

# 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*

2021-11-26

# 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)

# 1. Introduction

## 1.2. LUAD

# LUAD

# 1. Introduction

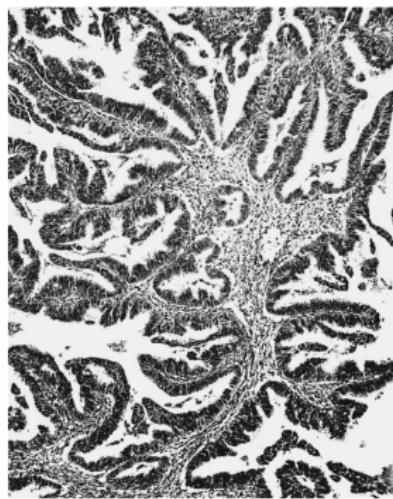
## 1.3. LUSC



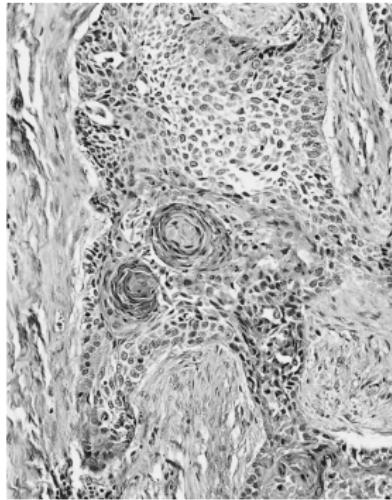
# 1. Introduction

## 1.4. LUAD vs. LUSC

# LUAD vs. LUSC I



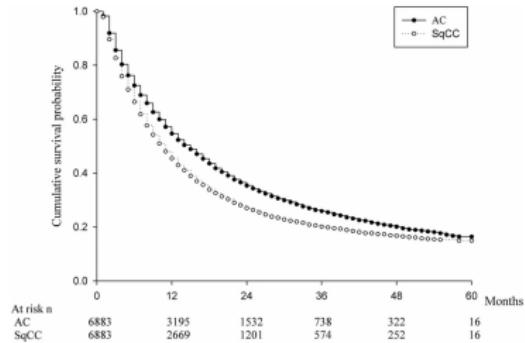
(a) LUAD



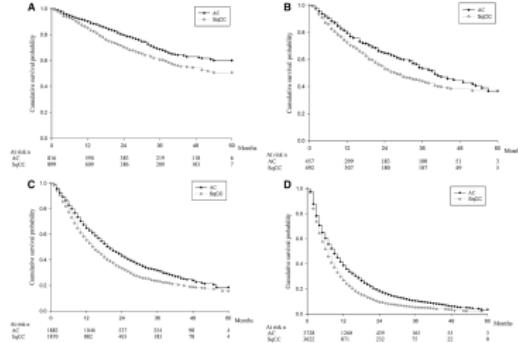
(b) LUSC

**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

## 1.5. 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

- Exome (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)

## 2. Materials

### 2.1. WES Data

# WES Data Composition

Table: Number of WES samples

Cancer Subtype	Stage	Number of Samples
LUSC	Normal	77
	Dysplasia	5
	AAH	8
	CIS+AIS	73
	Primary	77
	Total	240
LUAD	Normal	18
	AAH	15
	CIS+AIS	9
	MIA	1
	Primary	18
	Total	61

# WES Data Composition with Recurrence I

Table: LUSC WES Data with Recurrence

Recurrence?	Stage	Number of Samples	
		Normal	Dysplasia
Recurrence	Normal	14	
	Dysplasia		4
	CIS+AIS	12	
	Primary	14	
	Total	44	
Non-recurrence	Normal	63	
	Dysplasia		1
	AAH	8	
	CIS+AIS	61	
	Primary	63	
	Total	196	

# WES Data Composition with Recurrence II

Table: LUAD WES Data with Recurrence

Recurrence?	Number of Samples	
	Stage	
Recurrence	Normal	5
	AAH	8
	CIS+AIS	2
	Primary	5
	Total	20
Non-recurrence	Normal	13
	AAH	7
	CIS+AIS	7
	MIA	1
	Primary	13
	Total	41

# WES Data Composition with Smoking I

Table: LUSC WES Data with Smoking

Smoking?	Stage	Number of Samples	
		Normal	CIS+AIS
Never	Normal	3	3
	CIS+AIS	3	3
	Primary	3	3
	Total	9	9
Ex	Normal	41	41
	Dysplasia	1	1
	AAH	4	4
	CIS+AIS	40	40
	Primary	41	41
	Total	127	127
Current	Normal	33	33
	Dysplasia	4	4
	AAH	4	4
	CIS+AIS	30	30
	Primary	33	33
	Total	104	104

# WES Data Composition with Smoking II

Table: LUAD WES Data with Smoking

Smoking?	Stage	Number of Samples	
		Normal	Total
Never	Normal	1	
	CIS+AIS	1	
	Primary	1	
	Total	3	
Ex	Normal	10	
	AAH	9	
	CIS+AIS	6	
	Primary	10	
	Total	35	
Current	Normal	7	
	AAH	6	
	CIS+AIS	2	
	MIA	1	
	Primary	7	
	Total	23	

## 2. Materials

### 2.2. WTS Data

# WTS Data Composition

Table: Number of WTS samples

Cancer Subtype	Stage	Number of Samples	
		Normal	17
LUSC	Dysplasia		2
	CIS+AIS		34
	Primary		36
	Total		89
LUAD	Normal		13
	AAH		1
	CIS+AIS		5
	Primary		6
	Total		25

# WTS Data Composition with Recurrence I

Table: LUSC WTS Data with Recurrence

Recurrence?	Number of Samples	
	Stage	
Recurrence	Normal	1
	Dysplasia	1
	CIS+AIS	5
	Primary	6
	Total	13
Non-recurrence	Normal	16
	Dysplasia	1
	CIS+AIS	29
	Primary	30
	Total	76

# WTS Data Composition with Recurrence II

Table: LUAD WTS Data with Recurrence

Recurrence?	Stage	Number of Samples	
		Normal	Total
Recurrence	Normal	2	2
	CIS+AIS	1	1
	Primary	1	1
	Total	4	4
Non-recurrence	Normal	11	11
	AAH	1	1
	CIS+AIS	4	4
	Primary	5	5
	Total	21	21

# WTS Data Composition with Smoking I

Table: LUSC WTS Data with Smoking

Smoking?	Stage	Number of Samples	
		Normal	Total
Never	Normal	1	1
	CIS+AIS	1	1
	Primary	2	2
	Total	4	4
Ex	Normal	8	8
	Dysplasia	1	1
	CIS+AIS	21	21
	Primary	22	22
	Total	52	52
Current	Normal	8	8
	Dysplasia	1	1
	CIS+AIS	12	12
	Primary	12	12
	Total	33	33

# WTS Data Composition with Smoking II

Table: LUAD WTS Data with Smoking

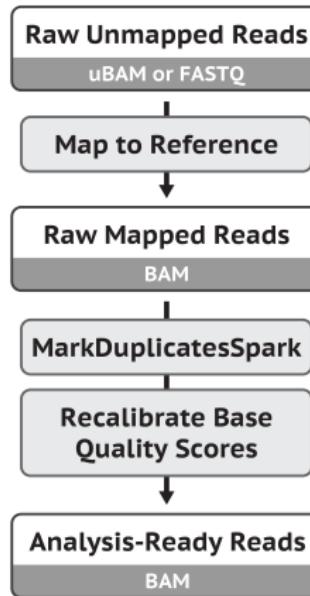
Smoking?	Stage	Number of Samples	
Never	Normal	10	
	AAH	1	
	CIS+AIS	3	
	Primary	4	
	Total	18	
Ex	Normal	3	
	CIS+AIS	1	
	Primary	1	
	Total	5	
Current	CIS+AIS	1	
	Primary	1	
	Total	2	

### 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

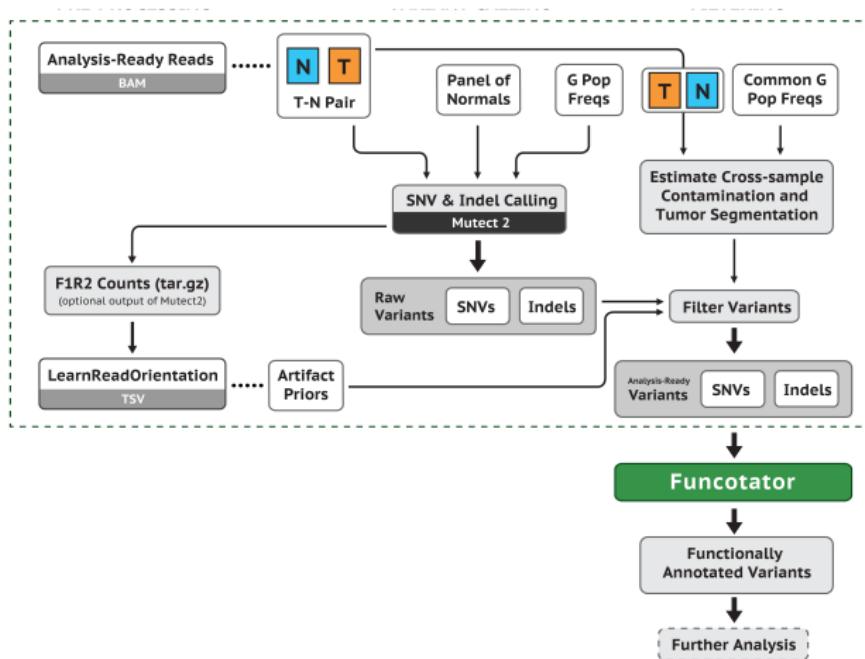
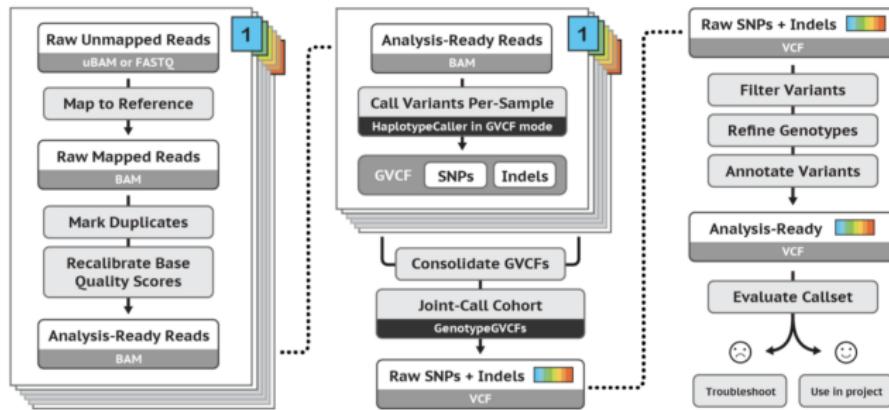


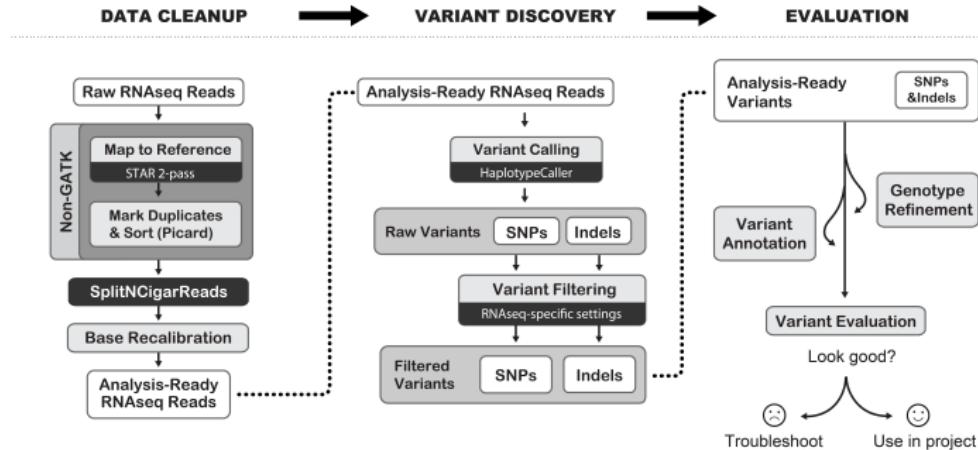
Figure: Somatic short variant (SNVs + Indels) discovery workflow (Van der Auwera et al., 2013; DePristo et al., 2011)

# 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 with FastQC

# FastQC?

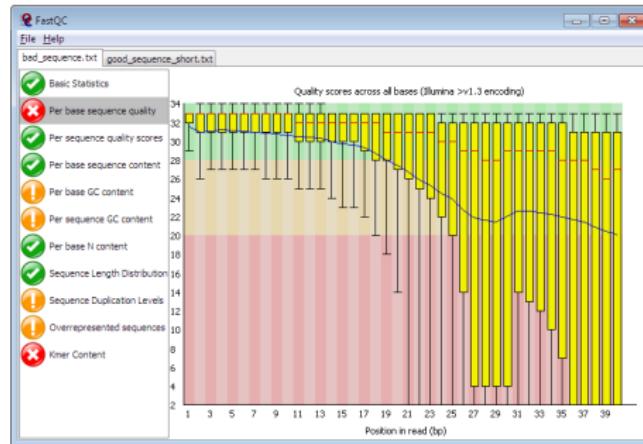


Figure: Example of FastQC Result (Andrews et al., 2012)

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

# FastQC on WES

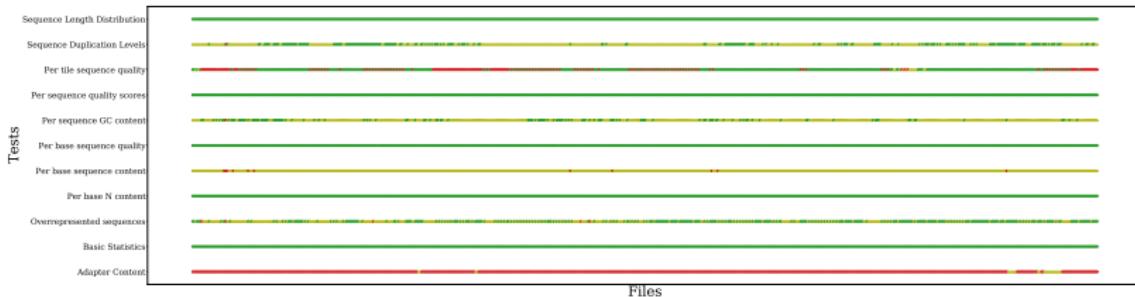
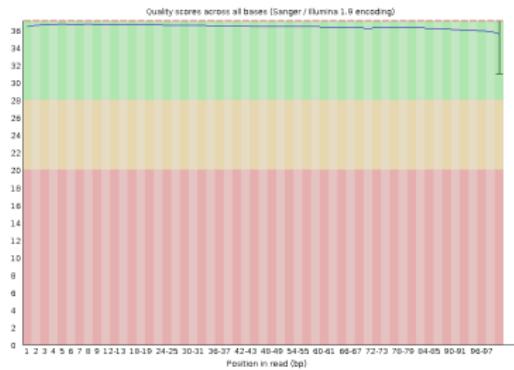


Figure: FastQC with WES data

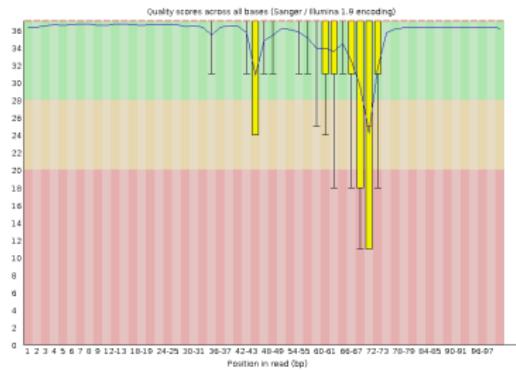
Failure on 33P1 sample

33P1 is excluded at further analysis.

# Failure on 33P1 I



(a) 33N



(b) 33P1

Figure: Per Base Sequence Quality Results

# Failure on 33P1 II

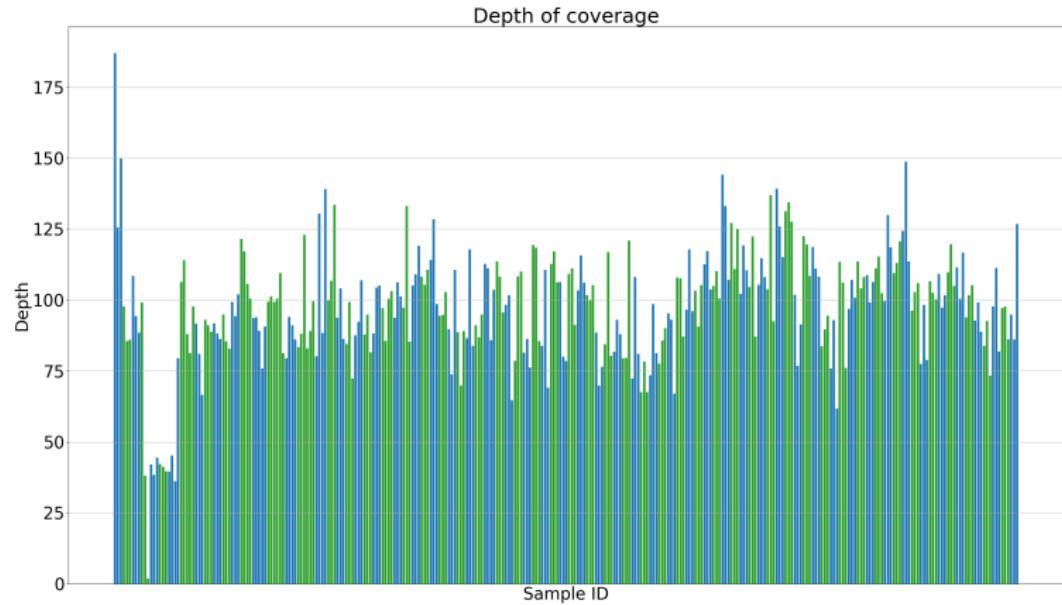


Figure: Coverage Depth Plot

# FastQC on WTS

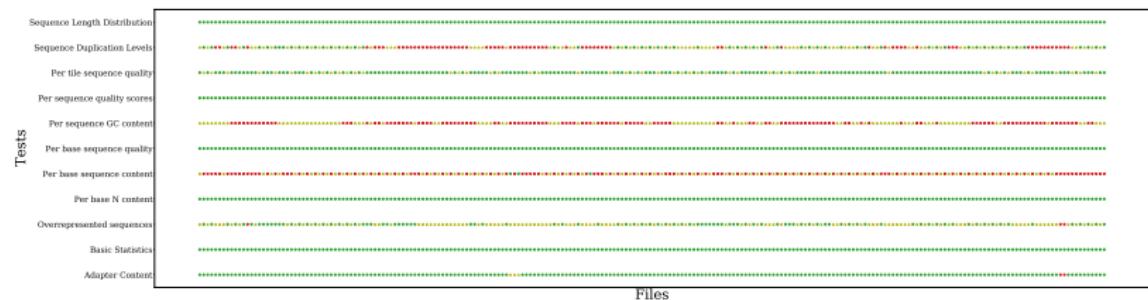


Figure: FastQC with WTS data

All sample are good to analysis

∴ No sample has more than 5 failures.

## 4. Results

### 4.2. Quality Checks with Picard

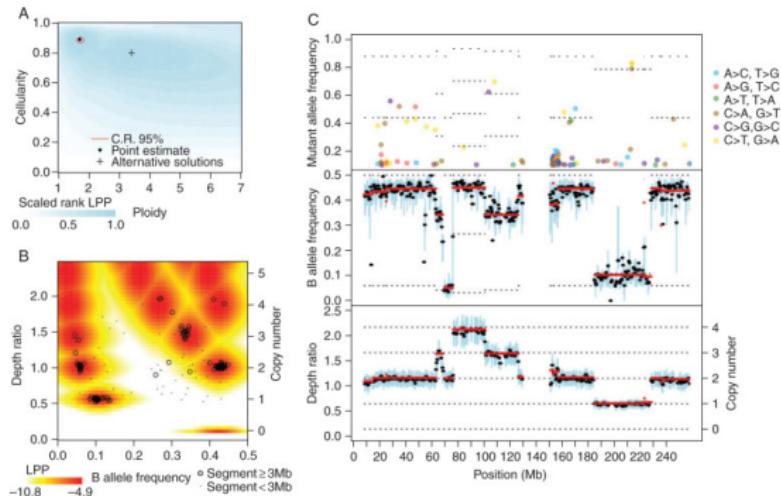
# Picard?

