# RESEARCH

# **Open Access**

# Identification of BTK as an immune-related biomarker for Hashimoto's thyroiditis by integrated bioinformatic analysis



Yingzhao Liu<sup>1†</sup>, Zhangwei Zhu<sup>1†</sup>, Qian Xu<sup>1†</sup>, Juan Xu<sup>2</sup>, Jie Xing<sup>3</sup>, Shengjun Wang<sup>3\*</sup> and Huiyong Peng<sup>3,4\*</sup>

# Abstract

**Background** Hashimoto's thyroiditis (HT) is one of the most common autoimmune disorders characterized by diffuse enlargement of the thyroid gland, lymphocyte infiltration, and thyroid-specific autoantibodies. Cellular and humoral immune disorders have been implicated in the development of HT. However, little is known regarding the role of immune-related molecules in HT. This study was aimed to identify key immune-related biomarkers in HT by using bioinformatic analysis.

**Method** Integration of the sequencing data from HT and normal control (NC) in the GSA and GTEx databases yielded a dataset named NGS. The GSE138198 dataset from the GEO database was downloaded as a validation set. WGCNA analysis was performed to identify key modules associated with HT. Lasso regression analysis (LASSO) and random forest (RF) were performed to determine potential diagnostic biomarkers. The potential value was assessed by using receiver operating characteristic (ROC) curve analysis. CIBERSORT algorithm was used to evaluate the infiltration of immune cells in HT and NC samples. The transcript levels of verified genes from expanded samples were detected by quantitative real-time PCR.

**Results** A total of 1,401 differentially expressed genes (DEGs) were identified in HT patients. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses indicated that these DEGs were predominantly enriched in immune-related pathways. Furthermore, 192 immune-related genes were identified in HT through the intersection of WGCNA modules, DEGs, and the IRGs. Among them, two upregulated genes ((Bruton's tyrosine kinase, BTK) and CD19) showed the potential diagnostic value for HT by using machine learning. The ROC curve analysis revealed that BTK had a higher diagnostic value than CD19 across two datasets. Intriguingly, only BTK expression was upregulated in the peripheral blood mononuclear cells of HT patients, and was significantly positively correlated with the serum levels of thyroid autoantibodies. Further studies confirmed a significant positive correlation between BTK and increased proportions of plasma cells in HT patients.

<sup>†</sup>Yingzhao Liu, Zhangwei Zhu and Qian Xu contributed equally to this work.

\*Correspondence: Shengjun Wang sjwjs@ujs.edu.cn Huiyong Peng penghuiyong33815@163.com

Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

Page 2 of 14

**Conclusion** This study identified BTK was significantly increased in HT patients, which might be the involved in the pathogenesis of HT by regulating plasma cells and represented a potential immune-related biomarker of HT.

**Keywords** Hashimoto's thyroiditis, Immune-related biomarker, Bioinformatics analysis, Machine learning, Immune cells, BTK

# Introduction

Hashimoto's thyroiditis (HT) is an autoimmune thyroid disease characterized by the production of autoantibodies against thyroid gland. The incidence of HT varies by race and gender, with the incidence in females being 5 to 10 times higher than in males [1]. The pathological characteristics of HT are diffuse inflammatory lymphocytic infiltration of the thyroid interstitium, forming tertiary lymphoid structures, which causes varying degrees of thyroid cell damage and atrophy, thyroid interstitial fibrosis, and progressive hypothyroidism [2]. Similar to other autoimmune diseases, HT is generally believed to be the result of genetic, environmental and epigenetic factors, ultimately leading to thyroid microenvironment disorders [1]. However, the pathogenesis of HT is not broadly known.

The diagnosis of HT is based on clinical symptoms of hypothyroidism, elevated levels of anti-thyroglobulin antibody (TgAb) and anti- thyroid peroxidase antibody (TPOAb), high thyroid-stimulating hormone (TSH) levels, and characteristic findings on thyroid ultrasound [3]. However, these clinical manifestations are not specific and may also be present in other thyroid diseases [4]. Currently, levothyroxine monotherapy remains the main treatment for hypothyroidism caused by HT [5]. This treatment is symptomatic and requires lifelong medication. Moreover, excessive or insufficient treatment can increase risk of cardiovascular diseases [6]. To improve the diagnostic accuracy of HT and overcome the limitations of current treatment methods, novel biomarkers were identified, including oxidative stress markers [7], noncoding RNAs [8, 9], and metabolic biomarkers [10]. In recent years, bioinformatics methods have been applied to identified the key genes shared by HT and papillary thyroid carcinoma [11, 12]. However, the study of immune-related biomarkers for HT remains poorly understood.

In this study, we integrated existing high-throughput sequencing data from HT patients and identified the potential immune-related genes by using bioinformatics analysis. Then, the screened genes were further validated by expanding the sample size. Through this approach, we aim to identify the potential biomarkers and provide insights into the development of novel immune therapeutic targets for HT.

