prevent issues with gene aliases (nomenclature problems).

Following the identification of subnetworks, PANOGA evaluates whether these subnetworks are biologically meaningful via comparing the constituent genes with known KEGG pathways. The details of these steps can be found elsewhere.[8,19,20,26,27]

RESULTS

To identify the biological pathways with the genes responsible for BD susceptibility, two data sets of GWASs that are conducted on BD case–control groups of Turkish (TR) and Japanese (JP) populations were studied. Following the study of Baranzini *et al*,[9] the SNPs showing nominal evidence of association ($P<0.05$) in GWASs were investigated. Then, PANOGA was applied separately on each data set, as shown in Figure 1. A total of 18 479 SNPs and 20 594 SNPs identified in GWASs were mapped to 3869 and 4076 genes for the TR and JP populations, respectively. Next, the network-oriented steps of our methodology were conducted. The genes in a subnetwork are likely to be functionally associated and might be underlying a potential genetic interaction that underlies the manifestation of the disease. To identify possible pathogenic pathways of BD, for each identified subnetwork, we computed the proportion of the genes in an identified subnetwork that were also found in a specific human biochemical pathway, compared with the overall proportion of genes

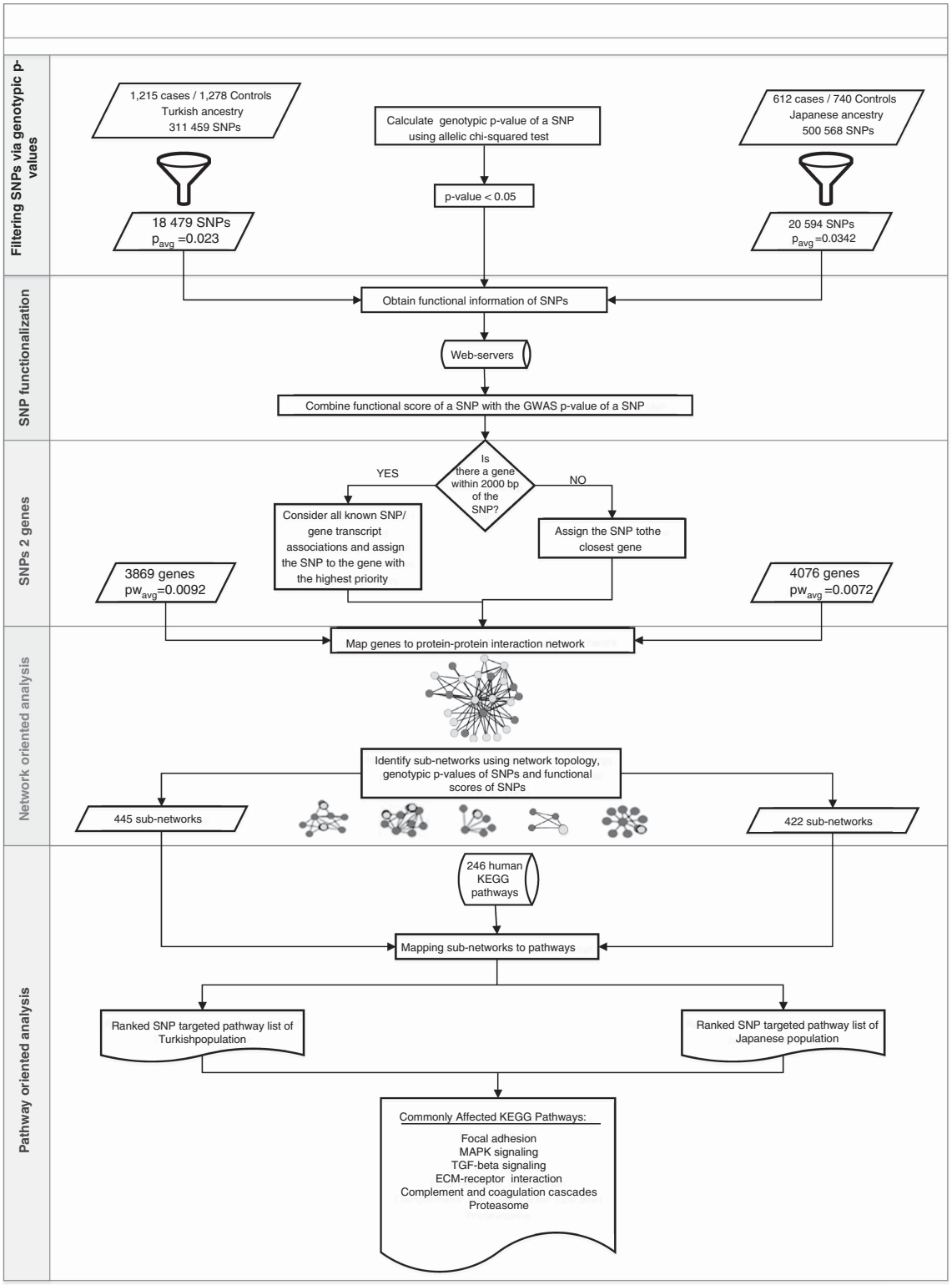

Figure 1 Overview of methodological pipeline.

described for that pathway. At this step, our subnetworks were tested against 246 available human KEGG pathways. If a KEGG pathway is found to be statistically significant for at least one of the identified subnetworks, PANOGA adds this pathway into our final list of significant KEGG pathways as associated with disease. Here, PANOGA calculates the significance of a pathway in relation to the disease as the minimum *P*-value given to that pathway among all subnetworks that are found to be associated with that KEGG pathway. The pathways are ranked according to the significance scores and are named as SNP-targeted pathways.

