

# Tumour heterogeneity and intercellular networks of nasopharyngeal carcinoma at single cell resolution

Yang Liu<sup>1,11</sup>, Shuai He<sup>1,2,3,11</sup>, Xi-Liang Wang<sup>4,11</sup>, Wan Peng<sup>1</sup>, Qiu-Yan Chen<sup>1</sup>, Dong-Mei Chi<sup>1,5</sup>, Jie-Rong Chen<sup>1,6</sup>, Bo-Wei Han<sup>1</sup>, Guo-Wang Lin<sup>1,7</sup>, Yi-Qi Li<sup>1</sup>, Qian-Yu Wang<sup>1</sup>, Rou-Jun Peng<sup>1</sup>, Pan-Pan Wei<sup>1</sup>, Xiang Guo<sup>1</sup>, Bo Li<sup>8,9</sup>, Xiaojun Xia<sup>1</sup>, Hai-Qiang Mai<sup>1</sup>, Xue-Da Hu<sup>4</sup>, Zemin Zhang<sup>4,10</sup>, Yi-Xin Zeng<sup>1</sup> & Jin-Xin Bei<sup>1,2,3</sup>

The heterogeneous nature of tumour microenvironment (TME) underlying diverse treatment responses remains unclear in nasopharyngeal carcinoma (NPC). Here, we profile 176,447 cells from 10 NPC tumour-blood pairs, using single-cell transcriptome coupled with T cell receptor sequencing. Our analyses reveal 53 cell subtypes, including tumour-infiltrating CD8<sup>+</sup> T, regulatory T (Treg), and dendritic cells (DCs), as well as malignant cells with different Epstein-Barr virus infection status. Trajectory analyses reveal exhausted CD8<sup>+</sup> T and immune-suppressive TNFRSF4<sup>+</sup> Treg cells in tumours might derive from peripheral CX3CR1<sup>+</sup>CD8<sup>+</sup> T and naïve Treg cells, respectively. Moreover, we identify immune-regulatory and tolerogenic LAMP3<sup>+</sup> DCs. Noteworthy, we observe intensive inter-cell interactions among LAMP3<sup>+</sup> DCs, Treg, exhausted CD8<sup>+</sup> T, and malignant cells, suggesting potential cross-talks to foster an immune-suppressive niche for the TME. Collectively, our study uncovers the heterogeneity and interacting molecules of the TME in NPC at single-cell resolution, which provide insights into the mechanisms underlying NPC progression and the development of precise therapies for NPC.

<sup>1</sup> Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Guangdong Key Laboratory of Nasopharyngeal Carcinoma Diagnosis and Therapy, Guangzhou 510060, People's Republic of China. <sup>2</sup> The Eighth Affiliated Hospital, Sun Yat-sen University, Shenzhen 518033, People's Republic of China. <sup>3</sup> Center for Precision Medicine, Sun Yat-sen University, Guangzhou 510080, People's Republic of China. <sup>4</sup> BIOPIIC, Beijing Advanced Innovation Center for Genomics, School of Life Sciences, Peking University, Beijing 100871, People's Republic of China. <sup>5</sup> Department of Anesthesiology, Sun Yat-sen University Cancer Center, Guangzhou 510060, People's Republic of China. <sup>6</sup> Guangdong General Hospital, Guangdong Academy of Medical Sciences, Guangzhou 510080, People's Republic of China. <sup>7</sup> Department of Laboratory Medicine, Zhujiang Hospital, Southern Medical University, Guangzhou, Guangdong 510282, People's Republic of China. <sup>8</sup> Department of Biochemistry and Molecular Biology, Zhongshan School of Medicine, Sun Yat-sen University, Guangzhou 510080, People's Republic of China. <sup>9</sup> RNA Biomedical Institute, Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou 510120, People's Republic of China. <sup>10</sup> Peking-Tsinghua Center for Life Sciences, Academy for Advanced Interdisciplinary Studies, Peking University, Beijing 100871, People's Republic of China. <sup>11</sup> These authors contributed equally: Yang Liu, Shuai He, Xi-Liang Wang. ✉email: [zemin@pku.edu.cn](mailto:zemin@pku.edu.cn); [zengyx@susucc.org.cn](mailto:zengyx@susucc.org.cn); [beijx@susucc.org.cn](mailto:beijx@susucc.org.cn)

**N**asopharyngeal carcinoma (NPC) is a distinct type of head and neck cancer, which has been closely linked with the infection of Epstein-Barr virus (EBV)<sup>1</sup>. NPC has a remarkable ethnic and geographic prevalence, where high incidence rate of 15–50 cases per 100,000 people was reported in Southern China and Southeast Asia, as compared to 0.4 per 100,000 in western populations<sup>2</sup>. In general, patients with NPC are diagnosed with advanced stages largely due to non-specific symptoms. Radiotherapy is the primary treatment modality for patients with NPC because of the radiosensitive nature of its tumour cells<sup>2</sup>. Survival outcomes of patients with NPC improve substantially, reaching a 5-year overall survival rate of 85.6%<sup>3,4</sup>, mainly benefited from the evolution of radiotherapy techniques and the addition of platinum-based chemotherapy in patients with loco-regionally advanced disease. Nevertheless, more than 10% of patients develop recurrent and metastatic NPC<sup>2</sup>, for whom recent studies showed an overall response rate of 11.7% and 25.9–34% to targeted therapy with the addition of an inhibitor for epidermal growth factor receptor<sup>5</sup> and immunotherapies using the immune checkpoint blockade<sup>6,7</sup>, respectively. These variations in treatment responses and survival outcomes indicate the heterogeneous nature of NPC.

Individual genetic makeup and genomic instability foster genetic diversity of cancer cells that contribute to tumour heterogeneity. Genome sequencing analyses have revealed diverse profiles of somatic alterations in NPC tumours, with high mutational frequencies at *CYLD*, *NFKBIA*, *TP53*, and *CDKN2A/B*, as well as accumulated mutations in MHC class I genes and chromatin modification genes, which were associated with poor overall survival of the patients<sup>8,9</sup>. Apart from heterogeneous cancer cells, tumours exhibit another dimension of heterogeneity, which contain diverse normal cells creating the tumour microenvironment (TME) for the maintenance of cancer hallmarks<sup>10</sup>. Heterogeneous immune cells and stromal cells have been characterised using transcriptional profiling at single-cell resolution in several cancers, revealing that certain subtypes of immune cells and gene signatures in TME are important for tumour progression and sustained treatment responses<sup>11,12</sup>. Profound infiltration of lymphocytes has been observed in histological biopsies of NPC, amid other stromal cells and tumour cells of different morphology<sup>13,14</sup>. Moreover, high density of tumour infiltrating lymphocytes (TILs) was associated with favourable survival outcomes of patients with NPC<sup>15,16</sup>. However, the composition of diverse cell populations in the TME has not been well illustrated in NPC. Two recent studies demonstrated the presence of T cells with various functional states and different immune cells in NPC tumours, using single-cell transcriptome analysis<sup>17,18</sup>.

In this work, we aim to provide a comprehensive global view of tumour heterogeneity of NPC, by analysing the single-cell transcriptional profiles of 176,447 cells from 10 treatment-naïve patients with NPC. With the combination of T cell receptor (TCR) repertoire information and individual tumour-blood sample pairs, we further characterise the clonality and migrations of T cells. In addition, we generate a potential cellular interaction network of cell populations in the TME of NPC.

## Results

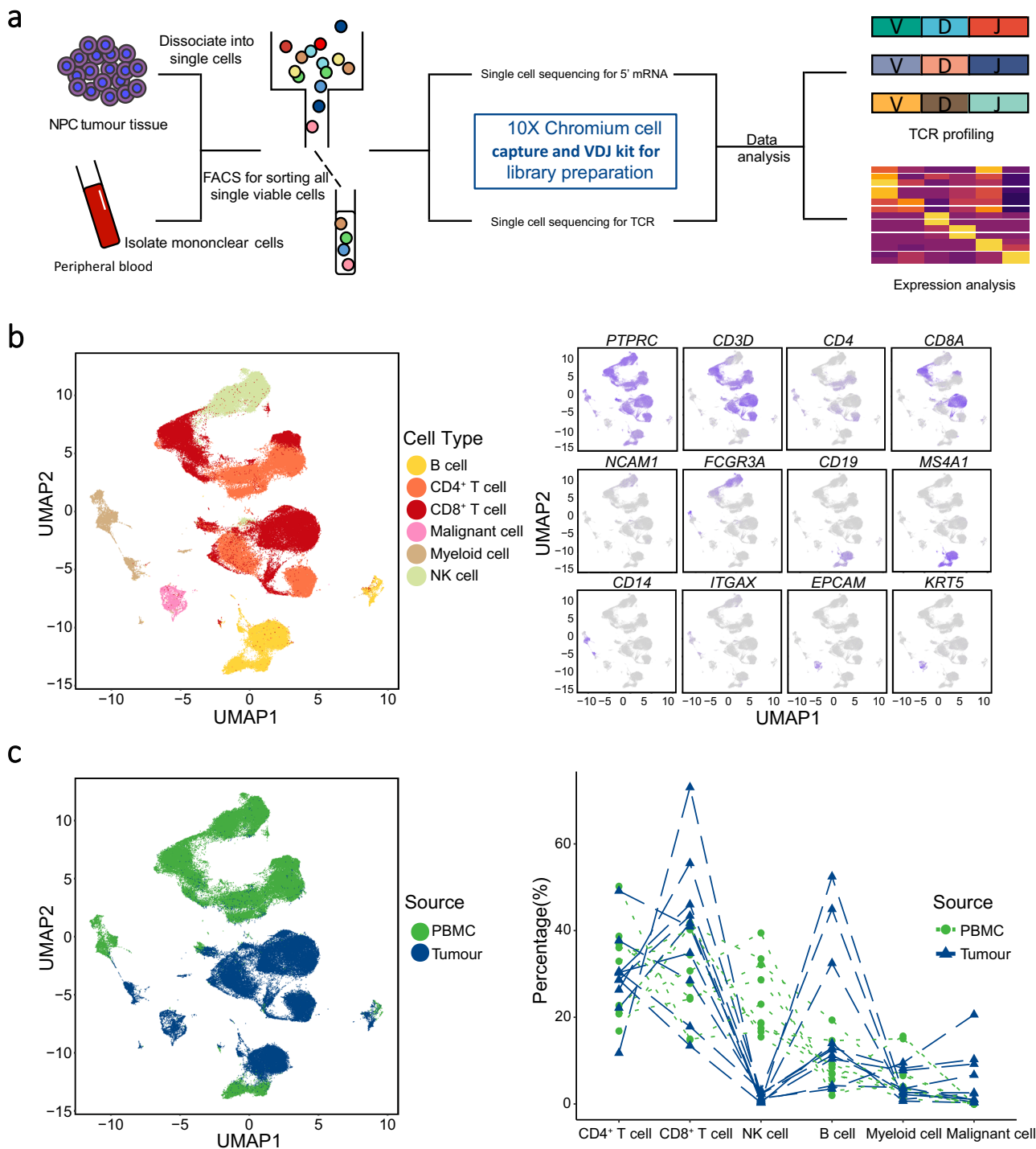
**Landscape view of cell composition in tumour biopsy and PBMC in patients with NPC.** To shed light on the complexity of tumour microenvironment (TME) in NPC, we performed single-cell RNA sequencing in combination of TCR repertoire sequencing on viable cells derived from tumour biopsies and matched peripheral blood mononuclear cells (PBMC) for 10 patients with EBV-positive NPC prior to any anti-cancer treatment (Fig. 1a, Supplementary Fig. 1a, b, and Supplementary Table 1). On

average, we obtained more than 380 million sequencing reads for each sample, with the median of sequencing saturation (covering the fraction of library complexity) at 90.75% (75.90%–94.50%; Supplementary Table 2). After strict quality control filters (low expression of representative genes and inferred doublets; see Methods; Supplementary Fig. 1c and Supplementary Table 3), a total of 176,447 cells were identified from the 10 patients (including 82,622 and 93,825 for tumours and PBMC, respectively; Supplementary Data 1). We obtained about 1500 genes and 4950 unique molecular identifiers (UMIs) on average for each cell, indicating sufficient coverage and representative of transcripts (Supplementary Fig. 1d and Supplementary Data 1).

Next, to define groups of cells with similar expression profiles, we performed unsupervised clustering analysis implemented in Seurat software<sup>19</sup>. The distribution of cell clusters for each patient matched well with that of other patients, suggesting that the potential variation of expression due to batch effect of sample processing was minimal (see Methods; Supplementary Fig. 1e). Each cluster was further identified as a specific cell subpopulation according to the expression of the most variable genes and the canonical markers, including CD4<sup>+</sup> T cells (gene markers: *PTPRC*, *CD3D*, and *CD4*), CD8<sup>+</sup> T cells (*PTPRC*, *CD3D*, and *CD8A*), myeloid cells (*CD14*, *ITGAX* for CD11C), malignant cells (*EPCAM* and *KRT5*), B cells (*CD19* and *MS4A1*), and NK cells (*FCGR3A* and *NCAM1*; Fig. 1b). Besides, we detected 56 fibroblasts and seven endothelial cells with sparse distribution among TME cells (Supplementary Fig. 1f), which might reflect their intrinsic nature in NPC or their low representation due to technical limitations. All these cell types were widespread in tumour samples, indicating the heterogeneous cell composition of TME in NPC (Fig. 1b), consistent with a recent single-cell transcriptome study of NPC<sup>17</sup>. We observed that the proportions of CD8<sup>+</sup> T and B cells were increased with 1.34 and 2.33 times on average, respectively, while the NK cells were decreased in the tumours compared to the PBMC, suggesting two distinct immune landscapes between tumour and peripheral blood (Fig. 1c). Moreover, we compared cell compositions between NPC and other types of cancers with single-cell data publicly available, including non-small-cell lung cancer (NSCLC), colorectal cancer (CRC), and pancreatic ductal adenocarcinoma (PDAC). We observed the common occurrence of infiltrating immune cells amid individual heterogeneity of cell composition in all types of cancers (Supplementary Fig. 1g). Notably, we observed a significantly higher proportion of T cells in NPC compared to any other cancers (Supplementary Fig. 1h), which is consistent with a previous finding that the tumour infiltrating leucocytes was the main characteristic of NPC stroma<sup>20</sup>.

