

# Lung Precancer Analysis

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*

2021-12-17

# 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

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 year)

- Thyroid: 68.4 %
- Lung: 35.4 %

# Type of Lung Cancer

Types of lung cancer (Collins, Haines, Perkel, & Enck, 2007):

- 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)
- ⑧ ...

# 1. Introduction

## 1.2. Non-small cell lung cancer

# Non-small cell lung cancer (NSCLC)

Types of NSCLC (Goldstraw et al., 2011):

- Adenocarcinoma (ICDO 8140/3)
- Squamous cell carcinoma (ICDO 8070/3)
- Large-cell carcinoma (ICDO 8012/3)
- ...

# 1. Introduction

## 1.3. LUAD

## TCGA LUAD (Duhig et al., 2014)

- 81 % patients reported past/present tobacco smoking.
- Candidate driver genes: RTK, RAS, and RAF (38 %)
- Cancer-associated mutations: KRAS (32 %), EGFR (11 %), and BRAF (7 %)
- Enriched mutations: TP53, KEAP1, NF1, and RIT1 ( $p < 0.01$ )
- Fusions: ROS1 and RET

# 1. Introduction

## 1.4. LUSC

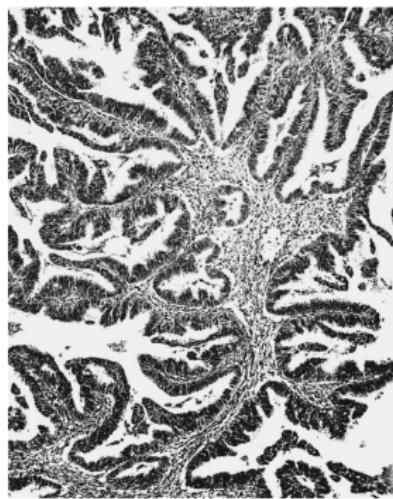
## TCGA LUSC (Dosenbach et al., 2007)

- 96 % patients reported past/present tobacco smoking.
- Not present EGFR and ALK fusions.
- Recurrent mutations: TP53, NFE2L2, KEAP1, BAI3, FBXW7, GRM8, MUC16, RUNX1T1, STK11, and ERBB4
- High rate of copy number alteration compared with other TCGA projects.
- Amplification of NFE2L2, MYC, CDK6, MDM2, BCL2L1, and EYS.
- Deletion of FOXP1, PTEN, and NF1

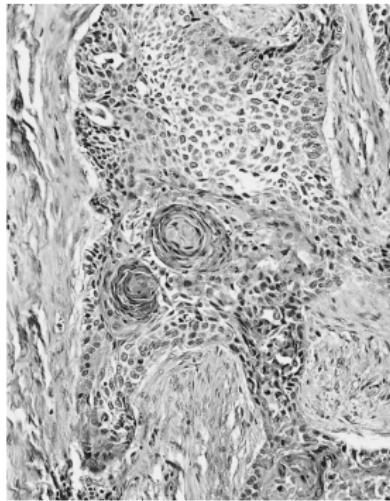
# 1. Introduction

## 1.5. LUAD vs. LUSC

# LUAD vs. LUSC I



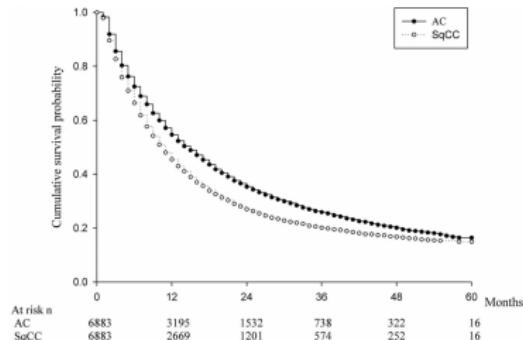
(a) LUAD



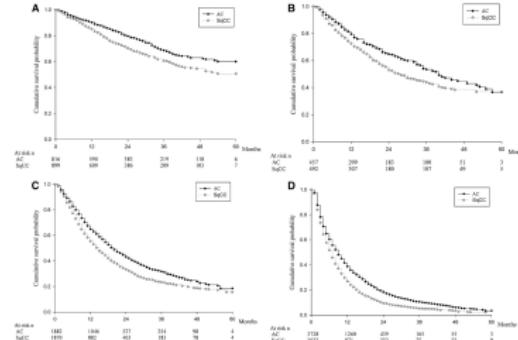
(b) LUSC

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

# LUAD vs. LUSC II



(a) All patients



(b) By cancer stages

**Figure:** Kaplan-Meiere survival curves for LUAD & LUSC (B.-Y. Wang et al., 2020)

## Findings

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

# 1. Introduction

## 1.6. Study Objectives

# Study Objectives

## Find different mutations

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

## 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 (WES) (sample n=289) + Transcriptome (WTS) (sample n=166)
- Normal + {AAH, CIS + AIS, Dysplasia, MIA} + Primary
  - Adenocarcinoma *in situ*
  - Atypical adenomatous hyperplasia
  - Carcinoma *in situ*
  - Dysplasia
  - Minimally invasive adenocarcinoma
- Adenocarcinoma (LUAD) & Squamous cell carcinoma (LUSC)
  - ① Normal → AAH → AIS → MIA → LUAD (patient n=28)
  - ② Normal → Dysplasia → CIS → LUSC (patient 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?	Stage	Number of Samples	
		Normal	AAH
Recurrence	Normal	5	8
	AAH	2	5
	CIS+AIS	2	5
	Total	20	
Non-recurrence	Normal	13	7
	AAH	7	1
	CIS+AIS	7	13
	MIA	1	
	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	
	CIS+AIS	3	
	Primary	3	
	Total	9	
Ex	Normal	41	
	Dysplasia	1	
	AAH	4	
	CIS+AIS	40	
	Primary	41	
	Total	127	
Current	Normal	33	
	Dysplasia	4	
	AAH	4	
	CIS+AIS	30	
	Primary	33	
	Total	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?	Stage	Number of Samples	
		Normal	Dysplasia
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	AIS
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	
	Primary	12	
	Total	33	

# WTS Data Composition with Smoking II

Table: LUAD WTS Data with Smoking

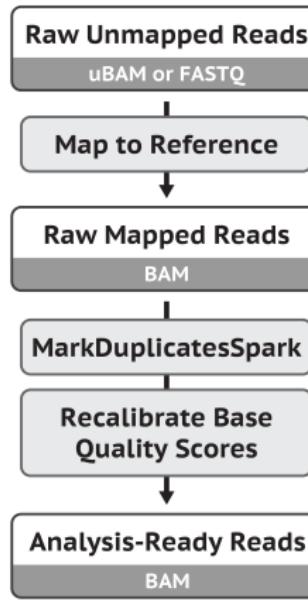
Smoking?	Stage	Number of Samples	
Never	Normal	10	
	AAH	1	
	CIS+AIS	3	
	Primary	4	
	Total	18	
Ex	Normal	3	
	CIS+AIS	1	
	Primary	1	
	Total	5	
Current	CIS+AIS	1	
	Primary	1	
	Total	2	

### 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

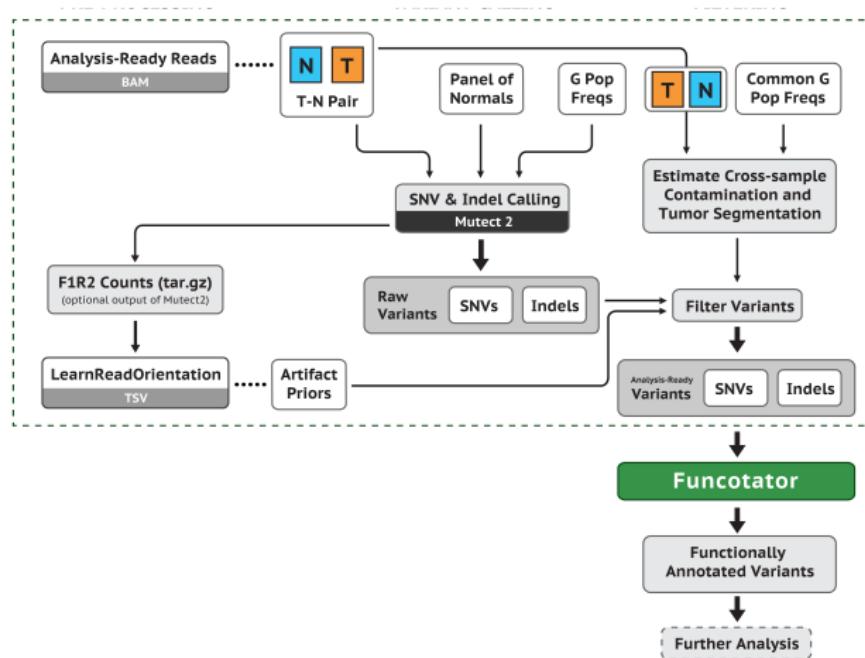
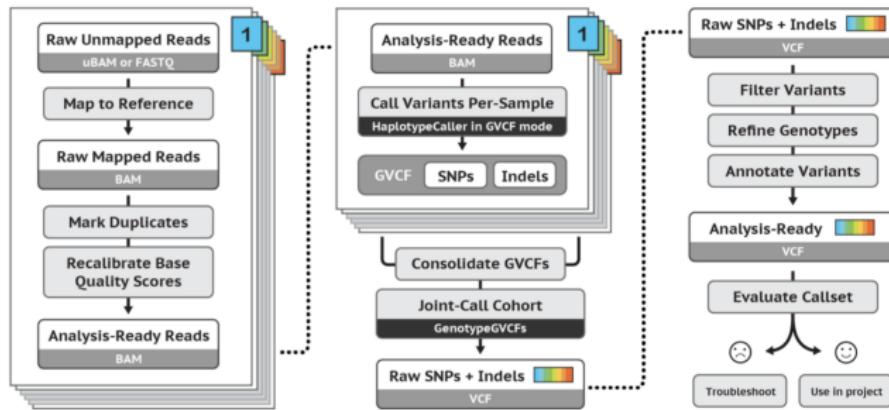


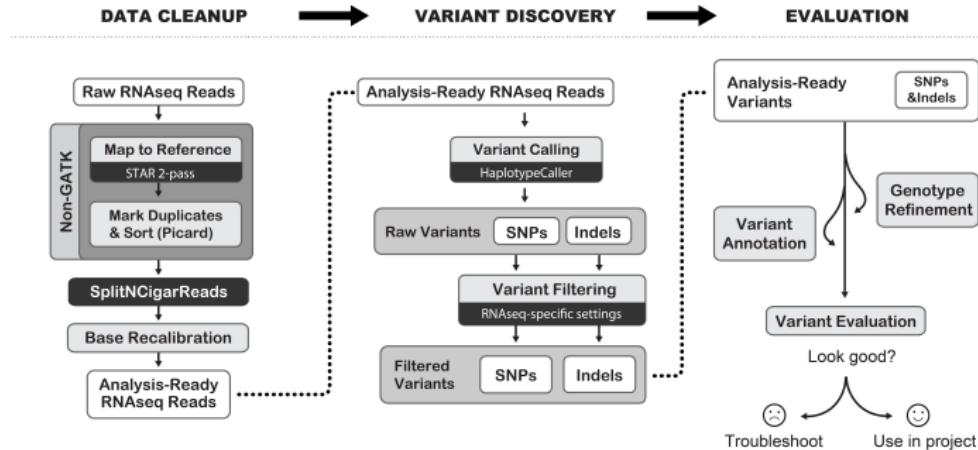
Figure: Somatic short variant (SNVs + Indels) discovery workflow (Van der Auwera et al., 2013; DePristo et al., 2011)

# 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

# FastQC?

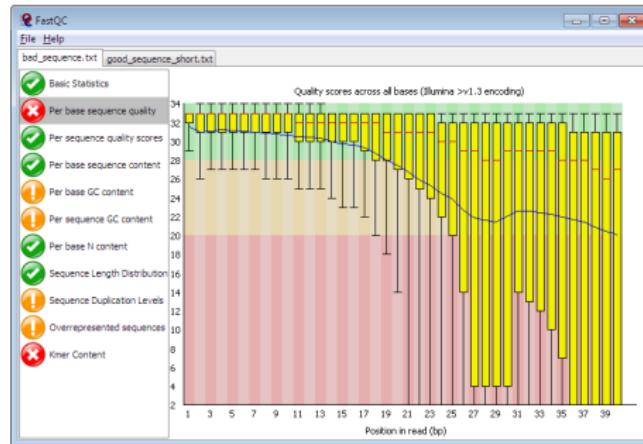


Figure: Example of FastQC Result (Andrews et al., 2012)

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

# FastQC on WES

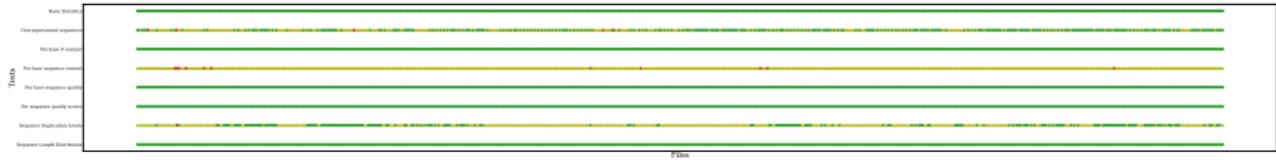
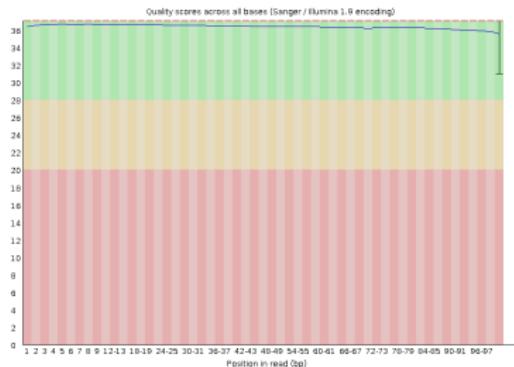


Figure: FastQC with WES data

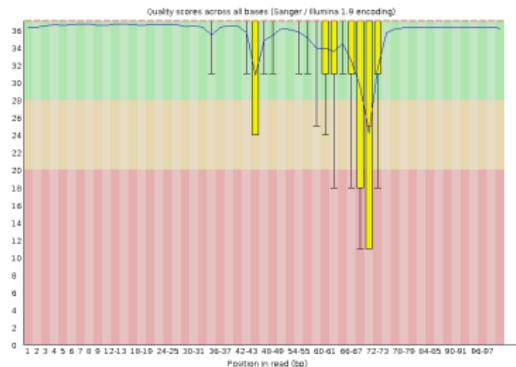
## Failure on 33P1 sample

33P1 is excluded from further analyses.

# Failure on 33P1 I



(a) 33N



(b) 33P1

Figure: Per Base Sequence Quality Results

# Failure on 33P1 II

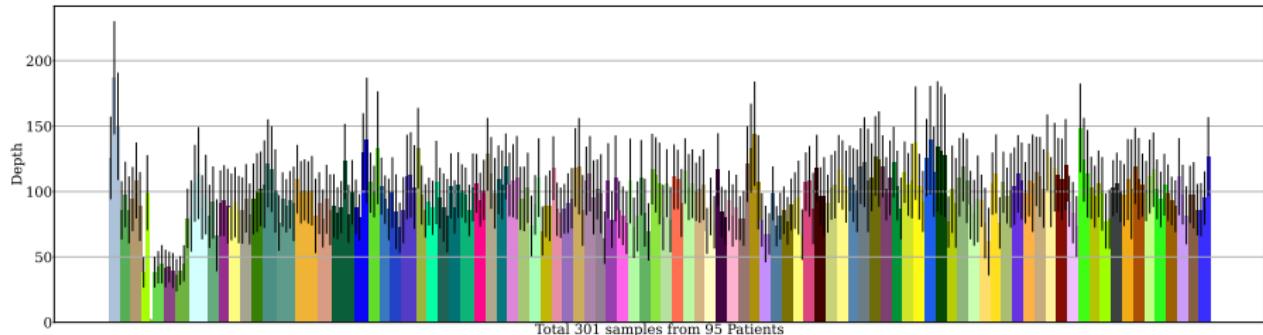


Figure: Coverage Depth Plot

- Tumor DNA:  $97.6\times$ ; Germline DNA:  $95.8\times$  in TCGA LUAD (Duhig et al., 2014)
- Mean  $121\times$ , with 83 % of target bases above  $30\times$  in TCGA LUSC (Dosenbach et al., 2007)

# FastQC on WTS

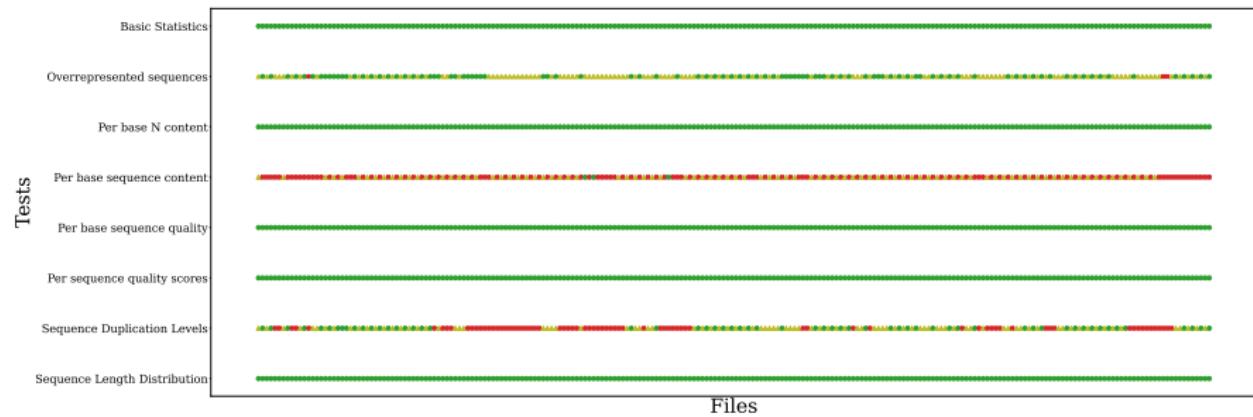


Figure: FastQC with WTS data

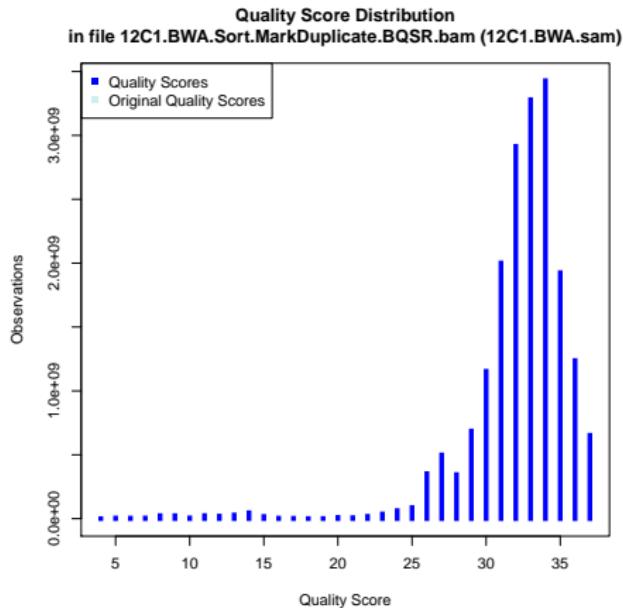
All sample are good to analysis

∴ No sample has more than 5 failures.

## 4. Results

### 4.2. Quality Checks with Picard

# Picard?



**Figure:** Quality Distribution of 12C1 sample

# Quality Distribution Plot

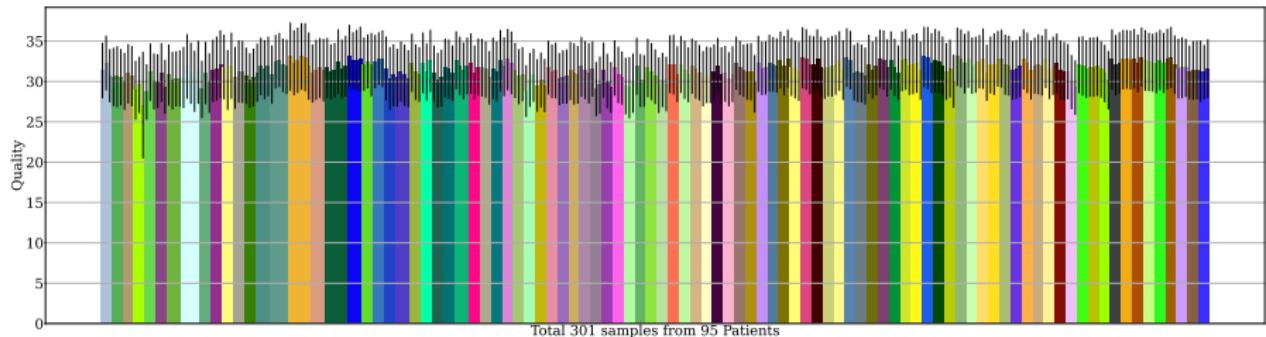


Figure: Quality Distribution by Samples

# Findings in Picard

## 4. Results

### 4.3. Copy Number Variations

# PureCN?

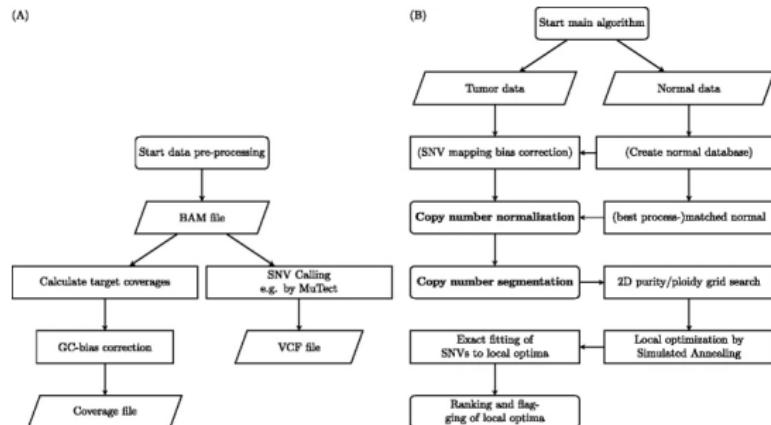
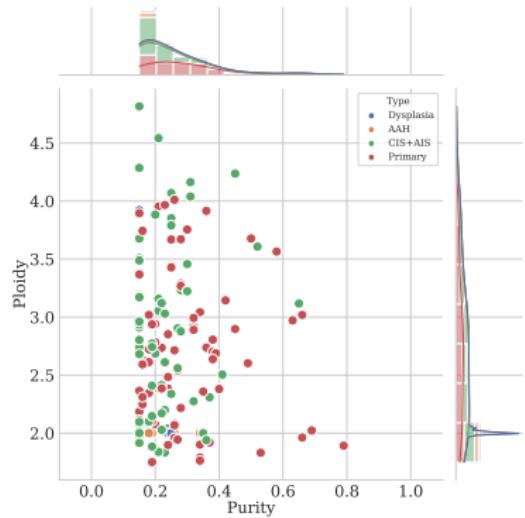
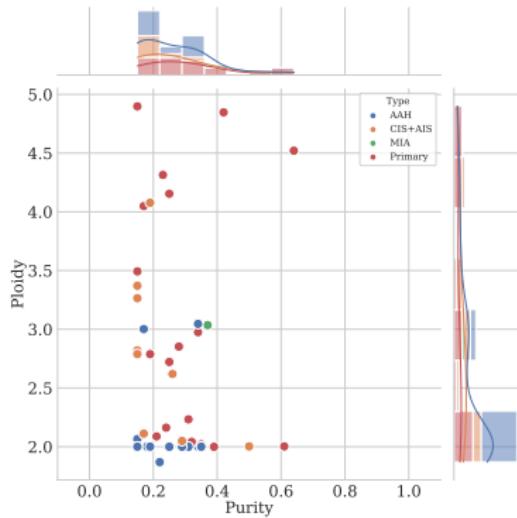


Figure: Flowchart of the PureCN data pre-processing pipeline (Riester et al., 2016)

# Purity & Ploidy on WES



(a) LUSC Samples



(b) LUAD Samples

Figure: Cellularity and Ploidy from PureCN

# LUSC in CNV Plot I

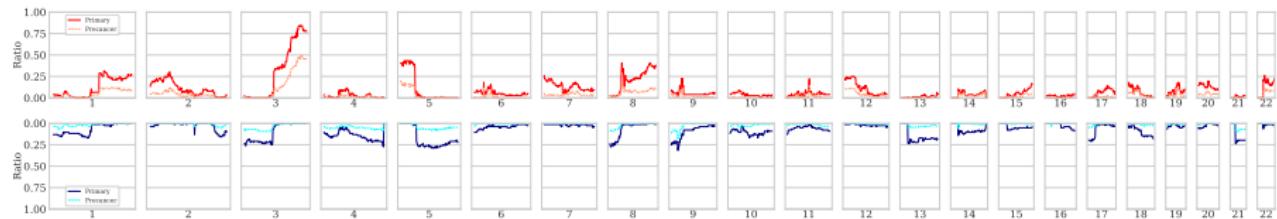


Figure: LUSC in CNV Plot

# LUSC in CNV Plot II

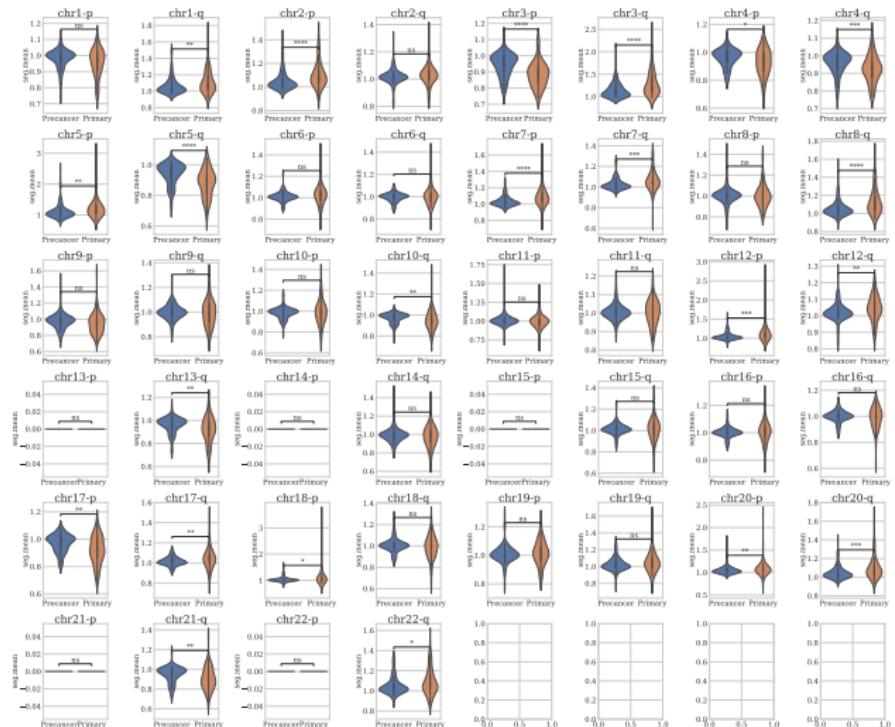


Figure: LUSC in Violin Plots

# LUSC with Recurrence in CNV Plot I

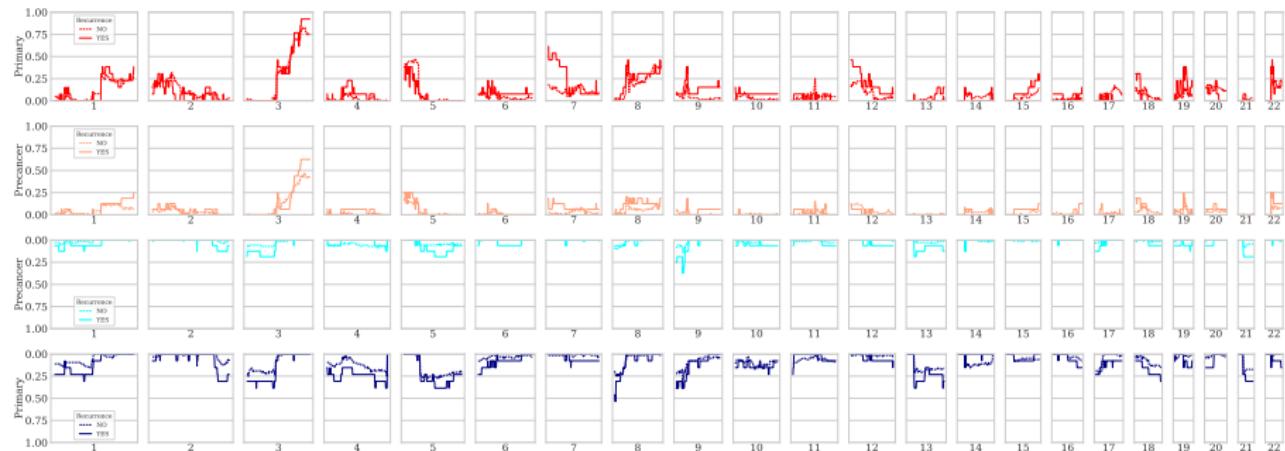


Figure: LUSC with Recurrence in CNV Plot

# LUSC with Recurrence in CNV Plot II



Figure: LUSC with Recurrence in Violin Plots

# LUSC with Smoking in CNV Plot I

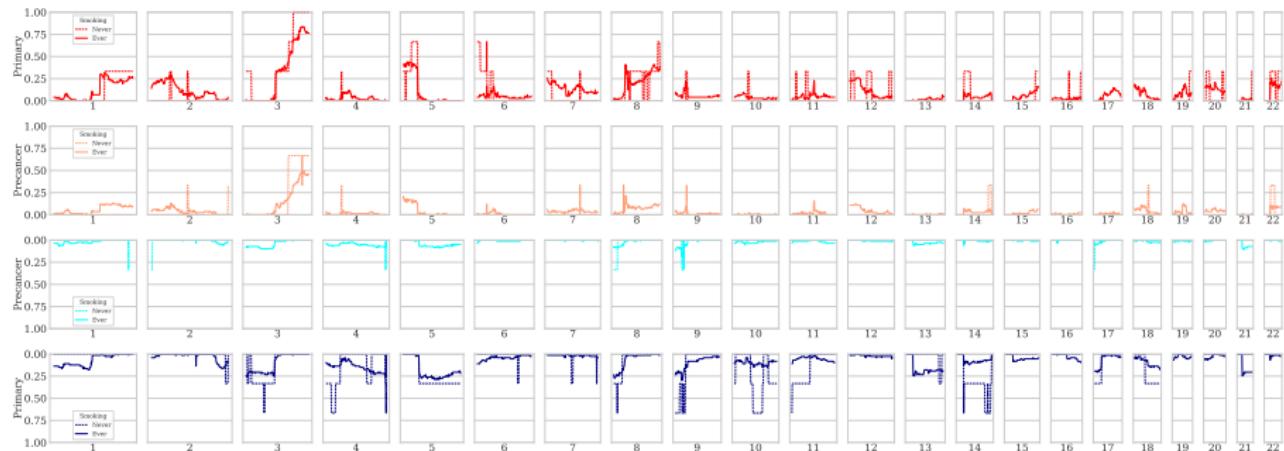


Figure: LUSC with Smoking in CNV Plot

# LUSC with Smoking in CNV Plot II

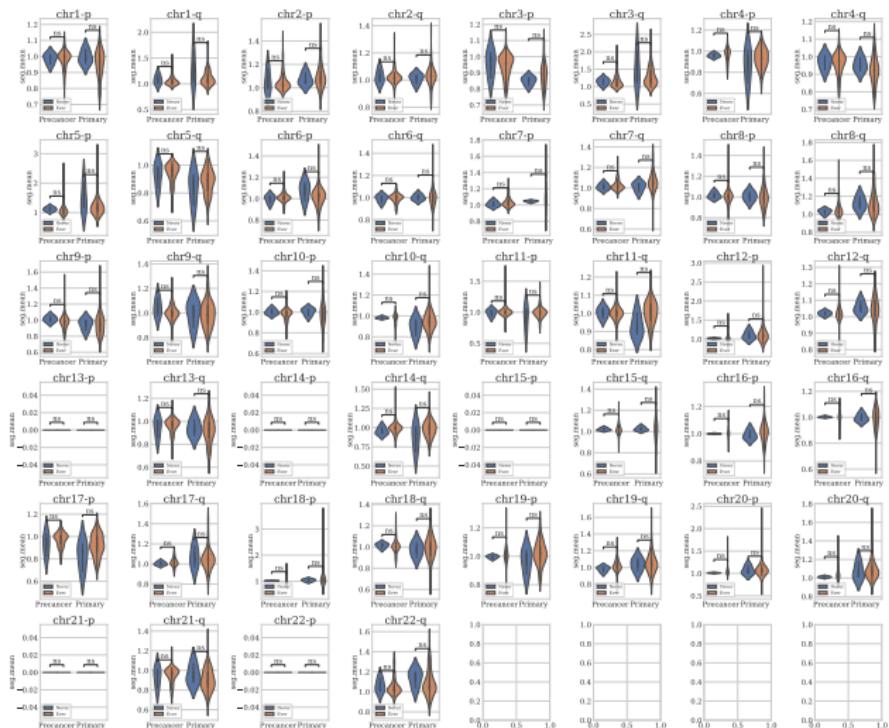


Figure: LUSC with Smoking in Violin Plots

# Findings in PureCN with LUSC I

## Chr2 p-arm gain

- Chr2 p-arm gain in Primary.