Table 1 The top 10 KEGG pathways identified in Turkish (TR) and Japanese (JP) Behçet's disease GWAS sets

	P-values		*Rank*		*No. of associated SNPs in GWAS*		*No. of SNP-targeted genes*		
KEGG term	*TR*	*JP*	*TR*	*JP*	*TR*	*JP*	*TR*	*JP*	*No. of commonly associated SNP-targeted genes*
Focal adhesion	9.92E−27	9.47E−23	1	2	102	131	29	24	8
MAPK signaling pathway	2.05E−23	2.14E−17	2	6	72	121	20	27	4
Jak/STAT signaling pathway	3.68E−21	6.36E−14	3	17	49	68	20	14	6
TGF-β signaling pathway	4.05E−21	1.87E−21	4	3	43	71	15	16	12
ECM–receptor interaction	1.43E−20	1.26E−18	5	5	56	49	18	15	9
Axon guidance	7.68E−19	5.02E−7	6	74	49	99	11	15	2
Complement and coagulation cascades	1.00E−18	2.35E−16	7	10	22	29	10	8	3
Antigen processing and presentation	1.79E−18	1.37E−9	8	43	161	53	14	10	7
Proteasome	1.97E−18	1.34E−24	9	1	17	9	4	6	1
Autoimmune thyroid disease	5.75E−18	7.15E−7	10	76	162	44	15	8	6
Neurotrophin signaling pathway	1.63E−11	2.66E−19	31	4	38	81	10	16	3
Wnt signaling pathway	2.27E−13	2.68E−17	22	7	45	93	15	17	7
Pathways in cancer	9.61E−11	2.68E−18	34	8	125	118	21	31	5
Neuroactive ligand–receptor interaction	1.26E−17	2.68E−19	11	9	109	151	25	26	11

Six out of the top 10 pathways shared in both populations are shown in bold.

In all, 102 pathways were detected in TR population (as shown in Supplementary Table S2) and 96 pathways were detected in JP population (as shown in Supplementary Table S3) with corrected *P*-values of $<$E−4. Of these pathways, 85 were commonly found in both populations (as shown in Supplementary Table S4). We found that the correlation between the two studies was significant (Pearson's $r^2 = 0.72$, $P < 5 \times 10 - 15$). A pairwise correlation of pathway statistics between two studies (which were carried out on independent populations with different ethnicities) should indicate common genetic variation associated with BD. As shown in Table 1, six pathways were found to be shared in the top 10 of both the TR and JP populations. The probability of random occurrence of such an event was $2.24E^{-31}$. These pathways are: (1) focal adhesion, (2) mitogen-activated protein kinase (MAPK) signaling, (3) transforming growth factor-β (TGF-β) signaling, (4) extracellular matrix (ECM)–receptor interaction, (5) complement and coagulation cascades, and (6) proteasome pathways. As shown in Figure 2a, among the 265 SNPs targeting these 6 pathways, only 2 SNPs were found in both TR and JP populations. In addition, as shown in Figure 2b and Table 2, in these 6 pathways, 37 genes out of 137 identified genes were commonly targeted by disease-predisposing SNPs in both populations. However, in the different populations, different SNP-targeted genes were found to be affected in these commonly found pathways. In another words, the same pathways can be targeted via independent genes in different populations. Even though there are few common disease-predisposing SNPs and even commonly targeted genes between these two populations, the identification of 6 common pathways in the top 10 pathways shows the relevance of our approach.

As a control for our interpretation of these pathways, we checked the nonrandomness of our results. We started with a list of 22 836 protein-coding genes available in Ensembl BioMart. We randomly picked a number between 1 and 22 836. We obtained the corresponding gene from the gene list. We repeated this procedure 3869 times for the Turkish population and 4076 times for the Japanese population. Hence, we randomly picked 3869 and 4076 genes (the number of SNP-targeted genes in the TR and JP GWAS data sets respectively). We created 10 such random sets of genes for Turkish population and 10 such random sets of genes for Japanese

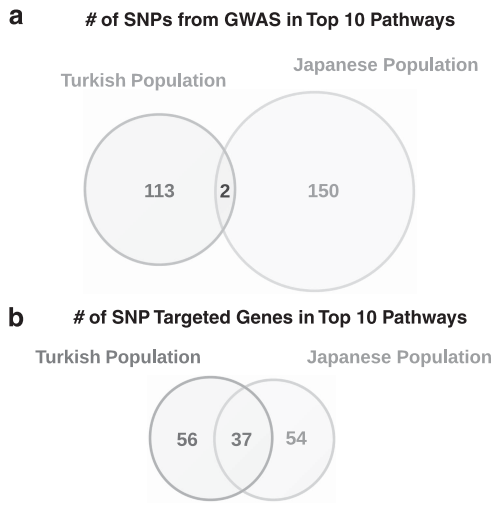

Figure 2 (**a**) SNP counts within the commonly identified top 10 pathways are shown as unique to TR population, unique to JP population, and common between these two populations. (**b**) SNP-targeted gene counts within the commonly identified top 10 pathways are shown as unique to Turkish (TR) population, unique to Japanese (JP) population, and common between these two populations.

population. We followed the same PANOGA procedure for these random gene sets. In other words, we identified active subnetworks and then searched for enriched pathways within these subnetworks. The list of pathways identified for random gene set was quite different from the pathways identified for the BD data set. There was no correlation (Pearson's $r^2 = 0.2757$, $P = 0.0200$) between the rankings of the identified pathways between the TR BD study and random gene set. In addition, there was no correlation (Pearson's $r^2 = -0.0312$, $P = 0.7894$) between the rankings of the identified pathways between the JP BD study and random gene set. For random gene sets, the complete list of the identified pathways is shown in Supplementary Tables S5 and S6. On the other hand, when we repeat this experiment independently with two real data sets, we found that

Table 2 Commonly identified SNP-targeted genes within the top 10 overrepresented KEGG pathways of each population

