



ChIP-seq analysis found IL21R, a target gene of GTF2I—the susceptibility gene for primary biliary cholangitis in Chinese Han

Ziyan Wu¹ · Haolong Li¹ · Honglin Xu² · Futai Feng² · Fengchun Zhang² · Shulan Zhang² · Li Wang² · Yongzhe Li¹

Received: 22 June 2023 / Accepted: 17 August 2023 / Published online: 15 September 2023
© Asian Pacific Association for the Study of the Liver 2023

Abstract

Aims Aimed to identify a new susceptibility gene associated with primary biliary cholangitis (PBC) in Chinese Han and investigate the possible mechanism of that gene in PBC.

Methods A total of 466 PBC and 694 healthy controls (HC) were included in our study, and genotyping GTF2I gene variants by Sequenom. CD19+B cells were isolated for Chromatin immunoprecipitation sequencing (ChIP-seq). Additionally, MEME-ChIP was utilized to perform searches for known motifs and de novo motif discovery. The GTF2I ChIP-seq of hematopoietic cell line (K562) results were obtained from ENCODE (GSE176987, GSE177691). The Genomic HyperBrowser was used to determine overlap and hierarchical clustering between ours and ENCODE datasets.

Results The frequency of the rs117026326 variant T allele was significantly higher in PBC patients than that in HC (20.26% compared with 13.89%, $P = 1.09E-04$). Furthermore, we observed an elevated proportion of GTF2I binding site located in the upstream and 5' UTR of genes in PBC in comparison with HC. Additionally, an in-depth analysis of IL21R region revealed that GTF2I might bind to the IL21R promoter to regulate the expression of the IL21R, with four peaks of GTF2I binding sites, including three increased binding sites in upstream, one increased binding site in 5' UTR. Motif analysis by MEME-ChIP uncovered five significant motifs. A significant overlap between our ChIP and GSE176987, GSE177691 were found by the Genomic HyperBrowser.

Conclusions Our study confirmed that GTF2I was associated with PBC in Chinese Han. Furthermore, our gene function analysis indicated that IL21R may be the target gene regulated by GTF2I.

Keywords PBC · GTF2I · Susceptibility gene · ChIP-seq · IL21R

Introductions

Primary biliary cholangitis (PBC) was a chronic autoimmune liver disease characterized by T lymphocyte-mediated intrahepatic bile duct injury, causing intrahepatic bile duct epithelial cells that may lead to cholestasis and eventually to cirrhosis and liver failure [1]. PBC was a relatively rare disease with an annual prevalence of 19–402 per million [2]. In China, a 2019 meta-analysis conducted found that PBC had an annual prevalence of 191.18 per million [3]. Ursodeoxycholic acid (UDCA) was one of the primary drugs used for PBC treatment [1]. PBC was a complex disease, caused by multiple microgenes and environmental factors, demonstrating genetic heterogeneity, phenotypic complexity and ethnic differences [4]. The natural history of PBC could be broadly divided into four stages. The first stage was the pre-clinical phase, characterized by AMA positivity but no significant abnormalities in biochemical markers. The second

Ziyan Wu and Haolong Li have equally contributed to this work.

- ✉ Shulan Zhang
shulanpumch@126.com
- ✉ Li Wang
wangli2221@sina.com
- ✉ Yongzhe Li
yongzhelipumch@126.com

- ¹ Department of Clinical Laboratory, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, 1 Shuaifuyuan Hutong, Dongcheng District, Beijing 100730, China
- ² Department of Rheumatology and Clinical Immunology, Key Laboratory of Rheumatology and Clinical Immunology, Ministry of Education, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China

stage was the asymptomatic phase, where biochemical abnormalities were the main manifestation without noticeable clinical symptoms. The third stage was the symptomatic phase, in which patients experienced clinical symptoms such as fatigue and pruritus. The fourth stage was the decompensated phase, during which patients developed clinical manifestations such as gastrointestinal bleeding, ascites, and hepatic encephalopathy [5]. The presence of fatigue and/or pruritus at the beginning of PBC indicated a specific group of patients that were predominantly female, younger in age, had a more aggressive form of the disease, shown less response to UDCA, and had a higher tendency to progress to cirrhosis and other complications [6]. The presence of antimitochondrial antibodies (AMA), which was the characteristic markers of PBC, could be detected in 90–95% PBC patients [7]. Antinuclear antibodies (ANA) were also found in PBC patients. Notably, specific antibodies targeting sp100 (characterized by ANA pattern of multiple nuclear dots) and gp210 (ANA pattern rim-like membrane) had been observed in approximately 20% of all PBC patients and 40%–50% of AMA-negative PBC patients [8]. In addition, antibodies against SS-A/Ro-52 and centromere demonstrated a high specificity for PBC and could help the diagnosis of AMA-negative PBC [9]. Currently, over 20 susceptibility genes for PBC had been identified through various genome wide association studies (GWAS), including HLA-DRB1*0801, HLA-DQB1, IL12A, IL12RB2, TNFSF15, IRF5, SOCS1, ARID3A, CD80, CTLA-4, SPIB, IRF8, PLC-L2, IKZF3 and PRKCB [10–15].

