

# Lung Precancer Study

Jaewoong Lee   S. Park   Y. Choi   I. Yun   Semin Lee

Department of Biomedical Engineering  
Ulsan National Institute of Science and Technology

*jwlee230@unist.ac.kr*

2021-10-05

# Overview

1 Introduction

2 Materials

3 Methods

4 Results

5 Discussion

6 References

# 1. Introduction

# 1. Introduction

## 1.1. Lung Cancer

# Lung Cancer?

**The most common cancer**

The most common form of cancer:

12.3 % of all cancers (Minna, Roth, & Gazdar, 2002)

**The most important factor**

**Tobacco**

# Cancer Survival Rate in Korea



Figure: Common cancer survival rates (S. Hong et al., 2021)

## Survival rate (More than 5 yr)

- Thyroid: 68.4 %
- Lung: 35.4 %

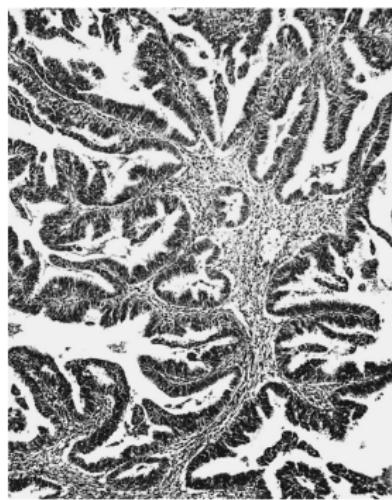
# Type of Lung Cancer

Types of lung cancer:

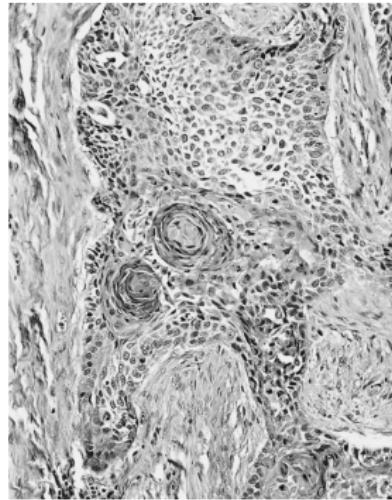
- ① Adenocarcinoma (LUAD) (40 %) ★
- ② Squamous cell carcinoma (LUSC) (25 %) ★
- ③ Small cell carcinoma (20 %)
- ④ Large cell carcinoma (10 %)
- ⑤ Adenosquamous carcinoma (< 5 %)
- ⑥ Carcinoid (< 5 %)
- ⑦ Bronchioalveolar (Bronchial gland carcinoma)

(Vincent et al., 1977; Collins, Haines, Perkel, & Enck, 2007)

# LUAD vs. LUSC I



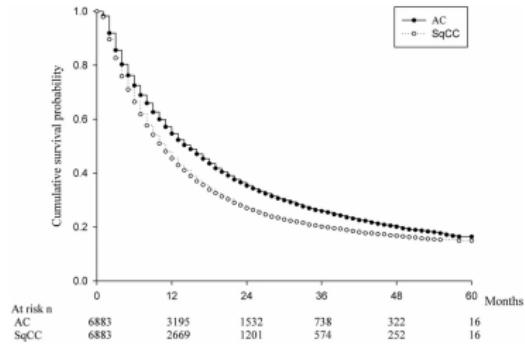
(a) LUAD



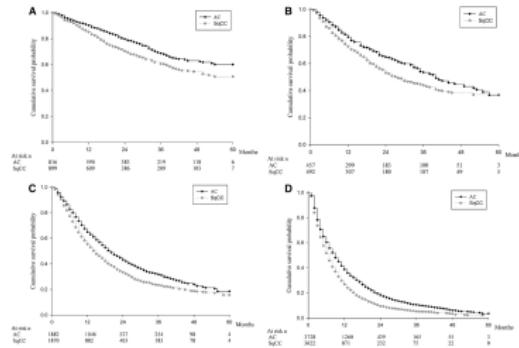
(b) LUSC

**Figure:** LUAD and LUSC histology in Lung cancer (Travis, 2002)

# LUAD vs. LUSC II



(a) All patients



(b) By cancer stages

Figure: Kaplan-Meier survival curves for LUAD & LUSC (Wang et al., 2020)

## Findings

LUSC is more dangerous than LUAD.  $\therefore p < 0.001$

## 1. Introduction

### 1.2. Study Objectives

# Study Objectives

## Find different mutations

- between WES vs. WTS
- from cancer vs. precancer

## Pathway examine

- with the mutation of WES & RNA-seq
- with immune-depleted animal models

## Ultra-deep sequencing

to find an *infinitesimal* quantity of Non-Circulating Tumor DNA

- from blood
- from urine
- from bronchus

## 2. Materials

# Lung Cancer Data

- WES (n=289) + Transcriptome (n=166)
- Normal + {Primary, CIS + AIS, AAH, Dysplasia, MIA}
  - Carcinoma in situ
  - Adenocarcinoma in situ
  - Atypical adenomatous hyperplasia
  - Dysplasia
  - Minimally invasive adenocarcinoma
- Adenocarcinoma (LUAD) & Squamous cell carcinoma (LUSC)
  - ① Normal → AAH → AIS → MIA → LUAD (n=28)
  - ② Normal → Dysplasia → CIS → LUSC (n=80)

### 3. Methods

### 3. Methods

#### 3.1. Workflows

# Data pre-processing for variant discovery



**Figure:** Data pre-processing for variant discovery (Van der Auwera et al., 2013; DePristo et al., 2011)

# Somatic short variant discovery

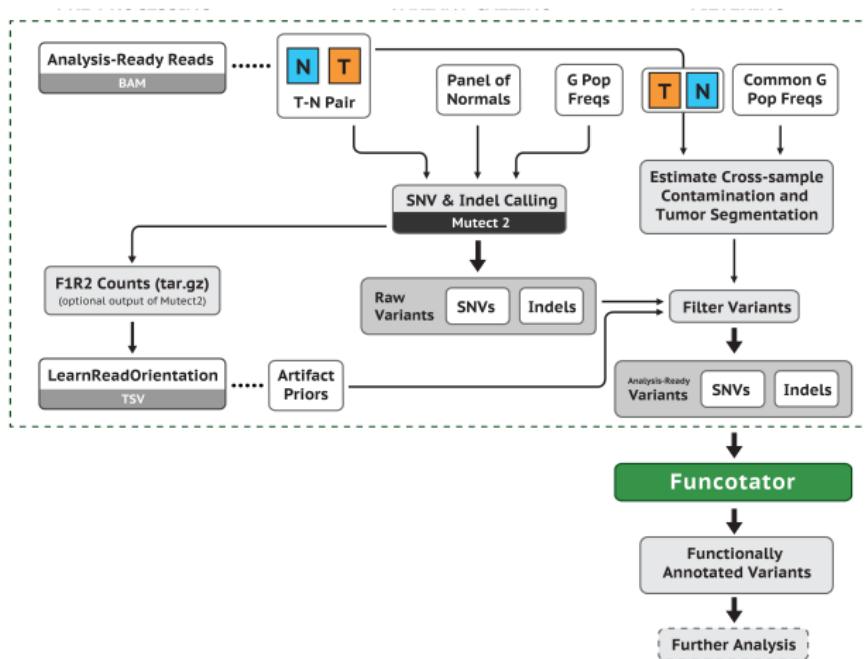


Figure: Somatic short variant (SNVs + Indels) discovery workflow (Van der Auwera et al., 2013; DePristo et al., 2011)

# Germline short variant discovery



**Figure:** Germline short variant (SNVs + Indels) discovery workflow (Van der Auwera et al., 2013; DePristo et al., 2011)

# RNA-seq short variant discovery



**Figure:** RNA-seq short variant (SNVs + Indels) discovery workflow (Van der Auwera et al., 2013; DePristo et al., 2011)

## 4. Results

## 4. Results

### 4.1. Quality Checks

# FastQC?



Figure: Example of FastQC Result (Andrews et al., 2012)

- A quality check tool for sequence data
- Give an overview that which test may be problems

# FastQC on WES

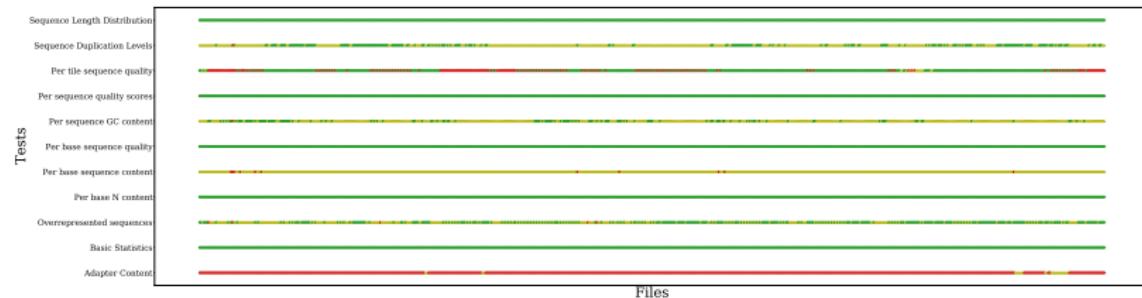
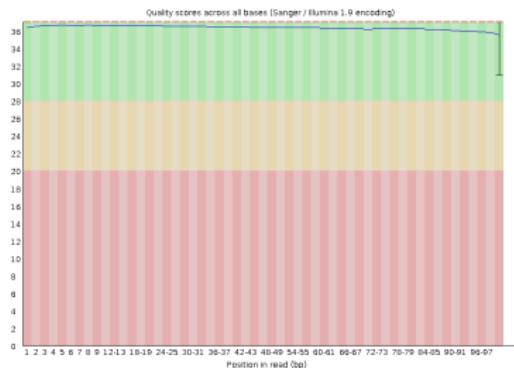


Figure: FastQC with WES data

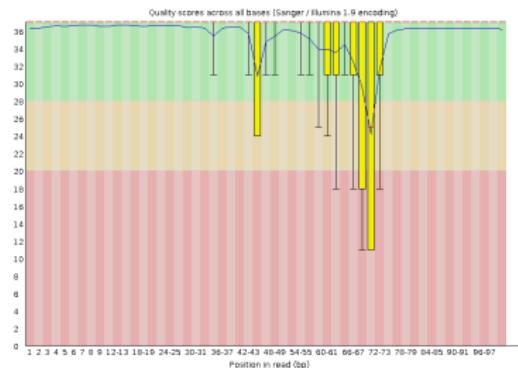
Failure on 33P1 sample

33P1 is excluded at further analysis.

