

Research



Cite this article: Huang C *et al.* 2022 Multi-omics analysis for potential inflammation-related genes involved in tumour immune evasion via extended application of epigenetic data. *Open Biol.* **12**: 210375.
<https://doi.org/10.1098/rsob.210375>

Received: 12 December 2021
Accepted: 18 July 2022

Subject Area:
bioinformatics/immunology

Keywords:
STAT2, PD-L1, immune evasion, ATAC-seq, immune checkpoint therapy, epigenetics

Authors for correspondence:
Bujun Ge
e-mail: gebujun@126.com
Qi Huang
e-mail: hqh007@hotmail.com
Bing Du
e-mail: bdu@bio.ecnu.edu.cn

†Co-first authorship: Chenshen Huang, Ning Wang, Na Zhang, and Zhizhong Chen have contributed equally to this work.

Electronic supplementary material is available online at <https://doi.org/10.6084/m9.figshare.c.6139248>.

Multi-omics analysis for potential inflammation-related genes involved in tumour immune evasion via extended application of epigenetic data

Chenshen Huang^{1,2,3,4,†}, Ning Wang^{2,5,†}, Na Zhang^{2,†}, Zhizhong Chen^{4,†}, Zhizhan Ni³, Xiaohong Liu², Hao Xiong², Huahao Xie³, Boxu Lin², Bujun Ge³, Qi Huang³ and Bing Du²

¹Department of Gastrointestinal Surgery, Fujian Provincial Hospital, Fuzhou, People's Republic of China
²Shanghai Key Laboratory of Regulatory Biology, Institute of Biomedical Sciences and School of Life Sciences, East China Normal University, Shanghai, People's Republic of China
³Department of General Surgery, Tongji Hospital, School of Medicine, Tongji University, Shanghai, People's Republic of China
⁴Shengli Clinical Medical College of Fujian Medical University, Fujian Provincial Hospital, Fuzhou, People's Republic of China
⁵Huzhou Central Hospital, Affiliated Hospital of Zhejiang University, Huzhou, People's Republic of China

CH, 0000-0001-6107-8264; QH, 0000-0002-7917-717X

Accumulating evidence suggests that inflammation-related genes may play key roles in tumour immune evasion. Programmed cell death ligand 1 (PD-L1) is an important immune checkpoint involved in mediating anti-tumour immunity. We performed multi-omics analysis to explore key inflammation-related genes affecting the transcriptional regulation of PD-L1 expression. The open chromatin region of the PD-L1 promoter was mapped using the assay for transposase-accessible chromatin using sequencing (ATAC-seq) profiles. Correlation analysis of epigenetic data (ATAC-seq) and transcriptome data (RNA-seq) were performed to identify inflammation-related transcription factors (TFs) whose expression levels were correlated with the chromatin accessibility of the PD-L1 promoter. Chromatin immunoprecipitation sequencing (ChIP-seq) profiles were used to confirm the physical binding of the TF STAT2 and the predicted binding regions. We also confirmed the results of the bioinformatics analysis with cell experiments. We identified chr9:5449463–5449962 and chr9:5450250–5450749 as reproducible open chromatin regions in the PD-L1 promoter. Moreover, we observed a correlation between STAT2 expression and the accessibility of the aforementioned regions. Furthermore, we confirmed its physical binding through ChIP-seq profiles and demonstrated the regulation of PD-L1 by STAT2 overexpression *in vitro*. Multiple databases were also used for the validation of the results. Our study identified STAT2 as a direct upstream TF regulating PD-L1 expression. The interaction of STAT2 and PD-L1 might be associated with tumour immune evasion in cancers, suggesting the potential value for tumour treatment.

1. Introduction

Immune evasion is an essential mechanism for cancer cells to circumvent immune-system mediated destruction and acquire resistance to treatment. Both laboratory and clinical studies have revealed that PD-L1 plays a key role in immune evasion. PD-L1, also known as CD274, is a co-inhibitory receptor expressed on the surface of multiple cell types, including cancer cells [1,2]. It

can bind to programmed death-1 (PD-1) and inhibit anti-tumour immune reactions, enabling cancer cells to escape immunosurveillance.

Based on the importance of the PD-1/PD-L1 axis in immune evasion, many studies have demonstrated the remarkable clinical efficacy of anti-PD-1/PD-L1 therapy [3]. However, the clinical effects of these treatments are less efficient for certain tumour types, such as non-microsatellite instability (non-MSI) colorectal cancer [4]. To date, the level of PD-L1 expression in cancer cells is regarded as one of the most important factors for determining the effects of immune checkpoint therapy. Therefore, an improved understanding of PD-L1 regulation in cancer cells might be helpful for clinical cancer treatment.

Accumulating evidence has demonstrated the upregulation of PD-L1 expression during cancer pathogenesis. Inflammatory signalling is regarded as a primary mechanism involved in this complex regulatory network [5]. For instance, certain pro-inflammatory factors, including type I and type II interferons, induce PD-L1 expression efficiently. Furthermore, the expression of PD-L1 could be regulated via multiple inflammation-related transcription factors (TFs), such as IRF1, STAT1, STAT3 and NF- κ B [6]. Given that cancer-related inflammation is observed in a substantial proportion of patients, a better understanding of the relationship between inflammation and PD-L1 is currently required. Furthermore, a search for novel inflammatory TFs that regulate PD-L1 expression is warranted.

Assay for transposase-accessible chromatin using sequencing (ATAC-seq) approach uses hyperactive Tn5 transposase to comprehensively recognize chromatin accessibility at the genome level and could map open chromatin regions in gene promoters, reflecting the possibilities of TF binding [7,8]. Although ATAC-seq analysis only indicates the necessity of TF binding, its results could be further confirmed through other experiments.

In this study, we aimed to perform a multi-omics analysis to screen novel inflammation-related TFs involved in PD-L1 regulation, followed by laboratory verification studies. Besides transcriptome data (RNA-seq), we also aimed to analyze epigenetic data (ATAC-seq) to explore the binding of TFs. We hypothesized that STAT2 (signal transducer and activator of transcription 2) could directly bind at open chromatin regions of the PD-L1 promoter and regulate PD-L1 expression in cancer cells. The predicted binding was then validated physically through ChIP assay, and its influence in translational regulation was further confirmed in cell experiments.

2. Material and methods

2.1. Data collection

The Cancer Genome Atlas (TCGA) datasets were accessed through the UCSC Xena database (<https://xenabrowser.net/>). Htseq-count profiles of 514 colon adenocarcinoma (COAD) samples were retrieved, and the corresponding clinical demographic information was also acquired. We also downloaded the Fragments Per Kilobase per Million mapped read (FPKM) profiles of the aforementioned patients with COAD. Publicly available ATAC-seq profiles were obtained from the NCI Genomic Data Commons (<https://gdc.cancer.gov/about-data/publications/ATACseq-AWG>). The ChIP-seq

profiles were acquired from the Cistrome database (<http://cistrome.org/>) [9]. RNA-seq profiles were obtained from the Gene Expression Omnibus (GEO, GSE137155, <https://www.ncbi.nlm.nih.gov/geo/>).

2.2. Identification of inflammation-related TFs

We first downloaded the list of 1639 transcription factors (TF) from the Human Transcription Factors database (<http://humantfs.cbr.utoronto.ca/>) [10]. Then, via literature search, we identified three TF families that were crucial in mediating inflammation, including nuclear factor- κ B (NF- κ B), interferon regulatory factors (IRFs), and signal transducers and activators of transcription (STATs) [11–13]. Accordingly, a total of 21 TFs were identified as inflammatory TFs: NF- κ B 1–2, RelB, c-Rel, IRF 1–9, STATs 1–4, 5a, 5b and 6.