# Findings in Picard

## 4. Results

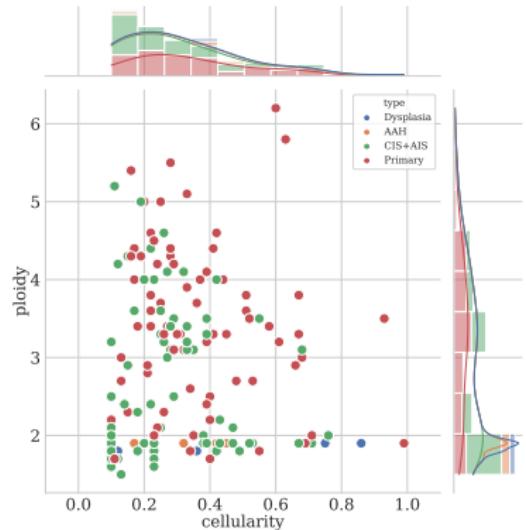
### 4.3. Copy Number Variations

# Sequenza?

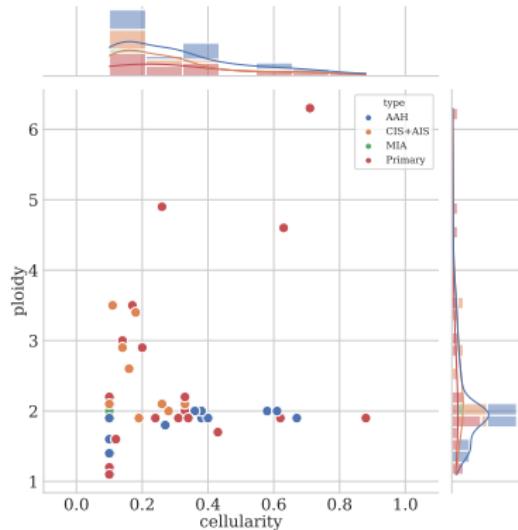


**Figure:** Representative Output of the Sequenza (Favero et al., 2015)

# Cellularity & Ploidy on WES



(a) LUSC Samples



(b) LUAD Samples

Figure: Cellularity and Ploidy from Sequenza

# LUSC in CNV Plot

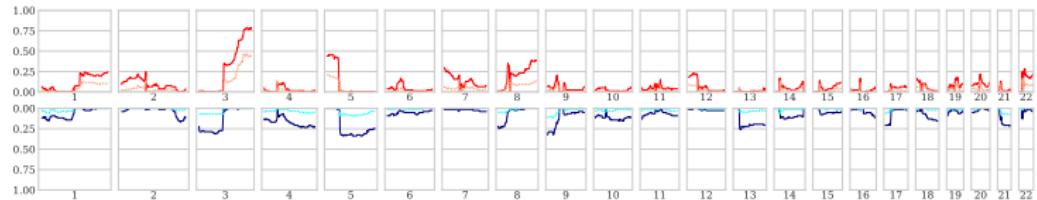


Figure: LUSC in CNV Plot

# LUSC with Recurrence in CNV Plot

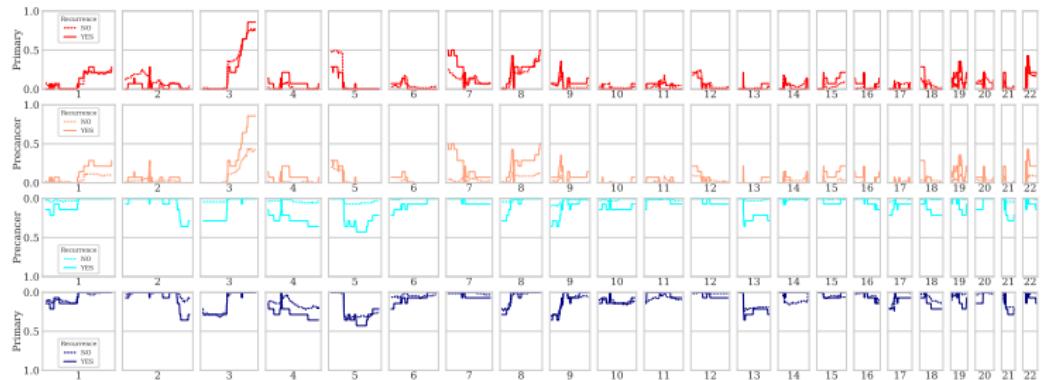


Figure: LUSC with Recurrence in CNV Plot

# LUAD in CNV Plot

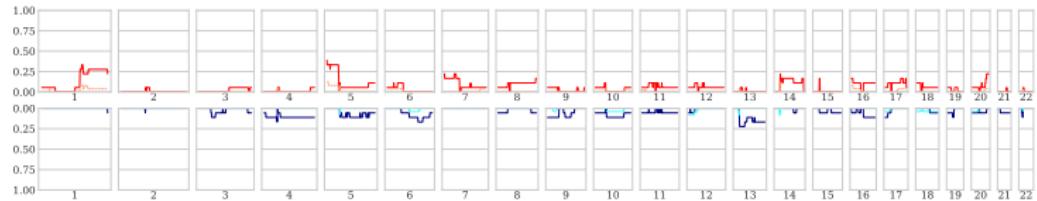


Figure: LUAD in CNV Plot

# LUAD with Recurrence in CNV Plot

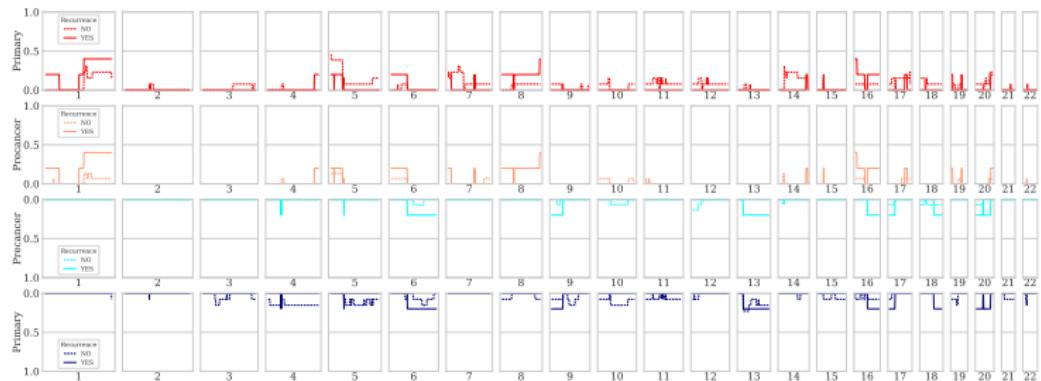


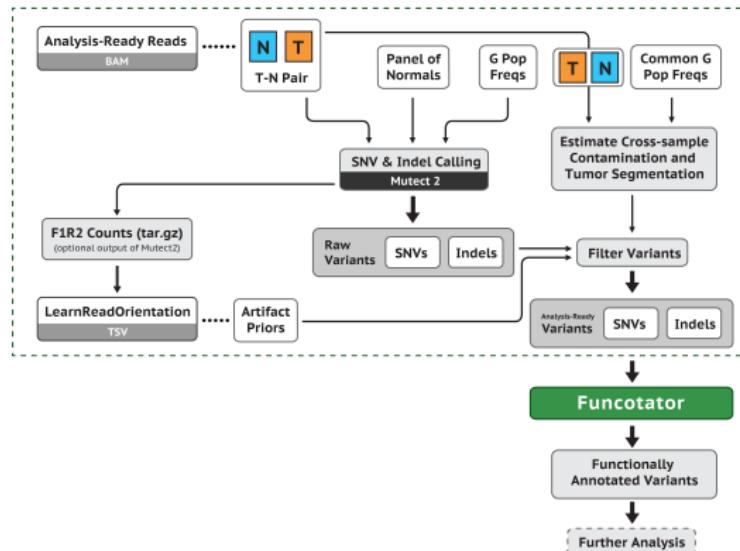
Figure: LUAD with Recurrence in CNV Plot

# Findings in Sequenza

## 4. Results

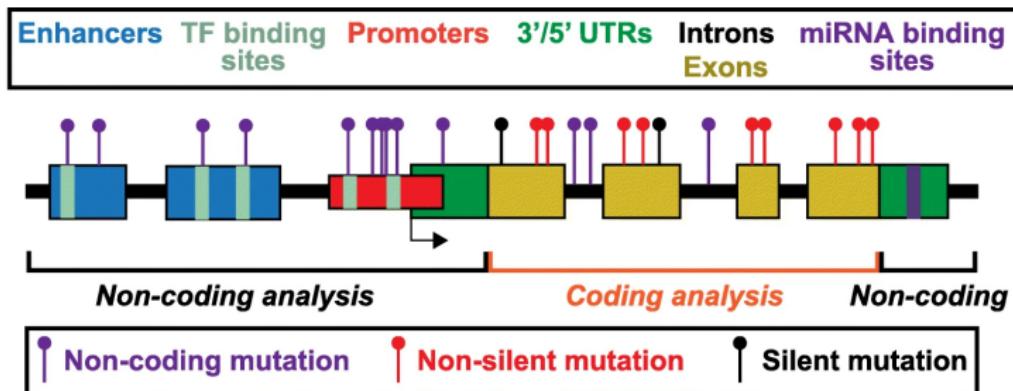
### 4.4. SNVs Analysis

# 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)

# CoMut?

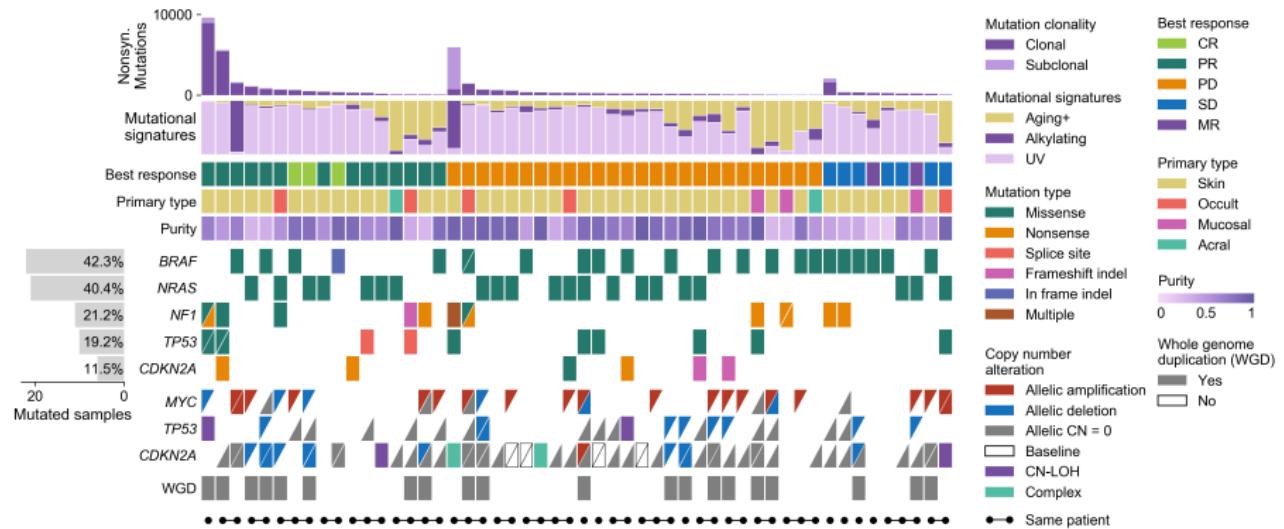


Figure: A comutation plot generated with CoMut (Crowdis et al., 2020)

# 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

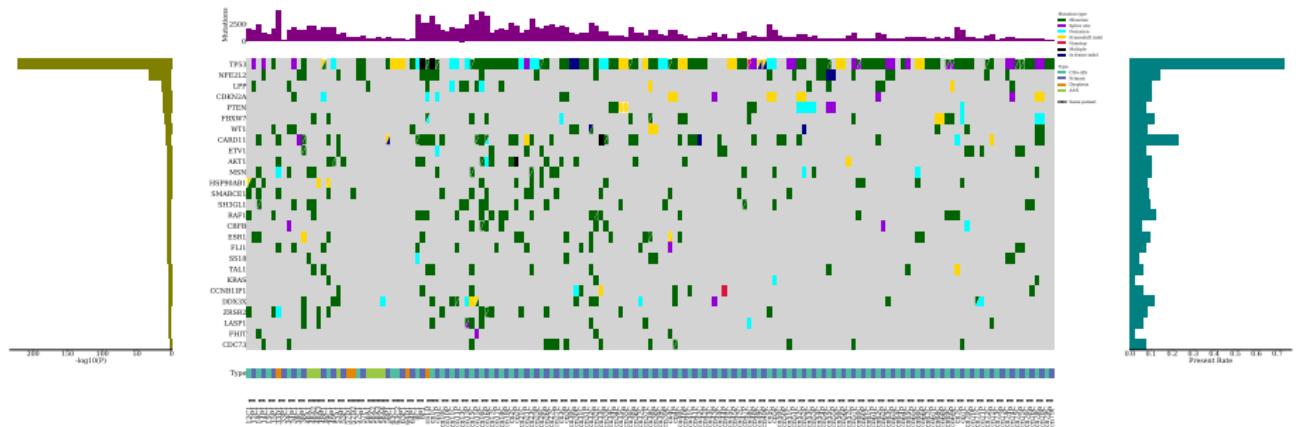


Figure: CoMut Plot with LUSC Patients

# Somatic Variant in LUSC with Recurrence

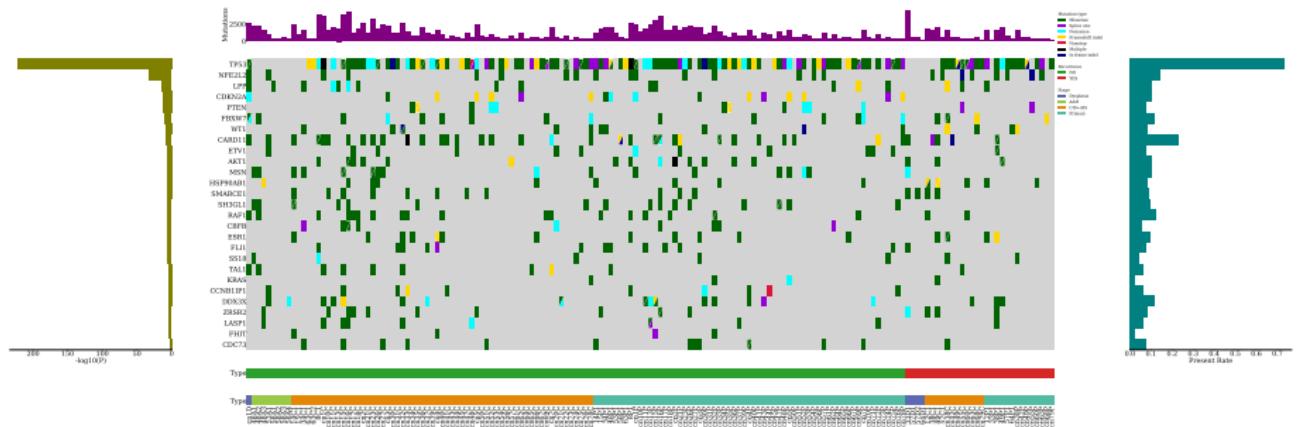


Figure: CoMut Plot in LUSC Patients with Recurrence

# Somatic Variant in LUAD

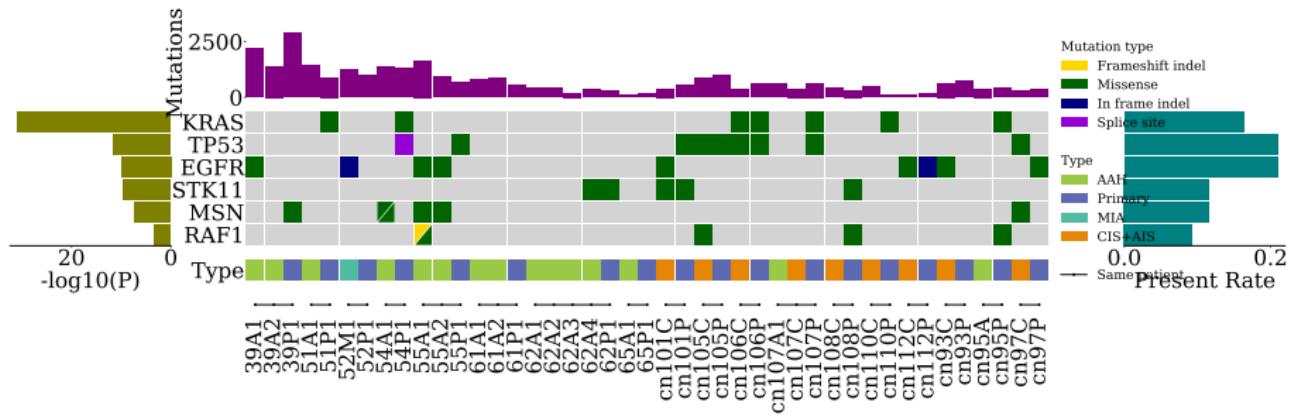


Figure: CoMut Plot with LUAD Patients

# Somatic Variant in LUAD with Recurrence

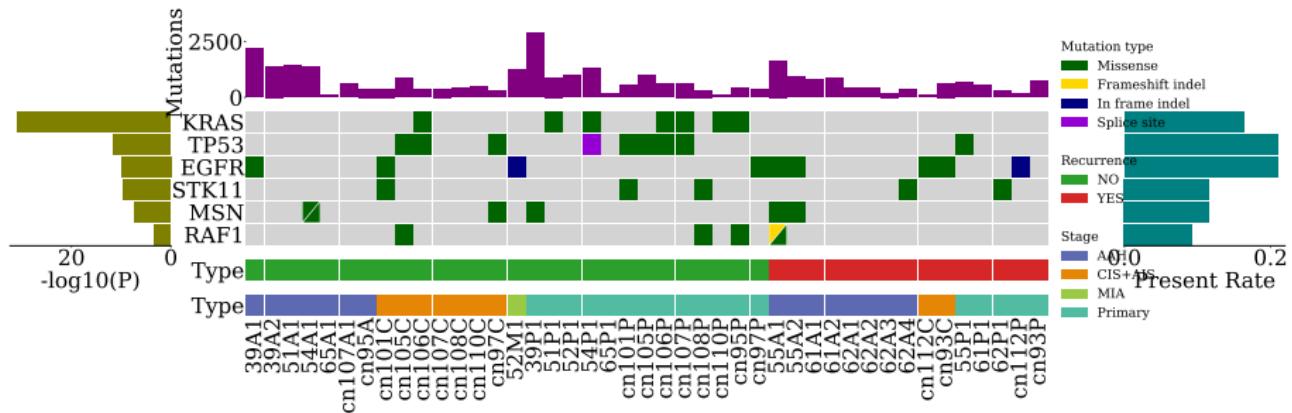


Figure: CoMut Plot in LUAD Patients with Recurrence

# Findings in SNVs Analysis

## 4. Results

### 4.5. VAF Analysis

# VAF?

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

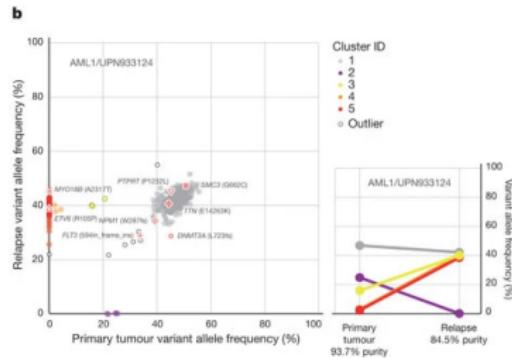


Figure: VAF distribution of validated mutations (Ding et al., 2012)

# VAF Plots I

# PyClone?

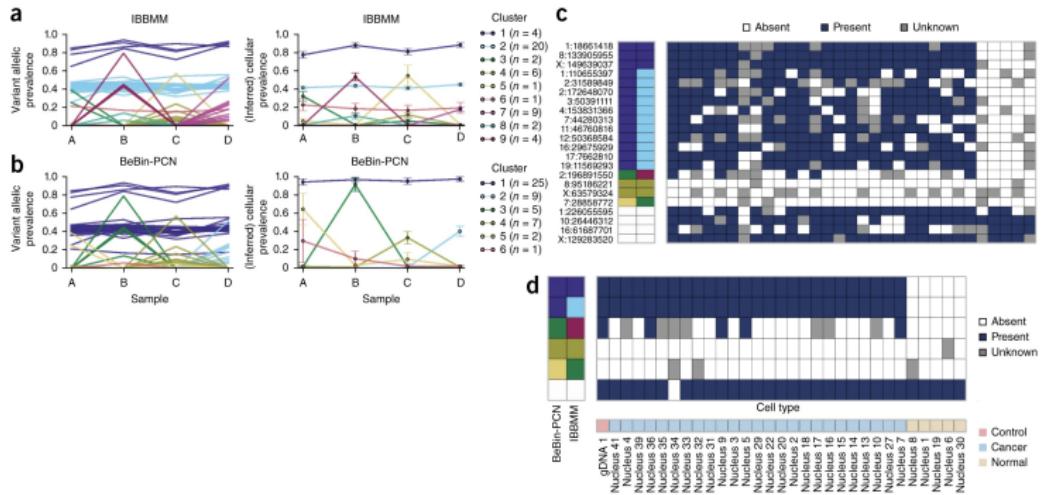


Figure: Analysis of multiple samples by PyClone (Roth et al., 2014)

# PyClone Plots I

# Findings in VAF Analysis

## 4. Results

### 4.6. Tumor Evolution Trajectories Analysis

# Revolver?

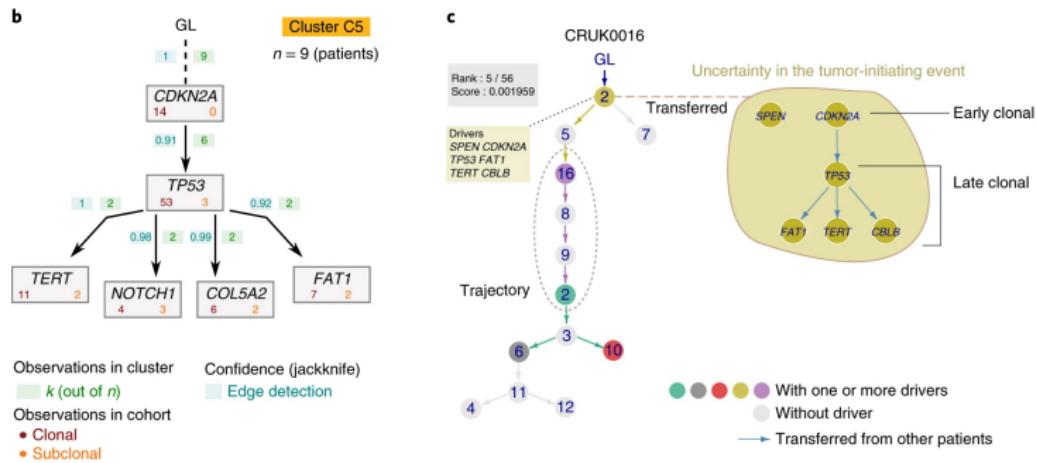


Figure: Repeated Evolutionary Trajectories (Caravagna et al., 2018)

# Findings in Tumor Evolution Trajectories Analysis

## 4. Results

### 4.7. Differences in Gene Expression Levels

# RSEM?

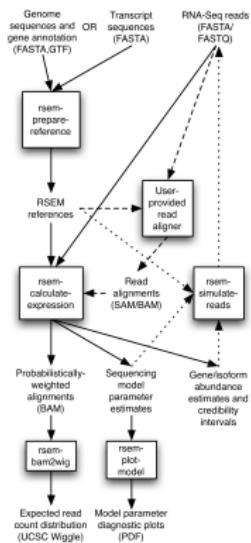


Figure: RSEM workflow (Li & Dewey, 2011)