# Failure on 33P1 I



(a) 33N



(b) 33P1

Figure: Per Base Sequence Quality Results

# Failure on 33P1 II

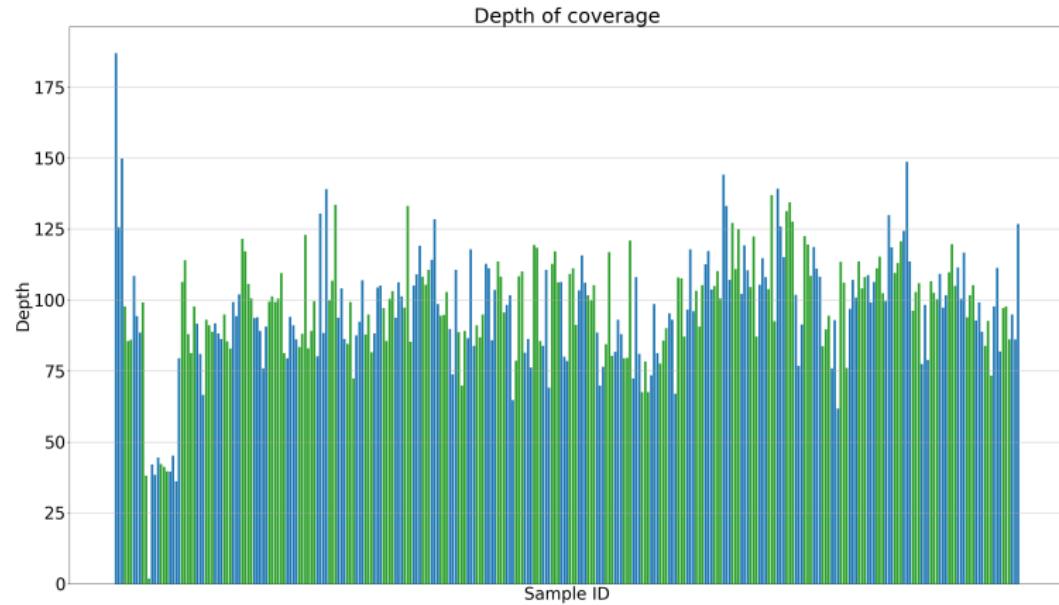


Figure: Coverage Depth Plot

## FastQC on WTS

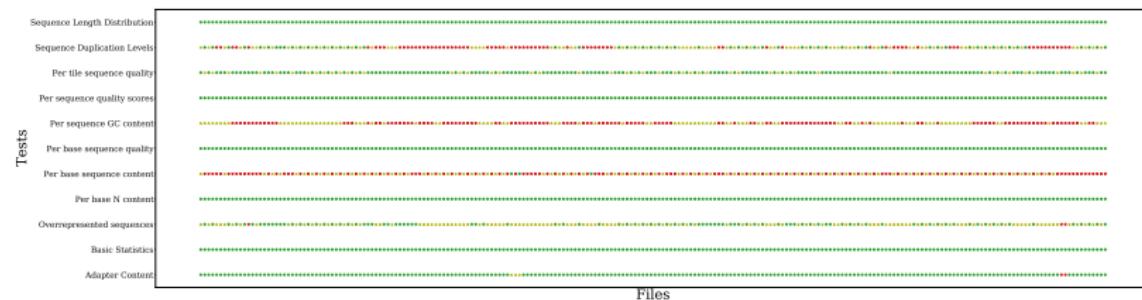


Figure: FastQC with WTS data

All sample are good to analysis

∴ No sample has more than 5 failures.

## 4. Results

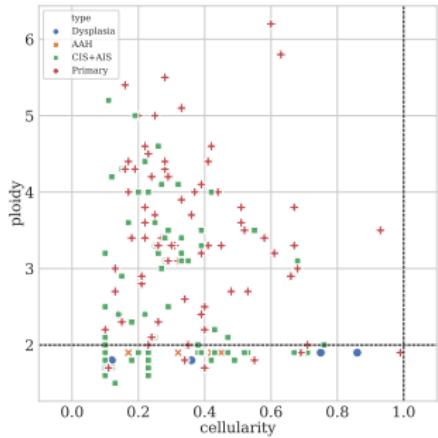
### 4.2. Copy Number Variations

# Sequenza?

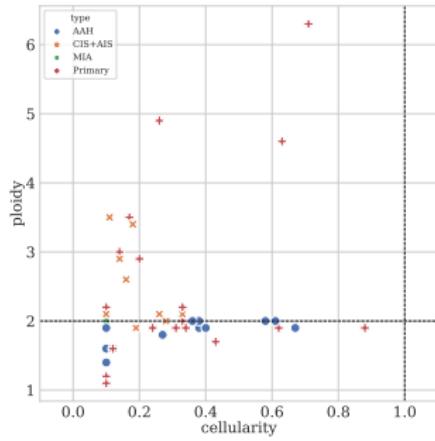


**Figure:** Representative Output of the Sequenza (Favero et al., 2015)

# Cellularity & Ploidy on WES



(a) LUSC Samples



(b) LUAD Samples

Figure: Cellularity and Ploidy from Sequenza

# Genome View on Patient #57

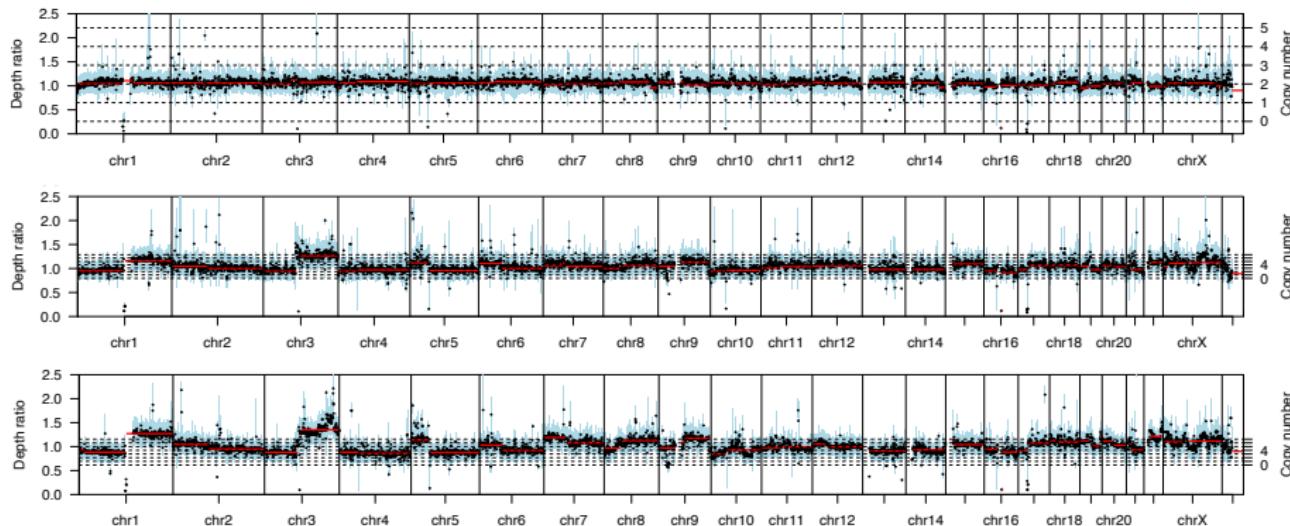


Figure: Dysplasia-CIS-Primary Tumor on Patient #57

# CNVs of LUSC

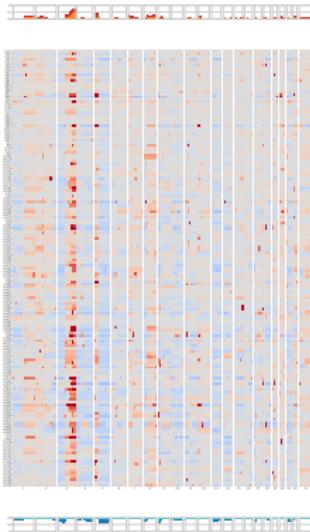
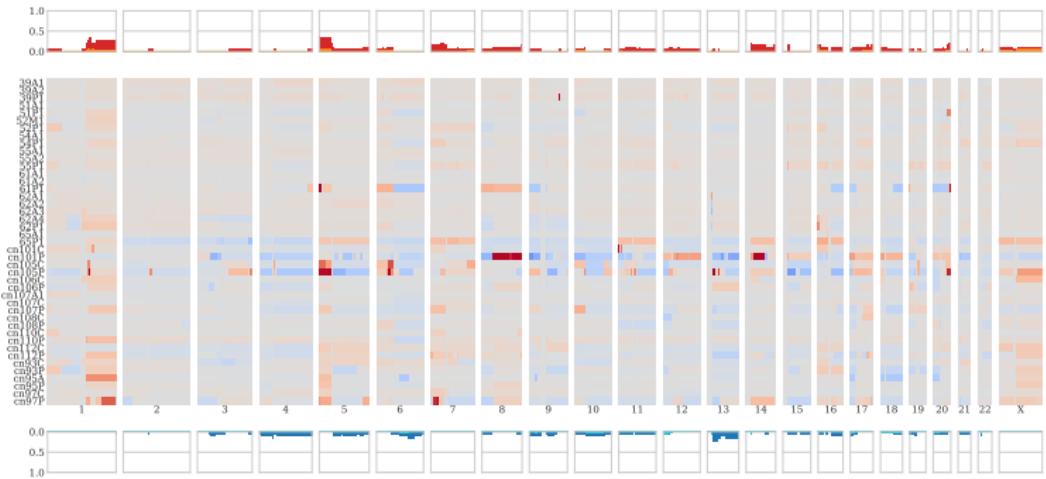


Figure: CNV Plot with LUSC Patients

## CNVs of LUAD



## Figure: CNV Plot with LUAD Patients

# LUSC vs. LUAD in CNV Plot

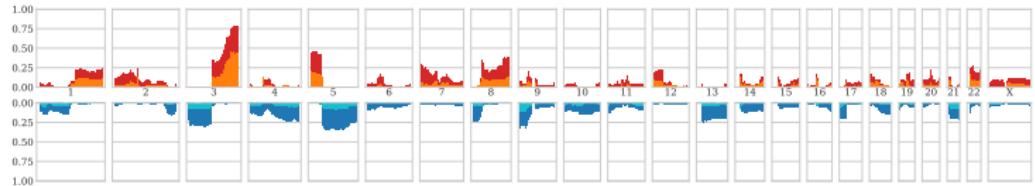


Figure: Simple CNV Plot with LUSC Patients

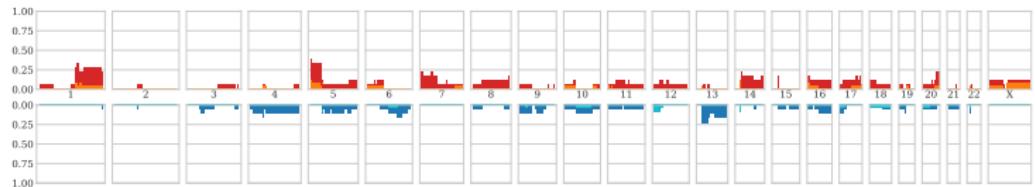


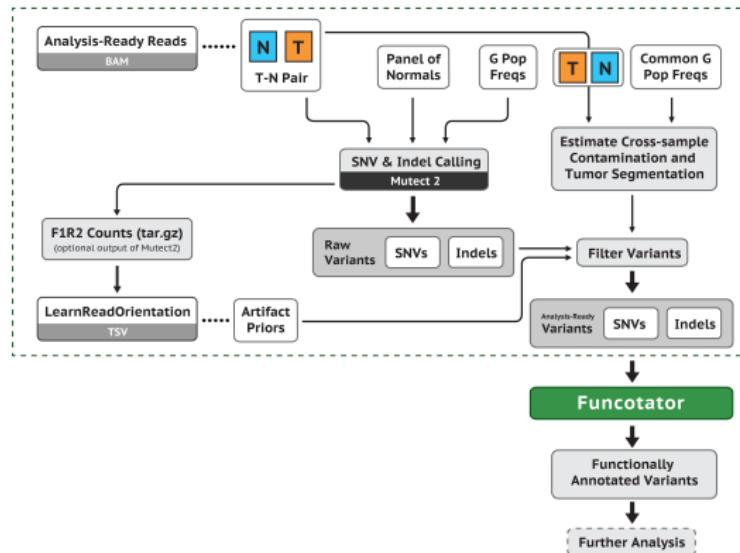
Figure: Simple CNV Plot with LUAD Patients

