Published in final edited form as:
Genes Immun. 2011 April ; 12(3): 176–182. doi:10.1038/gene.2010.64.

Sequencing of TNFAIP3 and Association of Variants with Multiple Autoimmune Diseases

Stacy L Musone[1], **Kimberly E Taylor**[2], **Joanne Nititham**[2], **Catherine Chu**[1], **Annie Poon**[1], **Wilson Liao**[1,3], **Ernest T Lam**[1], **Averil Ma**[4], **Pui-Yan Kwok**[1,3], and **Lindsey A Criswell**[2]

[1]Institute for Human Genetics, University of California San Francisco, San Francisco, California 94143, USA.

[2]Rosalind Russell Medical Research Center for Arthritis, Department of Medicine, University of California San Francisco, San Francisco, California 94143, USA.

[3]Department of Dermatology, University of California San Francisco, San Francisco, California 94143, USA.

[4]Colitis and Crohn's Disease Center, Department of Medicine, University of California San Francisco, San Francisco, California 94143, USA.

Abstract

The *TNFAIP3* locus at 6q23, encoding A20, has been associated with multiple autoimmune diseases (AIDs). Here, we sequence the coding portions of the gene in order to identify contributing causal polymorphisms that may explain some of the observed associations. A collection of 123 individuals from the Multiple Autoimmune Disease Genetics Consortium (MADGC) collection, each with multiple AIDs (mean=2.2 confirmed diagnoses), and 397 unrelated healthy controls were used for initial sequencing. Thirty-two polymorphisms were identified in the sequencing experiments, including 16 novel and 11 coding variants. Association testing in the entire MADGC collection (1,008 Caucasians with one or more AIDs and 770 unaffected family controls) revealed association of a novel intronic insertion/deletion polymorphism with rheumatoid arthritis (OR = 2.48, p = 0.041). Genotyping of the most common coding polymorphism, rs2230926, in the MADGC collection and additional control individuals revealed significant association with Sjögren's syndrome (OR = 3.38, p = 0.038), Crohn's disease (OR = 2.25, p = 0.041), psoriasis (OR = 0.037, p = 0.036) and rheumatoid arthritis (OR = 1.9, p = 0.025). Lastly, haplotype and additional testing of polymorphisms revealed that cases were enriched for 5' and 3' untranslated region (UTR) variants (one-sided p-value=0.04), but not specifically for common (>2% MAF), rare, exonic, intronic, non-synonymous, or synonymous variants.

Keywords

resequencing; genetic association study; autoimmunity; a20; TNFAIP3

Introduction

Autoimmune diseases (AIDs) are characterized by the misidentification of self as foreign with a resultant immune response that attacks one's own cells and organs. Inheritance

Conflict of Interest
The authors declare no conflicts of interest.

patterns have been studied for many of these disorders and they are generally accepted as having a genetic component to susceptibility. Genetic predisposition is multifactorial and disease incidence varies from rare (e. g. idiopathic thrombocytopenic purpura has a population prevalence of .08% in U. S. adults) to common (rheumatoid arthritis (RA) has a U.S. population prevalence of 1%[1]). Although AIDs affect different systems or organs, it has been frequently observed that these diseases can cluster in families and even within individuals. One example is a study of AID clustering in families with multiple sclerosis (MS) described by Barcellos et al.[2]

With such overlapping disease prevalence, it is not surprising that several genetic loci have been associated with more than one AID. The hallmark locus is the major histocompatibility complex (MHC) region on chromosome six, which plays an important role in autoimmunity. Another locus is the *TNFAIP3* gene and surrounding genomic locus which has to date been associated with RA[3–5], systemic lupus erythematosus (SLE)[6–12], psoriasis[13], celiac disease[14, 15], type 1 diabetes[16], ulcerative colitis[17], Crohn's disease[18] and juvenile idiopathic arthritis[19]. This gene encodes A20, a protein involved in inhibiting signals from the tumor necrosis factor, toll-like receptor and nucleotide-binding oligomerization domain pathways[20–23]. Dysregulation of these pathways results in inflammation and programmed cell death.

With the exception of missense polymorphism rs2230926 (F127C) in SLE[6, 7, 11] and RA[11], associations to date have been identified outside of coding regions of the gene. One explanation for such associations is that the polymorphisms are in linkage disequilibrium with putatively causal polymorphisms that were not genotyped directly. We sought to identify such variants for these AID associations by sequencing the coding portions of the gene in individuals from the Multiple Autoimmune Disease Genetics Consortium (MADGC) collection[24]. This collection includes families affected by more than one AID, and here we perform sequencing in individuals who are themselves affected by more than one disease. This sample set provides an opportunity to search for variants that may be relevant to more than one AID, as shown for the *PTPN22* AID associations using this same collection[24].

TNFAIP3 (NM_006290), at 6q23, is composed of nine exons with a non-coding exon one and partially coding exon nine. The 790 amino acids include an N-terminal cysteine protease OTU domain (Cys103) and seven C-terminal zinc finger motifs that perform its deubiquitination and ubiquitination functions[25], respectively. In this study, we sequence 123 individuals each affected with multiple AIDs and 397 healthy controls. We also perform genotyping of the most common coding polymorphism, rs2230926, which has previously been associated with SLE and RA, and a novel insertion/deletion variant identified here in the entire MADGC collection.