# DESeq2?

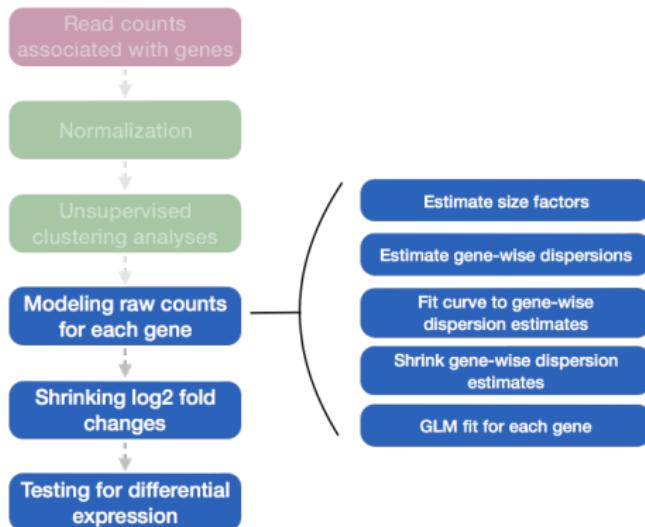


Figure: DESeq2 workflow (Love, Huber, & Anders, 2014)

# 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?

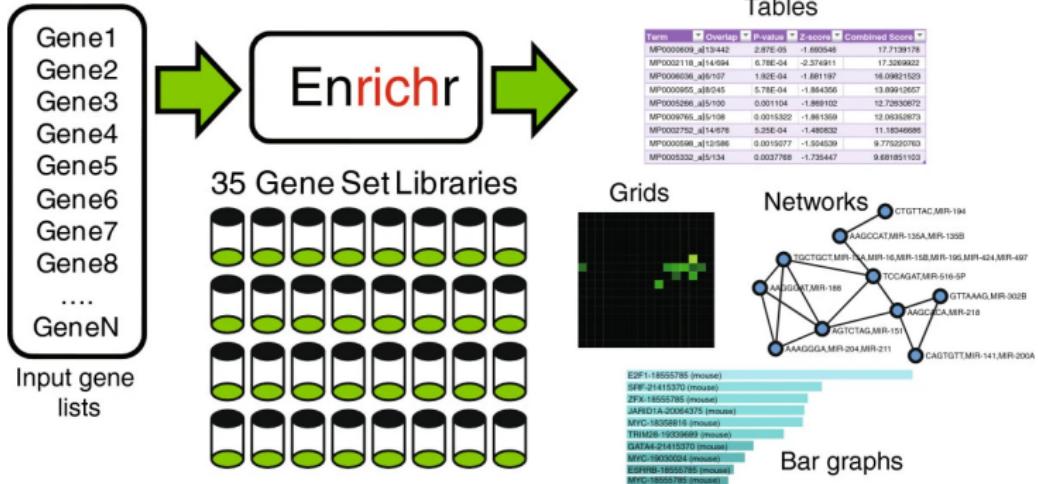


Figure: Enrichr workflow (Chen et al., 2013; Kuleshov et al., 2016)

# Gene-set Library

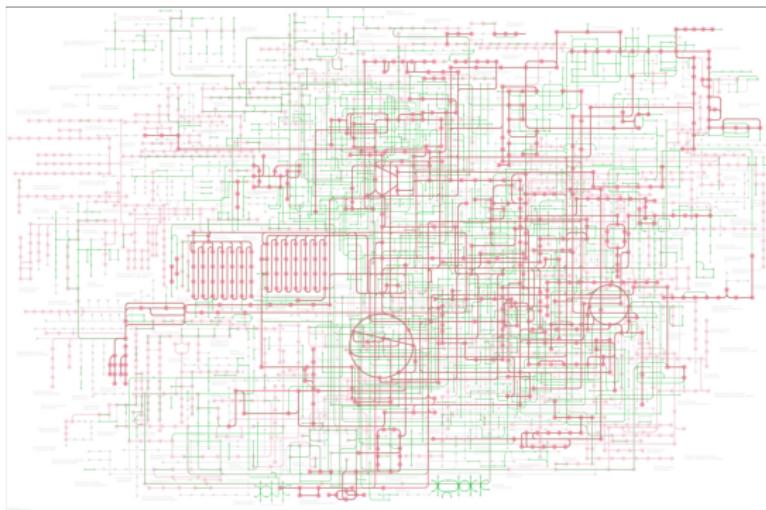


Figure: The global map of metabolic pathways by KEGG (Kanehisa et al., 2021)

KEGG

KEGG 2021 Human

# WTS Data Composition

Table: Number of WTS samples

Cancer Subtype	Stage	Number of Samples	
		Normal	17
LUSC	Dysplasia		2
	CIS+AIS		34
	Primary		36
	Total		89
LUAD	Normal		13
	AAH		1
	CIS+AIS		5
	Primary		6
	Total		25

# WTS Data Composition by Recur |

Table: Number of WTS LUSC samples

Recurrence?	Number of Samples	
	Stage	
Recurrence	Normal	1
	Dysplasia	1
	CIS+AIS	5
	Primary	6
	Total	13
Non-recurrence	Normal	16
	Dysplasia	1
	CIS+AIS	29
	Primary	30
	Total	76

# WTS Data Composition by Recur II

Table: Number of WTS LUAD samples

Recurrence?	Stage	Number of Samples	
		Normal	CIS+AIS
Recurrence	Normal	2	
	CIS+AIS		1
	Primary		1
	Total	4	
Non-recurrence	Normal	11	
	AAH		1
	CIS+AIS		4
	Primary		5
	Total	21	

# WTS Data Composition by Smoking I

Table: Number of WTS LUSC samples

Smoking?	Number of Samples	
	Stage	
Never	Normal	1
	CIS+AIS	1
	Primary	2
	Total	4
Ex	Normal	8
	Dysplasia	1
	CIS+AIS	21
	Primary	22
	Total	52
Current	Normal	8
	Dysplasia	1
	CIS+AIS	12

## WTS Data Composition by Smoking II

Table: Number of WTS LUAD samples

Smoking?	Stage	Number of Samples	
Never	Normal	10	
	AAH	1	
	CIS+AIS	3	
	Primary	4	
	Total	18	
Ex	Normal	3	
	CIS+AIS	1	
	Primary	1	
	Total	5	
Current	CIS+AIS	1	
	Primary	1	
	Total	2	

## 4. Results

### 4.7. Differences in Gene Expression Levels

#### 4.7.1. Comparing cancer stage in LUSC

# DEG List in LUSC

Table: Up-regulated DEG in LUSC

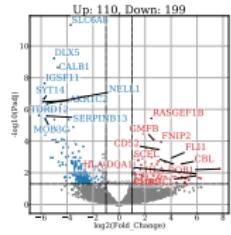
gene	log2FoldChange	pvalue	padj
AKR1C1	6.18e+00	5.14e-26	5.01e-23
AKR1C2	6.06e+00	1.19e-22	5.04e-20
CYP4F11	5.58e+00	1.51e-20	4.36e-18

Table: Down-regulated DEG in LUSC

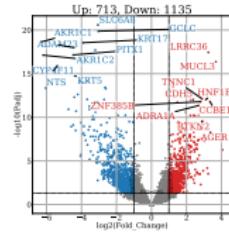
gene	log2FoldChange	pvalue	padj
SFTPC	-5.85e+00	9.16e-21	2.83e-18
FAM107A	-4.62e+00	2.27e-33	9.60e-30
LRRC36	-4.53e+00	5.49e-36	3.48e-32

## DEG Volcano Plots in LUSC

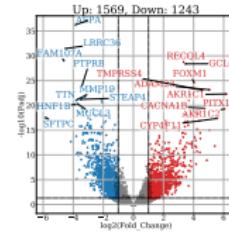
Normal → Dysplasia → CIS → Primary (LUSC)



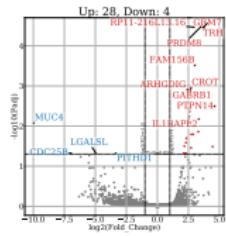
(a) Normal-Dysplasia



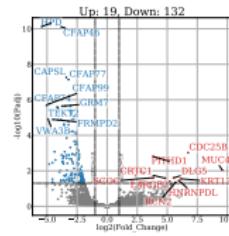
(b) Normal-CIS



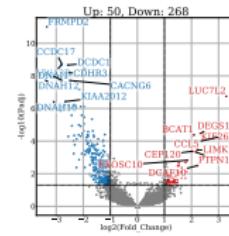
(c) Normal-Primary



(d) Dysplasia-CIS



(e) Dysplasia-Primary

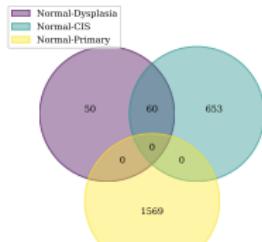


(f) CIS-Primary

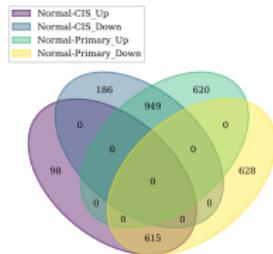
## Figure: DEG Volcano Plots in LUSC

# DEG Venn Diagram in LUSC

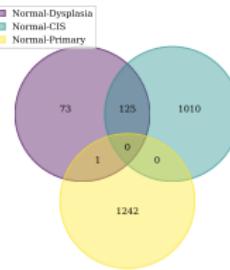
Normal → Dysplasia → CIS → Primary (LUSC)



(a) Up-regulated



(b) Both



(c) 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	ASAHI,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

### 4.7. Differences in Gene Expression Levels

#### 4.7.2. Comparing cancer stage in LUAD

# DEG List in LUSC

Table: Up-regulated DEG in LUAD

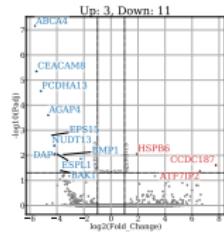
gene	log2FoldChange	pvalue	padj
ABCA4	4.95e+00	3.01e-12	2.58e-09
HMGA2	4.79e+00	8.06e-08	1.46e-05
KIF12	4.48e+00	1.33e-06	1.46e-04

Table: Down-regulated DEG in LUAD

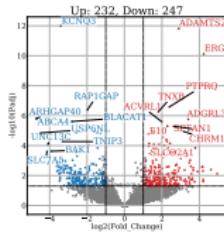
gene	log2FoldChange	pvalue	padj
SLC6A4	-6.20e+00	5.80e-10	2.36e-07
IL1RL1	-4.20e+00	7.47e-06	5.82e-04
RNF185	-4.06e+00	4.75e-05	2.45e-03

## DEG Volcano Plots in LUAD

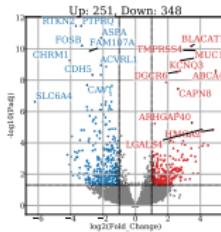
Normal → AAH → AIS → Primary (LUAD)



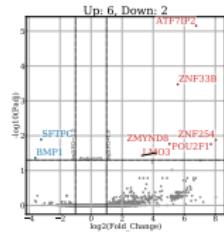
(a) Normal-AAH



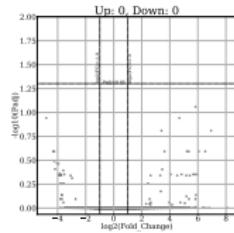
(b) Normal-AIS



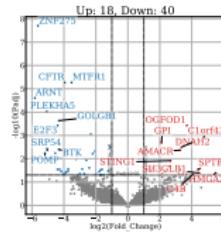
(c) Normal-Primary



(d) AAH-AIS



(e) AAH-Primary

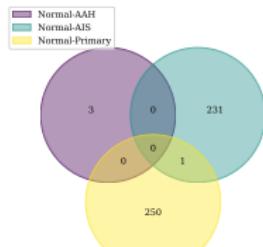


(f) AIS-Primary

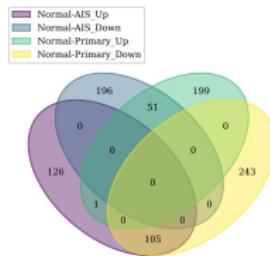
## Figure: DEG Volcano Plots in LUAD

# DEG Venn Diagram in LUAD

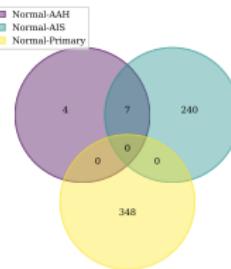
Normal → AAH → AIS → Primary (LUAD)



(a) Up-regulated



(b) Both



(c) 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

## ABCA4

- ① Down-regulated in AAH & AIS, but up-regulated in Primary.
- ② It is associated with ophthalmology (Maugeri et al., 2000).
- ③ It shows lung cancer susceptibility in Korean patients (Lee, Lee, Yoon, & Lee, 2013).

# Finding in Comparing cancer stage in LUAD II

## KCNQ3

- ① Down-regulated in AIS, but up-regulated in Primary.
- ②  $K^+$  voltage-dependent channels  $\Rightarrow$  Various physiological functions (Schroeder, Kubisch, Stein, & Jentsch, 1998; Surti, Huang, Jan, Jan, & Cooper, 2005; Singh et al., 2003).
- ③ Up-regulated microRNAs in hypoxia-induced LUAD (Geng et al., 2016).
- ④ KCNQ gene family is associated with lung disease (Mondejar-Parreño, Perez-Vizcaino, & Cogolludo, 2020).

## CHRM1

- ① Up-regulated in AIS, but down-regulated in Primary.
- ② Various cellular responses ⇒ neurodevelopmental disorders (Marcé-Grau et al., 2021), schizophrenia (Dean & Scarr, 2021), and Alzheimer's disease (Counts et al., 2007).
- ③ Reported down-regulation in LUSC & LUAD (Ma et al., 2019).

## 4. Results

### 4.7. Differences in Gene Expression Levels

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

Table: Number of WTS LUSC samples

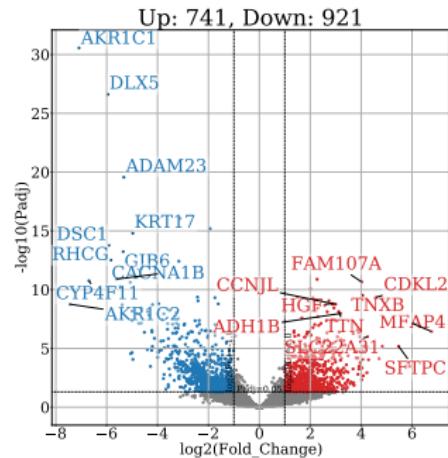
Recurrence?	Stage	Number of Samples	
		Normal	Total
Recurrence	Normal	1	1
	Dysplasia	1	1
	CIS+AIS	5	5
	Primary	6	6
	Total	13	13
Non-recurrence	Normal	16	16
	Dysplasia	1	1
	CIS+AIS	29	29
	Primary	30	30
	Total	76	76

## Pooled normal samples

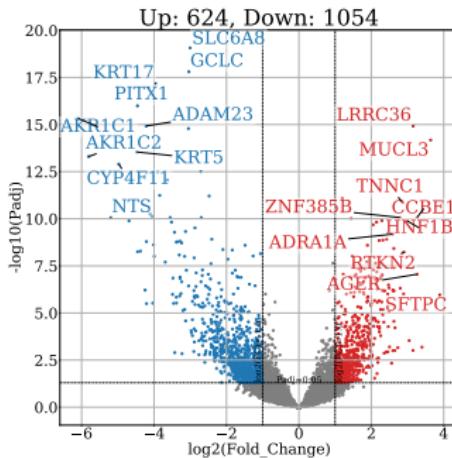
In order to compare with Normal stage, merging Normal samples.

∴ Insufficient number of Normal samples in Recur.

# DEG Volcano Plots for R vs. NR with CIS in LUSC



(a) Recur



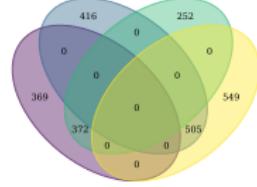
(b) Non-recur

Figure: DEG Volcanot Plot with CIS in LUSC

# DEG Venn Diagram for R vs. NR with CIS in LUSC



(a) Up-regulated



(b) Both



(c) Down-regulated

Figure: DEG Venn Diagram for R vs. NR with CIS in LUSC

# Enrichment test for Recur-specific with CIS in LUSC

Table: Up-regulated Pathways for Recur-specific

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

Table: Down-regulated Pathways for Recur-specific

Term name	Overlapping genes...	Adjusted p-value
Huntington disease	COX8A,DCTN5,COX7B,...(24)	6.36e-06
Amyotrophic lateral sclerosis	DCTN5,COX7B,TOMM40,...(25)	1.62e-05
Parkinson disease	COX8A,COX7B,NDUFA12,...(20)	1.62e-05

# Enrichment test for Non-recur-specific with CIS in LUSC

Table: Up-regulated Pathways for Non-recur-specific

Term name	Overlapping genes...	Adjusted p-value
Malaria	IL6,ITGB2,KLRC4-KLRK1,...(6)	7.76e-03
Th1 and Th2 cell differentiation	STAT5B,MAML2,MAML3,...(7)	1.15e-02
Transcriptional misregulation in cancer	PTCRA,CSF1R,IL6,...(10)	1.15e-02

Table: Down-regulated Pathways for Non-recur-specific

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

# Enrichment test for Intersected with CIS in LUSC

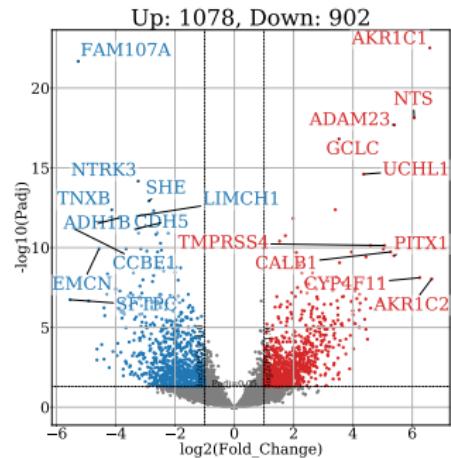
Table: Up-regulated Pathways for Intersected

Term name	Overlapping genes...	Adjusted p-value
Hypertrophic cardiomyopathy	EDN1,CACNB4,ACE,...(12)	1.44e-05
Cell adhesion molecules	CADM1,ICAM2,SELP,...(15)	1.44e-05
Hematopoietic cell lineage	CSF3,HLA-DMA,MME,...(12)	2.57e-05

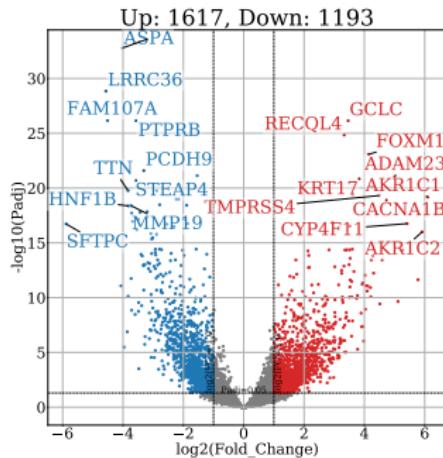
Table: Down-regulated Pathways for Intersected

Term name	Overlapping genes...	Adjusted p-value
Metabolism of xenobiotics by cytochrome P450	GSTM4,CBR1,GSTM3,...(14)	1.53e-06
Drug metabolism	GSTM4,GSTM3,RRM1,...(16)	1.69e-06
Steroid hormone biosynthesis	UGT1A1,SRD5A1,HSD17B1,...(11)	2.77e-05

# DEG Volcano Plots for R vs. NR with Primary in LUSC



(a) Recur



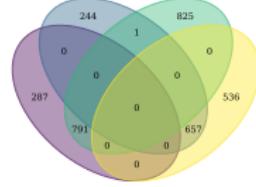
(b) Non-recur

Figure: DEG Volcanot Plot with Primary in LUSC

# DEG Venn Diagram for R vs. NR with Primary in LUSC



(a) Up-regulated



(b) Both



(c) Down-regulated

Figure: DEG Venn Diagram for R vs. NR with Primary in LUSC

# Enrichment test for Recur-specific with Primary in LUSC

Table: Up-regulated Pathways for Recur-specific

Term name	Overlapping genes...	Adjusted p-value
Amyotrophic lateral sclerosis	COX7B,HSPA5,NDUFA1,...(17)	4.85e-03
RNA transport	NUP93,EIF5,EIF1AY,...(11)	6.11e-03
mRNA surveillance pathway	PABPN1,PPP2R1A,NCBP2,...(8)	6.11e-03

Table: Down-regulated Pathways for Recur-specific

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

# Enrichment test for NR-specific with Primary in LUSC

Table: Up-regulated Pathways for Non-recur-specific

Term name	Overlapping genes...	Adjusted p-value
Homologous recombination	RAD51D,POLD1,RPA3,...(9)	1.00e-02

Table: Down-regulated Pathways for Non-recur-specific

Term name	Overlapping genes...	Adjusted p-value
Staphylococcus aureus infection	IL10,CFD,ITGB2,...(14)	5.37e-05
Hematopoietic cell lineage	CR1,MME,ITGB3,...(14)	5.37e-05
Leishmaniasis	IL10,C3,NFKBIA,...(11)	4.30e-04

# Enrichment test for Intersected with Primary in LUSC

Table: Up-regulated Pathways for Intersected

Term name	Overlapping genes...	Adjusted p-value
Glycolysis / Gluconeogenesis	GPI,TPI1,PDHA1,...(16)	1.09e-06
Drug metabolism	GSTM4,GSTM3,GSTM2,...(20)	1.09e-06
Metabolism of xenobiotics by cytochrome P450	GSTM4,CBR1,GSTM3,...(15)	2.27e-05

Table: Down-regulated Pathways for Intersected

Term name	Overlapping genes...	Adjusted p-value
Dilated cardiomyopathy	LAMA2,TNNC1,ADCY4,...(15)	1.44e-04
Hypertrophic cardiomyopathy	LAMA2,TNNC1,CACNA2D2,...(14)	1.81e-04
Malaria	SELP,CSF3,IL6,...(10)	2.91e-04

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

## NTS

- ① Highly up-regulated in Recur patients.
- ② Neurotensin.
- ③ Association with non-gastrointestinal cancers (Nikolaou et al., 2020).
- ④ Modulate lung cancer cell plasticity and heterogeneity (Wu et al., 2019).

