Single cell transcriptional profiling identifies heterogeneity in pleuropulmonary blastoma

Zhi-Xue Chen[1], Wen-Bo Zhang[1], Chun-Jing Ye[1], Yong Zhan[1], Ran Yang[1], Jia Liu[1], Yi Li[1], Bo Hong[1], Jia Wang[2], Wei-Qiang Tan[1], Xiao Shen[1], Ming Ye[1], and Rui Dong[1]

[1]Children's Hospital of Fudan University
[2]Shanghai Jiao Tong University

June 4, 2020

Abstract

Background: Pleuropulmonary blastoma (PPB) is a rare lung malignancy in children derived from mesenchyme. Single-cell RNA sequencing is used to explore the heterogeneity of tumors. Here, we used this technology to classify PPB subpopulations and predict their differentiation trajectory. Procedure: This study included 10 007 single cells from a girl with PPB. After choosing malignant tumor cells with an inferring copy number variation, we used non-negative matrix factorization and Seurat analysis to cluster the cells and divided the subgroups by their differentially expressed genes and Gene Ontology enrichment analysis. Additionally, pseudotime trajectory analysis of PPB was conducted with Monocle. Results: Tumor cells were divided into two major categories including muscle (DEShiTNNT1hiTNNI1hi) and cartilage (TWIST1hiHTRA1hiBMP4hi). In muscle lineage, satellite myogenic cells (PAX7hiMYF5hiMSChi), which were subdivided into more primitive FABP7hi one and later ITM2Ahi one, were the origin of muscle in PPB and gradually differentiated into myocytes (MYOGhiTTNhiMYL4hi). And in the cartilage lineage, SOX9hiPAX1hiPAX9hi prechondrocytes generated MGPhiOGNhi chondrocytes and DKK2hiTNMDhi chondrocytes. Conclusions: Our results demonstrated unrecognized heterogeneity in PPB at the single cell level and defined its muscle and cartilage subpopulations together with their trajectories, which would provide some basis for the molecular mechanism and treatment of PPB.

Introduction

Pleuropulmonary blastoma (PPB) is a rare ($<2/1\,000\,000$) primary pulmonary malignancy with a poor prognosis in children[1], which is a pulmonary soft tissue sarcoma. It is generally believed that PPB originates from mesenchyme and might be related to pulmonary or pleural development[2]. PPB is related to *DICER1* gene mutations (mostly biallelic*DICER1* mutations, nonsense mutations, and code shift mutations)[3] accompanied by overexpression of Neural cell adhesion molecule (NCAM) 1 and Fibroblast growth factor (FGF) 9[4] relevant to fetal lung development[5]. However, the cellular components of this tumor and the interactions between the various cell types are unknown.

Single-cell RNA sequencing (scRNA-seq) is a sequencing technology that obtains genetic information of a single cell. Compared with traditional technologies conducted at the multi-cell level, scRNA-seq has a higher resolution that determines precise gene expression patterns of thousands of single cells, putting more emphasis on the heterogeneity of genetic information. This technology is used in tumor research and facilitates the isolation and definition of cell subpopulations. Moreover, the differentiation of histologically similar cells, evolution of the tumor, and relationship between the tumor and its microenvironment can be well understood and explored using scRNA-seq[6,7].

Although some progress has been made in the study of gene mutations and pathological changes of PPB, the heterogeneity of the tumor and its specific developmental process remain unclear. Here, we applied a

PPB sample from the pediatric hospital affiliated to Fudan University to single cell sequencing to define the approximate landscape of this tumor and explore the differentiation trajectory of PPB.

Methods

Patient

The sample was obtained from one girl with stage III pleuropulmonary blastoma (PPB). Imaging data (CT and MRI) and pathology reports confirmed the clinical findings. The clinical information was shown in detail: girl, 46 months, III pathological stage with scattered polygonal neoplastic cells. The patient's privacy was protected and the methods of the study were conducted in accordance with relevant guidelines and rules.

