Genetics of Behçet’s Disease: Functional Genetic Analysis and Estimating Disease Heritability

Lourdes Ortiz-Fernández and Amr H. Sawalha

Behçet’s disease is a chronic multisystemic inflammatory disorder characterized by recurrent oral and genital ulcers. Although its etiology remains unclear, it is thought that both genetic and environmental factors contribute to the onset and progression of Behçet’s disease. Here, we provide an updated view of the genetic landscape and architecture of Behçet’s disease. Large-scale genetic studies performed to date revealed 21 genetic susceptibility loci associated with the disease at a GWAS level of significance ($p$-value $= 5 \times 10^{-8}$). We performed epigenetic pattern enrichment analysis in Behçet’s disease associated loci, providing new insights into the molecular mechanisms underlying its pathophysiology. Our data suggest the crucial involvement of several immune cell types, including natural killer cells, monocytes, and B cells in the pathogenesis of the disease. Pathway enrichment analysis identified important biological processes involved. Using large-scale genetic data available from $\sim$200 immune-related loci (Immunochip), we estimate Behçet’s disease heritability to be at least 16%. We further used the same approach to estimate the heritability explained by the known Behçet’s disease-associated loci, suggesting that they explain $\sim 60\%$ of the genetic component underlying Behçet’s disease. These results indicate a significant role of non-genetic factors in causing Behçet’s disease and that additional genetic variation influencing the risk of Behçet’s disease remains to be identified. Finally, we calculated a cumulative genetic risk score across populations reinforcing the link between geographic variations in disease prevalence with its genetic component.

Keywords: Behçet’s disease, genetics, heritability, genetic risk score, epigenetic

INTRODUCTION

Behçet’s disease is a chronic inflammatory disease with a relapsing-remitting course characterized by recurrent oral and genital ulcers. It is a debilitating systemic vasculitis that can affect the eyes, skin, blood vessels, central nervous system, and gastrointestinal tract, leading to a wide range of clinical manifestations (1, 2). Behçet’s disease onset is typically in the third or fourth decade of life, although younger adults and children can be affected. Both genders are equally affected, however, a more severe course of the disease has been described in men and in the young (3). Behçet’s disease patients have been diagnosed worldwide, although its highest prevalence coincides with...
the countries stretching from East Asia to the Mediterranean region along the ancient trading route “silk road” (4, 5).

Currently, the pathogenesis of Behçet’s disease remains unclear; however, the consensus is that the disease is triggered by exposure to environmental factors, such as infectious agents and others, in individuals with a genetic susceptibility background (6). The role that infections might play in Behçet’s disease pathogenesis has long been interrogated. Poor oral health is associated with a more severe course of the disease and the results of recent studies investigating the oral microbiome revealed that the salivary microbial community in Behçet’s disease is less diverse than in healthy controls (7, 8).

The contribution of genetic factors to the pathogenesis of Behçet’s disease is strongly supported. In the early 1970’s, the identification of the human leukocyte antigen (HLA) region as the first susceptibility genetic region for Behçet’s disease was reported. By using serological methods Ohno and collaborators detected the association of HLA-B51 with Behçet’s disease (9). Almost fifty years later, the challenging task of unraveling the genetic architecture of Behçet’s disease is still ongoing.

Here we examined our current knowledge of the genetic basis of Behçet’s disease paying special attention to the findings obtained in the large-scale genetic studies performed over the recent years. We further used these genetic information to yield new insights into the genetic architecture of this disorder. We first summarized the loci that have been previously associated with Behçet’s disease. Next, we used bioinformatic tools and publicly available data to perform functional annotation and enrichment analyses, to gain insights into the molecular mechanisms driven by the genetic associations in Behçet’s disease. We pinpoint key cell types and molecular pathways that may be involved in the disease pathophysiology. Finally, we estimated the heritability of Behçet’s disease for the first time and calculated a cumulative genetic risk score across various populations for this disease.

THE CONTRIBUTION OF LARGE-SCALE GENETIC STUDIES IN DECIPHERING THE COMPLEX GENETIC ARCHITECTURE OF BEHÇET’S DISEASE

The genetic component of Behçet’s disease was initially interrogated by candidate gene association studies, which were mainly focused on investigating genes with known associations with other immune-mediated diseases. Over several decades, investigators have conducted multiple candidate gene studies analyzing a myriad of candidate genes in different populations (10). This strategy had significant limitations, and the associations identified by candidate gene studies were mostly observed in single cohorts, and generally reported weak genetic effects that were difficult to replicate in subsequent studies.

Large-scale genetic studies, and genome-wide association studies (GWAS) in particular, have provided a powerful unbiased approach to investigate the genetics of complex diseases allowing for the identification of thousands of genetic variants associated with hundreds of human complex traits and diseases (11). Large-scale genetic studies performed in Behçet’s disease are summarize in Table 1.