2.3. Chromatin accessibility analysis of PD-L1

To identify the open chromatin regions of PD-L1, peaks were visualized using the R package karyoploteR [14] and ChIPseeker [15], and were annotated using TxDb.Hsapiens.UCSC.Hg38.knownGene. The details of the aforementioned methods have been described in our former study [13,16].

2.4. Identification of potential TFs involved in PD-L1 regulation

In order to analyze transcriptional regulation of PD-L1, we used the workflow reported by Huang *et al.* [16,17]. Briefly, gene expression data of TFs were retrieved from the TCGA datasets. Then, we performed correlation analysis between the TF expression and the chromatin accessibility of the PD-L1 promoter region. The TFs with a p -value < 0.05 were further filtered using the Cistrome database and the GEO database.

2.5. Cell culture

The human colon cancer cell line DLD-1 and cervical carcinoma cell line HeLa were obtained from Shanghai Key Laboratory of Regulatory Biology, East China Normal University, Shanghai, China. Both cell lines were cultured in DMEM supplemented with 1% streptomycin–penicillin and 10% fetal bovine serum.

2.6. Plasmids and transfection

PcDNA3.1-STAT2 plasmid was purchased from Youbio Biological Technology Co., Ltd. (China). The transfections were performed via the calcium phosphate-DNA coprecipitation method for both DLD-1 and HeLa cells, as described previously. Equal amounts of empty vectors were transfected in the negative control group.

2.7. Real-time qPCR

Total RNA was extracted using the TRIzol reagent (Takara). The PrimeScript RT Master Mix Kit (Takara) was used for cDNA generation. Then, real-time qPCR was performed using SYBR Green PCR Master Mix (Yeasten). The primer

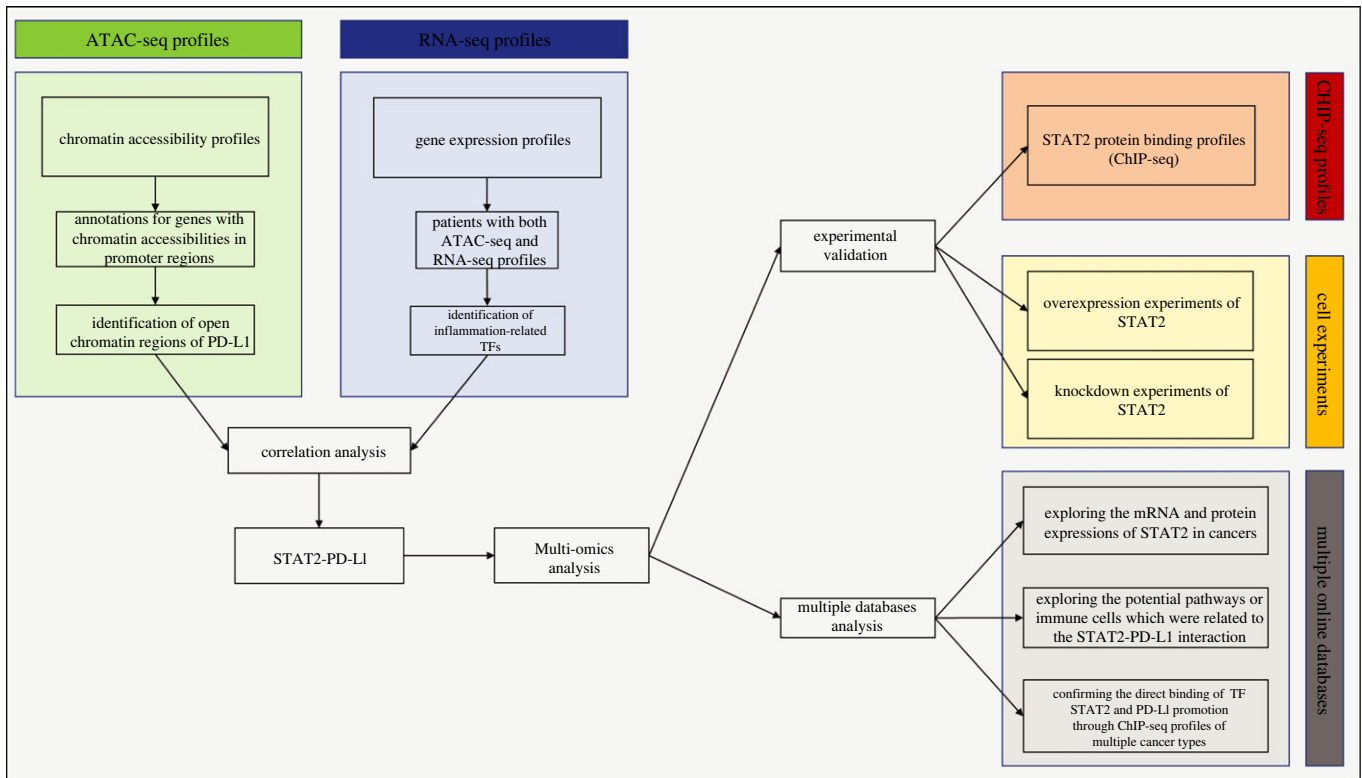


Figure 1. An overview of the study design.

sequences for each gene are listed in electronic supplementary material, table S1.

2.8. Multi-database validation

To minimize the bias in bioinformatics analysis, multiple databases were used for validation, including the Timer database (<http://timer.comp-genomics.org/>) [18], Human Protein Atlas (HPA) database (<https://www.proteinatlas.org/>) [19,20], LinkedOmics database (<https://linkedomics.org/>) [21] and GEPIA database (<http://gepia.cancer-pku.cn/>) [22].

2.9. Statistical analysis

p -values < 0.05 were regarded as statistically significant. Pearson and Spearman analysis was used to calculate correlation coefficients. All statistical analyses were conducted using the R software (v. 3.5.1; www.r-project.org).

3. Results

3.1. Identification of open chromatin regions of the PD-L1 promoter

The overview of our study is presented in figure 1. The chromatin accessibility landscape of patients with COAD was gauged from ATAC-profiles using the workflow reported by Huang *et al.* [16]. The open chromatin regions were widely expressed across the genome. The upset plot and the vennpie plot indicated that a considerable fraction of open chromatin regions was found in gene promoters (electronic supplementary material, figure S1A). Furthermore, we identified chr9:5449463–5449962 (Region 1) and chr9:

5450250–5450749 (Region 2) as reproducible open chromatin regions in the PD-L1 promoter across 41 patients with COAD. We also found that PD-L1 expression was significantly correlated with both Region 1 ($r = 0.6$, $p < 0.05$, figure 2*a*) and Region 2 ($r = 0.4$, $p < 0.05$, figure 2*b*).

3.2. Identification of STAT2 as an upstream factor of PD-L1 by integrative analysis

In order to explore the upstream factors of PD-L1, we combined transcriptome (RNA-seq) and epigenetics profiles (ATAC-seq) for co-analysis. Based on the list of TFs associated with inflammation, we obtained the mRNA expression of 21 inflammatory TFs from RNA-seq profiles in the TCGA database (figure 3*a*). These TFs were then filtered by correlation analysis with chromatin accessibilities of Region 1 and Region 2 (figure 3*b*). Among the 21 inflammatory TFs, STAT2 was found to have the strongest correlation with PD-L1 promoter accessibility. STAT2 demonstrated a remarkable correlation with both Region 1 and Region 2 (figure 3*c*). Additionally, in order to confirm that STAT2 had a high ranking even among all the TFs, we also used correlation analysis between PD-L1 promoter accessibility and expression of all database-recorded TFs (electronic supplementary material, table S2). Among all the TFs, STAT2 was only secondary to TF MAX with respect to correlation with Region 1 (electronic supplementary material, figure S2A, left). Similarly, STAT2 also had a significant correlation with Region 2 (electronic supplementary material, figure S2A, right).

