Identification of novel genes associated with dysregulation of B cells in patients with primary Sjögren's syndrome

CURRENT STATUS: UNDER REVIEW

Jun Inamo
Keio University School of Medicine

ORCiD: https://orcid.org/0000-0002-9927-7936

Katsuya Suzuki
Keio Gijuku Daigaku Rikogakubu Daigakuin Rikogaku Kenkyuka

Masaru Takeshita
Keio University School of Medicine

Yoshiaki Kassai
Takeda Pharamaceutical Company Limited

Maiko Takiguchi
Takeda Pharamaceutical Company Limited.

Rina Kurisu
Takeda Pharamaceutical Company Limited.

Yuumi Okuzono
Takeda Pharamaceutical Company Limited.

Shinya Tasaki
Takeda Pharamaceutical Company Limited.

Akihiko Yoshimura
Keio University School of Medicine

Tsutomu Takeuchi
tsutake@z5.keio.jp Corresponding Author
ORCiD: https://orcid.org/0000-0003-1111-8218

10.21203/rs.3.rs-26390/v1
SUBJECT AREAS
Rheumatology

KEYWORDS
primary Sjögren's syndrome, B cells, long non-coding RNA, transcriptome, gene co-expression network
Abstract

Background

The aim of this study was to identify at the transcriptome level the molecular mechanism of dysregulation of B cell subpopulations of primary Sjögren's syndrome (pSS).

Methods

We enrolled patients with pSS (n=6) and healthy controls (HC) (n=6) in the discovery cohort using microarray and pSS (n=14) and HC (n=12) in the validation cohort using quantitative PCR (qPCR). Peripheral B cells acquired from these subjects were separated by cell sorting into four subsets: CD38 - IgD + (Bm1), CD38 + IgD + (naïve B cells), CD38 high IgD + (pre-germinal centre B cells) and CD38 ± IgD - (memory B cells).

Results

Expression of the long non-coding RNA LINC00487 was significantly upregulated in all B cell subsets, as was that of HLA and interferon (IFN) signature genes. Moreover, the normalized intensity value of LINC00487 significantly correlated with the disease activity score of all pSS B cell subsets. Studies of human B cell lines revealed that the expression of LINC00487 was strongly induced by IFNα. Weighted gene co-expression network analysis revealed six gene clusters associated with the B cell subpopulation of pSS. Further, SOX4 was identified as an inter-module hub gene.

Conclusion

Our transcriptome analysis revealed key genes involved in the dysregulation of B cell subpopulations associated with pSS.

Background

Primary Sjögren's syndrome (pSS) is a systemic autoimmune disease characterized by exocrine gland dysfunction, which leads to dryness of the eyes and mouth [1]. Upregulation of the interferon (IFN) pathway [2] and dysregulation of B cells play a critical role in the pathogenesis of pSS. First, the loss of B cell tolerance can lead to overproduction of anti-Sjögren’s syndrome-related antigen A (anti-SSA) autoantibodies. Autoreactive B cells are important in the development of clinical disease, because serum autoantibodies may precede the onset of dryness. Second, IFN-stimulated genes (ISGs) are
upregulated in B cells in peripheral blood as well as in the salivary gland lesions of patients with pSS [3, 4]. Third, IFNα promotes loss of tolerance and development of autoreactive B cells [5]. Genome-wide association studies identified single nucleotide polymorphisms of ISGs that are associated with the risk of pSS [6, 7]. Further, viral infection, which enhances the IFN pathway, may contribute to the pathogenesis of pSS [8].

Non-coding RNAs, which are emerging as critical regulators of signal transduction pathways, may be involved in the immune system. Non-protein coding microRNAs are involved in the dysregulation in B cells [9] and in the upregulation of the IFN pathway in patients with pSS [10]. Long non-coding RNAs (IncRNAs) are involved in diverse regulatory functions, including the modulation of chromatin structure and post-transcriptional regulation affecting the stability of mRNAs and proteins [11]. Recent evidence indicates that IncRNAs, which have become the focus of studies on autoimmune diseases, may contribute to the pathogenesis of pSS through multiple signal transduction pathways [12]. For example, differentially expressed IncRNAs in the labial salivary glands of pSS patients were identified using microarray analysis [13]. While multi-omics studies of the whole-blood transcriptome highlight the importance of cytotoxic CD8 T cells in the pathogenesis of pSS, the B cell subpopulation is associated with clinical traits [14].

Considering that gene expression pattern is tissue-specific [15], it is better to focus on the target cell subpopulation to avoid noise from other cells in transcriptome analysis. However, we are unaware of studies that focus on the effects of dysregulation of transcriptomes, including IncRNAs, of B cells on the pathogenesis of pSS. We previously reported that proteome analysis of patient with pSS identified B cell activating factor as one of the serum proteins which positively correlate with disease activity [16]. In addition, the distribution of peripheral B cell subsets is profoundly altered in patients with pSS. Staining for IgD/CD38 is helpful for studying circulating B cell subsets, separating developmental stages from naive to memory B cells (Bm1 to Bm5) [17]. Using this gating strategy, we previously reported that the proportion of pre-GC B cells (IgD+/CD38high) was higher in patients with pSS compared with healthy controls and associated with clinical traits, such as disease activity [18], indicating the importance of evaluating the involvement of each B cell subset in the pathogenesis of
To determine the role of the B cell subset in the pathogenesis of pSS, we investigated potentially significant genes according to their differential levels of expression and connectivity with other genes. Here we describe the upregulation of the interferon signalling pathway and the differential expression of genes encoding human leukocyte antigen (HLA) molecules and LINC00487 in B cell subpopulations of patients with pSS compared with healthy controls (HC). The expression levels of LINC00487 correlated with the disease activity of pSS and IFN-signature genes; and LINC00487 was induced by IFNα. Further, we used weighted gene co-expression network analysis (WGCNA) to identify genes of co-expression networks specific to a B cell subset of patients with pSS, suggesting that aberrant molecular interactions in B cells contribute to the aetiology of pSS.