Acquisition of single cells

Tumor tissue obtained by surgery was enriched by Percoll-based gradient centrifugation. The specific experimental steps were as follows. Fresh tumor samples were fragmented and digested with collagenase IV (Gibco) and DNase I (Sigma) to remove elastase at 37 °C for 30 minutes with stirring. After filtration, washing, and centrifugation, cell fragments and red blood cells in the single cell suspension were removed with Human lymphocyte separation fluid (CL5020; Cedarlane). The obtained single cells were placed in PBS with 1% BSA. The viability and cell counts were determined using a Countess II FL instrument.

Single cell RNA sequencing

These single cells were loaded on a Chromium Controller instrument to generate single cell gel bead-in emulsions. As the single cells lysed, the gel beads automatically dissolved and released a large number of barcode sequences, followed by reverse transcription of the mRNA to generate cDNA with barcode and UMI information[8,9]. The sequencing library was constructed after quality inspection and amplification of cDNA, and then sequenced on an Illumina NovaSeq 6000 at a depth of 400 M reads. Using Cell ranger software (10X Genomics, version 2.1.1), the data of raw basecall files were split, counted, combined mathematically, and reanalyzed. Cells with less than 10% UMI in whole genes were filtered out.

Seurat analysis

For basic data visualization, a log-transformed gene matrix was imported into R Studio and analyzed by the Seurat package (http://www.satijalab.org/seurat). Cells with fewer than 500 detected genes and greater than 10% mitochondrial gene expression were removed. Moreover, we eliminated genes detected (UMI count >0) in less than three cells and deleted mitochondrial and ribosomal genes. The remaining genes were reduced in dimension by principle component analysis. After quality control and normalization, we used T-distributed Stochastic Neighbor Embedding (t-SNE) to visualize the characteristics of the continuity of subsets. Using other functions of Seurat, including FindVariableFeatures, FindAllMarkers, and Doheatmap, every cluster of collected cells was user-redefined for further analysis.

Comparison and definition of clusters

To distinguish tumor cells, we used inferring copy number variation (CNV)[10,11] to compare our PPB data with normal fibroblasts whose expression profiles were obtained from the open access repository of the Gene expression omnibus (Gse 128169)[12]. To further analyze tumor heterogeneity, the obtained clusters were identified by cluster-specific marker genes. Gene Ontology enrichment analysis (GO analysis) was performed with the ClusterProfiler package (https://bioconductor.org/packages/clusterProfiler) for functional comparison of similar clusters. Moreover, secondary sparse-nonnegative matrix factorization (NMF) analysis was applied to PPB to improve identification of both broad and rare cell clusters[13].

Pseudotime trajectory analysis with monocle

Using Monocle 2 (http://cole-trapnell-lab.github.io/monocle-release) with differential genes of each clusters, we reshaped the process of cellular changes over time by constructing trajectories between tumor cells. After exclusion of non-tumor cells, the remaining cells were reduced dimensionality using the DDRTree method, sequenced in pseudotime, and finally visualized.

Results

Droplet-mediated scRNA-seq reveals cancer and non-cancer components of PPB

To fully understand intra-tumoral and inter-tumor heterogeneity of pleuropulmonary blastoma (PPB), deep transcriptional profiles of the cancer cell status and gene expression were obtained by droplet-mediated scRNA-seq platform (10× Genomics Chromium)[14] to observe whether cell diversity existed (Fig.1A). We sequenced 10 007 cells from the individual at an average of 400 M reads. After quality control filtering to remove cells with low gene detection (<500 genes) and high mitochondrial gene coverage (>10%) with the Seurat pipeline, we obtained a rough cell atlas of this sample, which was clustered into 14 clusters (Fig.1D). This whole specimen had several classes of cells, including normal immune cells (cluster 12, $PTPRC^{hi}$), fibroblasts (cluster 13, $TAGLN^{hi}$), endothelial cells (cluster 14, $CLDN5^{hi}$), and tumor cells (cluster 0-11,$NCAM1^{hi}DICER1^{hi}IGF1R^{hi}$)[4,15] (Supporting Information Figure S1) with differentially expressed genes (DEGs). To further verify such cell classification,$TAGLN^{hi}$ normal fibroblasts from a healthy adult lung were used as a reference to analyze the inferring copy number variation (inferCNV) of all single cells in PPB, because fibroblasts are derived from mesenchyme. We found coherent chromosomal aberrations in tumor clusters (clusters 0–11) but not in the other defined non-cancer cells (clusters 12–14). Tumor clusters contained chromosome 19 gain and chromosome 1, 5, 6, and 19 loss (Fig.1C)[15]. Interestingly, there were both distinct gain and loss in chromosome 19, which were probably closely connected to tumorigenesis of the chromosome variation.

