Lymph nodes molecular subtypes unravel lymph nodes heterogeneity and clinical implications in colorectal cancer

Xu Guan[1,2,9], Yueyang Zhang[1,9], Ruifang Sun[3,9], Guiyu Wang[4,9], Xiaoman Bi[5,9], Zhen Zhang[6], Likun Zan[6], Yongsheng Meng[3], Yunxiao Liu[4], Jinyuan Guo[7], Jian Ma[1], Jiale Li[2], Hanqing Hu[4], Yixun Zhang[2], Xin Zhang[2], Maoxi Liu[2], Bo Jiang[2], Haiyi Liu[2], Shuangmei Zou[8], Yanfeng Xi[6] ✉, Haitao Zhou [1] ✉, Wenqi Bai[2] ✉, Deng Wu[5] ✉ & Xishan Wang [1] ✉

Lymph nodes (LNs) play a pivotal role in colorectal cancer (CRC) progression and immunity, yet their molecular and functional diversity remains poorly understood. By analyzing 630 LNs and 88 primary tumors from 200 CRC patients across four independent cohorts using bulk and single-cell RNA sequencing, we identify four non-metastatic negative LNs (NLN) subtypes (NLN_C1-C4) exhibiting obviously different immune function and stromal expansion. NLN_C3/C4 are characterized by diminished T and B cell activity and fibroblast-driven fibrosis, with follicular dendritic cell loss contributing to B cell dysfunction. Immune checkpoint inhibitors partially reverse these effects, restoring FDC and B cell activity. LNs subtypes demonstrate heterogeneity across patients and within individuals, with higher NLN_C3/C4 proportions associated with advanced tumor stages, poorer survival, and recurrence. Here, we report LNs subtypes as critical manifestations of LN heterogeneity in CRC, providing a basis for improved clinical stratification and LN-targeted therapeutic strategies.

Colorectal cancer (CRC) is one of the most prevalent malignancies globally, with a leading cause of cancer-related death[1]. Lymph node (LN) status, traditionally evaluated by the presence or absence of metastatic cancer cells, is a key factor in the staging of CRC and has long been associated with prognosis[2-4]. However, beyond their role as indicators of metastatic spread, LNs serve as critical sites for antigen presentation and the activation of adaptive immune responses[5,6]. These immune hubs facilitate the coordination of T and B cell responses against cancer cells. Recent studies have shown that tumor draining LNs exhibit diminished immune function compared to non-

[1]Department of Colorectal Surgery, National Cancer Center, National Clinical Research Center for Cancer, Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China. [2]Department of Colorectal Surgery, Shanxi Province Cancer Hospital, Hospital Affiliated to Cancer Hospital, Chinese Academy of Medical Sciences, Cancer Hospital Affiliated to Shanxi Medical University, Taiyuan, China. [3]Department of Tumor Biobank, Shanxi Province Cancer Hospital, Shanxi Hospital Affiliated to Cancer Hospital, Chinese Academy of Medical Sciences, Cancer Hospital Affiliated to Shanxi Medical University, Taiyuan, China. [4]Department of Colorectal Cancer Surgery, The Second Affiliated Hospital of Harbin Medical University, Nangang District, Harbin, China. [5]College of Biomedical Information and Engineering, Hainan Medical University, Haikou, China. [6]Department of Pathology, Shanxi Province Cancer Hospital, Shanxi Hospital Affiliated to Cancer Hospital, Chinese Academy of Medical Sciences, Cancer Hospital Affiliated to Shanxi Medical University, Taiyuan, China. [7]The Second Clinical Medical School, Shanxi Medical University, Taiyuan, China. [8]Department of Pathology, National Cancer Center, National Clinical Research Center for Cancer, Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China. [9]These authors contributed equally: Xu Guan, Yueyang Zhang, Ruifang Sun, Guiyu Wang, Xiaoman Bi. ✉e-mail: xiyanfeng1998@163.com; zhouhaitao@cicams.ac.cn; baiwenqi1998@163.com; wudeng@muhn.edu.cn; wxshan_1208@126.com

tumor draining LNs, highlighting their role in immune evasion and providing a basis for developing novel therapeutic strategies aimed at reactivating immune responses[7-9]. In the context of CRC, LNs not only act as barriers to metastatic spread but also play an active role in modulating the immune environment[10-12]. Therefore, the simplistic binary classification of categorizing LNs into "metastatic LNs" and "non-metastatic LNs" fails to capture the complexity of LN immune function, which is increasingly critical in clinical practice of tumor nodal staging.

In 2023, Paulina et al. introduced the concept of LN heterogeneity, emphasizing that LNs differ not only in their morphology and immune cell composition but also in their functional states across different individuals and disease stages[13]. In the case of tumor draining LNs, this heterogeneity manifests as progressive immune suppression, driven by both structural and molecular changes that impair lymphatic transport and immune surveillance[11]. However, despite the growing focus on LNs heterogeneity, the issue of how to reveal the characteristics of LNs heterogeneity, especially in clinical practice, remains an urgent topic for exploration. In the era of immunotherapy, this issue significantly hinders our ability to comprehensively elucidate the intricate relationship between LNs and tumor dynamics. Molecular subtypes, as a critical manifestation of heterogeneity, has become a pivotal framework for characterizing molecular features of cancers, profoundly influencing therapeutic strategies across various malignancies[14,15]. In contrast, the LNs subtypes, which are an essential foundation for understanding their interaction with tumors, remain conspicuously absent from current research and literature.

In this work, we establish a LNs molecular subtypes (LyMoS) in CRC based on whole-transcriptome sequencing of LNs with pathological validation, and further characterize these intrinsic subtypes to better explain LNs heterogeneity, and explore their relationship with clinical pathological features, patients' outcomes and immune checkpoint inhibitors (ICIs) treatment.

Results

The establishment and validation of LNs subtypes
We collected 173 LNs and 85 primary tissues from 78 CRC patients in BJ cohort (Fig. 1A and Supplementary Data 1). The principal component analysis (PCA) discovered LNs and tumor tissues exhibited strong intra-tissue similarities and inter-tissue heterogeneity in gene expression (Supplementary Fig. 1A). As expected, non-metastatic LNs and tumor tissues displayed the most significant gene expression differences. Unsupervised clustering analysis confirmed six distinct gene expression patterns among metastatic and non-metastatic LNs from BJ cohort (Fig. 1B), which suggested significant LNs heterogeneity. The subtypes with a majority of non-metastatic negative LNs (NLNs) were labeled as NLN_C1, NLN_C2, NLN_C3 and NLN_C4, and the subtypes with a majority of metastatic positive LNs (PLNs) were defined as PLN_C1 and PLN_C2 based on different gene expression profiles. Interestingly, while NLN and PLN subtypes showed general consistency with histopathological assessments, some pathologically PLNs were classified as NLN based on transcriptomic features. Distinct distance measures and agglomeration methods were performed in unsupervised clustering and obtained the same results (Supplementary Fig. 1B). To validate the LNs subtypes, the TY cohort including 227 LNs and the HB cohort including 162 LNs were used (Supplementary Data 2 and 3). Notably, four analogous NLN subtypes were also identified in two validation cohorts (Fig. 1C, D). Additionally, PCA revealed significant gene expression heterogeneity among the different LNs subtypes. Notably, a distinct structured clustering was observed across the four subtypes, from NLN_C1 to NLN_C4. This finding suggests that gene expression heterogeneity exists not only between metastatic and non-metastatic LNs but also within the subtypes of non-metastatic LNs. This pattern was consistently validated across three independent cohorts. Based on these findings, we established the LNs subtypes and

confirmed the robustness and stability of these subtypes across all three cohorts.

The molecular heterogeneity of LNs subtypes
Then, we compared the gene expression differences between different LNs subtypes, and found the number of up-regulated and down-regulated differentially expressed genes (DEGs) in NLN_C2, NLN_C3, and NLN_C4 gradually increased when compared to NLN_C1 (Fig. 2A). Together with the distinct structured clustering from NLN_C1 to NLN_C4 in PCA result (Fig. 1E–G), we preferred that the emergence of LNs subtypes reflected their distinct molecular heterogeneity. To validate this hypothesis, we first examined the expression patterns of DEGs across the four non-metastatic LNs subtypes (Fig. 2B, C). We observed consistent trends of gradual upregulation or downregulation in gene expression from NLN_C1 to NLN_C4. Supporting this, we performed an exploratory trajectory analysis, which confirmed a molecular heterogeneity among non-metastatic LNs, shifting from NLN_C1 to NLN_C4 (Supplementary Fig. 2A). To rule out technical confounders, we performed batch correction and confirmed that the clustering structure remained robust, thereby minimizing the likelihood of batch-related artifacts (Supplementary Fig. 2B). In addition, when the data were re-analyzed within established immune cell state frameworks[16], we found that certain cell states—such as tumor-associated fibroblasts, pro-migratory-like fibroblasts, and normal-like endothelial cells—differed across LNs subtypes and contributed to the separation in PCA space (Supplementary Fig. 2C). These findings support a significant molecular heterogeneity among LNs subtypes, driven by underlying changes in stromal and endothelial cell states, rather than a strict developmental or lineage progression.

Functional analysis of LNs subtypes
As expected, in metastatic LNs subtypes, the epithelial cell differentiation and extracellular matrix organization-associated functions were significantly activated (Supplementary Fig. 3A and Supplementary Data 5), and the immune response-associated biological processes were significantly inhibited (Supplementary Fig. 3B and Supplementary Data 6). Concerning the non-metastatic LNs subtypes, we found that the DEGs between NLN_C4 and NLN_C1 were more pronounced than those between metastatic and non-metastatic LNs (Fig. 2A, D, Supplementary Fig. 3C and Supplementary Data 7 and 8). In an attempt to further determine the salient and unique features of the NLN_C4, we found 3,713 up-regulated DEGs in NLN_C4 were significantly enriched in the extracellular matrix organization and the vascular-associated functions (Fig. 2E and Supplementary Data 9), and 2,365 down-regulated DEGs in NLN_C4 were involved in immune response regulation (Fig. 2F and Supplementary Data 10). Together, the above results suggested the decline of immune response ability together with the changes of stromal environment among non-metastatic LNs subtypes.