KEGG term	No. of common SNPs in commonly associated genes	Commonly associated, SNP-targeted genes and their SNP counts: (common) (TR GWAS) (JP GWAS)
Focal adhesion	0	*ITGA9* (0) (3) (9); *PTK2* (0) (4) (2); *FLT1* (0) (3) (2); *DOCK1* (0) (4) (4); *RASGRF1* (0) (1) (1); *FLNB* (0) (3) (6); *PTEN* (0) (2) (6); *PIK3R1* (0) (9) (14);
MAPK signaling pathway	0	*MAP3K5* (0) (3) (2); *FASLG* (0) (3) (3); *RAPGEF2* (0) (5) (13); *MYC* (0) (2) (2);
Jak/STAT signaling pathway	0	*IL2RB* (0) (2) (1); *IL7* (0) (1) (8); *CSF3R* (0) (2) (3); *CSF2RB* (0) (1) (2); *JAK2* (0) (3) (14); *PIK3R1* (0) (9) (14);
TGF-β signaling pathway	0	*INHBB* (0) (1) (1); ACVR2A (0) (5) (5); *INHBA* (0) (3) (6); *BMP2* (0) (7) (9); *TGFBR2* (0) (7) (4); *SMAD3* (0) (1) (13); *BMP7* (0) (1) (3); *BMPR1B* (0) (1) (5); *BMP6* (0) (6) (1); *ACVR1* (0) (2) (1); *ACVR1C* (0) (2) (2); *TGFB2* (0) (1) (3);
ECM–receptor interaction	2	*LAMA1* (0) (3) (1); *COL4A2* (1) (3) (12); *CD44* (0) (2) (1); *LAMA5* (0) (2) (1); TNC (0) (1) (2); *COL1A2* (0) (4) (1); *ITGA1* (0) (6) (3); *LAMB1* (1) (1) (1); *COL5A1* (0) (4) (1);
Axon guidance	0	*PTK2* (0) (4) (2); *EFNA5* (0) (6) (8);
Complement and coagulation cascades	0	*PLAT* (0) (1) (1); *F5* (0) (3) (2); *F13A1* (0) (5) (8);
Antigen processing and presentation	3	*TAP1* (0) (2) (2); *HLA-A* (0) (3) (4); *HLA-C* (2) (44) (14); *HLA-B* (1) (29) (7); *HLA-E* (0) (22) (2); *HLA-G* (0) (2) (10); *HLA-DRA* (0) (20) (4);
Proteasome	0	*PSMA6* (0) (1) (1);
Autoimmune thyroid disease	3	*HLA-DQB1* (0) (4) (1); *HLA-A* (0) (3) (4); *HLA-C* (2) (44) (14); *HLA-B* (1) (29) (7); *HLA-E* (0) (22) (2); *HLA-G* (0) (2) (10);
Neurotrophin signaling pathway	1	*NTRK3* (1) (3) (10); *MAP3K5* (0) (3) (2); *KIDINS220* (0) (2) (4);
Wnt signaling pathway	1	*FZD8* (1) (6) (10); *CSNK2B* (0) (8) (3); *SMAD3* (0) (1) (13); *PLCB1* (0) (5) (17); *MYC* (0) (2) (2); *TCF7L2* (0) (3) (3); *CTNNB1* (0) (1) (9);
Pathways in cancer	0	*PRKCA* (0) (4) (6); *TGFBR2* (0) (7) (4); FASLG (0) (3) (3); *MYC* (0) (2) (2); *CTNNB1* (0) (1) (9);
Neuroactive ligand–receptor interaction	0	*HTR1B* (0) (5) (23); *S1PR1* (0) (3) (3); *HTR1A* (0) (3) (3); *CHRM3* (0) (2) (8); *GRIN2B* (0) (15) (10); *GRIA1* (0) (13) (21); *DRD2* (0) (1) (6); *GRM7* (0) (13) (22); *HRH4* (0) (4) (5); *LPAR1* (0) (13) (8); *GRM1* (0) (4) (1)

The number of commonly typed SNPs in both populations, only in Turkish (TR) population, and only in Japanese (JP) population are shown in parentheses. Six out of the top 10 pathways shared by both TR and JP populations are shown in bold.

6 of the top 10 identified pathways in TR and JP populations were overlapping. The dysfunction of these pathways cause common functional problems and, hence, may cause BD.

We also searched for known BD-related pathways in KEGG Disease Database using 'Behçet' as a keyword. This search returned one disease term (H00106) and one pathway (complement and coagulation cascades) as associated with this disease term. This pathway is identified in 7th and 10th rankings with $P = 1.00\text{E}-18$ and $P = 2.35\text{E}-16$ in TR and JP populations, respectively.

DISCUSSION

Analysis of the Turkish and Japanese GWAS data using the PANOGA methodology identified 6 overlapping pathways out of the top 10 identified pathways in the individual data sets as possible mechanisms involved in the pathogenesis of BD. The shared pathways included those involved in antigen processing (proteasome), inflammation (focal adhesion, MAPK signaling pathway, TGF-β signaling pathway), coagulation (complement and coagulation cascade), and interactions with extracellular environment (ECM–receptor interaction, focal adhesion).

Genetic tendencies play a critical role in complex diseases including BD, and the GWAS approach has enabled us to define the roles of common functional genetic variants in the disease pathogenesis. Analysis of a large series of patients with BD with GWAS revealed the contributions of *IL10* and *IL23R/IL12RB2* polymorphisms in addition to already known HLA-B51 and other HLA Class I associations, and further analysis of the same data set by imputation provided evidences for *ERAP1*, *CCR1*, and *STAT4* gene associations.[32,33] However, these findings can only explain part of the overall genetic tendency to BD, and different approaches and analysis methods and/or the collection of a much larger series of patients are necessary for the identification of remaining genetic susceptibility factors contributing to the disease pathogenesis.[34] BD is a rare disease in most populations, and collection of larger sample sets of well-defined patients has become a major challenge for genetic studies. Therefore, alternative analysis methods for exploiting available data obtained from different ethnic groups are expected to provide some help for identification of novel genetic associations or interaction pathways involved in its pathogenesis.

The pathways found with the PANOGA method support the available GWAS findings. Identification of the proteasome pathway is very complementary for the known HLA class I and *ERAP1* associations and supports the importance of antigen processing and peptide loading to *HLA-B* antigens in the disease pathogenesis. The identification of inflammatory pathways (focal adhesion, MAPK signaling pathway, TGF-β signaling pathway) extends the spectrum of *IL10* and *IL23R/IL12RB2* associations to other pro-inflammatory and regulatory cytokines that may be critical in deregulated immune response observed in BD. Recently, to elucidate the mechanisms of innate immune responses, Ture-Ozdemir *et al*[35] studied inflammasome activation in dendritic cells and neutrophils following stimulation with two different pattern recognition receptors (RIG-1-like and NOD-like) in BD patients. In this study, the dendritic cells from patients with BD were found to be activated as a response to *NOD2* stimulus that was shown by the slightly defective ($P < 0.05$) expression levels of *RIP2* and p38 as well as *IL18*.[35] Activation of *RIP2* results in the activation of nuclear factor-κB (NF-κB) and the MAPKs. This cascade is known to trigger the

production of pro-inflammatory cytokines, including *IL1B* and *IL18*.[35] As shown in Table 1, the MAPK signaling pathway is identified in second and sixth rankings with $P = 2.05E^{-23}$ and $P = 2.14E-17$ in TR and JP populations, respectively.