GTF2I was located at chromosome 7q11.23 and encoded a 112 kDa phosphorylated protein—universal transcription factor Iii [16]. The inversions, microdeletions or structural variations of GTF2I had been closely associated with Williams–Beuren Syndrome (WBS) [17]. Our group had demonstrated an association between GTF2I and Sjögren’s syndrome (SS), systemic sclerosis (SSc), systemic lupus erythematosus (SLE) in Chinese Han [18–20]. In addition, the association was also found in rheumatoid arthritis (RA), SLE and SSc [12, 21–23]. Considering that both PBC and SS belonged to autoimmune epithelitis [24], we aimed to investigate the possible association and mechanism of GTF2I in PBC.

Methods

Patients and healthy controls

A consecutive series of 466 patients with PBC and 694 geographically and ethnically matched healthy controls (HC) were recruited at the Peking Union Medical College Hospital. All subjects were Han Chinese and unrelated. Diagnosis of PBC was established based on the 2009 American

Association for the Study of Liver Diseases practice guideline [25, 26]. To establish a diagnosis of PBC, it was necessary for any two of the three criteria mentioned below to be met: positive AMA serologic testing, consistently elevated cholestatic liver biochemistry (specifically increased ALP levels), and/or histological evidence of non-suppurative inflammation in the interlobular bile ducts. The medical records of each patient were reviewed. Patients with other autoimmune diseases including SLE, RA, type 1 diabetes mellitus, autoimmune hepatitis, primary sclerosing cholangitis, overlap syndrome or SS were excluded.

The Ethics Committee of Peking Union Medical College Hospital approved the study protocol (JS-2156).

Genotyping

Genomic DNA was extracted from the blood samples using a Tiangen DNA kit (Tiangen, Beijing, China) and stored at -80°C following standard procedures. The single nucleotide polymorphisms (SNPs) were genotyped using a Sequenom MassARRAY System (Sequenom iPLEX assay, San Diego, CA), in accordance with the manufacturer’s instructions.

DNA samples were then amplified by multiplex PCR, and PCR products were used for locus-specific single-base extension reactions. The final products were desalted and transferred to a 384-element SpectroChIP array. Allele detection was performed by matrix-assisted laser desorption ionization time-of-flight mass spectrometry.

The isolation of CD19+ B cells

Peripheral blood mononuclear cells (PBMCs) from 3 PBC patients and 3 HC were prepared according to standard Ficoll-Hypaque procedures (Pharmacia Biotech, Sweden). Then, the CD19+ B cells were sorted by magnetic bead kits according to the instructions (Becton, Dickinson and Company, USA).

Chromatin immunoprecipitation sequencing (ChIP-seq)

ChIP-seq was performed according to the manufacturer’s instructions. Briefly, cells were cross-linked in a final concentration of 1% formaldehyde followed by glycine quenching. Cells were lysed with lysis buffer (0.2% SDS; 10 mM Tris-HCl, pH 8.0; 10 mM EDTA, pH 8.0; proteinase inhibitor cocktail) and sonicated to fragments of approximately 300–500 bp (Bioruptor, Diagenode). Dynabeads Protein A was washed twice with ChIP buffer (10 mM Tris-HCl pH7.5, 140 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 1% Triton X-100, 0.1% SDS, 0.1% Na-deoxycholate, cocktail proteinase inhibitor) and was incubated with antibody at 4°C for 2–3 h. The fragmented chromatin was transferred

to bead-antibody complex tubes and rotated at 4 °C overnight. The beads were washed once with low salt buffer (10 mM Tris–HCl pH7.5, 250 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 1% Triton X-100, 0.1% SDS, 0.1% Na-deoxycholate, cocktail proteinase inhibitor) and twice with high salt buffer (10 mM Tris–HCl pH7.5, 500 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 1% Triton X-100, 0.1% SDS, 0.1% Na-deoxycholate, Cocktail proteinase inhibitor). After crosslink reversal, DNA was treated with the End Repair/dA-Tailing Module (NEB, E7442L) and Ligation Module (NEB, E7445L) following the user manual. The Chip library was amplified for approximately 11 cycles of PCR using the Q5 master mix (NEB, M0492L). The DNA was size-selected, purified and sequenced on an Illumina sequencing platform.

Statistical analysis

Statistical analysis was performed using the PLINK 1.9 software, GraphPad Prism 5 and R 4.2.3. The Hardy–Weinberg equilibrium (HWE) was assessed by chi-squared tests for each SNP. Any SNP was excluded from subsequent analysis if there were significant deviations from the HWE (i.e., $p < 0.05$) in HC. The Chi-squared test was used to analyze allele and genotype distributions between cases and controls with a significance threshold of 0.05 (two-tailed). The odds ratio (OR) and 95% confidence interval (95% CI) were calculated, and p values were corrected for multiple comparisons by the Bonferroni method ($P_c = P * n$, n was the number of tested SNPs).

