

## ORIGINAL RESEARCH

# Single-cell Long Non-coding RNA Landscape of T Cells in Human Cancer Immunity



Haitao Luo<sup>1,2,3,#,\*</sup>, Dechao Bu<sup>4,#</sup>, Lijuan Shao<sup>1,2,3,#</sup>, Yang Li<sup>5</sup>, Liang Sun<sup>4</sup>, Ce Wang<sup>1,2</sup>, Jing Wang<sup>1,2,3</sup>, Wei Yang<sup>1,2</sup>, Xiaofei Yang<sup>1,2</sup>, Jun Dong<sup>3,\*</sup>, Yi Zhao<sup>4,\*</sup>, Furong Li<sup>1,2,\*</sup>

<sup>1</sup>Translational Medicine Collaborative Innovation Center, The Second Clinical Medical College (Shenzhen People's Hospital), Jinan University, Shenzhen 518020, China

<sup>2</sup>Shenzhen Key Laboratory of Stem Cell Research and Clinical Transformation, Shenzhen 518020, China

<sup>3</sup>Integrated Chinese and Western Medicine Postdoctoral Research Station, Jinan University, Guangzhou 510632, China

<sup>4</sup>Bioinformatics Research Group, Key Laboratory of Intelligent Information Processing, Advanced Computing Research Center, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China

<sup>5</sup>Department of Gastrointestinal Surgery, The Second Clinical Medical College (Shenzhen People's Hospital), Jinan University, Shenzhen 518020, China

Received 27 March 2020; revised 3 December 2020; accepted 6 March 2021

Available online 18 July 2021

Handled by Yun-Gui Yang

**Abstract** The development of new biomarkers or therapeutic targets for cancer immunotherapies requires deep understanding of T cells. To date, the complete landscape and systematic characterization of long noncoding RNAs (lncRNAs) in T cells in cancer immunity are lacking. Here, by systematically analyzing full-length single-cell RNA sequencing (scRNA-seq) data of more than 20,000 libraries of T cells across three cancer types, we provided the first comprehensive catalog and the functional repertoires of lncRNAs in human T cells. Specifically, we developed a custom pipeline for *de novo* transcriptome assembly and obtained a novel lncRNA catalog containing 9433 genes. This increased the number of current human lncRNA catalog by 16% and nearly doubled the number of lncRNAs expressed in T cells. We found that a portion of expressed genes in single T cells were lncRNAs which had been overlooked by the majority of previous studies. Based on metacell maps constructed by the MetaCell algorithm that partitions scRNA-seq datasets into disjointed and homogenous groups of cells (metacells), 154 signature lncRNA genes were identified. They were associated with effector, exhausted, and regulatory T cell states. Moreover, 84 of them were functionally annotated based on the co-expression networks, indicating that lncRNAs might broadly participate in the regulation of T cell functions. Our findings provide a new point of view and resource for investigating the mechanisms of T cell regulation in cancer immunity as well as for novel cancer-immune biomarker development and cancer immunotherapies.

**KEYWORDS** Long non-coding RNA; Transcriptome assembly; Metacell; Immune regulation; Functional annotation

## Introduction

T cell checkpoint inhibition therapies, such as targeting exhausted CD8<sup>+</sup> T cells and regulatory T cells (Tregs), have

\*Corresponding authors.

E-mail: [luoht1985@gmail.com](mailto:luoht1985@gmail.com) (Luo H), [dongjunbox@163.com](mailto:dongjunbox@163.com) (Dong J), [bio-zy@ict.ac.cn](mailto:bio-zy@ict.ac.cn) (Zhao Y), [frli62@163.com](mailto:frli62@163.com) (Li F).

#Equal contribution.

Peer review under responsibility of Beijing Institute of Genomics, Chinese Academy of Sciences / China National Center for Bioinformation and Genetics Society of China. <https://doi.org/10.1016/j.gpb.2021.02.006>

shown remarkable clinical benefit in many cancers [1–3]. Nevertheless, the mechanisms underlying therapy response or resistance are largely unknown, which leads to different therapeutic efficacies among cancer patients [4–8]. To better understand the mechanisms that underlie successful response to immunotherapy, more comprehensive studies to explore the whole transcriptome of individual T cells in tumor ecosystems are desired. Long non-coding RNAs (lncRNAs), defined as a class of non-coding RNAs longer than 200 nt with no or low protein-coding potential, comprise a large proportion of the mammalian transcriptome [9–12]. Accumulating evidence has suggested that lncRNAs are widely expressed in immune cells and play crucial roles in cancer immunity by regulating the differentiation and function of T cells [13–17]. For example, overexpression of *NKILA*, an NF- $\kappa$ B-interacting lncRNA, correlates with T cell apoptosis and shorter patient survival [18], and an enhancer-like lncRNA *NeST* regulates expression of *IFN- $\gamma$*  and induces its synthesis in CD8<sup>+</sup> T cells [19]. However, previous studies seem to be somewhat scattered, and the landscape and comprehensive functional analysis of lncRNAs in T cells in cancer immunity are still lacking.

The dramatic advances of single-cell RNA sequencing (scRNA-seq) technologies have gained unprecedented insight into the high diversity in T cell types and states compared to bulk RNA sequencing (RNA-seq) methods, which do not address the complex structures of the tumor microenvironment [20–25]. Despite the advantages of the single-cell resolution, currently, most scRNA-seq studies of cancer immunology have generally focused on coding genes, overlooking the large amounts of lncRNAs. Detailed understanding of lncRNAs at the single-cell level is challenging owing to their relatively low and cell-specific expression [26–28]. As a widely used scRNA-seq approach, 3'-end sequencing technologies such as droplet-based 10 $\times$  Genomics have lower RNA capture efficiencies, leading to dropout events and technological noise for lowly expressed lncRNAs [29]. Furthermore, accurate identification of novel lncRNAs is not suitable for the 3'-end sequencing technologies, but such analysis could be achieved by using full-length scRNA-seq technologies such as SMART-seq2 [30]. In addition, the sampling noise in scRNA-seq is generated through the sampling of limited RNA transcripts from each cell [31], leading to a highly noisy estimation for most lncRNAs. Therefore, to effectively characterize the lncRNA landscape at the single-cell level, attention should be paid to choosing the appropriate scRNA-seq data and analytical approaches.

Here, using unprecedentedly large-scale full-length single-cell transcriptome data of more than 20,000 T cells from various tissues across three cancer types, we created a full annotation of the T cell lncRNA transcriptome and analyzed the functional roles associated with different T cell

states. Our study aimed to provide a basic and valuable resource for the future exploration of lncRNA regulatory mechanisms in T cells, which may facilitate novel cancer-immune biomarker development.

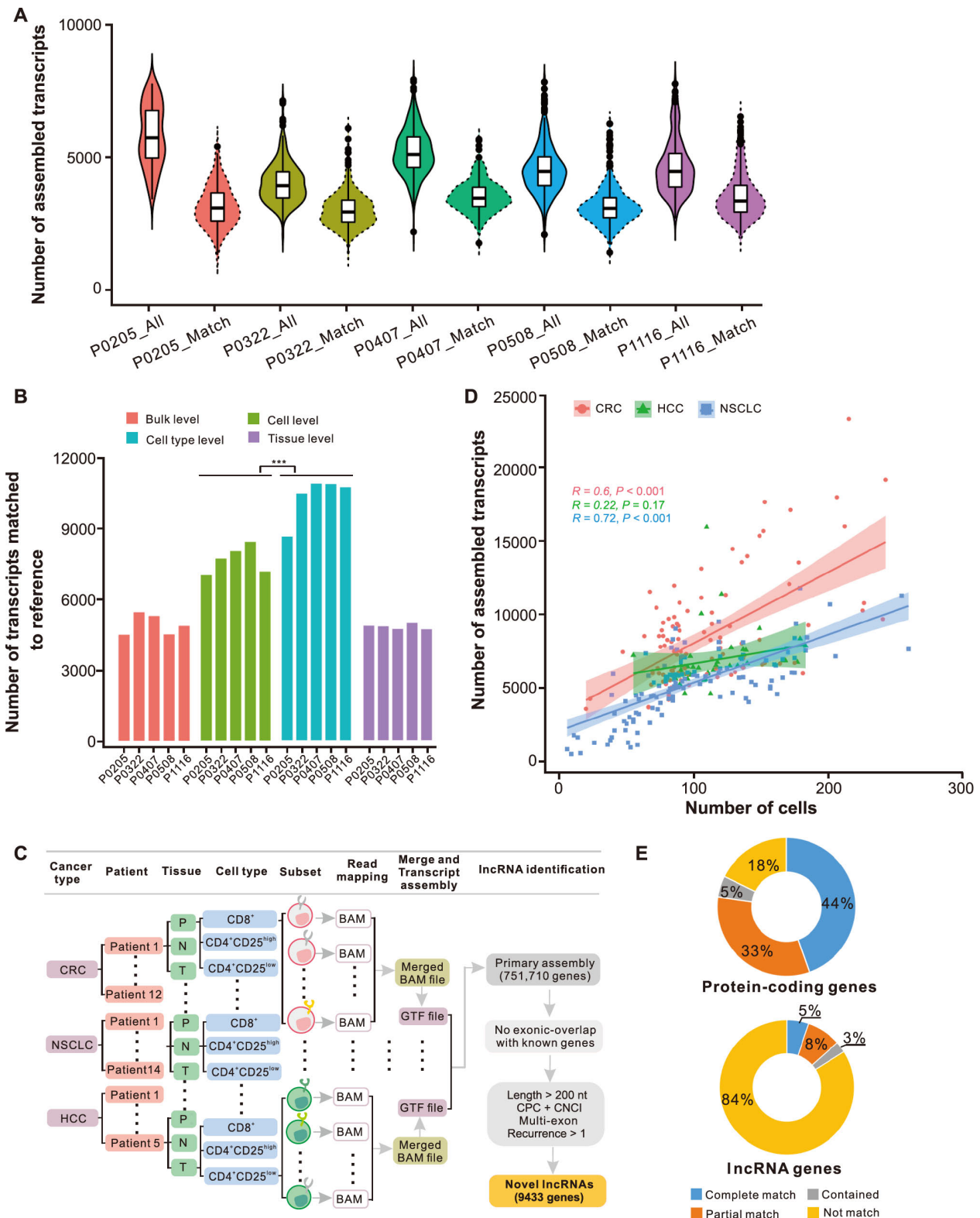
## Results

### *De novo* transcriptome assembly of lncRNAs from scRNA-seq data of T cells

To investigate the landscape of human lncRNAs in T cells across different tissues, patients, and cancer types, we collected the data of 24,068 T cells (the size of the gzip-compressed FASTQ file was 7.5 TB) generated by full-length scRNA-seq with SMART-seq2. This included the raw data of 9878 cells from 12 colorectal cancer (CRC) patients (2.8 TB), 10,188 cells from 14 non-small-cell lung cancer (NSCLC) patients (3.1 TB), and 4002 cells from 5 hepatocellular carcinoma (HCC) patients (1.6 TB) [32–34] (Figure S1A; Table S1). These cells were collected from peripheral blood, adjacent normal tissue, and tumor tissue from each patient and sorted into CD3<sup>+</sup>CD8<sup>+</sup> (CD8) and CD3<sup>+</sup>CD4<sup>+</sup> (CD4) T cells. The reads of each cell were mapped to the human reference genome (hg38/GRCh38), and the cells with unique mapping rates of less than 20% were removed. The remaining cells with on average 1.04 million uniquely mapped read pairs (0.63 million splices on average) and at least one pair of T cell receptor (TCR)  $\alpha$  and  $\beta$  chains enabled us to detect the expressed lncRNAs (Figure S1B–D).