The TGF-β/Smad signaling pathway is also shown as overactive in BD patients.[36] This pathway is ranked as fourth ($P = 4.05E-21$) and third ($P = 1.87E-21$) in the TR and JP data sets, as shown in Table 1. Shimizu *et al*[36] observed that after stimulation, the expressions of *TGFBR1*, *IL12* receptor β2, and *SOCS1* on peripheral blood mononuclear cell were significantly enhanced in BD patients as compared with the normal controls ($P<0.05$). They also reported that CD4+ T cells infiltrating into BD skin lesions expressed *TGFB1* much more than those infiltrating into non-BD erythema nodosum.[36] Similarly, in our study, *TGFB1* and *SOCS1* genes are found to be targeted by a SNP in TR population (shown in Table 3). Th17 cells (a novel subset of T cells) are suspected to play a fundamental role in pathogenesis of BD.[37,38] Shahneh *et al*[39] stated that Th17 cells predominantly produce *IL17A-F*, *IL21*, *IL22*, and TNF-α. In our study, the *tumor necrosis factor-*α *receptor* (*TNFRSF1A*) gene is found to be targeted by two SNPs in TR population (shown in Table 3). As this gene has a role in innate immunity, nonsynonymous variants identified by deep exonic resequencing of *TNFRSF1A* gene has been recently evaluated for BD association.[40] The association is shown to be significant[40] in the C-α test, the adaptive sum test, and the step-up test. Shahneh *et al*[39] also emphasized possible roles of *IL6* and TGF-β during the differentiation of Th17 cells from naive T cells.[39] In another study by Shimizu *et al*,[41] it is shown that *IL6*, *IL21*, and TGF-β play a role in the differentiation of Th17 cells that proliferated in the presence of IL23. In our study, the *IL21* gene is found to be targeted by two SNPs in JP population (shown in Table 3). Our results indicate that the propagation of a signal from TGF-β to MAPK signaling pathway has an important role in BD development and progression mechanisms. In both of these pathways, there are not so many commonly targeted genes between the two populations (shown in red in Figure 3), but there are several genes affected in individual populations (shown in yellow and blue in Figure 3). This finding supports our hypothesis that although there might be individual disease development mechanisms among populations, the affected pathways show a higher level of conservation. In the literature, there is also evidence that TGF-β activates p38 MAPK that regulates Th17 cell differentiation.[42,43] These improvements in the knowledge of BD pathogenesis pave the way for innovative therapy. While presenting the new approaches in immunotherapy of BD, Shahneh *et al*[39] suggested that if subsequent studies focus on these, it might eventually lead to the development of better treatment modalities for BD patients.[39]

The detection of the complement and coagulation cascade as one of the shared pathways provides further insights for the genetic background of vasculopathy and thrombophilia associated with BD. As shown in Table 1, this pathway is identified in 7th and 10th rankings

Table 3 The unique SNP-targeted genes of each population within the 10 overrepresented KEGG pathways for BD

KEGG term	*TR-specific SNP-targeted genes in pathway*	*JP-specific SNP-targeted genes in pathway*
Focal adhesion	*ACTB, CAV3, PRKCA, COL4A2, GRB2, TLN2, TNC, ITGA1, ITGA10, HGF, ITGA4, LAMA1, PAK4, TNR, COL1A2, COL6A2, RAP1A, PDGFC, PAK1, LAMB1, FN1*	*EGFR, VAV3, ROCK1, ITGB5, ACTN2, CAPN2, ITGB1, CTNNB1, PAK7, PPP1CA, FYN, GSK3B, PDGFRA, MAPK8, PDGFD, PARVA*
MAPK signaling pathway	*TRAF2, GRB2, MAP2K4, DUSP10, NR4A1, FGF12, FLNB, MAP3K7, TNFRSF1A, MAP3K4, MAP3K3, GADD45G, JUND, FAS, NFATC2, GADD45A*	*EGFR, PRKCA, BRAF, MRAS, TGFBR2, NF1, CACNG2, STK3, CDC25B, MAPK13, RASGRF1, RPS6KA2, MAPK14, MAP3K1, PDGFRA, CACNA1E, MAPK8, PPP3CA, IKBKB, CACNA1C, TRAF6, RASA1, NGF*
Jak/STAT signaling pathway	*SOCS3, OSMR, GRB2, STAM2, SOCS1, CBL, IL7R, IFNAR1, STAT6, TSLP, JAK1, JAK3, IFNGR1, GHR*	*IFNAR2, CBLB, IL2RA, IL4R, STAT5A, IL15, IL21, STAT1*
TGF-β signaling pathway	*SMAD7, SMURF1, TGFB1*	*NOG, FST, SMAD2, BMPR1A*
ECM–receptor interaction	*CD36, COL3A1, COL6A2, COL6A1, COL11A2, THBS2, COL11A1, COL5A2, FN1*	*VWF, COL4A1, ITGA2, THBS1, ITGB1, SDC2*
Axon guidance	*NGEF, EPHA4, EPHA7, EPHA6, EFNB3, EFNA1, EFNB2, EPHB1, EPHA3*	*DCC, NRP1, ROCK1, PLXNA2, ITGB1, SEMA6A, NCK2, PAK7, UNC5B, FYN, GSK3B, PPP3CA, RASA1*
Complement and coagulation cascades	*KNG1, F13B, C7, FGB, SERPINC1, SERPING1, F2R*	*F11, VWF, KLKB1, SERPINA1, CFD*
Antigen processing and presentation	*HLA-DQB1, CD4, HLA-DPA1, HLA-DPB1, HLA-DMB, HLA-DQA1, HLA-F*	*CD8A, HLA-DOB, HSPA8*
Proteasome	*PSMD14, PSMD7, PSMB9*	*PSMA2, PSMB4, PSMB7, PSMD1, POMP*
Autoimmune thyroid disease	*TG, PRF1, HLA-DPA1, HLA-DPB1, HLA-DMB, TSHR, HLA-DQA1, HLA-DRA, HLA-F*	*HLA-DQA2, CD28*
Neurotrophin signaling pathway	*YWHAZ, RPS6KA4, MAP3K3, GRB2, RPS6KA2, GAB1, YWHAQ*	*ZNF274, IRS2, NFKBIB, IRS1, BDNF, NTRK2, RIPK2, NGFR, ABL1, TRAF6, IKBKB, PIK3R1, NGF*
Wnt signaling pathway	*CSNK1A1, CSNK1E, SFRP2, PPP2R5C, LRP6, FZD1, TCF7L1, LRP5*	*PRKCA, VANGL1, PPP2CA, WNT3A, MAPK8, CXXC4, SMAD2, RUVBL1, APC, PRKCB*
Pathways in cancer	*DCC, FGFR2, CTBP2, GRB2, CBL, MITF, GLI3, TGFB1, TCF7L1, DAPK1, LAMA1, ETS1, JAK1, PIK3R1, FN1, FH*	*FGFR1, PDGFB, STAT5A, KITLG, NFKB1, KIT, ZBTB16, ITGB1, MMP2, BCL2, LAMB1, TRAF6, EGFR, PIK3CG, BMP2, RXRA, IGF1, SMAD3, ITGA2, SMAD2, STAT1, PRKCB, LAMA5, MAPK8, IKBKB, ABL1*
Neuroactive ligand–receptor interaction	*PARD3, GABBR2, CRHR1, AGTR1, EDNRB, S1PR3, GRM4, PTGIR, ADRB2, CNR1, P2RY1, ADRA1B, MC4R, HTR2A, OPRD1*	*PARD3, GABBR2, CRHR1, AGTR1, EDNRB, S1PR3, GRM4, PTGIR, ADRB2, CNR1, P2RY1, ADRA1B, MC4R, HTR2A, OPRD1*