# Findings in Sequenza

## 4. Results

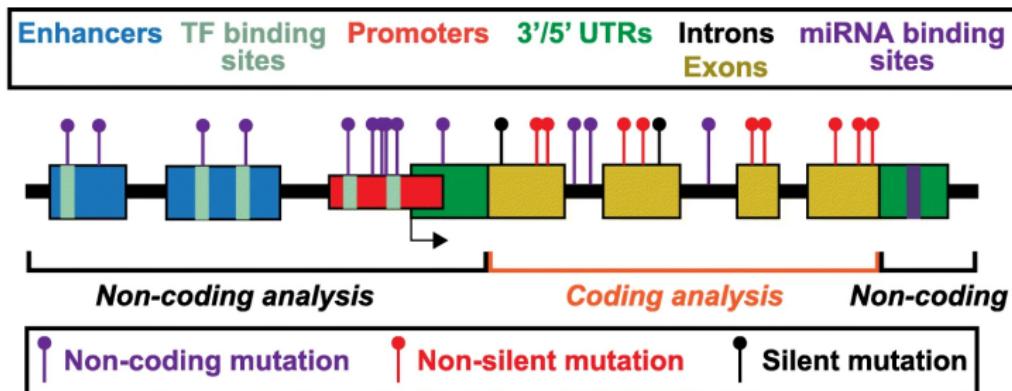
### 4.3. SNVs Analysis

# Mutect2?



**Figure:** Somatic short variant discovery workflow (Van der Auwera et al., 2013; DePristo et al., 2011)

# MutEnricher?



## Analysis summary:

### Inputs:

- Somatic mutations
- Features of interest:
  - Coding genes
  - Non-coding regions
- Genomic covariates (optional)

### Analyses:

- Background calculations:
  - global, local, or covariate clustered
- Mutation enrichments:
  - coding/non-coding modules

### Outputs:

- Gene or non-coding region enrichments:
  - Overall genes/regions
  - Hotspots
  - Combined

**Figure:** Schematic representation of MunEnricher's analysis procedures (Soltis et al., 2020)

# Driver Gene Selection Strategy

COSMIC Cancer Gene Census (Tate John et al., 2018)

Gene  $\in$  CGC Tier 1 set

Fisher FDR

Fisher FDR  $< 0.05$

Fisher P-value

Fisher P-value  $< 0.05$

Gene P-value

Gene P-value  $< 0.05$

# Somatic Variant in LUSC

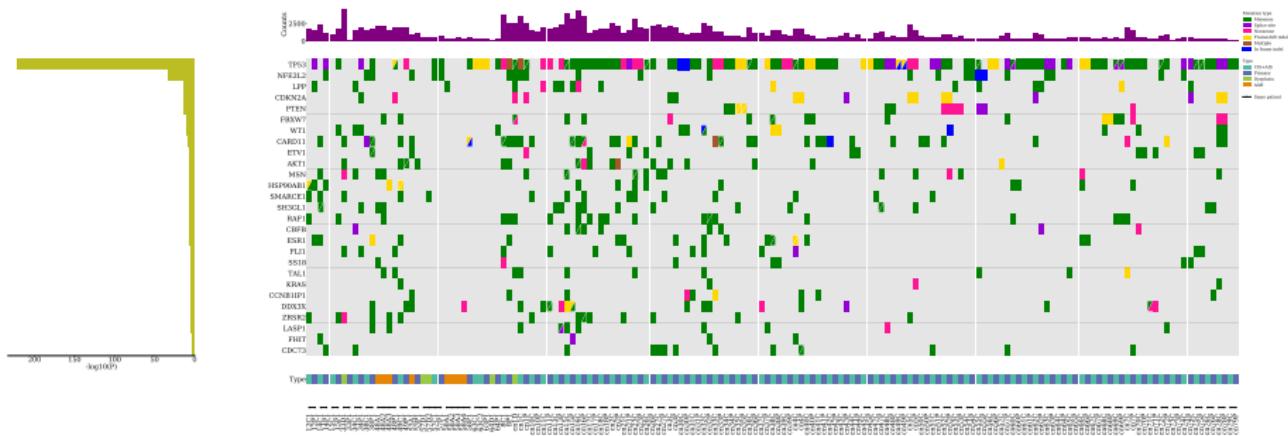


Figure: CoMut Plot with LUSC Patients

# Somatic Variant in LUAD

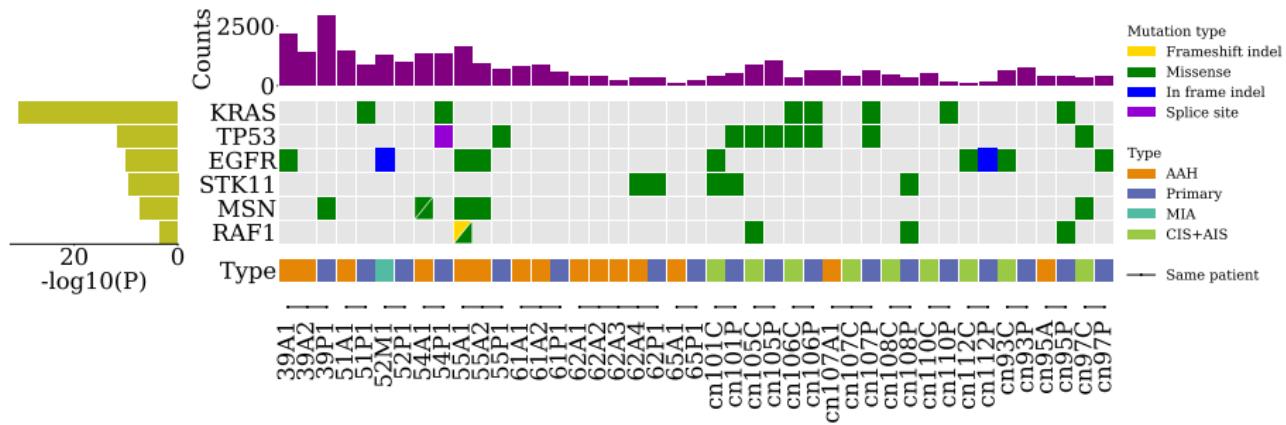


Figure: CoMut Plot with LUAD Patients

# Findings in SNVs Analysis

## 4. Results

### 4.4. VAF Analysis

# VAF?

- Variant allele frequency
- VAF = Alternative allele read count/Total read count
- To find tumor evolution

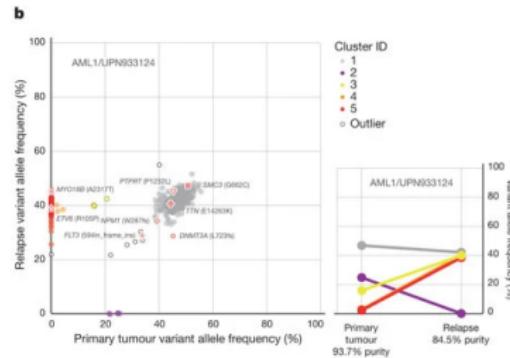


Figure: VAF distribution of validated mutations (Ding et al., 2012)

# VAF Plots I

# PyClone?

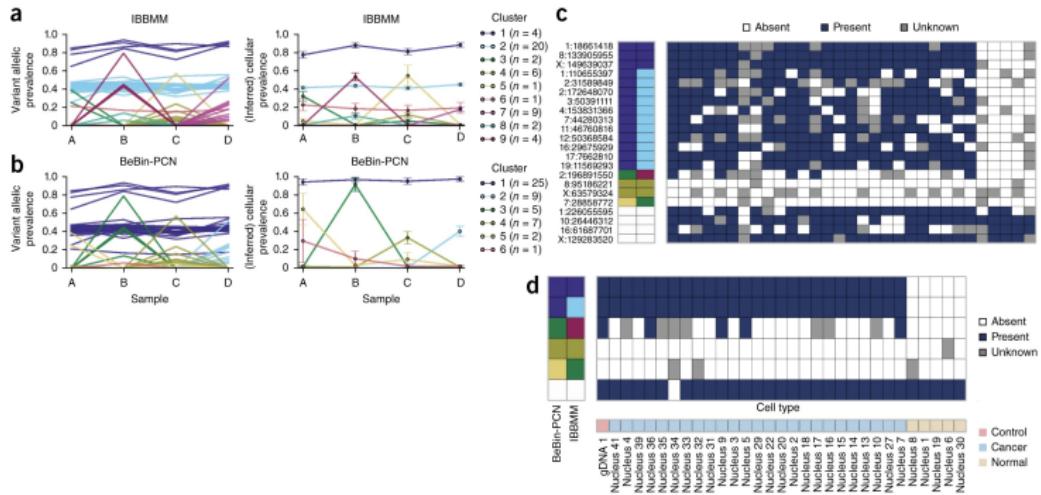


Figure: Analysis of multiple samples by PyClone (Roth et al., 2014)

# PyClone Plots I

# Findings in VAF Analysis

## 4. Results

### 4.5. Tumor Evolution Trajectories Analysis

# Revolver?

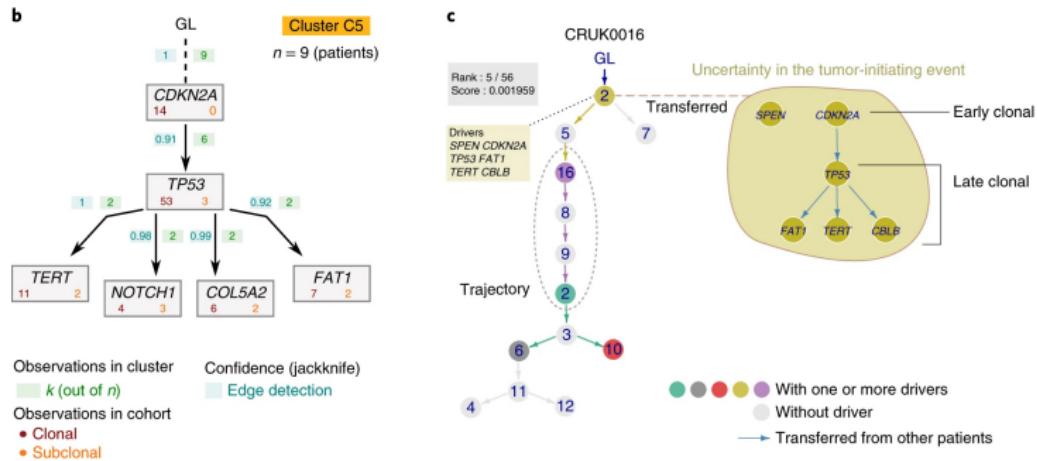


Figure: Repeated Evolutionary Trajectories (Caravagna et al., 2018)

# Findings in Tumor Evolution Trajectories Analysis

## 4. Results

### 4.6. Differences in Gene Expression Levels

# RSEM?

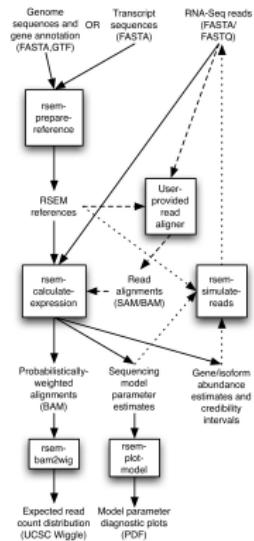


Figure: RSEM workflow (Li & Dewey, 2011)

# DESeq2?

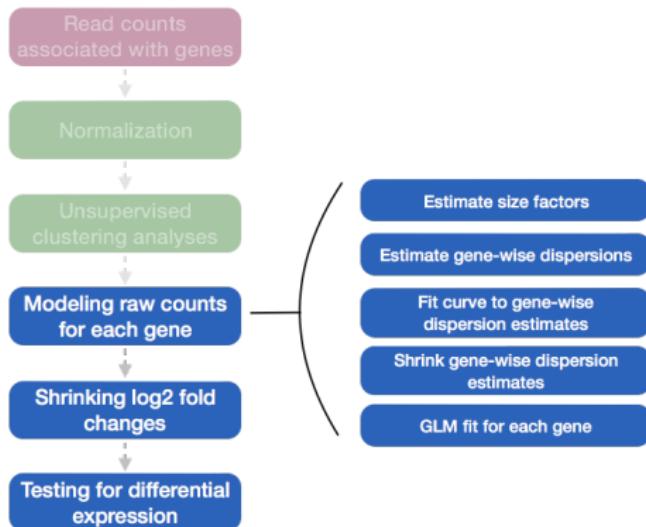


Figure: DESeq2 workflow (Love, Huber, & Anders, 2014)