Next, to generate the comprehensive T cell transcriptome beyond the current reference annotation, we performed *de novo* transcriptome assembly using the StringTie method [35]. Although StringTie could be run by providing the reference annotation to guide the transcript construction, in the current study, we focused on to what extent it could assemble the whole transcriptome without the prior annotation. Based on the T cell dataset from HCC patients, we first measured the extent of assembly in each T cell and found that an average of 4752 transcripts could be assembled at the single-cell level. An average of 69.8% (3318/4752) of those was matched to reference models (including reference protein-coding genes from GENCODE v31 and reference lncRNA genes from RefLnc database) (Figure 1A).

To explore the best way to obtain novel transcripts, we compared the assembly results using three different approaches based on the HCC dataset: 1) map and assemble transcripts for every single cell individually (cell-level); 2) assemble transcripts based on merged mapping results from each cell type of each patient (cell type-level); 3) assemble transcripts based on merged mapping results from each tissue of each patient (tissue-level). The transcripts assembled from each approach were merged independently



**Figure 1** Statistical analysis of assembled transcripts and workflow for novel lncRNA identification process in T cells during cancer immunity

**A.** Violin plots showing the number of assembled transcripts (All) and the number of those matched to the reference (Match) at single-cell level across five HCC patients (P0205, P0322, P0407, P0508, and P1116). **B.** Number of assembled transcripts matched to reference across five HCC patients based on four different strategies. \*\*\*,  $P < 0.001$  (Wilcoxon rank-sum test). **C.** Scheme of the pipeline used to identify novel lncRNAs expressed in T cells during cancer immunity using three full-length scRNA-seq datasets. **D.** Correlation of the number of cells and the number of assembled transcripts across different subsets for CRC, HCC, and NSCLC. 95% confidence intervals were added and shown as colored shades. **E.** The statistics of assembled transcripts that matched to reference protein-coding genes and reference lncRNA genes. CRC, colorectal cancer; HCC, hepatocellular carcinoma; NSCLC, non-small-cell lung cancer; P, peripheral blood; N, adjacent normal tissue; T, tumor tissue; CPC, Coding Potential Calculator; CNCI, Coding-Non-Coding Index.

and compared with reference genes respectively (Figure S1E). We found that the number of assembled transcripts matched to reference genes based on the cell type-level strategy (average 105,527 transcripts) was significantly higher than those based on the cell-level and tissue-level methods (average 77,860 and 49,689 transcripts, respectively;  $P < 0.001$ , Wilcoxon rank-sum test) (Figure 1B). Furthermore, the average number of matched transcripts from the cell type-level was more than twice that from the bulk-seq method (average 48,854 transcripts) (Figure 1B).

According to the cell type-level pooling strategy, the cells from all patients across three cancer types were partitioned into 266 subsets (Figure 1C, Figure S1A), and the mapping results of cells from the same subset were merged and fed into the assembling program. We found that the number of assembled transcripts across different subsets showed positive correlations with the number of cells in these subsets in both CRC and NSCLC datasets (CRC:  $R = 0.6$  and  $P = 4.3E-11$ ; NSCLC:  $R = 0.72$ ,  $P < 2.2E-16$ ), but not in the HCC dataset ( $R = 0.22$ ,  $P = 0.17$ ) (Figure 1D, Figure S1F). Then, assembled transcripts from all subsets were merged together, and a total of 751,710 primary genes were obtained. Next, we compared our assembled transcriptome with reference gene models. The results showed that reference lncRNA genes had a lower detection rate than protein-coding genes. Specifically, 82% (16,399/19,938) of the known protein-coding genes in GENCODE v31 could be verified, with 44% (8893/19,938) completely matched with the same intron chain, while 16% (9567/59,489) of known lncRNA genes were verified, with 5% (3140/59,489) completely matched (Figure 1E). These findings suggest that lncRNAs are expressed in a much more cell-specific manner than protein-coding genes, and further studies to uncover novel lncRNAs specifically expressed in human T cells are needed.

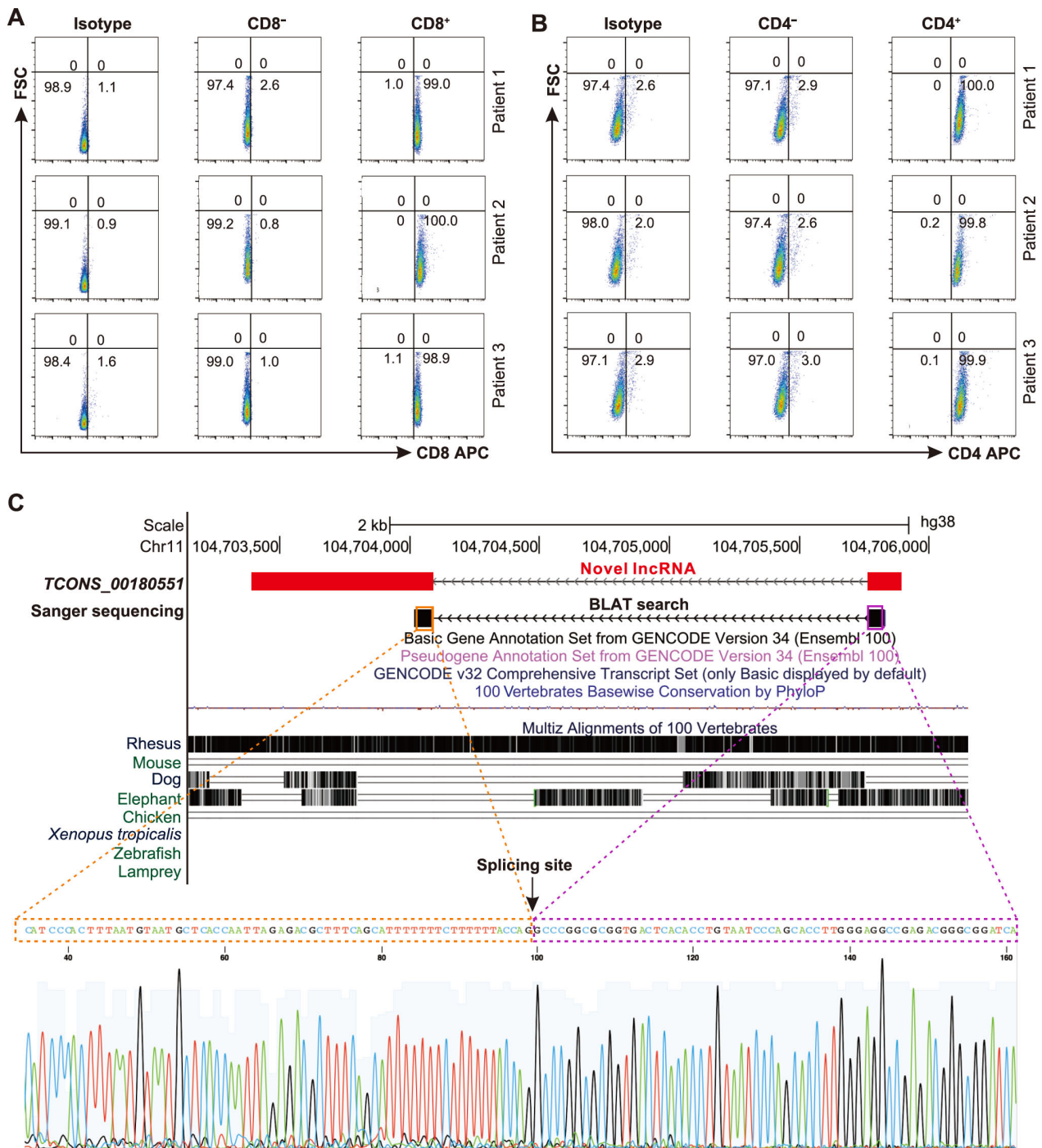
From the primary assembly, we developed a custom pipeline to identify novel lncRNAs. Briefly, we first selected transcripts that are more than 200 nt and have multiple exons. The transcripts that overlapped with known protein-coding or lncRNA genes were filtered out. Then, the transcripts that lacked coding potential predicted by both Coding Potential Calculator (CPC) [36] and Coding-Non-Coding Index (CNCI) [37] were retained. Finally, the remaining transcripts that were reconstructed in at least two subsets with complete match were defined as the novel lncRNA catalog (Figure 1C). Through this multi-layer analysis, we identified 9433 previously unknown lncRNA genes (13,025 transcripts with mean length of 1112 nt), which increased the number of current human lncRNA catalog [38] by 16% and nearly doubled the number of lncRNAs expressed in human T cells.

Finally, we performed experimental validation to evaluate the robustness of our identified novel lncRNAs.

First, fresh peripheral blood samples were collected from three CRC patients (Table S2). Then, mononuclear cells were isolated from each sample. The CD8 and CD4 T cells were separated by immunomagnetic beads, and the separation efficiency was verified by flow cytometry (Figure 2A and B). We then selected 50 novel lncRNA transcripts for qRT-PCR analysis and Sanger sequencing across T cell samples. As a result, 38 novel lncRNA transcripts were verified successfully by Sanger sequencing (Table S3). As an example, for a novel lncRNA *TCONS\_00180551* located in an intergenic region of chromosome 11, the BLAT search result of Sanger sequencing exactly matched the junction of this novel lncRNA (Figure 2C).