## Heterogeneity of T cells and the diversity of TCR repertoire.

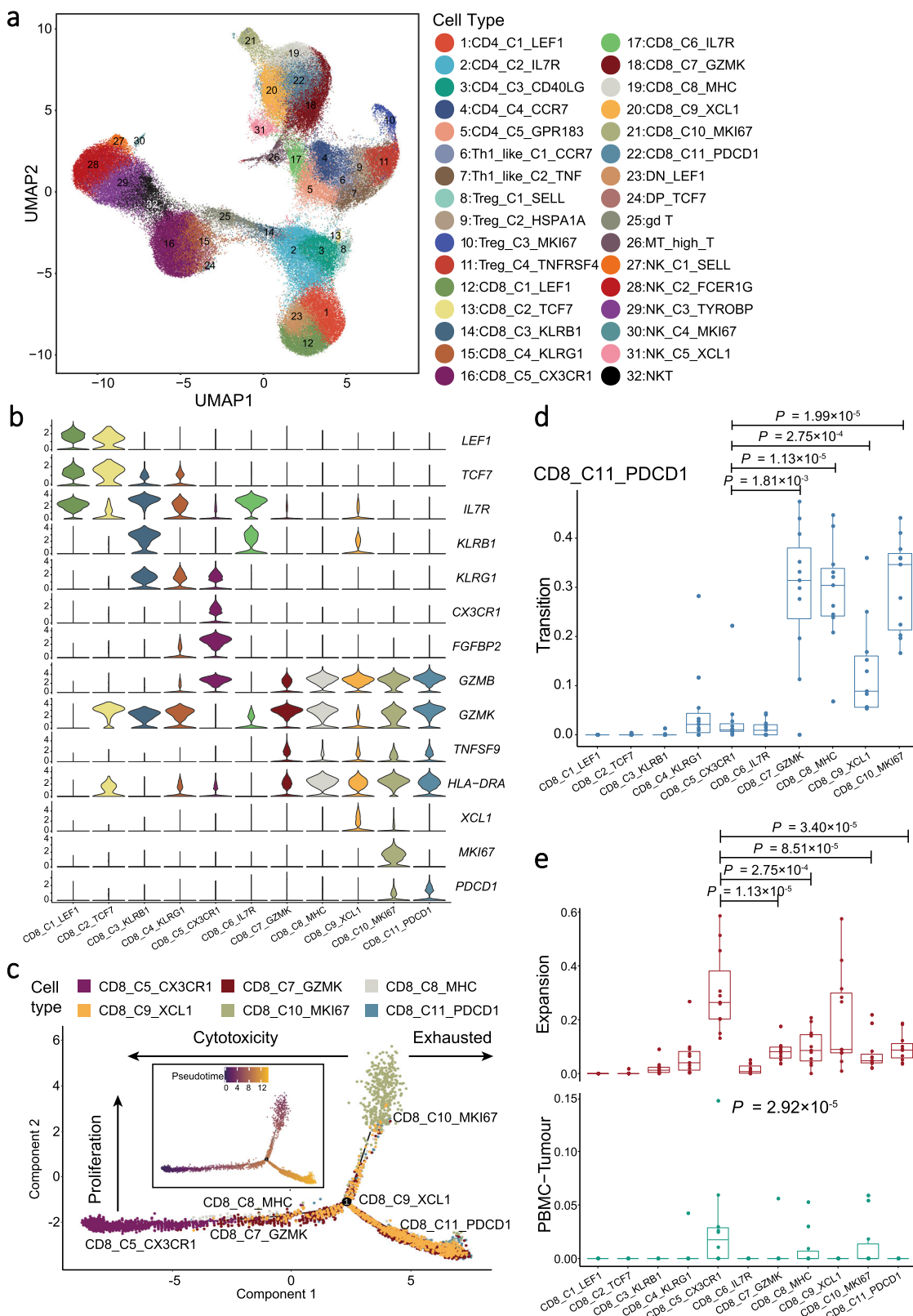
Considering the abundance of T and NK cells in NPC tumour samples and their anti-tumour capabilities, we explored the intrinsic structure and potential functional subtypes of the overall T and NK cell populations. We grouped all 141,875 T and NK cells into 32 subgroups using clustering analysis, of which the majority were CD4<sup>+</sup> and CD8<sup>+</sup> T cells (Fig. 2a, Supplementary Fig. 2a, Supplementary Data 1 and 2). To identify any gene with a specific expression on a cell type, we performed differential gene expression (DEG) analysis of T cell clusters. We observed that CD4<sup>+</sup> and CD8<sup>+</sup> T cell clusters in tumour samples had widespread overexpression of exhaustion markers (*LAG3*, *TIGIT*, *PDCD1*, *HAVCR2*, and *CTLA4*) and effector molecules (*GZMB*, *GZMK*, *INFG*, *NKG7*, *GNLY*, and *IL2*; Supplementary Fig. 2b), with remarkably high expression of the proliferative signatures for *CD8\_C10\_MKI67* and *Treg\_C3\_MKI67* (Supplementary Fig. 2c and Supplementary Data 3). The co-expression of



**Fig. 1** The landscape profiling of single cells in NPC tumours and matching PBMC. **a** An experimental scheme diagram highlights the overall study design. Single viable cells were collected using flow cytometry sorting (FACS) and subjected for cell barcoding. The cDNA libraries of 5'-mRNA expression and TCR were constructed independently, followed by high throughput sequencing and downstream analyses. **b** UMAP plot of 176,447 single cells grouped into six major cell types (left panel) and the normalised expression of marker genes for each cell type (right panel). Each dot represents one single cell, coloured according to cell type (left panel), and the depth of colour from grey to blue represents low to high expression (right panel). **c** UMAP plot of the above single cells coloured according to their origins from peripheral blood or tumour (left panel), and the fraction of cell types originating from each patient (right panel). Each dot represents one single cell, coloured according to sample origin.

exhaustion and effector genes in tumour infiltrating T cells has been also demonstrated recently in NPC<sup>17</sup>. Together, these observations suggest that T cells exhibited anti-tumour effects against antigens, but their effector functions were somehow suppressed in the TME of NPC. By contrast, we observed naïve

gene signatures (high expression of *TCF7*, *SELL*, *CCR7* and *LEF1*) in the resting T cells in PBMC, including CD4<sup>+</sup>C1\_LEF1, CD8<sup>+</sup>C1\_LEF1, CD8<sup>+</sup>C2\_TCF7, and DN\_LEF1 (CD4<sup>-</sup>CD8<sup>-</sup>) cell clusters, and partially in Treg\_C1\_SELL and DP\_TCF7 (CD4<sup>+</sup>CD8<sup>+</sup>; Supplementary Fig. 2b). Besides, we identified two



clusters of CD4<sup>+</sup> Th1-like cells in tumours, including Th1\_like\_C1\_CCR7 and Th1\_like\_C2\_TNF, with specific expression of naïve T cell markers and pro-inflammatory cytokines, respectively, as well as a common expression of Th1-like cell markers<sup>21</sup> (CXCL13, BHLHE40, and CXCR3; Supplementary Fig. 2d).

Next, we performed T cell receptor (TCR) repertoire analysis based on the sequences of  $\alpha$  and  $\beta$  chains of TCR, which revealed

38,720 (32.97%; out of 117,447) T cells with detectable TCR  $\alpha$ - $\beta$  pairs or clonotypes after strict quality control (Supplementary Fig. 2e and Supplementary Data 4). We observed no sharing of any identical TCR clonotype among different patients with NPC, although they had certain preferences of V and J fragments as well as V-J pairs (Supplementary Fig. 3a, b). Interestingly, we observed the sharing of the most variable CDR3 sequences across



**Fig. 2 Expression profile and development of CD8<sup>+</sup> T cells.** **a** UMAP plot of 141,875 T and NK cells grouped into 32 cell types. Each dot represents a cell, coloured according to the cell types indicated at the right legend. **b** Violin plots showed the normalised expression of CD8<sup>+</sup> T cell markers (rows) in each CD8<sup>+</sup> T cell cluster (columns). Cell clusters and the expression level of a gene are indicated at the *x*- and *y*-axis, respectively. **c** Pseudotime trajectory analysis of selected CD8<sup>+</sup> T cells (CD8\_C5, CD8\_C7, CD8\_C8, CD8\_C9, CD8\_C10, and CD8\_C11; *n* = 10,000) with high variable genes. Each dot represents one single cell, coloured according to its cluster label. The inlet plot showed each cell with a pseudotime score from dark blue to yellow, indicating early and terminal states, respectively. For CD8<sup>+</sup> T cell clusters, 10,000 cells were randomly selected for the analysis. **d** Box plots showed the transition-index scores of exhausted CD8<sup>+</sup> T cells (CD8\_C11\_PDCD1) and other CD8<sup>+</sup> T cells (*n* = 10). Comparison was made using a two-sided Wilcoxon test. Cell clusters and transition-index scores are indicated at the *x*- and *y*-axis, respectively. Endpoints depict minimum and maximum values; centre lines denote median values; whiskers denote 1.5× the interquartile range; coloured dots denote each patient. **e** Box plots showed the expansion- (top panel) and PBMC-Tumour migration-index (bottom panel) scores of each CD8<sup>+</sup> T cell cluster (*n* = 10). Each comparison was made using either a two-sided Wilcoxon test (top panel) or Kruskal-Wallis test (bottom panel). Cell clusters are indicated at the *x*-axis, and the *y*-axis shows the expansion- and PBMC-Tumour migration-index scores at the top and bottom panel, respectively. Endpoints depict minimum and maximum values; centre lines denote median values; whiskers denote 1.5× the interquartile range; coloured dots denote each patient.

the patient samples (Supplementary Fig. 3b), among which CAVRGTGTASKLTF and CASSFSGANVLTF have been associated with the recognition of MLANA and EBV antigens in the VDJ database<sup>22</sup>, respectively (Supplementary Table 4). Moreover, we observed that both CD4<sup>+</sup> and CD8<sup>+</sup> T cells have more clonal T cells, which are derived from identical TCR clonotypes and consistent with a previous study<sup>17</sup>, in the tumours compared to the PBMC, suggesting the clonal expansion of certain dominant clones of tumour infiltrating T cells upon continuous stimulations by tumour antigens (Supplementary Fig. 3c).

**Diversity of CD8<sup>+</sup> T cells and the development of exhausted intratumoral CD8<sup>+</sup> T cells.** We identified a total of 62,244 CD8<sup>+</sup> T cells in all NPC samples, which were grouped into 11 clusters based on their expression of canonical markers, including two naïve (CD8\_C1\_LEF1 and CD8\_C2\_TCF7), blood central memory (CD8\_C3\_KLRB1), blood effector memory (CD8\_C4\_KLRG1), high migration (CD8\_C5\_CX3CR1), tumour central memory (CD8\_C6\_IL7R), tumour effector memory (CD8\_C7\_GZMK and CD8\_C8\_MHC), tissue resident memory (CD8\_C9\_XCL1), high proliferating (CD8\_C10\_MKI67), and exhausted (CD8\_C11\_PDCD1) T cells (Fig. 2a, b, Supplementary Data 1 and 2). The majority (>97.68%) of CD8\_C6\_IL7R, CD8\_C7\_GZMK, CD8\_C8\_MHC, CD8\_C9\_XCL, CD8\_C10\_MKI67, and CD8\_C11\_PDCD1 were found in NPC tumours, whereas the majority (>94.85%) of CD8\_C1\_LEF1, CD8\_C2\_TCF7, CD8\_C3\_KLRB1, CD8\_C4\_KLRG1, and CD8\_C5\_CX3CR1 were in the peripheral blood (Supplementary Data 1).

To evaluate the functional status of CD8<sup>+</sup> T cells, we calculated cytotoxicity, proliferation, and exhaustion scores for all CD8<sup>+</sup> T cell clusters. We observed the highest cytotoxicity score for CD8\_C5\_CX3CR1, the highest proliferation score for CD8\_C10\_MKI67, and the highest exhaustion score for CD8\_C11\_PDCD1 (Supplementary Fig. 2c and 4a), suggesting their potential cytotoxic, proliferation, and exhausted functions, respectively. Next, the DEG analysis revealed high expression of chemokine receptors (*CX3CR1*, *CXCR1*, and *CXCR2*), S1P receptors (*S1PR1*, *S1PR4*, and *S1PR5*), and integrins (*ITGB2*, *ITGA4*, *ITGAL*, and *ITGB7*) in CD8\_C5\_CX3CR1, which were responsible for the regulation of CD8<sup>+</sup> T cell migration<sup>23</sup> (Supplementary Fig. 4b). Moreover, signalling pathway enrichment analyses of the genes with differential expression revealed that tumour cytotoxic CD8<sup>+</sup> T cell clusters (CD8\_C7\_GZMK, CD8\_C8\_MHC, and CD8\_C9\_XCL1) were enriched with the pathways related to cytokine production and lymphocyte activation; and CD8\_C5\_CX3CR1 was enriched with the pathways related to leukocyte trans-endothelial migration and leukocyte migration (Supplementary Fig. 4c), which are

consistent with their capability in peripheral circulation and infiltrating to tumour<sup>24</sup>.

To further explore the development of CD8<sup>+</sup> T cells in NPC, we first performed pseudotime trajectory analysis using Monocle2 to order each CD8<sup>+</sup> T cell along trajectories according to their expression and transition profiles. We observed the developmental trajectories from CD8\_C5\_CX3CR1 cells at the initial state or CD8\_C10\_MKI67 cells at the intermediate state to CD8\_C11\_PDCD1 cells at the terminal state (Fig. 2c). Supportively, compared to the exhausted CD8\_C11\_PDCD1, CD8\_C10\_MKI67 had an intermediate exhaustion score (Supplementary Fig. 4a) and lower expression of known exhaustion markers including *PDCD1*, *LAG3*, and *HAVCR2* (Supplementary Fig. 2b). TCR repertoire sequencing revealed 21,099 CD8<sup>+</sup> T cells (out of 62,244) with TCR clonotypes (Supplementary Fig. 4d and Supplementary Data 4). We observed that CD8\_C11\_PDCD1 shared considerable proportions of identical TCRs with other CD8<sup>+</sup> T cell clusters, ranging from 17.68% to 41.67% for infiltrating T cell clusters and 5.31% for peripheral CD8\_C5\_CX3CR1 (Supplementary Fig. 4e). To track the dynamic relationships among T cell clusters from NPC tumour and peripheral blood, we quantitated the expansion (exp, clonal expansion), migration (migr) and transition (tran, developmental transition or differentiation) of T cells using gene expression and TCR information with STARTRAC method<sup>24</sup>. Consistently, we observed the highest transition mobility of CD8\_C11\_PDCD1 with CD8\_C10\_MKI67, followed by CD8\_C7\_GZMK, CD8\_C8\_MHC, and CD8\_C9\_XCL1 (Fig. 2d). These observations strongly suggest that CD8\_C11\_PDCD1 cells were mainly expanded by proliferating pre-exhausted intratumoral CD8<sup>+</sup> T cells. Moreover, we observed that CD8\_C5\_CX3CR1 had the largest number of clonal T cells (Supplementary Fig. 4d) and the highest expansion mobility in CD8<sup>+</sup> T cell clusters (Fig. 2e). Furthermore, CD8\_C5\_CX3CR1 had the highest proportion of shared TCR between peripheral blood and tumour (Fig. 2e). The proportions of TCRs shared with peripheral CD8\_C5\_CX3CR1 ranged from 4.76% to 12.77% for infiltrating CD8<sup>+</sup> T cell clusters (Supplementary Fig. 4e). These data suggest a common origin of the intratumoral CD8<sup>+</sup> T cells in NPC tumour from CD8<sup>+</sup> T cells in peripheral blood including CD8\_C5\_CX3CR1 cells.

**The diversity and trajectory of Treg cells in NPC.** Treg cells are potent suppressors of immune cells and are essential to maintaining immunological tolerance and homeostasis. We identified a total of 11,631 Treg cells based on their transcription of canonical markers (*CD4*, *IL2RA*, and *FOXP3*), which were grouped into four cell clusters, including Treg\_C1\_SELL, Treg\_C2\_HSPA1A, Treg\_C3\_MKI67, and Treg\_C4\_TNFRSF4 (Fig. 2a, Supplementary Fig. 5a, Supplementary Data 1 and 2). We observed that the proportion of Treg cells among CD4<sup>+</sup>

T cells in the tumours was much higher than that in the PBMC (Supplementary Fig. 5b). All Treg\_C4\_TNFRSF4 cells and the majority of Treg\_C2\_HSPA1A (99.5%; 4,762 out of 4,786) and Treg\_C3\_MKI67 (90.0%; 1,187 out of 1,319) cells were found in the tumours, while Treg\_C1\_SELL cells were in the PBMC (Supplementary Data 1). To explore the immune-regulatory functions of Treg cells, we first calculated the IL2R scores for each cell based on their expression level of *CD25* (IL2RA), *CD122* (IL2RB), and *CD132* (IL2RG) using the *AddModuleScore* function implemented in Seurat software. The three genes encode transmembrane proteins that form a receptor complex competitively binding IL2 (the T cell growth factor) with high affinity so as to inhibit effector T cells<sup>25</sup>. We observed the highest IL2R score for Treg\_C4\_TNFRSF4 among all Treg clusters (Fig. 3a), suggesting the strongest IL-2 binding potential of Treg\_C4\_TNFRSF4 cells. Similarly, we also observed the highest inhibitory and co-stimulatory scores for Treg\_C4\_TNFRSF4 cells based on their expression levels of genes with immune-inhibitory functions and co-stimulatory receptors, respectively (Fig. 3a and Supplementary Fig. 5a), suggesting that Treg\_C4\_TNFRSF4 cells had a stronger suppression potential on immune response and were much activated than the other Treg cells. Consistently, such a subset of Treg cells were also identified in CRC, NSCLC, and hepatocellular carcinoma (HCC), with high activation and immune-suppressive potential as indicated by the high IL2R, inhibitory, and co-stimulatory scores (Supplementary Fig. 5c). Besides, we observed elevated expression levels of chemokine receptors in Treg\_C4\_TNFRSF4 cells, including *CXCR3*, *CXCR6*, and *CCR8* that have been implicated in several cancers<sup>26</sup> (Supplementary Fig. 5a).