Sequencing raw data quality was assessed using FastQC and MultiQC. FASTQ files were aligned to hg19 reference genome using Burrows–Wheeler Alignment tool (BWA). Mapped bam files were used for peak calling by MACS2. To annotate the peaks to the nearest genes, we utilized PAVIS and ChIPseeker. Furthermore, we carried out enrichment analyses of Gene Ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes pathways (KEGG) and genomes using clusterProfiler. Searches for known motifs and de novo motif discovery were performed using MEME–ChIP. The expressions of genes were queried in the Human Cell Atlas (www.humancellatlas.org). Additionally, ChIP-seq results of GTF2I (GSE176987, GSE177691) were searched in ENCODE from hematopoietic cell line (K562) [27]. Using the Genomic HyperBrowser, we determined the overlap and hierarchical clustering between our data and ENCODE datasets. Overlap was determined through segment-segment analysis with either 1,000 or 10,000 Monte-Carlo randomizations while maintaining the empiric distribution of segment and inter-segment lengths but randomizing positions. To perform hierarchical clustering analysis in the Genomic HyperBrowser, we obtained pairwise overlap enrichment values for each samples and computed the distance between samples as the inverse of

these values. Cooperative TFs were found by the Genomic HyperBrowser with sources of potential TFs for “UCSC tfbs conserved” datasets [28]. Finally, the overview of our study was accomplished by bioRender.com (2023).

Results

Clinical features of the participants

Figure 1 depicted the workflow for our study. Additionally, clinical data for each patients were gathered in this study (Table 1). A total of 466 PBC and 694 HC were recruited for SNPs analysis in this study. The average age of the PBC and HC was 55.70 ± 10.84 years and 49.40 ± 11.00 years, respectively. The gender ratio (female/male) in the PBC group was 9.98, while that for the controls was 10.57.

SNP analysis of PBC and HC

Among the 3 SNPs genotyped, SNP rs800879 of NCF1 deviated from the HWE in the controls ($p < 0.05$) and was excluded from further analysis.

The allele and genotype frequency distributions of the remaining 2 SNPs are presented in Table 2 and Supplementary Table 1. The frequency of the rs117026326 variant T allele was significantly higher in PBC patients than that in HC (20.26% compared with 13.89%, OR 1.56, 95% CI 1.27–2.96, $P_c = 1.09E-04$, Table 2). Similarly, the genotype distributions of rs117026326 were significantly different between PBC patients and controls (Supplementary Fig. 1, Supplementary Table 1). Logistic regression analyses using genetic additive, dominant, and recessive models yielded similar results, with the strongest association found in rs117026326 in dominant models (OR 1.64, 95% CI 1.28–2.13, $P_c = 2.31E-04$; Supplementary Table 2). The frequency of the rs73366469 variant T allele was significantly lower in PBC patients than in the controls (77.01% compared with 83.43%, OR 1.49, 95% CI 1.22–1.85, $P_c = 3.22E-04$, Table 2). Likewise, the genotype distributions of rs73366469 were significantly different between PBC patients and controls (Fig. 1, Supplementary Fig. 1, Supplementary Table 1). Logistic regression analyses using genetic additive, dominant, and recessive models yielded similar results, with the strongest association found in rs73366469 in additive models (OR 1.74, 95% CI 1.24–2.44, $P_c = 3.00E-03$; Supplementary Table 2).

ChIP-seq analysis of GTF2I

We observed an elevated proportion of binding sites located upstream and 5' UTR of genes in PBC patients in comparison with HC (11.3% VS 2.3%, 6.1% VS 0.4%,