## NTRK3

- ① Highly down-regulated in Recur patients.
- ② Activation of NTRK3 in LUSC (Bollig-Fischer et al., 2021).
- ③ NTRK3 mutation has association with immunotherapy in LUAD (Niu et al., 2020).

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

## RECQL4

- ① Highly up-regulated in Non-recur patients.

## ASPA

- ① Highly down-regulated in Non-recur patients.

## 4. Results

### 4.7. Differences in Gene Expression Levels

#### 4.7.4. 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

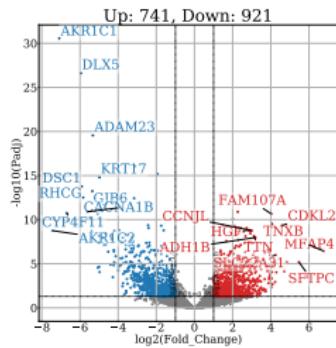
Table: Up-regulated DEG for Primary within Recur in LUSC

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

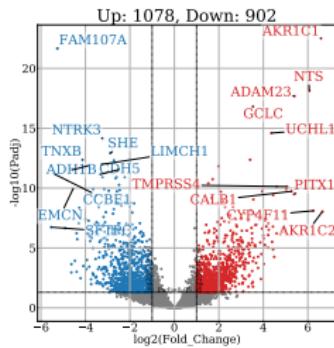
Table: Down-regulated DEG for Primary within Recur in LUSC

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

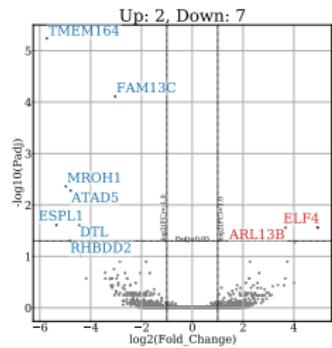
# DEG Volcano Plots with Recur in LUSC



(a) Normal-CIS



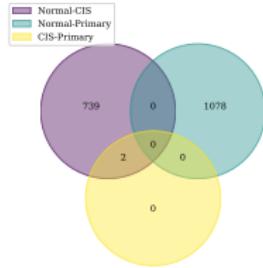
(b) Normal-Primary



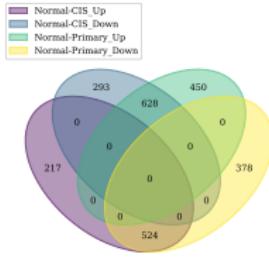
(c) CIS-Primary

Figure: DEG Volcano Plots with Recur samples in LUSC

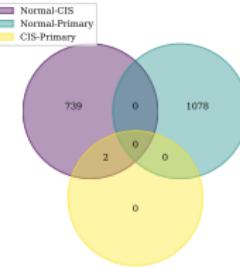
# DEG Venn Diagram with Recur in LUSC



(a) Up-regulated



(b) Both



(c) 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 I

## 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

### 4.7. Differences in Gene Expression Levels

#### 4.7.5. 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

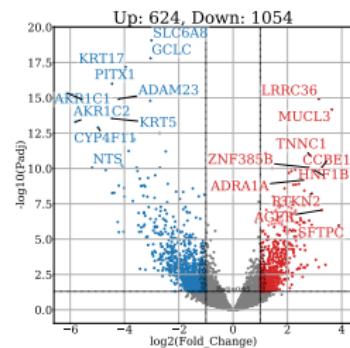
Table: Up-regulated DEG for Primary within Non-recr in LUSC

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

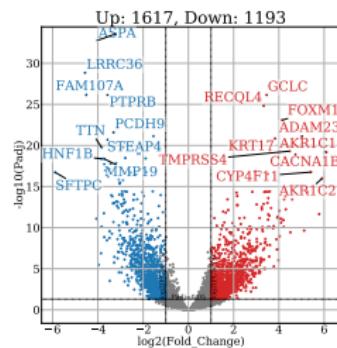
Table: Down-regulated DEG for Primary within Non-recr in LUSC

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

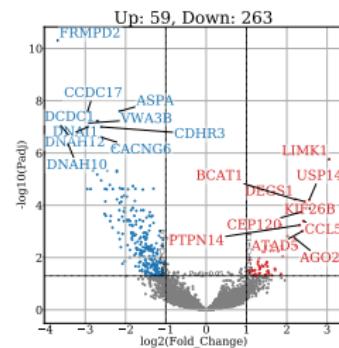
## DEG Volcano Plots with Non-recur in LUSC



(a) Normal-CIS



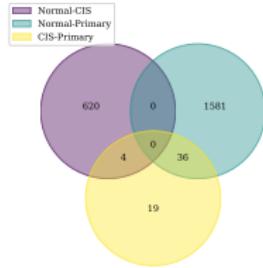
(b) Normal-Primary



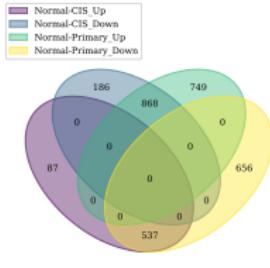
(c) CIS-Primary

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

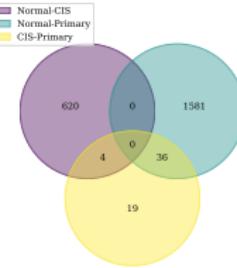
# DEG Venn Diagram with Non-recur in LUSC



(a) Up-regulated



(b) Both



(c) Down-regulated

Figure: DEG Venn Diagram with Non-recur 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 Non-recr in LUSC I

## 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

### 4.7. Differences in Gene Expression Levels

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

# LUAD Data Composition

Table: Number of WTS LUAD samples

Recurrence?	Stage	Number of Samples
Recurrence	Normal	2
	CIS+AIS	1
	Primary	1
	Total	4
Non-recurrence	Normal	11
	AAH	1
	CIS+AIS	4
	Primary	5
	Total	21

# DEG List for AIS within Non-recur in LUAD

Table: Up-regulated DEG for AIS within Non-recur in LUAD

gene	log2FoldChange	pvalue	padj
MUC4	4.83e+00	2.55e-04	1.68e-02
SIPA1	4.77e+00	4.87e-05	6.37e-03
C11orf45	4.68e+00	2.86e-04	1.85e-02

Table: Down-regulated DEG for AIS within Non-recur in LUAD

gene	log2FoldChange	pvalue	padj
ABCA4	-5.02e+00	2.44e-10	5.29e-07
UNC13C	-4.08e+00	6.49e-06	1.88e-03
SLC7A5	-3.93e+00	1.40e-06	6.76e-04

# DEG List for Primary within Non-recur in LUAD

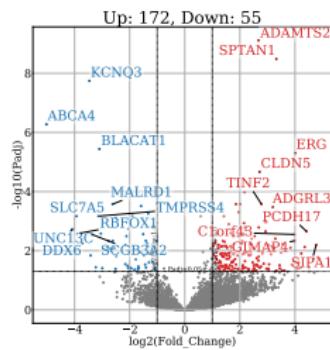
Table: Up-regulated DEG for Primary within Non-recur in LUAD

gene	log2FoldChange	pvalue	padj
ABCA4	5.22e+00	1.67e-11	3.32e-08
HMGA2	5.03e+00	4.39e-07	9.62e-05
KIF12	4.54e+00	2.62e-06	3.91e-04

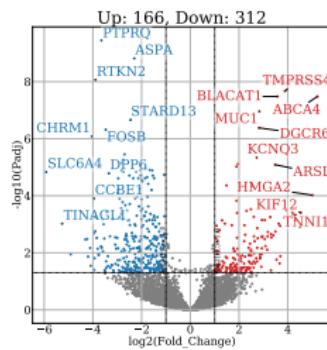
Table: Down-regulated DEG for Primary within Non-recur in LUAD

gene	log2FoldChange	pvalue	padj
SLC6A4	-5.92e+00	3.83e-08	1.47e-05
TINAGL1	-5.27e+00	9.47e-06	9.57e-04
SFTPA1	-4.91e+00	2.69e-04	1.13e-02

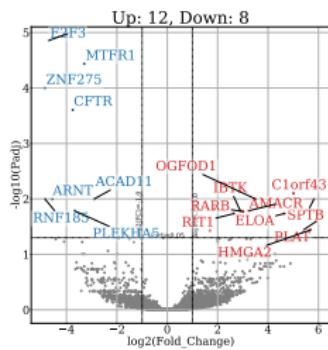
# DEG Volcano Plots with Non-recr in LUAD



(a) Normal-AIS



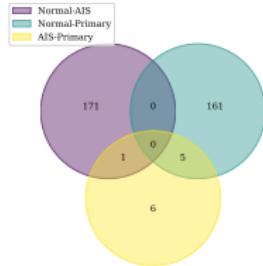
(b) Normal-Primary



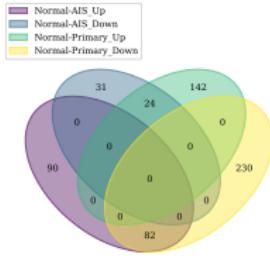
(c) AIS-Primary

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

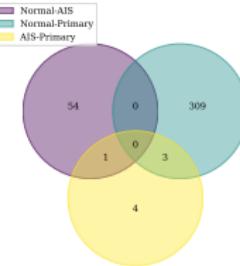
# DEG Venn Diagram with Non-recur in LUAD



(a) Up-regulated



(b) Both



(c) Down-regulated

Figure: DEG Venn Diagram with Non-recur 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

# Finding in Comparing within Non-recur in LUAD I

## KCNQ3

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

## BLACAT1

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

# Findings in DEG Analysis

## 4. Results

### 4.8. Bulk Cell Deconvolution

# BisqueRNA?

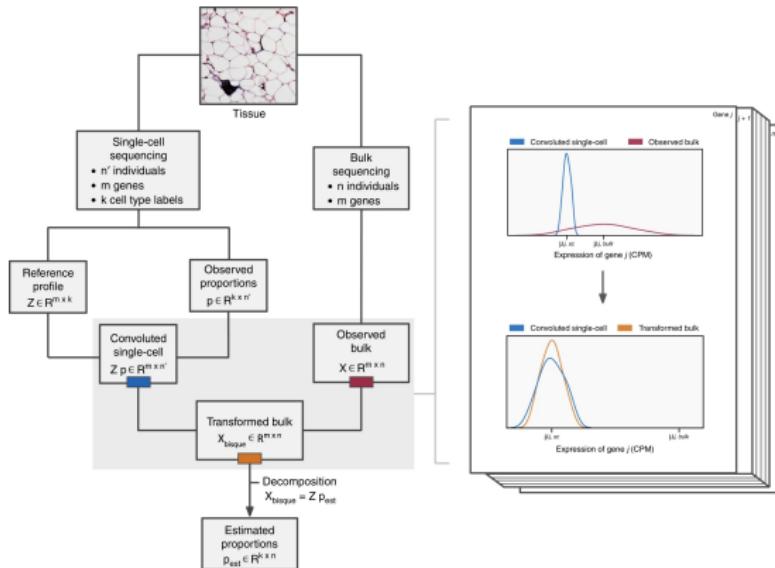


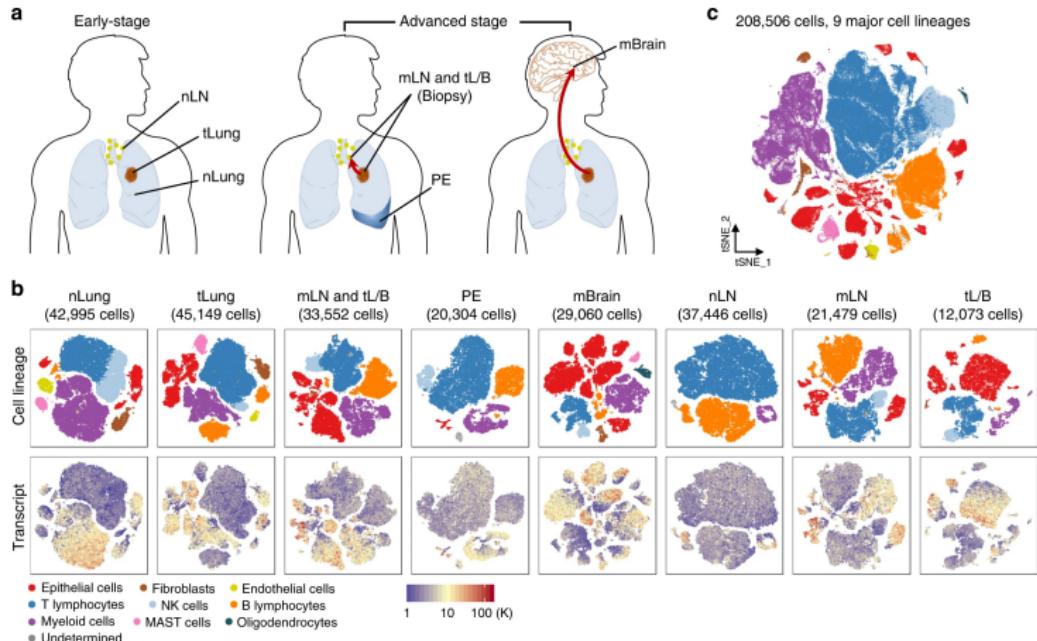
Figure: Workflow for BisqueRNA (Jew et al., 2020)

## 4. Results

### 4.8. Bulk Cell Deconvolution

#### 4.8.1. Reference by Kim et al. (2020)

# Reference Single-cell Data



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

# Cluster Plot in LUSC

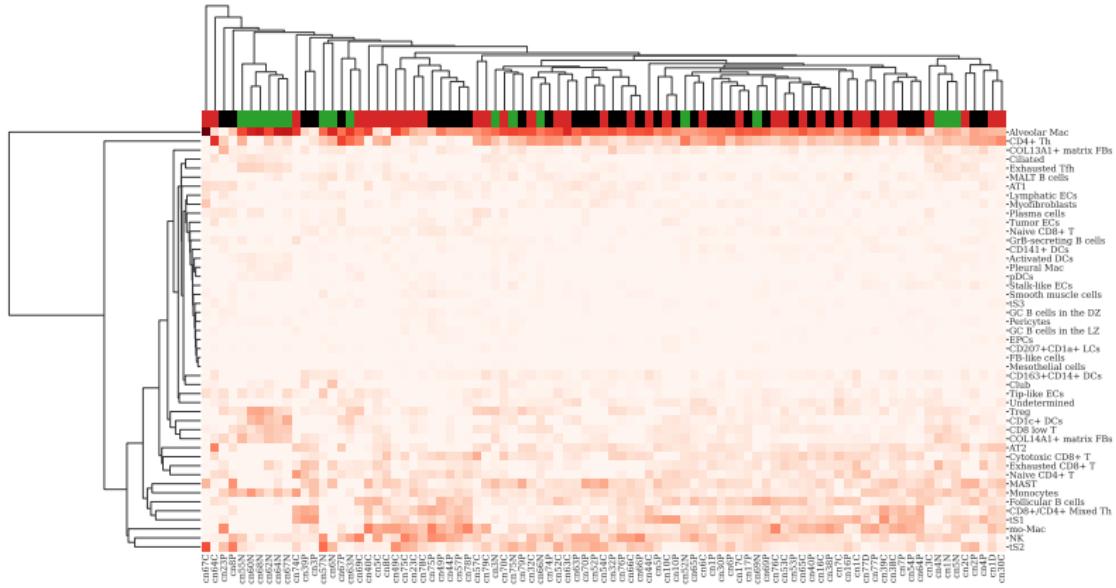
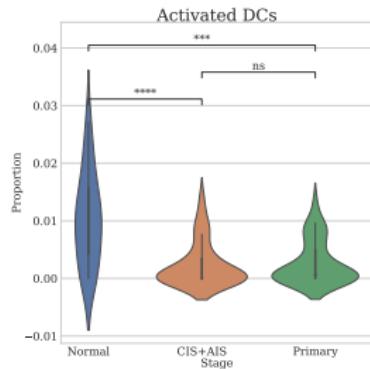
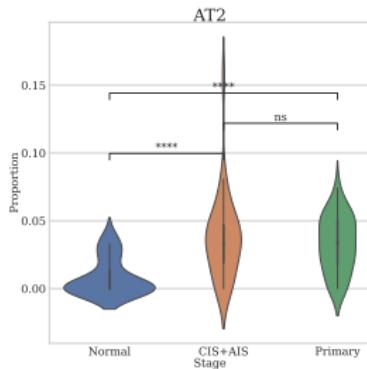


Figure: Cluster Plot in LUSC

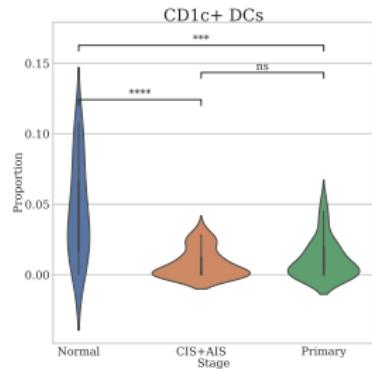
# Violin Plots in LUSC I



(a) Activated DCs



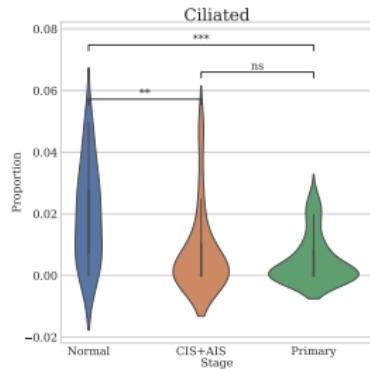
(b) Alveolar type II



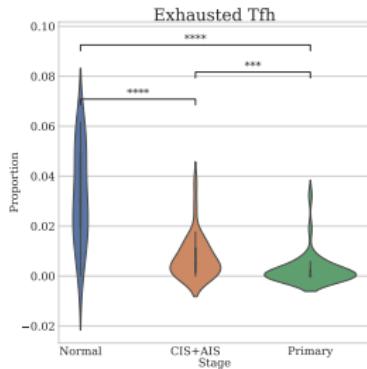
(c) Langerhans cells

Figure: Violin Plots in LUSC

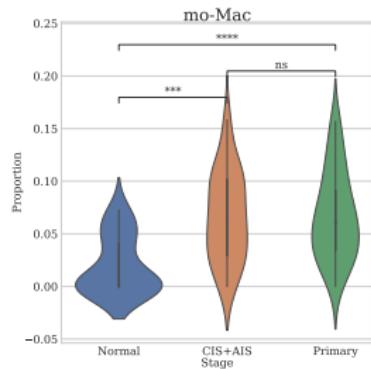
# Violin Plots in LUSC II



(d) Ciliated cells



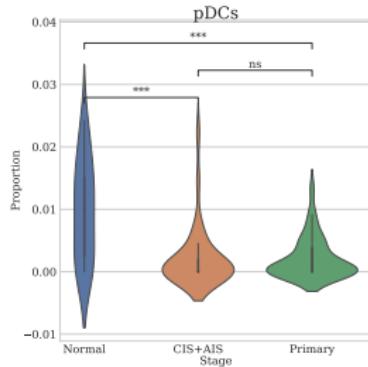
(e) Exhausted T follicular helper



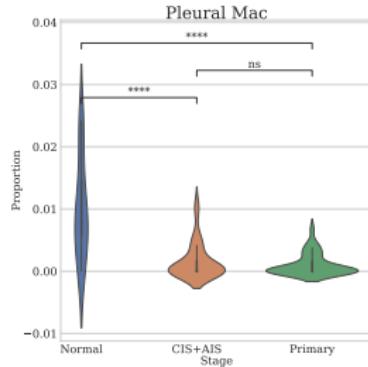
(f) Mo & Mac

Figure: Violin Plots in LUSC

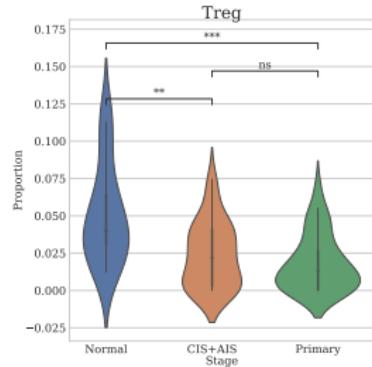
# Violin Plots in LUSC III



(g) Plasmacytoid DCs



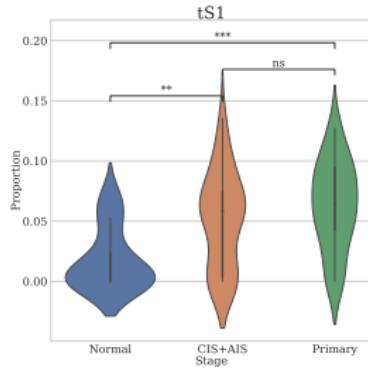
(h) Pleural Mac



(i) Regulatory T cells

Figure: Violin Plots in LUSC

# Violin Plots in LUSC IV



(j) Transcriptional states 1

Figure: Violin Plots in LUSC

# Findings in Bulk Cell Deconvolution with LUSC I

## Activated DCs

- ① Activated DCs have higher proportion in Normal samples.

## Alveolar type II

- ① Alveolar type II have lower proportion in Normal samples.

## CD1c+ DCs (Langerhans cells; LCs)

- ① LCs have higher proportion in Normal samples.

## Ciliated cells

- ① Ciliated cells have higher proportion in Normal samples.

# Findings in Bulk Cell Deconvolution with LUSC II

## Exhausted T follicular help cells

- ① Exhausted T follicular help cell is gradually decreased along cancer worsen.

## Monocyte & Macrophage

- ① Monocyte & Macrophage have lower proportion in Normal samples.

## Plasmacytoid DCs

- ① Plasmacytoid DCs have higher proportion in Normal samples.

## Pleural Macrophages

- ① Pleural Macrophages have higher proportion in Normal samples.