Results

Sequencing of *TNFAIP3* in Cases and Controls

We identified 33 polymorphisms through the sequencing of 246 case and 794 control chromosomes (Table 1). One was dropped from analysis for being out of Hardy-Weinberg equilibrium (HWE) (rs3214646) and probably does not represent a true polymorphic locus. Eleven were in protein coding regions; eight of these were non-synonymous and three were synonymous. The synonymous SNP, Leu725Leu, is located in zinc-finger motif six. Sixteen were novel, or not in the public database dbSNP Human Build 131, including nine of the coding variants. Seven variants failed to sequence and thus were missing from the control sequencing data and one from the case sequencing data and were not included in comparisons between cases and controls. For the two variants detected in cases only,

rs5029964 was in one of two family members sequenced and novel_2 was in an individual with no other family members sequenced.

Association Testing of Sequenced Variants

Fisher's exact tests of association were performed for 24 SNPs and insertion/deletion polymorphisms and empirical p-values calculated through permutation (Table 2). Comparison of 91 Caucasian, unrelated, multiply affected individuals to 397 controls revealed significant association for one intronic insertion/deletion polymorphism with multiple AID diagnoses [Novel_4; Empirical p=0.005 after 4700 permutations; OR (95% CI) = 7.05 (1.67 – 29.79)]. One SNP was not polymorphic in this restricted dataset (rs5029964) and was therefore not tested for association.

An omnibus test for association of 24-marker haplotypes with a frequency at least 1% was highly significant, with a p-value of 2.9×10^{-05}. Keeping the haplotype frequency threshold of 1% or greater revealed three significant haplotypes, none of which contained the risk allele of Novel_4 (data not shown). When we included haplotypes of all frequencies, eight reached significance given an alpha of 0.05 and one was borderline significant (Table 3 contains results for these nine haplotypes). Two of these haplotypes contained the risk allele of Novel_4 and none contained the risk allele for rs2230926, the previously associated coding polymorphism. Additionally, we tested for differences in polymorphisms between cases and controls with weighted sums analysis and found cases to be enriched for 5' and 3' untranslated region (UTR) variants (one-sided p-value=0.04), but not for common (>2% MAF), rare, exonic, intronic, non-synonymous, or synonymous variants.

Association Testing of novel_4 & rs2230926

We further evaluated the novel_4 and rs2230926 SNPs by testing associations in the full MADGC collection, in particular for the presence of multiple AIDs, any AID, and individual AIDs versus healthy controls. In order to prevent false positives due to population stratification, we also tested significance of three ancestry informative markers (AIMs)[28] with rs2230926 and novel_4 genotypes and between control groups. None of the AIMS were significantly associated with rs2230926 or novel_4 genotypes, suggesting that stratification is not a problem in these analyses.

First, to confirm the association of novel_4 with multiple AID diagnoses, the insertion-deletion polymorphism was genotyped in all Caucasian individuals in the MADGC collection, which included 1008 affected participants and 770 unaffected family controls. Software which can account for the relatedness among MADGC cases and controls was used for association testing (see Methods). Significant association with multiple AID diagnoses was not observed (p= 0.337; N cases = 142), but the odds ratio remained greater than 1 (OR = 1.67 (95% CI: 0.61–4.57). However, the novel_4 variant was significantly associated with RA (OR = 2.48, p = 0.041; N cases = 146).

Since the coding SNP rs2230926 was previously associated with SLE, we genotyped this variant in the same group of Caucasian individuals in the MADGC collection, as described above for the novel_4 variant. We also genotyped this coding SNP in an additional 743 unrelated healthy Caucasian controls from a study of MS[26] and the Study Of PHarmacogenetics In Ethnically diverse populations (SOPHIE) collection[27] (Table 5). The largest OR observed was for Sjögren's syndrome (N=18; p= 0.038; OR= 3.38), followed by Crohn's disease (N=50; p= 0.041; OR= 2.25), psoriasis (N=78; p= 0.037; OR= 2.17), and RA (N=148; p=0.025; OR= 1.9). However, these results should be interpreted with caution given the limited sample sizes for individual AIDs.

Discussion

This study represents, to our knowledge, the first comprehensive screening of coding exons of *TNFAIP3*, which encodes A20. We have sequenced a population affected by multiple AIDs given published association with several individual diseases. Additionally, we performed more extensive genotyping and association testing of the previously associated coding SNP rs2230926 and a novel insertion/deletion polymorphism among a larger group of Caucasian AID patients and control individuals.

We identified 32 polymorphisms, 16 novel and 11 coding, in cases and controls. One of the novel polymorphisms (novel_4), which is an intronic insertion/deletion polymorphism, was significantly associated with risk of RA based on our analysis of individuals enrolled in the MADGC collection. These results warrant validation in other collections.

The previously identified coding SNP in exon 3 (rs2230926) was significantly associated with risk of Sjögren's syndrome, Crohn's disease, psoriasis and RA among individuals enrolled in the MADGC collection. We did not observe significant association with SLE and this may be due to lack of power as we had very low power (< 25%, for $\alpha = 0.05$) given our observed OR and the number of SLE cases in this family collection. This could also reflect differences between SLE that arises in the context of such multiplex (for diverse AIDs) families compared to SLE cases from individual AID collections. Of interest, the difference in ORs for the autoimmune thyroid diseases Hashimoto's thyroiditis (0.84) and Graves' disease (1.55) are quite striking, suggesting differential association with these distinct diseases. Our results do not provide any evidence of association of this variant with MS, also suggesting differential effects across autoimmune diseases as has been observed with the *PTPN22* missense SNP.