Table: CGC Tier1 genes in Chr2 p-arm

Gene Symbol (15)	Name
ALK	anaplastic lymphoma kinase (Ki-1)
BCL11A	B-cell CLL/lymphoma 11A
DCTN1	dynactin 1
DNMT3A	DNA (cytosine-5-)-methyltransferase 3 alpha
EML4	echinoderm microtubule associated protein like 4

## Chr2 q-arm loss

- ① Chr2 q-arm loss in Recurrence & Primary.

Table: CGC Tier1 genes in Chr2 q-arm

Gene Symbol (23)	Name
ACKR3	atypical chemokine receptor 3
ACSL3	acyl-CoA synthetase long-chain family member 3
ACVR1	activin A receptor, type I
ACVR2A	activin A receptor type 2A
AFF3	AF4/FMR2 family, member 3

# Findings in PureCN with LUSC III

## Chr3 p-arm loss

- ① Chr3 p-arm loss in Primary.

Table: CGC Tier1 genes in Chr3 p-arm

Gene Symbol (17)	Name
BAP1	BRCA1 associated protein-1 (ubiquitin carboxy-t...
CACNA1D	calcium channel, voltage-dependent, L type, alp...
CTNNB1	catenin (cadherin-associated protein), beta 1
FANCD2	Fanconi anemia, complementation group D2
FHIT	fragile histidine triad gene

# Findings in PureCN with LUSC IV

## Chr3 q-arm gain

- ① Chr3 q-arm gain in Primary.

Table: CGC Tier1 genes in Chr3 q-arm

Gene Symbol (21)	Name
ATR	ATR serine/threonine kinase
BCL6	B-cell CLL/lymphoma 6
CBLB	Cas-Br-M (murine) ecotropic retroviral transfor...
CNBP	CCHC-type zinc finger, nucleic acid binding pro...
EIF4A2	eukaryotic translation initiation factor 4A, is...

# Findings in PureCN with LUSC V

## Chr5 q-arm loss

### ① Chr5 q-arm loss in Primary.

Table: CGC Tier1 genes in Chr5 q-arm

Gene Symbol (15)	Name
AFF4	AF4/FMR2 family, member 4
APC	adenomatous polyposis of the colon gene
ARHGAP26	Rho GTPase activating protein 26
CD74	CD74 molecule, major histocompatibility complex...
EBF1	early B-cell factor 1

# Findings in PureCN with LUSC VI

## Chr7 p-arm gain

- ① Chr7 p-arm gain in Primary.
- ② Chr7 p-arm gain in Recurrence.

Table: CGC Tier1 genes in Chr7 p-arm

Gene Symbol (11)	Name
CARD11	caspase recruitment domain family, member 11
EGFR	epidermal growth factor receptor (erythroblasti...
ETV1	ets variant gene 1
HNRNPA2B1	heterogeneous nuclear ribonucleoprotein A2/B1
HOXA11	homeo box A11

# Findings in PureCN with LUSC VII

## Chr8 q-arm gain

### ① Chr8 q-arm gain in Primary.

Table: CGC Tier1 genes in Chr8 q-arm

Gene Symbol (16)	Name
CHCHD7	coiled-coil-helix-coiled-coil-helix domain cont...
EIF3E	eukaryotic translation initiation factor 3, sub...
EXT1	multiple exostoses type 1 gene
HEY1	hairy/enhancer-of-split related with YRPW motif 1
MYC	v-myc myelocytomatosis viral oncogene homolog (...)

# Findings in PureCN with LUSC VIII

## Chr9 q-arm gain

- Chr9 q-arm gain in Recurrence & Primary.

Table: CGC Tier1 genes in Chr9 q-arm

Gene Symbol (16)	Name
ABL1	v-abl Abelson murine leukemia viral oncogene homolog
BRD3	bromodomain containing 3
CNTRL	centriolin
FANCC	Fanconi anemia, complementation group C
GNAQ	guanine nucleotide binding protein (G protein), alpha Q subunit

# Findings in PureCN with LUSC IX

## Chr14 q-arm gain

- ① Chr14 q-arm gain in Recurrence & Primary.

Table: CGC Tier1 genes in Chr14 q-arm

Gene Symbol (18)	Name
AKT1	v-akt murine thymoma viral oncogene homolog 1
BCL11B	B-cell CLL/lymphoma 11B (CTIP2)
CCNB1IP1	cyclin B1 interacting protein 1, E3 ubiquitin p...
DICER1	dicer 1, ribonuclease type III
FOXA1	forkhead box A1

# Findings in PureCN with LUSC X

## Chr19 q-arm gain

- ① Chr19 p-arm gain in Recurrence & Primary.

Table: CGC Tier1 genes in Chr19 p-arm

Gene Symbol (19)	Name
BRD4	bromodomain containing 4
CALR	calreticulin
CRTC1	CREB regulated transcription coactivator 1
DNAJB1	DnaJ heat shock protein family (Hsp40) member B1
DNM2	dynamin 2

# LUAD in CNV Plot I

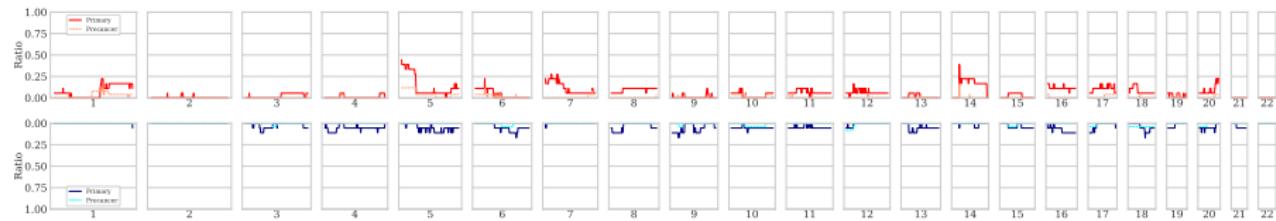


Figure: LUAD in CNV Plot

# LUAD in CNV Plot II

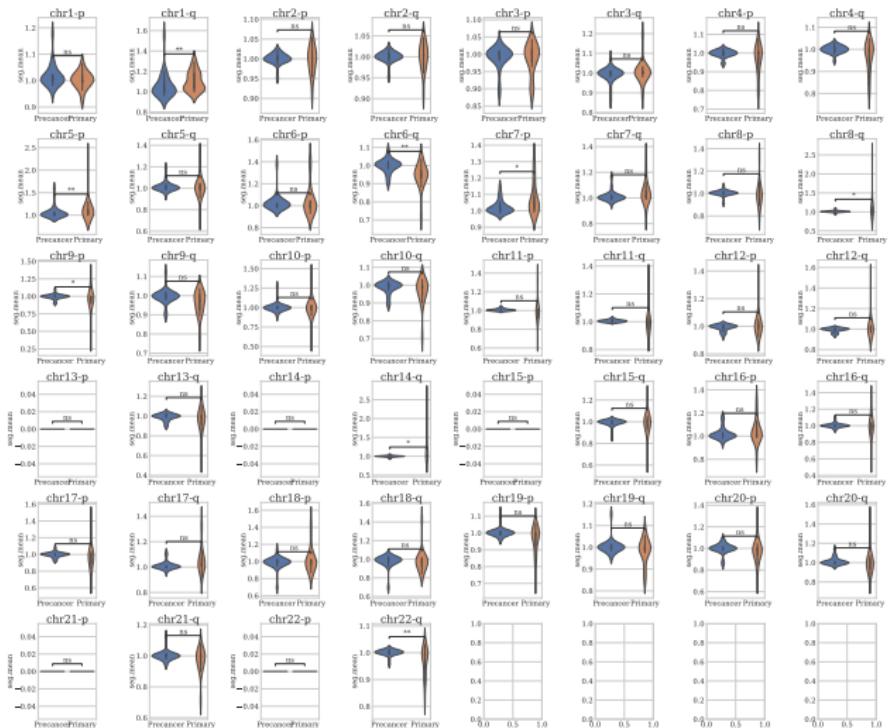


Figure: LUAD in Violin Plots

# LUAD with Recurrence in CNV Plot I

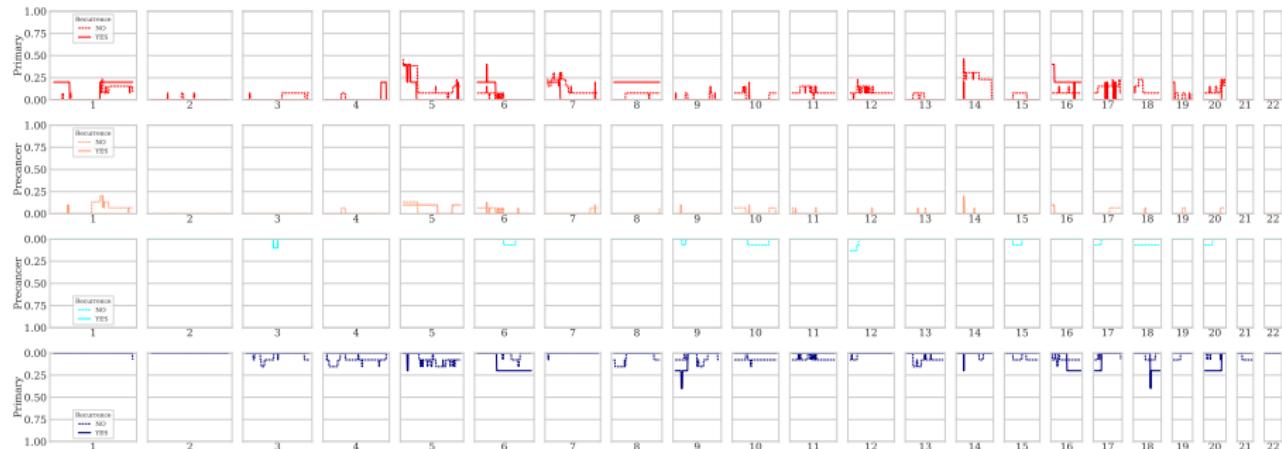


Figure: LUAD with Recurrence in CNV Plot

# LUAD with Recurrence in CNV Plot II

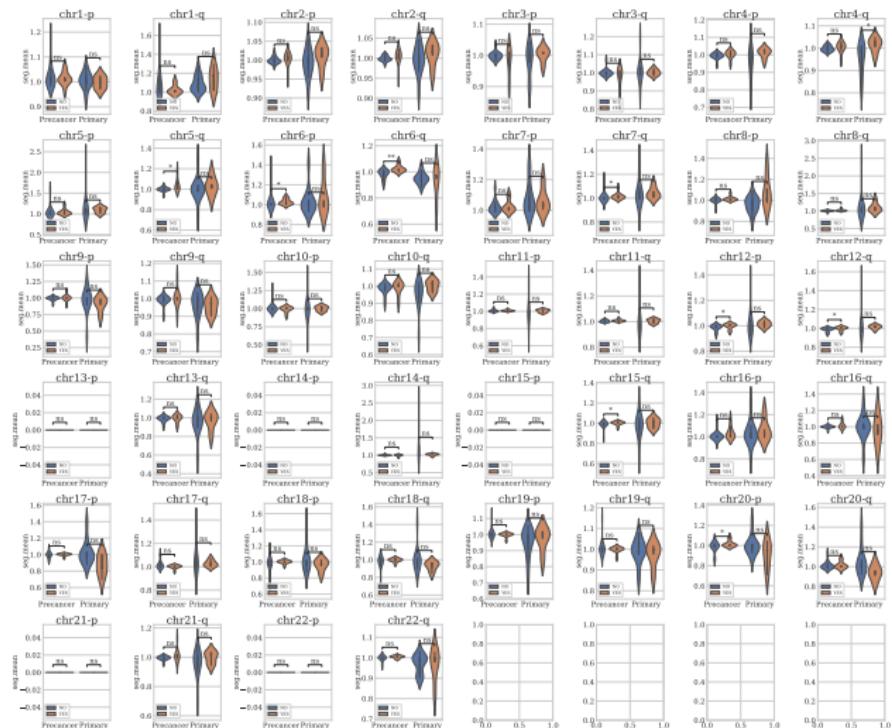


Figure: LUAD with Recurrence in Violin Plots

# LUAD with Smoking in CNV Plot I

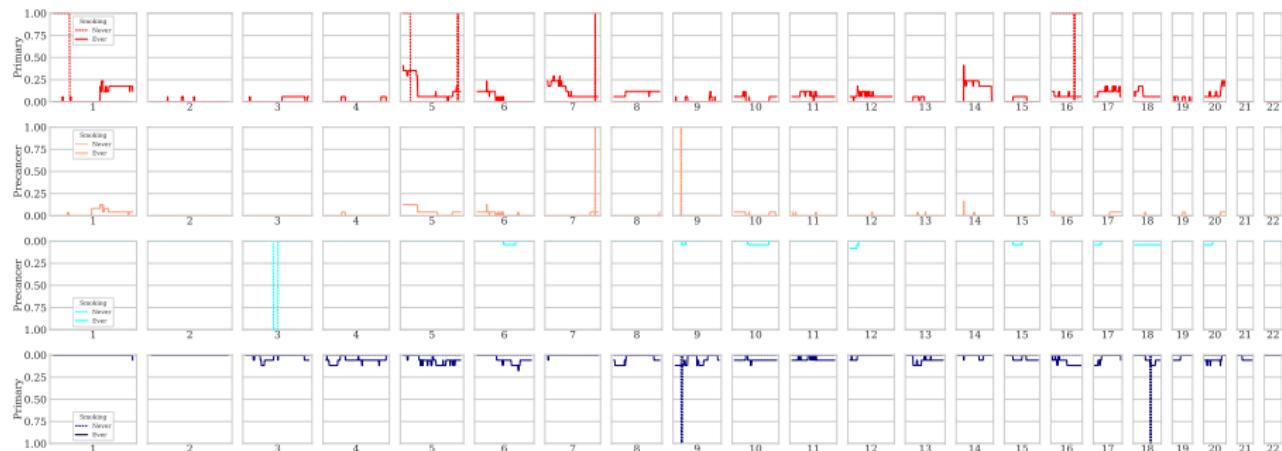


Figure: LUAD with Smoking in CNV Plot

# LUAD with Smoking in CNV Plot II

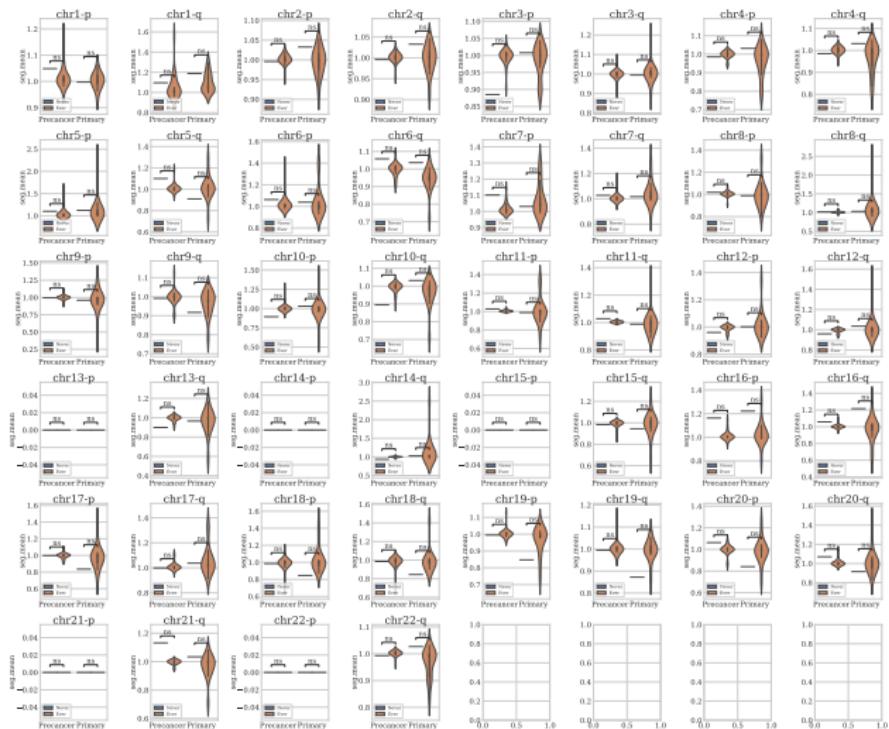


Figure: LUAD with Smoking in Violin Plots

# Findings in PureCN with LUAD I

## Chr1 q-arm gain

- ① Chr1 q-arm gain in Primary.

Table: CGC Tier1 genes in Chr1 q-arm

Gene Symbol (24)	Name
ABL2	c-abl oncogene 2, non-receptor tyrosine kinase
ARNT	aryl hydrocarbon receptor nuclear translocator
BCL9	B-cell CLL/lymphoma 9
CDC73	cell division cycle 73
DDR2	discoidin domain receptor 2

# Findings in PureCN with LUAD II

## Chr4 q-arm gain

- ① Chr5 q-arm gain in Recurrence & Primary.

Table: CGC Tier1 genes in Chr4 q-arm

Gene Symbol (13)	Name
AFF1	AF4/FMR2 family, member 1
FAT1	FAT atypical cadherin 1
FAT4	FAT atypical cadherin 4
FBXW7	F-box and WD-40 domain protein 7 (archipelago homolog)
FIP1L1	FIP1 like 1 ( <i>S. cerevisiae</i> )

# Findings in PureCN with LUAD III

## Chr5 q-arm gain

- ① Chr5 q-arm gain in Non-recurrence & Precancer.

Table: CGC Tier1 genes in Chr5 q-arm

Gene Symbol (15)	Name
AFF4	AF4/FMR2 family, member 4
APC	adenomatous polyposis of the colon gene
ARHGAP26	Rho GTPase activating protein 26
CD74	CD74 molecule, major histocompatibility complex...
EBF1	early B-cell factor 1

# Findings in PureCN with LUAD IV

## Chr6 p-arm gain

- ① Chr6 p-arm gain in Non-recurrence & Precancer.

Table: CGC Tier1 genes in Chr6 p-arm

Gene Symbol (16)	Name
CCND3	cyclin D3
DAXX	death-domain associated protein
DEK	DEK oncogene (DNA binding)
FANCE	Fanconi anemia, complementation group E
HIST1H3B	histone cluster 1, H3b

# Findings in PureCN with LUAD V

## Chr6 q-arm gain

- ① Chr6 q-arm gain in Non-recurrence & Precancer.

Table: CGC Tier1 genes in Chr6 q-arm

Gene Symbol (15)	Name
AFDN	myeloid/lymphoid or mixed-lineage leukemia (tri...)
ARID1B	AT rich interactive domain 1B
ESR1	estrogen receptor 1
EZR	ezrin
FGFR1OP	FGFR1 oncogene partner (FOP)

# Findings in PureCN with LUAD VI

## Chr6 q-arm loss

- ① Chr6 q-arm loss in Primary.

Table: CGC Tier1 genes in Chr6 q-arm

Gene Symbol (15)	Name
AFDN	myeloid/lymphoid or mixed-lineage leukemia (tri...)
ARID1B	AT rich interactive domain 1B
ESR1	estrogen receptor 1
EZR	ezrin
FGFR1OP	FGFR1 oncogene partner (FOP)

# Findings in PureCN with LUAD VII

## Chr7 q-arm gain

- ① Chr7 q-arm gain in Non-recurrence & Precancer.

Table: CGC Tier1 genes in Chr7 q-arm

Gene Symbol (14)	Name
BRAF	v-raf murine sarcoma viral oncogene homolog B1
CDK6	cyclin-dependent kinase 6
CREB3L2	cAMP responsive element binding protein 3-like 2
CUX1	cut-like homeobox 1
EZH2	enhancer of zeste homolog 2

# Findings in PureCN with LUAD VIII

## Chr12 p-arm gain

- ① Chr12 p-arm loss in Non-recurrence & Primary.

Table: CGC Tier1 genes in Chr12 p-arm

Gene Symbol (10)	Name
CCND2	cyclin D2
CDKN1B	cyclin-dependent kinase inhibitor 1B (p27, Kip1)
CHD4	chromodomain helicase DNA binding protein 4
ERC1	ELKS/RAB6-interacting/CAST family member 1
ETNK1	ethanolamine kinase 1

# Findings in PureCN with LUAD IX

## Chr12 q-arm gain

- ① Chr12 q-arm loss in Non-recurrence & Primary.

Table: CGC Tier1 genes in Chr12 q-arm

Gene Symbol (26)	Name
ARID2	AT rich interactive domain 2
ATF1	activating transcription factor 1
BCL7A	B-cell CLL/lymphoma 7A
BTG1	B-cell translocation gene 1, anti-proliferative
CDK4	cyclin-dependent kinase 4

# Findings in PureCN with LUAD X

## Chr22 q-arm loss

- ① Chr22 q-arm loss in Primary.

Table: CGC Tier1 genes in Chr22 q-arm

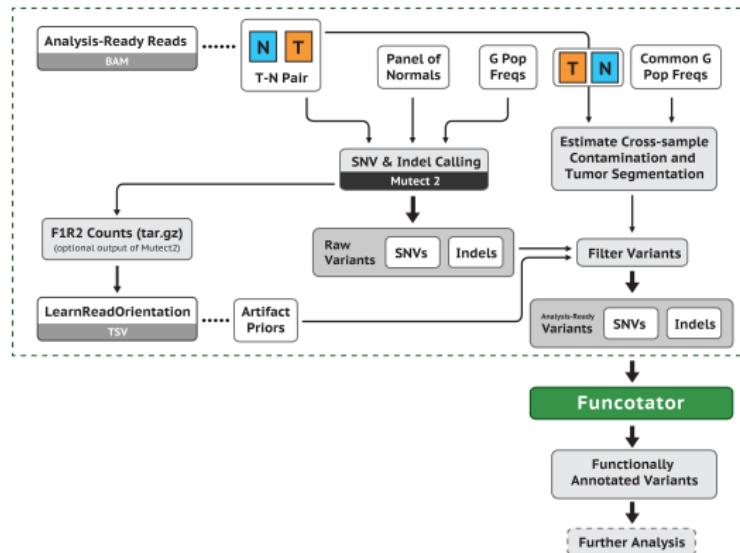
Gene Symbol (15)	Name
APOBEC3B	apolipoprotein B mRNA editing enzyme catalytic ...
BCR	breakpoint cluster region
CHEK2	CHK2 checkpoint homolog (S. pombe)
CLTCL1	clathrin, heavy polypeptide-like 1
EP300	300 kd E1A-Binding protein gene

# Findings in PureCN

## 4. Results

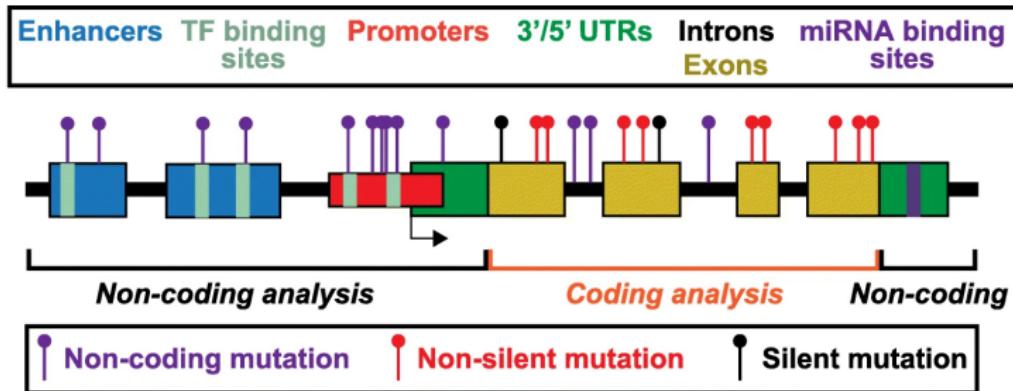
### 4.4. Single Nucleotide Variations Analysis

# 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?

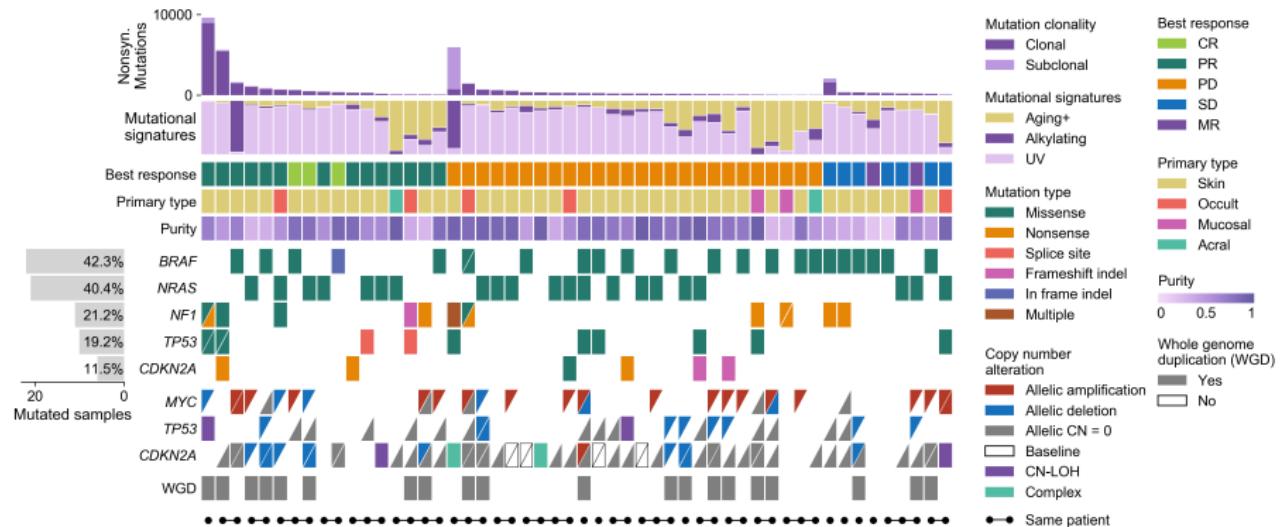


Figure: A comutation plot generated with CoMut (Crowdis et al., 2020)

# 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

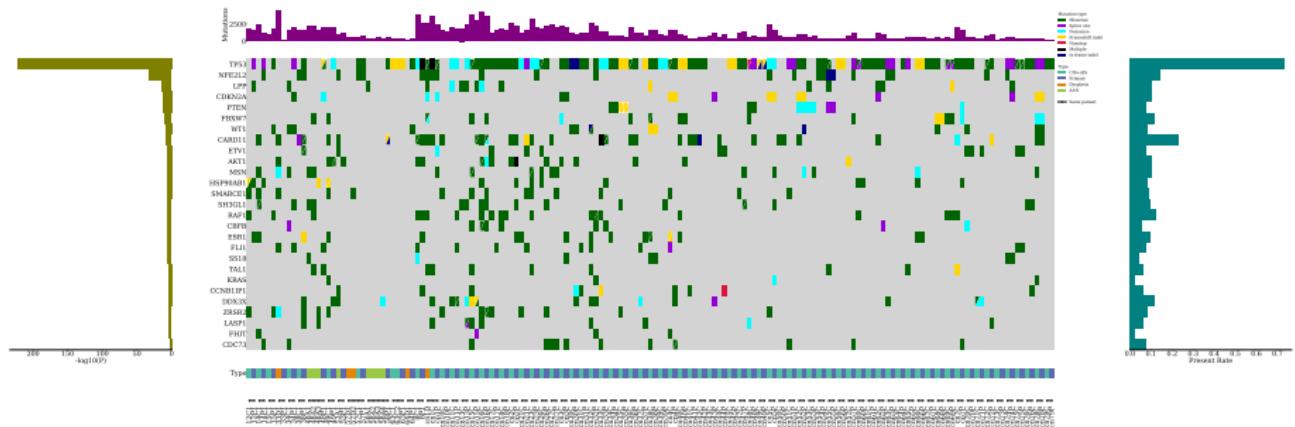


Figure: CoMut Plot with LUSC Patients

# Somatic Variant in LUSC with Recurrence

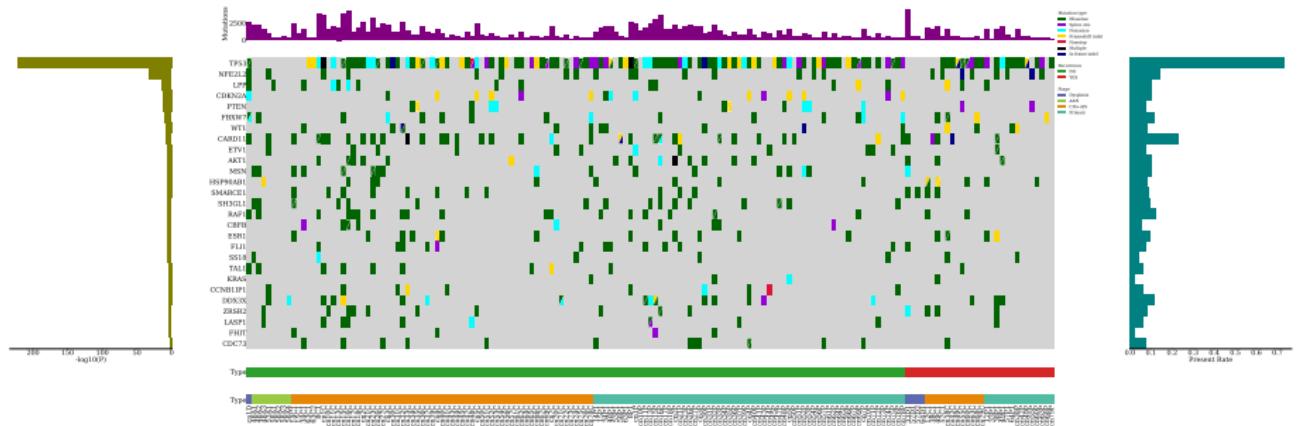


Figure: CoMut Plot in LUSC Patients with Recurrence

# Somatic Variant in LUAD

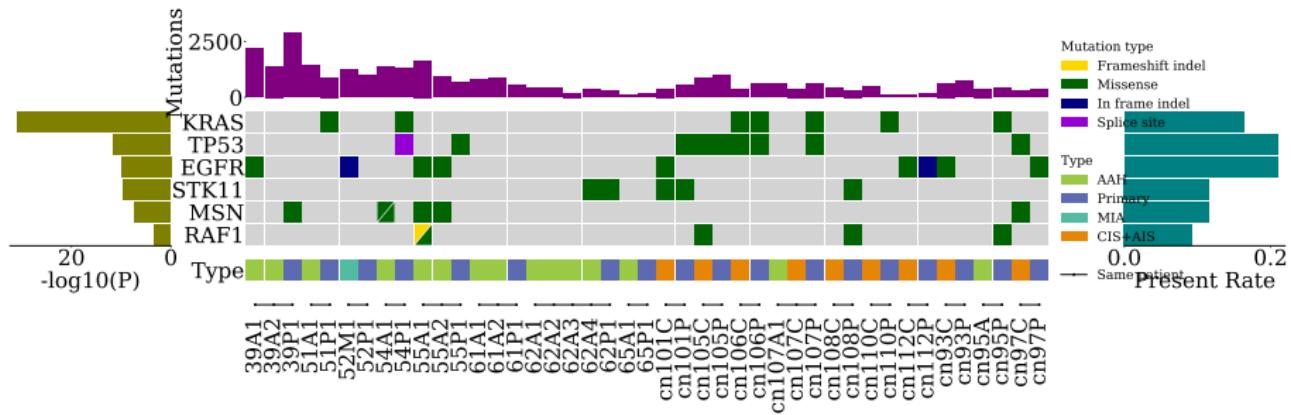


Figure: CoMut Plot with LUAD Patients

# Somatic Variant in LUAD with Recurrence

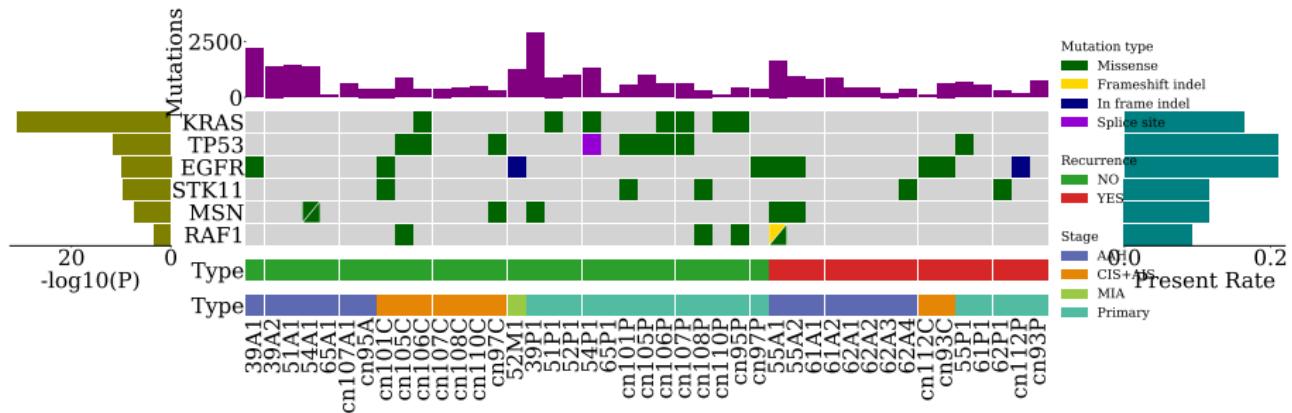


Figure: CoMut Plot in LUAD Patients with Recurrence

# 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

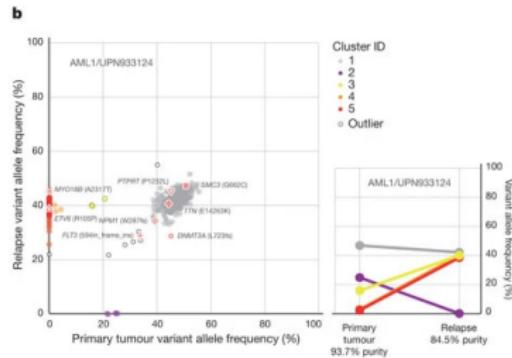
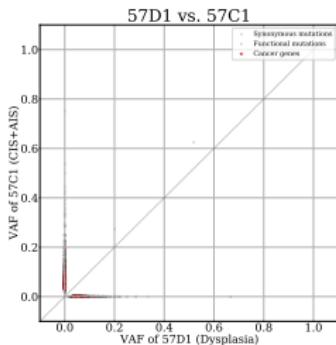
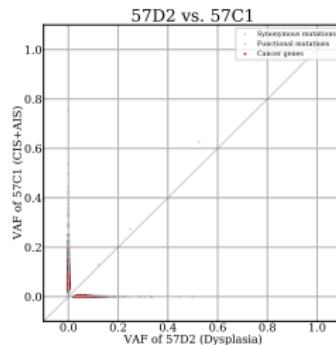