## Regulatory T cells

- ① Regulatory T cells have higher proportion in Normal samples.

## Transcriptional states 1 (tS1)

- ① tS1 have lower proportion in Normal samples.

# Cluster Plot in LUAD

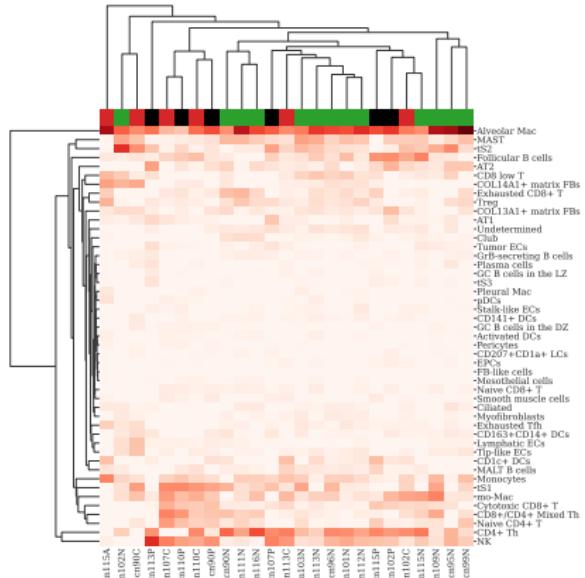
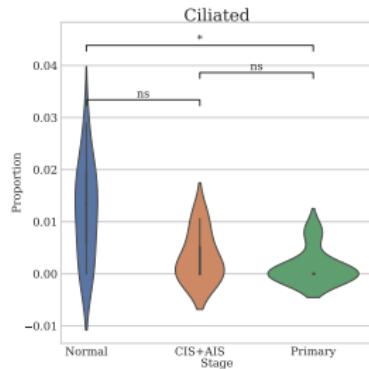
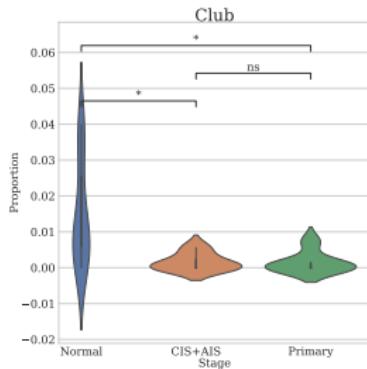


Figure: Cluster Plot in LUAD

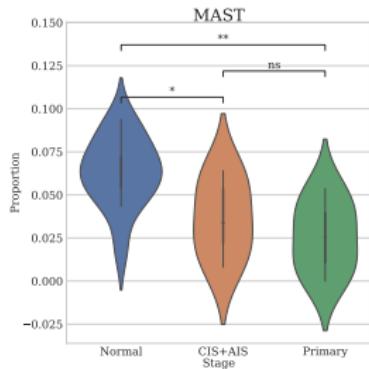
# Violin Plots in LUAD I



(a) Ciliated cells



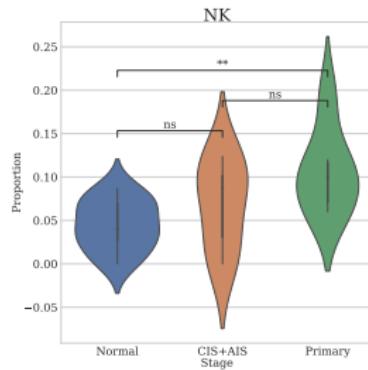
(b) Club Cell



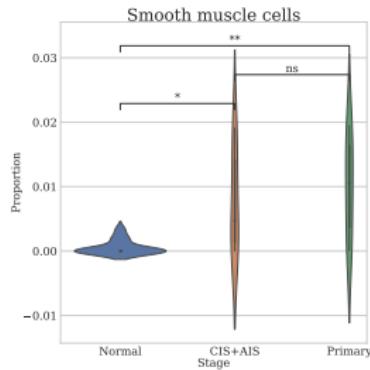
(c) Mast cell

Figure: Violin Plots in LUAD

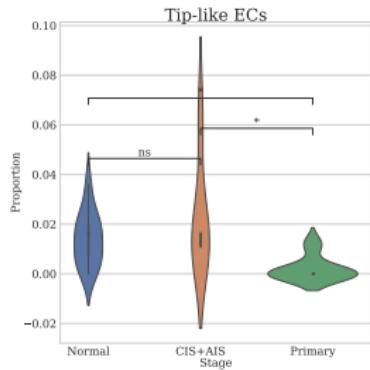
# Violin Plots in LUAD II



(d) NK cells



(e) Smooth muscle cells



(f) Tip-like ECs

Figure: Violin Plots in LUAD

# Findings in Bulk Cell Deconvolution with LUAD I

## Ciliated cells

- ① Ciliated cells have higher proportion in Normal than Primary samples.

## Club cells

- ① Club cells have higher proportion in Normal than Primary samples.

## Mast cells

- ① Mast cells have higher proportion in Normal than Primary samples.

## Natural Killer cells

- ① NK cells have higher proportion in Primary than Normal samples.

# Findings in Bulk Cell Deconvolution with LUAD II

## Smooth muscle cells

- ① Smooth muscle cells have higher proportion in Primary than Normal samples.

## Tip-like ECs

- ① Tip-like ECs have lower proportion in Primary than Normal samples.

## 4. Results

### 4.8. Bulk Cell Deconvolution

#### 4.8.2. Reference by Gueguen et al. (2021)

# Reference Single-cell Data

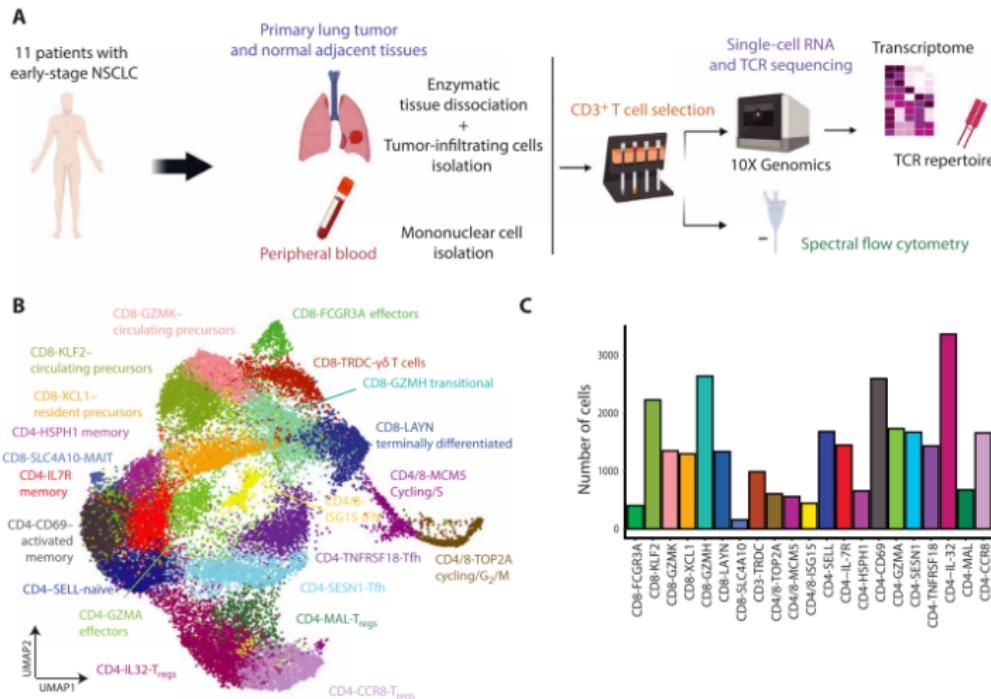


Figure: Characterization of CD3<sup>+</sup> TILs in NSCLC (Gueguen et al., 2021)

# Cluster Plots in LUSC

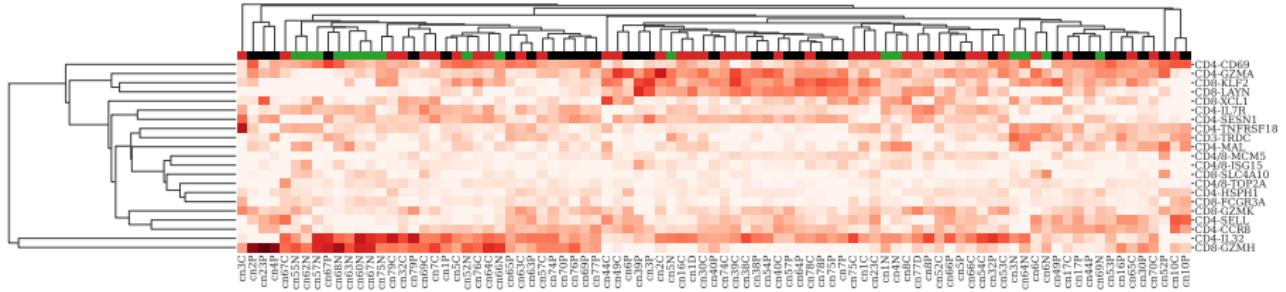


Figure: Cluster Plot in LUAD

# Violin Plots in LUSC I

# Findings in Bulk Cell Deconvolution with LUSC I

# Cluster Plots in LUAD

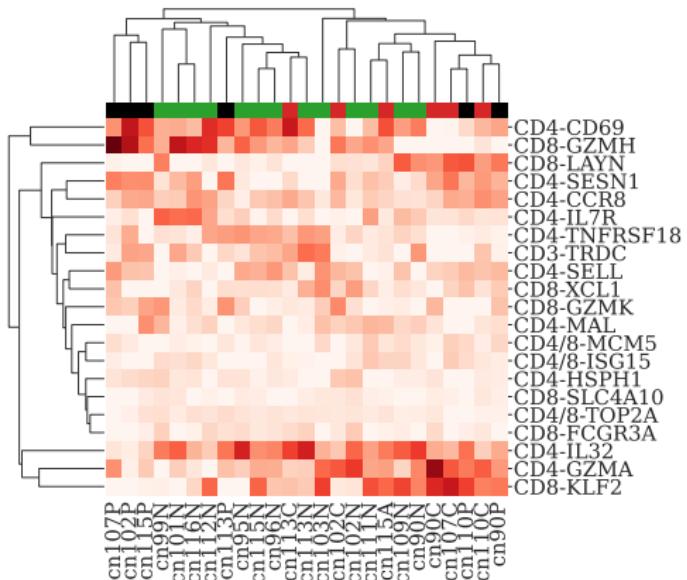


Figure: Cluster Plot in LUAD

# Violin Plots in LUAD I

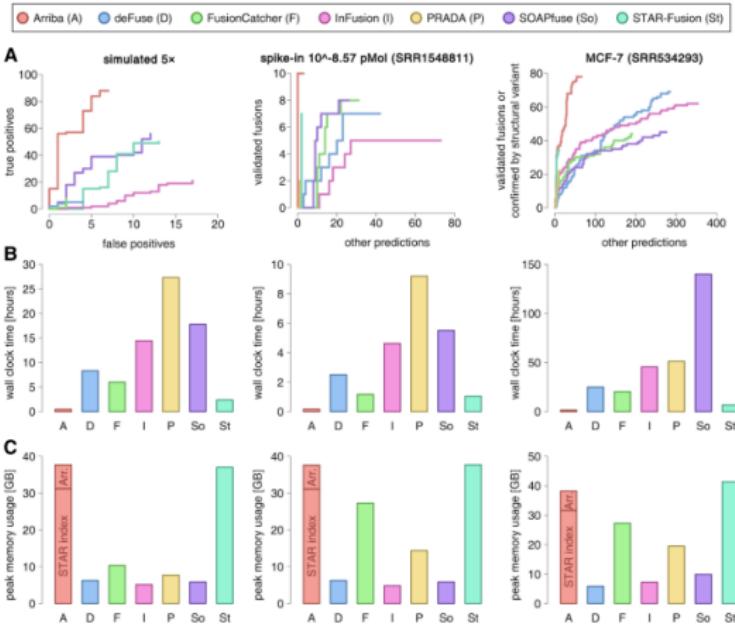
# Findings in Bulk Cell Deconvolution with LUAD I

# Findings in Bulk Cell Deconvolution

## 4. Results

### 4.9. Discovery of Gene Fusion

# Arriba?



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

# Findings in Gene Fusion Discovery

## 4. Results

### 4.10. Discovery of Mutational Signature

# Mutation Signature?

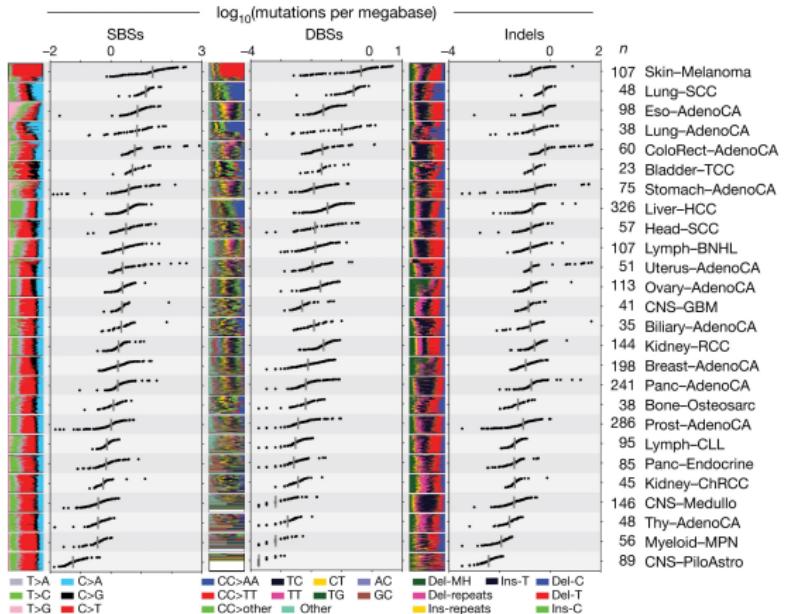
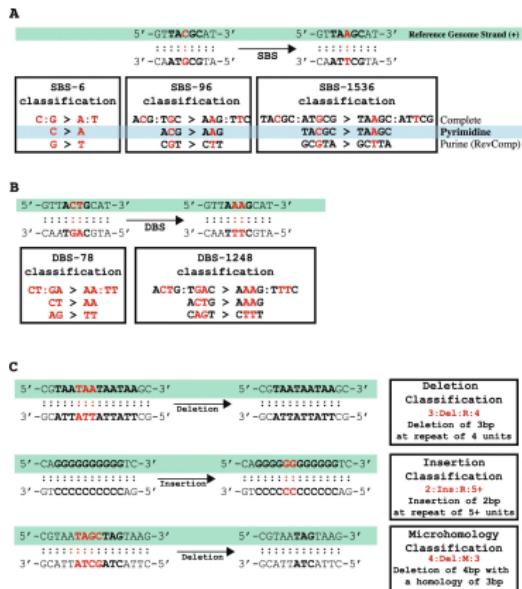


Figure: Mutation Burdens across PCAWG tumor types (Alexandrov et al., 2020)

# SigProfiler?



**Figure:** Classification of mutation signatures by SigProfiler (Bergstrom et al., 2019; Islam et al., 2021; Bergstrom et al., 2020)

## 4. Results

### 4.10. Discovery of Mutational Signature

#### 4.10.1. Single Base Substitutions (SBS)

# SBS Signatures I

SBS1

content...

SBS2

content...

SBS4

content...

SBS5

content...

SBS10b

content...

# SBS Signatures II

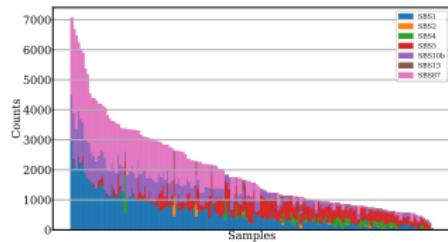
SBS13

content...

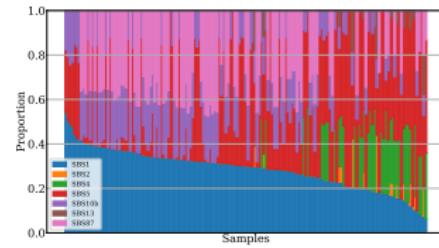
SBS87

content...

# SBS in LUSC I



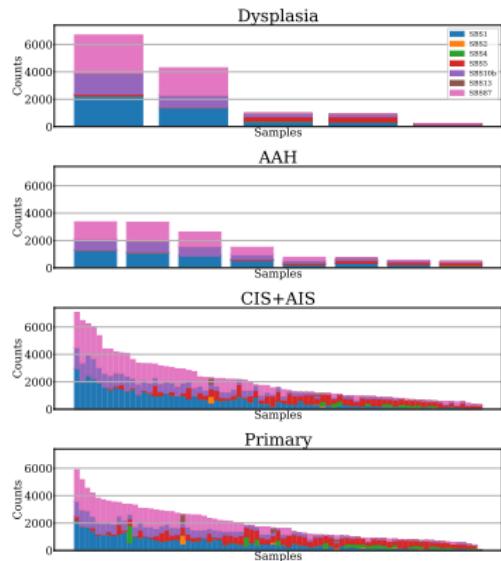
(a) Absolute



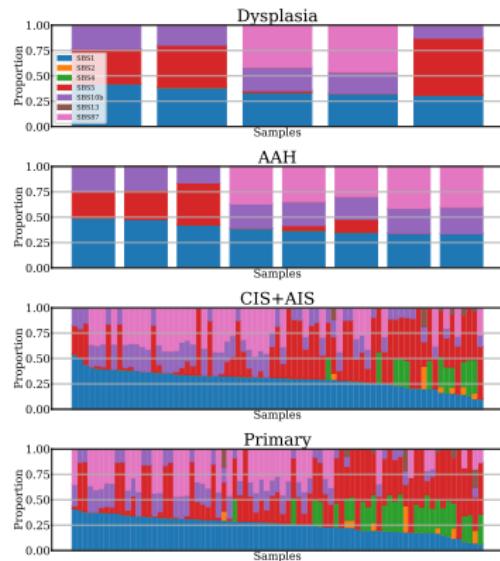
(b) Relative

Figure: SBS Bar Plot in LUSC

# SBS in LUSC II



(a) Absolute



(b) Relative

Figure: SBS Bar Plot by Cancer Subtype in LUSC

# SBS in LUSC with Smoking I

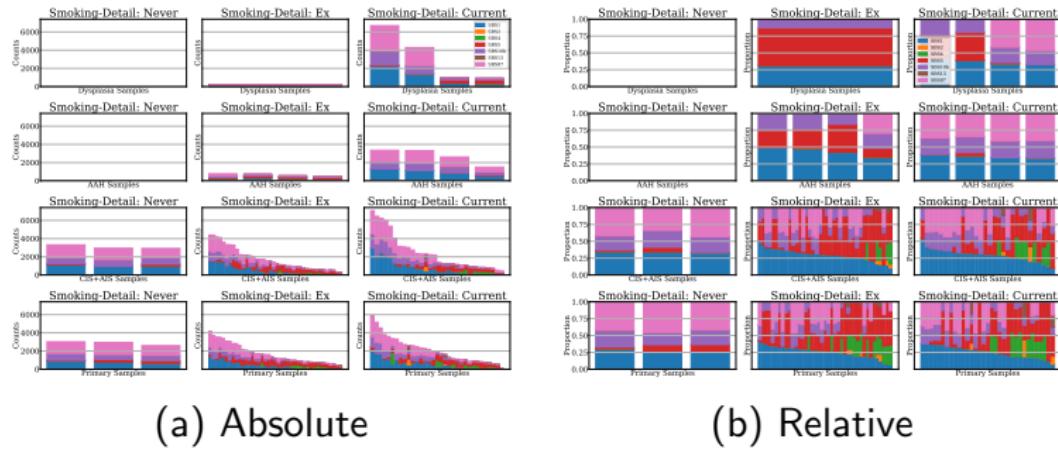
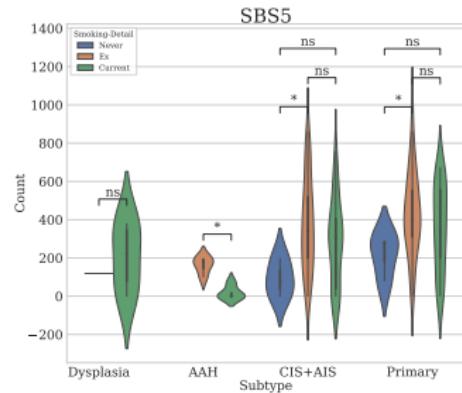
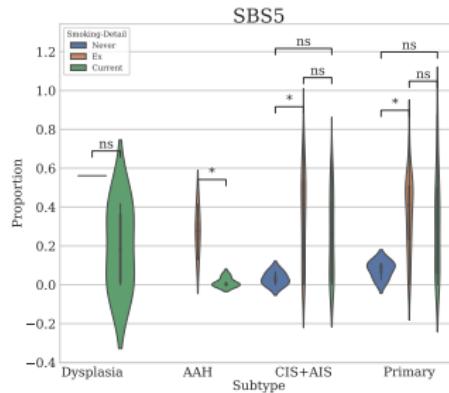