Methods

Patients and controls

The study protocol is shown in Fig. 1A. We enrolled patients with pSS (n = 6) and HC (n = 6) matched for age and sex (Supplementary Table 1). The patients with pSS fulfilled the 2002 American-European criteria for SS [19] and the 2012 American College of Rheumatology classification criteria for SS [20]. All patients were female and had symptom of dryness, anti-SSA antibodies and biopsy-proven sialadenitis. The patients had not received immunosuppressive therapy, and the HCs did not have immunological disorders. Clinical and serological information of patients with pSS were collected from their medical records. The disease activity of pSS was assessed according to the guidelines of the European League against Rheumatism and the Primary Sjögren's syndrome disease activity index (ESSDAI) [21, 22]. The distribution of ESSDAI domain was described in Supplementary Table 17. This study was performed in accordance with relevant guidelines and regulations. The Ethics Committee of Keio University School of Medicine approved this study (IRB No. 20110258), and written informed consent was obtained from each subject before blood collection.

Cell sorting

Peripheral blood mononuclear cells from patients with pSS and HC were separated using gradient centrifugation with Lymphoprep (Axis-Shield; Oslo, Norway). Gating strategy was shown in Supplementary Fig. 8. Peripheral CD19+ B cells were prepared with anti-CD19 antibody-coated
microbeads (Miltenyi Biotec). The peripheral CD19+ B cells were incubated with anti-IgD and CD38 antibodies for fluorescence-activated cell sorting (FACS) analysis (FACS Aria III flow cytometer, BD Biosciences). We defined subsets of B cells as follows: Bm1 cells; CD38−IgD+, naïve B cells; CD38+IgD+, pre-germinal centre (pre-GC) B cells; CD38highIgD+ and memory B cells; CD38±IgD−.

Microarray analysis
Total RNA was extracted from B cell subsets and transcribed into cDNA using NucleoSpin RNA (Macherey Nagel) and ReverTra Ace qPCR RT Master Mix (Toyobo). Gene expression was measured using the Human Genome U133 Plus 2.0 Array (Affymetrix). We applied percentile shift normalization to the raw signal data acquired from a microarray and annotated each probe with its gene symbol using GeneSpring software (Agilent Technologies). Probes with interquartile ranges in the lowest 20% were excluded. We next selected probes with > 2.0-changes for pSS vs HC in any one B cell subset to identify differentially expressed genes (DEGs). We controlled for the false discovery rate using the Bonferroni multiple testing-corrected p value < 0.05.

Further, to explore novel gene co-expression networks and common hub genes, we created another gene set by removing genes with coefficients of variation greater than 0.3 after intensity filtering and applied them to the WGCNA R package [23] following the package’s tutorial. Briefly, we applied step-by-step construction of the gene network and identification of modules to fileted gene expression dataset. First, we chose the proper soft-thresholding power based on the criterion of approximate scale-free topology. After calculating the adjacencies with selected soft-thresholding power, we transformed the adjacency into Topological Overlap Matrix, and calculate the corresponding dissimilarity. Then, we used hierarchical clustering to produce a hierarchical clustering tree of genes. Module identification amounts to the identification of individual branches; we cut the branches off the dendrogram with setting minimum module size 30. As a result, 6 and 5 modules were generated in B cells of patients with pSS and HC, respectively.

GO term enrichment and pathway analysis
To functionally characterize DEGs identified in each B cell subpopulation from microarray analysis, we performed enrichment analysis according to GO terms and KEGG pathways using the ClueGO [24] and
CluePedia [25] plugins of Cytoscape [26]. Briefly, ClueGO creates and visualizes a functionally grouped network of GO terms and KEGG pathways. CluePedia is a GlueGO plugin that integrates a gene set into the network derived from ClueGO. The enrichment is based on the percentage calculated according to the total number of genes in DEGs list as a two-sided hypergeometric test. To correct for multiple testing, we used the Benjamini-Hochberg procedure and the Bonferroni correction for KEGG pathways and GO terms, respectively, with adjusted p values in each case of p = 0.05. The connection between nodes is based on Cohen's kappa statistic (≥ 0.4), which depends on genes shared between nodes. Further, we investigated upregulated or downregulated pathways according to the expression level of each DEG using Ingenuity Pathway Analysis (IPA). In IPA, Z scores greater or less than 1.96 were considered significant.

**Cell culture and real-time quantitative PCR (qPCR)**

B cell lines (Ramos, CCRF, U266B1) (each 1.0 × 10^5 /mL) in RPMI 1640 medium (American Type Culture Collection) supplemented with 10% fetal calf serum were treated with different concentrations of IFNα, IFNγ and TNFα for 48 h. We next isolated total RNA that was transcribed into cDNAs using NucleoSpin RNA (Macherey Nagel) and ReverTra Ace qPCR RT Master Mix (Toyobo). qPCR was performed using PowerUp SYBR Green Master Mix (Thermo Fisher Scientific). The levels of total cellular RNA were similar among the cell lines. At least two biological repeats were performed for each experiment. The levels of GAPDH mRNA were used to adjust the values of target transcripts determined using qPCR analyses. We validated gene expression levels in CD19^+^ B cells of patients with pSS (n = 14) and HC (n = 12) (Supplementary Table 2) using quantitative PCR (qPCR) analysis. Sequence of primers for qPCR analysis was described in Supplementary Table 3.