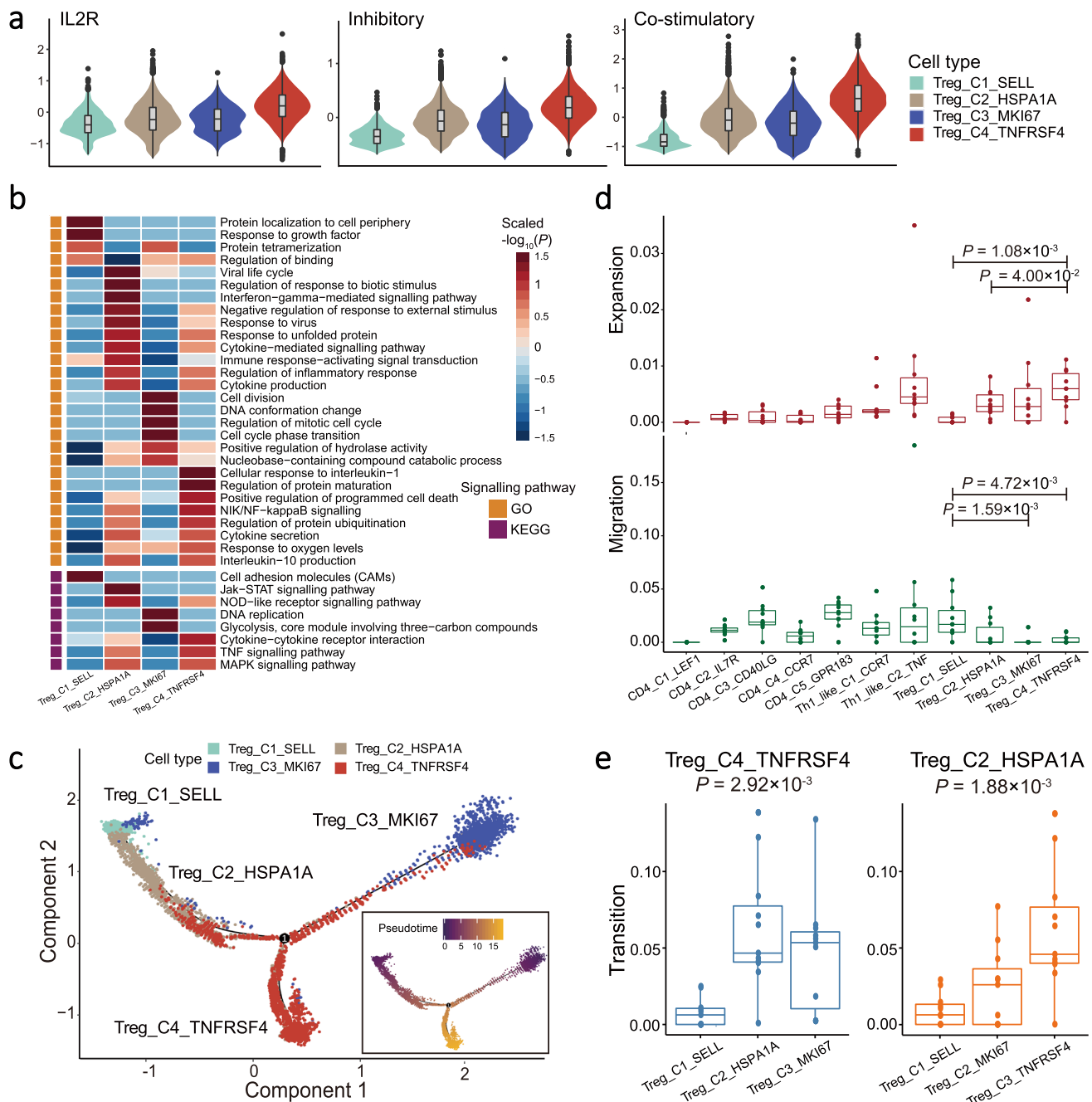
To characterise the potential functions of Treg cells, we first performed signalling pathway enrichment analyses for each Treg cluster based on the expression levels of genes implicated in each pathway. We observed a distinct pattern of pathway enrichment for each Treg cluster, suggesting their various functions. Particularly, the ‘cytokine-cytokine receptor interaction’ was highly enriched in Treg\_C4\_TNFRSF4 (Fig. 3b), consistent with their chemotactic potentials as mentioned earlier. Moreover, the ‘interleukin-10 production’, ‘TNF signalling pathway’ and ‘NF- $\kappa$ B signalling pathway’ were enriched in both Treg\_C4\_TNFRSF4 and Treg\_C2\_HSPA1A (Fig. 3b). As such, we further compared the major pathways between Treg\_C4\_TNFRSF4 and Treg\_C2\_HSPA1A using Gene Set Enrichment Analysis (GSEA), which revealed a higher enrichment of pathways related to cell cycle, chemokine, TGF- $\beta$ , and negative regulation of T cell proliferation, as well as transcription factors activating NF- $\kappa$ B and STAT pathways in Treg\_C4\_TNFRSF4 (Supplementary Fig. 5d). Since Treg\_C4\_TNFRSF4 expressed *CCR8* specifically (Supplementary Fig. 5a), we used the normalised mRNA ratio of *CCR8/FOXP3* to estimate the fraction of Treg\_C4\_TNFRSF4 cells in Treg cells (*FOXP3*<sup>+</sup>). Survival analysis showed that the higher ratio of *CCR8/FOXP3* was associated with the decreased progression-free survival (PFS; Supplementary Fig. 5e), suggesting that a higher fraction of Treg\_C4\_TNFRSF4 cells with activated potential in Treg cells had a strong immune-suppressive function in NPC.

To trace the origin of intratumoral Treg cells, we first performed pseudotime trajectory analysis using Monocle2, which revealed the most terminal status with the highest pseudotime scores for Treg\_C4\_TNFRSF4 cells and two developmental trajectories of Treg\_C4\_TNFRSF4 cells from Treg\_C1\_SELL cells in PBMC and Treg\_C3\_MKI67 cells in tumours (Fig. 3c). We next examined the expression of tissue resident markers (*CD69*, *ITGAE*, and *BHLHE40*; Supplementary Table 5) in Treg cells and observed much higher expression of *ITGAE* and *BHLHE40* in Treg\_C4\_TNFRSF4 and Treg\_C3\_MKI67 cells than

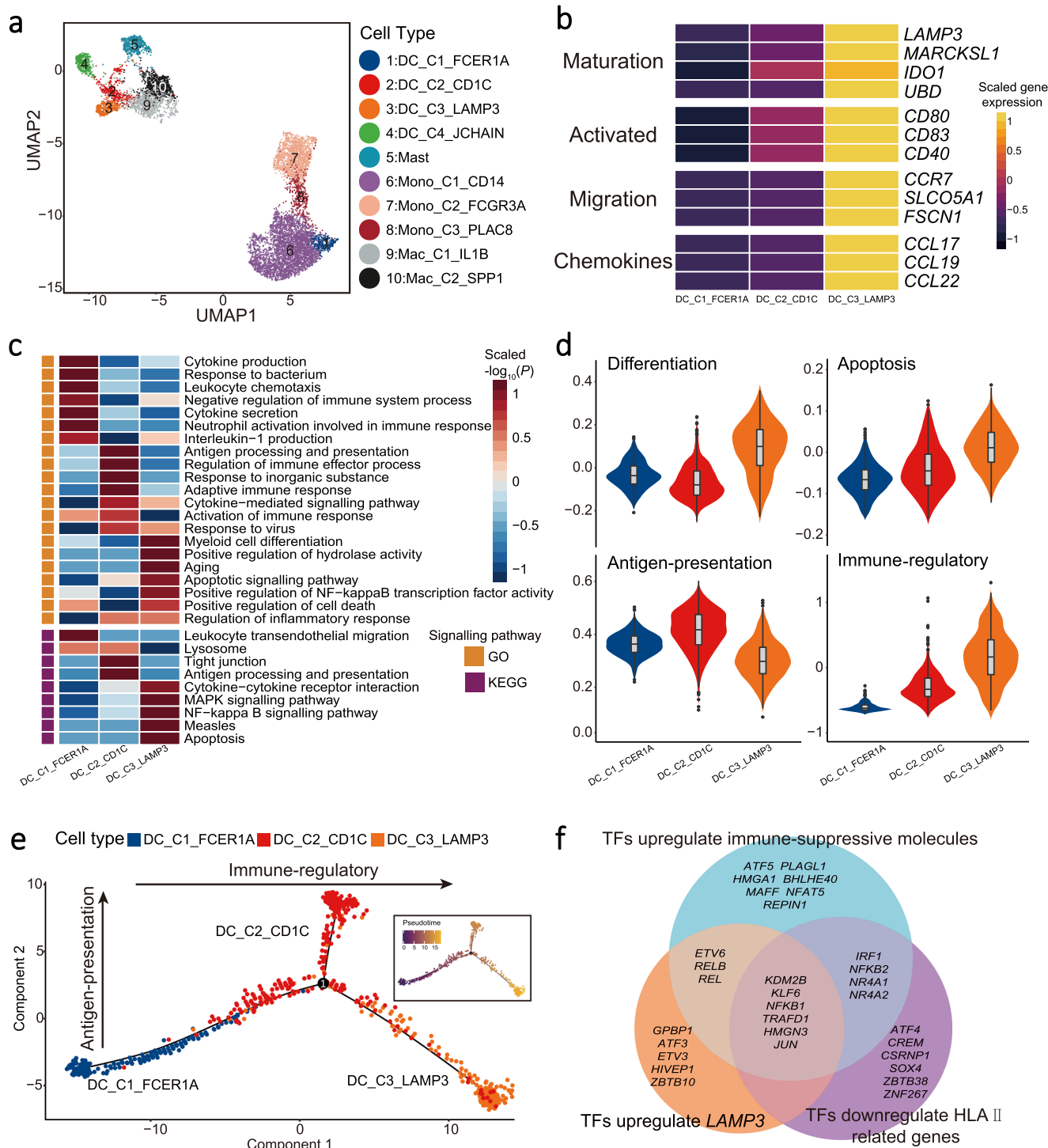
Treg\_C2\_HSPA1A and Treg\_C1\_SELL cells (Supplementary Fig. 6a). Given that the majority of Treg\_C2\_HSPA1A cells was in tumours and originated from Treg\_C1\_SELL cells according to the pseudotime trajectory analysis (Fig. 3c), the scarce expression of the resident markers might suggest the most recent recruitment of Treg\_C2\_HSPA1A cells from peripheral blood.

TCR repertoire analysis revealed 17,621 (out of 47,384) CD4<sup>+</sup> T cells assigned with clonotypes, among which Treg\_C2\_HSPA1A and Treg\_C4\_TNFRSF4 had intermediate numbers of clonotypes (Supplementary Fig. 6b and Supplementary Data 4). Noteworthy, Treg\_C4\_TNFRSF4 had the largest proportion of clonal cells, meaning the highest clonality, among all CD4<sup>+</sup> T cells (Supplementary Fig. 6b). Consistently, we observed that Treg\_C4\_TNFRSF4 had the highest expansion score, meaning the highest degree of clonal expansion, among the Treg cell clusters (Fig. 3d). We also observed the highest migration score, meaning the highest mobility, for Treg\_C1\_SELL derived from the PBMC (Fig. 3d). DEG analysis revealed that Treg\_C1\_SELL had high expression of chemokine receptors *CCR4*, which are chemotactic counterparts for *CCL5*, *CCL17*, and *CCL22* produced by intratumoral CD8<sup>+</sup> T, NK, and myeloid cells in NPC (Supplementary Figs. 5a and 6c). These observations suggest that the migration capability and chemotactic interaction potentials with intratumoral cells make the movement of peripheral Treg\_C1\_SELL cells to tumour site possible. Indeed, we observed a small number of shared TCRs between Treg cells from tumour and peripheral blood (Supplementary Fig. 6d), which is consistent with a previous finding that intratumoral Treg cells were partially recruited from peripheral blood<sup>27</sup>. We further examined the transition mobility of Treg\_C2\_HSPA1A and Treg\_C4\_TNFRSF4 with other Treg cells. We observed that Treg\_C4\_TNFRSF4 cells had the highest transition mobility with Treg\_C3\_MKI67 cells, followed by Treg\_C2\_HSPA1A and Treg\_C1\_SELL cells; and Treg\_C2\_HSPA1A cells had high transition mobility with Treg\_C4\_TNFRSF4 and Treg\_C3\_MKI67 cells (Fig. 3e). These observations again supported the developmental trajectory of intratumoral Treg\_C4\_TNFRSF4 cells from naïve Treg\_C1\_SELL cells through intermediate Treg\_C2\_HSPA1A or Treg\_C3\_MKI67 cells (Fig. 3c).

**Diversity of B cells in NPC.** We identified a total of 22,892 B cells, which were grouped into nine clusters (Supplementary Fig. 7a, Supplementary Data 1 and 2). Among them, B\_C1\_TCL1A, B\_C2\_FCRL3, and Plasma\_C1\_IgA clusters were derived from PBMC and the other six clusters were from tumour samples (Supplementary Fig. 7a). DEG analysis revealed unique gene signatures for B cell clusters in tumour samples, including B\_C5\_ISG15 with interferon induced genes, B\_C6\_HSPA1A with stressful gene expression, and Plasma\_C2\_IgG with elevated expression level of IgH genes (Supplementary Fig. 7b). We further identified two B cell clusters (B\_C1\_TCL1A and B\_C4\_IFITM3) before the class switch recombination based on the expression of *IGHM* and *IGHD* (Supplementary Fig. 7b and Supplementary Table 5). Moreover, correlation analysis revealed that the expression of *TCL1A* was highly correlated with that of *IGHM* and *IGHD* in the two clusters, which could be a sufficient marker to classify B cells before the class switch recombination (Supplementary Fig. 7c). Signalling pathway enrichment analyses of the genes with differential expression revealed that B cell clusters were enriched with various pathways related to immune regulation (Supplementary Fig. 7d). Particularly, B\_C4\_IFITM3, B\_C5\_ISG15, and B\_C6\_HSPA1A had the enrichment of ‘EBV infection’, ‘defence response to virus’, ‘viral carcinogenesis’, and ‘response to interferon-gamma’ pathways, suggesting that the



**Fig. 3 Expression profile and development of Treg cells.** **a** Violin plots showed the IL2R (left panel), inhibitory (middle panel), and co-stimulatory (right panel) scores for each Treg cell cluster ( $n = 11,631$ ). Box plots inside the violins indicated the quartiles of corresponding score levels. Endpoints depict minimum and maximum values; centre lines denote median values; whiskers denote  $1.5 \times$  the interquartile range; black dots denote each cell. Violin plots are coloured according to cell types, and signature scores are indicated at the y-axis. **b** Heatmap showed the selected signalling pathways (rows) that were significantly enriched in GO and KEGG analyses for each Treg cell cluster (columns). Filled colours from blue to red represent scaled expression levels (normalised  $-\log_{10}P$  values) from low to high.  $P$  values were calculated by one-sided hypergeometric test and adjusted for multiple comparisons. Orange and purple squares on the left column represent the results derived from GO and KEGG signalling pathways analysis, respectively. **c** Pseudotime trajectory analysis of Treg cells (Treg\_C1, Treg\_C2, Treg\_C3, and Treg\_C4;  $n = 11,631$ ) with high variable genes. Each dot represents one single cell, coloured according to its cluster label. The inset plot showed each cell with a pseudotime score from dark blue to yellow, indicating early and terminal states, respectively. **d** Box plots showed the expansion-index (top panel) and migration-index (bottom panel) scores of each  $CD4^+$  T cell cluster ( $n = 10$ ). Comparison was made using two-sided Wilcoxon test. Cell clusters are indicated at the x-axis, and the y-axis shows the expansion- or migration-index at the top or bottom panel, respectively. Endpoints depict minimum and maximum values; centre lines denote median values; whiskers denote  $1.5 \times$  the interquartile range; coloured dots denote each patient. **e** Box plots showed the transition-index scores of Treg\_C4\_TNFRSF4 (left panel) and Treg\_C2\_HSPA1A (right panel) with other Treg cells ( $n = 10$ ). Comparison was made using two-sided Kruskal-Wallis test. Cell clusters and the transition-index scores are indicated at the x- and y-axis, respectively. Endpoints depict minimum and maximum values; centre lines denote median values; whiskers denote  $1.5 \times$  the interquartile range; coloured dots denote each patient.



three groups of cells might be responsible for the immune response against EBV infection (Supplementary Fig. 7d).