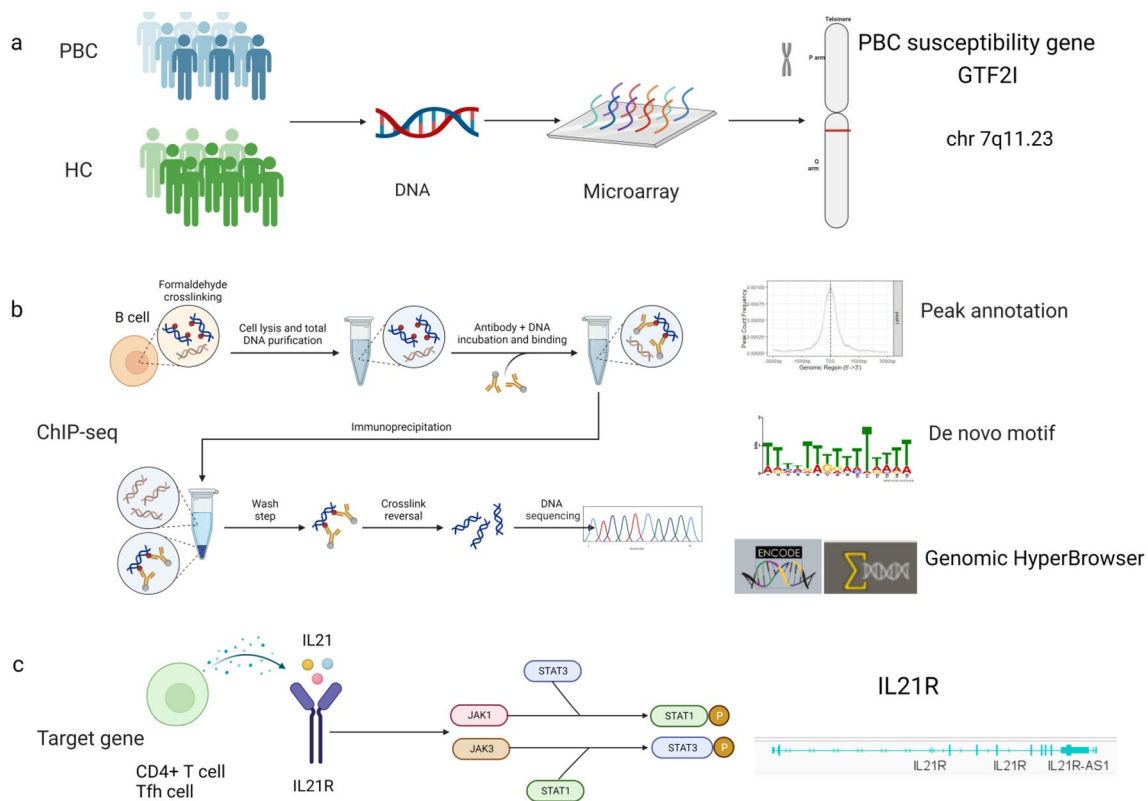


Fig. 1 The overview of our study. **a** The discovery of PBC susceptibility gene—GTF2I. **b** The conduction of the ChIP-seq of GTF2I gene in PBC and HC. **c** The identification of target gene—IL21R

Supplementary Fig. 2). Additionally, difference peaks annotation revealed that 65.34% of the significant increase sites in PBC compared to HC were located in the upstream and promoter regions (Fig. 2). While, for the significant decrease loci in PBC compared to HC, 73.45% of the peaks were located in the distal intergenic region (Supplementary Fig. 3).

In addition, we conducted gene enrichment and KEGG analysis on the genes with the upstream and 5' UTR binding sites, which were significant increase ones in PBC compared to HC. Gene enrichment analysis revealed that DNA-binding transcription factor binding, N-methyltransferase activity, transcription coactivator activity were the top three MF (Molecular Function) pathways for the upstream (Supplementary Fig. 4). As for the 5' UTR, the top three BP (Biological Process) pathways were G1/S transition of mitotic cell cycle, ribonucleoprotein complex biogenesis, regulation of protein stability, while those for CC (Cellular Component) and MF pathways were spliceosomal complex, ubiquitin ligase complex, ubiquitin ligase complex and ubiquitin-like protein transferase activity, ubiquitin-protein transferase activity, ubiquitin protein ligase binding, respectively (Supplementary Fig. 5). KEGG analysis of the upstream identified Epstein-Barr virus infection, lipid and

atherosclerosis and regulation of actin cytoskeleton were the top three significant pathways (Supplementary Fig. 6). Similarly, KEGG analysis demonstrated that the ubiquitin mediated proteolysis was the only significant pathway for the 5' UTR (p value 1.62E-04).

Our search for the overlaps of the significant increase sites of PBC in our ChIP-seq with the gene list of autoimmune disease in NHGRI-EBI Catalog of human genome-wide association studies [29] identified several susceptibility genes, including IL21R for PBC, BLK for SLE, ICAM1, ICOSLG, IFNAR1 for ankylosing spondylitis, psoriasis, ulcerative colitis, Crohn's disease, primary sclerosing cholangitis and MAPK1 for ulcerative colitis, Crohn's disease.

The distribution of our ChIP-seq peaks of the significant increase sites in PBC was throughout the entire genome. In-depth analysis of IL21R region, we found that GTF2I may bind to the IL21R promoter to regulate the IL21R expression, with four peaks of GTF2I binding sites, including three increased binding sites in upstream and one increased binding site in 5' UTR (Supplementary Fig. 7). By searching the Human Cell Atlas database, we found the expressions of GTF2I and IL21R varied among different cells. GTF2I had higher expression in naïve B cells, memory B cells, plasmacytoid dendritic cells than others. IL21R had higher

Table 1 Summary of clinical data for PBC subjects and controls

Characteristics	PBC	HC
Number	466*	694
Mean age ± SD	55.70 ± 10.84	49.40 ± 11.00
Female/male	419/47	634/60
Clinical manifestation (%)		
Fatigue	349 (74.89)	–
Pruritus	224 (48.07)	–
Jaundice	52 (11.16)	–
Poor appetite	107 (22.96)	–
Arthralgia	186 (39.91)	–
Hepatalgia	144 (30.90)	–
Liver biopsy	18 (3.86)	–
Treatments (%)		–
UDCA	466 (100.00)	–
Glucocorticoid	139 (29.83)	–
Immunosuppressant	69 (14.81)	–
Other hepatoprotectants	46 (9.87)	–
Biochemical characteristics		–
ALT U/L	27.00 (19.00–45.75)	–
TP	76.36 ± 7.24	–
ALB g/L	41.57 ± 5.07	–
TBIL µmol/L	12.70 (9.60–18.20)	–
DBIL µmol/L	4.25 (2.90–6.70)	–
GGT U/L	59.00 (31.00–149.30)	–
ALP U/L	117.0 (80.00–183.3)	–
AST U/L	33.00 (25.00–49.00)	–
TBA µmol/L	8.90 (4.28–21.30)	–
Autoantibodies (%)		–
ANA [#]	366 (88.62)	–
AMA [§]	309 (80.68)	–
AMA-M2 [§] RU/mL	554 (162–800)	–
gp210 ^{&}	63 (28.64)	–
sp100 ^{&}	34 (15.60)	–