Intratumoral heterogeneity of tumor cells

We divided the tumor cluster into two cell subtypes of muscle and cartilage based on pathological characteristics of the sample and DEGs of each cell clustered by Seurat (mentioned more specifically in the next parts). To further explore its intratumoral heterogeneity, we used nonnegative matrix factorization (NMF) to rebuild 20 modules of each subtype and scored each cell for these expression modules using AddModuleScore in Seurat. There were four programs (program A–D) with a relatively high Pearson correlation (Fig.1E) associated with skeletal system development (such as $SFRP1$ and $HTRA1$), DNA packaging (such as $MKi67$ and $PTTG1$), RNA transport (such as $MALAT1$ and $NK1R$), and muscle tissue development (such as $MYLPF$ and $ACTC1$), respectively (Supporting Information Figure S2). In the scope definition of program D, we proactively ignored some ribosomal gene-related modules (such as Mus_v5 and Carti_v14) to obtain a clear function of cell subsets (Fig.1E). The classification of tumor cells was validated again by NMF clustering. The top 30 highly expressed genes in bone- and muscle-related programs were separately and highly distributed in both subtypes. Other cell cycle programs were expressed in several clusters of both subtypes, which may be the only similarity between cell functions of the two subtypes (Fig.1F).

Skeletal muscle cells in PPB

Clusters 2, 4, 5, 6, and 9 (4 038 cells) were kinds of skeletal muscle tumor cells. Using t-SNE projection, we found that classic muscle cell genes DES, NEB, $TNNT1$, $MYOD1$, $TNNI1$,$ACTC1$, and $RYR1$ were highly expressed in these clusters (Fig.2A, Supporting Information Figure S3). Immunohistochemical analysis of DES and MyoD1 protein expression confirmed their presence in these cells (Fig.2B). We continued to subdivide this group and found that it was roughly divided into two subpopulations including mature differentiated myocytes (cluster 5,$MYOG^{hi}TTN^{hi}ENO3^{hi}MYL4^{hi}$, 726 cells) and proliferating satellite myogenic cells (clusters 2, 4, 6, and 9,$PAX7^{hi}MYF5^{hi}MSC^{hi}STC1^{hi}$, 3 312 cells). We found positivity for anti-MYOG antibody staining in these minor differentiated myocytes, showing scattered brown nuclear muscle cells (Fig.2C). Furthermore, these satellite myogenic cells, which were perhaps cells in transition to myoblasts, were named$PAX7^{hi}MYF5^{hi}$ satellite myogenic cells, and clusters 6, 9 (1 149 cells) among these satellite cells clusters expressed metabolic regulator $FABP7$, while the other clusters (2 and 4, 2 163 cells) had high expression of$ITM2A$, a gene expressed in a muscle cell line and associated with myogenesis[16]. GO analysis verified the potential biological processes of these clusters. Cluster 5 had more precise muscle characteristics such as participating in the muscle system process. The other clusters contained more cell cycle genes at G2/M phase with the biological process of chromosome segregation (Fig.2C).

Cartilage subpopulation in PPB

The other clusters were developing in the direction toward chondrocytes. Clusters 0, 1, 3, 7, 8, 10, and 11 (5 806 cells in total) exhibited high expression of *TWIST1* , *ANK3* , *HTRA1* , and *BMP4* , which are markers of osteogenesis (Fig.3A, Supporting Information Figure S3). Subsets of chondrocytes were distinguished based on further markers. A mature minor subset (cluster 7, 617 cells) was defined by chondrocyte markers that emerge in late differentiation, such as *ITM2A* , *MGP* , and *OGN* . Cluster 10 (402 cells) was characterized by *DKK2* and *TNMD* , which was possibly also at the terminal stage of their trajectory with the same status as cluster 7. Another major stem cell subset, clusters 1, 3, 8, and 11, expressed decisive genes in early chondrogenesis, including *SOX9* and *SOX9* -coordinated factor-like *PAX1* , *PAX9* and they maybe were a group of prechondrocytes. Notably, these cartilages were more naïve in their differentiation compared with the muscle subpopulation. Additionally, the characteristic cartilage structure was not observed pathologically and the tumor was negative for immunohistochemical staining of S100 protein. GO analysis of these clusters validated our presumed results. These cells participated in development of the skeletal system and functions such as maintenance and rebuilding of the extracellular matrix[17] (Fig.3B).