In 2009, Fei et al. performed the first GWAS in Behçet’s disease in a Turkish population using a DNA pooling approach (12). Although no loci were identified at a GWAS level of significance, this study constituted a milestone in studying the genetics of Behçet’s disease, and identified novel genetic susceptibility loci for the disease including genetic effects that were replicated in subsequent studies and in other populations. The main limitation of this work was the relatively low sample size to achieve the established threshold of significance for these studies ($p$-value $= 5 \times 10^{-8}$). One year later, GWAS studies including larger cohorts were performed revealing two loci associated with Behçet’s disease at GWAS level of significance: interleukin 23 receptor (IL23R) and the interleukin 10 (IL10) (*Tables 1, 2*) (13, 14). Both loci have been replicated in multiple populations and are considered established susceptibility factors for Behçet’s disease (22, 23, 25–32). The development of faster and more accurate imputation methods allowed the evaluation and statistical interrogation of additional genetic variants in complex diseases using pre-existing genetic data. In this regard, Kirino et al. detected three new associated loci after imputing from the genetic data obtained from a previous GWAS, and provided important insight into the pathogenesis of Behçet’s disease including the possibility of genetic interaction between HLA-B*51 and ERAP1 (14, 17) (*Tables 1, 2*).

The development of the Immunochip array represented another landmark in the efforts to study the genetics of immune-mediated diseases. The Immunochip is a custom array from Illumina designed to perform fine-mapping of almost 200 genetic loci relevant to multiple different immune-mediated disorders. This cost-effective genotyping platform was demonstrated to be a successful strategy in revealing and fine-mapping genetic associations in multitude of diseases (33). Several Immunochip studies have been performed, including the largest and most diverse genetic study reported to date in Behçet’s disease in which 3,477 patients and 5,967 healthy individuals from seven different populations were analyzed (24). In this recent work, eight genetic susceptibility loci were found to be associated with Behçet’s disease at a GWAS level of significance, two of which corresponded with novel genetic susceptibility loci in interferon gamma receptor 1 (IFNGR1) and the intergenic regions between LNCAROD, lncRNA activating regulator of DKK, and DKK1, dikkopf WNT signaling pathway inhibitor 1 (24).

The recent advances in the design of the genotyping arrays as well as in the statistical genetics and imputation algorithms have also proven successful in terms of analyzing the complex HLA region. The association of the classical allele HLA-B*51 is well-established and is considered the strongest genetic risk factor for Behçet’s disease. More recently, a robust genetic association between a genetic variant between HLA-B and MICA was suggested to explain the association with HLA-B*51 in Behçet’s disease. However, whether the classical allele HLA-B*51 or the HLA-B/MICA intergenic locus is causal in Behçet’s disease requires further investigation (18, 19, 26, 34, 35). It is also likely
### TABLE 1 | Summary of the large-scale genetic studies performed in Behçet’s disease.

| Year | First Author | Ethnicity           | Case/Control | Discovery stage | Replication stage | References |
|------|--------------|---------------------|--------------|-----------------|-------------------|------------|
| 2009 | Fei Y        | Turkish             | 152/172      | Affymetrix GeneChip® Human Mapping 500 K Array§ | NO (12)         |
| 2010 | Mizuki N     | Japanese            | 612/740      | Affymetrix GeneChip® Human Mapping 500 K Array | NO (13)         |
| 2010 | Remmers EF   | Turkish             | 1,321/1,336  | HumanCNV370-Duo v1.0 and Human CNV370-Quad v3.0 | NO (14)         |
| 2012 | Lee YJ       | Korean              | 379/800      | Affymetrix genome-wide human single nucleotide polymorphism (SNP) array 6.0 | NO (15)         |
| 2012 | Hou S        | Han Chinese         | 149/951      | Affymetrix genome-wide human single nucleotide polymorphism (SNP) array 6.0 | NO (16)         |
| 2013 | Kirino Y     | Turkish             | 1,209/1,278†| HumanCNV370-Duo v1.0 and Human CNV370-Quad v3.0 | YES (17)        |
| 2013 | Hughes T     | Turkish             | 503/504      | Infinium ImmunoArray-24 V.1.0 BeadChip | HLA imputation (18) |
| 2014 | Ombrello MJ  | Turkish             | 1,190/1,257 | Infinium Human CNV370 arrays (Illumina) | HLA imputation (19) |
| 2015 | Xavier JM    | Iranian             | 292/294      | Affymetrix genome-wide human single nucleotide polymorphism (SNP) array 6.0§ | NO (20)         |
| 2015 | Kappen JH    | Multi-ethnic†       | 369/5,843    | Illumina HumanHap 610K and/or 660 K arrays | NO (21)         |
| 2016 | Ortiz-Fernandez L | Spanish    | 286/1,517   | Infinium ImmunoArray-24 V.1.0 BeadChip | HLA imputation (22) |
| 2016 | Ortiz-Fernandez L | Spanish    | 286/1,517   | Infinium ImmunoArray-24 V.1.0 BeadChip | HLA imputation (22) |

(Continued)
that a haplotype effect that includes both coding and non-coding genetic variants within the HLA-B region is involved in Behçet’s disease. In addition, dense genotyping followed by imputation have revealed several independent genetic effects within the complex HLA region in Behçet’s disease, including in HLA-A and HLA-C (18, 19, 22, 23).

Considering all above, large-scale genetic studies have been the main contributors to our current knowledge of the genetic component of Behçet’s disease. To date, a total of 21 genetic susceptibility loci for Behçet’s disease that surpass the threshold for genome-wide level of significance have been discovered. These loci are summarized in Table 2 and will be subjected to additional functional analysis below.

