

## Whole Exome Sequencing Identifies Rare Protein-Coding Variants in Behçet's Disease

Mikhail Ognenovski,<sup>1</sup> Paul Renauer,<sup>1</sup> Elizabeth Gensterblum,<sup>1</sup> Ina Kötter,<sup>2</sup>  
Theodoros Xenitidis,<sup>3</sup> Jörg C. Henes,<sup>3</sup> Bruno Casali,<sup>4</sup> Carlo Salvarani,<sup>4</sup> Haner Direskeneli,<sup>5</sup>  
Kenneth M. Kaufman,<sup>6</sup> and Amr H. Sawalha<sup>1</sup>

**Objective.** Behçet's disease (BD) is a systemic inflammatory disease with an incompletely understood etiology. Despite the identification of multiple common genetic variants associated with BD, rare genetic variants have been less explored. We undertook this study to investigate the role of rare variants in BD by performing whole exome sequencing in BD patients of European descent.

**Methods.** Whole exome sequencing was performed in a discovery set comprising 14 German BD patients of European descent. For replication and validation, Sanger sequencing and Sequenom genotyping were performed in the discovery set and in 2 additional independent sets of 49 German BD patients and 129 Italian BD patients of European descent. Genetic association analysis was then performed in BD patients and 503 controls of European descent. Functional effects of associated genetic variants were assessed using bioinformatic approaches.

Exome variant calls used in the current study were provided through the National Heart, Lung, and Blood Institute Grand Opportunity (GO) Exome Sequencing Project and its ongoing NIH-funded studies: the Lung GO Sequencing Project (HL-102923), the Women's Health Initiative Sequencing Project (HL-102924), the Broad GO Sequencing Project (HL-102925), the Seattle GO Sequencing Project (HL-102926), and the Heart GO Sequencing Project (HL-103010).

<sup>1</sup>Mikhail Ognenovski, BS, Paul Renauer, BS, Elizabeth Gensterblum, BS, Amr H. Sawalha, MD: University of Michigan, Ann Arbor; <sup>2</sup>Ina Kötter, MD: Asklepios Clinic Altona, Hamburg, Germany; <sup>3</sup>Theodoros Xenitidis, MD, Jörg C. Henes, MD: University of Tübingen and University of Tübingen Hospital, Tübingen, Germany; <sup>4</sup>Bruno Casali, MD, Carlo Salvarani, MD: Azienda Ospedaliera Arcispedale Santa Maria Nuova–IRCCS di Reggio Emilia, Reggio Emilia, Italy; <sup>5</sup>Haner Direskeneli, MD: Marmara University, Istanbul, Turkey; <sup>6</sup>Kenneth M. Kaufman, PhD: Cincinnati Children's Hospital Medical Center and Cincinnati VA Medical Center, Cincinnati, Ohio.

Address correspondence to Amr H. Sawalha, MD, Division of Rheumatology, Department of Internal Medicine, University of Michigan, Room 5520 MSRB-1, SPC 5680, 1150 West Medical Center Drive, Ann Arbor, MI 48109. E-mail: asawalha@umich.edu.

Submitted for publication September 6, 2015; accepted in revised form December 3, 2015.

**Results.** Using whole exome sequencing, we identified 77 rare variants (in 74 genes) with predicted protein-damaging effects in BD. These variants were genotyped in 2 additional patient sets and then analyzed to reveal significant associations with BD at 2 genetic variants detected in all 3 patient sets that remained significant after Bonferroni correction. We detected genetic association between BD and *LIMK2* (rs149034313), involved in regulating cytoskeletal reorganization, and between BD and *NEIL1* (rs5745908), involved in base excision DNA repair ( $P = 3.22 \times 10^{-4}$  and  $P = 5.16 \times 10^{-4}$ , respectively). The *LIMK2* association is a missense variant with predicted protein damage that may influence functional interactions with proteins involved in cytoskeletal regulation by Rho GTPase, inflammation mediated by chemokine and cytokine signaling pathways, T cell activation, and angiogenesis (Bonferroni-corrected  $P = 5.63 \times 10^{-14}$ ,  $P = 7.29 \times 10^{-6}$ ,  $P = 1.15 \times 10^{-5}$ , and  $P = 6.40 \times 10^{-3}$ , respectively). The genetic association in *NEIL1* is a predicted splice donor variant that may introduce a deleterious intron retention and result in a noncoding transcript variant.

**Conclusion.** We used whole exome sequencing in BD for the first time and identified 2 rare putative protein-damaging genetic variants associated with this disease. These genetic variants might influence cytoskeletal regulation and DNA repair mechanisms in BD and might provide further insight into increased leukocyte tissue infiltration and the role of oxidative stress in BD.

Behçet's disease (BD) is a chronic inflammatory disease characterized by recurrent oral and genital ulcers, skin lesions, and uveitis (1). The disease can present with systemic manifestations including cardiovascular, gastrointestinal, musculoskeletal, and neurologic involvement (1,2). The etiology of BD is incompletely

understood; however, the disease is thought to be triggered by a dysregulated immune response to environmental stimuli from viruses, certain bacterial species, and microbial products such as heat-shock proteins (3). Aberrant epigenetic mechanisms may also contribute to BD pathogenesis with disease-associated patterns in DNA methylation that have been identified at several cytoskeletal genes in monocytes and CD4+ T cells from BD patients (4).

BD also has an important genetic component, which can be gleaned from an increased disease prevalence along the “ancient Silk Road”—a region stretching from the Mediterranean basin to Japan (5). This genetic component is further evidenced by familial aggregation of the disease, with a relative risk in siblings of affected patients compared to the general population ( $\lambda_s$ ) ranging from 11.4 to 52.5 in Turkey (6). In addition, genetic susceptibility for BD has been identified within the HLA region in multiple ethnic groups (7). Recent genome-wide association studies have expanded our understanding of the genetics underlying BD pathogenesis by facilitating the discovery of new HLA and non-HLA susceptibility loci. Indeed, multiple independent genetic associations have recently been identified in the HLA region, with the most robust association in BD localized to an intergenic region between *HLAB* and *MICA* (8–10). Additional genetic risk loci for BD include a number of non-HLA loci (*UBAC2*, *IL10*, *IL23R/IL12RB2*, *CCR1*, *STAT4*, *KLRC4*, *ERAP1*, and the *GIMAP* gene cluster) (7).