Moreover, we performed correlation analysis between PD-L1 expression and the expression of potential TFs. Interestingly, among the top 5 TFs associated with PD-L1 promoter Region 1 or Region 2, STAT2 was found to have the strongest correlation with PD-L1 expression (electronic supplementary material, figure S2B, figure 4*a*), which also suggested that

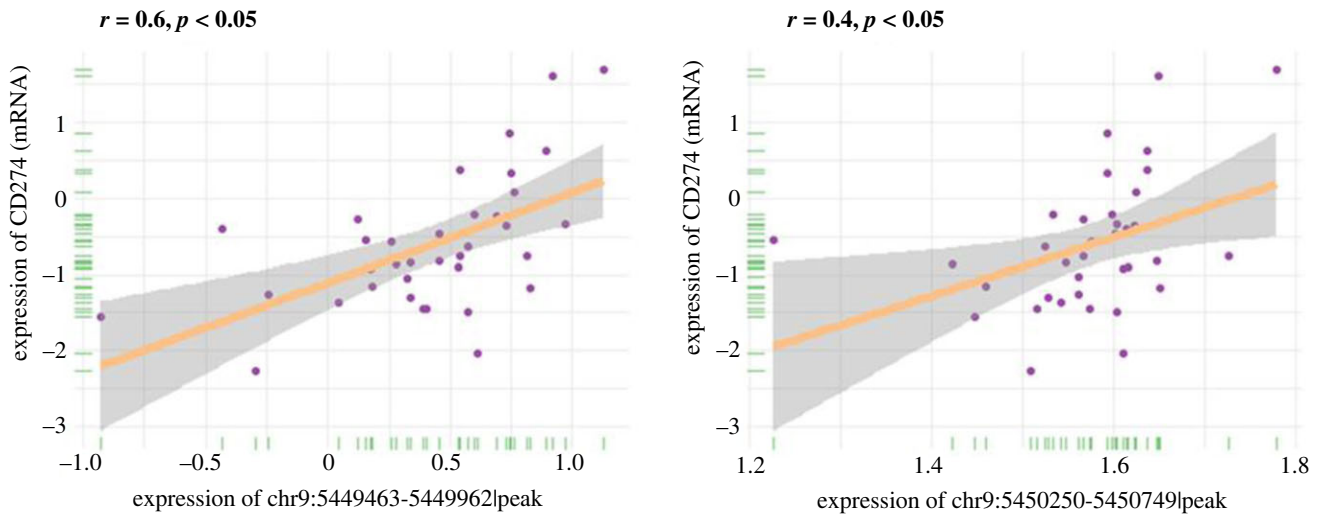


Figure 2. Correlation analysis between PD-L1 expression and PD-L1 promoter accessibility. (a) Correlation analysis revealed that PD-L1 expression was significantly correlated with the chromatin accessibility of Region 1 ($r = 0.6$, $p < 0.05$). (b) Correlation analysis revealed that PD-L1 expression was significantly correlated with the chromatin accessibility of Region 2 ($r = 0.4$, $p < 0.05$).

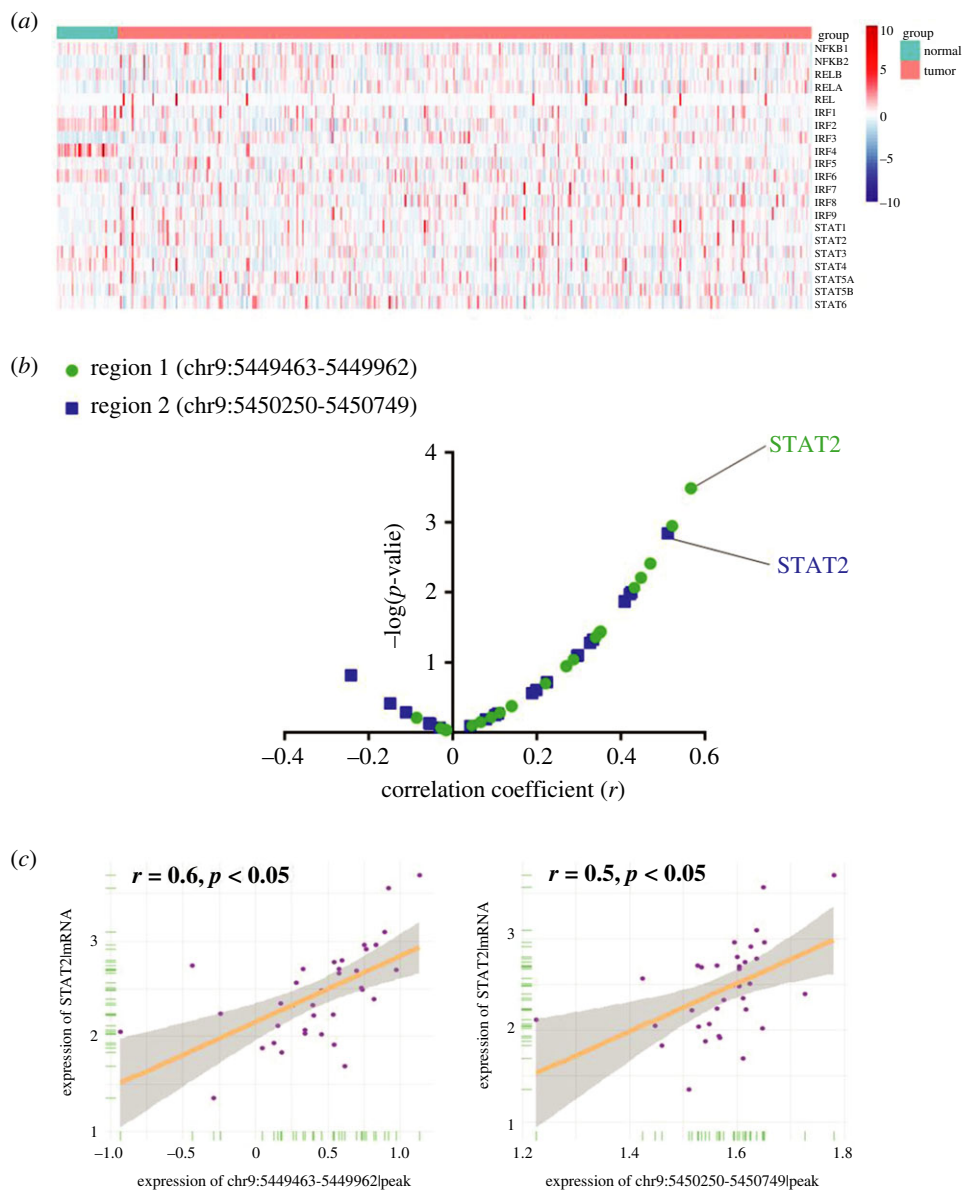


Figure 3. Integrative analysis of RNA-seq and ATAC-seq profiles identified STAT2 as a potential upstream for PD-L1. (a) Heatmap for gene expression of the identified 21 inflammation-related TFs, which were from NF- κ B, IRFs or STATs families. The mRNA expression of all the 21 TFs could be detected in colon cancer tissues. (b) Correlation analysis revealed that STAT2 had a stronger association with chromatin accessibilities of PD-L1 promoter. (c) The STAT2 expression was significantly correlated with the chromatin accessibility of Region 1 ($r = 0.6$, $p < 0.05$). The STAT2 expression was significantly correlated with the chromatin accessibility of Region 2 ($r = 0.5$, $p < 0.05$).

