Thesis for the Master of Science

A computational detection of lncRNA signatures in tumor-immune microenvironment at single-cell level

Eun-Gyeong Park

Graduate School of Hanyang University

February 2022

Thesis for the Master of Science

A computational detection of lncRNA signatures in tumor-immune microenvironment at single-cell level

Thesis Supervisor: Jin-Wu Nam

A Thesis submitted to the graduate school of
Hanyang University in partial fulfillment of the
requirements for the degree of Master of Science

Eun-Gyeong Park

February 2022

Department of Life Science
Graduate School of Hanyang University

This thesis, written by Eun-Gyeong Park,
has been approved as a thesis for the degree of
Master of Science.

February 2022

Committee Chairman: Prof. Je-Min Choi

Committee member: Prof. Junho Choe

Committee member: Prof. Jin-Wu Nam

Graduate School of Hanyang University

Table of Contents

List of Figures

List of Tables

Abstract

A computational detection of lncRNA signatures in tumor-immune microenvironment at single-cell level

Eun-Gyeong Park
Department of Life Science
Graduate School of
Hanyang University

Long non-coding RNAs (lncRNAs) are transcripts without protein-coding capacity and accumulating reports indicate that lncRNAs are key factors of tumorigenesis. In several cancer types, dysregulation of lncRNAs often causes enhanced cancer progression, metastasis, or even tumor suppression, however, the detailed functional mechanisms of these lncRNAs have not yet been studied. One of the issues is that, although lncRNAs are known to exhibit high specificity in different cell types or cellular states, many cancer-associated lncRNAs were studied based on bulk RNA-sequencing (RNA-seq) data which represents a mixed signal of cancer cells and diverse non-cancerous cells coexisting and interacting with each other in the tumor microenvironment (TIME). In this reason, lncRNA studies should be performed not only at the bulk sample level, but also at the single-cell RNA-sequencing (scRNA-seq) level to decode the precise source of the expression. Here, non-small cell lung cancer scRNA-seq data produced using

two popular scRNA-seq platforms, 10X Chromium and SMART-seq2, were analyzed with novel lncRNA transcriptome to identify differentially expressed lncRNAs in TIME. 812 cell type-specific lncRNA markers were identified and 74 of these lncRNAs were shared between the two platforms. Well-known oncogenic or tumor suppressive lncRNAs, such as *LUCAT1* or *MEG3*, were specifically expressed by non-malignant cells. Moreover, 12 and 57 of cell type-specific lncRNAs were differentially expressed between cells from tumor and nontumor samples in 10X Chromium and SMART-seq2 data, respectively. As a result, this study provides not only a comprehensive lncRNA signature in terms of cell type- and tumor-specificity, but also suggests lncRNA targets that could help to investigate mechanisms in cancer development.

1. Introduction

Long non-coding RNA (lncRNA) is a non-coding transcript of longer than 200 nucleotides. Thousands of lncRNAs have been identified to exhibit specific expression patterns in various types of cancer, and these lncRNAs exert cancer-related functions, such as cancer cell survival, proliferation, invasion, migration, and anti-apoptosis (Prensner and Chinnaiyan 2011, Du et al. 2013, Huarte 2015, Schmitt and Chang 2016). Oncogenic lncRNAs that promote cancer progression tend to be upregulated in cancer patients with poor prognosis, hence, inactivation of these lncRNAs often leads to hinder tumorigenesis and cancer cell survival which makes lncRNAs as potential targets for therapeutic interventions. For examples of these cancer-related lncRNAs, lung cancer associated transcript 1 (*LUCAT1*) has tumor-promoting roles in several types of cancer. (Xing et al. 2021). In non-small cell lung cancer (NSCLC), *LUCAT1* regulates the expression of p21 and p57 by interacting with PRC2 and EZH2 complex (Sun et al. 2017). In esophagus squamous cell carcinoma (ESCC), *LUCAT1* also inhibits the expression of tumor suppressors by regulating the stability of DNA methyltransferase 1 (DNMT1) (Yoon et al. 2018). Highly Expressed lncRna in ESophageal squamous cell carcinoma (*HERES*) is another oncogenic lncRNA detected in ESCC which controls tumor growth by regulating the Wnt signaling pathways interacting with EZH2 (You et al. 2019). On the other hand, tumor suppressive lncRNAs are generally downregulated in tumor, and when these lncRNAs are dysregulated,

tumorigenesis tends to be promoted. For example, *NKILA* functions as a tumor suppressor lncRNA in breast cancer. *NKILA* is activated by TNF-α and IL-1β which in turn inhibits NF-κB pathway to limit tumorigenesis, angiogenesis, immunosuppression, and EMT signaling (Liu et al. 2015). However, most of these lncRNA studies were conducted in bulk tumor samples which is an averaged expression of heterogeneous cell mixtures and experimental validations were performed in cancer cell lines with or without limited confirmation of which cell type in TIME expresses these lncRNAs.

Tumor mass is composed of various cell types including malignant, resident or infiltrating immune, and stromal cells. Many key features in cancer, such as progression, invasion, or drug resistance, are known to be affected by the composition and interaction of non-cancerous cells in the tumor-immune microenvironment (TIME) (Binnewies et al. 2018, Taube et al. 2018, Galli et al. 2020). Therefore, it is critical to understand in which cell type specifically expresses oncogenic or tumor suppressing lncRNAs to survey the functions. Some of the cancer-related lncRNAs also function in immune cells. In case of *LUCAT1*, it functions in myeloid cells by binding to STAT1 and inhibiting transcription of interferon-stimulated genes (ISGs) (Agarwal et al. 2020). *NKILA* also specifically functions in T cells inducing hyper activated immune cell death (Huang et al. 2018). Therefore, these multifaceted lncRNAs suggest the possibility of cancer-driving or -suppressing lncRNAs may actually be highly expressed and function in non-malignant cells. However, majority of lncRNA studies

exploited bulk RNA-sequencing data where various cell types contribute to an averaged signal, limiting discovery of cell type-specific lncRNA markers in TIME. Because multi-functional lncRNAs were confirmed at the bulk cancer level, the precise molecular mechanisms remain vague. Therefore, it is necessary to study lncRNAs at the level of individual cell types to understand characterization of their functional mechanisms in cancer.

Advances in single-cell RNA-sequencing (scRNA-seq) enable profiling of transcriptome in cancer at the single-cell level. Therefore, it allowed us to study lncRNAs that are expressed specifically in a certain cell type in TIME. In fact, there are a few studies analyzed lncRNA expression using scRNA-seq. For example, Liu *et al.* studied changes in lncRNA expression during the development of human neocortex at the single-cell level and discovered various lncRNAs which were dynamically expressed (Liu et al. 2016). The mechanism of *H19* was also revealed by using scRNA-seq, which acts as a regulator in the development of hematopoietic stem cells (HSCs) (Zhou et al. 2019). Additionally, the lncRNA landscape in T cell subtypes across various of cancers was analyzed using scRNA-seq data (Luo et al. 2021). However, because majority of single-cell studies were focused on protein-coding genes (PCGs), there are only few published studies profiling lncRNAs at single-cell resolution. Moreover, since the reference transcriptome of lncRNAs was annotated using bulk RNA-seq data, there is limited information on cell type compositions and cell type-specific lncRNAs.

Many single-cell sequencing techniques were introduced so far, however, majority of studies use droplet-based sing-cell platform, such as 10X Chromium 3'-seq (10X). 10X data capture cells in small droplets and produces single-cell reads at the 3'-end of RNAs with poly-A tails (Zheng et al. 2017). Because droplet-based approach can process a large number of cells, 10X single-cell data has the advantage of detecting minor cell types. On the other hand, SMART-seq2 (SS2) data is sequenced by plate-based approach that produces full-length transcripts by cell similar to that of bulk RNA-seq (Picelli et al. 2014). Therefore, with higher read coverages and sequencing depth of SS2 data, lncRNA with low expression can be detected well, suggesting that SS2 would provide a more sensitive means to identify cell type-specific lncRNAs in the TIME. Analysis was conducted using both data to reflect the advantages of 10X and SS2. Since the two scRNA-seq platforms have features that can affect lncRNA detection, it will be helpful for further lncRNA research using other cancer single-cell data.

In this study, lncRNA expression profiling was performed using scRNA-seq using BIGTranscriptome (v1.2) augmented with novel lncRNA annotations specific for lung cancer, colon cancer, and 33 immune cell types (Figure 2A) (You et al. 2017). With newly annotated lncRNAs, I searched for marker genes that are cell type-specifically expressed in TIME. I next confirmed the commonly expressed marker genes detected between two different scRNA-seq platforms, 10X and SS2. By profiling the expression of novel lncRNAs in lung cancer, the landscape of new lncRNA marker sets at the cell type

level in TIME were established. Many lncRNAs known as oncogenes or tumor suppressors were highly expressed in non-malignant cells. 12 and 57 of cell type-specific lncRNAs were differentially expressed between cells from tumor and nontumor samples in 10X and SS2 data, respectively. Accordingly, this study suggests cell type- and tumor type-specifically expressed lncRNA candidates and offers novel treatment target lncRNAs.

2. lncRNA profiling in lung cancer scRNA-seq data

2.1. lncRNA detection in single-cell platforms

I analyzed non-small cell lung cancer (NSCLC) scRNA-seq data produced using 10X (Qian et al. 2020) and lung adenocarcinomas produced using SS2 (Maynard et al. 2020) together comprising 80 samples: 42 tumor samples (24 in 10X; 18 in SS2), 10 nontumor samples (7 in 10X; 3 in SS2), and 28 metastasis samples (SS2). These two representative single-cell platforms, 10X and SS2, were compared with public lung cancer bulk RNA-seq datasets (The Cancer Genome Atlas; TCGA) to benchmark the sensitivity of lncRNA detection. I compared the transcripts per million (TPM) distribution of lncRNA expressing more than 1 TPM in bulk with single-cell data (Figure 1A). ~37.6% and ~65.4% of lncRNAs which were expressed in bulk RNA-seq data were also detected at the same level (≥ 1 TPM) in both 10X and SS2 platforms, respectively. In detail, 12,738 PCGs and 1146 lncRNAs expressing over 1 TPM; 1277 PCGs and 1133 lncRNAs expressing over 0.1 TPM in 10X data. Also, 12,629 PCGs and 3026 lncRNAs were detected at over 1 TPM, and 1743 PCGs and 1518 lncRNAs were detected at over 0.1 TPM in SS2. On the other hands, genes with less than 0.1 TPM in bulk which are regarded as “not expressed” genes were detected at similar amounts in single-cell data. In SS2 data, more lncRNAs were expressed (≥1 TPM) indicating that sequencing depth of the platform heavily affects the lncRNA detection rates. In fact, by downsampling 10X and SS2 to different sequencing depths, I checked the detection rate of PCGs and lncRNAs according to the sequencing depth, and the detection rate of lncRNAs in the SS2 platform

showed about 10% higher than that of the 10X platform (Figure 1B). As expected, SS2 appeared to be more sensitive for the detection of lncRNAs, although the sensitivity depends on the sequencing depth. As a result, although lncRNAs generally show low expression, they are detected in single-cell data and well agreed with bulk RNA-seq data.

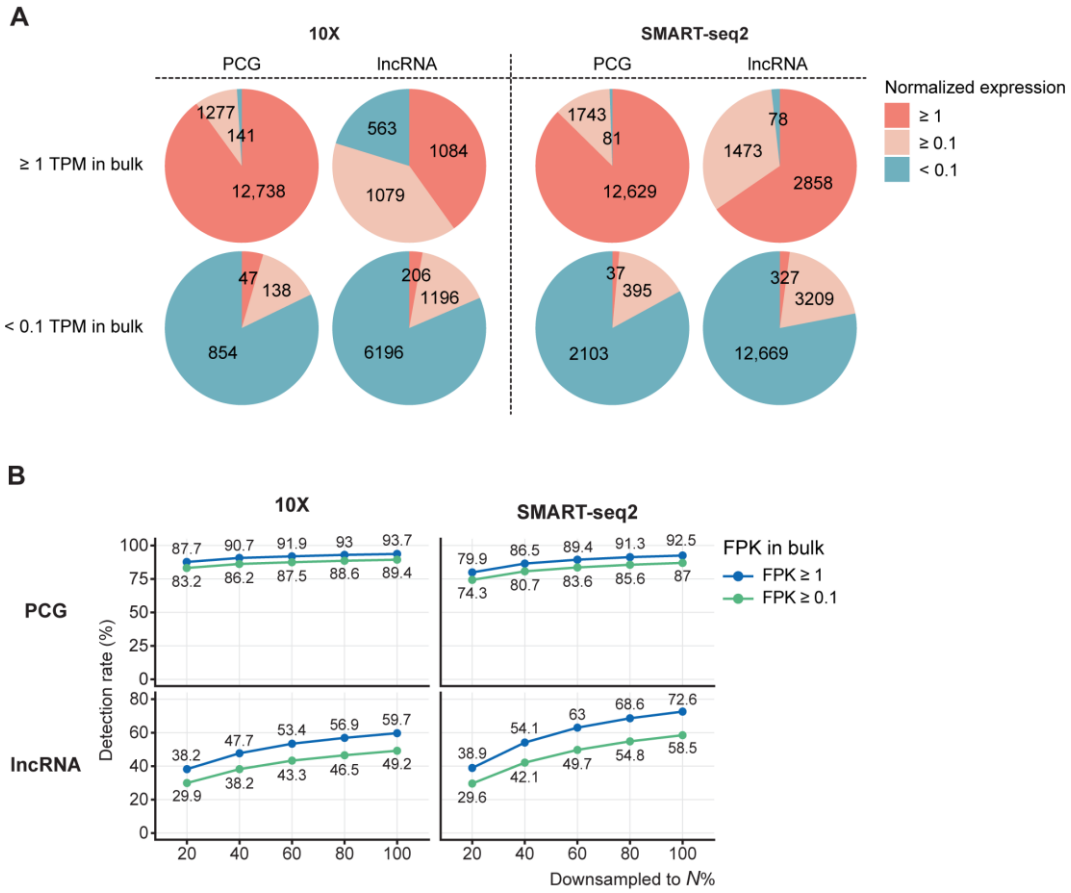


Figure 1. lncRNA detection in bulk RNA-seq vs scRNA-seq. (A) Transcripts per million (TPM) distribution for genes represented in scRNA-seq datasets that were detected at ≥ 1 TPM and < 0.1 TPM in bulk RNA-seq data from TCGA (lung adenocarcinoma and squamous cell carcinoma for 10X and lung adenocarcinoma for SS2) in each scRNA-seq platform. (B) Fraction of expressing genes (FPK ≥0.1) in downsampled single-dell data. lncRNAs were detected more sensitively in high sequencing depth, and detection rate in SS2 platform found about 10% more lncRNAs than 10X platform.

2.2. Basic analysis for lung cancer scRNA-seq data

To identify cellular heterogeneity of lung cancer, both lung cancer datasets were analyzed using a scRNA-seq software, Seurat (v3.0) (Stuart et al. 2019). After quantification using BIGTranscriptome v1.2, libraries in each dataset were individually filtered for low-quality cells based on UMI (≥ two standard deviations from the mean), gene counts (< 200), mitochondrial gene contents (≥ 20%). Both preprocessed datasets were merged and cell annotations were assigned by label transfer using a publicly available lung cancer reference from the Blueprint project (Qian et al. 2020) (Figure 2B). After additional quality controls based on prediction score of label transfer (prediction score > 0.5) and mismatches of cell annotations between cell type and subtype (cells that are annotated as a cell type A but assigned as a subtype of a different cell type B), the median of 2358 UMIs and 972 genes per cell were obtained in 10X data, median of 8367.3 reads, 3650 genes per cell are obtained in SS2 data. Much higher number of genes were detected in SS2 compared to 10X (median of 972 genes from 10X compared to 3650 genes from SS2) indicating the characteristics of each data were well represented. Consequently, 14 major clusters were identified; alveolar, epithelial, malignant, B, plasma B, mast cells, macrophages, fibroblasts, endothelial cells (ECs), dendritic cells (DCs), plasmacytoid DCs (pDCs), neutrophils, T cells, and erythroblasts (Figure 2C). Of these, T cells and macrophages were the most frequent cell types in both 10X and SS2. Most cells in T cells and

B cells clusters consist of tumor cells suggesting immune cells are infiltrated in tumor with affecting cancer development. Alveolar and ECs clusters have many cells originating from nontumor samples and neutrophils and pDCs clusters exhibit more metastatic cells than other clusters. (Figure 2D). I validated clusters by exhibiting gene expression of canonical marker genes for each cell type (Figure 2E, F). In summary, I reconstructed heterogeneous TIME of lung cancer produced by different platforms based on reference datasets and clustered well-known cell types using BIGTranscriptome v1.2 with various novel lncRNA transcripts.