**Tumour-associated LAMP3<sup>+</sup> DCs display a tolerogenic phenotype in NPC.** A total of 8,893 myeloid cells were identified and clustered into 10 subsets including one for mast cells, five for monocyte or macrophage cells, three for conventional dendritic cells, and one for plasmacytoid dendritic cells (Fig. 4a, Supplementary Fig. 8a and 8b, Supplementary Data 1 and 2). Among the four clusters of dendritic cells, DC\_C2\_CD1C, DC\_C3\_LAMP3, and DC\_C4\_JCHAIN were derived from tumours, and DC\_C1\_FCER1A was derived from peripheral blood, which was

assigned as monocyte-like DC because of the expression of monocyte marker *S100A8* (Supplementary Fig. 8a, b). Noteworthy, we identified DC\_C3\_LAMP3 cells as a group of DCs with high maturation, activation, and migration potentials in NPC, based on the expression levels of the signature genes related to maturation (*LAMP3*, *MARCKSL1*, *IDO1*, and *UBD*), activation (*CD80*, *CD83*, and *CD40*), and migration (*CCR7*, *FSCN1*, and *SLCO5A1*; Fig. 4b, Supplementary Fig. 8c and Supplementary Table 5), respectively. Moreover, DC\_C3\_LAMP3 cells had high expression of special chemokine ligands (*CCL17*, *CCL19*, and *CCL22*), which are known to recruit immune cells expressing chemokine receptors *CCR4*, *CCR7*, and *CXCR3* (Fig. 4b). We also



**Fig. 4 Expression and development of dendritic cells.** **a** UMAP plot of 8,893 myeloid cells grouped into 10 cell types. Each dot represents a cell, coloured according to cell types. **b** Heatmap showed the normalised mean expression of genes associated with maturation, activation, migration, and chemokine ligand (rows) in three dendritic cell clusters (DC\_C1, DC\_C2, and DC\_C3; columns). Filled colours from black to yellow represent scaled gene expression levels from low to high. **c** Heatmap showed the selected signalling pathways (rows) with significant enrichment of GO and KEGG terms for three dendritic cell clusters (DC\_C1, DC\_C2, and DC\_C3; columns). Filled colours from blue to red represent scaled expression levels (normalised  $-\log_{10}P$  values) from low to high.  $P$ -values were calculated by one-sided hypergeometric test and adjusted for multiple comparisons. Orange and purple squares on the left column represent the results derived from GO and KEGG signalling pathways analysis, respectively. **d** Violin plots showed the differentiation, apoptosis, antigen presentation, and dysfunction scores of three dendritic cell cluster (DC\_C1, DC\_C2, and DC\_C3;  $n = 1134$ ). Box plots inside the violins indicated the quartiles of corresponding score levels. Endpoints depict minimum and maximum values; centre lines denote median values; whiskers denote  $1.5 \times$  the interquartile range; black dots denote each cell. Cell clusters and the signature scores are indicated at the x- and y-axis, respectively. **e** Pseudotime trajectory analysis of three dendritic cell clusters (DC\_C1, DC\_C2, and DC\_C3;  $n = 1134$ ) with high variable genes. Each dot represents one single cell, coloured according to its cluster label. The inlet plot showed each cell with a pseudotime score from dark blue to yellow, indicating early and terminal states, respectively. **f** Venn diagram showed overlapped transcription factors regulating *LAMP3* gene, immune-suppressive molecules, and HLA-II in DC\_C3\_LAMP3 cells.

observed significant correlations of expression between the marker gene *LAMP3* and other functional genes related to maturation, migration, activation, and chemokine ligands in DC\_C3\_LAMP3 (Supplementary Fig. 8d). These observations suggest that DC\_C3\_LAMP3 cells might be *LAMP3*<sup>+</sup> DCs, which are featured with high migration, activation, and maturation in several cancers as reported previously<sup>28,29</sup>.

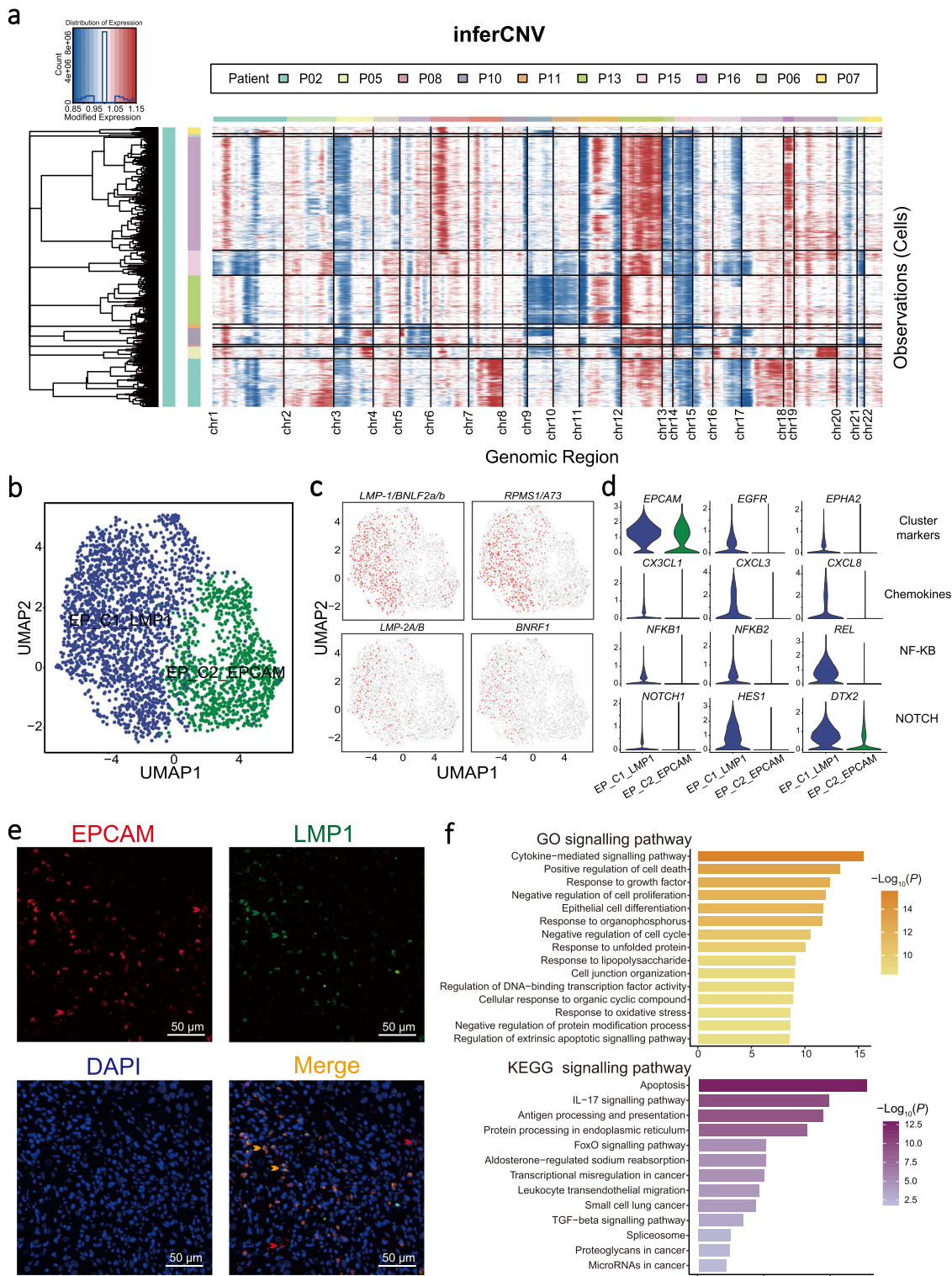
Signalling pathway enrichment analyses using GO and KEGG revealed a specific pattern of enriched pathways among the three conventional DC cell clusters, where the ‘antigen processing and presentation’ was significantly upregulated in DC\_C2\_CD1C but downregulated in DC\_C3\_LAMP3 (Fig. 4c). Moreover, apoptosis, NF- $\kappa$ B, and MAPK signalling pathways as well as myeloid cell differentiation were also upregulated in DC\_C3\_LAMP3 compared to other two clusters (Fig. 4c). These observations are consistent with the GSEA analyses (Supplementary Fig. 8e). As such, we scored the expression levels of genes related to these pathways in each cluster (Supplementary Data 3), which revealed the highest levels of differentiation and apoptosis but the lowest antigen presentation for DC\_C3\_LAMP3 (Fig. 4d). Furthermore, we observed that the gene signatures corresponding to the activation of immune response was reduced in DC\_C3\_LAMP3 (Fig. 4c), which was consistent with the highest immune-regulatory score and the increased expression of a subset of immune-suppressive genes, including *CD274* (PD-L1), *PDCD1LG2* (PD-L2), *CD200*, *EBI3*, *IDO1*, *IL4I1*, *SOCS1*, *SOCS2*, and *SOCS3* (Supplementary Fig. 9a and Supplementary Data 3). Besides, we observed a similar expression profile of *LAMP3*<sup>+</sup> DC among NPC, HCC, and NSCLC (Supplementary Fig. 9a). These observations suggest that DC\_C3\_LAMP3 cells could be considered as a group of regulatory and tolerogenic DCs, which restrain the activation of T cells<sup>30</sup>.

Next, we performed pseudotime trajectory analysis and observed that DC\_C1\_FCER1A cells developed into two branches including DC\_C2\_CD1C and DC\_C3\_LAMP3 cells, and DC\_C3\_LAMP3 cells had the highest pseudotime score meaning the most differentiated and matured DC (Fig. 4e). Combined with their immune-regulatory and antigen-presenting scores (Fig. 4d), these data suggest that DC\_C1\_FCER1A cells in peripheral blood might infiltrate to tumour, convert to DC\_C2\_CD1C cells with increased antigen-presenting capacity and to immune-suppressive DC\_C3\_LAMP3 cells (Fig. 4e). Consistently, we observed the similar pattern of changes in the expression of transcription factors (TFs) specific for the genes with differential expression from DC\_C1\_FCER1A to DC\_C2\_CD1C and then DC\_C3\_LAMP3 cells (Supplementary Fig. 9b). We further constructed an accurate cellular network to infer the regulons associated with transcription factors and signalling molecules in DC\_C3\_LAMP3 using ARACNe (Supplementary Data 5). We

observed that the upregulation of *LAMP3* was linked with multiple TFs, including *ETV3*, *ETV6*, *HMG3*, *GPBP1*, *TRAFD1*, *ATF3*, *KDM2B*, *JUN*, *HIVEP1*, *KLF6*, *ZBTB10*, and *NFKB1* that have been related to the maturation of DC in mouse<sup>31</sup>, while the downregulation of *LAMP3* was linked with *CREM* (Supplementary Fig. 9c). We also observed that among these TFs *KDM2B*, *KLF6*, *ETV6*, *JUN*, *HMG3*, and *TRAFD1*, as well as *NFKB1*, *REL*, and *RELB* in the NF- $\kappa$ B pathway were linked with the upregulated expression of immune-suppressive molecules like *CD274*, *PDCD1LG2*, *CD200*, and *IDO1*, but the downregulated expression of HLA-class II genes (Fig. 4f and Supplementary Fig. 9c). By contrast, *SOX4* and *CREM* were associated with the downregulation of *CD274*, *CD200*, and *IDO1* (Supplementary Fig. 9c). These observations suggest that multiple TFs regulate the immune-suppressive function, antigen-presenting capacity, and maturation of DC\_C3\_LAMP3 in NPC.

#### Heterogeneity of malignant cells with different EBV infection status.

We identified a total of 2,787 malignant epithelial cells in NPC tumours based on their presence of large-scale chromosomal copy number variation (CNV) compared to a reference data of normal epithelial cells<sup>32</sup> (see Methods; Fig. 5a). Given that EBV is a known factor responsible for the malignant transformation and tumorigenesis of NPC<sup>33</sup>, we examined the expression of EBV molecules in the malignant cells and divided them into EBV<sup>+</sup> (EP\_C1\_LMP1) and EBV<sup>-</sup> (EP\_C2\_EPCAM) malignant cells according to their detectable or not EBV transcripts (*LMP-1/BNLF2a/b*, *RPMS1/A73*, *LMP-2A/B*, and *BNRF1*; Fig. 5b, c, Supplementary Fig. 10a, Supplementary Data 1 and 2). We observed higher expression of *EPHA2* and *EGFR* in EP\_C1\_LMP1 cells (Fig. 5d), which have been related to the susceptibility of EBV infection<sup>34</sup>. Moreover, immunofluorescence staining of an EBV-encoded protein (LMP1) confirmed the presence of EBV<sup>+</sup> malignant cells (LMP1<sup>+</sup>EPCAM<sup>+</sup>) and EBV<sup>-</sup> malignant cells (LMP1<sup>-</sup>EPCAM<sup>+</sup>) in NPC (Fig. 5e). We also observed specifically high activations of the major genes involved in NF- $\kappa$ B and Notch pathways as well as chemokines including *CX3CL1* in EP\_C1\_LMP1 compared to EP\_C2\_EPCAM cells (Fig. 5d). Consistently, we observed high expression of *CX3CL1* in an independent collection of NPC tumours ( $n = 113$ ) compared to non-cancerous samples (rhinitis,  $n = 10$ ; Supplementary Fig. 10b). Interestingly, we noted that the overexpression of *CX3CR1*, the receptor of *CX3CL1*, in multiple types of immune cells in peripheral blood (Supplementary Fig. 10c). Signalling pathway enrichment analyses revealed that EP\_C1\_LMP1 were enriched with cytokine-mediated, regulation of cell death, apoptosis, and cancer-related pathways (Fig. 5f). Taken together, these observations suggest that malignant NPC cells exhibit different



susceptibility to EBV infection, leading to distinct expression profiles.

To further explore the intra-tumour and inter-tumour heterogeneity of malignant NPC cells, we first divided them into five prominent cell subgroups (C1–C5; Supplementary Fig. 10d, e), using clustering analysis without EBV information. Next, we deciphered the variations of gene expression in malignant cells for

different clusters using gene set variation analysis (GSVA), which revealed a distinct enrichment of signalling pathway for each cluster (Supplementary Fig. 10f). We observed variable proportions of cell subtypes among different tumour samples (Supplementary Fig. 10e), which might contribute to the inter-tumour heterogeneous expression profiles. Indeed, the GSVA data showed different enrichments of signalling pathways among

**Fig. 5 Heterogeneity of malignant cells with distinct EBV infection in tumour tissues.** **a** Heatmap showed the large-scale CNVs for epithelial cells (rows along y-axis) from 10 NPC tumours. CNVs were inferred according to the average expression of 100 genes spanning each chromosomal position (x-axis). Red: gains; blue: losses. Malignant NPC cells from different patients and the range of different chromosomes are indicated as different colour bars on the left and top to the heatmap, respectively. **b** UMAP plot of 2,787 malignant cells grouped into two cell clusters (EP\_C1\_LAMP1 and EP\_C2\_EPCAM). Each dot represents a cell, coloured according to a cell cluster. **c** UMAP plots showed the expression of EBV-encoded genes (*LMP-1/BNLF2a/b*, *RPMS1/A73*, *LMP-2A/B*, and *BNRF1*) in malignant cells. Each dot represents a single cell, and the depth of colour from grey to red represents low to high expression. **d** Violin plots showed the normalised expression of cluster markers, chemokines, and genes associated with NF- $\kappa$ B and Notch pathways in each cluster. In each plot, cell clusters and the expression level of a gene as the chart tile are indicated at the x- and y-axis, respectively. **e** Representative images of multiplex immunofluorescence staining of malignant cells in NPC tissues. Proteins detected using respective antibodies in the assays are indicated on top. The red, green, and orange arrows indicated the representative cells positive for EPCAM, LMP1, and co-expression of EPCAM and LMP1 proteins in malignant cells, respectively. Images are representative of three biological replicates. Scale bars, 50  $\mu$ m. **f** Bar plots showed the selected signalling pathways with significant enrichment of GO (top panel) and KEGG (bottom panel) terms for EBV<sup>+</sup> malignant cells (EP\_C1\_LAMP1) compared to EBV<sup>-</sup> malignant cells (EP\_C2\_EPCAM), coloured from light to dark according to their  $-\log_{10}(P\text{-values})$  from low to high. *P*-values were calculated by one-sided hypergeometric test and adjusted for multiple comparisons.

tumour samples (Supplementary Fig. 10f). Noteworthy, the P02, P08, P11, and P15 samples with a higher proportion of C4 cluster compared to other samples showed enrichment of cell cycle (E2F, MYC, and G2M checkpoint) related pathways (Supplementary Fig. 10e–g), although C4 cluster had the highest proliferation scores but overall low content in NPC tumours compared to the other three clusters. These observations suggest the intra- and inter-tumour heterogeneity of the malignant cells in NPC.