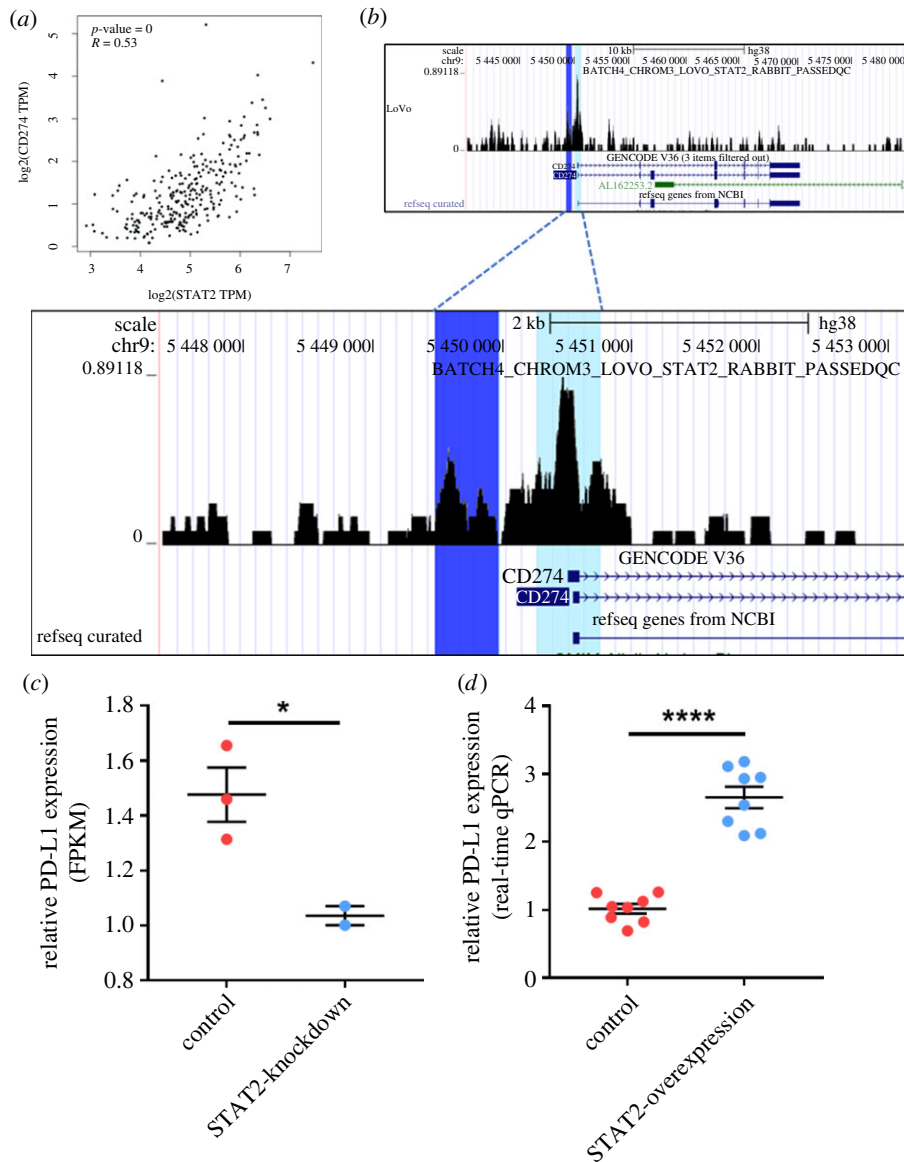


Figure 4. Validation of the direct regulation of STAT2 on PD-L1 through ChIP-seq profiles and cell experiments. (a) The mRNA expressions of STAT2 and PD-L1 were significantly correlated ($r = 0.53$, $p < 0.05$). (b) ChIP-seq profiles of STAT2 in colon cancer cell line revealed that STAT2 could directly bind to PD-L1 promoter. And there was a strong overlap between STAT2 binding sites and the predicted regions (Region 1 in dark blue, or Region 2 in light blue). (c) The knockdown of STAT2 could lead to downregulation of PD-L1 significantly ($p < 0.05$). (d) The overexpression of STAT2 could lead to upregulation of PD-L1 significantly ($p < 0.05$). * $p < 0.05$; **** $p < 0.0001$.

STAT2 might have a close relationship with PD-L1 expression. Collectively, STAT2 showed a significant correlation with both PD-L1 promoter accessibility and PD-L1 expression. Hence, we proposed that STAT2 might be a potential upstream of PD-L1, and we chose STAT2 for further validation.

3.3. Validation of the physical interactions between STAT2 and PD-L1 promoter

Considering that the TF binding motif is an important index for predicting TF binding, we used the online Human Transcription Factor Database (<http://bioinfo.life.hust.edu.cn/HumanTFDB/>) to explore the STAT2 binding motifs on Region 1 and Region 2. As shown in electronic supplementary material, table S3, STAT2 binding motifs could be found in both two regions, indicating that STAT2 might physically bind at Region 1 and Region 2. To confirm the physical binding, we further obtained STAT2 ChIP-seq profiles of colon cancer cells for validation. Figure 4b shows that

STAT2 protein could specifically bind at Region 1 (highlighted in dark blue) and Region 2 (highlighted in light blue), confirming the physical interaction between STAT2 and PD-L1 promoter. Furthermore, we obtained the RNA-seq data of colon cancer cells with STAT2-knockdown from the GEO dataset GSE137155. PD-L1 expression was significantly downregulated in the STAT2-knockdown group (figure 4c). The analysis of the protein-protein interaction network also supported the correlation observed between STAT2 and PD-L1 (electronic supplementary material, figure S3A). Collectively, the integrated analysis indicated that the TF STAT2 could physically bind at the open chromatin regions of the PD-L1 promoter and might regulate PD-L1 expression.

3.4. Validation of the results of bioinformatics analysis in colon cancer cells

To minimize the bias, we also confirmed the results of the aforementioned bioinformatics analysis with colon cancer