*19(4.08%) newly diagnosed PBC, 447(95.92%) revisited PBC

[#]Only 413 ANA

[§]383 AMA, 409 AMA-M2

[&]220 gp210, 218 sp100 were available

expression in naïve B cells, non-classic monocyte cells than others (Supplementary Fig. 8).

Motif analysis by MEME-ChIP identified five significant motifs (*E* value ≤ 0.05) (Fig. 3, Table 3). Comparison with JASPAR2022 CORE non-redundant v2 revealed that the most significant motif (motif id:TTHTTKTWTTTWTT, *E*-value 2.20E-90) had 67 matches with known motifs, for instance: CDF5, DOF3.6, SoxC, DOF5.8, FoxM, DOF5.1 and so on. Additionally, the one with motif id:AGRRRGAGA AGGARR (*E*-value 4.70E-050) had 15 matches, including RAMOSA1, BPC1, BPC6, PRDM1, BPC5, Spi1 and Trl.

The genomic HyperBrowser analysis

We found significant overlap between our ChIP-seq peaks and GSE176987, GSE17769 (*p* value 0.020, 0.020, respectively). We performed hierarchical clustering analysis to assess the similarity between samples within the same cell group (in this case, GSE176987 and GSE17769) and ours. Our analysis revealed that there were closer similarities between samples within the same cell group (GSE176987 and GSE17769) than when comparing samples among different groups (Supplementary Fig. 9). We also drew a Venn diagram to illustrate the overlap genes of the upstream and 5' UTR of our ChIP peaks with those of GSE176987, GSE17769 (Supplementary Fig. 9). Among the three datasets, we identified a total of 854 overlap genes, with four autoimmune disease susceptibility genes: IL21R, BLK, ICAM1 and IFNAR1. There were 256 transcription factors (TFs) targeting our regions of interest, using “UCSC tfbs conserved” as source of TF occurrences. The top ten cooperative TFs were PAX5, ELK1, SP1, AP2, PAX4, NRF2, CETS1P54, EGR1, MAZR and NGFIC.

Discussion

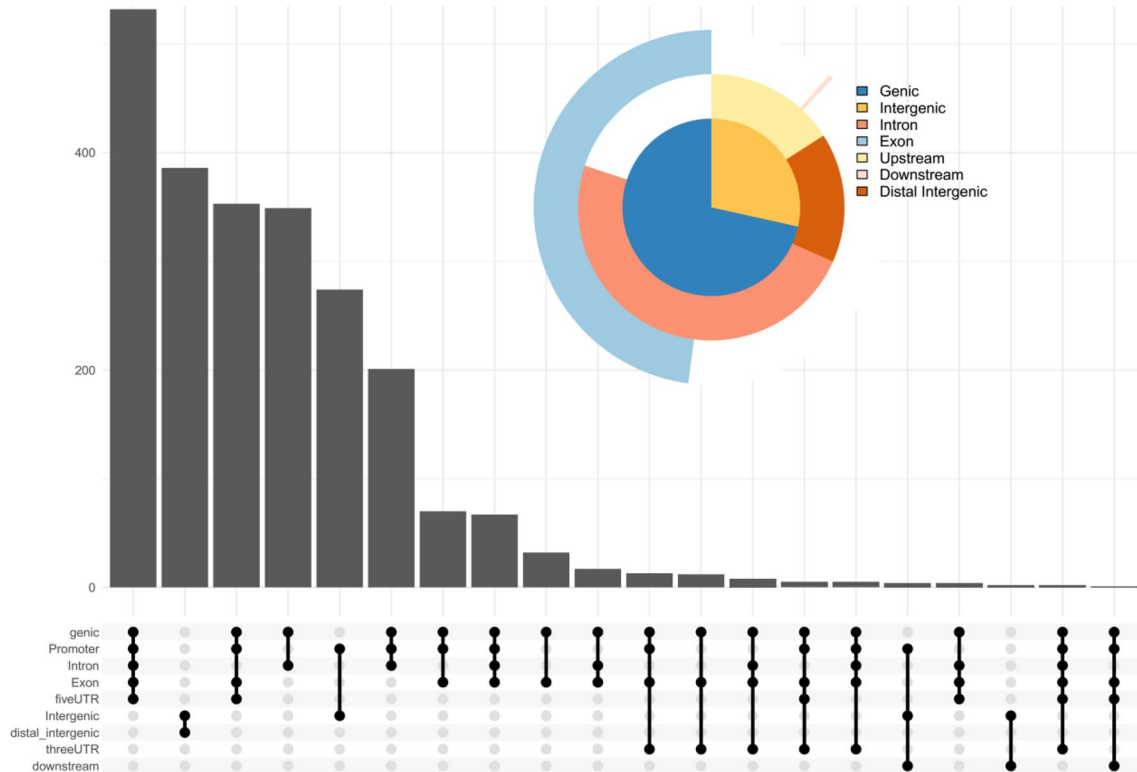
The study represented the initial investigation conducted in Chinese Han to confirm the association between GTF2I and PBC. Furthermore, our analysis of CD19 + B cells from PBC patients via ChIP-seq had revealed that IL21R was the