Cellular landscape of LNs subtypes
To further decipher the cellular components among LNs subtypes, the LN-associated cell type-specific molecular characters were defined using 463,054 high quality cells from a single-cell RNA sequencing (scRNA-seq) cohort of 68 LNs and 3 tumor tissues (Fig. 1A and Supplementary Data 11). A total of 14 major cell types including epithelial, fibroblast, endothelial and leukocyte lineages were identified (Supplementary Fig. 4A) using their established markers (Supplementary Fig. 4B). Depending on this cell type specific molecular features, pre-defined LNs subtypes associated DEGs (Fig. 3A and Supplementary Data 8) were mapped to different cell types. Consistent to the over-representation analysis (Fig. 2E, F and Supplementary Fig. 2B, C), the up-regulated genes in PLNs and NLN_C3/C4 were found to be enriched in the stromal lineage including epithelial cells, endothelial cells and fibroblast cells (Fig. 3A). The down-regulated genes were highly
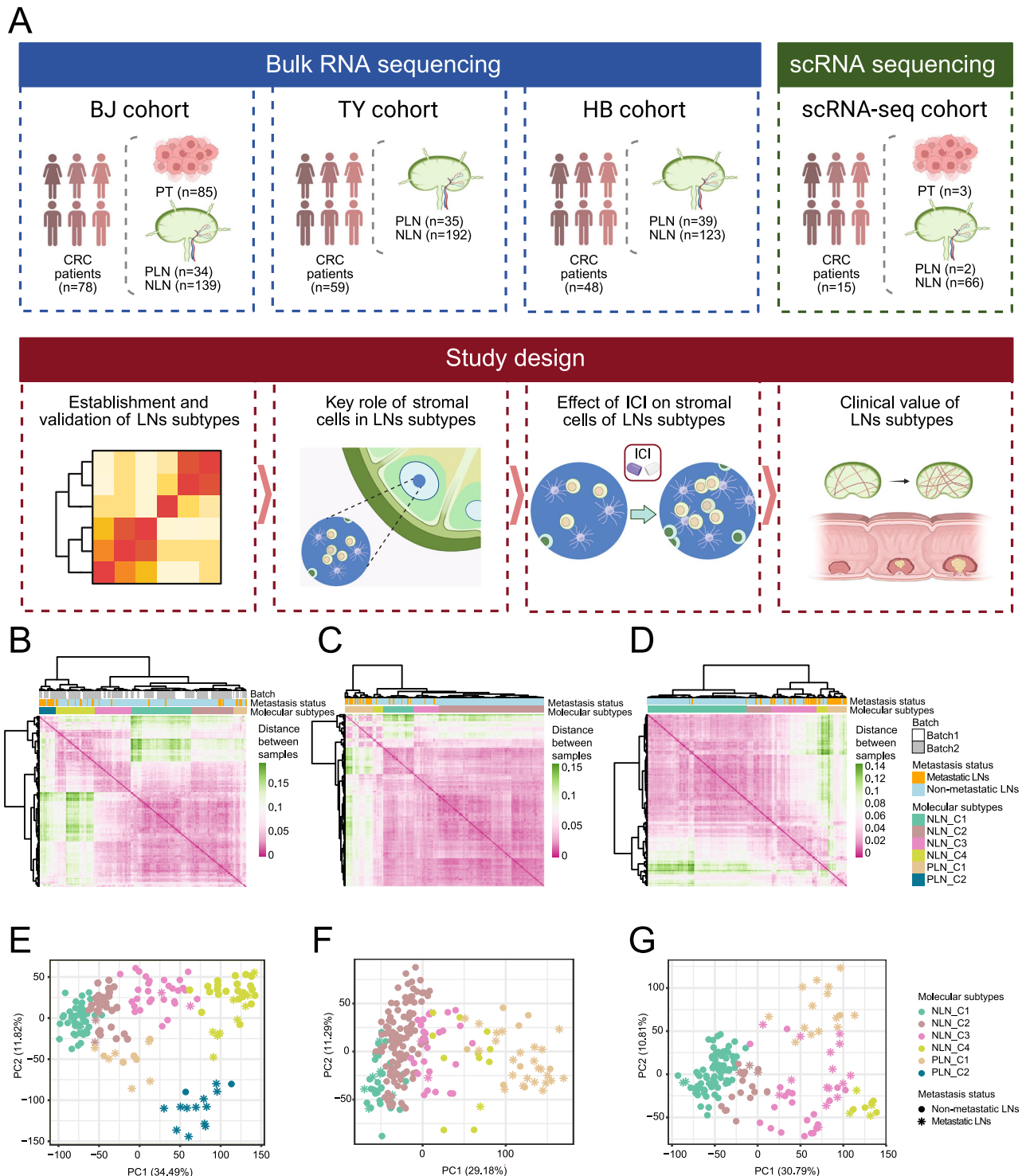
Fig. 1 | The establishment and validation of LNs subtypes. A The introduction of patient cohorts and study design. This panel was created in BioRender. •i, fK. (2025) https://BioRender.com/4tsnyv. Heatmaps showing unsupervised clustering result of bulk-RNA sequencing data from BJ cohort (*n* = 173) (**B**), TY cohort (*n* = 227) (**C**), and HB cohort (*n* = 162) (**D**). Dot plot of PCA quantifying the gene expression similarity between samples from BJ cohort (*n* = 173) (**E**), TY cohort (*n* = 227) (**F**), and HB cohort (*n* = 162) (**G**). CRC colorectal cancer, PT primary tumor, scRNA single-cell RNA, LNs lymph nodes, ICI immune checkpoint inhibitor, PC principal component, PCA principal component analysis.

expressed in the immune cells, especially in B cells and MKI67 progenitor cells (Fig. 3A).

To further estimate the cell type specific variation across the LNs subtypes, we deconvolved the bulk-profiled LNs subtypes to estimate their relative quantities of cell types by the CIBERSORTx[17]. The similar cell type compositions differences of immune cells and stromal cells

from NLN_C1 to NLN_C4 were observed in three cohorts (Fig. 3B and Supplementary Fig. 4C–E). Among which, CD4 T cells, CD8 T cells and B cells displayed the decreased cell types from NLN_C1 to NLN_C4, while fibroblasts and endothelial cells as the dominant cell types occupied the NLN_C3/C4 (Fig. 3C and Supplementary Fig. 4C–E). Meanwhile, the quantitative immunohistochemistry (IHC) analyses
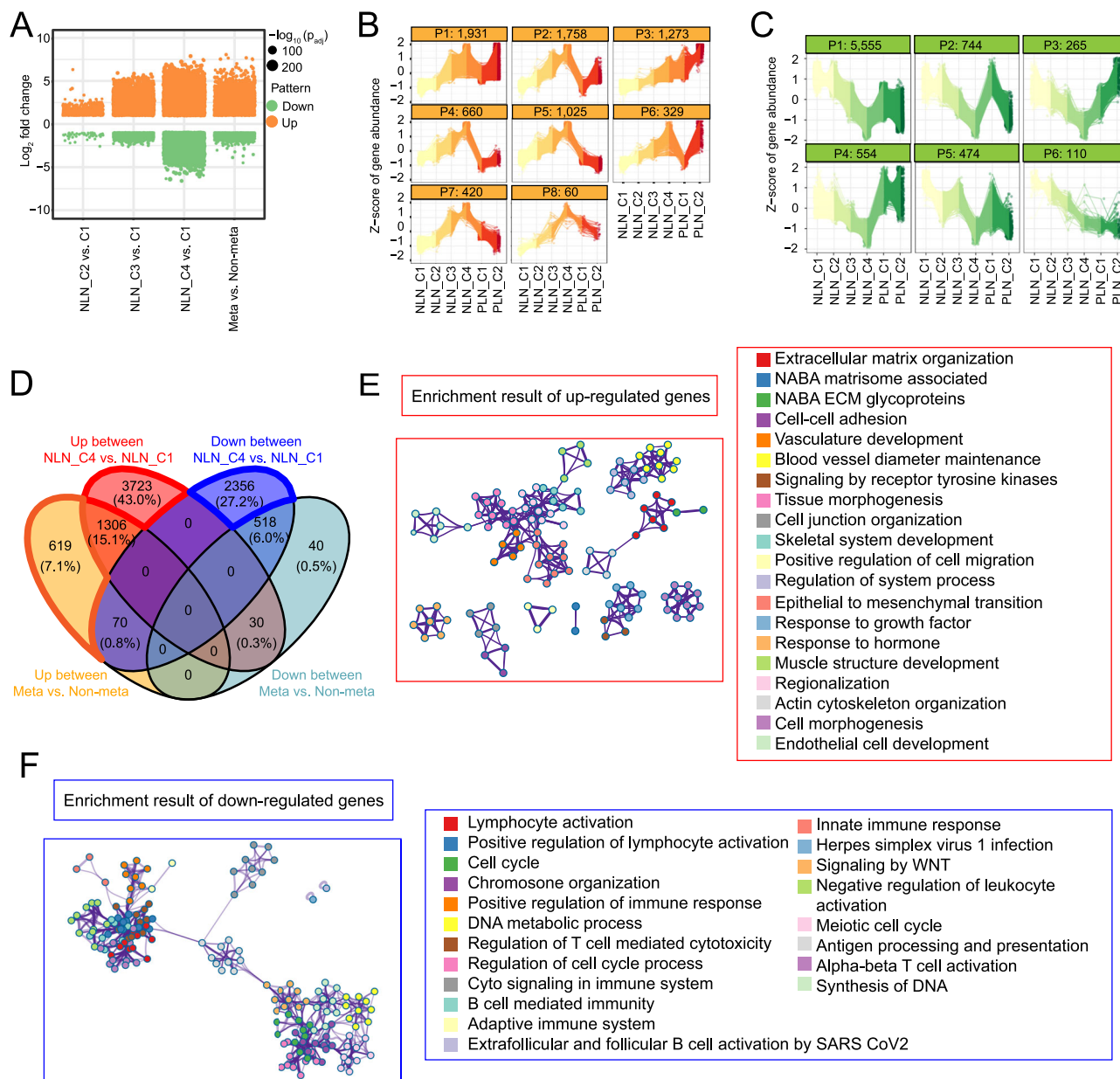
Fig. 2 | The molecular heterogeneity and functional analysis of the LNs subtypes. A Gene expression differences between LNs subtypes from BJ cohort ($n = 173$). Adjustment p value was obtained using DESeq2 which utilized Wald test. Gene expression patterns of target DEGs across LNs subtypes ($n = 173$): **B** up-regulated DEGs; **C** down-regulated DEGs. **D** Venn diagram of DEGs between different subgroups ($n = 173$). Over-representation enrichment analysis for the target DEGs based on the Metascape enrichment method ($n = 173$): **E** up-regulated DEGs between NLN_C4 and NLN_C1; **F** down-regulated DEGs between NLN_C4 and NLN_C1. LNs lymph nodes, DEGs differentially expressed genes, Meta metastatic, Non-meta Non-metastatic.

confirmed the transcriptomic differences between NLN_C1 and NLN_C4 across multiple LNs (Fig. 3D, E). Specifically, NLN_C1 displayed significantly higher densities of CD4 and CD8 T cells, as well as increased CD20 B cell infiltration, compared to NLN_C4. Collectively, these data confirmed the decline of immune cell types as well as the expansion of stromal cell types from NLN_C1 to NLN_C4.

Stromal cells are the key cells that determine LNs subtypes
Identifying the key cellular components that determine LNs subtypes is crucial for understanding the underlying mechanisms of LNs molecular heterogeneity. The scRNA-seq analysis supported our bulk-based classification by independently identifying NLN_C1 and NLN_C4 subtypes in non-metastatic LNs. Specifically, two non-metastatic LNs were classified as NLN_C4, while the remaining were assigned to NLN_C1. Cell type specific signatures were defined by scRNA-seq with a stricter

cutoff (Supplementary Fig. 5A and Supplementary Data 12) and then was used to distinguish the LNs subtypes by PCA. Consistent to the ability of the whole expressed gene list on LNs subtypes identification (Fig. 1E–G and Supplementary Fig. 5B), the integration of the whole cell type specific markers preserves the ability to reproduce the LNs subtypes (Supplementary Fig. 5C), which further confirmed that the emergence of LNs subtypes was decided via the cell type composition alteration.