# Materials and methods

# Data collection and processing

The HRA001684 dataset was obtained from the GSA-Human database (https://ngdc.cncb.ac.cn/gsa-human/), which is a subset of the GSA database contained human genetic resources [13]. In our analysis, we used the transcriptome sequencing data of 16 HT and 50 normal control (NC) samples from the HRA001684 dataset. The raw data was preprocessed through Trim\_galore, STAR, Picard, and RNA-SeQC tools, including trimming, alignment, sorting, deduplication, and expression calculation along with gene reference and annotation files from GTEx V8 version (https://github.com/broadinstitute/gt ex-pipeline/tree/master/rnaseq). In addition, Qiu et al. screened HT patients from the thyroid pathology atlas in the GTEx database [14]. Based on this, the transcriptome sequencing data of 14 HT and 234 NC were obtained from GTEx V8 version by matching the high-throughput sequencing data in count format. We then used the 'ComBat' function in the 'sva' package to eliminate the batch effect in both sets of sequencing data and created a new dataset named NGS, including 30 HT and 284 NC samples [15]. Further, the validation set was obtained from the NCBI database (GEO ID: GSE138198), which contained the sequencing data of 13 HT and 3 NC samples. A total of 2,843 immune-related genes (IRGs) were obtained from the ImmPort database (https://www.imm port.org/shared/home).

# Identification of differentially expressed genes (DEGs)

The differentially expressed HT-related genes in NGS were determined by using the 'limma' package according to the criteria  $\log_2$  foldchange (FC) > 1 and adjusted p < 0.05 [16]. The FC and adjusted p -value were calculated by the false discovery rate (FDR) procedure and the Benjamini-Hochberg method, respectively [11]. The 'pheatmap' and 'ggplot2' packages were used to draw the heatmap and volcano plots of DEGs, respectively.

# Annotation and functional characterization of DEGs

Gene ontology (GO) term enrichment analysis (http://w ww.geneontology.org) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis (https://www.ge nome.jp/kegg/) were performed to analyze the biological processes and signaling pathways of DEGs by using the 'clusterProfiler' package, respectively [17]. Among them, GO annotation of DEGs were categorized into three categories: biological process (BP), cellular component (CC), and molecular function (MF). Both GO term and KEGG pathway enrichment analyses were performed with a threshold of adjusted p < 0.05.

# Weighted gene co-expression network analysis (WGCNA)

To identify the highly correlated gene clusters, the 'WGCNA' package was used to generate a signed coexpression network cluster for filtering NGS dataset [18]. Firstly, the 'pickSoftThreshold' function was performed to calculate an optimal soft-thresholding parameter to achieve a scale-free network with an R<sup>2</sup> value close to 0.90. Next, hierarchical clustering analysis was performed to transform the adjacency matrix into the topological overlap matrix (TOM) and calculate the corresponding dissimilarity. Genes with similar expression profiles were classified into the same gene modules by using the Dynamic Tree Cut algorithm with a minimum module size of 50 and merge CutHeight of 0.25. Finally, the HT and NC phenotypes were correlated with the modules, and the key modules were significantly associated with HT.

# Screening potential biomarkers by machine learning

To identify the candidate genes, we used 'VennDiagram' package to indicate the intersection of DEGs, IRGs, and key genes associated with HT [19]. Subsequently, two machine learning algorithms (LASSO and RF) were used to identify the potential biomarkers. LASSO analysis was performed to select important variables in high-dimensional data by using the 'glmnet' package with penalty parameter and 10-fold cross-validation [20]. In order to determine the optimal number of variables, the 'Random-Forest' package was used to calculate the average error rate of candidate genes, where the number of trees was determined based on the lowest error rate. Then, genes with feature importance scores were selected greater than 2 [20]. Finally, the intersection genes of LASSO and RF algorithms were identified as the potential biomarkers of HT.

 Table 1
 Clinical characteristics of HT patients and normal controls

Variables	нт	NC	Range	P-value
Number	24	20	-	-
Gender (M/F)	9/15	7/13	-	-
Age (year)	53±16	46±12	-	0.126
FT3 (pmol/liter)	$4.83 \pm 0.76$	$4.95\pm0.56$	3.28-6.47	0.538
FT4 (pmol/ liter)	$9.36 \pm 2.67$	$9.66 \pm 1.28$	7.64–16.03	0.646
TSH (uIU/ml)	$8.90 \pm 12.74$	$2.27 \pm 1.13$	0.56-5.91	0.024
TgAb (IU/ml)	$244.55 \pm 299.99$	$0.20\pm0.20$	0–4	< 0.001
TPOAb (IU/ml)	888.89±1746.14	$0.97 \pm 1.03$	0–9	< 0.001

HT: hashimoto's thyroiditis; NC: normal control; M: male; F: female; FT3, free triiodothyronine; FT4, free thyroxine; TSH, thyroid-stimulating hormone; TgAb, anti-thyroglobulin antibody; TPOAb, anti-thyroid peroxidase antibody. Data correspond to the arithmetic mean ±SD

# Subjects and samples for validation of selected mRNA

A total of 24 adult HT patients, including 15 females and 9 males, and 20 adult age- and sex-matched NC, including 13 females and 7 males, were enrolled in the Affiliated People's Hospital of Jiangsu University, which were used to verify the screened genes. All HT patients were referred to the department of endocrinology, and the diagnosis was based on clinical manifestations and auxiliary examinations. The clinical features of subjects were shown in Table 1. All healthy volunteers were from the physical examination center. Individuals with tumors, allergies, infectious diseases, acute or chronic visceral diseases, or other autoimmune diseases were excluded.