Table 2 Allele distributions in the subjects

Gene	SNPs	Groups	Allele <i>n</i> , (%)		OR (95% CI)	<i>P</i> _c
			T	C		
GTF2I	rs117026326	PBC	186 (20.26)	732 (79.74)	1.56 (1.27–1.96)	1.09E-04
		HC	192 (13.89)	1190 (86.11)		
GTF2I	rs73366469	PBC	200 (22.99)	670 (77.01)	1.49(1.22–1.85)	3.22E-04
		HC	228 (16.57)	1148 (83.43)		

Table 3 Information of the five de novo motifs discovered by MEME-ChIP

Motif index	Motif id	E value	Matches	Most similar motif
1	TTTHTTKTWTTTWT	2.20E-90	32	CDF5, DOF3.6, SoxC, DOF5.8, FoxM, DOF5.1, DOF3.4, DOF1.5, ZNF384, Fox E, et al
2	AGRRRGAGAAGGARR	4.70E-50	15	RAMOSA1, BPC1, BPC6, PRDM1, BPC5, Spi1, Trl, eor-1, GATA4, SPIB, GATA2, STAT2, et al
3	CATCTGTRAAATGGG	2.10E-43	0	NA
4	GCCGCCGCCGCKCC	5.50E-40	67	LOB, ERF091, ERF104, ERF2, ERF9, LEP, ERF10, ERF5, ERF095, RAP2-6, et al
5	TACTCRGGAGGCTGA	9.00E-08	1	ZNF460

**Fig. 2** Different peaks of significant increase sites in PBC compared with HC annotated by Chip seeker. There were 65.34% of those binding sites located in the upstream and promoter

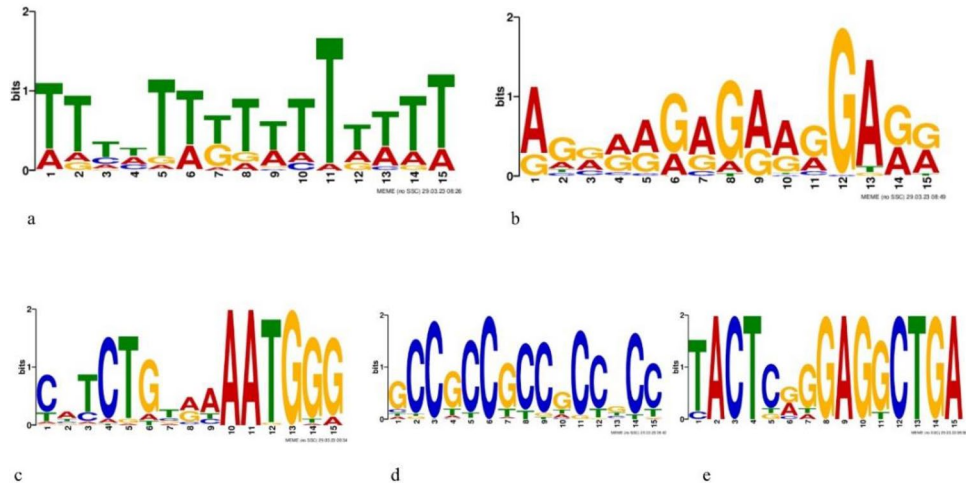
downstream target gene of GTF2I. Notably, it was important to note that our ChIP-seq results were compared to those obtained through the ENCODE project, revealing a significant overlap.

GTF2I was located at a complicated gene region that encompassed multiple CNV and repetitive sequences. Interestingly, the analysis of "the 1000 genome" data revealed that rs117026326 had a very low frequency of C polymorphism among European and African American populations (2%), in comparison to a relatively high rate within East Asian populations (China and Japan) (9%) and the northern Chinese Han (12%). Our previous research had revealed the association of GTF2I and multiple autoimmune disease in

Chinese Han, including SS, SSc and SLE [18–20]. Similar association had also been established between GTF2I and RA, SLE and SSc in East Asian [12, 21–23]. Hence, it appeared that GTF2I maybe a general susceptibility gene for autoimmune diseases in East Asian. Notably, Bruton's tyrosine protein kinase (BTK) could phosphorylate the GTF2I's tyrosine within the B cell receptor signaling (B cell receptor, BCR) pathway. The phosphorylation of GTF2I was transported into the nucleus to regulate the expression of related genes, including immunoglobulin heavy chains [30].

