

Supplementary Table 1. Information on 3'-end scRNA-seq dataset used for gene expression profiling analysis.

Cell line	Raw data from NGS		After processing raw data		After read mapping with processed data	
	No. of cells	No. of reads	No. of cells (%) ^a	No. of reads (%) ^b	No. of cells ^c (%)	No. of reads (%)
A549	400	44,009,436	391 (97.8)	25,708,488 (58.4)	391 (97.8)	20,016,549 (45.5)
H460	400	45,299,075	383 (95.8)	26,463,657 (58.4)	381 (95.3)	20,455,852 (45.2)
H1299	400	42,205,130	328 (82.0)	24,391,885 (57.8)	319 (79.8)	19,780,315 (46.9)
Calu3	400	48,437,279	354 (88.5)	29,925,212 (61.8)	350 (87.5)	22,678,217 (46.8)
Total (%)	1,600	179,950,920	1,456 (91.0)	106,489,242 (59.2)	1,441 (90.1)	82,930,933 (46.1)

^{a, b}Against raw data and ^c15 read sets (or cells) containing <2,000 sequence reads were excluded after processing raw read sets.

Supplementary Table 2. The number of single cells regrouped from four human NSCLC epithelial cell lines into four clusters.

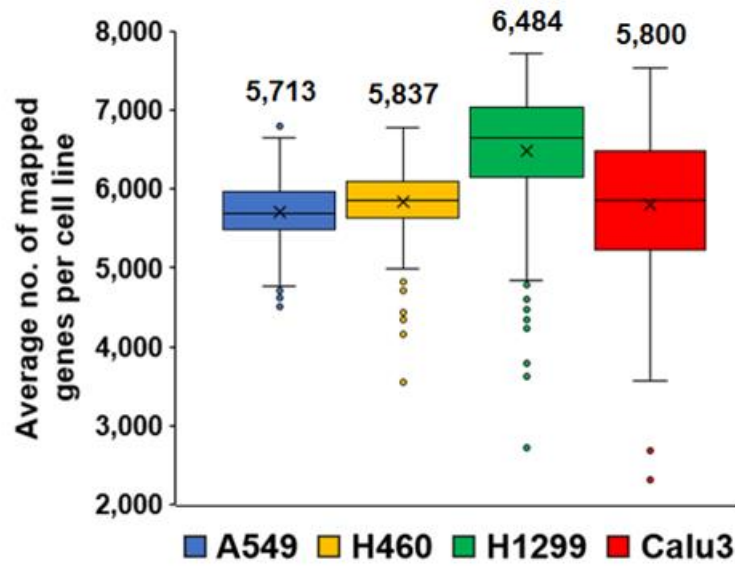
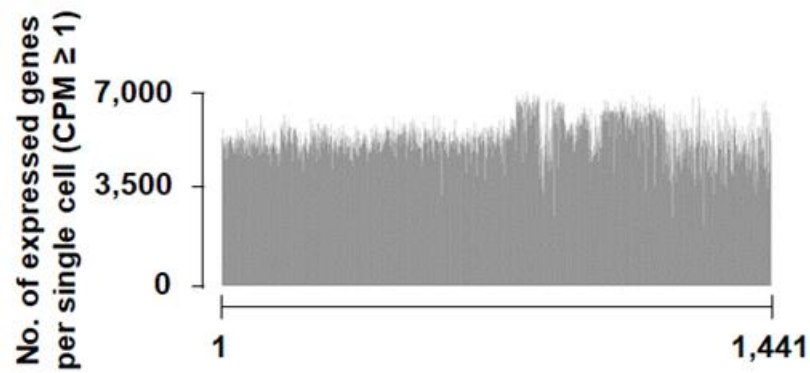
Clusters / Cell lines (%) ^a	A549	H460	Calu3	H1299	Total per cluster (%) ^c
Cluster 1	317 (52.0)	77 (12.6)	46 (7.5)	170 (27.9)	610 (42.3)
Cluster 2	74 (19.6)	304 (80.4)	0	0	378 (26.2)
Cluster 3	0	0	277 (80.3)	68 (19.7)	345 (23.9)
Cluster 4	0	0	27 (25.0)	81 (75.0)	108 (7.5)
Total per cell line (%)^b	391 (27.1)	381 (26.4)	350 (24.3)	319 (22.1)	1,441