Figure: VAF distribution of validated mutations (Ding et al., 2012)

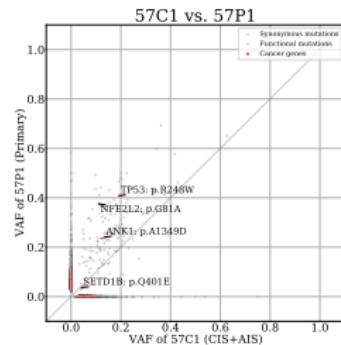
# VAF Plots



(a) Dysplasia + CIS



(b) Dysplasia + CIS



(c) CIS + Primary

Figure: VAF plots in patient #57

# PyClone?

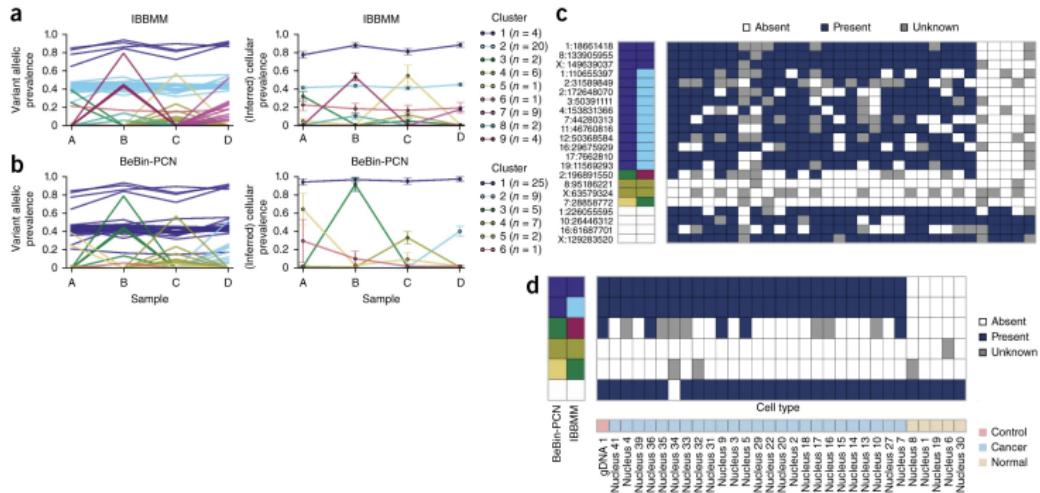
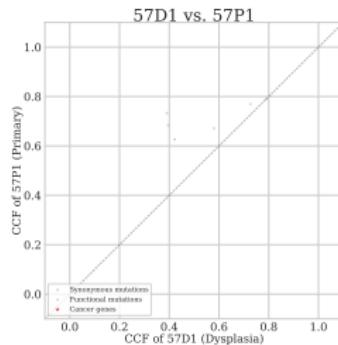
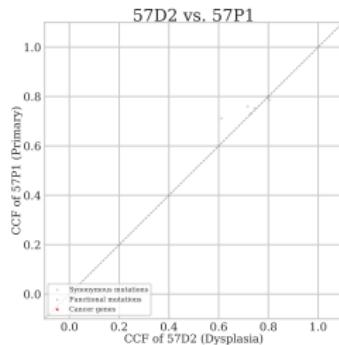


Figure: Analysis of multiple samples by PyClone (Roth et al., 2014)

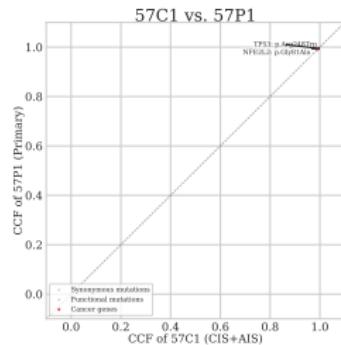
# PyClone Plots I



(a) 57D1



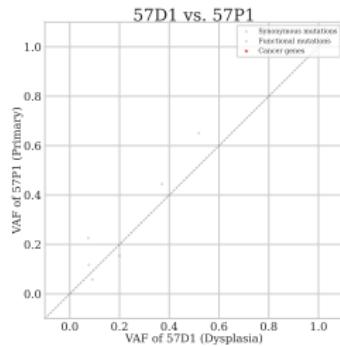
(b) 57D2



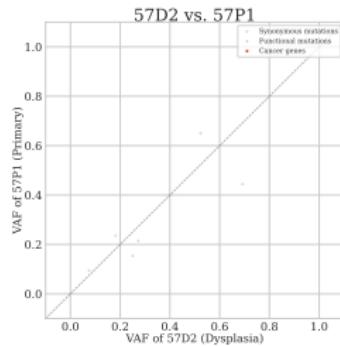
(c) 57C1

Figure: CCF plot in patient #57

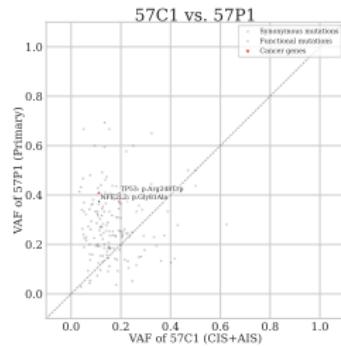
# PyClone Plots II



(a) 57D1



(b) 57D2



(c) 57C1

Figure: VAF plot in patient #57

# Findings in VAF Analysis

## 4. Results

### 4.6. Differences in Gene Expression Levels

# RSEM?

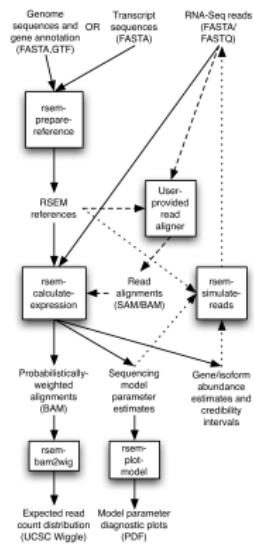


Figure: RSEM workflow (Li & Dewey, 2011)

# DESeq2?

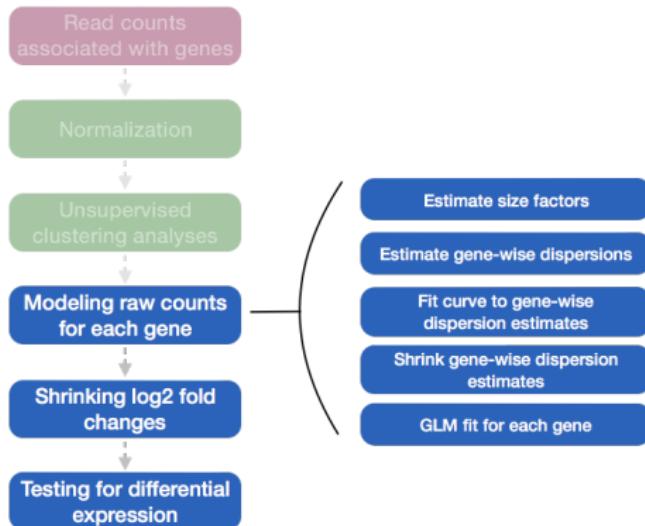


Figure: DESeq2 workflow (Love, Huber, & Anders, 2014)

# 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?

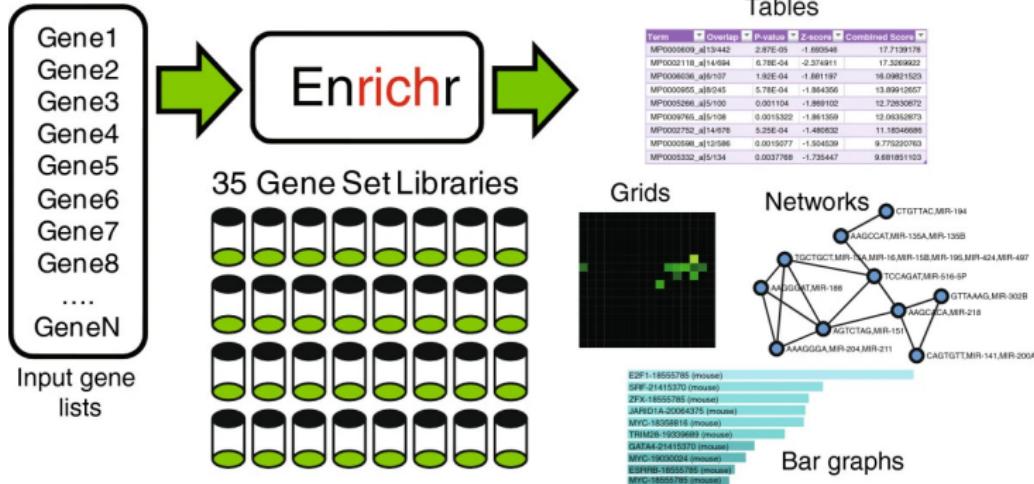


Figure: Enrichr workflow (E. Y. Chen et al., 2013; Kuleshov et al., 2016)

# Gene-set Library

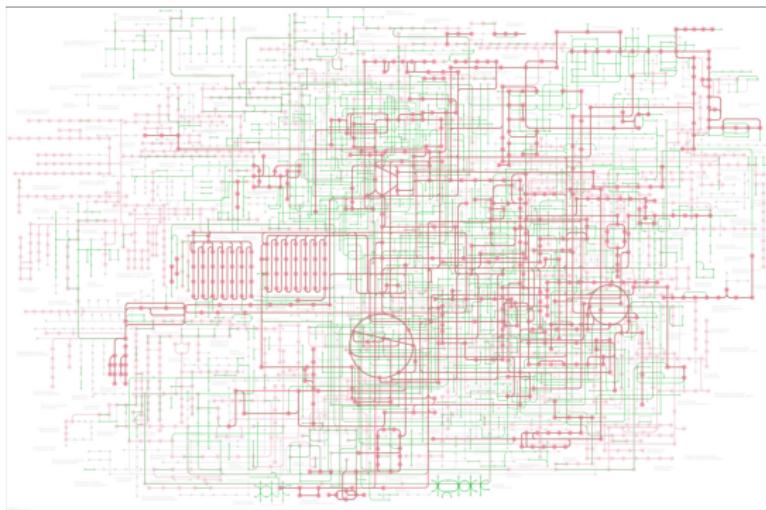


Figure: The global map of metabolic pathways by KEGG (Kanehisa et al., 2021)

KEGG

KEGG 2021 Human

# WTS Data Composition

Table: Number of WTS samples

Cancer Subtype	Stage	Number of Samples	
		Normal	Dysplasia
LUSC	Normal	17	
	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?	Stage	Number of Samples	
		Normal	AIS
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	
	Primary	12	
	Total	33	

# 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	

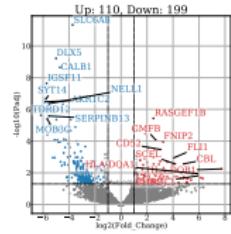
## 4. Results

### 4.6. Differences in Gene Expression Levels

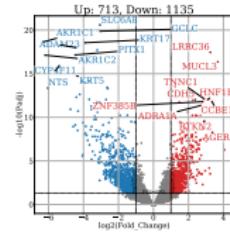
#### 4.6.1. Comparing cancer stage in LUSC

# DEG Volcano Plots in LUSC

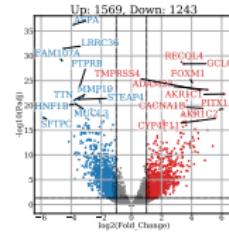
Normal → Dysplasia → CIS → Primary (LUSC)



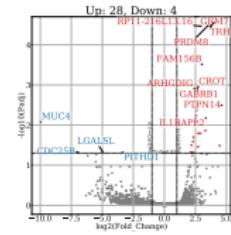
(a) Normal-Dysplasia



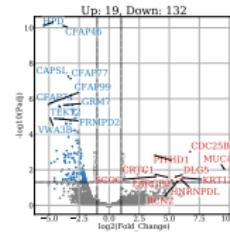
(b) Normal-CIS



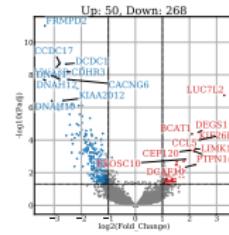
(c) Normal-Primary



(d) Dysplasia-CIS



(e) Dysplasia-Primary

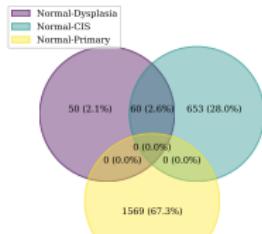


(f) CIS-Primary

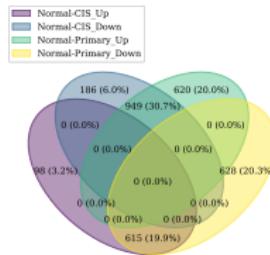
Figure: DEG Volcano Plots in LUSC

# DEG Venn Diagram in LUSC

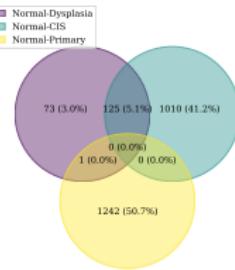
Normal → Dysplasia → CIS → Primary (LUSC)



(a) Up-regulated



(b) Both



(c) Down-regulated

Figure: DEG Venn Diagram in LUSC

# Enrichment test with Normal vs. Dysplasia in LUSC

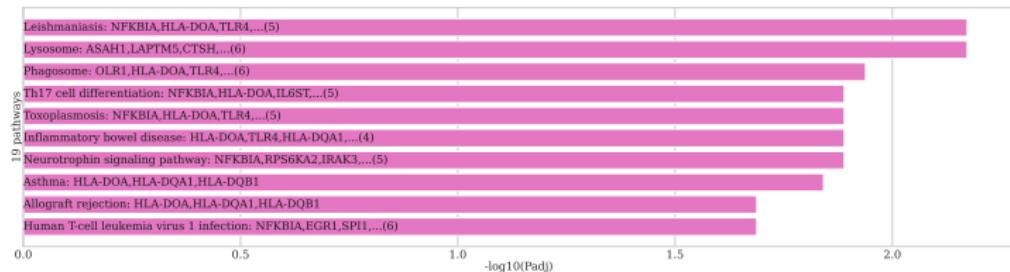


Figure: Up-regulated Pathways on Normal vs. Dysplasia

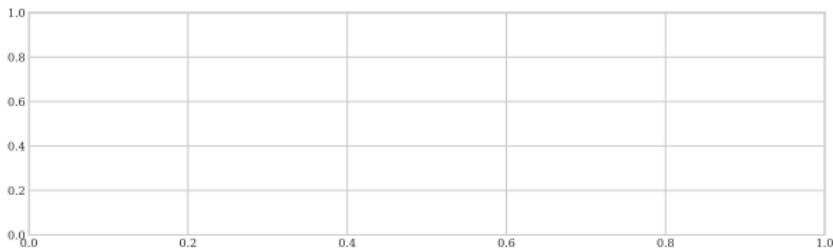


Figure: Down-regulated Pathways on Normal vs. Dysplasia

# Enrichment test with Normal vs. CIS in LUSC

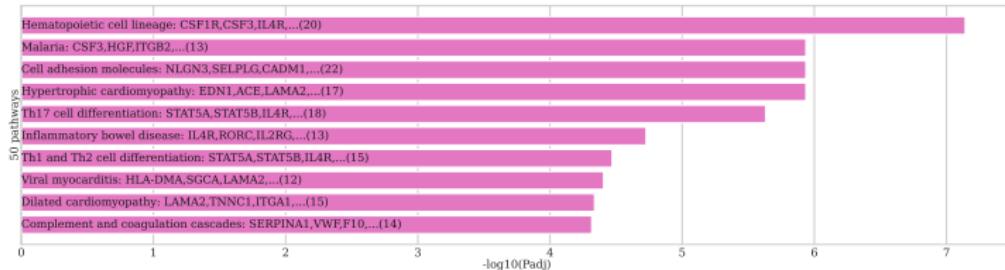


Figure: Up-regulated Pathways on Normal vs. CIS

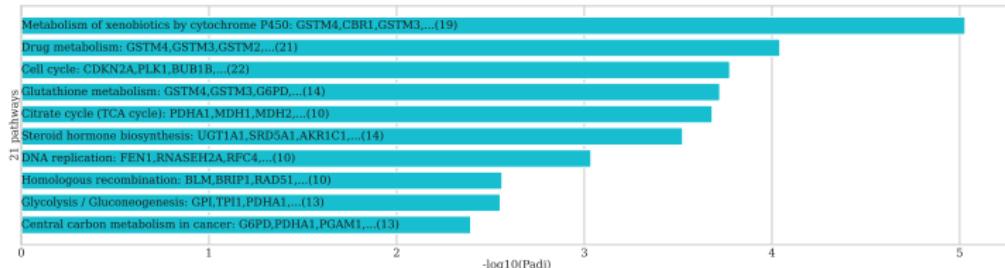


Figure: Down-regulated Pathways on Normal vs. CIS

# Enrichment test with Normal vs. Primary in LUSC

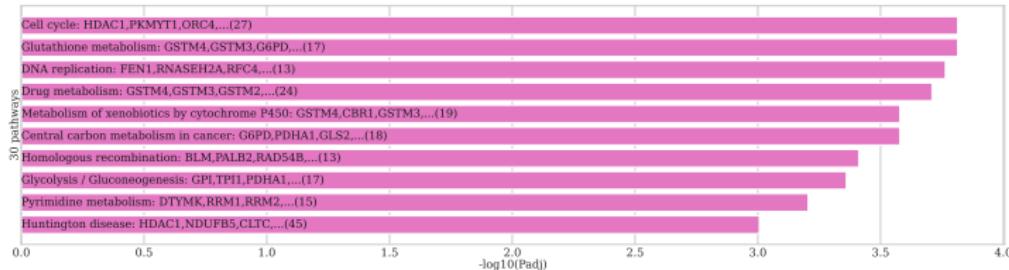


Figure: Up-regulated Pathways on Normal vs. Primary

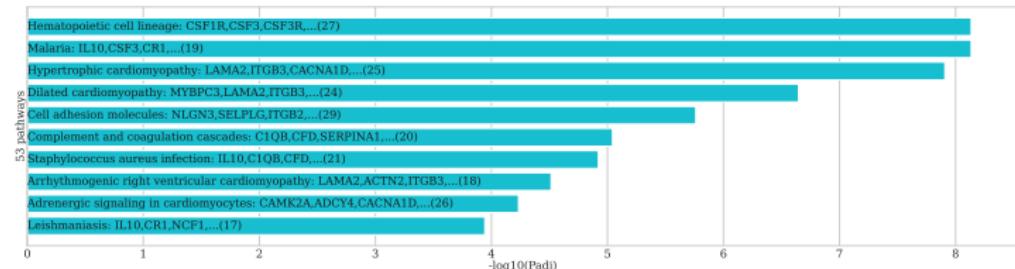


Figure: Down-regulated Pathways on Normal vs. Primary

# Findings in Comparing cancer stage in LUSC

## AKR1C1 & AKR1C2

- ① Down-regulated in CIS, but up-regulated in Primary.
- ② Regulate steroids (Jin et al., 2009) and hormones (Penning et al., 2000).
- ③ Promote the metastasis of NSCLC (Z. Hong et al., 2018).

## SFTPC

- ① Down-regulate in Primary than Normal.
- ② A pulmonary surfactant associated protein (Z. Lin et al., 2018).
- ③ SFTPC  $\downarrow \Rightarrow$  Poor survival in LUAD (Li et al., 2019).
- ④ Associated with lung disease in adult (Henderson et al., 2013) and baby (Brasch et al., 2004).

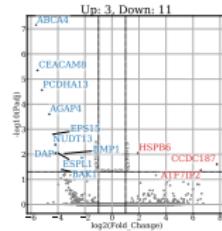
## 4. Results

### 4.6. Differences in Gene Expression Levels

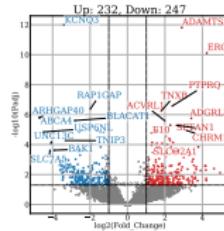
#### 4.6.2. Comparing cancer stage in LUAD

# DEG Volcano Plots in LUAD

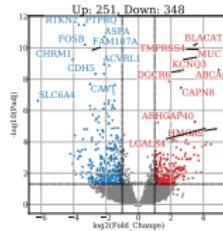
Normal → AAH → AIS → Primary (LUAD)



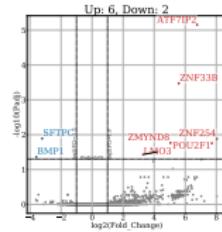
(a) Normal-AAH



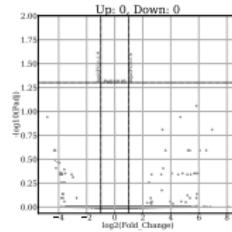
(b) Normal-AIS



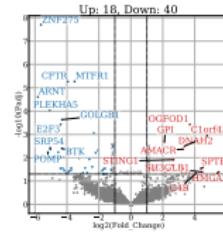
(c) Normal-Primary



(d) AAH-AIS



(e) AAH-Primary

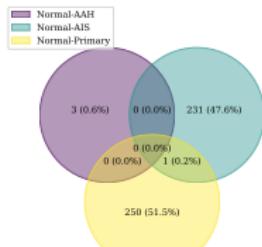


(f) AIS-Primary

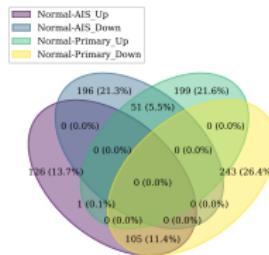
Figure: DEG Volcano Plots in LUAD

# DEG Venn Diagram in LUAD

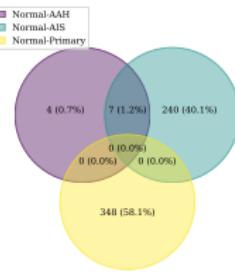
Normal → AAH → AIS → Primary (LUAD)



(a) Up-regulated



(b) Both



(c) Down-regulated

Figure: DEG Venn Diagram in LUAD

# Enrichment test with Normal vs. AAH in LUAD

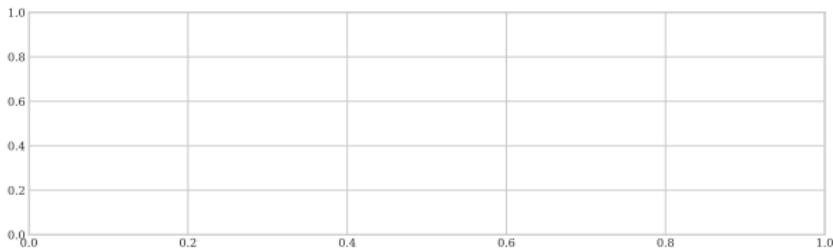


Figure: Up-regulated Pathways on Normal vs. AAH

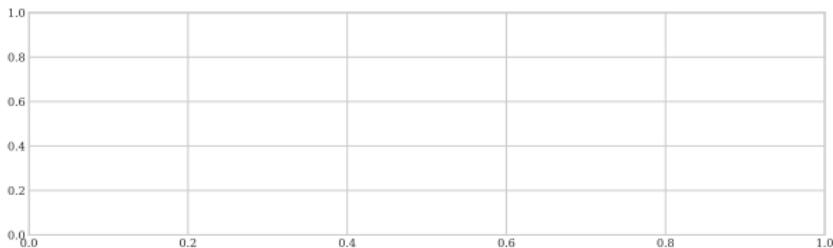


Figure: Down-regulated Pathways on Normal vs. AAH

# Enrichment test with Normal vs. AIS in LUAD

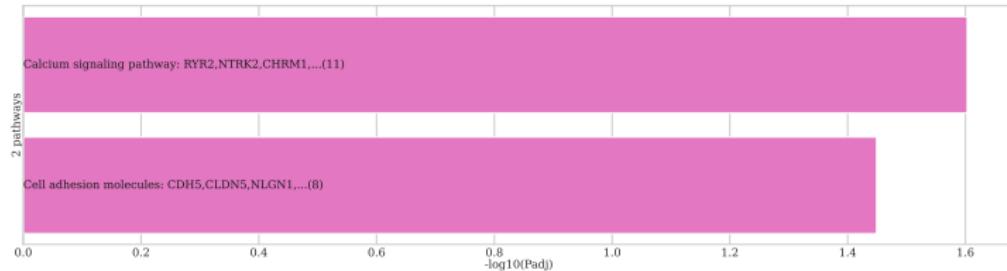


Figure: Up-regulated Pathways on Normal vs. AIS

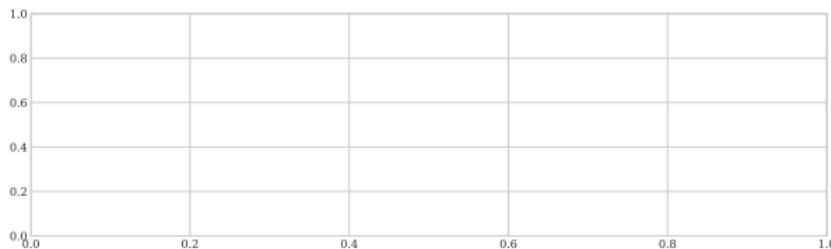


Figure: Down-regulated Pathways on Normal vs. AIS

# Enrichment test with Normal vs. Primary in LUAD

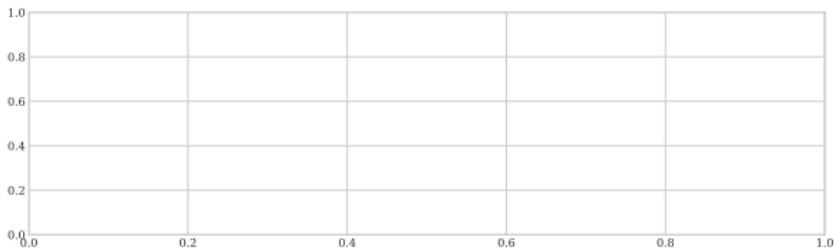


Figure: Up-regulated Pathways on Normal vs. Primary

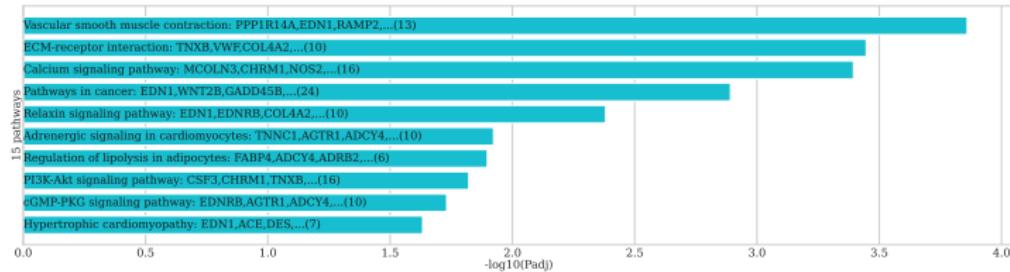


Figure: Down-regulated Pathways on Normal vs. Primary

# Finding in Comparing cancer stage in LUAD I

## 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 diseases (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

### 4.6. Differences in Gene Expression Levels

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

# LUSC Data Composition

Table: Number of WTS LUSC samples

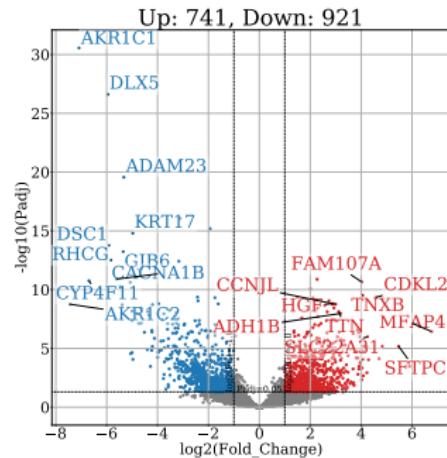
Recurrence?	Stage	Number of Samples	
		Normal	Dysplasia
Recurrence	Normal	1	
	Dysplasia		1
	CIS+AIS	5	
	Primary	6	
	Total	13	
Non-recurrence	Normal	16	
	Dysplasia		1
	CIS+AIS	29	
	Primary	30	
	Total	76	

## Pooled normal samples

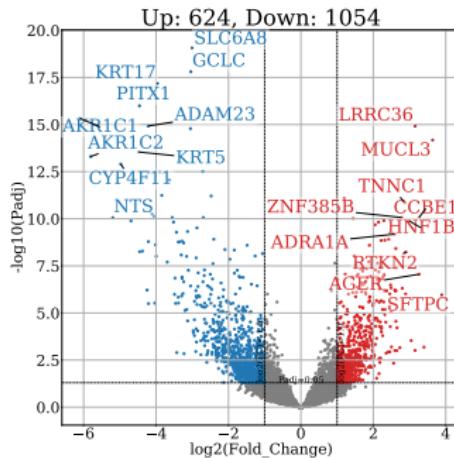
In order to compare with Normal stage, merging Normal samples.

∴ Insufficient number of Normal samples in Recur.

