

SCIENTIFIC REPORTS



OPEN

Network-assisted analysis of primary Sjögren's syndrome GWAS data in Han Chinese

Kechi Fang, Kunlin Zhang & Jing Wang

Received: 27 July 2015
Accepted: 05 November 2015
Published: 21 December 2015

Primary Sjögren's syndrome (pSS) is a complex autoimmune disorder. So far, genetic research in pSS has lagged far behind and the underlying biological mechanism is unclear. Further exploring existing genome-wide association study (GWAS) data is urgently expected to uncover disease-related gene combination patterns. Herein, we conducted a network-based analysis by integrating pSS GWAS in Han Chinese with a protein-protein interactions network to identify pSS candidate genes. After module detection and evaluation, 8 dense modules covering 40 genes were obtained for further functional annotation. Additional 31 MHC genes with significant gene-level P-values (sigMHC-gene) were also remained. The combined module genes and sigMHC-genes, a total of 71 genes, were denoted as pSS candidate genes. Of these pSS candidates, 14 genes had been reported to be associated with any of pSS, RA, and SLE, including *STAT4*, *GTF2I*, *HLA-DPB1*, *HLA-DRB1*, *PTTG1*, *HLA-DQB1*, *MBL2*, *TAP2*, *CFLAR*, *NFKBIE*, *HLA-DRA*, *APOM*, *HLA-DQA2* and *NOTCH4*. This is the first report of the network-assisted analysis for pSS GWAS data to explore combined gene patterns associated with pSS. Our study suggests that network-assisted analysis is a useful approach to gaining further insights into the biology of associated genes and providing important clues for future research into pSS etiology.

Sjögren's syndrome (SS) is a chronic autoimmune disease characterized by exocrine gland dysfunction, specifically the salivary and lacrimal glands, resulting in oral and ocular dryness¹. The disease may occur alone as primary Sjögren's syndrome (pSS) or in connection with other systemic rheumatic conditions as secondary Sjögren's syndrome (sSS)¹. In China, the prevalence of pSS is estimated to be 0.77%². Although pSS is one of the most common autoimmune diseases, scientific and medical research in pSS has lagged far behind and the pathogenic mechanisms of pSS are not yet fully known³. An interaction between genetic predisposition and environmental factors is believed to cause pSS⁴.

In recent years, genome-wide association studies (GWAS) have become a promising approach to unravelling common variants associated with human complex disorders including pSS^{5,6}. The pSS GWASs have uncovered a few risk loci conferring susceptibility to pSS^{5,6}. In spite of these successes, as with other complex diseases, GWAS analysis of pSS is limited by the use of a genome-wide significance cutoff SNP P-value of 5×10^{-8} needed for multiple testing correction⁷. Except the strongest genetic markers, many modest loci that each contributes in small part to the genetics of the disease may be ignored under this stringent strategy⁸. The reported loci by GWAS account for only a small proportion of pSS genetic risk. The underlying genes remain largely unknown, especially the interactions among these susceptibility genes are elusive. Moreover, how to translate the GWAS observations into any biological function is still a challenge for pSS. Hence there is an urgent need to apply new method that can integrate GWAS data with high-throughput datasets to examine the combined effect of multiple variants for pSS.

As human protein interaction data become more and more abundant, protein-protein interaction (PPI) networks are increasingly serving as tools to discover the molecular basis of diseases. PPI network provides a convenient framework for exploring relationships of disease-related genes and can be integrated with other various biological data. An integrative analysis of GWAS data with PPI network opens a new avenue for promoting the identification of true genetic signals and has been widely applied in many diseases⁹⁻¹¹. The rationale behind network-assisted analysis is "guilt by association"¹², *i.e.* different causal genes for the same phenotypes often interact, either directly or via common interaction partners.

Along these lines, the present study applied a network-assisted method by integrating pSS GWAS data in Han Chinese with human PPI network to investigate whether a set of genes, whose protein products closely interact

Key Laboratory of Mental Health, Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, China. Correspondence and requests for materials should be addressed to J.W. (email: wangjing@psych.ac.cn)

