

Single-cell RNA-seq reveals the transcriptional program underlying tumor progression and metastasis in neuroblastoma

Zhe Nian^{1,*}, Dan Wang^{1,*}, Hao Wang^{1,*}, Wenxu Liu¹, Zhenyi Ma², Jie Yan³, Yanna Cao³, Jie Li³, Qiang Zhao (✉)³, Zhe Liu (✉)^{1,2,4,#}

¹Department of Immunology, School of Basic Medical Sciences, Tianjin Medical University, Tianjin 300070, China; ²Zhejiang Key Laboratory of Medical Epigenetics, Department of Cell Biology, School of Basic Medical Sciences, Hangzhou Normal University, Hangzhou 311121, China; ³Department of Pediatric Oncology, Tianjin Medical University Cancer Institute and Hospital, National Clinical Research Center for Cancer, Key Laboratory of Cancer Prevention and Therapy, Tianjin's Clinical Research Center for Cancer, Tianjin 300060, China; ⁴Collaborative Innovation Center for Cancer Personalized Medicine, Nanjing Medical University, Nanjing 211166, China

© Higher Education Press 2024

Abstract Neuroblastoma (NB) is one of the most common childhood malignancies. Sixty percent of patients present with widely disseminated clinical signs at diagnosis and exhibit poor outcomes. However, the molecular mechanisms triggering NB metastasis remain largely uncharacterized. In this study, we generated a transcriptomic atlas of 15 447 NB cells from eight NB samples, including paired samples of primary tumors and bone marrow metastases. We used time-resolved analysis to chart the evolutionary trajectory of NB cells from the primary tumor to the metastases in the same patient and identified a common ‘starter’ subpopulation that initiates tumor development and metastasis. The ‘starter’ population exhibited high expression levels of multiple cell cycle-related genes, indicating the important role of cell cycle upregulation in NB tumor progression. In addition, our evolutionary trajectory analysis demonstrated the involvement of partial epithelial-to-mesenchymal transition (p-EMT) along the metastatic route from the primary site to the bone marrow. Our study provides insights into the program driving NB metastasis and presents a signature of metastasis-initiating cells as an independent prognostic indicator and potential therapeutic target to inhibit the initiation of NB metastasis.

Keywords single-cell RNA sequencing; metastasis; neuroblastoma; epithelial-to-mesenchymal transition

Introduction

Neuroblastoma (NB) is the most common extracranial solid tumor occurring during childhood and accounts for ~15% of childhood cancer-related deaths [1]. NB arises from aberrant development of the sympathetic nervous system from neural crest cells [2]. It is characterized as a highly heterogeneous disease with multiple clinical presentations, ranging from localized asymptomatic primary tumors to widely disseminated diseases with poor outcomes [3,4]. Sixty percent of patients have high-risk NB with metastasis at diagnosis [5,6]. Despite the significant advances in the treatment of pediatric cancer

over the past two decades, NB remains a highly refractory malignancy, with a 5-year survival rate of less than 50% for the majority of patients who are diagnosed with high-risk disease [5]. Previous studies with diverse NB animal models have allowed the identification of the molecular mechanisms driving NB metastatic dissemination. For instance, Teitz *et al.* showed that caspase-8 deletion significantly enhanced bone marrow metastasis of neuroblastoma in mice [7]. Delloye-Bourgeois *et al.* developed a model in chick embryos and found that NB dissemination is induced by the shutdown of a pro-cohesion autocrine signal, SEMA3C [8]. An additional study focused on the LIM-domain-only transcriptional cofactor LMO1, which elevates the expression of genes affecting tumor cell-extracellular matrix interactions and promotes NB cell invasion in zebrafish [9]. Although many molecular cues controlling NB dissemination have been deciphered, the biology of NB metastasis remains less well understood. It has been proposed that only a

Received January 8, 2024; accepted April 18, 2024

Correspondence: Zhe Liu, zheliu@tmu.edu.cn;

Qiang Zhao, zhaoqiang@tjmuch.com

*These authors contributed equally to this work.

#Leading contact.

small portion of primary tumor cells are metastasis-initiating cells that can migrate from the primary site to initiate metastasis [10]. Therefore, identification of this crucial subpopulation of tumor cells in NB may uncover the molecular players involved in the onset of the NB metastatic process and provide novel targets for therapy.

Single-cell RNA sequencing (scRNA-seq) can help identify cell types, cell states, and the drivers that govern cell state transitions, allowing us to better understand the mechanisms underlying tumor formation and progression. In recent years, many studies have used scRNA-seq to reveal intratumor heterogeneity [11–13]. Recently, several scRNA-seq studies of primary NB tumors and a scRNA-seq study of bone marrow metastases of NB tumors have been performed and provided insight into the cell of origin, the plasticity, and the cell–cell interactions between NB cells and the immune microenvironment [14–21]. Although these studies have significantly improved our understanding of NB tumor biology, transcriptomic comparisons between primary tumors and matched metastases at single-cell resolution remain elusive. The onset of the NB metastatic process is still a mystery. In this study, we employed scRNA-seq to determine the molecular features of primary tumors and their matched metastases. By using time-resolved analyses of scRNA-seq data, we determined the potential transitional trajectories of tumor cells and identified the metastasis-initiating subpopulations. Our work provides important insights into the program driving tumor progression and metastasis in NB and is potentially valuable for future therapeutic strategies in clinical NB patients with metastases.

Materials and methods

Ethics statement and collection of human samples

Tissues were collected from Tianjin Medical University Cancer Institute and Hospital. The use of clinical samples was reviewed and approved by the Institutional Ethics Committee of Tianjin Medical University Cancer Institute and Hospital, Tianjin, China (Approve number: E20210664). Informed written consent was obtained from the patients. A total of six patients with NB were included in this study. Tumor specimens were collected from untreated primary tumors and untreated paired metastatic tumors after pathological confirmation of malignancy. In total, eight NB tumor samples were collected. Clinical information including patient age, patient sex, localization and NB subgroup, is provided in Table S1.

Tissue dissociation and library preparation

Tumor tissue was enzymatically digested with a MACS tumor dissociation kit (130-108-339, Miltenyi Biotec) for

60 min on a rotor at 37 °C, following the manufacturer's instructions. After digestion, the cells were centrifuged at 400× *g* and 4 °C for 5 min. They were subsequently resuspended in 100 U/mL DNase I (D4527, Sigma Aldrich), and incubated at 37 °C for 15 min. After filtration and centrifugation, the pellet was suspended in red blood cell lysis buffer (R1010, Solarbio) and incubated on ice for 2 min to lyse the red blood cells. Finally, the cells were centrifuged and resuspended in culture medium (DMEM/F12) (C11330500BT, Gibco) supplemented with 10% FBS. The isolated cells were stained with DAPI (564907, BD Biosciences) and CD45 (H20451, SUNGENE BIOTECH) and then washed with PBS (02-024-1ACS, Biological Industries) with 2% FBS (04-001-1ACS, Biological Industries). The CD45[−] and DAPI[−] cells were sorted on a BD FACS Aria II for further analysis. For droplet-based scRNA-seq, single cells were captured and barcoded with a 10x Chromium Controller (10x Genomics). Single-cell RNA sequencing libraries were prepared using the Chromium Single Cell 3'v3 Reagent Kit (10x Genomics). Sequencing libraries were loaded on an Illumina NovaSeq with 2 × 150 paired-end kits.

scRNA-seq data processing

The raw sequencing data were processed using the Cell Ranger Software Suite (v.4.0.0) to align them to the GRCh38 human reference genome and generate gene-barcode matrices. To visualize the processed data from the Cell Ranger analysis, Seurat (v.4.0.0) was employed to perform cell clustering. First, the filtered gene-barcode matrix of the sample identified by Cell Ranger Count pipeline was used as input for the Seurat object. Low-quality cells/dying cells were removed, and the cells with proper read count and gene number ($nFeature_RNA > 200$), as well as a low percentage of mitochondrial transcripts (percent. mt < 25) were retained; in addition, cells with more than 1% of reads mapped to hemoglobin genes were filtered out. Subsequently, potential cell doublets were predicted and removed using DoubletFinder (v.2.0.3) [22]. Then, the data were normalized, scaled, and subjected to linear dimensionality reduction using principal component analysis (PCA). Finally, the cell distribution was clustered and visualized via UMAP analysis. Batch effects across different samples were eliminated by the Harmony (v.1.0) [23] method.

Inferred CNV analysis

To distinguish malignant cells from nonmalignant cells, syngeneic cells from published single cell RNA-seq data [16] were used as normal reference cells to estimate CNVs in the potential tumor cell populations. Initial

CNVs from expression profiles were inferred by calculating the average Transcripts Per kilobase Million (TPM) expression value of 100 adjacent genes on chromosomes within a sliding window. To prevent the inflation of expression level differences caused by generally undetectable genes, the TPM values were adjusted by dividing them with the multiple of the difference between the actual size of the single-cell library and one million transcripts, ensuring that each transcript is not counted repeatedly [24]. To avoid the considerable impact of any particular gene on the moving average of the relative expression values, we limited the expression levels within a specific range for different cells: positive values above the upper 99th quantile of all positive values were replaced by the expression value of that quantile, and the values less than the lower 99th quantile of all negative values were similarly replaced. Inferred CNVs were finally clustered and visualized by the R package pheatmap based on the CNVs.

WGS data analysis

The WGS reads of T57 were aligned to the GRCh38 human reference genome using BWA (v.0.7.17) and subsequently processed by the Picard pipelines. However, WGS data of the other samples were generated using Oxford Nanopore Technologies (ONT) and were aligned to the GRCh38 human reference genome according to the method of Li *et al.* [25]. Furthermore, we performed copy number variant calling in both the WGS data obtained from tumor samples and their corresponding peripheral blood cells with CNVkit (v.0.9.9) by the cnvkit.py batch command. The regions of the genome used for copy number variant calling were limited to the gene loci filtered in the analysis of inferred CNVs and were subsequently matched to the inferred CNVs. Structural variants affecting TERT rearrangement were determined using Sniffles [26].

Cell type determination

The cell type identified in each cluster was manually annotated based on the expression of canonical markers. We used AddModuleScore function in the Seurat package to visualize the expression of cell-type markers.

Comparison of malignant cells and developing adrenal cell populations

To quantify cellular similarities between the malignant cells and normal cells in the developing adrenal gland, malignant cells in all eight NB samples were inferred by mapping our scRNA-seq data to the well-annotated single-cell database of adrenal medullary cell population [15,16] with the R package SingleR (v.1.0.5) (de.method =

‘Wilcox’) [27].

Identification of differentially expressed genes (DEGs) and biomarker clustering

The FindAllMarkers function with the parameter ‘thresh.test = 1.5’ in the Seurat package was used to identify the DEGs in the cell clusters. For computing the DEGs, we analyzed all genes that were expressed in at least 20% of cells in either of the two comparison populations and exhibited a minimum difference in expression of 0.25 on a natural log scale. The top 10 genes ordered by the average expression level in each cluster were visualized using DoHeatmap function of Seurat package.

RNA velocity estimation of tumor cells

To analyze the time-resolved developmental sequence of tumor cells from the 10x Genomics Chromium samples, we estimated RNA velocity by distinguishing between unspliced and spliced mRNAs. We performed the analysis of single-cell RNA velocity using the velocity software (v. 0.17.17) by running 10x command [28]. The output loom file was used to calculate RNA velocity values for each gene in each cell, and subsequently created a low-dimensional representation of the RNA velocity vector by following the scVelo (v.0.1.20) python pipeline. The results were visualized on UMAP plots obtained from Seurat clustering analysis to ensure consistency.

Trajectory analysis of tumor cells

Pseudotime trajectory analysis was conducted using Monocle2 (v.2.26.0) [29] and Slingshot (v.2.4.0) [30] with default parameters to determine the trajectories of tumor cells in each sample. We used the expression matrix derived from Seurat object to build a CellDataSet for the Monocle2 pipeline. After obtaining size factor and dispersion estimates, trajectory ordering genes were called by $\text{mean_expression} \geq 0.1$ and $\text{dispersion_empirical} \geq 10$. Data dimensionality was reduced using the reduceDimension function with max_components set to 2 and reduction_method set to DDRTree. Dynamically expressed genes were identified by the VGAM function with a full model of ‘~sm.ns(Pseudo-time)’, and a q-values greater than 0.01 was used as the filtering criterion. Expression patterns were visualized with pheatmap, and the enrichGO function implemented in ClusterProfiler (v.3.18.1) was used to determine the enriched Gene Ontology biological processes within each module. Gene expression plots along pseudotime were generated using the plot_genes_in_pseudotime function in the Monocle package and the plot_density function in

the *Nebulosa* (v.1.8.0) package. For Slingshot cellular trajectory analysis of tumor cells in T57 and T62, the input matrix was filtered and normalized by the R package *Seurat*. The *slingPseudotime* scores were added as metadata to Single Cell Experiment objects and visualized using the *FeaturePlot* function in *Seurat* package.