**Intercellular interaction network in NPC.** To explore the cellular communication network in NPC, we examined potential ligand-receptor binding among different cell clusters derived from NPC tumours and PBMC, using CellPhoneDB software<sup>35</sup> (Supplementary Data 6). We observed intensive cellular interactions among the DC\_C3\_LAMP3 cells, Treg cells, and exhausted CD8<sup>+</sup> T cells (CD8\_C11\_PDCD1) via inhibitory, co-stimulatory molecules, or chemokines (Fig. 6a, b). Among them, DC\_C3\_LAMP3 cells were predicted to interact with Treg\_C1\_SELL cells in peripheral blood through CCL17-CCR4 and CCL22-CCR4, which are known for recruiting Treg cells into tumour tissue<sup>36</sup> (Fig. 6a). Treg\_C4\_TNFRSF4 cells had high expression of CTLA4, ENTPD1, and CSF1, which showed ligand-receptor bindings to CD80/CD86, ADORA2A, and SIRPA on DC\_C3\_LAMP3 cells, suggesting the potential interaction between Treg\_C4\_TNFRSF4 and DC\_C3\_LAMP3 cells (Fig. 6a). DC\_C3\_LAMP3 cells were also predicted to interact with CD8\_C11\_PDCD1 cells through CD200-CD200R signalling, a non-classical immune-suppressive pathway involved in the suppression of anti-tumour responses<sup>37</sup> (Fig. 6b). Potential ligand-receptor interactions were observed between Treg\_C4\_TNFRSF4 and CD8\_C11\_PDCD1 cells, including those of chemokines (CCL4-CCR8), adhesive connection (ITGAL-ICAM1 and SELPLG-SELL), and immune regulation (HAVCR2-LGALS9; Fig. 6b and Supplementary Data 6), which are well-known in the TME of tumour and promote the immune-suppressive activity of Treg cells and CD8<sup>+</sup> T cells exhaustion<sup>26,38</sup>. Notably, these potential interactions were commonly observed in our NPC cohort (Supplementary Fig. 11a). Consistently, in another independent NPC sample collection ( $n = 113$ ), we observed strong correlations of expression among the gene signatures for DC\_C3\_LAMP3 cells, Treg cells, and exhausted CD8<sup>+</sup> T cells (CD8\_C11\_PDCD1;  $r > 0.8$ ,  $P < 2.2 \times 10^{-16}$ ; Supplementary Fig. 11b). These observations suggest the widespread occurrence of the immune-regulatory interactions among DC\_C3\_LAMP3, Treg\_C4\_TNFRSF4, and CD8\_C11\_PDCD1 cells in NPC tumours. We further performed multiplex immunohistochemistry (IHC) staining of NPC biopsies and confirmed the physical juxtapositions of CD80-expressing DC\_C3\_LAMP3 cells (CD80<sup>+</sup>) and CTLA4-expressing Treg cells (CD3<sup>+</sup>CD4<sup>+</sup>FOXP3<sup>+</sup>), as well as PD-L1-expressing DC\_C3\_LAMP3 cells (CD80<sup>+</sup>) and PD-1-expressing CD8<sup>+</sup> T cells (CD3<sup>+</sup>CD8<sup>+</sup>; Fig. 6c, d).

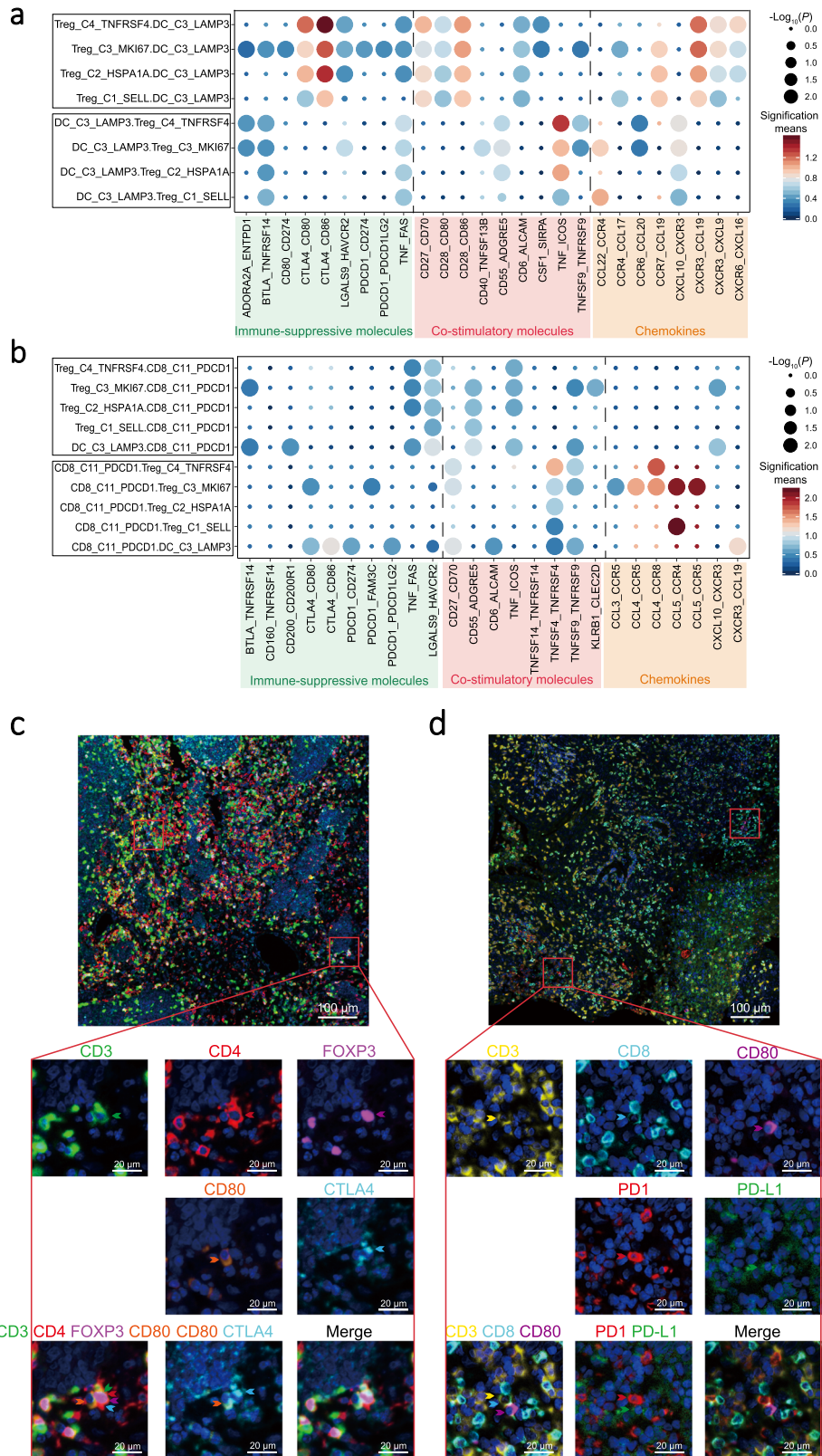
Between malignant NPC cells and immune cells, we observed that EBV<sup>+</sup> EP\_C1\_LAMP1 cells had significantly more receptor-ligand interactions than EBV<sup>-</sup> EP\_C2\_EPCAM cells in each NPC patient (Supplementary Fig. 12a). We noted that EP\_C1\_LAMP1-cells uniquely expressed CX3CL1 in tumour, which was predicted to interact with CX3CR1 expressed on peripheral immune cells including CD8\_C5\_CX3CR1 cells, DC\_C1\_FCER1A cells, NK cells, and monocytes (Supplementary Fig. 12b), suggesting the chemotactic potential of EP\_C1\_LAMP1 cells to immune cells from peripheral blood. Moreover, we observed that EGFR on EP\_C1\_LAMP1 cells was predicated to bind TGFB1 on multiple cell types, which has been reported to regulate the EBV life cycle<sup>39</sup> (Supplementary Fig. 12c). We also observed potential interacting pairs between EP\_C1\_LAMP1 cells with activated Notch pathway and multiple cell types through NOTCH1-TNF and NOTCH2-JAG2, which have been related to radiation sensitivity<sup>40</sup> and cancer stem-like side population cells<sup>41</sup> in NPC (Supplementary Fig. 12d). In addition, the above-mentioned interactions were commonly determined but with individual variable intensity among the patients (Supplementary Data 6), suggesting that the interactions are widespread phenomena and heterogeneous in NPC (Supplementary Fig. 12b–d).

## Discussion

Through the comprehensive single-cell transcriptome study on NPC, we provided a landscape view of the heterogeneous cell composition and complex interacting network in the tumour microenvironment and peripheral circulating blood of NPC at single-cell resolution. Transcriptome analyses of more than 176,000 individual cells of 53 subtypes revealed two distinct microenvironments between tumour and peripheral blood in NPC. With such large-scale single-cell data, we identified novel cell populations with specific gene signatures in NPC. Moreover, in combination of TCR repertoire sequencing, we delineated the potential developmental trajectories of intratumoral immune cells. Furthermore, we dissected a multiple intercellular network in NPC using ligand-receptor pairing analyses (Fig. 7).

CD8<sup>+</sup> T cells are the key effector of anti-tumour immunity with cytotoxicity to kill tumour cells<sup>42</sup>. We observed an abundance of tumour infiltrating CD8<sup>+</sup> T cells in NPC, which exhibited clonal expansion, effector, proliferation, and exhausted status, suggesting that the CD8<sup>+</sup> T cells were largely suppressed amid being stimulated by the tumour neoantigens in the TME of NPC. This observation is consistent with the previous findings in other cancer types<sup>11,43</sup>, suggesting a common immunosuppressed state of CD8<sup>+</sup> T cells in tumours. Combining TCR repertoire and transcriptome analyses for each T cell, we revealed, for the first time to our knowledge, the differentiation trajectory of CD8<sup>+</sup> T cells in NPC, by which CX3CR1<sup>+</sup>CD8<sup>+</sup> T cells





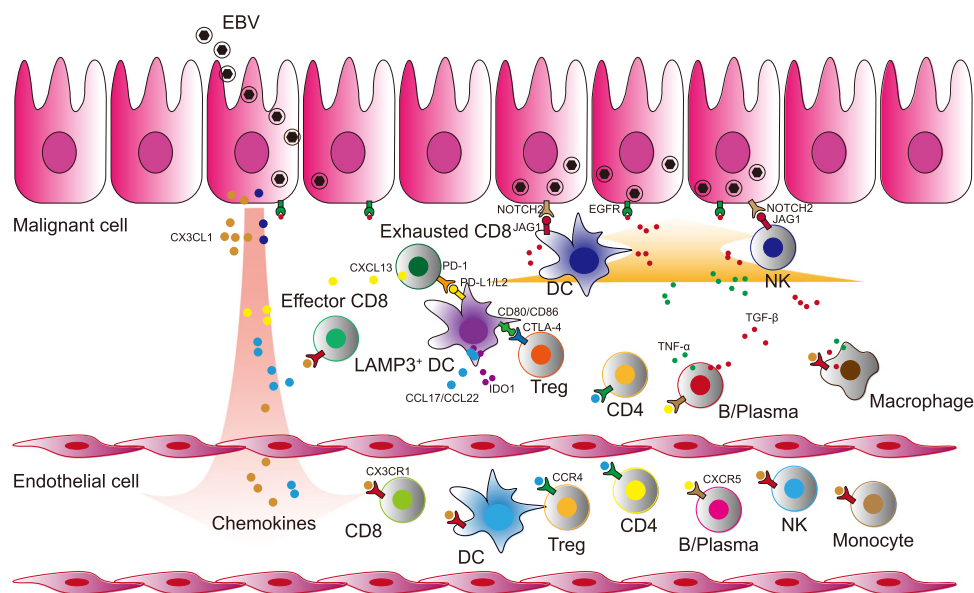
(CD8\_C5\_CX3CR1) in the peripheral blood infiltrated and transformed to the exhausted CD8<sup>+</sup> T cells (CD8\_C11\_PDCD1) in tumours. Consistently, the CX3CR1<sup>+</sup>CD8<sup>+</sup> T cells of highly cytotoxic potential shared TCR clonotypes with tumour infiltrating CD8<sup>+</sup> T cells, which are responsible for recognising antigens. In light of the previous findings that CX3CR1<sup>+</sup>CD8<sup>+</sup> T cells are essential for viral control and infiltrating to tumour site

to reduce tumour growth<sup>44,45</sup>, our findings raise possible T cell therapies for NPC by infusing the CX3CR1<sup>+</sup>CD8<sup>+</sup> T cells from peripheral blood after ex vivo expansion or engaging chimeric TCR of T cells with CX3CR1 chemotactic potential towards tumour site.

As cancer cells grow with high immunogenicity, inflammatory cells that are actively immunosuppressive, including regulatory



**Fig. 6 Intercellular interactions among immune and malignant cells in NPC.** **a, b** Dot plots showed selected ligand-receptor interactions between two cell clusters, for Treg and DC\_C3\_LAMP3 cells (**a**) and for exhausted CD8<sup>+</sup> T (CD8\_C11\_PDCD1) and DC\_C3\_LAMP3 cells (**b**). The ligand-receptor interactions and cell-cell interactions are indicated at columns and rows, respectively. The means of the average expression levels of two interacting molecules are indicated by colour heatmap (right panel), with blue to red representing low to high expression. The log<sub>10</sub>(*P*-values) were indicated by circle size in one-sided permutation test. Different colour boxes at the bottom represent different function modules of receptor-ligand interactions. **c** Representative images of multiplex IHC staining for the juxtaposition of CTLA4-expressing Treg cells (CD3<sup>+</sup>CD4<sup>+</sup>FOXP3<sup>+</sup>) and CD80-expressing DC\_C3\_LAMP3 cells in NPC tissue samples. Proteins detected using respective antibodies are indicated on top. The green, red, magenta, cyan, and orange arrows indicated positive cells with the expression of CD3, CD4, FOXP3, CTLA4, and CD80 proteins in NPC tissue, respectively (bottom panel). Images are representative of three biological replicates. Scale bars, 100 μm and 20 μm for top and bottom panels, respectively. **d** Representative images of multiplex IHC staining for the juxtaposition of PD1-expressing CD8<sup>+</sup> T cells (CD3<sup>+</sup>CD8<sup>+</sup>) and PD-L1-expressing DC\_C3\_LAMP3 cells (CD80<sup>+</sup>) in NPC tissue samples. Proteins detected using respective antibodies are indicated on top. The yellow, cyan, magenta, red, and green arrows indicated positive cells with the expression of CD3, CD8, CD80, PD1, and PD-L1 proteins in NPC tissue, respectively (bottom panel). Images are representative of three biological replicates. Scale bars, 100 and 20 μm for top and bottom panels, respectively.