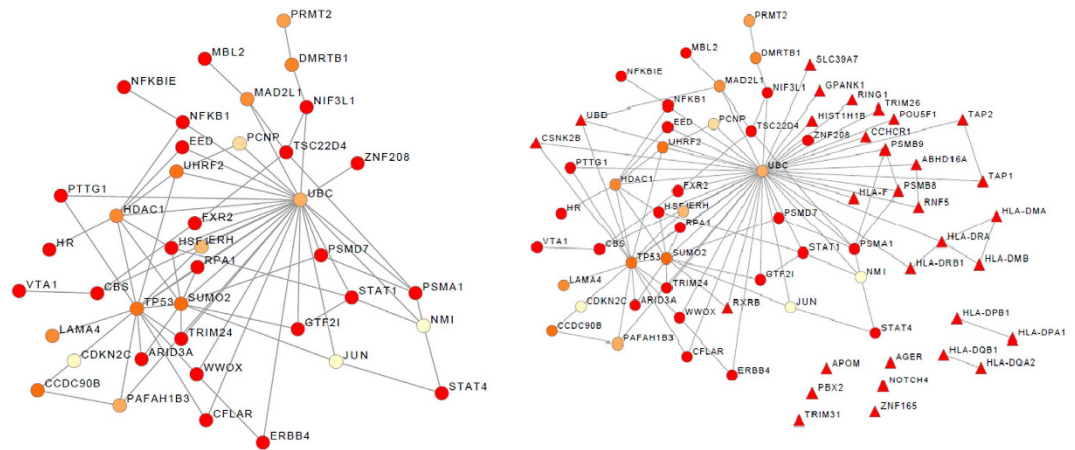


Figure 1. (a) The subnetwork formed by identified module genes; (b) the subnetwork formed by sigMHC-genes and identified module genes. The triangle-shaped nodes represent sigMHC-genes and circular-shaped nodes represent DMS-identified module genes. The color of the node was proportioned with the gene *P*-value. The most significant gene *P*-value was red color and the most non-significant gene *P*-value was yellow color.

with each other might collectively contribute to pSS risk. We highlighted 71 pSS candidate genes including 40 module genes identified by dense module searching (DMS) algorithm and additional 31 MHC genes with small gene-level *P*-values (sigMHC-genes). Of these candidates, 14 genes had been reported to be associated with any of pSS, RA, and SLE. The results also obtained gene-gene interactions among these candidates. Our network-assisted analysis of pSS GWAS would facilitate the understanding of genetic mechanism of pSS.

Results

Identification of sigMHC-genes and modules enriched for pSS-associated genes. To perform network-assisted analysis, pSS GWAS data in Han Chinese was applied and gene-level *P*-values were computed with VEGAS (see Methods). A total of 26,929 genes with *P*-values were obtained. Then, the gene *P*-values were integrated with a high confident PPI network (see Methods), resulting in a pSS specific node-weighted network of 9,203 proteins and 31,908 interactions. The involved interactions were listed in Supplementary Table S1.

Particularly, there were 31 genes located in MHC region and with gene *P*-values < 0.05 , defined as sigMHC-genes (see Methods). In order to reduce the influence of sigMHC-genes on module searching and primarily focus on genes outside MHC, the sigMHC-genes were not set as seed nodes to search modules.

Dense module searching was performed within the node-weighted pSS network to identify modules enriched for genes with significant pSS genetic signals. A total of 8,594 independent modules were preliminarily generated. After estimating the significance of module scores, 127 modules met the criterion of $P(Z_m) < 0.05$ were left for further estimating the topological properties. As a result, eight modules were finally remained as significant modules. The union of resultant modules was finally computed, resulting in a single connected subnetwork of 40 non-redundant genes and 70 interactions (Fig. 1a). Of these 40 module genes, 24 had nominally significant gene *P*-values (< 0.05). To further validate whether module genes were significantly physically interacted, DAPPLE was used¹³. The results showed that the direct PPI network of module genes had more significant edges than expected by chance (permutation *P*-value = 9.9×10^{-5}) (Supplementary Figure S1), suggesting that the interactions among the module genes were statistically significantly connected. The detailed information for the DMS-identified module genes and sigMHC-genes were listed in Supplementary Table S2.

Biological annotation for the identified module genes. To better understand the biological functions of the DMS-identified module genes, we conducted a Gene Ontology (GO) enrichment analysis by using DAVID. The enriched biological processes were related with negative regulation of protein metabolic process and proteasomal protein catabolic process. The enriched molecular functions were about transcription regulator activity and transcription factor activity. Further information of GO enrichment analysis of module genes was summarized in Supplementary Table S3.

We also computed the tissue specificity of module genes by using the Gene Enrichment Profiler. In the transcript expression heatmap (Supplementary Figure S2), approximately two-thirds of these genes were highly expressed in immune-related cell types (specifically, B, T and myeloid cells). In the transcript enrichment heatmap (Supplementary Figure S3), we found module genes were preferentially expressed in the immune cell types.