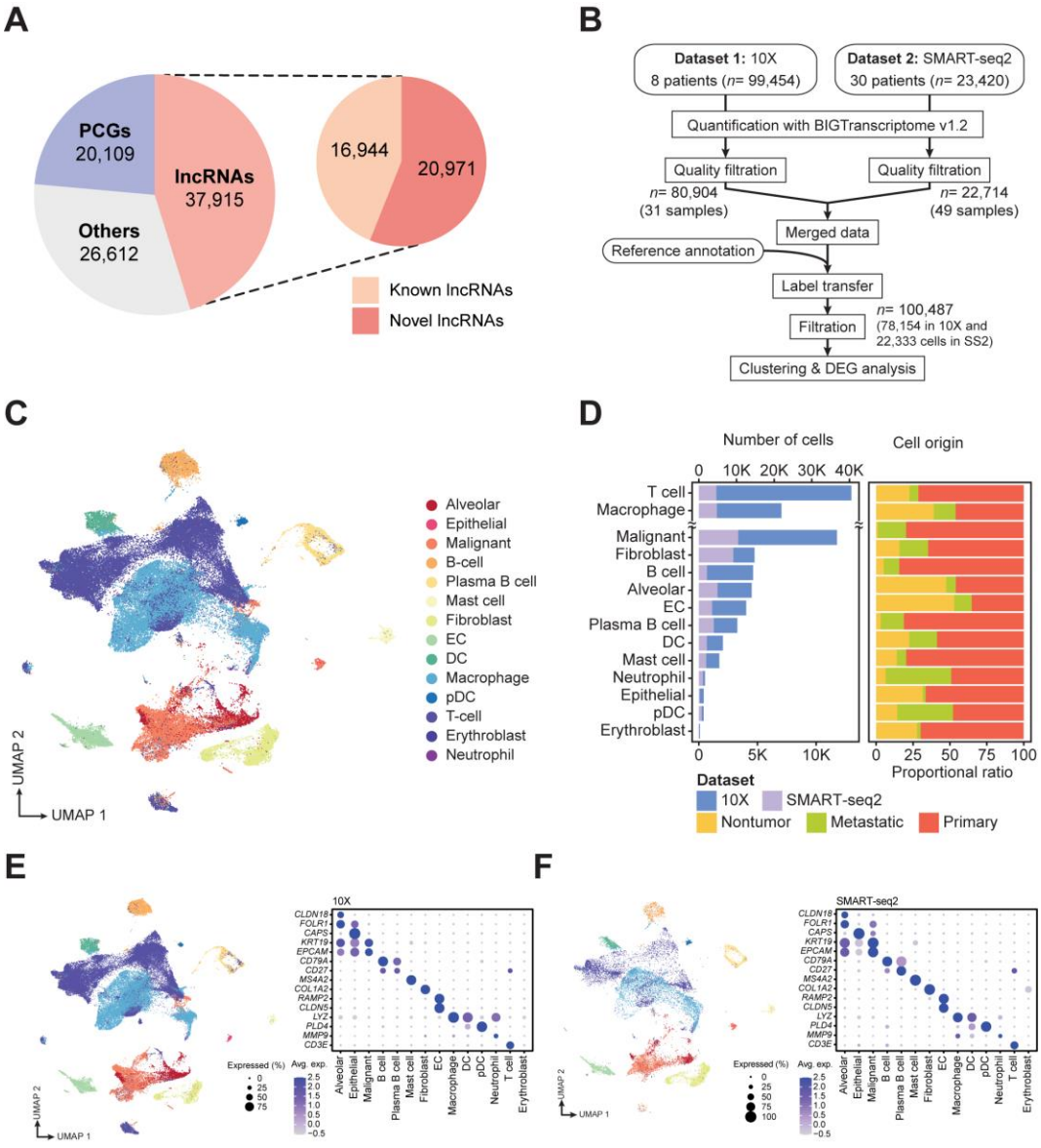
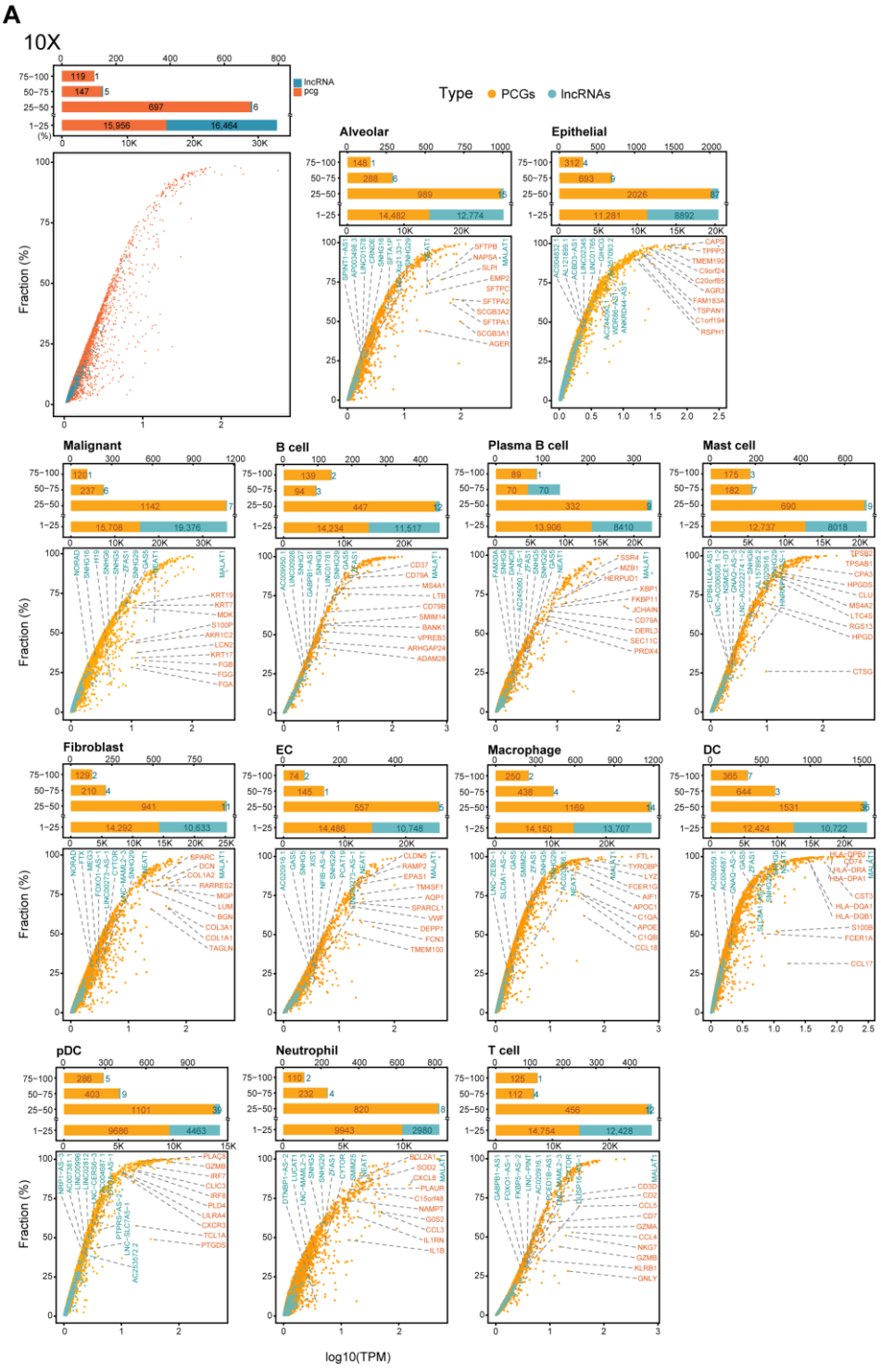


Figure 2. Study pipeline and clustering of lung cancer scRNA-seq data. (A) The proportion of genes in BIGTranscriptome v1.2. (B) Single-cell analysis procedure of two lung cancer datasets for lncRNA marker detection. (C) UMAP of 100,487 cells, color coded by each 14 cell types. (D) Cell count in each cell type clusters with a fraction of dataset (left) and fraction of cell origination (right). (E) 10X; 78,154 cells clusters (left) and expression dot plot of cell type-specific canonical markers (right). Dot size indicates expressed cell fraction and color means averaged expression values. (F) SS2; 22,333 cells clusters (left) and expression dot plot of canonical marker genes of each cell type (right).

2.3. Profiling of lncRNAs at the single-cell resolution

Since it is known that lncRNA has a very low expression level compared to PCGs, I checked whether lncRNAs could be observed cell type-specifically at the single-cell level by comparing the expression levels (mean TP10K; transcripts per 10k transcripts) and the fraction of expressing cells (Figure 3). In case of lncRNAs (n = 37,915), there are exceptional lncRNAs that have higher expression values and fractions than PCGs, such as *MALAT1*, and genes expressed in more than 50% of cells, such as *NEAT1*, *GAS5*, *SNHG29*, *SNHG5*, and *ZFAS1* (Figure 3A). However, most of lncRNAs were expressed in less than 25% of the cells and have lower expression levels compared with PCGs. Although the expression of lncRNA was very low in pooled scRNA-seq data, it was confirmed that the fraction and/or expression was largely increased at the individual cell type level while PCGs remained similarly (Figure 3A, B). These results indicate that lncRNAs are specifically expressed in certain cell types or the result from dropout effect of lowly expressed transcripts prevalent in scRNA-seq data. In fact, when comparing the ratio of specifically expressed lncRNA or PCGs in each cell type, lncRNAs showed much higher specificity compared to PCGs (Figure 4). 17% of the lncRNAs were specifically expressed in epithelial cells, 29% were in malignant cells, 55% were in lymphoid cells, and 25% were in myeloid cells. On the other hand, only 16~22% of PCGs were expressed cell type specifically indicating the discrete expression patterns of lncRNAs (Figure 4A; left). Similar results were also observed

in SS2 data. In SS2, 275 known and 632 novel lncRNAs were expressed in pooled data at 1 TPM or higher. PCGs that were specifically expressed were found to be less than 13% of cells in each cell types, and lncRNA had a higher proportion than that of 10X (Figure 4A; right). As a result, lncRNAs are expressed cell type-specifically in scRNA-seq using 10X and SS2 platforms, and these features could be sufficiently detected at single-cell resolution.


A
10X
lncRNA
pcg
Type PCGs lncRNAs
Alveolar
Epithelial
Malignant
B cell
Plasma B cell
Mast cell
Fibroblast
EC
Macrophage
DC
pDC
Neutrophil
T cell
Fraction (%)
log10(TPM)

(Continued)

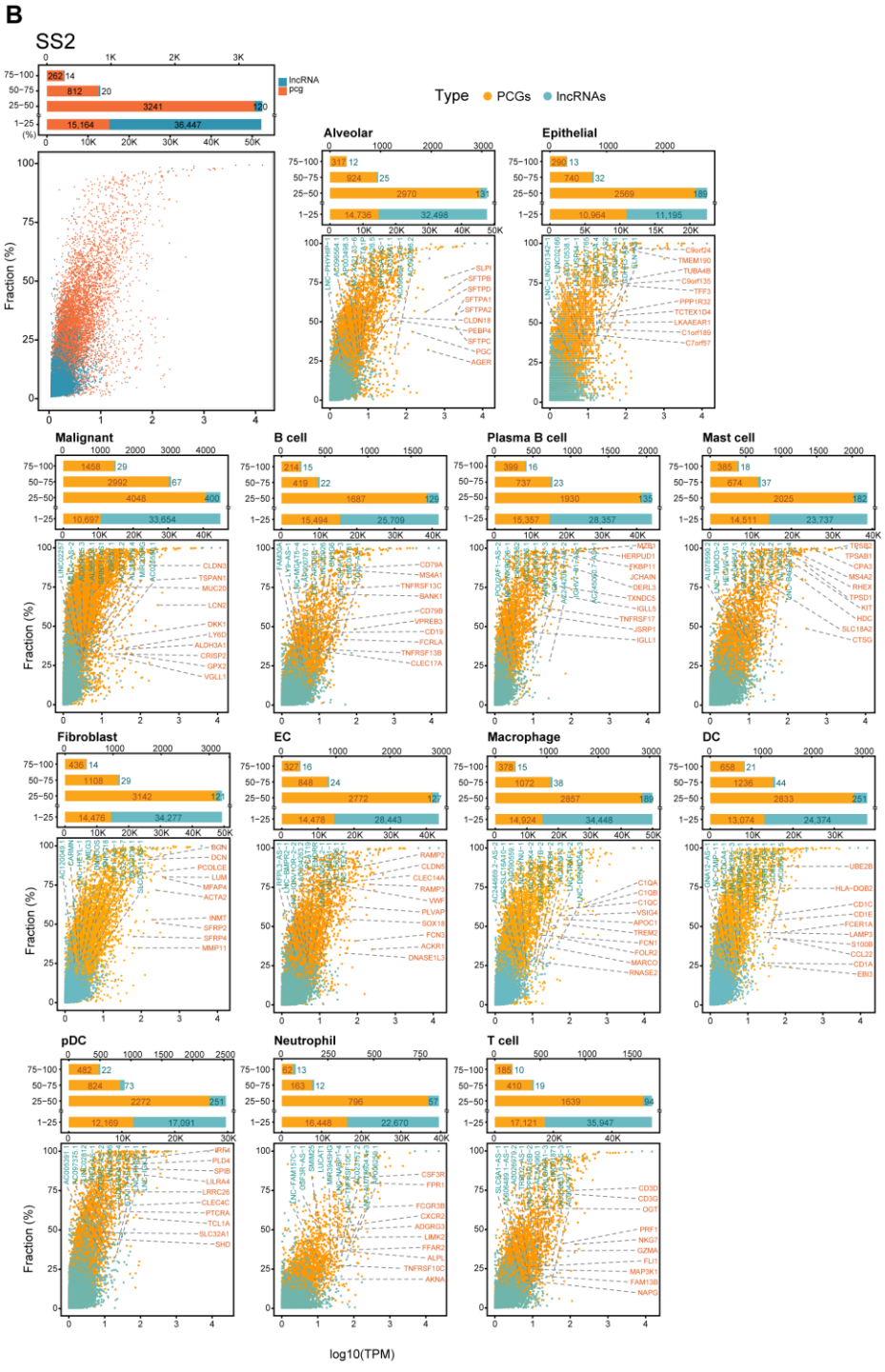


Figure 3. Plots of lncRNA expression and fraction in total and each 13 cell types. Scatter plots of the log-transformed mean expression of PCGs and lncRNAs (X-axis), and expressed cell fraction (Y-axis) combined with bar plots of gene count in each fraction. The scatter plot labeled 10 lncRNAs and 10 PCGs that are known to be highly expressed for that cell type.; (A) 10X, (B) SS2.

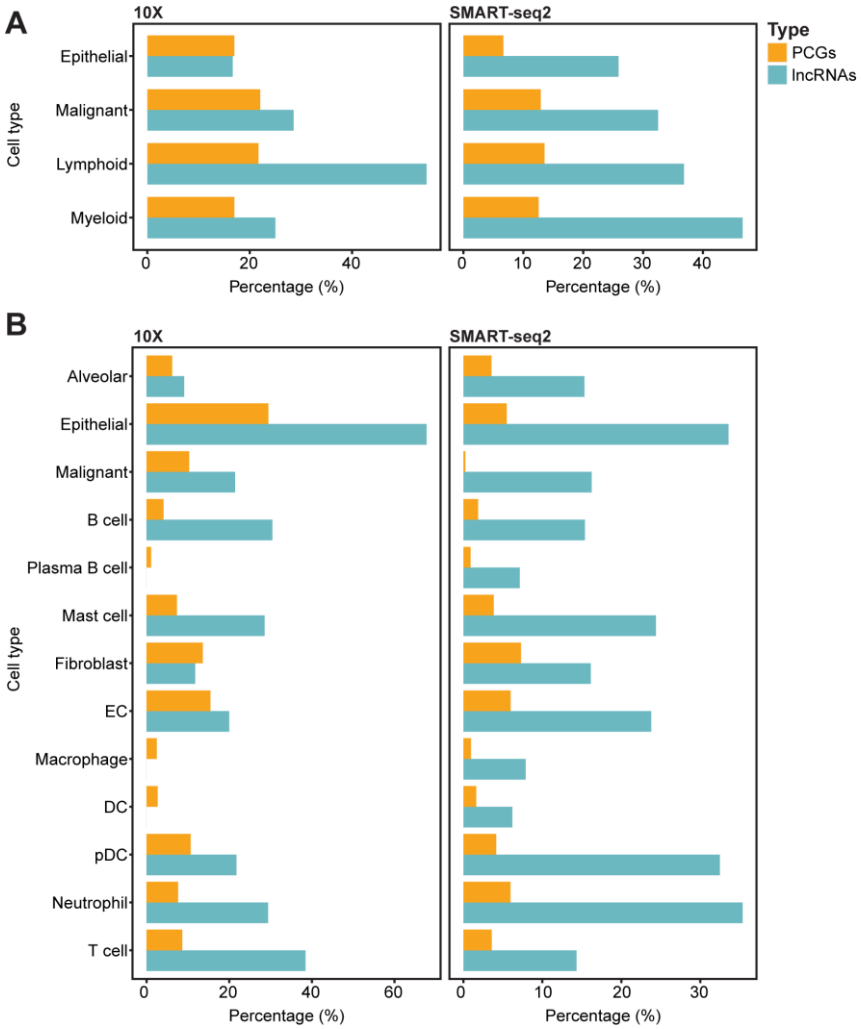


Figure 4. Cell type-specifically expressed gene ratio in each cell type. (A) Specifically expressed genes in 4 major cell types; alveolar and epithelial cells included in Epithelial, malignant cells in Malignant, T cells, B cells and plasma B cells in Lymphoid, mast cells, macrophages and neutrophils in Myeloid. (B) Cell types are divided into 13 that are presented in Figure 2C except for erythroblast.