Evaluation of differentiation potential

The R package *CytoTRACE* (v.0.3.3) was used to predict the differentiation state of cells from using the scRNA-seq data from samples T57 and T62. *CytoTRACE* scores were computed for each malignant cell in samples T57 and T62, ranging from 0 to 1. Higher scores indicate increased stemness and decreased differentiation.

Identification of variant information from scRNA-seq data

To establish clonal relationships between cells based on the 10x Genomics scRNA-seq data, we referred to the method of Zhang *et al.* [28]. First, we separated the .bam file generated with *Cell Ranger* into an individual .bam file by *samtools* (v.1.7) software according to the barcode of each cell. Next, the *bcftools* (v.1.8) was used to generate VCF format files for each cell. To select informative variants, we applied specific criteria, including a minimum base coverage threshold of 80% in cells, $VAF > 0.1$, and $DP > 10$, to filter out irrelevant variants. Selected informative variants were sorted and visualized in a heatmap by *seaborn* (v.0.11.0). We used *bcftools* to determine the number of mutations in each cell. All these bases were concatenated into a DNA sequence for each cell, and then the identical sequences in each *Seurat* cluster were removed. Multiple sequence alignment was performed by *MUSCLE* integrated in the *MEGA-X* software. The phylogenetic tree was constructed utilizing the Maximum Likelihood method within *MEGA-X*. The original reference sequence (GRCh38) was adopted as the outgroup of the phylogenetic tree analysis. These selected variants were utilized for conducting PCA using the PCA function available in the decomposition module from *scikit-learn* (v.0.24.2). To verify the existence of these selected variants, *Mutect2* was employed for calling somatic mutations using the WGS data of T57.

Immunostaining

NB tissues were obtained from the patients undergoing surgical resection with a histological diagnosis of high-risk poor-differentiated NB at Tianjin Medical University Cancer Institute and Hospital in China. Samples were fixed with 4% paraformaldehyde at 4 °C overnight, then

dehydrated in gradient alcohol and xylene, and finally embedded in paraffin. Paraffin blocks were cut into 5- μ m sections. Tumor sections were immunostained with antibodies against NCAM1(ab220360, Abcam), VIM (V5255, Sigma), and EPCAM (14452s, Cell Signaling Technology), the secondary antibodies used were Alexa Fluor 555, goat anti-mouse IgG (H + L) (A-21422, Invitrogen) and Alexa Fluor 488, goat anti-rabbit IgG (H + L) (A-11008, Invitrogen). Nuclei were counterstained with DAPI (1 μ g/mL). The fluorescence images were observed under a confocal microscope (ZEISS, LSM800).

Biomarker identification and pathway analysis of the ‘starter’ subpopulations

DEGs were identified by the function *FindMarkers* in *Seurat* package, applying the Wilcoxon rank-sum test. To identify DEGs in the ‘starter’ cluster, UMAP plots presented in Fig. 2B were used to compare their expression levels with those of all other clusters, requiring a difference in expression of ≥ 0.25 on a natural logarithmic scale. Subsequently, \log_2 (fold changes) were determined for each gene using *FindMarkers*, and GSEA was performed using the *ClusterProfiler* package through *gseGO* function, with the GO biological process ontology gene sets. The enriched pathways were visualized using a bar plot. The gene expression signature of ‘starter’ cells was created using the top 150 genes ranked by average log fold change (*avg_log2FC*), as computed by *FindMarkers*.

Analysis of TF activity in single cells

The TF activity in individual cells was inferred using *DoRothEA* (v.1.9.0) [31] package. Following the *DoRothEA* manual, we estimated the TF activities for each cell using *Viper* [32], a GSEA-like approach, as implemented in the *DoRothEA* R package and tutorial [33]. TFs were identified with confidence levels ranging from A (high confidence) to C (low confidence). Heatmap was generated to visualize the TFs with the most pronounced differences in activity across each cluster.

Expression of the ‘starter’-cell signature in public data sets

We obtained two scRNA-seq data sets (GEO data set GSE137804, GSE147766) from publicly available sources. After quality control and data harmonization, we clustered cells with UMAP. By utilizing the metadata available in the data sets [14,34], we annotated malignant cells and non-malignant cells. The expression of ‘starter’-cell signature was computed using the *AddModuleScore* function in *Seurat*. *Monocle* analysis was similar as

described above.

Analysis of cell–cell interactions in scRNA-seq data

CellChat (v1.5.0) [35] was used to predict potential cell–cell interactions on the basis of the specific, normalized single-cell expression profiles. The scRNA-seq samples containing cells from the microenvironment were sourced from previously published works [14,34]. All graphs of visualized cell communications were generated using the functions implemented in the CellChat package.

Applying scMLnet to analyze interactions between microenvironment and tumor cells

We applied scMLnet [36] method with default parameters to analyze interactions between microenvironment and tumor cells based on the NB scRNA-seq data previously examined in the CellChat analysis.

Validation of the ‘starter’-cell prognostic signature

The signature was subjected to validation with publicly available data sets. We included one scRNA-seq data set (GSE137804) and three bulk transcriptome sequencing (GSE62564 from GEO data set, Dgc2102 and E-MTAB-8248 from the R2 database) data sets to validate the ‘starter’-cell prognostic signature. These data sets included expression data and clinical information such as patient sex, patient age, INSS stage, MYCN amplification status, and TERT status. For the scRNA-seq data set, the signature scores were computed using AddModuleScore function in Seurat. For the bulk RNA-seq data sets, the scores were calculated by averaging the z-scores of all genes in the ‘starter’-cell signature. Then, the samples were categorized into the low- or high-expression groups based on the corresponding signature scores: \leq median or $>$ median, respectively. Survival curves were obtained with the Kaplan–Meier method. *P*-values between the comparison groups were calculated using the two-sided log-rank test. Additionally, hazard ratios were calculated using the Cox proportional-hazards model, with the first level serving as the reference for each variable. The block at the center of the error bars represents the weighted mean. Whiskers of error bars represent the 95% confidence intervals.

Statistical analysis

Statistical analysis was performed using the language of Python (v.3.8.6), and R (v.4.0.0), as well as the software of SPSS 25.0, and GraphPad Prism (v.8.0) software. Statistical significance was defined as $P < 0.05$. If not detailed above, specific statistical analyses undertaken in

this manuscript are detailed in the results section and in the figure legends.

Results

Single-cell expression atlas of NB

Recent scRNA-seq data related to the normal development of murine and human adrenal glands identified the cellular origin of NB [15–17,19]. To explore the transcriptional profiles during the progression of NB, we isolated individual cells from eight freshly resected and dissociated human NB samples from six untreated patients: two matched primary tumors and bone marrow metastases (patients 57 and 62), three high-risk primary tumors (patients 59, 60, and 63), and one primary ganglioneuroblastoma (GNB) tumor (patient 55). In addition, we performed whole genomic sequencing (WGS) on T57-P, T57-M, T59-P, T60-P, and T62-P samples and determined that there is no *TERT* rearrangement present in these samples. No NB tumor except for T57 harbored *MYCN* amplification (Table S1). After removing inflammatory infiltrate and DAPI-positive dead cells, we generated scRNA-seq profiles using the 10x Genomics Chromium platform (Figs. 1A, S1A, and S1B). After quality filtering, we acquired single-cell transcriptomes for a total of 15 447 cells (1168 cells from T55-P, 1003 cells from T57-P, 2091 cells from T57-M, 2146 cells from T59-P, 1425 cells from T60-P, 4716 cells from T62-P, 1385 cells from T62-M, and 1513 cells from T63-P) for further analysis. The gene numbers, total read count, percentages of mitochondrial gene expression, and percentages of hemoglobin gene expression in each sample were shown in violin plots (Fig. S1C–S1F). After normalization of gene expression, we conducted combined dimensionality reduction and clustering using principal component analysis (PCA) and uniform manifold approximation and projection (UMAP), respectively, and identified 11 clusters with different transcriptional characteristics (Figs. 1B, 1C, and S1G). High heterogeneity was observed between the GNB and NBs (Fig. 1D).

To distinguish malignant cells from nonmalignant cells, we inferred large-scale chromosomal copy number variations (CNVs) based on transcriptomes as previously described [24,37], using the reported scRNA-seq data of fetal sympathoblasts as a control [16]. The inferred CNVs were confirmed by WGS for T57, T59, T60, and T62. Copy number gains at 17q were frequently observed in malignant cells (5 of 6 cases), consistent with previous reports [38]. All malignant cells within the same tumor exhibited a common CNV pattern, and the primary tumor and its corresponding metastases exhibited a common CNV pattern, suggesting their single clone origination

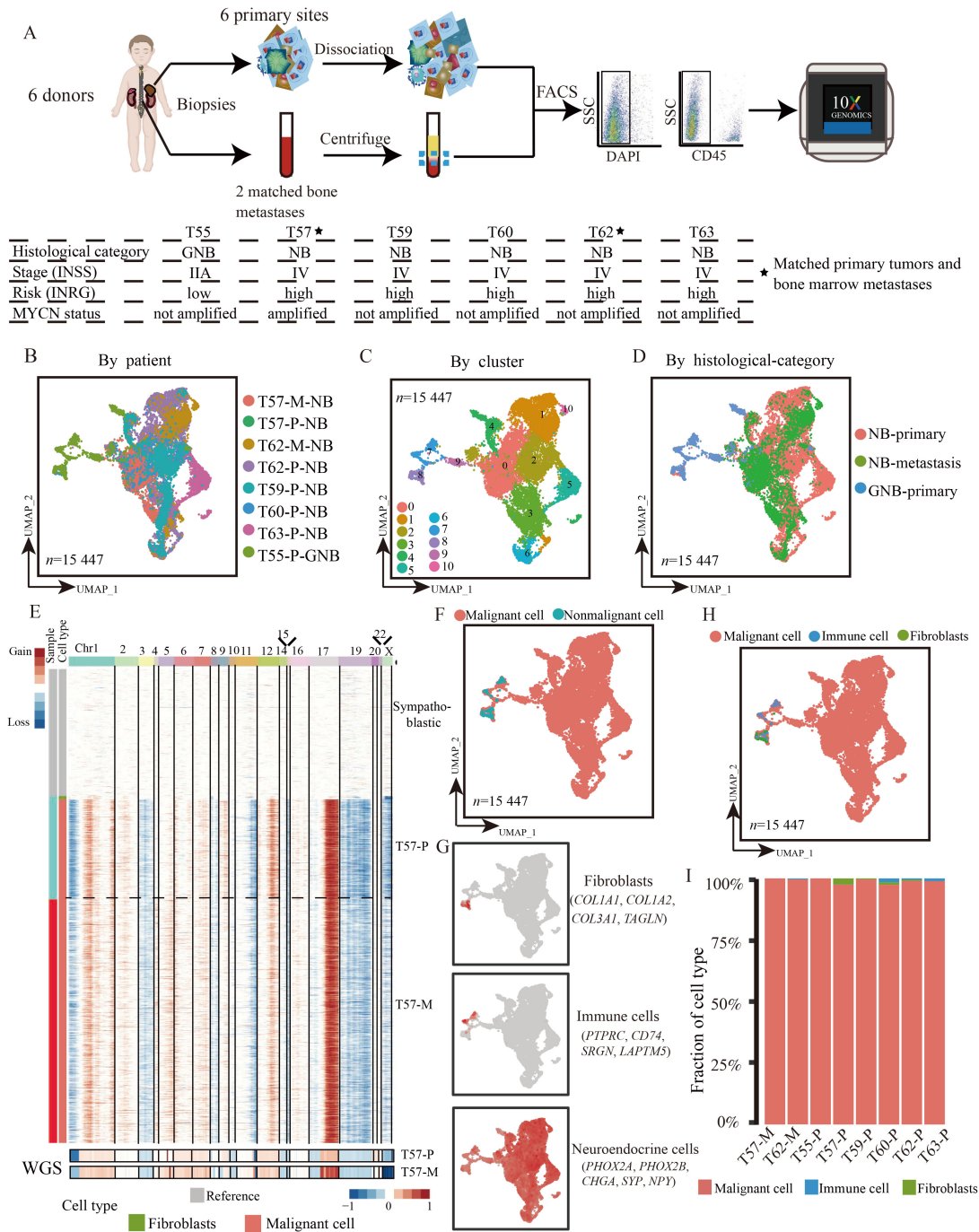


Fig. 1 Single-cell RNA sequencing of NB cells. (A) Overview of the experimental workflow. Eight NB samples (six primary NB tissues and two matched bone marrow tissues) were collected from six NB patients. Single DAPI⁺CD45⁻ human tumor cells were isolated by flow cytometry, and single-cell cDNA libraries were prepared using 10x Genomics technology (upper). NB samples with molecular and clinical annotations (bottom). (B–D) UMAP plots of the 15 447 high-quality cells showing the sample origin (B), transcriptomic clusters (C), and histological category (D). (E) Inferred CNV profiling (upper) and WGS (bottom) were performed to identify malignant and nonmalignant cells in scRNA-seq data for T57. (F) UMAP plots of the 15 447 high-quality cells showing the cell type classification based on the CNV classification. (G) UMAP plots showing the fibroblast signature scores (*COL1A1*, *COL1A2*, *COL3A1*, *TAGLN*) (upper), the immune-cell signature scores (*PITPRC*, *CD74*, *SRGN*, *LAPTM5*) (middle), and the neuroendocrine signature scores (*PHOX2A*, *PHOX2B*, *CHGA*, *SYP*, *NPY*) (bottom) (gray to red). (H) UMAP plots of the 15 447 high-quality cells showing the cell type based on marker gene expression shown in Fig. 1G. (I) The proportion of each cell type in the eight samples.