Module genes and sigMHC-genes as candidates for pSS. To explore whether exist interactions between the DMS-identified module genes and sigMHC-genes, we extracted these genes and according interactions from the node-weighted pSS network, resulting in a subnetwork as shown in Fig. 1b. Most of sigMHC-genes directly or indirectly connected with module genes except 6 singletons and two isolated PPI pairs (*HLA-DPB1* vs. *HLA-DPA1*, and *HLA-DQB1* vs. *HLA-DQA2*). The combined module genes and sigMHC-genes were defined as candidate genes (71 genes) for pSS. Of these candidates, four genes (*STAT4*, *GTF2I*, *HLA-DPB1*, and *HLA-DRB1*) had been reported their

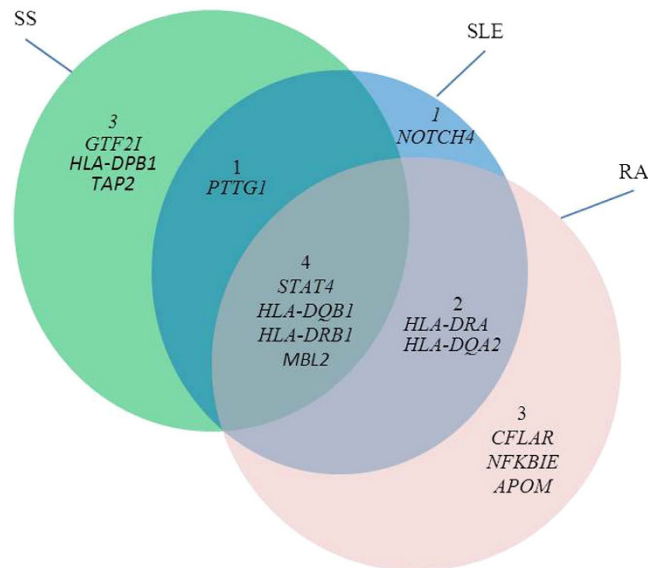


Figure 2. The overlaps of candidate pSS genes among other autoimmune diseases. The green, blue, and pink circles indicate the candidate pSS genes that have positive evidence to be associated with SS, SLE, and RA, respectively.

association with pSS in the original GWAS dataset⁶, and gene *PTTG1* (pituitary tumor-transforming 1) had been reported as suggestive association (rs2431098, allelic meta P-value = 2.28×10^{-7} ; rs2431697, allelic meta P-value = 3.76×10^{-6}) with SS in another GWAS study of SS in European descent⁵. In addition, three genes (*HLA-DQB1*¹⁴, *MBL2*¹⁵, and *TAP2*¹⁶) had been previously reported their association with SS/pSS by candidate gene studies.

Given the overlap of certain clinical and serologic features between pSS and other autoimmune diseases (AIDs), such as Rheumatoid arthritis (RA) and Systemic lupus erythematosus (SLE), it is reasonably assumed that pSS might share some genetic signatures with other AIDs¹⁷. Hence, we also investigated how many candidate pSS genes had been reported their susceptibility to either SLE or RA by searching GWAS studies collected in GWAS Catalog⁸. The overlaps of candidate pSS genes among SLE and RA was shown in Fig. 2. Another six genes with positive evidence associated with SLE or RA were obtained, including five genes (*HLA-DRA*¹⁸, *HLA-DQA2*¹⁹, *CFLAR*²⁰, *NFKBIE*^{20–22}, and *APOM*²³) reported to be associated with RA, and two genes (*HLA-DQA2*^{24,25} and *NOTCH4*²⁴) reported to be associated with SLE. Gene *HLA-DQA2* was overlapped between RA and SLE. These six SLE and RA susceptibility genes might be the shared genetic signatures between AIDs. In addition, four genes (*STAT4*^{4,6,20,22,24–27}, *HLA-DQB1*^{14,19,24,27–29}, *HLA-DRB1*^{6,20,26,27}, and *MBL2*^{15,30,31}) were overlapped among SS, SLE and RA, and gene *PTTG1* was overlapped between SS and SLE^{5,24}.

Discussion

Unlike the extensive GWAS experiment in other AIDs, such as RA and SLE, there have been only two GWAS studies in SS/pSS until now^{5,6}. Genetic studies of pSS have lagged behind. To further mining the existing genetic data, network-assisted analysis of pSS GWAS in Han Chinese was performed in order to explore the joint effects of multiple genetic association signals on pSS and discover additional candidate genes associated with pSS pathogenesis. First, all SNPs were first mapped into genes and gene-based association was performed by using VEGAS. Then, dense modules were dynamically searched in the context of the node-weighted pSS network via DMS, yielding thousands of functional modules. To avoid false positive results and the topology bias, a strict criterion was applied to select modules with significant genetic signals. After two stepwise significance tests (see Methods), 8 modules covering 40 genes were screened out. These 40 module genes had a high proportion of significant genes (60%) and preferentially expressed in immune-related cell types. The proteins encoded by the DMS-identified module genes were more closely interconnected than what would be expected by random cases, as suggested by DAPPLE analysis. In addition, there were 31 MHC genes with significant gene P-values, defined as sigMHC-genes. To avoid the results of module search focusing on MHC region, the sigMHC-genes were not set as “seed genes” (see methods) and directly remained as part of final results. By merging all module genes and sigMHC-genes, a total of 71 genes were involved and denoted as candidate pSS genes.