3. Cell type signature lncRNAs in TIME

3.1. Identification of specific lncRNAs in each cell type

To identity lncRNA signatures in lung cancer microenvironment, differentially expressed gene (DEG) analysis was performed on 13 broad cell types in both platforms, separately (Figure 2; exclude erythroblasts). In summary, 3760 cell type markers (3623 PCGs and 137 lncRNAs) from 10X data (Figure 5A) and 8890 cell type markers (8141 PCGs and 749 lncRNAs) from SS2 data were identified (Figure 6A). In 10X data, 137 lncRNA markers were consisted of 104 known and 33 novel lncRNAs. In these markers, many lncRNAs known as regulators of tumor progression and suppression were included, such as, *MALAT1, NEAT1, MEG3, LUCAT1, GAS5,* and *PCAT19* (Figure 5B). *MALAT1* and *NEAT1* were reported to enhance cell proliferation and invasion in several cancers (Ji et al. 2003, Hirata et al. 2015, Sun et al. 2016, Yu et al. 2017, Li et al. 2018). *GAS5* functions as a tumor suppressor in multiple types of cancer including breast, colorectal, gastric, and liver cancers (Mourtada-Maarabouni et al. 2009, Yang et al. 2020). However, *MALAT1*, *NEAT1*, and *GAS5* are non-specifically expressed in most cell types. Similarly, *LUCAT1* is known as a oncogene that is identified as upregulated in lung cancer and promotes cancer progression (Sun et al. 2017, Yoon et al. 2018) and *MEG3* is a tumor-suppressing lncRNA in liver and gastric cancers (Zhou et al. 2007, Sun et al. 2014, Zhuo et al. 2016), however, in the single-cell data, *LUCAT1* and *MEG3* were specifically found in neutrophils and fibroblasts, respectively. Novel lncRNA markers were more specifically expressed in each cell type than

known lncRNAs. Most of these marker genes were expressed in fraction ranging from 25% to 75% of cells in each cell type (Figure 5C). Interestingly, several novel lncRNAs were associated to the genes regulating functions or differentiation of the cell types. For example, *DUSP16* is a PCG which is selectively upregulated in T cells and regulates balance between helper T cell type 1 and 2 and a novel lncRNA in the antisense strand, *DUSP16-AS-1*, is highly upregulated in T cells (Figure 5B) (Musikacharoen et al. 2011). *FOXO1* is previously known regulatory T cell marker (Luo et al. 2016) and antisense novel lncRNA *FOXO1-AS-1* is also specifically expressed in T cells (Figure 5B). Additionally, *SLC8A1* was one of the PCGs markers that were highly expressed in DCs (data not shown), it's antisense lncRNA, *SLC8A1-AS-2*, was also specifically detected in DCs (Figure 5B). These antisense lncRNAs can be studied whether they have passenger or eRNA-like functions by analyzing the co-expression pattern with known markers. The SS2 data detected a much higher number of DEGs than that of 10X data especially in the novel lncRNAs (296 known and 453 novel lncRNAs) (Figure 6). In summary, Union set of 74 known/novel lncRNAs were identified in both platforms (Figure 7A). These common lncRNAs shared the similar expression levels and mostly expressed from the identical cell types (Figure 7B, C). Genes that were missed in 10X data could be resulted from low expression levels because of the technical limitation of the droplet-based platform, such as shallow sequencing depths and prevalent dropouts. In fact, it was confirmed that the expressions of commonly expressed genes were generally higher than

that of specifically expressed genes (Figure 7D). In particular, SS2-specific marker genes were often could not be detected in 10X because of their low expression levels in cells. In consequently, several lncRNA markers are reproducible between different scRNA-seq platforms that can be used as confident lncRNA markers.

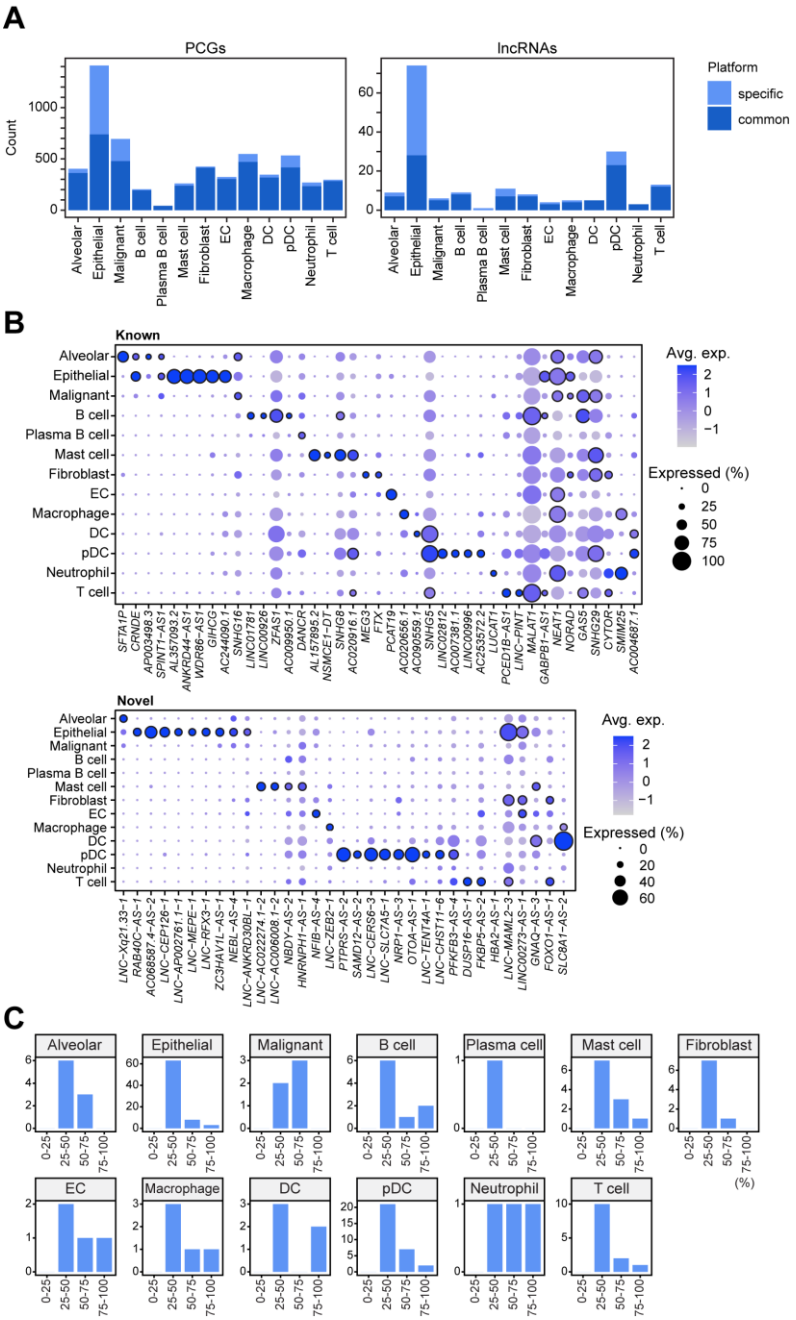


Figure 5. DE lncRNAs in each cell type in 10X data. (A) DEG counts in each cell type. 10X dataset specifically detected markers are colored light blue (specific), and markers that both detected in 10X and SS2 colored dark blue (common). (B) Expression dot plot of each DE lncRNAs. Known lncRNAs were indicated top 5 genes for each cell type (upper) and newly annotated novel lncRNAs have fully appeared (lower). (C) Distribution of lncRNA markers according to the cell fraction being expressed. Most of the genes are expressed in 25% to 75% of the corresponding cell type.

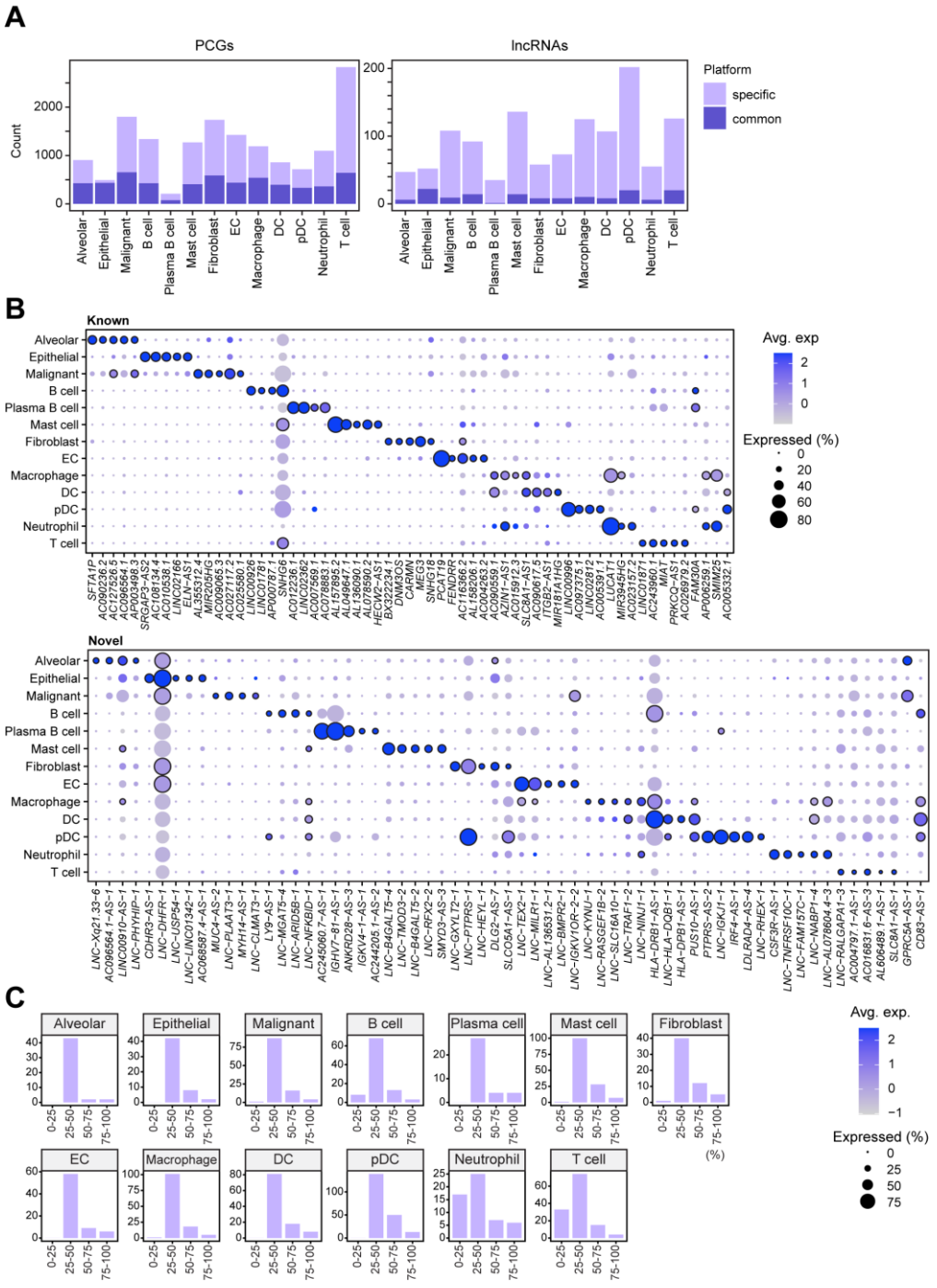


Figure 6. DE lncRNAs in each cell type in SMART-seq2 data. (A) DEG counts in each cell type. SS2 dataset specifically detected markers are colored light purple (specific), and markers that both detected in platforms colored dark purple (common). (B) Expression dot plot of each DE lncRNAs. Known lncRNAs (upper) and novel lncRNAs (lower) were indicated top 5 genes for each cell type. (C) Distribution of lncRNA markers according to the cell fraction being expressed.

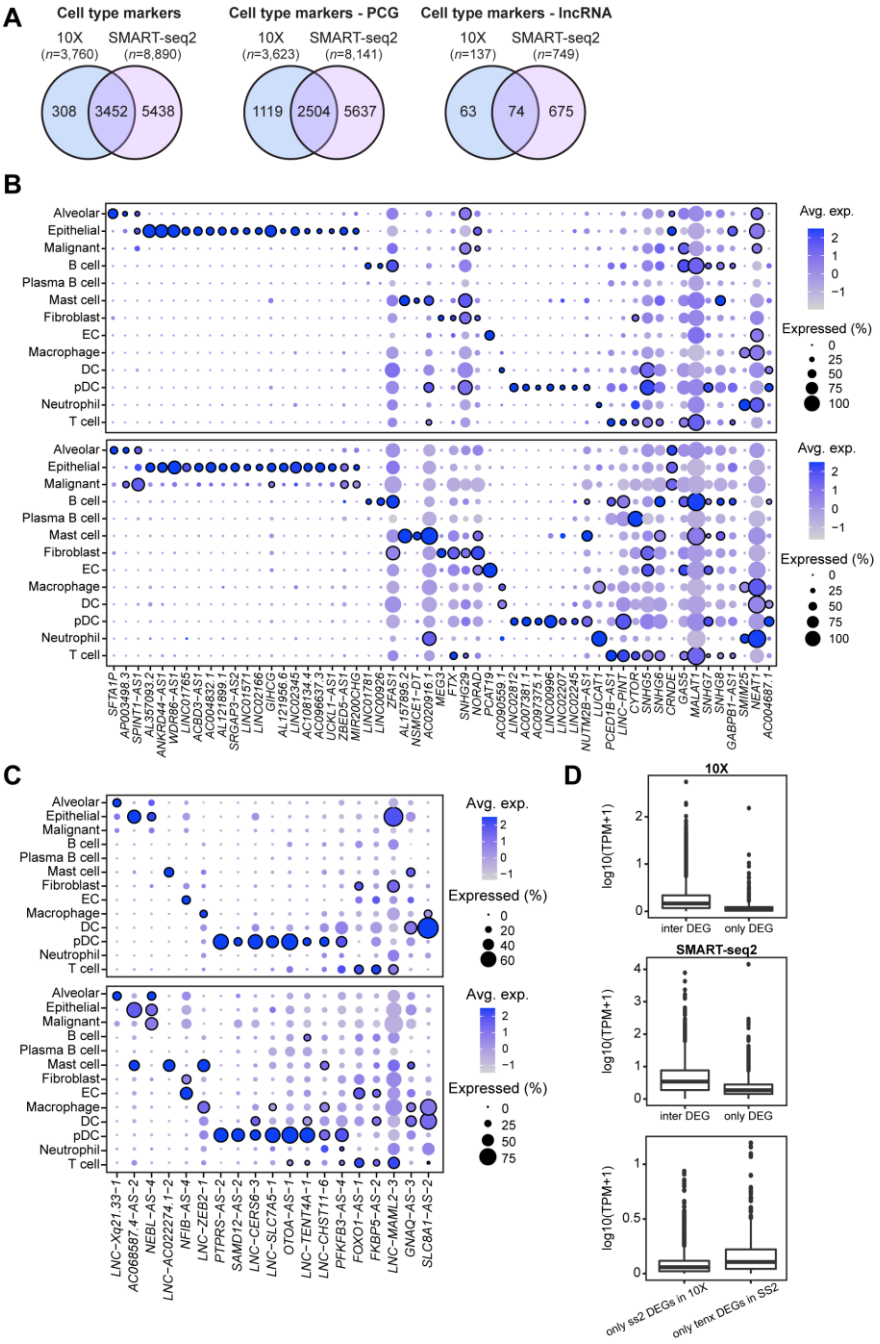


Figure 7. lncRNA expressions in different scRNA-seq platforms. (A) Counts of overlapped cell type markers between 10X and SMART-seq2 datasets (B) Expression of known lncRNAs (n=55). (C) Novel lncRNAs (n=19) (D) Box plot of log-transformed expression value (TPM+1.0; Y-axis) of lncRNA markers that are common to both platforms (inter DEG) and those detected in only a single platform (only DEG); 10X (top) and SS2 (middle) respectively. Expression of SS2-specific markers in the 10X platform, vice versa (bottom).

4. Tumor-specific lncRNAs in TIME

4.1. lncRNAs differentially expressed in tumor sites

Next, I tried to identify lncRNAs that are specifically dysregulated in tumors. Each cell type was first divided into group of cells from tumor or nontumor and differentially expressed genes were identified between the two conditions. To compare the expression with malignant cluster, epithelial cell from nontumor was integrated with alveolar cells.