### Lower detection rates and expression levels of lncRNAs in T cells

Based on the relative genomic locations to reference protein-coding genes, the novel lncRNA transcripts were classified into three locus biotypes, including 6525 as intergenic, 3187 as intronic, and 3313 as antisense lncRNAs. As in the case of reference lncRNA genes, these novel lncRNA genes showed fewer exons (the average number of exons was 2) and lower detection rates and average gene abundance than protein-coding genes at the single-cell level (Figure 3A and B). Specifically, by using pseudoalignment of scRNA-seq reads to both reference and novel lncRNA transcriptomes, on average 5902 genes were detected (counts larger than 1) in each cell, 41% (2397) of which were lncRNA genes, including 1258 reference and 1139 novel lncRNA genes (Figure 3A). Furthermore, for both reference and novel lncRNA genes, the average number of expressed genes across T cells was significantly lower than that of protein-coding genes. More precisely, we found that an average of 5596 protein-coding and 2093 lncRNA genes were expressed in at least 25% of cells. In such a situation, novel lncRNA genes exhibited a higher average expression number and expression rate than reference lncRNA genes [1489 vs. 604; 15.8% (1489/9433) vs. 1% (604/59,489)] (Figure 3B), suggesting that novel lncRNAs exhibited more enrichment than known lncRNAs in T cells in cancer. Moreover, we performed further analysis to investigate the specifically expressed lncRNAs in different tissues for each cancer type. In brief, for CD4 and CD8 T cells of three cancer types, we totally identified 96 and 90 lncRNA genes (including 44 and 40 novel lncRNA genes) that were expressed in a tissue-specific pattern, respectively (Table S4). For example, some novel lncRNA genes such as *XLOC-301694* and *XLOC-126527* were significantly expressed in CD4 T cells from tumor tissue of CRC (adjusted  $P$  value =  $3.17E-68$  and  $1.72E-64$ , respectively), while others such as *XLOC-302096* and *XLOC-502999* were significantly enriched in CD4 T cells from normal tissue



**Figure 2** Validation of novel lncRNAs using qRT-PCR

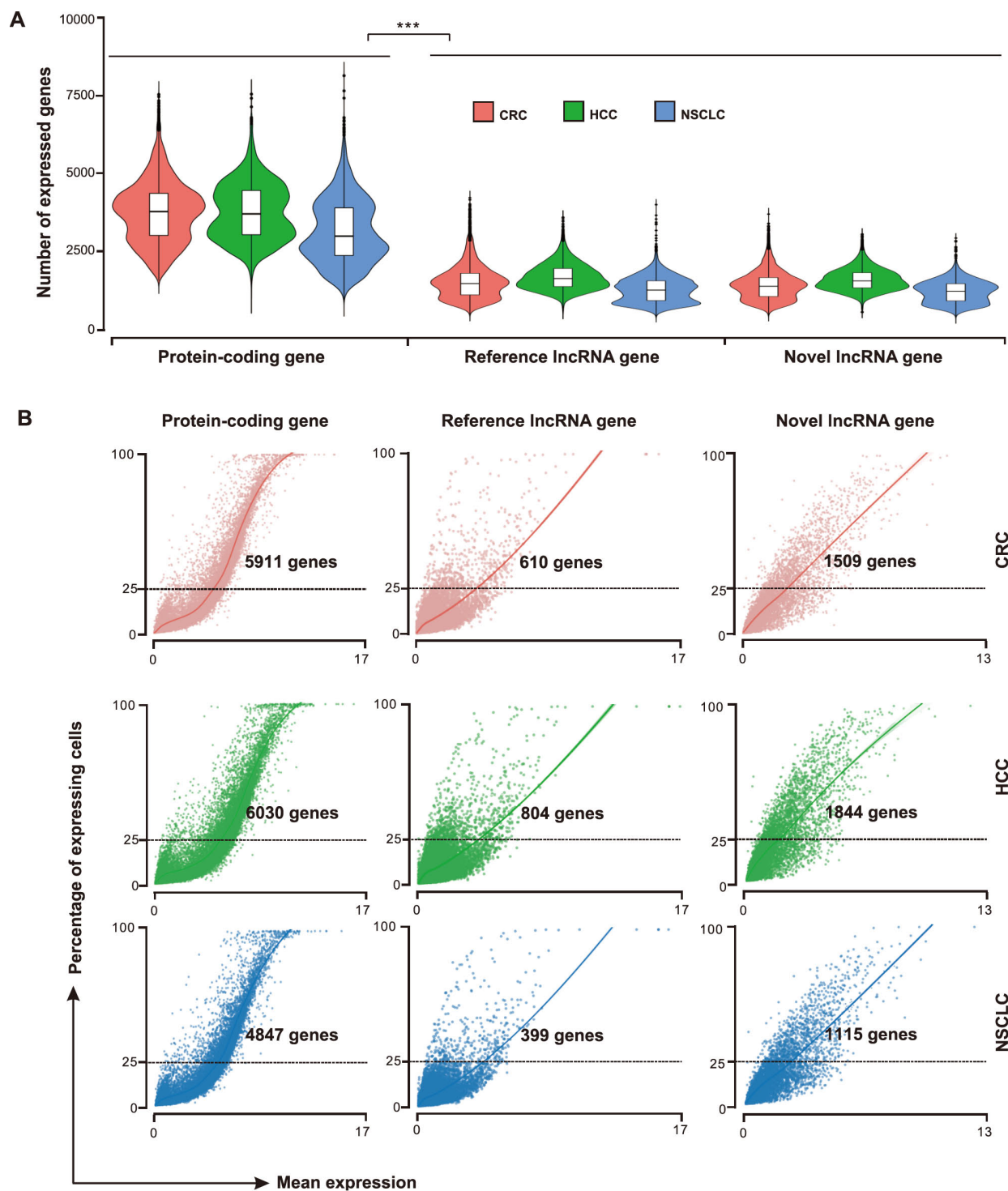
**A.** Flow cytometric analysis for CD8 T cells. T cells from three NSCLC patients were separated by magnetic beads and stained with antibody CD8-APC. **B.** Flow cytometric analysis for CD4 T cells. T cells from three NSCLC patients were separated by magnetic beads and stained with antibody CD4-APC. Isotype was used as a negative control. **C.** An example of novel intergenic lncRNA that was validated by Sanger sequencing. The genomic views are generated from the UCSC Genome Browser. The spliced sequence outputted by Sanger sequencing is shown.

and peripheral blood of CRC, respectively (adjusted  $P$  value =  $9.18E-82$  and  $1.35E-44$ , respectively) (Table S4). Finally, we assessed the evolutionary conservation of these novel lncRNA transcripts and found that, on average, 61.2% have orthologous regions in the primate genomes, while only 3.4% are mapped to the mouse genome, suggesting the poor

sequence conservation of these novel lncRNA transcripts.

### More than one hundred signature lncRNAs associated with T cell states were identified based on metacell maps

To explore signature lncRNAs associated with T cell states



**Figure 3** Expression features of reference genes and novel lncRNA genes at the single-cell level

**A.** Number of protein-coding, reference lncRNA, and novel lncRNA genes expressed in T cells across three cancer types. \*\*\*,  $P < 0.001$  (Wilcoxon rank-sum test). **B.** Plots showing the percentage of expressing cells against the mean expression level (log counts) for protein-coding, reference lncRNA, and novel lncRNA genes across three cancer types. The number of genes that are expressed in at least 25% of the cells is provided in the figure.

in cancer immunity, we used the MetaCell method [31] that partitioned the scRNA-seq datasets into disjointed and homogeneous cell groups (namely metacells) using the non-parametric  $K$ -nn graph algorithm. For the low and specific expression nature of lncRNA genes, metacells

pooling together data from cells derived from the same transcriptional states could serve as building blocks for approximating the distributions of lncRNA gene expression and minimizing the technical variance and noise. After quality control, 19,572 cells with predefined cluster

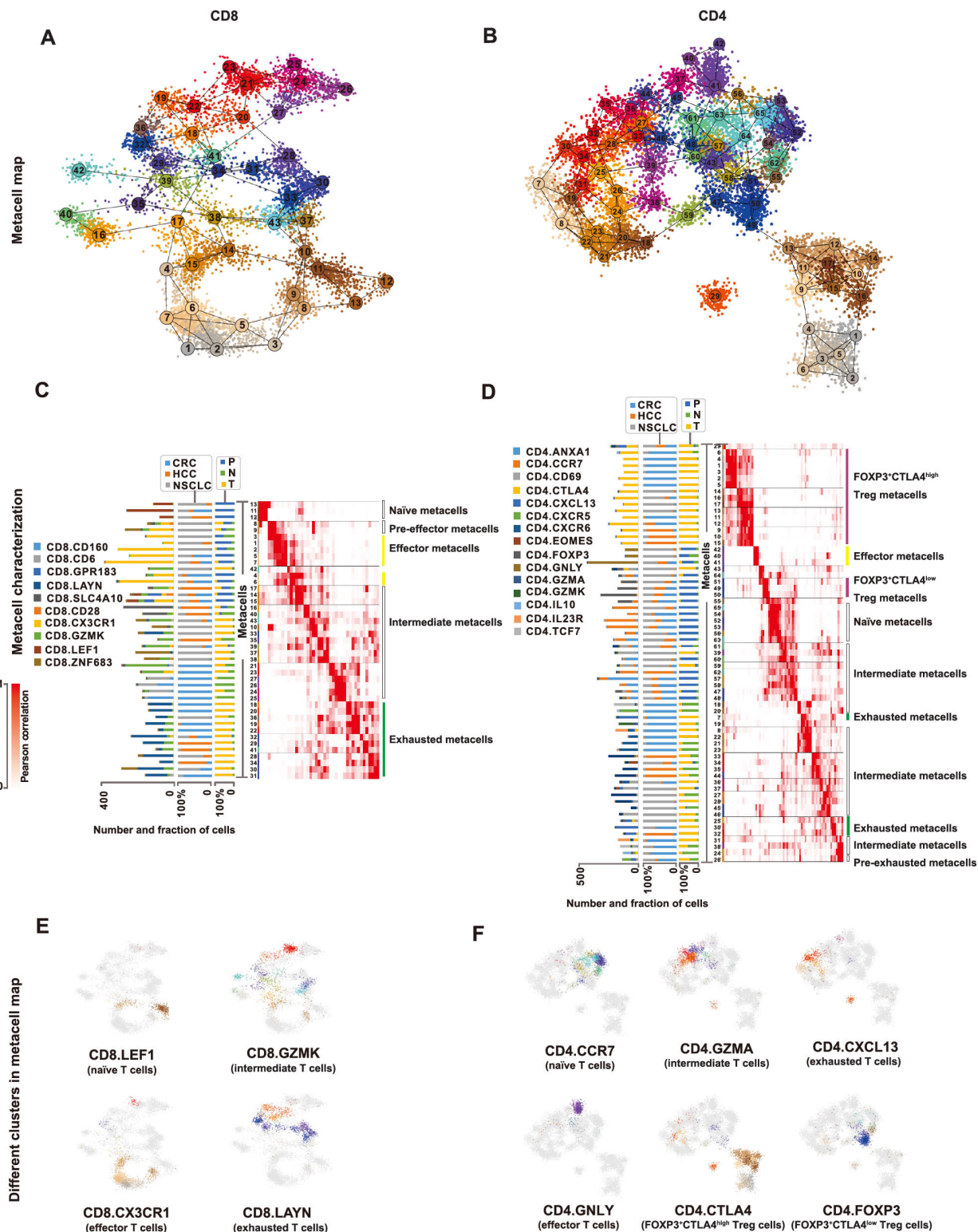
annotations and 21,205 genes, including both protein-coding and lncRNA genes, were retained and used for the following analyses. The expression tables of CD8 and CD4 T cells across three cancers (Tables S5 and S6) were fed into the MetaCell pipeline separately, resulting in detailed maps of 43 and 65 metacells, respectively (Figure 4A and B).