Of the 71 candidate pSS genes, eight genes had been previously reported their association with SS/pSS, as well as another six genes that had positive evidence associated with SLE or RA (Fig. 2). All the 14 reported genes had significant gene P-values (see Supplementary Table S2). Considering that pSS might share genetic signatures with SLE and RA to some extent, these six SLE or RA susceptibility genes were very likely associated with pSS. For example, one of reported RA susceptibility genes was *NFKBIE* (nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, epsilon)^{20–22}, which is a non-MHC gene with gene-level P-value of 0.00801. *NFKBIE*, also known as *I κ B ϵ* , is part of the I κ B family of proteins that regulates NF- κ B-dependent transcription by inhibiting DNA binding and localizing these factors to the cell cytoplasm³². It has been demonstrated that *I κ B α* (another I κ B

family of protein) promoter polymorphisms are associated with susceptibility to SS³³. In addition, NF- κ B plays an important role in inflammatory diseases and in the development of autoimmunity³⁴. Experimental studies have shown an activation of NF- κ B in pSS. These clues implied that *NFKBIE* might be also associated with pSS.

Due to disease genes execute their functions are not alone, the interactions among the candidate pSS genes might play an important role. It has been observed that causal genes for the same Mendelian disease often physically interact^{35,36}. There were two direct interactions among the reported genes, *i.e.* (*HLA-DRB1* vs. *HLA-DRA*), and (*HLA-DQB1* vs. *HLA-DQA2*). It was worth noting that most of 14 reported genes were closely connected by gene *UBC* (ubiquitin C) except for some sigMHC-genes. As the theory of “guilt by association”, it is possible that a few highly connected nodes (hub genes) bring together several disease-associated genes, even though the hubs themselves are not relevant³⁷. In this study, although the gene-level P-value of *UBC* was not significant (gene P-value = 0.463), it was involved by all the DMS-identified modules. As shown in Fig. 1a,b, *UBC* was centered both in the subnetwork formed by DMS-identified genes and in the subnetwork formed by the combined candidate genes. *UBC* directly or indirectly (via one or two nodes) connected with most of pSS candidate genes.

Of the rest of 57 genes, some genes with non-significant gene P-values might act as connectors to connect with other disease genes. For example, the gene P-value of *JUN* (jun proto-oncogene) was not significant (gene P-value = 0.918), but it interacted with *STAT4* (signal transducer and activator of transcription 4), which was the most significant gene in this study (gene P-value = 1E-07). In addition, *STAT4* was the only gene that had been confirmed its association with SS/pSS in both Han Chinese and European descent^{5,6}. In addition, other novel candidate pSS genes with significant gene P-values were also valuable, such as *STAT1* (signal transducer and activator of transcription 1, 91kDa). *STAT1* (gene P-value = 0.00267) directly interacted with *GTF2I*, which was a reported pSS associated gene in the Chinese cohort⁶. *STAT1* phosphorylation at serine 708 is a key event in the interferon signalling pathway³⁸. In many SS patient, interferon activation plays an important role in the immune attack and destruction of salivary and lacrimal glands at some stage in the course of the disease. These evidence suggested that *STAT1* was likely related with pSS.

The present study has some limitations that require consideration. First, only one pSS GWAS data in Han Chinese was available for this study. It would be more valuable if we could make a comparison between multiple GWAS datasets from different population. In our study, only gene *PTTG1* could be cross evaluated between two SS/pSS GWAS. *PTTG1* was only reported as suggestive association with SS in European descent⁵ and not found to be a risk factor in the Chinese cohort⁶. However, this gene was significant at gene-level (gene P-value = 0.00143) and identified by module search, and it was also reported to be associated with SLE²⁴. These lines of evidence implied that *PTTG1* might be associated with pSS in Han Chinese. Second, calculation of gene-level P-value is a key step in the network-assisted analysis of GWAS. VEGAS is comparable with other tools to compute gene P-values, however, it could only deal with autosomal SNPs. Due to women are nine times more likely than men to be affected with SS, it would be interesting to evaluate SNPs located on the sex chromosomes (X and Y).

In summary, this is the first use of network analysis of pSS GWAS data to further mine genetic signals at a molecular level instead of analyzing each of single locus (SNP). Complementary to the traditional GWAS analysis, it was more powerful that gene-level P-value was considered by calculating the combined effect of all SNPs within a gene and subsequently integrated with a pSS-specific PPI network to search for gene combination patterns contributed to pSS. Our findings included 40 non-MHC genes identified by DMS algorithm and 31MHC-region genes with significant gene-level P-values. These candidates and interactions among them were more likely to be associated with pSS.

Deciphering the mechanism of pSS pathogenesis is still challenging, although certain progress has been made, much remains to be understood. Deriving a pSS-specific PPI network and identification of dense module genes and sigMHC-genes, as described herein, offers new targets for further functional assessment for this chronic and complex condition.