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



(a) Recur

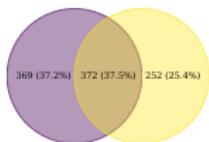


(b) Non-recur

Figure: DEG Volcanot Plot with CIS in LUSC

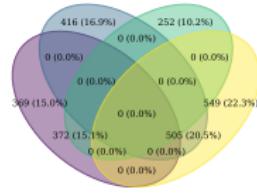
# DEG Venn Diagram for R vs. NR with CIS in LUSC

Recur  
Non-recur



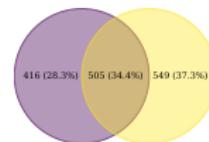
(a) Up-regulated

Recur\_Up  
Recur\_Down  
Non-recur\_Up  
Non-recur\_Down



(b) Both

Recur  
Non-recur



(c) Down-regulated

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

# Enrichment test for Recur-specific with CIS in LUSC

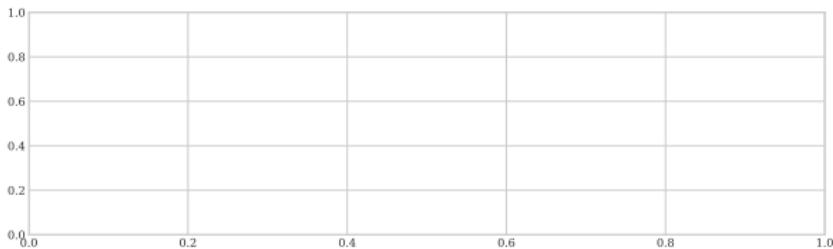


Figure: Up-regulated Pathways for Recur-specific

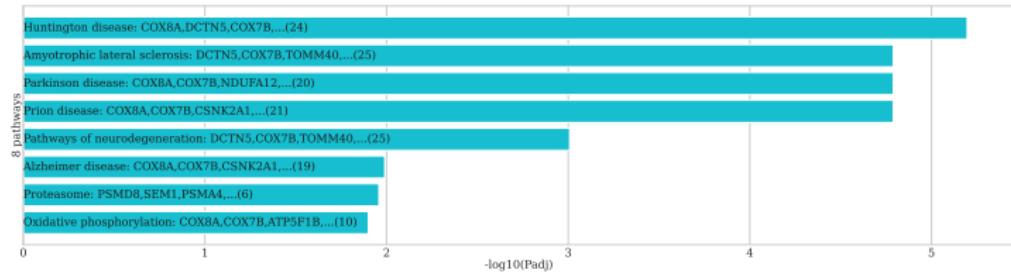


Figure: Down-regulated Pathways for Recur-specific

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

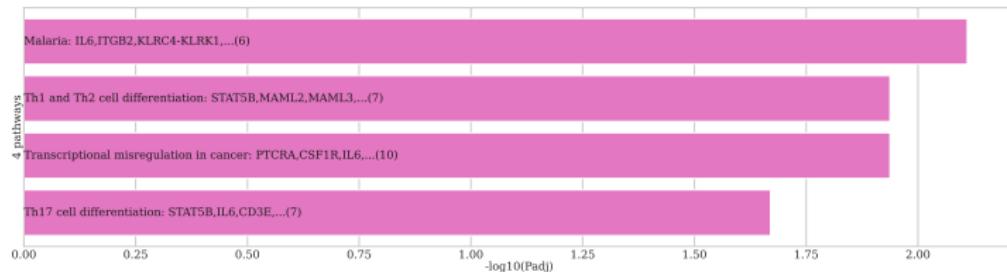


Figure: Up-regulated Pathways for Non-recr-specific

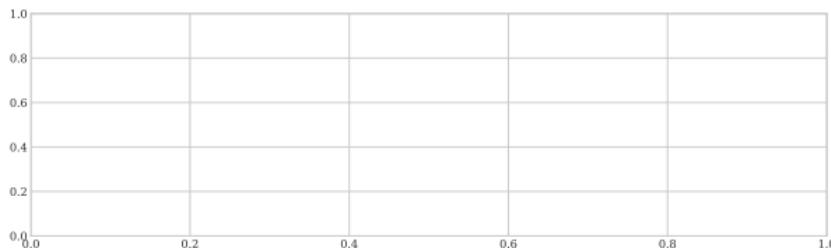


Figure: Down-regulated Pathways for Non-recr-specific

# Enrichment test for Intersected with CIS in LUSC

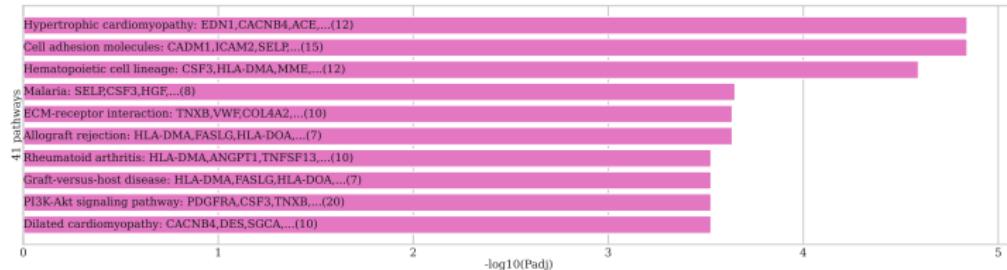


Figure: Up-regulated Pathways for Intersected

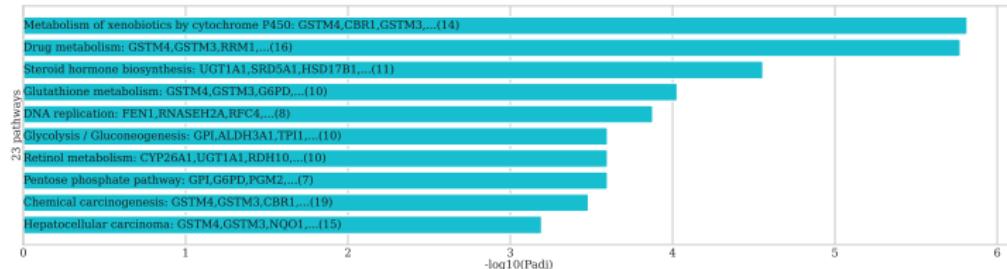
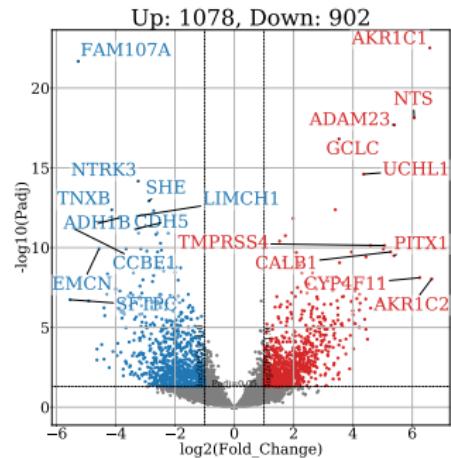
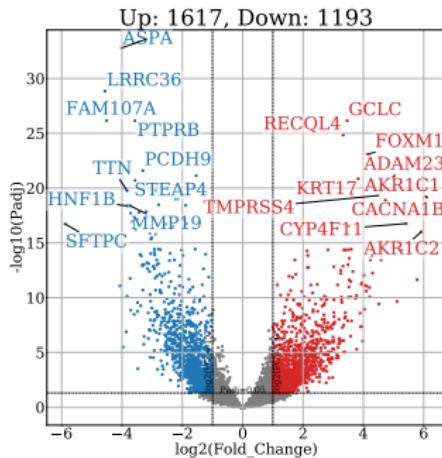


Figure: Down-regulated Pathways for Intersected

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



(a) Recur

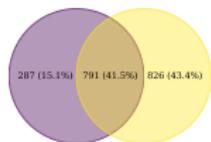


(b) Non-recur

Figure: DEG Volcanot Plot with Primary in LUSC

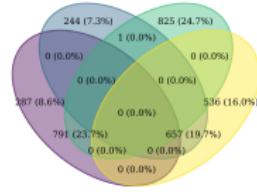
# DEG Venn Diagram for R vs. NR with Primary in LUSC

Recur  
Non-recur



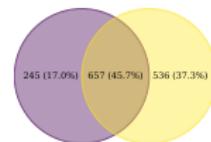
(a) Up-regulated

Recur\_Up  
Recur\_Down  
Non-recur\_Up  
Non-recur\_Down



(b) Both

Recur  
Non-recur



(c) Down-regulated

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

# Enrichment test for Recur-specific with Primary in LUSC

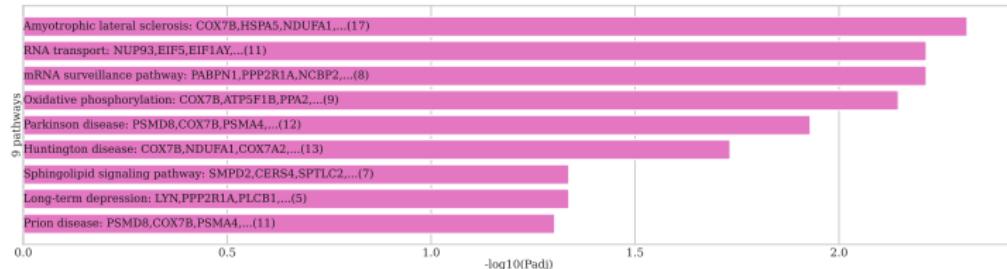


Figure: Up-regulated Pathways for Recur-specific

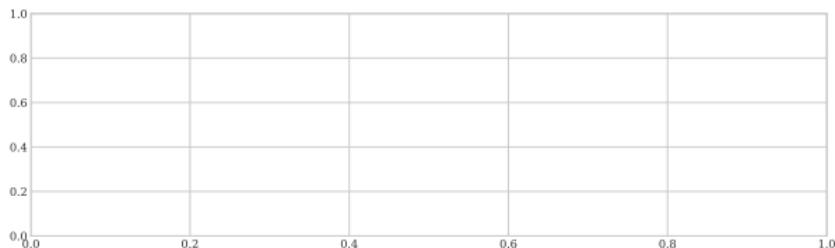


Figure: Down-regulated Pathways for Recur-specific

# Enrichment test for NR-specific with Primary in LUSC

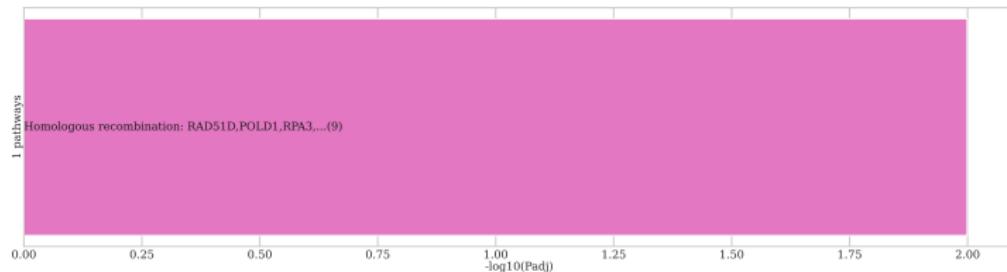


Figure: Up-regulated Pathways for Non-recur-specific

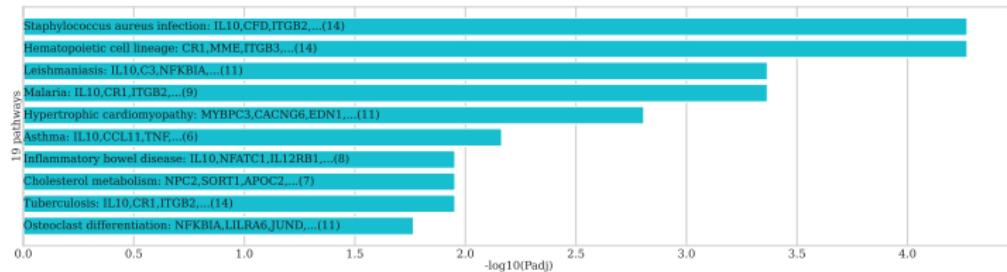


Figure: Down-regulated Pathways for Non-recur-specific

# Enrichment test for Intersected with Primary in LUSC

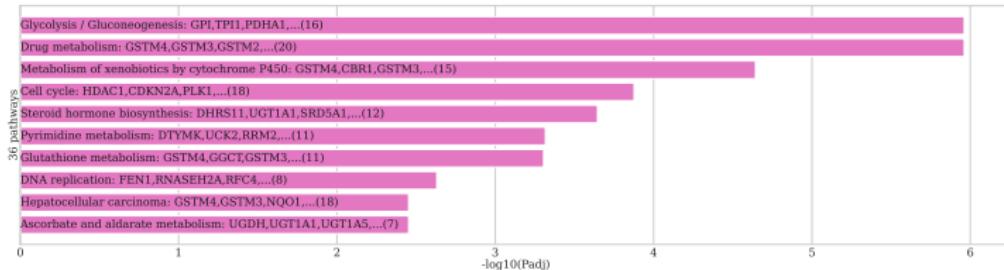


Figure: Up-regulated Pathways for Intersected

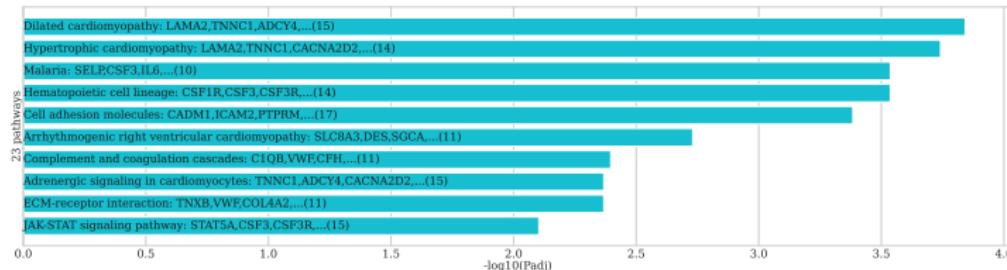


Figure: Down-regulated Pathways for Intersected

# 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

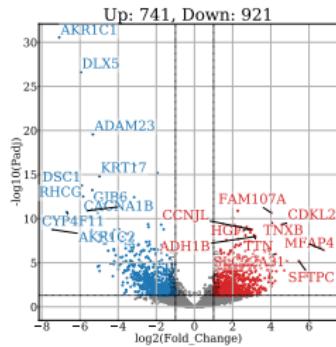
- ① Highly up-regulated in Non-recur patients.
- ② DNA-dependent ATPase (Yin, Kwon, Varshavsky, & Wang, 2004)
- ③ RECQL4 modulate chromosome segregation (Yin et al., 2004)
- ④ RECQL5 promotes metastasis & resistance in NSCLC (Xia, Zhang, Yuan, & Niu, 2021)
- ⑤ RECQL4 ↑ ⇒ Poor prognosis in breast cancer (Zhu et al., 2018)
  - ① Overall survival
  - ② Distant metastasis-free survival
  - ③ Relapse-free survival

## 4. Results

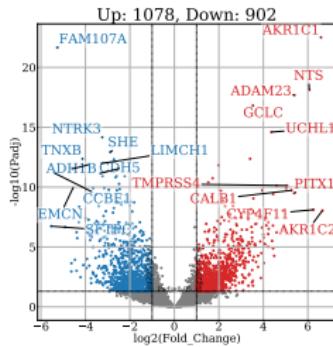
### 4.6. Differences in Gene Expression Levels

#### 4.6.4. Within Recur in LUSC

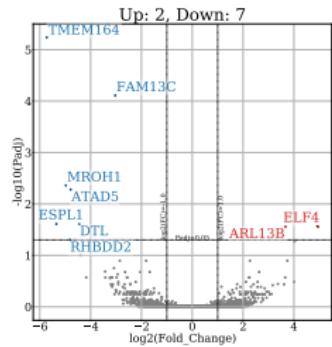
# DEG Volcano Plots with Recur in LUSC



(a) Normal-CIS



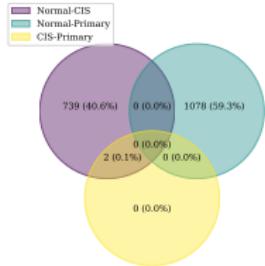
(b) Normal-Primary



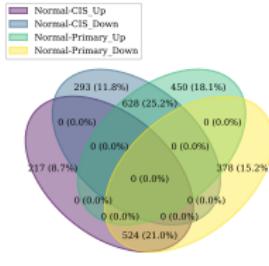
(c) CIS-Primary

Figure: DEG Volcano Plots with Recur samples in LUSC

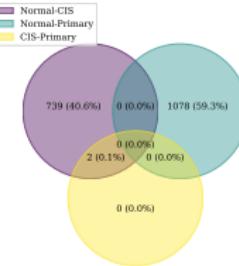
# DEG Venn Diagram with Recur in LUSC



(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

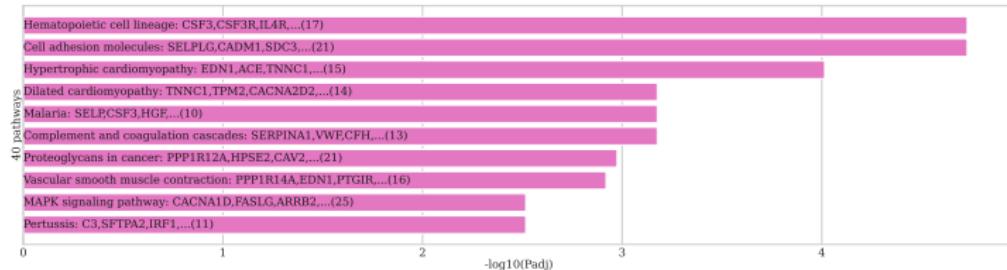


Figure: Up-regulated Pathways on Normal vs. CIS for Recur in LUSC

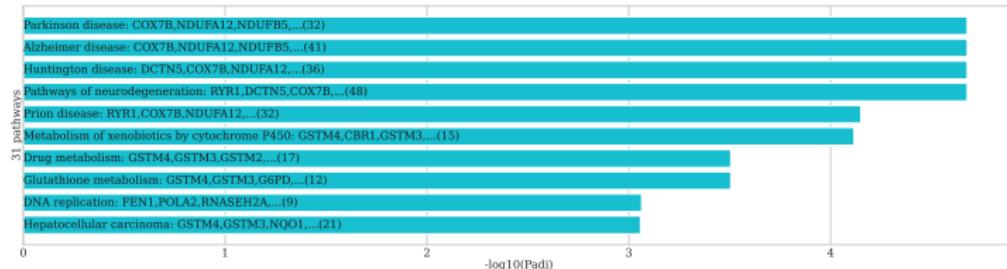


Figure: Down-regulated Pathways on Normal vs. CIS for Recur in LUSC

# Enrichment test with Normal vs. Primary for Recur

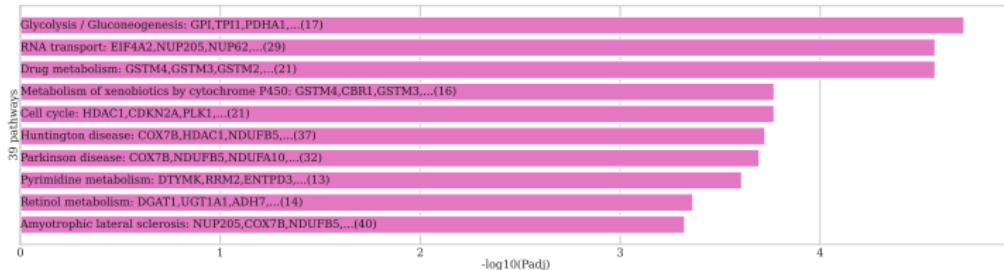


Figure: Up-regulated Pathways on Normal vs. Primary for Recur in LUSC

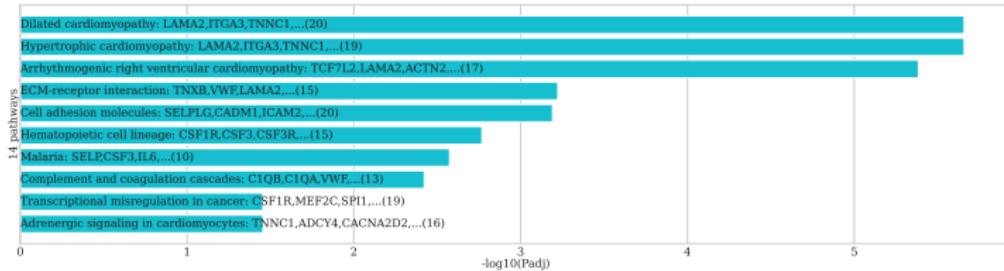


Figure: Down-regulated Pathways on Normal vs. Primary for Recur in LUSC

# Finding in Comparing within Recur in LUSC I

## AKR1C1

- ① Down-regulated in CIS, but up-regulated in Primary.
- ② Regulate steroids (Jin et al., 2009) and hormones (Penning et al., 2000).
- ③ Promote the metastasis of NSCLC (Z. Hong et al., 2018)

## ADAM23

- ① Down-regulated in CIS, but up-regulated in Primary.
- ② Play a role in cell-cell and cell-matrix interactions (Cal, Freije, López, Takada, & Lopez-Otin, 2000)
- ③ Suppresses metastasis in lung carcinoma cells (Ota et al., 2016)
- ④ ADAM protein was lower in NSCLC than in normal tissue & benign pulmonary lesions (Hu et al., 2011)

## FAM107A

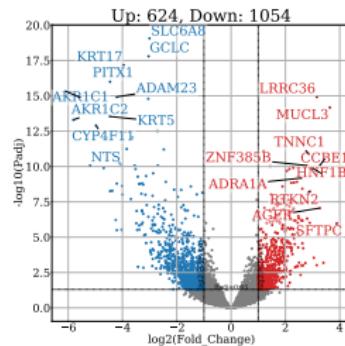
- ① Up-regulated in CIS, but down-regulated in Primary.
- ② May play a role in tumor development (L. Wang et al., 2000)
- ③ Negatively regulates focal adhesion assembly (Le et al., 2010)

## 4. Results

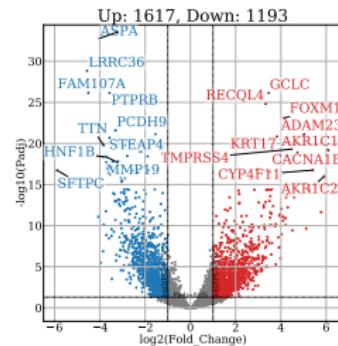
### 4.6. Differences in Gene Expression Levels

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

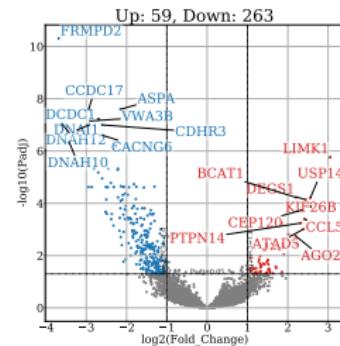
# DEG Volcano Plots with Non-recr in LUSC



(a) Normal-CIS



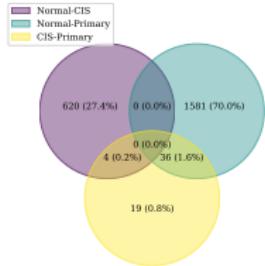
(b) Normal-Primary



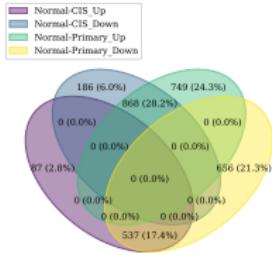
(c) CIS-Primary

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

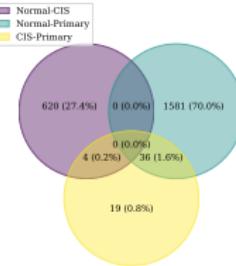
# DEG Venn Diagram with Non-recur in LUSC



(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

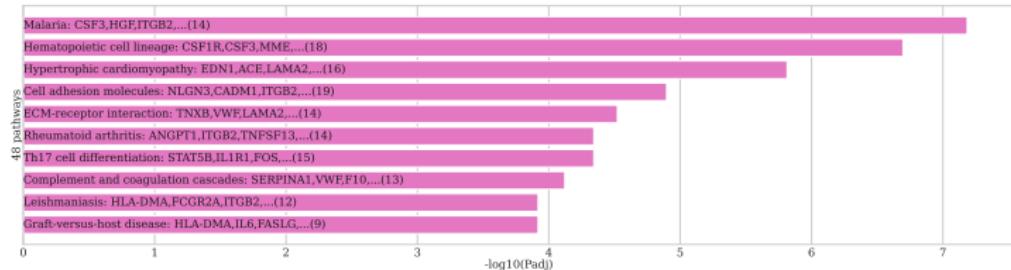


Figure: Up-regulated Pathways on Normal vs. CIS for Non-recur in LUSC

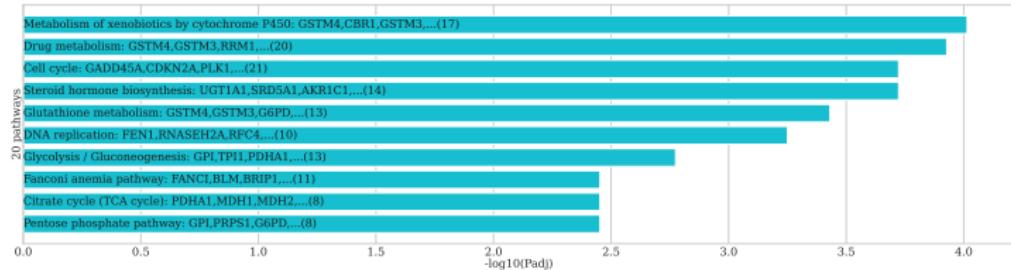


Figure: Down-regulated Pathways on Normal vs. CIS for Non-recur in LUSC

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

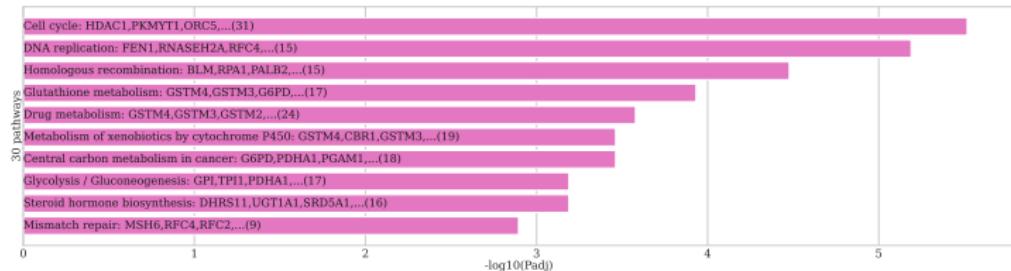


Figure: Up-regulated Pathways on Normal vs. Primary for Non-recur in LUSC

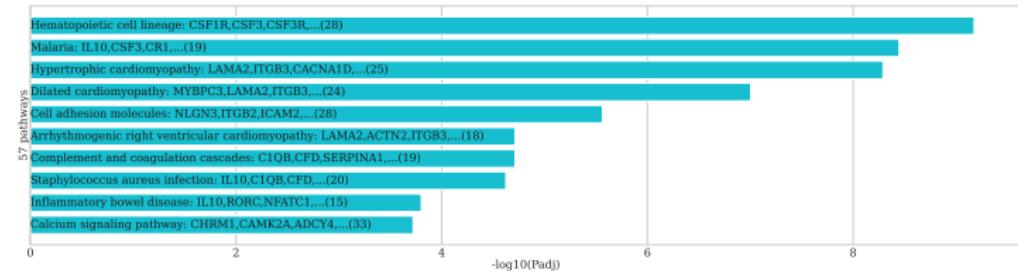


Figure: Down-regulated Pathways on Normal vs. Primary for Non-recur in LUSC

# Finding in Comparing within Non-recur in LUSC I

## AKR1C1 & AKR1C2

- ① Down-regulated in CIS, but up-regulated in Primary.
- ② Regulate steroids (Jin et al., 2009) and hormones (Penning et al., 2000)
- ③ Promote the metastasis of NSCLC (Z. Hong et al., 2018)

## CYP4F11

- ① Down-regulated in CIS, but up-regulated in Primary.
- ② Involved in the metabolism, including fatty acid and their derivatives (Edson et al., 2013; Kalsotra, Turman, Kikuta, & Strobel, 2004; Dhar, Sepkovic, Hirani, Magnusson, & Lasker, 2008)
- ③ CYP4F11 showed a strong association with survival in colorectal cancer (Alnabulsi, Swan, Cash, Alnabulsi, & Murray, 2017).

## LRRC36

- ① Up-regulated in CIS, but down-regulated in Primary.
- ② Leucine-rich repeat-containing protein 36
- ③ LRRC36 is positively correlated with survival in LUAD (Zhang et al., 2017).

## 4. Results

### 4.6. Differences in Gene Expression Levels

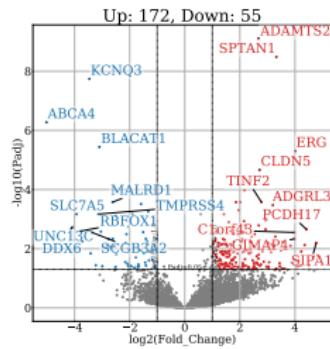
#### 4.6.6. Within Non-recur in LUAD

# LUAD Data Composition

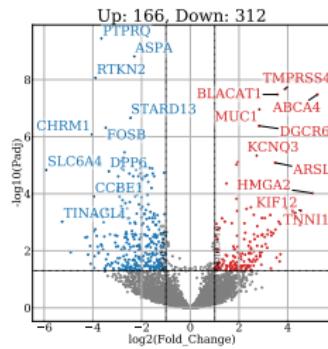
Table: Number of WTS LUAD samples

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

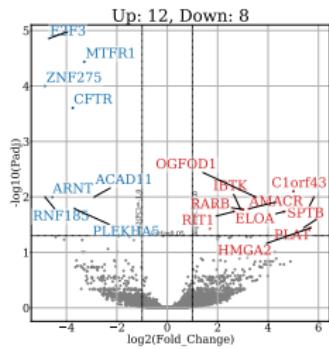
# DEG Volcano Plots with Non-recr in LUAD



(a) Normal-AIS



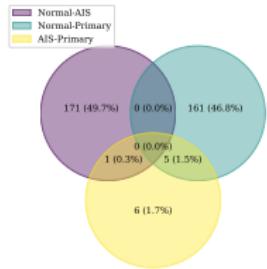
(b) Normal-Primary



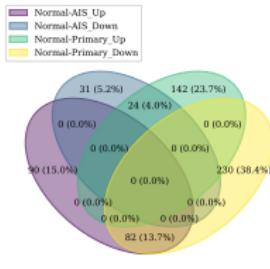
(c) AIS-Primary

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

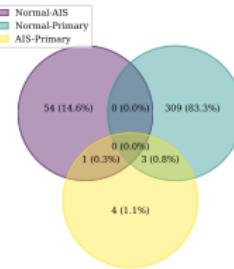
# DEG Venn Diagram with Non-recur in LUAD



(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

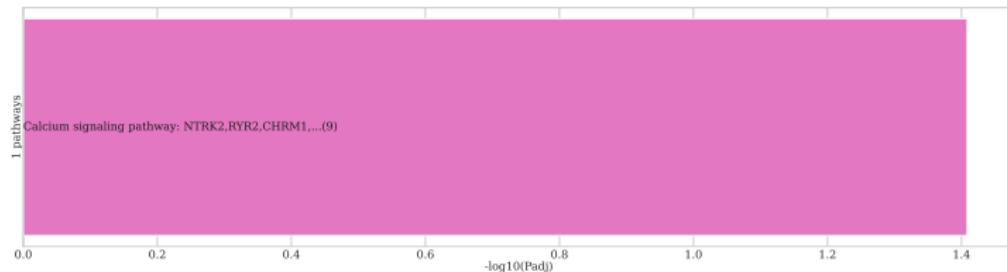


Figure: Up-regulated Pathways on Normal vs. AIS for Non-recur in LUAD

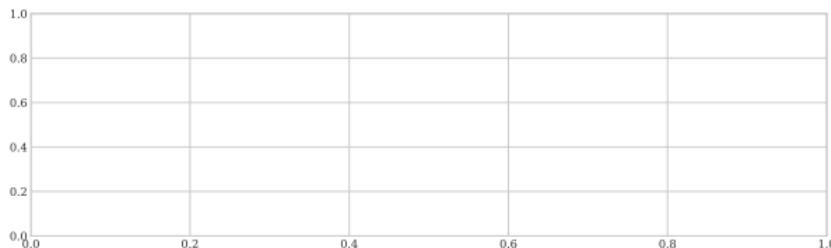


Figure: Down-regulated Pathways on Normal vs. AIS for Non-recur in LUAD

# Enrichment test with Normal vs. Primary in LUAD

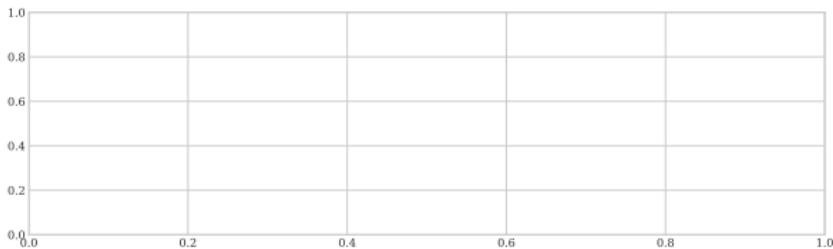


Figure: Up-regulated Pathways on Normal vs. Primary for Non-recur in LUAD

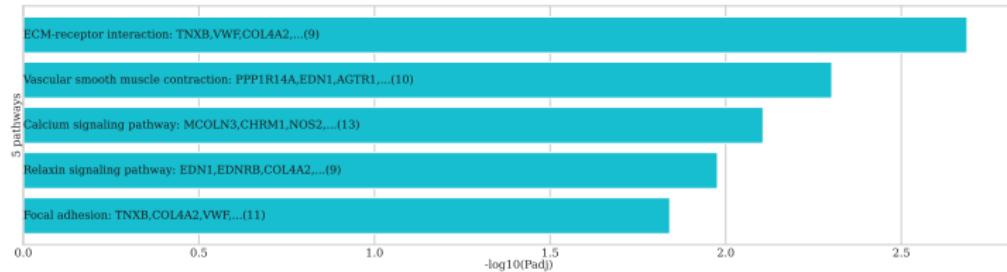


Figure: Down-regulated Pathways on Normal vs. Primary for Non-recur in LUAD

## KCNQ3

- ① Down-regulated in AIS, but up-regulated in Primary.
- ②  $K^+$  voltage-dependent channels  $\Rightarrow$  Various physiological functions (Schroeder et al., 1998; Surti et al., 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 diseases (Mondejar-Parreño et al., 2020)

## BLACAT1

- ① Down-regulated in AIS, but up-regulated in Primary.
- ② Bladder cancer-associated transcript 1
- ③ Chemo-resistance of NSCLC (Huang et al., 2019)
- ④ Predicts poor prognosis in SCLC (W. Chen et al., 2019)
- ⑤ Up-regulated in many human cancers (Ye, Yang, Liu, Lv, & Ye, 2020)