The serum samples were collected by centrifugation at 2,000 rpm for 10 min, and then thyroid function was determined by an LDX-800 system (Beckman Coulter, CA, USA), including free triiodothyronine (FT3), free thyroxine (FT4), thyrotropin (TSH), TgAb, and TPOAb.

# Quantitative real-time PCR (qRT-PCR)

Fresh human peripheral blood was collected by using an EDTA-K<sup>2</sup> anticoagulant tube (Becton Dickinson, Sparks, USA). Then, the peripheral blood mononuclear cells (PBMCs) were separated by lymphocyte separation medium (Tianjin Haoyang Biological Technology Co., Tianjin, China) according to the manufacturer's instructions. The cDNA reverse transcription and qRT-PCR were performed as previously described [21]. The sequences of primers were as follows: Bruton's tyrosine kinase (BTK), sense 5' - GCTCAAAAACGTAATCCGG TACA - 3'; antisense 5' - GTCTTCCGGTGAGAACTC CC - 3'; CD19, sense 5' - CCCAAGGGGCCTAAGTCA TTG - 3'; antisense 5' - AACAGACCCGTCTCCATTA CC – 3'; β actin, sense 5' – CCTGGCACCCAGCACAA T - 3'; antisense 5' - GGGCCGGACTCGTCATAC - 3'; The relative expression of BTK and CD19 mRNA were normalized to  $\beta$  actin.

# Receiver operating characteristic (ROC) curves analysis

ROC curves and the area under the curve (AUC) values were used to evaluate the diagnostic efficacy of BTK and CD19 in the NGS and GSE138198 datasets, which were performed by using the 'pROC' package [22]. The Y-axis represented the true positive rate, indicated by sensitivity. The X-axis represented the false positive rate, indicated by 100%-specificity%.

# Immune cells infiltration analysis

The CIBERSORT algorithm immune cell LM22 gene set was used to evaluate the infiltration of 22 immune cells in HT and NC samples [23]. The Wilcoxon test was used to compare the proportions of immune cells between HT and NC samples. The spearman correlation was performed to determine the correlations between the potential biomarkers and immune cells.

# Statistical analysis

The unpaired Student's t-test was used to analyze significant differences in the relative expression of BTK and CD19 mRNA between HT and NC when variables met the normal distribution criteria. The relationships between BTK and CD19 mRNA levels in the PBMCs and clinical indexes were performed by using Pearson's correlation coefficient. A *p* value less than 0.05 was considered statistically significant (\*p<0.05, \*\*p<0.01, \*\*\*p<0.001). NS: no significance.

# Results

# Data processing and identification of DEGs

The process of the study was depicted in Fig. 1. To identify the DEGs, we merged the raw expression matrices of the HRA001684 dataset and the GTEx thyroid dataset, and filtered them using annotation files to retain mRNA genes. Then, we filtered out low-quality expressed genes by removing those with a sample proportion of zero expression greater than 0.3 in both the HRA001684 and GTEx thyroid datasets, and then performed a  $log_2$ transformation on the raw gene counts, adding 1 to each count to generate  $log_2$  (counts + 1) values. We performed batch correction and differential analysis using the "sva" and "limma" package, respectively (Fig. 2A, B). In total,





Fig. 2 Data processing and identification of DEGs. (**A**, **B**) The raw expression matrices of HRA001684 and merged GTEx dataset before and after batch correction, respectively. The volcano plots (**C**) and heatmaps (**D**) were used to distinguish the DEGs in terms of fold change. Red and blue represented upregulated and downregulated genes, respectively

15,426 DEGs were determined between HT patients and controls. The volcano plots (Fig. 2C) and hierarchical clustering analysis (Fig. 2D) were performed to identified significantly dysregulated genes, including 1,055 significantly upregulated genes and 346 significantly down-regulated genes. These results were used for preliminary analysis of the sequencing data.