In 10X data, a total of 32 genes were identified among which 9 were tumor-enriched genes and 25 were nontumor-enriched genes (Figure 8A). In epithelial, *SNHG29* was expressed in cells from tumor while *SFTA1P* that suppresses cell migration and invasion (Ma et al. 2018) and *MALAT1* were detected in cells from nontumor samples. *MALAT1* that is known to be involved in tumorigenesis in lung cancer, was highly upregulated in most cell types, especially in nontumor site of epithelial cells, fibroblasts, and DCs, but largely depleted in macrophages. Three lncRNAs (*GAS5, SNHG6,* and *SNHG5*) were highly expressed in the malignant cells, and eight lncRNAs (*SFTA1P, AP003498.3, MALAT1, CRNDE, LINC01578,* and *AC020916.1*, and novel lncRNAs *LNC-Xq21.33-1* and *LNC-MAML2-1*) were depleted. For example, *LINC01578* is known to be correlates with metastasis and prognosis in colon cancer although the expression was found in T cells (Liu et al. 2020). In the tumor B cells, genes related to antigen processing and presentation pathways were enriched (e.g., *HSPA1A*, *HSPA1B*, or *HSP90AA1*) but no lncRNAs were enriched compared to the cells from nontumor samples. In plasma B cells, *HSPA1A* was highly upregulated in tumor like B cells and *VIM*, *JCHAIN*,

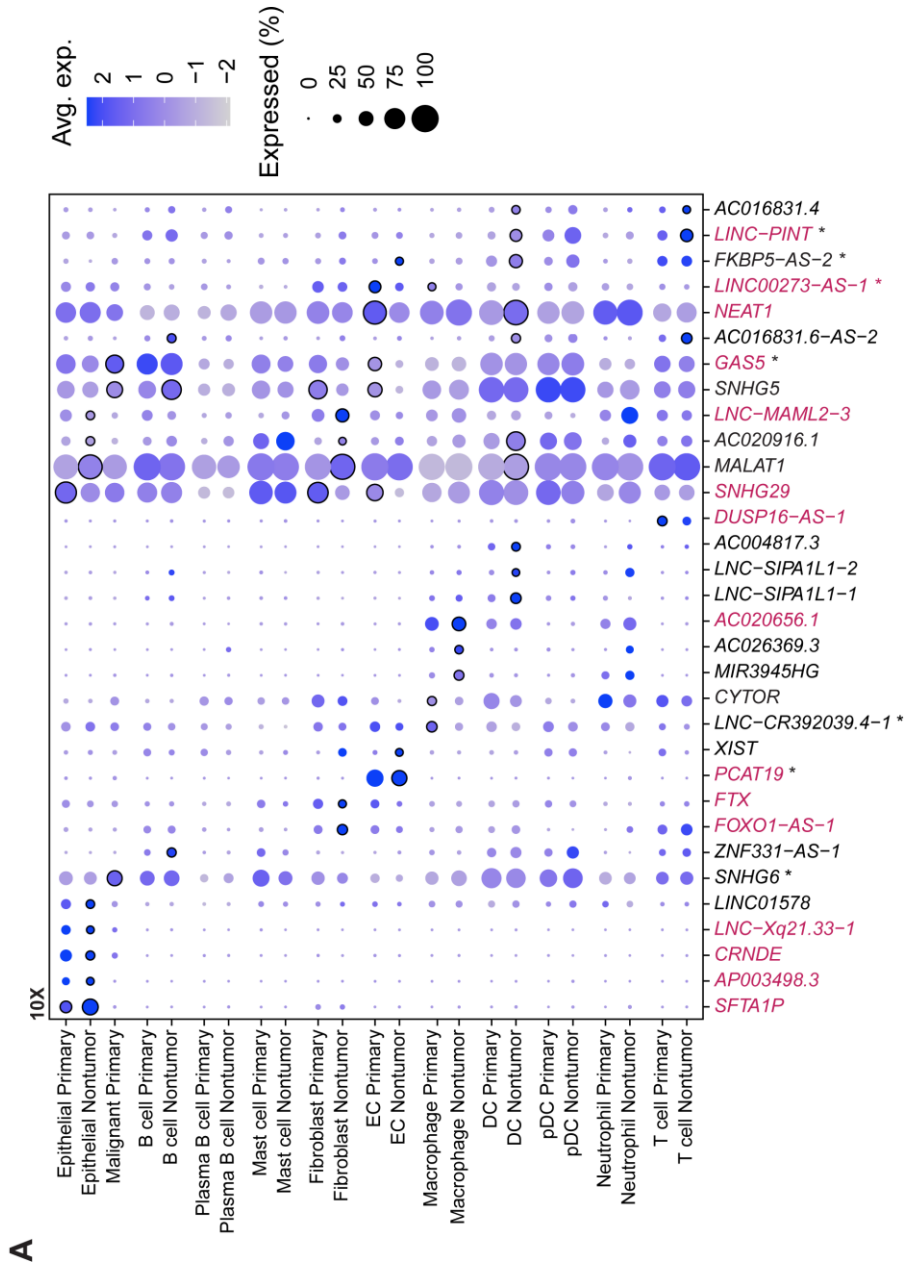
SFTPC, and *ZFP36* were upregulated in nontumor, but there were no specifically expressed lncRNAs. In fibroblasts, expression of SNHG family genes (*SNHG5* and *SNHG29*) was found in cells from tumor samples. In nontumor, *MALAT1*, *AC020916*.1, and *FTX* were expressed, and *LNC-MAML2-1* and *FOXO1-AS-1* which was also found as a cell type marker were identified as novel lncRNA markers. In ECs, *SNHG5*, *SNHG29*, and *NEAT1* were identified as tumor site markers, and *LINC00273-AS-1* was detected as a novel lncRNA. Interestingly, *GAS5*, known as a tumor suppressor gene, was highly expressed in ECs from tumor although the expression was less specific to a certain cell type. In nontumor, *PCAT19* was a very specific marker gene of ECs, and it was highly expressed in nontumor site than tumor. *FKBP5-AS-2* is a novel lncRNA antisense to a gene *FKBP5* which is known to regulate tumorigenesis and chemoresistance (Li et al. 2011) and detected with a low cell fraction but high expression level in ECs. Macrophages expressed a lot of tumor associated macrophage (TAM)-related PCGs such as *SPP1*, *CCL18*, and *GPNMB*, especially in tumor samples. As lncRNA markers, *LNC-CR392039.4-1*, *CYTOR*, and *LINC00273-AS-1* were identified which could be associated to the TAM-related functions in the tumor. In DCs, some non-specific marker genes were identified only in nontumor site, such as *NEAT1* and *MALAT1* which are lncRNAs known to be involved in tumor progression, and *LINC-PINT*, *AC020916*.1, *AC016831*.4, *AC004817*.3, and *FKBP5-AS-2* were detected. Most of the lncRNAs in T cells were observed at a similar level of expression regardless of context. As marker genes, one in tumor and three in nontumor were identified. *DUSP16-AS-1* that was a marker gene in

T cells was highly expressed in tumor and in nontumor, *AC016831.6-AS-2*, *LINC-PINT*, and *AC016831*.4 were detected.

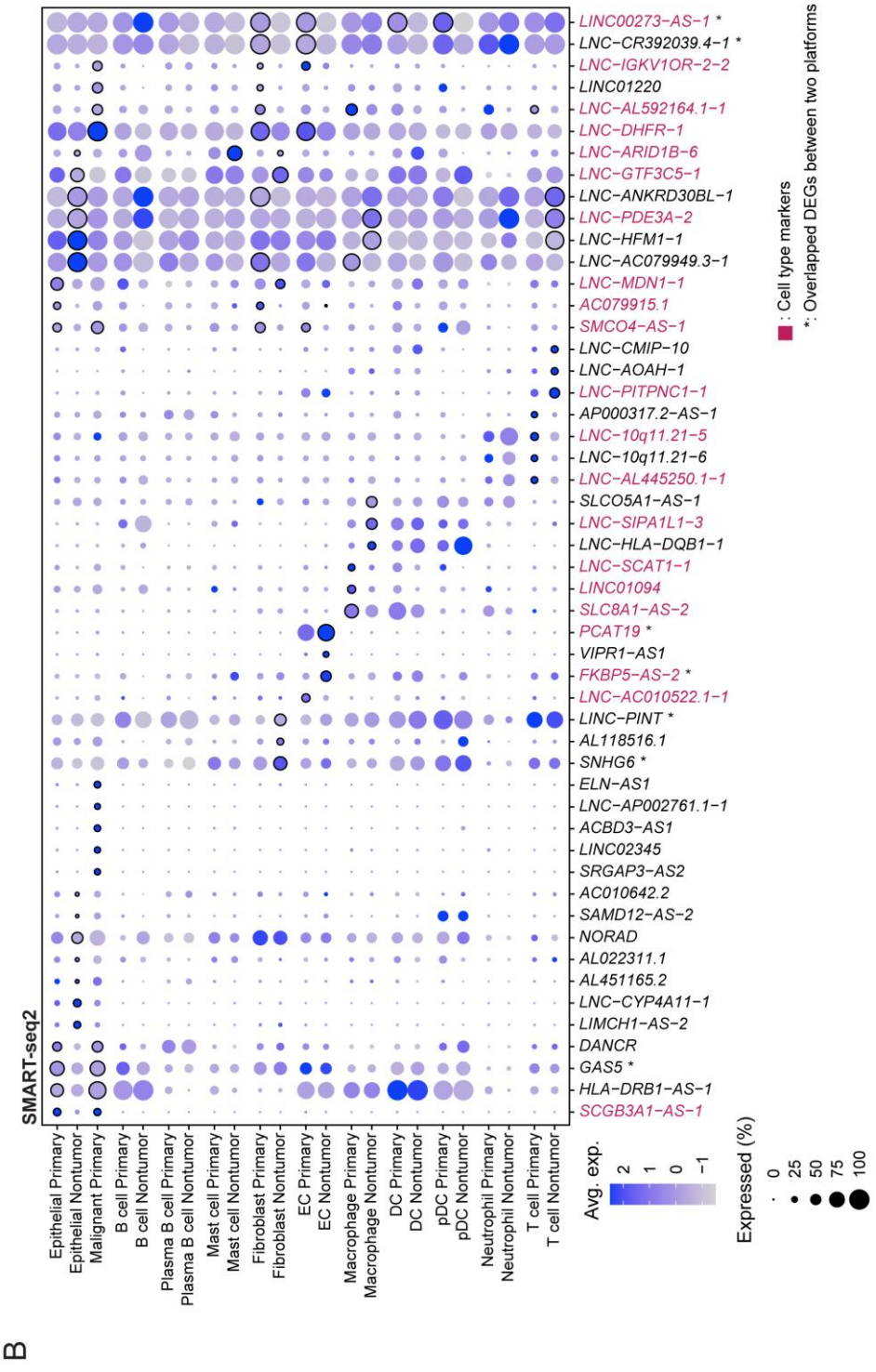
In the result, 15 lncRNAs (*SFTA1P, AP003498.3, CRNDE, LNC-Xq21.33-1, FOXO1-AS-1, FTX, PCAT19, AC020656.1, DUSP16-AS-1, SNHG29, LNC-MAML2-3, GAS5, NEAT1, LINC00273-AS-1,* and *LINC-PINT*) are cell type markers simultaneously dysregulated in tumor context. For example, *SNHG29* is a cell type marker that is specifically expressed in fibroblasts and is particularly expressed in those cells from a tumor site. As a result, further studies regarding functional validation of lncRNA related to tumor progression or suppression should focus on the actual cell types expressing lncRNAs in cancer.

In SS2 data, because it has higher sequencing depth than those of 10X data, much more marker genes were identified (Figure 8B). Total of 241 lncRNAs were found as differentially expressed between tumor and nontumor samples among which 146 were highly expressed in tumor and 101 were upregulated in nontumor. All tumor/nontumor markers from 10X were detected in SS2, but most were filtered out by insignificant q-value (adjusted by false discovery rate). The following seven marker genes, *SNHG6*, *PCAT19*, *LNC-CR392039.4-1*, *GAS5*, *LINC00273-AS-1*, *FKBP5-AS-2*, and *LINC-PINT* were identified in both datasets, and four of these lncRNAs were also matched the expressing cell types: *PCAT19* in ECs nontumor, *GAS5* in malignant cells, *LINC00273-AS-1* in ECs tumor, and *FKBP5-AS-2* in ECs nontumor. These markers can be used as confident candidates for further studies.

DEGs between tumor and nontumor at single-cell level (single-cell DEGs) could be different with those at the bulk level because single-cell DEGs were identified using the individual cell type rather than mixture of heterogeneous cells. In fact, more than half of the single-cell DEGs (65.62% for 10X and 67.22% for SS2) were only found in the scRNA-seq data, indicating that scRNA-seq would be specialized to the detection of markers expressed by minor cell types in TIME (Figure 9B). Only 15.6% of 10X and 30.3% of SS2 single-cell DEGs were restated with bulk DEG analysis, suggesting that the SS2 platform is more sensitive to detect DEGs than the 10X platform. Therefore, the cell type-specific expression and function of many lncRNAs previously studied in bulk tumor samples can be validated using the variety of available scRNA-seq datasets (Table 5). This process could provide new insights about known and novel lncRNA functions and regulatory roles in specific cell types in the TIME.


A
Avg. exp.
2
1
0
−1
−2
Expressed (%)
0
25
50
75
100
10X
Epithelial Primary
Epithelial Nontumor
Malignant Primary
B cell Primary
B cell Nontumor
Plasma B cell Primary
Plasma B cell Nontumor
Mast cell Primary
Mast cell Nontumor
Fibroblast Primary
Fibroblast Nontumor
EC Primary
EC Nontumor
Macrophage Primary
Macrophage Nontumor
DC Primary
DC Nontumor
pDC Primary
pDC Nontumor
Neutrophil Primary
Neutrophil Nontumor
T cell Primary
T cell Nontumor
AC016831.4
LINC-PINT *
FKBP5-AS-2 *
LINC00273-AS-1 *
NEAT1
AC016831.6-AS-2
GAS5 *
SNHG5
LNC-MAML2-3
AC020916.1
MALAT1
SNHG29
DUSP16-AS-1
AC004817.3
LNC-SIPA1L1-2
LNC-SIPA1L1-1
AC020656.1
AC026369.3
MIR3945HG
CYTOR
LNC-CR392039.4-1 *
XIST
PCAT19 *
FTX
FOXO1-AS-1
ZNF331-AS-1
SNHG6 *
LINC01578
LNC-Xq21.33-1
CRNDE
AP003498.3
SFTA1P

(Continued)



B

Figure 8. DE lncRNAs between tumor and nontumor in each cell type. (A) Dot plot of tumor vs nontumor DE lncRNAs in 10X data. lncRNAs that cell type specifically expressed and showed differences in expression by tumor context were colored in pink. (B) Differentially expressed lncRNAs in SS2 data.

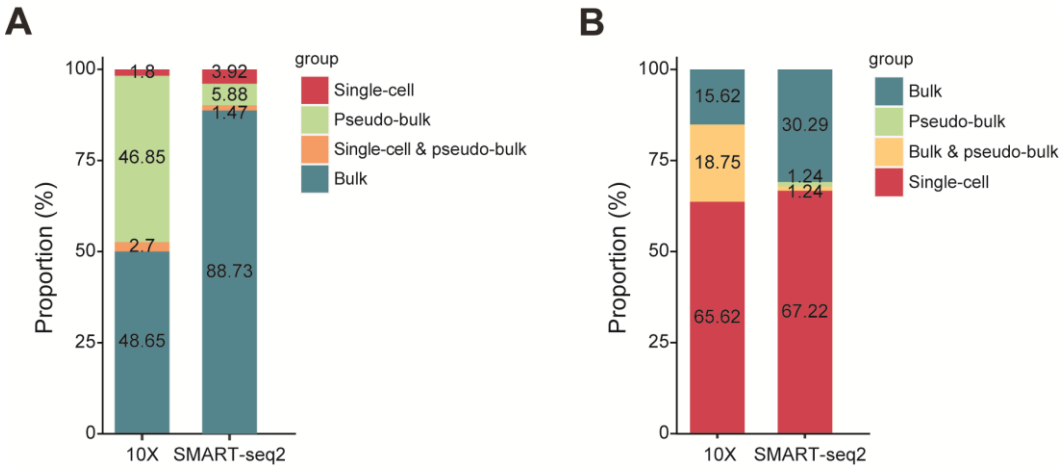


Figure 9. Comparison of DE lncRNAs in scRNA-seq and bulk RNA-seq. (A) The proportion of bulk DE lncRNAs between tumor and nontumor samples overlapped with those from single-cell RNA-seq data or pseudo-bulk RNA-seq data. (B) The proportion of single-cell DE lncRNAs between tumor and nontumor samples overlapped with those from bulk RNA-seq data or pseudo-bulk RNA-seq data. Single-cell DEGs were acquired by comparing gene expression between single cells from tumor and nontumor samples for each cell type. Bulk and pseudo-bulk DEGs were acquired by comparing gene expression between bulk RNA-seq data from paired tumor and nontumor samples and between pseudo-bulk data transformed from scRNA-seq data from tumor and nontumor tissue, respectively.

4.2. Validation of lncRNA expression patterns in breast and colon cancer

I additionally analyzed lncRNA cell type markers in publicly available breast cancer and colon cancer scRNA-seq datasets. Breast cancer is consisted of 14 treatment-naïve breast cancer patients sequenced by using a 10X Chromium 5'-seq platform (Qian et al. 2020). As a result, 26 lncRNAs were identified in breast cancer (Figure 10A), among which eight were identified from the similar cell types in lung cancer datasets; *LINC00926, LINC00863, NUTM2A-AS1, NUTM2B-AS1, EPB41L4A-AS1, MIR22HG, LINC01094,* and *PCAT19* (Figure 10B). Moreover, reanalysis of BIGTranscriptome v1.2 using a colon cancer data (10X Chromium 3'-seq) (Lee et al. 2020) showed that several lncRNAs were also expressed in the same cell types (Figure 10C; unpublished data). Following lncRNAs; *LINC00926, ZFAS1, MEG3, PCAT19, LINC-PINT, MALAT1, SNHG,* and *NEAT1*, were defined as shared cell type markers in colon and breast cancer (Figure 10B, C). These lncRNAs, which are detected as cell type markers in various cancer types, are mostly found in infiltrating immune cell types such as myeloids or ECs. Taken together, lncRNA can be used as a significant cell type marker regardless of sequencing platforms or cancer types. This indicates that lncRNAs are not only expressed in cell type-specific manner but are reproducible indicators that can be used pan-cancer.