(Figs. 1E,1F, and S1H–S1L). The cells expressing characteristic markers of fibroblasts or immune cells and showing different CNV patterns with cells expressing neuroendocrine markers were annotated as fibroblasts or

immune cells (Fig. 1G and 1H). The proportions of each cell type in eight samples were shown in Fig. 1I (Table S2).

Previous studies have noted that NB cells share transcriptional similarity with their cells of origin, i.e., neuroblasts or sympathoblasts [15,16]. We then calculated the transcriptomic similarities between our malignant tumor cells and the reported normal cell types in the developing adrenal gland [15,16] (see Methods). As expected, the malignant cells from eight samples were transcriptionally similar to normal neuroblasts or sympathoblasts (Fig. S1M and S1N).

Evolutionary trajectory analysis identified the ‘starter’ of NB at primary and metastatic sites

To further transcriptionally characterize the primary NBs and their corresponding distant metastases, we excluded nonmalignant cells and performed another integrated UMAP analysis for tumors T57 and T62, which contained matched pairs of primary tumor and bone marrow metastases. The malignant cells were mainly clustered according to their tumor of origin (Fig. 2A). The cells in Cluster 1 and Cluster 4 were mainly derived from T57, and the cells of Clusters 0, 2, and 5 were mainly derived from T62. However, the cells of Cluster 3 were evenly distributed across T57 and T62 (Fig. 2B). For the same tumor, each of the clusters contained malignant cells from both the primary and metastatic sites (Figs. 2C, 2D, S2A, and S2B), suggesting substantial conservation of cell types and expression states between primary tumors and their corresponding bone marrow metastases. This finding further supported the preservation of tumor phenotype during metastasis, consistent with previous study [21]. The heatmap showed the molecular features of the different clusters (Fig. 2E). For example, genes associated with neural features (neurotransmitter transmission, *SLC5A7*, *IGFBP5*, and *PDE10A*; neuron projection guidance, *EPHB6*, *NRP2*, and *CNTNAP2*) were preferentially expressed in Clusters 0 and 2. Other differentially expressed genes were related to the cell cycle (*MKI67*, *TOP2A*, and *FOXM1*), cellular response to stimulus (*DNAJA4*, *HSPB1*, and *DNAJB1*), and immune activity (*HLA-C*, *HLA-A*, and *IRF1*) (Table S3).

To explore the timing and nature of the events leading to dissemination of NB cells, we used three unsupervised algorithms to reveal cell state-related developmental trajectories in the primary tumors and their matched bone marrow metastases derived from the same patient. First, we evaluated the scRNA-seq data using UMAP analysis and RNA velocity analysis, which can distinguish the relative abundances of nascent (unspliced) and mature (spliced) mRNA to predict the development of individual cells on a timescale of hours [39,40]. In patient T57, two clear directional flows were observed. One directional flow was from Cluster 3 to Cluster 1 and the other

directional flow was from Cluster 4 to Cluster 1, suggesting that the cells in Cluster 3 or Cluster 4 might be the initiating cell subpopulation (Fig. 2F). The primary tumor and the corresponding bone marrow metastases exhibited similar flow patterns (Fig. 2G).

Then, we used the Monocle algorithm to organize NB cells along a pseudotime trajectory [29]. When we projected the tumor cells along the pseudotime trajectory onto the tumor cells clustered by differential gene expression using UMAP, we found that Cluster 3 but not Cluster 4 was distributed at the starting point in both the primary tumor and metastases of T57 (Fig. 2H–2J).

We next performed trajectory inference on the integrated embedding from an integration strategy using Slingshot [30]. Consistent with the above findings, Slingshot analysis revealed that Cluster 3 was located at an early stage of the trajectory course in T57 (Fig. 2K). Again consistent with the above findings, when we applied the same strategies above to analyze tumor cells from the primary site and metastases of T62, we also identified Cluster 3 as the initiation subpopulation of the primary tumor and metastases in T62 (Fig. S2C–S2G). Additionally, the cells in Cluster 3 showed higher stemness than the other malignant cells, based on their signaling entropy and transcriptomic diversity (Fig. S2H and S2I).

Tracing metastasis-initiating cells in primary tumors

We next sought to reconstruct the clonal relationships in primary tumors and metastases and trace the source of metastasis-initiating cells. It is well known that prospective ‘mutate-and-record’ methods that rely on an engineered genetic label to reconstruct cell lineage relationships in model organisms provide powerful tools to determine the developmental origin of cells [41,42]. However, this approach has largely been limited to model organisms. Retrospective lineage tracing, which detects and utilizes naturally occurring somatic mutations as lineage markers, is an alternative solution to infer clonal dynamics in humans [43,44]. We subsequently focused on exploring the tumor evolutionary trajectories in primary and metastatic samples by analyzing the available variants collected from the scRNA-seq (10x Genomics) data. Because both bulk WGS and scRNA-seq data were available for primary and metastatic samples of T57, we performed this study on T57. The .bam file generated by Cell Ranger was initially separated into an individual .bam file for each cell. Due to the technical characteristics of 10x Genomics sequencing (only partial-length transcripts could be captured from single cells), we collected the available genomic positions that covered more than 80% of the cells. Because sequencing of individual cells allows only a low resolution of the underlying genetic diversity, our subsequent rules for

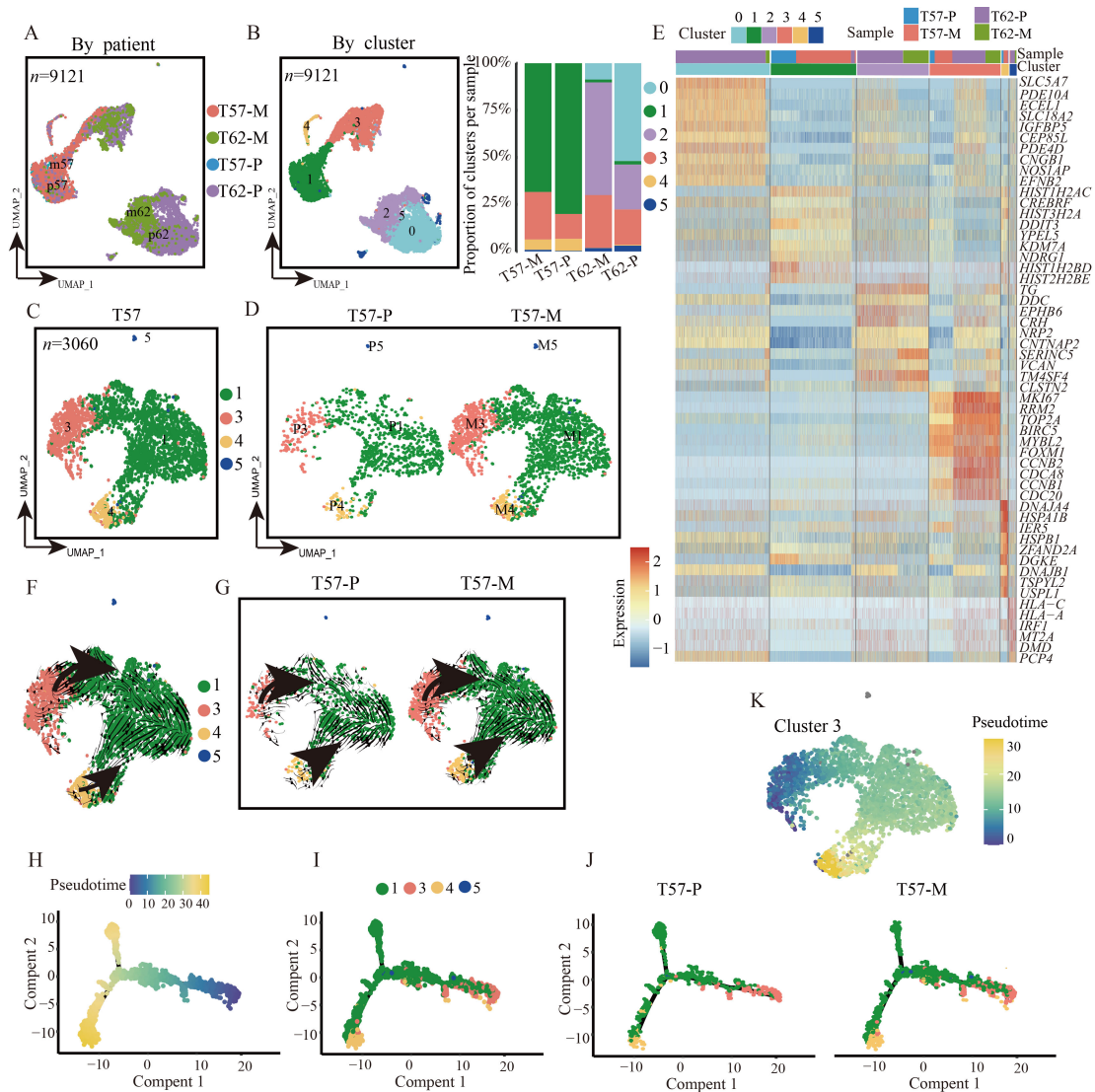


Fig. 2 Trajectory modeling of malignant NB cells identifies a ‘starter’ cluster in T57. (A) UMAP plots of 9121 malignant NB cells from T57 and T62 colored by the sample origin. (B) UMAP plots of 9121 malignant NB cells from T57 and T62 colored by the transcriptomic cluster (left). The proportions of cell clusters within each sample (right). (C) UMAP plots of 3060 malignant NB cells from T57 colored by the transcriptomic cluster. (D) UMAP plots split by the sample origin in T57. (E) Heatmap showing the top ten differentially expressed genes in each cluster based on the UMAP plots shown in Fig. 2B. (F) UMAP plots showing the RNA velocity analysis of the malignant NB cells in T57. (G) UMAP plots showing the separate RNA velocity analyses of the malignant NB cells in the primary tumor and metastases of T57. The velocity of cells is shown by black arrows indicating the predicted direction of development. (H–J) Pseudotime analysis of malignant NB cells from T57 by Monocle 2 showing the timing sequence of tumor cell progression colored by pseudotime (H), colored by intratumoral subpopulation (I), and split by the sample origin in T57 (J). (K) NB cell state trajectories during tumor progression in T57 were inferred by Slingshot and are shown using UMAP plots according to the slingPseudotime score.

screening mutations were relatively less strict: variant allele frequency (VAF) > 0.1 and read depth (DP) > 10 [28]. The obtained variants were then verified by WGS of blood cells, primary tumor, and the bone marrow metastases. Nine variants met the verification criteria and were visualized and sorted to construct the phylogenetic relationships and lineage tracks in primary and metastatic NB from T57. As shown in Fig. 3A–3C, Cluster 3 possessed the most diverse and numerous mutations, suggesting that Cluster 3 may act as a common ancestor

among these tumor cell clusters, supporting the above developmental trajectory analysis. In addition, the metastases had a higher proportion of clonal mutations than the primary tumor (Fig. 3D), consistent with previous studies [45] and our WGS data.