IL21R was a heterodimeric receptor for IL21, a class I cytokine, consisting of the common γ -chain (CD132) and the IL21 specific α -chain. The receptor was expressed on

Fig. 3 Five de novo motifs discovered by MEME-ChIP: **a** motif id: TTTHTTKTWTTTWT, *E*-value 2.20E-90, **b** motif id: AGRRRRGAGAAGGARR, *E*-value 4.70E-50, **c** motif id: CATCTGTRAAATGGG, *E*-value 2.10E-43, **d** motif id: GCCGCCGCCGCKCC, *E*-value 5.50E-40, **e** motif id: TACTCRGGAGGCTGA, *E*-value 9.00E-08



CD4 + T, B, NK and dendritic cells, with activated B cells presenting the highest levels. The binding of IL21 with its receptor IL21R activated multiple downstream signaling molecules, including STAT1 and STAT3, which had a crucial role in the proliferation and differentiation of T cells, B cells and NK cells [31–33]. A significant association was noted between multiple SNPs in the IL21 and IL21R loci with PBC, accompanied with significantly elevated levels of IL21 and IL21R expression within the livers of PBC patients, in contrast to chronic hepatitis B, autoimmune liver hepatitis or healthy controls [9]. Furthermore, the severity of inflammation positively correlated with the numbers of IL21 + and IL21R + cells. The follicular helper T (Tfh) cells with the expression of CXCR5, ICOS, PD-1, Bcl-6 and IL-21, played an important role in regulating germinal center formation or immune responses and B cell activation, contributing to the pathogenesis of PBC [34]. In particular, a study reported that elevations of both serum and intrahepatic IL-21 levels in patients with PBC, which promoted B cell proliferation, STAT3 phosphorylation and AMA production in vitro [35].

In conclusion, GTF2I was identified as the susceptibility gene for PBC in Chinese Han. Subsequent gene function analysis had established that IL21R represented a target gene regulated by GTF2I. Further experiments were required to elucidate the specific pathogenesis of GTF2I and IL21R in PBC.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s12072-023-10586-x>.

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

Funding This research was supported by grants from the National Natural Science Foundation of China Grants (81871302, 81801631),

and the National Key Research and Development Program of China (2018YFE0207300), Beijing Municipal Science & Technology Commission (Z211100002521021), the National High Level Hospital Clinical Research Funding(2022-PUMCH-B-124,2022-PUMCH-A-007).

Data availability The data underlying this article will be shared on reasonable request to the corresponding author.

Declarations

Conflict of interest ZW, HL, HX, FF, FZ, SZ, LW, YL declare no conflicts of interest.

Ethical approval The Ethics Committee of Peking Union Medical College Hospital approved the study protocol (JS-2156).

Informed consent The informed consent to participate in the ChIP-seq study was provided by the enrolled 3 PBC patients and 3 healthy controls or their families.

References

1. Gulamhusein AF, Hirschfield GM. Primary biliary cholangitis: pathogenesis and therapeutic opportunities. *Nat Rev Gastroenterol Hepatol*. 2020;17(2):93–110
2. Liu H, Liu Y, Wang L, et al. Prevalence of primary biliary cirrhosis in adults referring hospital for annual health check-up in Southern China. *BMC Gastroenterol*. 2010;10:100
3. Zeng N, Duan W, Chen S, et al. Epidemiology and clinical course of primary biliary cholangitis in the Asia-Pacific region: a systematic review and meta-analysis. *Hepatol Int*. 2019;13(6):788–799
4. Lleo A, Leung PSC, Hirschfield GM, Gershwin EM. The pathogenesis of primary biliary cholangitis: a comprehensive review. *Semin Liver Dis*. 2020;40(1):34–48
5. Selmi C, Bowlus CL, Gershwin ME, Coppel RL. Primary biliary cirrhosis. *Lancet (London, England)*. 2011;377(9777):1600–1609
6. Quarneti C, Muratori P, Lalanne C, et al. Fatigue and pruritus at onset identify a more aggressive subset of primary biliary cirrhosis. *Liver Int*. 2015;35(2):636–641
7. Colapietro F, Lleo A, Generali E. Antimitochondrial antibodies: from bench to bedside. *Clin Rev Allergy Immunol*. 2022;63(2):166–177