Given the fact that multiple variants within the *TNFAIP3* genomic region have been associated with AIDs in the current and previous studies, we examined haplotypes across this region. We identified nine haplotypes significantly or marginally associated with multiple AID diagnoses, and thus combinations of alleles appear to be important in conferring risk to multiple AIDs. Cases were found to be enriched for 5' and 3' UTR variants compared to controls, indicating that perhaps splicing or translational control are important for the function of this gene and its role in disease. In choosing to focus on coding exons in this study, we may have missed potentially important polymorphisms in non-coding regulatory regions.

In summary, overall our results support an important role for variants in the *TNFAIP3* gene region in the development of human autoimmune disease. However, the pattern of genetic association appears to be complex, with multiple variants contributing differentially across the spectrum of AID. Additional work will be required to confirm novel associations reported here and to identify polymorphisms in non-coding regions that may contribute to risk of autoimmunity.

Materials and Methods

DNA Collections

We sequenced 123 individuals from the MADGC collection who were affected by two or more individual AIDs. To be eligible for enrollment in the MADGC collection, families had to have at least two members with confirmed diagnoses of at least one of nine "core" AIDs. Details of MADGC recruitment and enrollment procedures have been described previously[24]. The four most common AID combinations among the 123 individuals sequenced were the presence of Hashimoto's thyroiditis with one of the following: RA

(N=19), SLE (N=11), MS (N=15), or type I diabetes (TID; N=12). The mean AID count was 2.2 with a maximum of six. Numbers of affected individuals per disease are listed in Table 6 for the nine core diseases plus an extra category termed “other AIDs” that includes other confirmed diagnoses outside the core nine. Hashimoto’s thyroiditis and Grave’s disease were counted as a single core disease, autoimmune thyroid disease, but were analyzed separately in this study. Most subjects were Caucasian (N=108), 11 were Caucasian/Native American, one was Caucasian/Asian, and three were Hispanic. Eighteen families had multiple members sequenced (38 individuals) while the remaining 85 individuals had no relatives sequenced in this study. For association testing, one member of each family was randomly selected and the 91 unrelated Caucasian cases (of the 123 sequenced individuals) were compared to 397 healthy Caucasian controls who were enrolled at the University of California San Francisco and included individuals from the Study Of Pharmacogenetics in Ethnic populations (SOPHIE) collection (N=262). All subjects gave informed consent in accordance with the IRB at their respective institutions.

Sequencing

To sequence all protein coding bases, eight sequencing reactions were performed for each DNA sample. Four sets of PCR primers were from SeattleSNPs (http://pga.gs.washington.edu/) while the other four were designed using Primer3[29]. Detailed primer information can be found in Table 7. Primer sets were checked through ePCR on the UCSC genome browser to ensure one unique genomic hit and were also inspected for a lack of known SNPs with dbSNP.

PCR was performed with 8ng DNA, 0.4μM each forward and reverse primer, 1× buffer, 4mM dNTPs, and 0.3U Qiagen (Valencia, CA USA) HotStar Taq in a 10μL reaction. PCR was cleaned up by incubation with 1× SAP PCR Clean-Up Reagent (PerkinElmer Life Sciences, Inc. Waltham, MA USA) at 37°C for one hour. Sequencing reactions contained 2.5μL of clean PCR product, 0.375μM primer and 8.3% Applied Biosystems (ABI; Foster City, CA USA) BigDye Terminator v3.1 in a 12μL reaction. Excess dye terminator removal was performed with genCLEAN (Genetix; New Milton, Hampshire, United Kingdom) plates following manufacturer’s instructions before sequencing on an ABI 3730×L DNA Analyzer. Sequencing was performed in one direction, except for regions with insertion-deletion polymorphisms and novel polymorphisms which were confirmed by sequencing the other strand.

Genotyping

Genotyping of indel Novel_4 and SNP rs2230926 was conducted in the entire MADGC collection (Caucasians; N=1778), including 1008 affected participants and 770 unaffected family controls. For rs2230926, additional healthy Caucasian controls were from the SOPHIE collection (N=257) and from a study of MS, which included northern European derived individuals with no family history of MS (N=486). Three AIMs were genotyped to assess population stratification. Genotyping was performed with a predesigned ABI TaqMan assay for SNP rs2230926 and for the three AIMs (rs1042712, rs7696175, rs9378805) and a custom TaqMan design for Novel_4 following the manufacturer’s protocol. We used 2× PCR Universal Master Mix and 4.5ng DNA in a 5μL reaction. Duplicates and no template controls were checked for quality control purposes.