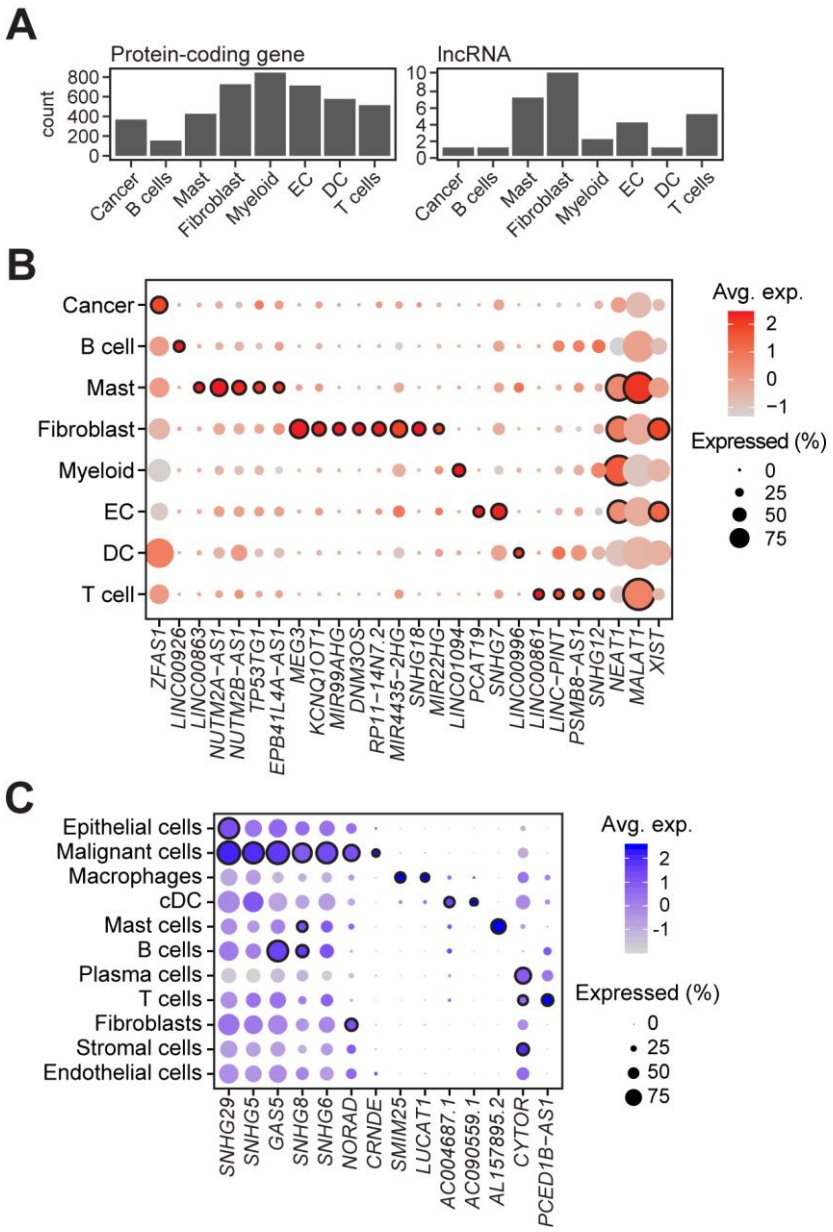


Figure 10. Cell type markers of BRCA and CRC scRNA-seq data. (A) Bar plots represent marker genes in each cell type. (B) Expression dot plot of lncRNA markers. *LINC00926, LINC00863, NUTM2A-AS1, NUTM2B-AS1, EPB41L4A-AS1, IR22HG, LINC01094, PCAT19* were identified as lung cancer cell type-specific markers. *ZFAS1, KCNQ1OT1, RP11-14N7.2, SNHG18, XIST* were detected markers in colon cancer. (C) Common cell type-specific lncRNA markers in lung cancer and colon cancer (unpublished data)

4.3. Expression profiling of well-known lncRNA markers at lung cancer single-cell resolution

Additionally, lncRNAs that were reported to play crucial roles in cancer progression related mechanisms, such as oncogenic or tumor suppressive roles, were confirmed in NSCLC single-cell level. I also included lncRNAs specifically expressed in lymphoid and myeloid cells from literatures. Overall, the expression patterns were generally agreed between the two scRNA-seq platforms (Figure 11). Some of the oncogenic lncRNAs such as *ZFAS1*, *PVT1*, or *HIF1A-AS3* were expressed in multiple cell types in TIME, while others were specifically expressed in various non-cancerous cells, such as *PACERR* in mast cells, *HAND2-AS1*, *MIR100HG*, and *TEX41* in fibroblasts, *CRNDE*, and *HULC* in alveolar, epithelial, and malignant, or *IFNG-AS1* in plasma B cells. There were a couple of disagreements in the expression patterns between the datasets. For example, *H19* was highly expressed in the malignant cell cluster from the 10X data, but not in the SS2 data. In SS2, *ZFAS1* and *SNHG6* are particularly highly expressed in B cells than in others. Tumor suppressive lncRNAs, such as *GAS5*, *XIST*, or *MEG3,* were expressed in higher levels from non-malignant cell types than malignant cells. Moreover, *H19* and *LUCAT1* that play oncogenic roles in cancer cells while also having functional roles in immune cells were expressed malignant and neutrophils. As a result, by confirming the expression of previously known lncRNAs in scRNA-seq data, it was possible to trace the cell type in which cancer-related lncRNAs are actually being

expressed. Because scRNA-seq would be more specific to the detection of markers for major and minor cell types in the TIME, lncRNAs previously studied in bulk tumor samples can be validated using scRNA-seq data.

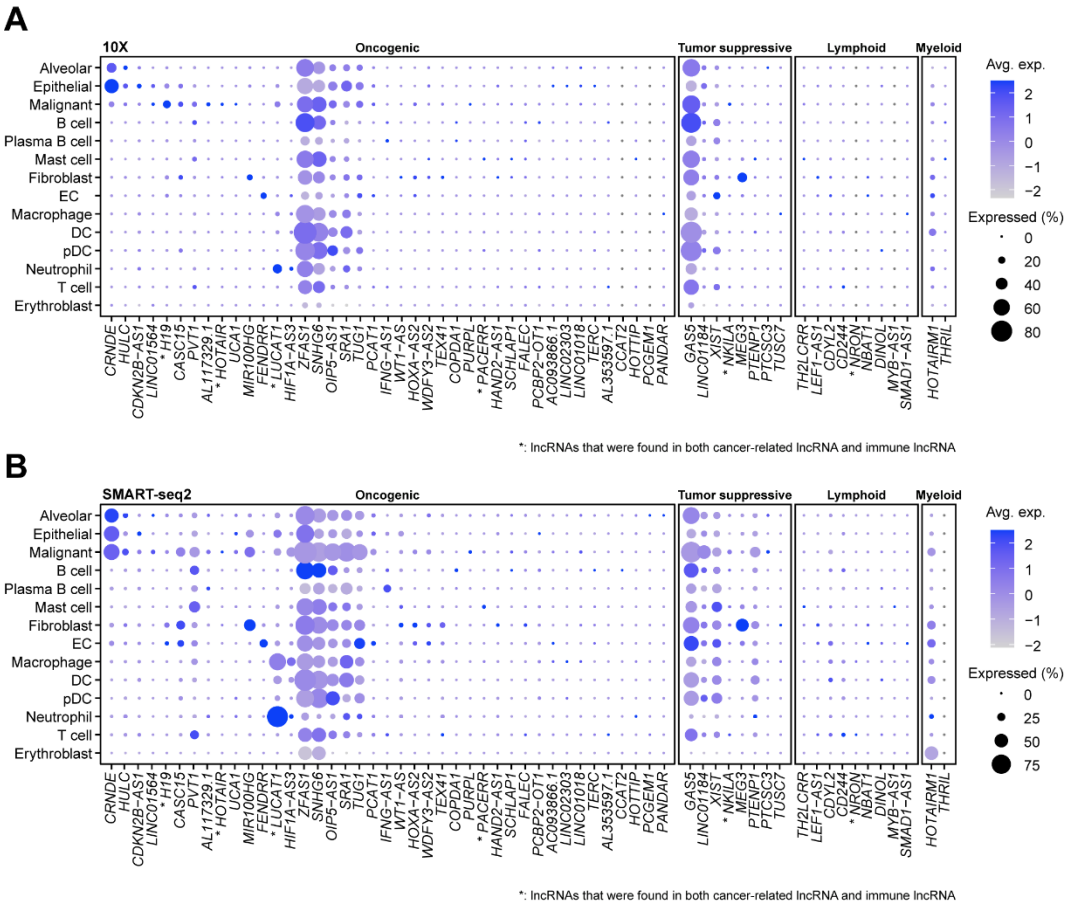


Figure 11. Expression landscape of lncRNAs from literatures. lncRNAs list representing from left to right: oncogenic, tumor suppressive, lymphoid- and myeloid-lineage cell type-specific lncRNAs. (A) 10X (B) SS2. lncRNAs are summarized in eg park *et al*.

Conclusion

Here, I identified cell type- and tumor/nontumor-specific lncRNAs in TIME using scRNA-seq data. By establishing additional cancer/immune cell type specific lncRNA transcriptome, it was possible to supplement insufficient cancer- and cell type-specific lncRNA annotation. Although lncRNA has a relatively lower expression than PCGs, they are detected in single-cell data and well agreed with bulk RNA-seq data, and also cell type-specific expression was confirmed at the single-cell level. Furthermore, SS2 platform which has a higher sequencing depth than the 10X data was more sensitive to lncRNA detection and identification of DEGs in the TIME. After validating that lncRNA was sufficiently detected at the single-cell level, cell type-specific lncRNAs were identified. In case of lncRNA markers, it was possible to validate in which cell types express well-known cancer-related lncRNAs that were previously studied in bulk tumor samples using scRNA-seq. More lncRNA markers were found in SS2 than 10X and more than 90% of 10X markers were reproducible in the SS2 data. Several lncRNA markers that are reproducible between different scRNA-seq platforms of lung cancer data can be used as helpful cancer-specific lncRNA markers. Additionally, some lncRNAs that are expressed in cell type- or tumor/nontumor-specific manner would indicate dynamic expression patterns depending on cell states. It is also needed to study novel lncRNAs named AS- and LNC-, which are newly annotated positioned in near known genes, whether coexpressed with PCGs and how they function in

TIME. Moreover, shared lncRNA markers detected in both platforms were expressed not only in lung cancer but also in other cancer types, breast and colon cancer. This study suggests cell type- and tumor type-specifically expressed lncRNA candidates and offers novel treatment target lncRNAs.

Based on these results, experimental validations of known lncRNAs that were previously studied in bulk tumor samples and novel lncRNA markers could be designed to target cell type in TIME not in only cancer cell line. In addition, these newly constructed cell type- and tumor/nontumor-specific lncRNA markers can be used as promising biomarkers for investigation of cancer-related mechanisms and development of new therapeutic targets. Single-cell analysis of lncRNA would highlight new area of TIME and help overcome the limitations that previous lncRNA studies had.

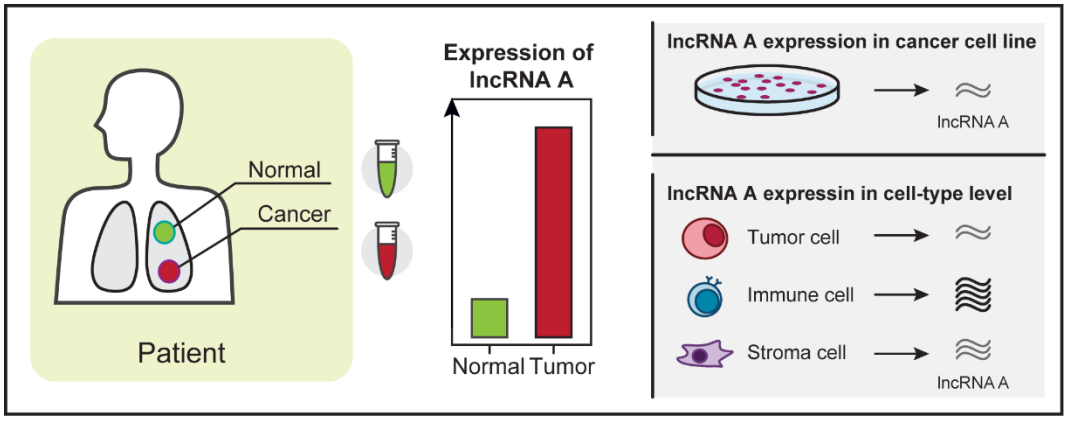


Figure 12. Precision profiling of lncRNAs using high resolution single-cell data. Expression of lncRNAs dysregulated in bulk tumor samples were validated in cancer cell lines. The single-cell analysis can determine a specific cell type in which lncRNA is actually expressed and dysregulated in TIME. Based on the results of cell type- and tumor/nontumor-specific markers, scRNA-seq would be a powerful strategy to detect markers of diverse cell types in TIME.

Tables

Table 1. 10X cell type markers top 100

Gene	p-value	Avg logFC	Pct.1	Pct.2	Adjust p-value	Cell type	Type
SFTA1P	0	1.595	0.583	0.012	0	Alveolar	knownLnc
LNC-Xq21.33-1	0	0.646	0.279	0.015	0	Alveolar	novelLnc
CRNDE	0	0.612	0.315	0.024	0	Alveolar	knownLnc
AP003498.3	0	0.557	0.254	0.003	0	Alveolar	knownLnc
NEAT1	3.4E-189	0.560	0.718	0.703	2.84E-184	Alveolar	knownLnc
AL357093.2	0	1.995	0.786	0.003	0	Epithelial	knownLnc
ANKRD44-AS1	0	1.734	0.78	0.014	0	Epithelial	knownLnc
WDR86-AS1	0	1.574	0.749	0.016	0	Epithelial	knownLnc
LINC01765	0	0.915	0.471	0.001	0	Epithelial	knownLnc
ACBD3-AS1	0	0.875	0.494	0.001	0	Epithelial	knownLnc
AC004832.1	0	0.842	0.46	0.001	0	Epithelial	knownLnc
AL121899.1	1.78e-322	0.844	0.46	0.001	1.47e-317	Epithelial	knownLnc
SRGAP3-AS2	9.14e-322	0.837	0.41	0	7.54e-317	Epithelial	knownLnc
LNC-LBCS	6.3E-294	0.719	0.41	0.001	5.20E-289	Epithelial	knownLnc
LINC01571	6.5E-286	0.712	0.367	0	5.32E-281	Epithelial	knownLnc
LINC02166	1E-267	0.677	0.39	0.001	8.58E-263	Epithelial	knownLnc
GIHCG	9.7E-253	1.218	0.688	0.037	8.00E-248	Epithelial	knownLnc
AC130456.2	5.5E-233	0.727	0.379	0.002	4.51E-228	Epithelial	knownLnc
RAB40C-AS-1	1.2E-229	0.533	0.303	0	9.67E-225	Epithelial	novelLnc
AC068587.4-AS-2	6.8E-217	0.832	0.477	0.012	5.64E-212	Epithelial	novelLnc
CASC2	4.3E-216	0.661	0.37	0.003	3.54E-211	Epithelial	knownLnc
LINC01513	1.3E-212	0.492	0.292	0	1.07E-207	Epithelial	knownLnc
LNC-CEP126-1	2.2E-210	0.706	0.384	0.004	1.85E-205	Epithelial	novelLnc
AC244090.1	1.1E-205	1.188	0.659	0.074	8.90E-201	Epithelial	knownLnc
AC027237.5	1.1E-203	0.500	0.318	0.001	8.99E-199	Epithelial	knownLnc
LINC02345	2.2E-195	0.859	0.494	0.019	1.85E-190	Epithelial	knownLnc
AC108134.4	4.9E-194	0.507	0.289	0.001	4.06E-189	Epithelial	knownLnc
MUC12-AS1	2.4E-187	0.768	0.439	0.012	2.00E-182	Epithelial	knownLnc
AC234582.1	6.3E-183	0.580	0.306	0.002	5.19E-178	Epithelial	knownLnc
LNC-MEPE-1	3.8E-175	0.526	0.266	0.001	3.12E-170	Epithelial	novelLnc
SERTAD4-AS1	2.1E-159	0.636	0.399	0.013	1.73E-154	Epithelial	knownLnc