Whole exome sequencing is quickly becoming a commonly used high-throughput method for detecting protein-coding and splice-altering variants, which are more likely to be protein damaging and thus might have a more clear effect on disease pathogenesis (11). Furthermore, this technique can be used to detect rare variants which are proposed to contribute to the heritability of complex disease that remains largely unexplained by common variants (12). In addition, the low cost of whole exome sequencing compared to other sequencing techniques has led to the rapid expansion of exome databases that make it easy to compare allele frequencies, identify de novo variants, and retrieve previously reported exome data to use in new studies. Although whole exome sequencing is already a widely used approach for detecting genetic variants associated with monogenic disorders, it is becoming a viable alternative technique for elucidating genetic susceptibility in more complex genetic conditions (11).

Recent discoveries of disease-associated genetic variants have begun to illuminate the complex genetic component of BD. However, more research is required

to identify additional susceptibility loci and to determine the functional relationships between genetic associations and disease pathogenesis (13). In this study, we performed the first whole exome sequencing in BD and identified novel genetic associations by investigating rare variants with potentially protein-damaging effects in BD patients of European descent.

## PATIENTS AND METHODS

**Patient selection and DNA isolation.** Whole blood samples were isolated from BD patients of German or Italian descent, and genomic DNA was extracted using a QIAamp DNA Blood Maxi kit in accordance with the instructions of the manufacturer (Qiagen). All patients fulfilled the 1990 International Study Group criteria for diagnosis of Behçet’s disease (14). Written informed consent was obtained from all study participants prior to enrollment, and the study was approved by the institutional review boards and ethics committees from each institution involved. A total of 192 BD patients were studied. Whole exome sequencing was performed in 14 sporadic nonfamilial BD patients from Germany. Two additional cohorts of 49 German BD patients and 129 Italian BD patients were included for replication and genetic association analysis.

**Whole exome sequencing.** Whole exome sequencing was performed on genomic DNA that was isolated from whole blood samples obtained from BD patients. Exome enrichment was prepared using a TruSeq Exome Enrichment kit (Illumina), and paired-end 100-bp reads were sequenced on a HiSeq 2000 instrument (Illumina). Sequence alignment, quality assurance measures, and data analysis were performed using a DNA-Seq Analysis Package implemented in SNP & Variation Suite 7 (Golden Helix) and a Cincinnati Analytical Suite for Sequencing Informatics (15). Variant filtering was performed using the quality control measures for read depth, quality score, and a calculated ratio of the alternate allele depth to the alternate allele read depth plus the reference allele read depth, as previously described (16). Synonymous variants were excluded from further analysis.

**Sanger sequencing.** Sanger sequencing was performed in our discovery cohort for select variants in *ALOX15B*, *IL21R*, *MYO15A*, *PKP1*, and *PKP3*. Sanger sequencing confirmed whole exome sequencing results 100% of the time. Primers were designed using Primer3 software (Premier Biosoft), and primer quality was assessed using NetPrimer software (Premier Biosoft) (17). The following primers were used: for *ALOX15B*, 5'-ATTGAGTGACCCCGTTCC-3' (forward) and 5'-GAGG-AGTGTGTTTCGGGAGA-3' (reverse); for *IL21R*, 5'-GCCT-TGAAACATTGGTGCTC-3' (forward) and 5'-AGTCCTG-TGAGCCCTGAGTG-3' (reverse); for *MYO15A*, 5'-GGAC-CTGGGCGAGTATTATG-3' (forward) and 5'-GGGAGGT-CATAGGGTGGGTA-3' (reverse); for *PKP1*, 5'-CCCTCTT-CCACCCTCTTCTC-3' (forward) and 5'-CCACTCCTGTAG-CCACCACT-3' (reverse); and, for *PKP3*, 5'-GCAAAGGCAG-AAGCAGAGG-3' (forward) and 5'-ACCACCAGGTTGTT-GAGCAC-3' (reverse). Polymerase chain reaction (PCR) was performed on a T100 thermocycler (Bio-Rad) using ZymoTaq one-step master mix (Zymo) and the following settings: 95°C for 10 minutes; 40 cycles of 95°C for 30 seconds, 58.3°C for 40

seconds, and 72°C for 1 minute; and 72°C for 7 minutes. PCR amplification was then confirmed by agarose gel electrophoresis (2% agarose gel). The PCR-amplified DNA was purified and then sequenced using a 3730 XL sequencer (Applied Biosystems).

**Sequenom genotyping.** Sequenom genotyping was performed on all BD patient samples at the University of Michigan DNA Sequencing Core. Genotyping assays were designed for 73 of the 77 target variants using an Assay Design Suite (Agena Bioscience) and MassArray Assay Design 4.0 (Sequenom). Assays could not be created for rs80096349, rs143211074, rs145837446, and rs138771398 within *AGER*, *BTNL2*, *SPPL2C*, and *TNXB*, respectively. Genotyping assays were performed using the MassArray System according to the standard protocols of the manufacturer (Agena Bioscience). Single-nucleotide polymorphism (SNP) calls were automatically generated using TyperAnalyzer software (Sequenom) and then validated manually to exclude SNPs with ambiguous spectra.

**Statistical analysis of genotyping data.** Quality control measures were applied to Sequenom genotyping data to include individuals with a genotyping success rate of >90% for all variants, and then to include variants with a genotyping success rate of >90% for all individuals. Genotyping data were obtained from 503 healthy controls of European descent included in the 1000 Genomes Project (18). Association analyses were then performed using a basic allelic chi-square test with 1df via a Plink version 1.07 whole-genome data analysis toolset (19).