8. Granito A, Muratori P, Muratori L, et al. Antinuclear antibodies giving the ‘multiple nuclear dots’ or the ‘rim-like/membranous’ patterns: diagnostic accuracy for primary biliary cirrhosis. *Aliment Pharmacol Ther.* 2006;24(11–12):1575–1583
9. Granito A, Muratori P, Muratori L, et al. Antibodies to SS-A/Ro-52kD and centromere in autoimmune liver disease: a clue to diagnosis and prognosis of primary biliary cirrhosis. *Aliment Pharmacol Ther.* 2007;26(6):831–838
10. Nakamura M, Nishida N, Kawashima M, et al. Genome-wide association study identifies TNFSF15 and POU2AF1 as susceptibility loci for primary biliary cirrhosis in the Japanese population. *Am J Hum Genet.* 2012;91(4):721–728
11. Cordell HJ, Han Y, Mellis GF, et al. International genome-wide meta-analysis identifies new primary biliary cirrhosis risk loci and targetable pathogenic pathways. *Nat Commun.* 2015;6:8019
12. Sun C, Molineros JE, Looger LL, et al. High-density genotyping of immune-related loci identifies new SLE risk variants in individuals with Asian ancestry. *Nat Genet.* 2016;48(3):323–330
13. Kawashima M, Hitomi Y, Aiba Y, et al. Genome-wide association studies identify PRKCB as a novel genetic susceptibility locus for primary biliary cholangitis in the Japanese population. *Hum Mol Genet.* 2017;26(3):650–659
14. Qiu F, Tang R, Zuo X, et al. A genome-wide association study identifies six novel risk loci for primary biliary cholangitis. *Nat Commun.* 2017;8:14828
15. Wang C, Zheng X, Jiang P, et al. Genome-wide association studies of specific antinuclear autoantibody subphenotypes in primary biliary cholangitis. *Hepatology.* 2019;70(1):294–307
16. Nathany S, Tripathi R, Mehta A. Gene of the month: GTF2I. *J Clin Pathol.* 2021;74(1):1–4
17. Kozel BA, Barak B, Kim CA, et al. Williams syndrome. *Nat Rev Dis Primers.* 2021;7(1):42
18. 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(11):1361–1365
19. Liu C, Yan S, Chen H, et al. Association of GTF2I, NFKB1, and TYK2 regional polymorphisms with systemic sclerosis in a Chinese Han population. *Front Immunol.* 2021;12: 640083
20. Li Y, Li P, Chen S, et al. Association of GTF2I and GTF2IRD1 polymorphisms with systemic lupus erythematosus in a Chinese Han population. *Clin Exp Rheumatol.* 2015;33(5):632–638
21. Zhao J, Ma J, Deng Y, et al. A missense variant in NCF1 is associated with susceptibility to multiple autoimmune diseases. *Nat Genet.* 2017;49(3):433–437
22. Kim K, Bang SY, Ikari K, et al. Association-heterogeneity mapping identifies an Asian-specific association of the GTF2I locus with rheumatoid arthritis. *Sci Rep.* 2016;6:27563
23. Yokoyama N, Kawasaki A, Matsushita T, et al. Association of NCF1 polymorphism with systemic lupus erythematosus and systemic sclerosis but not with ANCA-associated vasculitis in a Japanese population. *Sci Rep.* 2019;9(1):16366
24. Wang L, Zhang FC, Chen H, et al. Connective tissue diseases in primary biliary cirrhosis: a population-based cohort study. *World J Gastroenterol.* 2013;19(31):5131–5137
25. Heathcote EJ. Management of primary biliary cirrhosis. The American association for the study of liver diseases practice guidelines. *Hepatology.* 2000;31(4):1005–1013
26. Lindor KD, Gershwin ME, Poupon R, Kaplan M, Bergasa NV, Heathcote EJ. Primary biliary cirrhosis. *Hepatology.* 2009;50(1):291–308
27. Abascal F, Acosta R, Addleman NJ, et al. Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature.* 2020;583(7818):699–710
28. Sandve GK, Gundersen S, Johansen M, et al. The genomic HyperBrowser: an analysis web server for genome-scale data. *Nucleic Acids Res.* 2013;41(W1):W133–W141
29. Buniello A, MacArthur JAL, Cerezo M, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* 2019;47(D1):D1005–D1012
30. Rajaiya J, Nixon JC, Ayers N, Desgranges ZP, Roy AL, Webb CF. Induction of immunoglobulin heavy-chain transcription through the transcription factor Bright requires TFII-I. *Mol Cell Biol.* 2006;26(12):4758–4768
31. Mehta DS, Wurster AL, Grusby MJ. Biology of IL-21 and the IL-21 receptor. *Immunol Rev.* 2004;202:84–95
32. Long D, Chen Y, Wu H, Zhao M, Lu Q. Clinical significance and immunobiology of IL-21 in autoimmunity. *J Autoimmun.* 2019;99:1–14
33. Ren HM, Lukacher AE, Rahman ZSM, Olsen NJ. New developments implicating IL-21 in autoimmune disease. *J Autoimmun.* 2021;122: 102689
34. Wang L, Sun Y, Zhang Z, et al. CXCR5+ CD4+ T follicular helper cells participate in the pathogenesis of primary biliary cirrhosis. *Hepatology (Baltimore, MD).* 2015;61(2):627–638
35. Li Y, Wang W, Tang L, et al. Chemokine (C-X-C motif) ligand 13 promotes intrahepatic chemokine (C-X-C motif) receptor 5+ lymphocyte homing and aberrant B-cell immune responses in primary biliary cirrhosis. *Hepatology (Baltimore, MD).* 2015;61(6):1998–2007

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.