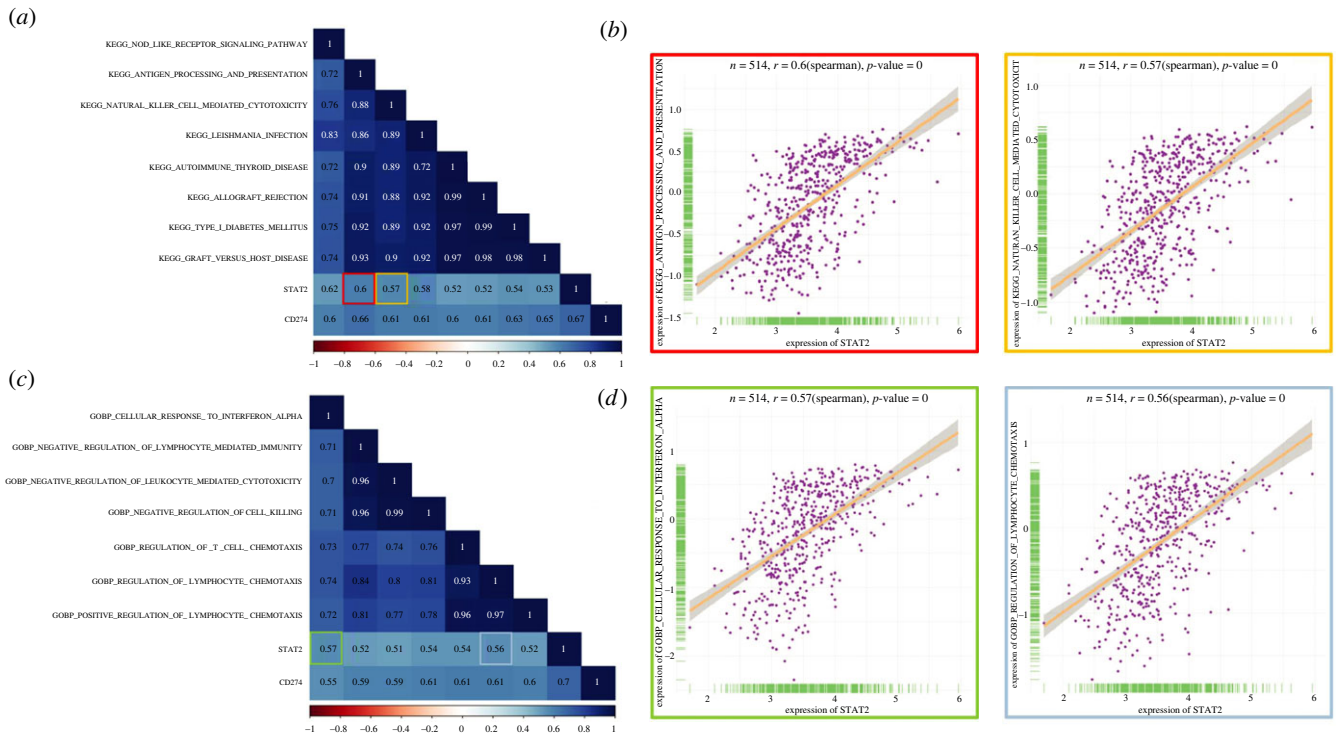


Figure 5. Exploring the potential pathways which were related with the interaction of STAT2 and PD-L1. (a) Correlation heatmap of STAT2, PD-L1, and the significant KEGG pathways. The pathways, which had a strong correlation with both STAT2 and PD-L1 ($r > 0.4$, $p < 0.05$), were displayed. (b) The dot plots showed the correlation between STAT2 expression, and the KEGG_antigen_processing_and_presentation pathway (marked in red, $r = 0.6$, $p < 0.05$) or KEGG_natural_killer_cell_mediating_cytotoxicity pathway (marked in yellow, $r = 0.57$, $p < 0.05$). (c) Correlation heatmap of STAT2, PD-L1, and the significant GO pathways. The pathways, which had a strong correlation with both STAT2 and PD-L1 ($r > 0.4$, $p < 0.05$), were displayed. (d) The dot plots showed the correlation between STAT2 expression, and the GOBP_cellular_response_to_interferon_alpha pathway (marked in green, $r = 0.57$, $p < 0.05$) or GOBP_regulation_of_lymphocyte_chemotaxis pathway (marked in blue, $r = 0.56$, $p < 0.05$).

cell line DLD-1. As expected, the transfection and over-expression of STAT2 resulted in significant upregulation of PD-L1 expression (figure 4d). Similar results were also observed with HeLa cells (electronic supplementary material, figure S3B). Thus, via *in vitro* experiments, we validated that overexpression of STAT2 could upregulate PD-L1 expression in cancer cells, which was in line with the results of the bioinformatics analysis.

3.5. Identification of significant pathways and immune cells associated with STAT2 and PD-L1

Considering that STAT2 could be a direct upstream factor of PD-L1, we further explored associated pathways in COAD. We used the gene set variation analysis algorithm [23] to identify the expression level of genes enriched in the GO and KEGG pathway analysis (figure 5a–d). Correlation analysis was applied to explore significant pathways that were correlated with both STAT2 and PD-L1 expression (figure 5a,c). We found that ‘KEGG_antigen_processing_and_presentation’ and ‘GOBP_cellular_response_to_interferon_alpha’ pathways were most significantly correlated (figure 5b, d). Therefore, we hypothesized that the interaction of STAT2 and PD-L1 might influence colon cancers in an immune-related way.

Considering this hypothesis, we analysed tumour-infiltrating immune cells in colon cancers. With the clustering function of R package corrplot, we found that macrophages might be potentially associated with both STAT2 and PD-L1 (figure 6a). Furthermore, besides clustering function, we

also explored the estimated immune cells using correlation analyses. And we found that algorithm-estimated macrophages and T cells were statistically correlated with both STAT2 and PD-L1 ($r > 0.4$, $p < 0.05$, figure 6b). Thus, based on the analysis above, macrophages might play a role in the interaction between STAT2 and PD-L1. However, limited to bioinformatics methods, we only observed a correlation among macrophages, STAT2, and PD-L1. The underlying mechanisms still required further elucidation through laboratory experiments.

3.6. Multi-database validation

To confirm the bioinformatics results, we used multiple databases for validation. Profiles from Xena database showed that STAT2 was widely expressed in multiple cancer types, including colon cancers (figure 7a). Moreover, the HPA database indicated that in the tumour microenvironment of colon cancers, STAT2 could be detected in multiple cell types, including cancer cells. Although there is no significant difference of STAT2 protein expression between colon cancer tissues and normal tissues, we found that the expression level of STAT2 in normal endothelial cells was relatively lower ($p < 0.05$, figure 7b). In addition, the LinkedOmics database demonstrated that the STAT2 expression was significantly lower in patients with MSS (non-MSI) colon cancer compared to that in patients with MSI-H colon cancer (figure 7c). To confirm the direct binding of TF STAT2 and PD-L1 promoter, we also obtained STAT2 ChIP-seq profiles of multiple types of cancer cell lines, including GM12878,