**Statistics**

Continuous values are shown as the mean ± standard deviation (SD). Differences between the groups were analysed using the Mann–Whitney test for continuous variables. Correlations were analysed using Spearman’s correlation coefficient, unless otherwise noted. P < 0.05 indicates a significant difference. Statistical tests were conducted using R software version 3.5.2.

**Results**

Identification of gene signatures in B cell subsets of pSS
First, we visualized the gene expression profile after intensity filtration. Using principal component and hierarchical clustering analyses, we found that clinical status, namely pSS or HC, had less influence on phenotype than the type of B cell subset (Fig. 1B and C). Memory B cells, in particular, had a distinct expression pattern compared with those of other cell types.

To determine the gene signatures of B cells subset of patients with pSS, we identified DEGs and performed GO and pathway analyses. Gene selection yielded 623 significant genes (Supplementary Table 4). Analysis of changes in expression indicated that 23, 92, 18 and 32 genes were upregulated in Bm1, naïve, pre-GC and memory B cell subsets, respectively (Fig. 1D).

In immunological GO analysis, the type 1 interferon response was independent of cell subpopulation (Fig. 2A-D). GO annotations were enriched in the earlier developmental stages of B cell subsets such as Bm1 and naïve B cells, and this trend was apparent in enrichment analysis according to GO term biological process (Supplementary Fig. 1) and KEGG pathways (Supplementary Fig. 2). Consistent with previous reports, ISGs were upregulated in B cells of pSS compared with those of HC (Fig. 2E).

Further, HLA genes were similarly upregulated. In pathway analysis using IPA, the IFN signalling pathway was the most upregulated pathway in B cell subsets of patients with pSS (Fig. 2F).

We found that a lncRNA, LINC00487, was among the 15 probes significantly upregulated in all B cell subsets of pSS compared with those of HC (Supplementary Table 5). The fold-changes of LINC00487 expression were as follows: Bm1, 8.4 (p = 0.038); naïve, 11.3 (p = 0.014); pre-GC, 8.5 (p = 0.089); memory, 6.37 (p = 0.078) (Fig. 2E). There was no significant difference among cell types (Supplementary Fig. 3).

**Expression of LINC00487 correlates with IFN signature genes and disease activity score of pSS**

To determine the characteristics of LINC00487, we identified the top three genes that correlated with LINC00487 (Fig. 3A). These genes significantly correlated with LINC00487, and all are ISGs. Further, the clinical disease activity score significantly correlated with the expression level of LINC00487 in all pSS B cell subsets: Bm1, r = 0.98 (p < 0.01); naïve, r = 0.93 (p < 0.01); pre-GC, r = 0.84 (p = 0.03); memory, r = 0.93 (p < 0.01) (Fig. 3B).
Using validation cohort, we tried to conduct qPCR analysis using each B cell subset. However, because LINC00487 expressed at very low level in HCs, we couldn’t sort enough each B cell subset from them to detect expression of LINC00487. Therefore, considering that upregulation of LINC00487 was common in all B cell subsets in microarray cohort, we used a bulk of CD19+ B cells in validation cohort. As a result, the expression of LINC00487 in CD19+ B cells of patients with pSS was significantly upregulated compared with HC (p = 0.035) (Fig. 3C) and significantly correlated with interferon induced protein 44-like gene (IFI44L) in patients with pSS (r = 0.83, p < 0.01) (Fig. 3D) as well as in microarray analysis. Although there was a significant correlation between them in HC, that in pSS was more apparent (r = 0.60, p = 0.04 in HC, Supplementary Fig. 4A). To confirm the induction of LINC00487 in response to IFNα, we used three types of B cell lines and measured LINC00487 expression using qPCR. As with IFI44L, LINC00487 was strongly induced by treatment with IFNα (Fig. 3E and F), unlike TNFα and IFNγ (Supplementary Fig. 4B), suggesting that IFNα is an upstream regulator of LINC00487. Compared with clinical features, the expression of LINC00487 tended to be higher in patients with higher serum IgG levels and lower lymphocyte counts which were characteristic for pSS (Supplementary Fig. 4C).

Co-expression gene networks constructed for each B cell subset

To identify transcriptional networks in the B cells of patients with pSS, we performed WGCNA. WGCNA can identify novel gene interactions which might be missed in the DEGs analysis, because WGCNA is based on correlations of gene expression [23].

After selection, we identified a new set comprising 3634 genes (Supplementary Table 6–7). In principal component analysis using these gene sets, Bm1 and naïve B cell subsets were clearly distinguishable between pSS and HC (Fig. 4A). WGCNA identified 6 modules in pSS and an association between cell subsets and modules (Fig. 4B). For example, the yellow and brown modules are associated with the pre-GC B cell subset, whereas the turquoise module is associated with the memory B cell subset. The grey module of pSS was enriched in Bm1/naïve B cell subsets with various ranges of eigengenes. In the analysis to detect an association between modules and clinical traits, the grey module significantly correlated with the clinical disease activity score (Fig. 4C). Similarly,
another five modules were identified in HC (Supplementary Fig. 5A).

To reveal the characteristics of modules, we performed functional analysis. The cluster of enriched pathway, upstream regulator and disease/biological functions in modules of pSS were distinguished from those of HC (Fig. 5A and Supplementary Tables 8-13). The top five most significantly enriched pathways, regulators and disease/biological functions in modules of pSS are shown in Fig. 5B. In upstream regulator analysis, transcriptional regulators such as Sry-related high mobility group box (Sox) family such as SOX4 and SOX11 were identified in the yellow and blue modules of pSS. Focusing on specificity (Supplementary Fig. 6 and Supplementary Table 14-16), mir-21-5p was identified as the top significant regulator in the yellow module of pSS. To search for hub genes, we described a co-expressed gene network using top highly interconnected genes defined by topological overlap in each module (Fig. 5C). Several genes bridged inter-modules, namely hub genes, such as SOX4. SOX4 was not identified in the co-expression network of HC (Supplementary Fig. 5B), suggesting its involvement in the dysregulation of B cells.