Based on the 2D projections (Figure 4A and B), the pre-defined cell cluster annotations (Table S1), and the metacell similarity matrices (similarity among 43 or 65 metacells for CD8 or CD4 T cells) (Figure 4C and D, Figure S2A and B), we organized the complex transcriptional landscape of CD8 T cells into naïve, effector/pre-effector, intermediate, and exhausted metacell groups and that of CD4 T cells into naïve, effector, intermediate, exhausted/pre-exhausted, and regulatory (including FOXP3<sup>+</sup>CTLA4<sup>low</sup> and FOXP3<sup>+</sup>CTLA4<sup>high</sup>) metacell groups (Figure 4C and D). To evaluate the composition of metacells, we mapped tissue- and cancer-specific patterns in all metacells and achieved results in accordance with previous studies [32–34] (Figure 4C and D, Figures S3 and S4). As an example, exhausted metacells were preferentially enriched in tumors, while effector metacells were prevalent in peripheral blood. Although some metacells were enriched in different cancer types, they were organized into the same functional groups (Figure 4C and D). Notably, effector metacell groups (cytotoxic state) and exhausted metacell groups (dysfunctional state) were located in different directions in the metacell maps, while the diffuse border was observed between the intermediate state and the cytotoxic or dysfunctional state (Figure 4E and F). These intermediate cells exhibited remarkable transcriptional heterogeneity indicating functional divergence of these cells (Figure 4E and F, Figures S3 and S4). The observed cluster distribution in both CD8 and CD4 metacell maps might suggest a relative transition from activation to exhaustion that began with naïve cells, followed by intermediate cells, such as central memory (CM), effector memory (EM), and tissue-resident memory (RM) cells, and ended with exhausted cells. Moreover, the CD4 metacell map revealed that Tregs were subdivided into FOXP3<sup>+</sup>CTLA4<sup>low</sup> Tregs and FOXP3<sup>+</sup>CTLA4<sup>high</sup> Tregs that were preferentially enriched in peripheral blood and tumors, respectively (Figure 4D and F). These observations demonstrated that the diversity and dynamics of T cell states in cancer immune infiltrates could be controlled by complex and intricate gene regulatory mechanisms. The association between these cell states and lncRNAs is still poorly characterized, prompting us to subsequently investigate the potential roles of lncRNA genes in T cells. Currently, a few cell groups, such as FOXP3<sup>+</sup>CTLA4<sup>high</sup> Tregs and exhausted T cells (both expressing inhibitory receptors, e.g., PDCD1 and TIGIT), have been used as therapeutic targets for anti-cancer immunotherapies [1–3], thus we focused on these cells in the following analyses.

To explore signature lncRNAs associated with effector T cells, exhausted T cells, and Tregs, we performed a systematic analysis of these metacell groups based on well-defined anchor genes [39], such as the genes associated with CD8 effector functions (*CX3CR1*, *FGFBP2*, *GZMH*, and *PRF1*) or with the CD8 exhausted state (*HAVCR2*, *LAG3*, *PDCD1*, *TIGIT*, and *CTLA4*). As a result, 154 lncRNA genes were identified to be significantly correlated to the anchor genes, which were involved in a set of co-expressed gene modules, including effector, exhausted, and Treg gene modules (Figure 5A and B; Table S7). Interestingly, a putative CTLA4<sup>high</sup> Treg gene subset was observed in the Treg gene module, suggesting its specific functional role in tumor-infiltrating Treg cells (Figure 5B). Overall, combined with the expression profile across metacell groups, we found 47 and 79 lncRNA genes correlated with effector and exhausted states in CD8/CD4 T cells, respectively, which were designated as effector and exhausted signature lncRNAs, respectively (Figure 5C, Figure S5). Similarly, 49 lncRNA genes were highly associated with Tregs and were designated as Treg signature lncRNAs (Figure S5). Among these signature lncRNA genes, 14 were shared between CD8 and CD4 effector states; 7 were shared between CD8 and CD4 exhausted states; 21 lncRNA genes associated with Tregs overlapped with those characteristics in the exhausted CD4 T cells (Table S7), indicating the presence of shared regulatory roles of these lncRNAs. In contrast, no signature lncRNA was shared between exhausted and effector states.

### The functions of 84 signature lncRNAs were annotated by co-expression networks

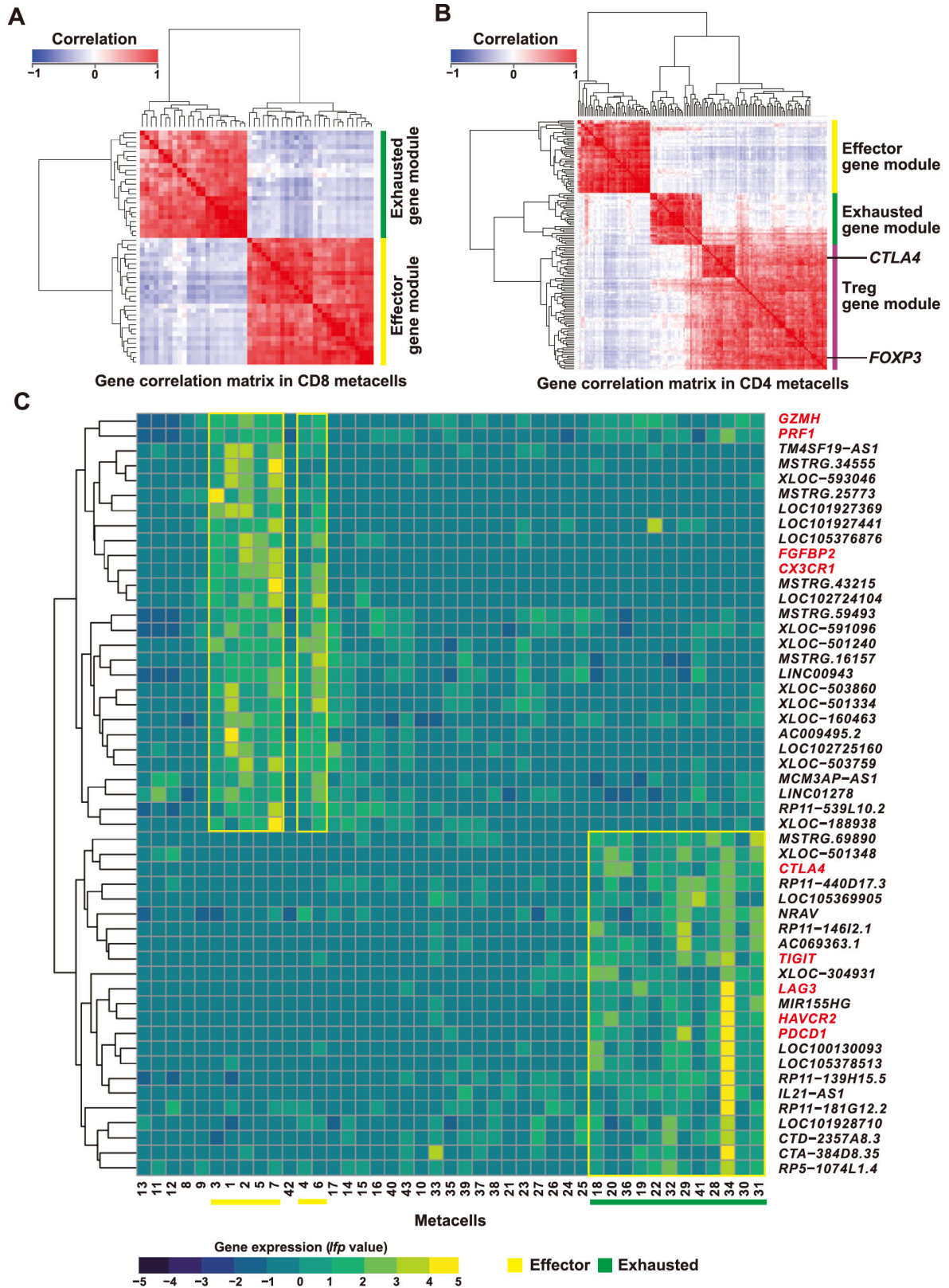
To gain further insights into the functional roles of signature lncRNAs in different T cell states in cancer, we built a coding-noncoding network (CNC), as we previously reported [40,41], using linear correlation over all metacells. Applying this strategy, the functions of 54% (84/154) signature lncRNAs were annotated (Table S8). As expected, both CD8 and CD4 exhausted T cells had the functional enrichments of signature lncRNAs that were markedly different from CD8 and CD4 effector T cells, respectively, including regulation manners in immune system processes and several signaling pathways (Figure 6A and B). For example, exhausted signature lncRNAs were significantly enriched in immunoinhibitory functions such as negative regulation of immune response (adjusted *P* value = 2.96E–14), negative regulation of T cell activation (adjusted *P* value = 1.24E–06), and positive regulation of interleukin-10 biosynthetic process (adjusted *P* value = 1.02E–18). In comparison, effector signature lncRNAs were enriched in cytotoxic programs such as T cell proliferation involved in immune response (adjusted *P* value = 8.16E–09), positive regulation of cytokine secretion (adjusted *P* value =



**Figure 4** Characterization of T cell states based on the 2D projection of T cells and the annotation of metacell maps

**A.** 2D projection of CD8 T cells from three cancer types into 43 metacells. **B.** 2D projection of CD4 T cells from three cancer types into 65 metacells. In panels A and B, metacells are indicated as nodes, which are numbered (1–43) and color-coded for differentiation. Cells are shown as small data points, which are color-coded according to the metacells they belong to. **C.** CD8 metacells (rows) ordered by groups and organized within each group. **D.** CD4 metacells (rows) ordered by groups and organized within each group. The left bar plot shows the number of cells of different clusters in each metacell. The middle and right bar plots show the percentage of cells from different cancer types (CRC, HCC, and NSCLC) and tissues (P, N, and T) in each metacell, respectively. Heatmaps show the confusion matrices (the pairwise similarities between metacells) for CD8 or CD4 metacells. The annotations of different metacell groups are shown on the right. **E.** 2D projections of the composition of CD8 T cells from different clusters. **F.** 2D projections of the composition of CD4 T cells from different clusters. In panels E and F, cells from different clusters are colored-coded according to the metacells they belong to.