FROM STATISTICS TO THE FUNCTION

The primary goal of identifying the loci that contribute to Behçet’s disease is to better understand the molecular mechanisms underlying its pathophysiology, which might result in the identification of novel drug targets. However, since most of the Behçet’s disease-associated polymorphisms reside in non-coding regions of the genome, linking the statistical genetic associations to the molecular functional consequences remains a challenge. Non-coding SNPs might influence the disease by disrupting regulatory elements, and although genetic variants are inherited identically across tissues, allele-specific regulatory effects might be cell type-specific. Over the last years, new bioinformatic approaches and analysis methods have allowed the integration of large-scale genetic studies with functional genomics datasets (36). In this context, cells types that are functionally relevant in the disease pathogenesis can be inferred by using methods that test for enrichment of disease-risk variants in genomic annotations.

As the cell types involved in the molecular mechanisms driving Behçet’s disease pathology are not yet completely elucidated, here we aimed to detect the specific cell subsets that are more likely to be altered by the disease-associated variants. We carried out a regulatory enrichment analysis in which we checked if the presence of epigenetic marks was more frequent than expected by chance at Behçet’s disease susceptibility loci (lead SNPs highlighted in bold in Table 2 were included in this analysis). We used GenomeRunner software (37) to analyze seven different histone marks (H3K4me1, H3K4me2, H3K4me3, H3K9ac, H3K27ac, H3K27me3, and H3K9me3) across 127 reference epigenomes obtained from the Encyclopedia of DNA Elements (ENCODE) and the Roadmap Epigenomics Consortium projects (38, 39). We selected the processed gapped peak annotation panels (38) and all common SNPs from dbSNP v142 were used as background to control for functional enrichments occurring by chance. The False Discovery Rate (FDR) calculation were applied to correct the enrichment p-values for multiple testing (FDR p-values < 0.05 were considered significant). Regarding the selected histone marks, H3K4me1 is associated with enhancers and regions downstream of transcription factor binding sites, and H3K4me2 has been related with both promoters and enhancers. H3K4me3 marks active promoters. Acetylation in H3K9 and H3K27 has been
associated with transcriptional initiation and open chromatin structure. Finally, H3K27me3 marks silenced promoters and H3K9me3 has been associated with repressive heterochromatic state (38, 39).

The results of this analysis revealed 40 significantly overrepresented histone marks in 27 cell types/tissues, which are illustrated in Figure 1. A complete list of the results is shown in Supplementary Table 1. Most epigenetic mark enrichment patterns were observed in blood cells which include a wide repertoire of immune cells. In addition, the histone marks detected represent cell activation. Therefore, the histone enrichment findings in the reported Behçet’s disease-susceptibility loci reflected a robust immunological signature which is consistent with our understanding of the disease. Specifically, the stronger enrichment patterns were observed in natural killer (NK): H3K27ac (p-value = 3.71 × 10^{-4}), H3K4me1 (p-value = 3.35 × 10^{-3}) and H3K4me3 (p-value = 1.39 × 10^{-2}); in a lymphoblastoid cell line (GM12878): H3K9ac (p-value = 9.87 × 10^{-5}), H3K27ac (p-value = 6.72 × 10^{-3}) and H3K4me2 (p-value = 2.47 × 10^{-2}); and in monocytes, including primary monocytes: H3K27ac (p-value = 5.65 × 10^{-4}), H3K4me3 (p-value = 2.18 × 10^{-3}) and H3K4me1 (p-value = 1.39 × 10^{-2}) and monocytes-CD14+: H3K4me2 (p-value = 7.65 × 10^{-3}), H3K4me3 (p-value = 1.55 × 10^{-2}) and H3K27ac (p-value = 3.67 × 10^{-2}). These results pinpoint to NK, monocytes, and B cells as important cell types likely implicated in the pathophysiology of Behçet’s disease and indicate that further investigation to elucidate their role should be conducted. In addition, the enrichments detected in T cells are remarkable as well: H3K9ac (p-value = 4.91 × 10^{-3}) in primary naïve CD8+ T cells and H3K27ac (p-value = 5.56 × 10^{-4}) in primary T cells. These findings reinforce the known involvement of T cells in the pathogenesis of Behçet’s disease (40). Finally, additional epigenetic mark co-localizations were also observed in other immune cells that have been suggested to be implicated in the disease, such as other subsets of T cells and neutrophils. Specifically, neutrophils have been postulated to be involved in the immunologic dysfunction in Behçet’s disease but the mechanism underlying their hyperfunction remains to be explored (41). Further functional studies to decipher the