Association among SOX4 and GC B cells-related genes

As expected from the results of WGCNA, SOX4 was significantly upregulated in pre-GC B cells of pSS and HCs compared with other subsets (Fig. 6A). In a comparison between pSS and HC, the difference in expression levels was significant only in naïve B cells.

Accordingly, we predicted that SOX4 was associated with GC-related biological function. Principal component analysis showed that the expression profiles of GC-B cells-related genes [25] in Bm1, naïve and pre-GC B cells of pSS were more similar compared with those of HC (Fig. 6B), suggesting that Bm1 and naïve B cells of patients with pSS already have a GC-related biological function. Expression patterns of GC-B cells-related genes were distinct according to the maturation stage of B cells (Fig. 6C). AURKA and NFKB1 were significantly correlated with SOX4 in pre-GC B cells negatively and memory B cells positively of pSS, respectively (Fig. 6D and Fig. 6E). In contrast, CD40 and MYC were significantly correlated with SOX4 in Bm1 negatively and pre-GC B cells positively of HC, respectively (Supplementary Fig. 7).

Discussion
In the present study, we conducted transcriptome analysis of B cell subpopulations. First, we found that expression of *LINC00487* was upregulated in B cell subsets derived from patients with pSS. Further, *LINC00487* expression significantly correlated with disease activity and was induced by IFNα stimulation. Next, using WGCNA, we identified several key networks and hub genes in the B cells of patients with pSS. Our findings deepen our understanding about the mechanism of pathogenesis of aberrant B cells in pSS.

IFN signalling is a central component of the pathogenesis of pSS [27]. While we found significant upregulation of IFN signalling in B cells of patients with pSS, to our knowledge, such differences among subsets have not been reported. Type 1 IFN signalling plays a crucial role in the development of autoreactive B cells in a mouse model of autoimmune disease [5]. Further, more immunological GO terms including type 1 IFN response were enriched in earlier maturation stages of B cells. Moreover, primary component analysis using gene sets of WGCNA and GC-B cells-related genes revealed that Bm1/naïve B cells of pSS were clearly distinguished from those of HC. These results suggest that molecular dysfunction, such as disruption of peripheral tolerance, might start at an early stage in the maturation of B cells. Indeed, the frequencies of naïve B cells expressing autoreactive antibodies are significantly increased in patients with pSS [28].

Interestingly, expression of the HLA class II gene was upregulated. GWAS studies of pSS found associations of *HLA-DQA1* and *HLA-DQB1* loci [6, 7]. Further, IFNα induces the expression of *HLA-DQA1* [29]. In patients with pSS, *HLA-DQA1* and *HLA-DQB1* alleles are associated with higher concentrations of anti-SSA and SSB antibodies [30]. These findings suggest that aberrant interactions among IFN signalling and HLA class II genes may trigger the breakdown of B cell tolerance, leading to the development of pSS.

*LINC00487* was upregulated in all B cell subsets of pSS and its expression significantly correlated with the disease activity scores of pSS and ISGs. Moreover, we found that IFNα was an upstream regulator of *LINC00487* in B cells. The sequence of *LINC00487*, which belongs to the class of long intergenic non-coding RNAs and resides on human chromosome 2, is atypically long (> 40,000 bases). Although *LINC00487* is one of the hub genes in the normal development of human B cells [31], many of its
properties, including its function, are unexplained. Further, there is no ortholog or paralog of this gene in species other than humans that may provide clues to its function in human cells. However, according to AceView, one of three transcriptional variants derived from \textit{LINC00487} has high potential to encode a protein [32].

Four genomic locations are considered candidate enhancers of \textit{LINC00487} transcription. The targets of the regulators of \textit{LINC00487} overlap with those of other ISGs [33]. Moreover, referring to the public transcriptome database of microarray analyses of healthy humans, expression of \textit{LINC00487} is higher specifically in centroblasts and centrocytes of the germinal centre [34]. IFNα promotes the autoreactivity of B cells via germinal centre pathways [5]. Further, \textit{LINC00487} expression is upregulated in the subgroup of diffuse large B-cell lymphoma with molecular characteristics of germinal centre B cells, compared with other subgroups, and is associated with the efficacy of B cell depletion therapy [35]. Therefore, our study suggests that upregulation of \textit{LINC00487} expression in all B cell subsets may reflect or regulate the enrichment of a germinal centre-like reaction by IFNα from an early stage of B cell development, leading to B cell autoreactivity in patients with pSS. Thus, \textit{LINC00487} may serve as a predictive biomarker for B cell target therapy in autoimmune diseases.

Indeed, we show here that the expression of \textit{LINC00487} was associated with higher IgG levels.

Gene co-expression analysis revealed an aberrant network in B cell subpopulations. The top significant upstream regulator of the grey module of pSS, which was associated with clinical disease activity score and enriched in the early stage of B cell development in pSS, was the gene encoding T-cell acute lymphocytic leukemia protein 1 (\textit{TAL1}). B cell development is stringently controlled by stage-specific transcription factors. The transcription factor \textit{TAL1} regulates genes such as IKAROS family zinc finger 3 (\textit{IKZF3}), which is one of hub genes in the grey module of pSS (Fig. 5C). \textit{IKZF3} is a lineage-specific transcription factor that is important in the regulation of B cell proliferation and development [36].