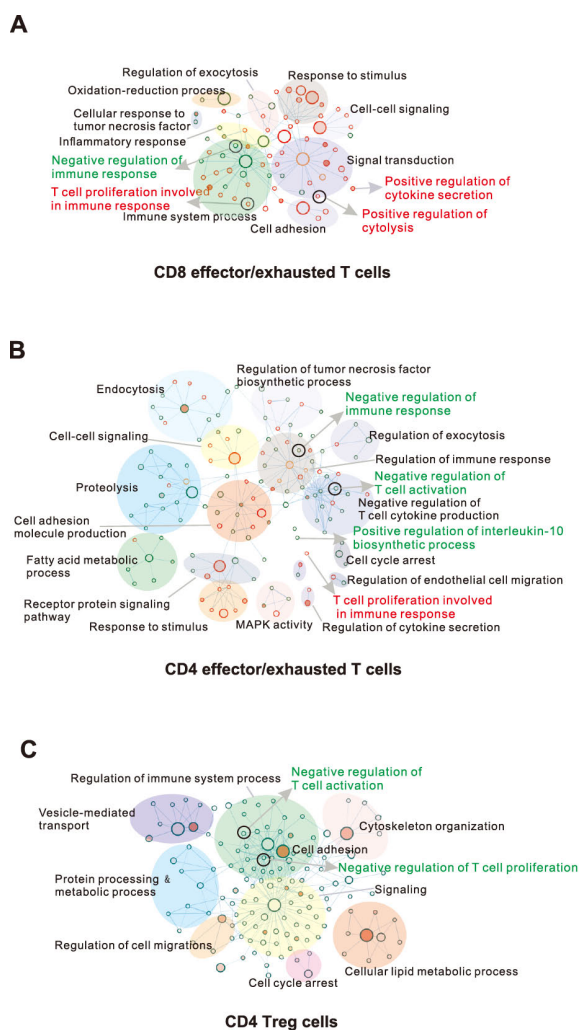
**Figure 5** Correlation and expression analyses of signature lncRNAs associated with different T cell states

**A.** Pearson correlation heatmaps for signature lncRNA genes and anchor genes in CD8 metacells. **B.** Pearson correlation heatmaps for signature lncRNA genes and anchor genes in CD4 metacells. The signature gene modules and two anchor genes (*CTLA4* and *FOXP3*) are labeled on the right. **C.** Expression (log fold enrichment values; *lfp* values) of signature lncRNA genes and anchor genes across CD8 metacells. Signature lncRNA genes and anchor genes are marked in black and red on the right, respectively. Metacells are numbered at the bottom. Metacell groups associated with effector and exhausted functions are underlined with yellow and green bars, respectively.

4.65E–05), and positive regulation of cytolysis (adjusted  $P$  value = 1.59E–19) (Figure 6A and B; Tables S9 and S10). These results are consistent with the phenotypes of exhausted or effector states of T cells as described in previous studies [1,32–34,42]. In addition, the enriched functions of Treg signature lncRNAs were similar with those of CD4 exhausted signature lncRNAs involving multiple immunosuppressive programs (Figure 6C; Table S11), suggesting the shared regulatory roles of these lncRNAs in CD4 Tregs and exhausted T cells. Further analysis of the functions of co-signature lncRNAs that were shared between CD8 and CD4 exhausted or effector states, as well as between CD4 exhausted and Treg states (Figure S6), sug-

gested that the signature lncRNAs might broadly participate in the regulation of T cell functions within the human tumor microenvironment.

For example, a known lncRNA *TM4SF19-AS1*, defined as a signature lncRNA for both CD8 and CD4 effector T cells, was transcribed in the antisense orientation to the *TM4SF19* gene and co-expressed with 66 protein-coding genes and 11 lncRNA genes (Figure 7A and B). Of note, *TM4SF19-AS1* was highly correlated and located in the same topologically associated domain (TAD) with its host gene *TM4SF19* (Pearson correlation coefficient = 0.88) (Figure 7A), a member of the four-transmembrane L6 superfamily participating in various cellular processes including cell proliferation, motility, and cell adhesion [43–46]. Consistently, *TM4SF19-AS1* was significantly enriched in several effector T cell-associated processes such as cellular response to cholesterol (adjusted  $P$  value = 1.09E–30), cell adhesion (adjusted  $P$  value = 5.25E–27), and regulation of tumor necrosis factor biosynthetic process (adjusted  $P$  value = 3.75E–11) (Figure 7C). Interestingly, a recent study suggested that the anti-tumor response of CD8 T cells could be enhanced by regulating cholesterol metabolism [47]. For another example, a novel lncRNA *XLOC-633950*, defined as a signature lncRNA for both CD4 exhausted T cells and Tregs, was an intergenic gene and transcribed from the promoter-enhancer cluster region of the *SLA* and *CCN4* genes (Figure 7D). Furthermore, *XLOC-633950*, as a novel gene whose expression was supported by multiple expressed sequence tags (ESTs), was located in the same TAD with the *SLA* gene, which acted as an inhibitor of antigen receptor signaling by negative regulation of positive selection and mitosis of T cells [48–51] (Figure 7D). In accordance with *SLA* functions, the functional enrichments of *XLOC-633950* according to its co-expressed protein-coding genes were mainly associated with immunoinhibitory processes, such as negative regulation of T cell cytokine production (adjusted  $P$  value = 4.56E–13) and negative regulation of T cell proliferation and activation (adjusted  $P$  value = 7.25E–11 and 5.85E–08, respectively) (Figure 7E and F). These results provided a starting point for future dissecting the mechanisms of signature lncRNAs.

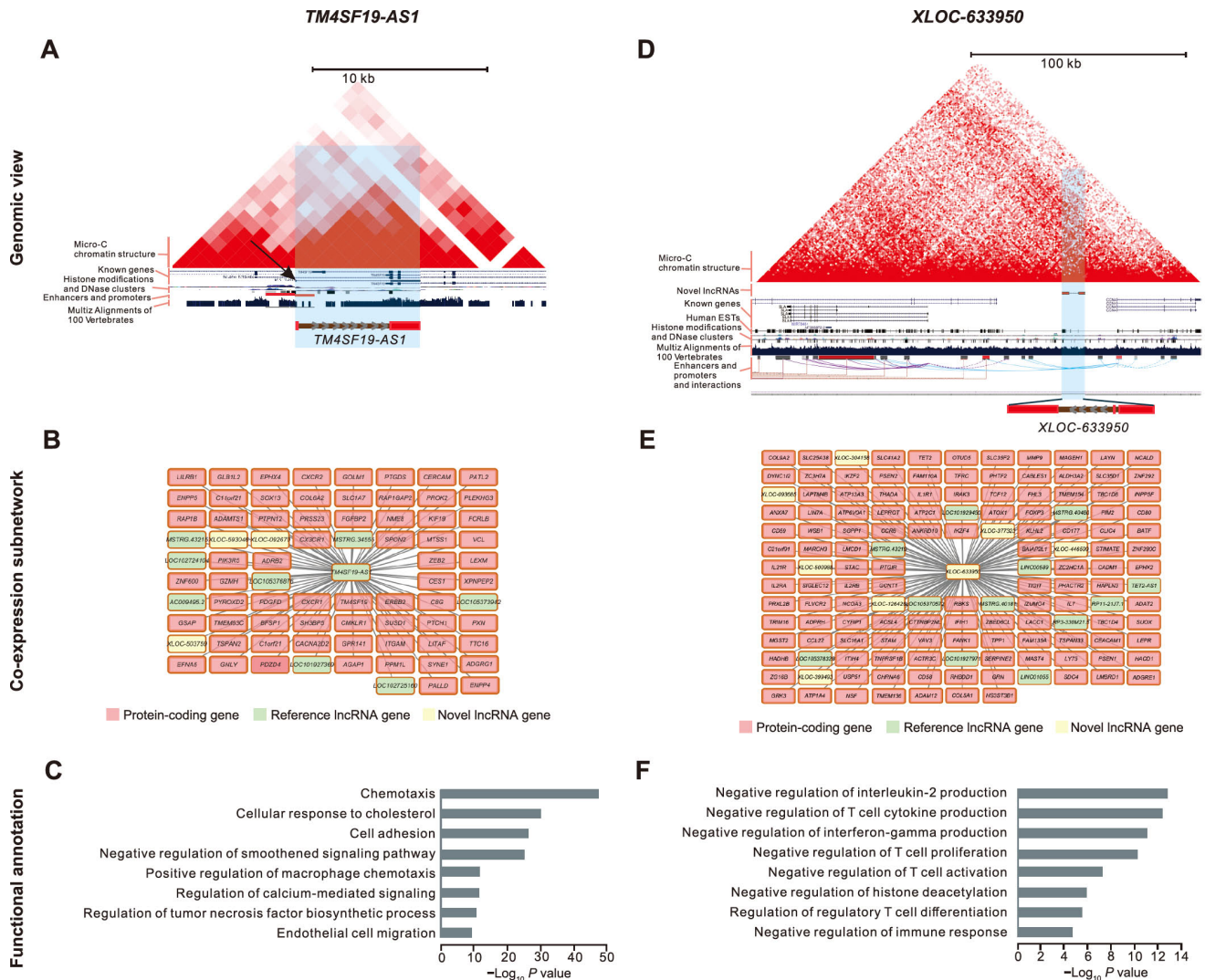


**Figure 6** Functional annotation of signature lncRNAs

**A.** Functional enrichment maps of signature lncRNAs for CD8 effector/exhausted T cells. **B.** Functional enrichment maps of signature lncRNAs for CD4 effector/exhausted T cells. **C.** Functional enrichment maps of signature lncRNAs for CD4 Treg cells. The enriched gene sets from Gene Ontology based on the predicted functions of signature lncRNA genes are visualized by Cytoscape plugin Enrichment Map. Each node represents a gene set; the size of the node is indicative of the number of genes and the color intensity of the node reflects the level of significance. Effector signature gene sets are shown in red circles, exhausted or Treg ones in green circles, and the common gene sets in orange circles. Maps are magnified differently for easy visualization.