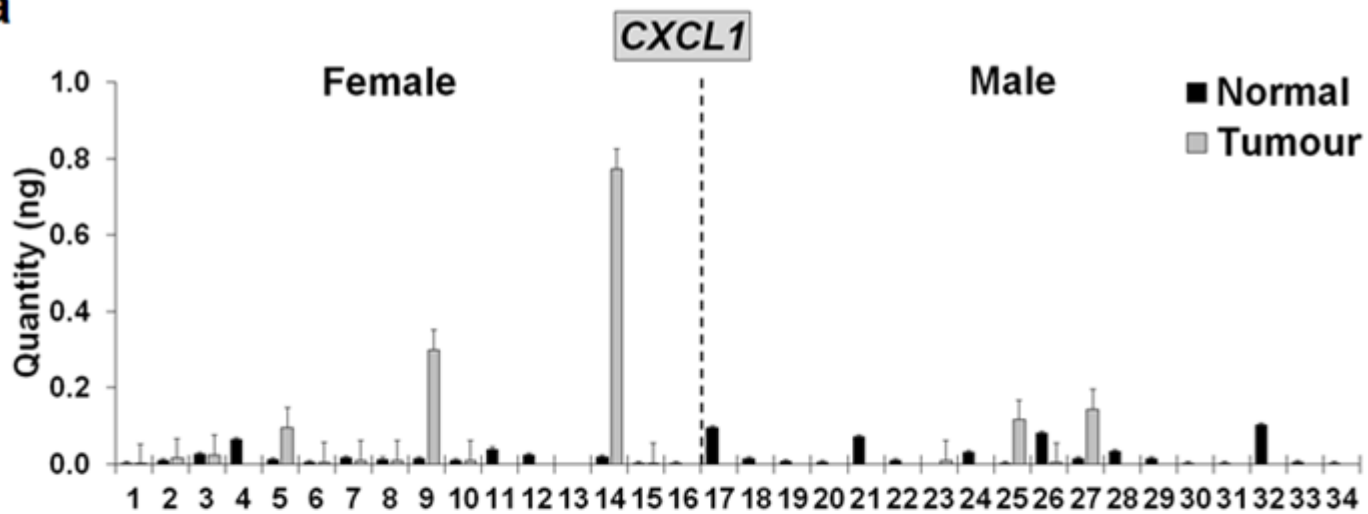
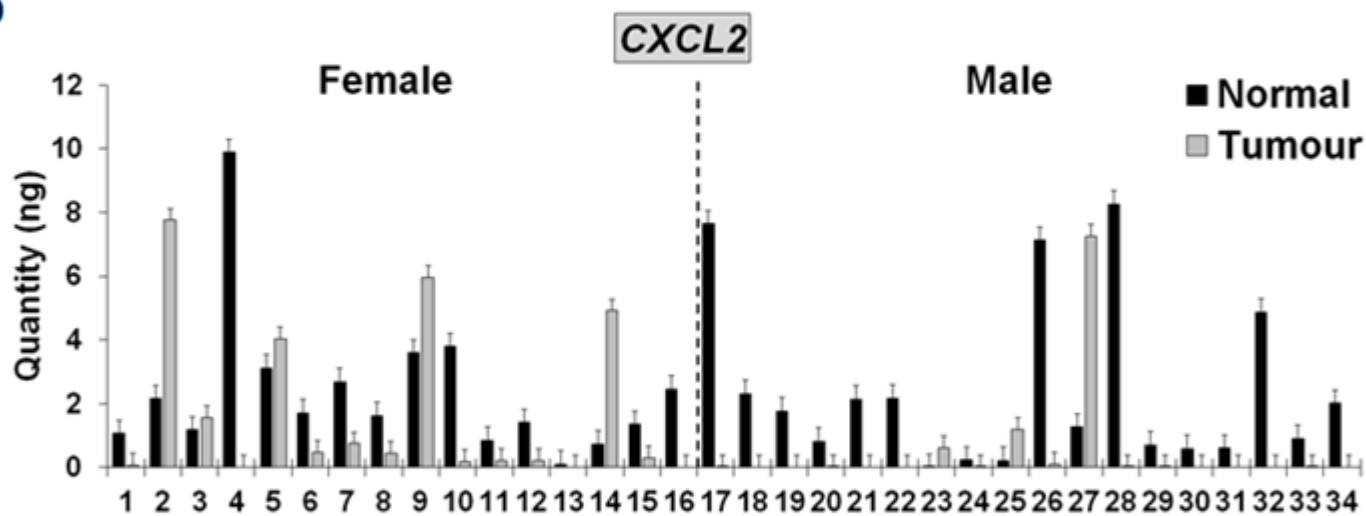
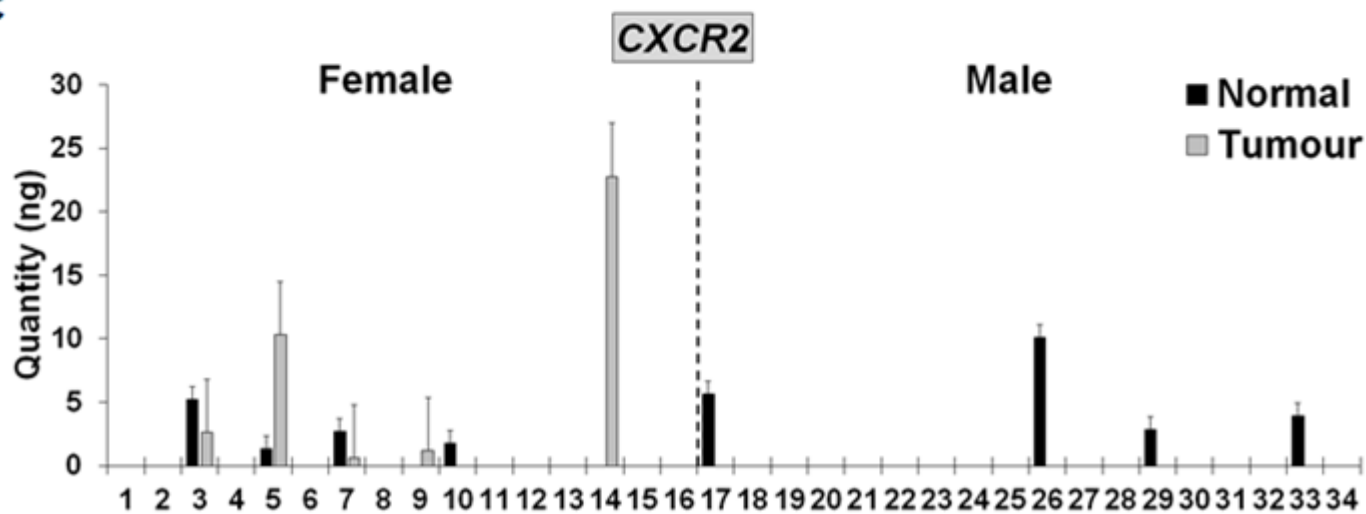
^aAgainst the number of single cells per cluster and ^{b,c}against the total number of 1,441 single cells.