Constructing lineage hierarchies of the PPB tumor

To further understand the internal correlation of tumor cells in both subtypes, we constructed differentiation trajectories of cells in pseudotemporal order using Monocle after renaming the clusters of two subtypes to obtain two tightly connected differentiation tumor lineages. In the skeletal muscle lineage, $FABP7^{hi}$ satellite myogenic cells (clusters 6 and 9, $FABP7^{hi}PAX7^{hi}MYF5^{hi}MSC^{hi}$) were more primordial and at the start of their trajectory. Highly expressed genes, such as *PAX7* and *MYF5* , were stable in the early and middle stages of the pseudotime and then declined rapidly (Fig.4B). $ITM2A^{hi}$ satellite myogenic cells (clusters 2 and 4, $ITM2A^{hi}PAX7^{hi}MYF5^{hi}MSC^{hi}$) emerged slightly later, but were still more primitive than $MYOG^{hi}$ differentiated myocytes (cluster 5) that dominated on the terminal branch of the trajectory (Fig.4A,4B). The cartilage lineage was similar to the muscle cells. Naïve $SOX9^{hi}PAX9^{hi}PAX1^{hi}$ prechondrocytes (clusters 1, 3, 8, and 11) drove sequential differentiation and generated $MGP^{hi}OGN^{hi}$ chondrocytes (cluster 7) and $DKK2^{hi}TNMD^{hi}$ chondrocytes (cluster 10) (Fig.4C,4D). However, there was no evidence to order the time sequence of these two chondrocyte subpopulations and we could not define the detailed difference between them, although they converged on the same branch in the trajectory. Interestingly, whether in cartilage or muscle subtypes, when primitive cells occupying the trunk of the trajectory developed into differentiated daughter cells, they first produced a small branch filled with daughter cells and then a large branch again from the trunk. The mechanism was not well understood and might be related to cell activation.

Discussion

Pleuropulmonary blastoma (PPB) is a rare malignant mesenchymal tumor in children. It usually presents with sarcoma-like pathological changes[18,19]. However, there is a gap in the understanding of PPB heterogeneity at the level of single genes. A better understanding of its heterogeneity may facilitate further exploration of the correlation with fetal lung development and the connection between PPB and other diseases such as congenital pulmonary adenomatoid malformation. We used scRNA-seq to roughly define the subtypes of muscle and cartilage in PPB as well as the trajectory of tumor development.

The development of muscle cells in PPB may be related to the differentiation of muscle satellite cells. By clustering, we have easily found striated muscle cells defined by classic genes such as *DES* . These muscle cells are refined into $MYOG^{hi}MYOD1^{hi}$ and $PAX7^{hi}MYF5^{hi}MYOD1^{lo}$ subpopulations by DEGs. The direction of PPB differentiation is controlled by $PAX7^{hi}MYF5^{hi}$ active satellite myogenic cells with the capacity for self-renewal[20,21]. Highly expressed *PAX7* is essential for maintenance of satellite stem cells and targets its downstream gene *MYF5* for conversion into $PAX7^{hi}MYF5^{hi}$ satellite myogenic cells without *MYOD1* . More precisely, this subpopulation in our sample is at the transition stage from satellite myogenic cells to myoblasts ($PAX7^{hi}MYF5^{hi}MYOD1^{hi}$) because of its low expression of *MYOD1* . These satellite myogenic cells in PPB express *MSC* , which generally emerges between embryonic day 10.5 and 16.5 in mice [22] and coordinates with *MYF5* gene[23] to participate in the proceeding myogenesis[21,24]. Cells expressing integral membrane protein

4

2A ($ITM2A$) marks a subset that emerges late among satellite myogenic cells and are in the middle of the entire PPB differentiation trajectory. It controls calcium channels to regulate expression of MYOG protein[25] and prepares for subsequent mature $MYOG^{hi}$ cells. The $MYOG^{hi}$ subpopulation comprises mature myocytes ($MYOG^{hi}MYL4^{hi}TTN^{hi}$) that exit the cell cycle[25]. MYOG protein is essential in late myogenesis[26] without functional overlap with MYLF or MYOD1 protein[27]. We have found that these $MYOG^{hi}$ cells are at the terminal stage of PPB tumor muscle cell development (Fig.4B). The trajectory of skeletal muscle cells in PPB is similar to the differentiation and proliferation of active satellite cells, which is consistent with the fact that the tumor originates from mesoderm.