## Discussion

Despite the obvious advantages, most scRNA-seq datasets were still limited in their ability to study lncRNAs, which have emerged as central players and key regulators in a number of biological processes such as anti-tumor immune response [52,53]. In comparison with many scRNA-seq methods that amplify only the 3' end of transcripts, the SMART-seq2 protocol could generate full-length cDNA from polyadenylated transcripts and thus is suitable for



**Figure 7 Genomic and functional characterization of example signature lncRNAs**

A. Genomic view of a known effector signature lncRNA *TM4SF19-AS1*. The genomic view is generated from the UCSC Genome Browser. B. Co-expressed genes of *TM4SF19-AS1*. Co-expressed protein-coding genes, reference lncRNA genes, and novel lncRNA genes are colored in pink, light green, and light yellow, respectively. C. Functional annotations of *TM4SF19-AS1* based on co-expression network. D. Genomic view of a novel exhausted signature lncRNA *XLOC-633950*. The genomic view is generated from the UCSC Genome Browser. E. Co-expressed genes of *XLOC-633950*. Co-expressed protein-coding genes, reference lncRNA genes, and novel lncRNA genes are colored in pink, light green, and light yellow, respectively. F. Functional annotations of *XLOC-633950* based on co-expression network.

analysis of lncRNAs [30,54]. In the current study, we performed systematic analyses of SMART-seq2 full-length scRNA-seq datasets and provided the first comprehensive atlas of lncRNAs in T cells of human cancer.

Recently, Jiang et al. [38] presented a comprehensive human lncRNA catalog (RefLnc) containing 77,900 lncRNAs based on analysis of 14,166 polyA(+) RNA-seq libraries and previously known annotations. Among the RefLnc lncRNAs, only 16% could be assembled and expressed in T cells. In addition, compared with bulk RNA-seq data, scRNA-seq data could detect more known and novel transcripts. These observations suggested that despite the vast number of lncRNAs that have been identified using bulk RNA-seq data [10,12,26,38,55], the catalog of human

lncRNAs is still far from being complete at single-cell resolution due to their low and cell-specific expression patterns. Based on the cell-pooling strategy and more than 20,000 scRNA-seq libraries from 31 patients across three cancer types, we identified 9433 previously non-annotated lncRNA genes. These results significantly expand the current lncRNA catalog and enable us to carry out in-deep analysis of the T cell context-specific lncRNA transcriptome. It should be noted that the StringTie method we used to assemble transcripts was not designed for single-cell data and may decrease the sensitivity of lncRNA detection at the single-cell level. To effectively identify lncRNAs, the reads from 90 cells on average were pooled together and then were fed into the StringTie program. The number of

unique mapped reads for each cell set were equivalent to those produced by canonical RNA-seq data. Nevertheless, further evaluation of StringTie performance on such pooling reads is needed. Furthermore, all the scRNA-seq data used in the current study were generated by sequencing the polyadenylated transcriptome, in which non-polyadenylated lncRNAs were absent.

Several previous studies have applied full-length scRNA-seq to unleash tumor-infiltrating lymphocytes in HCC [34], NSCLC [32], and CRC [33], providing a deep understanding of the immune landscape of T cells in cancer. Nevertheless, the physiological functions of lncRNAs in different T cell states during the cancer immune response remain elusive. Although the abundance of lncRNAs is relatively low and hard to distinguish from technical noise in single T cells, pooling the transcripts from multiple cells that are derived from the same cell state allows more accurate quantification of lncRNAs, making it feasible to explore their signatures and putative regulatory mechanisms associated with T cell states in cancer immunity. Based on such partitioning and pooling strategies, in this study, we used the MetaCell method to identify homogeneous T cell groups from scRNA-seq data and derived detailed maps of 43 and 65 metacells for CD8 and CD4 T cells, respectively. These metacells with higher homogeneity allowed a more accurate quantification of lncRNAs as well as identification of T cell differentiation gradients. For example, we observed 7 metacells involved in CD8 effector cell cluster, which might reflect the transcriptional heterogeneity in this cluster (Figure 4C). The roles of lncRNAs in these different subsets (metacells) of CD8 effector T cells need further investigation. While MetaCell was not designed to perform single-cell lncRNA analysis, the MetaCell partitioning algorithm facilitated robust cell grouping of scRNA-seq data, which enabled us to study lncRNAs more accurately.

According to the metacell maps (Figure 4E and F), in contrast to the pool of intermediate T cells with diffuse borders with other cell states, a discrete pool of effector T cells, exhausted T cells, and Tregs was observed that showed clear gaps among them, thus facilitating unbiased analysis of signature lncRNAs in these cell states. In total, 154 signature lncRNA genes were obtained, providing a useful reference lncRNA resource to further investigate their functions in T cell-mediated cancer immunity. Since lncRNAs generally interact with protein-coding genes, and highly correlated genes generally have similar functions, the putative functions of these signature lncRNA genes could be predicted by the co-expressed protein-coding genes. Therefore, by constructing the ‘two color’ co-expression network in which both protein-coding and lncRNA genes were involved, the functions of 84 signature lncRNA genes were annotated. Some lncRNAs were genomically co-located with their host genes, which revealed the compli-

cated regulation mechanisms of lncRNAs in cancer immunity. For example, as described above, *TM4SF19-AS1* was both co-expressed and co-located with its host gene *TM4SF19*, whose family had functions in various biological processes, including cell proliferation and cell adhesion that were consistent with the characteristics of effector T cells [43–46].

Our study has a few limitations which could be addressed in the future. On the one hand, limited by the sources of high-quality SMART-seq2 data, only three cancer types were studied in this work. On the other hand, other types of immune cells involved in tumor, such as dendritic cells, natural killer cells, and tumor-associated macrophages, were not included. To depict a more comprehensive atlas of lncRNAs in cancer immune regulation at single-cell level, more cancer types and immune cells should be further explored in the future.

In summary, the current study provides the first comprehensive catalog and the functional repertoires of lncRNAs in human cancer T cells. Although the expression patterns and exact mechanisms of these signature lncRNAs in regulating T cell states need further experimental validation, we provide the groundwork for future studies of the functional mechanisms of lncRNAs in the T cell-mediated cancer immunity, especially in two of the essential states of T cells: effector state and exhausted state. The signature lncRNAs of CD8 exhausted T cells and tumor Tregs may serve as new targets for novel cancer-immune biomarker development and cancer immunotherapies.

## Materials and methods

### Full-length scRNA-seq and bulk RNA-seq datasets from cancer patients

Raw sequencing data of three compendium datasets used in the current study were authorized by the European Genome-phenome Archive (EGA) and obtained from the EGA database (EGA: EGAS00001002791, EGAS00001002430, and EGAS00001002072). The CRC scRNA-seq dataset contains the raw data of 11,138 single T cells isolated from different tissues (peripheral blood, adjacent normal, and tumor tissues) of 12 CRC patients [33]. The NSCLC scRNA-seq dataset contains the raw data of 12,346 single T cells from 14 NSCLC patients [32]. The HCC scRNA-seq dataset contains the raw data of 5063 single T cells from 6 HCC patients [34]. All the data were generated by Illumina HiSeq 2500 sequencer with 100 bp pair-end reads or Illumina HiSeq 4000 sequencer with 150 bp pair-end reads. The cells from HCC patient P1202 (TCRs could not be assembled in those cells) were not analyzed in the current study. After preliminary filtration, 24,068 T cells with at least one pair of TCR  $\alpha$  and  $\beta$  chains were retained. The bulk

RNA-seq data of five tumor samples from HCC patients were obtained from the HCC dataset.

According to the cell annotations from original papers [32–34], these T cells were classified into different subtypes (Figure S1A; Table S1). PTC, NTC, and TTC represent CD3<sup>+</sup>CD8<sup>+</sup> T cells that were isolated from peripheral blood, adjacent normal, and tumor tissues, respectively. PTH, NTH, and TTH represent CD3<sup>+</sup>CD4<sup>+</sup>CD25<sup>low</sup> T cells that were isolated from the three tissues. PTR, NTR, and TTR represent CD3<sup>+</sup>CD4<sup>+</sup>CD25<sup>high</sup> T cells that were isolated from the three tissues. PPQ, NPQ, and TPQ represent CD4<sup>+</sup> T cells that were isolated from the three tissues. PTY, NTY, and TTY represent CD4<sup>+</sup>CD25<sup>int</sup> T cells that were isolated from the three tissues.

### Read mapping and transcript assembling

Clean reads from each T cell were mapped to the human reference genome (version hg38/GRCh38) using STAR aligner (v2.7.1) [56] with the *twopassMode* set as Basic. The BAM files of T cells from each cell type of each patient were merged using SAMtools merge [57]. StringTie (v2.0.3) [35] was used to assemble transcripts based on genomic read alignments. Assembled transcripts of all cell types across all patients were merged together using the Cuffmerge utility of the Cufflinks package [58].

### Comparison with reference gene annotation

For reference gene annotation, lncRNA genes were collected from RefLnc [38], and other genes were collected from GENCODE v31 [59]. According to the “class code” information outputted by Cuffcompare, the merged assembly was classified into four categories by comparison with the reference gene annotation, including known protein-coding genes, known lncRNA genes, potentially novel genes (class code is “i, x, u”), and others.

### Identification of novel lncRNAs

Based on the potentially novel gene catalog derived from single-cell data, we developed a custom pipeline for the identification of reliable novel lncRNAs that included the following steps: 1) transcripts that were more than 200 nt and had more than one exon were selected for downstream analysis (for intergenic transcripts, at least 1 kb away from known protein-coding genes); 2) transcripts that were predicted to lack coding potential by both CPC [36] and CNCI [37] were regarded as candidate non-coding transcripts and retained; 3) the remaining transcripts that were assembled and had the same intron chain of at least two cell types were retained as the final novel lncRNA catalog. The final lncRNA catalog was obtained by combining the reference

lncRNA and novel lncRNA genes directly. The UCSC liftOver tool ([http://genome.ucsc.edu/cgi-bin/hgLiftOver?hgid=806106955\\_h2xhcK2iPRI7SiMkxkB41I2mwF90](http://genome.ucsc.edu/cgi-bin/hgLiftOver?hgid=806106955_h2xhcK2iPRI7SiMkxkB41I2mwF90)) was used to identify the orthologous locations of human novel lncRNAs in the mouse genome and in the genomes of primates such as chimpanzee and gorilla, with the parameters: Minimum ratio of bases that must remap = 0.1 and Minimum ratio of alignment blocks or exons that must map = 0.5.