Methods

The workflow of the network-assisted analysis for pSS GWAS data was shown in Supplementary Figure S4, and the sections below were labeled in correspondence with this figure.

pSS GWAS dataset. The pSS GWAS data is composed of samples of Han Chinese⁶. There are 642,832 SNPs in 597 pSS cases and 1,090 controls genotyped with the Affymetrix Axiom Genome-Wide CHB 1 Array Plate. Details of this pSS GWAS data and process of quality control are provided in ref.⁶. After quality control filtering, a total of 542 cases, 1,050 controls and 556,134 autosomal SNPs were remained for subsequent analysis.

Computing gene-level P-values. To perform network analysis and examine functional correlation between genes, gene-level P-value representing the significance of association with phenotype for each gene was needed to be considered. The gene-level P-value was calculated with VEGAS³⁹, which can incorporate information from all SNPs mapped to a gene and take into account the linkage disequilibrium (LD) patterns between SNPs for the specific samples (a custom set of individuals) or ethnic background (HapMap data). In VEGAS, all SNPs were mapped to human protein-coding genes according to positions on the UCSC Genome Browser. In this study, an off-line version of VEGAS was applied and improved in some aspects. First, gene position was updated from original hg18 to hg19 downloaded from Ensembl (GRCh37.P11)⁴⁰. Second, in order to capture SNPs in regulatory region and simultaneously avoid too many genes with overlapped SNPs, gene boundary was extended to 20kb upstream/downstream of the gene coordinates instead of 50kb by default. Third, when estimating the LD patterns, the pSS GWAS data was used.

Building a node-weighted pSS interactome. A consolidated human protein-protein interaction (PPI) network data was obtained from iRefIndex database (version 13.0)⁴¹, which collected nine interaction databases

and computed the union of data sets. Among this large network, many interactions were either predicted or supported by a single experiment. In order to reduce the rate of false positives, we included only those interactions supported by at least two publications for all subsequent analyses, resulting in a highly reliable network of 10,163 nodes (genes) and 36,680 interactions. Then, the nodes involved in this network was annotated with gene P-value as a node attribute, and extracted to derive a node-weighted pSS network.

sigMHC-genes. Currently, the reported susceptibility genes associated with pSS mainly focused on immune-related genes and the MHC region^{5,6,42,43}. To unravel more risk genes outside MHC region and avoid the complexity of the MHC region⁴⁴, we did not assign gene P-values for nodes in pSS network if the corresponding genes located in MHC region and their gene P-values < 0.05, named as significant MHC genes (sigMHC-genes). Since sigMHC-genes might interplay with other significant genes and play an important role in the disease-related biological functions, these genes were still left in the pSS network to maintain the integrity of the network.

Module detection. A dense module search (DMS) method implemented in dmGWAS was applied to search for modules that were enriched with significant P-value genes in the context of the node-weighted pSS network⁴⁵. DMS starts by transferring each gene P-value into a Z score (z_i) by using the inverse normal cumulative distribution function⁴⁶. For a module with k genes, the module score (Z_m) was computed by summing the z_i over all genes in the module, *i.e.* $Z_m = \sum z_i / \sqrt{k}$. The detailed process of module search can be found in this study⁴⁵. Briefly, for a given “seed gene”, module grows by adding the neighboring nodes that can generate the maximum increment of a module score (Z_m). Module growth will stop if the increment is not greater than $Z_m \times 0.1$. The process of module searching was conducted taking each node in the pSS interactome as the seed gene except for sigMHC-genes.

Module evaluation. For the proper capture of the connection between genetic association and network topology, two steps of tests were performed.

First, to assess the significance of the resultant modules, module P-values were calculated based on the module scores by empirically estimating the null distribution, which is assumed to be a normal distribution⁴⁷. Specifically, module scores were median-centered, and then the parameters of mean δ and standard deviation σ were estimated for the empirical null distribution using the R package *locfdr*. The standardized module scores Z_s were computed and converted to P-values by using the normal cumulative density function. The modules with P-values $P(Z_m) < 0.05$ were selected for the further estimation.

Second, to avoid the bias that nodes with many interactors in the PPI are more probably to be chose by DMS, the topology of resultant modules were evaluated as suggested in⁴⁸. All nodes in the network were divided into four groups according to nodes degree, *i.e.* $0-2^2$, 2^2-2^4 , 2^4-2^6 , and $>2^6$. For a given module with k genes, 10,000 modules with the same number of genes were generated by considering which group each gene located in and then randomly picking one gene from the corresponding group (*i.e.* structurally equivalent random networks). An empirical P-value was calculated by $P_{topo} = P\{Z_m(\pi) \geq Z_m\} = \frac{\# \text{ of resamples } \{Z_m(\pi) \geq Z_m\} + 1}{\text{total } \# \text{ resamples} + 1}$, where $Z_m(\pi)$ is the score of the random module for the π^{th} resample. The modules with P-values $P_{topo} < 0.05$ were selected as the final results.