Figure: SBS Bar Plot by Cancer Subtype & Smoking in LUSC

# SBS in LUSC with Smoking II



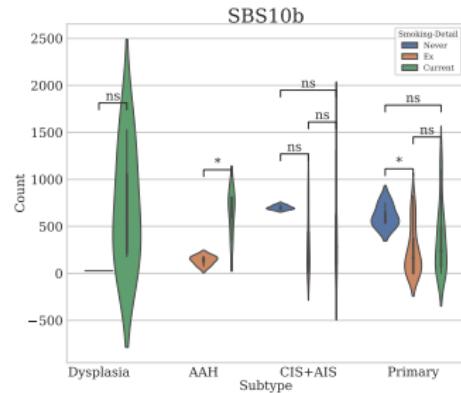
(a) Absolute



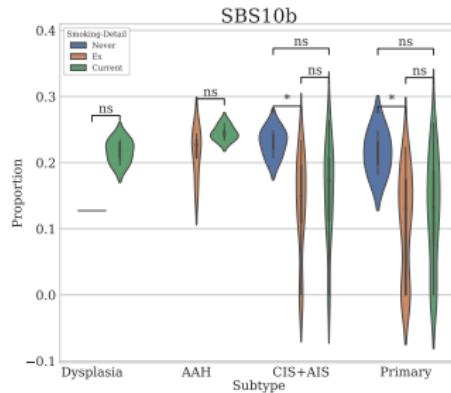
(b) Relative

**Figure:** SBS5 Signature in LUSC with Smoking

# SBS in LUSC with Smoking III



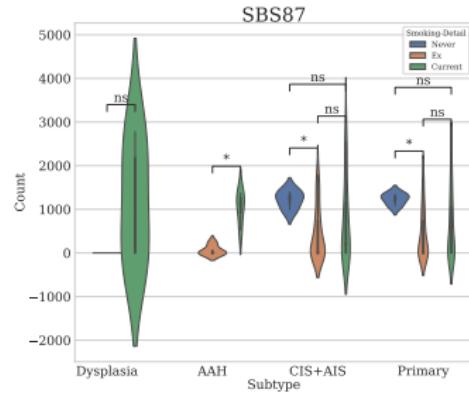
(a) Absolute



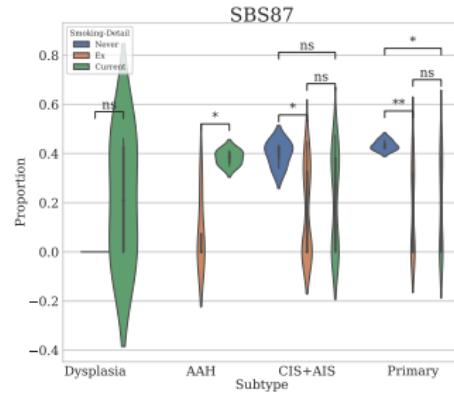
(b) Relative

Figure: SBS10b Signature in LUSC with Smoking

# SBS in LUSC with Smoking IV



(a) Absolute



(b) Relative

Figure: SBS87 Signature in LUSC with Smoking

# SBS in LUSC with Recurrence I

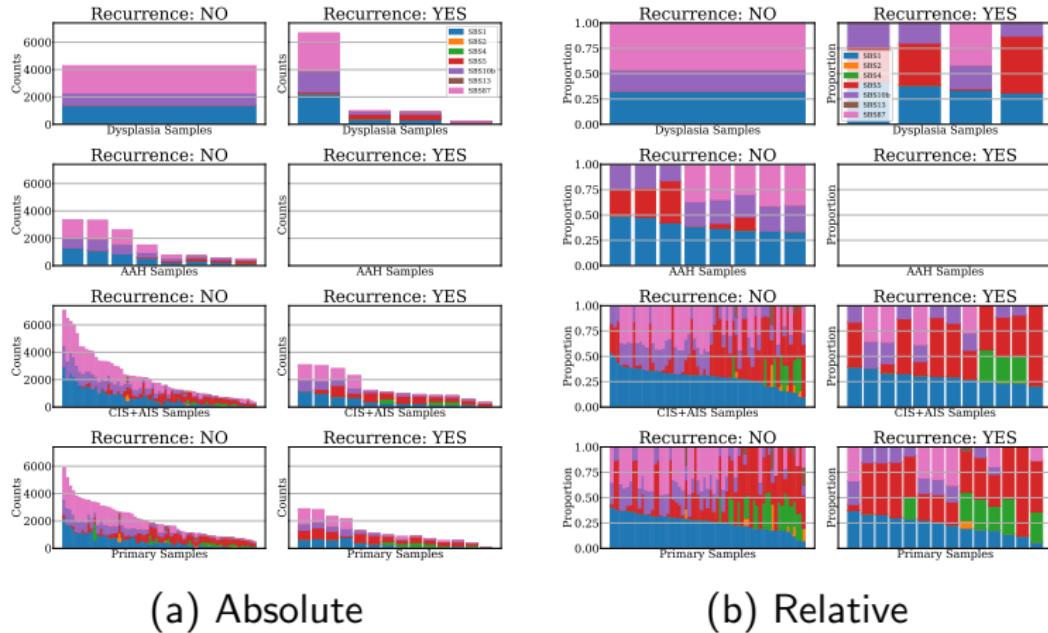
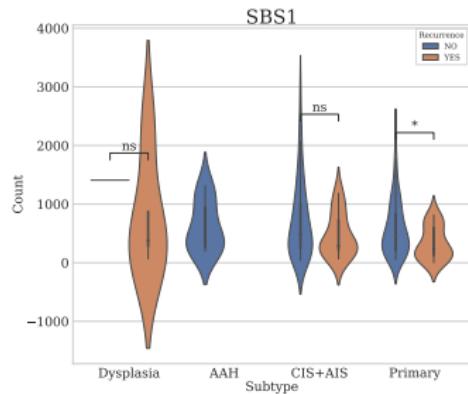
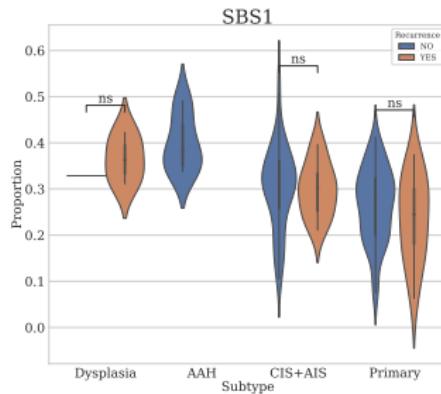


Figure: SBS Bar Plot by Cancer Subtype & Recurrence in LUSC

# SBS in LUSC with Recurrence II



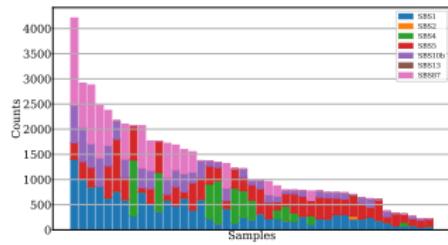
(a) Absolute



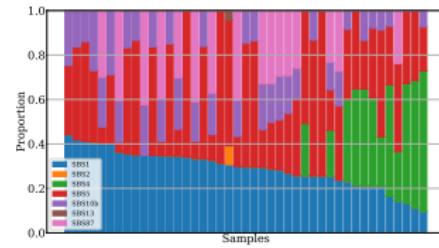
(b) Relative

**Figure:** SBS1 Signature in LUSC with Recurrence

# SBS in LUAD I



(a) Absolute



(b) Relative

Figure: SBS Bar Plot in LUSC

# SBS in LUAD II

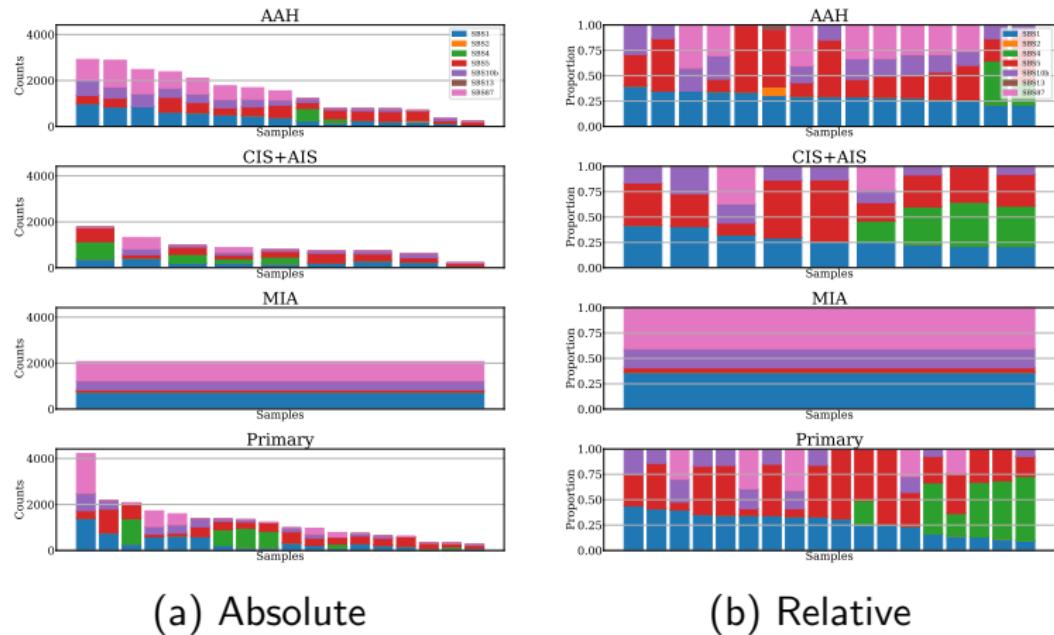


Figure: SBS Bar Plot by Cancer Subtype in LUSC

# SBS in LUAD with Smoking I

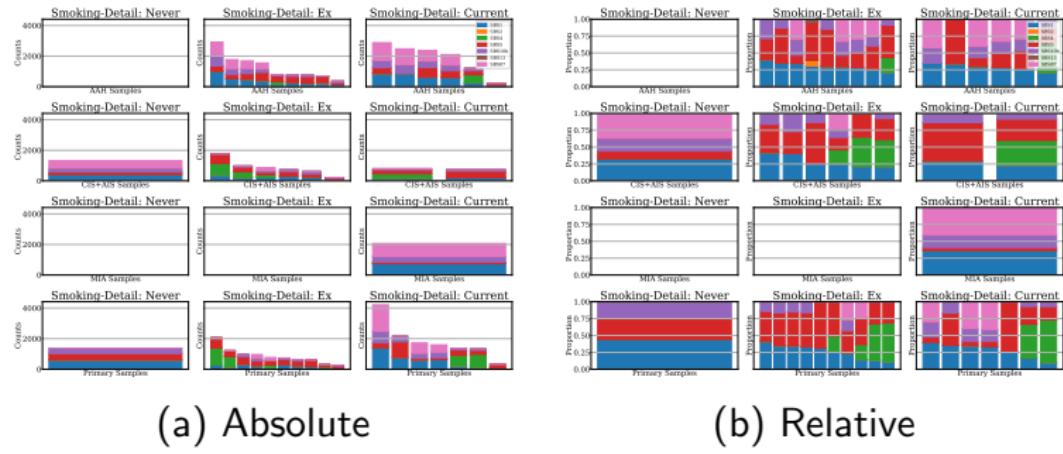
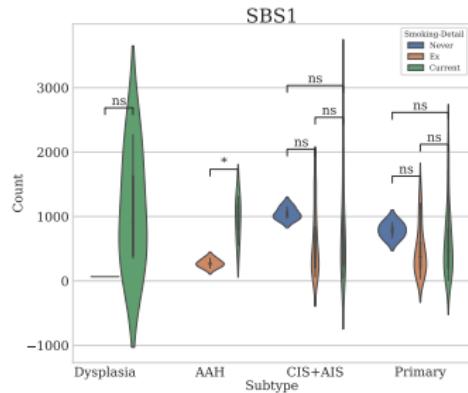
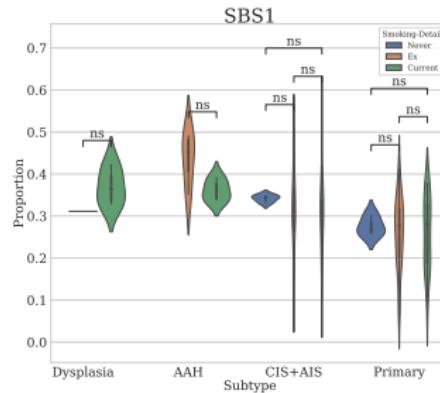


Figure: SBS Bar Plot by Cancer Subtype & Smoking in LUAD

# SBS in LUAD with Smoking II



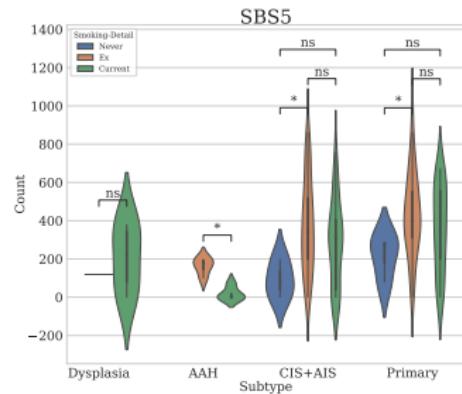
(a) Absolute



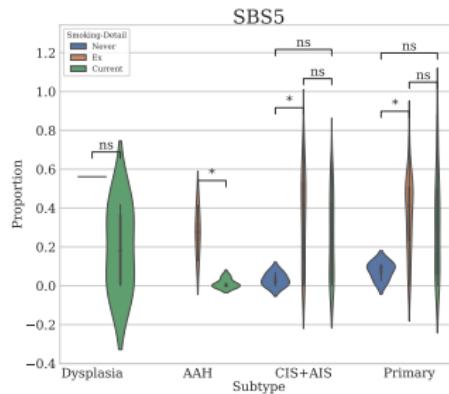
(b) Relative

**Figure:** SBS1 Signature in LUAD with Smoking

# SBS in LUAD with Smoking III



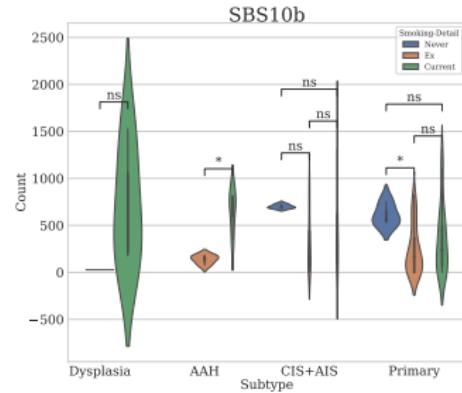
(a) Absolute



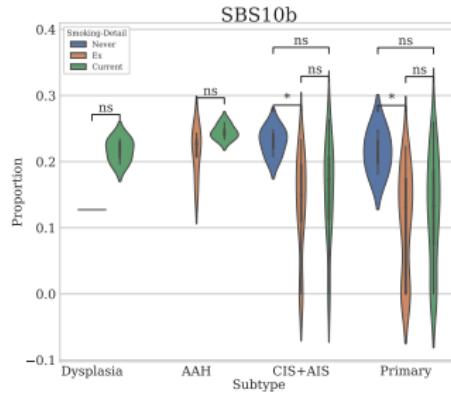
(b) Relative

**Figure:** SBS5 Signature in LUAD with Smoking

# SBS in LUAD with Smoking IV



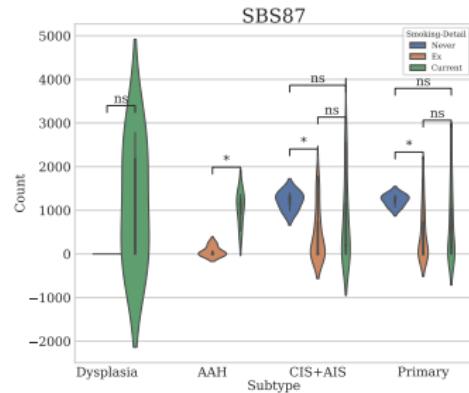
(a) Absolute



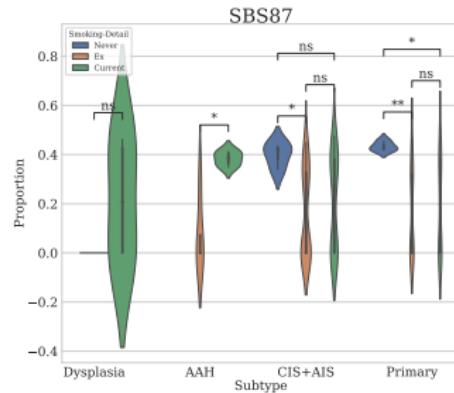
(b) Relative

Figure: SBS10b Signature in LUAD with Smoking

# SBS in LUAD with Smoking V



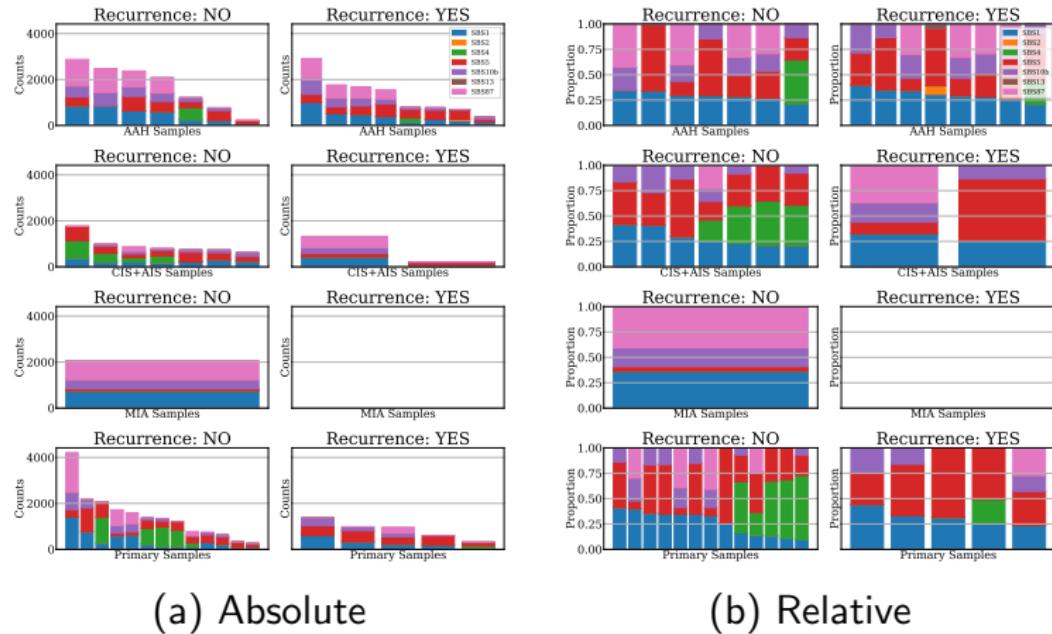
(a) Absolute



(b) Relative

Figure: SBS87 Signature in LUAD with Smoking

# SBS in LUAD with Recurrence I

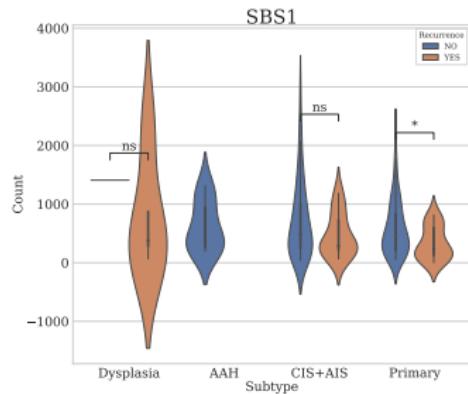


(a) Absolute

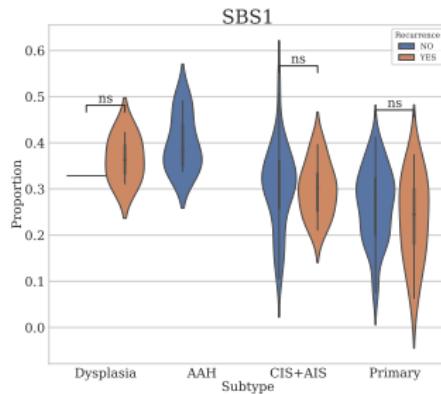
(b) Relative

Figure: SBS Bar Plot by Cancer Subtype & Recurrence in LUAD

# SBS in LUAD with Recurrence II



(a) Absolute



(b) Relative

Figure: SBS1 Signature in LUAD with Recurrence

## 4. Results

### 4.10. Discovery of Mutational Signature

#### 4.10.2. Double Base Substitutions (DBS)

# DBS Signatures I

DBS2

content...

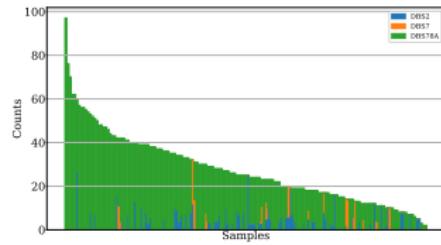
DBS7

content...

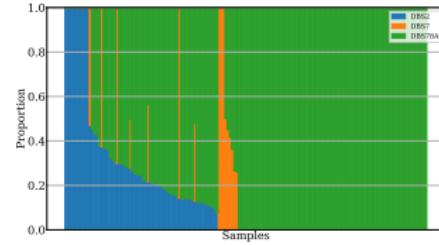
DBS78A

content...