In the pre-GC B cells-associated module of pSS, \textit{SOX4} was identified as an upstream regulator and a hub gene. In mice, \textit{SOX4} regulates the differentiation of early-stage B cells by activating the expression of \textit{Rag1} and \textit{Rag2} [37]. Further, \textit{SOX4} contributes to the formation of ectopic lymphoid-like
structures via promoting CXCL13, which is ligand of CXCR5 on naïve B cells and critical for migration into light zone of germinal centre undergoing somatic hypermutation, production from PD-1hiCXCR5−CD4+ T cells [38]. Additionally, in proteome analysis, CXCL13 positively correlates with the disease activity score and serum IgG levels of patients with pSS [16]. However, the role of SOX4 in mature B cells is unclear. Our data showed that the expression of SOX4 was significantly associated with AURKA and NFKB1 in B cell subpopulations of patients with pSS, and CD40 and MYC in those of HC, respectively. In the initiation of GC reaction, naïve B cells are activated by antigen within follicles. Then, activated B cells migrate to interfollicular region, where they interact with follicular helper T cells (Tfh) via CD40. After fully activation by Tfh, selected B cells enter the GC pathway to expand and accumulate mutations [39]. A mature GC is composed by two compartments, a dark zone (DZ) and a light zone (LZ) [40]. DZ B cells undergo rapid proliferation and affinity maturation. LZ B cells undergo class switch recombination, and selected B-cells in LZ can recirculate to the DZ for further rounds of division and affinity maturation, or differentiate into memory B-cells or plasma cells and exit the GC. AURKA is serine/threonine kinase which plays an essential role in mitosis [41], and its expression is highly increased in DZ B cells [40]. Additionally, in vivo study, AURKA inhibitor triggers mitotic arrest and apoptosis of B cell lymphoma, indicating AURKA has critical role in GC homeostasis [42]. In cancer-focused protein-protein interaction screens, AURKA and SOX4 are identified as partners [43], although further study is needed to explore the significant function in autoimmune diseases. NFKB signalling is activated by CD40-stimulation in LZ, and aberrant activation of NFKB signalling contributes to the GC lymphogenesis [44]. Previous silico analysis of human breast cancer also identified SOX4 as a activator of Phosphatidyl Inositol-3 kinases/Akt signalling, which is upstream of NFKB [45], supporting our results about association between SOX4 and NFKB1. Regarding CD40 and MYC, their expressions are upregulated in LZ B cells [40]. MYC regulates the expression of key regulators of cell cycle transit to coordinate cell growth and metabolism with cell division, including AURKA [42]. In addition, MYC+-B cells is required for GC formation and maintenance [46, 47]. Given that several studies have suggested that SOX4 function is context-dependent [48], SOX4 may be
involved in GC response as important regulator in a complexed gene network of mature B cells, not only T cells.

Further, we identified miR-21 as an upstream regulator of the yellow module of pSS. Although miR-21 is upregulated in peripheral blood mononuclear cells of pSS [49], the present study is the first to indicate its involvement in pSS via B cell dysregulation. miR-21 regulates the immune response of memory T cells via induction of transcription networks, such as SOX4 [50], supporting our hypothesis that interactions between miR-21 and SOX4 play a critical role in B cell dysfunction in pSS.

The current study suffers from several limitations. First, patients in our cohort have low disease activity/severity. Because there were few patients with high disease activity before immunosuppressive treatment, it was difficult to include such patients in this study. Second, sample size was limited. To overcome limitation about small sample size, we validated by qPCR using another cohort, supporting results derived from microarray analysis. Nevertheless, our focus on B cell subpopulation using a multi-level approach employing analysis of DEGs and WGCNA identified significant genes and networks as novel players in the pathogenesis of pSS. To confirm our results, functional studies are required.

Abbreviations
DZ, dark zone: ESSDAI, EULAR Sjögren's Syndrome Disease Activity Index: FACS, fluorescence-activated cell sorting: GC-B, germinal centre B cell: HC, healthy control: HLA, human leukocyte antigen: IFN, interferon: IKZF3, IKAROS family zinc finger 3: ISGs, interferon-stimulated genes: LZ, light zone: IncRNA, long non-coding RNA: pSS, primary Sjögren's syndrome: TAL1, T-cell acute lymphocytic leukemia protein 1: WGNCA, weighted gene co-expression network analysis.

Declarations
Ethics approval and consent to participate
Ethics approval was obtained from the Institutional Review Board of Keio University School of Medicine (IRB No. 20110258).

Consent for publication
Not required.
Availability of data and materials

The transcriptome data are available at the GEO database. The accession codes are GSE135809. All custom computer codes in the generation or processing of the described data are available upon reasonable request. Supplementary Table 4-17 and R script for WGCNA are deposited in figshare (https://figshare.com/articles/Supplementary_Table_4-14/11683959).

Competing interests

YK, YO, MTaki and RK are employed by Takeda Pharmaceutical Company Limited. S.T. was employed by Takeda Pharmaceutical Company Limited. KS has received research grants from Eisai, Bristol-Myers Squibb, Kissei Pharmaceutical, and Daiichi Sankyo, and speaking fees from Abbie Japan, Astellas Pharma, Bristol-Myers Squibb, Chugai Pharmaceutical, Eisai, Fuji Film Limited, Janssen Pharmaceutical, Kissei Pharmaceutical, Mitsubishi Tanabe Pharmaceutical, Pfizer Japan, Shionogi, Takeda Pharmaceutical, and UCB Japan, consulting fees from Abbie, and Pfizer Japan. AY has received speaking fees from Chugai Pharmaceutical, Mitsubishi Tanabe Pharmaceutical, Pfizer Japan, Ono Pharmaceutical, Maruho, and Novartis, and consulting fees from GSK Japan. TT has received research grants from Astellas Pharma Inc, Bristol-Myers KK, Chugai Pharmaceutical Co. Ltd., Daiichi Sankyo Co. Ltd, Takeda Pharmaceutical Co. Ltd, Teijin Pharma Ltd, AbbVie GK, Asahikasei Pharma Corp, Mitsubishi Tanabe Pharma Co, Pfizer Japan Inc, and Taisho Toyama Pharmaceutical Co. Ltd, Eisai Co. Ltd, AYUMI Pharmaceutical Corporation, and Nipponkayaku Co. Ltd, and speaking fees from AbbVie GK, Bristol-Myers KK, Chugai Pharmaceutical Co. Ltd, Mitsubishi Tanabe Pharma Co, Pfizer Japan Inc, and Astellas Pharma Inc, and Diaichi Sankyo Co. Ltd, and consultant fees from Astra Zeneca KK, Eli Lilly Japan KK, Novartis Pharma KK, Mitsubishi Tanabe Pharma Co, Abbvie GK, Nipponkayaku Co. Ltd, Janssen Pharmaceutical KK, Astellas Pharma Inc, and Taiho Pharmaceutical Co. Ltd. JI and MT declare no potential conflict of interest.