# Findings in DEG Analysis

## 4. Results

### 4.7. Bulk Cell Deconvolution

# BisqueRNA?

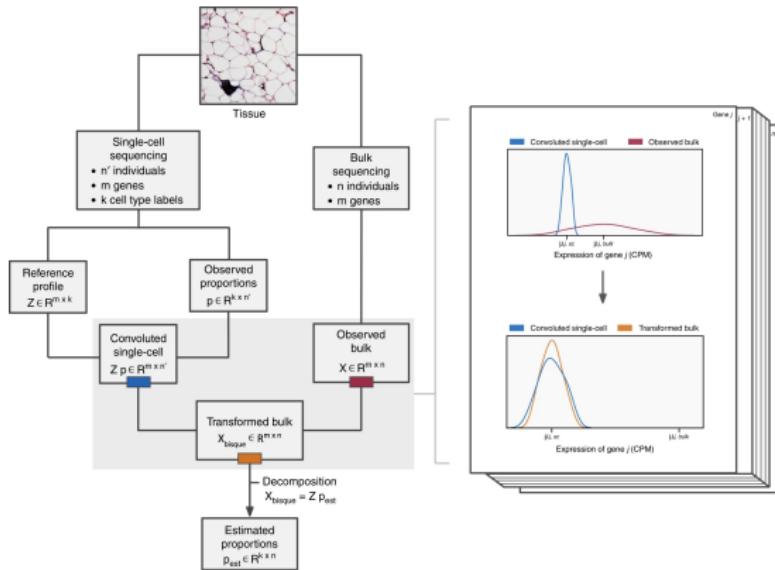


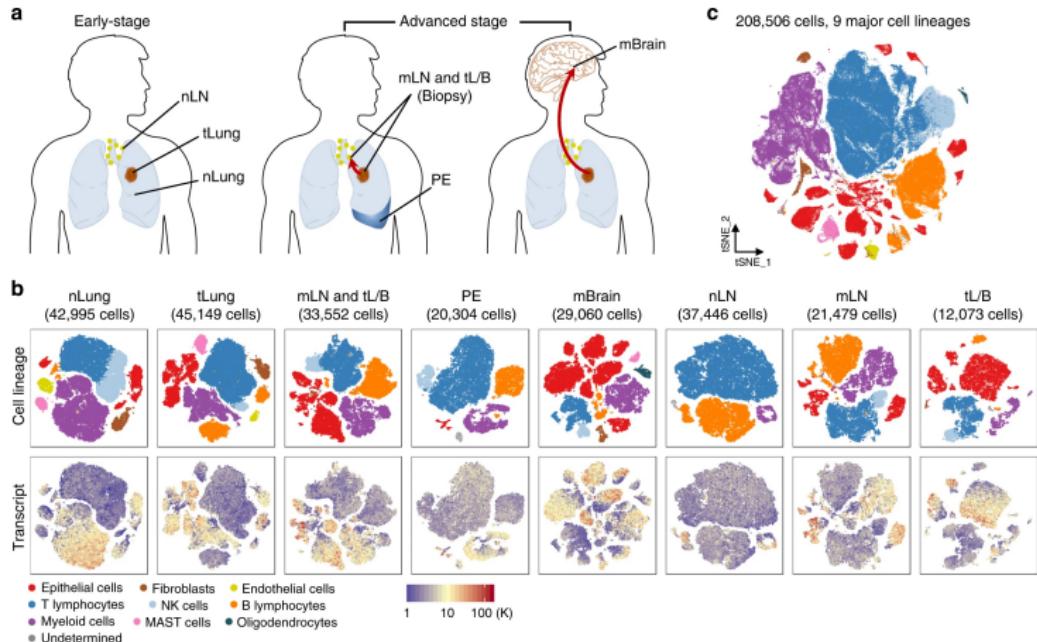
Figure: Workflow for BisqueRNA (Jew et al., 2020)

## 4. Results

### 4.7. Bulk Cell Deconvolution

#### 4.7.1. 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

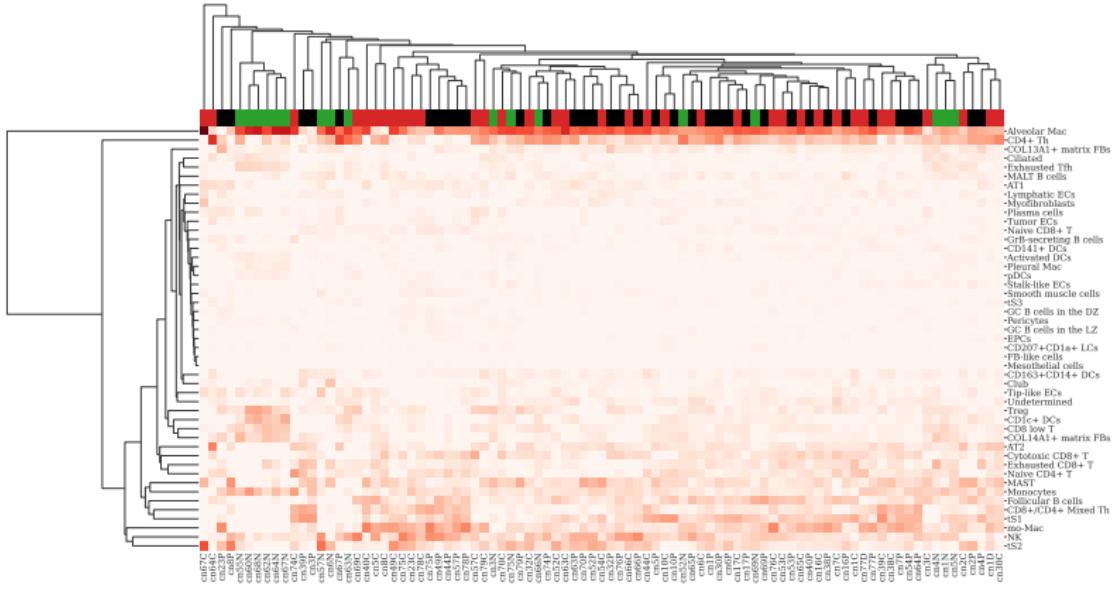
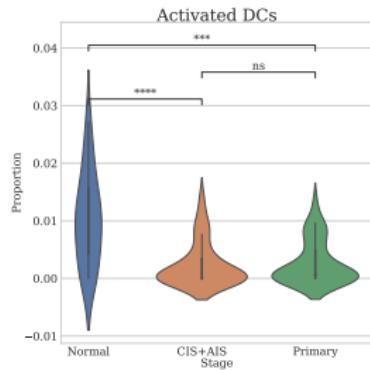
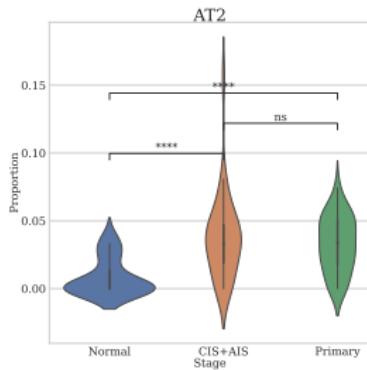


Figure: Cluster Plot in LUSC

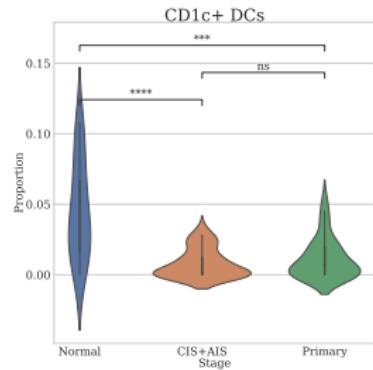
# Violin Plots in LUSC I



(a) Activated DCs



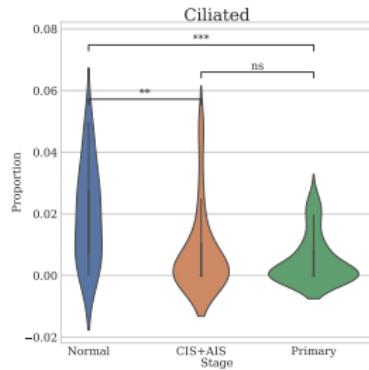
(b) Alveolar type II



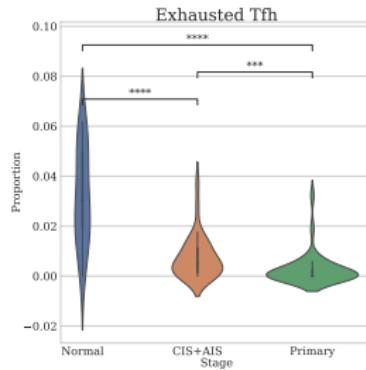
(c) Langerhans cells

Figure: Violin Plots in LUSC

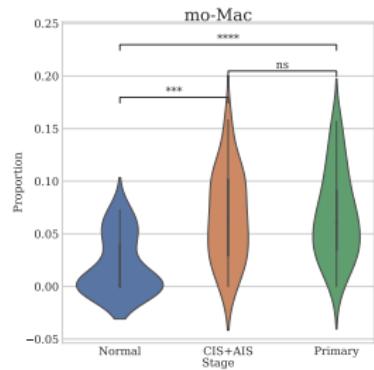
# Violin Plots in LUSC II



(d) Ciliated cells



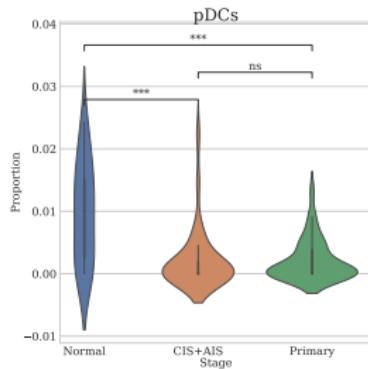
(e) Exhausted T follicular helper



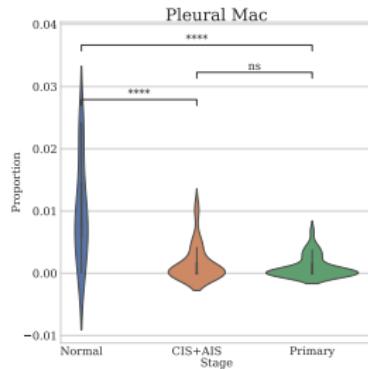
(f) Mo & Mac

Figure: Violin Plots in LUSC

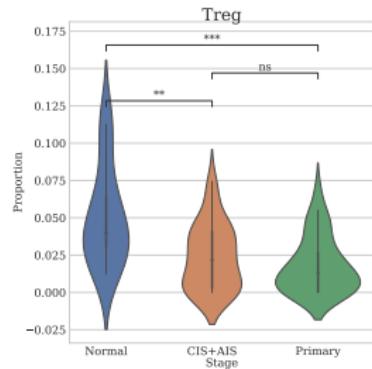
# Violin Plots in LUSC III



(g) Plasmacytoid DCs



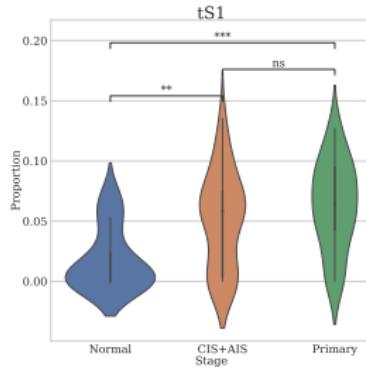
(h) Pleural Mac



(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.
- ② DCs are central regulators of adaptive immune response, e.g. anti-tumoral responses .
- ③ DCs establish a rare immune cell population in tumors .

## Alveolar type II

- ① Alveolar type II have lower proportion in Normal samples.
- ② Alveolar type II proliferate to restore epithelium, and participate in innate immune response (Mason, 2006).
- ③ Capability of initiating lung cancer development (C. Lin et al., 2012).

# Findings in Bulk Cell Deconvolution with LUSC II

## CD1c+ DCs (Langerhans cells; LCs)

- ① LCs have higher proportion in Normal samples.
- ② LCs impact on pathology by inducing tolerance or mediating inflammation (Deckers, Hammad, & Hoste, 2018)
- ③ LCs facilitate DNA damage and squamous cell carcinoma (Modi et al., 2012)

## Ciliated cells

- ① Ciliated cells have higher proportion in Normal samples.
- ② A terminally differentiated population in lung epithelial cells (Rawlins & Hogan, 2008).
- ③ Generated under homeostatic condition or response to epithelial injury (Sutherland et al., 2011).

# Findings in Bulk Cell Deconvolution with LUSC III

## Exhausted T follicular helper cells (Tfh)

- ① Exhausted Tfh is gradually decreased along cancer worsen.
- ② Tfh cell response is critical for viral infection (Greczmiel et al., 2017; Poonia, Ayithan, Nandi, Masur, & Kottilil, 2018a).
- ③ Down-regulated Tfh exhaustion correlate with compromise CD8 T-cell immunity (Poonia, Ayithan, Nandi, Masur, & Kottilil, 2018b)

## Monocyte & Macrophage

- ① Monocyte & Macrophage have lower proportion in Normal samples.
- ② Monocyte is a regulator of tumor development & progression (Olingy, Dinh, & Hedrick, 2019).
- ③ Macrophage is a regulator of link between inflammation & cancer (Sica, Allavena, & Mantovani, 2008).

## Plasmacytoid DCs (pDCs)

- ① pDCs have higher proportion in Normal samples.
- ② pDCs bring capacities of innate & adaptive immunity (Vermi, Soncini, Melocchi, Sozzani, & Facchetti, 2011).
- ③ Infiltrated pDCs in neoplasms ⇒ Poor prognosis (Pinto, Rega, Crother, & Sorrentino, 2012).

## Pleural Macrophages

- ① Pleural macrophages have higher proportion in Normal samples.
- ② Neutrophil recruitment in pleural inflammation (Cailhier et al., 2006).

# Findings in Bulk Cell Deconvolution with LUSC V

## Regulatory T cells (Tregs)

- ① Tregs have higher proportion in Normal samples.
- ② Elevation of Tregs ↑ in solid tumors & hematologic malignancies (Beyer & Schultze, 2006).
- ③ Increasing Tregs ↑ along metastatic stage in NSCLC (Erfani et al., 2012)

## Tumor cell states 1 (tS1)

- ① tS1 have lower proportion in Normal samples.
- ② Represent a de-regulation of normal differentiation programs (Kim et al., 2020).

# Cluster Plot in LUAD

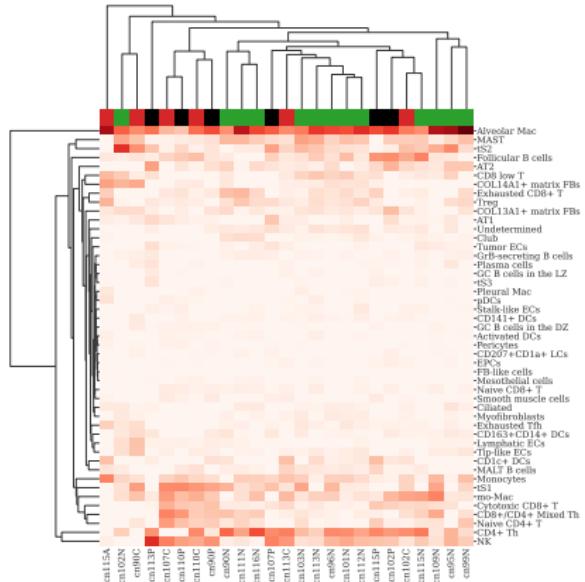
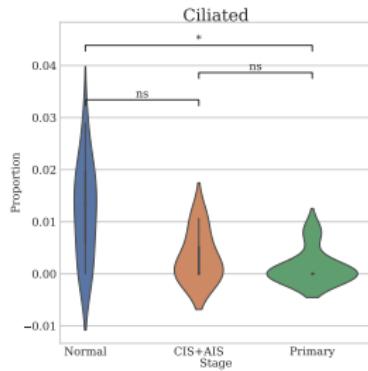
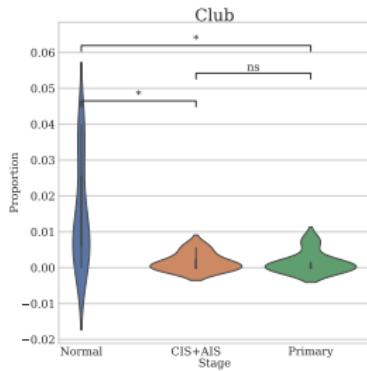


Figure: Cluster Plot in LUAD

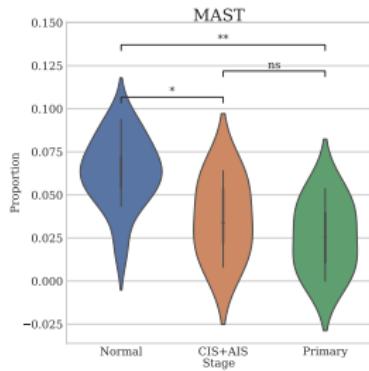
# Violin Plots in LUAD I



(a) Ciliated cells



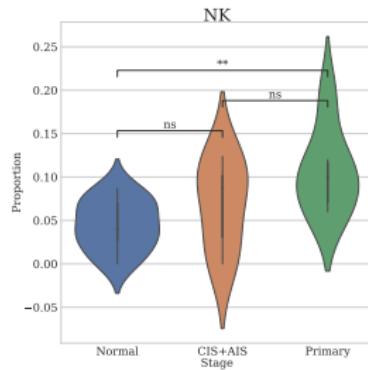
(b) Club Cell



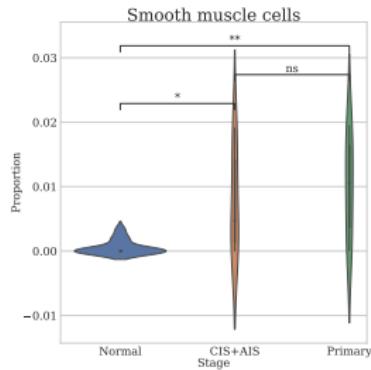
(c) Mast cell

Figure: Violin Plots in LUAD

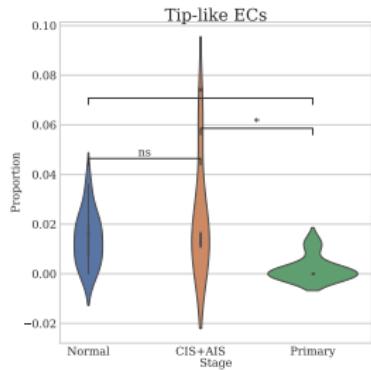
# Violin Plots in LUAD II



(d) NK cells



(e) Smooth muscle cells



(f) Tip-like ECs

Figure: Violin Plots in LUAD

## Ciliated cells

- ① Ciliated cells have higher proportion in Normal than Primary samples.
- ② A terminally differentiated population in lung epithelial cells (Rawlins & Hogan, 2008).
- ③ Generated under homeostatic condition or response to epithelial injury (Sutherland et al., 2011).

# Findings in Bulk Cell Deconvolution with LUAD II

## Club cells

- ① Club cells have higher proportion in Normal than Primary samples.
- ② Club cells form LUAD in adult mice (Spella et al., 2019).
- ③ Club cells in smoking-associated LUAD (Behrend, Giotopoulou, Spella, & Stathopoulos, 2021).
- ④ Increasing club cells ⇒ Good indicator of advanced bronchopulmonary dysplasia (Rokicki, Rokicki, Wojtacha, & Dżelijjli, 2016).

## Mast cells

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

## Natural Killer cells (NK cells)

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

# Findings in Bulk Cell Deconvolution with LUAD III

## 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

### 4.7. Bulk Cell Deconvolution

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

# Reference Single-cell Data

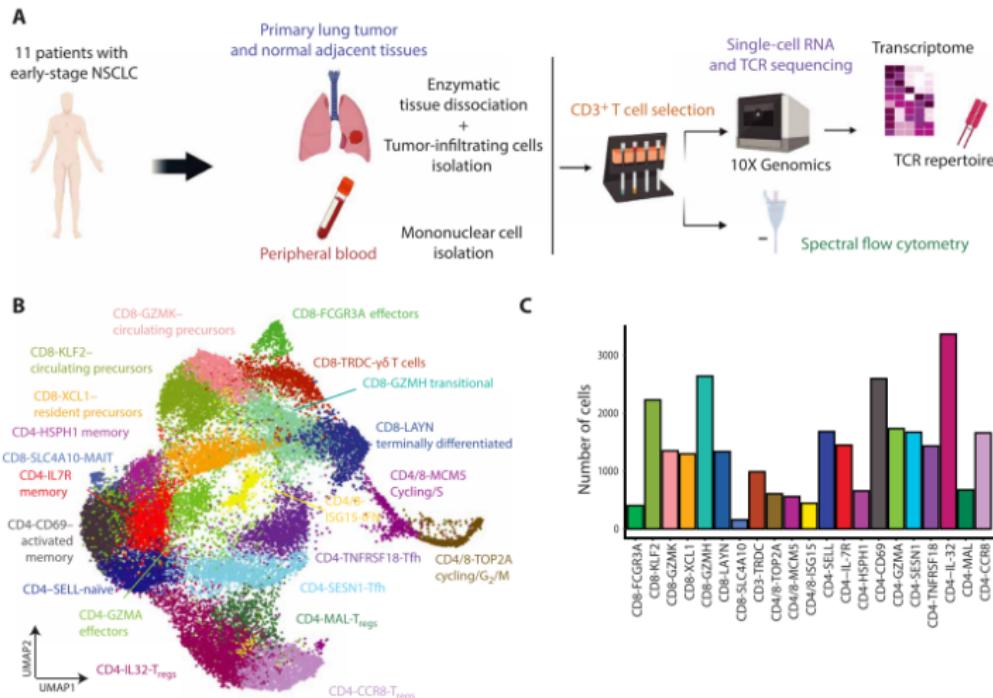


Figure: Characterization of CD3<sup>+</sup> TILs in NSCLC (Gueguen et al., 2021)

# Cluster Plots in LUSC

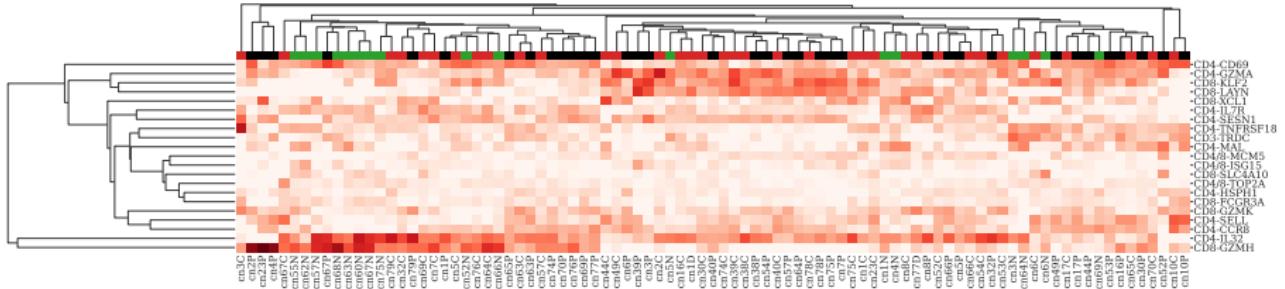
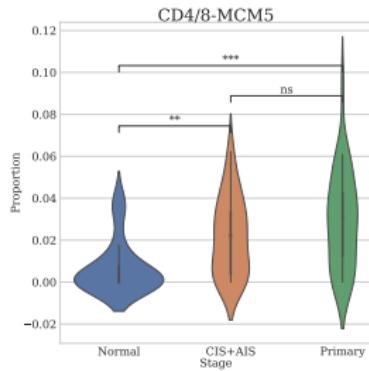
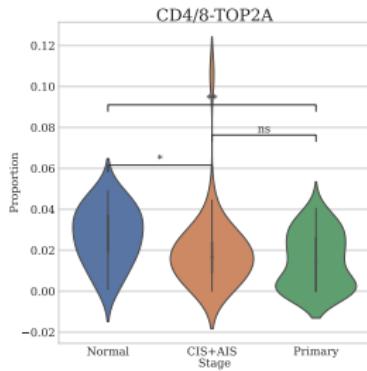


Figure: Cluster Plot in LUAD

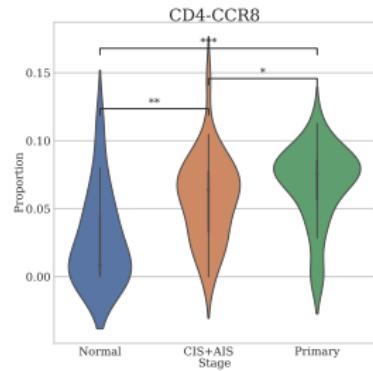
# Violin Plots in LUSC I



(a) CD4/8-MCM5



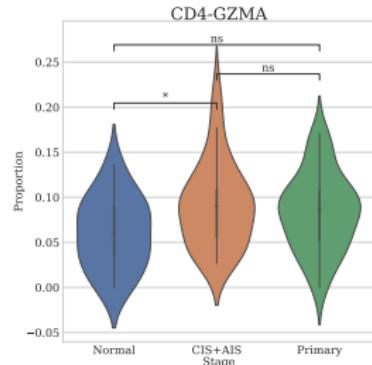
(b) CD4/8-TOP2A



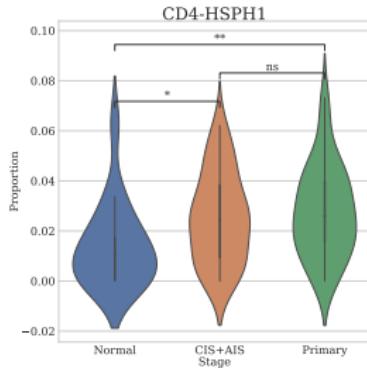
(c) CD4-CCR8

Figure: Violin Plots in LUSC

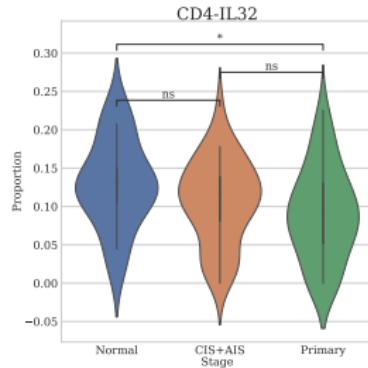
# Violin Plots in LUSC II



(d) CD4-GZMA



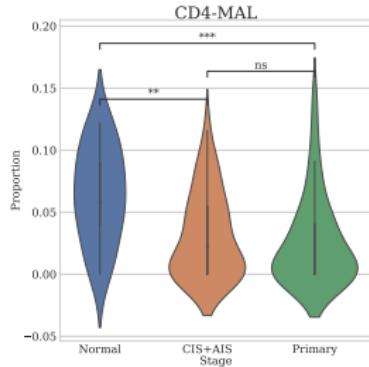
(e) CD4-HSPH1



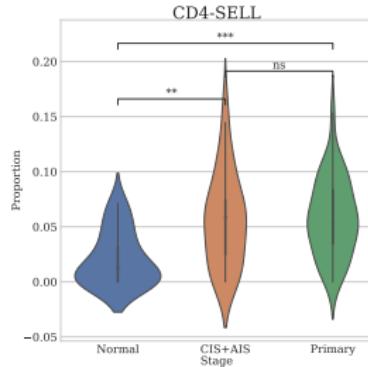
(f) CD4-IL32

Figure: Violin Plots in LUSC

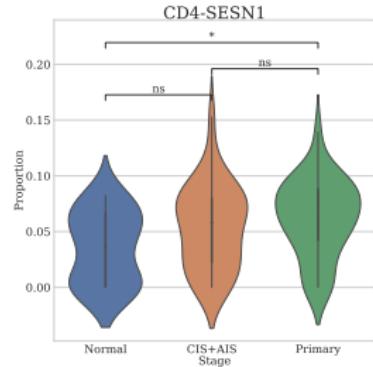
# Violin Plots in LUSC III



(g) CD4-MAL



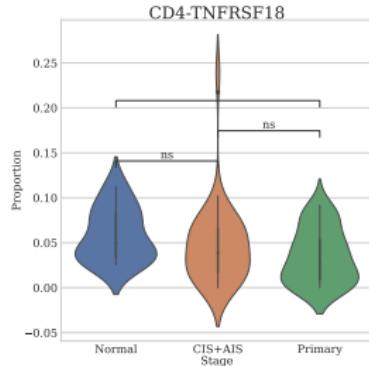
(h) CD4-SELL



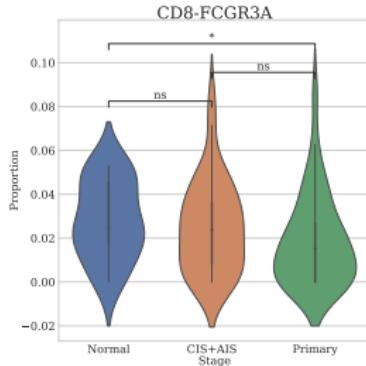
(i) CD4-SESN1

Figure: Violin Plots in LUSC

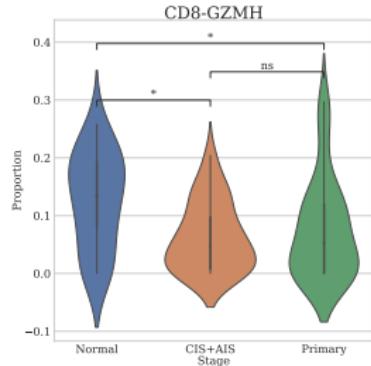
# Violin Plots in LUSC IV



(g) CD4-TNFRSF18



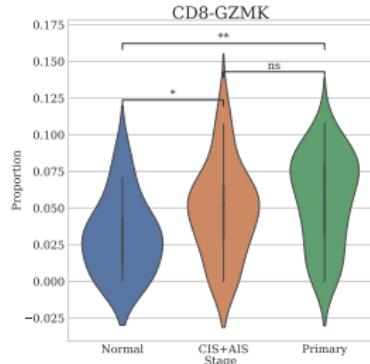
(h) CD8-FCGR3A



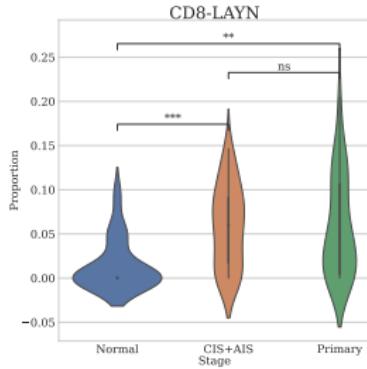
(i) CD8-GZMH

Figure: Violin Plots in LUSC

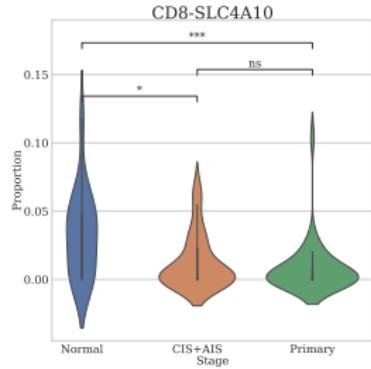
# Violin Plots in LUSC V



(g) CD8-GZMK



(h) CD8-LAYN



(i) CD8-SLC4A10

Figure: Violin Plots in LUSC

# Findings in Bulk Cell Deconvolution with LUSC I

## CD4/8-MCM5

- ① CD4/8-MCM5 have higher proportion in Primary.

## CD4/8-TOP2A

- ① CD4/8-TOP2A have lower proportion in Primary.

## CD4-CCR8

- ① CD4-CCR8 is gradually increased along tumor progression.

## CD4-GZMA

- ① CD4-GZMA have higher proportion in CIS than Normal.

# Findings in Bulk Cell Deconvolution with LUSC II

## CD4-HSPH1

- ① CD4-HSPH1 have higher proportion in Primary.

## CD4-IL32

- ① CD4-IL32 have lower proportion in Primary than Normal.

## CD4-MAL

- ① CD4-MAL have lower proportion in Primary.

## CD4-SELL

- ① CD4-SELL have higher proportion in Primary.

## CD4-SESN1

- ① CD4-SESN1 have higher proportion in Primary than Normal.

# Findings in Bulk Cell Deconvolution with LUSC III

## CD4-TNFRSF18

- ① CD4-TNFRSF18 have lower proportion in Primary than Normal.

## CD8-FCGR3A

- ① CD8-FCGR3A have lower proportion in Primary than Normal.

## CD8-GZMH

- ① CD8-GZMH have lower proportion in Primary.

## CD8-GZMK

- ① CD8-GZMK have higher proportion in Primary.

## CD8-LAYN

- ① CD8-LAYN have higher proportion in Primary.

## CD8-SLC4A10

- ① CD8-SLC4A1 have lower proportion in Primary.

# Cluster Plots in LUAD

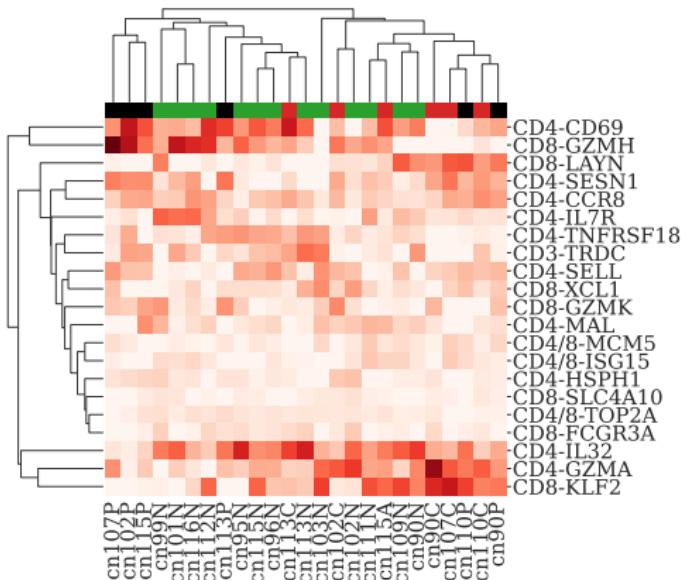
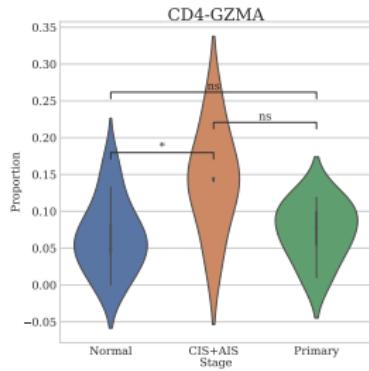
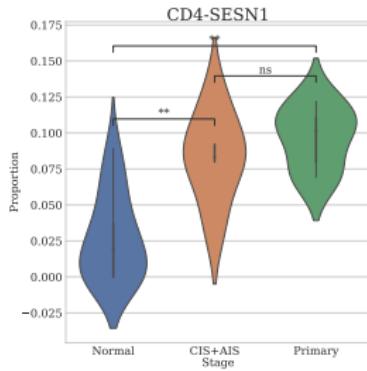


Figure: Cluster Plot in LUAD

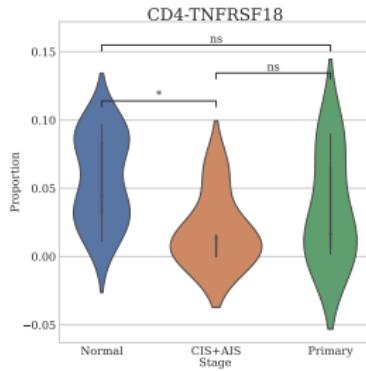
# Violin Plots in LUAD



(a) CD4-GZMA



(b) CD4-SESN1



(c) CD4-TNFRSF18

Figure: Violin Plots in LUAD

# Findings in Bulk Cell Deconvolution with LUAD I

## CD4-GZM

- ① CD4-GZM have higher proportion in AIS than Normal.