| Reported gene       | Chr | SNPs                                                                 | Candidate genes | References   |
|---------------------|-----|----------------------------------------------------------------------|-----------------|--------------|
| IL12Rβ2, IL23R     | 1   | rs10889664, rs6660226, rs1496965, rs924080                          | IL37            | (13, 14, 22, 24) |
| IL10                | 1   | rs1518111, rs1518110, rs3024490, rs1800871                          |                 | (13, 14, 23, 24) |
| IL1A, IL1B          | 2   | rs3783550                                                           | IL37            | (23)          |
| TFCP2L1             | 2   | rs17006292                                                          |                 | (16)          |
| STAT4               | 2   | rs7574070, rs987200                                                  |                 | (16, 17)      |
| CCR1, CCR3          | 3   | rs7616215, rs2087726                                                 | CCR2, RFP3, FYCO1| (17, 24)      |
| IL12A-AS1, IL12A    | 3   | rs76830965, rs7753641, rs17810546, rs1874866                        | SCHIP1, IQCJ-SCHIP1| (21-24)      |
| ERAPI               | 5   | rs17482078                                                          | CAST            | (17)          |
| HLA-A               | 6   | rs9260997, rs112166694, rs114854070                                  | HLA-F, RNF39, TRIM31, PPP1R11 | (14, 18, 19) |
| HLA-B, MICA         | 6   | rs4959053, rs4947296, rs79556279, rs266490, rs1050502, rs2848713, rs7777216, rs2442736, rs116799036 | POU5F1         | (13-23)       |
| HLA-C               | 6   | rs12525170                                                          |                 | (18)          |
| IFNγR1              | 6   | rs4896243                                                           |                 | (24)          |
| RIPK2               | 8   | rs2230801                                                           |                 | (23)          |
| Intergenic LNCAROD/DKK1 | 10 | rs1660760                                                           |                 | (24)          |
| ADO, EGR2           | 10  | rs224127, rs12220700                                                 |                 | (23, 24)      |
| Intergenic JPKL/ CNTN5 | 11 | rs2848479                                                          |                 | (22)          |
| KLRC4               | 12  | rs2617170                                                           | KLRC3, KLRC2, KLRC1, KLRC1 | (17)          |
| LACC1               | 13  | rs2121034, rs2121033                                                 | Ccdc122         | (23, 24)      |
| IRF8                | 16  | rs7203487                                                           |                 | (23)          |
| IRF8                | 16  | rs11117433                                                          |                 | (23)          |
| FUT2                | 19  | rs681343                                                            | RASIP1, IZUMO1, NTN5 | (20)          |
| Intergenic CEBPB/PTPN1 | 20 | rs913678                                                           |                 | (23)          |

*Lead SNPs (or the most significant SNP in each genetic region in the largest single cohort for each associated locus) are highlighted in bold. Candidate genes suggested by functional annotation are also included.*
Immunological mechanisms involved in the development and progression of Behçet’s disease will help in better management of this disease. Our findings also revealed enrichment patterns across non-immune related tissues, such as brain and digestive tissues, which is in agreement with known organ involvement in Behçet’s disease.

Identifying the causal genes responsible for the genetic association signals detected in genetic studies represents another challenge for translating genetic findings into disease mechanisms and new treatments. This task might be relatively more straightforward for coding variants that directly disrupt the structure of a protein. However, as stated earlier, most of

| Cell_ID | Cell/tissue |
|---------|-------------|
| E029    | Primary monocytes from peripheral blood |
| E030    | Primary neutrophils from peripheral blood |
| E031    | Primary B cells from cord blood |
| E034    | Primary T cells from peripheral blood |
| E036    | Primary hematopoietic stem cells short term culture |
| E038    | Primary T helper naive cells from peripheral blood |
| E046    | Primary Natural Killer cells from peripheral blood |
| E047    | Primary T CD8+ naive cells from peripheral blood |
| E048    | Primary T CD8+ memory cells from peripheral blood |
| E051    | Primary hematopoietic stem cells G-CSF-mobilized Male |
| E062    | Primary mononuclear cells from peripheral blood |
| E115    | Dnd41 TCell Leukemia Cell Line |
| E116    | GM12878 Lymphoblastoid Cells |
| E124    | Monocytes-CD14+ RO01746 Primary Cells |
| E074    | Brain Substantia Nigra |
| E077    | Duodenum Mucosa |
| E085    | Fetal Intestine Small |
| E101    | Rectal Mucosa Donor 29 |
| E102    | Rectal Mucosa Donor 31 |
| E106    | Sigmoid Colon |
| E120    | HSMM Skeletal Muscle Myoblasts Cells |
| E005    | H1 BMP4 Derived Trophoblast Cultured Cells |
| E011    | hESC Derived CD184+ Endoderm Cultured Cells |
| E078    | Duodenum Smooth Muscle |
| E091    | Placenta |
| E093    | Fetal Thymus |
| E103    | Rectal Smooth Muscle |

FIGURE 1 | Histone marks enrichment analysis among genetic susceptibility loci previously reported at GWAS level of significance ($p$-value < $5 \times 10^{-8}$) for Behçet’s disease. GenomeRunner was used to perform the enrichment of seven histone marks in 127 cell and tissue types included in Roadmap Epigenomics Consortium and the Encyclopedia of DNA Elements (ENCODE). Dark gray color squares denote statistical significance (corrected $p$-value < 0.05). Only the 27 cell and tissue types for which significant enrichment was identified for at least one histone mark are displayed.
the disease-associated genetic variants in Behçet’s disease are in non-coding regions of the genome which makes it difficult to link them to a candidate causal gene. Genes located near the risk variants are the most immediate obvious candidates for further investigation but the gene nearest to the polymorphism might not necessarily be relevant or causal.