The cartilage subpopulation is defined by some bone markers emerging in osteogenesis, such as $TWIST1$ (crucial regulator of the differentiation of skeletal progenitor/stem cell[28,29]) and $HTRA1$ (generally expresses in hypertrophic chondrocytes[30], but expresses in a large range in PPB). Such a population comprises naïve prechondrocytes and relatively mature chondrocytes. The prechondrocytes with high expression of $SOX9$, $PAX1$, $PAX9$, and $SOX5$ are at the departure stage of PPB chondrogenesis. $Sox9$ is a crucial gene in chondrogenesis and initiates the chondrocyte differentiation program[31]. $SOX9$ together with $SOX5$ activated by $SOX9$ gene[32] are regulatory markers of mesenchymal cells and prechondrocytes, but disappear after the formation of prehypertrophic chondrocytes[33]. Paralogous transcription factors $PAX1/9$ regulate vertebral column development[34], activate chondrogenic mesenchymal cells to differentiate into prechondrocytes under the combined action of $SOX9$, and are required for maintenance of $BMP4$, a osteogenesis marker[35,36]. Unfortunately, we currently couldn't subdivide the large group of prehondrocytes with DEGs further. Located downstream of PPB chondrogenesis are $DKK2^{hi}$ chondrocytes and $OGN^{hi}MGP^{hi}$ chondrocytes, as indicated by the pseudotime trajectory and gene markers. MGP protein is a kind of calcification inhibitor[37] synthesized by late, but not telophase, chondrocytes[38]. $DKK2$ also plays a role in the late stages of osteoblast differentiation to produce mineralized matrices[39,40]. However, the occurrence time and relationship between the two mature chondrocytes subtypes downstream remain unclear, perhaps because two subtypes have different functions.

The cellular components of pleuropulmonary blastoma and its development trajectory suggest strong intratumoral heterogeneity and the ability to differentiate in multiple directions to skeletal muscle and cartilage. Our atlas of PPB may form the basis for further tumor detection and a more defined molecular mechanism of PPB tumorigenesis. And more detailed cell subtypes and associations between subtypes require our further research.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgements

Supporting information

Additional Supporting Information may be found online in the supporting information section for this article.

Availability of Data

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

[1] A. Ferrari, I.B. Brecht, G. Gatta, et al. Defining and listing very rare cancers of paediatric age: consensus of the Joint Action on Rare Cancers in cooperation with the European Cooperative Study Group for Pediatric

Rare Tumors. *Eur. J. Cancer* 2019;110:120-126.

[2] D.A. Hill, J. Ivanovich, J.R. Priest, et al. DICER1 mutations in familial pleuropulmonary blastoma. *Science* 2009;325(5943):965.

[3] M. Seki, K. Yoshida, Y. Shiraishi, et al. Biallelic DICER1 mutations in sporadic pleuropulmonary blastoma. *Cancer Res.*2014;74(10):2742-2749.

[4] R. Shukrun, H. Golan, R. Caspi, et al. NCAM1/FGF module serves as a putative pleuropulmonary blastoma therapeutic target.*Oncogenesis* 2019;8(9):48.

[5] A.C. White, J. Xu, Y. Yin, et al. FGF9 and SHH signaling coordinate lung growth and development through regulation of distinct mesenchymal domains. *Development* 2006;133(8):1507-1517.

[6] S.S. Potter. Single-cell RNA sequencing for the study of development, physiology and disease. *Nat Rev Nephrol*2018;14(8):479-492.

[7] X. Ren, B. Kang, Z. Zhang. Understanding tumor ecosystems by single-cell sequencing: promises and limitations. *Genome Biol.*2018;19(1):211.