Unexpectedly, the immune cells, such as NK/T cells, B cells, macrophages, neutrophils and mast cells-associated signatures, did not exhibit subtypes differentiation capabilities (Supplementary Fig. 5D). Interestingly, the highly variable gene sets of fibroblasts and endothelial cells possess strong abilities to distinguish LNs subtypes (Fig. 4A, B and Supplementary Fig. 5E). Specifically, there are notable differences in gene expression among the LNs subtypes for fibroblasts
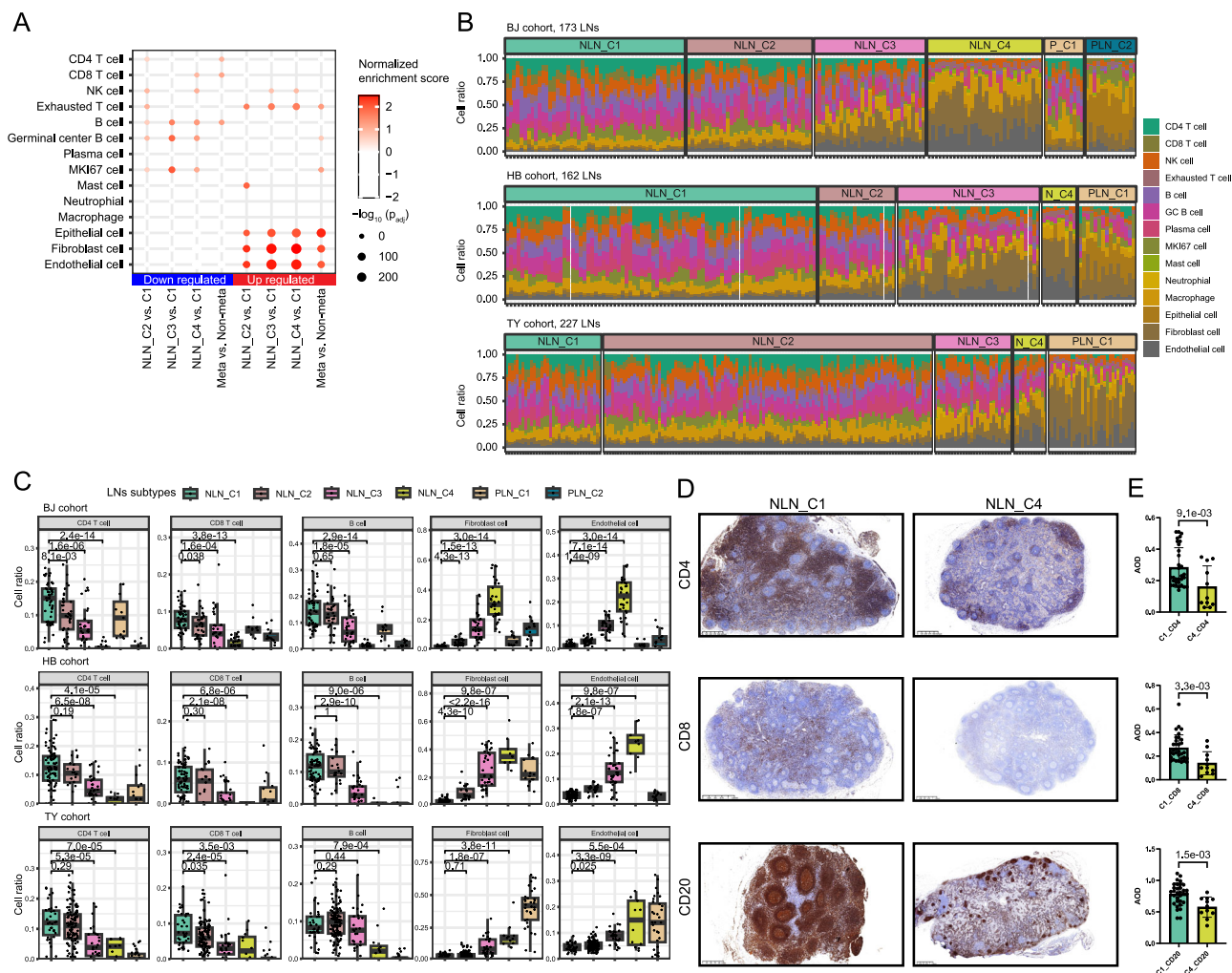
Fig. 3 | Cellular landscape of the LNs subtypes. A Dot plot showing cell types highly expressed DEGs by the expression of up-regulated and down-regulated DEGs across cell types ($n = 173$). Adjustment p value was obtained using GSEA with default setting. **B** Bar plots illustrating the differences in cell proportions among LNs subtypes from BJ cohort ($n = 173$), HB cohort ($n = 162$) and TY cohort ($n = 227$). **C** Box plots illustrating the differences in cell proportions among LNs subtypes, using LN scRNA-seq data as a reference based on CIBERSORTx, from the BJ cohort ($n = 50, 35, 31, 32, 11,$ and 14), HB cohort ($n = 81, 20, 37, 9,$ and 15), and TY cohort ($n = 35, 120, 28, 12,$ and 32). Each data point represents an individual LN (biological replicates). Box plots represent the median (center line), 25th and 75th percentiles (bounds of the box), and minimum and maximum values (whiskers). The p value was obtained using Wilcox test. The IHC staining (**D**) and quantification results (**E**) reveal differential expression of CD4, CD8 and CD20 between NLN_C1 ($n = 32$) and NLN_C4 ($n = 11$). Each data point represents an individual LN (biological replicates). Data are presented as mean values ± SD. The p value was calculated using a two-sided Student's t test. Scale bars indicate 625 μm and 1.25 mm respectively. Source data are provided as a Source Data file. Meta metastatic, Non-meta non-metastatic, LNs lymph nodes, DEGs differentially expressed genes, GSEA gene set enrichment analysis, IHC immunohistochemistry, SD standard deviation, AOD average optical density.

and endothelial cells. For example, the expression levels of fibroblast marker genes COL1A1, ACTA2 and FN1, endothelial cell marker gene PECAM1, and lymphatic endothelial cell marker gene LYVE1 increase progressively from NLN_C1 to NLN_C4 (Fig. 4C). We further performed multiplex immunofluorescence (mIF) staining on LNs samples from NLN_C1 and NLN_C4. The semi-quantitative results confirmed increased stromal cell proliferation in NLN_C4 compared to NLN_C1 (Fig. 4D, E). Furthermore, by using Sirius Red staining, we found that fibrosis, representing the terminal stage of stromal proliferation, was also more pronounced in NLN_C4 subtype (Fig. 4F). These findings collectively demonstrate that stromal cells play an important role in regulating LNs subtypes.

The regulation of immune cells by stromal cells across LNs subtypes

There was a significant reduction in the number of immune cells and their functional impairment from NLN_C1 to NLN_C4, which represents a key manifestation of LN immune dysfunction (Fig. 3C–E and Supplementary Fig. 4C–E). Given the importance of stromal cells in regulating LN immune function, we hypothesized that certain stromal cells may be pivotal in the changes observed in immune cells across different LNs subtypes. To investigate this, we performed dimensionality reduction clustering on stromal cells within LNs from scRNA-seq cohort, identifying six fibroblast subtypes: follicular dendritic cells (FDC), fibroblastic reticular cell (FRC), smooth muscle cell (SMC), Fib-ADH1B, Fib-IL7R and Fib-NTRK3, as well as five endothelial cells subtypes: vein blood endothelial cell (vBEC), artery blood endothelial cell (aBEC), capillaries blood endothelial cell (cBEC), lymphatic endothelial cell (LEC) and floor lymphatic endothelial cell (fLEC) according to their established markers[7,18,19] (Fig. 5A, B).

The LNs subtypes from scRNA-seq cohort were identified by cell type ratio similarity (Supplementary Fig. 6A). Two non-metastatic LNs were classified as NLN_C4, characterized by a higher proportion of endothelial and fibroblast cells and a lower proportion of B cells, while the remaining 64 non-metastatic LNs were grouped as NLN_C1 (Supplementary Fig. 6A, B). Then, we compared the proportion of stromal
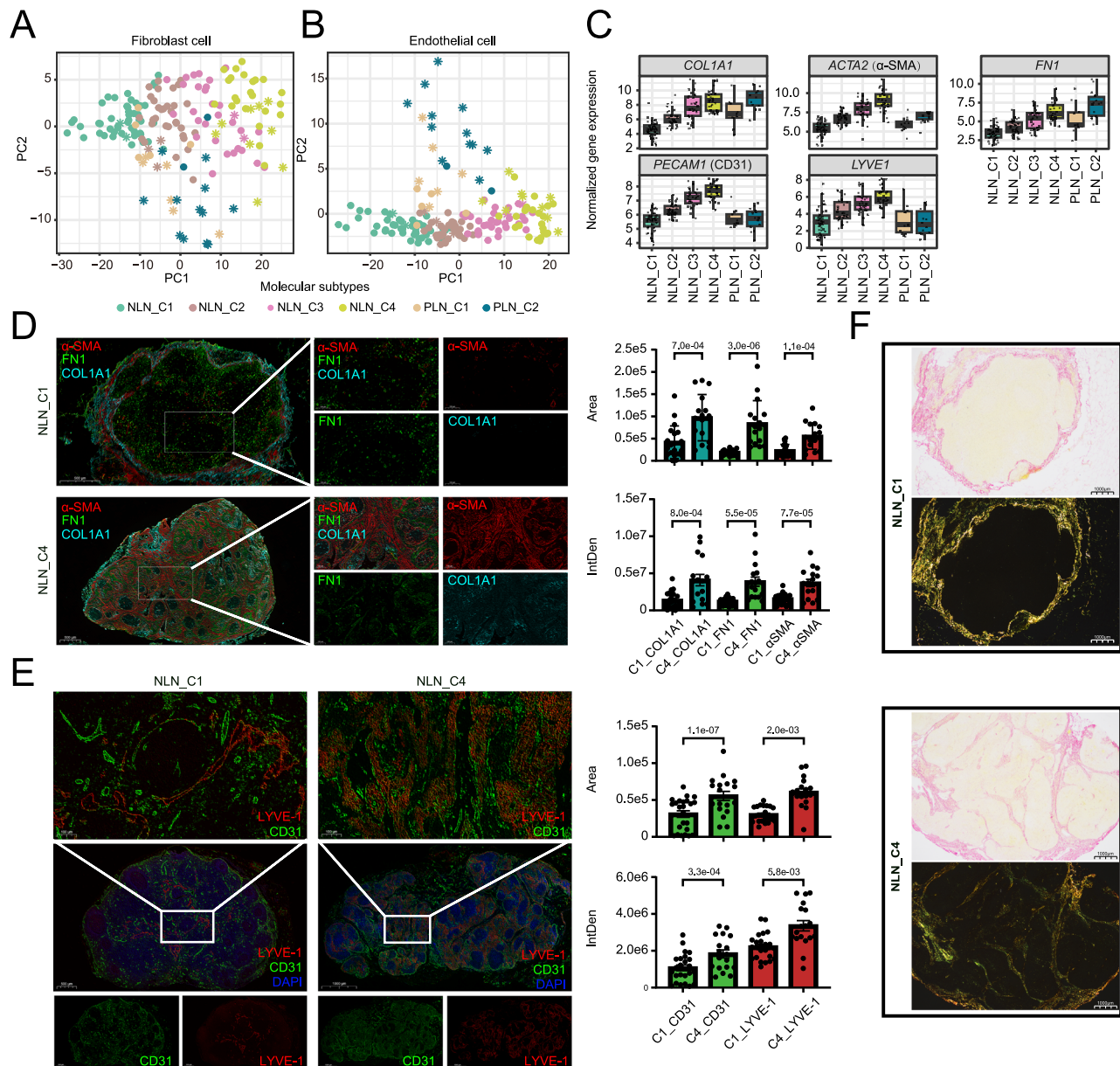
Fig. 4 | Identification of key cellular components of the LNs subtypes. Dot plot of PCA based on fibroblasts-specific (**A**) and endothelial cell-specific (**B**) gene markers to estimate their ability in distinguish the LNs subtypes ($n = 68$). **C** Box plots illustrating the differences in COL1A1, α-SMA, FN1, CD31, and LYVE1 gene expression between LNs subtypes ($n = 50, 35, 31, 32, 11,$ and 14). Each data point represents an individual LN (biological replicates). Box plots represent the median (center line), 25th and 75th percentiles (bounds of the box), and minimum and maximum values (whiskers). The mIF staining and quantification reveal differential cell subtypes across different LNs subtypes (Fig. 5C, D). Strikingly, the

expression of α-SMA, FN1, COL1A1 ($n = 20$ and 16) (**D**) and LYVE-1, CD31 ($n = 24$ and 18) (**E**) between NLN_C1 and NLN_C4. Each data point represents an individual LN (biological replicates). Data are presented as mean values ± SD. The p value was calculated using a two-sided Student's t test. Scale bars indicate 100 μm, 500 μm, and 1000 μm respectively. Source data are provided as a Source Data file. **F** Sirius Red staining results of differences in fibrosis levels between NLN_C1 and NLN_C4. Scale bars indicate 1000 μm. PC principal component, PCA principal component analysis, mIF multiplex immunofluorescence, SD standard deviation.

cell subtypes across different LNs subtypes (Fig. 5C, D). Strikingly, the proportions of FRC, FDC and aBEC were significantly decreased in NLN_C4. To further validate this trend, we compared the expression of FDC, FRC, and LEC-associated marker genes across LNs subtypes using bulk-RNA sequencing data (Fig. 5E and Supplementary Fig. 7A, B). Only the trends observed for FDC were corroborated at the bulk-RNA level and further supported by mIF staining (Fig. 5F, G and Supplementary Fig. 7C, D). Based on previous literature highlighting the critical regulatory role of FDC in B cell function[20,21], as well as the results of our study, we confirm that the reduction in FDC numbers is the primary cause of the B cell immune dysfunction from NLN_C1 to NLN_C4. Therefore, it is crucial to understand which factors regulate FDCs and, in turn, influence the immune status of LNs.