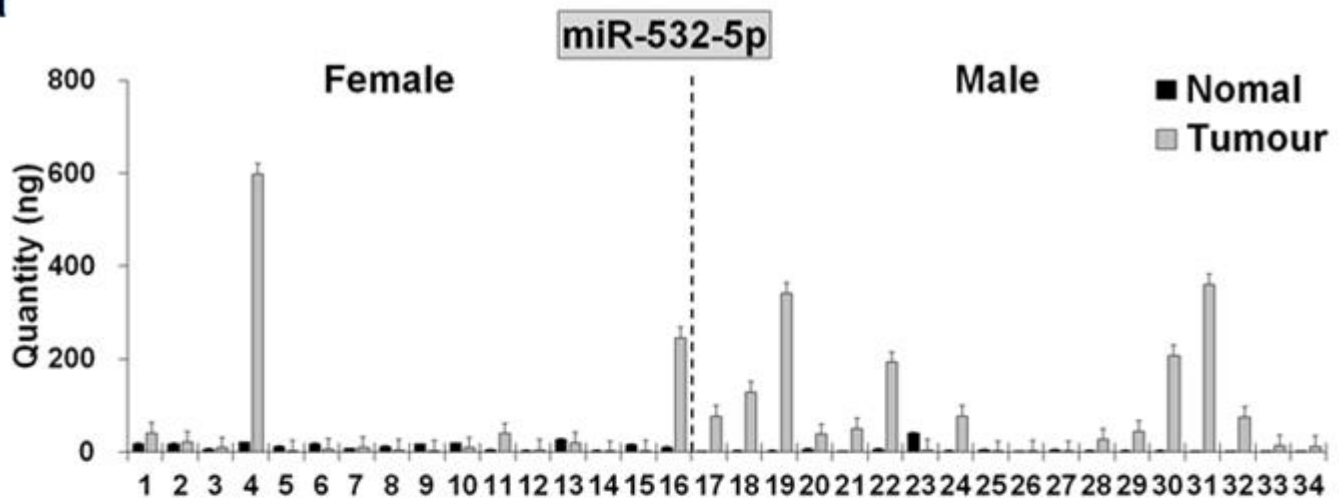
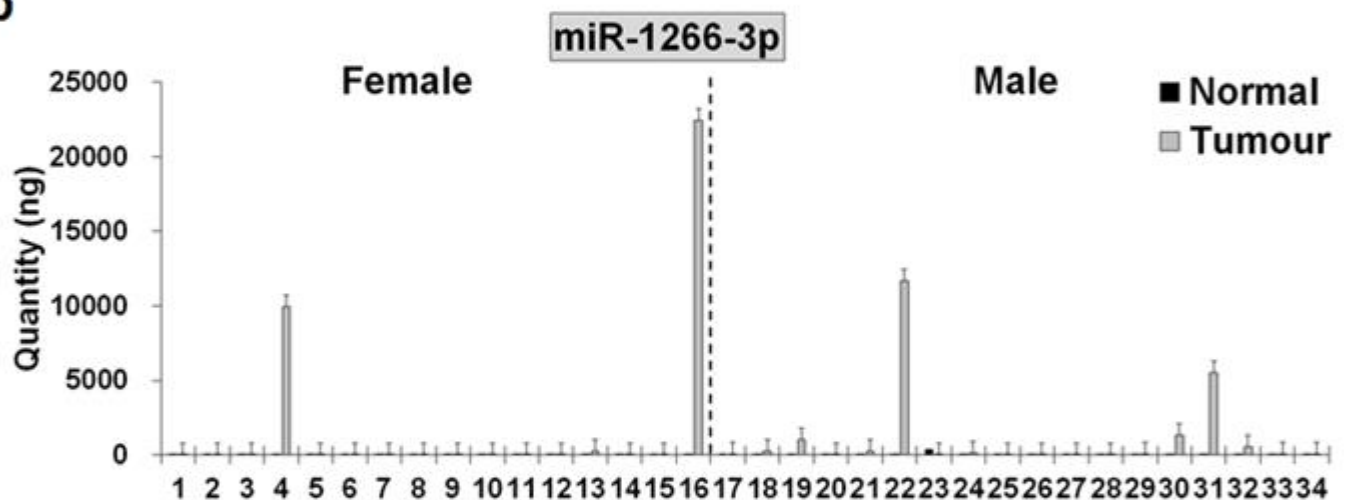
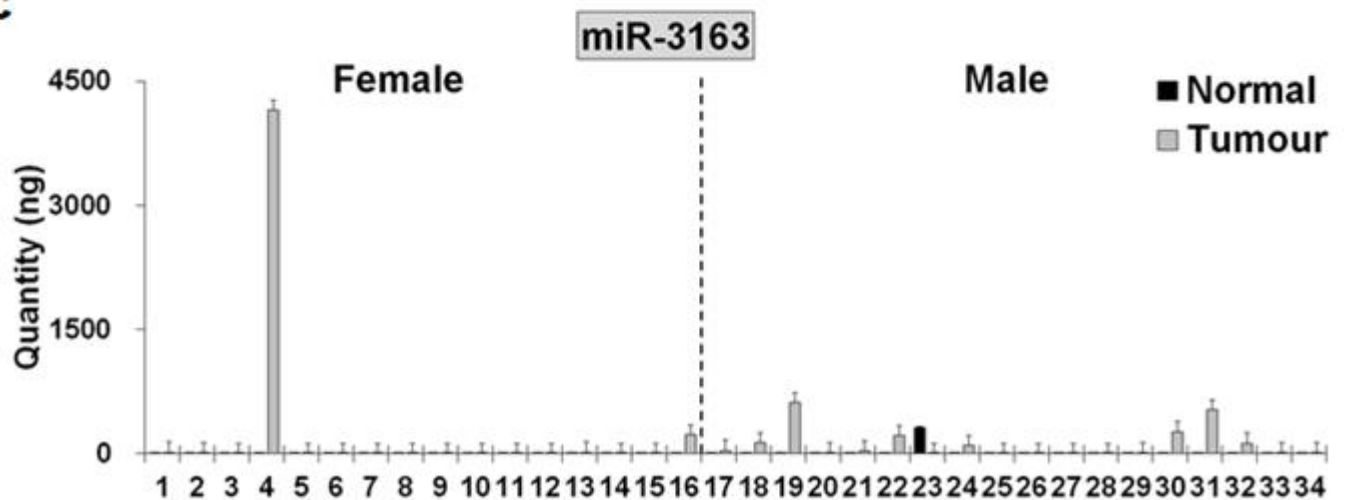
Supplementary Table 3. The number of GO terms, KEGG pathways and Molecular Signatures (oncogenic gene sets) overrepresented from GSEA.