Funding

This study was funded by Keio University (grunt number: 04-013-0188) and Takeda Pharmaceutical Company Limited (grunt number: 04-078-0067).

Authors’ Contributions
Conceptualization: JI, KS, MTake, YK, YO, MTaki, RK and ST. Funding acquisition: KS and TT. Data acquisition: YK, MTaki and RK. Formal analysis: JI, YO and ST. Supervision: KS. Writing and original draft preparation: JI. Writing review and editing: KS, AY and TT.

Acknowledgements

This work was supported by Takeda Pharmaceutical Company Limited, Kanagawa, Japan (grant number 04-078-0067) and Keio University (grant number 04-013-0188).

We thank Ms. Harumi Kondo, Ms. Mayumi Ota, Ms. Yoshiko Yogiashi, Ms. Yuki Otomo, Mr. Fumitsugu Yamane for helping with the experiments, and Digital Medical Communications Corp. for proofreading sentences for English proofreading.

References

1. Qin B, Wang J, Yang Z, et al. Epidemiology of primary Sjögren's syndrome: a systematic review and meta-analysis. Ann Rheum Dis. 2015;74:1983–9.

2. Muskardin TLW, Niewold TB. Type I interferon in rheumatic diseases. Nat Rev Rheumatol. 2018;14:214–28.

3. Imgenberg-Kreuz J, Sandling JK, Björk A, et al. Transcription profiling of peripheral B cells in antibody-positive primary Sjögren's syndrome reveals upregulated expression of CX3CR1 and a type I and type II interferon signature. Scand J Immunol. 2018;87:e12662.

4. Devauchelle-Pensec V, Cagnard N, Pers JO, et al. Gene expression profile in the salivary glands of primary Sjögren's syndrome patients before and after treatment with rituximab. Arthritis Rheum. 2010;62:2262–71.

5. Domeier PP, Chodisetti SB, Schell SL, et al. B-Cell-Intrinsic Type 1 Interferon Signaling Is Crucial for Loss of Tolerance and the Development of Autoreactive B Cells. Cell Rep. 2018;24:406–18.

6. Lessard CJ, Li H, Adrianto I, et al. Variants at multiple loci implicated in both innate and adaptive immune responses are associated with Sjögren’s syndrome. Nat Genet.
7. Li Y, Zhang K, Chen H, et al. A genome-wide association study in Han Chinese identifies a susceptibility locus for primary Sjögren's syndrome at 7q11.23. Nat Genet. 2013;45:1361-5.

8. Triantafyllopoulou A, Moutsopoulos H. Persistent viral infection in primary Sjögren's syndrome: review and perspectives. Clin Rev Allergy Immunol. 2007;32:210-4.

9. Reale M, D'Angelo C, Costantini E, et al. MicroRNA in Sjögren's Syndrome: Their Potential Roles in Pathogenesis and Diagnosis. J Immunol Res. 2018:7510174.

10. Wang-Renault SF, Boudaoud S, Nocturne G, et al. Deregulation of microRNA expression in purified T and B lymphocytes from patients with primary Sjögren's syndrome. Ann Rheum Dis. 2018;77:133-40.

11. Tang Y, Zhou T, Yu X. The role of long non-coding RNAs in rheumatic diseases. Nat Rev Rheumatol. 2017;13:657-69.

12. Sandhya P, Joshi K, Scaria V. Long noncoding RNAs could be potential key players in the pathophysiology of Sjögren's syndrome. Int J Rheum Dis. 2015;18:898-905.

13. Shi H, Cao N, Pu Y, et al. Long non-coding RNA expression profile in minor salivary gland of primary Sjögren's syndrome. Arthritis Res Ther. 2016;18:109.

14. Tasaki S, Suzuki K, Nishikawa A, et al. Multiomic disease signatures converge to cytotoxic CD8 T cells in primary Sjögren's syndrome. Ann Rheum Dis. 2017;76:1458-66.

15. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science. 2015;348:648-60.

16. Nishikawa A, Suzuki K, Kassai Y, et al. Identification of definitive serum biomarkers associated with disease activity in primary Sjögren's syndrome. Arthritis Res Ther. 2016;18:106.
17. Bohnhorst J, Bjørgan MB, Thoen JE, et al. Bm1-Bm5 classification of peripheral blood B cells reveals circulating germinal center founder cells in healthy individuals and disturbance in the B cell subpopulations in patients with primary Sjögren's syndrome. J Clin Invest. 2005;115:3205–16.

18. Ishioka-Takei E, Yoshimoto K, Suzuki K, et al. Increased proportion of a CD38highIgD + B cell subset in peripheral blood is associated with clinical and immunological features in patients with primary Sjögren's syndrome. Clin Immunol. 2018;187:85–91.