**Fig. 7 Schematic diagram of cross-talks among multiple immune cells in the TME of NPC.** EBV infects nasopharyngeal epithelial cells and participates in the tumorigenic process of NPC. EBV-positive malignant NPC cells secrete a variety of chemokines (CX3CL1, etc.) and initiate the recruitment and tumoral infiltration of multiple immune cells with the chemokines receptors from the peripheral blood. Multiple tumour infiltrating immune cells activate EGFR and Notch pathway in EBV-positive malignant NPC cells. Naive CD8<sup>+</sup> cells infiltrate to the lesion and develop to effector and further exhausted CD8<sup>+</sup> cells. Peripheral DCs infiltrate to the tumour and differentiate into LAMP3<sup>+</sup> DCs. The mature LAMP3<sup>+</sup> DCs with the expression of PD-L1/PD-L2 interact with PD1 on CD8<sup>+</sup> T cells whereby the signalling restrains the activation of CD8<sup>+</sup> T cells and promotes their exhaustion. Treg cells interact with LAMP3<sup>+</sup> DCs through CTLA4-CD80/CD86, which might limit the antigen presentation process of DCs and promote the secretion of IDO1 to induce the proliferation of Treg cells. The intensive cell-cell interactions among LAMP3<sup>+</sup> DCs, Treg cells, exhausted CD8<sup>+</sup> T cells, and malignant cells foster an immune-suppressive niche for the tumour microenvironment of NPC.

T cells are recruited to help evade immune destruction by suppressing cytotoxic lymphocytes<sup>46</sup>. We observed three heterogeneous Treg cell clusters in NPC tumours, each with specific gene signatures and functions. Treg\_C4\_TNFRSF4 cells featured with high expression of TNFRSF family genes (*TNFRSF4*, *TNFRSF9*, and *TNFRSF18*) as well as *CCR8* and exhibited the strongest immune-suppressive function compared to other Treg cell clusters in NPC. Consistently, it has been reported previously that Treg cells with the expression of TNFRSF family genes facilitate tumour immune evasion and promote cancer development<sup>36,47</sup>. Moreover, high expression of *CCR8* has been demonstrated as a signature of Treg cells that restrain immunity, of which their amount in tumours is significantly associated with poor prognosis in several cancers<sup>26,48</sup>. Indeed, our present study revealed that the fraction of Treg\_C4\_TNFRSF4 in Treg cells might contribute to the poor survival of NPC. Given that *CCR8* can promote immune-suppressive Treg cells<sup>26</sup>, a blockade of *CCR8* signalling in Treg cells might abolish their specific

suppressive effect on cytotoxic lymphocytes and thus inhibit tumour growth. While the presence of tumoral Treg cells in NPC has been confirmed using IHC staining assays in our present study and previously by other group<sup>15</sup>, the origin of tumour infiltrating Treg cells in NPC remains exclusive. Pseudotime trajectory and TCR repertoire analyses revealed that Treg\_C4\_TNFRSF4 cells were differentiated from naive Treg\_C1\_SELL cells in peripheral blood through intermediate Treg\_C2\_HSPA1A or Treg\_C3\_MKI67 cells. We suspected that reducing the migration of Treg cells in the peripheral blood to tumour sites might attenuate the immune suppressions endowed by Treg cells. With this, we note that the high expression of *CCR4* in Treg\_C1\_SELL cells might be a potential therapeutic target, since it has been considered as the key molecule for Treg cells migrating into tumour<sup>49</sup>.

Tumour-associated myeloid cells are heterogeneous as reported previously<sup>50</sup>. We observed multiple clusters of myeloid cells in NPC tumours, among which DC\_C3\_LAMP3 could be

considered as a group of regulatory and tolerogenic DC, showing high expression of the migration (*CCR7*) and maturation (*LAMP3*) related genes. Both are the signatures of *LAMP3*<sup>+</sup> DCs that have been recently demonstrated in multiple cancers<sup>28,29</sup>. Moreover, DC\_C3\_LAMP3 cells had the elevated expression of *CD274*, *PDCD1LG2*, *CD200*, *IDO1*, *EBI3*, *SOCS1*, *SOCS2*, and *SOCS3*, which are immune-suppressive related genes with similar expression pattern as *LAMP3*<sup>+</sup> DCs in lung cancer<sup>51</sup>. These observations suggest that DC\_C3\_LAMP3 cells are a group of *LAMP3*<sup>+</sup> DCs in NPC, which are a ubiquitous cell population in tumours and exert immune-regulatory function and the control of T cell activation<sup>52</sup>. Our data also revealed the developmental trajectory of the *LAMP3*<sup>+</sup> DCs in NPC and the master transcription factors that are potentially crucial for the promoting maturation, decreased antigen presentation capability, and increased immune regulatory capability of *LAMP3*<sup>+</sup> DCs. These transcription factors are connected to each other, forming a robust regulatory gene expression network in *LAMP3*<sup>+</sup> DCs. We suspect that targeting these transcription factors might reshape the dendritic cells towards a normal antigen presentation phenotype, leading to potential therapeutic benefits for NPC.

EBV infection is the key feature of NPC in the endemic regions, and its role in malignant transformation and tumorigenesis has been implicated in NPC<sup>33</sup>. We identified two groups of malignant cells with (EP\_C1\_LMP1) or without (EP\_C2\_EP-CAM) EBV infection in NPC tumours. The EBV<sup>+</sup> NPC cells showed a distinct transcriptional state compared to the EBV<sup>-</sup> NPC cells, with the specific activation of major genes implicated in EBV entry and cancer-related pathways, such as NF- $\kappa$ B<sup>53</sup> and Notch pathways<sup>54</sup>. Moreover, the EBV<sup>+</sup> NPC cells had more chemotactic interactions with immune cells derived from tumours and peripheral blood, which are important for the regulation of EBV life cycle and the shaping of tumour behaviours including radiation sensitivity and stemness<sup>40,41</sup>. The potential chemotactic interactions might also explain the abundant infiltration of immune cells in NPC tumour stroma<sup>55</sup>. Given that all endemic NPC are EBV positive, our findings are of particular interests for future studies on the role of EBV infection in NPC development and the persistence of EBV in body cells compared to its gradual loss during in vitro culture. The specific presence of EBV molecules in NPC cells has been harnessed as an adoptive immunotherapy for NPC using cytotoxic T lymphocyte recognising EBV in clinical trials<sup>56,57</sup>. We note that the presence of TCRs recognising EBV antigen among multiple patients with NPC may engage T cells targeting NPC cells with higher specificity. However, it's noteworthy to explore whether and how the heterogeneous malignant NPC cells especially the EBV negative cells and other subtypes of cells with different proliferative capabilities contribute to the variable treatment outcomes. In addition, we didn't observe EBV transcripts in any B cells, although B cells have been known as a primary host for EBV<sup>58</sup>. It might be explained by the low expression level of EBV genes in infected cells and the small fraction of B cells captured in our study.

Our study revealed an intercellular network among *LAMP3*<sup>+</sup> DCs (DC\_C3\_LAMP3), Treg cells, and exhausted CD8<sup>+</sup> T cells (CD8\_C11\_PDCD1) in NPC, suggesting potential cross-talks among multiple immune cells to foster an immune-suppressive niche for the TME of NPC (Fig. 7). Indeed, the link between *LAMP3*<sup>+</sup> DCs and Treg cells through the interaction between CCL17-CCR4 and CCL22-CCR4 has also been demonstrated in other cancers<sup>36</sup>, by which *LAMP3*<sup>+</sup> DCs potentiate the chemotactic recruitment of peripheral Treg cells and promote their infiltration in tumours<sup>59</sup>. Moreover, *LAMP3*<sup>+</sup> DCs also had abundant expression of *IDO1* in NPC, which could induce the proliferation of tumour infiltrating Treg cells as reported previously<sup>60</sup>. On the other hand, our observation of the interaction

between tumoral Treg cells and *LAMP3*<sup>+</sup> DCs through CTLA4 and CD80/CD86 has been consistently reported in other cancers<sup>61</sup>, whereby Treg cells regulate the maturation of tolerogenic *LAMP3*<sup>+</sup> DCs. By contrast, the cross-talks between cytotoxic CD8<sup>+</sup> T cells and either DCs or Treg cells have been widely addressed in multiple cancers<sup>28,62,63</sup>. The mutual interactions between *LAMP3*<sup>+</sup> DCs and Treg cells may enhance immune-suppressive effects on the exhausted CD8<sup>+</sup> T cells in NPC. Together with the observation of chemotactic potential of EBV<sup>+</sup> NPC cells to recruit peripheral immune cells, these findings suggest that these cross-talks among diverse cell types play important roles in maintaining the homeostasis of TME in NPC. It would be plausible that disrupting the interactions might break the balance of TME and thus cure the tumour. Besides the promising results of PD-1/PD-L1 blockade immunotherapy in NPC, antibodies targeting EGFR have been reported to enhance the current treatment paradigms for locoregionally advanced NPC<sup>64</sup>. We suspect that immune-suppressive interaction of CD200-CD200R1 and LGALS9-HAVCR2 among *LAMP3*<sup>+</sup> DCs, Treg cells, and exhausted CD8<sup>+</sup> T cells might be also potential immunotherapeutic targets for NPC.

Taken together, through uncovering the heterogeneous tumour microenvironment of NPC at a high resolution, we identified the essential cells and molecules with potential contributions to NPC tumorigenesis, and thus provide insights into the mechanisms underlying NPC progression and the development of potential therapeutic strategies for NPC. We acknowledge that our study has several limitations. First, we observed some trends of the associations of cell compositions with clinical characteristics (Supplementary Fig. 13). Particularly, patients with more advanced NPC had a higher proportion of peripheral CD8\_C10\_MKI67 and more intensive cellular interactions in DC\_C4\_JCHAIN and NK\_C2\_FCER1G cells (Supplementary Fig. 13c, e). However, we are not yet to draw a conclusion with such a limited cohort size that any components of the TME could be associated with the clinical outcomes. Second, we believe that NPC is one of the malignancies that are shaped together with immunity and neoantigens endowed by the somatic alterations in malignant cells. However, we are not able to predict any neoantigens, because we have no DNA level (whole genome or exome sequencing) data with our patient samples for HLA estimations and somatic mutations callings. Lastly, although our findings of the key interaction network and molecules of the TME in NPC were obtained using robust bioinformatic analyses and additional immunostaining assays, further functional experiments are awaited to explore the biological consequences and underlying mechanisms.

## Methods

**Patient recruitment and sample collection.** Ten male individuals with nasopharyngeal carcinoma (NPC) were recruited from a local hospital in Guangzhou, China, an endemic region with high prevalence of NPC, between June 2018 and September 2018. The patients were histopathologically diagnosed with primary NPC by at least two pathologists according to the World Health Organization (WHO) classification. No history of cancer and any anti-tumour therapy prior to the primary diagnosis was self-reported. Clinical staging of NPC was determined according to the 8th edition of the International Union against Cancer (UICC) and American Joint Committee on Cancer (AJCC) staging system. Fresh tumour sample was obtained using endoscopic nasopharyngeal biopsy and matching peripheral blood sample was collected for each patient, followed immediately by single cell preparation as described below. All patients were EBV positive as confirmed using in situ hybridisation of EBV encoded small RNAs (EBERs) in tumour tissue. The average age was 50.6 and the patient's characteristics were listed in the supplementary (Supplementary Data 1). For immunostaining assays, additional NPC biopsies were collected. The specimens were collected within 30 min after the tumour resection and fixed in formalin for 48 h. Written informed consent was obtained from all participants, and the study was approved by the Institutional Review Boards at the Sun Yat-sen University Cancer Center (SYSUCC).

**Preparation of single cell suspensions.** Fresh tumour samples were processed independently with enzymatic digestion and mechanical dissociation immediately after collection to generate single cell suspensions. Briefly, each tumour was cut into small pieces with approximately 1-mm<sup>3</sup> in a D10 resuspension buffer, containing culture medium (DMEM medium; Gibco™, USA; Cat. no. 11965092) with 10% foetal bovine serum (FBS; Gibco; cat. no. 10099141), followed by enzymatic type II (Thermo Fisher, USA; cat. no. 17101015) and IV (Thermo Fisher; cat. no. 17104019) digestion for 30 min on a rotator at 37 °C. The digested mixture was passed through a 40 µm cell-strainer (BD Biosciences, USA; Cat. no. 352340) to obtain dissociated cells. The filtered mixture was centrifuged at 400 g for 5 min, and after removal of the supernatant, the pelleted cells were resuspended in 0.8% NH<sub>4</sub>Cl red blood cell lysis buffer and incubated on ice for 10 min. After washing twice with DPBS (Gibco; cat. no. 14190250), the dissociated cells from tumour were resuspended in a sorting buffer, consisting of 1X DPBS supplemented with 0.04% BSA (Sigma-Aldrich, USA; cat. no. 9048468). Viable cells were collected using fluorescence activated cell sorting (FACS; BD FACSAria III; BD Biosciences) with negative staining of propidium iodide (PI; Thermo Fisher, cat. no. P1304MP). At least 300,000 cells were collected for each tissue sample.

From blood sample, PBMCs were isolated using a leukocyte separation solution, following the manufacturer's instruction (HISTOPAQUE-1077; Sigma-Aldrich; cat. no. 10771). Briefly, 5-ml of fresh peripheral blood was collected in EDTA anticoagulant tubes (BD; Cat. no. 366643) and subsequently transferred onto the solution. After density gradient centrifugation for 20 min at 750 X g, PBMCs settled at the interphase were carefully collected and washed twice with DPBS. Residual red blood cells were lysed using the same procedure above-mentioned. Viable cells were collected using FACS with PI staining.

#### Library construction for single cell gene expression and TCR profiling.

Immune repertoire measurement and gene expression at single cell resolution were conducted using Chromium Single Cell V(D)J Reagent Kit (10x Genomics, USA) following the manufacturer's instructions. Briefly, the sorted cells were washed twice with the sorting buffer. Cell viability and number were determined using Trypan Blue (Thermo Fisher; Cat. no. 15250061) exclusion assay. Appropriate volume of cell suspension with a concentration of 700–1200 cells/µl were loaded in each channel, targeting a capture of 8,000 cells per sample, which were further mixed with barcoded gel beads on a Chromium Controller (10x Genomics). After reverse transcription reaction, cDNA amplification for 14 cycles was conducted on a thermal cycler (C1000; Bio-Rad, USA). The post-amplification cDNA was used as template to further enrich TCR fragments. Sequencing libraries for cDNA and TCR were separately constructed according to the instructions. The average fragment size of a library was quantitated using Qseq100 (Bioptic; Taiwan).

**Next generation sequencing and data processing.** Each DNA library was loaded into a sequencing lane on a HiSeq X system (Illumina, USA) and was sequenced with pair-end reads of 150 bp. Raw data of Binary Base Call (BCL) format was converted to FASTQ files using bcl2fastq (version v2.19.0.316, Illumina). Next, Cell Ranger pipelines (version 3.0.1; 10x Genomics) were used to align sequencing reads in the FASTQ files to reference genomes of interest and generate feature-barcode matrices. Single-cell 5'-gene expression data and TCR enriched data from the same cDNA library were processed using Cell Ranger count and Cell Ranger vj implemented in the pipelines, respectively. The gene expression data was mapped to human genome reference sequence (GRCh38; <http://cf.10xgenomics.com/supp/cell-exp/refdata-cellranger-GRCh38-1.2.0.tar.gz>) and EBV reference sequence<sup>65</sup> (Akata; <https://github.com/flemingtonlab/public/tree/master/annotation>) for cDNA sequencing reads. The TCR enriched data were mapped to the VDJ reference sequence (<http://cf.10xgenomics.com/supp/cell-vdj/refdata-cellranger-vdj-GRCh38-alt-ensembl-2.0.0.tar.gz>) for TCR sequencing reads.