**Bioinformatic analyses.** Protein-damage predictions were made using SIFT, Polyphen2, MutationTaster, Mutation Assessor, and FATHMM algorithms included in SNP & Variation Suite 7 (20–24). For noncoding variants, an additional prediction was made using the FATHMM-MKL algorithm (24,25). Conservation was assessed using PhastCons and PhyloP algorithms for which we defined highly conserved loci as having a PhastCons score of >0.95 and a PhyloP score of >3.00 (26,27). Structural data were acquired for LIM domain kinase 2 (*LIMK2*) (UniProtKB accession no. P53671; Research Collaboratory for Structural Bioinformatics Protein Data Bank [RCSB PDB] ID no. 4TPT) and Nei-like DNA glycosylase 1 (*NEIL1*) (UniProtKB accession no. Q96FI4) from the RCSB PDB (28,29), and crystal structures were visualized using Visual Molecular Dynamics software version 1.9.2 (30). The predicted effects of the S605C residue change on *LIMK2* secondary structures and disulfide bond formation were determined using PsiPred and PredictProtein software, respectively (31–33). Functional protein interactions were determined using STRING 10 software with a confidence threshold of 0.800, and results were filtered to include activation, inhibition, binding, catalysis, post-translational modifications, and reaction-based interactions (34,35). Pathway enrichment analyses were performed using Panther Pathway annotations and DAVID (36,37).

## RESULTS

**Enrichment of rare genetic variants with putative protein-damaging potential.** In a discovery cohort comprising 14 BD patients of European descent, we explored exon and splice site variants using whole exome sequencing. Variants were initially filtered using the quality control measures as previously described (16). Variants that can alter the amino acid protein

sequence were then filtered to include those observed in at least 2 of the 14 sequenced patients and previously reported with a minor allele frequency (MAF) of <0.01 in a population of 6,503 individuals (Exome Variant Server; National Heart, Lung, and Blood Institute [NHLBI] Grand Opportunity [GO] Exome Sequencing Project) (<http://evs.gs.washington.edu/EVS/>). Variants were then analyzed for the potential to induce protein-damaging effects using SIFT, Polyphen2, MutationTaster, Mutation Assessor, and FATHMM functional prediction algorithms (20–24). Our analyses identified 77 rare variants in 74 genes with protein-damaging potential as predicted by at least 1 prediction algorithm. These variants included 73 nonsynonymous variants, 1 stop-gain variant, and 3 splice site variants (see Supplementary Table 1, available on the *Arthritis & Rheumatology* web site at <http://onlinelibrary.wiley.com/doi/10.1002/art.39545/abstract>). We identified multiple putative protein-damaging variants in the *SNX8* and *TTN* genes, which contained 2 and 3 variants, respectively.

**Genetic association analysis.** The putative protein-damaging variants were genotyped in the original 14 sequenced patients and 2 additional cohorts of 49 German BD patients and 129 Italian BD patients by Sequenom genotyping. In addition, genotyping data were obtained for 503 healthy controls of European descent included in the 1000 Genomes Project (18). We then applied quality control filters and performed a genetic association analysis for a total of 61 successfully genotyped variants in 178 BD patients and 503 healthy controls using a basic allelic chi-square test with 1df. Our results confirmed 2 novel genetic associations that remained significant after Bonferroni correction for multiple testing ( $P < 8 \times 10^{-4}$ ). The Bonferroni-corrected  $P$  value threshold was derived using an alpha level of 0.05 and 61 independent genetic variants tested ( $0.05/61 = 8 \times 10^{-4}$ ). These 2 BD-associated rare variants were rs149034313 (odds ratio [OR] 17.23,  $P = 3.22 \times 10^{-4}$ ), a missense variant within *LIMK2*, and rs5745908 (OR 4.35,  $P = 5.16 \times 10^{-4}$ ), a splice site variant within *NEIL1* (Table 1) (see Supplementary Table 2, available on the *Arthritis & Rheumatology* web site at <http://onlinelibrary.wiley.com/doi/10.1002/art.39545/abstract>).

**Analyses of the putative protein-damaging effects of *LIMK2* and *NEIL1* associations.** We explored the putative protein-damaging effects of rs149034313 and rs5745908 within *LIMK2* and *NEIL1*, respectively. The association in *LIMK2* (rs149034313) is found at a highly conserved locus with a high probability of inducing a protein-damaging effect (MutationTaster2 probability of 0.99999). Specifically, this variant introduces an S605C residue change within the *LIMK2* protein tyro-

**Table 1.** Genetic association analysis results for rare variants that were significantly associated with BD following Bonferroni correction ( $P < 8 \times 10^{-4}$ )\*

	Gene name	
	<i>LIMK2</i>	<i>NEIL1</i>
SNP (minor allele)	rs149034313 (G)	rs5745908 (C)
Genomic position (HG19)	Chr. 22: 31674324	Chr. 15: 75641682
Codon change	TCC → TGC	NA
MAF in cases	0.017	0.034
MAF in controls	0.001	0.008
<i>P</i>	$3.22 \times 10^{-4}$	$5.16 \times 10^{-4}$
OR (95% CI)	17.23 (2.07–143.60)	4.35 (1.76–10.74)