To further investigate the potential evolutionary relationship between the initiating cells (Cluster 3) at the primary (P3) and metastatic (M3) sites, we performed separate trajectory analyses on tumor cells from P3 and tumor cells from M3 in T57 and T62. A potential

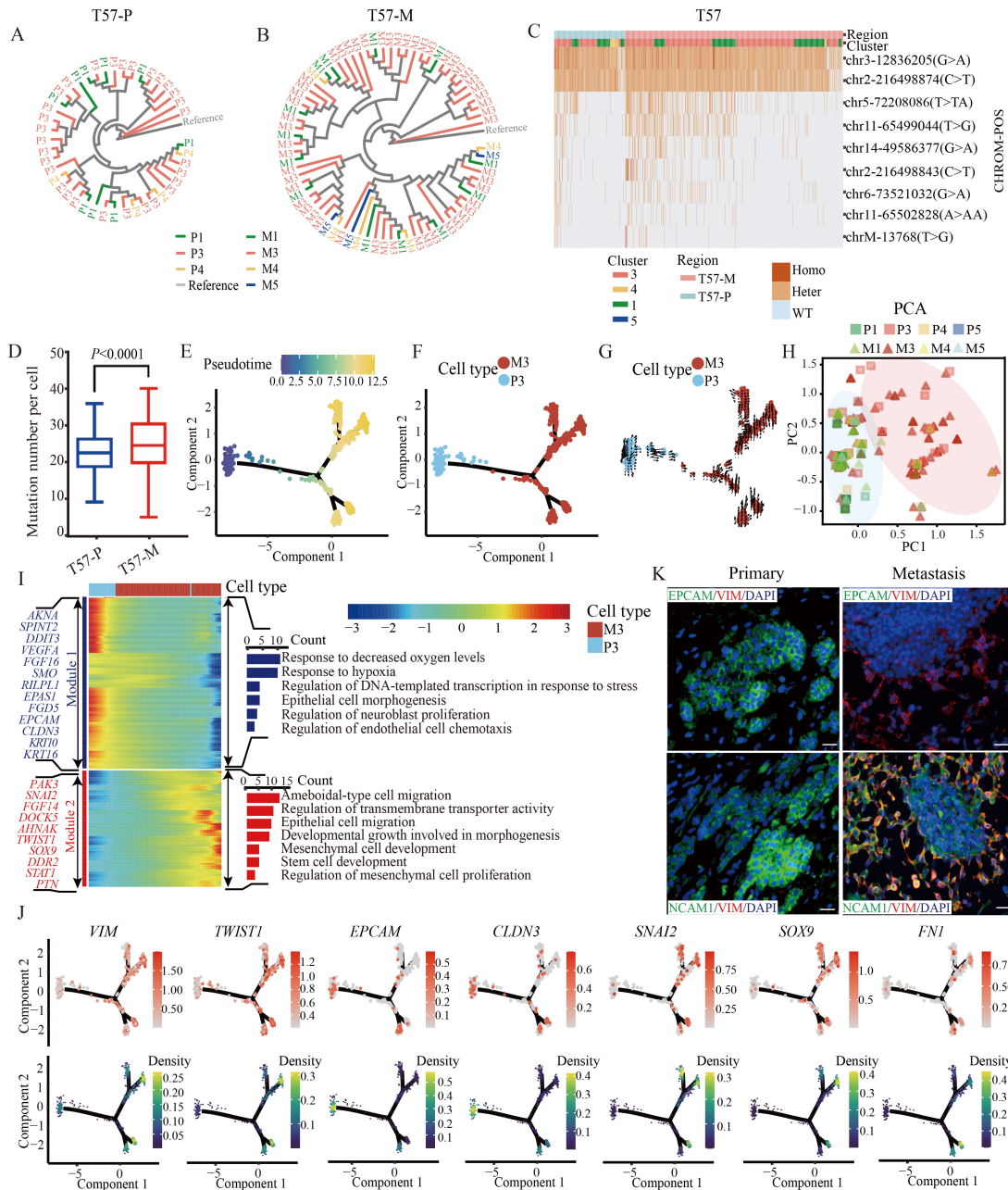


Fig. 3 Tracing of the Cluster 3 cells in T57. (A and B) Unrooted phylogenetic tree of matched pairs of primary (A) and metastatic (B) samples based on the nine selected mutations shown in Fig. 3C. Every cluster of unique cell mutation types is displayed in the same color shown in each branch. Bases were extracted from the reference genome as the contrast branch (Reference). (C) Heatmap showing the nine selected mutations in T57. Dark orange and light orange represent homozygous and heterozygous mutations, respectively. Blue represents wild type (WT). (D) Boxplot showing the mutation number in each cell from paired primary and metastatic samples in T57. $P < 0.0001$ (Wilcoxon test). (E and F) Pseudotime analysis of malignant NB cells from Cluster 3 in T57 by Monocle 2, showing the timing sequence of tumor cell progression colored by pseudotime (E) and sample origin (F). (G) RNA velocity analysis of malignant NB cells from Cluster 3 in the primary tumor and metastases of T57. (H) PCA of every cell subcluster based on the nine selected mutations shown in Fig. 3C. (I) The dynamically expressed genes and enriched gene signatures were significantly associated with pseudotime from P3 to M3 in T57. (J) The expression levels (upper) and the densities in Monocle space (bottom) of epithelial and mesenchymal signatures along the pseudotime trajectory from P3 to M3 in T57. (K) Representative immunofluorescence staining of (primary and metastatic) human NB sections for EPCAM and VIM expression (upper), and NCAM1 and VIM expression (bottom) in patient 1. Scale bars, 20 μm .

evolutionary route from P3 to M3 was observed in the pseudotime trajectories (Figs. 3E, 3F, S3A, and S3B) and in the RNA velocity analysis (Figs. 3G and S3C) in both

T57 and T62. This primary-to-metastatic evolutionary relationship was not found in any other clusters in the Monocle analysis (Fig. S3D–S3I). Additionally, in T57,

M3 cells mainly clustered with P3 cells in the PCA of different clusters based on the nine mutations above, indicating that M3 and P3 share similar somatic mutation patterns and in turn suggesting the evolutionary linkage between M3 and P3 (Fig. 3H). Collectively, these results suggest that the P3 cells may be ‘ancestors’ of the M3 cells.

Dynamically expressed genes along the P3-M3 trajectory were identified and divided into two modules named module 1 and module 2 (see Methods), representing the signatures of malignant cells at the two extremes of the evolutionary trajectory. GO term analysis showed that the genes in module 1 (signature genes of the ‘starter’ cells along the P3-M3 trajectory) were involved in the response to stress, endothelial chemotaxis, neuroblast proliferation, and epithelial cell morphogenesis, while the genes in module 2 (signature genes of the ‘end’ cells along the P3-M3 trajectory) were involved in amoeboid migration, mesenchymal proliferation, and stem cell development (Figs. 3I and S3J). Given that P3 exhibited more epithelial features and M3 exhibited more mesenchymal features, we speculated that epithelial-to-mesenchymal transition (EMT) may occur during NB metastasis. We therefore closely examined the EMT program along the P3-M3 trajectory. The mesenchymal markers *VIM* and *TWIST1* were weakly expressed in a subset of P3 cells at the starting point, and their expression gradually increased along the P3-M3 trajectory and peaked in the majority of M3 cells at the endpoint. In contrast, epithelial markers, such as *EPCAM* and *CLDN3*, were significantly downregulated with progression along the P3-M3 trajectory. Additionally, the expression of the EMT transcription factors *SNAI2* and *SOX9* and the extracellular matrix (ECM) gene *FNI* were rarely detected in P3 cells but were greatly upregulated in M3 cells (Fig. 3J). These expression patterns suggested that the program driving NB metastasis had key features of EMT. However, the program lacked other EMT hallmarks. For example, other ECM genes, including MMPs, laminins, and integrin β -1 (*ITGB1*) were universally expressed in both P3 and M3 cells, and integrin α -5 (*ITGA5*) was not detected in any malignant cells (Fig. S3K). Recently, EMT has been increasingly recognized as a continuous and variable process [46–48]. We therefore suggest that the *in vivo* program identified here reflects a partial EMT-like state. Consistent with this hypothesis, epithelial genes were downregulated and mesenchymal genes were upregulated along the P3-M3 trajectory in T62 (Fig. S3L). These results suggest that EMT occurs during NB metastasis. Indeed, immunofluorescence of primary NB tissue sections and matched lymphatic metastases from four patients showed that EPCAM-expressing NB cells were mainly detected in primary NB tumors, whereas VIM-expressing NB cells were mainly detected in their matched lymphatic

metastases in two patients (Figs. 3K and S3M). Thus, partial-EMT is involved in metastasis of some NB tumors, consistent with previous studies [49,50].

Identifying malignant programs in the ‘starter’ populations

To determine the molecular features of Cluster 3 that were distinct from those of the other clusters in T57 and T62, we ranked the differentially expressed genes in Cluster 3 compared with the other clusters by the log₂-transformed fold change values and applied gene set enrichment analysis (GSEA) to identify the differential functional gene sets (Table S4). The most enriched upregulated signaling pathways in Cluster 3 were related to the cell cycle including DNA replication, mitotic cell cycle checkpoint signaling, and mitotic sister chromatid segregation. Additionally, repair of DNA damage and telomere maintenance pathways were upregulated in Cluster 3 (Fig. 4A). Consistently, the majority of cells in Cluster 3 were found to be in the S or G2/M stage, suggesting a significantly shorter doubling time compared to other malignant cells (Fig. 4B). Genes related to cell proliferation, such as *MKI67*, *PCNA*, *TOP2A*, *CDK1*, *CCNBI*, and *MYBL2*, and genes related to DNA repair, such as *BRCA1*, *BRCA2*, *FOXMI*, were upregulated in Cluster 3 cells compared to other malignant cells (Fig. 4C). However, the well-known markers of NB metastasis, such as *HIF1A*, *LIN28B* [51], *NME1* [52], and *SDCBP* [53], were universally expressed in all malignant NB cells (Fig. S4A and S4B).

We next explored the characteristic transcription factors in Cluster 3. We inferred transcription factor (TF) activity using the DoRothEA algorithm [33] and scored the activity of each regulon using the Viper inference tool [32]. We found higher activity scores for *MYCN*, *CEBPA*, *E2F1*, *MYBL2*, *FOXMI*, and *TFDP1* in the Cluster 3 cells than in other malignant cells in T57 (Fig. 4D). Among these TFs, *E2F1*, *TFDP1*, *FOXMI*, and *MYBL2* were highly expressed in Cluster 3 and gradually downregulated along the trajectory (Fig. 4E and 4F). These genes are predominantly involved in cell proliferation, DNA repair, and stem cell development signaling pathways [54–58], consistent with our GSEA results. *MYCN* and *CEBPA* were expressed across all clusters in T57 and T62 (Fig. S4C and S4D). Consistent with this pattern, *E2F1*, *TFDP1*, *FOXMI*, and *MYBL2* also had high scores and were specifically expressed in Cluster 3 in T62 (Fig. S4E and S4F). Moreover, those TFs were mainly expressed early in pseudotime in T62 (Fig. S4G). Thus, *E2F1*, *FOXMI*, *MYBL2*, and *TFDP1* were identified as potential key TFs in ‘starter’ cells. This is consistent with the notion that, in clinical NB cases, high expression of *E2F1* [59], *FOXMI* [60], or *MYBL2* [61] is associated with poor survival of patients.