To identify the main factors responsible for the reduction in FDC, we conducted Single Cell Regulatory Network Inference and Clustering (SCENIC) analysis to identify transcription factors within each cell subtype (Fig. 5H). The ETS family transcription factor SPIB was found to have the highest transcriptional activity in FDC, regulating the development and function of FDC, particularly in B cell differentiation and immune responses to antigen stimulation. Additionally, the NF-κB family transcription factor RELB was active in FDC regulation, influencing B cell survival, maturation, and lymphoid organ development and maintenance. Notably, the expression of FDC-associated transcription factors was highly correlated with B cells (Fig. 5I), which suggested a strong link between FDC and B cells. Then the expression of the FDC-associated transcription factors was examined across the
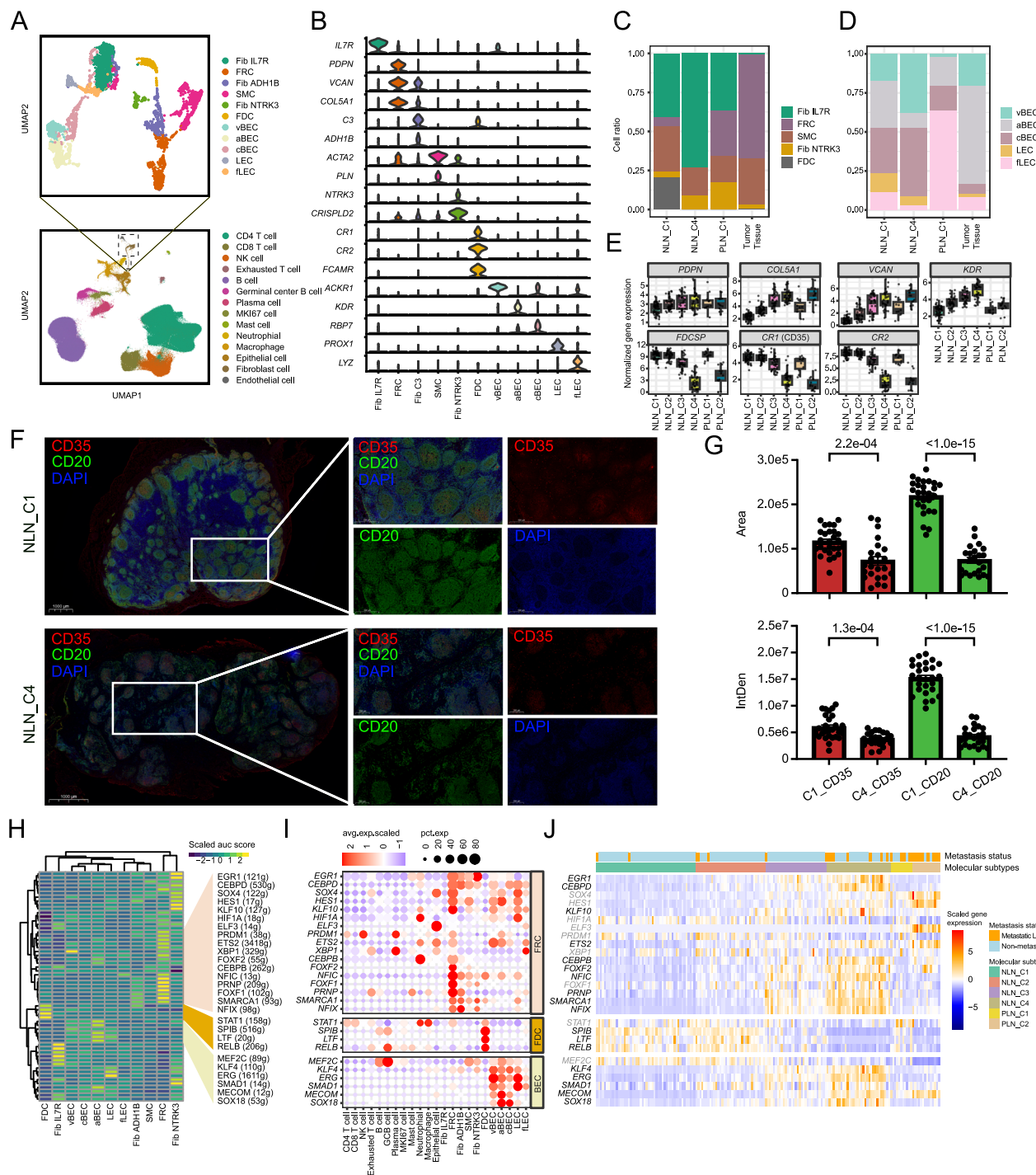
Fig. 5 | Dynamics of key stromal cell subsets across the LNs subtypes. A The UMAP analysis of transcriptional profiles of fibroblasts and endothelial cell (3436 cells) from the scRNA-seq cohort, colored by cell type (*n* = 68). **B** Violin plots showing marker gene expression for each fibroblast and endothelial cell subtype (*n* = 68). **C** Bar plot showing changes in each fibroblasts subtype proportions among different LNs subtypes (*n* = 68). **D** Bar plot showing changes in each endothelial cell subtype proportions among different LNs subtypes (*n* = 68). **E** Box plots illustrating the differences in PDPN, COL5A1, VCAN, KDR, FDCSP, CR1, and CR2 gene expression between LNs subtypes from BJ cohort (*n* = 50, 35, 31, 32, 11, and 14). Each data point represents an individual LN (biological replicates). Box plots represent the median (center line), 25th and 75th percentiles (bounds of the box), and minimum and maximum values (whiskers). The mIHC staining (**F**) and

quantification (**G**) reveal differential expression of CD20 and CD35 between NLN_C1 (*n* = 27) and NLN_C4 (*n* = 23). Each data point represents an individual LN (biological replicates). Data are presented as mean values ± SD. The *p* value was calculated using a two-sided Student's *t* test. Scale bars indicate 200 μm, 500 μm, and 1000 μm respectively. Source data are provided as a Source Data file. **H** The heatmap of fibroblasts and endothelial cell-specific gene regulatory networks based on the SCENIC analysis. **I** Dot plot showing the expression profiles of upstream regulatory transcription factors across different cell types. **J** Heatmap displaying the expression trends of upstream regulatory transcription factors across different LNs subtypes. UMAP Uniform manifold approximation and projection, mIF multiplex immunofluorescence, SD standard deviation, SCENIC Single Cell Regulatory Network Inference and Clustering.

LNs subtypes. As expected, FDC-associated transcription factors showed continuously down-regulated from NLN_C1 to NLN_C4 (Fig. 5J and Supplementary Fig. 8). Furthermore, the shift from immune cells to stromal cells in LNs could be influenced by accumulated cellular senescence, which is commonly associated with fibrosis[22]. Specific details can be found in the Supplementary materials (Supplementary Fig. 9A–D).

Effects of ICIs on stromal cell across LNs subtypes

The regulation of cellular components of LNs by ICIs is a critical factor influencing the therapeutic efficacy[8,9,23]. Additionally, given the important role of LNs stromal cells in immune regulation[7,18,24,25], it is also crucial to investigate the impact of ICIs on stromal cell within LNs. The uniform manifold approximation and projection (UMAP) plots based on ICI treatment status revealed the impact of ICI on the cellular composition of LNs (Supplementary Fig. 10A). We compared the proportions of fibroblast subtypes (Supplementary Fig. 10B) and endothelial cells (Supplementary Fig. 10C) between the ICI+ and ICI− groups. Notably, the proportion of FDC in LNs increased in the ICI+ group compared to the ICI− group. Additionally, the DEGs between ICI+ and ICI− showed significantly up-regulated in FDC (Supplementary Fig. 10D, E), including the cell type specific markers including CR2, FDCSP and FCAMR (Supplementary Fig. 10F) and the other types of fibroblast-associated markers including COL1A1 and PDPN which increased from NLN_C1 to NLN_C4 decreased after ICI treatment. Furthermore, we compared the expression of FDC-associated transcription factors between the ICI+ and ICI- groups, finding that RELB together with its target genes significantly up-regulated (Supplementary Fig. 10G, H). Pathological validation of these results was conducted by additional mIF staining (Supplementary Fig. 10I, J).

To address the potential confounding effects of mismatch repair (MMR) status on LN immune features, we performed stratified analyses in both the bulk-RNA sequencing and scRNA-seq cohorts. In the scRNA-seq cohort, a total of 15 CRC patients were included: 4 patients were proficient MMR (pMMR) and did not receive ICI therapy, while 11 patients were deficient MMR (dMMR), of whom 8 received ICI treatment. MMR status and ICI treatment information for each patient and corresponding LN sample are detailed in Supplementary Data 11.

In the three bulk-RNA cohort, we stratified patients by microsatellite instability (MSI)/microsatellite stability (MSS) status and compared the expression of FDC-related markers (CR1, CR2, FDCSP), B cell marker (MS4A1), and extracellular matrix marker (COL1A1). No significant differences were observed (Supplementary Fig. 11A). Similarly, no differences were found in B cell, fibroblast, or endothelial cell infiltration between MSI and MSS patients (Supplementary Fig. 11B). After excluding all pMMR patients, the same results were observed within the dMMR subgroup (Supplementary Fig. 11C). This suggests that the alterations in FDC and B cells are more likely to be a direct effect of ICI therapy, independent of MMR status. Further supporting this, mIF staining revealed no significant difference in CD20 or CD35 expression between untreated dMMR and pMMR patients (Supplementary Fig. 11D, E). However, a marked increase in both markers was observed in dMMR patients following ICI treatment (Supplementary Fig. 11D, E).

Together, ICI treatment leads to the expansion of FDC in LNs and promotes B cell development and function through the up-regulation of transcription factors. In other words, ICIs can exert antitumor effects by modulating LNs, providing a significant addition to the current understanding of ICI antitumor mechanisms.

Clinical significance of LNs subtypes

The distribution of LNs subtypes inter- and intra-patients was presented in Fig. 6A, B. The total number of resected LNs per patient, along with the number of metastatic and non-metastatic LNs were provided in Supplementary Data 1–3. In patients with multiple sampled LNs, each LN could belong to a distinct molecular subtype, thereby reflecting intra-patient LN heterogeneity. We further investigate the association between the clinical characteristics and LNs subtypes (Fig. 6A and Supplementary Fig. 12A–D). The unsupervised clustering of LNs subtypes indicated that NLN_C1 and NLN_C2 preferred from the same patient (Fig. 6A). Consistent to the transcriptomic similarity, we found that NLN_C3/C4 displayed a significantly higher co-detection ratio in the patients (Fig. 6B and Supplementary Fig. 12B, D). Accordingly, NLN_C1 and NLN_C2 were combined into one group for subsequent analyses, while NLN_C3 and NLN_C4 were classified as a separate group.