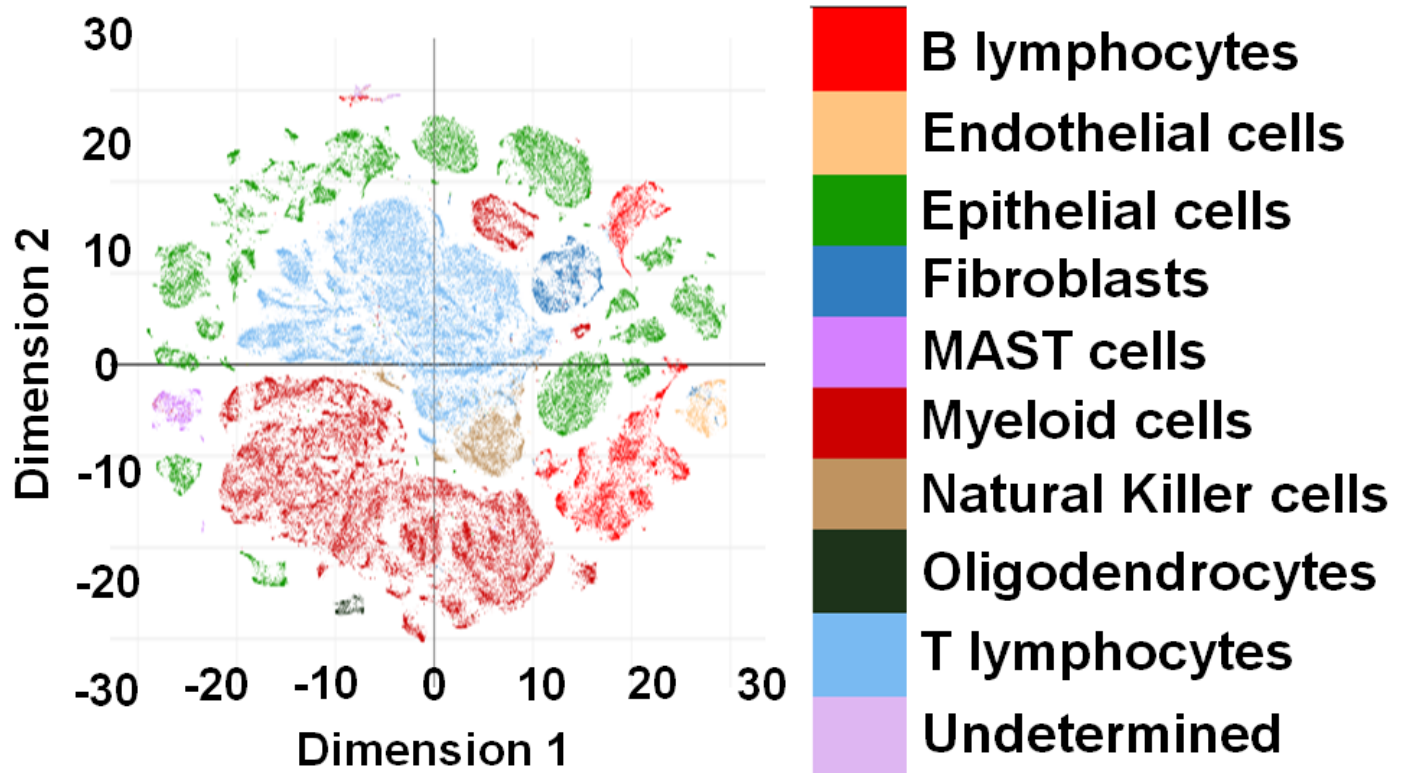
Gene set (Cluster)^a	No. of DEGs	No. of GO terms^b	No. of pathways	No. of signatures
Up (Cluster 2)	355	55	6	18
Down (Cluster 2)	305	40	0	12
Up (Cluster 3)	939	161	26	38
Down (Cluster 3)	1,172	157	51	27
Up (Cluster 4)	199	3	0	3
Down (Cluster 4)	226	68	6	15
Total^c	2,655	441	79	72