# GO and KEGG analysis of DEGs

GO enrichment analysis was performed to determine the potential biological functions of DEGs. The top 5 GO enrichment terms were presented in Fig. 3A. Among them, the significantly enriched GO terms of DEGs were 'T cell activation' in BP, 'extracellular side of the plasma membrane' in CC, and 'immune receptor activity' in MF. According to KEGG pathway enrichment analysis, a total of 24 signaling pathways were significantly associated with DEGs. The top 10 KEGG pathways were shown in



Fig. 3 GO and KEGG enrichment analysis of DEGs: (A) The top 5 enriched terms of BP, CC, and MF in the GO enrichment analysis. (B) The top 10 significantly KEGG pathways of upregulated genes (left) and all genes (right), respectively. (C) Hierarchical enrichment pathway network plots showing all 24 enriched KEGG pathways of DEGs, which are divided into four major categories: cellular processes, organismal systems, environmental information processing, and human diseases. The size of each circle indicated the number of genes enriched in each pathway, and the larger the circle, the more genes were enriched

Fig. 3B. The KEGG network plots showed all enriched 24 KEGG pathways of DEGs (Fig. 3C). These 24 pathways were subsequently into categorized four major classes: cellular processes, organismal systems, environmental information processing, and human diseases. The organismal systems category had the most enriched KEGG pathways (Fig. 3C). These data showed that the DEGs were closely associate with HT.

# Identification of immune-related genes in HT

We performed WGCNA analysis on genes with adjusted p < 0.05 in the matrix. To ensure that the interactions between the genes maximally conformed to a scale-free distribution, the optimal soft threshold ( $R^2 = 0.90$ ) was selected as 10 (Fig. 4A). A total of 27 modules were determined based on a gene clustering tree and a Dynamic Tree Cut algorithm (Fig. 4B, C). Then, we analyzed the correlation between modules and traits (HT and NC) and found that the blue module with 1,454 genes had the highest correlation (r = 0.85). Therefore, the 1,454 genes were considered to be associated with HT. To further identify the core immune-related genes in HT, a Venn diagram was performed to indicate the intersection of the genes in the blue module, DEGs, and IRGs. Based on these, 192 immune-related genes were characterized (Fig. 4D).

# Machine learning for the selection of immune-related biomarkers

Two machine learning algorithms (LASSO and RF) were applied to screen feature genes from 192 immune-related genes in HT patients. LASSO analysis showed 32 nonzero coefficient feature genes related to HT (Fig. 5A, B), while RF analysis identified 15 feature genes with relative importance greater than 2 (Fig. 5C, D). A Venn diagram analysis revealed that two shared genes (BTK and CD19) meet the requirement for potential immune-related biomarkers in HT (Fig. 5E).

# ROC curve analysis of the selected genes

To evaluate the potential diagnostic value of BTK and CD19, we performed ROC curve analysis for BTK and CD19 in the NGS dataset and GSE138198 dataset, respectively. The AUCs of BTK and CD19 in the NGS dataset were 0.98 and 0.99, respectively (Fig. 6A, B). While the AUCs of BTK and CD19 were 0.85 and 0.77, respectively (Fig. 6C, D). In addition, we found that BTK and CD19 were significantly upregulated in HT patients in both NGS and GSE138198 datasets (Fig. 6E). These data suggested that BTK and CD19 might distinguish HT patients from normal controls, and BTK had a higher diagnostic value than CD19.



Fig. 4 Identification of immune-related genes. (A) Visualization of the scale-free fit index (left) and the average network connectivity (right) for choosing various soft-thresholding powers. (B) The module-trait relationships chart showing the correlations between modules and traits (HT and NC). (C) Module dendrogram with different colors representing different modules, and each module contains a group of genes with similar expression. (D) Venn diagram showing the intersection of the genes in the blue module, DEGs, and IRGs



Fig. 5 Machine learning algorithms for 192 immune-related genes. (**A**, **B**) LASSO regression coefficient distribution path and cross-validation curves. (**C**, **D**) Error rate curves of RF algorithm at different numbers of trees, and genes with relative importance greater than 2. (**E**) Venn diagram of genes selected by LASSO and RF algorithms



Fig. 6 ROC curve analysis of BTK and CD19. (A, B) ROC curves of BTK and CD19 in NGS dataset. (C, D) ROC curves of BTK and CD19 in GSE138198 dataset. (E) Violin plots of relative expression of BTK mRNA and CD19 mRNA. \*p < 0.05, \*\*\*p < 0.001

# Validation of the selected genes

To further confirm the results of bioinformatic analysis, we verified the relative expression of BTK mRNA and CD19 mRNA by using qRT-PCR with an expanded sample size. Our data revealed that the transcript levels of BTK were significantly increased in the PBMCs of HT patients compared with that in healthy controls (Fig. 7A). However, there was no significant difference in the relative expression of CD19 mRNA between HT and NC groups (Fig. 7B). Then, we also analyzed the ROC curves of BTK and CD19. The ROC curve showed that BTK could distinguish the HT group from the NC group with the AUC was 0.69 (Fig. S1A). However, CD19 could not distinguish the HT group from the NC group, and the AUC was 0.51 (Fig. S1B). These data showed that the results of qRT-PCR validation were inconsistent with the results of dataset analysis, and BTK might be as the potential immune-related biomarkers of HT.