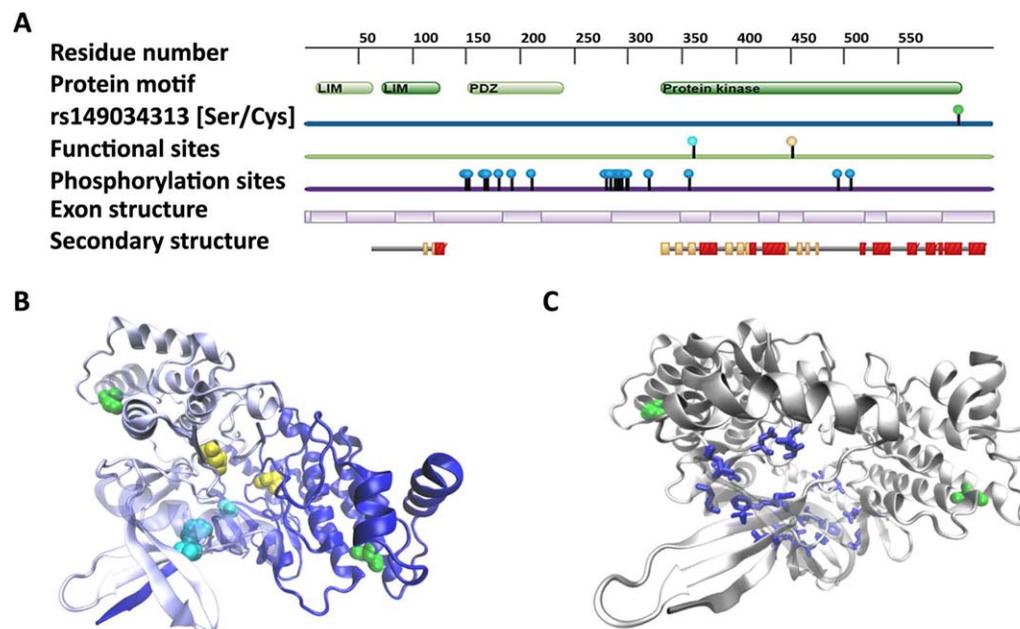
\* The analysis included 178 patients with Behçet's disease (BD) and 503 healthy controls of European descent. SNP = single-nucleotide polymorphism; HG19 = human genome version 19; Chr. = chromosome; NA = not applicable; MAF = minor allele frequency; OR = odds ratio; 95% CI = 95% confidence interval.

sine kinase domain that is responsible for serine- and threonine-specific protein phosphorylation (38). We examined how S605C interacts with residues involved in LIMK2 catalysis using the crystal structure of the homodimeric protein kinase domains. The structure shows that S605C is localized to a C-terminal  $\alpha$ -helix that is distal to the ATP binding site and active site at residues 360 and 451, respectively (38,39) (Figure 1). The S605C  $\alpha$ -helix is also distant from key structural residues related to LIMK2 activity (Leu<sup>337</sup>, Phe<sup>342</sup>, Ala<sup>345</sup>, Val<sup>358</sup>,

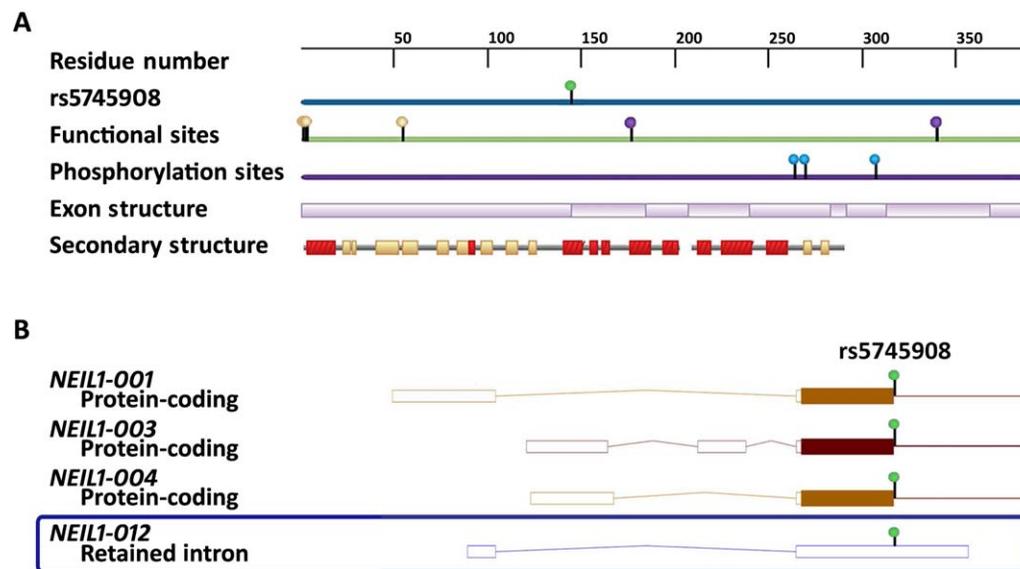
Lys<sup>360</sup>, Leu<sup>389</sup>, Ile<sup>408</sup>, Leu<sup>458</sup>, Asp<sup>469</sup>, Gly<sup>471</sup>, and Leu<sup>472</sup>) (40) (Figure 1). We then assessed whether the residue change affected LIMK2 secondary structures and disulfide bond formation. However, there was no predicted change to the S605C  $\alpha$ -helix, and no new disulfide bonds were predicted to form at S605C.

The association in *NEIL1* (rs5745908) is located 2-bp downstream of an exon splice site in an intron between the first 2 exons of the *NEIL1* exon structure (UniProtKB accession no. Q96FI4) (41) (Figure 2A). This variant was predicted to be deleterious to *NEIL1* with high probability (FATHMM-MKL probability of 0.96693; MutationTaster2 probability of 1.00000) (22,25). In addition, this *NEIL1* association is a predicted splice donor variant, a type of variant that can induce intron retentions that lead to non-coding transcripts via mechanisms of nuclear sequestration and nonsense-mediated decay, as introns often contain premature stop codons. Similarly, intron retention by this *NEIL1* variant is predicted to result in a premature intron stop codon, such as in *NEIL1-012* (Ensembl transcript ID: ENST00000564951), a noncoding *NEIL1* transcript (Figure 2B).