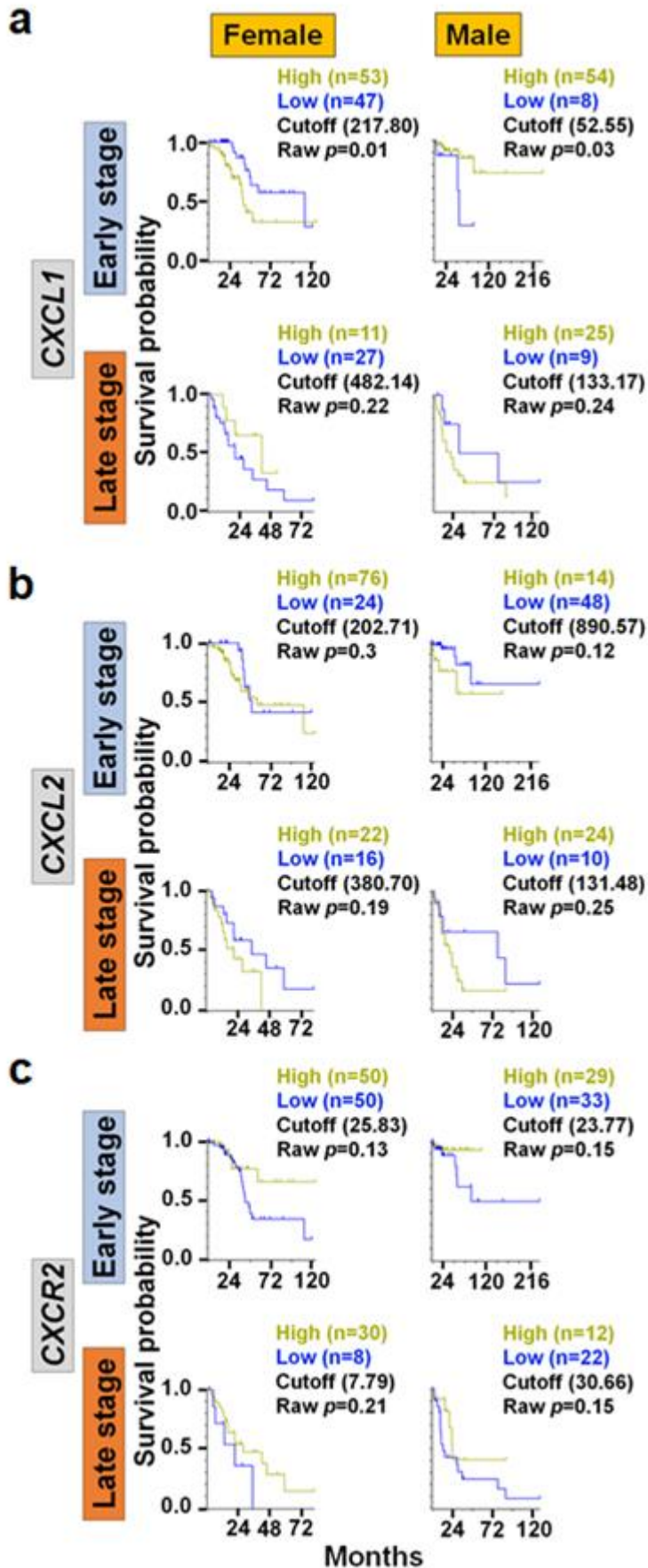
^aAgainst Cluster 1, ^bincluding overlapping GO terms and ^cunique DEGs or GO terms.