We evaluated the potential causality of the Behçet’s disease-associated genetic signals to identify candidate causal genes. We first explored expression quantitative trait loci (eQTL) across different human tissues by querying Genotype-Tissue Expression (GTEx) project (42). We found that 72.7% of Behçet’s disease associated genetic variants have been identified to alter expression of at least one gene (Supplementary Table 2). Next, under the assumption that chromatin architecture may influence transcriptional regulation, we investigated if interactions via long-range chromatin loops between Behçet’s disease risk variants and distal elements exist. We used the Capture HiC Plotter webtool to detect and visualize physical chromatin interactions between Behçet’s disease-associated variants and gene promoter regions across multiple cell types (43). Most of inspected signals showed multiple interactions as shown in Supplementary Table 2. Of particular interest are those interactions with the promoters of genes whose expression levels were also altered by the same genetic variant in our eQTL analysis. These genes are highlighted in bold in Supplementary Table 2 and are listed in Table 2. Further investigation should be performed to assess the potential role these identified candidate genes might play in the pathogenesis of Behçet’s disease.

Finally, we performed a pathway enrichment analysis to yield new insights into the biological processes that might be relevant in the pathogenesis of Behçet’s disease. The identification of key pathways was performed using Metascape (44) by analyzing both the annotated and candidate genes (Table 2). The Gene Ontology (GO) Biological Process branch and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways were selected and only terms with p-value < 1 × 10⁻⁴ and at least three genes were considered. Bonferroni corrected p-values (q values) were calculated as well. A detailed list of all results, including genes and p-values are displayed in Supplementary Table 3. Since ontologies are often redundant, Metascape categorized the results using a cluster-based approach to simplify their interpretation and to avoid reporting overlapping terms. 13 different clusters were significantly enriched (corrected p-values < 0.05): positive regulation of interferon-gamma production (p-value = 2.42 × 10⁻¹¹), regulation of lymphocyte mediated immunity (p-value = 4.45 × 10⁻¹⁰), inflammatory bowel disease (p-value = 6.38 × 10⁻¹⁰), cytokine-cytokine receptor interaction (p-value = 9.98 × 10⁻⁹), regulation of innate immune response (p-value = 1.35 × 10⁻⁶), regulation of interleukin-12 production (p-value = 7.20 × 10⁻⁵), cell chemotaxis (p-value = 1.38 × 10⁻⁴), cellular defense response (p-value = 2.56 × 10⁻⁴), negative regulation of leukocyte activation (p-value = 1.55 × 10⁻³), positive regulation of JNK cascade (p-value = 7.34 × 10⁻⁰³), regulation of osteoclast differentiation (p-value = 1.49 × 10⁻²), positive regulation of gliogenesis (p-value = 2.09 × 10⁻²), and response to virus (p-value = 2.31 × 10⁻²). The top clusters, including genes within each cluster, are represented in Figure 2. These results are in line with the hypothesis that Behçet’s disease can result from a disruption of the immune system after an infectious trigger. These results also highlight the close genetic relationship between Behçet’s disease and inflammatory bowel disease (45).

MISSING HERITABILITY

Despite the considerable advances made in our understanding of the genetics underlying Behçet’s disease, it is believed that additional susceptibility factors remain unidentified. However, the complete role of genetics in Behçet’s disease is unclear as no accurate estimation of the relative contribution of genetic factors vs. non-genetic factors in this disease has been reported.

Familial aggregation of Behçet’s disease has been reported repeatedly showing differences among populations (46–48). In addition, a higher familial incidence was also observed in juvenile patients with the disease compared with patients diagnosed during adulthood (49). One approach to measure the magnitude of genetic factors in complex disease is to calculate the sibling recurrence risk ratio (srs), which is defined as the ratio of the risk of being affected among the siblings of patients and the risk of being affected in the general population. In this sense, Gul and collaborators reported a 3.s of 11.4–52.5 for Behçet’s disease in Turkish population (50). In addition, a higher concordance rate of monozygotic twins compared to dizygotic twins was suggested in Turkish population (51). All these data support the important role of genetic factors in the pathogenesis of Behçet’s disease, however, they do not clarify what portion of the disease is attributable to genetic factors and how many susceptibility loci might exist.

Heritability is a population parameter that can be defined in broad-sense (H²) as the proportion of total phenotypic variation among individuals that can be attributed to genetic factors. This includes additive and non-additive genetic effects such as interaction between alleles on the same loci (dominance) or at different genetic loci (epistasis). Heritability in a narrow-sense (h²) estimates the genetic component of a phenotype that is due to additive genetic variance, and therefore, does not include possible genetic interactions (52). New approaches have been developed recently to use genome-scale SNPs data to estimate heritability in unpedigreed populations. These methods are based on the assumption that a sample of unrelated individuals is unlikely to be confounded by common environmental factors and, therefore, phenotypic similarity should correlate with genetic similarity (53). In this context, the genomic era has brought new insights into the heritability of multitude of complex human traits.