# DBS in LUSC I



(a) Absolute



(b) Relative

Figure: DBS Bar Plot in LUSC

# DBS in LUSC II

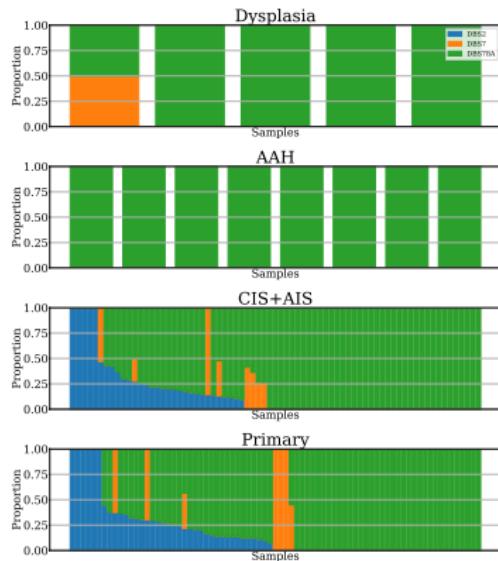
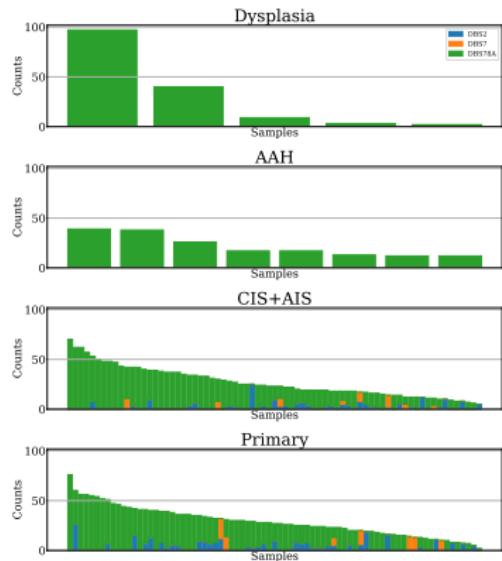


Figure: DBS Bar Plot by Cancer Subtype in LUSC

# DBS in LUSC with Smoking I

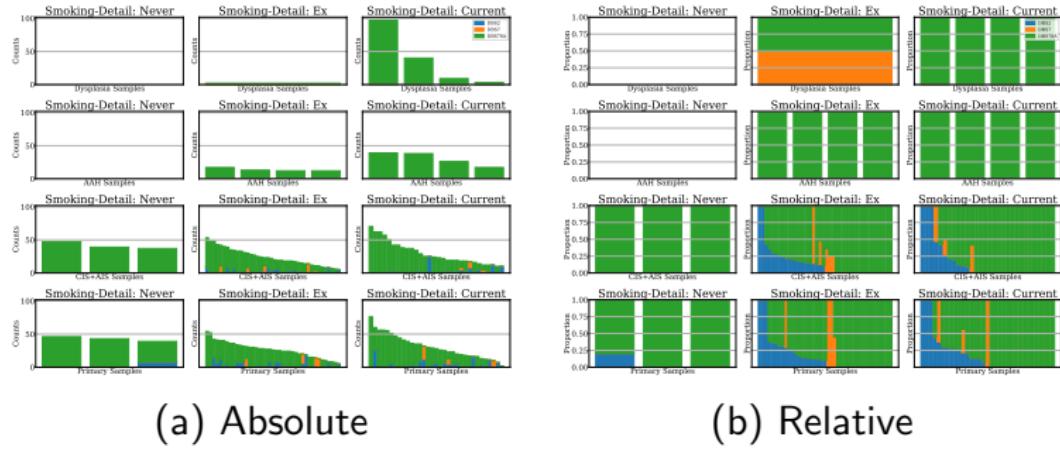


Figure: DBS Bar Plot by Cancer Subtype & Smoking in LUSC

# DBS in LUSC with Recurrence I

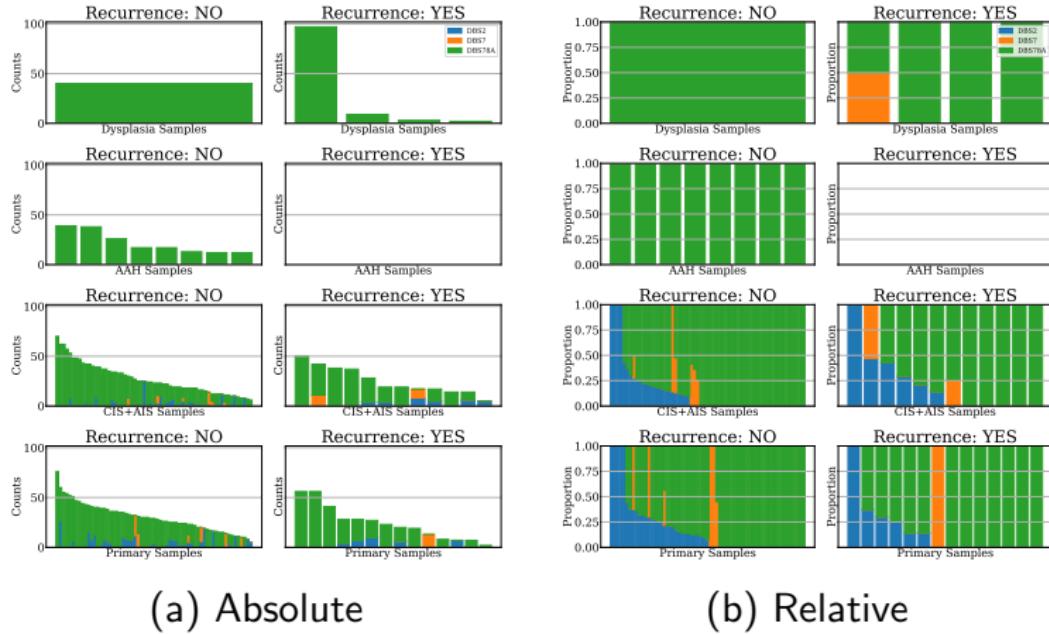
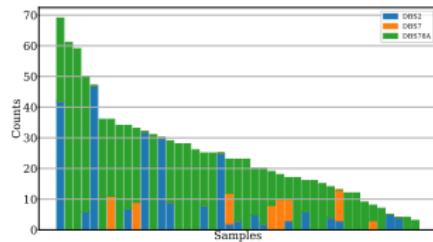
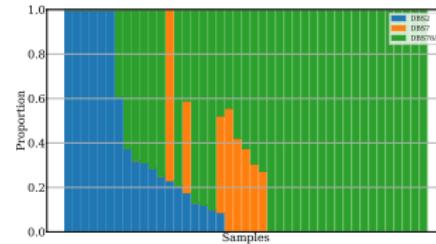


Figure: DBS Bar Plot by Cancer Subtype & Recurrence in LUSC

# DBS in LUAD I



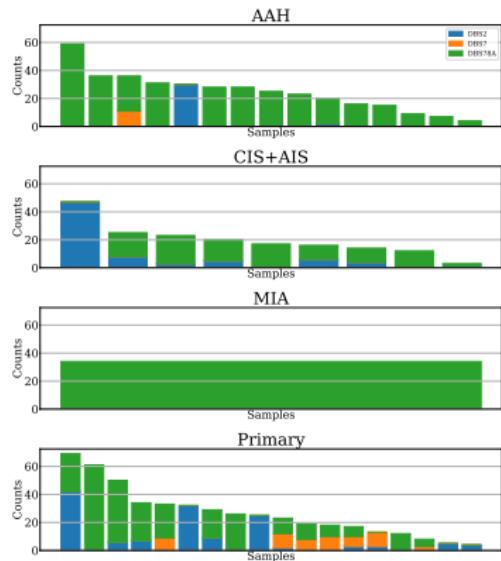
(a) Absolute



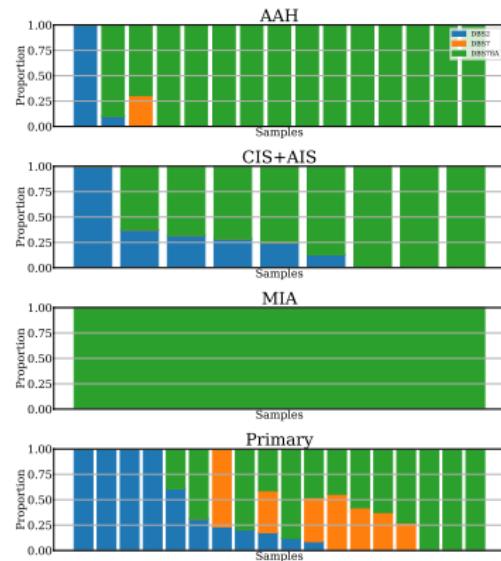
(b) Relative

Figure: DBS Bar Plot in LUAD

# DBS in LUAD II



(a) Absolute



(b) Relative

Figure: DBS Bar Plot by Cancer Subtype in LUAD

# DBS in LUAD with Smoking I

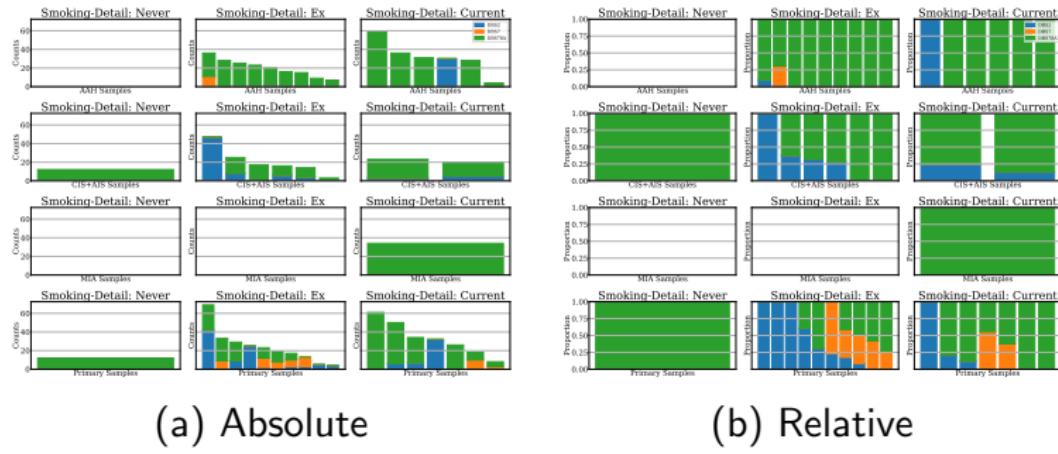


Figure: DBS Bar Plot by Cancer Subtype & Smoking in LUAD

# DBS in LUAD with Recurrence I

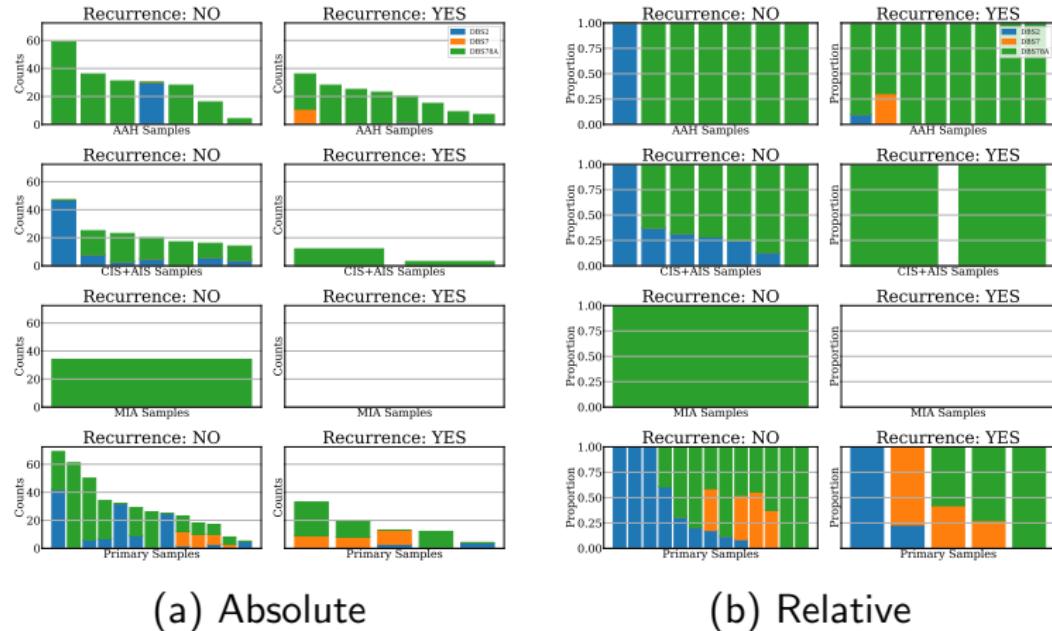


Figure: DBS Bar Plot by Cancer Subtype & Recurrence in LUAD

## 4. Results

### 4.10. Discovery of Mutational Signature

#### 4.10.3. Short insertions & Deletions (Indels)

# Indel signatures I

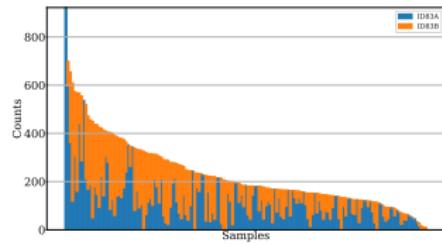
ID83A

content...

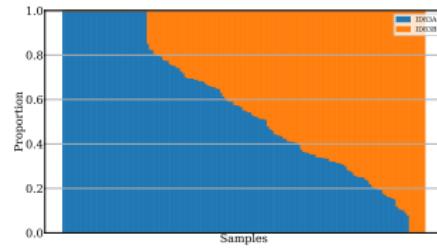
ID83B

content...

# Indels in LUSC I



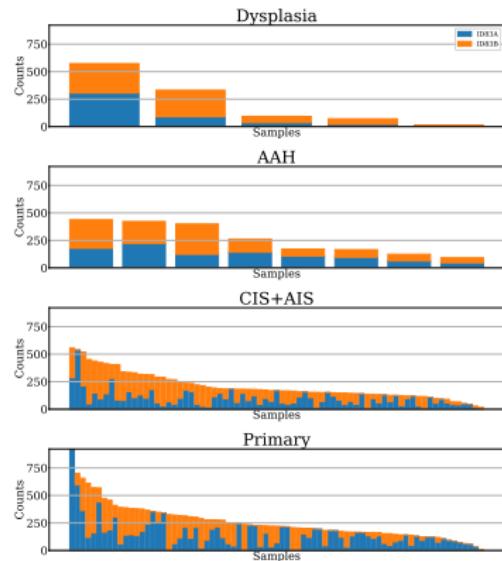
(a) Absolute



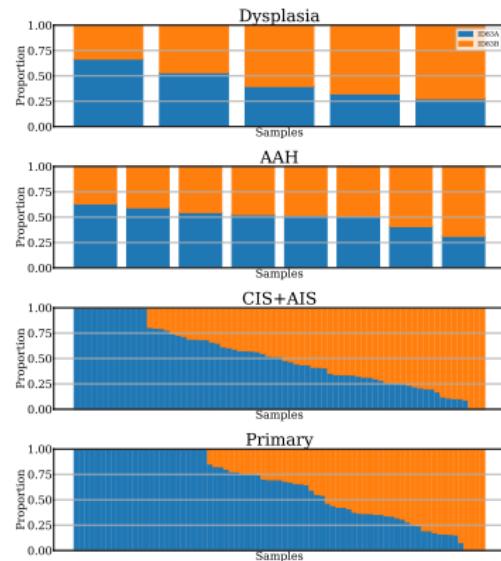
(b) Relative

Figure: Indel Bar Plot in LUSC

# Indels in LUSC II



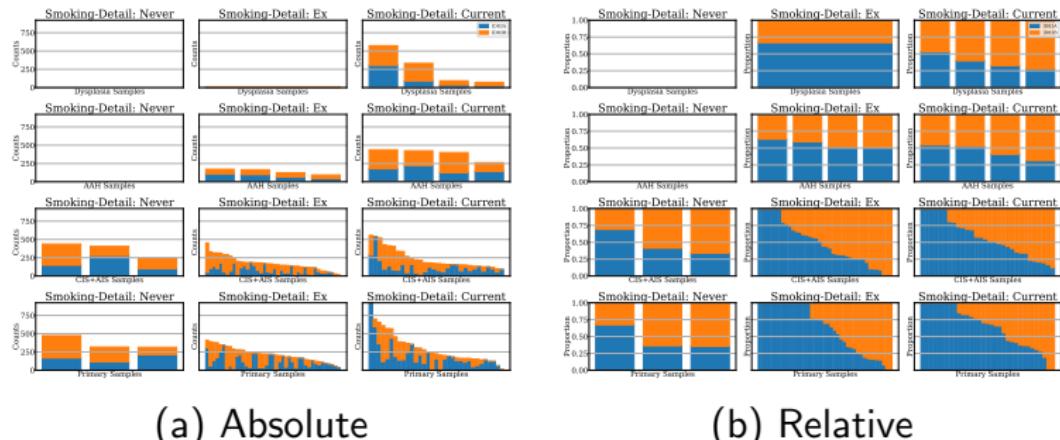
(a) Absolute



(b) Relative

Figure: Indel Bar Plot by Cancer Subtype in LUSC

# Indel in LUSC with Smoking I



(a) Absolute

(b) Relative

Figure: Indel Bar Plot by Cancer Subtype & Smoking in LUSC

# Indel in LUSC with Recurrence I

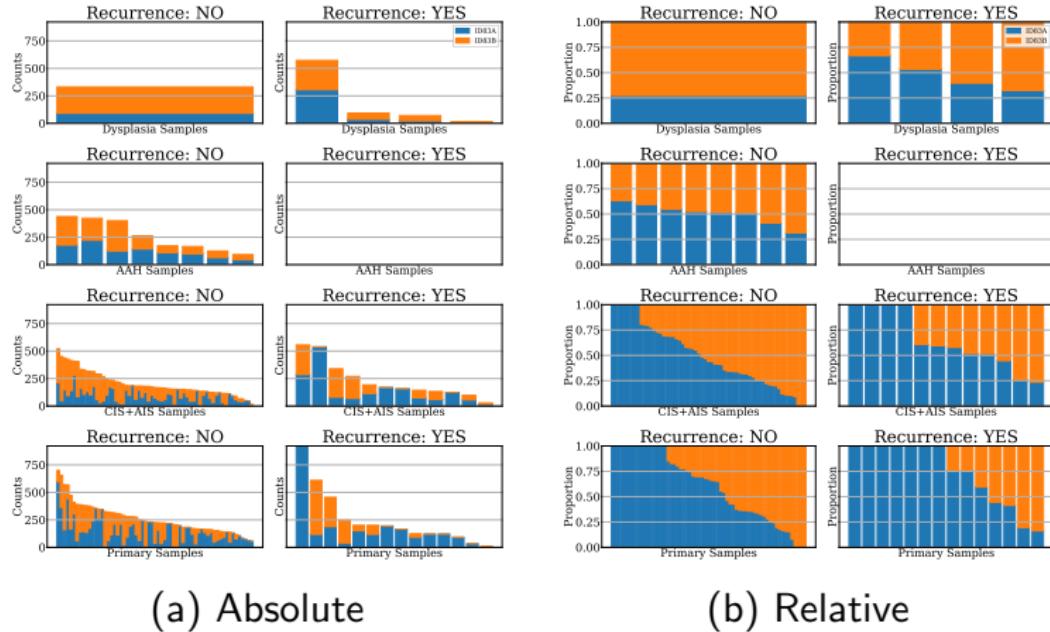
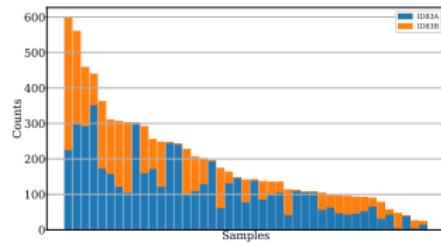
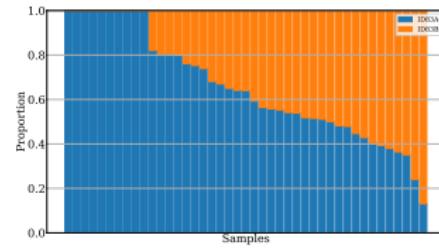