# The relationship between BTK expression and clinical parameters

To analyze the relationship between BTK and HT, we subsequently assessed the correlation between BTK expression and the clinical parameters of HT. Our results revealed that transcript levels of BTK were significantly positively correlated with the serum levels of TgAb (r=0.4241; p=0.0389) (Fig. 8D) and TPOAb (r=0.5810; p=0.0029) (Fig. 8E), but not with the serum levels of FT3 (r = -0.0849; p=0.6932) (Fig. 8A), FT4 (r=0.0686; p=0.7500) (Fig. 8B), and TSH (r = -0.0309; p=0.8861) (Fig. 8C). These data suggested that increased BTK expression was associated with the autoantibody levels of HT patients.

# Correlation analysis between BTK and immune cells

To investigate the potential biological function of BTK, we first analyzed the infiltration of immune cells in HT. CIBERSORT algorithm was used to determine the proportions of 22 immune cells in NGS dataset. Wilcox tests were used to compare the proportions of immune cells between HT and NC groups. We found that the proportions of immature B cells, plasma cells, memory B cells, M1 macrophages, CD8<sup>+</sup> T cells, activated memory CD4<sup>+</sup> T cells, T follicular helper cells, and resting mast cells were significantly increased in HT, while the proportions of eosinophils, M0 macrophages, monocytes, activated NK cells, and immature CD4<sup>+</sup> T cells were significantly decreased compared with healthy individuals (Fig. 9B). Moreover, M0 macrophages showed the most significant negative correlation with M2 macrophages, while activated NK cells showed the most significant positive correlation with resting mast cells (Fig. 9A). These results showed that there was an imbalance of immune cells in HT.

To further investigate the relationship between BTK and immune cells, we performed the correlation analysis between BTK expression and the proportions of immune cells in the HT group. The data showed that BTK expression was positively correlated with M2 macrophages, plasma cells, and M1 macrophages, and negatively correlated with M0 macrophages, activated NK cells, and eosinophils (Fig. 10). These results showed that BTK might be involved in the pathogenesis of HT by regulating immune cells.

# Discussion



As a common organ-specific autoimmune thyroid disease, the laboratory diagnostic indicators of HT are easily confused with other thyroid diseases [24, 25]. Moreover, the underlying mechanisms of HT are not

**Fig. 7** Validation of the transcript levels of BTK and CD19. (**A**) The relative expression of BTK mRNA in the PBMCs from HT patients (n = 24) and normal controls (n = 20). (**B**) The relative expression of CD19 mRNA in the PBMCs from HT patients (n = 24) and normal controls (n = 20). Each data point represents an individual subject, and the horizontal lines show the mean. \*p < 0.05, NS: no significance



Fig. 8 The relationship between BTK expression and clinical parameters. Pearson's correlation analysis was performed to assess the correlation between the relative expression of BTK and the serum levels of FT3 (A), FT4 (B), TSH (C), TgAb (D), and TPOAb (E). Each data point represents an individual subject

broadly known. Hence, identifying novel checkpoints and investigating their roles are essential for finding new diagnostic and therapeutic targets for HT. Although previous studies have reported a number of potential biomarkers associated with HT (e.g. noncoding RNA), the methods of these studies were single and ROC curve analysis was only performed on screened molecules [8, 26, 27]. In the present study, a total of 1,401 observably dysregulated genes were identified between HT group and NC group by using the public databases. Unlike the previous analysis of individual database, we performed the same consistent data processing pipeline to integrate the original sequencing data from different databases into a new dataset. The benefit of this approach was to ensure the consistency in gene annotation, filtering, and counting methods between the two databases. Moreover, we analyzed the potential biological functions of these DEGs. According to GO enrichment analysis of these DEGs, we found that the DEGs were mainly related to immune cells differentiation, activation, and interaction, which may be involved in the process of HT. Among the 24 signaling pathways analyzed by KEGG enrichment, the 'JAK-STAT signaling pathway', T cell receptor signaling pathway', and 'B cell receptor signaling pathway' have been reported to be involved in the immune disorder of HT [9, 24]. Intriguingly, there was no statistical significance in downregulated genes enriched signaling pathways. A possible explanation for this phenomenon is that few downregulated genes have been identified in HT patients. These findings demonstrated that the pathogenesis of HT may be primarily driven by the upregulated genes.

To date, there were no specific markers for the diagnosis of HT. The role of immune-related biomarkers in immune-related adverse events has recently come into the spotlight with the publication of a number of studies in the past few years [28, 29]. To further investigate the immune-related biomarkers of HT, we used WGCNA analysis to determine the module most relevant to HT patients and focused on analyzing the genes overlapping with DEGs and immune-related genes in that module. Then, two machine learning methods were performed to analyze these overlapping genes, ultimately identifying two potential biomarkers, BTK and CD19, that were associated with HT. Bioinformatics analysis showed that BTK and CD19 were significantly upregulated in HT group. ROC diagnostic curves revealed that these two biomarkers have high diagnostic efficacy for HT diagnosis. Intriguingly, there was only BTK was significantly upregulated in HT patients with the potential diagnosis valuable showing in ROC curve analysis in the subsequent validation experiments. One possible explanation