[8] S. Picelli. Single-cell RNA-sequencing: The future of genome biology is now. *RNA Biol.* 2017;14(5):637-650.

[9] F. Tang, C. Barbacioru, Y. Wang, et al. mRNA-Seq whole-transcriptome analysis of a single cell. *Nature methods*2009;6(5):377-382.

[10] W. Chung, H.H. Eum, H.O. Lee, et al. Single-cell RNA-seq enables comprehensive tumour and immune cell profiling in primary breast cancer. *Nat Commun* 2017;8:15081.

[11] A.P. Patel, I. Tirosh, J.J. Trombetta, et al. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma.*Science* 2014;344(6190):1396-1401.

[12] E. Valenzi, M. Bulik, T. Tabib, et al. Single-cell analysis reveals fibroblast heterogeneity and myofibroblasts in systemic sclerosis-associated interstitial lung disease. *Ann. Rheum. Dis.*2019;78(10):1379-1387.

[13] M. Venkatasubramanian, K. Chetal, D. Schnell, et al. Resolving single-cell heterogeneity from hundreds of thousands of cells through sequential hybrid clustering and NMF. *Bioinformatics* 2020;

[14] C. Trapnell. Defining cell types and states with single-cell genomics. *Genome Res.* 2015;25(10):1491-1498.

[15] C. Vokuhl, L. de Leon-Escapini, I. Leuschner. Strong Expression and Amplification of IGF1R in Pleuropulmonary Blastomas. *Pediatr. Dev. Pathol.* 2017;20(6):475-481.

[16] M. Lagha, A. Mayeuf-Louchart, T. Chang, et al. Itm2a is a Pax3 target gene, expressed at sites of skeletal muscle formation in vivo.*PLoS One* 2013;8(5):e63143.

[17] C. Prein, F. Beier. ECM signaling in cartilage development and endochondral ossification. *Curr. Top. Dev. Biol.* 2019;133:25-47.

[18] M.K. Yadav, M. Singhal, A. Bhatia, et al. Pleuropulmonary blastoma in adolescence: A rare tumor beyond first decade of life.*Lung India* 2015;32(3):281-284.

[19] Y.H. Messinger, D.R. Stewart, J.R. Priest, et al. Pleuropulmonary blastoma: a report on 350 central pathology-confirmed pleuropulmonary blastoma cases by the International Pleuropulmonary Blastoma Registry. *Cancer* 2015;121(2):276-285.

[20] P.S. Zammit. Function of the myogenic regulatory factors Myf5, MyoD, Myogenin and MRF4 in skeletal muscle, satellite cells and regenerative myogenesis. *Semin. Cell Dev. Biol.* 2017;72:19-32.

[21] M.A. Rudnicki, F. Le Grand, I. McKinnell, et al. The molecular regulation of muscle stem cell function. *Cold Spring Harb. Symp. Quant. Biol.* 2008;73:323-331.

[22] P. Zhao, E.P. Hoffman. Musculin isoforms and repression of MyoD in muscle regeneration. *Biochem. Biophys. Res. Commun.*2006;342(3):835-842.

[23] G.H. Lee, M.Y. Chang, C.H. Hsu, et al. Essential roles of basic helix-loop-helix transcription factors, Capsulin and Musculin, during craniofacial myogenesis of zebrafish. *Cell. Mol. Life Sci.*2011;68(24):4065-4078.

[24] L. Robb, L. Hartley, C.C. Wang, et al. musculin: a murine basic helix-loop-helix transcription factor gene expressed in embryonic skeletal muscle. *Mech. Dev.* 1998;76(1-2):197-201.

[25] E.J. Lee, J.H. Nam, I. Choi. Fibromodulin modulates myoblast differentiation by controlling calcium channel. *Biochem. Biophys. Res. Commun.* 2018;503(2):580-585.

[26] J.M. Venuti, J.H. Morris, J.L. Vivian, et al. Myogenin is required for late but not early aspects of myogenesis during mouse development. *J. Cell Biol.* 1995;128(4):563-576.

[27] A. Rawls, J.H. Morris, M. Rudnicki, et al. Myogenin's functions do not overlap with those of MyoD or Myf-5 during mouse embryogenesis.*Dev. Biol.* 1995;172(1):37-50.