Statistical methods

Sequencing traces were analyzed with Sequencher (Gene Codes; Ann Arbor, MI USA). HWE p-values were calculated in Haploview[30] to assess sequencing quality and a p-value of 0.001 was used as the significance threshold for exclusion. Individual polymorphism tests for association between sequenced cases and controls were conducted in Plink[31]. We used

Fisher's exact test and also conducted adaptive permutation tests by swapping case-control status to calculate empirical p-values for each variant. In order to mitigate the potential for false positives due to population stratification, we restricted the analysis to Caucasian samples, and we trimmed the panel to unrelated individuals, which reduced the number of cases from 123 to 91.

A single haplotype block was defined using the spine of LD definition in Haploview. Haplotype tests for association were conducted in Plink for 24-variant combinations for all frequencies and also restricted to those with a frequency greater than one percent. The 24 variants were polymorphic in cases or controls when the cases were restricted to unrelated Caucasians.

A weighted sum statistic was calculated to test for association with disease for a group of variants, each of which might have an independent influence on a genetically heterogeneous trait. This approach of combining independent signals has been shown to be significantly more powerful than variant-by-variant approaches. Also, mechanistic interpretations can be more readily made when the variants are grouped based upon the functional elements in which they reside. We used a custom script according to the method of Madsen et al.[32]. We checked for differences in common (>2% MAF), rare, exonic, intronic, non-synonymous, synonymous, and UTR variants between cases and controls.

Association testing for novel_4 and rs2230926 in the larger collection was conducted using PedGenie[33] in order to account for the relatedness among the MADGC collection cases and controls. AIM association testing was performed in Plink and HWE p-values were checked in Haploview with criteria as above. Since many AIDs are represented in the MADGC collection, only the most common were individually tested for association, with Sjögren's syndrome being the least frequent.

Acknowledgments

The authors would like to thank all of the participants in the studies featured in this manuscript. Jorge R. Oksenberg provided control DNA samples and Stacy J. Caillier performed genotyping of these samples for rs2230926. W.L. was supported by a grant from the Dermatology Foundation. The authors also wish to thank Drs. Peter Gregersen and Timothy Behrens for their important contributions to the MADGC study (supported by N01 AI95386). Dr. Criswell was also supported by K24 AR02175 and a Kirkland Scholar Award.

References

1. Helmick CG, Felson DT, Lawrence RC, Gabriel S, Hirsch R, Kwoh CK, et al. Estimates of the prevalence of arthritis and other rheumatic conditions in the United States. Part I. Arthritis and rheumatism. 2008; 58(1):15–25. [PubMed: 18163481]
2. Barcellos LF, Kamdar BB, Ramsay PP, DeLoa C, Lincoln RR, Caillier S, et al. Clustering of autoimmune diseases in families with a high-risk for multiple sclerosis: a descriptive study. Lancet neurology. 2006; 5(11):924–931. [PubMed: 17052659]
3. Plenge RM, Cotsapas C, Davies L, Price AL, de Bakker PI, Maller J, et al. Two independent alleles at 6q23 associated with risk of rheumatoid arthritis. Nature genetics. 2007; 39(12):1477–1482. [PubMed: 17982456]
4. Thomson W, Barton A, Ke X, Eyre S, Hinks A, Bowes J, et al. Rheumatoid arthritis association at 6q23. Nature genetics. 2007; 39(12):1431–1433. [PubMed: 17982455]
5. Dieguez-Gonzalez R, Calaza M, Perez-Pampin E, Balsa A, Blanco FJ, Canete JD, et al. Analysis of TNFAIP3, a feedback inhibitor of nuclear factor-kappaB and the neighbor intergenic 6q23 region in rheumatoid arthritis susceptibility. Arthritis research & therapy. 2009; 11(2):R42. [PubMed: 19292917]