Here we aimed to estimate, the heritability of Behçet’s disease for the first time. We used data from a large case-control cohort of Turkish population including a total of 2,344 Behçet’s disease patients and 1,920 unrelated healthy individuals genotyped or imputed to include genetic data on loci represented on the Immunochip array (24). We applied the GREML-LDMS method implemented in the genome-wide complex trait analysis (GCTA) software (54, 55). To control for population stratification, we
included the first five ancestry principal components as covariates in the model. We found that Behçet’s disease heritability in narrow-sense ($h^2$) on the observed scale, which is the proportion of variance in case-control status (0 or 1), was 0.36 [Standard Error (SE) = 0.03] in our Turkish dataset. Since it is usual that the proportion of cases included in genetic studies is higher than the general prevalence of the disease, the software implements the option to correct for the ascertainment bias and the observed scale heritability can be transformed to a liability scale (56). The advantage of working on a liability scale is that heritability values can be compared across traits or populations. The proportion of cases in our cohort was 55% and a prevalence of 420/100,000 of Behçet’s disease in Turkey has been reported (57). Our estimates for Behçet’s disease liability reached 0.16 (SE = 0.02). According to the statistical power calculator implemented in GCTA (58), the probability of estimating heritability on a liability scale of $\geq 0.1$ in our cohort was $>0.83$.

We then estimated the heritability of Behçet’s disease explained by the genetic risk loci identified to date in this disease. We selected all SNPs included within $\pm 1\text{ Mb}$ of the
lead SNP of each Behçet’s disease-associated locus, as previously described (59). Unfortunately, no SNPs within the TFCP2L1 and RIPK2 associated loci were available in our dataset. Our analysis revealed that the estimated heritability of the disease using the known associated genetic regions in Behçet’s disease was 0.20 (SE = 0.02) on the observed scale and 0.094 (SE = 0.01) on the liability scale. These data suggest that the total variance explained by the currently known genetic susceptibility loci for Behçet’s disease accounts for 58.75% of the genetic contribution to the development of Behçet’s disease. This implies that ~40% of the genetic etiology of Behçet’s disease remains to be identified.

It should be noted that the SNP-based methods to estimate heritability provide a lower bound of $h^2$ since other genetic factors that influence $h^2$, such as copy number variation and insertion/deletions, are not considered. In addition, it is also worth to highlight that heritability values we obtained are minimum values of suggested heritability as they reflect the immune-related loci included on the Immunochip. In addition, we did not have data from other populations sufficiently powered to estimate heritability, and despite that heritability values are often similar between populations, our heritability estimation should not be used to predict heritability in other populations.

**CUMULATIVE GENETIC RISK SCORE**

The genetic risk score (GRS) is a simple and intuitive approach used to estimate the genetic propensity to a trait at the individual level. It is calculated by summing risk alleles corresponding to a phenotype of interest, with each allele weighted by its effect size estimated from an independent GWAS in the phenotype. Here, we aimed to calculate the genetic risk score of healthy individuals from different populations to check if genetic variance in the loci associated with Behçet’s disease correlate with the differences in the prevalence of the disease.
We followed the method described by Hughes et al. to calculate a cumulative genetic risk score (60). The lead SNP in each locus was included in the model. When a lead SNP is not clear, the most significant SNP in a given locus in the largest single cohort reported was used. In case of a clear evidence for multiple independent associations within the same genetic region, all the independent signals were represented. Association odds ratios were derived from published data in Behçet’s disease (Supplementary Table 2). We examined the GRS of a total of 2,504 individuals from the five major populations included in the 1000 Genomes Project phase 3 release: African, n = 661; Admixed American, n = 347; East Asian, n = 504; European, n = 503; South Asian, n = 489 (61). To determine if the cumulative GRS was different across populations, one-way ANOVA followed by Tukey’s multiple comparison test was performed using GraphPad Prism version 8.1.1 (GraphPad Software, La Jolla California USA). ANOVA p-values < 0.05 and Tukey’s adjusted p-values < 0.05 were considered significant.

The results obtained were in line with the prevalence data that have been reported in Behçet’s disease (4, 5). Our results revealed significant differences between the genetic risk scores among populations (ANOVA p-value < 0.0001). The highest GRS was observed in East Asian, followed by European, and South Asian populations, while Admixed Americans and African populations showed the lowest values (Figure 3). In addition, we further used the same approach to calculate the GRS considering only the HLA genetic associations to test the contribution of the HLA in the ancestral differences identified. Overall, the genetic risk within the HLA region appears to trend along the total genetic risk for Behçet’s disease across populations, with the highest HLA genetic risk detected in Asian populations and the lowest in African populations. Notably, the highest genetic risk arising from the HLA region was detected in South Asian populations (Figure 3). These data support the role of genetics in contributing to the differences in the prevalence of Behçet’s disease across populations.

FUTURE PERSPECTIVES AND CHALLENGES

Large-scale genetic studies have resulted in a significantly better understanding of the genetic predisposition to Behçet’s disease, however, additional genetic susceptibility loci remain to be identified. The lack of large sample sizes to detect disease-associations with low to medium effects remains a limitation. Increasing the sample size in genetics studies has been shown to be effective in other complex diseases (62). However, in diseases with low prevalence such as Behçet’s disease, this represents an enormous challenge and huge collaborative international efforts are needed. In this context, another important consideration is the development of well-powered patient cohorts from diverse populations. As in most of the complex traits, there is an under-representation of participants from diverse populations in Behçet’s disease genetic studies (63). The inclusion of diverse populations and transancestral mapping have been shown helpful to pinpoint causal genetic variants in complex diseases.