# DEG Selection Strategy

DEG: differentially expressed genes

## Fold Change

$$\log_2(\text{Fold Change}) > 1 \vee \log_2(\text{Fold Change}) < -1$$

## P-value

$$P\text{-value} < 0.05$$

## Adjusted P-value

$$P_{adj} < 0.05$$

# Enrichr?

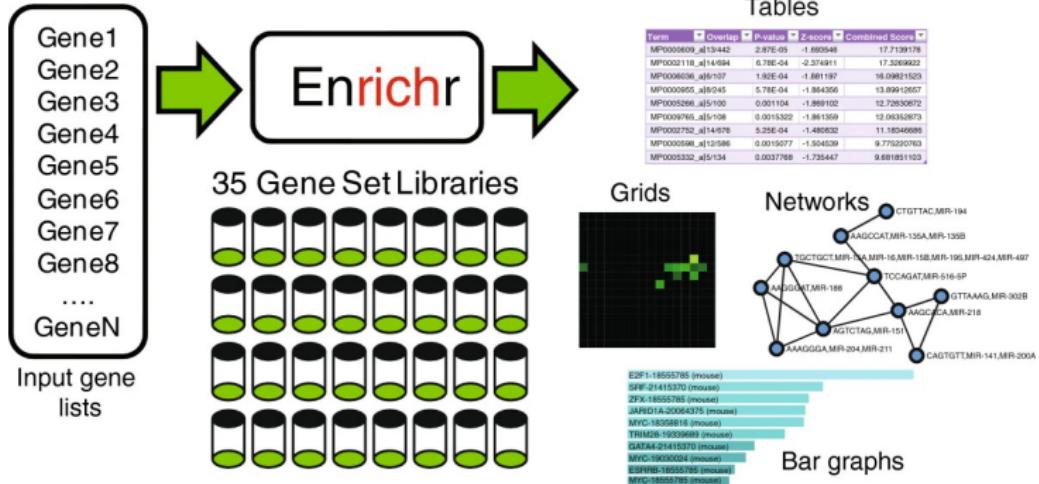


Figure: Enrichr workflow (Chen et al., 2013; Kuleshov et al., 2016)

# Gene-set Library

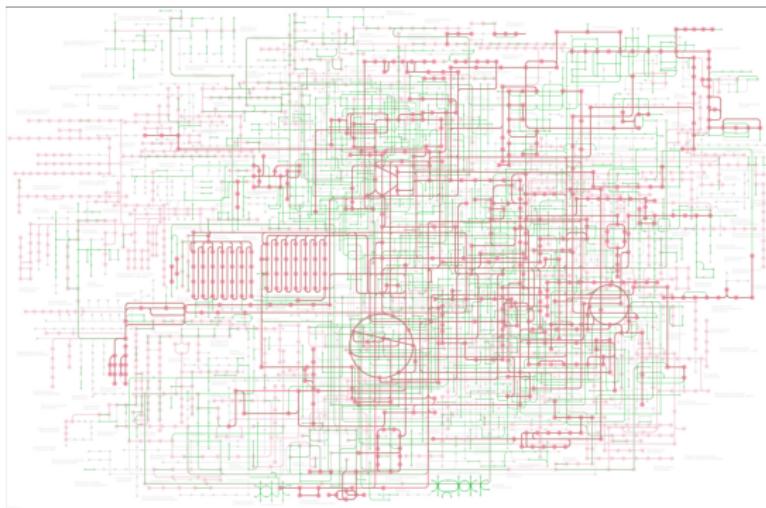


Figure: The global map of metabolic pathways by KEGG (Kanehisa et al., 2021)

KEGG

KEGG 2021 Human

# WTS Data Composition I

Table: Number of WTS samples

Cancer Subtype	Stage	Number of Samples
SQC (n=89)	Normal	17
	Dysplasia	2
	CIS	33
	Primary	35
ADC (n=30)	Normal	12
	AAH	1
	AIS	9
	MIA	0
	Primary	8

# WTS Data Composition II

Table: Number of WTS LUSC samples

Recurrence?	Stage	Number of Samples
Recurrence (n=13)	Normal	1
	Dysplasia	1
	CIS	5
	Primary	6
Non-recurrence (n=74)	Normal	16
	Dysplasia	1
	CIS	28
	Primary	29

# WTS Data Composition III

Table: Number of WTS LUAD samples

Recurrence?	Stage	Number of samples
Recurrence (n=4)	Normal	1
	AAH	0
	AIS	2
	MIA	0
	Primary	1
Non-recurrence (n=26)	Normal	11
	AAH	1
	AIS	7
	MIA	0
	Primary	7

## 4. Results

### 4.6. Differences in Gene Expression Levels

#### 4.6.1. Comparing cancer stage in LUSC

# DEG List in LUSC

Table: Up-regulated DEG in LUSC

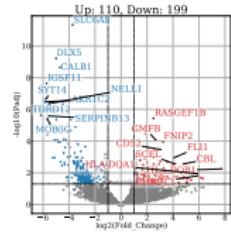
gene	log2FoldChange	pvalue	padj
AKR1C1	6.18e+00	5.14e-26	5.01e-23
AKR1C2	6.06e+00	1.19e-22	5.04e-20
CYP4F11	5.58e+00	1.51e-20	4.36e-18

Table: Down-regulated DEG in LUSC

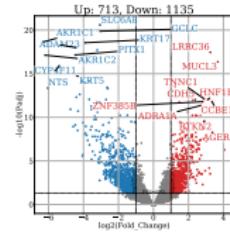
gene	log2FoldChange	pvalue	padj
SFTPC	-5.85e+00	9.16e-21	2.83e-18
FAM107A	-4.62e+00	2.27e-33	9.60e-30
LRRC36	-4.53e+00	5.49e-36	3.48e-32

# DEG Volcano Plots in LUSC

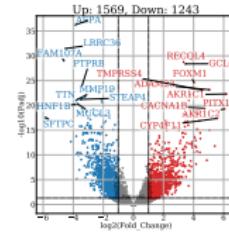
Normal → Dysplasia → CIS → Primary (LUSC)



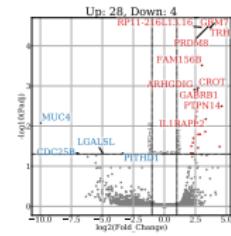
(a) Normal-Dysplasia



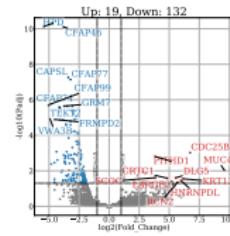
(b) Normal-CIS



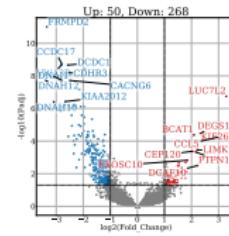
(c) Normal-Primary



(d) Dysplasia-CIS



(e) Dysplasia-Primary

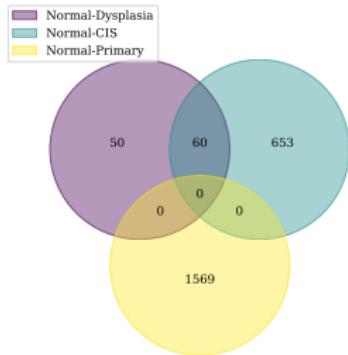


(f) CIS-Primary

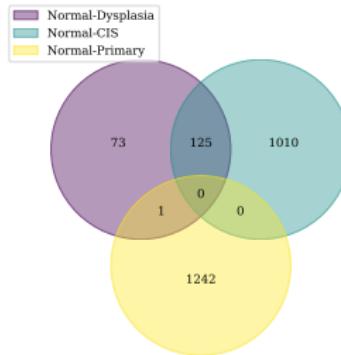
Figure: DEG Volcano Plots in LUSC

# DEG Venn Diagram in LUSC

Normal → Dysplasia → CIS → Primary (LUSC)



(a) Up-regulated



(b) Down-regulated

Figure: DEG Venn Diagram in LUSC

# Enrichment test with Normal vs. Dysplasia in LUSC

Table: Up-regulated Pathways on Normal vs. Dysplasia

Term name	Overlapping genes...	Adjusted p-value
Leishmaniasis	NFKBIA,HLA-DOA,TLR4,...(5)	6.72e-03
Lysosome	ASAH1,LAPTM5,CTSH,...(6)	6.72e-03
Phagosome	OLR1,HLA-DOA,TLR4,...(6)	1.15e-02

Table: Down-regulated Pathways on Normal vs. Dysplasia

Term name	Overlapping genes...	Adjusted p-value
None		

# Enrichment test with Normal vs. CIS in LUSC

Table: Up-regulated Pathways on Normal vs. CIS

Term name	Overlapping genes...	Adjusted p-value
Hematopoietic cell lineage	CSF1R,CSF3,IL4R,...(20)	7.22e-08
Malaria	CSF3,HGF,ITGB2,...(13)	1.16e-06
Cell adhesion molecules	NLGN3,SELPLG,CADM1,...(22)	1.16e-06

Table: Down-regulated Pathways on Normal vs. CIS

Term name	Overlapping genes...	Adjusted p-value
Metabolism of xenobiotics by cytochrome P450	GSTM4,CBR1,GSTM3,...(19)	9.34e-06
Drug metabolism	GSTM4,GSTM3,GSTM2,...(21)	9.06e-05
Cell cycle	CDKN2A,PLK1,BUB1B,...(22)	1.68e-04

# Enrichment test with Normal vs. Primary in LUSC

Table: Up-regulated Pathways on Normal vs. Primary

Term name	Overlapping genes...	Adjusted p-value
Cell cycle	HDAC1,PKMYT1,ORC4,...(27)	1.53e-04
Glutathione metabolism	GSTM4,GSTM3,G6PD,...(17)	1.53e-04
DNA replication	FEN1,RNASEH2A,RFC4,...(13)	1.72e-04

Table: Down-regulated Pathways on Normal vs. Primary

Term name	Overlapping genes...	Adjusted p-value
Hematopoietic cell lineage	CSF1R,CSF3,CSF3R,...(27)	7.33e-09
Malaria	IL10,CSF3,CR1,...(19)	7.33e-09
Hypertrophic cardiomyopathy	LAMA2,ITGB3,CACNA1D,...(25)	1.24e-08

# Findings in Comparing cancer stage in LUSC

## AKR1C1 & AKR1C2

- ① They are down-regulated in CIS, but up-regulated in Primary.
- ② They regulate steroids (Jin et al., 2009) and hormones (Penning et al., 2000) .
- ③ They promote the metastasis of NSCLC (Z. Hong et al., 2018).

## Down-regulation of SFTPC

- ① A pulmonary surfactant associated protein (Lin et al., 2018).
- ② SFTPC  $\downarrow \Rightarrow$  Poor survival in LUAD (Li et al., 2019).
- ③ Associated with lung disease in adult (Henderson et al., 2013) and baby (Brasch et al., 2004).