B

group 🖲 HT 💽 NC



**Fig. 9** Immune infiltration analysis of NGS dataset. (**A**) Radar plots showing the proportions of 22 immune cells in HT and NC groups, with the size of outer circle indicating the magnitude of the differences. Orange and blue represented significantly increased and decreased immune cells in HT group, respectively (p < 0.05). The inner curves showing the correlation between immune cells with significant positive correlation in red and significant negative correlation in blue. (**B**) Violin plots showing the proportions of 22 immune cells in HT and NC groups from NGS dataset



Fig. 10 Correlation analysis between BTK expression and immune cells in the HT group from NGS dataset

for the inconsistent data between validation and bioinformatics analysis is that the samples were sourced from different parts of the body. The samples for bioinformatics analysis were derived from thyroid tissues, while the samples for validation were derived from PBMCs. Another possible explanation for this phenomenon is that the sample size included for verification might be insufficient. There data suggested that elevated BTK expression might be as a potential biomarker of HT.

BTK, belonged to the Tec family of protein tyrosine kinases, is a non-receptor cytoplasmic tyrosine kinase. It is expressed in hematopoietic cells excluded T cells and natural killer cells [30]. As a key enzyme of B-cell receptor signaling, BTK activation is essential for B-cell proliferation, survival, differentiation, and trafficking [31, 32]. Meanwhile, BTK can induce the polarization of M1 macrophages by responding to bacterial lipopolysaccharide stimulation [33, 34]. In addition, BTK inhibitors have been used in the treatment of B -cell malignancies [35]. However, the potential role of BTK in HT remains enigmatic. As expected, the infiltration proportions of immature B cells, plasma cells, memory B cells, and M1

macrophages were significantly increased in HT patients by CIBERSORT algorithm. Combining the relationship between BTK expression and the proportions of immune cells, BTK was associated with the imbalance of plasma cells and M1 macrophages in HT. Taken together, these findings suggest that BTK expression may be associated with an imbalance of plasma cells and M1 macrophages in HT.

Given the potential role of BTK in HT, we analyzed the relationship between BTK and thyroid function parameters. Interestingly, the transcript levels of BTK were positively correlated with the serum levels of TgAb and TPOAb, which are produced by plasma cells in response to antigen stimulation. Furthermore, BTK kinase inhibitors have shown efficacy in clinical trials for various autoimmune diseases, supporting the notion that BTK plays a pivotal role in mediating pathogenic processes [36]. The effects of BTK inhibition in lupus and rheumatoid arthritis suggested that it may be a good target for controlling autoreactive B cells [37]. Together, we speculate that BTK may play a role in the pathogenesis of HT by influencing plasma cell activity. However, there are some

limitations in the study. The sample size for verifying the levels of BTK and CD19 was too small. HT patients are more common in Whites and Asians than in African Americans [38]. However, the verified samples in the study were conducted among the Chinese Han population, which could not reflect all ethnic groups. In addition, we only conducted a preliminary analysis of the HT datasets to investigate the relationship between BTK and plasma cells. The detailed mechanisms need to be further investigated with large cohorts of different regions, race and ethnicity and in vitro and in vivo experiments.

# Conclusion

Collectively, we integrated existing publicly available transcriptomic data of HT and performed various bioinformatic methods to identify immune-related molecular markers of HT. BTK was ultimately screened as a potential immune-related biomarker of HT, which might be involved in the pathogenesis of HT by regulating plasma cells.

# **Supplementary Information**

The online version contains supplementary material available at https://doi.or g/10.1186/s12865-025-00691-x.

Supplementary Material 1

# Acknowledgements

The authors thank all the patients who were involved in the study.

# Author contributions

Y.L. and Z.Z. wrote code, analyzed data, and wrote manuscript; Q.X. helped with analyzed data; J.X. participated in the design of experiments. J.X. conducted the project guidance. S.W. and H.P. were responsible for the conceptualization, funding acquisition, project administration, and supervision for all the work on this manuscript. All authors discussed the results and agreed to publish the manuscript.

## Funding

This work was supported by Jiangsu Provincial Medical Key Discipline Cultivation Unit (Grant No. JSDW202241), the Research Project of Jiangsu Commission of Health (Grant No. ZD2021049, H2023053), Zhenjiang science and technology planning project (Grant No. SH2023006, SH2023008).

#### Data availability

The datasets generated and/or analysed during the current study are available in the GSA-Human database/HRA001684 (https://ngdc.cncb.ac.cn/gsa-huma n/browse/HRA001684), GTEx database/phs000424. vpn. pn (https://www.gtex portal.org/home/downloads/adult-gtex/bulk\_tissue\_expression), and GEO/G SE138198 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE138198) repository. All the raw data and source code are available. The link: https://ww wjjanguoyun.com/p/DeJocHYQ3OLUCRj7p6gFIAA.

# Declarations

# Ethics approval and consent to participate

The Ethics Committee of Zhenjiang First People's Hospital (No. K-20200012-Y) authorized this study. All volunteers provided written informed consent in the Ethics approval and consent to participate section. All methods were conducted in accordance with the ethical standards of the Declaration of Helsinki and relevant guidelines and regulations.