Furthermore, analysis across these three cohorts revealed that perineural invasion (PNI), TNM stage, and vascular thrombosis were more prevalent in the NLN_C3/C4 groups (Fig. 6C, D and Supplementary Fig. 12E, F).There is no significant difference in LNs subtypes between different sex (Supplementary Fig. 12G). These clinical features suggest that a higher proportion of NLN_C3/C4 is associated with increased tumor progression. However, no clear association was observed between LNs subtypes and MSI/MSS status (Supplementary Fig. 13A). Additionally, we did not identify a significant enrichment of specific LNs subtypes within any individual consensus molecular subtypes (CMS) classification (Supplementary Fig. 13B–E), suggesting that individual CMS subtypes could give rise to multiple distinct LN transcriptional phenotypes. These findings highlight the complexity of tumor–LN interactions.

We also examined LN-specific phenotypes, including LN size and LN location (Fig. 6E). Interestingly, the results show that LN location relative to the tumor and LN size did not significantly impact LNs subtypes classification across the three cohorts (Fig. 6E and Supplementary Fig. 14A). The PCA analyses stratified by LN size did not reveal clustering driven solely by size, suggesting that LNs subtypes capture transcriptomic differences beyond physical dimensions (Supplementary Fig. 14B). Immune and stromal cell infiltration patterns remained consistent across LN size groups, with a gradual decrease in immune infiltration and increase in stromal signatures from NLN_C1 to C4 (Supplementary Fig. 14C). These trends were more pronounced among small LNs, suggesting that transcriptional subtype may evolve independently of size but with differential biological consequences.

Most notably, the NLN_C1/C2 demonstrated significantly improved 3-year overall survival (OS) and recurrence-free survival (RFS) compared to NLN_C3/C4 (Fig. 6F and Supplementary Fig. 15A, B, all log-rank $p < 0.05$) Taken together, these findings indicate that LNs subtypes are strongly associated with tumor progression and patient outcomes, offering valuable insights for future clinical research.

Discussion

In this study, we initially established LyMoS in CRC by analyzing gene expression profiles from a comprehensive series of LNs. We then further characterized these subtypes to explore the cellular and molecular mechanisms underlying their heterogeneity, while also assessing the potential impact of ICIs on LNs subtypes. Finally, we examined the correlation between LNs subtypes and clinical features. Our findings offer a perspective on the molecular characteristics of LNs, providing valuable insights into their role in prognostic assessment and clinical decision-making for CRC patients.

Unlike the traditional classification of metastatic and non-metastatic LNs, LyMoS present significant heterogeneity among the four non-metastatic LNs subtypes, even surpassing the heterogeneity observed between metastatic and non-metastatic LNs. This finding challenges the conventional dichotomous framework and highlights the complexity of LN responses in the tumor microenvironment.

Specifically, our results show that NLN_C1/C2 retain a more immunologically active profile, while NLN_C3/C4 exhibit signs of immune dysfunction and stromal fibrosis. The heterogeneity between NLN_C1/C2 and NLN_C3/C4 appeared to reflect a distinct loss of
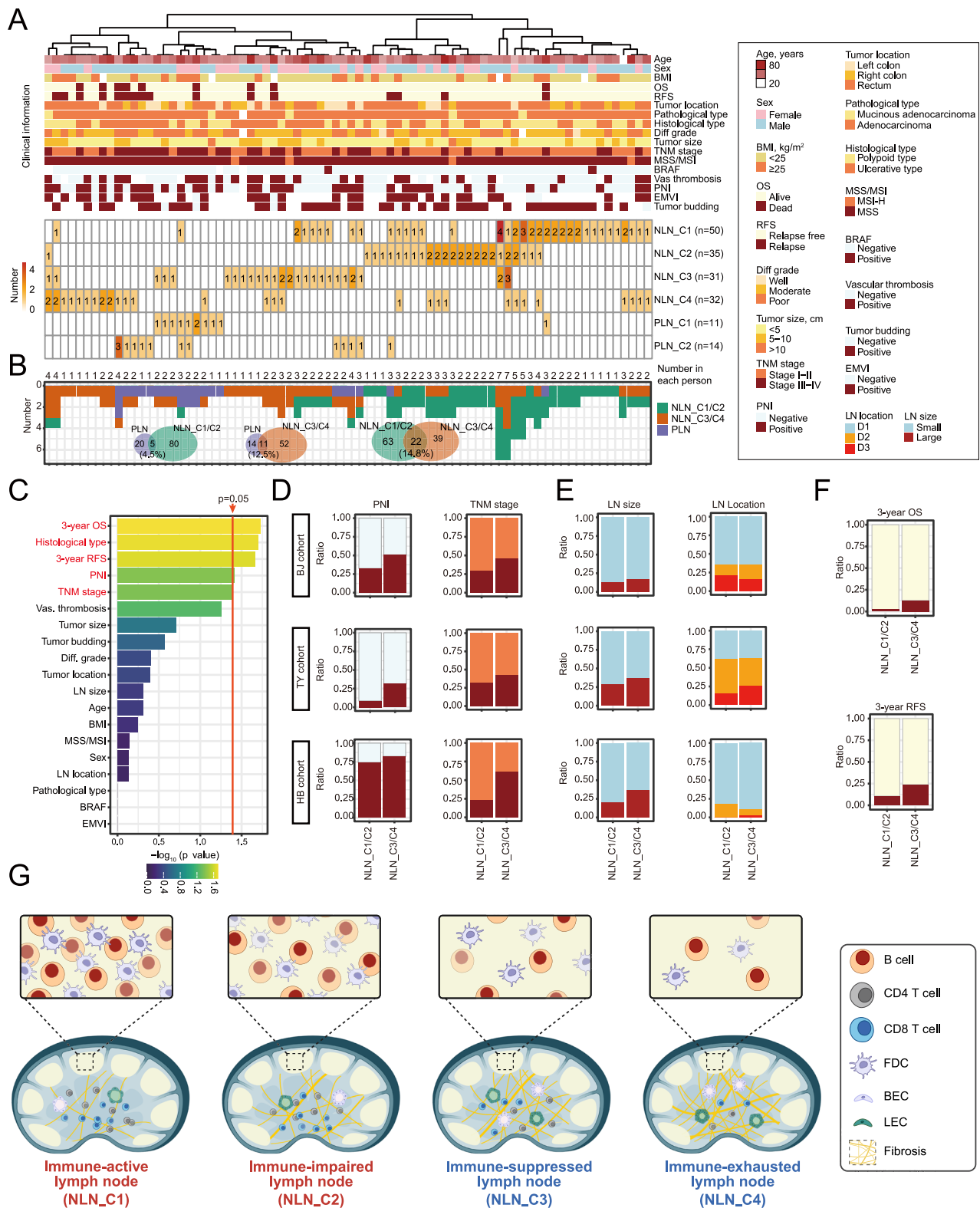
Fig. 6 | Clinical significance of the LNs molecular subtypes. A Heatmap revealing the LNs subtypes status for individual patient and its relationship with clinical pathological features from BJ cohort (*n* = 173). Each column represents an individual patient, and each row corresponds to clinical-pathological characteristics, the number of sampled LNs, and the distribution of the respective subtypes. **B** Bar plot revealing the LNs subtypes status for individual patient from BJ cohort (*n* = 173). Each column represents the total number of sampled LNs. The Venn diagram illustrating the overlap of different subtypes across individual patients. Bar plot revealing the statistical significance (*n* = 173) which is indicated by red text and red vertical lines (**C**) and the relationship between LNs subtypes and pathological characteristics (**D**), LNs size/location (**E**), and survival outcomes (**F**). The *p* value was calculated using fisher's exact test. **G** The schematic of LNs subtypes. This panel was created in BioRender. •í, fK. (2025) https://BioRender.com/va7fsdu. LNs lymph nodes, BMI body mass index, OS overall survival, RFS recurrence free survival, MSS microsatellite stability, MSI microsatellite instability, MSI-H microsatellite instability-high, PNI perineural invasion, EMVI extramural vascular invasion, Diff differentiation.

immune function, particularly in T cells and B cells, which are critical for the generation and maintenance of effective antitumor immunity (Fig. 6G). Furthermore, we carefully evaluated alternative explanations for the clustering patterns, including batch effects and cell state heterogeneity. Batch effects were ruled out through within-cohort correction and cross-cohort validation. However, cell state PCA using established frameworks[16] identified that certain cell states, such as tumor-associated fibroblasts, pro-migratory-like fibroblasts, and normal-like endothelial cells, exhibited distinct distributions across LNs subtypes. These results indicate that the observed clustering is at least partially driven by underlying cell state differences among LNs subtypes. While we cannot entirely exclude the influence of microenvironmental stimuli, the convergence of evidence of LNs subtypes suggests that non-metastatic LNs may undergo disruption of the stromal environment under tumor-derived stimuli before tumor cell invasion, leading to immune dysfunction.

The functional differences between subtypes were also confirmed at the cellular level. Immune cells, such as CD4 T cells, CD8 T cells and B cells, were down-regulated in NLN_C4, while stromal cells such as fibroblasts and endothelial cells significantly increased. Additionally, NLN_C4 exhibited prominent stromal fibrosis. This finding is consistent with previous studies highlighting the role of LN stromal cells in modulating immune responses[7,18]. LN stromal cells, including FRC, FDC, HEV and LEC, create a complex microenvironment within LNs that influences the behavior of immune cells[19]. For example, FRC produce chemokines that guide T cell migration, while FDC play a critical role in maintaining B cell follicles and supporting humoral immunity[21,26,27]. The loss of these critical stromal cell functions likely contributes to a diminished antitumor response, allowing for tumor progression and metastasis.

The mechanisms driving LN heterogeneity remain complex and multifaceted. It is commonly believed that the diminished immune function of LNs is predominantly regulated by tumor-induced immunosuppression. This concept forms the basis of most current studies, particularly those using animal models, which frequently conclude that the immune microenvironment of tumor-draining LNs is characterized by immunosuppression[11,28–31]. In humans, however, LN immune function exhibits substantial heterogeneity. This heterogeneity is evident not only between individuals but also among different LNs within the same patient. Strikingly, in the same tumor-draining region, both NLN_C1 and NLN_C4 can coexist. This observation underscores the significant heterogeneity between LNs, even within a single patient, which is likely driven by the complex anatomical structure of the lymphatic network and variations in the lymphatic drainage content among LNs. Therefore, the factors driving LN heterogeneity at the patient level are highly intricate and extend far beyond the scope of what can be captured through animal models. This complexity highlights the need for more nuanced studies that reflect the diverse immunological landscape of human LNs.

The ICIs implications of our findings remain profound. Following ICIs treatment, there was an up-regulation of the FDC development-related transcription factor SPIB, along with an increase in FDC and B cell numbers. Previous studies have demonstrated the critical role of tumor draining LNs in PD-1/PD-L1 checkpoint therapy[29]. Our results further suggest that ICI treatment could partially reverse the immunosuppression and stromal fibrosis in non-metastatic LNs, particularly by modulating FDC-association upstream transcription factors. Targeting the molecular pathways involved in LN stromal cell regulation may therefore enhance the effectiveness of ICIs, offering a potential strategy to improve outcomes for CRC patients. For example, therapies that promote the restoration of FDC and B cell function within LNs could potentially reverse the immune dysfunction. Additionally, our data suggest that combining ICIs with agents that target stromal cell proliferation and extracellular matrix (ECM) remodeling could disrupt the tumor-supportive microenvironment within the LNs, thereby enhancing the antitumor response. Further studies are necessary to elucidate the molecular mechanisms underlying LNs subtypes heterogeneity and identifying additional biomarkers for predicting patient response to ICI therapy.