## 4. Results

### 4.6. Differences in Gene Expression Levels

#### 4.6.2. Comparing cancer stage in LUAD

# DEG List in LUSC

Table: Up-regulated DEG in LUAD

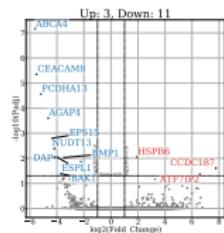
gene	log2FoldChange	pvalue	padj
ABCA4	4.95e+00	3.01e-12	2.58e-09
HMGA2	4.79e+00	8.06e-08	1.46e-05
KIF12	4.48e+00	1.33e-06	1.46e-04

Table: Down-regulated DEG in LUAD

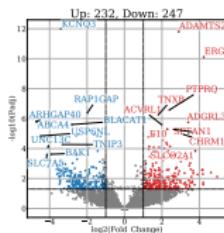
gene	log2FoldChange	pvalue	padj
SLC6A4	-6.20e+00	5.80e-10	2.36e-07
IL1RL1	-4.20e+00	7.47e-06	5.82e-04
RNF185	-4.06e+00	4.75e-05	2.45e-03

## DEG Volcano Plots in LUAD

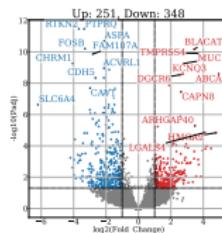
Normal → AAH → AIS → Primary (LUAD)



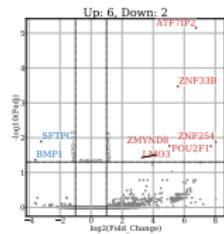
(a) Normal-AAH



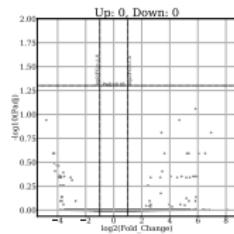
(b) Normal-AIS



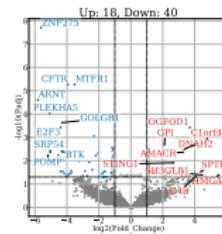
(c) Normal-Primary



(d) AAH-AIS



(e) AAH-Primary

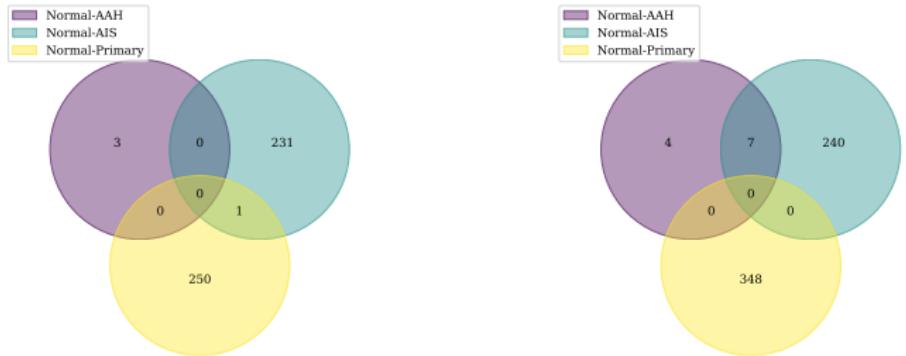


(f) AIS-Primary

## Figure: DEG Volcano Plots in LUAD

# DEG Venn Diagram in LUAD

Normal → AAH → AIS → Primary (LUAD)



(a) Up-regulated

(b) Down-regulated

Figure: DEG Venn Diagram in LUAD

## Enrichment test with Normal vs. AAH in LUAD

Table: Up-regulated Pathways on Normal vs. AAH

Term name	Overlapping genes...	Adjusted p-value
None		

Table: Down-regulated Pathways on Normal vs. AAH

Term name	Overlapping genes...	Adjusted p-value
None		

# Enrichment test with Normal vs. AIS in LUAD

Table: Up-regulated Pathways on Normal vs. AIS

Term name	Overlapping genes...	Adjusted p-value
Calcium signaling pathway	RYR2,NTRK2,CHRM1,...(11)	2.49e-02
Cell adhesion molecules	CDH5,CLDN5,NLGN1,...(8)	3.55e-02

Table: Down-regulated Pathways on Normal vs. AIS

Term name	Overlapping genes...	Adjusted p-value
None		

# Enrichment test with Normal vs. Primary in LUAD

Table: Up-regulated Pathways on Normal vs. Primary

Term name	Overlapping genes...	Adjusted p-value
None		

Table: Down-regulated Pathways on Normal vs. Primary

Term name	Overlapping genes...	Adjusted p-value
Vascular smooth muscle contraction	PPP1R14A, EDN1, RAMP2, ... (13)	1.38e-04
ECM-receptor interaction	TNXB, VWF, COL4A2, ... (10)	3.58e-04
Calcium signaling pathway	MCOLN3, CHRM1, NOS2, ... (16)	4.03e-04

# Finding in Comparing cancer stage in LUAD

## ABCA4

- ① Down-regulated in AAH & AIS, but up-regulated in Primary.

## KCNQ3

- ① Down-regulated in AIS, but up-regulated in Primary.

## CHRM1

- ① Up-regulated in AIS, but down-regulated in Primary.

## 4. Results

### 4.6. Differences in Gene Expression Levels

#### 4.6.3. Recur vs. Non-recur in LUSC

# LUSC Data Composition

Table: Number of WTS LUSC samples

Recurrence?	Stage	Number of Samples
Recurrence (n=13)	Normal	1
	Dysplasia	1
	CIS	5
	Primary	6
Non-recurrence (n=74)	Normal	16
	Dysplasia	1
	CIS	28
	Primary	29

## Normal samples

In order to compare with Normal stage, merging Normal samples.  
∴ Insufficient number of Normal samples in Recur.

# DEG List for Recur in CIS of LUSC

Table: Up-regulated DEG for Recur in CIS of LUSC

gene	log2FoldChange	pvalue	padj
ATAD5	3.62e+00	1.97e-07	6.43e-04
RGPD2	3.28e+00	1.28e-06	3.33e-03
KCNN3	3.25e+00	1.37e-07	6.43e-04

Table: Down-regulated DEG for Recur in CIS of LUSC

gene	log2FoldChange	pvalue	padj
ARL13B	-4.64e+00	1.05e-05	1.36e-02
PARP11	-4.18e+00	1.60e-07	6.43e-04
ADCK2	-3.57e+00	6.41e-05	4.91e-02

# DEG List for Recur in Primary of LUSC

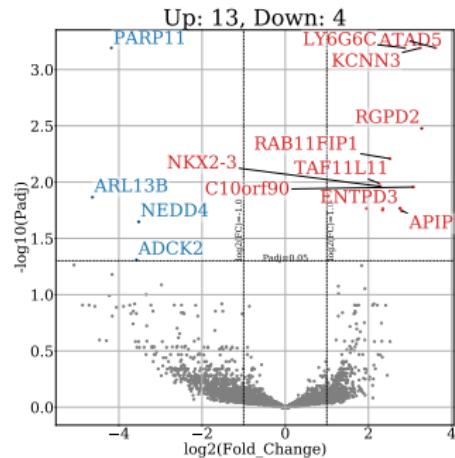
Table: Up-regulated DEG for Recur in Primary of LUSC

gene	log2FoldChange	pvalue	padj
GPC5	2.88e+00	1.58e-06	9.87e-03
AR	2.54e+00	5.21e-05	5.00e-02
COPE	2.30e+00	1.43e-06	9.87e-03

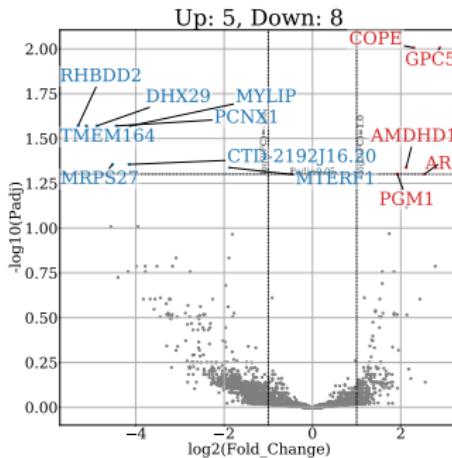
Table: Down-regulated DEG for Recur in Primary of LUSC

gene	log2FoldChange	pvalue	padj
RHBDD2	-5.31e+00	1.08e-05	2.69e-02
TMEM164	-5.12e+00	1.51e-05	2.69e-02
DHX29	-4.88e+00	9.47e-06	2.69e-02

# DEG Volcano Plots Recur vs. Non-recur in LUSC



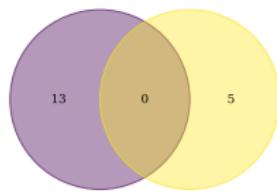
(a) CIS



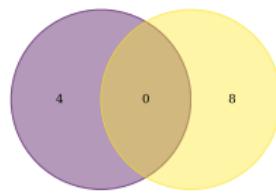
(b) Primary

Figure: DEG Volcano Plots Recur vs. Non-recur

# DEG Venn Diagram Recur vs. Non-recur with CIS in LUSC



(a) Up-regulated



(b) Down-regulated

Figure: DEG Venn Diagram Recur vs. Non-recur

# Finding in Comparing Recur vs. Non-recur in LUSC

## 4. Results

### 4.6. Differences in Gene Expression Levels

#### 4.6.4. Within Recur in LUSC

# DEG List for CIS within Recur in LUSC

Table: Up-regulated DEG for CIS within Recur in LUSC

gene	log2FoldChange	pvalue	padj
MFAP4	6.77e+00	2.70e-09	3.72e-07
TBX2	5.90e+00	1.19e-05	3.40e-04
SFTPC	5.47e+00	8.66e-08	6.57e-06

Table: Down-regulated DEG for CIS within Recur in LUSC

gene	log2FoldChange	pvalue	padj
AKR1C2	-7.44e+00	4.70e-12	1.74e-09
AKR1C1	-7.09e+00	2.35e-35	2.74e-31
CYP4F11	-6.70e+00	1.95e-14	1.75e-11

# DEG List for Primary within Recur in LUSC

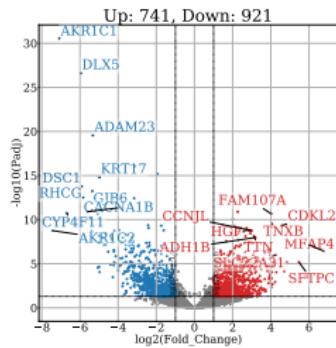
Table: Up-regulated DEG for Primary within Recur in LUSC

gene	log2FoldChange	pvalue	padj
AKR1C2	6.66e+00	4.34e-11	9.07e-09
AKR1C1	6.59e+00	2.62e-27	3.06e-23
CYP4F11	6.25e+00	3.61e-11	7.67e-09

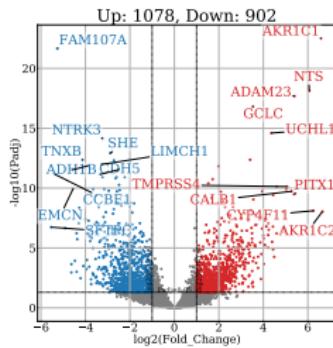
Table: Down-regulated DEG for Primary within Recur in LUSC

gene	log2FoldChange	pvalue	padj
SFTPC	-5.54e+00	1.56e-09	1.83e-07
CCBE1	-5.36e+00	9.73e-15	7.11e-12
FAM107A	-5.27e+00	3.64e-26	2.13e-22

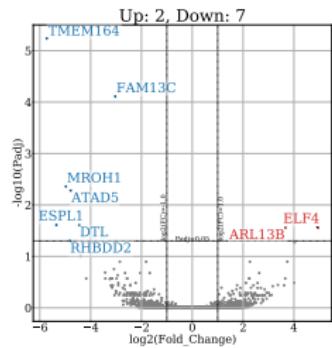
# DEG Volcano Plots with Recur in LUSC



(a) Normal-CIS



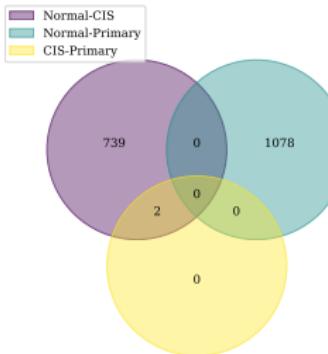
(b) Normal-Primary



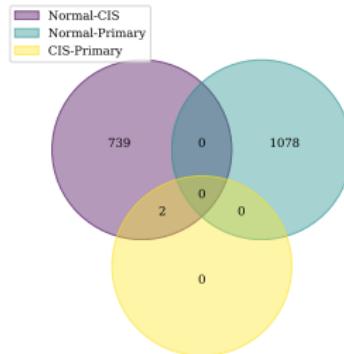
(c) CIS-Primary

Figure: DEG Volcano Plots with Recur samples in LUSC

# DEG Venn Diagram with Recur in LUSC



(a) Up-regulated



(b) Down-regulated

Figure: DEG Venn Diagram with Recur samples in LUSC

## Enrichment test with Normal vs. CIS for Recur

Table: Up-regulated Pathways on Normal vs. CIS for Recur in LUSC

Term name	Overlapping genes...	Adjusted p-value
Hematopoietic cell lineage	CSF3,CSF3R,IL4R,...(17)	1.87e-05
Cell adhesion molecules	SELPLG,CADM1,SDC3,...(21)	1.87e-05
Hypertrophic cardiomyopathy	EDN1,ACE,TNNC1,...(15)	9.66e-05