Assessment of the significance of connectivity between module genes. To evaluate whether module genes were densely connected via PPI network, a permutation test was performed to assess the significance of connectivity between module genes by using DAPPLE (Disease Association Protein-Protein Link Evaluator) algorithm¹³. Briefly, this approach first generated a random network that has nearly the exact same structure as the original one that is derived from the InWeb database³⁶. The node labels (*i.e.* the protein names) were then randomly re-assigned to nodes of equal binding degree. DAPPLE assumes a null distribution of connectivity that is entirely a function of the binding degree of individual proteins. We built 10,000 different random networks and each of them had the same number of proteins, connectivity and per-protein binding degree as InWeb. The significance of our real PPI network formed by module genes was then assessed through permutation. For more details, please refer to the article¹³.

GO enrichment and cell-specific expression of module genes. In order to discern biological attributes of the identified module genes, we performed Gene Ontology (GO) enrichment analysis by using DAVID⁴⁹. DAVID bioinformatics resources consist of an integrated biological knowledgebase and analytic tools aimed at systematically extracting biological meaning from a list of genes or proteins. Fisher Exact tests were conducted in DAVID to compute the P-value for each GO term. In this case study, only GO terms with an adjusted P-value (Benjamini & Hochberg) of less than 0.25 were selected.

Cell-specific expression was assessed with an online tool Gene Enrichment Profiler⁵⁰. This tool computes the expression and enrichment of any set of query genes on the basis of a reference set obtained from 126 normal tissues and cell types (represented by 557 microarrays).

References

1. Fox, R. I. Sjogren's syndrome. *Lancet* **366**, 321–331, doi:10.1016/S0140-6736(05)66990-5 (2005).
2. Zhang, N. Z. *et al.* Prevalence of primary Sjogren's syndrome in China. *J Rheumatol* **22**, 659–661 (1995).
3. Ice, J. A. *et al.* Genetics of Sjogren's syndrome in the genome-wide association era. *J Autoimmun* **39**, 57–63, doi:10.1016/j.jaut.2012.01.008 (2012).
4. Voulgarelis, M. & Tzioufas, A. G. Pathogenetic mechanisms in the initiation and perpetuation of Sjogren's syndrome. *Nat Rev Rheumatol* **6**, 529–537, doi:10.1038/nrrheum.2010.118 (2010).