It is perhaps not too distant in the future that data obtained from large-scale genetic studies can be routinely applied to health care of individual patients. Nevertheless, there is still much to be learned about the genetics of Behçet’s disease and, importantly, how the associated genetic variants lead to pathogenic consequences predisposing to the disease. Novel insights can be gathered by applying new analysis methods to existing genetic data as we have demonstrated in this publication. For example, the integration of genetic data with other large functional datasets, such as epigenetic features and gene expression datasets, among others, can help to decipher which cell types are relevant for the disease. In this sense, further functional studies can be prioritized to relevant cell types involved to elucidate the molecular effects of disease-associated genetic variants. A complete understanding of the genetic contribution to the development of Behçet’s disease, the pathways and cell types involved, and the resulting functional disturbances, will help to achieve better management strategies in this complex disease.

AUTHOR CONTRIBUTIONS

All authors wrote and critically revised the manuscript.

FUNDING

This work was supported by the National Institute of Arthritis and Musculoskeletal and Skin Diseases of the National Institutes of Health (NIH) grant number R01 AR070148 to AS.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2021.625710/full#supplementary-material

REFERENCES

1. Bettiol A, Prisco D, Emmi G. Behcet: the syndrome. Rheumatology (Oxford). (2020) 59(Suppl_3):iii101–7. doi: 10.1093/rheumatology/kez626
2. Bernabe E, Marcenos W, Mather J, Phillips C, Fortune F. Impact of Behcet’s syndrome on health-related quality of life: influence of the type and number of symptoms. Rheumatology (Oxford). (2010) 49:2165–71. doi: 10.1093/rheumatology/keq251
3. Yarici H, Tuzun Y, Pazarli H, Yardakul S, Ozyazgan Y, Ozdogan H, et al. Influence of age of onset and patient’s sex on the prevalence and severity of manifestations of Behcet’s syndrome. Ann Rheum Dis. (1984) 43:783–9. doi: 10.1136/ard.43.6.783
4. Verity DH, Marr JE, Ohno S, Wallace GR, Stanford MR. Behcet’s disease, the Silk Road and HLA-B51: historical and geographical perspectives. Tissue Antigens. (1999) 54:213–20. doi: 10.1034/j.1399-0039.1999.540301.x
23. Takeuchi M, Mizuki N, Meguro A, Ombrello MJ, Kirino Y, Satorius C, et al. Genetics of Behçet’s Disease with an emphasis on the HLA region and new susceptibility loci. *PLoS ONE*. (2015) 10:e0119085. doi: 10.1371/journal.pone.0119085

24. Ortiz-Fernández L, Coit P, Yilmaz V, Ventur SP, Alibaz-Oner F, Akus K, et al. Genetic association of a gain of function interferon gamma receptor 1 (IFNGR1) polymorphism and the intergenic region LNCAROD/DKK1 with Behçet’s disease. *Arthritis Rheumatol*. (2021). doi: 10.1002/art.41637

25. Meguro A, Inoko H, Ota M, Katsuyama Y, Oka A, Okada E, et al. Genetics of Behçet disease inside and outside the MHC. *Ann Rheum Dis*. (2010) 69:747–54. doi: 10.1136/ard.2009.108571

26. Carapito R, Shahram F, Michel S, Le Gentil M, Radosavljevic M, Meguro A, et al. On the genetics of the Silk Route: association analysis of HLA, IL10, and IL23R-IL12RB2 regions with Behçet’s disease in an Iranian population. *Immunogenetics*. (2015) 67:289–93. doi: 10.1007/s00251-015-0841-6

27. Xavier JM, Shahram F, Davatchi F, Rosa A, Crespo J, Abdollahi BS, et al. Association study of IL10 and IL23R-IL12RB2 in Iranian patients with Behçet’s disease. *Arthritis Rheum*. (2012) 64:2761–72. doi: 10.1002/art.34437

28. Jiang Z, Yang P, Hou S, Du L, Xie L, Zhou H, et al. IL-23 gene confers susceptibility to Behçet’s disease in a Chinese Han population. *Ann Rheum Dis*. (2010) 69:1325–8. doi: 10.1136/ard.2009.119420

29. Kim ES, Kim SW, Moon CM, Park JJ, Kim TI, Kim WH, et al. Interactions between IL17A, IL23R, and STAT4 polymorphisms confer susceptibility to intestinal Behçet’s disease in Korean population. *Life Sci*. (2012) 90:740–6. doi: 10.1016/j.lfs.2012.03.017

30. Yalcin B, Atakan N, Dogan S. Association of interleukin-23 receptor gene polymorphism with Behçet disease. *Clin Exp Dermatol*. (2014) 39:3881–7. doi: 10.1111/ced.12400

31. Sousa I, Shahram F, Francisco D, Davatchi F, Abdollahi BS, Ghaideribarni F, et al. Brief report: association of CCR1, KLRC4, IL12A-AS1, STAT4, and ERAP1 with Behçet’s disease in Iranians. *Arthritis Rheumatol*. (2015) 67:2742–8. doi: 10.1002/art.39240

32. Wu Z, Zheng W, Xu J, Sun F, Chen H, Li P, et al. IL10 polymorphisms associated with Behçet’s disease in Chinese Han. *Hum Immunol*. (2014) 75:271–6. doi: 10.1016/j.humimm.2013.11.009