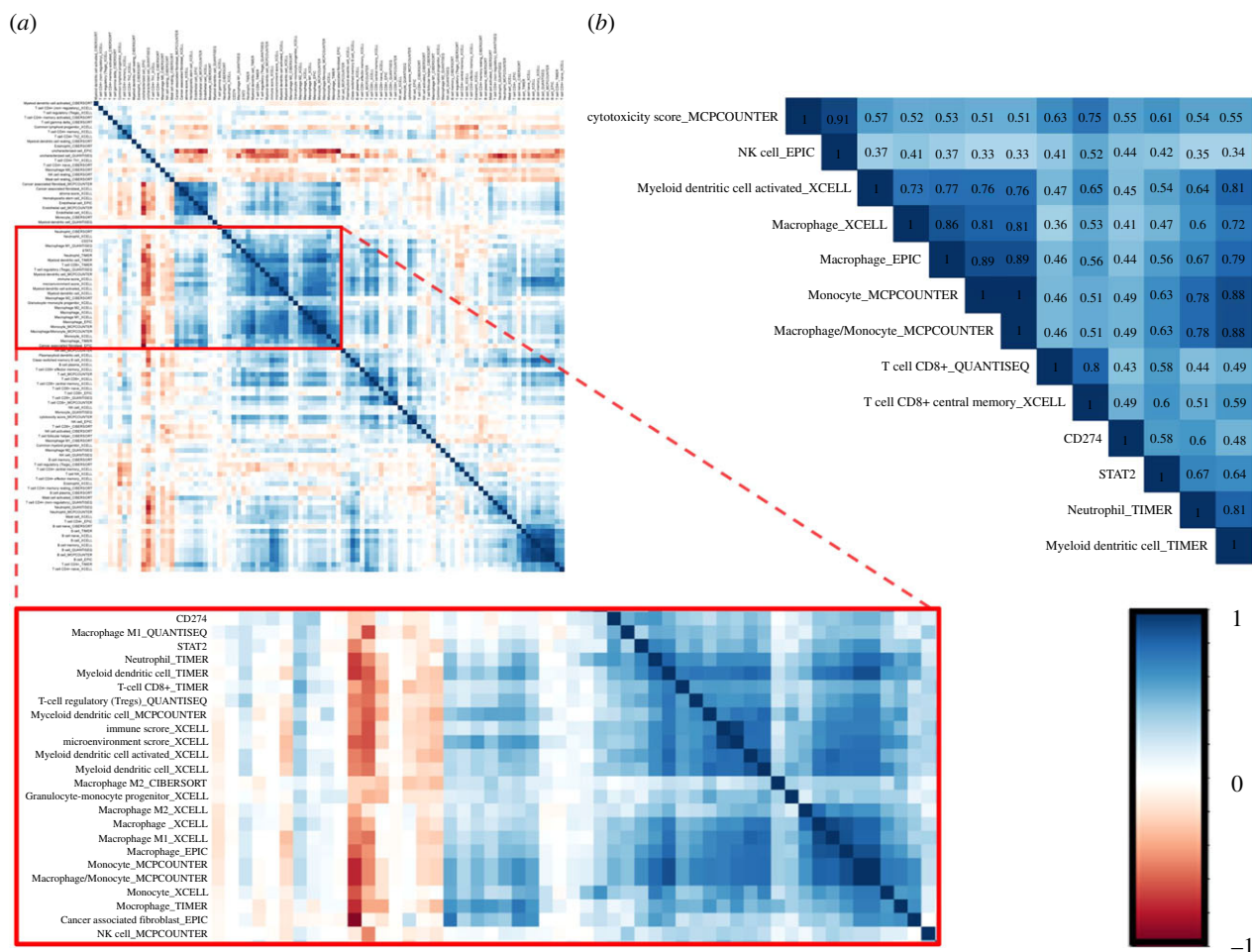


Figure 6. Exploring the potential immune cells which were related with the interaction of STAT2 and PD-L1. (a) Correlation heatmap of STAT2, PD-L1, and the significant tumour infiltrating immune cells. After the clustering through R package corplot, STAT2 and PD-L1 were found to be potentially associated with multiple cell types, especially macrophages. (b) The correlation heatmap showed the immune cells that had a strong correlation with both STAT2 and PD-L1 ($r > 0.4$, $p < 0.05$).

K562 and LoVo. The TF STAT2 were confirmed to directly bind to the predicted regions of PD-L1 promoter (figure 7d).

4. Discussion

With the understanding of immune evasion mechanisms, immune checkpoint therapy has been developed as an important clinical strategy for cancer treatment. Among various such strategies, anti-PD-1/PD-L1 therapy is one of the most extensively examined strategies. Considering that the level of PD-L1 expression in cancer cells is closely associated with clinical efficacy, there is an urgent need to elucidate mechanisms underlying PD-L1 regulation. Cancer-induced chronic inflammation is very common among patients and affects PD-L1 expression via multiple pathways, including transcriptional regulation. However, to date, its regulatory mechanism has not been fully clarified. Therefore, we aimed to screen undiscovered inflammatory TFs that were direct upstream factors of PD-L1 and to further confirm the results via laboratory validation.

The integrated analysis of transcriptome and epigenetic profiles was applied for screening the potential inflammatory TFs. First, we performed correlation analysis between PD-L1 promoter accessibility and expression of inflammatory TFs. Among the 21 database-recorded inflammatory TFs, STAT2 was found to be most correlated with PD-L1 promoter

accessibility. Then, we performed correlation analysis between PD-L1 promoter accessibility and expression of all the database-recorded TFs, confirming that STAT2 had a high ranking even among all the database-recorded TFs. Moreover, we further performed correlation analysis between PD-L1 expression and the expression of several TFs, which were also significantly correlated with PD-L1 promoter accessibility. Among these TFs, STAT2 was found to have the strongest correlation with PD-L1 expression. Collectively, the inflammatory TF STAT2 was found to be significantly correlated with both PD-L1 promoter accessibility and PD-L1 expression. Hence, we proposed that STAT2 might be a potential upstream of PD-L1, and chose STAT2 for further validation.

Based on the integrated analysis of transcriptome and epigenetic profiles, we proposed that the TF STAT2 might directly bind at the PD-L1 promoter. We also identified the potential binding sites chr9:5449463–5449962 and chr9:5450250–5450749. Upon analysing ChIP-seq profiles of STAT2, we confirmed the physical binding within the predicted region. Subsequently, we further verified our results at the cellular level. We demonstrated that the overexpression of STAT2 can significantly upregulate PD-L1 expression in cancer cell lines DLD-1 and HeLa. Based on the GEO database, we found that STAT2-knockdown could significantly inhibit PD-L1 expression. Taken together, based on the results of this study, we propose that STAT2 is a direct upstream factor of PD-L1.