5. Lessard, C. J. *et al.* Variants at multiple loci implicated in both innate and adaptive immune responses are associated with Sjogren's syndrome. *Nat Genet* **45**, 1284–1292, doi:10.1038/ng.2792 (2013).
6. Li, Y. *et al.* A genome-wide association study in Han Chinese identifies a susceptibility locus for primary Sjogren's syndrome at 7q11.23. *Nat Genet* **45**, 1361–1365, doi:10.1038/ng.2779 (2013).
7. Dudbridge, F. & Gusnanto, A. Estimation of significance thresholds for genomewide association scans. *Genet Epidemiol* **32**, 227–234, doi:10.1002/gepi.20297 (2008).
8. Hindorf, L. A. *et al.* Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A* **106**, 9362–9367, doi:10.1073/pnas.0903103106 (2009).
9. Han, S. *et al.* Integrating GWASs and human protein interaction networks identifies a gene subnetwork underlying alcohol dependence. *Am J Hum Genet* **93**, 1027–1034, doi:10.1016/j.ajhg.2013.10.021 (2013).
10. International Multiple Sclerosis Genetics, C. Network-Based Multiple Sclerosis Pathway Analysis with GWAS Data from 15,000 Cases and 30,000 Controls. *Am J Hum Genet*, doi:10.1016/j.ajhg.2013.04.019 (2013).
11. Pedroso, I. *et al.* Common genetic variants and gene-expression changes associated with bipolar disorder are over-represented in brain signaling pathway genes. *Biol Psychiatry* **72**, 311–317, doi:10.1016/j.biopsych.2011.12.031 (2012).
12. Ideker, T. & Sharan, R. Protein networks in disease. *Genome Res* **18**, 644–652, doi:10.1101/gr.071852.107 (2008).
13. Rossin, E. J. *et al.* Proteins encoded in genomic regions associated with immune-mediated disease physically interact and suggest underlying biology. *PLoS Genet* **7**, e1001273, doi:10.1371/journal.pgen.1001273 (2011).
14. Loiseau, P. *et al.* HLA class I and class II are both associated with the genetic predisposition to primary Sjogren syndrome. *Hum Immunol* **62**, 725–731 (2001).
15. Mullighan, C. G. *et al.* Lack of association between mannose-binding lectin gene polymorphisms and primary Sjogren's syndrome. *Arthritis Rheum* **43**, 2851–2852, doi:10.1002/1529-0131(200012)43:12<2851::Aid-Anr28>3.0.Co;2-W (2000).
16. Kumagai, S. *et al.* Association of a new allele of the TAP2 gene, TAP2*Bky2 (Val577), with susceptibility to Sjogren's syndrome. *Arthritis Rheum* **40**, 1685–1692, doi:10.1002/1529-0131(199709)40:9<1685::Aid-Art19>3.0.Co;2-1 (1997).
17. Fox, R. I. Clinical features, pathogenesis, and treatment of Sjogren's syndrome. *Curr Opin Rheumatol* **8**, 438–445 (1996).
18. Eleftherohorinou, H., Hoggart, C. J., Wright, V. J., Levin, M. & Coin, L. J. Pathway-driven gene stability selection of two rheumatoid arthritis GWAS identifies and validates new susceptibility genes in receptor mediated signalling pathways. *Hum Mol Genet* **20**, 3494–3506, doi:10.1093/hmg/ddr248 (2011).
19. Negi, S. *et al.* A genome-wide association study reveals ARL15, a novel non-HLA susceptibility gene for rheumatoid arthritis in North Indians. *Arthritis Rheum* **65**, 3026–3035, doi:10.1002/art.38110 (2013).
20. Okada, Y. *et al.* Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* **506**, 376–381, doi:10.1038/nature12873 (2014).
21. Myouzen, K. *et al.* Functional variants in NFKBIE and RTKN2 involved in activation of the NF-kappaB pathway are associated with rheumatoid arthritis in Japanese. *PLoS Genet* **8**, e1002949, doi:10.1371/journal.pgen.1002949 (2012).
22. Okada, Y. *et al.* Meta-analysis identifies nine new loci associated with rheumatoid arthritis in the Japanese population. *Nat Genet* **44**, 511–516, doi:10.1038/ng.2231 (2012).
23. Hu, H. J. *et al.* Common variants at the promoter region of the APOM confer a risk of rheumatoid arthritis. *Exp Mol Med* **43**, 613–621, doi:10.3858/em.2011.43.11.068 (2011).
24. Chung, S. A. *et al.* Differential genetic associations for systemic lupus erythematosus based on anti-dsDNA autoantibody production. *PLoS Genet* **7**, e1001323, doi:10.1371/journal.pgen.1001323 (2011).
25. Lee, Y. H., Bae, S. C., Choi, S. J., Ji, J. D. & Song, G. G. Genome-wide pathway analysis of genome-wide association studies on systemic lupus erythematosus and rheumatoid arthritis. *Mol Biol Rep* **39**, 10627–10635, doi:10.1007/s11033-012-1952-x (2012).
26. Han, J. W. *et al.* Genome-wide association study in a Chinese Han population identifies nine new susceptibility loci for systemic lupus erythematosus. *Nat Genet* **41**, 1234–1237, doi:10.1038/ng.472 (2009).
27. Yang, W. *et al.* Meta-analysis followed by replication identifies loci in or near CDKN1B, TET3, CD80, DRAM1, and ARID5B as associated with systemic lupus erythematosus in Asians. *Am J Hum Genet* **92**, 41–51, doi:10.1016/j.ajhg.2012.11.018 (2013).
28. Anaya, J. M., Correa, P. A., Mantilla, R. D. & Arcos-Burgos, M. TAP, HLA-DQB1, and HLA-DRB1 polymorphism in Colombian patients with primary Sjogren's syndrome. *Semin Arthritis Rheum* **31**, 396–405 (2002).
29. Nakken, B. *et al.* Associations of MHC class II alleles in Norwegian primary Sjogren's syndrome patients: implications for development of autoantibodies to the Ro52 autoantigen. *Scand J Immunol* **54**, 428–433 (2001).
30. Garred, P. *et al.* Mannose-binding lectin polymorphisms and susceptibility to infection in systemic lupus erythematosus. *Arthritis Rheum* **42**, 2145–2152, doi:10.1002/1529-0131(199910)42:10<2145::Aid-Anr15>3.0.Co;2-# (1999).
31. Graudal, N. A. *et al.* The association of variant mannose-binding lectin genotypes with radiographic outcome in rheumatoid arthritis. *Arthritis Rheum* **43**, 515–521, doi:10.1002/1529-0131(200003)43:3<515::Aid-Anr6>3.0.Co;2-T (2000).
32. Tam, W. F. & Sen, R. IkappaB family members function by different mechanisms. *J Biol Chem* **276**, 7701–7704, doi:10.1074/jbc.C000916200 (2001).
33. Ou, T. T. *et al.* IkappaBalpha promoter polymorphisms in patients with primary Sjogren's syndrome. *J Clin Immunol* **28**, 440–444, doi:10.1007/s10875-008-9212-5 (2008).
34. Tak, P. P. & Firestein, G. S. NF-kappaB: a key role in inflammatory diseases. *J Clin Invest* **107**, 7–11, doi:10.1172/JCI11830 (2001).
35. Brunner, H. G. & van Driel, M. A. From syndrome families to functional genomics. *Nat Rev Genet* **5**, 545–551, doi:10.1038/nrg1383 (2004).
36. Lage, K. *et al.* A human phenome-interactome network of protein complexes implicated in genetic disorders. *Nat Biotechnol* **25**, 309–316, doi:10.1038/nbt1295 (2007).
37. Barabasi, A. L., Gulbahce, N. & Loscalzo, J. Network medicine: a network-based approach to human disease. *Nat Rev Genet* **12**, 56–68, doi:10.1038/nrg2918 [pii]10.1038/nrg2918 (2011).
38. Perwitasari, O., Cho, H., Diamond, M. S. & Gale, M. Jr. Inhibitor of kappaB kinase epsilon (IKK(epsilon)), STAT1, and IFIT2 proteins define novel innate immune effector pathway against West Nile virus infection. *J Biol Chem* **286**, 44412–44423, doi:10.1074/jbc.M111.285205 (2011).
39. Liu, J. Z. *et al.* A versatile gene-based test for genome-wide association studies. *Am J Hum Genet* **87**, 139–145, doi:10.1016/j.ajhg.2010.06.009 (2010).
40. Flicek, P. *et al.* Ensembl 2013. *Nucleic Acids Res* **41**, D48–55, doi:10.1093/nar/gks1236 (2013).
41. Razick, S., Magklaras, G. & Donaldson, I. M. iRefIndex: a consolidated protein interaction database with provenance. *BMC Bioinformatics* **9**, 405, doi:10.1186/1471-2105-9-405 (2008).
42. Cobb, B. L., Lessard, C. J., Harley, J. B. & Moser, K. L. Genes and Sjogren's syndrome. *Rheum Dis Clin North Am* **34**, 847–868, vii, doi:10.1016/j.rdc.2008.08.003 (2008).
43. Cruz-Tapias, P., Rojas-Villarraga, A., Maier-Moore, S. & Anaya, J. M. HLA and Sjogren's syndrome susceptibility. A meta-analysis of worldwide studies. *Autoimmun Rev* **11**, 281–287, doi:10.1016/j.autrev.2011.10.002 (2012).
44. Complete sequence and gene map of a human major histocompatibility complex. The MHC sequencing consortium. *Nature* **401**, 921–923, doi:10.1038/44853 (1999).
45. Jia, P., Zheng, S., Long, J., Zheng, W. & Zhao, Z. dmGWAS: dense module searching for genome-wide association studies in protein-protein interaction networks. *Bioinformatics* **27**, 95–102, doi:10.1093/bioinformatics/btq615 (2011).