6. Musone SL, Taylor KE, Lu TT, Nititham J, Ferreira RC, Ortmann W, et al. Multiple polymorphisms in the TNFAIP3 region are independently associated with systemic lupus erythematosus. Nature genetics. 2008; 40(9):1062–1064. [PubMed: 19165919]
7. Graham RR, Cotsapas C, Davies L, Hackett R, Lessard CJ, Leon JM, et al. Genetic variants near TNFAIP3 on 6q23 are associated with systemic lupus erythematosus. Nature genetics. 2008; 40(9): 1059–1061. [PubMed: 19165918]
8. Han JW, Zheng HF, Cui Y, Sun LD, Ye DQ, Hu Z, et al. Genome-wide association study in a Chinese Han population identifies nine new susceptibility loci for systemic lupus erythematosus. Nature genetics. 2009; 41(11):1234–1237. [PubMed: 19838193]
9. Bates JS, Lessard CJ, Leon JM, Nguyen T, Battiest LJ, Rodgers J, et al. Meta-analysis and imputation identifies a 109 kb risk haplotype spanning TNFAIP3 associated with lupus nephritis and hematologic manifestations. Genes and immunity. 2009; 10(5):470–477. [PubMed: 19387456]
10. Cai LQ, Wang ZX, Lu WS, Han JW, Sun LD, Du WH, et al. A single-nucleotide polymorphism of the TNFAIP3 gene is associated with systemic lupus erythematosus in Chinese Han population. Molecular biology reports. 2010; 37(1):389–394. [PubMed: 19774492]
11. Shimane K, Kochi Y, Horita T, Ikari K, Amano H, Hirakata M, et al. The association of a nonsynonymous single-nucleotide polymorphism in TNFAIP3 with systemic lupus erythematosus and rheumatoid arthritis in the Japanese population. Arthritis and rheumatism. 2010; 62(2):574–579. [PubMed: 20112363]
12. Yang W, Shen N, Ye DQ, Liu Q, Zhang Y, Qian XX, et al. Genome-wide association study in Asian populations identifies variants in ETS1 and WDFY4 associated with systemic lupus erythematosus. PLoS genetics. 2010; 6(2):e1000841. [PubMed: 20169177]
13. Nair RP, Duffin KC, Helms C, Ding J, Stuart PE, Goldgar D, et al. Genome-wide scan reveals association of psoriasis with IL-23 and NF-kappaB pathways. Nature genetics. 2009; 41(2):199–204. [PubMed: 19169254]
14. Trynka G, Zhernakova A, Romanos J, Franke L, Hunt K, Turner G, et al. Coeliac disease associated risk variants in TNFAIP3 and REL implicate altered NF-{kappa}B signalling. Gut. 2009
15. Coenen MJ, Trynka G, Heskamp S, Franke B, van Diemen CC, Smolonska J, et al. Common and different genetic background for rheumatoid arthritis and coeliac disease. Human molecular genetics. 2009; 18(21):4195–4203. [PubMed: 19648290]
16. Fung EY, Smyth DJ, Howson JM, Cooper JD, Walker NM, Stevens H, et al. Analysis of 17 autoimmune disease-associated variants in type 1 diabetes identifies 6q23/TNFAIP3 as a susceptibility locus. Genes and immunity. 2009; 10(2):188–191. [PubMed: 19110536]
17. Wang K, Baldassano R, Zhang H, Qu HQ, Imielinski M, Kugathasan S, et al. Comparative genetic analysis of inflammatory bowel disease and type 1 diabetes implicates multiple loci with opposite effects. Human molecular genetics.
18. Consortium TWTCC. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature. 2007; 447(7145):661–678. [PubMed: 17554300]
19. Prahalad S, Hansen S, Whiting A, Guthery SL, Clifford B, McNally B, et al. Variants in TNFAIP3, STAT4, and C12orf30 loci associated with multiple autoimmune diseases are also associated with juvenile idiopathic arthritis. Arthritis and rheumatism. 2009; 60(7):2124–2130. [PubMed: 19565500]
20. Dixit VM, Green S, Sarma V, Holzman LB, Wolf FW, O'Rourke K, et al. Tumor necrosis factor-alpha induction of novel gene products in human endothelial cells including a macrophage-specific chemotaxin. The Journal of biological chemistry. 1990; 265(5):2973–2978. [PubMed: 2406243]
21. Lee EG, Boone DL, Chai S, Libby SL, Chien M, Lodolce JP, et al. Failure to regulate TNF-induced NF-kappaB and cell death responses in A20-deficient mice. Science. 2000; 289(5488): 2350–2354. [PubMed: 11009421]
22. Boone DL, Turer EE, Lee EG, Ahmad RC, Wheeler MT, Tsui C, et al. The ubiquitin-modifying enzyme A20 is required for termination of Toll-like receptor responses. Nat Immunol. 2004; 5(10):1052–1060. [PubMed: 15334086]

23. Hitotsumatsu O, Ahmad RC, Tavares R, Wang M, Philpott D, Turer EE, et al. The ubiquitin editing enzyme A20 restricts NOD2 triggered signals. Immunity. 2008; 28:381–390. [PubMed: 18342009]
24. Criswell LA, Pfeiffer KA, Lum RF, Gonzales B, Novitzke J, Kern M, et al. Analysis of families in the multiple autoimmune disease genetics consortium (MADGC) collection: the PTPN22 620W allele associates with multiple autoimmune phenotypes. American journal of human genetics. 2005; 76(4):561–571. [PubMed: 15719322]
25. Wertz IE, O'Rourke KM, Zhou H, Eby M, Aravind L, Seshagiri S, et al. De-ubiquitination and ubiquitin ligase domains of A20 downregulate NF-kappaB signalling. Nature. 2004; 430(7000): 694–699. [PubMed: 15258597]
26. Baranzini SE, Wang J, Gibson RA, Galwey N, Naegelin Y, Barkhof F, et al. Genome-wide association analysis of susceptibility and clinical phenotype in multiple sclerosis. Human molecular genetics. 2009; 18(4):767–778. [PubMed: 19010793]
27. Hesselson SE, Matsson P, Shima JE, Fukushima H, Yee SW, Kobayashi Y, et al. Genetic variation in the proximal promoter of ABC and SLC superfamilies: liver and kidney specific expression and promoter activity predict variation. PloS one. 2009; 4(9):e6942. [PubMed: 19742321]
28. Cui J, Saevarsdottir S, Thomson B, Padyukov L, van der Helm-van Mil AH, Nititham J, et al. Rheumatoid arthritis risk allele PTPRC is also associated with response to anti-tumor necrosis factor alpha therapy. Arthritis and rheumatism. 62(7):1849–1861. [PubMed: 20309874]
29. Rozen S, Skaletsky H. Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol. 2000; 132:365–386. [PubMed: 10547847]
30. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics (Oxford, England). 2005; 21(2):263–265.
31. 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. American journal of human genetics. 2007; 81(3):559–575. [PubMed: 17701901]
32. Madsen BE, Browning SR. A groupwise association test for rare mutations using a weighted sum statistic. PLoS genetics. 2009; 5(2):e1000384. [PubMed: 19214210]
33. Curtin K, Wong J, Allen-Brady K, Camp NJ. Meta-genetic association of rheumatoid arthritis and PTPN22 using PedGenie 2.1. BMC proceedings. 2007; 1 Suppl 1:S12. [PubMed: 18466461]

Table 1

Polymorphism Discovery Summary for Cases and Controls.