Figure: Indel Bar Plot by Cancer Subtype & Recurrence in LUSC

# Indels in LUAD I



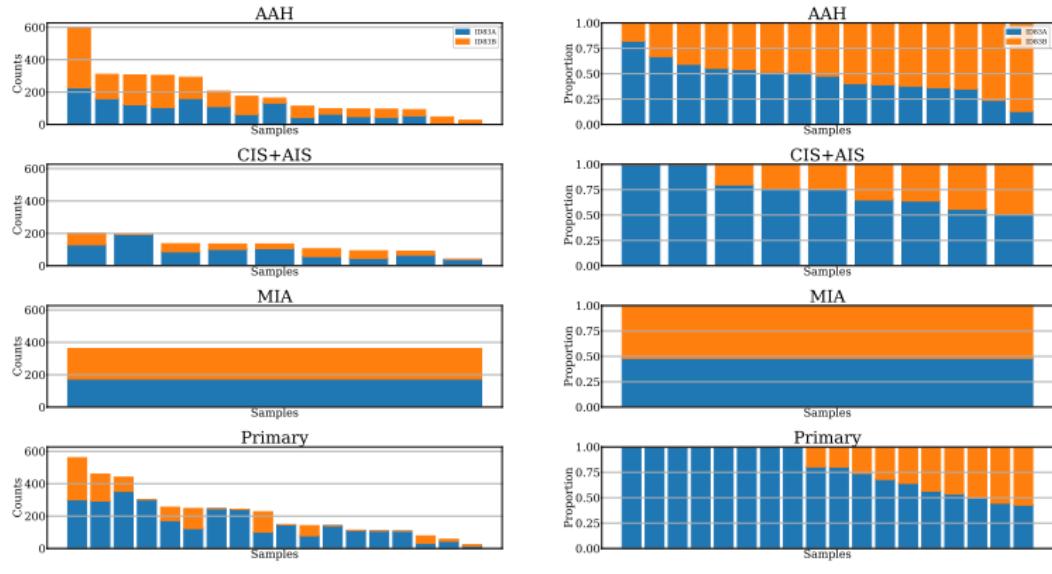
(a) Absolute



(b) Relative

Figure: Indel Bar Plot in LUAD

# Indels in LUAD II



(a) Absolute

(b) Relative

Figure: Indel Bar Plot by Cancer Subtype in LUAD

# Indel in LUAD with Smoking I

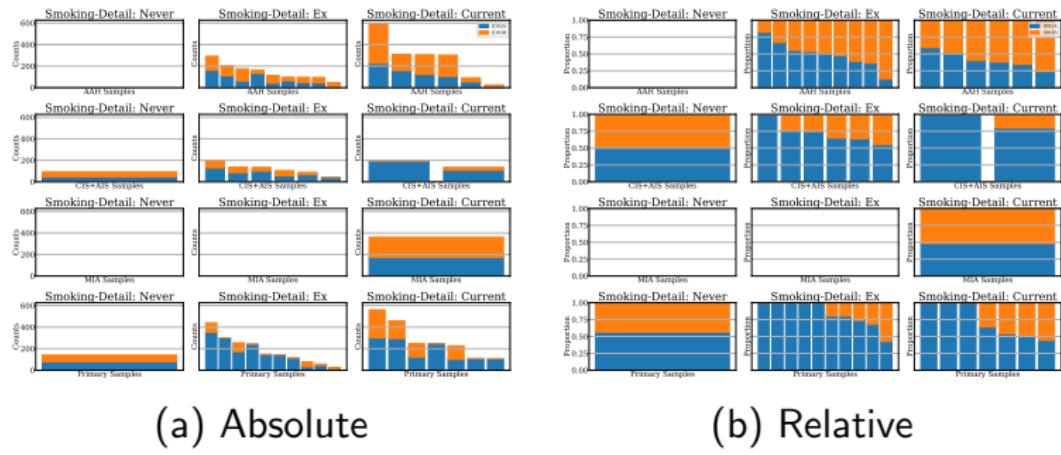


Figure: Indel Bar Plot by Cancer Subtype & Smoking in LUAD

# Indel in LUAD with Recurrence I

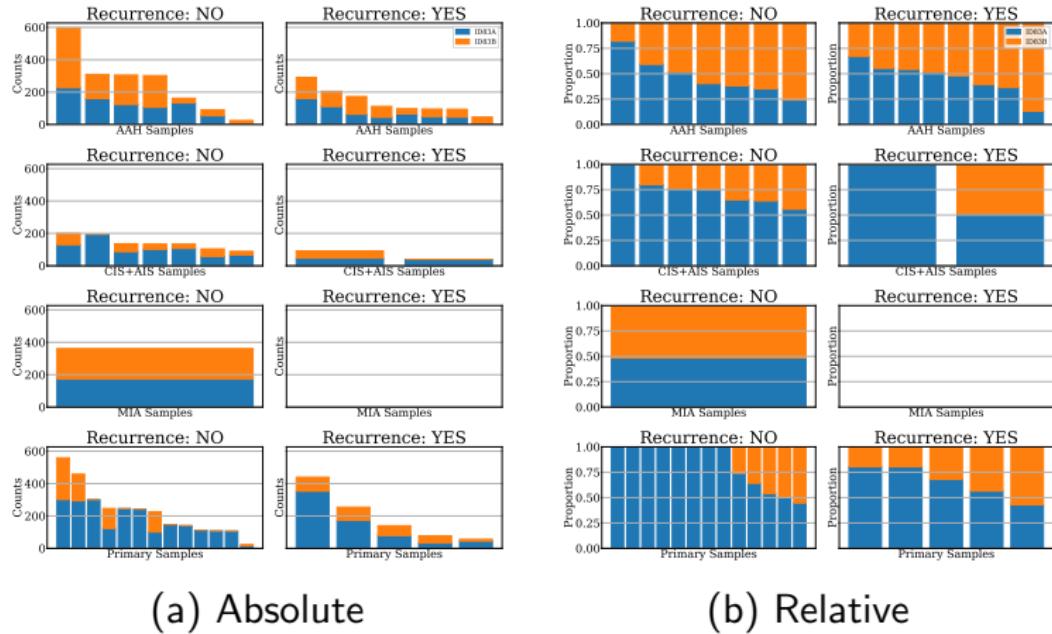


Figure: Indel Bar Plot by Cancer Subtype & Recurrence in LUAD

# Findings in Mutation Signature

## 4. Results

### 4.11. Clinical Data with Point Mutation

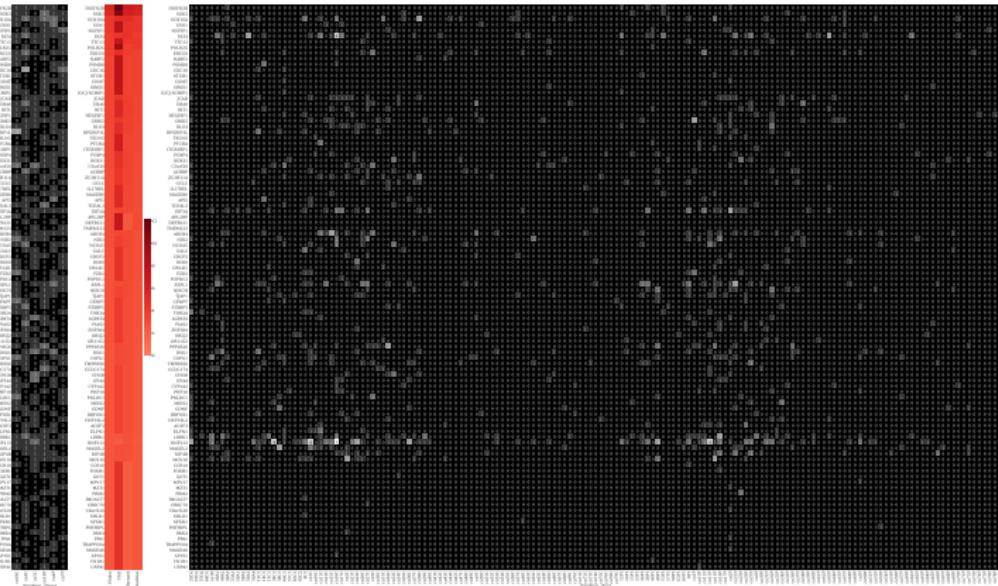
# Mutect2?

## 4. Results

### 4.11. Clinical Data with Point Mutation

#### 4.11.1. For Smoking

# Clinical Data about LUSC for Smoking I



## Figure: Clinical Data about LUSC for Smoking

# Clinical Data about LUSC for Smoking II

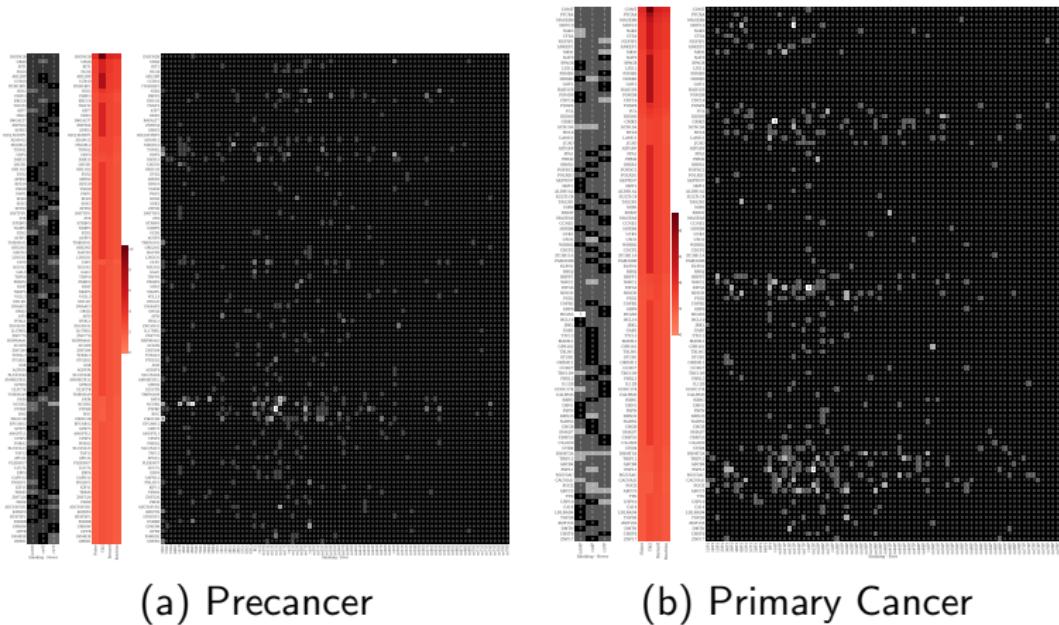
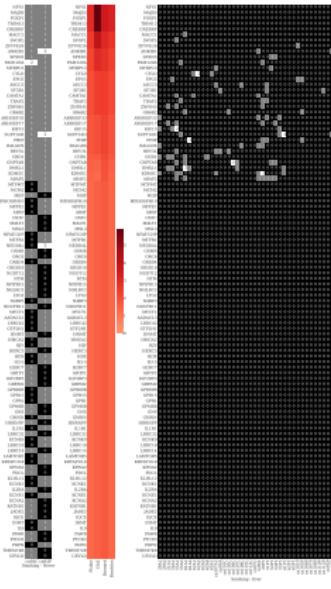


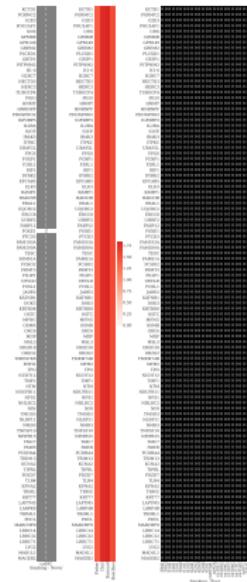
Figure: Clinical Data about LUSC for Smoking with Precancer/Primary

# Clinical Data about LUAD for Smoking I



## Figure: Clinical Data about LUAD for Smoking

# Clinical Data about LUAD for Smoking II



(a) Precancer



(b) Primary Cancer

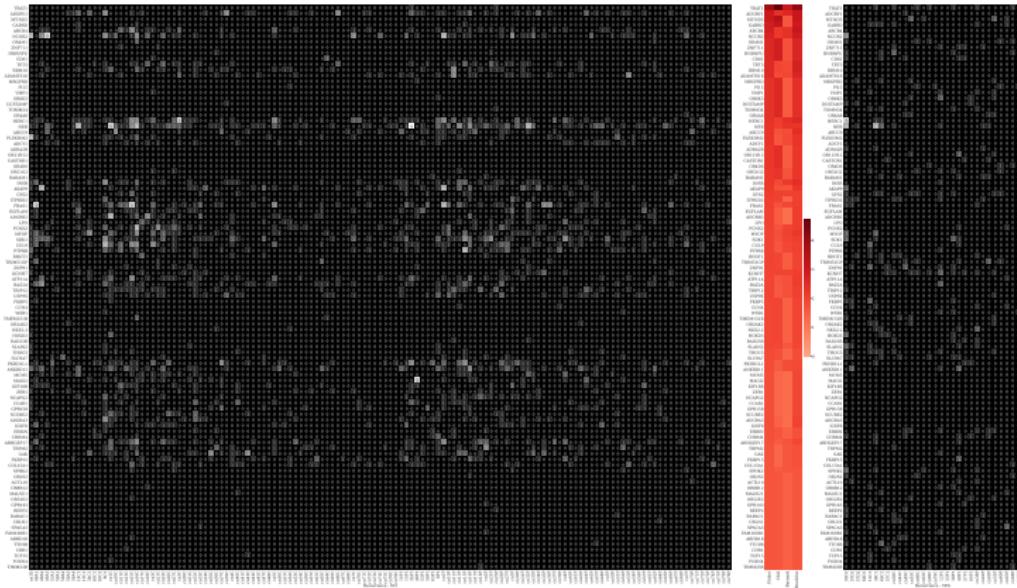
Figure: Clinical Data about LUAD for Smoking with Precancer/Primary

## 4. Results

### 4.11. Clinical Data with Point Mutation

#### 4.11.2. For Recurrence

## Clinical Data about LUSC for Recurrence I



## Figure: Clinical Data about LUSC for Recurrence

## Clinical Data about LUSC for Recurrence II

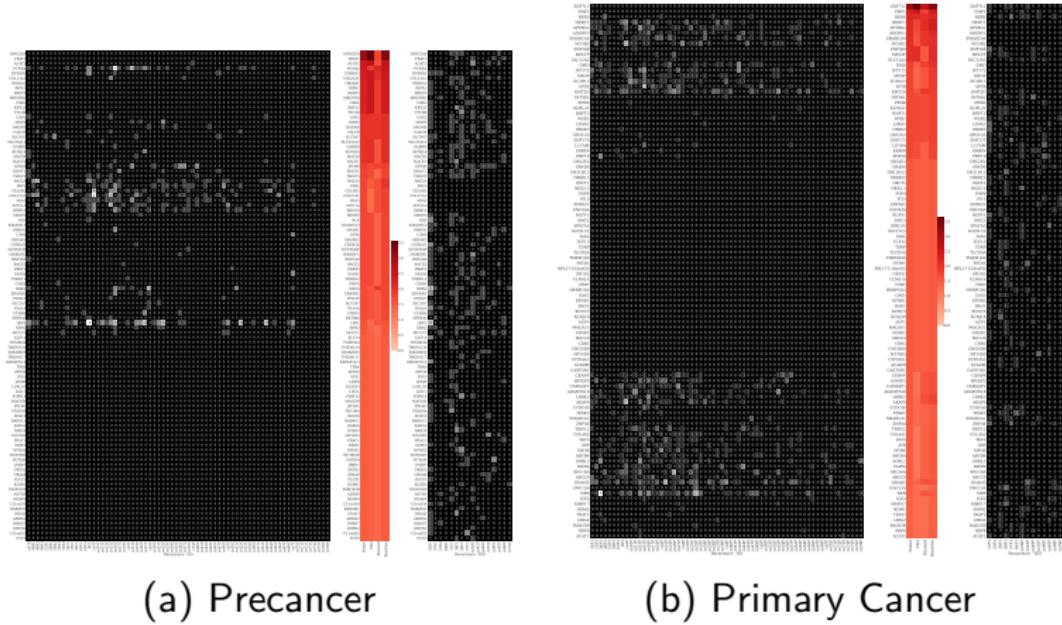


Figure: Clinical Data about LUSC for Recurrence with Precancer/Primary

# Clinical Data about LUAD for Recurrence I

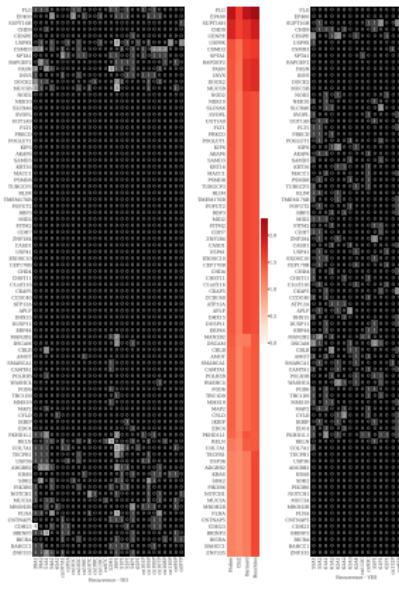


Figure: Clinical Data about LUAD for Recurrence

# Clinical Data about LUAD for Recurrence II

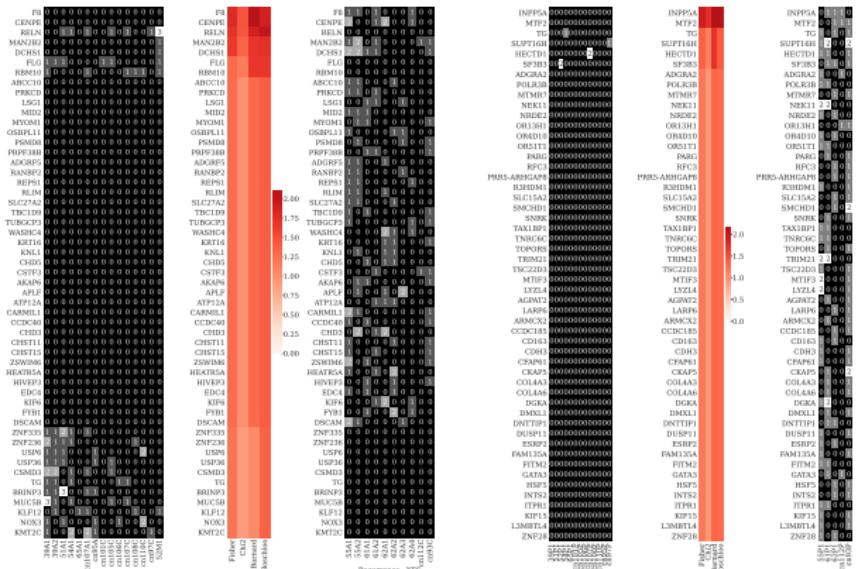


Figure: Clinical Data about LUAD for Recurrence with Precancer/Primary

# Findings in Clinical Data with Point Mutation

## 5. Discussion

## 6. References

## References I

- Alexandrov, L. B., Kim, J., Haradhvala, N. J., Huang, M. N., Ng, A. W. T., Wu, Y., ... others (2020). The repertoire of mutational signatures in human cancer. *Nature*, 578(7793), 94–101.
- Andrews, S., Krueger, F., Segonds-Pichon, A., Biggins, L., Krueger, C., & Wingett, S. (2012, January). *FastQC*. Babraham Institute. Babraham, UK.
- Bergstrom, E. N., Barnes, M., Martincorena, I., & Alexandrov, L. B. (2020). Generating realistic null hypothesis of cancer mutational landscapes using sigprofilersimulator. *BMC bioinformatics*, 21(1), 1–10.
- Bergstrom, E. N., Huang, M. N., Mahto, U., Barnes, M., Stratton, M. R., Rozen, S. G., & Alexandrov, L. B. (2019). Sigprofilermatrixgenerator: a tool for visualizing and exploring patterns of small mutational events. *BMC genomics*, 20(1), 1–12.

## References II

- Bollig-Fischer, A., Bao, B., Manning, M., Dyson, G., Michelhaugh, S. K., Mittal, S., ... Mamdani, H. (2021). Role of novel cancer gene slitrk3 to activate ntrk3 in squamous cell lung cancer. *Molecular Biomedicine*, 2(1), 1–12.
- 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 III

- Collins, L. G., Haines, C., Perkel, R., & Enck, R. E. (2007). Lung cancer: diagnosis and management. *American family physician*, 75(1), 56–63.
- Counts, S. E., He, B., Che, S., Ikonomovic, M. D., DeKosky, S. T., Ginsberg, S. D., & Mufson, E. J. (2007).  $\alpha 7$  nicotinic receptor up-regulation in cholinergic basal forebrain neurons in alzheimer disease. *Archives of neurology*, 64(12), 1771–1776.
- Crowdis, J., He, M. X., Reardon, B., & Van Allen, E. M. (2020). Comut: visualizing integrated molecular information with comutation plots. *Bioinformatics*, 36(15), 4348–4349.
- Dean, B., & Scarr, E. (2021). Changes in cortical gene expression in the muscarinic m1 receptor knockout mouse: potential relevance to schizophrenia, alzheimer's disease and cognition. *npj Schizophrenia*, 7(1), 1–7.

## References IV

- 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 V

- Geng, Y., Deng, L., Su, D., Xiao, J., Ge, D., Bao, Y., & Jing, H. (2016). Identification of crucial micrornas and genes in hypoxia-induced human lung adenocarcinoma cells. *OncoTargets and therapy*, 9, 4605.
- Gueguen, P., Metoikidou, C., Dupic, T., Lawand, M., Goudot, C., Baulande, S., ... others (2021). Contribution of resident and circulating precursors to tumor-infiltrating cd8+ t cell populations in lung cancer. *Science Immunology*, 6(55), eabd5778.
- 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.

## References VI

- 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.
- Islam, S. M. A., Wu, Y., Díaz-Gay, M., Bergstrom, E. N., He, Y., Barnes, M., ... Alexandrov, L. B. (2021). Uncovering novel mutational signatures by de novo extraction with sigprofilerextractor. *bioRxiv*. Retrieved from <https://www.biorxiv.org/content/early/2021/05/16/2020.12.13.422570> doi: 10.1101/2020.12.13.422570

## References VII

- 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.
- 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.

## References VIII

- 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.
- Lee, D., Lee, G. K., Yoon, K.-A., & Lee, J. S. (2013). Pathway-based analysis using genome-wide association data from a korean non-small cell lung cancer study. *PloS one*, 8(6), e65396.
- 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.

## References IX

- 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, sftpib, 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.
- Ma, G., Ji, D., Qu, X., Liu, S., Yang, X., Wang, G., ... Du, J. (2019). Mining and validating the expression pattern and prognostic value of acetylcholine receptors in non-small cell lung cancer. *Medicine*, 98(20).
- Marcé-Grau, A., Elorza-Vidal, X., Pérez-Rius, C., Ruiz-Nel·lo, A., Sala-Coromina, J., Gabau, E., ... Macaya, A. (2021). Muscarinic acetylcholine receptor m1 mutations causing neurodevelopmental disorder and epilepsy. *Human Mutation*.

## References X

- Maugeri, A., Klevering, B. J., Rohrschneider, K., Blankenagel, A., Brunner, H. G., Deutman, A. F., ... Cremers, F. P. (2000). Mutations in the abca4 (abcr) gene are the major cause of autosomal recessive cone-rod dystrophy. *The American Journal of Human Genetics*, 67(4), 960–966.
- Minna, J. D., Roth, J. A., & Gazdar, A. F. (2002). Focus on lung cancer. *Cancer cell*, 1(1), 49–52.
- Mondejar-Parreño, G., Perez-Vizcaino, F., & Cogolludo, A. (2020). Kv7 channels in lung diseases. *Frontiers in Physiology*, 11, 634.
- Nikolaou, S., Qiu, S., Fiorentino, F., Simillis, C., Rasheed, S., Tekkis, P., & Kontovounisios, C. (2020). The role of neuropeptides and their receptors in non-gastrointestinal cancers: a review. *Cell Communication and Signaling*, 18(1), 1–10.

## References XI

- Niu, Y., Lin, A., Luo, P., Zhu, W., Wei, T., Tang, R., ... Zhang, J. (2020). Prognosis of lung adenocarcinoma patients with ntrk3 mutations to immune checkpoint inhibitors. *Frontiers in pharmacology*, 11, 1213.
- 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.

## References XII

- Schroeder, B. C., Kubisch, C., Stein, V., & Jentsch, T. J. (1998). Moderate loss of function of cyclic-amp-modulated kcnq2/kcnq3 k<sup>+</sup> channels causes epilepsy. *Nature*, 396(6712), 687–690.
- Singh, N. A., Westenskow, P., Charlier, C., Pappas, C., Leslie, J., Dillon, J., ... Leppert, M. F. (2003). Kcnq2 and kcnq3 potassium channel genes in benign familial neonatal convulsions: expansion of the functional and mutation spectrum. *Brain*, 126(12), 2726–2737.
- 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.
- Surti, T. S., Huang, L., Jan, Y. N., Jan, L. Y., & Cooper, E. C. (2005). Identification by mass spectrometry and functional characterization of two phosphorylation sites of kcnq2/kcnq3 channels. *Proceedings of the National Academy of Sciences*, 102(49), 17828–17833.

## References XIII

- 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 XIV

- 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.
- Wu, Z., Fournel, L., Stadler, N., Liu, J., Boullier, A., Hoyeau, N., ... others (2019). Modulation of lung cancer cell plasticity and heterogeneity with the restoration of cisplatin sensitivity by neurotensin antibody. *Cancer letters*, 444, 147–161.