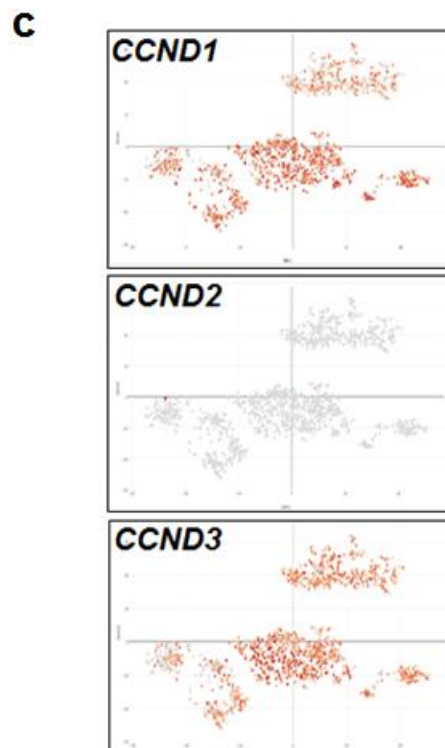
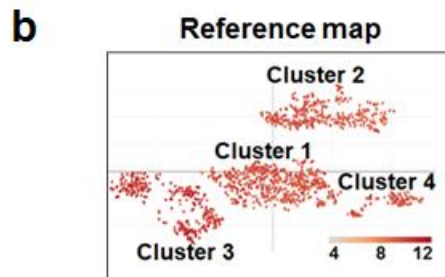
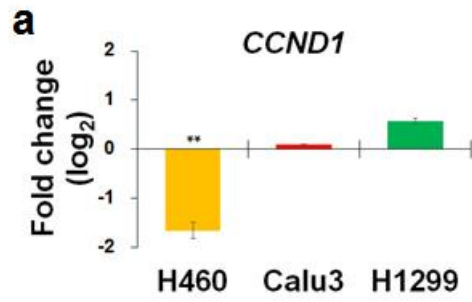
a**b**