This study has several limitations. First, despite the inclusion of diverse patient cohorts from multiple centers, LN heterogeneity remains a major challenge, driven by patient-specific factors, tumor microenvironmental differences and anatomical variations, which may influence the characteristics and clinical significance of LyMoS. Second, although transcriptomic profiling using bulk-RNA sequencing and scRNA-seq enabled the identification of distinct LNs subtypes, the lack of a simple biomarker panel for practical classification has hampered its clinical translation. Third, the current classification is based primarily on gene expression, which may not fully reflect the biological complexity of LNs subtypes. Integration of additional omics layers, such as proteomics or metabolomics, may offer deeper mechanistic insights. Fourth, the application of Monocle to bulk-RNA sequencing data has inherent limitations, including the lack of single-cell resolution and potential confounding due to mixed cell populations. Nevertheless, previous studies have demonstrated the utility of trajectory inference in bulk datasets to uncover meaningful biological gradients[32–34]. Importantly, the LN is a highly heterogeneous microenvironment comprising immune, stromal and endothelial components, and even single-cell approaches may not fully capture its functional diversity. Bulk transcriptomic profiling thus provides a complementary, integrated view of LN biology. Moreover, as LNs are critical immune organs predominantly populated by immune cells, the limited total number of cells captured in scRNA-seq may result in the sparse detection of stromal cell subsets, thereby potentially leading to their underrepresentation in the dataset. It is plausible that LNs subtypes, such as NLN_C2/C3, were not adequately captured in the single-cell dataset due to their reliance on stromal cell signatures. Therefore, the Monocle-based analysis was not intended to infer actual cellular transitions, but rather to reconstruct transcriptional continua among LN samples with shared molecular features. Finally, while the study provides molecular and pathological support for the existence of LyMoS, clinical validation in larger, real-world cohorts is needed, and the prognostic and therapeutic implications warrant further investigation in prospective trials.

In summary, we introduce the LyMoS, a framework that categorizes LNs into distinct molecular subtypes based on gene expression profiles in CRC. We also delineate the underlying biological mechanism of non-metastatic LNs, characterizing their heterogeneity and its relationship with clinical features. As our understanding of LN biology deepens, LyMoS has the potential to play a pivotal role in advancing personalized therapeutic strategies, ultimately enhancing outcomes for CRC patients.

Methods
Ethical statements
The clinicopathological data collection for this study was conducted according to the principles of the Declaration of Helsinki. All tissue samples, including primary tumor and LNs samples, were obtained in accordance with national guidelines. These tissues were collected from surgical specimens by surgeons and pathologists. All LNs were processed using a standardized protocol designed to preserve both research and clinical diagnostic integrity. Fresh LN specimens were bisected after surgical resection. A measured 1/3 to 1/2 portion was allocated for research analysis. The remaining 1/2 to 2/3 portion was used to generate formalin-fixed and paraffin-embedded (FFPE) samples for routine pathological evaluation. All patients provided signed informed consent and agreed to be reported and shared individual-level data. The study was evaluated and approved by the institutional review board (IRB) of the National Cancer Center/Cancer Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College

(Approval ID: 24/571–4851), Shanxi Hospital Affiliated to Cancer Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College (Approval ID: KY2023166) and The Second Affiliated Hospital of Harbin Medical University (Approval ID: YJSKY2024-269). This study is fully compliant with the "Guidance of the Ministry of Science and Technology (MOST) for the Review and Approval of Human Genetic Resources", and have obtained all necessary approvals from MOST of China for the export and use of genetic and sequencing data related to this work (approval ID: 2025BAT00719).

Patient cohorts

The LNs subtypes were identified on an independent cohort of 78 CRC patients in the National Cancer Center (BJ cohort), and then were validated from two independent cohorts from the Shanxi Hospital Affiliated to Cancer Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College (TY cohort) and the Second Affiliated Hospital of Harbin Medical University (HB cohort) (Fig. 1A). All these tissues were utilized for comprehensive bulk-RNA sequencing analysis. All included patients who underwent surgery did not receive pre-operative chemotherapy and/or radiation therapy.

To further interpret the characteristics of the LyMoS, another scRNA-seq cohort of 15 CRC patients was developed, including a total of 66 non-metastatic LNs, two metastatic LNs and three primary tissues (Fig. 1A). Among these CRC patients, seven received no treatment prior to surgery (ICI- group), eight received treatment with PD-1 inhibitors either as monotherapy or in combination with CTLA-4 inhibitors (ICI+ group), all of whom achieved a pathological response. The overall study design is shown in Fig. 1A.

Bulk-RNA sequencing samples preparation and processing

We extracted RNA from Fresh-frozen specimens using TRIzol reagent. The Bioanalyzer (Agilent Technologies) was employed to check the quality of the total RNA. The libraries were generated using Epicentre's Ribo-Zero rRNA Removal kit to remove rRNA from total RNA. The Illumina TruSeq™ RNA Sample Prep Kit (Illumina, San Diego, CA, USA) was used to construct the library, utilizing 30–100 ng of Ribo-Zero RNA. For quality analysis prior to sequencing, the generated cDNA libraries were quantified using a Qubit dsDNA HS Assay Kit and a Qubit 3.0 fluorometer (both from Thermo Fisher Scientific, Wilmington, DE, USA). The RNA libraries were subjected to sequencing as 2×150 bp paired-end reads on Illumina HiSeq2500 sequencers. Reads obtained from the sequencing were aligned using STAR (version 2.5.2b)[35], and aligned against the human reference genome (Ensemble GRCh38). Transcripts per kilobase million (TPM) were computed and adopted as the gene expression levels throughout the study.

Definition of LNs subtypes and metastasis status

Unsupervised clustering analysis was performed using transcriptomic profiles from the BJ cohort to define LNs subtypes, with the aim of capturing the intrinsic molecular heterogeneity of LNs irrespective of their pathological metastasis status. The DEGs distinguishing these subtypes were identified within BJ cohort using pairwise comparisons between clusters. To independently validate this classification framework, unsupervised clustering was also conducted in the TY and HB cohorts. Subtype assignment in these external datasets was performed independently, without reference to the BJ cohort, to assess the reproducibility of the LN classification. Additionally, we compared gene expression profiles and cellular compositions of the resulting subtypes across the three cohorts to evaluate cross-cohort consistency.

Accordingly, LNs were classified into transcriptionally distinct subtypes—namely, NLN and PLN—based on gene expression patterns rather than histopathological findings. The subtypes with a majority of non-metastatic negative LNs were labeled as NLN_C1, NLN_C2, NLN_C3 and NLN_C4, and the subtypes with a majority of metastatic positive LNs were defined as PLN_C1 and PLN_C2 based on different gene

expression profiles. Each LN was independently classified into a molecular subtype based solely on its gene expression profile, without aggregating to patient-level categories.

The metastasis status of each LN was determined separately by conventional pathological examination. A notable observation in our study was the partial discordance between transcriptional LNs subtypes and pathological metastasis status. This discrepancy may stem from two plausible factors. First, certain metastatic LNs may preserve an immune microenvironment and cellular architecture that remain largely intact despite the presence of tumor cells, leading to gene expression patterns that closely resemble those of non-metastatic LNs. Second, technical limitations, particularly sampling bias, may have contributed to misclassification. In cases of micro-metastases (i.e., tumor foci <2 mm), it is possible that the tissue section used for RNA sequencing did not include the metastatic area, whereas the adjacent section used for histopathological evaluation did capture tumor cells. These findings underscore the importance of integrating spatial and single-cell resolution approaches in future studies to better resolve the heterogeneity of LN at both cellular and spatial levels.

Identification of DEGs and pathway analysis

The raw gene counts were determined and normalized via HTSeq (version 2.0.5)[36]. We set a corrected adjust p value < 0.01 and |log2-fold change| >1 as the significance criteria to identify the DEGs by the "DESeq2" package (version 1.36.0). PLN-associated DEGs were defined by comparing the gene expression between metastatic LNs and all the other non-metastatic LNs regardless of their subtypes. The LNs subtypes-associated DEGs were defined by comparing the NLN_C2/C3/C4 to NLN_C1. The "degpattern" package was used to identify the expression patterns of target DEGs across LNs subtypes. Over-representation enrichment analysis for the top3000 protein coding target genes was performed on Metascape which is a comprehensive and biologist-oriented resource with default settings. The 20 most significantly enriched terms were visualized and analyzed as a network.

To assess potential batch effects, PCA was conducted on all samples stratified by sequencing batches. Batch-specific clustering was examined visually and quantitatively to determine whether batch artifacts influenced sample distribution. The lack of batch-driven segregation in PCA plots suggested that batch effects were negligible.

To evaluate whether the observed transcriptomic heterogeneity reflected intrinsic cell state variations, we reanalyzed the subtype clusters within the framework of defined immune cell states as described in previous studies[16]. This was done by mapping each cluster's transcriptomic profile to annotated immune cell state signatures and assessing the distribution in PCA space.

Cluster trajectory analysis

Due to the transition process from non-metastatic LNs to metastatic LNs under tumor stimulation, we exploratively applied trajectory analysis to investigate the LNs subtypes using Monocle 2 (2.10.1)[37,38]. Although Monocle was originally developed for scRNA-seq data, we applied it to bulk-RNA sequencing profiles by treating each LN sample as an independent unit, consistent with previously published methodologies that applied Monocle to bulk tumor datasets[32–34]. Nevertheless, as Monocle's trajectory inference relies on cell-to-cell variability—which is absent in bulk RNA-seq data—the inferred trajectories may be driven by inter-sample variability. Therefore, this approach was used in an exploratory manner and cannot confirm a progressive or sequential nature of the LNs subtypes. In our study, pseudotime analysis was performed on the set of DEGs distinguishing LNs subtypes. Principal component-based dimensionality reduction followed by pseudotime ordering allowed us to reconstruct a continuum of transcriptional states across LN samples.

Based on bulk-RNA sequencing data from discovery cohort, we treated each LN sample as "an independent cell" to align with the

Monocle algorithm framework. Genes for trajectory inference were selected using the dispersionTable() function to calculate a smooth function describing how variance in each gene's expression across clusters varies according to the mean. Only genes with mean expression greater than or equal to 0.1 were used for the analysis. The reduceDimension() function was utilized with the DDRTree[37] reduction method and the following parameters modified: max_components = 2, and num_dim = 20. Results were visualized using the plot_cell_trajectory() and plot_complex_cell_trajectory() functions and annotated with LNs subtypes labels and pseudotime.

scRNA-seq samples preparation and processing

These tissues were cut into approximately 1 mm³ pieces in RPMI-1640 medium (Invitrogen) with 10% fetal bovine serum (FBS; ScienCell) and enzymatically digested for 30 min on a rotor at 37 °C using a MACS tumor dissociation kit (Miltenyi Biotec), according to the manufacturer's instructions. The suspended cells were centrifuged at $400 \times g$ for 5 min after filtration with a 70 m Cell-Strainer (BD) in RPMI-1640 media (Invitrogen). The pelleted cells were suspended in red blood cell lysis buffer (Solarbio) and incubated on ice for 2 min to lyse red blood cells after the supernatant was removed. After washing twice with PBS (Invitrogen), the cell pellets were resuspended in sorting buffer (PBS mixed with 2% FBS). The single cell suspensions were stained for flow cytometry (FACS) on a BD Aria III equipment with 7-AAD Viability Staining Solution (Cat# 00-6993-50, eBioscience). Live cells were sorted into 1.5 ml tubes with sorting buffer and manually counted under the microscope based on FACS analysis. The GemCode Single Cell Platform was then used to process single cells using the 3' GemCode Gel Bead, Chip, and Library Kits (10 x Genomics) according to the manufacturer's methodology. For each sample, the loaded cell numbers were 10,000. The cells were then partitioned into Emulsion Gel Beads in the GemCode instrument, where they were lysed and barcodes were ligated by reverse transcription; the RNA was then amplified and sheared, and 3' adapters and sample indexes were ligated. The libraries were sequenced on an Illumina HiSeq 4000 using a 150 bp paired-end approach.