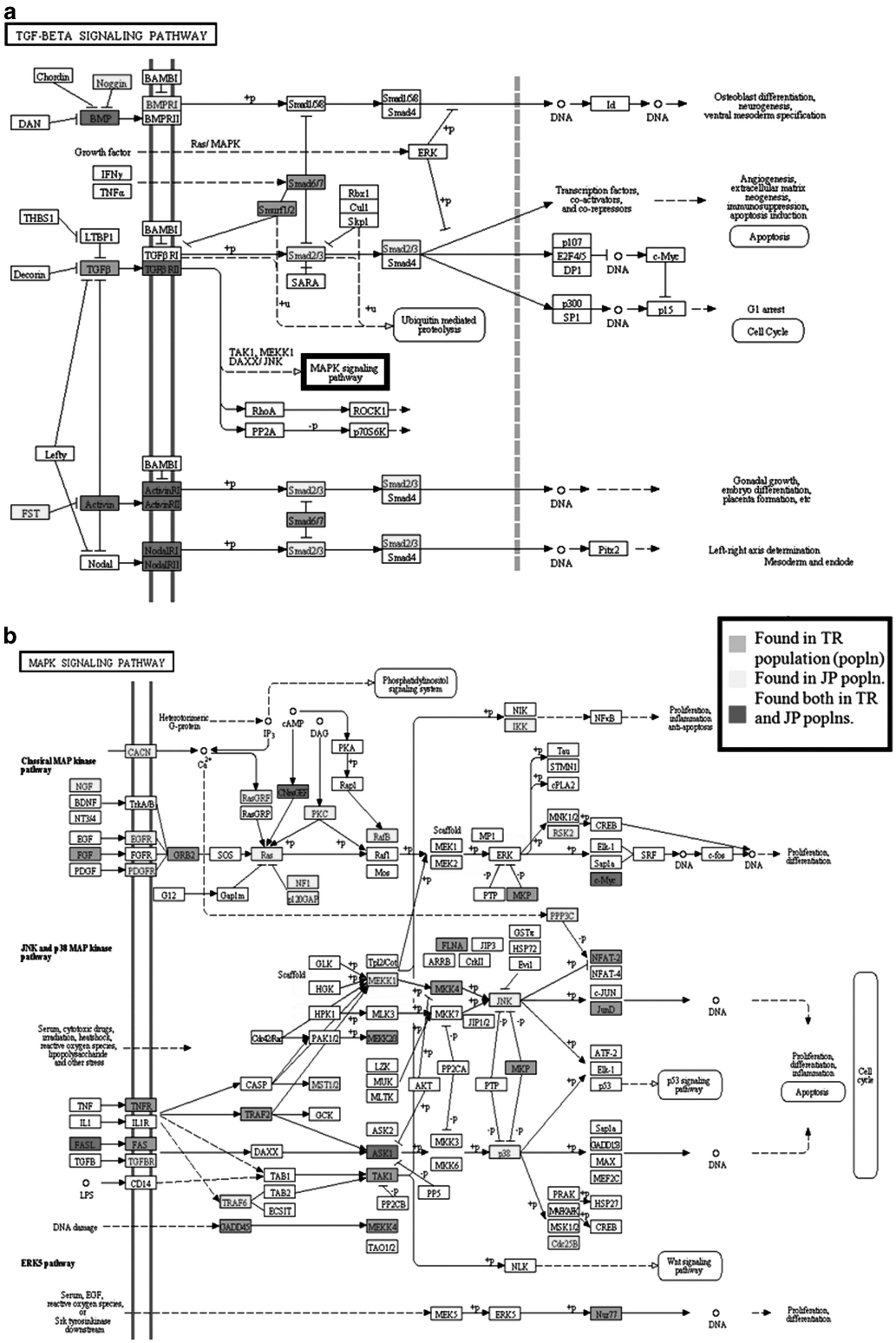

Figure 3 KEGG pathway map for TGF-β and MAPK signaling pathways. The set of genes shown in blue includes genes that are found for Turkish (TR) data set; yellow includes genes that are found for Japanese (JP) data set; red includes genes that are found both by TR and JP GWASs of Behçet's disease. A full color version of this figure is available at the *European Journal of Human Genetics* journal online.

with $P = 1.00E-18$ and $P = 2.35E-16$ in TR and JP populations, respectively. As shown in Figure 4 in red and in Table 2, in this pathway, *PLAT* (plasminogen activator, tissue), *F5* (coagulation factor V), and *F13A1* (coagulation factor XIII) genes are identified in our method by both TR and JP GWASs. Previously, abnormalities in the coagulation cascade, such as significantly higher levels of tissue