**Functional pathway analyses of LIMK2 and NEIL1.** We investigated downstream effects of the putative protein damage from the disease-associated variants by exploring the affected functional pathways. For LIMK2, we



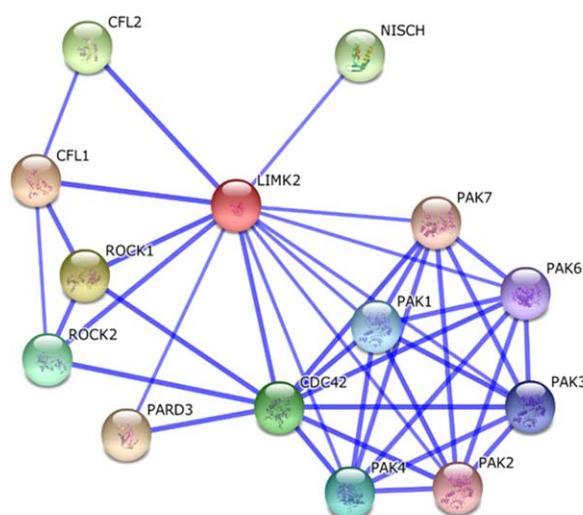
**Figure 1.** A, Protein map depicting functional and structural features of LIM domain kinase 2 (LIMK2) relative to rs149034313. Functional sites are shown with the ATP-binding site and the active site as cyan and yellow markers, respectively. In addition, secondary structures are shown with  $\alpha$ -helices and  $\beta$ -sheets as red and yellow blocks, respectively. B and C, Crystal structures depicting the homodimeric protein kinase domains of LIMK2 to show the S605C residue change (green) relative to the active site (yellow) and ATP-binding site (cyan) (B) and to key structural residues (purple) (C). LIM = LIM protein structural domain; PDZ = PDZ protein structural domain.



**Figure 2.** A, Protein map depicting functional and structural features of Nei-like DNA glycosylase 1 (NEIL1) relative to rs5745908. Functional sites are shown with DNA binding sites and active sites as purple and yellow markers, respectively. In addition, secondary structures are shown with  $\alpha$ -helices and  $\beta$ -sheets as red and yellow blocks, respectively. B, Transcript variants for *NEIL1* with respect to rs5745908, highlighting the retained intron of noncoding transcript *NEIL1-012*.

identified functional interactions with 13 proteins with high confidence ( $\geq 0.800$ ) (Figure 3). We then analyzed the LIMK2-interacting proteins for biologic pathways to identify a highly significant enrichment for cytoskeletal regulation by Rho GTPase (Bonferroni-corrected  $P = 5.63 \times 10^{-14}$ , false discovery rate [FDR] of  $2.86 \times 10^{-12}$ ) (Table 2).

Significant enrichment was also found in pathways involved in inflammation mediated by chemokine and cytokine signaling (Bonferroni-corrected  $P = 7.29 \times 10^{-6}$ , FDR of  $3.71 \times 10^{-4}$ ), T cell activation (Bonferroni-corrected  $P = 1.15 \times 10^{-5}$ , FDR of  $5.88 \times 10^{-4}$ ), and angiogenesis (Bonferroni-corrected  $P = 6.40 \times 10^{-3}$ ,



Protein	Confidence	Functional Interaction			
		Activation	Inhibition	Binding	Modification
CFL1	0.997	X	X	X	X
ROCK1	0.987	X		X	X
CFL2	0.968		X		X
CDC42	0.957	X			
ROCK2	0.932	X			X
PAK4	0.863	X			
PAK1	0.863	X			X
PAK3	0.858	X			
PAK6	0.857	X			X
PAK2	0.857	X			X
PAK7	0.855	X			X
PARD3	0.815			X	
NISCH	0.800			X	

**Figure 3.** Diagram depicting LIM domain kinase 2 (LIMK2) functional interactions with proteins based on activation, inhibition, binding, and posttranslational modifications. Interactions are represented by lines with a thickness that corresponds to the confidence level. The diagram was generated using STRING 10 software. CFL1 = cofilin 1; ROCK1 = Rho-associated coiled-coil-containing protein kinase 1; PAK4 = p21 protein (Cdc42/Rac)-activated kinase 4; PARD3 = par-3 family cell polarity regulator; NISCH = nischarin.

**Table 2.** Results of the pathway analysis of proteins that functionally interact with LIM domain kinase 2\*

Pathway	Fold enrichment	Bonferroni-corrected <i>P</i>	FDR	Proteins
Cytoskeletal regulation by Rho GTPase	26.21	$5.63 \times 10^{-14}$	$2.86 \times 10^{-12}$	PAK6, CDC42, PAK7, PAK2, ROCK1, ROCK2, PAK3, CFL2, PAK4, CFL1, PAK1
Ras pathway	20.20	$2.04 \times 10^{-6}$	$1.04 \times 10^{-4}$	PAK6, CDC42, PAK7, PAK2, PAK3, PAK4, PAK1
Axon guidance mediated by semaphorins	31.80	$3.70 \times 10^{-6}$	$1.89 \times 10^{-4}$	PAK6, PAK7, PAK2, PAK3, PAK4, PAK1
Inflammation mediated by chemokine and cytokine signaling pathway	7.69	$7.29 \times 10^{-6}$	$3.71 \times 10^{-4}$	PAK6, CDC42, PAK7, PAK2, ROCK1, ROCK2, PAK3, PAK4, PAK1
T cell activation	15.15	$1.15 \times 10^{-5}$	$5.88 \times 10^{-4}$	PAK6, CDC42, PAK7, PAK2, PAK3, PAK4, PAK1
Angiogenesis	7.02	$6.40 \times 10^{-3}$	$3.27 \times 10^{-1}$	PAK6, PAK7, PAK2, PAK3, PAK4, PAK1

\* Enrichment was performed using DAVID with Panther Pathway annotations (36,37). FDR = false discovery rate; PAK6 = p21 protein (Cdc42/Rac)-activated kinase 6; ROCK1 = Rho-associated coiled-coil-containing protein kinase 1; CFL2 = cofilin 2.