Table: Down-regulated Pathways on Normal vs. CIS for Recur in LUSC

Term name	Overlapping genes...	Adjusted p-value
Parkinson disease	COX7B,NDUFA12,NDUFB5,...(32)	2.11e-05
Alzheimer disease	COX7B,NDUFA12,NDUFB5,...(41)	2.11e-05
Huntington disease	DCTN5,COX7B,NDUFA12,...(36)	2.11e-05

# Enrichment test with Normal vs. Primary for Recur

Table: Up-regulated Pathways on Normal vs. Primary for Recur in LUSC

Term name	Overlapping genes...	Adjusted p-value
Glycolysis / Gluconeogenesis	GPI,TPI1,PDHA1,...(17)	1.90e-05
RNA transport	EIF4A2,NUP205,NUP62,...(29)	2.66e-05
Drug metabolism	GSTM4,GSTM3,GSTM2,...(21)	2.66e-05

Table: Down-regulated Pathways on Normal vs. Primary for Recur in LUSC

Term name	Overlapping genes...	Adjusted p-value
Dilated cardiomyopathy	LAMA2,ITGA3,TNNC1,...(20)	2.19e-06
Hypertrophic cardiomyopathy	LAMA2,ITGA3,TNNC1,...(19)	2.19e-06
Arrhythmogenic right ventricular cardiomyopathy	TCF7L2,LAMA2,ACTN2,...(17)	4.12e-06

# Finding in Comparing within Recur in LUSC

## AKR1C1

- ① Down-regulated in CIS, but up-regulated in Primary.

## ADAM23

- ① Down-regulated in CIS, but up-regulated in Primary.

## FAM107A

- ① Up-regulated in CIS, but down-regulated in Primary.

## 4. Results

### 4.6. Differences in Gene Expression Levels

#### 4.6.5. Within Non-recur in LUSC

# DEG List for CIS within Non-recr in LUSC

Table: Up-regulated DEG for CIS within Non-recr in LUSC

gene	log2FoldChange	pvalue	padj
SFTPC	3.89e+00	1.33e-08	1.10e-06
MUCL3	3.64e+00	4.99e-18	6.71e-15
HLA-DRB1	3.40e+00	3.00e-05	6.50e-04

Table: Down-regulated DEG for CIS within Non-recr in LUSC

gene	log2FoldChange	pvalue	padj
AKR1C1	-6.10e+00	1.95e-19	4.73e-16
AKR1C2	-5.81e+00	4.57e-17	5.03e-14
NTS	-5.19e+00	1.60e-13	8.61e-11

# DEG List for Primary within Non-recr in LUSC

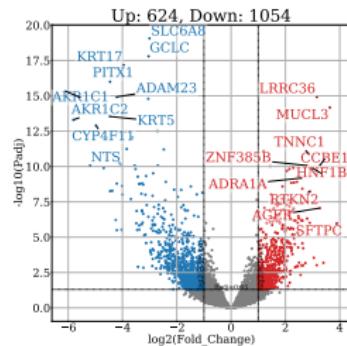
Table: Up-regulated DEG for Primary within Non-recr in LUSC

gene	log2FoldChange	pvalue	padj
AKR1C1	6.10e+00	9.04e-23	6.57e-20
AKR1C2	5.91e+00	3.13e-19	9.92e-17
NTS	5.78e+00	2.01e-14	2.28e-12

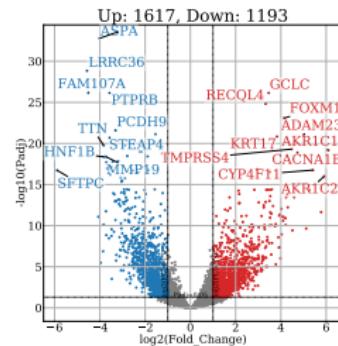
Table: Down-regulated DEG for Primary within Non-recr in LUSC

gene	log2FoldChange	pvalue	padj
SFTPC	-5.89e+00	5.10e-20	1.86e-17
LRRC36	-4.57e+00	2.29e-33	1.42e-29
FAM107A	-4.51e+00	2.49e-30	7.01e-27

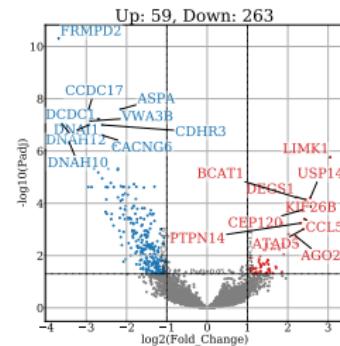
# DEG Volcano Plots with Non-recr in LUSC



(a) Normal-CIS



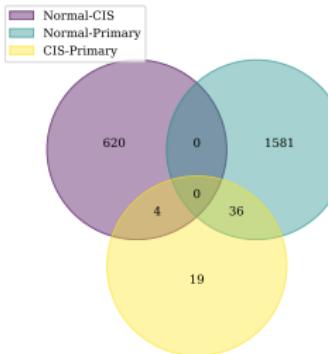
(b) Normal-Primary



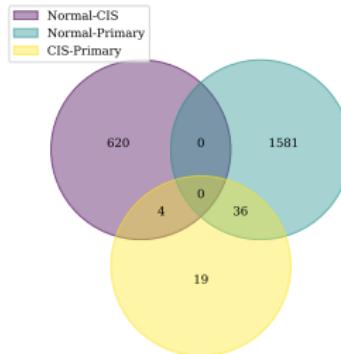
(c) CIS-Primary

Figure: DEG Volcano Plots with Non-recr samples in LUSC

# DEG Venn Diagram with Non-recr in LUSC



(a) Up-regulated



(b) Down-regulated

Figure: DEG Venn Diagram with Non-recr in LUSC

# Enrichment test with Normal vs. CIS for Non-recur

Table: Up-regulated Pathways on Normal vs. CIS for Non-recur in LUSC

Term name	Overlapping genes...	Adjusted p-value
Malaria	CSF3,HGF,ITGB2,...(14)	6.53e-08
Hematopoietic cell lineage	CSF1R,CSF3,MME,...(18)	2.01e-07
Hypertrophic cardiomyopathy	EDN1,ACE,LAMA2,...(16)	1.53e-06

Table: Down-regulated Pathways on Normal vs. CIS for Non-recur in LUSC

Term name	Overlapping genes...	Adjusted p-value
Metabolism of xenobiotics by cytochrome P450	GSTM4,CBR1,GSTM3,...(17)	9.67e-05
Drug metabolism	GSTM4,GSTM3,RRM1,...(20)	1.18e-04
Cell cycle	GADD45A,CDKN2A,PLK1,...(21)	1.89e-04

## Enrichment test with Normal vs. Primary for Non-recur

Table: Up-regulated Pathways on Normal vs. Primary for Non-recur in LUSC

Term name	Overlapping genes...	Adjusted p-value
Cell cycle	HDAC1,PKMYT1,ORC5,...(31)	3.04e-06
DNA replication	FEN1,RNASEH2A,RFC4,...(15)	6.47e-06
Homologous recombination	BLM,RPA1,PALB2,...(15)	3.33e-05

Table: Down-regulated Pathways on Normal vs. Primary for Non-recur in LUSC

Term name	Overlapping genes...	Adjusted p-value
Hematopoietic cell lineage	CSF1R,CSF3,CSF3R,...(28)	6.65e-10
Malaria	IL10,CSF3,CR1,...(19)	3.57e-09
Hypertrophic cardiomyopathy	LAMA2,ITGB3,CACNA1D,...(25)	5.12e-09

# Finding in Comparing within Recur in LUSC

## AKR1C1 & AKR1C2

- ① Down-regulated in CIS, but up-regulated in Primary.

## CYP4F11

- ① Down-regulated in CIS, but up-regulated in Primary.

## LRRC36

- ① Up-regulated in CIS, but down-regulated in Primary.

## 4. Results

### 4.6. Differences in Gene Expression Levels

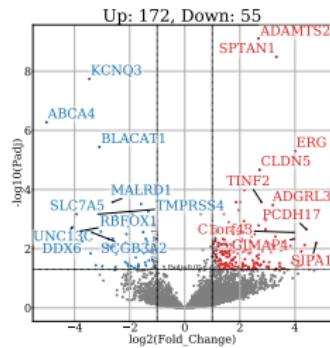
#### 4.6.6. Within Non-recur in LUAD

# LUAD Data Composition

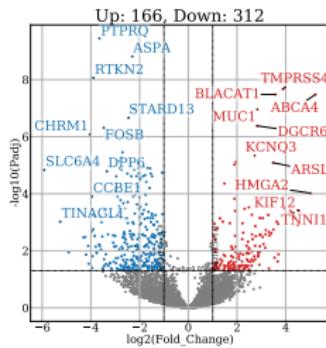
Table: Number of WTS LUAD samples

Recurrence?	Stage	Number of samples
Recurrence (n=4)	Normal	1
	AAH	0
	AIS	2
	MIA	0
	Primary	1
Non-recurrence (n=26)	Normal	11
	AAH	1
	AIS	7
	MIA	0
	Primary	7

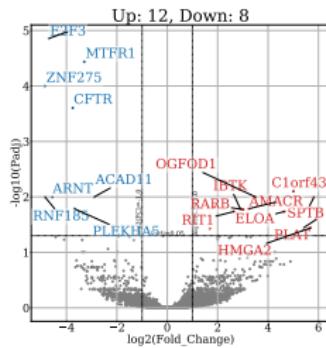
# DEG Volcano Plots with Non-recr in LUAD



(a) Normal-AIS



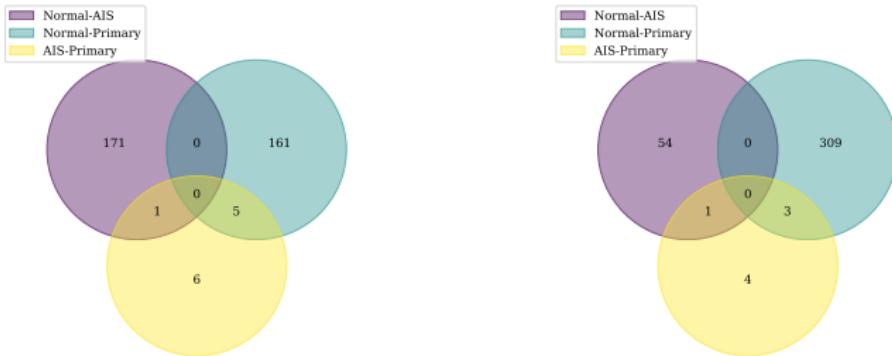
(b) Normal-Primary



(c) AIS-Primary

Figure: DEG Volcano Plots with Non-recr samples in LUAD

# DEG Venn Diagram with Non-recr in LUAD



(a) Up-regulated

(b) Down-regulated

Figure: DEG Venn Diagram with Non-recr in LUAD

# Enrichment test with Normal vs. AIS in LUAD

Table: Up-regulated Pathways on Normal vs. AIS for Non-recur in LUAD

Term name	Overlapping genes...	Adjusted p-value
Calcium signaling pathway	NTRK2, RYR2, CHRM1, ... (9)	3.90e-02