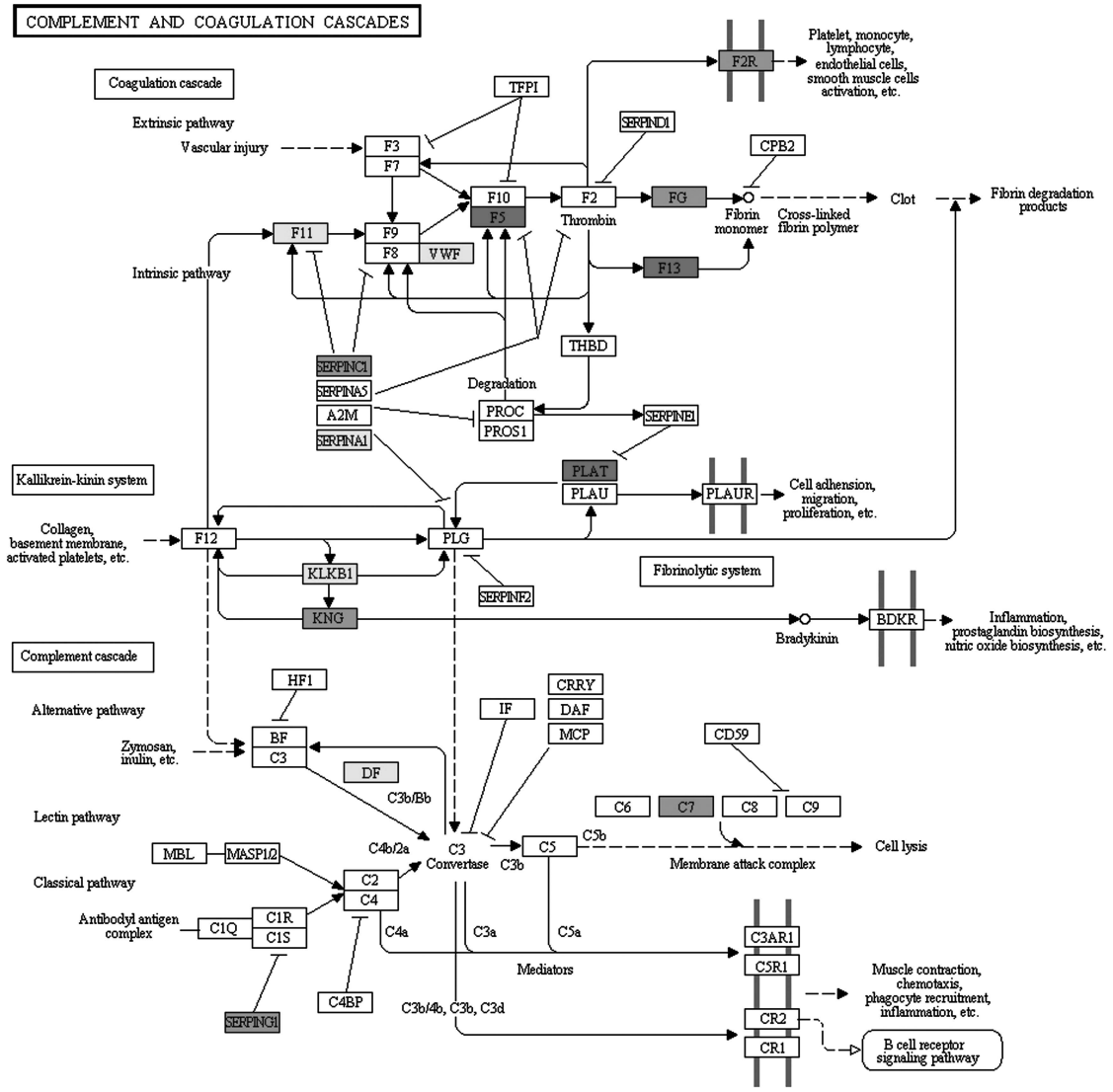

Figure 4 KEGG pathway map for complement and coagulation pathway. The set of genes shown in blue includes genes that are found for Turkish (TR) data set; yellow includes genes that are found for Japanese (JP) data set; red includes genes that are found both by TR and JP GWASs of Behçet's disease. A full color version of this figure is available at the *European Journal of Human Genetics* journal online.

plasminogen activator in patients with BD compared with healthy age-matched controls, have been reported.[44–46] Moreover, high levels of serum concentration of von Willebrand factor (vWF, an endothelial product that is shown in yellow in Figure 4), plasminogen activator inhibitor-1, and thrombomoduline were found in patients with BD.[47,48] Along this line, Demirer *et al*,[46] Ozoran *et al*,[47] and Beyan *et al*[49] showed that the level of vWF was significantly higher in patients with BD, supporting endothelial destruction because of vasculitis related with BD.[50] Factor V Leiden (*F5*) mutation was also reported as associated with thrombosis in BD.[51] Similarly, the study of Gul *et al*[52] concluded that factor V gene mutation may play a major role in the development of venous thrombosis in BD. In parallel, Batioğlu *et al*[34] indicated that the prevalence of factor V Leiden mutation was significantly higher in ocular Behçet patients as compared with healthy control subjects. Hence, they argued that factor V Leiden may be an additional risk factor in ocular BD.[34]

In this study, the Turkish and Japanese sets of patients fulfilled different but established criteria for the diagnosis of BD that overlap to a great extent and define a similar phenotype criteria.[1,2] However, depending on the features of the referral centers, some differences were observed in the distribution of BD manifestations between two sets, like more frequent ocular inflammation in the Japanese and more frequent vascular involvement in the Turkish patient groups.[1,2] These differences may have played a role in the rankings of the identified pathways in Turkish and Japanese sets, that is, proteasome pathway was ranked as the first in the Japanese and ninth in the Turkish sets, and the complement and coagulation cascade was ranked as seventh in the Turkish and tenth in the Japanese sets.

The selection of top 10 pathways for comparison was an arbitrary approach, and it may have resulted in overlooking some potentially important pathways, such as the Janus kinase (JAK)-signal transducer and activator of transcription (STAT) signaling pathway. In our study, this pathway is identified in 3rd and 17th rankings with $P = 3.68E-21$ and $P = 6.36E-14$ in TR and JP populations, respectively. The *JAK2* gene is targeted by 14 genotyped SNPs in Japanese population, and none of these SNPs target the same gene in Turkish

population. Hence, if one searches for conserved SNPs between populations, such important clues, illuminating an aspect of disease etiology, might have been missed. The genes such as *JAK2* and *STAT3* in the IL23 signaling pathway contribute to a surprising range of autoimmune diseases, including BD and inflammatory bowel disease.[53–55] As part of this pathway, low-frequency and rare variants in the FN3, fibronectin tenth type III domain of *IL23R* gene were detected by targeted resequencing of TR and JP samples.[40] As shown in Table 2, *IL2RB* gene, which has the same FN3 domain, is targeted by the SNPs in both TR and JP populations. Considering the fact that the elucidation of the JAK/STAT pathway has provided many insights into disease mechanisms and has become the basis for new pharmacologic agents, here we would like to emphasize once again the importance of the pathway-oriented approach. If we were to proceed with the strict cutoff of traditional GWAS (P-value $< 10^{-8}$) and did not integrate network and pathway-oriented analysis, we could not detect any affected genes (shown in red, yellow, and blue in Figures 3 and 4) within the identified pathways. As illustrated in Figures 3 and 4, in TGF-β signaling, MAPK signaling pathways, and complement and coagulation cascades, only a few genes are commonly targeted between the two populations. There is striking evidence that each population has its own disease mechanisms. But at the same time, the affected pathways show higher conservation among different populations. That is why we propose that GWAS data should be analyzed in this manner to extract intrinsic information hidden in the data.