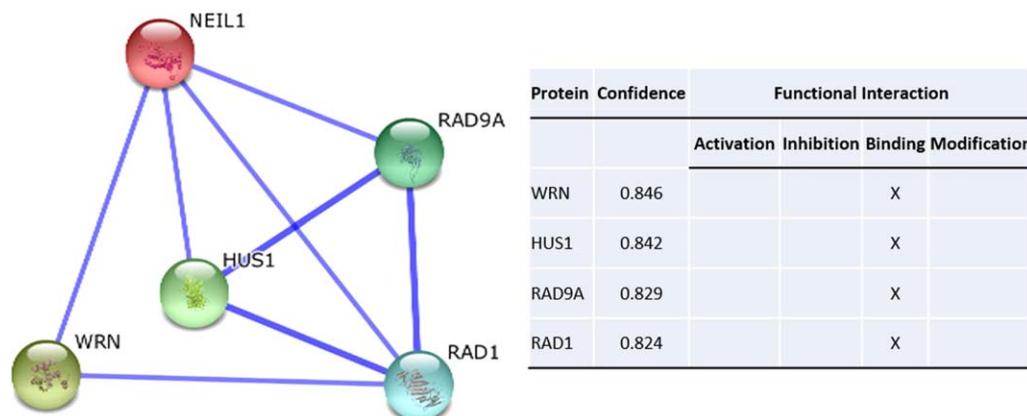
FDR of 0.327). In our analysis of NEIL1 functional interactions, we identified 4 proteins with high confidence ( $\geq 0.800$ ) (Figure 4). However, pathway analysis of these proteins revealed no significantly enriched pathways.

**DISCUSSION**

Accumulating evidence suggests that rare variants make an important contribution to complex disease. Although these variants are infrequent, separate rare variants present in different individuals may lead to the same disease by affecting related pathways (42). In addition, rare variation is proposed to contribute to the heritability of complex diseases, which is mostly unexplained by common variants (12). In this first whole exome sequencing study in BD, we identified and verified the

presence of rare variants with predicted protein-damaging effects in patients of European descent in a discovery cohort and in 2 additional independent cohorts. Our sample size had >90% power to detect a genetic association of a variant with an MAF of 0.01 and an OR of 4.0 ( $\alpha = 0.05$ , disease prevalence of 5.2/100,000).

We identified and replicated significant genetic associations with protein-damaging potential in *LIMK2* and *NEIL1*. The most significant association was localized to a highly conserved locus within *LIMK2* (rs149034313), an important regulator of actin cytoskeletal reorganization (43). Characterization of this association showed that the variant induces an S605C residue change in the LIMK2 protein kinase domain. However, the location of S605C suggests a low likelihood that it will affect the catalytic mechanism or function-related



**Figure 4.** Diagram depicting Nei-like DNA glycosylase 1 (NEIL1) functional interactions with proteins based on activation, inhibition, binding, and posttranslational modifications. Interactions are represented by lines with a thickness that corresponds to the confidence level. The diagram was generated using STRING 10 software. WRN = Werner syndrome RecQ-like helicase; HUS1 = HUS1 checkpoint clamp component; RAD9A = RAD9 checkpoint clamp component A; RAD1 = RAD1 checkpoint DNA exonuclease.

structural features due to its distance from the nucleotide binding region, ATP binding site, active site, and key structural residues (38–40). In addition, protein-modeling algorithms showed that S605C would not interrupt secondary protein structures, and the cysteine residue change is unlikely to form new disulfide bonds within the known crystallized portion of the protein. However, the location of S605C on the surface of LIMK2 may render it more susceptible to damage by posttranslational modifications. Cysteine thiol groups are known targets of both reactive oxygen species (ROS) and reactive nitrogen species, which can induce a number of reversible and irreversible posttranslational modifications to proteins at critical cysteines (44). These modifications could potentially affect a protein's charge and/or structure, thereby changing its function, location, or association with other proteins.

Of course, further studies will be required to elucidate the specific effect of this residue change on LIMK2 and to determine whether this amino acid change will affect LIMK2 interactions with other proteins. Due to the high number of LIMK2 activators and modifiers, there is a strong possibility that functional interactions are affected by the residue change.

We next investigated the impact of putative damage of LIMK2 protein on cellular pathways. Our analysis revealed LIMK2 interactions with 13 proteins that were significantly enriched for several BD-associated pathways including angiogenesis, inflammation, T cell activation, and cytoskeletal regulation. Indeed, BD pathogenesis has been linked to abnormal levels of several angiogenic factors such as angiopoietin 1 (45,46), monocyte chemoattractant protein 1 (47), serum vascular endothelial growth factor (48), and soluble VE-cadherin (49). BD is also associated with tissue inflammation, which is accompanied by leukocyte infiltration, predominantly from T cells and monocytes (50–52). In addition, cytoskeletal remodeling is central to leukocyte adhesion and infiltration in BD, and the genes involved in this remodeling exhibit dysfunctional DNA methylation that reverses upon disease remission (4). Cytoskeletal remodeling is also directly regulated by LIMK2 through the phosphorylation/inactivation of cofilin, a protein responsible for actin filament depolymerization. This induces the formation of stress fibers and focal complexes in response to LIMK2 activation by Rho and Cdc42 pathways, respectively (43). Importantly, this process is directly relevant to the leukocyte motility and tissue infiltration that are increased in BD patients.

In addition, LIMK2 has been shown to play a role in extracellular matrix adhesion in the epidermal basal layer, and the expression of a related LIM kinase

(LIMK1) has been shown to be down-regulated and involved in the pathogenesis of psoriasis (53). Up-regulation of cofilin and LIMK2 has been described in a model of pulmonary artery inflammation (54), a characteristic and serious clinical feature in some patients with BD (55). Recent studies have also identified a role for LIMK2 in the stress response pathway as a transcriptional target of p53. In this role, genotoxic stress up-regulated LIMK2 to cause a “pro-survival” cell cycle arrest at the G<sub>1</sub> phase (56). Conversely, LIMK2 inhibition rendered cells more susceptible to apoptosis when exposed to genotoxic stress. Therefore, putative damage of LIMK2 protein may increase apoptosis susceptibility in response to the oxidative stress that is increased in active BD (57). Taken together, LIMK2 has a central role in the complex cellular pathways that are implicated in BD pathogenesis, and the association detected at rs149034313 might present a novel disease target for future studies.