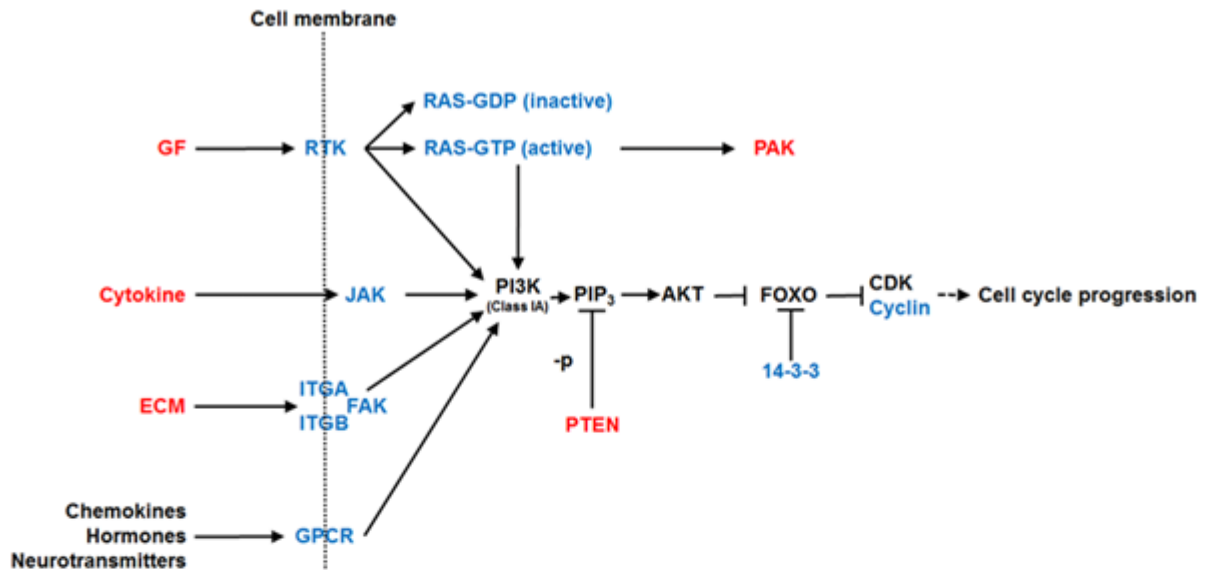
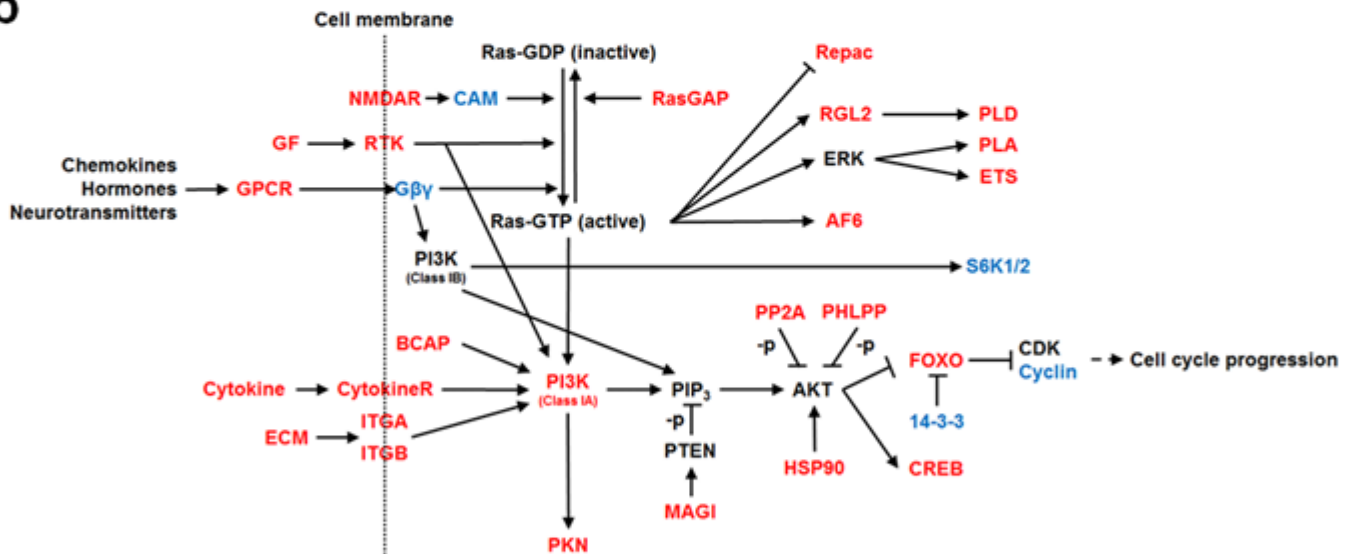
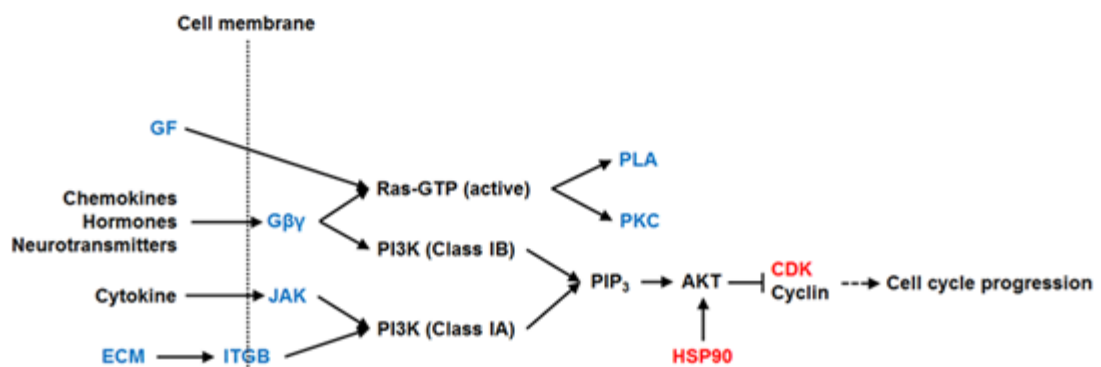
a**b****c**

a**b****c**







a**b****c**

Supplementary Figure Legends

Supplementary Figure 1: Human genes mapped by Fluidigm 3'-end scRNA-seq dataset from four NSCLC cell lines. a The average number of genes mapped with sequence reads per cell line. Lines and x marks in individual boxes indicate mean and median values, respectively. Data is expressed mean $\pm$ SEM. **b** The number of mapped genes per single cell after normalizing read counts per gene at counts per million reads (CPM) ≥ 1 .