MIR99AHG	2.5E-158	0.682	0.422	0.016	2.07E-153	Epithelial	knownLnc
CRNDE	5.8E-157	0.827	0.506	0.033	4.80E-152	Epithelial	knownLnc
AZIN1-AS1	3.9E-150	0.771	0.399	0.015	3.23E-145	Epithelial	knownLnc
NRAV	4.6E-146	0.610	0.37	0.012	3.79E-141	Epithelial	knownLnc
LINC00271	8.7E-143	0.485	0.303	0.005	7.15E-138	Epithelial	knownLnc
ERICH6-AS1	6.9E-142	0.565	0.355	0.012	5.71E-137	Epithelial	knownLnc
COLCA1	8.8E-125	0.493	0.312	0.009	7.27E-120	Epithelial	knownLnc
AC006230.1	1.9E-123	0.486	0.303	0.009	1.56E-118	Epithelial	knownLnc
AC007405.4	5.8E-115	0.497	0.318	0.013	4.80E-110	Epithelial	knownLnc
ZBED5-AS1	5.8E-106	0.658	0.445	0.047	4.75E-101	Epithelial	knownLnc
MIR200CHG	1.1E-105	0.573	0.335	0.019	9.23E-101	Epithelial	knownLnc
TP53TG1	9.6E-103	0.803	0.572	0.101	7.90E-98	Epithelial	knownLnc
ZC3HAV1L-AS-1	5.4E-102	0.540	0.312	0.016	4.43E-97	Epithelial	novelLnc
AC060780.1	2.03E-94	0.501	0.344	0.026	1.68E-89	Epithelial	knownLnc
LNC-MAML2-3	7.03E-47	0.519	0.702	0.319	5.80E-42	Epithelial	novelLnc
NEAT1	0	0.532	0.572	0.719	0	Malignant	knownLnc
LINC01781	0	1.372	0.342	0.004	0	B cell	knownLnc
LINC00926	0	0.887	0.269	0.004	0	B cell	knownLnc
GAS5	1.74e-318	0.538	0.754	0.541	1.43e-313	B cell	knownLnc
AC009950.1	3.4E-232	0.533	0.257	0.081	2.81E-227	B cell	knownLnc
AL157895.2	0	1.638	0.61	0.001	0	Mast cell	knownLnc
LNC-AC022274.1-2	0	0.936	0.332	0	0	Mast cell	novelLnc
NSMCE1-DT	0	0.829	0.293	0.012	0	Mast cell	knownLnc
LNC-AC006008.1-2	8.9E-304	0.747	0.281	0.012	7.34E-299	Mast cell	novelLnc
SNHG8	7.9E-124	0.694	0.598	0.319	6.47E-119	Mast cell	knownLnc
SNHG29	2.7E-120	0.577	0.839	0.609	2.26E-115	Mast cell	knownLnc
GNAQ-AS-3	1.35E-88	0.569	0.298	0.092	1.11E-83	Mast cell	novelLnc
AC020916.1	1.53E-82	0.594	0.557	0.302	1.26E-77	Mast cell	knownLnc
MEG3	0	1.018	0.309	0.004	0	Fibroblast	knownLnc
PCAT19	0	1.934	0.571	0.012	0	EC	knownLnc
NFIB-AS-4	0	1.234	0.288	0.023	0	EC	novelLnc
LINC00273-AS-1	1.2E-145	0.645	0.303	0.185	9.65E-141	EC	novelLnc
AC020656.1	0	1.566	0.473	0.024	0	Macrophage	knownLnc
SMIM25	0	1.298	0.612	0.018	0	Macrophage	knownLnc
SLC8A1-AS-2	0	0.552	0.264	0.031	0	Macrophage	novelLnc
NEAT1	0	0.516	0.904	0.647	0	Macrophage	knownLnc
SLC8A1-AS-2	0	1.685	0.764	0.07	0	DC	novelLnc

PTPRS-AS-2	8.6E-243	1.445	0.552	0	7.12E-238	pDC	novelLnc
LINC02812	3.6E-147	0.996	0.491	0.003	2.98E-142	pDC	knownLnc
AC007381.1	2.7E-136	0.888	0.382	0.001	2.24E-131	pDC	knownLnc
AC097375.1	1.4E-114	0.652	0.261	0	1.17E-109	pDC	knownLnc
SAMD12-AS-2	1.92E-85	0.591	0.279	0.001	1.58E-80	pDC	novelLnc
LNC-CERS6-3	1.63E-82	0.999	0.521	0.034	1.34E-77	pDC	novelLnc
AC008764.7	1.19E-80	0.582	0.352	0.007	9.83E-76	pDC	knownLnc
LINC00996	4.46E-76	0.893	0.388	0.012	3.68E-71	pDC	knownLnc
LNC-SLC7A5-1	9.69E-53	0.948	0.43	0.04	7.99E-48	pDC	novelLnc
NRP1-AS-3	9.4E-52	0.704	0.388	0.032	7.75E-47	pDC	novelLnc
LINC02207	4.17E-48	0.513	0.267	0.01	3.44E-43	pDC	knownLnc
OTOA-AS-1	4.88E-47	0.857	0.588	0.114	4.03E-42	pDC	novelLnc
SNHG5	1.47E-30	0.643	0.921	0.594	1.21E-25	pDC	knownLnc
AL135925.1	1.95E-29	0.583	0.273	0.029	1.61E-24	pDC	knownLnc
LNC-TENT4A-1	2.72E-28	0.539	0.267	0.029	2.24E-23	pDC	novelLnc
AC004687.1	5.05E-28	0.661	0.467	0.115	4.17E-23	pDC	knownLnc
AC253572.2	4.36E-25	0.735	0.382	0.082	3.60E-20	pDC	knownLnc
SNHG7	8.51E-22	0.508	0.552	0.202	7.02E-17	pDC	knownLnc
LNC-CHST11-6	2.19E-13	0.484	0.315	0.1	1.81E-08	pDC	novelLnc
SMIM25	1.28E-90	1.891	0.674	0.147	1.05E-85	Neutrophil	knownLnc
LUCAT1	2.18E-32	0.912	0.289	0.045	1.80E-27	Neutrophil	knownLnc
NEAT1	1.68E-14	0.731	0.879	0.703	1.38E-09	Neutrophil	knownLnc
DUSP16-AS-1	0	1.323	0.313	0.016	0	T cell	novelLnc
PCED1B-AS1	0	1.037	0.437	0.178	0	T cell	knownLnc
LINC-PINT	0	0.895	0.34	0.154	0	T cell	knownLnc
FOXO1-AS-1	0	0.826	0.309	0.129	0	T cell	novelLnc
FKBP5-AS-2	0	0.822	0.312	0.146	0	T cell	novelLnc
CYTOR	0	0.774	0.371	0.234	0	T cell	knownLnc
MALAT1	0	0.729	0.994	0.952	0	T cell	knownLnc
GABPB1-AS1	0	0.668	0.276	0.163	0	T cell	knownLnc
AC020916.1	0	0.505	0.318	0.296	0	T cell	knownLnc
LNC-MAML2-3	0	0.504	0.335	0.308	0	T cell	novelLnc

Table 2. SMART-seq2 cell type markers top 100

Gene	p-value	Avg logFC	Pct.1	Pct.2	Adjust p-value	Cell type	Type
LNC-Xq21.33-6	0	3.399	0.284	0.004	0	Alveolar	novelLnc
SFTA1P	0	2.589	0.394	0.038	0	Alveolar	knownLnc
AC090236.2	1.8E-291	4.442	0.296	0.038	1.52E-286	Alveolar	knownLnc
AC096564.1	2.1E-226	2.436	0.324	0.074	1.74E-221	Alveolar	knownLnc
AC096564.1-AS-1	1E-223	3.065	0.333	0.07	8.56E-219	Alveolar	novelLnc
WDR86-AS1	3.81E-50	2.591	0.708	0.046	3.14E-45	Epithelial	knownLnc
CDHR3-AS-1	5.04E-38	3.525	0.472	0.019	4.16E-33	Epithelial	novelLnc
ANKRD44-AS1	4.92E-36	2.593	0.514	0.031	4.06E-31	Epithelial	knownLnc
SRGAP3-AS2	1.03E-35	3.313	0.444	0.018	8.47E-31	Epithelial	knownLnc
LINC01765	3.13E-34	2.720	0.431	0.017	2.59E-29	Epithelial	knownLnc
AC113349.1	6.05E-34	2.459	0.403	0.014	4.99E-29	Epithelial	knownLnc
LINC02832	5.84E-29	2.480	0.389	0.02	4.81E-24	Epithelial	knownLnc
AC108134.4	9.4E-29	3.821	0.431	0.029	7.76E-24	Epithelial	knownLnc
AC010538.1	7.13E-22	2.748	0.403	0.052	5.88E-17	Epithelial	knownLnc
LINC02166	1.87E-21	2.816	0.333	0.024	1.54E-16	Epithelial	knownLnc
ELN-AS1	2.37E-20	2.920	0.375	0.035	1.96E-15	Epithelial	knownLnc
LNC-USP54-1	7.15E-19	2.881	0.278	0.015	5.9E-14	Epithelial	novelLnc
LNC-LINC01342-1	1.22E-14	2.672	0.306	0.051	1.01E-09	Epithelial	novelLnc
MIR200CHG	4.35E-11	2.489	0.278	0.072	3.59E-06	Epithelial	knownLnc
AL355312.4	0	2.648	0.391	0.007	0	Malignant	knownLnc
LINC00926	3.4E-279	3.499	0.393	0.03	2.81E-274	B cell	knownLnc
LINC01781	3.6E-211	4.470	0.267	0.008	2.98E-206	B cell	knownLnc
AC245060.7-AS-1	0	6.983	0.898	0.057	0	Plasma B cell	novelLnc
IGHV7-81-AS-1	0	6.556	0.998	0.125	0	Plasma B cell	novelLnc
ANKRD28-AS-3	0	4.541	0.597	0.098	0	Plasma B cell	novelLnc
IGKV4-1-AS-1	0	4.254	0.272	0.001	0	Plasma B cell	novelLnc
AC012236.1	0	3.165	0.549	0.014	0	Plasma B cell	knownLnc
LNC-IGKJ1-1	2.7E-303	2.552	0.307	0.016	2.19E-298	Plasma B cell	novelLnc
AC244205.1-AS-2	7.2E-298	5.963	0.285	0.012	5.92E-293	Plasma B cell	novelLnc
AL157895.2	0	4.504	0.754	0.008	0	Mast cell	knownLnc
AL049647.1	0	3.713	0.47	0.009	0	Mast cell	knownLnc
LNC-AC022274.1-2	0	2.622	0.514	0.003	0	Mast cell	novelLnc
LNC-B4GALT5-4	6E-279	2.747	0.661	0.135	4.93E-274	Mast cell	novelLnc
AL136090.1	1.3E-253	5.367	0.291	0.002	1.04E-248	Mast cell	knownLnc

AL078590.2	1.3E-253	2.843	0.425	0.033	1.11E-248	Mast cell	knownLnc
LNC-TMOD3-2	5.7E-192	2.766	0.451	0.092	4.70E-187	Mast cell	novelLnc
HECW2-AS1	1.8E-187	3.667	0.328	0.018	1.46E-182	Mast cell	knownLnc
LNC-B4GALT5-2	3.3E-146	3.116	0.405	0.07	2.71E-141	Mast cell	novelLnc
LNC-RFX2-2	7.3E-146	2.749	0.386	0.063	6.04E-141	Mast cell	novelLnc
LNC-RFX2-1	5.5E-135	2.629	0.426	0.091	4.58E-130	Mast cell	novelLnc
SMYD3-AS-3	6.8E-135	3.027	0.433	0.102	5.64E-130	Mast cell	novelLnc
LNC-B4GALT5-5	8.4E-117	2.517	0.388	0.072	6.95E-112	Mast cell	novelLnc
MAPK1-AS-1	1.6E-99	2.597	0.254	0.025	1.32E-94	Mast cell	novelLnc
ANKRD28-AS-2	1.1E-77	2.521	0.281	0.06	9.09E-73	Mast cell	novelLnc
BX322234.1	0	3.119	0.294	0.013	0	Fibroblast	knownLnc
LNC-TEX2-1	0	2.979	0.795	0.161	0	EC	novelLnc
PCAT19	0	2.851	0.782	0.029	0	EC	knownLnc
FENDRR	1.9E-272	2.903	0.256	0.01	1.54E-267	EC	knownLnc
LNC-KYNU-1	0	2.680	0.254	0.032	0	Macrophage	novelLnc
LNC-RASGEF1B-2	0	2.468	0.294	0.04	0	Macrophage	novelLnc
PTPRS-AS-2	2.5E-304	4.323	0.633	0.002	2.08E-299	pDC	novelLnc
LNC-IGKJ1-1	3.7E-244	3.689	0.783	0.024	3.03E-239	pDC	novelLnc
LINC00996	4.9E-191	3.353	0.671	0.027	4.06E-186	pDC	knownLnc
IRF4-AS-1	3.8E-165	3.668	0.546	0.015	3.11E-160	pDC	novelLnc
AC097375.1	4.2E-162	4.135	0.358	0.001	3.50E-157	pDC	knownLnc
LDLRAD4-AS-4	2.3E-148	3.918	0.579	0.051	1.90E-143	pDC	novelLnc
LINC02812	1.4E-127	3.926	0.383	0.006	1.16E-122	pDC	knownLnc
FCHSD2-AS-3	1.4E-120	2.787	0.6	0.056	1.16E-115	pDC	novelLnc
LNC-FCHSD2-3	2E-109	2.578	0.512	0.038	1.6E-104	pDC	novelLnc
LNC-RHEX-1	1.8E-103	4.504	0.304	0.005	1.46E-98	pDC	novelLnc
LNC-SLC7A5-1	1.55E-88	2.514	0.671	0.16	1.28E-83	pDC	novelLnc
INPP4A-AS-1	5.9E-86	2.534	0.538	0.083	4.86E-81	pDC	novelLnc
STX17-AS1-AS-1	1.77E-74	3.161	0.408	0.04	1.46E-69	pDC	novelLnc
AC005391.1	2.87E-74	3.195	0.325	0.019	2.37E-69	pDC	knownLnc
LNC-TGFBR1-1	2.78E-69	2.484	0.45	0.072	2.3E-64	pDC	novelLnc
LNC-SULF2-1	5E-60	3.108	0.3	0.021	4.13E-55	pDC	novelLnc
AC005332.1	1.09E-54	2.890	0.408	0.07	8.97E-50	pDC	knownLnc
LNC-GALNT2-2	4.72E-48	2.652	0.317	0.047	3.89E-43	pDC	novelLnc
LNC-SPN-2	4.14E-47	2.970	0.362	0.068	3.41E-42	pDC	novelLnc
AC104365.2	6.89E-39	2.444	0.267	0.032	5.68E-34	pDC	knownLnc
LNC-RPA2-1	1.65E-36	2.609	0.296	0.049	1.37E-31	pDC	novelLnc

MIR155HG	1.81E-28	2.596	0.292	0.151	1.5E-23	pDC	knownLnc
MIR155-AS-1	1.11E-26	2.497	0.262	0.095	9.17E-22	pDC	novelLnc
LUCAT1	1.6E-153	2.580	0.832	0.225	1.34E-148	Neutrophil	knownLnc
CSF3R-AS-1	5.1E-123	2.827	0.513	0.067	4.24E-118	Neutrophil	novelLnc
LNC-TNFRSF10C-1	5.6E-112	3.137	0.41	0.041	4.60E-107	Neutrophil	novelLnc
TREML4-AS-1	3.08E-73	2.448	0.277	0.023	2.54E-68	Neutrophil	novelLnc
LNC-LITAF-1	4.69E-63	2.530	0.69	0.265	3.87E-58	Neutrophil	novelLnc
SMIM25	2.12E-62	2.968	0.478	0.138	1.75E-57	Neutrophil	knownLnc
LNC-FAM157C-1	2.56E-49	3.176	0.265	0.081	2.11E-44	Neutrophil	novelLnc
AP006259.1	1.39E-44	2.844	0.31	0.148	1.15E-39	Neutrophil	knownLnc
MIR3945HG	7.98E-25	3.624	0.263	0.087	6.58E-20	Neutrophil	knownLnc
LNC-NABP1-4	2.76E-22	3.169	0.327	0.193	2.28E-17	Neutrophil	novelLnc
AC023157.2	1.19E-21	2.910	0.336	0.149	9.8E-17	Neutrophil	knownLnc
LNC-AL078604.4-3	3.47E-19	2.707	0.428	0.225	2.86E-14	Neutrophil	novelLnc
LINC01871	0	3.225	0.275	0.012	0	T cell	knownLnc
AC243960.1	0	2.812	0.323	0.058	0	T cell	knownLnc
TRBC2-AS-1	0	2.711	0.562	0.035	0	T cell	novelLnc
LNC-PPP1R16B-2	0	2.611	0.386	0.066	0	T cell	novelLnc
MIAT	0	2.473	0.277	0.065	0	T cell	knownLnc
LNC-BCL11B-3	0	2.435	0.286	0.023	0	T cell	novelLnc
AC026979.2	3.6E-267	2.699	0.287	0.104	2.99E-262	T cell	knownLnc
LNC-RALGAPA1-3	5.9E-134	3.969	0.28	0.18	4.87E-129	T cell	novelLnc
AC004797.1-AS-1	4.2E-114	3.354	0.196	0.27	3.45E-109	T cell	novelLnc
AC016831.6-AS-3	5.2E-113	3.309	0.3	0.188	4.27E-108	T cell	novelLnc
BORCS5-AS-1	4E-108	2.576	0.292	0.224	3.3E-103	T cell	novelLnc
AL606489.1-AS-1	1.3E-104	3.650	0.137	0.28	1E-99	T cell	novelLnc
LNC-AC027644.4-2	1.22E-94	2.434	0.145	0.272	1E-89	T cell	novelLnc
KAT6A-AS-1	8.51E-75	2.540	0.267	0.21	7.02E-70	T cell	novelLnc
SLC8A1-AS-1	5.67E-21	3.633	0.223	0.29	4.67E-16	T cell	novelLnc

Table 3. 10X tumor/nontumor-specific markers (n=32)