SNPID	SNP Coordinate	Assayed In Controls	Assayed In Cases	SNP property	Alleles	MAF
rs5029933	138192062	N	Y	Intron 1	A/G	0.049
rs79608867	138192270	N	Y	Intron 1	G/C	0.008
rs3214646*	138192325	N	Y	Intron 1	*T*/-	0.500
novel_8	138192351	Y	Y	Exon 2, 5' UTR	T/G	0.003
novel_2	138192601	Y	Y	Exon 2, Ser79Arg	G/C	0.001
rs5029938	138195633	N	Y	Intron 2	C/T	0.049
rs643177	138195693	N	Y	Intron 2	C/T	0.248
rs5029939	138195723	N	Y	Intron 2	C/G	0.041
novel_3	138195726	N	Y	Intron 2	A/C	0.004
rs5029940	138195964-6	Y	Y	Intron 2 (−15 to −18 from Ex.	3) -/CCT	0.352
novel_9	138195991	Y	Y	Exon 3, Asn102Ser	A/G	0.001
rs2230926	138196066	Y	Y	Exon 3, Phe127Cys	T/G	0.029
novel_10	138196156	Y	Y	Exon 3, Leu157Pro	T/C	0.001
rs5029947	138196817	Y	Y	Intron 3 (-8bp from Ex. 4)	C/G	0.004
rs5029948	138197329	Y	Y	Intron 5	C/T	0.052
rs661561	138197331	Y	Y	Intron 5	C/A	0.342
rs5029964	138197341	Y	Y	Intron 5	A/G	0.001
rs582757	138197824	Y	Y	Intron 5	T/C	0.268
novel_4	138197889	Y	Y	Intron 5	C/-	0.010
novel_11	138199316	Y	N	Intron 6	A/G	0.001
rs610604	138199417	Y	Y	Intron 6	T/G	0.323
novel_12	138199898	Y	Y	Exon 7, Arg439Gln	G/A	0.001
novel_13	138200220	Y	Y	Exon 7, Glu546Glu	G/A	0.001
rs5029953	138200760	Y	Y	Intron 7	G/A	0.009
rs5029965	138200852	Y	Y	Intron 7	G/A	0.011
novel_5	138201240	Y	Y	Exon 8, Thr647Pro	A/C	0.004

SNPID	SNP Coordinate	Assayed In Controls	Assayed In Cases	SNP property	Alleles	MAF
novel_14	138202130	Y	Y	Intron 8	G/A	0.006
novel_15	138202223	Y	Y	Exon 9, Pro714Ser	C/T	0.001
novel_6**	138202258	Y	Y	Exon 9, Leu725Leu	G/A	0.004
novel_16	138202314	Y	Y	Exon 9, Gly744Asp	G/A	0.001
rs5029956	138202378	Y	Y	Exon 9, Pro765Pro	C/T	0.003
novel_7	138202557	Y	Y	Exon 9, 3' UTR	G/T	0.013
novel_17	138202630	Y	Y	Exon 9, 3' UTR	G/A	0.001

Coordinates obtained from hg19. Flanking sequences are on the positive strand of the genome and SNP alleles are shown as Major/Minor. Assayed in Controls and Assayed in Cases indicate if sequence data was obtained at that base in the control and case groups, respectively.

MAF- Minor Allele Frequency.

* rs3214646 removed for violation of Hardy-Weinberg Equilibrium ($P=3.7009\times10^{-36}$).

** Novel SNP 6 is located within zinc-finger motif 6.

Table 2

Association Testing of Sequenced Variants.

N	SNP	A1	F_A	F_U	A2	OR	L95	U95	Fisher P	EMP1	NP
1	novel_8	G	0	0.004	T	0	0	NA	1	1	6
2	novel_2	C	0.005	0	G	NA	NA	NA	0.194	0.182	98
3	rs5029940	C1CT	0.341	0.354	-	0.943	0.671	1.327	0.795	1	6
4	novel_9	G	0	0.001	A	0	0	NA	1	0.857	6
5	rs2230926	G	0.027	0.024	T	1.142	0.419	3.119	0.791	0.778	8
6	novel_10	C	0	0.001	T	0	0	NA	1	0.857	6
7	rs5029947	G	0	0.001	C	0	0	NA	1	0.326	45
8	rs5029948	T	0.055	0.054	C	1.026	0.503	2.093	1	1	6
9	rs661561	A	0.341	0.345	C	0.983	0.699	1.384	1	1	6
	rs5029964	0	0	0	A	---	---	---	1	1	---
10	rs582757	C	0.258	0.274	T	0.923	0.638	1.334	0.711	1	6
11	novel_4	-	0.027	0.004	C	7.053	1.67	29.79	0.009	0.005	4700
12	rs610604	G	0.313	0.324	T	0.951	0.671	1.348	0.860	1	6
13	novel_12	A	0	0.001	G	0	0	NA	1	0.326	45
14	novel_13	A	0	0.001	G	0	0	NA	1	0.246	68
15	rs5029953	A	0	0.008	G	0	0	NA	0.601	0.55	19
16	rs5029965	A	0.011	0.012	G	0.922	0.198	4.305	1	0.857	6
17	novel_5	C	0.005	0.004	A	1.387	0.143	13.41	0.579	0.625	15
18	novel_14	A	0	0.008	G	0	0	NA	0.603	0.727	10
19	novel_15	T	0	0.001	C	0	0	NA	1	0.182	98
20	novel_6	A	0.011	0.001	G	8.367	0.755	92.78	0.098	0.142	133
21	novel_16	A	0	0.001	G	0	0	NA	1	0.214	83
22	rs5029956	T	0	0.004	C	0	0	NA	1	0.778	8
23	novel_7	T	0	0.013	G	0	0	NA	0.224	0.262	60
24	novel_17	A	0	0.001	G	0	0	NA	1	0.212	84