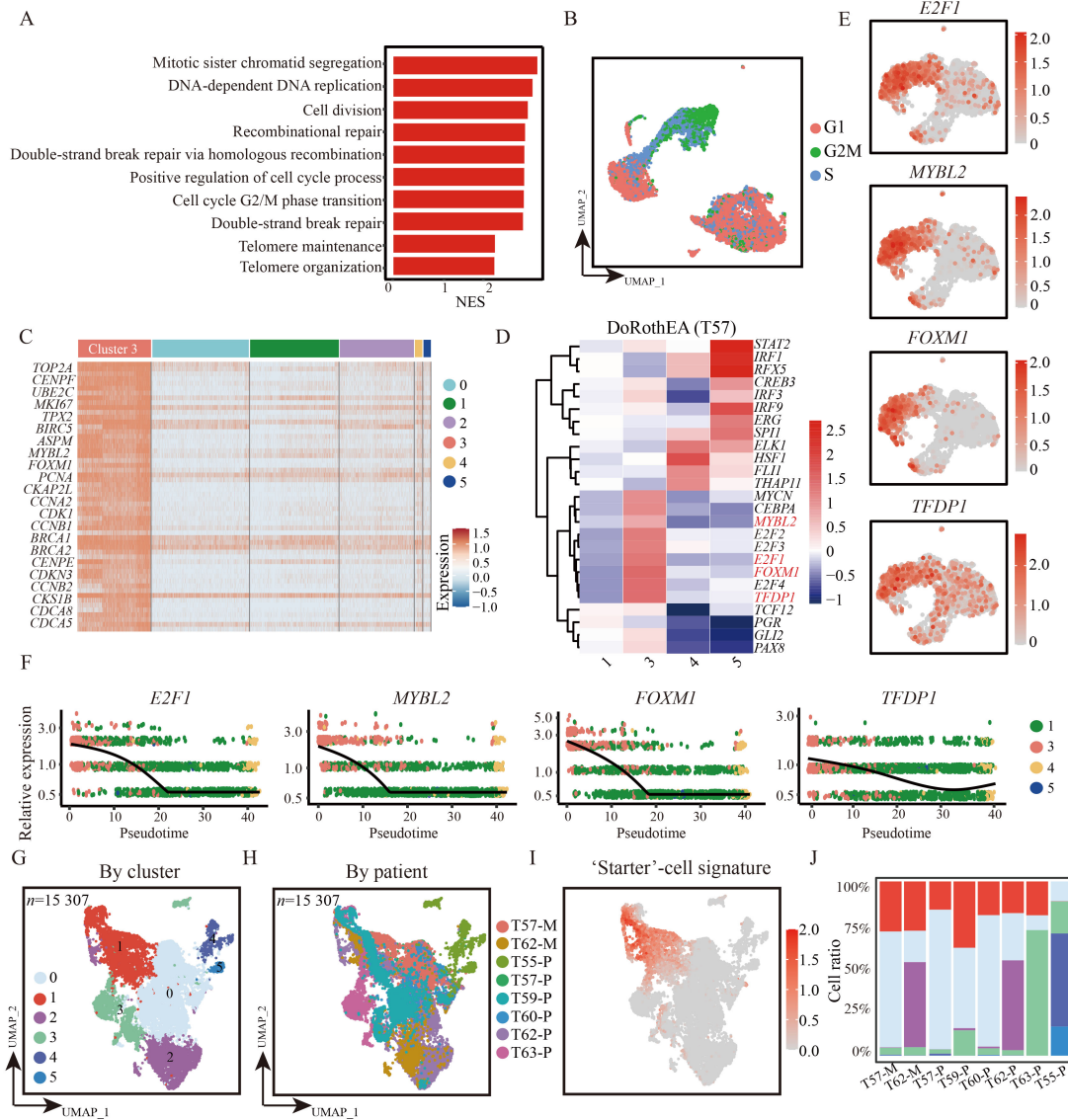


Fig. 4 Characterization of the 'starter' populations. (A) Functional enrichment analysis showing the ten enriched pathways positively correlated with cells in Cluster 3. Normalized enrichment scores (NES) were shown below. (B) UMAP plots showing the cell cycle states of the tumor cells in T57 and T62 (color code in the right). (C) Heatmap showing the scaled expression values of genes upregulated in Cluster 3 of T57 and T62. Biologically important genes are listed on the left. (D) Heatmap of the highest-scoring TFs inferred by DoRothEA in each cluster in T57. Genes colored in red are the TFs of interest. (E) UMAP plots showing the expression levels and distribution of TFs in T57 (gray to red) based on Fig. 4D. (F) Expression of the four TFs in individual cells ordered by pseudotime along the arms of the trajectory shown in Fig. 2I. (G and H) UMAP plots of 15 307 malignant NB cells from scRNA-seq data for the eight NB samples colored by the transcriptomic cluster (G), and sample origin (H). (I) UMAP plots showing the expression levels and distribution of the 'starter'-cell signature (gray to red). (J) The proportions of cell clusters within each sample.

To validate the biological characteristics of 'starter' cells across a broader range of NB tumors, we generated the molecular signature of these cells by identifying the top 150 upregulated genes based on their average log fold change ($\text{avg_log}_2\text{FC}$) using FindMarkers (Table S5). Then, we determined the expression pattern of the 'starter'-cell signature in all eight NB samples. We regrouped the malignant cells from the eight NB tumors using UMAP and identified six clusters with distinct transcriptional profiles (Fig. 4G and 4H). The 'starter'-

cell signature was preferentially expressed in Cluster 1 in NB tumors but not in the GNB tumor (Fig. 4I and 4J). In addition, cells expressing the 'starter'-cell signature were distributed in the earliest branch of the pseudotime trajectory for high-risk NBs (Fig. S4H). Consistent with this finding, analysis of the published scRNA-seq data sets [14] showed that the tumor cells expressing the 'starter'-cell signature were distributed in the earliest branch of the pseudotime trajectory for stage IV primary NB tumors with metastases (Fig. S4I–S4N). Furthermore,

we found that cells expressing ‘starter’-cell signature were also relapse-initiators in relapse tumors (Fig. S4O–S4R). These results confirmed the role of ‘starter’-cells as tumor-initiators and suggested that intratumoral development and progression in high-risk NBs are conferred by a universal transcriptional regulatory network.

Interaction between the NB microenvironment and NB cells

We next interrogated scRNA-seq profiles from malignant, immune, and stromal cells generated by Dong *et al.* [14] for characterizing the tumor-microenvironment interactions. Following clustering of the scRNA-seq data, we identified malignant cells and non-malignant cells according to the metadata in the data set (GSE137804). Based on the expression characteristics [34], we further classified the non-malignant cells into eight subtypes: T cells, B cells, mDC, monocytes, macrophages, fibroblasts, endothelium and Schwann cells. The ‘starter’ NB cells were identified according to the expression of ‘starter’-cell signature (Figs. 5A and S5A).

Then, we conducted receptor–ligand interaction analysis between the microenvironment and NB cells. In general, stronger signal intensities were detected between the microenvironment and the ‘starter’ cells compared to those between the microenvironment and other tumor cells (Fig. 5B). Among these interactive molecular pairs, TGF β and its receptor were the most intensive interactions between the ‘starter’ cells and the microenvironment both in forward (NB to microenvironment) and in backward (microenvironment to NB) directions (Fig. 5C and 5D). The forward and backward interactions mainly occurred between NB cells and T cells, fibroblasts, as well as macrophages (Fig. 5E and 5F). Further analysis revealed that the interactions were mainly mediated by TGF β 1–TGF β R1 and TGF β 1–TGF β R2 (Fig. 5G and 5H). The increased TGF β interactions from the microenvironment to the ‘starter’ cells support our finding that partial-EMT is involved in NB metastasis, because TGF β is a well-known signaling pathway to induce EMT [62]. Consistently, our scRNA-seq analysis revealed that the expression levels of *TGF β R1* and the critical downstream effectors of TGF β signaling [63], *SMAD2* and *SMAD3*, were higher in the ‘starter’ cells compared to in other NB cells (Fig. 5I), suggesting that increased TGF β signaling may serve as the causal mechanism triggering EMT. Given that TGF β signaling suppresses the anti-cancer activities of T cells and macrophages [64], the increased TGF β interactions from the ‘starter’ cells to the microenvironment suggested that the ‘starter’ cells may possess a heightened immunosuppressive function. The strong TGF β interactions between the ‘starter’ cells and the

microenvironment were also observed using another cohort [34] (Fig. S5B–S5F). Furthermore, we constructed a multilayer signaling network by integrating the intercellular pathways and intracellular sub-networks using scMLnet [65–67]. We also verified TGF β pathway between the microenvironment and the ‘starter’ cells. Moreover, the intracellular TF sub-networks showed that the expression of *TGF β I* is activated in forward direction (‘starter’ cluster to T cell and macrophages), and the activated genes were associated with EMT program and cell cycle in backward direction (microenvironment to ‘starter’ cluster) (Figs. 5J, 5K, S5G, and S5H). Overall, our results demonstrated that the TGF β signaling plays a vital role in regulating intercellular communication between microenvironment and the ‘starter’ cells, which could trigger the malignancy of the ‘starter’ cells.

‘Starter’-cell signature predicts poor prognosis

To further validate and characterize the ‘starter’-cell signature, we analyzed the scRNA-seq profiles from malignant cells generated by Dong *et al.* [14], and four bulk RNA-seq data sets (GSE62564 and GSE16476 from GEO data set, Dgc2102 and E-MTAB-8248 from the R2 database), which included extensive clinical information. The high ‘starter’-cell signature showed positive correlations with *MYCN* amplification, INSS 4 disease, *TERT* rearrangement and distant metastasis (Figs. 6A, 6B, S6A, and S6B), all of which predict poor prognosis [4,68,69]. Moreover, compared with *HIF1A*, *LIN28B*, *NME1*, and *SDCBP*, the ‘starter’-cell signature showed a greater statistical difference in the two groups (split by the distant metastasis), suggesting that the ‘starter’-cell signature may have a more predictive power of NB metastasis (Fig. S6C). Kaplan–Meier analysis showed that the high ‘starter’-cell signature group exhibited shorter event-free survival time (high vs. low group, hazard ratio = 1.970, 95% CI 1.296–2.994, log-rank test, $P = 0.0009$; high vs. low group, hazard ratio = 3.580, 95% CI 2.865–4.473, log-rank test, $P < 0.0001$; high vs. low group, hazard ratio = 3.320, 95% CI 2.478–4.448, log-rank test, $P < 0.0001$) (Figs. 6C, S6D, and S6F) and overall survival time (high vs. low group, hazard ratio = 5.405, 95% CI 2.939–9.940, log-rank test, $P < 0.0001$; high vs. low group, hazard ratio = 5.305, 95% CI 4.064–6.925, log-rank test, $P < 0.0001$; high vs. low group, hazard ratio = 7.159, 95% CI 4.868–10.53, log-rank test, $P < 0.0001$) compared to the low ‘starter’-cell signature group (Figs. 6D, S6E, and S6G). We next assessed the ability of the ‘starter’-cell signature to predict prognosis using Cox regression analysis. Similar to age [70] (younger than 15–18 months was preferable), *MYCN* amplification, *TERT* rearrangement, and stage [69], the ‘starter’-cell signature in NB samples was also a strong independent prognostic factor for a short survival