Gene	p-value	Avg logFC	Pct.1	Pct.2	Adjust p-value	Cell type - Site	Type
SNHG29	6.84E-25	0.348	0.804	0.677	5.64E-20	Epithelial Tumor	knownLnc
SFTA1P	2.69E-36	-0.556	0.59	0.42	2.22E-31	Epithelial Nontumor	knownLnc
MALAT1	1.76E-20	-0.497	0.918	0.901	1.45E-15	Epithelial Nontumor	knownLnc
SNHG6	1.24E-70	0.498	0.553	0.417	1.03E-65	Malignant Tumor	knownLnc
GAS5	1.43E-52	0.444	0.672	0.578	1.18E-47	Malignant Tumor	knownLnc
SNHG5	3.31E-20	0.276	0.59	0.547	2.73E-15	Malignant Tumor	knownLnc
SFTA1P	0	-1.606	0.59	0.035	0	Malignant Nontumor	knownLnc
AP003498.3	5.1E-291	-0.555	0.262	0.014	4.2E-286	Malignant Nontumor	knownLnc
MALAT1	3.6E-82	-0.370	0.918	0.892	2.97E-77	Malignant Nontumor	knownLnc
CRNDE	3.15E-79	-0.396	0.311	0.137	2.6E-74	Malignant Nontumor	knownLnc
LNC-Xq21.33-1	8.05E-72	-0.390	0.258	0.099	6.64E-67	Malignant Nontumor	novelLnc
LINC01578	3.9E-68	-0.351	0.295	0.141	3.22E-63	Malignant Nontumor	knownLnc
AC020916.1	3.95E-64	-0.433	0.308	0.158	3.26E-59	Malignant Nontumor	knownLnc
LNC-MAML2-3	7.56E-57	-0.393	0.301	0.148	6.24E-52	Malignant Nontumor	novelLnc
AC016831.6-AS-2	1.76E-13	-0.620	0.308	0.134	1.45E-08	B cell Nontumor	novelLnc
ZNF331-AS-1	1.01E-08	-0.517	0.316	0.16	0.000834	B cell Nontumor	novelLnc
SNHG5	7.89E-08	-0.317	0.735	0.613	0.006508	B cell Nontumor	knownLnc
SNHG29	7.11E-35	0.610	0.78	0.49	5.87E-30	Fibroblast Tumor	knownLnc
SNHG5	1.94E-28	0.386	0.693	0.397	1.6E-23	Fibroblast Tumor	knownLnc
LNC-MAML2-3	2.37E-22	-0.617	0.477	0.41	1.95E-17	Fibroblast Nontumor	novelLnc
MALAT1	7.01E-19	-0.560	0.962	0.957	5.79E-14	Fibroblast Nontumor	knownLnc
FOXO1-AS-1	2.06E-18	-0.725	0.364	0.248	1.7E-13	Fibroblast Nontumor	novelLnc
AC020916.1	1.52E-09	-0.535	0.253	0.198	0.000126	Fibroblast Nontumor	knownLnc
FTX	5.24E-09	-0.271	0.272	0.303	0.000433	Fibroblast Nontumor	knownLnc
SNHG29	2.44E-76	0.762	0.613	0.264	2.01E-71	EC Tumor	knownLnc
SNHG5	2.37E-68	0.678	0.522	0.201	1.96E-63	EC Tumor	knownLnc
GAS5	1.04E-55	0.516	0.489	0.205	8.59E-51	EC Tumor	knownLnc
NEAT1	3.22E-51	0.609	0.88	0.727	2.66E-46	EC Tumor	knownLnc
LINC00273-AS-1	3.91E-29	0.307	0.427	0.237	3.22E-24	EC Tumor	novelLnc
PCAT19	4.27E-45	-0.313	0.56	0.592	3.52E-40	EC Nontumor	knownLnc
FKBP5-AS-2	1.03E-38	-0.806	0.269	0.121	8.47E-34	EC Nontumor	novelLnc
XIST	5.68E-29	-0.685	0.261	0.092	4.68E-24	EC Nontumor	knownLnc
LNC-CR392039.4-1	1.8E-160	0.436	0.37	0.22	1.5E-155	Macrophage Tumor	novelLnc
CYTOR	2.15E-91	0.345	0.314	0.234	1.77E-86	Macrophage Tumor	knownLnc

LINC00273-AS-1	2.12E-57	0.284	0.27	0.181	1.75E-52	Macrophage Tumor	novelLnc
MIR3945HG	1.5E-235	-0.445	0.346	0.138	1.2E-230	Macrophage Nontumor	knownLnc
AC026369.3	1.5E-210	-0.344	0.288	0.11	1.3E-205	Macrophage Nontumor	knownLnc
AC020656.1	3.58E-43	-0.400	0.499	0.448	2.95E-38	Macrophage Nontumor	knownLnc
NEAT1	1.48E-57	-0.715	0.94	0.897	1.22E-52	DC Nontumor	knownLnc
LINC-PINT	3.31E-19	-0.377	0.431	0.214	2.73E-14	DC Nontumor	knownLnc
AC020916.1	2.4E-17	-0.424	0.701	0.552	1.98E-12	DC Nontumor	knownLnc
AC016831.6-AS-2	1.66E-15	-0.268	0.32	0.134	1.37E-10	DC Nontumor	novelLnc
LNC-SIPA1L1-1	7.08E-15	-0.313	0.376	0.213	5.84E-10	DC Nontumor	novelLnc
AC016831.4	2.29E-14	-0.285	0.287	0.122	1.89E-09	DC Nontumor	knownLnc
LNC-SIPA1L1-2	1.09E-12	-0.276	0.255	0.111	9.02E-08	DC Nontumor	novelLnc
MALAT1	4.47E-12	-0.283	0.978	0.973	3.68E-07	DC Nontumor	knownLnc
FKBP5-AS-2	1.11E-10	-0.292	0.477	0.341	9.17E-06	DC Nontumor	novelLnc
AC004817.3	2.78E-07	-0.264	0.301	0.191	0.022969	DC Nontumor	knownLnc
DUSP16-AS-1	1.52E-68	0.331	0.336	0.244	1.26E-63	T cell Tumor	novelLnc
AC016831.6-AS-2	7.4E-296	-0.600	0.375	0.191	6.1E-291	T cell Nontumor	novelLnc
LINC-PINT	3.1E-167	-0.423	0.45	0.304	2.6E-162	T cell Nontumor	knownLnc
AC016831.4	1.1E-116	-0.388	0.261	0.156	9.4E-112	T cell Nontumor	knownLnc

Table 4. SMART-seq2 tumor/nontumor-specific markers top 100 (n=241)

Gene	p-value	Avg logFC	Pct.1	Pct.2	Adjust p-value	Cell type - Site	Type
SYT1-AS-3	2.29E-18	-1.290	0.404	0.175	1.89E-13	Epithelial Nontumor	novelLnc
LNC-AC079949.3-1	6.03E-08	-1.400	0.988	0.918	4.97E-03	Epithelial Nontumor	novelLnc
HLA-DRB1-AS-1	1.61E-40	1.400	0.881	0.568	1.33E-35	Malignant Primary	novelLnc
ANKRD44-AS1	2.5E-35	2.400	0.315	0.004	2.06E-30	Malignant Primary	knownLnc
WDR86-AS1	7.99E-34	3.330	0.333	0.016	6.59E-29	Malignant Primary	knownLnc
SRGAP3-AS2	9.26E-34	4.480	0.29	0.008	7.64E-29	Malignant Primary	knownLnc
LINC01765	7.98E-32	2.970	0.27	0	6.58E-27	Malignant Primary	knownLnc
AC025154.2	1.02E-31	1.550	0.317	0.012	8.41E-27	Malignant Primary	knownLnc
AC004832.1	4.15E-31	3.080	0.309	0.02	3.42E-26	Malignant Primary	knownLnc
AL121899.1	4.91E-31	2.860	0.288	0.004	4.05E-26	Malignant Primary	knownLnc
LINC02345	7.44E-31	3.510	0.287	0.004	6.14E-26	Malignant Primary	knownLnc
ACBD3-AS1	1.09E-30	3.890	0.308	0.012	9.03E-26	Malignant Primary	knownLnc
LINC01571	1.18E-30	1.160	0.261	0	9.77E-26	Malignant Primary	knownLnc
NRAV	1.67E-30	1.900	0.371	0.04	1.38E-25	Malignant Primary	knownLnc
AC006230.1	4.45E-30	1.330	0.281	0.004	3.67E-25	Malignant Primary	knownLnc
LINC02166	4.97E-30	2.590	0.292	0.008	4.10E-25	Malignant Primary	knownLnc
AC096637.3	1.02E-29	2.890	0.28	0.012	8.43E-25	Malignant Primary	knownLnc
PCCA-DT	1.84E-29	1.480	0.329	0.024	1.52E-24	Malignant Primary	knownLnc
AC004130.2	3.49E-29	2.280	0.269	0.004	2.88E-24	Malignant Primary	knownLnc
AL357093.2	7.2E-29	1.990	0.366	0.052	5.94E-24	Malignant Primary	knownLnc
AL121956.6	1.5E-28	2.450	0.262	0.004	1.24E-23	Malignant Primary	knownLnc
AC108134.4	6.67E-28	2.600	0.308	0.02	5.50E-23	Malignant Primary	knownLnc
MUC12-AS1	1.86E-27	2.520	0.258	0.004	1.53E-22	Malignant Primary	knownLnc
AC105052.5	3.04E-26	1.920	0.285	0.016	2.51E-21	Malignant Primary	knownLnc
AC010538.1	3.61E-26	1.720	0.325	0.044	2.98E-21	Malignant Primary	knownLnc
AC008875.3	6.86E-26	1.110	0.266	0.012	5.66E-21	Malignant Primary	knownLnc
LNC-AP002761.1-1	1.56E-24	3.430	0.271	0.016	1.29E-19	Malignant Primary	novelLnc
ELN-AS1	6.19E-24	3.360	0.317	0.052	5.11E-19	Malignant Primary	knownLnc
GIHCG	2.61E-23	1.530	0.395	0.096	2.15E-18	Malignant Primary	knownLnc
LINC01513	2.84E-23	2.420	0.262	0.016	2.35E-18	Malignant Primary	knownLnc
UCKL1-AS1	2.43E-21	1.690	0.344	0.068	2.00E-16	Malignant Primary	knownLnc
LINC00342	1.24E-18	1.250	0.388	0.124	1.02E-13	Malignant Primary	knownLnc
WAKMAR2	5.16E-18	1.220	0.314	0.068	4.26E-13	Malignant Primary	knownLnc
HLA-C-AS-1	1.15E-17	1.340	0.905	0.848	9.51E-13	Malignant Primary	novelLnc

AC130456.2	5.39E-17	1.020	0.328	0.08	4.45E-12	Malignant Primary	knownLnc
AC023157.2	9.5E-17	1.690	0.411	0.14	7.84E-12	Malignant Primary	knownLnc
VIM-AS1	1.07E-16	1.380	0.256	0.04	8.82E-12	Malignant Primary	knownLnc
LNC-EIF3C-1	1.13E-16	1.360	0.261	0.044	9.33E-12	Malignant Primary	novelLnc
AC068587.4-AS-2	2.5E-13	1.190	0.396	0.2	2.07E-08	Malignant Primary	novelLnc
LNC-SPDYE16-1	3.32E-13	1.720	0.251	0.056	2.74E-08	Malignant Primary	novelLnc
AL162734.1	5.16E-13	1.420	0.272	0.072	4.25E-08	Malignant Primary	knownLnc
LNC-IFFO2-1	5.7E-10	1.290	0.271	0.092	4.70E-05	Malignant Primary	novelLnc
AC044849.1	5E-08	1.110	0.27	0.108	4.12E-03	Malignant Primary	knownLnc
LNC-Xq21.33-6	1.58E-78	-3.350	0.456	0.015	1.30E-73	Malignant Nontumor	novelLnc
SFTPC-AS-1	3.09E-56	-2.130	0.328	0.008	2.55E-51	Malignant Nontumor	novelLnc
LNC-Xq21.33-1	1.06E-46	-1.460	0.516	0.12	8.76E-42	Malignant Nontumor	novelLnc
SFTA1P	4.73E-24	-2.040	0.548	0.33	3.90E-19	Malignant Nontumor	knownLnc
LIMCH1-AS-1	9.34E-24	-1.790	0.428	0.223	7.70E-19	Malignant Nontumor	novelLnc
LNC-CCSER1-3	8.68E-23	-0.979	0.984	0.947	7.16E-18	Malignant Nontumor	novelLnc
SYT1-AS-3	1.7E-20	-2.020	0.404	0.232	1.40E-15	Malignant Nontumor	novelLnc
LHFPL3-AS2	4.2E-20	-1.780	0.372	0.129	3.47E-15	Malignant Nontumor	knownLnc
AC090236.2	2.76E-16	-2.380	0.388	0.213	2.27E-11	Malignant Nontumor	knownLnc
RPARP-AS1	8.11E-16	-1.180	0.4	0.463	6.69E-11	Malignant Nontumor	knownLnc
HEIH	1.08E-14	-1.140	0.204	0.291	8.94E-10	Malignant Nontumor	knownLnc
DLG2-AS-7	2.09E-13	-1.090	0.604	0.399	1.72E-08	Malignant Nontumor	novelLnc
LRRK2-DT	2.72E-12	-1.140	0.3	0.112	2.25E-07	Malignant Nontumor	knownLnc
LNC-ZFAND3-7	1.95E-11	-1.480	0.336	0.414	1.61E-06	Malignant Nontumor	novelLnc
LNC-FAM214A-1	4.57E-11	-1.480	0.268	0.5	3.77E-06	Malignant Nontumor	novelLnc
MPRIP-AS1	1.52E-10	-1.540	0.308	0.265	1.26E-05	Malignant Nontumor	knownLnc
LIMCH1-AS-2	5.62E-10	-1.170	0.332	0.176	4.63E-05	Malignant Nontumor	novelLnc
LNC-IMMP2L-5	6.81E-10	-1.400	0.172	0.296	5.62E-05	Malignant Nontumor	novelLnc
AL035071.1	2.82E-09	-1.430	0.348	0.326	2.32E-04	Malignant Nontumor	knownLnc
LNC-CYP4A11-1	7.18E-09	-1.300	0.376	0.234	5.93E-04	Malignant Nontumor	novelLnc
AC108010.1-AS-1	1.03E-08	-1.110	0.284	0.41	8.49E-04	Malignant Nontumor	novelLnc
LNC-AC092139.1-1	3.27E-08	-1.130	0.188	0.333	2.70E-03	Malignant Nontumor	novelLnc
AC005332.6	3.47E-08	-1.240	0.22	0.357	2.86E-03	Malignant Nontumor	knownLnc
LNC-LHFPL3-1	5.72E-08	-1.980	0.312	0.326	4.72E-03	Malignant Nontumor	novelLnc
LNC-SMURF1-1	1.32E-07	-0.989	0.24	0.348	1.09E-02	Malignant Nontumor	novelLnc
DLEU2	2.04E-07	-1.530	0.236	0.301	1.68E-02	Malignant Nontumor	knownLnc
AL133268.4	2.34E-07	-2.030	0.208	0.256	1.93E-02	Malignant Nontumor	knownLnc
AC008268.1	4.9E-07	-1.270	0.268	0.196	4.04E-02	Malignant Nontumor	knownLnc

LNC-ARID1B-6	8.08E-08	-1.620	0.75	0.557	6.67E-03	Mast cell Nontumor	novelLnc
LNC-AL592164.1-1	4.93E-14	1.350	0.477	0.27	4.07E-09	Fibroblast Primary	novelLnc
LNC-AC079949.3-1	7.04E-11	1.020	0.959	0.956	5.81E-06	Fibroblast Primary	novelLnc
LINC01220	1.29E-09	1.010	0.296	0.128	1.06E-04	Fibroblast Primary	knownLnc
AC116366.2	2.95E-14	-1.160	0.485	0.266	2.43E-09	Fibroblast Nontumor	knownLnc
LINC00910-AS-1	1.28E-11	-1.970	0.42	0.244	1.06E-06	Fibroblast Nontumor	novelLnc
LNC-LHFPL3-1	7.64E-10	-1.100	0.343	0.276	6.30E-05	Fibroblast Nontumor	novelLnc
LNC-ARID1B-6	2.88E-08	-1.030	0.252	0.273	2.38E-03	Fibroblast Nontumor	novelLnc
AL035071.1	5.82E-08	-0.975	0.354	0.206	4.80E-03	Fibroblast Nontumor	knownLnc
SMCO4-AS-1	1.83E-10	1.090	0.428	0.196	1.51E-05	EC Primary	novelLnc
LNC-IGKV1OR-2-2	3.97E-07	1.410	0.413	0.229	3.28E-02	EC Primary	novelLnc
LNC-AC010522.1-1	4.97E-07	0.985	0.4	0.2	4.10E-02	EC Primary	novelLnc
FENDRR	7.81E-21	-1.350	0.654	0.275	6.44E-16	EC Nontumor	knownLnc
VIPR1-AS1	3.06E-10	-1.110	0.267	0.074	2.53E-05	EC Nontumor	knownLnc
PCAT19	5.52E-10	-1.050	0.85	0.801	4.56E-05	EC Nontumor	knownLnc
LNC-AL592164.1-1	2.37E-09	1.060	0.569	0.485	1.96E-04	Macrophage Primary	novelLnc
LNC-AC079949.3-1	2.04E-07	1.220	0.862	0.905	1.68E-02	Macrophage Primary	novelLnc
LNC-RNF144B-1	2.1E-13	-1.520	0.358	0.158	1.73E-08	Macrophage Nontumor	novelLnc
AC079753.2	1.12E-10	-1.180	0.274	0.109	9.21E-06	Macrophage Nontumor	knownLnc
LNC-HLA-DQB1-1	9.13E-10	-1.020	0.38	0.221	7.54E-05	Macrophage Nontumor	novelLnc
LINC00273-AS-1	2.21E-07	1.250	1	1	1.83E-02	pDC Primary	novelLnc
LNC-AL445250.1-1	9.5E-09	2.380	0.319	0.306	7.84E-04	T cell Primary	novelLnc
LNC-10q11.21-6	5.06E-08	2.420	0.306	0.229	4.18E-03	T cell Primary	novelLnc
LNC-10q11.21-5	5.95E-08	3.020	0.373	0.384	4.91E-03	T cell Primary	novelLnc
AP000317.2-AS-1	4.66E-07	1.920	0.287	0.148	3.84E-02	T cell Primary	novelLnc
LNC-PDE3A-2	5.97E-23	-1.150	0.98	0.927	4.93E-18	T cell Nontumor	novelLnc
LNC-SLA2-1	1.97E-17	-1.510	0.316	0.113	1.62E-12	T cell Nontumor	novelLnc
LNC-ZEB2-1	1.3E-09	-0.981	0.323	0.172	1.08E-04	T cell Nontumor	novelLnc
LNC-PRKCH-2	3.22E-07	-1.110	0.296	0.195	2.66E-02	T cell Nontumor	novelLnc