Table: Down-regulated Pathways on Normal vs. AIS for Non-recur in LUAD

Term name	Overlapping genes...	Adjusted p-value
None		

## Enrichment test with Normal vs. Primary in LUAD

Table: Up-regulated Pathways on Normal vs. Primary for Non-recur in LUAD

Term name	Overlapping genes...	Adjusted p-value
None		

Table: Down-regulated Pathways on Normal vs. Primary for Non-recur in LUAD

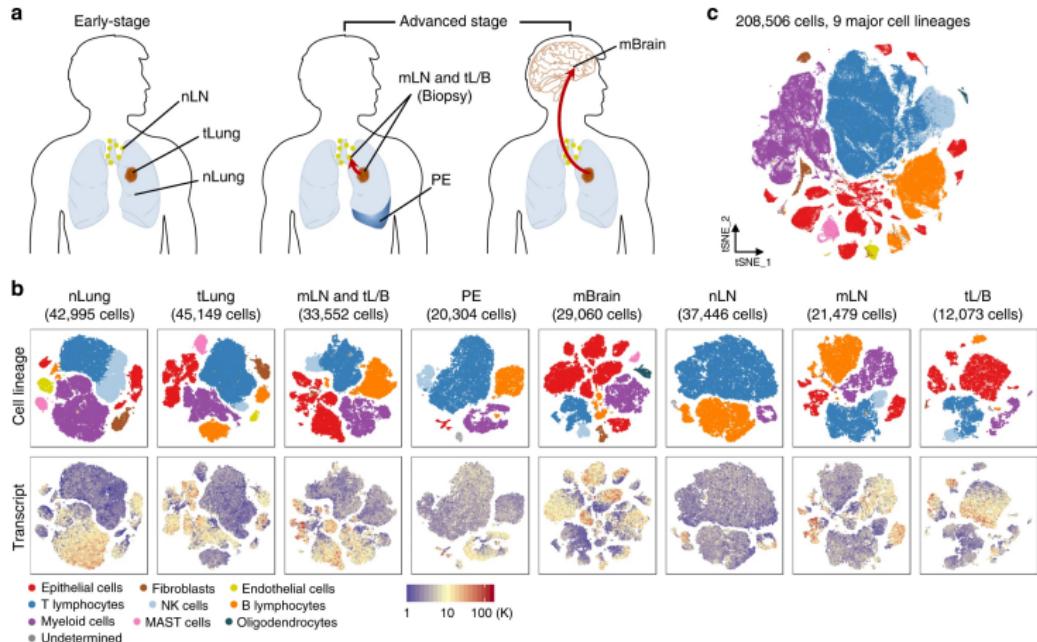
Term name	Overlapping genes...	Adjusted p-value
ECM-receptor interaction	TNXB,VWF,COL4A2,...(9)	2.05e-03
Vascular smooth muscle contraction	PPP1R14A,EDN1,AGTR1,...(10)	4.98e-03
Calcium signaling pathway	MCOLN3,CHRM1,NOS2,...(13)	7.82e-03

# Findings in DEG Analysis

## 4. Results

### 4.7. Bulk Cell Deconvolution

# Single-cell data as Reference



**Figure:** Comprehensive dissection and clustering of 208,506 single cells from LUAD patients (Kim et al., 2020)

# BisqueRNA?

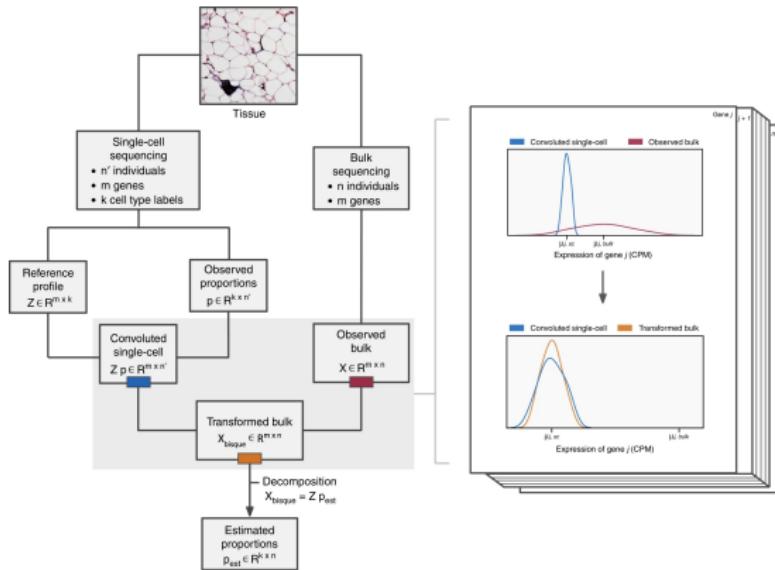


Figure: Workflow for BisqueRNA (Jew et al., 2020)