## CD4-SESN1

- ① CD4-SESN1 have higher proportion in AIS than Normal.

## CD4-TNFRSF18

- ① CD4-TNFRSF18 have lower proportion in AIS than Normal.

# Findings in Bulk Cell Deconvolution

## 4. Results

### 4.8. Discovery of Mutational Signature

# Mutational Signature?

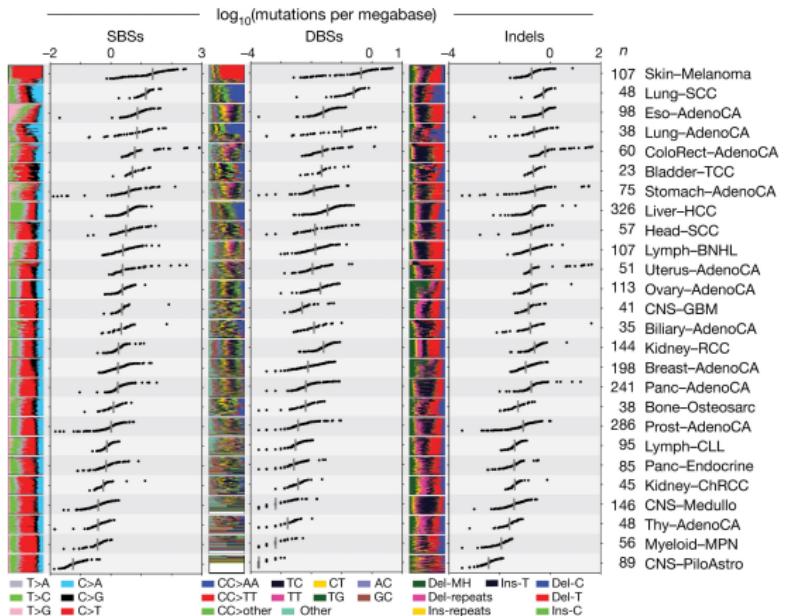
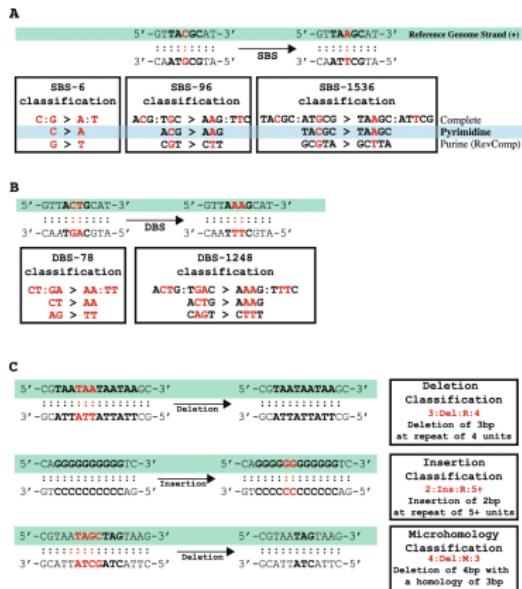


Figure: Mutation Burdens across PCAWG tumor types (Alexandrov et al., 2020)

# SigProfiler?



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

## 4. Results

### 4.8. Discovery of Mutational Signature

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

# SBS Signatures I

## SBS1

- An endogenous mutational process (Nik-Zainal et al., 2012a)
- generates G>T mismatches in double-stranded DNA
- Failure ↓ to detect & remove these mismatches

## SBS2

- Activity of the AID/APOBEC family of cytidine deaminases (Nik-Zainal et al., 2012a)
  - ① APOBEC3A is probably responsible in human cancer
  - ② APOBEC3B may also contribute
- may be generated directly by DNA replication

## SBS4

- Tobacco smoking (Alexandrov et al., 2013)
- Exposed to tobacco carcinogens e.g. benzopyrene

## SBS5

- Unknown (Alexandrov et al., 2013)
- SBS5 ↑ in bladder cancer
- SBS5 ↑ in many cancer types ∵ Tobacco smoking

## SBS10b

- Polymerase ε exonuclease domain mutations (Alexandrov et al., 2020)

# SBS Signatures III

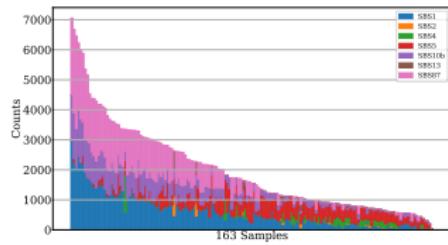
## SBS13

- Activity of the AID/APOBEC family of cytidine deaminases (Nik-Zainal et al., 2012b)
- SBS13 is usually found with SBS2

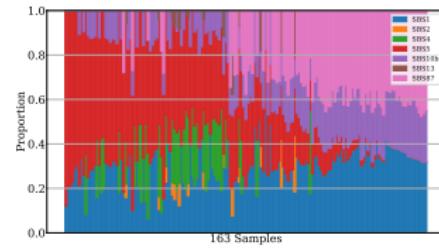
## SBS87

- Thiopurine chemotherapy treatment (Li et al., 2020)

# SBS in LUSC I



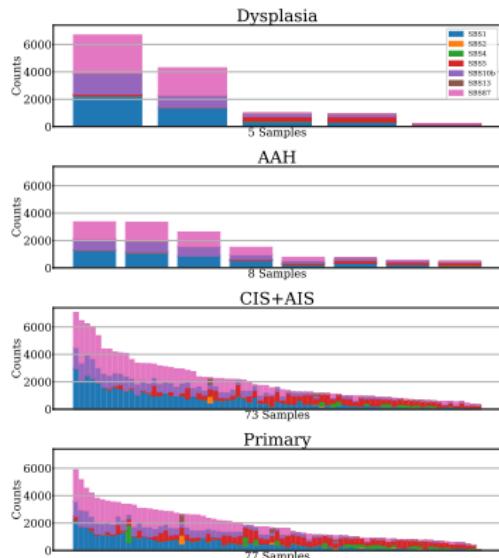
(a) Absolute



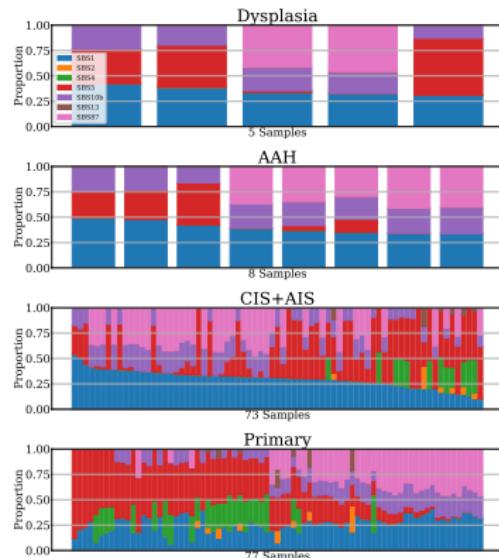
(b) Relative

Figure: SBS Bar Plot in LUSC

# SBS in LUSC II



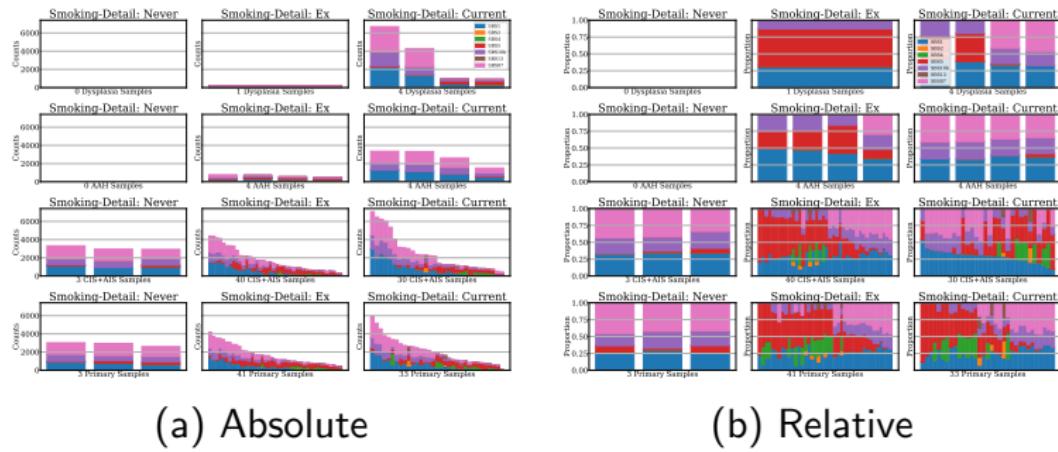
(a) Absolute



(b) Relative

Figure: SBS Bar Plot by Cancer Subtype in LUSC

# SBS in LUSC with Smoking I

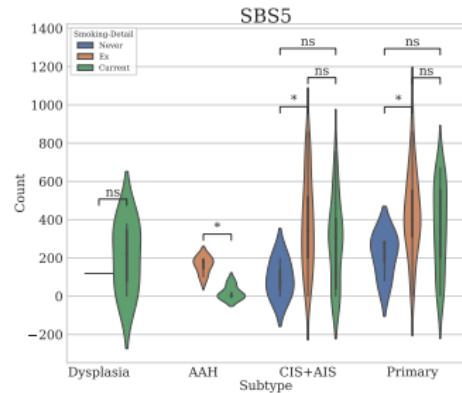


(a) Absolute

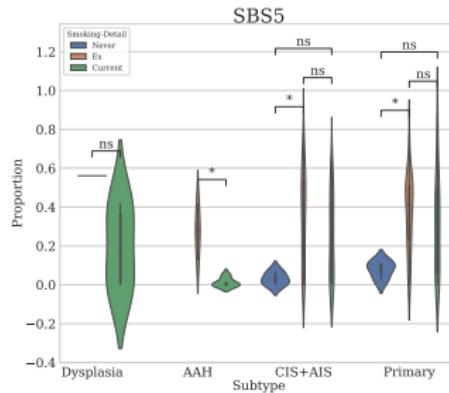
(b) Relative

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

# SBS in LUSC with Smoking II



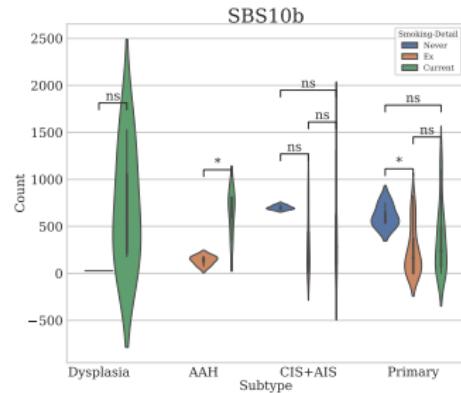
(a) Absolute



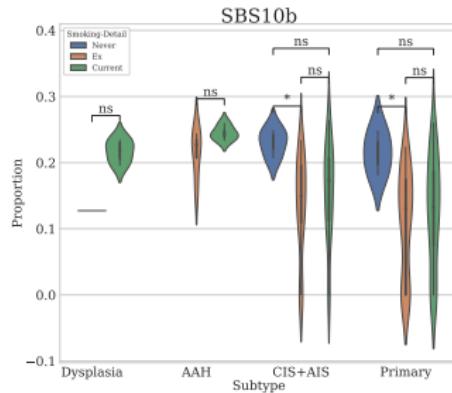
(b) Relative

Figure: SBS5 Signature in LUSC with Smoking

# SBS in LUSC with Smoking III



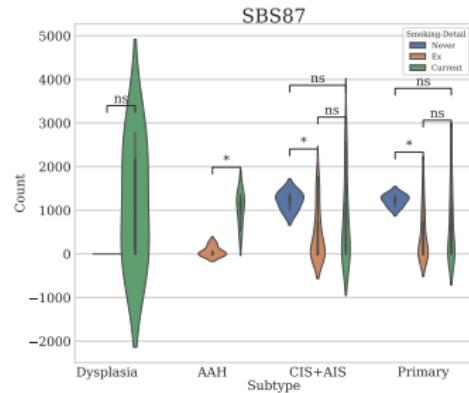
(a) Absolute



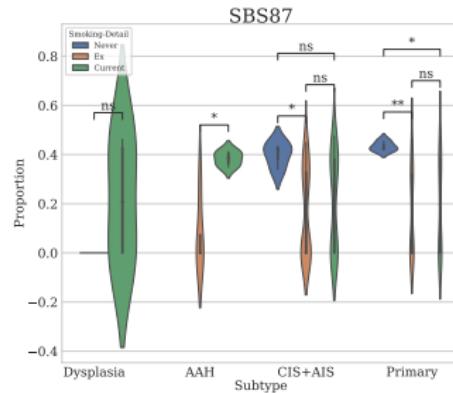
(b) Relative

Figure: SBS10b Signature in LUSC with Smoking

# SBS in LUSC with Smoking IV



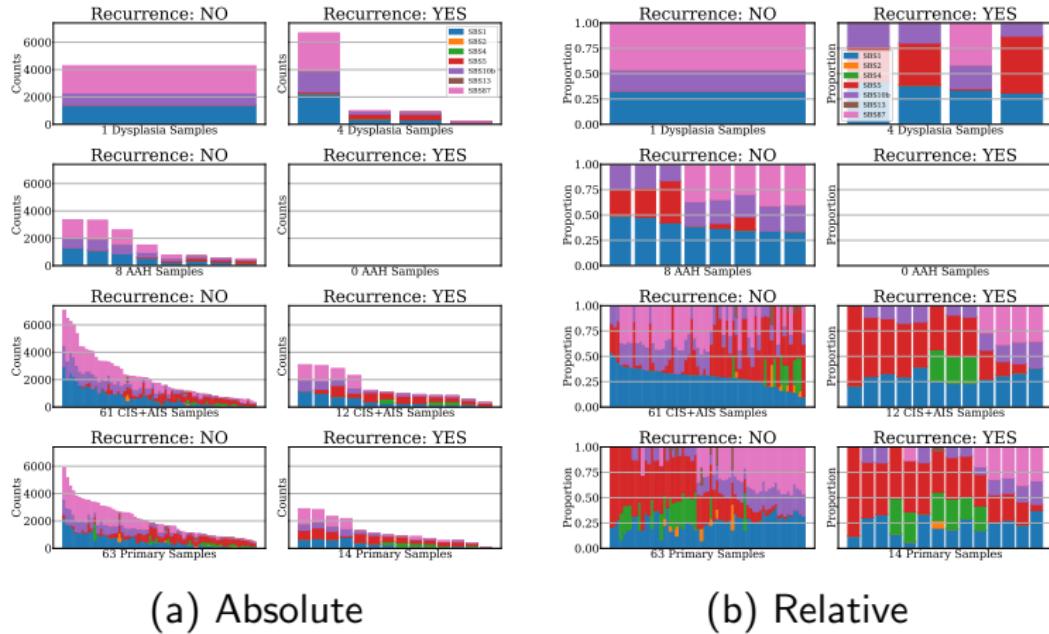
(a) Absolute



(b) Relative

Figure: SBS87 Signature in LUSC with Smoking

# SBS in LUSC with Recurrence I

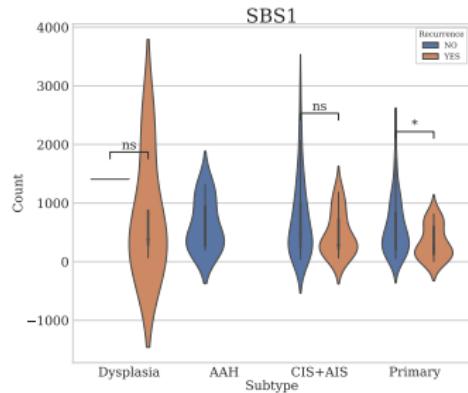


(a) Absolute

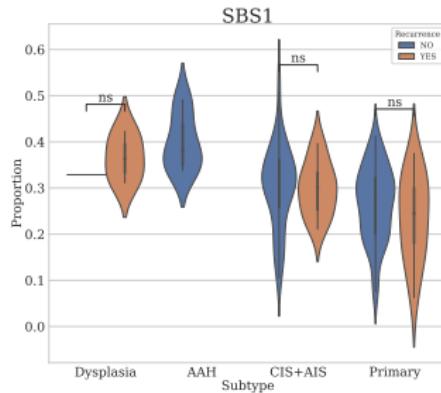
(b) Relative

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

# SBS in LUSC with Recurrence II



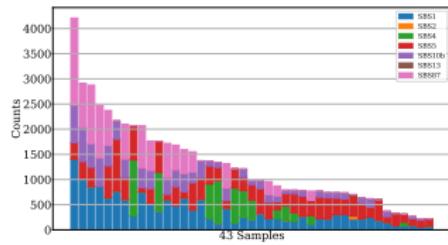
(a) Absolute



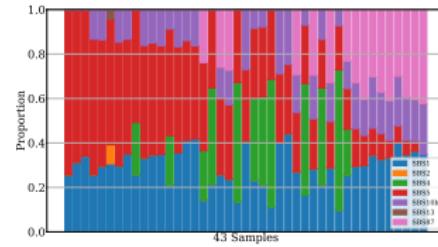
(b) Relative

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

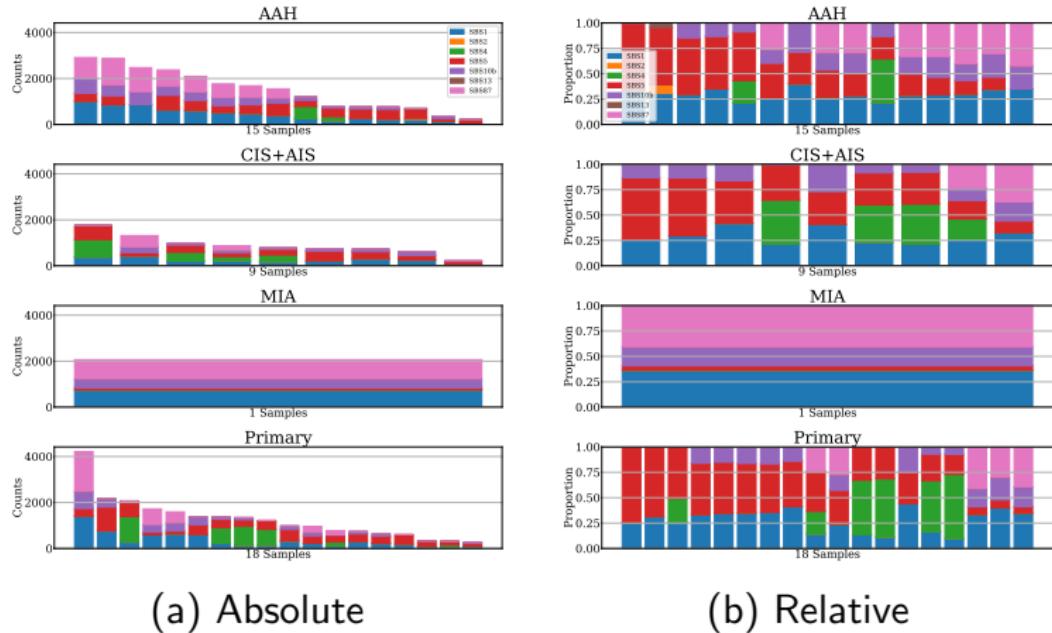


Figure: SBS Bar Plot by Cancer Subtype in LUSC

# SBS in LUAD with Smoking I

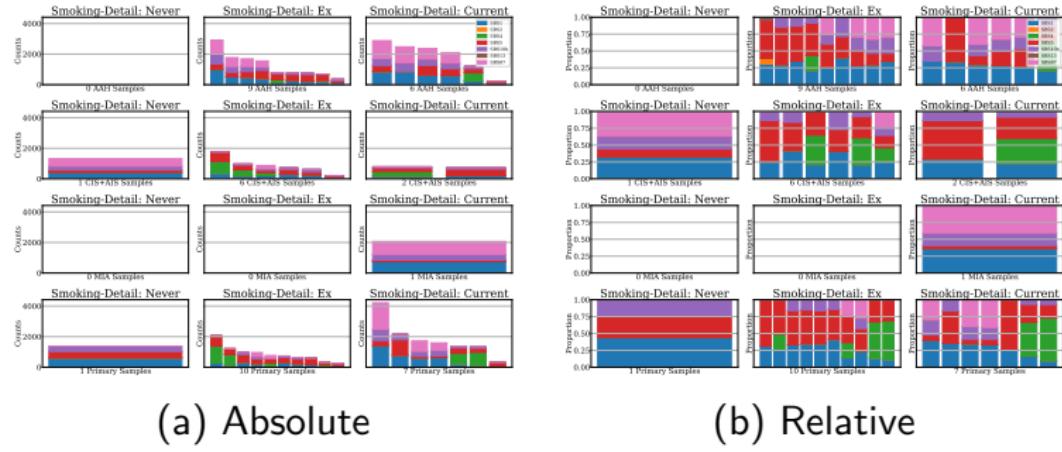
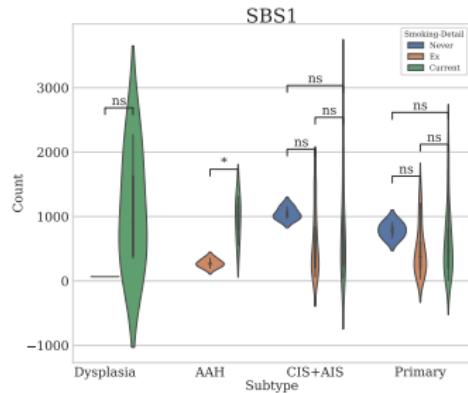
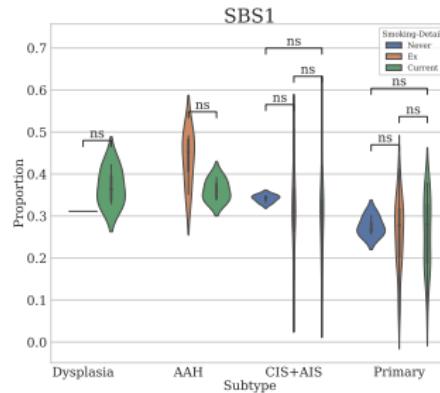


Figure: SBS Bar Plot by Cancer Subtype & Smoking in LUAD

# SBS in LUAD with Smoking II



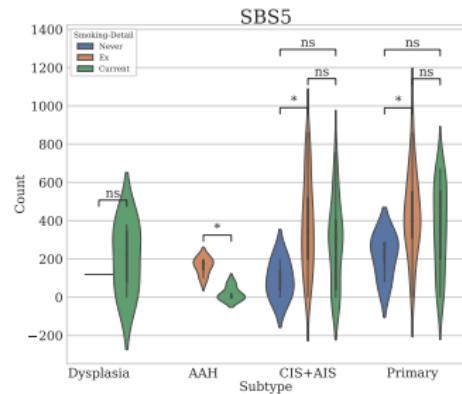
(a) Absolute



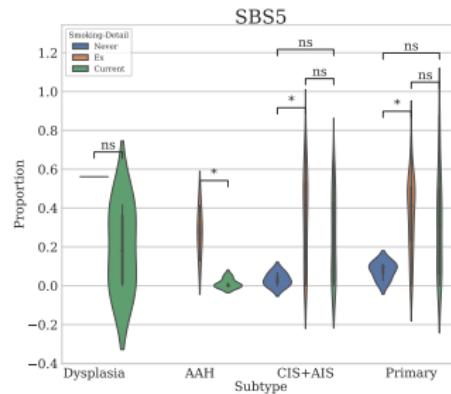
(b) Relative

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

# SBS in LUAD with Smoking III



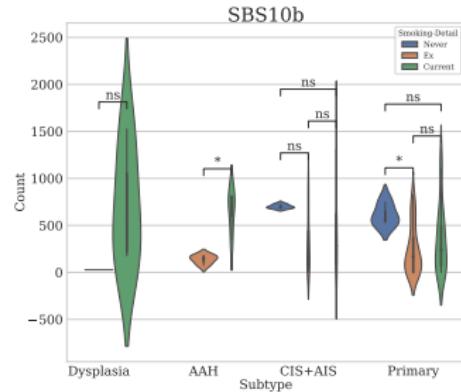
(a) Absolute



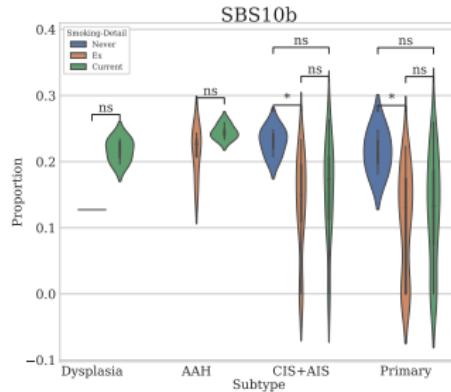
(b) Relative

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

# SBS in LUAD with Smoking IV



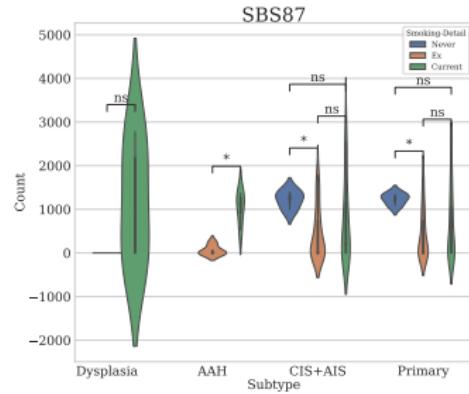
(a) Absolute



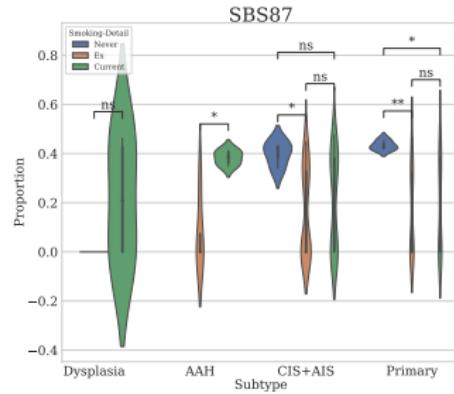
(b) Relative

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

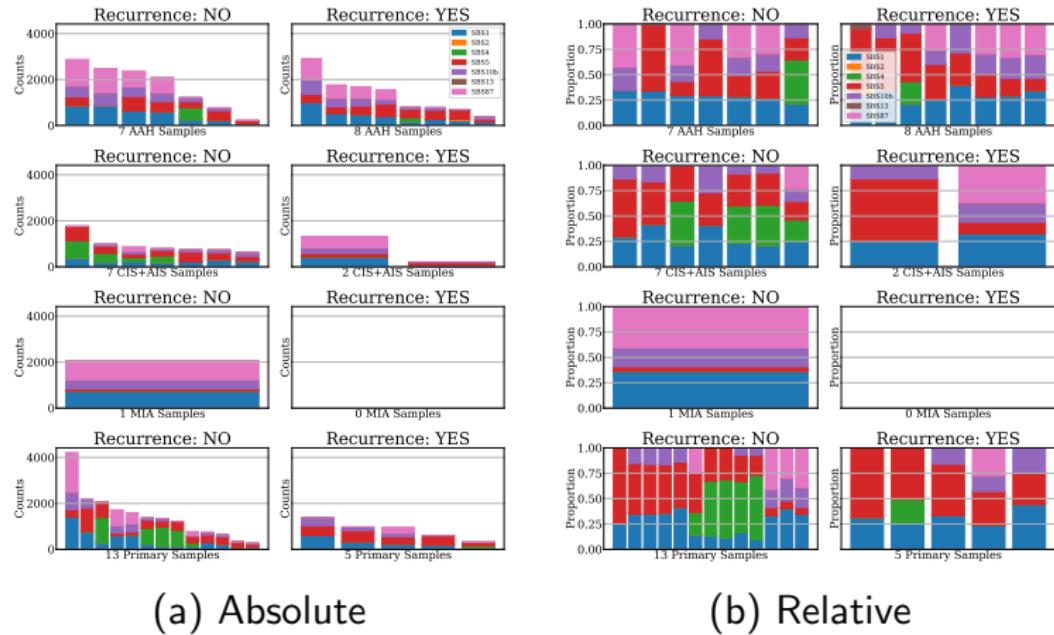
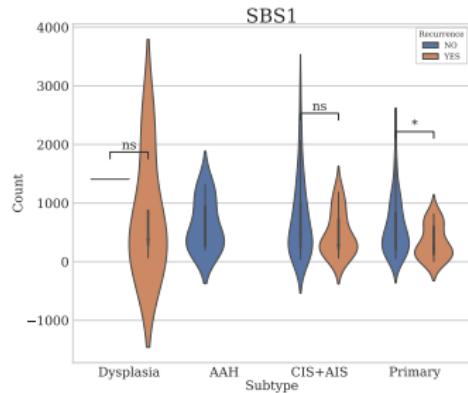
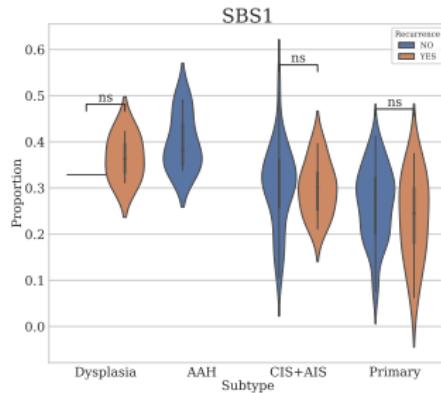


Figure: SBS Bar Plot by Cancer Subtype & Recurrence in LUAD

# SBS in LUAD with Recurrence II



(a) Absolute



(b) Relative

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

## 4. Results

### 4.8. Discovery of Mutational Signature

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

# DBS Signatures I

## DBS2

- Tobacco smoking (J.-M. Chen, Férec, & Cooper, 2013)
- Other endogenous/exogenous mutagens e.g. acetaldehyde

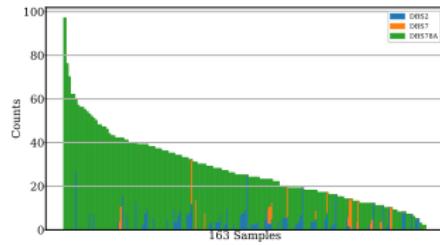
## DBS7

- Defective ↓ DNA mismatch repair (Alexandrov et al., 2020)

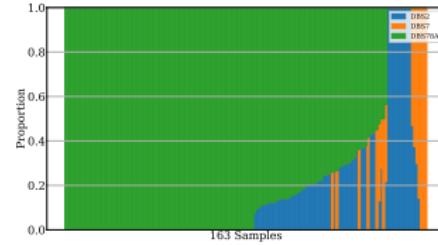
## DBS78A

content...