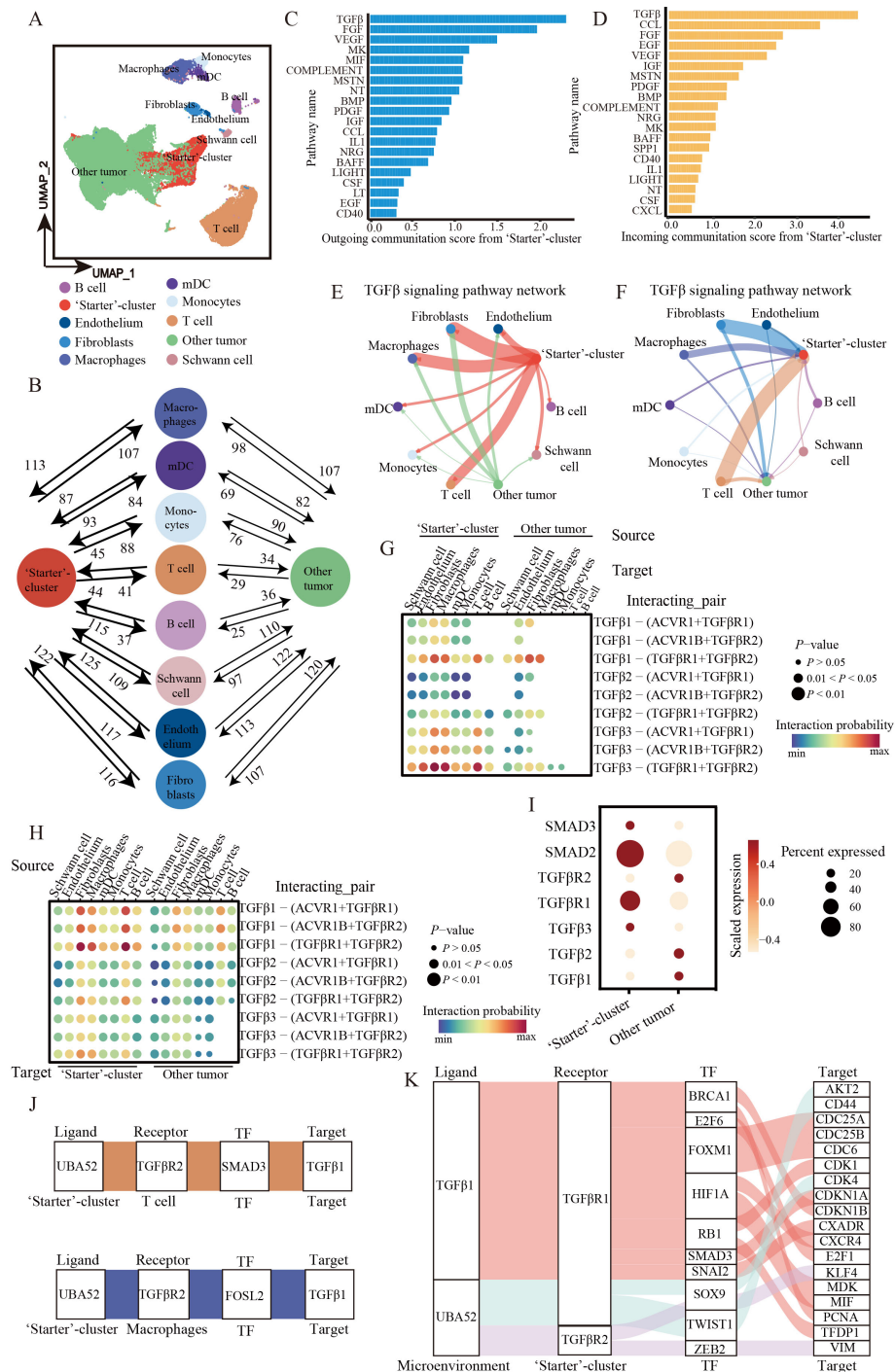


Fig. 5 Cell-cell interactions between the microenvironment and NB cells. (A) UMAP plots of scRNA-seq data from GSE137804 data set colored by cell types. (B) Schematic diagram illustrating the number of receptor-ligand interactions between the microenvironment and the 'starter' cells or between the microenvironment and the other tumor cells; the arrows indicate the direction of the interaction. (C and D) Relative contribution of each signaling pathway to the outgoing communication (C) and the incoming communication (D) network in the 'starter' cells. (E and F) The inferred TGFβ signaling networks from NB cells to the microenvironment (E) and from the microenvironment to NB cells (F). Circle sizes are proportional to the number of cells in each cell group and edge width represents the communication probability. (G and H) Dot plot showing the receptor-ligand pairs of the TGFβ pathway from NB cells to the microenvironment (G) and from the microenvironment to NB cells (H). The dot color and size represent the calculated communication probability and *P*-values. *P*-values are computed from the one-sided permutation test. (I) Dot plot showing the expression levels of representative receptor-ligand pair genes of TGFβ pathway in our scRNA-seq data from the eight NB samples. (J and K) Multilayer intercellular/intracellular TGFβ signaling network between the 'starter' cells and the microenvironment both in forward (NB to microenvironment) (J) and in backward (microenvironment to NB) (K) directions in the GSE137804 data set.

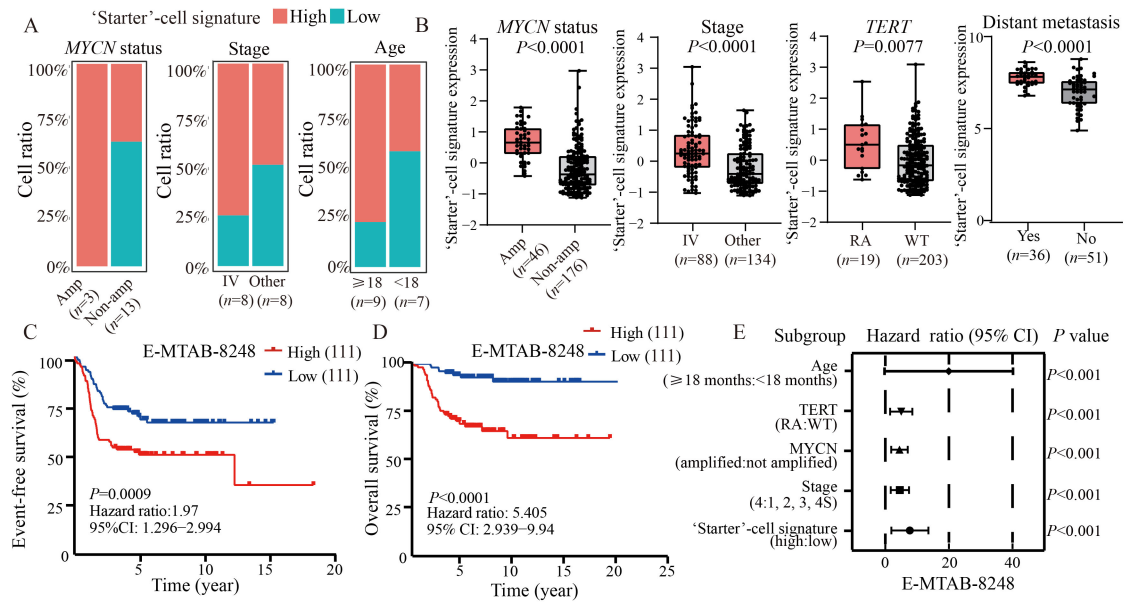


Fig. 6 Validation of the 'starter'-cell signature. (A) The proportion of high- or low-expression groups within major clinical subgroups in the 16 NB patients of previous study (GSE137804). (B) Boxplots depict the expression of the 'starter'-cell signature in patients of the R2 data set (E-MTAB-8248), categorized based on MYCN status (amplified versus non-amplified), INSS staging (stage 4 versus stages 1, 2, 3, and 4S) or TERT status (TERT rearrangement versus TERT wild-type). Boxplot of the 'starter'-cell signature expression in patients of the GSE16476 data set, categorized based on the distant metastasis. The *P* values are reported from two-sided *t*-tests. (C and D) Kaplan–Meier plots of event-free survival (C) and overall survival (D) for 222 NB patients from the previously published R2 data sets (E-MTAB-8248) based on the expression levels of the 'starter'-cell signature. Patients were divided into two groups by the median expression level of the signature ($n = 111$ and 111 tumors in the two groups). *P* values were determined using the log-rank test to compare both groups. (E) The Cox proportional regression analysis of the data set described in Fig. 6C and 6D, with the 'starter'-cell signature included.

time (HR = 6.283, 95% CI 2.788–14.161, $P < 0.001$) (Fig. 6E). Altogether, these findings suggest that the 'starter'-cell signature in NB cells promotes malignant progression and independently predicts unfavorable outcome.

Discussion

Metastasis is the major cause of death in pediatric patients with neuroblastoma, and understanding the molecular programs driving metastasis will facilitate the development of a molecular framework to efficiently target metastasis during NB progression. In this study, we generated single-cell transcriptomic data sets of human NB tumors, including two primary NBs and the corresponding bone marrow metastases. Based on these scRNA-seq data, we demonstrated the evolutionary trajectory of NB progression and metastasis and identified a subset of NB cells as the common 'starter' subset of tumor cells that initiate tumors at the primary site and metastasis. We further characterized the 'starter' cells as highly proliferative cells, consistent with previous reports that upregulation of the cell-cycle pathway is a significant predictor of unfavorable outcomes in patients with NB [14,18].

By analyzing somatic mutations called from 10x

Genomics sequencing data, we found that the bone marrow metastases contained more mutations than the majority of primary tumor cells and shared similar mutation patterns with a subset of primary tumor cells, suggesting that the metastasis-component clones arise late during tumorigenesis, which supports the linear progression model of the evolution of metastatic disease [71].

Although EMT programs have been considered potential drivers of drug resistance, invasion, and metastasis, their patterns and significance in human tumors *in vivo* remain unclear [48,72,73]. Several scRNA-seq studies identified malignant cells with epithelial features and malignant cells with mesenchymal features within the same tumor [12,74,75]. However, given that EMT is also considered to contribute to tumor initiation [76,77], whether differentiation from a mesenchymal state to an epithelial state or transdifferentiation from an epithelial state to a mesenchymal state occurred remained unknown. In this study, we observed a partial-EMT program along the route from the primary NB tumor to its bone marrow metastases. Primary NB tumors had lower expression of *TWIST1* and *VIM* than the corresponding bone marrow metastases. Concurrently, the primary NB tumors exhibited a more epithelial phenotype, while the bone

marrow metastases exhibited a more mesenchymal phenotype. These transcriptomic features suggest that a *TWIST1*-driven partial-EMT program may be involved in NB metastasis, similar to observations in carcinomas. Analysis of other scRNA-seq data sets confirmed the presence of the ‘starter’ cluster and, additionally, showed an upregulation of TGF β signaling in the ‘starter’ cells compared to the other NB cells. TGF β signaling is one of the most crucial drivers to cell invasiveness and tumor metastasis [63]. Core EMT transcription factors like *SNAI1*, *SNAI2*, *ZEB1/2*, and *TWIST1* can be directly induced by the TGF β pathway [78]. Activation of TGF β signaling promotes bone marrow metastasis of NB in Th-MYCN/Caspase8^{-/-} mouse model [7]. Thus, targeting TGF β signaling may prevent metastasis and benefit high risk NB patients.

EMT-TFs in cancer are often nonredundant and tissue specific. It has been demonstrated that *SNAI1* triggers metastasis in breast cancer [79], whereas *SNAI1* and *ZEB1* promote metastasis in pancreatic cancer [80]. We showed here that the expression of *TWIST1* gradually increased along the metastatic route, suggesting that *TWIST1* may play an important role in driving NB metastasis, consistent with a previous study [81].

The collection of paired samples during NB metastasis presents technical challenges in the clinic, as a result, the sample size in our study was relatively small and included only two samples with matched primary and metastatic tissues. However, the main discoveries were validated either by experiments in other paired NB primary and metastatic tumors or by public NB data sets. For example, we confirmed the finding that partial-EMT is involved in NB metastasis by immunostaining of four paired NB primary and lymphatic metastases, two of which showed enrichment of EPCAM-expressing NB cells in primary tumors but enrichment of VIM-expressing NB cells in the corresponding lymphatic metastases. We generated a 150-gene signature of the ‘starter’ cells, identified the presence of ‘starter’ cells in other published scRNA-seq data sets of primary NB tumors and relapse NB tumors using this signature, and confirmed their tumor-initiator function using trajectory analysis. Moreover, we confirmed the positive correlation between a high ‘starter’ cell signature and the unfavorable outcome by analyzing publicly available bulk sequencing data obtained from different platforms (GEO and R2). Overall, our study provides insights into transcriptional programs that drive NB metastasis and serves as a starting point to further understand the biology of NB metastasis.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Nos. 81825017, 81773034 to Zhe Liu, 8217113342, 81872350 to Zhenyi Ma), the Ministry of Science and

Technology of China (No. 2018YFC1313002 to Zhe Liu), the Tianjin Municipal Science and Technology Commission (No. 20JCZDJC00110 to Zhe Liu), the Interdisciplinary Research Project of Hangzhou Normal University (No. 2024JCXK03 to Zhe Liu), and the Haihe Laboratory of Cell Ecosystem Innovation Fund (No. HH22KYZX0025 to Zhe Liu).

Compliance with ethics guidelines

Conflicts of interest Zhe Nian, Dan Wang, Hao Wang, Wenxu Liu, Zhenyi Ma, Jie Yan, Yanna Cao, Jie Li, Qiang Zhao, and Zhe Liu declare that they have no conflict of interest.

This study was approved by the Institutional Ethics Committee of Tianjin Medical University Cancer Institute and Hospital, Tianjin, China (Approve number: E20210664). The study was performed in accordance with the ethical standards as laid down in the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards. Informed consent was obtained from all patients for being included in the study.

Electronic Supplementary Material Supplementary material is available in the online version of this article at <https://doi.org/10.1007/s11684-024-1081-7> and is accessible for authorized users.