46. Ideker, T., Ozier, O., Schwikowski, B. & Siegel, A. F. Discovering regulatory and signalling circuits in molecular interaction networks. *Bioinformatics* **18** Suppl 1, S233–240 (2002).
47. Efron, B. Correlated z-values and the accuracy of large-scale statistical estimates. *J Am Stat Assoc* **105**, 1042–1055, doi:10.1198/jasa.2010.tm09129 (2010).
48. Jia, P. *et al.* Network-assisted investigation of combined causal signals from genome-wide association studies in schizophrenia. *PLoS Comput Biol* **8**, e1002587, doi:10.1371/journal.pcbi.1002587 (2012).
49. Dennis, G. Jr. *et al.* DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol* **4**, P3 (2003).
50. Benita, Y. *et al.* Gene enrichment profiles reveal T-cell development, differentiation, and lineage-specific transcription factors including ZBTB25 as a novel NF-AT repressor. *Blood* **115**, 5376–5384, doi:10.1182/blood-2010-01-263855 (2010).

Acknowledgements

This study was supported by Strategic Priority Research Program (B) of the Chinese Academy of Sciences (Grant No. XDB02030002), State Administration of Foreign Experts Affairs of Chinese Academy of Sciences (CAS/SAFEA) International Partnership Program for Creative Research Teams (Grant No. Y2CX131003), National Natural Science Foundation of China (NSFC, Grant No. 81101545) and Key Laboratory of Mental Health, Institute of Psychology, Chinese Academy of Sciences (to K.Z.).

Author Contributions

K.F. designed the study and performed the analyses; K.Z. contributed to and supported GWAS data analysis; J.W. supervised the research. All authors read and approved the final manuscript.

Additional Information

Supplementary information accompanies this paper at <http://www.nature.com/srep>

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Fang, K. *et al.* Network-assisted analysis of primary Sjögren's syndrome GWAS data in Han Chinese. *Sci. Rep.* **5**, 18855; doi: 10.1038/srep18855 (2015).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>