[28] H. Miraoui, P.J. Marie. Pivotal role of Twist in skeletal biology and pathology. *Gene* 2010;468(1-2):1-7.

[29] C.F. Liu, W.E. Samsa, G. Zhou, et al. Transcriptional control of chondrocyte specification and differentiation. *Semin. Cell Dev. Biol.* 2017;62:34-49.

[30] A. Tsuchiya, M. Yano, J. Tocharus, et al. Expression of mouse HtrA1 serine protease in normal bone and cartilage and its upregulation in joint cartilage damaged by experimental arthritis. *Bone*2005;37(3):323-336.

[31] L. Zeng, H. Kempf, L.C. Murtaugh, et al. Shh establishes an Nkx3.2/Sox9 autoregulatory loop that is maintained by BMP signals to induce somitic chondrogenesis. *Genes Dev.* 2002;16(15):1990-2005.

[32] H. Akiyama, M.C. Chaboissier, J.F. Martin, et al. The transcription factor Sox9 has essential roles in successive steps of the chondrocyte differentiation pathway and is required for expression of Sox5 and Sox6. *Genes Dev.* 2002;16(21):2813-2828.

[33] V. Lefebvre, P. Smits. Transcriptional control of chondrocyte fate and differentiation. *Birth Defects Res C Embryo Today*2005;75(3):200-212.

[34] H. Peters, B. Wilm, N. Sakai, et al. Pax1 and Pax9 synergistically regulate vertebral column development.*Development* 1999;126(23):5399-5408.

[35] H. Peters, A. Neubüser, K. Kratochwil, et al. Pax9-deficient mice lack pharyngeal pouch derivatives and teeth and exhibit craniofacial and limb abnormalities. *Genes Dev.*1998;12(17):2735-2747.

[36] V. Sivakamasundari, P. Kraus, W. Sun, et al. A developmental transcriptomic analysis of Pax1 and Pax9 in embryonic intervertebral disc development. *Biol Open* 2017;6(2):187-199.

[37] K. Yagami, J.Y. Suh, M. Enomoto-Iwamoto, et al. Matrix GLA protein is a developmental regulator of chondrocyte mineralization and, when constitutively expressed, blocks endochondral and intramembranous ossification in the limb. *J. Cell Biol.* 1999;147(5):1097-1108.

[38] G. Luo, P. Ducy, M.D. McKee, et al. Spontaneous calcification of arteries and cartilage in mice lacking matrix GLA protein.*Nature* 1997;386(6620):78-81.

[39] X. Li, P. Liu, W. Liu, et al. Dkk2 has a role in terminal osteoblast differentiation and mineralized matrix formation. *Nat. Genet.* 2005;37(9):945-952.

[40] H. Oh, J.H. Ryu, J. Jeon, et al. Misexpression of Dickkopf-1 in endothelial cells, but not in chondrocytes or hypertrophic chondrocytes, causes defects in endochondral ossification. *J. Bone Miner. Res.*2012;27(6):1335-1344.

Legends

Fig.1. NMF and Seurat analysis verify heterogeneity in PPB. (A) General process of droplet-enabled single cell sequencing. (B). Heatmap of the top five marker genes for each cluster using Seurat. Yellow indicates high expression of a particular gene, and purple indicates low expression. (C) Tumor inferCNV in comparison with normal fibroblasts (fibroN) indicates the CNV changes of tumor cells. A reference group is shown at the top, while the other clusters are below. Brown is chromosome gain, and blue is chromosome loss. (D) Clustering of 10 007 cells and definition of cell types in PPB are displayed in t-SNE plot. (E) Pairwise correlations between the expression modules of the two subtypes (rows, columns) were made to acquire four positive correlated programs. Brown indicates a high Pearson correlation. (F) Top 30 genes in each program were re-clustered and shown in a heatmap. Programs are shown on the y-coordinate and previous clusters of cell types are shown on the x-coordinate.

Fig.2. Subsets of smooth muscle cells in PPB. (A) Feature plots of muscle-related genes revealed the subpopulations of muscle cells. (B) Immunohistochemical staining of Desmin expression (brown) in combination with other muscle proteins, MyoD1 (brown) and MYOG (brown), within PPB tissue sections show striated muscle cells. (C) GO analysis of the selected top 20 GO terms for clusters 5 and 6 is shown in bar plots.