Table 5. Cancer scRNA-seq datasets

Platform	Cancer*	Database	Samples	Library type	Reference
10X 3'-seq (v2 Chemistry)	AML	dbGaP	5	5'-seq	(Petti et al. 2019)
	GBM	SRA	11	3'-seq	(Bhaduri et al. 2020)
	HCC	SRA	12	3'-seq	(Losic et al. 2020)
	LC	ERA	31	3'-seq	(Lambrechts et al. 2018)
	CRC	EGA	33	3'-seq	(Lee et al. 2020)
		EBI	27	3'-seq	
	BRCA	EBI	14	5'-seq	(Qian et al. 2020)
	CRC		21	3'-seq	
	LC		36	3'-seq	
	OVC		10	3'-seq	
	HCC	dbGaP	19	3'-seq	(Ma et al. 2019)
	B-ALL	SRA	33	3'-seq	(Witkowski et al. 2020)
	Uveal melanoma	dbGaP	11	3'-seq	(Durante et al. 2020)
	LC	ERA	58	3'-seq	(Kim et al. 2020)
	EGC	SRA	13	3'-seq	(Zhang et al. 2019)
SMART-seq2	HNSCC	Not available	18	full-length	(Puram et al. 2017)
	LC	SRA	49	full-length	(Puram et al. 2017)
	Melanoma	DUOS	19	full-length	(Tirosh et al. 2016)
	Melanoma	DUOS/dbGaP	31	full-length	(Tirosh et al. 2016)
	Melanoma	dbGaP	32	full-length	(Sade-Feldman et al. 2018)
	Oligodendroglioma	dbGaP	6	full-length	(Tirosh et al. 2016)
	BRCA	SRA	6	full-length	(Tirosh et al. 2016)
Drop-seq	OVC	SRA	6	3'-seq	(Olalekan et al. 2021)
C1	BRCA	SRA	33	full-length	(Chung et al. 2017)
Seq-Well	AML	SRA	83	3'-seq	(van Galen et al. 2019)

*: AML (acute myeloid leukemia); GBM (glioblastoma); HCC (hepatocellular carcinoma); LC (lung cancer); CRC (colorectal cancer); BRCA (breast cancer); OVC (ovarian cancer); B-ALL (acute B lymphoblastic leukemia); EGC (early gastric cancer); HNSCC (head and neck squamous cell carcinoma)

References

Agarwal, S., et al. (2020). The long non-coding RNA LUCAT1 is a negative feedback regulator of interferon responses in humans. Nat Commun 11(1): 6348.

Bhaduri, A., et al. (2020). Outer Radial Glia-like Cancer Stem Cells Contribute to Heterogeneity of Glioblastoma. Cell Stem Cell 26(1): 48-63 e46.

Binnewies, M., et al. (2018). Understanding the tumor immune microenvironment (TIME) for effective therapy. Nat Med 24(5): 541-550.

Chung, W., et al. (2017). Single-cell RNA-seq enables comprehensive tumour and immune cell profiling in primary breast cancer. Nat Commun 8: 15081.

Du, Z., et al. (2013). Integrative genomic analyses reveal clinically relevant long noncoding RNAs in human cancer. Nat Struct Mol Biol 20(7): 908-913.

Durante, M. A., et al. (2020). Single-cell analysis reveals new evolutionary complexity in uveal melanoma. Nat Commun 11(1): 496.

Galli, F., et al. (2020). Relevance of immune cell and tumor microenvironment imaging in the new era of immunotherapy. J Exp Clin Cancer Res 39(1): 89.

Hirata, H., et al. (2015). Long Noncoding RNA MALAT1 Promotes Aggressive Renal Cell Carcinoma through Ezh2 and Interacts with miR-205. Cancer Res 75(7): 1322-1331.

Huang, D., et al. (2018). NKILA lncRNA promotes tumor immune evasion by sensitizing T cells to activation-induced cell death. Nat Immunol 19(10): 1112-1125.

Huarte, M. (2015). The emerging role of lncRNAs in cancer. Nat Med 21(11): 1253-1261.

Ji, P., et al. (2003). MALAT-1, a novel noncoding RNA, and thymosin beta4 predict metastasis and survival in early-stage non-small cell lung cancer. Oncogene 22(39): 8031-8041.

Kim, N., et al. (2020). Single-cell RNA sequencing demonstrates the molecular and cellular reprogramming of metastatic lung adenocarcinoma. Nat Commun 11(1): 2285.

Lambrechts, D., et al. (2018). Phenotype molding of stromal cells in the lung

tumor microenvironment. Nat Med 24(8): 1277-1289.

Lee, H. O., et al. (2020). Lineage-dependent gene expression programs influence the immune landscape of colorectal cancer. Nat Genet 52(6): 594-603.

Li, L., et al. (2011). The role of FKBP5 in cancer aetiology and chemoresistance. Br J Cancer 104(1): 19-23.

Li, S., et al. (2018). Pan-cancer analysis of long non-coding RNA NEAT1 in various cancers. Genes Dis 5(1): 27-35.

Liu, B., et al. (2015). A cytoplasmic NF-kappaB interacting long noncoding RNA blocks IkappaB phosphorylation and suppresses breast cancer metastasis. Cancer Cell 27(3): 370-381.

Liu, J., et al. (2020). Long noncoding RNA LINC01578 drives colon cancer metastasis through a positive feedback loop with the NF-kappaB/YY1 axis. Mol Oncol 14(12): 3211-3233.

Liu, S. J., et al. (2016). Single-cell analysis of long non-coding RNAs in the developing human neocortex. Genome Biol 17: 67.

Losic, B., et al. (2020). Intratumoral heterogeneity and clonal evolution in liver cancer. Nat Commun 11(1): 291.

Luo, C. T., et al. (2016). Graded Foxo1 activity in Treg cells differentiates tumour immunity from spontaneous autoimmunity. Nature 529(7587): 532-536.

Luo, H., et al. (2021). Single-cell Long Non-coding RNA Landscape of T Cells in Human Cancer Immunity. Genomics Proteomics Bioinformatics.

Ma, H., et al. (2018). The pseudogene-derived long non-coding RNA SFTA1P suppresses cell proliferation, migration, and invasion in gastric cancer. Biosci Rep 38(2).

Ma, L., et al. (2019). Tumor Cell Biodiversity Drives Microenvironmental Reprogramming in Liver Cancer. Cancer Cell 36(4): 418-430 e416.

Maynard, A., et al. (2020). Therapy-Induced Evolution of Human Lung Cancer Revealed by Single-Cell RNA Sequencing. Cell 182(5): 1232-1251 e1222.

Mourtada-Maarabouni, M., et al. (2009). GAS5, a non-protein-coding RNA, controls apoptosis and is downregulated in breast cancer. Oncogene 28(2): 195-208.

Musikacharoen, T., et al. (2011). Functional involvement of dual specificity

phosphatase 16 (DUSP16), a c-Jun N-terminal kinase-specific phosphatase, in the regulation of T helper cell differentiation. J Biol Chem 286(28): 24896-24905.

Olalekan, S., et al. (2021). Characterizing the tumor microenvironment of metastatic ovarian cancer by single-cell transcriptomics. Cell Rep 35(8): 109165.

Petti, A. A., et al. (2019). A general approach for detecting expressed mutations in AML cells using single cell RNA-sequencing. Nat Commun 10(1): 3660.

Picelli, S., et al. (2014). Full-length RNA-seq from single cells using Smart-seq2. Nat Protoc 9(1): 171-181.

Prensner, J. R. and A. M. Chinnaiyan (2011). The emergence of lncRNAs in cancer biology. Cancer Discov 1(5): 391-407.

Puram, S. V., et al. (2017). Single-Cell Transcriptomic Analysis of Primary and Metastatic Tumor Ecosystems in Head and Neck Cancer. Cell 171(7): 1611-1624 e1624.

Qian, J., et al. (2020). A pan-cancer blueprint of the heterogeneous tumor microenvironment revealed by single-cell profiling. Cell Res 30(9): 745-762.

Sade-Feldman, M., et al. (2018). Defining T Cell States Associated with Response to Checkpoint Immunotherapy in Melanoma. Cell 175(4): 998-1013 e1020.

Schmitt, A. M. and H. Y. Chang (2016). Long Noncoding RNAs in Cancer Pathways. Cancer Cell 29(4): 452-463.

Stuart, T., et al. (2019). Comprehensive Integration of Single-Cell Data. Cell 177(7): 1888-1902 e1821.

Sun, C., et al. (2016). Long non-coding RNA NEAT1 promotes non-small cell lung cancer progression through regulation of miR-377-3p-E2F3 pathway. Oncotarget 7(32): 51784-51814.

Sun, M., et al. (2014). Downregulated long noncoding RNA MEG3 is associated with poor prognosis and promotes cell proliferation in gastric cancer. Tumour Biol 35(2): 1065-1073.

Sun, Y., et al. (2017). Long non-coding RNA LUCAT1 is associated with poor

prognosis in human non-small lung cancer and regulates cell proliferation via epigenetically repressing p21 and p57 expression. Oncotarget 8(17): 28297-28311.

Taube, J. M., et al. (2018). Implications of the tumor immune microenvironment for staging and therapeutics. Mod Pathol 31(2): 214-234.

Tirosh, I., et al. (2016). Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. Science 352(6282): 189-196.

Tirosh, I., et al. (2016). Single-cell RNA-seq supports a developmental hierarchy in human oligodendroglioma. Nature 539(7628): 309-313.

van Galen, P., et al. (2019). Single-Cell RNA-Seq Reveals AML Hierarchies Relevant to Disease Progression and Immunity. Cell 176(6): 1265-1281 e1224.

Witkowski, M. T., et al. (2020). Extensive Remodeling of the Immune Microenvironment in B Cell Acute Lymphoblastic Leukemia. Cancer Cell 37(6): 867-882 e812.

Xing, C., et al. (2021). Role of lncRNA LUCAT1 in cancer. Biomed Pharmacother 134: 111158.

Yang, X., et al. (2020). Long non-coding RNA GAS5 in human cancer (Review). Oncol Lett 20(3): 2587-2594.

Yoon, J. H., et al. (2018). The long noncoding RNA LUCAT1 promotes tumorigenesis by controlling ubiquitination and stability of DNA methyltransferase 1 in esophageal squamous cell carcinoma. Cancer Lett 417: 47-57.

You, B.-H., et al. (2017). High-confidence coding and noncoding transcriptome maps. Genome research 27(6): 1050-1062.

You, B. H., et al. (2019). HERES, a lncRNA that regulates canonical and noncanonical Wnt signaling pathways via interaction with EZH2. Proc Natl Acad Sci U S A 116(49): 24620-24629.

Yu, X., et al. (2017). NEAT1: A novel cancer-related long non-coding RNA. Cell Prolif 50(2).

Zhang, P., et al. (2019). Dissecting the Single-Cell Transcriptome Network Underlying Gastric Premalignant Lesions and Early Gastric Cancer. Cell Rep 27(6): 1934-1947 e1935.

Zheng, G. X., et al. (2017). Massively parallel digital transcriptional profiling of single cells. Nat Commun 8: 14049.

Zhou, J., et al. (2019). Combined Single-Cell Profiling of lncRNAs and Functional Screening Reveals that H19 Is Pivotal for Embryonic Hematopoietic Stem Cell Development. Cell Stem Cell 24(2): 285-298 e285.

Zhou, Y., et al. (2007). Activation of p53 by MEG3 non-coding RNA. J Biol Chem 282(34): 24731-24742.

Zhuo, H., et al. (2016). The aberrant expression of MEG3 regulated by UHRF1 predicts the prognosis of hepatocellular carcinoma. Mol Carcinog 55(2): 209-219.

국문 요지

단일 세포 수준에서 확인한 종양 면역 미세 환경 내에서 특이적인 lncRNA signatures

박은경

자연과학대학 생명과학과

한양대학교

긴 비번역 RNA (long non-coding RNA; lncRNA)는 200 뉴클레오타이드 이상의 비번역 유전자이며, 악성 종양의 형성에 핵심 인자로 작동하고 있다는 연구가 보고되고 있다. 여러 암 종류에서 비이상적인 발현을 보이는 lncRNA들은 종양 악성화 (tumorigenesis)나 종양 형성을 억제하는 (tumor suppressive) 특징을 보인다고 밝혀져 왔지만, 이러한 lncRNA의 자세한 작용 메커니즘은 아직까지 많이 연구된 바가 없다. 그 이유 중 하나로는, lncRNA가 여러 세포 타입이나 상태에 따라서 특이적인 발현을 나타내는 것으로 알려져 있지만, 대부분의 lncRNA분석이 여러 가지 세포 타입이 섞여있는 RNA-seq(bulk RNA-sequencing) 데이터를 기반으로 이루어졌다는 것이다. 따라서 암 특이적인 발현은 종양 세포와 종양 면역 미세 환경 (TIME)내에서 공존하고 서로 상호작용 하는 여러 비종양 세포들의 평균 발현량을 확인한 결과이다. 따라서 lncRNA 연구는 TIME 내의 다양한 세포 타입 각각에서의 발현 패턴을 파악하고 이에 따른 정확한 표적 세포에서의 기능 연구를 위해, 벌크 샘플 수준에서뿐만 아니라 single-cell 수준에서의 분석이 필요한 상황이다. 본 연구에서는 scRNA-seq 플랫폼인 10X Chromium (10X) 및 SMART-seq2 (SS2) 에서 생산된 폐암 scRNA-seq데이터를 novel lncRNA annotation 정보를 추가한 정밀한 전사체 지도 (BIGTranscriptome v1.2)를 이용해 분석하여 TIME에서 차등 발현되는 lncRNA를 식별하였다. 세포 타입 별로 특이적

인 발현을 보이는 812개의 lncRNA마커를 확인하였고, 이 lncRNA중 74개가 두 플랫폼에서 모두 나타났다. 잘 알려진 발암성 또는 종양 억제 lncRNA인 LUCAT1 또는 MEG3는 비악성세포에서 특이적으로 발현하고 있음을 확인하였다. 또한, 10X와 SS2 데이터에서 종양 샘플과 비종양 샘플 간에 차등발현하는 유전자를 분석하였고, 각각 12개와 57개의 lncRNA가 확인되었다. 결과적으로, 이 연구에서는 lncRNA 프로파일링을 통해 세포 유형 및 종양 특이성 측면에서 포괄적인 lncRNA signature를 제공할 뿐만 아니라 암 발달의 메커니즘을 조사하는데 도움이 될 수 있는 lncRNA 타겟을 제안할 수 있다.

Acknowledgements

During the time of research, I was able to complete my degree thesis by learning various studies and research skills with the help of many people. I would like to take this opportunity to thank everyone for their excellent teaching and encouragement.

First of all, I would like to express my deepest gratitude to Prof. Nam for his guidance and encouragement in my training. I learned the researcher's mindset through the professor's passion and energy for research. Without the help of Prof. Nam, it would have been impossible to the writing of this thesis.

Furthermore, I would like to thank Prof. Je-Min Choi and Jun-Ho Choe for their helpful comments of broad insight into research. I also thank the other lab members of BIG lab in Hanyang University for their feedback and helpful discussions.

Lastly, I am especially grateful to my family who gave support and encouragement to give me the chance to focus on the thesis.

Declaration of Ethical Conduct in Research

I, as a graduate student of Hanyang University, hereby declare that I have abided by the following Code of Research Ethics while writing this dissertation thesis, during my degree program.

"First, I have strived to be honest in my conduct, to produce valid and reliable research conforming with the guidance of my thesis supervisor, and I affirm that my thesis contains honest, fair and reasonable conclusions based on my own careful research under the guidance of my thesis supervisor.

Second, I have not committed any acts that may discredit or damage the credibility of my research. These include, but are not limited to : falsification, distortion of research findings or plagiarism.

Third, I need to go through with Copykiller Program(Internet-based Plagiarism-prevention service) before submitting a thesis."

DECEMBER 16, 2021

Degree :	Master
Department :	DEPARTMENT OF LIFE SCIENCE
Thesis Supervisor :	Nam, Jin-Wu
Name :	PARK EUNGYEONG (Signature)

연구 윤리 서약서

본인은 한양대학교 대학원생으로서 이 학위논문 작성 과정에서 다음과 같이 연구 윤리의 기본 원칙을 준수하였음을 서약합니다.

첫째, 지도교수의 지도를 받아 정직하고 엄정한 연구를 수행하여 학위논문을 작성한다.

둘째, 논문 작성시 위조, 변조, 표절 등 학문적 진실성을 훼손하는 어떤 연구 부정행위도 하지 않는다.

셋째, 논문 작성시 논문유사도 검증시스템 "카피킬러"등을 거쳐야 한다.

2021년12월16일

학위명 : 석사

학과 : 생명과학과

지도교수 : 남진우

성명 : 박은경 (서명)

한 양 대 학 교 대 학 원 장 귀 하