References

- Zafar A, Wang W, Liu G, Wang X, Xian W, McKeon F, Foster J, Zhou J, Zhang R. Molecular targeting therapies for neuroblastoma: progress and challenges. *Med Res Rev* 2021; 41(2): 961–1021
- Furlan A, Adameyko I. Schwann cell precursor: a neural crest cell in disguise? *Dev Biol* 2018; 444(Suppl 1): S25–S35
- Brodeur GM, Iyer R, Croucher JL, Zhuang T, Higashi M, Kolla V. Therapeutic targets for neuroblastomas. *Expert Opin Ther Targets* 2014; 18(3): 277–292
- Ackermann S, Cartolano M, Hero B, Welte A, Kahlert Y, Roderwieser A, Bartenhagen C, Walter E, Gecht J, Kerschke L, Volland R, Menon R, Heuckmann JM, Gartlgruber M, Hartlieb S, Henrich KO, Okonechnikov K, Altmüller J, Nürnberg P, Lefever S, de Wilde B, Sand F, Ikram F, Rosswog C, Fischer J, Theissen J, Hertwig F, Singhi AD, Simon T, Vogel W, Perner S, Krug B, Schmidt M, Rahmann S, Achter V, Lang U, Vokuhl C, Ortman M, Büttner R, Eggert A, Speleman F, O’Sullivan RJ, Thomas RK, Berthold F, Vandesompele J, Schramm A, Westermann F, Schulte JH, Peifer M, Fischer M. A mechanistic classification of clinical phenotypes in neuroblastoma. *Science* 2018; 362(6419): 1165–1170
- Kholodenko IV, Kalinovskiy DV, Doronin II, Deyev SM, Kholodenko RV. Neuroblastoma origin and therapeutic targets for immunotherapy. *J Immunol Res* 2018; 2018: 7394268
- Matthay KK, Maris JM, Schleiermacher G, Nakagawara A, Mackall CL, Diller L, Weiss WA. Neuroblastoma. *Nat Rev Dis Primers* 2016; 2(1): 16078
- Teitz T, Inoue M, Valentine MB, Zhu K, Rehg JE, Zhao W, Finkelstein D, Wang YD, Johnson MD, Calabrese C, Rubinstein M, Hakem R, Weiss WA, Lahti JM. Th-MYCN mice with caspase-

- 8 deficiency develop advanced neuroblastoma with bone marrow metastasis. *Cancer Res* 2013; 73(13): 4086–4097
8. Zheng T, Ménard M, Weiss WA. Neuroblastoma metastases: leveraging the avian neural crest. *Cancer Cell* 2017; 32(4): 395–397
 9. Zhu S, Zhang X, Weichert-Leahey N, Dong Z, Zhang C, Lopez G, Tao T, He S, Wood AC, Oldridge D, Ung CY, van Ree JH, Khan A, Salazar BM, Lummertz da Rocha E, Zimmerman MW, Guo F, Cao H, Hou X, Weroha SJ, Perez-Atayde AR, Neuberg DS, Meves A, McNiven MA, van Deursen JM, Li H, Maris JM, Look AT. LMO1 synergizes with MYCN to promote neuroblastoma initiation and metastasis. *Cancer Cell* 2017; 32(3): 310–323.e5
 10. Massagué J, Obenauf AC. Metastatic colonization by circulating tumour cells. *Nature* 2016; 529(7586): 298–306
 11. Peng J, Sun BF, Chen CY, Zhou JY, Chen YS, Chen H, Liu L, Huang D, Jiang J, Cui GS, Yang Y, Wang W, Guo D, Dai M, Guo J, Zhang T, Liao Q, Liu Y, Zhao YL, Han DL, Zhao Y, Yang YG, Wu W. Single-cell RNA-seq highlights intra-tumoral heterogeneity and malignant progression in pancreatic ductal adenocarcinoma. *Cell Res* 2019; 29(9): 725–738
 12. Puram SV, Tirosh I, Parkh AS, Patel AP, Yizhak K, Gillespie S, Rodman C, Luo CL, Mroz EA, Emerick KS, Deschler DG, Varvares MA, Mylvaganam R, Rozenblatt-Rosen O, Rocco JW, Faquin WC, Lin DT, Regev A, Bernstein BE. Single-cell transcriptomic analysis of primary and metastatic tumor ecosystems in head and neck cancer. *Cell* 2017; 171(7): 1611–1624.e24
 13. Kan T, Zhang S, Zhou S, Zhang Y, Zhao Y, Gao Y, Zhang T, Gao F, Wang X, Zhao L, Yang M. Single-cell RNA-seq recognized the initiator of epithelial ovarian cancer recurrence. *Oncogene* 2022; 41(6): 895–906
 14. Dong R, Yang R, Zhan Y, Lai HD, Ye CJ, Yao XY, Luo WQ, Cheng XM, Miao JJ, Wang JF, Liu BH, Liu XQ, Xie LL, Li Y, Zhang M, Chen L, Song WC, Qian W, Gao WQ, Tang YH, Shen CY, Jiang W, Chen G, Yao W, Dong KR, Xiao XM, Zheng S, Li K, Wang J. Single-cell characterization of malignant phenotypes and developmental trajectories of adrenal neuroblastoma. *Cancer Cell* 2020; 38(5): 716–733.e6
 15. Jansky S, Sharma AK, Körber V, Quintero A, Toprak UH, Wecht EM, Gartlgruber M, Greco A, Chomsky E, Grünewald TGP, Henrich KO, Tanay A, Herrmann C, Höfer T, Westermann F. Single-cell transcriptomic analyses provide insights into the developmental origins of neuroblastoma. *Nat Genet* 2021; 53(5): 683–693
 16. Kildisiute G, Kholosy WM, Young MD, Roberts K, Elmentaite R, van Hooff SR, Pacyna CN, Khabirova E, Piapi A, Thevanesan C, Bugallo-Blanco E, Burke C, Mamanova L, Keller KM, Langenberg-Ververgaert KPS, Lijnzaad P, Margaritis T, Holstege FCP, Tas ML, Wijnen M, van Noesel MM, Del Valle I, Barone G, van der Linden R, Duncan C, Anderson J, Achermann JC, Haniffa M, Teichmann SA, Rampling D, Sebire NJ, He X, de Krijger RR, Barker RA, Meyer KB, Bayraktar O, Straathof K, Molenaar JJ, Behjati S. Tumor to normal single-cell mRNA comparisons reveal a pan-neuroblastoma cancer cell. *Sci Adv* 2021; 7(6): eabd3311
 17. Hanemaaijer ES, Margaritis T, Sanders K, Bos FL, Candelli T, Al-Saati H, van Noesel MM, Meyer-Wentrup FAG, van de Wetering M, Holstege FCP, Clevers H. Single-cell atlas of developing murine adrenal gland reveals relation of Schwann cell precursor signature to neuroblastoma phenotype. *Proc Natl Acad Sci USA* 2021; 118(5): e2022350118
 18. Bedoya-Reina OC, Li W, Arceo M, Plescher M, Bullova P, Pui H, Kaucka M, Kharchenko P, Martinsson T, Holmberg J, Adameyko I, Deng Q, Larsson C, Juhlin CC, Kogner P, Schlisio S. Single-nuclei transcriptomes from human adrenal gland reveal distinct cellular identities of low and high-risk neuroblastoma tumors. *Nat Commun* 2021; 12(1): 5309
 19. Kameneva P, Artemov AV, Kastriti ME, Faure L, Olsen TK, Otte J, Erickson A, Semsch B, Andersson ER, Ratz M, Frisén J, Tischler AS, de Krijger RR, Boudierlique T, Akkuratova N, Vorontsova M, Gusev O, Fried K, Sundström E, Mei S, Kogner P, Baryawno N, Kharchenko PV, Adameyko I. Single-cell transcriptomics of human embryos identifies multiple sympathoblast lineages with potential implications for neuroblastoma origin. *Nat Genet* 2021; 53(5): 694–706
 20. Yuan X, Seneviratne JA, Du S, Xu Y, Chen Y, Jin Q, Jin X, Balachandran A, Huang S, Xu Y, Zhai Y, Lu L, Tang M, Dong Y, Cheung BB, Marshall GM, Shi W, Carter DR, Zhang C. Single-cell profiling of peripheral neuroblastic tumors identifies an aggressive transitional state that bridges an adrenergic-mesenchymal trajectory. *Cell Rep* 2022; 41(1): 111455
 21. Fetahu IS, Esser-Skala W, Dnyansagar R, Sindelar S, Rifatbegovic F, Bileck A, Skos L, Bozsaky E, Lazic D, Shaw L, Tötzl M, Tarlungeanu D, Bernkopf M, Rados M, Weninger W, Tomazou EM, Bock C, Gerner C, Ladenstein R, Farlik M, Fortelny N, Taschner-Mandl S. Single-cell transcriptomics and epigenomics unravel the role of monocytes in neuroblastoma bone marrow metastasis. *Nat Commun* 2023; 14(1): 3620
 22. McGinnis CS, Murrow LM, Gartner ZJ. DoubletFinder: doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. *Cell Syst* 2019; 8(4): 329–337.e4
 23. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, Baglaenko Y, Brenner M, Loh PR, Raychaudhuri S. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods* 2019; 16(12): 1289–1296
 24. Tirosh I, Venteicher AS, Hebert C, Escalante LE, Patel AP, Yizhak K, Fisher JM, Rodman C, Mount C, Filbin MG, Neftel C, Desai N, Nyman J, Izar B, Luo CC, Francis JM, Patel AA, Onozato ML, Riggi N, Livak KJ, Gennert D, Satija R, Nahed BV, Curry WT, Martuza RL, Mylvaganam R, Iafrate AJ, Frosch MP, Golub TR, Rivera MN, Getz G, Rozenblatt-Rosen O, Cahill DP, Monje M, Bernstein BE, Louis DN, Regev A, Suvà ML. Single-cell RNA-seq supports a developmental hierarchy in human oligodendroglioma. *Nature* 2016; 539(7628): 309–313
 25. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 2018; 34(18): 3094–3100
 26. Sedlazeck FJ, Rescheneder P, Smolka M, Fang H, Nattestad M, von Haeseler A, Schatz MC. Accurate detection of complex structural variations using single-molecule sequencing. *Nat Methods* 2018; 15(6): 461–468
 27. Aran D, Looney AP, Liu L, Wu E, Fong V, Hsu A, Chak S, Naikawadi RP, Wolters PJ, Abate AR, Butte AJ, Bhattacharya M. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat Immunol* 2019; 20(2): 163–172
 28. Zhang Q, He Y, Luo N, Patel SJ, Han Y, Gao R, Modak M, Carotta S, Haslinger C, Kind D, Peet GW, Zhong G, Lu S, Zhu W,