Fig.3. Subsets of chondrocytes in PPB. (A) Feature plots of genes defining various subpopulations of tumor cartilage. (B) GO analysis of the selected top 20 GO terms for clusters 1, 3, 7, and 10 are shown on the bottom.
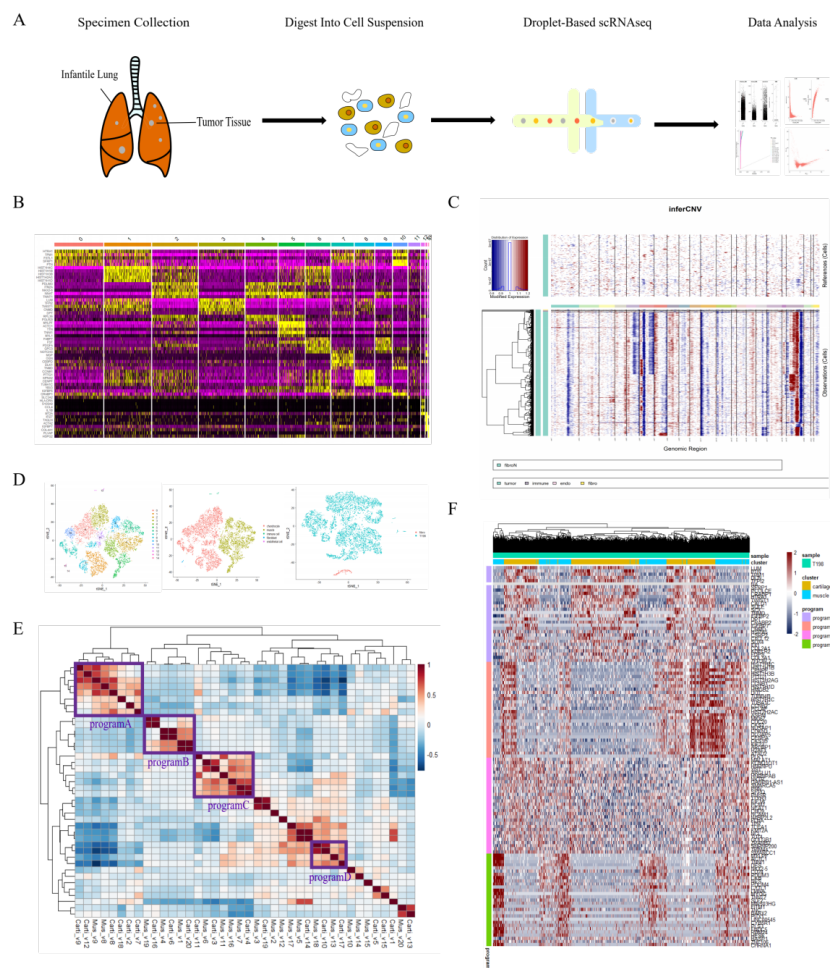
Fig.4.Two lineage hierarchies of the PPB tumor. (A), (C) Muscle and cartilage cells were respectively divided into three cell subtypes and automatically shown in a Monocle-generated pseudotime trajectory. Dark blue is the start of pseudotime. A proposed model summarizing the lineage hierarchies from relatively stem cells to differentiated cells is shown. PAX7^{hi}MYF5^{hi}MYOD1^{hi}myoblast in (A) are defined transitional cell during pseudotime and drawn by a dashed line. (B), (D) Trendgrams of highly expressed genes in the cell subtypes during pseudotime.

Supporting Information Figure

Supporting Information Figure S1 Feature and violin plots of genes defining the rough cell subtypes of 10 007 collected single cells.

Supporting Information Figure S2 (A)-(D) GO analysis of the selected top 10 GO terms for programs A-D is shown in bar plots.

Supporting Information Figure S3 Feature plots of genes defining the subpopulations in muscle and cartilage cells of PPB.

Hosted file

Fig.2.pdf available at https://authorea.com/users/329689/articles/456637-single-cell-transcriptional-profiling-identifies-heterogeneity-in-pleuropulmonary-blastoma