### Experimental validation of novel lncRNAs

Three CRC patients were enrolled at Shenzhen People’s Hospital, Shenzhen, China. The clinical characteristics of three patients are summarized in Table S2. Peripheral blood samples from those patients were obtained and treated with anticoagulation. Peripheral blood mononuclear cells (PBMCs) were extracted by Ficoll-Paque Plus (Catalog No. 17144003, GE Healthcare, Pittsburgh, PA). Then, CD8 and CD4 T cells were separated by immunomagnetic beads (Catalog No. 130045101, Miltenyi Biotec, Bergis-Gladbach, Germany). The separation efficiency was verified by flow cytometry. The sorted cells were dissolved in Trizol Reagent (Catalog No. 15596026, Ambion, Boston, MA) for RNA extraction according to the manufacturer’s protocol. cDNA was synthesized by PrimerScript RT Reagent Kit (Catalog No. AHG1552A, Takara, Kyoto, Japan). We chose 50 novel lncRNA transcripts to perform experimental validation according to the following criteria: 1) highly expressed in either CD8 or CD4 T cells; 2) reconstructed in at least ten subsets with complete match; 3) uniquely mapped to the human genome. For each lncRNA transcript, at least two pairs of primers for qRT-PCR were designed using NCBI Primer-BLAST (<https://www.ncbi.nlm.nih.gov/tools/primer-blast>). In order to ensure the specificity of primers, UCSC InSilicon PCR (<http://genome.ucsc.edu/cgi-bin/hgPcr>) was used to compare the primer pairs with the human genome (hg38). Some primer pairs were specifically designed to span splicing sites (exon junctions). qRT-PCR was performed with SYBR Green Master Mix (Catalog No. RR820A, Takara) on an ABI StepOnePlus (Applied Biosystems, Boston, MA). *GAPDH* as a housekeeping gene was used as the positive control. For each lncRNA, we selected one qRT-PCR product for Sanger sequencing.

### Data normalization

We calculated the read counts and transcripts per million (TPM) values using pseudoalignment of scRNA-seq reads to both protein-coding and lncRNA transcriptomes, as implemented in Kallisto (v0.46.0) [60] with default parameters, and summarized expression levels from the transcript level to the gene level.

Low-quality and doublet cells were removed if the number of expressed genes (counts more than 1) was fewer than 2000 or higher than the medians of all cells plus  $3\times$  the median absolute deviation. Moreover, the cells with the proportion of reads mapped to mitochondrial genes larger than 10% were discarded. Genes with average counts more than 1 and expressed in at least 1% of cells for each type of cancer were retained. The combined count tables from all T cells passing the aforementioned filtration were normalized using a pooling and deconvolution method implemented in the R package named ComputeSumFactors [61] with the sizes ranged from 80, 100, 120, to 140. According to the assumption that most genes were not differentially expressed, normalization was performed within each predefined cluster separately to compute cell size factors. The cell size factors were rescaled by normalization among clusters. Finally, the counts for each cell were normalized by dividing the cell counts by the cell size factor.

### Construction of metacell maps

The MetaCell method [31], which partitioned the scRNA-seq dataset into disjointed and homogeneous cell groups (metacells) using the  $K$ -nn graph algorithm, was performed for both the CD8 and CD4 T cells independently. We first removed specific mitochondrial genes (annotated with the prefix “MT-”) that typically mark cells as being stressed or dying, rather than cellular identity. Based on the count matrices of both protein-coding and lncRNA genes, feature genes whose scaled variance (variance/mean on down-sampled matrices) exceeded 0.08 were selected and used to compute cell-to-cell similarity using Pearson correlations. According to the cell-to-cell similarity matrices, two balanced  $K$ -nn similarity graphs for CD8 and CD4 T cells were constructed using the parameter  $K = 100$  (the number of neighbors for each cell was limited by  $K$ ). Next, we performed the resampling procedures (resampling 75% of the cells in each iteration with 500 iterations) and co-clustering graph construction (the minimal cluster size was 50). Finally, the graphs of metacells (and the cells belonging to them) were projected into 2D spaces to explore the similarities between cells and metacells.

### Annotation of metacells

Annotation of metacells was performed based on the metacell confusion matrix and predefined cluster annotations (Table S1) of T cells involved in the metacells. Briefly, we first created a hierarchical clustering of metacells according to the number of similarity relationships between their cells. Next, we generated clusters of metacells as confusion matrices based on the hierarchy results, and then annotated these clusters according to the annotations of T cells.

### Defining signature lncRNAs associated with T cell states

To identify signature lncRNAs associated with effector and exhausted T cells as well as Tregs, as described in a recent study [39], we adopted the anchor approach by identifying the lncRNAs that were significantly correlated to well-defined anchor genes, based on their expression levels (log fold enrichment scores;  $lfp$  values calculated by the MetaCell method) in metacells. The lncRNAs that significantly correlated with anchor genes (adjusted  $P$  value  $< 0.01$  and ranked in the top 5 percentile for each anchor gene) were regarded as signature lncRNAs. The anchor genes were defined as follows. The anchor genes of CD8 exhausted T cells include *HAVCR2*, *LAG3*, *PDCD1*, *TIGIT*, and *CTLA4*; the anchor genes of CD8 effector T cells include *CX3CR1*, *FGFBP2*, *GZMH*, and *PRF1*; genes associated with Tregs include *FOXP3*. The anchor genes of CD4 exhausted T cells include *CXCL13*, *PDCD1*, *TIGIT*, *CTLA4*, and *HAVCR2*; genes associated with CD4 effector T cells include *GNLY*, *GZMB*, *GZMH*, *PRF1*, and *NKG7*.

### Functional prediction of signature lncRNAs based on co-expression networks

Based on  $lfp$  values of both lncRNA and protein-coding genes across all metacells, we used a custom pipeline for large-scale prediction of signature lncRNA functions by constructing the coding-lncRNA gene co-expression networks [40,41]. Briefly, genes with log enrichment scores ranked in the top 75% of each metacell were retained. Then,  $P$  values of Pearson correlation coefficients for each gene pair were calculated based on the Fisher's asymptotic test using the WGCNA package of R.  $P$  values were adjusted based on the Bonferroni multiple test correction using the MULTTEST package of R. The gene pairs with an adjusted  $P$  value  $< 0.01$ , Pearson correlation coefficient  $> 0.7$ , and ranked in the top 5% for each gene were involved in the co-expression network.

Based on the co-expression network, lncRNA functions were predicted using module- and hub-based methods. Specifically, the Markov Cluster algorithm was adopted to identify co-expressed modules [40]. For each module, if the known genes were significantly enriched for at least one Gene Ontology (GO) term, the functions of the lncRNAs involved in the module were assigned as the same ones. For the hub-based method, the functions of a hub lncRNA (node degree  $> 10$ ) were assigned, if its immediate neighboring genes were significantly enriched for at least one GO term.

### Data availability

All the novel lncRNA genes identified in the current study and

their expression files are available in the NONCODE database ([http://noncode.org/datadownload/Novel\\_lncRNA\\_in\\_T\\_cells\\_gtf.gz](http://noncode.org/datadownload/Novel_lncRNA_in_T_cells_gtf.gz)).

## Ethical statement

This study was approved by the Institutional Ethics Committee at Shenzhen People's Hospital, Shenzhen, China. The written informed consent was obtained from patients.

## CRedit author statement

**Haitao Luo:** Conceptualization, Methodology, Software, Visualization, Formal analysis, Writing - original draft, Writing - review & editing, Funding acquisition. **Dechao Bu:** Conceptualization, Software, Visualization. **Lijuan Shao:** Investigation, Resources. **Yang Li:** Resources. **Liang Sun:** Software, Funding acquisition. **Ce Wang:** Data curation. **Jing Wang:** Data curation. **Wei Yang:** Data curation. **Xiaofei Yang:** Data curation. **Jun Dong:** Conceptualization, Validation, Supervision, Writing - review & editing. **Yi Zhao:** Conceptualization, Validation, Supervision, Writing - review & editing. **Furong Li:** Conceptualization, Validation, Supervision, Funding acquisition, Writing - review & editing. All authors have read and approved the final manuscript.

## Competing interests

The authors have declared no competing interests.

## Acknowledgments

This work was supported by the Science and Technology Project of Shenzhen, China (Grant Nos. JCYJ20190807145013281, JHZ20170310090257380, JCYJ20170413092711058, and JCYJ20170307095822325), the China Postdoctoral Science Foundation (Grant No. 2019M663369), and the National Natural Science Foundation of China (Grant No. 31970636). We thank Drs. Lei Zhang and Yao He (Peking University, China) for their assistance with the raw data download. The data analysis was performed on Amazon Web Services (AWS) China region, and we thank Mr. Hansen Huang, Chao Wu, and Fei Shi for providing the technical service and support. We thank Mr. Geoffrey Pearce for his suggestions on the writing of this paper.

## Supplementary material

Supplementary data to this article can be found online at

<https://doi.org/10.1016/j.gpb.2021.02.006>.

## ORCID

0000-0003-3671-7786 (Haitao Luo)  
 0000-0002-8833-5432 (Dechao Bu)  
 0000-0001-7980-2003 (Lijuan Shao)  
 0000-0001-8199-8527 (Yang Li)  
 0000-0002-5213-6941 (Liang Sun)  
 0000-0003-3048-7596 (Ce Wang)  
 0000-0001-7856-4533 (Jing Wang)  
 0000-0003-2138-5563 (Wei Yang)  
 0000-0001-6255-5181 (Xiaofei Yang)  
 0000-0002-6064-0814 (Jun Dong)  
 0000-0001-6046-8420 (Yi Zhao)  
 0000-0002-0606-8861 (Furong Li)

## References

- [1] Chen DS, Mellman I. Elements of cancer immunity and the cancer-immune set point. *Nature* 2017;541:321–30.
- [2] Bodor JN, Bumber Y, Borghaei H. Biomarkers for immune checkpoint inhibition in non-small cell lung cancer (NSCLC). *Cancer* 2020;126:260–70.
- [3] Houot R, Schultz LM, Marabelle A, Kohrt H. T-cell-based immunotherapy: adoptive cell transfer and checkpoint inhibition. *Cancer Immunol Res* 2015;3:1115–22.
- [4] Restifo NP, Dudley ME, Rosenberg SA. Adoptive immunotherapy for cancer: harnessing the T cell response. *Nat Rev Immunol* 2012;12:269–81.
- [5] Hilmi M, Vienot A, Rousseau B, Neuzillet C. Immune therapy for liver cancers. *Cancers* 2019;12:77.
- [6] Carter JA, Gilbo P, Atwal GS. IMPRES does not reproducibly predict response to immune checkpoint blockade therapy in metastatic melanoma. *Nat Med* 2019;25:1833–5.
- [7] Hugo W, Zaretsky JM, Sun L, Song C, Moreno BH, Hui-Lieskovan S, et al. Genomic and transcriptomic features of response to anti-PD-1 therapy in metastatic melanoma. *Cell* 2016;165:35–44.
- [8] Auslander N, Zhang G, Lee JS, Frederick DT, Miao B, Moll T, et al. Robust prediction of response to immune checkpoint blockade therapy in metastatic melanoma. *Nat Med* 2018;24:1545–9.
- [9] Rinn JL, Chang HY. Genome regulation by long noncoding RNAs. *Annu Rev Biochem* 2012;81:145–66.
- [10] Fang SS, Zhang LL, Guo JC, Niu YW, Wu Y, Li H, et al. NONCODEV5: a comprehensive annotation database for long non-coding RNAs. *Nucleic Acids Res* 2018;46:D308–14.
- [11] Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, et al. GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res* 2012;22:1760–74.
- [12] Iyer MK, Niknafs YS, Malik R, Singhal U, Sahu A, Hosono Y, et al. The landscape of long noncoding RNAs in the human transcriptome. *Nat Genet* 2015;47:199–208.
- [13] Yu WD, Wang H, He QF, Xu Y, Wang XC. Long noncoding RNAs in cancer-immunity cycle. *J Cell Physiol* 2018;233:6518–23.
- [14] Wang L, Felts SJ, Van Keulen VP, Scheid AD, Block MS, Markovic SN, et al. Integrative genome-wide analysis of long non-coding RNAs in diverse immune cell types of melanoma patients. *Cancer Res* 2018;78:4411–23.
- [15] Xu J, Cao X. Long noncoding RNAs in the metabolic control of inflammation and immune disorders. *Cell Mol Immunol* 2019;16:1–5.