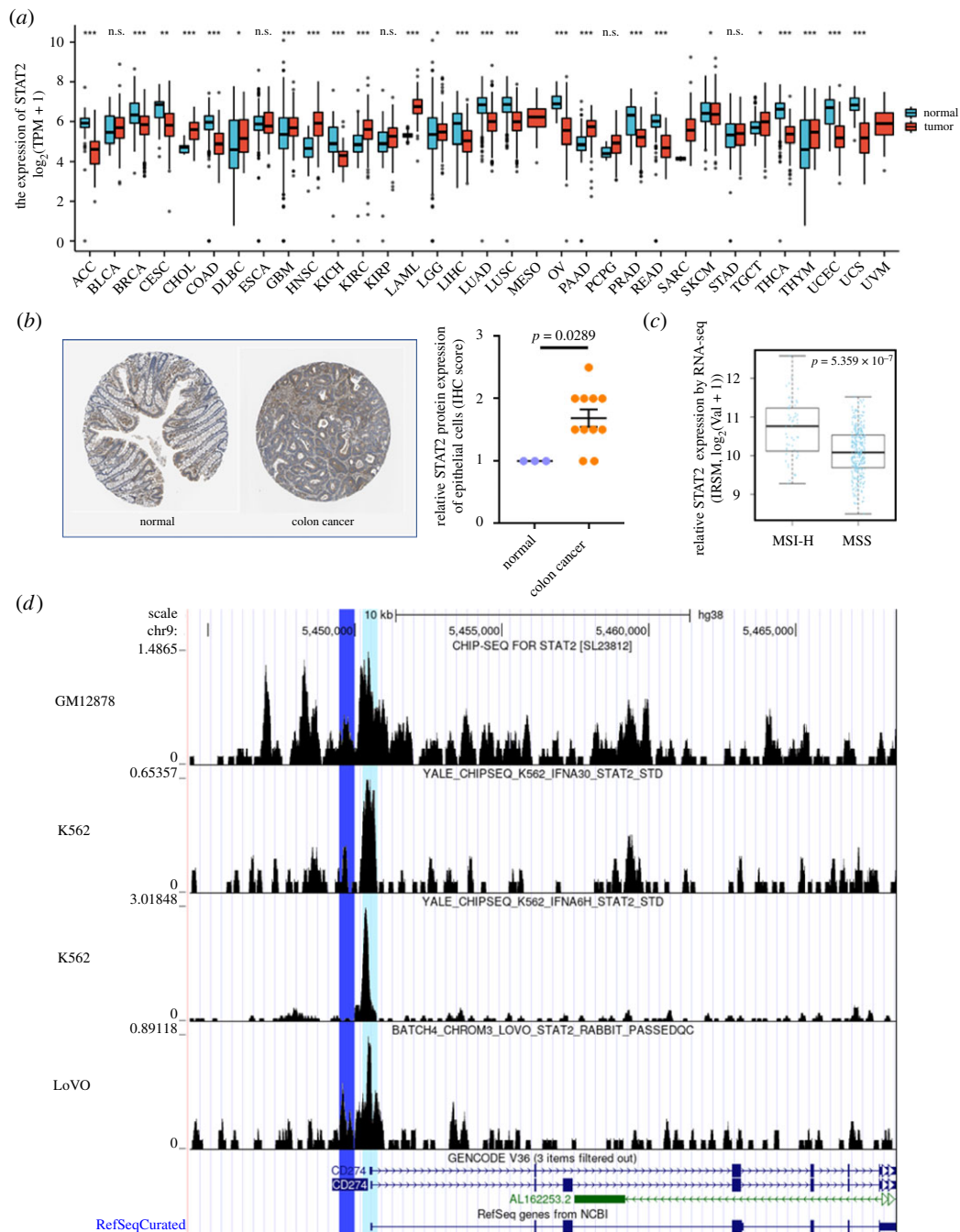


Figure 7. Multiple databases were used for validation. (a) Box plots of STAT2 expression in different cancers from the TCGA and GTEx databases accessed by Xena. STAT2 was widely expressed across multiple cancer types. (b) Immunohistochemical results of the protein expression of STAT2 in patients with colon cancers via the HPA database. The expression level of STAT2 in normal endothelial cells was relatively lower ($p < 0.05$). (c) Box plots revealed a significant difference of STAT2 expression between MSI-H and MSS (no-MSI) colon cancer ($p < 0.05$). (d) The overlap of STAT2 binding sites and the predicted region (Region 1 in dark blue, or Region 2 in light blue) was validated in STAT2 ChIP-seq profiles of multiple cell lines, including GM12878, K562 and LoVo.

STAT2 is known for its role in immunomodulatory reactions and anti-viral immunity. It is significantly different from other members of the STAT family. For instance, although other members can be activated by multiple cytokines, including type I and II interferons, STAT2 is primarily activated by type I interferon. Moreover, it is involved in mediating inflammatory pathways and acts as a cofactor [24]. Thus, the regulation of STAT2 might not have a significant impact on anti-tumour immunity. Therefore, if PD-L1 expression is regulated via STAT2, the combination therapy targeting STAT2 might be more effective.

In this study, we demonstrated the regulation of PD-L1 expression mediated by STAT2 via multi-omics analysis and laboratory validation. Over past decades, the regulation of PD-L1

expression by various inflammatory TFs has been reported; however, our study is the first to report that the TF STAT2 could directly regulate PD-L1 expression. We also confirmed the physical binding of the TF STAT2 and PD-L1 promoter based on ChIP-seq results. Interestingly, a study performed by Angel Garcia-Diaz *et al.* [25] used mutagenesis to delete the predicted binding sites of STAT2 in a firefly luciferase reporter plasmid comprising the PD-L1 promoter. The results demonstrated that interferon-induced luciferase expression was remarkably decreased in the transfected cells. To some extent, the results of this experiment supplemented the findings of our study, which provided strong evidence for the direct binding of the TF.

In addition, Angel Garcia-Diaz *et al.* [25] not only found that the deletion of STAT2 putative binding site could affect

luciferase expression but also observed similar results upon deletion of IRF1 putative site. Indicating the importance of the JAK1/JAK2-STAT1/STAT2/STAT3-IRF1 axis, they revealed that IRF1 was the potential TF that regulated PD-L1 directly. In our study, we focused on the direct regulation of PD-L1 expression by STAT2, providing an improved understanding of the JAK1/JAK2-STAT1/STAT2/STAT3-IRF1 axis. As a direct upstream factor of both PD-L1 and IRF1, STAT2 showed great promise for anti-PD-1/PD-L1 immunotherapy.

Despite the increasing number of studies examining STAT2, the effects of STAT2 in anti-tumour immunity remain controversial [26]. For instance, Yue *et al.* [27] demonstrated the acceleration of tumour growth in mice with STAT2 knockout. Wang *et al.* [28] confirmed the anti-tumour activity of STAT2 in a mouse model. Gamero *et al.* [29] used models of inflammation-induced cancers to demonstrate that STAT2 might promote colorectal and skin carcinogenesis. Considering the controversial roles of STAT2, our study might provide further clarifications with respect to the role of STAT2 in cancer. We found that since STAT2 can upregulate PD-L1 expression on the surface of cancer cells, it can aid the cancer cells in escaping immunosurveillance.

Currently, the anti-PD-1/PD-L1 treatment demonstrates low efficiency for patients with non-MSI colon cancers [4,5]. Based on the LinkedOmics database, we found that the STAT2 expression was significantly lower in no-MSI colon cancers. Thus, we hypothesized that the combination treatment targeting STAT2 might increase the efficacy of anti-PD-1/PD-L1 treatment in patients diagnosed with this cancer subtype.

Certain limitations of this study need to be addressed. First, most of our bioinformatics results were based on the data on colon cancers, which were used to perform integrated analysis. On the one hand, the multi-omics analysis would be more reliable when using profiles from the same cancer type. On the other hand, these results might be specific to colon cancers. To minimize the bias, we used multiple databases to validate our results at the pan-cancer level, and we also used other types of cell lines in subsequent experiments. Second, the association between the differential TF binding and the changes in PD-L1 promoter accessibility could be influenced by many potential events, such as the dynamics of histone modification, DNA methylation, and promoter-enhancer looping. This study did not include all these factors into consideration. However, although we were not able to provide evidence for all the potential factors, we provided the evidence that there was a significant association between PD-L1 promoter accessibility and PD-L1 expression.

Despite the aforementioned limitations, our study was the first to highlight the direct regulation of PD-L1 expression mediated by the TF STAT2. We performed both bioinformatics and laboratory analysis to validate our results. Future studies should further validate the interaction of STAT2 and PD-L1 with larger data sizes, different cancer cell lines, and the STAT2 knockdown mouse model. The potential therapeutic value of the combination treatment should also be analysed further.

Our study identified STAT2 as a direct upstream TF that regulates PD-L1 expression, suggesting its potential to be used as a therapeutic target for tumour treatment.

Data accessibility. Publicly available datasets were analysed in this study. Publicly available gene expression profiles can be acquired from the UCSC Xena database (<https://xenabrowser.net/>), and the GEO database (GSE137155, <https://www.ncbi.nlm.nih.gov/geo/>). Publicly available ATAC-seq profiles can be acquired from the NCI Genomic Data Commons (<https://gdc.cancer.gov/about-data/publications/ATACseq-AWG>). The ChIP-seq profiles can be acquired from the Cistrome database (<http://cistrome.org/>). Further inquiries can be directed to the corresponding authors.