This method of analysis provides opportunities for identification of new pathways critical for the disease development and as targets for better treatments. Despite the identification of the same pathways in two different populations, findings of this method need to be confirmed by other approaches such as gene expression analyses or resequencing of the genes involved in the pathways for rare variants. The strategy may also be applicable to other clinical phenotypes, although more testing will be needed to assess its potential for discovery in general.

To conclude, with the fast technological developments and continuous data production in the field of GWASs, more and more data sets are expected to be available in the near future. In this respect, we propose a novel method for the identification of disease-related pathways using the GWAS data from different populations. By applying our method on BD data set, we have shown that although the shared pathways between the TR and the JP populations explain the general mechanisms of BD development, the pathways that are identified by population-specific GWASs also need to be examined to gain a more comprehensive understanding of BD pathogenesis. Each population may search for disease-causing factors targeting the genes within these affected pathways. Rather than the population, the same method can be extended to individuals to identify modifications occurring on the genes within these pathways. Thus, we can determine individual reasons for disease development that can be exploited for drug development and personalized therapeutical applications. To understand individual disease development mechanisms, the identified disease-related pathways can be scanned for an individual for alterations in the functions of the genes contained within. Thus, determining the disease-causing factors will provide a valuable insight for individualized therapy targets that would rectify the impact of these function-altering factors.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

This work was supported by the Intensified Cooperation (IntenC): Promotion of German-Turkish Higher Education Research Grant of The Scientific and Technological Research Council of Turkey (TUBITAK; 109S218). The Abdullah Gul University Support Foundation (AGUV) supported the work of BB-G.

1 Remmers EF, Cosan F, Kirino Y *et al*: Genome-wide association study identifies variants in the MHC class I, IL10, and IL23R-IL12RB2 regions associated with Behcet's disease. *Nat Genet* 2010; **42**: 698–U678.
2 Mizuki N, Meguro A, Ota M *et al*: Genome-wide association studies identify IL23R-IL12RB2 and IL10 as Behcet's disease susceptibility loci. *Nat Genet* 2010; **42**: 703–U783.
3 Gul A: Behcet's disease: an update on the pathogenesis. *Clin Exp Rheumatol* 2001; **19**: S6–S12.
4 Seng KC, Seng CK: The success of the genome-wide association approach: a brief story of a long struggle. *Eur J Hum Genet* 2008; **16**: 554–564.
5 Visscher PM, Brown MA, McCarthy MI, Yang J: Five years of GWAS discovery. *Am J Hum Genet* 2012; **90**: 7–24.
6 Cirulli ET, Goldstein DB: Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nat Rev Genet* 2010; **11**: 415–425.
7 Fridley BL, Biernacka JM: Gene set analysis of SNP data: benefits, challenges, and future directions. *Eur J Hum Genet* 2011; **19**: 837–843.
8 Bakir-Gungor B, Sezerman OU: A new methodology to associate SNPs with human diseases according to their pathway related context. *PLoS One* 2011; **6**: e26277.
9 Baranzini SE, Galwey NW, Wang J *et al*: Pathway and network-based analysis of genome-wide association studies in multiple sclerosis. *Hum Mol Genet* 2009; **18**: 2078–2090.
10 Bebek G, Koyuturk M, Price ND, Chance MR: Network biology methods integrating biological data for translational science. *Brief Bioinform* 2012; **13**: 446–459.
11 Peng G, Luo L, Siu HC *et al*: Gene and pathway-based second-wave analysis of genome-wide association studies. *Eur J Hum Genet* 2010; **18**: 111–117.
12 Torkamani A, Topol EJ, Schork NJ: Pathway analysis of seven common diseases assessed by genome-wide association. *Genomics* 2008; **92**: 265–272.
13 Stranger BE, Bjorkegren J, Dolan ME, Ritchie MD: Systems and genome-wide approaches unite to provide a route to personalized medicine. *Genome Med* 2012; **4**: 29.
14 de las Fuentes L, Yang W, Davila-Roman VG, Gu CC: Pathway-based genome-wide association analysis of coronary heart disease identifies biologically important gene sets. *Eur J Hum Genet* 2012; **20**: 1168–1173.
15 Cho DY, Kim YA, Przytycka TM: Chapter 5: Network biology approach to complex diseases. *PLoS Comput Biol* 2012; **8**: e1002820.
16 Kim YA, Przytycka TM: Bridging the gap between genotype and phenotype via network approaches. *Front Genet* 2012; **3**: 227.
17 Barabasi AL, Gulbahce N, Loscalzo J: Network medicine: a network-based approach to human disease. *Nat Rev Genet* 2011; **12**: 56–68.
18 Schadt EE, Bjorkegren JLM: NEW: network-enabled wisdom in biology, medicine, and health care. *Sci Transl Med* 2012; **4**: 115rv1.
19 Bakir-Gungor B, Sezerman OU: The identification of pathway markers in intracranial aneurysm using genome-wide association data from two different populations. *PLoS One* 2013; **8**: e57022.
20 Bakir-Gungor B, Baykan B, Iseri SU, Tuncer FN, Sezerman OU: Identifying SNP targeted pathways in partial epilepsies with genome-wide association study data. *Epilepsy Res* 2013; **105**: 92–102.
21 Calabrese R, Capriotti E, Fariselli P, Martelli PL, Casadio R: Functional annotations improve the predictive score of human disease-related mutations in proteins. *Hum Mutat* 2009; **30**: 1237–1244.
22 Liu CT, Lin HW, Lin HH: Functional analysis of HapMap SNPs. *Gene* 2012; **511**: 358–363.
23 Hugot JP, Chamaillard M, Zouali H *et al*: Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* 2001; **411**: 599–603.
24 Palmer CN, Irvine AD, Terron-Kwiatkowski A *et al*: Common loss-of-function variants of the epidermal barrier protein filaggrin are a major predisposing factor for atopic dermatitis. *Nat Genet* 2006; **38**: 441–446.
25 Rosenberg NA, Huang L, Jewett EM, Szpiech ZA, Jankovic I, Boehnke M: Genome-wide association studies in diverse populations. *Nat Rev Genet* 2010; **11**: 356–366.
26 Bakir-Gungor B, Sezerman OU: Identification of SNP targeted pathways from genome-wide association study (GWAS) data. *Protoc Exch* 2012; **2012**: doi:10.1038/protex.2012.019.
27 Bakir-Gungor B: *Bioinformatics Approaches to Associate Single Nucleotide Polymorphisms with Human Diseases According to Their Pathway Related Context*, PhD thesis. Sabancı University: İstanbul, Turkey, 2012.
28 Bakir-Gungor B, Egemen E, Sezerman OU: PANOGA: a web-server for identification of SNP targeted pathways from genome-wide association study data. *Bioinformatics* 2013; **30**: 1287–1289.
29 Wu JM, Vallenius T, Ovaska K, Westermarck J, Makela TP, Hautaniemi S: Integrated network analysis platform for protein-protein interactions. *Nat Methods* 2009; **6**: 75–77.