Supplementary Figure 2: Measured quantities of CXCL1, CXCL2 and CXCR2 in primary lung tumors and normal lung tissues resected from Stage I LUAD patients. a-c Absolute quantification of CXCL1 in **a**, CXCL2 in **b** and CXCR2 in **c** from normal (black bars) and tumor (grey bars) tissues of female ($n = 16$) and male ($n = 18$) Stage I LUAD patients. X and y axes present individual patient IDs (female patients, 1-16, and male patients, 17-34) and measured quantities (ng), respectively. Data is expressed mean $\pm$ SEM from three separate experiments conducted in duplicate. FS2 supports Fig. 5.

Supplementary Figure 3: Measured quantities of microRNAs in primary lung tumors and adjacent normal lung tissues resected from Stage I LUAD patients. a-c Absolute quantification of miR-532-5p in **a**, miR-1266-3p in **b** and miR-3163 in **c** from normal (black bars) and tumor (grey bars) tissues of female ($n = 16$) and male ($n = 18$) Stage I LUAD patients. X and y axes present individual patient IDs (female patients, 1-16, and male patients, 17-34) and measured quantities (ng), respectively. Data is expressed mean $\pm$ SEM from three separate experiments conducted in duplicate. FS3 supports Fig. 5.

Supplementary Figure 4: Reference expression map mined from publicly available LUAD 10x Genomics 3'-end scRNA-seq dataset. Cluster plot of single cells ($n = 208,506$) from multiple-stage LUAD patients ($n = 58$). Clusters were generated from 10x Genomics 3'-end scRNA-seq dataset (GSE131907)¹⁶. FS4 supports Fig. 7.

Supplementary Figure 5: Survival analysis of chemokine genes in primary tumors from LUAD patients. a-c Kaplan-Meier survival plots for differential expression of *CXCL1* in **a**, *CXCL2* in **b** and *CXCR2* in **c** obtained from mined public bulk RNA-seq datasets of primary tumor tissues resected from female LUAD patients (early-Stage I; $n = 100$) and (late-Stage III & IV; $n = 38$), and male LUAD patients (early-Stage I; $n = 62$) and (late-Stage III & IV; $n = 34$). X and y axes indicate follow up in months and survival probability, respectively. The cutoff expression value between high and low expression per sex, stage and gene was determined by a log-rank test. FS5 supports Fig. 7.

Supplementary Figure 6: Validation of cyclin genes from human NSCLC epithelial cell lines. a Fold change validation of cyclin D1 (*CCND1*) in H460 (orange), Calu3 (red) and H1299 (green) compared with A549 (control cell line) and *ACTG1* (endogenous control gene) using a qRT-PCR analysis. Fold change values presented as mean $\pm$ SEM from three separate experiments conducted in duplicate. **** $P < 0.01$** from two-sample *t*-tests for measured quantities. **b** Reference clusters map generated from single cells ($n = 1,441$) from the four NSCLC cell lines. **c** Expression maps for *CCND1*, *CCND2* and *CCND3*.

Supplementary Figure 7: RAS-PI3K-AKT signaling pathways for cell cycle-associated DEGs. a-c RAS-PI3K-AKT signaling pathway encoded by up ($n = 10$) and down ($n = 12$)-regulated genes in Cluster 2 in **a**, up-regulated ($n = 44$) and down-regulated ($n = 10$) genes in Cluster 3 in **b**, and up-regulated ($n = 3$) and down-regulated ($n = 8$) genes in Cluster 4 in **c** compared with Cluster 1. Up- and down-regulated protein expression (Red and blue, respectively).

06-03-20
10 seconds

7% gel

15% gel

1st

A7CC

1. A549

2. H460

3. Calu3

4. H1299

STRATAGENE®

1 2 3 4

← E-cadherin

-1 α (rmc) 1:1000

-2 α (gdn) 1:1000

β-tubulin

-1 α (gpc) 1:2000

-2 α (gdn) 1:10000

← Actin

-1 α (gpc) 1:2000

-2 α (rac5) 1:10000

1 2 3 4

TBX7

-1 α (rmc) 1:3000

-2 α (gdn) 1:60000

1 2 3 4

CXCL2

-1 α (gpc) 1:1000

-2 α (rac5) 1:4000