19. Vitali C, Bombardieri S, Jonsson R, et al. Classification criteria for Sjögren's syndrome: a revised version of the European criteria proposed by the American-European Consensus Group. Ann Rheum Dis. 2002;61:554–8.

20. Shiboski SC, Shiboski CH, Criswell L, et al. American College of Rheumatology classification criteria for Sjögren's syndrome: A data-driven, expert consensus approach in the Sjögren's International Collaborative Clinical Alliance Cohort. Arthritis Care Res (Hoboken). 2012;64:475–87.

21. Seror R, Ravaud P, Bowman SJ, et al. EULAR Sjögren's syndrome disease activity index: development of a consensus systemic disease activity index for primary Sjögren's syndrome. Ann Rheum Dis. 2010;69:1103–9.

22. Seror R, Bowman SJ, Brito-Zeron P, et al. EULAR Sjögren's syndrome disease activity index (ESSDAI): a user guide. RMD Open. 2015;1:e000022.

23. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9:559.

24. Bindea G, Mlecnik B, Hackl H, et al. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. Bioinformatics. 2009;25:1091–3.

25. Bindea G, Galon J, Mlecnik B. CluePedia Cytoscape plugin: pathway insights using
integrated experimental and in silico data. Bioinformatics. 2013;29:661–3.

26. Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13:2498–504.

27. Nocturne G, Mariette X. B cells in the pathogenesis of primary Sjögren’s syndrome. Nat Rev Rheumatol. 2018;14:133–45.

28. Glauzy S, Sng J, Bannock JM, et al. Defective Early B Cell Tolerance Checkpoints in Sjögren’s Syndrome Patients. Arthritis Rheumatol. 2017;69:2203–8.

29. Harris PE, Malanga D, Liu Z, et al. Effect of interferon alpha on MHC class II gene expression in ex vivo human islet tissue. Biochim Biophys Acta. 2006;1762:627–35.

30. Harley JB, Reichlin M, Arnett FC, et al. Gene interaction at HLA-DQ enhances autoantibody production in primary Sjögren's syndrome. Science. 1986;232:1145–7.

31. Petri A, Dybkær K, Bøgsted M, et al. Long Noncoding RNA Expression during Human B-Cell Development. PLoS One. 2015;10:e0138236.

32. Thierry-Mieg D, Thierry-Mieg J. AceView: a comprehensive cDNA-supported gene and transcripts annotation. Genome Biol. 2006;7(Suppl 1):12.1–14.

33. Fishilevich S, Nudel R, Rappaport N, Hadar R, Plaschkes I, Iny Stein T, Rosen N, Kohn A, Twik M, Safran M, Lancet D, Cohen D. GeneHancer: genome - wide integration of enhancers and target genes in GeneCards Database (Oxford)(2017).

34. McCall MN, Uppal K, Jaffee HA, et al. The Gene Expression Barcode: leveraging public data repositories to begin cataloging the human and murine transcriptomes. Nucleic Acids Res. 2011;39(Database issue):D1011-5.

35. Yu J, Bouska A, Lone W, et al. Long Non Coding RNA Signatures with Diagnostic and Prognostic Significance in Aggressive B-Cell Lymphoma. Blood. 2016;128:1759.

36. Thompson EC, Cobb BS, Sabbattini P, et al. Ikaros DNA-binding proteins as integral
components of B cell developmental-stage-specific regulatory circuits. Immunity. 2007;26:335–44.

37. Mallampati S, Sun B, Lu Y, et al. Integrated genetic approaches identify the molecular mechanisms of Sox4 in early B-cell development: intricate roles for RAG1/2 and CK1ε. Blood. 2014;123:4064–76.

38. Yoshitomi H, Kobayashi S, Miyagawa-Hayashino A, et al. Human Sox4 facilitates the development of CXCL13-producing helper T cells in inflammatory environments. Nat Commun. 2018;9:3762.

39. De Silva NS, Klein U. Dynamics of B cells in germinal centres. Nat Rev Immunol. 2015;15:137–48.

40. Victora GD, Dominguez-Sola D, Holmes AB, et al. Identification of human germinal center light and dark zone cells and their relationship to human B-cell lymphomas. Blood. 2012;120:2240–8.

41. Andrews PD, Knatko E, Moore WJ, et al. Mitotic mechanics: the auroras come into view. Curr Opin Cell Biol. 2003;15:672–83.

42. den Hollander J, Rimpi S, Doherty JR, et al. Aurora kinases A and B are up-regulated by Myc and are essential for maintenance of the malignant state. Blood. 2010;116:1498–505.

43. Li Z, Ivanov AA, Su R, et al. The OncoPPi network of cancer-focused protein-protein interactions to inform biological insights and therapeutic strategies. Nat Commun. 2017;8:14356.

44. Zhang B, Calado DP, Wang Z, et al. An oncogenic role for alternative NF-κB signaling in DLBCL revealed upon deregulated BCL6 expression. Cell Rep. 2015;11:715–26.

45. Mehta GA, Parker JS, Silva GO, et al. Amplification of SOX4 promotes PI3K/Akt signaling in human breast cancer. Breast Cancer Res Treat. 2017;162:439–50.
46. Dominguez-Sola D, et al. The proto-oncogene MYC is required for selection in the germinal center and cyclic reentry. Nature Immunol. 2012;13:1083-91.

47. Calado DP, et al. The cell-cycle regulator c-Myc is essential for the formation and maintenance of germinal centers. Nature Immunol. 2012;13:1092-100.

48. Moreno CS. SOX4: The unappreciated oncogene. Semin Cancer Biol. 2019. pii: S1044-579X(18)30145-7. [Epub ahead of print].