91 unrelated, multiply affected, Caucasian MADGC cases vs. 397 healthy controls. ORs for variants with F_A or F_U of 0 cannot be calculated. SNPs are ordered by genomic position. rs5029964 was not polymorphic when restricted to these samples. NP determined by Plink adaptive permutation procedure. N - SNP number which can be applied to order in Table 3 haplotypes. A1 - Allele 1; F_A -

Frequency in cases (affected); F_U - Frequency in controls (unaffected); A2 - Allele 2; OR - Odds ratio; L95 - Lower 95% confidence interval; U95 - Upper 95% confidence interval; EMP1 - Empirical P-value; NP - Number of permutations; NA - not available (i.e., due to zero cells).

Table 3

Haplotype Testing Results between Sequenced Cases and Controls.

Haplotype	Hap Freq	F_A	F_U	P	Minor Alleles Present
TGAA**T**TCC**AC**CTGGGGAGCGGCGG	0.046	0.0039	0.057	0.0024	rs661561, rs582757
TG**C**A**T**TCCCTC**G**GGGGAGCGGCGG	0.058	0.0050	0.070	0.0007	rs5029940, rs610604
TGAA**T**TCCCTCTGGGGAGCGGCGG	0.53	0.61	0.51	0.014	
TG**C**A**T**TCC**AC**C**G**GGGGAGCGGCGG	0.19	0.25	0.18	0.048	rs5029940, rs661561, rs582757, rs610604
TGAA**T**TCCCTCTGGGGAGC**A**GCGG	0.0032	0.01	0.0013	0.037	novel_6
TGAA**T**TCCCT**A**TGGGGAGCGGCGG	0.0038	0.014	0.0014	0.012	novel_4
TG**C**A**T**TCC**ACAG**GGGGAGCGGCGG	0.0016	0.0084	0	0.011	rs5029940, rs661561, rs582757, novel_4, rs610604
T**C**AA**T**TCCCTCTGGGGAGCGGCGG	0.0011	0.0055	0	0.041	novel_2
TGAA**T**TCC**A**TCTGGGGAGCGGCGG	0.016	0	0.020	0.053	rs661561

Significant or borderline significant haplotypes listed for 24 polymorphisms sequenced in cases and controls listed in order as in Table 2 with rs5029940 coded as A/C and novel_4 coded as C/A. Minor alleles present in each haplotype are in bold red color and SNP rs2230926 is in bold black text (all haplotypes have the major, protective allele). 91 unrelated, multiply affected, Caucasian MADGC cases vs. 397 healthy controls. Hap Freq – haplotype frequency; F_A – frequency in cases (affected); F_U – frequency in controls (unaffected); P – unadjusted p-value.

Table 4

MADGC Collection Genotyping and Allelic Association of novel_4 Polymorphism.

Disease	N	N Deletions	MAF	Odds Ratio (95%CI)	P
All Cases	1008	27	0.013	1.04 (0.48 – 2.25)	0.876
SLE	116	6	0.026	2.19 (0.76 – 6.34)	0.116
RA	146	7	0.024	**2.48 (0.93 – 6.64)**	**0.041**
Hashimoto's	242	11	0.023	1.72 (0.62 – 4.81)	0.265
Multiply affected	142	6	0.021	1.67 (0.61 – 4.57)	0.337
JIA	30	1	0.017	1.61 (0.32 – 8)	0.844
TID	81	2	0.012	1.2 (0.28 – 5.08)	0.842
Crohn's Disease*	50	1	0.01	0.96 (0 – infinity)	0.963
MS	192	3	0.008	0.76 (0.13 – 4.33)	0.712
Graves	80	1	0.006	0.65 (0 - infinity)	0.713
IBD (IC, UC, CD)	88	1	0.006	0.57 (0 - infinity)	0.603
Psoriasis	79	0	0	NA	0.232
Ulcerative Colitis*	34	0	0	NA	0.407
Sjögren's	17	0	0	NA	0.604
Controls	770	16	0.01		

Individual disease results shown in descending order by odds ratio strength. P-values < 0.05 and corresponding odds ratios are in bold. N Deletions is the number of minor alleles for cases used in calculating chi-square tests.