# DBS in LUSC I



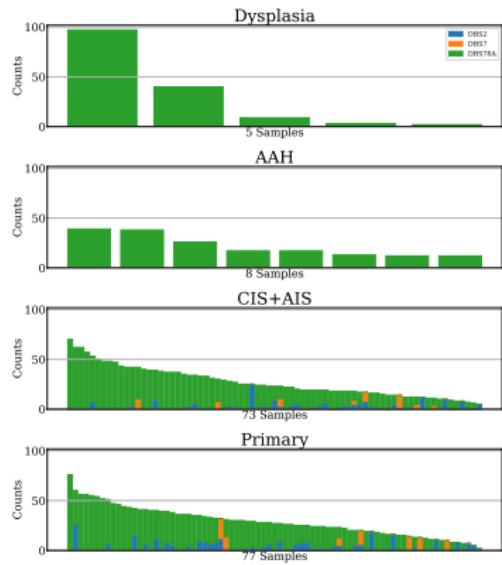
(a) Absolute



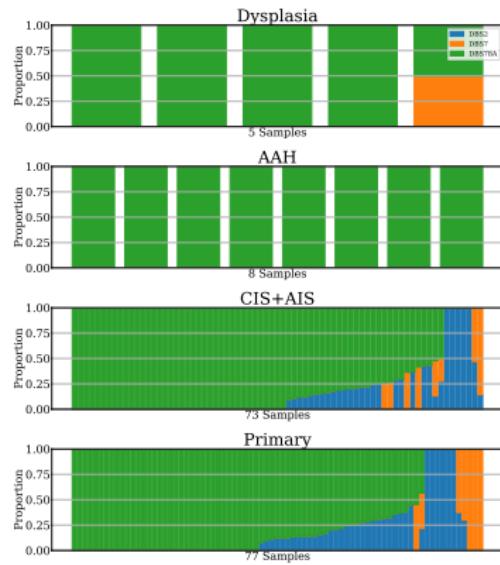
(b) Relative

Figure: DBS Bar Plot in LUSC

# DBS in LUSC II



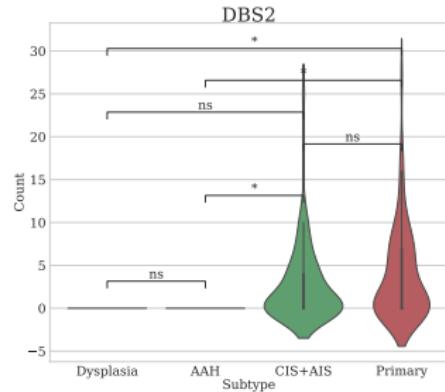
(a) Absolute



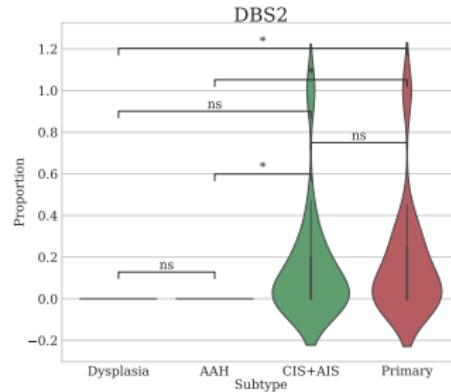
(b) Relative

Figure: DBS Bar Plot by Cancer Subtype in LUSC

# DBS in LUSC III



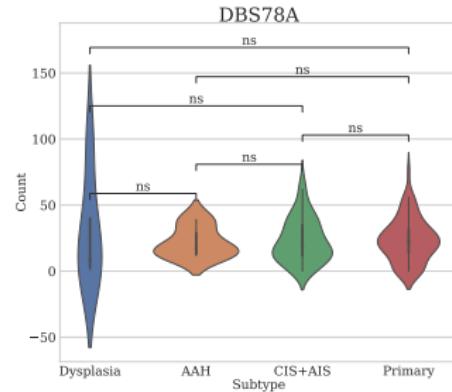
(a) Absolute



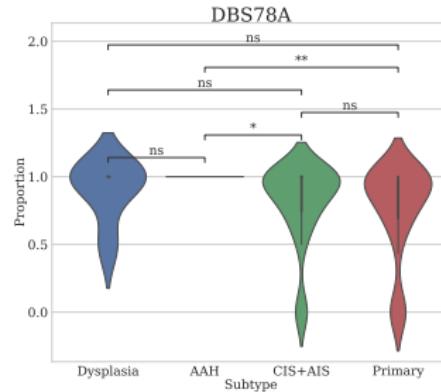
(b) Relative

Figure: DBS2 Signature in LUSC

# DBS in LUSC IV



(a) Absolute



(b) Relative

Figure: DBS78A Signature in LUSC

# DBS in LUSC with Smoking I

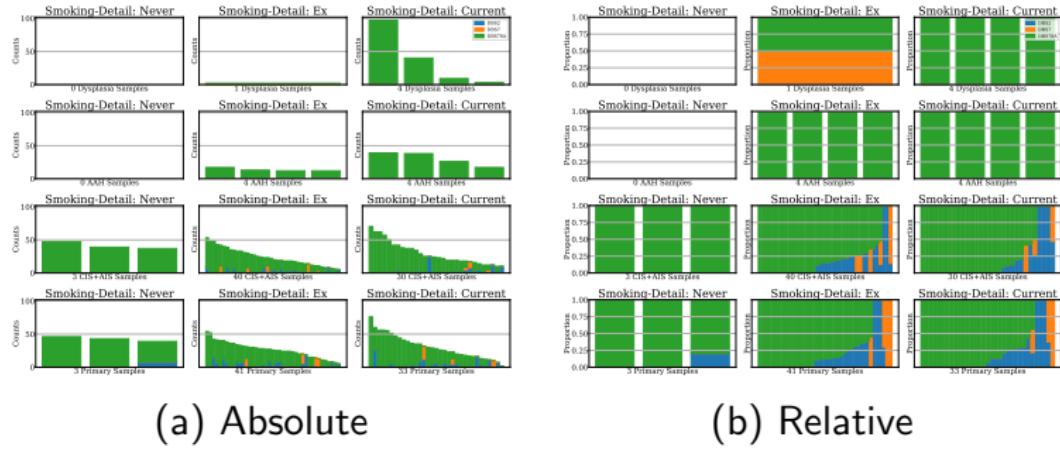
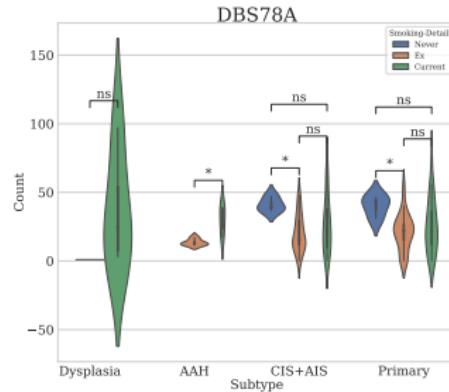
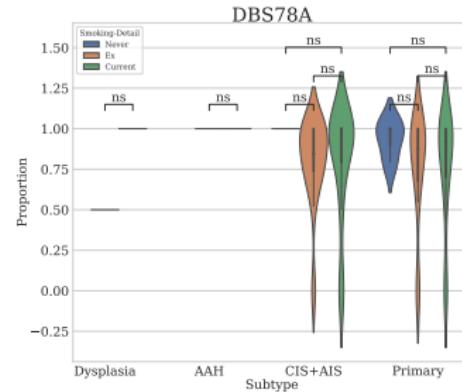


Figure: DBS Bar Plot by Cancer Subtype & Smoking in LUSC

# DBS in LUSC with Smoking II



(a) Absolute



(b) Relative

Figure: DBS78A Signature in LUSC with Smoking

# DBS in LUSC with Recurrence

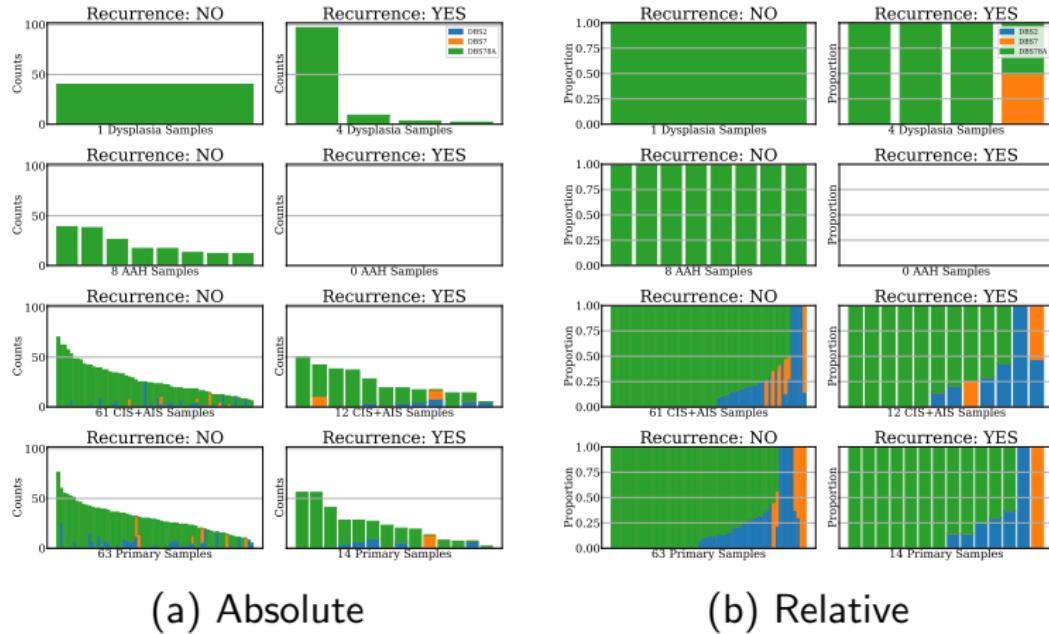
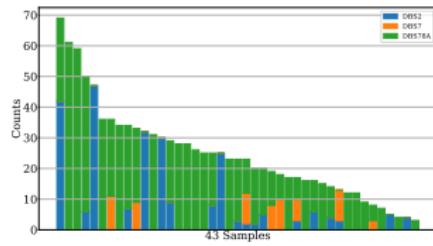
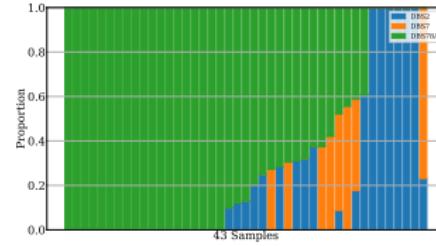


Figure: DBS Bar Plot by Cancer Subtype & Recurrence in LUSC

# DBS in LUAD I



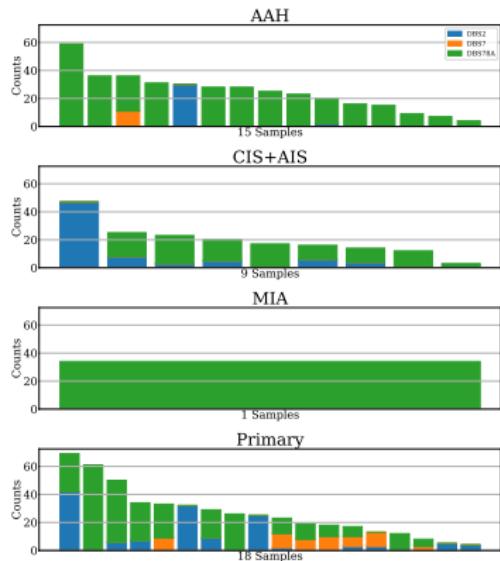
(a) Absolute



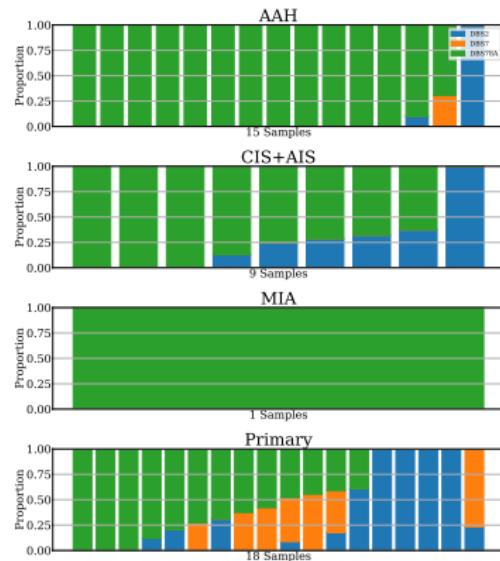
(b) Relative

Figure: DBS Bar Plot in LUAD

# DBS in LUAD II



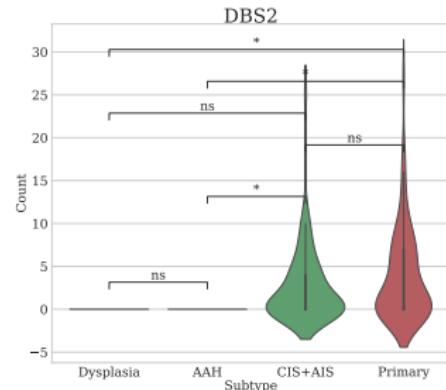
(a) Absolute



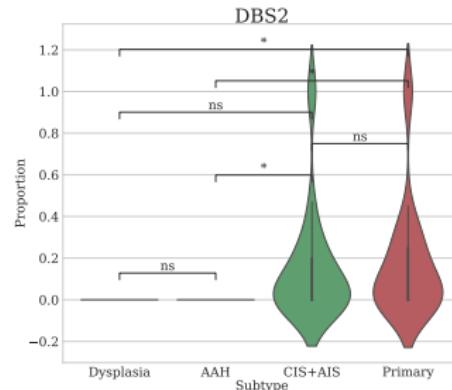
(b) Relative

Figure: DBS Bar Plot by Cancer Subtype in LUAD

# DBS in LUAD III



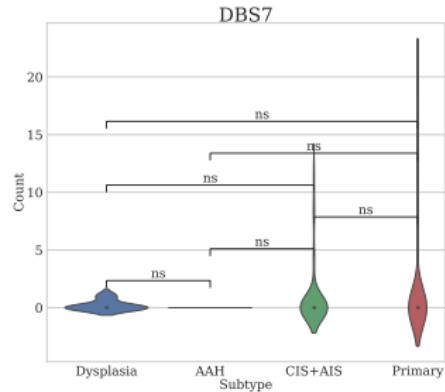
(a) Absolute



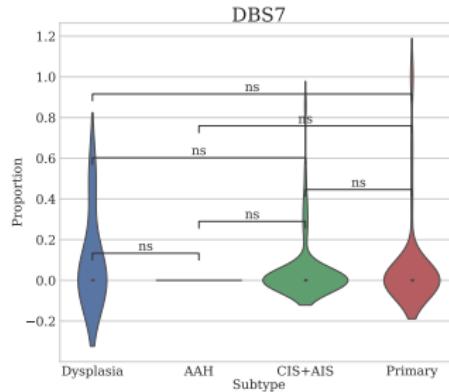
(b) Relative

Figure: DBS2 Signature in LUSC

# DBS in LUAD IV



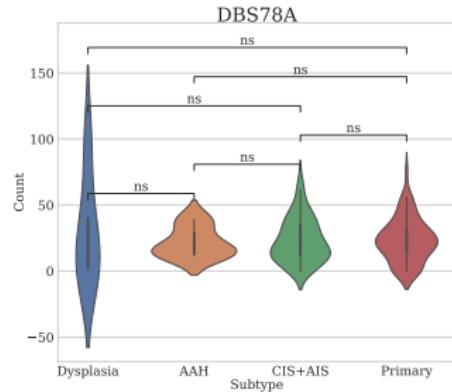
(a) Absolute



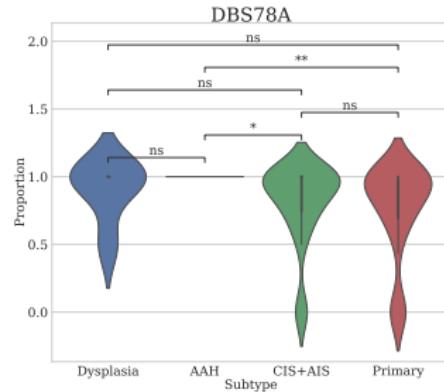
(b) Relative

Figure: DBS7 Signature in LUSC

# DBS in LUAD V



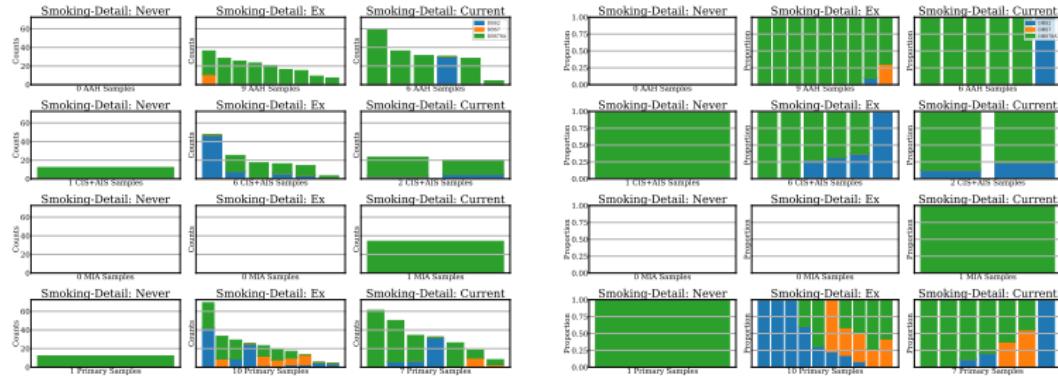
(a) Absolute



(b) Relative

Figure: DBS78A Signature in LUSC

# DBS in LUAD with Smoking I

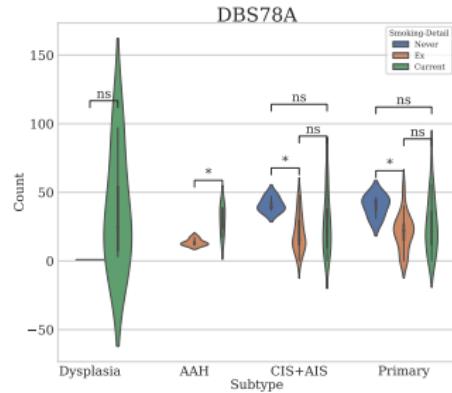


(a) Absolute

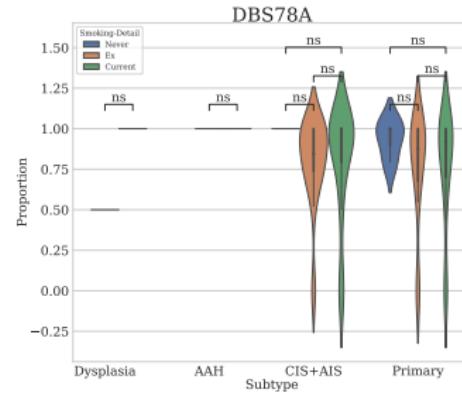
(b) Relative

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

# DBS in LUAD with Smoking II



(a) Absolute



(b) Relative

Figure: DBS78A Signature in LUSC in Smoking

# DBS in LUAD with Recurrence

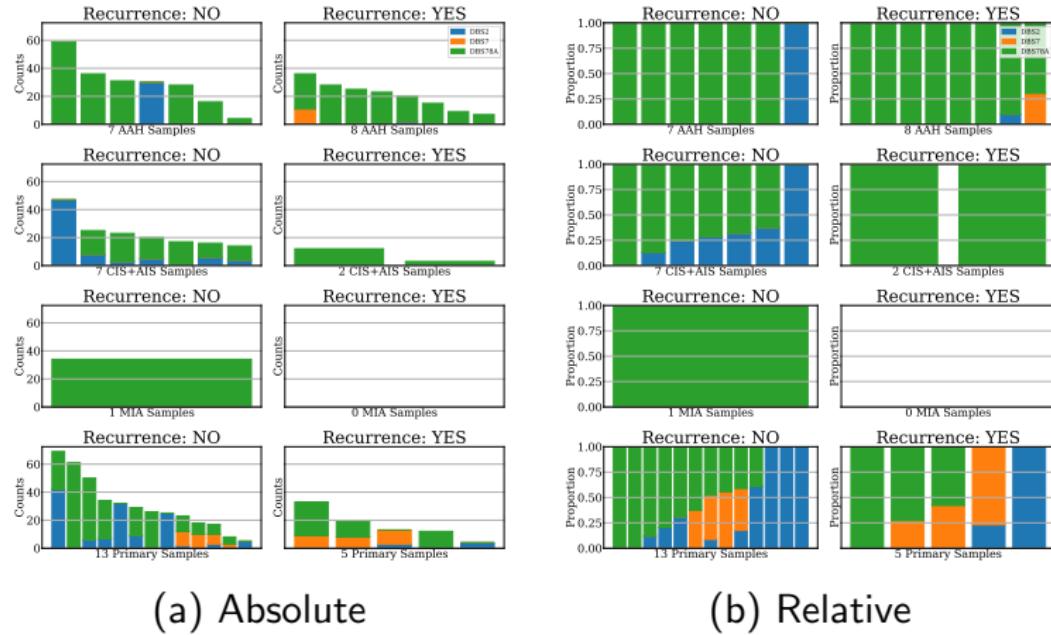


Figure: DBS Bar Plot by Cancer Subtype & Recurrence in LUAD

## 4. Results

### 4.8. Discovery of Mutational Signature

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

# Indel signatures I

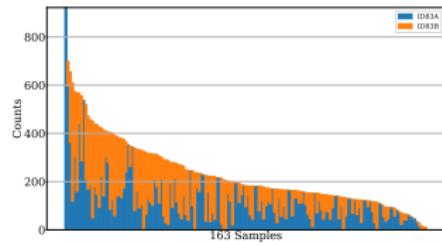
ID83A

content...

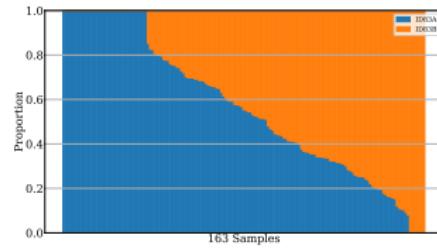
ID83B

content...

# Indels in LUSC I



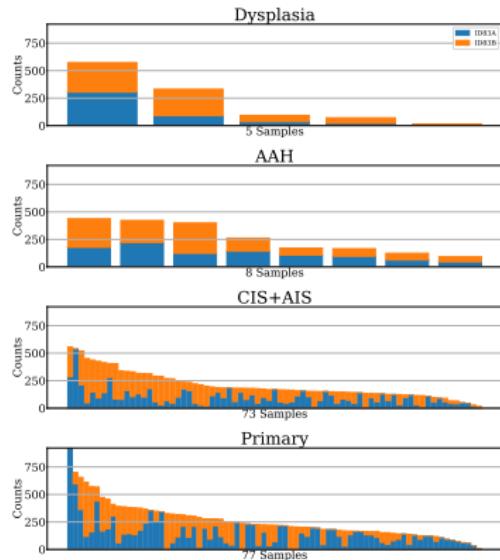
(a) Absolute



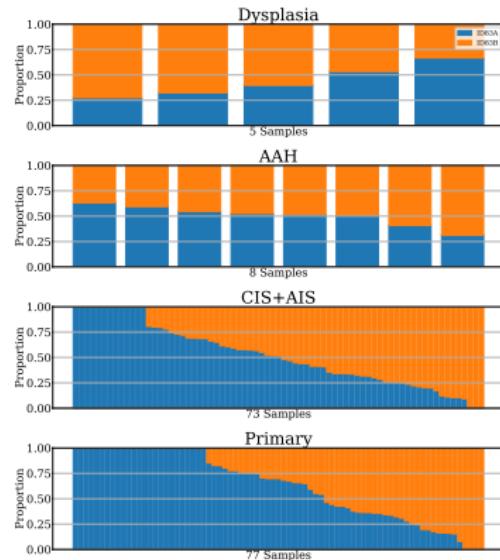
(b) Relative

Figure: Indel Bar Plot in LUSC

# Indels in LUSC II



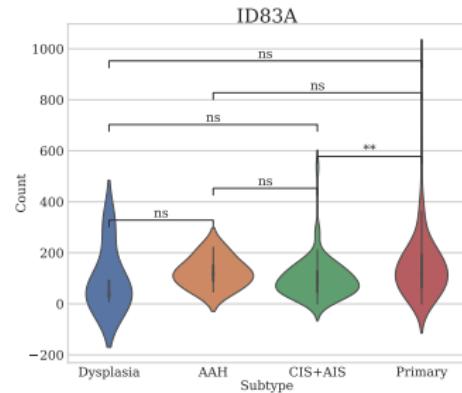
(a) Absolute



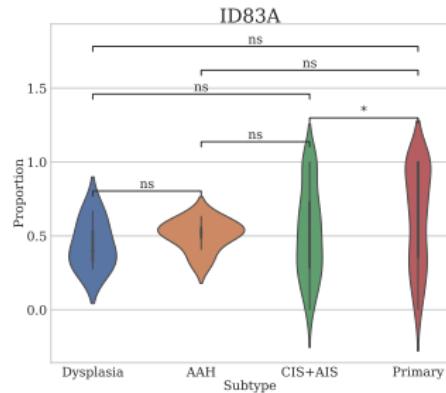
(b) Relative

Figure: Indel Bar Plot by Cancer Subtype in LUSC

# Indels in LUSC III



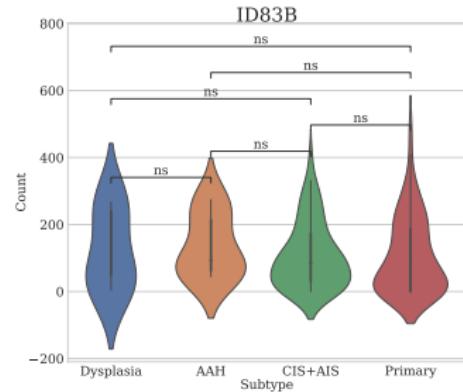
(a) Absolute



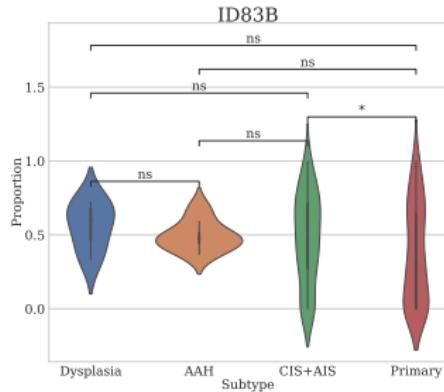
(b) Relative

Figure: Indel83A Signature in LUSC

# Indels in LUSC IV



(a) Absolute



(b) Relative

Figure: Indel83B Signature in LUSC

# Indel in LUSC with Smoking I

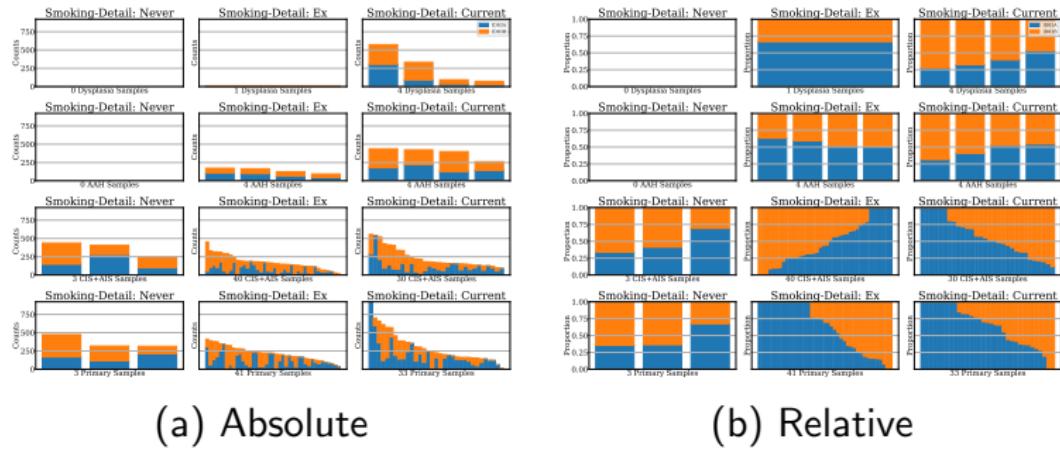
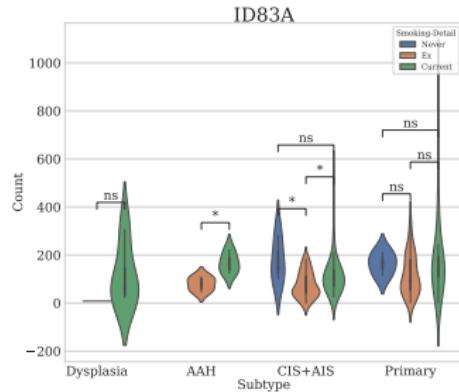
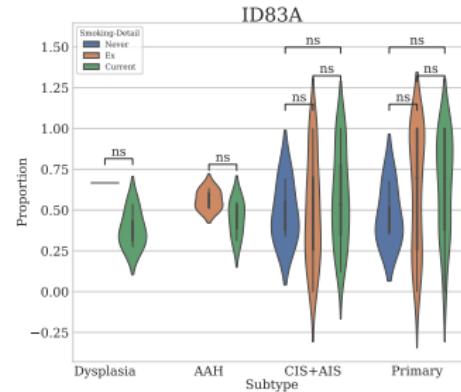


Figure: Indel Bar Plot by Cancer Subtype & Smoking in LUSC

# Indel in LUSC with Smoking II



(a) Absolute



(b) Relative

Figure: Indel83A Signature in LUSC with Smoking

# Indel in LUSC with Recurrence

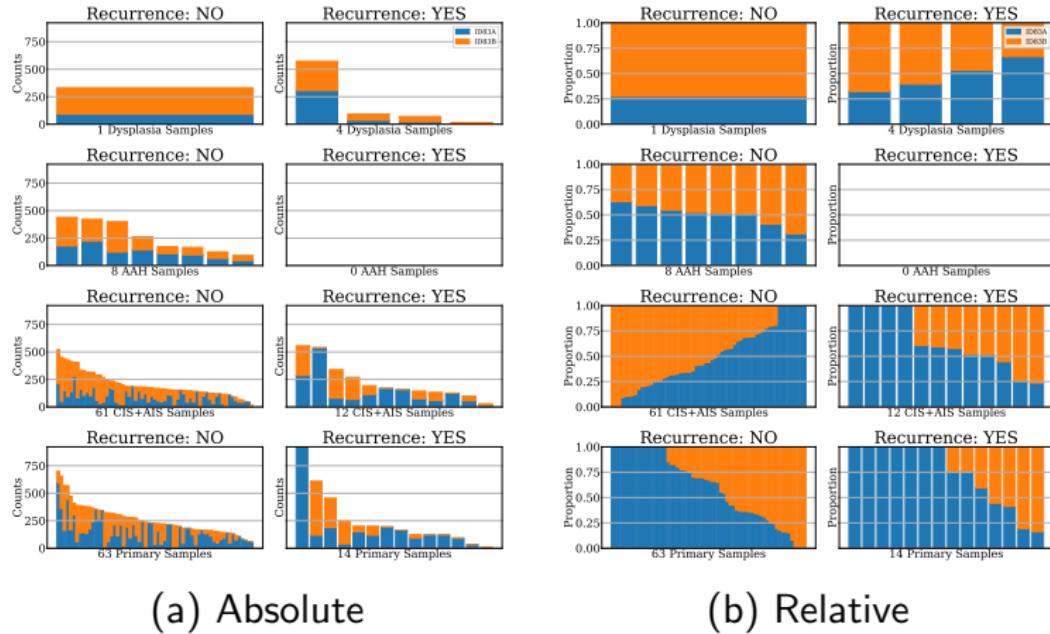
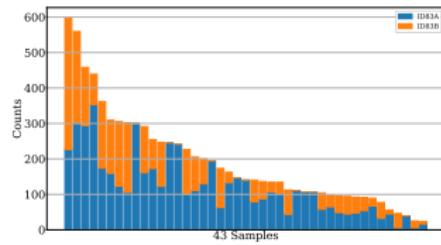
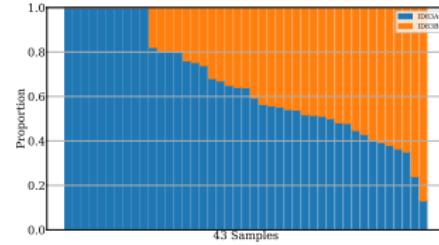