# Cluster Plot in LUSC

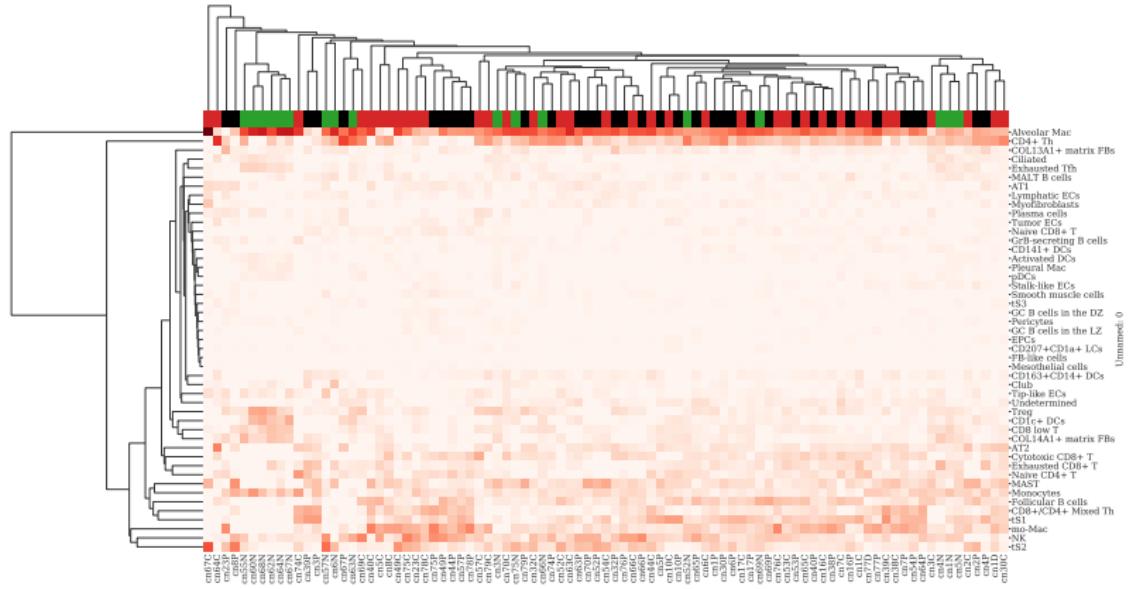


Figure: Cluster Plot in LUSC

# Violin Plots in LUSC I

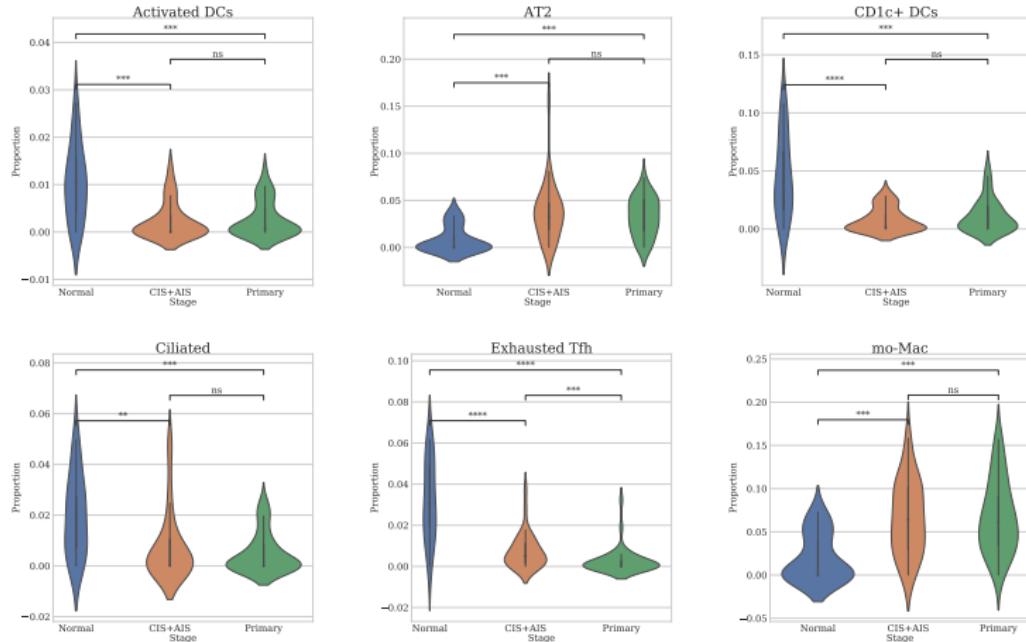


Figure: Violin Plots in LUSC

# Violin Plots in LUSC II

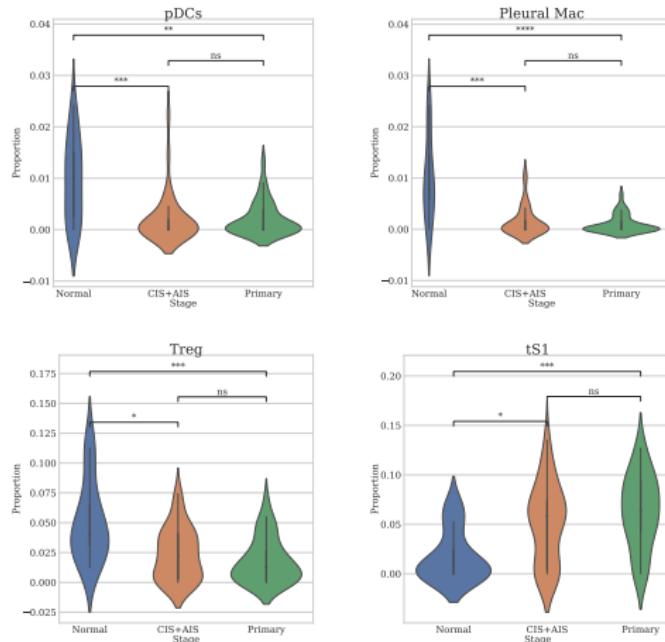


Figure: Violin Plots in LUSC

# Cluster Plot in LUAD

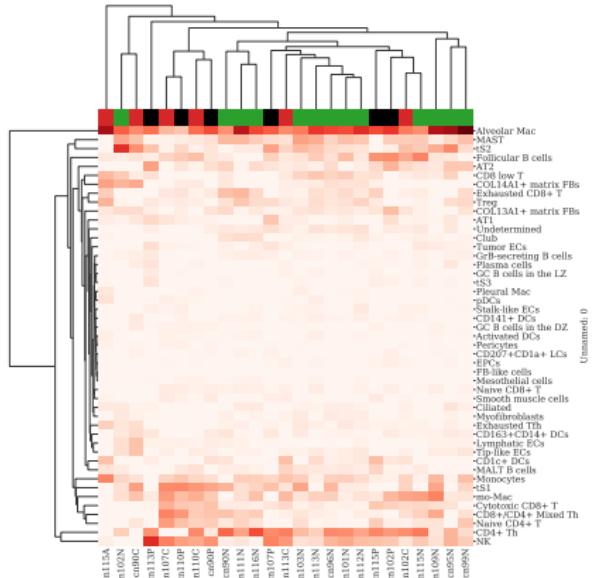


Figure: Cluster Plot in LUAD

# Violin Plots in LUAD

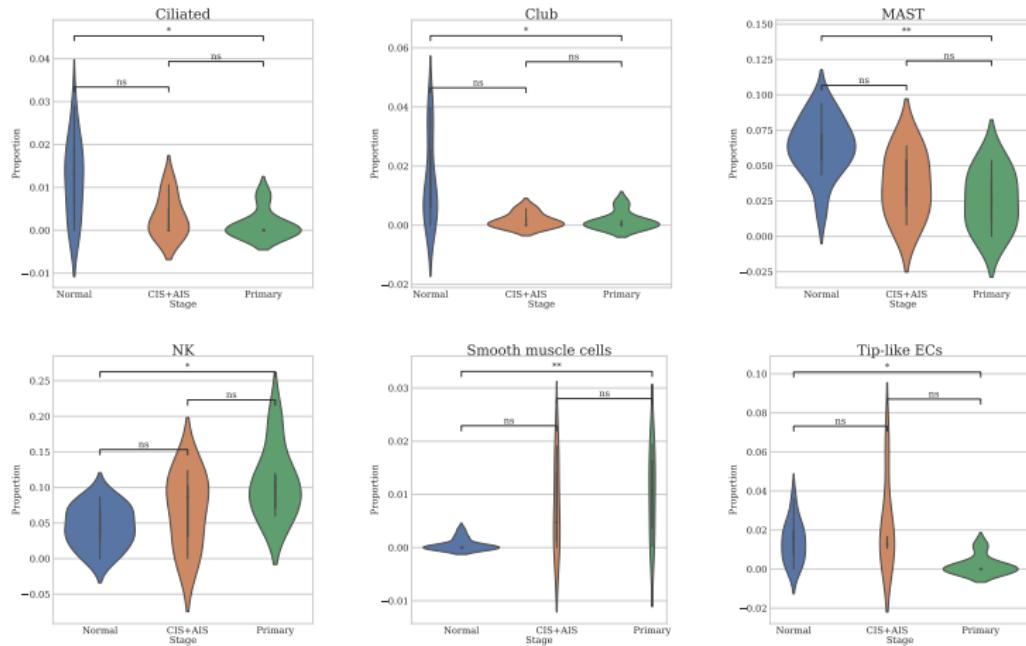


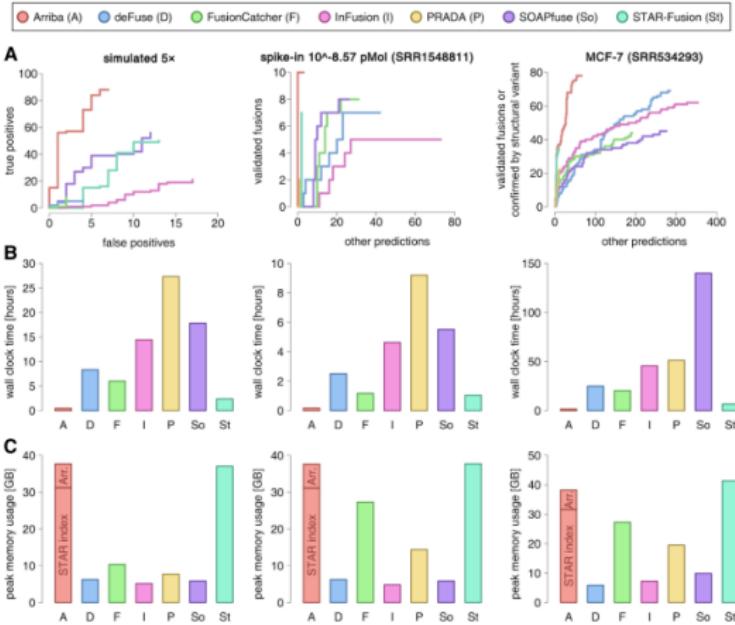
Figure: Violin Plots in LUAD

# Findings in Bulk Cell Deconvolution

## 4. Results

### 4.8. Discovery of Gene Fusion

# Arriba?



**Figure:** Benchmark of Arriba versus alternative methods (Uhrig et al., 2021)

# Findings in Gene Fusion Discovery

## 5. Discussion

## 6. References

# References I

- Andrews, S., Krueger, F., Segonds-Pichon, A., Biggins, L., Krueger, C., & Wingett, S. (2012, January). *FastQC*. Babraham Institute. Babraham, UK.
- Brasch, F., Griese, M., Tredano, M., Johnen, G., Ochs, M., Rieger, C., ... Beers, M. (2004). Interstitial lung disease in a baby with a de novo mutation in the sftpc gene. *European Respiratory Journal*, 24(1), 30–39.
- Caravagna, G., Giarratano, Y., Ramazzotti, D., Tomlinson, I., Graham, T. A., Sanguinetti, G., & Sottoriva, A. (2018). Detecting repeated cancer evolution from multi-region tumor sequencing data. *Nature methods*, 15(9), 707–714.
- Chen, E. Y., Tan, C. M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G. V., ... Ma'ayan, A. (2013). Enrichr: interactive and collaborative html5 gene list enrichment analysis tool. *BMC bioinformatics*, 14(1), 1–14.

## References II

- Collins, L. G., Haines, C., Perkel, R., & Enck, R. E. (2007). Lung cancer: diagnosis and management. *American family physician*, 75(1), 56–63.
- DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., ... others (2011). A framework for variation discovery and genotyping using next-generation dna sequencing data. *Nature genetics*, 43(5), 491.
- Ding, L., Ley, T. J., Larson, D. E., Miller, C. A., Koboldt, D. C., Welch, J. S., ... others (2012). Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature*, 481(7382), 506–510.
- Favero, F., Joshi, T., Marquard, A. M., Birkbak, N. J., Krzystanek, M., Li, Q., ... Eklund, A. C. (2015). Sequenza: allele-specific copy number and mutation profiles from tumor sequencing data. *Annals of Oncology*, 26(1), 64–70.

## References III

- Henderson, L. B., Melton, K., Wert, S., Couriel, J., Bush, A., Ashworth, M., & Nogee, L. M. (2013). Large abca3 and sftpc deletions resulting in lung disease. *Annals of the American Thoracic Society*, 10(6), 602–607.
- Hong, S., Won, Y.-J., Lee, J. J., Jung, K.-W., Kong, H.-J., Im, J.-S., ... others (2021). Cancer statistics in korea: Incidence, mortality, survival, and prevalence in 2018. *Cancer Research and Treatment: Official Journal of Korean Cancer Association*, 53(2), 301.
- Hong, Z., Chang, L.-L., Fang-Jie, Y., Yan, H., Chen-Ming, Z., Tian-Yi, Z., ... others (2018). Akr1c1 activates stat3 to promote the metastasis of non-small cell lung cancer. *Theranostics*, 8(3), 676.
- Jew, B., Alvarez, M., Rahmani, E., Miao, Z., Ko, A., Garske, K. M., ... Halperin, E. (2020). Accurate estimation of cell composition in bulk expression through robust integration of single-cell information. *Nature communications*, 11(1), 1–11.

## References IV

- Jin, Y., Duan, L., Lee, S. H., Kloosterboer, H. J., Blair, I. A., & Penning, T. M. (2009). Human cytosolic hydroxysteroid dehydrogenases of the aldo-ketoreductase superfamily catalyze reduction of conjugated steroids. *Journal of Biological Chemistry*, 284(15), 10013–10022.
- Kanehisa, M., Furumichi, M., Sato, Y., Ishiguro-Watanabe, M., & Tanabe, M. (2021). Kegg: integrating viruses and cellular organisms. *Nucleic acids research*, 49(D1), D545–D551.
- Kim, N., Kim, H. K., Lee, K., Hong, Y., Cho, J. H., Choi, J. W., ... others (2020). Single-cell rna sequencing demonstrates the molecular and cellular reprogramming of metastatic lung adenocarcinoma. *Nature communications*, 11(1), 1–15.
- Kuleshov, M. V., Jones, M. R., Rouillard, A. D., Fernandez, N. F., Duan, Q., Wang, Z., ... others (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic acids research*, 44(W1), W90–W97.

## References V

- Li, B., & Dewey, C. N. (2011). Rsem: accurate transcript quantification from rna-seq data with or without a reference genome. *BMC bioinformatics*, 12(1), 1–16.
- Li, B., Meng, Y.-Q., Li, Z., Yin, C., Lin, J.-P., Zhu, D.-J., & Zhang, S.-B. (2019). Mir-629-3p-induced downregulation of sftpc promotes cell proliferation and predicts poor survival in lung adenocarcinoma. *Artificial cells, nanomedicine, and biotechnology*, 47(1), 3286–3296.
- Lin, Z., Thorenoor, N., Wu, R., DiAngelo, S. L., Ye, M., Thomas, N. J., ... Floros, J. (2018). Genetic association of pulmonary surfactant protein genes, sftpa1, sftpa2, sftp, sftpc, and sftpd with cystic fibrosis. *Frontiers in immunology*, 9, 2256.
- Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for rna-seq data with deseq2. *Genome biology*, 15(12), 1–21.

## References VI

- Minna, J. D., Roth, J. A., & Gazdar, A. F. (2002). Focus on lung cancer. *Cancer cell*, 1(1), 49–52.
- Penning, T. M., Burczynski, M. E., Jez, J. M., Hung, C.-F., Lin, H.-K., Ma, H., ... RATNAM, K. (2000). Human  $3\alpha$ -hydroxysteroid dehydrogenase isoforms (akr1c1–akr1c4) of the aldo-keto reductase superfamily: functional plasticity and tissue distribution reveals roles in the inactivation and formation of male and female sex hormones. *Biochemical journal*, 351(1), 67–77.
- Roth, A., Khattra, J., Yap, D., Wan, A., Laks, E., Biele, J., ... Shah, S. P. (2014). Pyclone: statistical inference of clonal population structure in cancer. *Nature methods*, 11(4), 396–398.
- Soltis, A. R., Dalgard, C. L., Pollard, H. B., & Wilkerson, M. D. (2020). Mutenricher: a flexible toolset for somatic mutation enrichment analysis of tumor whole genomes. *BMC bioinformatics*, 21(1), 1–8.

## References VII

- Tate John, G., Sally, B., Jubb Harry, C., Zbyslaw, S., Beare David, M., Nidhi, B., ... Elisabeth, D. (2018). Stefancsik ray, thompson sam I, wang shicai, ward sari, campbell peter j, forbes simon a. cosmic: the catalogue of somatic mutations in cancer. *Nucleic Acids Research*, 47(D1), D941–D947.
- Travis, W. D. (2002). Pathology of lung cancer. *Clinics in chest medicine*, 23(1), 65–81.
- Uhrig, S., Ellermann, J., Walther, T., Burkhardt, P., Fröhlich, M., Hutter, B., ... others (2021). Accurate and efficient detection of gene fusions from rna sequencing data. *Genome research*, 31(3), 448–460.
- Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., ... others (2013). From fastq data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. *Current protocols in bioinformatics*, 43(1), 11–10.

## References VIII

- Vincent, R. G., Pickren, J. W., Lane, W. W., Bross, I., Takita, H., Houten, L., ... Rzepka, T. (1977). The changing histopathology of lung cancer. a review of 1682 cases. *Cancer*, 39(4), 1647–1655.
- Wang, B.-Y., Huang, J.-Y., Chen, H.-C., Lin, C.-H., Lin, S.-H., Hung, W.-H., & Cheng, Y.-F. (2020). The comparison between adenocarcinoma and squamous cell carcinoma in lung cancer patients. *Journal of cancer research and clinical oncology*, 146(1), 43–52.