Diseases marked * are subtypes of inflammatory bowel disease (IBD).

MAF - minor allele frequency in cases; OR - odds ratio; CI - confidence interval; SLE - systemic lupus erythematosus; RA - rheumatoid arthritis; JIA -juvenile idiopathic arthritis; TID - type I diabetes; MS - multiple sclerosis; IBD - inflammatory bowel disease; IC - idiopathic colitis; UC - ulcerative colitis; CD - Crohn's disease. NA - not available (i.e., due to zero cells).

Table 5

MADGC Collection Genotyping and Allelic Association of rs2230926.

Disease	N	G Alleles	MAF	Odds Ratio (95%CI)	P
All Cases	1008	72	0.036	1.37(0.95–1.97)	0.083
Sjögren's	18	3	0.083	3.38(0.91–12.58)	0.038
Crohn's Disease*	50	6	0.060	2.25 (0.92 – 5.5)	0.041
Psoriasis	78	9	0.058	2.17(0.96–4.88)	0.037
RA	148	15	0.051	1.9(1.05–3.45)	0.025
Graves	80	8	0.050	1.55(0.72–3.3)	0.323
SLE	117	9	0.038	1.31 (0.66–2.63)	0.467
Multiply affected	144	11	0.038	1.35(0.69–2.64)	0.422
IBD-IC, UC, CD	88	6	0.034	1.32(0.6–2.91)	0.536
TID	79	5	0.032	1.16(0.54–2.5)	0.770
Hashimoto's	244	12	0.025	0.84 (0.36 – 1.95)	0.664
MS	191	7	0.018	0.66(0.27 – 1.65)	0.290
Ulcerative Colitis*	34	0	0	NA	0.177
JIA	30	0	0	NA	0.194
Controls	1513	86	0.028		

Individual disease results shown in descending order by odds ratio strength. P-values < 0.05 and corresponding odds ratios are in bold. G Alleles is the number of minor alleles for cases used in calculating chi-square tests.

Diseases marked * are subtypes of inflammatory bowel disease (IBD).

MAF - minor allele frequency in cases; OR - odds ratio; CI - confidence interval; RA - rheumatoid arthritis; SLE - systemic lupus erythematosus; IC -idiopathic colitis; UC - ulcerative colitis; CD - Crohn's disease; TID - type I diabetes; MS - multiple sclerosis; JIA - juvenile idiopathic arthritis. NA - not available (i.e., due to zero cells).

Table 6

Autoimmune Disease Distribution in 123 Sequenced MADGC Participants.

Affection Status	Hash	RA	SLE	MS	TID	Graves'	Psoriasis	Sjög	IBD	JIA	Other AIDs
Affected	74	43	32	28	20	16	13	11	10	7	12
Unaffected	47	80	89	95	103	106	110	105	112	116	104
Reported, Unconfirmed	2	0	2	0	0	1	0	7	1	0	7
Fraction Affected	0.612	0.350	0.264	0.228	0.163	0.131	0.106	0.095	0.082	0.057	0.103

Autoimmune diseases present in participants listed in order of decreasing frequency from left to right. Fraction affected only includes confirmed cases. All analyses in this study exclude reported, but unconfirmed diagnoses.

Hash - Hashimoto's; RA - rheumatoid arthritis; SLE - systemic lupus erythematosus; MS - multiple sclerosis; TID - type I diabetes; Sjög - Sjögren's syndrome; IBD - inflammatory bowel disease; JIA - juvenile idiopathic arthritis; Other AIDs - other autoimmune diseases, which include autoimmune Addison's Disease, autoimmune hepatitis, CREST syndrome (limited scleroderma), idiopathic thromobocytopenic purpura, mixed or undifferentiated connective tissue disease, myasthenia gravis, pernicious or hemolytic anemia, polymyositis or dermatomyositis, scleroderma and vitiligo.

Table 7

Sequencing Primer Pairs.

Exon	Amp. Coordinates	Forward Primer, 5' to 3'	Reverse Primer, 5' to 3'	PD
2	138191752–138192730	GGGGCTAAAGAGGAAACACC	CTTCATGAATGGGGATCCAG	Primer3
3	138195539–138196331	CCCTGTGTGCTCCTCCTTAG	CCACTGGAGGTTTCTGGTGT	Primer3
4 & 5	138196610–138197370	TCCCCAACTTTTGAGTTTGC	AAGCAAAAAGGAAAACCCTGA	Primer3
6	138197565–138198544	CAAGTAAACGCCTGTCAGGTTAG	ACCATGCACAAGACTCTGAATTT	SeattleSNPs
7	138199065–138200179	CGTCTTAGTTACTCATGGCTGCT	TAAATGTCCTGGTAACATCCTGG	SeattleSNPs
7	138199989–138200922	GTTCAGTGAGACCACTGCCAT	TGAGAGATTTCCAAACCACATCT	SeattleSNPs
8	138200689–138201772	GCAGCTCCTAATATCACATTCCA	TCTGTCTGTTCGCTCCTTATGAT	SeattleSNPs
9	138202044–138202816	CCTTGCTCAGGCAGGTAAAG	AGCCAAGACGATGAAGCAGT	Primer3

Amp. Coordinates – amplicon positions on chromosome 6, UCSC hg19; PD – primer design software/source.