Authors' contributions. C.H.: conceptualization, data curation, formal analysis, methodology, resources, software, validation, visualization, writing—original draft, writing—review and editing; N.W.: data curation, methodology, validation, writing—review and editing; N.Z.: methodology, resources, validation, visualization, writing—review and editing; Z.C.: investigation, methodology, software, validation, writing—review and editing; Z.N.: conceptualization, data curation, methodology, software, visualization, writing—original draft; X.L.: data curation, investigation, validation, visualization, writing—review and editing; Ha.X.: validation, writing—review and editing; Hu.X.: methodology, writing—review and editing; B.L.: methodology, validation, visualization, writing—review and editing; B.G.: conceptualization, funding acquisition, supervision, writing—review and editing; Q.H.: conceptualization, funding acquisition, supervision, writing—original draft, writing—review and editing; B.D.: funding acquisition, methodology, supervision, writing—review and editing.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Conflict of interest declaration. The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Funding. We received no funding for this study.

Acknowledgements. The authors thank Jianming Zeng (University of Macau, Macau, China), and all the members of his bioinformatics team for generously sharing their experience and codes. The authors also thank Hong Chen (Fujian Provincial Hospital, Fuzhou, China), Yue Tian (East China Normal University, Shanghai, China) and Qi Pan (Tianjin Academy of Fine Arts, Tianjin, China) for generously sharing their experience.

References

- Doroshov DB *et al.* 2021 PD-L1 as a biomarker of response to immune-checkpoint inhibitors. *Nat. Rev. Clin. Oncol.* **18**, 345–362. (doi:10.1038/s41571-021-00473-5)
- Egen JG, Ouyang W, Wu LC. 2020 Human anti-tumour immunity: insights from immunotherapy clinical trials. *Immunity* **52**, 36–54. (doi:10.1016/j.immuni.2019.12.010)
- He X, Xu C. 2020 Immune checkpoint signaling and cancer immunotherapy. *Cell Res.* **30**, 660–669. (doi:10.1038/s41422-020-0343-4)
- Overman MJ *et al.* 2017 Nivolumab in patients with metastatic DNA mismatch repair-deficient or microsatellite instability-high colorectal cancer (CheckMate 142): an open-label, multicentre, phase 2 study. *Lancet Oncol.* **18**, 1182–1191. (doi:10.1016/S1470-2045(17)30422-9)
- Sun C, Mezzadra R, Schumacher TN. 2018 Regulation and function of the PD-L1 checkpoint. *Immunity* **48**, 434–452. (doi:10.1016/j.immuni.2018.03.014)
- Antonangeli F, Natalini A, Garassino MC, Sica A, Santoni A, Di Rosa F. 2020 Regulation of PD-L1 Expression by NF- κ B in Cancer. *Front. Immunol.* **11**, 584626. (doi:10.3389/fimmu.2020.584626)
- Corces MR *et al.* 2018 The chromatin accessibility landscape of primary human cancers. *Science* **362**, 6413. (doi:10.1126/science.aav1898)
- Nordstrom KJV *et al.* 2019 Unique and assay specific features of NOME-, ATAC- and DNase I-seq data. *Nucleic Acids Res.* **47**, 10 580–10 596. (doi:10.1093/nar/gkz799)

9. Liu T *et al.* 2011 Cistrome: an integrative platform for transcriptional regulation studies. *Genome Biol.* **12**, R83. (doi:10.1186/gb-2011-12-8-r83)
10. Lambert SA *et al.* 2018 The human transcription factors. *Cell* **172**, 650–665. (doi:10.1016/j.cell.2018.01.029)
11. Yu H, Pardoll D, Jove R. 2009 STATs in cancer inflammation and immunity: a leading role for STAT3. *Nat. Rev. Cancer* **9**, 798–809. (doi:10.1038/nrc2734)
12. Plataniotis E, Decker T. 2018 Regulatory Networks Involving STATs, IRFs, and NF κ B in Inflammation. *Front. Immunol.* **9**, 2542. (doi:10.3389/fimmu.2018.02542)
13. Ni Z, Huang C, Zhao H, Zhou J, Hu M, Chen Q, Ge B, Huang Q. 2021 PLXNC1: a novel potential immune-related target for stomach adenocarcinoma. *Front. Cell Dev. Biol.* **9**, 662707. (doi:10.3389/fcell.2021.662707)
14. Gel B, Serra E. 2017 karyoploteR: an R/Bioconductor package to plot customizable genomes displaying arbitrary data. *Bioinformatics* **33**, 3088–3090. (doi:10.1093/bioinformatics/btx346)
15. Yu G, Wang LG, He QY. 2015 ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* **31**, 2382–2383. (doi:10.1093/bioinformatics/btv145)
16. Huang C, Huang R, Chen H, Ni Z, Huang Q, Huang Z, Ge B. 2020 Chromatin accessibility regulates gene expression and correlates with tumor-Infiltrating Immune Cells in Gastric Adenocarcinoma. *Front. Oncol.* **10**, 609940. (doi:10.3389/fonc.2020.609940)
17. Huang C *et al.* 2021 Multi-omics analysis for transcriptional regulation of immune-related targets using epigenetic data: a new research direction. *Front. Immunol.* **12**, 741634. (doi:10.3389/fimmu.2021.741634)
18. Li T, Fu J, Zeng Z, Cohen D, Li J, Chen Q, Li B, Liu XS. 2020 TIMER2.0 for analysis of tumor-infiltrating immune cells. *Nucleic Acids Res.* **48**, W509–W514. (doi:10.1093/nar/gkaa407)
19. Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A *et al.* 2015 Tissue-based map of the human proteome. *Science* **347**, 1260419. (doi:10.1126/science.1260419)
20. Uhlen M *et al.* 2017 A pathology atlas of the human cancer transcriptome. *Science* **357**, 6352. (doi:10.1126/science.aan2507)
21. Vasaikar SV, Straub P, Wang J, Zhang B. 2018 LinkedOmics: analyzing multi-omics data within and across 32 cancer types. *Nucleic Acids Res.* **46**, D956–D963. (doi:10.1093/nar/gkx1090)
22. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. 2017 GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res.* **45**(W1), W98–W102. (doi:10.1093/nar/gkx247)
23. Hanzelmann S, Castelo R, Guinney J. 2013 GSEA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinform.* **14**, 7. (doi:10.1186/1471-2105-14-7)
24. Stark GR, Darnell Jr JE. 2012 The JAK-STAT pathway at twenty. *Immunity* **36**, 503–514. (doi:10.1016/j.immuni.2012.03.013)
25. Garcia-Diaz A *et al.* 2017 Interferon receptor signaling pathways regulating PD-L1 and PD-L2 expression. *Cell Rep.* **19**, 1189–1201. (doi:10.1016/j.celrep.2017.04.031)
26. Verhoeven Y *et al.* 2020 The potential and controversy of targeting STAT family members in cancer. *Semin Cancer Biol.* **60**, 41–56. (doi:10.1016/j.semcancer.2019.10.002)
27. Yue C, Xu J, Tan Estioko MD, Kotredes KP, Lopez-Otalora Y, Hilliard BA, Baker DP, Gallucci S, Gamero AM. 2015 Host STAT2/type I interferon axis controls tumor growth. *Int. J. Cancer* **136**, 117–126. (doi:10.1002/ijc.29004)
28. Wang J, Pham-Mitchell N, Schindler C, Campbell IL. 2003 Dysregulated Sonic hedgehog signaling and medulloblastoma consequent to IFN-alpha-stimulated STAT2-independent production of IFN-gamma in the brain. *J. Clin. Invest.* **112**, 535–543. (doi:10.1172/JCI18637)
29. Gamero AM, Young MR, Mentor-Marcel R, Bobe G, Scarzello AJ, Wise J, Colburn NH. 2010 STAT2 contributes to promotion of colorectal and skin carcinogenesis. *Cancer Prev. Res. (Phila)* **3**, 495–504. (doi:10.1158/1940-6207.CAPR-09-0105)