33. Cortes A, Brown MA. Promise and pitfalls of the immunochip. *Arthritis Res Ther*. (2011) 13:101. doi: 10.1186/ar3204

34. Gensterblum-Miller E, Wu W, Sawalha AH. Novel transcriptional activity and extensive allelic imbalance in the human MHC region. *J Immunol*. (2018) 200:1496–503. doi: 10.4049/jimmunol.1701061

35. Xavier JM, Davatchi F, Abade O, Shahram F, Francisco V, Abdollahi BS, et al. Characterization of the major histocompatibility complex locus association with Behçet’s disease in Iran. *Arthritis Res Ther*. (2015) 17:81. doi: 10.1186/s13075-015-0585-6

36. Cano-Gamez E, Trynka G. From GWAS to function: using functional genomics to identify the mechanisms underlying complex diseases. *Front Genet*. (2020) 11:624. doi: 10.3389/fgene.2020.00624

37. Dzomorov MG, Cara LR, Gules CB, Wren JD. GenomeRunner web server: regulatory similarity and differences define the functional impact of SNP sets. *Bioinformatics*. (2016) 32:2256–63. doi: 10.1093/bioinformatics/btw169

38. Roadmap Epigenomics C, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen Y, et al. An integrated encyclopedia of DNA elements in the human genome. *Nature*. (2012) 489:57–74. doi: 10.1038/nature11247

39. Gul A. Behçet’s disease: an update on the pathogenesis. *Clin Exp Rheumatol*. (2001) 19(Suppl 24):S6–12.

40. Neves FS, Spiller F. Possible mechanisms of neutrophil activation in Behçet’s disease. *Int Immunopharmacol*. (2013) 17:2016–2018. doi: 10.1016/j.intimp.2013.07.017

41. Consortium GT. Human genomics. The genotype-tissue expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science*. (2015) 348:648–60. doi: 10.1126/science.1262110

42. Schofield EC, Carver T, Achuthan P, Freire-Pritchett P, Spivakov M, Todd JA, et al. CHiCOP: a web-based tool for the integrative and interactive visualization of promoter capture Hi-C datasets. *Bioinformatics*. (2016) 32:2511–3. doi: 10.1093/bioinformatics/btw173

43. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun*. (2019) 10:15253. doi: 10.1038/s41467-019-09234-6
45. Valenti S, Gallizzi R, De Vivo D, Romano C. Intestinal Behcet and Crohn’s disease: two sides of the same coin. Pediatr Rheumatol Online J. (2017) 15:33. doi: 10.1186/s12969-017-0162-4

46. Zouboulis CC. Epidemiology of Adamantiades-Behcet’s disease. Ann Med Interne. (1999) 150:488–98.

47. Zouboulis CC, Kotter I, Diawari D, Kirch W, Kohl PK, Ochsendorf FR, et al. Epidemiological features of Adamantiades-Behcet’s disease in Germany and in Europe. Yonsei Med J. (1997) 38:411–22. doi: 10.3349/ymj.1997.38.6.411

48. Nishiyama M, Nakae K, Yukawa S, Hashimoto T, Inaba G, Mochizuki M, et al. A study of comparison between the nationwide epidemiological survey in 1991 and previous surveys on behcet’s disease in Japan. Environ Health Prev Med. (1999) 4:130–4. doi: 10.1007/BF02932268

49. Kone-Paut I, Geisler I, Wechsler B, Ozen S, Ozdogan H, Rozenbaum M, et al. Familial aggregation in Behcet’s disease: high frequency in siblings and parents of pediatric probands. J Pediatr. (1999) 135:89–93. doi: 10.1016/S0022-3476(99)70333-1

50. Gul A, Inanc M, Ocal L, Aral O, Konice M. Familial aggregation of Behcet’s disease in Turkey. Ann Rheum Dis. (2000) 59:622–5. doi: 10.1136/ard.59.8.622

51. Masatlioglu S, Seyahi E, Tahir Tunrali E, Fresko I, Gogus F, Senates E, et al. A twin study in Behcet’s syndrome. Clin Exp Rheumatol. (2010) 28(4 Suppl 60):S62–6.

52. Visscher PM, Hill WG, Wray NR. Heritability in the genomics era–concepts and misconceptions. Nat Rev Genet. (2008) 9:255–66. doi: 10.1038/nrg2322

53. Evans LM, Tahmasbi R, Vrieze SI, Abecasis GR, Das S, Gazal S, et al. Comparison of methods that use whole genome data to estimate the heritability and genetic architecture of complex traits. Nat Genet. (2018) 50:737–45. doi: 10.1038/s41588-018-0108-x

54. Yang JA, Lee SH, Goddard ME, Visscher PM. GCTA: a Tool for genome-wide complex trait analysis. Am J Hum Genet. (2011) 88:76–82. doi: 10.1016/j.ajhg.2010.11.011

55. Yang J, Bakshi A, Zhu Z, Hemani G, Vinkhuyzen AAE, Lee SH, et al. Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. Nat Genet. (2015) 47:1114–20. doi: 10.1038/ng.3390

56. Yang J, Zeng J, Goddard ME, Wray NR, Visscher PM. Concepts, estimation and interpretation of SNP-based heritability. Nat Genet. (2017) 49:1304–10. doi: 10.1038/ng.3941

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Ortiz-Fernández and Sawalha. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.