We detected a second rare variant BD association within *NEIL1*, which encodes a DNA glycosylase involved in initiating the base excision repair (BER) pathway (58). This association is a predicted splice donor variant which has deleterious effects on *NEIL1* by inducing intron retention and introducing a premature intron stop codon exactly like that in *NEIL1-012* (ENST00000564951), a noncoding *NEIL1* transcript. Given the role of *NEIL1* in BER initiation, this deleterious effect could have extensive consequences for DNA damage repair following oxidative stress, especially because *NEIL1* is important in repairing mutagenic oxidatively modified bases (58). Importantly, BD inflammation exposes cells to abnormal levels of ROS, and DNA damage repair machinery is essential for managing the effects of this oxidative stress.

In this study, we performed the first whole exome sequencing in BD and then identified and confirmed significant novel genetic associations in patients of European descent. Further analysis of these associations revealed potential protein-damaging effects in *LIMK2* and *NEIL1*, which play considerable roles in cell motility and DNA damage repair. As these functions are paramount to the leukocyte infiltration and oxidative stress that are increased in BD, the rare variant associations identified might present novel disease targets and key pathways to explore in future studies.

#### ACKNOWLEDGMENTS

The authors would like to thank the NHLBI GO Exome Sequencing Project and its ongoing studies which produced and provided exome variant calls for comparison: the Lung GO Sequencing Project, the Women's Health Initiative

Sequencing Project, the Broad GO Sequencing Project, the Seattle GO Sequencing Project, and the Heart GO Sequencing Project.

#### AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be published. Dr. Sawalha had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

**Study conception and design.** Sawalha.

**Acquisition of data.** Ognenovski, Renauer, Kötter, Xenitidis, Henes, Casali, Salvarani, Direskeneli, Sawalha.

**Analysis and interpretation of data.** Ognenovski, Renauer, Gensterblum, Kaufman, Sawalha.

#### REFERENCES

- Melikoglu M, Kural-Seyahi E, Tascilar K, Yazici H. The unique features of vasculitis in Behçet's syndrome. *Clin Rev Allergy Immunol* 2008;35:40–6.
- Mendes D, Correia M, Barbedo M, Vaio T, Mota M, Goncalves O, et al. Behçet's disease: a contemporary review. *J Autoimmun* 2009; 32:178–88.
- Hatemi G, Seyahi E, Fresko I, Talarico R, Hamuryudan V. Behçet's syndrome: a critical digest of the 2013-2014 literature. *Clin Exp Rheumatol* 2014;32 Suppl 84:S112–22.
- Hughes T, Ture-Ozdemir F, Alibaz-Oner F, Coit P, Direskeneli H, Sawalha AH. Epigenome-wide scan identifies a treatment-responsive pattern of altered DNA methylation among cytoskeletal remodeling genes in monocytes and CD4+ T cells from patients with Behçet's disease. *Arthritis Rheum* 2014;66:1648–58.
- Verity DH, Marr JE, Ohno S, Wallace GR, Stanford MR. Behçet's disease, the Silk Road and HLA-B51: historical and geographical perspectives. *Tissue Antigens* 1999;54:213–20.
- Gul A, Inanc M, Ocal L, Aral O, Konice M. Familial aggregation of Behçet's disease in Turkey. *Ann Rheum Dis* 2000;59: 622–5.
- Gul A. Genetics of Behçet's disease: lessons learned from genomewide association studies. *Curr Opin Rheumatol* 2014;26: 56–63.
- Hughes T, Coit P, Adler A, Yilmaz V, Aksu K, Duzgun N, et al. Identification of multiple independent susceptibility loci in the HLA region in Behçet's disease. *Nat Genet* 2013;45:319–24.
- 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.
- 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.
- Johar AS, Anaya JM, Andrews D, Patel HR, Field M, Goodnow C, et al. Candidate gene discovery in autoimmunity by using extreme phenotypes, next generation sequencing and whole exome capture. *Autoimmun Rev* 2015;14:204–9.
- Eichler EE, Flint J, Gibson G, Kong A, Leal SM, Moore JH, et al. Missing heritability and strategies for finding the underlying causes of complex disease. *Nat Rev Genet* 2010;11:446–50.
- Sawalha AH, Dozmorov MG. Epigenomic functional characterization of genetic susceptibility variants in systemic vasculitis. *J Autoimmun* 2016;67:76–81.
- Criteria for diagnosis of Behçet's disease. International Study Group for Behçet's Disease. *Lancet* 1990;335:1078–80.
- Patel ZH, Kottyan LC, Lazaro S, Williams MS, Ledbetter DH, Tromp H, et al. The struggle to find reliable results in exome sequencing data: filtering out Mendelian errors. *Front Genet* 2014;5:16.
- Kaufman KM, Linghu B, Szustakowski JD, Husami A, Yang F, Zhang K, et al. Whole-exome sequencing reveals overlap between macrophage activation syndrome in systemic juvenile idiopathic arthritis and familial hemophagocytic lymphohistiocytosis. *Arthritis Rheum* 2014;66:3486–95.
- Rozen S, Skaletsky H. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol* 2000;132:365–86.
- 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. *Nature* 2012;491:56–65.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007; 81:559–75.
- Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* 2009;4:1073–81.
- Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. *Nat Methods* 2010;7:248–9.
- Schwarz JM, Cooper DN, Schuelke M, Seelow D. MutationTaster2: mutation prediction for the deep-sequencing age. *Nat Methods* 2014;11:361–2.
- Reva B, Antipin Y, Sander C. Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Res* 2011;39:e118.
- Shihab HA, Gough J, Cooper DN, Stenson PD, Barker GL, Edwards KJ, et al. Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. *Hum Mutat* 2013;34:57–65.
- Shihab HA, Rogers MF, Gough J, Mort M, Cooper DN, Day IN, et al. An integrative approach to predicting the functional effects of non-coding and coding sequence variation. *Bioinformatics* 2015;31:1536–43.
- Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, et al. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res* 2005;15:1034–50.
- Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res* 2010;20:110–21.
- Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. *Nucleic Acids Res* 2000;28:235–42.
- Goodwin NC, Cianchetta G, Burgoon HA, Healy J, Mabon R, Strobel ED, et al. Discovery of a type III inhibitor of LIM kinase 2 that binds in a DFG-out conformation. *ACS Med Chem Lett* 2015;6:53–7.
- Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *J Mol Graph* 1996;14:33–8, 27–8.
- Buchan DW, Minnici F, Nugent TC, Bryson K, Jones DT. Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic Acids Res* 2013;41:W349–57.
- Jones DT. Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol* 1999;292:195–202.
- Rost B, Yachdav G, Liu J. The PredictProtein server. *Nucleic Acids Res* 2004;32:W321–6.
- Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguéz P, et al. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res* 2011;39:D561–8.
- Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res* 2015;43:D447–52.
- Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 2009;4:44–57.

37. Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* 2009;37:1–13.
38. Okano I, Hiraoka J, Otera H, Nunoue K, Ohashi K, Iwashita S, et al. Identification and characterization of a novel family of serine/threonine kinases containing two N-terminal LIM motifs. *J Biol Chem* 1995;270:31321–30.
39. Manetti F. LIM kinases are attractive targets with many macromolecular partners and only a few small molecule regulators. *Med Res Rev* 2012;32:968–98.
40. Shen M, Zhou S, Li Y, Li D, Hou T. Theoretical study on the interaction of pyrrolopyrimidine derivatives as LIMK2 inhibitors: insight into structure-based inhibitor design. *Mol Biosyst* 2013;9:2435–46.
41. Doublet S, Bandaru V, Bond JP, Wallace SS. The crystal structure of human endonuclease VIII-like 1 (NEIL1) reveals a zincless finger motif required for glycosylase activity. *Proc Natl Acad Sci U S A* 2004;101:10284–9.
42. McClellan J, King MC. Genetic heterogeneity in human disease. *Cell* 2010;141:210–7.
43. Sumi T, Matsumoto K, Takai Y, Nakamura T. Cofilin phosphorylation and actin cytoskeletal dynamics regulated by rho- and Cdc42-activated LIM-kinase 2. *J Cell Biol* 1999;147:1519–32.
44. Kiraz S, Ertenli I, Calguneri M, Ozturk MA, Haznedaroglu IC, Altun B, et al. Interactions of nitric oxide and superoxide dismutase in Behçet's disease. *Clin Exp Rheumatol* 2001;19 Suppl 24:S25–9.
45. Choe JY, Park SH, Kim SK. Serum angiopoietin-1 level is increased in patients with Behçet's disease. *Joint Bone Spine* 2010;77:340–4.
46. Bassyouni IH, Sharaf M, Wali IE, Mansour HM. Clinical significance of Angiopoietin-1 in Behçet's disease patients with vascular involvement. *Heart Vessels* 2015. E-pub ahead of print.
47. Bozoglu E, Dinc A, Erdem H, Pay S, Simsek I, Kocar IH. Vascular endothelial growth factor and monocyte chemoattractant protein-1 in Behçet's patients with venous thrombosis. *Clin Exp Rheumatol* 2005;23 Suppl 38:S42–8.
48. Cekmen M, Evreklioglu C, Er H, Inaloz HS, Doganay S, Turkoz Y, et al. Vascular endothelial growth factor levels are increased and associated with disease activity in patients with Behçet's syndrome. *Int J Dermatol* 2003;42:870–5.
49. Habibagahi Z, Habibagahi M, Heidari M. Raised concentration of soluble form of vascular endothelial cadherin and IL-23 in sera of patients with Behçet's disease. *Mod Rheumatol* 2010;20:154–9.
50. Tulunay A, Dozmorov MG, Ture-Ozdemir F, Yilmaz V, Eksioğlu-Demiralp E, Alibaz-Oner F, et al. Activation of the JAK/STAT pathway in Behçet's disease. *Genes Immun* 2015;16:170–5.
51. Hirohata S. Histopathology of central nervous system lesions in Behçet's disease. *J Neurol Sci* 2008;267:41–7.
52. Renauer P, Coit P, Sawalha AH. Epigenetics and vasculitis: a comprehensive review. *Clin Rev Allergy Immunol* 2015. E-pub ahead of print.
53. Honma M, Benitah SA, Watt FM. Role of LIM kinases in normal and psoriatic human epidermis. *Mol Biol Cell* 2006;17:1888–96.
54. Dai YP, Bongalon S, Tian H, Parks SD, Mutafova-Yambolieva VN, Yamboliev IA. Upregulation of profilin, cofilin-2 and LIMK2 in cultured pulmonary artery smooth muscle cells and in pulmonary arteries of monocrotaline-treated rats. *Vascul Pharmacol* 2006;44:275–82.
55. Celik S, Yazici Y, Sut N, Yazici H. Pulmonary artery aneurysms in Behçet's syndrome: a review of the literature with emphasis on geographical differences. *Clin Exp Rheumatol* 2015;33 Suppl 94:S54–9.
56. Croft DR, Crighton D, Samuel MS, Lourenco FC, Munro J, Wood J, et al. p53-mediated transcriptional regulation and activation of the actin cytoskeleton regulatory RhoC to LIMK2 signaling pathway promotes cell survival. *Cell Res* 2011;21:666–82.
57. Acikgoz N, Ermis N, Yagmur J, Cansel M, Karıncaoglu Y, Atas H, et al. Elevated oxidative stress markers and its relationship with endothelial dysfunction in Behçet disease. *Angiology* 2011;62:296–300.
58. Hazra TK, Izumi T, Boldogh I, Imhoff B, Kow YW, Jaruga P, et al. Identification and characterization of a human DNA glycosylase for repair of modified bases in oxidatively damaged DNA. *Proc Natl Acad Sci U S A* 2002;99:3523–8.