## **Consent for publication**

Not applicable.

#### **Competing interests**

The authors declare no competing interests.

#### Author details

<sup>1</sup>Department of Endocrinology, The Affiliated People's Hospital of Jiangsu University, Zhenjiang Medical School of Nanjing Medical University, Zhenjiang 212002, China

<sup>2</sup>Department of Critical Care Medicine, The Affiliated People's Hospital of Jiangsu University, Zhenjiang Medical School of Nanjing Medical University, Zhenjiang 212002, China

<sup>3</sup>Department of Laboratory Medicine, The Affiliated People's Hospital of Jiangsu University, Zhenjiang Medical School of Nanjing Medical University, Zhenjiang 212002, China

<sup>4</sup>Department of Genetic Toxicology, The Key Laboratory of Modern Toxicology of Ministry of Education, Center for Global Health, School of Public Health, Nanjing Medical University, Nanjing 211100, China

# Received: 25 December 2023 / Accepted: 18 February 2025 Published online: 28 February 2025

## References

- Tywanek E, Michalak A, Swirska J, Zwolak A, Autoimmunity. New Potential Biomarkers and the Thyroid Gland-The Perspective of Hashimoto's Thyroiditis and Its Treatment. Int J Mol Sci 2024; 25(9).
- 2. Zhang QY, Ye XP, Zhou Z, Zhu CF, Li R, Fang Y, et al. Lymphocyte infiltration and thyrocyte destruction are driven by stromal and immune cell components in Hashimoto's thyroiditis. Nat Commun. 2022;13(1):775.
- Mao L, Zheng C, Ou S, He Y, Liao C, Deng G. Influence of Hashimoto thyroiditis on diagnosis and treatment of thyroid nodules. Front Endocrinol (Lausanne). 2022;13:1067390.
- Cui Z, Wang Z, Liu X, Cai Y, Xu X, Yang T. Establishment of clinical diagnosis model of graves' disease and Hashimoto's thyroiditis. J Transl Med. 2019;17(1):11.
- Taylor PN, Medici MM, Hubalewska-Dydejczyk A, Boelaert K, Hypothyroidism. Lancet. 2024;404(10460):1347–64.
- Klubo-Gwiezdzinska J, Wartofsky L. Hashimoto thyroiditis: an evidence-based guide to etiology, diagnosis and treatment. Pol Archives Intern Med 2022; 132(3).
- Ruggeri RM, Giovinazzo S, Barbalace MC, Cristani M, Alibrandi A, Vicchio TM, et al. Influence of dietary habits on oxidative stress markers in Hashimoto's thyroiditis. Thyroid. 2021;31(1):96–105.
- Martinez-Hernandez R, Marazuela M. MicroRNAs in autoimmune thyroid diseases and their role as biomarkers. Best Pract Res Clin Endocrinol Metab. 2023;37(2):101741.
- Peng H, Xiong S, Ding X, Tang X, Wang X, Wang L, et al. Long non-coding RNA expression profiles identify IncRNA-XLOC\_12\_006631 as a potential novel blood biomarker for Hashimoto's thyroiditis. Int J Mol Med. 2020;46(6):2172–84.
- Sarandi E, Kruger Krasagakis S, Tsoukalas D, Rudofsky G, Tsatsakis A. A clinical trial for the identification of metabolic biomarkers in Hashimoto's thyroiditis and in psoriasis: study protocol. Pathophysiology. 2021;28(2):291–306.
- Liu TT, Yin DT, Wang N, Li N, Dong G, Peng MF. Identifying and analyzing the key genes shared by papillary thyroid carcinoma and Hashimoto's thyroiditis using bioinformatics methods. Front Endocrinol (Lausanne). 2023;14:1140094.
- Liu C, Pan Y, Li Q, Zhang Y. Bioinformatics analysis identified shared differentially expressed genes as potential biomarkers for Hashimoto's thyroiditisrelated papillary thyroid cancer. Int J Med Sci. 2021;18(15):3478–87.
- Chen T, Chen X, Zhang S, Zhu J, Tang B, Wang A, et al. Genom Proteom Bioinform. 2021;19(4):578–83. The Genome Sequence Archive Family: Toward Explosive Data Growth and Diverse Data Types.
- Álvarez-Sierra D, Marín-Sánchez A, Gómez-Brey A, Bello I, Caubet E, Moreno-Llorente P, et al. Lymphocytic thyroiditis transcriptomic profiles support the role of checkpoint pathways and B cells in pathogenesis. Thyroid: Official J Am Thyroid Association. 2022;32(6):682–93.

- Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The Sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinf (Oxford England). 2012;28(6):882–3.
- Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7):e47.
- Huang S, Cai S, Ling L, Zhang W, Xiao H, Yu D, et al. Investigating the molecular mechanism of traditional Chinese medicine for the treatment of placental syndromes by influencing inflammatory cytokines using the Mendelian randomization and molecular Docking technology. Front Endocrinol (Lausanne). 2023;14:1290766.
- Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9:559.
- Chen H, Boutros PC. VennDiagram: a package for the generation of highlycustomizable Venn and Euler diagrams in R. BMC Bioinformatics. 2011;12:35.
- Li J, Zhang Y, Lu T, Liang R, Wu Z, Liu M, et al. Identification of diagnostic genes for both Alzheimer's disease and metabolic syndrome by the machine learning algorithm. Front Immunol. 2022;13:1037318.
- Peng H, Xing J, Wang X, Ding X, Tang X, Zou J, et al. Circular RNA circNUP214 modulates the T helper 17 cell response in patients with rheumatoid arthritis. Front Immunol. 2022;13:885896.
- 22. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, et al. pROC: an open-source package for R and S + to analyze and compare ROC curves. BMC Bioinformatics. 2011;12:77.
- Chen B, Khodadoust MS, Liu CL, Newman AM, Alizadeh AA. Profiling tumor infiltrating immune cells with CIBERSORT. Methods Mol Biology (Clifton NJ). 2018;1711:243–59.
- 24. Zheng H, Xu J, Chu Y, Jiang W, Yao W, Mo S, et al. A global regulatory network for dysregulated gene expression and abnormal metabolic signaling in immune cells in the microenvironment of graves' disease and Hashimoto's thyroiditis. Front Immunol. 2022;13:879824.
- Ralli M, Angeletti D, Fiore M, D'Aguanno V, Lambiase A, Artico M, et al. Hashimoto's thyroiditis: an update on pathogenic mechanisms, diagnostic protocols, therapeutic strategies, and potential malignant transformation. Autoimmun Rev. 2020;19(10):102649.
- Luo J, Liu T, Teng W. LncRNA profile in Hashimoto's thyroiditis and potential function of NONHSAT079547.2. J Mol Endocrinol. 2020;64(4):259–70.
- 27. Peng H, Ding X, Xu J, Han Y, Yang J, Tang X et al. Elevated Expression of the Long Noncoding RNA MAFTRR in Patients with Hashimoto's Thyroiditis. *J Immunol Res* 2021; 2021:3577011.
- Meng XW, Cheng ZL, Lu ZY, Tan YN, Jia XY, Zhang M. MX2: identification and systematic mechanistic analysis of a novel immune-related biomarker for systemic lupus erythematosus. Front Immunol. 2022;13:978851.

- Chennamadhavuni A, Abushahin L, Jin N, Presley CJ, Manne A. Risk factors and biomarkers for immune-Related adverse events: A practical guide to identifying High-Risk patients and rechallenging immune checkpoint inhibitors. Front Immunol. 2022;13:779691.
- Smith CI, Islam TC, Mattsson PT, Mohamed AJ, Nore BF, Vihinen M. The tec family of cytoplasmic tyrosine kinases: mammalian Btk, Bmx, Itk, tec, Txk and homologs in other species. BioEssays. 2001;23(5):436–46.
- Gibert C, Blaize M, Fekkar A. Fungal infection in patients treated with Bruton's Tyrosine Kinase inhibitor, from epidemiology to clinical outcome: A systematic review. Clin Microbiol Infect 2024;S1198-743X(24):00639–6. https://doi.or g/10.1016/j.cmi.2024.12.032
- 32. Gruber RC, Wirak GS, Blazier AS, Lee L, Dufault MR, Hagan N, et al. BTK regulates microglial function and neuroinflammation in human stem cell models and mouse models of multiple sclerosis. Nat Commun. 2024;15(1):10116.
- Ni Gabhann J, Hams E, Smith S, Wynne C, Byrne JC, Brennan K, et al. Btk regulates macrophage polarization in response to lipopolysaccharide. PLoS ONE. 2014;9(1):e85834.
- 34. Qiu J, Fu Y, Chen Z, Zhang L, Li L, Liang D et al. BTK Promotes Atherosclerosis by Regulating Oxidative Stress, Mitochondrial Injury, and ER Stress of Macrophages. *Oxid Med Cell Longev* 2021; 2021:9972413.
- 35. Joseph RE, Wales TE, Jayne S, Britton RG, Fulton DB, Engen JR et al. Impact of the clinically approved BTK inhibitors on the conformation of full-length BTK and analysis of the development of BTK resistance mutations in chronic lymphocytic leukemia. Elife 2024;13:RP95488. https://doi.org/10.7554/eLife.9 5488
- Ringheim GE, Wampole M, Oberoi K. Bruton's tyrosine kinase (BTK) inhibitors and autoimmune diseases: making sense of BTK inhibitor specificity profiles and recent clinical trial successes and failures. Front Immunol. 2021;12:662223.
- Corneth OBJ, Klein Wolterink RGJ, Hendriks RW. BTK signaling in B cell differentiation and autoimmunity. Curr Top Microbiol Immunol. 2016;393:67–105.
- Caturegli P, De Remigis A, Rose NR. Hashimoto thyroiditis: clinical and diagnostic criteria. Autoimmun Rev. 2014;13(4–5):391–7.

# **Publisher's note**

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