The libraries for scRNA-seq were prepared using the Chromium Single Cell 3′ v3 according to the manufacturer's protocol (10x Genomics). A total of 7000 cells were targeted per sample. Libraries were sequenced on the NextSeq 500 platform (Illumina) with paired-end sequencing.

scRNA-seq and cell type determination

The raw fast-q files were mapped to the reference genome GRCh38 using Cell Ranger Single Cell (version 4.0.0) software with the default parameters. The filtered results were used as input for the second round of quality control and downstream analysis in Seurat (version 4.3.0)[39]. To exclude potential cell debris and doublets from the dataset, we filtered out cells with nFeature_RNA ≤ 100 and ≥6000 and DoubletFinder (2.0.2). Potential dead cells with high expression of mitochondrial genes (≥25%) were also eliminated from the following analysis. To reduce the potential effect on the cell clustering results, we removed genes expressed in fewer than 3 cells, as well as mitochondrial genes and ribosome-associated genes. The filtered genes were projected into principal component space using principal component analysis (RunPCA). The top 50 significant dimensions identified by ElbowPlot were subjected to graph-based clustering.

For validation of the LNs subtypes defined by bulk-RNA sequencing, the scRNA-seq cohort were analyzed. After characterizing each subtype's cellular composition and clustering features at the bulk-RNA level, LNs subtypes labels (C1–C4) in the scRNA-seq cohort were inferred based on the similarity of cell type proportions and the degree of clustering proximity to metastatic LNs in unsupervised analyses.

Uniform manifold approximation and projection (UMAP) was utilized in the dimensionality reduction of cells for visualization, and a shared nearest neighbor-based clustering algorithm was used to find the clusters. FindAllMarkers was used to identify the cell type specific genes with the parameters logfc.threshold = 0.25, only.pos = TRUE, and min.pct = 0.25. By integrating the cell type-specific markers identified in previous studies[40,41], validated cell markers curated in Cell-Marker and cell type-specific genes, we assigned a cell type identity to each cell cluster.

Cell type-specific variations between LNs subtypes

To estimate the cell type expression of positive- and LNs subtypes-associated DEGs, gene expression in each cell type was estimated using fast Wilcox functions in presto package and was set as pre-ranked gene set. And then fgsea package was used to map the DEGs to this pre-ranked gene set to obtain the cell type specific expression of DEGs. To estimate the cell type specific variation between LNs subtypes, we firstly down sampling the largest cell types including T cells, B cells, NK cells in LNs to 5000 cells to construct a down sampling dataset. The FindMarker function was used to examine the cell type specific markers in primary tumors and LNs with min.pct = 0.25, logfc.threshold = 0.25, and test.use = "wilcox". The output genes were defined as cell type specific markers. And their expression across the cells were set as input to CIBERSORTx to generate the signature matrix. The raw count matrix from bulk-RNA sequencing cohorts was then deconvoluted with default settings. The S-mode was selected to remove any potential batch effects.

Key cells identification in the LNs subtypes

Based on each cell type specific gene sets, we performed dimensionality reduction using PCA on the bulk-RNA sequencing cohorts. To obtain the high confident cell type specific signatures, cells marker with fold change greater than 1 were retained in the PCA analysis. The ability of cell type-specific signatures in reproducing the LNs subtypes were used to estimate the cell types in determination of the LNs subtypes.

Based on the clustering results from PCA, we found that fibroblasts and endothelial cells can effectively distinguish different clusters of LNs subtypes. Therefore, we further re-clustered fibroblasts and endothelial cells using the dimensionality reduction and clustering methods mentioned earlier. Fibroblasts were classified into six subgroups, among which, three subtypes were annotated as FRC, SMC, FDC by highly expressed PDPN, ACTA2, and CR2, respectively. Another three clusters were not annotated to any previous reported subtypes and were labeled as Fib IL7R, Fib ADH1B, and Fib NTRK3 according to their highly expressed genes. Endothelial cells were similarly classified into five subgroups: aBEC, vBEC, cBEC, LEC, and fLEC based on the expression of ACKR1, KDR, RBP7, PROX1, and LYZ. The subtypes-associated cell proportion variation and gene variation across LNs subtypes were quantify to estimate the influence of fibroblast and endothelial cells-associated subtypes on the emergence of LNs subtypes.

Transcription factors analysis

To identify cell type specific gene regulatory networks, SCENIC (version 1.1.2)[42] was performed on fibroblasts and endothelial cells dataset. SCENIC is a combination of 3 packages (GENIE3, RcisTarget and AUCell). First, we filtered for genes expressed in 3% of samples and cells that are expressing >0 UMI and normalized by log2(filteredexpr + 1). Filtering criteria was less stringent to prevent exclusion of rare cell types. Genes that are available in RcisTarget's human feather databases (hg19-500bp-upstream-7species.mc9nr.feather, hg19-tss-centered-10kb-7species.mc9nr.feather) are utilized. These genes are found in a region of 500 to 10,000 base pairs upstream in the hg19 human reference genome. Only the transcription factors highly expressed in any of fibroblast and endothelial cell subtypes were

retained and further removed if their activate score smaller than 0.01 in any subtypes.

ICI treatment correlative analysis

We divided the LNs from scRNA-seq cohort into ICI− and ICI+ groups. The DEGs between ICI+ and ICI− were obtained using Findmarkers using MAST method and set the batches as latent variable. Considering the limited cells within each subtype of fibroblast cells and endothelial cells, the target cell types used in FindMarkers was set as fibroblast cells and endothelial cells rather than subtypes. And then the DEGs were mapped to subtypes based on their expression level in each subtype. Furthermore, we compared the changes in FDC marker genes and transcriptional regulators, which were significantly downregulated from NLN_C1 to NLN_C4, between the ICI+ and ICI− groups.

Pathological validation

All LNs tissues used for pathological validation were extracted from FFPE samples of clinical pathological evaluation. These sections underwent processing steps, including deparaffinization, rehydration and endogenous peroxidase inactivation, followed by antigen retrieval using sodium citrate buffer (10 mM, pH 6.0) with microwave treatment at 95 °C for 20 min. All pathological evaluations were conducted by two independent pathologists. In cases of disagreement, a final decision was made by a pathology expert with over 10 years of experience.

Sirius Red staining was used to assess the collagen fiber content, distribution, and degree of fibrosis in the LNs. For IHC, CD4 antibodies were purchased from ZSGB-Bio (dilution 1:800, ZA0519, Beijing, China), CD8 antibodies were purchased from ZSGB-Bio (dilution 1:800, ZA0508, Beijing, China) and CD20 antibodies from Cell Signaling Technology (dilution 1:800, #92688, Massachusetts, USA). Positive controls, human tonsil tissues, were stained with these antibodies, while isotype-matched antibodies were used as negative controls. The IHC staining procedure was automated using the BenchMark ULTRA immunostainer from Ventana Medical Systems, Inc. The CD4, CD 8 and CD20 expression in LNs was evaluated. All NLN_C1 and NLN_C4 samples were used for semi-quantitative analysis, and the ImageJ software was utilized to analyses average optical density (AOD) values of stained areas.

For multiplex IHC (mIHC), fresh samples were fixed in cold acetone for 5 min, washed with PBS for 5 min, and blocked with 3% BSA solution. Then, the samples were stained with primary antibodies overnight at 4 °C and with secondary antibodies for 30 min at room temperature. DAPI (VECTASHIELD, Vector Laboratories) was used to stain the nuclei. The primary antibodies included rabbit polyclonal antibody CXCL13/BCA1 (Proteintech, 10927-1-AP); rabbit polyclonal antibody LYVE1 (Proteintech, 28321-1-AP); mouse monoclonal antibody CD20 (Proteintech, 60271-1-Ig); mouse monoclonal antibody FN1 (Proteintech, 66042-1-Ig); mouse monoclonal antibody CD31 (Proteintech, 66065-2-Ig); mouse monoclonal antibody COL1A1 (Proteintech, 67288-1-Ig); mouse monoclonal antibody PDPN (Proteintech, 67432-1-Ig); mouse monoclonal antibody CD35 (Proteintech, 68033-1-Ig); rabbit recombinant antibody α-SMA (Proteintech, 80008-1-RR); All images were captured using an EVOS FL Auto 2 Imaging System (Thermo Fisher Scientific). DAPI was used to stain the cell nuclei. For each sample, 5–6 random fields were used for semi-quantitative analysis, and ImageJ was used to measure the positively stained area and to quantify cells.

The correlation between LNs subtypes and clinicopathological characteristics

We reviewed and collected clinicopathological information of CRC patients from the bulk-RNA sequencing cohorts (BJ, TY, and HB cohorts) using the medical record system. This information includes demographic variables, such as gender, age, body mass index (BMI), and clinicopathological variables, such as American Joint Committee on Cancer (AJCC) TNM stage, tumor histology, tumor size, tumor location, tumor differentiation, and MSI/MSS status.

The CMS and iCMS subtypes were assigned based on the primary tumor tissues corresponding to the involved LNs, using the CMScaller algorithms (https://github.com/peterawe/CMScaller). We then assessed the correlation between primary tumor CMS/iCMS and LNs subtypes.

LN size was dichotomized using a 1 cm cutoff, based on preoperative imaging and pathology records. LNs were stratified into "small" (<1 cm) and "large" (≥1 cm) groups. Subtype frequency distribution was compared across size categories. PCA and cell-type deconvolution were performed within each size group to assess whether transcriptomic variation was primarily driven by size or underlying LNs subtypes.

Additionally, since BJ cohort had an adequate median follow-up time (more than 3 years), survival outcomes including OS and RFS were collected. Furthermore, the CRC patients were stratified based on clinicopathological characteristics, and then were used to compare the changes in LNs subtypes proportions between groups.

Statistical analysis

Categorical variables are presented as numbers with percentages, while quantitative variables are presented as medians with interquartile ranges (IQRs), unless indicated otherwise. The Student's t test and Mann–Whitney U test were used for continuous data. Fisher's exact test or Chi-square test were used for categorical data. All the statistical tests were two-sided, and $p < 0.05$ was considered to indicate statistical significance. A significance level of $p < 0.05$ was assumed for all the statistical evaluations. All the statistical analyses and data visualizations were performed using R 4.2.1 software.

To assess the prognostic relevance of LN molecular subtypes, we performed Kaplan–Meier survival analysis. Each LN was assigned OS and RFS outcomes based on the prognosis of its corresponding patient, and survival differences were evaluated using the log-rank test.

Reporting summary

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

Data availability

The raw sequencing data reported in this paper have been deposited in the Genome Sequence Archive[43] in National Genomics Data Center[44], China National Center for Bioinformation/Beijing Institute of Genomics, Chinese Academy of Sciences (GSA-Human: HRA010982) that are publicly accessible at https://ngdc.cncb.ac.cn/gsa-human/browse/HRA010982. The data generated and/or analyzed during the current study are available under restricted access in accordance with the "Guidance of the Ministry of Science and Technology (MOST) for the Review and Approval of Human Genetic Resources" of China. Qualified researchers can gain access to the data by submitting a formal application through the repository platform. Data requests will be reviewed and approved by the corresponding author within 4 weeks. The access will be granted only for academic research purposes. Once access has been granted, the data will remain available to the requestor for 6 months in accordance with institutional and ethical guidelines. All the data, analytic methods, and study materials that support the findings of this study are available from the corresponding author upon reasonable request. The remaining data are available within the Article, Supplementary Information or Source Data file. Source data are provided with this paper.

Code availability

All original codes for each figure can be found at https://github.com/DBlab-web/LNSubType. All analyses were carried out using freely available software and packages.