#### Single-cell gene expression quantification and determination of cell types.

Doublets are artefactual libraries generated from two cells arising due to errors in droplet encapsulation of cells, and thus commonly affect the quality of single-cell sequencing data. The R package “DoubletFinder” (<https://github.com/chris-mcginnis-ucsf/DoubletFinder>) was applied to predict doublets in our data. Basically, a doublet is defined as a single-cell library representing more than one cell, and a closer examination of some known markers would suggest that the offending cluster consists of doublets of more than one cell type, while no cell type is known to strongly express both markers at the same time. We removed doublets in each sample individually, with an expected doublet rate of 0.05 and default parameters used otherwise (Supplementary Table 3). The remaining cells survived from the filtering criteria were single cells. Then the gene expression matrices for all remaining PBMC and tumour cells were combined and converted to a Seurat object using the R package Seurat (version 2.3.4, <https://satijalab.org/seurat>). Next, any cells were removed for which had either less than 101 UMIs, or expression of less than 501 genes, or over 15% UMIs linked to mitochondrial genes. From the remaining cells, gene expression matrices were generated with log normalisation and linear regression using the NormalizeData and ScaleData function of the Seurat package.

Because the samples were processed independently and high-dimensional variables are common in single-cell sequencing data, which might introduce potential batch effect, we used canonical correlation analysis (CCA) and

RunUMAP function implemented in Seurat to reduce dimensionality and remove batch effect. Cell clusters were identified using the FindClusters function in Seurat, with a *K* parameter of 20 and default parameters used otherwise. We annotated the clusters as different major cell types based on their average gene expression of well-known markers, including CD4<sup>+</sup> T cell (*PTPRC*, *CD3D*, and *CD4*), CD8<sup>+</sup> T cell (*PTPRC*, *CD3D*, and *CD8A*), myeloid cell (*CD14* and *ITGAX* encoding CD11C), malignant cell (*EPCAM* and *KRT* family genes), B cell (*CD19* and *MS4A1*), and NK cell (*FCGR3A* and *NCAM1*).

Repeating the abovementioned steps (normalisation, dimensionality reduction, and clustering), we further identified sub-clusters and annotated them as different specific cell subtypes by the average expression of respective gene sets in each major cell type. To identify marker genes for each sub-cluster within the major cell types (CD4<sup>+</sup> T, CD8<sup>+</sup> T, NK, B, myeloid, and malignant cells), the expression profiles of the sub-cluster were contrasted with those of the other sub-clusters using the Seurat FindAllMarkers function. Differential expression analysis implemented in the function compared all the genes in the two datasets using the default two-sided non-parametric Wilcoxon rank sum test. A significant differentially expressed gene was determined if it had the Bonferroni-adjusted *P* value lower 0.05 and an average natural logarithm (ln) fold-change of expression of at least 0.1 and 0.25 for malignant cells and other cells, respectively. The cluster with multiple well-defined marker genes of different cell types and an elevated number of UMI was considered cell contamination and removed in downstream analysis. For each cluster (like C1) of a major cell type (like CD4<sup>+</sup> T cells), we assigned a cluster identifier with a marker gene (like *LEF1*) as “CD4\_C1\_LEF1”. The selection criteria for the marker gene included (1) with top ranking at the differential gene expression analysis for the corresponding cell cluster, (2) with strong specificity of gene expression meaning high expression ratio within the corresponding cell cluster but low in other clusters, and (3) with literature supports that it's either a marker gene or functional relevant to the type of cell.

**Collection of public single-cell datasets.** To compare the features of tumour microenvironment including cell compositions between NPC and other types of cancers, we collected the single-cell data publicly available for multiple cancers, including NSCLC<sup>12</sup> (downloaded from <https://gbiomed.kuleuven.be/scRNAseq-NSCLC>), CRC<sup>66</sup> (Gene Expression Omnibus; <https://www.ncbi.nlm.nih.gov/geo/>; GEO accession number: GSE132465), and PDAC<sup>67</sup> (Genome Sequence Archive; <https://bigd.big.ac.cn/gsa/>; GSA accession number: CRA001160), as well as that for specific cell types, including CD4<sup>+</sup> T cells in CRC<sup>68</sup> (GEO accession number: GSE146771), NSCLC<sup>11</sup> (GEO accession number: GSE99254) and HCC<sup>28</sup> (GEO accession number: GSE140228) and DCs in NSCLC<sup>29</sup> (GEO accession number: GSE127465) and HCC<sup>28</sup> (GEO accession number: GSE140228).

**Calculation of functional module scores.** To evaluate the potential functions of a cell cluster of interest, we calculated the scores of functional modules for the cell cluster, using the AddModuleScore function in Seurat at single cell level. The average expression levels of the corresponding cluster were subtracted by the aggregated expression of control feature sets. All analysed genes were binned based on averaged expression, and the control features were randomly selected from each bin<sup>69</sup>. The functional modules including proliferation score for T and malignant cells, cytotoxicity and exhausted scores for CD8<sup>+</sup> T cells, IL2R, inhibitory, and co-stimulatory scores for Treg cells, as well as maturation, activation, migration, differentiation, apoptosis, antigen presentation, and immune regulatory scores for dendritic cells. The involved genes were listed in the supplementary material (Supplementary Data 3).

**Pathway enrichment analysis.** To gain functional and mechanistic insights of a cell cluster, we performed Gene Ontology (GO) and KEGG Pathway enrichment analyses using Metascape (<http://metascape.org/>) to identify biological pathways that were enriched in a certain gene list more than that would be expected by chance. For malignant cells, the gene list imported to Metascape included the top 100 differentially expressed genes (DEGs) with a natural logarithm of fold changes of expression (lnFC) > 0.1 in clusters. For non-malignant cells, the gene list included the top 100 DEGs with lnFC > 0.25 in clusters. *P* value < 0.05 was considered to be a significant enrichment. To compare the difference of signalling pathway enrichment between two clusters (Treg\_C2\_HSPA1A versus Treg\_C4\_TNFRSF4 and DC\_C2\_CD11C versus DC\_C3\_LAMP3), we performed the gene set enrichment analysis (GSEA; version 3.0) using the selected molecular signatures database v7.0<sup>70</sup>. To explore the heterogeneous expression of malignant cells, we performed gene set variation analysis (GSVA, version 1.34.0), using 18 hallmark pathways described in the molecular signature database.

**Developmental trajectory inference.** To characterise the potential process of immune cell functional changes and determine the potential lineage differentiation among diverse immune cells, we performed trajectory analyses for Treg, CD8<sup>+</sup> T, and dendritic cells, using Monocle2 (version 2.8.0; <http://cole-trapnell-lab.github.io/monocle-release/monocle2/>). The data of the indicated clusters calculated in Seurat was fed directly into Monocle2. Next, we carried out density peak clustering (Monocle2 dpFeature procedure) to order cells based on the genes with differential expression between clusters, using the differentialGeneTest function in Monocle2.



The top 1,000–2,000 significant genes (ordered by  $q$  value) were used for ordering in all instances. Then the immune cell differentiation trajectory was inferred after dimension reduction and cell ordering with the default parameters of Monocle2.

**TCR repertoire analysis.** The outputs of CellRanger vdj included the assembled nucleotide sequences for both  $\alpha$  and  $\beta$  chains, the coding potential of the nucleotide sequences (that is productive or not), the translated amino acid sequence, the CDR3 sequences, and the estimated UMI value of  $\alpha$  or  $\beta$  chains. Only cells with UMI values larger than 1 for  $\alpha$  and  $\beta$  chains were kept. The dominant TCR of a single cell was defined based on an in-frame TCR  $\alpha$ - $\beta$  pair. If one clonotype defined as a unique in-frame TCR  $\alpha$ - $\beta$  pair was present in at least two cells, this clonotype would be considered clonal, and the number of cells with such dominant  $\alpha$ - $\beta$  pair indicated the degree of clonality of the clonotype.

R package STARTRAC (version 0.1.0) was used to assess the enrichment of TCR in various T cell clusters. The degree of clonal expansion, tissue migration, and state transition of T cell clusters upon TCR tracking were determined using three STARTRAC indices, STARTRAC-expa, STARTRAC-migr, and STARTRAC-tran, respectively. For detailed pipeline, please referred to the website (<https://github.com/Japrin/STARTRAC>).

**Bulk RNA sequencing and data analysis.** To identify the genes responsible for EBV infection and carcinogenesis, we compared the expression profiles between NPC and non-cancerous control cohorts. The data for the NPC cohort was retrieved from the public GEO database (Accession number: GSE102349), including 113 NPC tissue samples profiled by RNA-seq<sup>71</sup>. The control cohort was in-house RNA-seq data of 10 rhinitis samples that had been published recently<sup>72</sup>. For data analysis, pair-end reads with high quality were aligned to ribosome RNAs using Bowtie2<sup>73</sup>, and reads after removal of those being aligned as ribosome RNAs were realigned to the human genome (GRCh38) and EBV (Akata) reference sequence using HISAT2 with default settings<sup>74</sup>. HTseq was used to quantitate the read counts of each gene<sup>75</sup>. The expression levels of genes were normalised as Transcripts Per Kilobase Million (TPM), to minimise the potential effect of tumour purity.

We next assessed the associations of the immune signatures and the survival of NPC. Receiver operating characteristic (ROC) was used to determine the optimal cut-off value of gene expression for patient stratification. Kaplan-Meier analysis was conducted to reveal the prognostic ability of normalised mRNA ratio of *CCR8/FOXP3* in 88 NPC samples with prognostic information from the public cohort, and a log-rank test was performed to compare the survival between high and low normalised mRNA ratios of *CCR8/FOXP3*. For the correlation analyses of the expression of immune signature genes among LAMP3<sup>+</sup> DC (DC\_C3\_LAMP3), Treg cell, and exhausted CD8<sup>+</sup> T cell (CD8\_C11\_PDCD1), we first selected the signature genes based on the top 200 differentially expressed genes among all cell subsets (Supplementary Data 7) and then calculated the mean of the expression (TPM) for signature genes as signature scores. Pearson correlation between signatures was calculated by `cor()` function in R programme.

**Assembly of context specific regulatory models and master regulator analysis of LAMP3<sup>+</sup> dendritic cell (DC).** A LAMP3<sup>+</sup> DC context-specific network model of transcriptional regulation was assembled with the ARACNe<sup>76</sup> (<https://github.com/califano-lab/ARACNe-AP>), based on 357 LAMP3<sup>+</sup> DC (DC\_C3\_LAMP3) single cell RNA-sequencing expression profiles. ARACNe was run with 100 bootstraps, a  $P$  value threshold of  $10^{-8}$ , and 0 data processing inequality (DPI) tolerance, generating a network of 38 TFs associated with 2,495 target genes by 10,759 interactions. Pearson correlation between TFs and target genes was calculated by `cor()` function in R. We considered the correlation coefficient between TF and downstream target genes greater than 0 as positive regulation, otherwise negative regulation.

**inferCNV analysis.** To identify malignant cells, we identified evidence for somatic alterations of large-scale chromosomal copy number variants, either gains or losses, in a single cell using inferCNV (<https://github.com/broadinstitute/inferCNV>), in addition to the expression of *EPCAM*. The raw single-cell gene expression data was extracted from the Seurat object according to the software recommendation. A public single-cell data derived from normal epithelium cells was included as a control reference<sup>32</sup> (GEO accession number: GSE121600). We preformed inferCNV analysis with the default parameters.

**Cellular communication analysis.** To investigate the potential cell-cell communications between any two different cell types in NPC, we performed ligand-receptor analyses using CellPhoneDB software (version 2.0.6; <https://github.com/Teichlab/cellphonedb>). CellPhoneDB applies an algorithm that considers only receptors and ligands with broad expression among the tested cell types, followed by calculating the likelihood of cell-type specificity of a given receptor-ligand complex with a sufficient number of permutations. The gene expression matrixes of CD8<sup>+</sup> T, CD4<sup>+</sup> T, NK, B, myeloid, and malignant cells were selected as input for the CellPhoneDB analysis. We identified the most relevant cell-type specific ligand-receptor interactions and considered only ligands and receptors with expression in more than 20% of the cells in the corresponding sub-clusters. Moreover, we permuted the change of cell type label for each cell at 1000 times to calculate the

significance of each pair. The  $P$  value was calculated using the proportion of the mean value for specific receptor-ligand pairs compared to a randomly permuted mean distribution. Finally, we prioritised the interactions with a  $P$ -value greater than 0.05 and selected the interaction pairs with biological relevance.

**Immunostaining assays.** For tissue sample stored in the formalin, dehydration and embedding in paraffin were performed according to routine methods. The paraffin blocks were cut into 5  $\mu$ m slides and adhered on the glass slides. The paraffin-embedded sections were dewaxed, rehydrated, subjected to the blockade of endogenous peroxidase activity, and antigen retrieval at high-temperature. Subsequently, the sections were processed further for either multiplex immunofluorescence (IF) or immunohistochemistry (IHC) staining assays.

Multiplex IF staining assays were conducted to determine the presence of EBV<sup>+</sup> and EBV<sup>-</sup> NPC cells. The sections were permeabilized in PBS with 0.5% Triton X-100 (Sigma-Aldrich; Cat. no. T8787) and incubated for overnight at 4 °C with the following primary antibodies: anti-EPCAM (rabbit; Abcam, USA; Cat. no. ab71916; 1:100) and anti-LMP1 (mouse; Abcam; Cat. no. ab78113; 1:500). Subsequently, the sections were incubated with Cy3 conjugated Goat Anti-Rabbit IgG and FITC conjugated Goat Anti-Mouse IgG secondary antibodies (Servicebio, China; Cat. no. GB21303/GB22301; 1:300). Nuclei were counterstained with 4'-6'-diamidino-2-phenylindole (DAPI; Sigma-Aldrich; Cat. no. D9542). Images were captured using a confocal laser-scanning microscope (LSM880; Zeiss, Germany).

To determine the spatial contact of LAMP3<sup>+</sup> DCs, Treg cells, and CD8<sup>+</sup> T cells, we performed multiplex IHC staining assays using the PANO 7-plex IHC kit (Panovue, China) according to the manufacturer's instructions. The slides were incubated with blocking antibody diluent at room temperature for 10 min, and then incubated overnight at 4 °C with primary antibodies. The slides were then incubated with the secondary antibody (HRP polymer, anti-mouse/rabbit IgG) at room temperature for 10 min. Subsequently, fluorophore (tyramide signal amplification or TSA plus working solution) was applied to the sections, followed by heat-treatment with microwave. The primary antibodies were applied sequentially, followed by incubation with the secondary antibody and TSA treatment. Nuclei were stained with DAPI after all the antigens had been labelled. Multispectral images for each stained slide were captured using the Mantra System (PerkinElmer, USA). Primary antibodies included anti-CD3 (rabbit; Abcam; Cat. no. ab135372; 1:50), anti-CD4 (rabbit; Abcam; Cat. no. ab133616; 1:100), anti-CD8A (mouse; CST, USA; Cat. no. CST7030; 1:100), anti-FOXP3 (mouse; Abcam; Cat. no. ab22510; 1:100), anti-CD80 (mouse; R&D Systems, USA; Cat. no. MAB140; 1:80), anti-PD1 (mouse; CST; Cat. no. CST43248; 1:50), anti-PD-L1 (rabbit; CST; Cat. no. CST13684; 1:50), and anti-CTLA4 (rabbit; Abcam; Cat. no. ab237712; 1:100).

**Statistics analysis.** All statistical analyses were performed using R (<http://www.r-project.org>), including two-sided paired student  $t$ -test, two-sided Wilcoxon test, two-sided Pearson correlation test, and two-sided Kruskal-Wallis test.  $P < 0.05$  was considered as statistical significance. Multiplex IF and IHC staining assays were confirmed in at least three biological replicates.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

The raw sequence data reported in this study have been deposited in the Genome Sequence Archive (Genomics, Proteomics & Bioinformatics 2017) in the National Genomics Data Center (Nucleic Acids Res 2020), Beijing Institute of Genomics (China National Center for Bioinformation), Chinese Academy of Sciences, under accession number HRA000159 (accessible at <http://bigd.big.ac.cn/gsa-human>) and GEO dataset under the accession number GSE162025. The key data in this study has also been deposited in the Research Data Deposit (RDDB2020000980; <http://www.researchdata.org.cn/>). Other datasets described in the Methods could be downloaded from NCBI GEO under the accession numbers GSE132465<sup>66</sup>, GSE146771<sup>68</sup>, GSE99254<sup>11</sup>, GSE140228<sup>28</sup>, GSE127465<sup>29</sup>, GSE102349<sup>71</sup>, and GSE121600<sup>32</sup>, Genome Sequence Archive under the accession numbers CRA001160<sup>67</sup>, and the URL <https://biomed.kuleuven.be/scRNAseq-NSCLC><sup>12</sup>. The remaining data are available within the Article, Supplementary Information or available from the authors upon request.