- [16] Spurlock CF, Crooke PS, Aune TM. Biogenesis and transcriptional regulation of long noncoding RNAs in the human immune system. *J Immunol* 2016;197:4509–17.
- [17] Fanucchi S, Fok ET, Dalla E, Shibayama Y, Börner K, Chang EY, et al. Immune genes are primed for robust transcription by proximal long noncoding RNAs located in nuclear compartments. *Nat Genet* 2019;51:138–50.
- [18] Huang D, Chen J, Yang L, Ouyang Q, Li J, Lao L, et al. NKILA lncRNA promotes tumor immune evasion by sensitizing T cells to activation-induced cell death. *Nat Immunol* 2018;19:1112–25.
- [19] Gomez JA, Wapinski OL, Yang YW, Bureau JF, Gopinath S, Monack DM, et al. The NeST long ncRNA controls microbial susceptibility and epigenetic activation of the interferon- $\gamma$  locus. *Cell* 2013;152:743–54.
- [20] Ren X, Kang B, Zhang Z. Understanding tumor ecosystems by single-cell sequencing: promises and limitations. *Genome Biol* 2018;19:211.
- [21] Chung W, Eum HH, Lee HO, Lee KM, Lee HB, Kim KT, et al. Single-cell RNA-seq enables comprehensive tumour and immune cell profiling in primary breast cancer. *Nat Commun* 2017;8:15081.
- [22] Lavin Y, Kobayashi S, Leader A, Amir EAD, Elefant N, Bigenwald C, et al. Innate immune landscape in early lung adenocarcinoma by paired single-cell analyses. *Cell* 2017;169:750–65.e17.
- [23] Zhang Q, He Y, Luo N, Patel SJ, Han Y, Gao R, et al. Landscape and dynamics of single immune cells in hepatocellular carcinoma. *Cell* 2019;179:829–45.e20.
- [24] Savas P, Virassamy B, Ye C, Salim A, Mintoff CP, Caramia F, et al. Single-cell profiling of breast cancer T cells reveals a tissue-resident memory subset associated with improved prognosis. *Nat Med* 2018;24:986–93.
- [25] Azizi E, Carr AJ, Plitas G, Cornish AE, Konopacki C, Prabhakaran S, et al. Single-cell map of diverse immune phenotypes in the breast tumor microenvironment. *Cell* 2018;174:1293–308.e36.
- [26] Uszczyńska-Ratajczak B, Lagarde J, Frankish A, Guigó R, Johnson R. Towards a complete map of the human long non-coding RNA transcriptome. *Nat Rev Genet* 2018;19:535–48.
- [27] Stegle O, Teichmann SA, Marioni JC. Computational and analytical challenges in single-cell transcriptomics. *Nat Rev Genet* 2015;16:133–45.
- [28] Vallejos CA, Risso D, Scialdone A, Dudoit S, Marioni JC. Normalizing single-cell RNA sequencing data: challenges and opportunities. *Nat Methods* 2017;14:565–71.
- [29] Zheng GXY, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, et al. Massively parallel digital transcriptional profiling of single cells. *Nat Commun* 2017;8:14049.
- [30] Picelli S, Faridani OR, Björklund AK, Winberg G, Sagasser S, Sandberg R. Full-length RNA-seq from single cells using Smart-seq2. *Nat Protoc* 2014;9:171–81.
- [31] Baran Y, Bercovich A, Sebe-Pedros A, Lubling Y, Giladi A, Chomsky E, et al. MetaCell: analysis of single-cell RNA-seq data using *K*-nn graph partitions. *Genome Biol* 2019;20:206.
- [32] Guo X, Zhang Y, Zheng L, Zheng C, Song J, Zhang Q, et al. Global characterization of T cells in non-small-cell lung cancer by single-cell sequencing. *Nat Med* 2018;24:978–85.
- [33] Zhang L, Yu X, Zheng L, Zhang Y, Li Y, Fang Q, et al. Lineage tracking reveals dynamic relationships of T cells in colorectal cancer. *Nature* 2018;564:268–72.
- [34] Zheng C, Zheng L, Yoo JK, Guo H, Zhang Y, Guo X, et al. Landscape of infiltrating T cells in liver cancer revealed by single-cell sequencing. *Cell* 2017;169:1342–56.e16.
- [35] Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol* 2015;33:290–5.
- [36] Kong L, Zhang Y, Ye ZQ, Liu XQ, Zhao SQ, Wei L, et al. CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res* 2007;35:W345–9.
- [37] Sun L, Luo H, Bu D, Zhao G, Yu K, Zhang C, et al. Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic Acids Res* 2013;41:e166.
- [38] Jiang S, Cheng SJ, Ren LC, Wang Q, Kang YJ, Ding Y, et al. An expanded landscape of human long noncoding RNA. *Nucleic Acids Res* 2019;47:7842–56.
- [39] Li H, van der Leun AM, Yofe I, Lubling Y, Gelbard-Solodkin D, van Akkooi ACJ, et al. Dysfunctional CD8 T cells form a proliferative, dynamically regulated compartment within human melanoma. *Cell* 2019;176:775–89.e18.
- [40] Liao Q, Liu C, Yuan X, Kang S, Miao R, Xiao H, et al. Large-scale prediction of long non-coding RNA functions in a coding–non-coding gene co-expression network. *Nucleic Acids Res* 2011;39:3864–78.
- [41] Luo H, Bu D, Sun L, Fang S, Liu Z, Zhao Y. Identification and function annotation of long intervening noncoding RNAs. *Brief Bioinform* 2017;18:789–97.
- [42] Wherry EJ, Kurachi M. Molecular and cellular insights into T cell exhaustion. *Nat Rev Immunol* 2015;15:486–99.
- [43] Wright MD, Ni J, Rudy GB. The L6 membrane proteins—a new four-transmembrane superfamily. *Protein Sci* 2000;9:1594–600.
- [44] Chang YW, Chen SC, Cheng EC, Ko YP, Lin YC, Kao YR, et al. CD13 (aminopeptidase N) can associate with tumor-associated antigen L6 and enhance the motility of human lung cancer cells. *Int J Cancer* 2005;116:243–52.
- [45] Lekishvili T, Fromm E, Mujoomdar M, Berditchevski F. The tumour-associated antigen L6 (L6-Ag) is recruited to the tetraspanin-enriched microdomains: implication for tumour cell motility. *J Cell Sci* 2008;121:685–94.
- [46] Allioli N, Vincent S, Vlaeminck-Guillem V, Decaussin-Petrucci M, Ragage F, Ruffion A, et al. *TM4SF1*, a novel primary androgen receptor target gene over-expressed in human prostate cancer and involved in cell migration. *Prostate* 2011;71:1239–50.
- [47] Yang W, Bai Y, Xiong Y, Zhang J, Chen S, Zheng X, et al. Potentiating the antitumour response of CD8<sup>+</sup> T cells by modulating cholesterol metabolism. *Nature* 2016;531:651–5.
- [48] Marton N, Baricza E, Érsek B, Buzás EI, Nagy G. The emerging and diverse roles of Src-like adaptor proteins in health and disease. *Mediators Inflamm* 2015;2015:1–9.
- [49] Sosinowski T, Pandey A, Dixit VM, Weiss A. Src-like adaptor protein (SLAP) is a negative regulator of T cell receptor signaling. *J Exp Med* 2000;191:463–74.
- [50] Pandey A, Ibarrola N, Kratchmarova I, Fernandez MM, Constantinescu SN, Ohara O, et al. A novel Src homology 2 domain-containing molecule, Src-like adapter protein-2 (SLAP-2), which negatively regulates T cell receptor signaling. *J Biol Chem* 2002;277:19131–8.
- [51] Park SK, Beaven MA. Mechanism of upregulation of the inhibitory regulator, src-like adaptor protein (SLAP), by glucocorticoids in mast cells. *Mol Immunol* 2009;46:492–7.
- [52] Yao RW, Wang Y, Chen LL. Cellular functions of long noncoding RNAs. *Nat Cell Biol* 2019;21:542–51.
- [53] Agirre X, Meydan C, Jiang Y, Garate L, Doane AS, Li Z, et al. Long non-coding RNAs discriminate the stages and gene regulatory states of human humoral immune response. *Nat Commun* 2019;10:821.
- [54] Liu F, Zhang Y, Zhang L, Li Z, Fang Q, Gao R, et al. Systematic comparative analysis of single-nucleotide variant detection methods from single-cell RNA sequencing data. *Genome Biol* 2019;20:242.
- [55] Quek XC, Thomson DW, Maag JLV, Bartonicek N, Signal B, Clark MB, et al. lncRNAdb v2.0: expanding the reference database for functional long noncoding RNAs. *Nucleic Acids Res* 2015;43:D168–73.
- [56] Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 2013;29:15–21.



- 
- [57] Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. *Bioinformatics* 2009;25:2078–9.
- [58] Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* 2010;28:511–5.
- [59] Frankish A, Diekhans M, Ferreira AM, Johnson R, Jungreis I, Loveland J, et al. GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res* 2019;47:D766–73.
- [60] Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* 2016;34:525–7.
- [61] Lun AT, Bach K, Marioni JC. Pooling across cells to normalize single-cell RNA sequencing data with many zero counts. *Genome Biol* 2016;17:75.