References

1. Sung, H. et al. Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **71**, 209–249 (2021).
2. Amin, M. B. et al. The eighth edition AJCC cancer staging manual: continuing to build a bridge from a population-based to a more "personalized" approach to cancer staging. *CA Cancer J. Clin.* **67**, 93–99 (2017).
3. Dekker, E., Tanis, P. J., Vleugels, J. L. A., Kasi, P. M. & Wallace, M. B. Colorectal cancer. *Lancet* **394**, 1467–1480 (2019).
4. Locker, G. Y. et al. ASCO 2006 update of recommendations for the use of tumor markers in gastrointestinal cancer. *J. Clin. Oncol.* **24**, 5313–5327 (2006).
5. Rahim, M. K. et al. Dynamic CD8(+) T cell responses to cancer immunotherapy in human regional lymph nodes are disrupted in metastatic lymph nodes. *Cell* **186**, 1127–1143.e18 (2023).
6. Gasteiger, G., Ataide, M. & Kastenmüller, W. Lymph node—an organ for T-cell activation and pathogen defense. *Immunol. Rev.* **271**, 200–220 (2016).
7. Dieterich, L. C. et al. Tumor-associated lymphatic vessels upregulate PDL1 to inhibit T-cell activation. *Front. Immunol.* **8**, 66 (2017).
8. Goode, E. F., Roussos Torres, E. T. & Irshad, S. Lymph node immune profiles as predictive biomarkers for immune checkpoint inhibitor response. *Front. Mol. Biosci.* **8**, 674558 (2021).
9. Mishra, R., Sukhbaatar, A., Mori, S. & Kodama, T. Metastatic lymph node targeted CTLA4 blockade: a potent intervention for local and distant metastases with minimal ICI-induced pneumonia. *J. Exp. Clin. Cancer Res.* **42**, 132 (2023).
10. Esterházy, D. et al. Compartmentalized gut lymph node drainage dictates adaptive immune responses. *Nature* **569**, 126–130 (2019).
11. du Bois, H., Heim, T. A. & Lund, A. W. Tumor-draining lymph nodes: at the crossroads of metastasis and immunity. *Sci. Immunol.* **6**, eabg3551 (2021).
12. Grant, S. M. et al. The lymph node at a glance—how spatial organization optimizes the immune response. *J. Cell Sci.* **133**, jcs241828 (2020).
13. Cruz de Casas, P., Knöpper, K., Dey Sarkar, R. & Kastenmüller, W. Same yet different—how lymph node heterogeneity affects immune responses. *Nat. Rev. Immunol.* **24**, 358–374 (2024).
14. Guinney, J. et al. The consensus molecular subtypes of colorectal cancer. *Nat. Med.* **21**, 1350–1356 (2015).
15. Sadanandam, A. et al. A colorectal cancer classification system that associates cellular phenotype and responses to therapy. *Nat. Med.* **19**, 619–625 (2013).
16. Luca, B. A. et al. Atlas of clinically distinct cell states and ecosystems across human solid tumors. *Cell* **184**, 5482–5486.e28 (2021).
17. Newman, A. M. et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* **12**, 453–457 (2015).
18. Krishnamurty, A. T. & Turley, S. J. Lymph node stromal cells: cartographers of the immune system. *Nat. Immunol.* **21**, 369–380 (2020).
19. Rodda, L. B. et al. Single-cell RNA sequencing of lymph node stromal cells reveals niche-associated heterogeneity. *Immunity* **48**, 1014–1028.e6 (2018).
20. Allen, C. D. & Cyster, J. G. Follicular dendritic cell networks of primary follicles and germinal centers: phenotype and function. *Semin. Immunol.* **20**, 14–25 (2008).
21. Heesters, B. A., Myers, R. C. & Carroll, M. C. Follicular dendritic cells: dynamic antigen libraries. *Nat. Rev. Immunol.* **14**, 495–504 (2014).
22. O'Reilly, S., Tsou, P. S. & Varga, J. Senescence and tissue fibrosis: opportunities for therapeutic targeting. *Trends Mol. Med.* **30**, 1113–1125 (2024).
23. Zhang, C. et al. Localization and density of tertiary lymphoid structures associate with molecular subtype and clinical outcome in colorectal cancer liver metastases. *J. Immunother. Cancer* **11**, e006425 (2023).
24. Pezoldt, J. et al. Postnatal expansion of mesenteric lymph node stromal cells towards reticular and CD34(+) stromal cell subsets. *Nat. Commun.* **13**, 7227 (2022).
25. Simon, T. et al. Differential regulation of T-cell immunity and tolerance by stromal laminin expressed in the lymph node. *Transplantation* **103**, 2075–2089 (2019).
26. Fletcher, A. L., Acton, S. E. & Knoblich, K. Lymph node fibroblastic reticular cells in health and disease. *Nat. Rev. Immunol.* **15**, 350–361 (2015).
27. Krautler, N. J. et al. Follicular dendritic cells emerge from ubiquitous perivascular precursors. *Cell* **150**, 194–206 (2012).
28. Reticker-Flynn, N. E. et al. Lymph node colonization induces tumor-immune tolerance to promote distant metastasis. *Cell* **185**, 1924–1942.e23 (2022).
29. Fransen, M. F. et al. Tumor-draining lymph nodes are pivotal in PD-1/PD-L1 checkpoint therapy. *JCI Insight* **3**, e124507 (2018).
30. Li, Y. L. & Hung, W. C. Reprogramming of sentinel lymph node microenvironment during tumor metastasis. *J. Biomed. Sci.* **29**, 84 (2022).
31. Lal, N., Chan, D. K. H., Ng, M. E., Vermeulen, L. & Buczacki, S. J. A. Primary tumour immune response and lymph node yields in colon cancer. *Br. J. Cancer* **126**, 1178–1185 (2022).
32. Huang, X., Tang, T., Zhang, G. & Liang, T. Identification of tumor antigens and immune subtypes of cholangiocarcinoma for mRNA vaccine development. *Mol. Cancer* **20**, 50 (2021).
33. Li, B., Cui, Y., Nambiar, D. K., Sunwoo, J. B. & Li, R. The immune subtypes and landscape of squamous cell carcinoma. *Clin. Cancer Res.* **25**, 3528–3537 (2019).
34. Huang, X., Zhang, G., Tang, T. & Liang, T. Identification of tumor antigens and immune subtypes of pancreatic adenocarcinoma for mRNA vaccine development. *Mol. Cancer* **20**, 44 (2021).
35. Holm, J. B. et al. Ultrahigh-throughput multiplexing and sequencing of >500-base-pair amplicon regions on the illumina HiSeq 2500 platform. *mSystems* **4**, e00029-19 (2019).
36. Bao, Y. et al. Transcriptome profiling revealed multiple genes and ECM-receptor interaction pathways that may be associated with breast cancer. *Cell Mol. Biol. Lett.* **24**, 38 (2019).
37. Qiu, X. et al. Reversed graph embedding resolves complex single-cell trajectories. *Nat. Methods* **14**, 979–982 (2017).
38. Trapnell, C. et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386 (2014).
39. Hao, Y. et al. Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat. Biotechnol.* **42**, 293–304 (2024).
40. Che, L. H. et al. A single-cell atlas of liver metastases of colorectal cancer reveals reprogramming of the tumor microenvironment in response to preoperative chemotherapy. *Cell Discov.* **7**, 80 (2021).
41. Pelka, K. et al. Spatially organized multicellular immune hubs in human colorectal cancer. *Cell* **184**, 4734–4752.e20 (2021).
42. Aibar, S. et al. SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods* **14**, 1083–1086 (2017).
43. Chen, T. et al. The genome sequence archive family: toward explosive data growth and diverse data types. *Genomics Proteom. Bioinforma.* **19**, 578–583 (2021).
44. CNCB-NGDC Members and Partners. Database Resources of the National Genomics Data Center, China National Center for Bioinformation in 2024. *Nucleic Acids Res.* **52**, D18–d32 (2024).

Acknowledgements

Author contributions

Xu Guan (X.G.), MD (Conceptualization: supporting; Data curation: supporting; Formal analysis: supporting; Investigation: supporting; Methodology: lead; Validation: lead; Supervision: equal; Visualization: supporting; Resources: supporting; Funding acquisition: supporting; Writing—original draft: lead; Writing—review & editing: supporting). Yueyang Zhang (YY.Z.), MD (Conceptualization: supporting; Data curation: supporting; Formal analysis: supporting; Investigation: lead; Methodology: supporting; Visualization: lead; Validation: supporting; Writing—original draft: supporting). Ruifang Sun (R.S.), MD (Data curation: supporting; Formal analysis: supporting; Investigation: supporting; Validation: supporting; Resources: supporting; Writing—review & editing: supporting). Guiyu Wang (G.W.), MD (Conceptualization: supporting; Methodology: supporting; Project administration: supporting; Resources: supporting; Supervision: equal). Xiaoman Bi (X.B.), MD (Conceptualization: supporting; Data curation: supporting; Formal analysis: supporting; Methodology: supporting; Supervision: equal; Validation: supporting; Visualization: supporting; Writing—review & editing: supporting). Zhen Zhang (Z.Z.), MD (Data curation: supporting; Investigation: supporting; Resources: supporting). Likun Zan (L.Z.), MD (Data curation: supporting; Investigation: supporting; Resources: supporting). Yongsheng Meng (Y.M.), MD (Data curation: supporting; Investigation: supporting; Resources: supporting). Yunxiao Liu (Y.L.), MD (Data curation: supporting; Investigation: supporting). Jinyuan Guo (J.G.), MD (Data curation: supporting; Investigation: supporting; Visualization: supporting). Jian Ma, MD (Data curation: supporting; Investigation: supporting). Jiale Li (J.L.), MD (Data curation: supporting; Investigation: supporting). Hanqing Hu (H.H.), MD (Data curation: supporting; Investigation: supporting). Yixun Zhang (YX.Z.), MD (Data curation: supporting; Investigation: supporting). Xin Zhang (X.Z.), MD (Data curation: supporting; Investigation: supporting). Maoxi Liu (M.L.), MD (Data curation: supporting; Investigation: supporting). Bo Jiang (B.J.), MD (Data curation: supporting; Investigation: supporting). Haiyi Liu (H.L.), MD (Data curation: supporting; Investigation: supporting). Shuangmei Zou (S.Z.), MD (Data curation: supporting; Investigation: supporting; Resources: supporting). Yanfeng Xi (Y.X.), MD (Data curation: supporting; Investigation: supporting; Methodology: supporting; Resources: supporting; Supervision: equal; Writing—review & editing: supporting). Haitao Zhou (H.Z.), MD (Conceptualization: supporting; Funding acquisition: supporting; Methodology: supporting; Project administration: supporting; Resources: supporting; Supervision: equal; Writing—review & editing: supporting). Wenqi Bai (W.B.), MD (Conceptualization: supporting; Funding acquisition: supporting; Methodology: supporting; Project administration: supporting; Resources: supporting; Supervision: equal; Writing—review & editing: supporting). Deng Wu (D.W.), MD (Conceptualization: supporting; Data curation: supporting; Formal analysis: lead; Methodology: supporting; Supervision: equal; Validation: supporting; Visualization: supporting; Writing—review & editing: supporting). Xishan Wang (X.W.), MD (Conceptualization: lead; Data curation: lead; Funding acquisition: lead; Methodology: supporting; Resources: lead; Supervision: equal; Writing—review & editing: lead).

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/s41467-025-63200-z.

Correspondence and requests for materials should be addressed to Yanfeng Xi, Haitao Zhou, Wenqi Bai, Deng Wu or Xishan Wang.

Peer review information *Nature Communications* thanks Iain Beehuat Tan, Tomoharu Yoshizumi, who co-reviewed with Koji Ando, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions 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.