Figure: Indel Bar Plot by Cancer Subtype & Recurrence in LUSC

# Indels in LUAD I



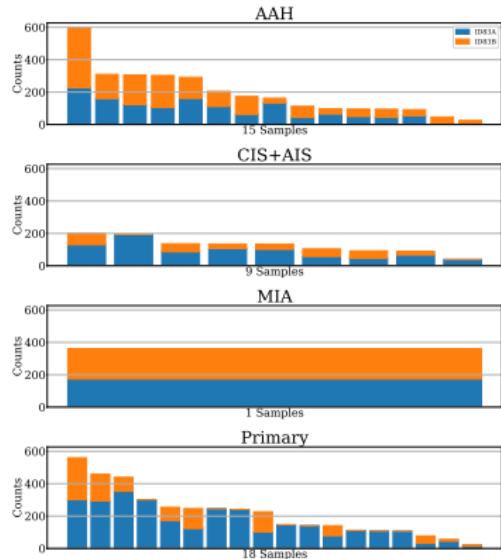
(a) Absolute



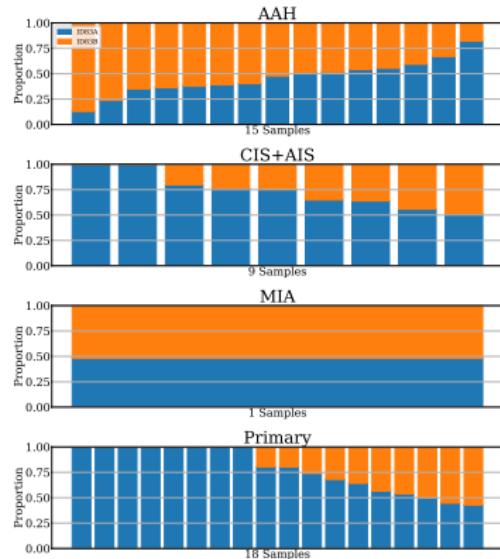
(b) Relative

Figure: Indel Bar Plot in LUAD

# Indels in LUAD II



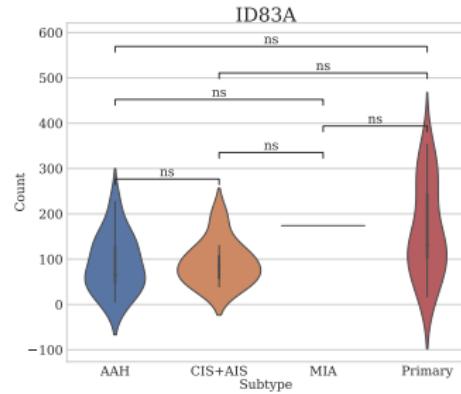
(a) Absolute



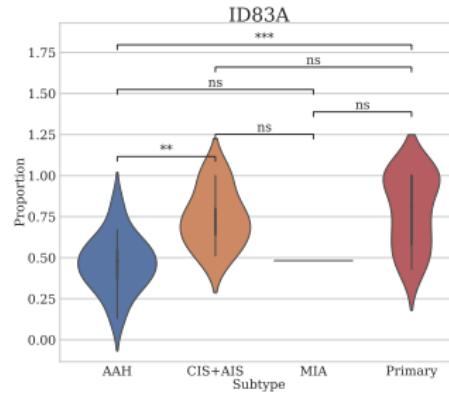
(b) Relative

Figure: Indel Bar Plot by Cancer Subtype in LUAD

# Indels in LUAD III



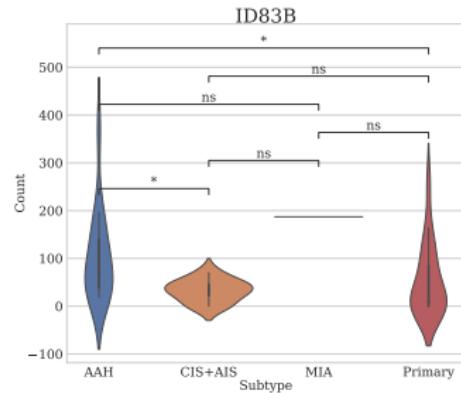
(a) Absolute



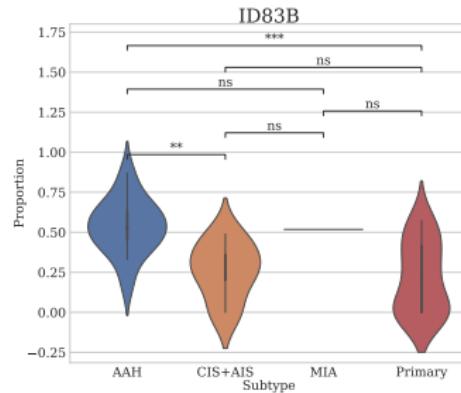
(b) Relative

Figure: Indel83A Signature in LUAD

# Indels in LUAD IV



(a) Absolute



(b) Relative

Figure: Indel83B Signature in LUAD

# Indel in LUAD with Smoking

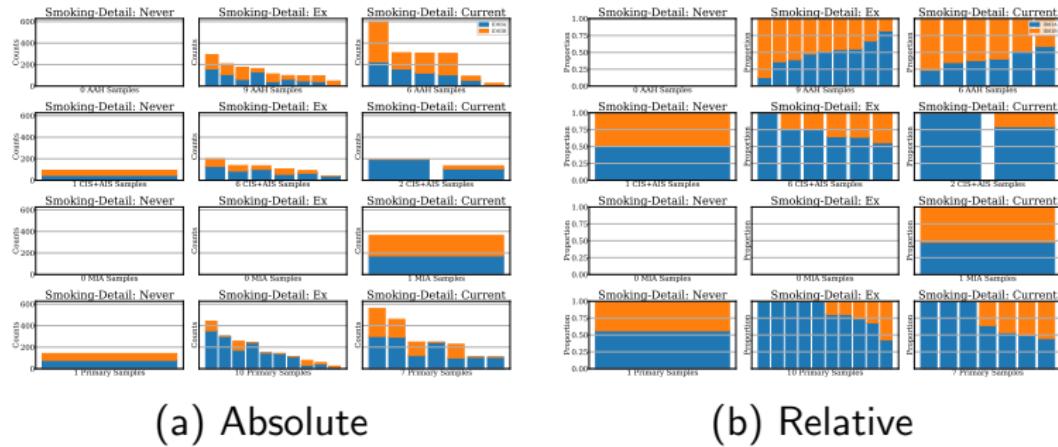


Figure: Indel Bar Plot by Cancer Subtype & Smoking in LUAD

# Indel in LUAD with Recurrence

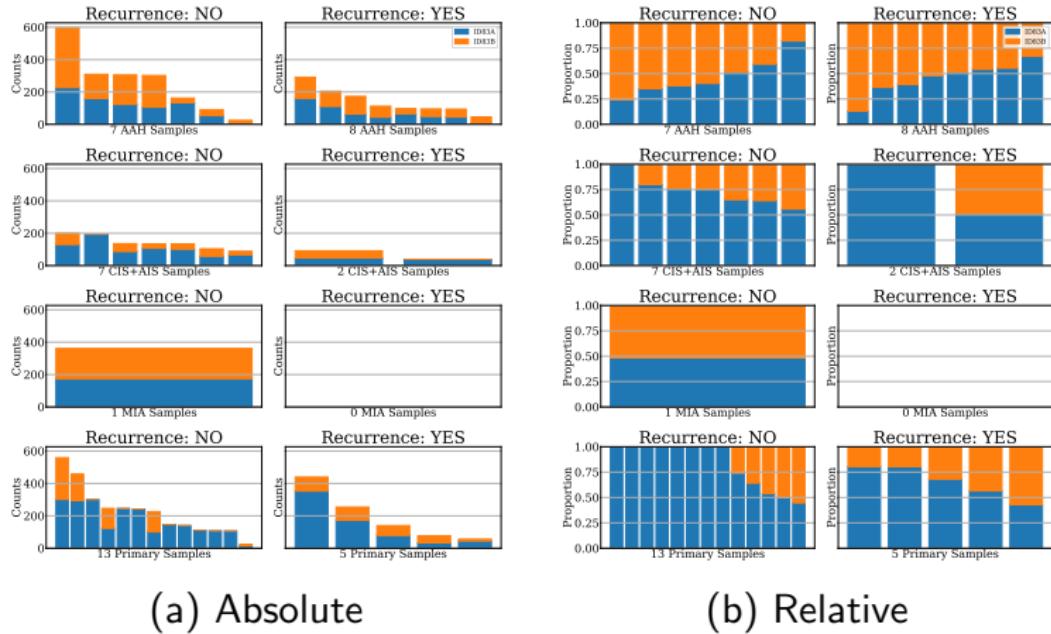


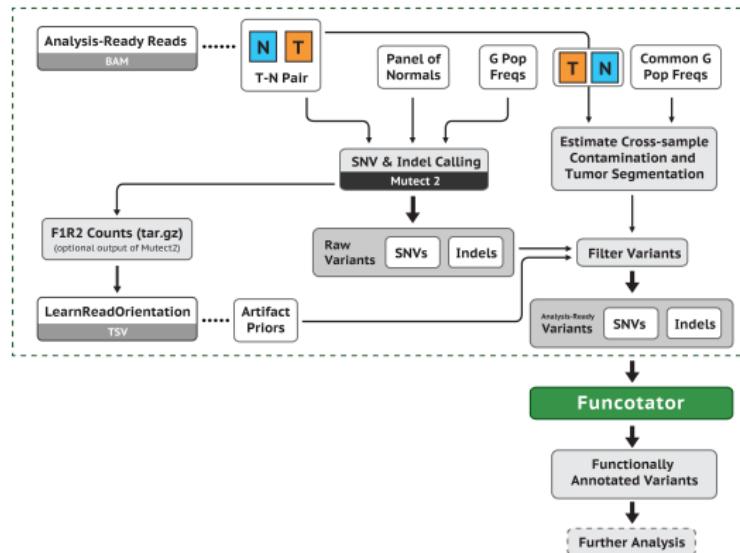
Figure: Indel Bar Plot by Cancer Subtype & Recurrence in LUAD

# Findings in Mutation Signature

## 4. Results

### 4.9. Clinical Data with Point Mutation

# Mutect2?



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

## 4. Results

### 4.9. Clinical Data with Point Mutation

#### 4.9.1. For Smoking

# LUSC with Smoking

Table: LUSC WES Data with Smoking

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

# Clinical Data about LUSC for Smoking I

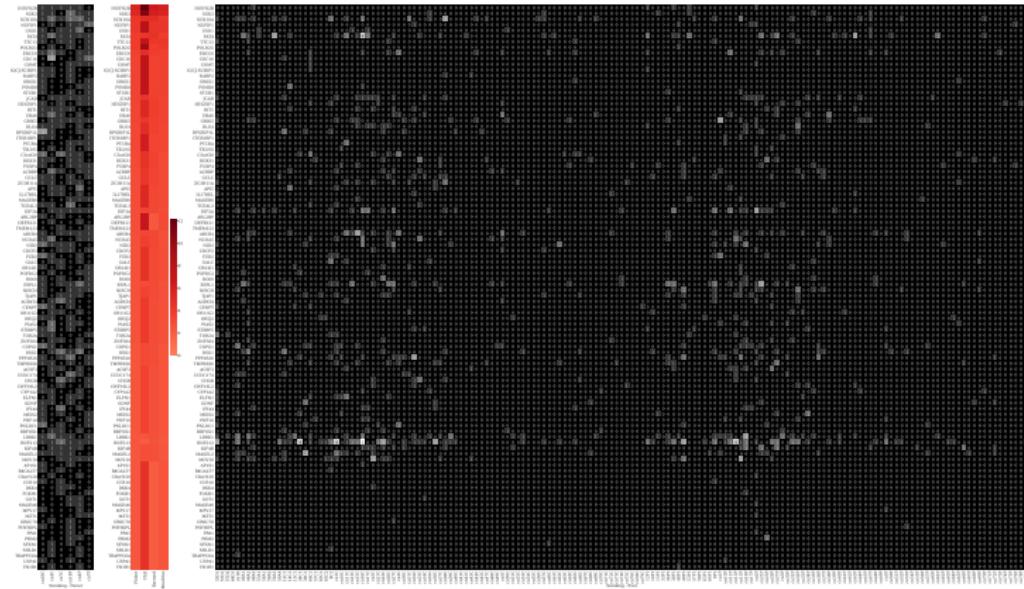


Figure: Clinical Data about LUSC for Smoking

# Clinical Data about LUSC for Smoking II

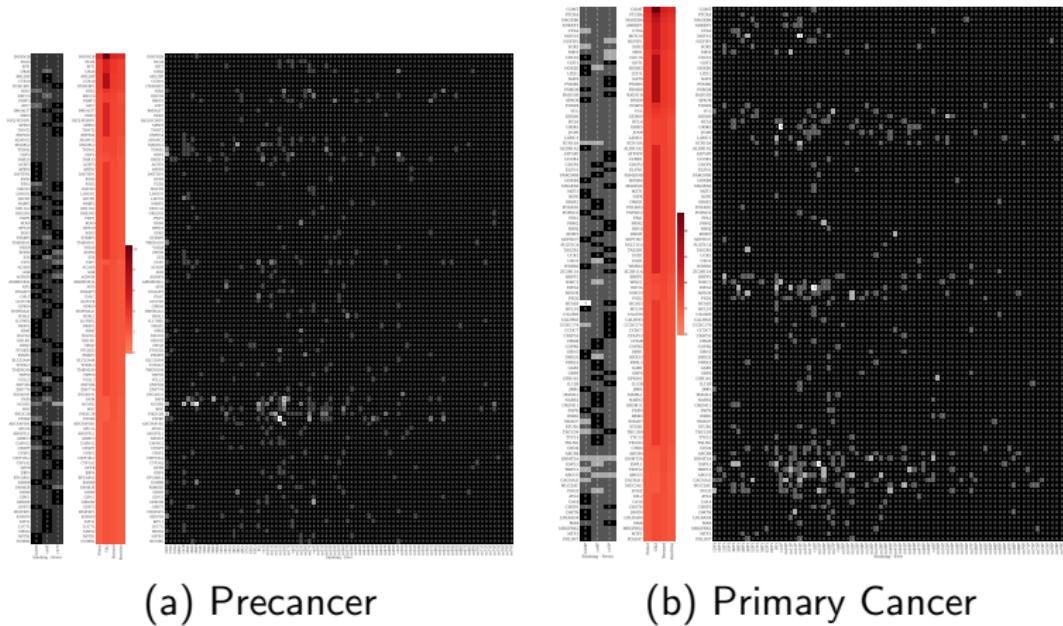


Figure: Clinical Data about LUSC for Smoking with Precancer/Primary

# Notable genes in LUSC for Smoking I

## INSYN2B

- ① INSYN2B is the best indicator for Smoking.
- ② INSYN2B is the best indicator for Smoking in Precancer.

## COMT

- ① COMT is the best indicator for Smoking in Primary.

# LUAD with Smoking

Table: LUAD WES Data with Smoking

Smoking?	Stage	Number of Samples
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

# Clinical Data about LUAD for Smoking I

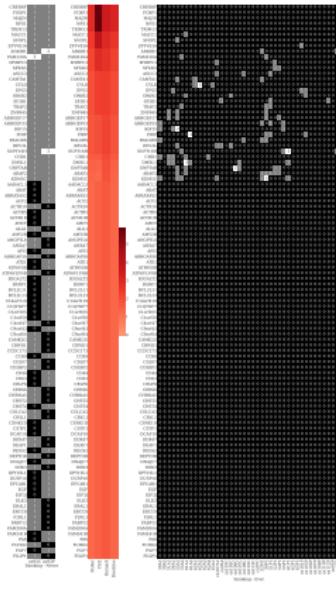
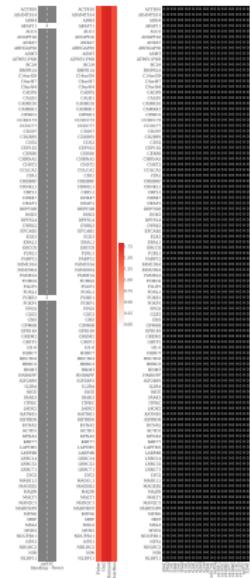
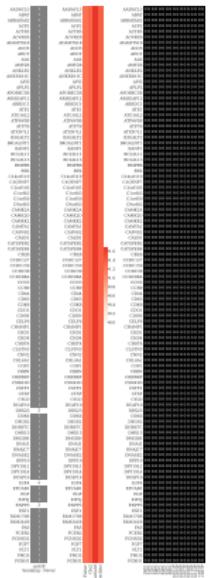


Figure: Clinical Data about LUAD for Smoking

# Clinical Data about LUAD for Smoking II



(a) Precancer



(b) Primary Cancer

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

# Notable genes in LUAD for Smoking I

## CREBRF

- ① CREBRF is the best indicator for Smoking.

## ACTR10

- ① ACTR10 is the best indicator for Smoking in Precancer.

## AADACL3

- ① AADACL3 is the best indicator for Smoking in Primary.

## 4. Results

### 4.9. Clinical Data with Point Mutation

#### 4.9.2. For Recurrence

# LUSC with Recurrence

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	

# Clinical Data about LUSC for Recurrence I

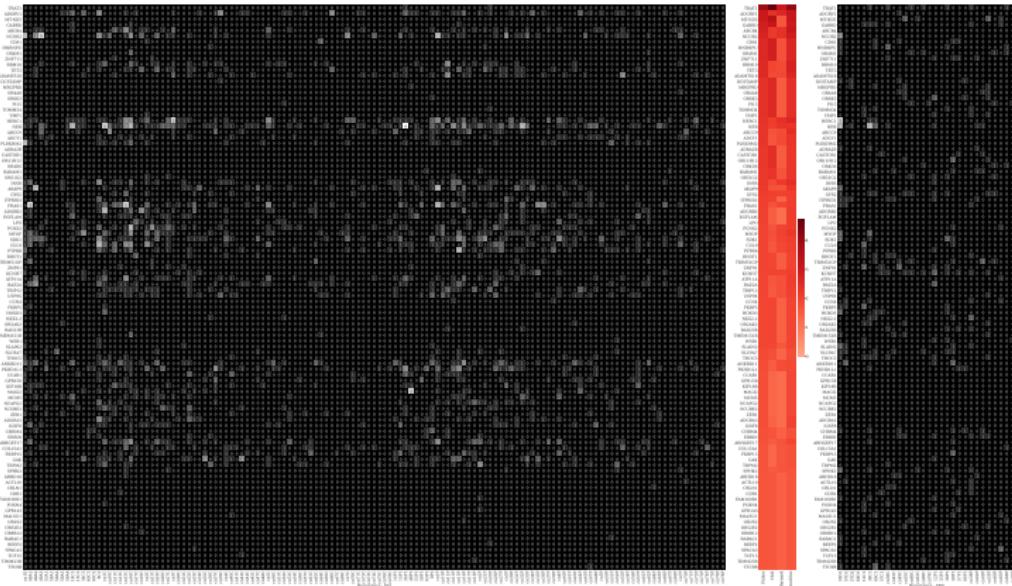


Figure: Clinical Data about LUSC for Recurrence

# Clinical Data about LUSC for Recurrence II

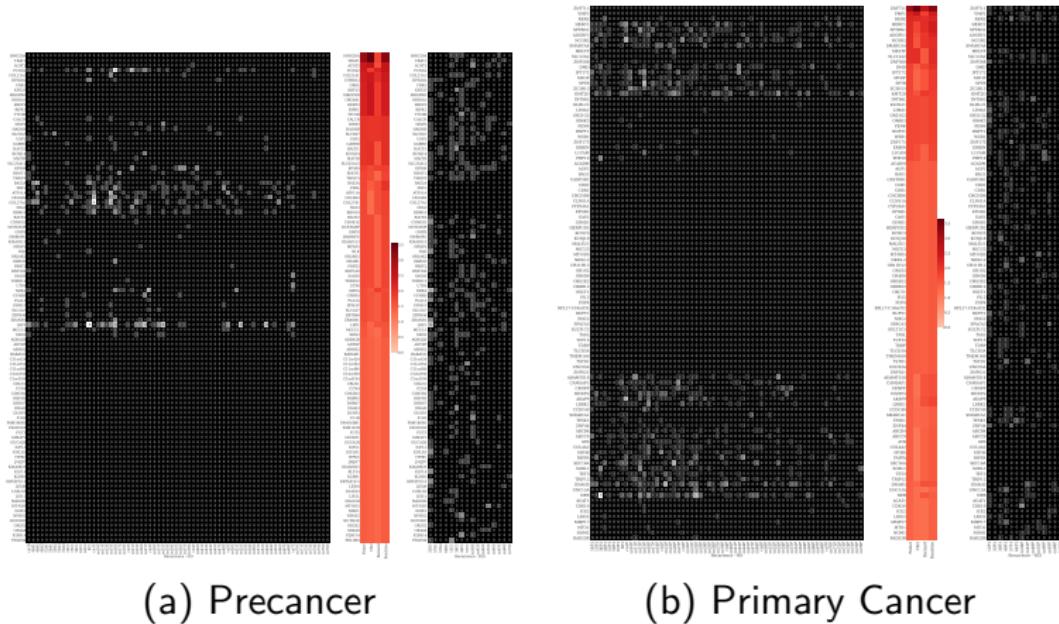


Figure: Clinical Data about LUSC for Recurrence with Precancer/Primary

# Notable genes in LUSC with Recurrence I

## TRAT1

- ① TRAT1 is the best indicator for Recurrence.

## HMG20A

- ① HMG20A is the best indicator for Recurrence in Precancer.

## ZNF711

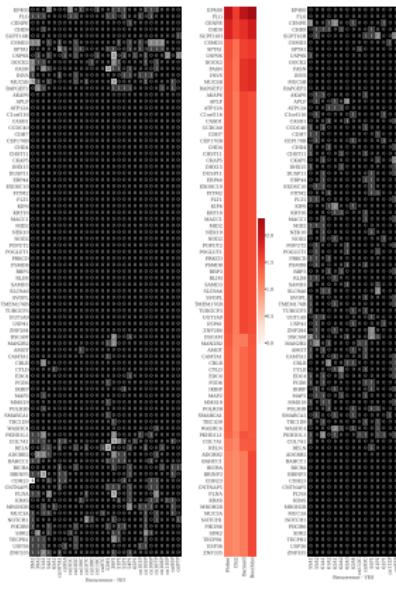
- ① ZNF711 is the best indicator for Recurrence in Primary.

# LUAD with Recurrence

Table: LUAD WES Data with Recurrence

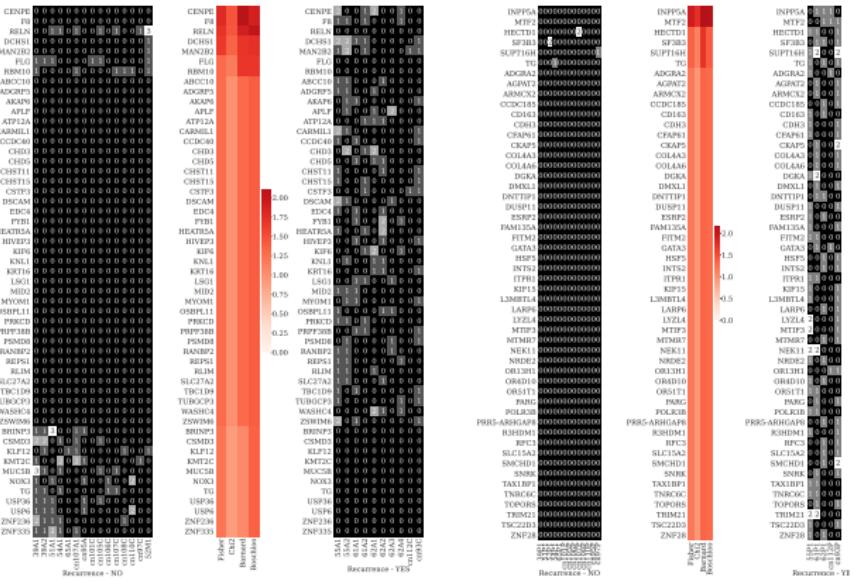
Recurrence?	Stage	Number of Samples	
		Normal	Affected
Recurrence	Normal	5	5
	AAH	8	8
	CIS+AIS	2	2
	Primary	5	5
	Total	20	20
Non-recurrence	Normal	13	13
	AAH	7	7
	CIS+AIS	7	7
	MIA	1	1
	Primary	13	13
	Total	41	41

## Clinical Data about LUAD for Recurrence I



## Figure: Clinical Data about LUAD for Recurrence

## Clinical Data about LUAD for Recurrence II



(a) Precancer

(b) Primary Cancer

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

# Notable genes in LUSC with Recurrence I

## EP400

- ① EP400 is the best indicator for Recurrence.

## CENPE

- ① CENPE is the best indicator for Recurrence in Precancer.

## INPP5A

- ① INPP5A is the best indicator for Recurrence in Primary.

# Findings in Clinical Data with Point Mutations

## 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.
- Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Aparicio, S. A., Behjati, S., Biankin, A. V., ... others (2013). Signatures of mutational processes in human cancer. *Nature*, 500(7463), 415–421.
- Alnabulsi, A., Swan, R., Cash, B., Alnabulsi, A., & Murray, G. I. (2017). The differential expression of omega-3 and omega-6 fatty acid metabolising enzymes in colorectal cancer and its prognostic significance. *British journal of cancer*, 116(12), 1612–1620.
- Andrews, S., Krueger, F., Segonds-Pichon, A., Biggins, L., Krueger, C., & Wingett, S. (2012, January). *FastQC*. Babraham Institute. Babraham, UK.

## References II

- Behrend, S. J., Giotopoulou, G. A., Spella, M., & Stathopoulos, G. T. (2021). A role for club cells in smoking-associated lung adenocarcinoma. *European Respiratory Review*, 30(162).
- 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.
- Beyer, M., & Schultze, J. L. (2006). Regulatory t cells in cancer. *Blood*, 108(3), 804–811.

## References III

- 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.
- Cailhier, J. F., Sawatzky, D. A., Kipari, T., Houlberg, K., Walbaum, D., Watson, S., ... others (2006). Resident pleural macrophages are key orchestrators of neutrophil recruitment in pleural inflammation. *American journal of respiratory and critical care medicine*, 173(5), 540–547.

## References IV

- Cal, S., Freije, J. M., López, J. M., Takada, Y., & Lopez-Otin, C. (2000). Adam 23/mdc3, a human disintegrin that promotes cell adhesion via interaction with the  $\alpha v \beta 3$  integrin through an rgd-independent mechanism. *Molecular biology of the cell*, 11(4), 1457–1469.
- 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.
- Chen, J.-M., Férec, C., & Cooper, D. N. (2013). Patterns and mutational signatures of tandem base substitutions causing human inherited disease. *Human mutation*, 34(8), 1119–1130.
- Chen, W., Hang, Y., Xu, W., Wu, J., Chen, L., Chen, J., ... Wang, H. (2019). Blacat1 predicts poor prognosis and serves as oncogenic lncrna in small-cell lung cancer. *Journal of cellular biochemistry*, 120(2), 2540–2546.

## References V

- 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 VI

- Deckers, J., Hammad, H., & Hoste, E. (2018). Langerhans cells: sensing the environment in health and disease. *Frontiers in immunology*, 9, 93.
- 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.
- Dhar, M., Sepkovic, D. W., Hirani, V., Magnusson, R. P., & Lasker, J. M. (2008). Omega oxidation of 3-hydroxy fatty acids by the human cyp4f gene subfamily enzyme cyp4f11. *Journal of lipid research*, 49(3), 612–624.
- 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.

## References VII

- Dosenbach, N. U., Fair, D. A., Miezin, F. M., Cohen, A. L., Wenger, K. K., Dosenbach, R. A., ... others (2007). Distinct brain networks for adaptive and stable task control in humans. *Proceedings of the National Academy of Sciences*, 104(26), 11073–11078.
- Duhig, E., Clarke, B., Yang, I. A., Fong, K. M., Hunter, L., Windsor, M., ... others (2014). Comprehensive molecular profiling of lung adenocarcinoma: the cancer genome atlas research network. *Nature*, 511(7511), 543–550.
- Edson, K. Z., Prasad, B., Unadkat, J. D., Suhara, Y., Okano, T., Guengerich, F. P., & Rettie, A. E. (2013). Cytochrome p450-dependent catabolism of vitamin k:  $\omega$ -hydroxylation catalyzed by human cyp4f2 and cyp4f11. *Biochemistry*, 52(46), 8276–8285.

## References VIII

- Erfani, N., Mehrabadi, S. M., Ghayumi, M. A., Haghshenas, M. R., Mojtabaei, Z., Ghaderi, A., & Amani, D. (2012). Increase of regulatory t cells in metastatic stage and cta-4 over expression in lymphocytes of patients with non-small cell lung cancer (nsclc). *Lung cancer*, 77(2), 306–311.
- 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.
- Goldstraw, P., Ball, D., Jett, J. R., Le Chevalier, T., Lim, E., Nicholson, A. G., & Shepherd, F. A. (2011). Non-small-cell lung cancer. *The Lancet*, 378(9804), 1727–1740.

## References IX

- Greczmiel, U., Kräutler, N. J., Pedrioli, A., Bartsch, I., Agnelli, P., Bedenikovic, G., ... Oxenius, A. (2017). Sustained t follicular helper cell response is essential for control of chronic viral infection. *Science Immunology*, 2(18).
- 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 X

- 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.
- Hu, C., Lv, H., Pan, G., Cao, H., Deng, Z., Hu, C., ... Zhou, J. (2011). The expression of adam23 and its correlation with promoter methylation in non-small-cell lung carcinoma. *International journal of experimental pathology*, 92(5), 333–339.
- Huang, F.-X., Chen, H.-J., Zheng, F.-X., Gao, Z.-Y., Sun, P.-F., Peng, Q., ... others (2019). Lncrna blacat1 is involved in chemoresistance of non-small cell lung cancer cells by regulating autophagy. *International journal of oncology*, 54(1), 339–347.

## References XI

- 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
- 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.

## References XII

- Kalsotra, A., Turman, C. M., Kikuta, Y., & Strobel, H. W. (2004). Expression and characterization of human cytochrome p450 4f11: Putative role in the metabolism of therapeutic drugs and eicosanoids. *Toxicology and applied pharmacology*, 199(3), 295–304.
- 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 XIII

- Le, P., Angers-Loustau, A., De Oliveira, R., Ajlan, A., Brassard, C., Dudley, A., ... others (2010). Drr drives brain cancer invasion by regulating cytoskeletal-focal adhesion dynamics. *Oncogene*, 29(33), 4636–4647.
- 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., Brady, S. W., Ma, X., Shen, S., Zhang, Y., Li, Y., ... others (2020). Therapy-induced mutations drive the genomic landscape of relapsed acute lymphoblastic leukemia. *Blood*, 135(1), 41–55.
- 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.

## References XIV

- 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, C., Song, H., Huang, C., Yao, E., Gacayan, R., Xu, S.-M., & Chuang, P.-T. (2012). Alveolar type ii cells possess the capability of initiating lung tumor development. *PloS one*, 7(12), e53817.
- 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, sftpb, 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 XV

- 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*.
- Mason, R. J. (2006). Biology of alveolar type ii cells. *Respirology*, 11, S12–S15.
- 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.

## References XVI

- Minna, J. D., Roth, J. A., & Gazdar, A. F. (2002). Focus on lung cancer. *Cancer cell*, 1(1), 49–52.
- Modi, B. G., Neustadter, J., Binda, E., Lewis, J., Filler, R. B., Roberts, S. J., ... others (2012). Langerhans cells facilitate epithelial dna damage and squamous cell carcinoma. *Science*, 335(6064), 104–108.
- 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.
- Nik-Zainal, S., Alexandrov, L. B., Wedge, D. C., Van Loo, P., Greenman, C. D., Raine, K., ... others (2012a). Mutational processes molding the genomes of 21 breast cancers. *Cell*, 149(5), 979–993.

## References XVII

- Nik-Zainal, S., Alexandrov, L. B., Wedge, D. C., Van Loo, P., Greenman, C. D., Raine, K., ... others (2012b). Mutational processes molding the genomes of 21 breast cancers. *Cell*, 149(5), 979–993.
- 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.
- Olingy, C. E., Dinh, H. Q., & Hedrick, C. C. (2019). Monocyte heterogeneity and functions in cancer. *Journal of leukocyte biology*, 106(2), 309–322.
- Ota, M., Mochizuki, S., Shimoda, M., Abe, H., Miyamae, Y., Ishii, K., ... Okada, Y. (2016). Adam 23 is downregulated in side population and suppresses lung metastasis of lung carcinoma cells. *Cancer science*, 107(4), 433–443.

## References XVIII

- 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.
- Pinto, A., Rega, A., Crother, T. R., & Sorrentino, R. (2012). Plasmacytoid dendritic cells and their therapeutic activity in cancer. *Oncoimmunology*, 1(5), 726–734.
- Poonia, B., Ayithan, N., Nandi, M., Masur, H., & Kottilil, S. (2018a). Hbv induces inhibitory fcrl receptor on b cells and dysregulates b cell-t follicular helper cell axis. *Scientific reports*, 8(1), 1–14.
- Poonia, B., Ayithan, N., Nandi, M., Masur, H., & Kottilil, S. (2018b). Hbv induces inhibitory fcrl receptor on b cells and dysregulates b cell-t follicular helper cell axis. *Scientific reports*, 8(1), 1–14.

## References XIX

- Rawlins, E. L., & Hogan, B. L. (2008). Ciliated epithelial cell lifespan in the mouse trachea and lung. *American Journal of Physiology-Lung Cellular and Molecular Physiology*, 295(1), L231–L234.
- Riester, M., Singh, A. P., Brannon, A. R., Yu, K., Campbell, C. D., Chiang, D. Y., & Morrissey, M. P. (2016). Purecn: copy number calling and snv classification using targeted short read sequencing. *Source code for biology and medicine*, 11(1), 1–13.
- Rokicki, W., Rokicki, M., Wojtacha, J., & Dżeljijli, A. (2016). The role and importance of club cells (clara cells) in the pathogenesis of some respiratory diseases. *Kardiochirurgia i torakochirurgia polska= Polish journal of cardio-thoracic surgery*, 13(1), 26.
- 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 XX

- 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.
- Sica, A., Allavena, P., & Mantovani, A. (2008). Cancer related inflammation: the macrophage connection. *Cancer letters*, 267(2), 204–215.
- 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.

## References XXI

- Spella, M., Lilis, I., Pepe, M. A., Chen, Y., Armaka, M., Lamort, A.-S., ... others (2019). Club cells form lung adenocarcinomas and maintain the alveoli of adult mice. *Elife*, 8, e45571.
- 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.
- Sutherland, K. D., Proost, N., Brouns, I., Adriaensen, D., Song, J.-Y., & Berns, A. (2011). Cell of origin of small cell lung cancer: inactivation of trp53 and rb1 in distinct cell types of adult mouse lung. *Cancer cell*, 19(6), 754–764.

## References XXII

- 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 XXIII

- Vermi, W., Soncini, M., Melocchi, L., Sozzani, S., & Facchetti, F. (2011). Plasmacytoid dendritic cells and cancer. *Journal of leukocyte biology*, 90(4), 681–690.
- 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.
- Wang, L., Darling, J., Zhang, J.-S., Liu, W., Qian, J., Bostwick, D., ... others (2000). Loss of expression of the drr 1 gene at chromosomal segment 3p21. 1 in renal cell carcinoma. *Genes, Chromosomes and Cancer*, 27(1), 1–10.

## References XXIV

- 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.
- Xia, H.-W., Zhang, Z.-Q., Yuan, J., & Niu, Q.-L. (2021). Human recql5 promotes metastasis and resistance to cisplatin in non-small cell lung cancer. *Life Sciences*, 265, 118768.
- Ye, T., Yang, X., Liu, H., Lv, P., & Ye, Z. (2020). Long non-coding rna blacat1 in human cancers. *OncoTargets and therapy*, 13, 8263.
- Yin, J., Kwon, Y. T., Varshavsky, A., & Wang, W. (2004). Recql4, mutated in the rothmund–thomson and rapadilino syndromes, interacts with ubiquitin ligases ubr1 and ubr2 of the n-end rule pathway. *Human molecular genetics*, 13(20), 2421–2430.

## References XXV

- Zhang, J., Shao, J., Zhu, L., Zhao, R., Xing, J., Wang, J., ... Yu, K. (2017). Molecular profiling identifies prognostic markers of stage ia lung adenocarcinoma. *Oncotarget*, 8(43), 74846.
- Zhu, X., Chen, H., Yang, Y., Xu, C., Zhou, J., Zhou, J., & Chen, Y. (2018). Distinct prognosis of mrna expression of the five recq dna-helicase family members—recql, blm, wrn, recql4, and recql5—in patients with breast cancer. *Cancer management and research*, 10, 6649.