## Code availability

The scripts are available at <https://github.com/bei-lab/scRNA-of-NPC77>.

Received: 28 April 2020; Accepted: 7 January 2021;

Published online: 02 February 2021

## References

1. Chung, G. T. et al. Constitutive activation of distinct NF- $\kappa$ B signals in EBV-associated nasopharyngeal carcinoma. *J. Pathol.* **231**, 311–322 (2013).



2. Chen, Y. P. et al. Nasopharyngeal carcinoma. *Lancet* **394**, 64–80 (2019).
3. Sun, Y. et al. Induction chemotherapy plus concurrent chemoradiotherapy versus concurrent chemoradiotherapy alone in locoregionally advanced nasopharyngeal carcinoma: a phase 3, multicentre, randomised controlled trial. *Lancet Oncol.* **17**, 1509–1520 (2016).
4. Li, W. F. et al. Concurrent chemoradiotherapy with/without induction chemotherapy in locoregionally advanced nasopharyngeal carcinoma: Long-term results of phase 3 randomized controlled trial. *Int. J. Cancer* **145**, 295–305 (2019).
5. Chan, A. T. et al. Multicenter, phase II study of cetuximab in combination with carboplatin in patients with recurrent or metastatic nasopharyngeal carcinoma. *J. Clin. Oncol.* **23**, 3568–3576 (2005).
6. Ma, B. B. Y. et al. Antitumor activity of nivolumab in recurrent and metastatic nasopharyngeal carcinoma: an International, Multicenter Study of the Mayo Clinic Phase 2 Consortium (NCI-9742). *J. Clin. Oncol.* **36**, 1412–1418 (2018).
7. Hsu, C. et al. Safety and antitumor activity of pembrolizumab in patients with programmed death-ligand 1-positive nasopharyngeal carcinoma: results of the KEYNOTE-028 Study. *J. Clin. Oncol.* **35**, 4050–4056 (2017).
8. Lin, D. C. et al. The genomic landscape of nasopharyngeal carcinoma. *Nat. Genet.* **46**, 866–871 (2014).
9. Li, Y. Y. et al. Exome and genome sequencing of nasopharynx cancer identifies NF- $\kappa$ B pathway activating mutations. *Nat. Commun.* **8**, 14121 (2017).
10. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646–674 (2011).
11. Guo, X. et al. Global characterization of T cells in non-small-cell lung cancer by single-cell sequencing. *Nat. Med.* **24**, 978–985 (2018).
12. Lambrechts, D. et al. Phenotype molding of stromal cells in the lung tumor microenvironment. *Nat. Med.* **24**, 1277–1289 (2018).
13. Wang, H. Y. et al. A new prognostic histopathologic classification of nasopharyngeal carcinoma. *Chin. J. Cancer* **35**, 41 (2016).
14. Lau, K. M. et al. Increase in circulating Foxp3+CD4+CD25(high) regulatory T cells in nasopharyngeal carcinoma patients. *Br. J. Cancer* **96**, 617–622 (2007).
15. Zhang, Y. L. et al. Different subsets of tumor infiltrating lymphocytes correlate with NPC progression in different ways. *Mol. Cancer* **9**, 4 (2010).
16. Wang, Y. Q. et al. Prognostic significance of tumor-infiltrating lymphocytes in non-disseminated nasopharyngeal carcinoma: a large-scale cohort study. *Int. J. Cancer* **142**, 2558–2566 (2018).
17. Zhao, J. et al. Single cell RNA-seq reveals the landscape of tumor and infiltrating immune cells in nasopharyngeal carcinoma. *Cancer Lett.* **477**, 131–143 (2020).
18. Chen, Y. P. et al. Single-cell transcriptomics reveals regulators underlying immune cell diversity and immune subtypes associated with prognosis in nasopharyngeal carcinoma. *Cell Res.* **30**, 1024–1042 (2020).
19. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33**, 495–502 (2015).
20. Perry, M. & Whyte, A. Immunology of the tonsils. *Immunol. Today* **19**, 414–421 (1998).
21. Xu, X. et al. IFN- $\gamma$ -producing Th1-like regulatory T cells may limit acute cellular renal allograft rejection: Paradoxical post-transplantation effects of IFN- $\gamma$ . *Immunobiology* **222**, 280–290 (2017).
22. Bagaev, D. V. et al. VDJdb in 2019: database extension, new analysis infrastructure and a T-cell receptor motif compendium. *Nucleic Acids Res* **48**, D1057–D1062 (2020).
23. Griffith, J. W., Sokol, C. L. & Luster, A. D. Chemokines and chemokine receptors: positioning cells for host defense and immunity. *Annu. Rev. Immunol.* **32**, 659–702 (2014).
24. Zhang, L. et al. Lineage tracking reveals dynamic relationships of T cells in colorectal cancer. *Nature* **564**, 268–272 (2018).
25. Chinen, T. et al. An essential role for the IL-2 receptor in Treg cell function. *Nat. Immunol.* **17**, 1322–1333 (2016).
26. Barsheshet, Y. et al. CCR8(+)FOXP3(+) Treg cells as master drivers of immune regulation. *Proc. Natl Acad. Sci. USA* **114**, 6086–6091 (2017).
27. Chaudhary, B. & Elkord, E. Regulatory T cells in the tumor microenvironment and cancer progression: role and therapeutic targeting. *Vaccines* **4**, 28 (2016).
28. Zhang, Q. et al. Landscape and dynamics of single immune. *Cells Hepatocell Carcinoma Cell* **179**, 829–845 e820 (2019).
29. Zilionis, R. et al. Single-cell transcriptomics of human and mouse lung cancers reveals conserved myeloid populations across individuals and species. *Immunity* **50**, 1317–1334 e1310 (2019).
30. Li, H. & Shi, B. Tolerogenic dendritic cells and their applications in transplantation. *Cell Mol. Immunol.* **12**, 24–30 (2015).
31. Miller, J. C. et al. Deciphering the transcriptional network of the dendritic cell lineage. *Nat. Immunol.* **13**, 888–899 (2012).
32. Ruiz Garcia, S. et al. Novel dynamics of human mucociliary differentiation revealed by single-cell RNA sequencing of nasal epithelial cultures. *Development* **146**, dev177428 (2019).
33. Young, L. S. & Dawson, C. W. Epstein-Barr virus and nasopharyngeal carcinoma. *Chin. J. Cancer* **33**, 581–590 (2014).
34. Zhang, H. et al. Ephrin receptor A2 is an epithelial cell receptor for Epstein-Barr virus entry. *Nat. Microbiol.* **3**, 1–8 (2018).
35. Vento-Tormo, R. et al. Single-cell reconstruction of the early maternal-fetal interface in humans. *Nature* **563**, 347–353 (2018).
36. Ohue, Y. & Nishikawa, H. Regulatory T (Treg) cells in cancer: can Treg cells be a new therapeutic target? *Cancer Sci.* **110**, 2080–2089 (2019).
37. Liu, J. Q. et al. CD200-CD200R Pathway in the Regulation of Tumor Immune Microenvironment and Immunotherapy. *Adv. Exp. Med. Biol.* **1223**, 155–165 (2020).
38. Fujihara, S. et al. Galectin-9 in cancer therapy. *Recent Pat. Endocr. Metab. Immune Drug Disco.* **7**, 130–137 (2013).
39. Velapasamy, S., Dawson, C. W., Young, L. S., Paterson, I. C. & Yap, L. F. The dynamic roles of TGF- $\beta$  signalling in EBV-associated cancers. *Cancers* **10**, 247 (2018).
40. Shen, E. D. & Zeng, Q. Inhibition of the Numb/Notch signaling pathway increases radiation sensitivity in human nasopharyngeal carcinoma cells. *Kaohsiung J. Med. Sci.* **35**, 474–485 (2019).
41. Yu, S. et al. Notch inhibition suppresses nasopharyngeal carcinoma by depleting cancer stem-like side population cells. *Oncol. Rep.* **28**, 561–566 (2012).
42. Apetoh, L. et al. Consensus nomenclature for CD8(+) T cell phenotypes in cancer. *Oncoimmunology* **4**, e998538 (2015).
43. Li, H. et al. Dysfunctional CD8 T cells form a proliferative, dynamically regulated compartment within human melanoma. *Cell* **176**, 775–789 e718 (2019).
44. Zander, R. et al. CD4(+) T cell help is required for the formation of a cytolytic CD8(+) T cell subset that protects against chronic infection and cancer. *Immunity* **51**, 1028–1042 e1024 (2019).
45. Siddiqui, I., Erreni, M., van Brakel, M., Debets, R. & Allavena, P. Enhanced recruitment of genetically modified CX3CR1-positive human T cells into Fractalkine/CX3CL1 expressing tumors: importance of the chemokine gradient. *J. Immunother. Cancer* **4**, 21 (2016).
46. Mouggiakakos, D., Choudhury, A., Lladser, A., Kiessling, R. & Johansson, C. C. Regulatory T cells in cancer. *Adv. Cancer Res.* **107**, 57–117 (2010).
47. Facciabene, A., Motz, G. T. & Coukos, G. T-regulatory cells: key players in tumor immune escape and angiogenesis. *Cancer Res.* **72**, 2162–2171 (2012).
48. De Simone, M. et al. Transcriptional landscape of human tissue lymphocytes unveils uniqueness of tumor-infiltrating T regulatory cells. *Immunity* **45**, 1135–1147 (2016).
49. Collins, A. T., Berry, P. A., Hyde, C., Stower, M. J. & Maitland, N. J. Prospective identification of tumorigenic prostate cancer stem cells. *Cancer Res.* **65**, 10946–10951 (2005).
50. Azizi, E. et al. Single-cell map of diverse immune phenotypes in the breast tumor microenvironment. *Cell* **174**, 1293–1308 e1236 (2018).
51. Maier, B. et al. A conserved dendritic-cell regulatory program limits antitumour immunity. *Nature* **580**, 257–262 (2020).
52. Oh, S. A. et al. PD-L1 expression by dendritic cells is a key regulator of T-cell immunity in cancer. *Nat. Cancer* **1**, 681–691 (2020).
53. de Oliveira, D. E., Ballon, G. & Cesarman, E. NF- $\kappa$ B signaling modulation by EBV and KSHV. *Trends Microbiol.* **18**, 248–257 (2010).
54. Cheng, F., Pekkonen, P. & Ojala, P. M. Instigation of Notch signaling in the pathogenesis of Kaposi’s sarcoma-associated herpesvirus and other human tumor viruses. *Future Microbiol.* **7**, 1191–1205 (2012).
55. Gourzones, C., Barjon, C. & Busson, P. Host-tumor interactions in nasopharyngeal carcinomas. *Semin Cancer Biol.* **22**, 127–136 (2012).
56. Li, F. et al. Delayed-type hypersensitivity (DTH) immune response related with EBV-DNA in nasopharyngeal carcinoma treated with autologous dendritic cell vaccination after radiotherapy. *J. Immunother.* **36**, 208–214 (2013).
57. Li, J. J. et al. Autologous cytokine-induced killer cell transfusion in combination with gemcitabine plus cisplatin regimen chemotherapy for metastatic nasopharyngeal carcinoma. *J. Immunother.* **35**, 189–195 (2012).
58. Zhang, J., Jia, L., Tsang, C. M. & Tsao, S. W. EBV infection and glucose metabolism in nasopharyngeal carcinoma. *Adv. Exp. Med. Biol.* **1018**, 75–90 (2017).
59. Mizukami, Y. et al. CCL17 and CCL22 chemokines within tumor microenvironment are related to accumulation of Foxp3+ regulatory T cells in gastric cancer. *Int. J. Cancer* **122**, 2286–2293 (2008).
60. Wu, H., Gong, J. & Liu, Y. Indoleamine 2, 3-dioxygenase regulation of immune response (Review). *Mol. Med. Rep.* **17**, 4867–4873 (2018).
61. Tanaka, A. & Sakaguchi, S. Regulatory T cells in cancer immunotherapy. *Cell Res.* **27**, 109–118 (2017).
62. Jang, J. E. et al. Crosstalk between regulatory T cells and Tumor-associated dendritic cells negates anti-tumor immunity in pancreatic cancer. *Cell Rep.* **20**, 558–571 (2017).

63. Sanchez-Paulete, A. R. et al. Antigen cross-presentation and T-cell cross-priming in cancer immunology and immunotherapy. *Ann. Oncol.* **28**, xii44–xii55 (2017).
64. Peng, L. et al. The efficacy and safety of anti-epidermal growth factor receptor monoclonal antibodies in nasopharyngeal carcinoma: literature-based meta-analyses. *J. Cancer* **9**, 4510–4520 (2018).
65. O'Grady, T., Baddoo, M. & Flemington, E. K. Analysis of EBV transcription using high-throughput RNA sequencing. *Methods Mol. Biol.* **1532**, 105–121 (2017).
66. Lee, H. O. et al. Lineage-dependent gene expression programs influence the immune landscape of colorectal cancer. *Nat. Genet.* **52**, 594–603 (2020).
67. Peng, J. et al. Single-cell RNA-seq highlights intra-tumoral heterogeneity and malignant progression in pancreatic ductal adenocarcinoma. *Cell Res.* **29**, 725–738 (2019).
68. Zhang, L. et al. Single-cell analyses inform mechanisms of myeloid-targeted therapies in colon cancer. *Cell* **181**, 442–459 e429 (2020).
69. Tirosh, I. et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189–196 (2016).
70. Liberzon, A. et al. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* **1**, 417–425 (2015).
71. Zhang, L. et al. Genomic analysis of nasopharyngeal carcinoma reveals TME-based subtypes. *Mol. Cancer Res.* **15**, 1722–1732 (2017).
72. Lin, G. W. et al. Genetic risk of extranodal natural killer T-cell lymphoma: a genome-wide association study in multiple populations. *Lancet Oncol.* **21**, 306–316 (2020).
73. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
74. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–360 (2015).
75. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
76. Margolin, A. A. et al. ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics* **7** (Suppl 1), S7 (2006).
77. Yang Liu, S. H. et al. Tumour heterogeneity and intercellular networks of nasopharyngeal carcinoma at single cell resolution. *Zenodo* <https://doi.org/10.5281/zenodo.4287763> (2020).

## Acknowledgements

This study is supported by the national Key research and development program of China (2016YFC0902001), the Sci-Tech Project Foundation of Guangzhou City (201707020039), Guangdong Innovative and Entrepreneurial Research Team Program (2016ZT06S638), the National High Technology Research and Development Program of China (2012AA02A206), the National Program for Support of Top-Notch Young Professionals (J.-X.B.), Chang Jiang Scholars Program (J.-X.B.), Special Support Program of Guangdong (J.-X.B.) and Sun Yat-sen University Young Teacher Key Cultivate Project

(17ykzd29). We thank all the participants in the study, and staffs at the High-Throughput Analysis Platform (HTAP) of SYSUCC for data generation and processing.

## Author contributions

J.X.B. and Y.X.Z. conceived and supervised this study. S.H., Y.L., W.P., Q.Y.C., J.R.C., Q. Y.W., R.J.P., P.P.W., X.G., B.L., X.J.X. and H.Q.M. performed sample preparation, library construction, and data generation. Y.L., X.L.W., S.H., B.W.H., D.M.C., G.W.L., Y.Q.L. and X.D.H. performed bioinformatics analyses under the supervision of J.X.B. and Z.Z., Y.L. and J.X.B. wrote the manuscript. All authors read and approved the final manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-021-21043-4>.

**Correspondence** and requests for materials should be addressed to Z.Z., Y.-X.Z. or J.-X.B.

**Peer review information** *Nature Communications* thanks Jianjun Zhang and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021