49. Chen JQ, Papp G, Póliska S, et al. MicroRNA expression profiles identify disease-specific alterations in systemic lupus erythematosus and primary Sjögren's syndrome. PLoS One. 2017;12:e0174585.

50. Kim C, Hu B, Jadhav RR, et al. Activation of miR-21-Regulated Pathways in Immune Aging Selects against Signatures Characteristic of Memory T Cells. Cell Rep. 2018;25:2148-62.e5.

51. Kim C, Hu B, Jadhav RR, et al. Activation of miR-21-Regulated Pathways in Immune Aging Selects against Signatures Characteristic of Memory T Cells. Cell Rep. 2018;25:2148-62.e5.

Figures
Gene expression profile of B cell subsets in patients with pSS and HC. (A) Study protocol. (B) Hierarchical clustering analysis and (C) principal component analysis to summarize the dissimilarity of the transcriptome of each B cell subset. Rows correspond to genes and columns to samples. The ellipse shows the 50% confidence interval of the value of principal component analysis. (D) Dysregulated genes are shown in the volcano plot. Horizontal green lines show the cut-off of the p-value indicating a significant difference, and the vertical green lines show a log2-fold change. DEG, differentially expressed gene; GC-B, germinal centre B cell; HC, healthy control; pSS, primary Sjögren's syndrome; WGCNA, weighted gene co-expression network analysis.
Gene signatures of B cell subsets of patients with primary Sjögren’s syndrome. Analysis of enriched immunological GO terms associated with differentially expressed genes of Bm1 (A), naïve B cells (B), pre-germinal centre B cells (C) and memory B cells (D). Each node represents a GO term, and the size of each node represents the significance of the enrichment. (E) Upregulated top five genes in each B cell subset of patients with pSS compared with those of the HC. (F) Dysregulated pathways in each B cell subset with pSS compared with those of the HC. The significant range of Z scores is greater or less than 1.96. GC-B, germinal centre B cell.
Figure 3

Characteristics of LINC00487. (A) Correlation between LINC00487 and each interferon signature gene, with bivariate scatter plots shown below the diagonal, histograms on the diagonal, and the Pearson correlation above the diagonal. The name of row and column were described in the side and top, respectively. ‘****’, p<0.001; ‘**’, p<0.01; ‘*’, p<0.05; Pearson’s correlation test. (B) Scatter plot of disease activity scores and normalized expression levels of LINC00487 in Bm1, naïve, pre-GC and memory subset of patients with pSS. (C-D) qPCR analysis of LINC00487 in primary human CD19+ B cells. Comparison of the expression of LINC00487 between patients with pSS and HC (C) (***, p<0.05; the Mann-Whitney test), and correlation plot of expression of the LINC00487 and interferon induced protein 44-like gene in patients with pSS (D). (E-F) qPCR analysis of B cell lines after treatment with IFNα. B cell lines were treated for 48 h. Results are represented as the mean ± standard deviation. ESSDAI, EULAR Sjögren's Syndrome Disease Activity Index: GC-B, germinal centre B cell; HC, healthy control; pSS, primary Sjögren's syndrome.
Figure 4

Weighted gene co-expression network analysis of B cell subsets of patients with pSS. (A) Primary component analysis using the gene set produced generated using WGCNA. The ellipse shows the 50% confidence interval of the value of principal component analysis. (B) Module eigengene of each B cell subset of patients with pSS. The y-axis displays the values of the module eigengene, and the x-axis displays each cell subset. (C) The heat map shows module-trait associations. Each row corresponds to a module’s first eigengene and columns to a trait. Each cell contains the corresponding correlation (p-value). CRP, C-reactive protein: ESSDAI, EULAR Sjögren’s Syndrome Disease Activity Index: GC-B, germinal centre B cell: Lym, lymphocyte count: SS.A, anti-Sjögren’s syndrome-related antigen A: ME, module
eigengene: SS.B, anti-Sjögren’s syndrome-related antigen B.

**Figure 5**

Gene network characteristics and hub genes in the gene networks of B cells of pSS. (A) Hierarchical clustering analysis with heatmap of enriched pathways (left), upstream regulators (middle) and disease/functions (right). (B) Top five significantly enriched canonical pathways (left), upstream regulators (middle) and disease & biological functions (right) in the modules of pSS. The colour of each frame corresponds to module-colour. (C)
Associations of intra- and inter-modular hub genes in patients with pSS. Node-colour corresponds to module-colour. The pink-coded edge represents the correlation, and the green-coded edge represents an inverse correlation. The width of the edge reflects the absolute weight of a correlation.
Figure 6

Association among SOX4 and GC-B cells-related genes. (A) Relative expression levels of SOX4. ‘†’, p<0.05, between each subset in pSS and HC; ‘∗’, p<0.05, between each subset only in HC; ‘□’, p<0.05, between pSS and HC in each subset; the Mann-Whitney test. (B) Primary component analysis using GC-B cells-related genes. The ellipse shows the 50% confidence interval of the value of principal component analysis. (C) Hierarchical clustering analysis with heatmap using GC-B cells-related genes. Rows correspond to genes and columns to samples. (D) Correlation networks among SOX4 and GC-B cells-related genes in B cells of patients with pSS. The pink-coded edge represents the correlation, and the green-coded edge represents an inverse correlation. The width of the edge reflects the absolute weight of a correlation. “∗”, genes which have significant correlation (p<0.05, Spearman’s correlation test) with SOX4. (E) Scatter plot of expression levels of SOX4 and AURKA in pre-GC (above) and NFkB1 in memory B cells (below) of patients with pSS. GC-B, germinal centre B cell.

Supplementary Files
This is a list of supplementary files associated with this preprint. Click to download.
SupplementaryInformation.pdf