30 Saccone SF, Bolze R, Thomas P *et al*: SPOT: a web-based tool for using biological databases to prioritize SNPs after a genome-wide association study. *Nucleic Acids Res* 2010; **38**: W201–W209.
31 Ideker T, Ozier O, Schwikowski B, Siegel AF: Discovering regulatory and signalling circuits in molecular interaction networks. *Bioinformatics* 2002; **18**(Suppl 1): S233–S240.
32 Kirino Y, Bertsias G, Ishigatsubo Y *et al*: Genome-wide association analysis identifies new susceptibility loci for Behcet's disease and epistasis between HLA-B*51 and ERAP1. *Nat Genet* 2013; **45**: 202–207.
33 Hughes T, Coit P, Adler A *et al*: Identification of multiple independent susceptibility loci in the HLA region in Behcet's disease. *Nat Genet* 2013; **45**: 319–324.
34 Batioglu F, Atmaca LS, Karabulut HG, Beyza Sayin D: Factor V Leiden and prothrombin gene G20210A mutations in ocular Behcet disease. *Acta Ophthalmol Scand* 2003; **81**: 283–285.
35 Ture-Ozdemir F, Tulunay A, Elbasi MO *et al*: Pro-inflammatory cytokine and caspase-1 responses to pattern recognition receptor activation of neutrophils and dendritic cells in Behcet's disease. *Rheumatology* 2013; **52**: 800–805.
36 Shimizu J, Izumi T, Arimitsu N *et al*: Skewed TGF beta/Smad signalling pathway of T cells in patients with Behcet's disease. *Clin Exp Rheumatol* 2012; **30**: S35–S39.
37 Hamzaoui K: Th17 cells in Behcet's disease: a new immunoregulatory axis. *Clin Exp Rheumat* 2011; **29**: S71–S76.
38 de Chambrun MP, Wechsler B, Geri G, Cacoub P, Saadoun D: New insights into the pathogenesis of Behcet's disease. *Autoimmun Rev* 2012; **11**: 687–698.
39 Shahneh FZ, Mohammadian M, Babaloo Z, Baradaran B: New approaches in the immunotherapy of Behçet disease. *Adv Pharm Bull* 2013; **3**: 9–11.
40 Kirino Y, Zhou Q, Ishigatsubo Y *et al*: Targeted resequencing implicates the familial Mediterranean fever gene MEFV and the toll-like receptor 4 gene TLR4 in Behcet disease. *Proc Natl Acad Sci USA* 2013; **110**: 8134–8139.
41 Shimizu J, Izumi T, Suzuki N: Aberrant activation of heat shock protein 60/65 reactive T cells in patients with Behcet's disease. *Autoimmune Dis* 2012; **2012**: 105205.
42 Lu L, Wang J, Zhang F *et al*: Role of SMAD and Non-SMAD signals in the development of Th17 and regulatory T cells. *J Immunol* 2010; **184**: 4295–4306.
43 Noubade R, Krementsov DN, del Rio R *et al*: Activation of p38 MAPK in CD4 T cells controls IL-17 production and autoimmune encephalomyelitis. *Blood* 2011; **118**: 3290–3300.
44 Haznedaroglu IC, Ozcebe OI, Ozdemir O, Celik I, Dundar SV, Kirazli S: Impaired haemostatic kinetics and endothelial function in Behcet's disease. *J Intern Med* 1996; **240**: 181–187.
45 Espinosa G, Font J, Tassies D *et al*: Vascular involvement in Behcet's disease: relation with thrombophilic factors, coagulation activation, and thrombomodulin. *Am J Med* 2002; **112**: 37–43.
46 Demirer S, Sengul N, Yerdel MA *et al*: Haemostasis in patients with Behcet's disease. *Eur J Vasc Endovasc* 2000; **19**: 570–574.
47 Ozoran K, Duzgun N, Gurler A, Tutkak H, Tokgoz G: Plasma von Willebrand factor, tissue plasminogen activator, plasminogen activator inhibitor, and antithrombin III levels in Behcet's disease. *Scand J Rheumatol* 1995; **24**: 376–382.
48 Hampton KK, Chamberlain MA, Menon DK, Davies JA: Coagulation and fibrinolytic-activity in Behcets-disease. *Thromb Haemostasis* 1991; **66**: 292–294.
49 Beyan E, Sadikoglu B, Ertugrul E, Beyan C: von Willebrand factor antigen levels in Behcet disease. *Am J Hematol* 2005; **79**: 70–72.
50 Shang H, Ye JJ, Ji M, Wang FF, Zhu YY, Qi XM: Anticoagulant and fibrinolytic disorders in patients with Behcet's disease and recurrent aphthous ulcer. *Chin J Physiol* 2011; **54**: 235–240.
51 Ricart JM, Vaya A, Todoli J *et al*: Thrombophilic risk factors and homocysteine levels in Behcet's disease in eastern Spain and their association with thrombotic events. *Thromb Haemostasis* 2006; **95**: 618–624.
52 Gul A, Ozbek U, Inanc M, Konice M, Ozcelik T: Coagulation factor V gene mutation increases the risk of venous thrombosis in Behcet's disease. *Brit J Rheumatol* 1996; **35**: 1178–1180.
53 O'Shea JJ, Holland SM, Staudt LM: JAKs and STATs in immunity, immunodeficiency, and cancer. *New Engl J Med* 2013; **368**: 161–170.
54 Polgar N, Csongei V, Szabo M *et al*: Investigation of JAK2, STAT3 and CCR6 polymorphisms and their gene-gene interactions in inflammatory bowel disease. *Int J Immunogenet* 2012; **39**: 247–252.
55 Cho JH, Gregersen PK: Genomics and the multifactorial nature of human autoimmune disease. *New Engl J Med* 2011; **365**: 1612–1623.

Supplementary Information accompanies this paper on European Journal of Human Genetics website (http://www.nature.com/ejhg)