- Mao Y, Xiao M, Bergmann M, Hu X, Kerkar SP, Vogt AB, Pflanz S, Liu K, Peng J, Ren X, Zhang Z. Landscape and dynamics of single immune cells in hepatocellular carcinoma. *Cell* 2019; 179(4): 829–845.e20
29. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, Lennon NJ, Livak KJ, Mikkelsen TS, Rinn JL. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol* 2014; 32(4): 381–386
30. Street K, Risso D, Fletcher RB, Das D, Ngai J, Yosef N, Purdom E, Dudoit S. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* 2018; 19(1): 477
31. Garcia-Alonso L, Holland CH, Ibrahim MM, Turei D, Saez-Rodriguez J. Benchmark and integration of resources for the estimation of human transcription factor activities. *Genome Res* 2019; 29(8): 1363–1375
32. Alvarez MJ, Shen Y, Giorgi FM, Lachmann A, Ding BB, Ye BH, Califano A. Functional characterization of somatic mutations in cancer using network-based inference of protein activity. *Nat Genet* 2016; 48(8): 838–847
33. Holland CH, Tanevski J, Perales-Patón J, Gleixner J, Kumar MP, Mereu E, Joughin BA, Stegle O, Lauffenburger DA, Heyn H, Szalai B, Saez-Rodriguez J. Robustness and applicability of transcription factor and pathway analysis tools on single-cell RNA-seq data. *Genome Biol* 2020; 21(1): 36
34. Verhoeven BM, Mei S, Olsen TK, Gustafsson K, Valind A, Lindström A, Gisselsson D, Fard SS, Hagerling C, Kharchenko PV, Kogner P, Johnsen JL, Baryawno N. The immune cell atlas of human neuroblastoma. *Cell Rep Med* 2022; 3(6): 100657
35. Jin S, Guerrero-Juarez CF, Zhang L, Chang I, Ramos R, Kuan CH, Myung P, Plikus MV, Nie Q. Inference and analysis of cell-cell communication using CellChat. *Nat Commun* 2021; 12(1): 1088
36. Zhang J, Guan M, Wang Q, Zhang J, Zhou T, Sun X. Single-cell transcriptome-based multilayer network biomarker for predicting prognosis and therapeutic response of gliomas. *Brief Bioinform* 2020; 21(3): 1080–1097
37. Patel AP, Tirosh I, Trombetta JJ, Shalek AK, Gillespie SM, Wakimoto H, Cahill DP, Nahed BV, Curry WT, Martuza RL, Louis DN, Rozenblatt-Rosen O, Suvà ML, Regev A, Bernstein BE. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* 2014; 344(6190): 1396–1401
38. Brodeur GM. Neuroblastoma: biological insights into a clinical enigma. *Nat Rev Cancer* 2003; 3(3): 203–216
39. Pan XW, Zhang H, Xu D, Chen JX, Chen WJ, Gan SS, Qu FJ, Chu CM, Cao JW, Fan YH, Song X, Ye JQ, Zhou W, Cui XG. Identification of a novel cancer stem cell subpopulation that promotes progression of human fatal renal cell carcinoma by single-cell RNA-seq analysis. *Int J Biol Sci* 2020; 16(16): 3149–3162
40. La Manno G, Soldatov R, Zeisel A, Braun E, Hochgerner H, Petukhov V, Lidschreiber K, Kastrioti ME, Lönnberg P, Furlan A, Fan J, Borm LE, Liu Z, van Bruggen D, Guo J, He X, Barker R, Sundström E, Castelo-Branco G, Cramer P, Adameyko I, Linnarsson S, Kharchenko PV. RNA velocity of single cells. *Nature* 2018; 560(7719): 494–498
41. Spanjaard B, Hu B, Mitic N, Olivares-Chauvet P, Janjuha S, Ninov N, Junker JP. Simultaneous lineage tracing and cell-type identification using CRISPR-Cas9-induced genetic scars. *Nat Biotechnol* 2018; 36(5): 469–473
42. Kester L, van Oudenaarden A. Single-cell transcriptomics meets lineage tracing. *Cell Stem Cell* 2018; 23(2): 166–179
43. Lee-Six H, Øbro NF, Shepherd MS, Grossmann S, Dawson K, Belmonte M, Osborne RJ, Huntly BJP, Martincorena I, Anderson E, O'Neill L, Stratton MR, Laurenti E, Green AR, Kent DG, Campbell PJ. Population dynamics of normal human blood inferred from somatic mutations. *Nature* 2018; 561(7724): 473–478
44. Miller TE, Lareau CA, Verga JA, DePasquale EAK, Liu V, Ssozi D, Sandor K, Yin Y, Ludwig LS, El Farran CA, Morgan DM, Satpathy AT, Griffin GK, Lane AA, Love JC, Bernstein BE, Sankaran VG, van Galen P. Mitochondrial variant enrichment from high-throughput single-cell RNA sequencing resolves clonal populations. *Nat Biotechnol* 2022; 40(7): 1030–1034
45. Nguyen B, Fong C, Luthra A, Smith SA, DiNatale RG, Nandakumar S, Walch H, Chatila WK, Madupuri R, Kundra R, Bielski CM, Mastrogiacomo B, Donoghue MTA, Boire A, Chandarlapaty S, Ganesh K, Harding JJ, Iacobuzio-Donahue CA, Razavi P, Reznik E, Rudin CM, Zamarin D, Abida W, Abou-Alfa GK, Aghajanian C, Cercek A, Chi P, Feldman D, Ho AL, Iyer G, Janjigian YY, Morris M, Motzer RJ, O'Reilly EM, Postow MA, Raj NP, Riely GJ, Robson ME, Rosenberg JE, Safonov A, Shoushtari AN, Tap W, Teo MY, Varghese AM, Voss M, Yaeger R, Zauderer MG, Abu-Rustum N, Garcia-Aguilar J, Bochner B, Hakimi A, Jarnagin WR, Jones DR, Molena D, Morris L, Rios-Doria E, Russo P, Singer S, Strong VE, Chakravarty D, Ellenson LH, Gopalan A, Reis-Filho JS, Weigelt B, Ladanyi M, Gonen M, Shah SP, Massague J, Gao J, Zehir A, Berger MF, Solit DB, Bakhom SF, Sanchez-Vega F, Schultz N. Genomic characterization of metastatic patterns from prospective clinical sequencing of 25,000 patients. *Cell* 2022; 185(3): 563–575.e11
46. Lambert AW, Pattabiraman DR, Weinberg RA. Emerging biological principles of metastasis. *Cell* 2017; 168(4): 670–691
47. Lundgren K, Nordenskjöld B, Landberg G. Hypoxia, Snail and incomplete epithelial-mesenchymal transition in breast cancer. *Br J Cancer* 2009; 101(10): 1769–1781
48. Nieto MA, Huang RY, Jackson RA, Thiery JP. EMT: 2016. *Cell* 2016; 166(1): 21–45
49. Li M, Sun C, Bu X, Que Y, Zhang L, Zhang Y, Zhang L, Lu S, Huang J, Zhu J, Wang J, Sun F, Zhang Y. ISL1 promoted tumorigenesis and EMT via Aurora kinase A-induced activation of PI3K/AKT signaling pathway in neuroblastoma. *Cell Death Dis* 2021; 12(6): 620
50. Ferronha T, Rabadán MA, Gil-Guiñón E, Le Dréau G, de Torres C, Martí E. LMO4 is an essential cofactor in the Snail2-mediated epithelial-to-mesenchymal transition of neuroblastoma and neural crest cells. *J Neurosci* 2013; 33(7): 2773–2783
51. Missios P, da Rocha EL, Pearson DS, Philipp J, Aleman MM, Pirouz M, Farache D, Franses JW, Kubaczka C, Tsanov KM, Jha DK, Pepe-Mooney B, Powers JT, Gregory RI, Lee AS, Dominguez D, Ting DT, Daley GQ. LIN28B alters ribosomal dynamics to promote metastasis in MYCN-driven malignancy. *J Clin Invest* 2021; 131(22): e145142
52. Tan CY, Chang CL. NDPKA is not just a metastasis suppressor—be aware of its metastasis-promoting role in neuroblastoma. *Lab Invest* 2018; 98(2): 219–227
53. Das SK, Maji S, Wechman SL, Bhoopathi P, Pradhan AK, Talukdar S, Sarkar D, Landry J, Guo C, Wang XY, Cavenee WK,

- Emdad L, Fisher PB. MDA-9/Syntenin (SDCBP): novel gene and therapeutic target for cancer metastasis. *Pharmacol Res* 2020; 155: 104695
54. Kent LN, Leone G. The broken cycle: E2F dysfunction in cancer. *Nat Rev Cancer* 2019; 19(6): 326–338
55. Barger CJ, Branick C, Chee L, Karpf AR. Pan-cancer analyses reveal genomic features of FOXM1 overexpression in cancer. *Cancers (Basel)* 2019; 11(2): 251
56. Musa J, Aynaud MM, Mirabeau O, Delattre O, Grünwald TG. MYBL2 (B-Myb): a central regulator of cell proliferation, cell survival and differentiation involved in tumorigenesis. *Cell Death Dis* 2017; 8(6): e2895
57. Polager S, Ginsberg D. E2F—at the crossroads of life and death. *Trends Cell Biol* 2008; 18(11): 528–535
58. Bandara LR, Lam EW, Sørensen TS, Zamanian M, Girling R, La Thangue NB. DP-1: a cell cycle-regulated and phosphorylated component of transcription factor DRTF1/E2F which is functionally important for recognition by pRb and the adenovirus E4 orf 6/7 protein. *EMBO J* 1994; 13(13): 3104–3114
59. Fang E, Wang X, Yang F, Hu A, Wang J, Li D, Song H, Hong M, Guo Y, Liu Y, Li H, Huang K, Zheng L, Tong Q. Therapeutic targeting of MZF1-AS1/PARP1/E2F1 axis inhibits proline synthesis and neuroblastoma progression. *Adv Sci (Weinh)* 2019; 6(19): 1900581
60. Wang Z, Park HJ, Carr JR, Chen YJ, Zheng Y, Li J, Tyner AL, Costa RH, Bagchi S, Raychaudhuri P. FoxM1 in tumorigenicity of the neuroblastoma cells and renewal of the neural progenitors. *Cancer Res* 2011; 71(12): 4292–4302
61. Raschella G, Cesi V, Amendola R, Negroni A, Tanno B, Altavista P, Tonini GP, De Bernardi B, Calabretta B. Expression of B-myb in neuroblastoma tumors is a poor prognostic factor independent from MYCN amplification. *Cancer Res* 1999; 59(14): 3365–3368
62. Xu J, Lamouille S, Derynck R. TGF-beta-induced epithelial to mesenchymal transition. *Cell Res* 2009; 19(2): 156–172
63. Padua D, Massagué J. Roles of TGFbeta in metastasis. *Cell Res* 2009; 19(1): 89–102
64. Derynck R, Turley SJ, Akhurst RJ. TGFβ biology in cancer progression and immunotherapy. *Nat Rev Clin Oncol* 2021; 18(1): 9–34
65. Cheng J, Zhang J, Wu Z, Sun X. Inferring microenvironmental regulation of gene expression from single-cell RNA sequencing data using scMLnet with an application to COVID-19. *Brief Bioinform* 2021; 22(2): 988–1005
66. Luo J, Deng M, Zhang X, Sun X. ESICCC as a systematic computational framework for evaluation, selection, and integration of cell-cell communication inference methods. *Genome Res* 2023; 33(10): 1788–1805
67. He X, Sun X, Shao Y. Multicellular network-informed survival model for identification of drug targets of gliomas. *IEEE J Biomed Health Inform* 2023; [Epub ahead of print] doi: 10.1109/JBHI.2023.3309825
68. Huang M, Weiss WA. Neuroblastoma and MYCN. *Cold Spring Harb Perspect Med* 2013; 3(10): a014415
69. Brodeur GM, Pritchard J, Berthold F, Carlsen NL, Castel V, Castelberry RP, De Bernardi B, Evans AE, Favrot M, Hedborg F. Revisions of the international criteria for neuroblastoma diagnosis, staging, and response to treatment. *J Clin Oncol* 1993; 11(8): 1466–1477
70. London WB, Castleberry RP, Matthay KK, Look AT, Seeger RC, Shimada H, Thorner P, Brodeur G, Maris JM, Reynolds CP, Cohn SL. Evidence for an age cutoff greater than 365 days for neuroblastoma risk group stratification in the Children’s Oncology Group. *J Clin Oncol* 2005; 23(27): 6459–6465
71. Turajlic S, Swanton C. Metastasis as an evolutionary process. *Science* 2016; 352(6282): 169–175
72. Thiery JP, Acloque H, Huang RY, Nieto MA. Epithelial-mesenchymal transitions in development and disease. *Cell* 2009; 139(5): 871–890
73. Ye X, Weinberg RA. Epithelial-mesenchymal plasticity: a central regulator of cancer progression. *Trends Cell Biol* 2015; 25(11): 675–686
74. Zhang Q, Fei L, Han R, Huang R, Wang Y, Chen H, Yao B, Qiao N, Wang Z, Ma Z, Ye Z, Zhang Y, Wang W, Wang Y, Kong L, Shou X, Cao X, Zhou X, Shen M, Cheng H, Yao Z, Zhang C, Guo G, Zhao Y. Single-cell transcriptome reveals cellular hierarchies and guides p-EMT-targeted trial in skull base chordoma. *Cell Discov* 2022; 8(1): 94
75. Grasset EM, Dunworth M, Sharma G, Loth M, Tandurella J, Cimino-Mathews A, Gentz M, Bracht S, Haynes M, Fertig EJ, Ewald AJ. Triple-negative breast cancer metastasis involves complex epithelial-mesenchymal transition dynamics and requires vimentin. *Sci Transl Med* 2022; 14(656): eabn7571
76. Pastushenko I, Blanpain C. EMT transition states during tumor progression and metastasis. *Trends Cell Biol* 2019; 29(3): 212–226
77. Krebs AM, Mitschke J, Lasierra Losada M, Schmalhofer O, Boerries M, Busch H, Boettcher M, Mougiakakos D, Reichardt W, Bronsert P, Brunton VG, Pilarsky C, Winkler TH, Brabletz S, Stemmler MP, Brabletz T. The EMT-activator Zeb1 is a key factor for cell plasticity and promotes metastasis in pancreatic cancer. *Nat Cell Biol* 2017; 19(5): 518–529
78. Hao Y, Baker D, Ten Dijke P. TGF-β-mediated epithelial-mesenchymal transition and cancer metastasis. *Int J Mol Sci* 2019; 20(11): 2767
79. Zhang Y, Zou X, Qian W, Weng X, Zhang L, Zhang L, Wang S, Cao X, Ma L, Wei G, Wu Y, Hou Z. Enhanced PAPSS2/VCAN sulfation axis is essential for Snail-mediated breast cancer cell migration and metastasis. *Cell Death Differ* 2019; 26(3): 565–579
80. Carstens JL, Yang S, Correa de Sampaio P, Zheng X, Barua S, McAndrews KM, Rao A, Burks JK, Rhim AD, Kalluri R. Stabilized epithelial phenotype of cancer cells in primary tumors leads to increased colonization of liver metastasis in pancreatic cancer. *Cell Rep* 2021; 35(2): 108990
81. Sepporta MV, Praz V, Balmes Bourlout K, Joseph JM, Jauquier N, Riggi N, Nardou-Auderset K, Petit A, Scoazec JY, Sartelet H, Renella R, Mühlethaler-Mottet A. TWIST1 expression is associated with high-risk neuroblastoma and promotes primary and metastatic tumor growth. *Commun Biol* 2022; 5(1): 42