# 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*

2022-06-24

# Overview

# 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

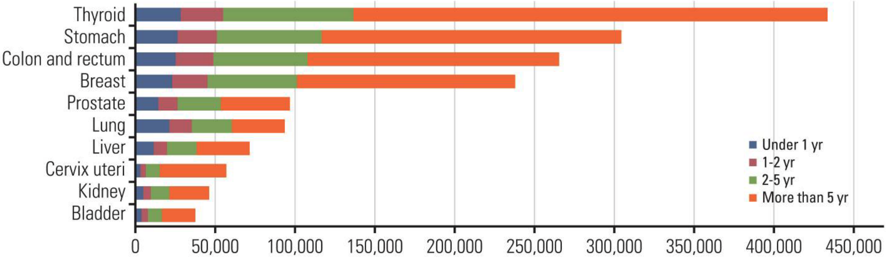



Figure: Common cancer survival rates (S. Hong et al., 2021)

## Survival rate (More than 5 year)

- Tyroid: 68.4 %
- Lung: 35.4 %

# Type of Lung Cancer I

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. ...

# Type of Lung Cancer II

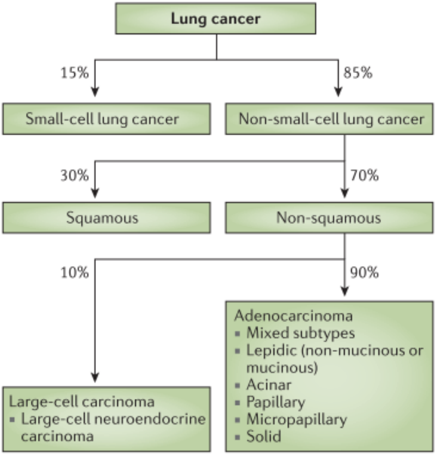



Figure: Lung cancer classification (Gridelli et al., 2015)

# 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

# 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

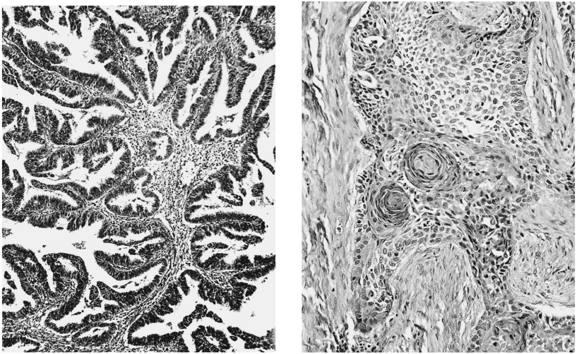
# LUSC

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.
- Amplication of NFE2L2, MYC, CDK6, MDM2, BCL2L1, and EYS.
- Deletion of FOXP1, PTEN, and NF1

# 1. Introduction

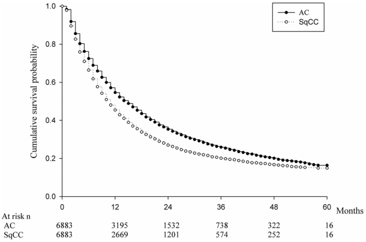
## 1.5. LUAD vs. LUSC



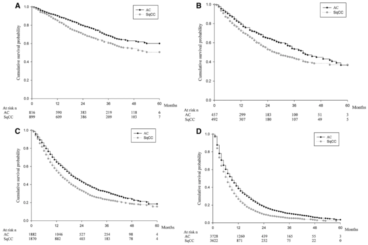
(a) LUAD (b) LUSC

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. $\because p < 0.001$

# 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 + {Dysplasia, AAH, CIS + AIS, 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 $n$=18)
  2. Normal → Dysplasia → CIS → LUSC (patient $n$=77)

# 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 |
|---|---|---|
| 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 |
|---|---|---|
| Recurrence | Normal | 5 |
| | AAH | 8 |
| | CIS+AIS | 2 |
| | Primary | 5 |
| | Total | 20 |
| Non-recurrence | Normal | 13 |
| | AAH | 7 |
| | CIS+AIS | 7 |
| | MIA | 1 |
| | Primary | 13 |
| | Total | 41 |

# WES Data Composition with Smoking I

Table: LUSC WES Data with Smoking

| Smoking? | Stage | Number of Samples |
|---|---|---|
| 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 |
|---|---|---|
| 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 |
|---|---|---|
| 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 with Recurrence I

Table: LUSC WTS Data with Recurrence

| Recurrence? | Stage | Number of Samples |
|---|---|---|
| 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 |
|---|---|---|
| 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 with Smoking I

Table: LUSC WTS Data with Smoking

| Smoking? | Stage | Number of Samples |
|---|---|---|
| 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

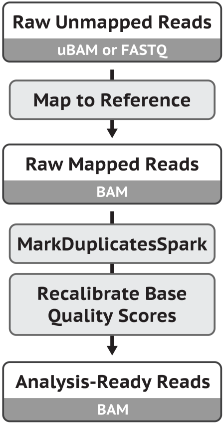



Figure: Data pre-processing for variant discovery (Van der Auwera et al., 2013; DePristo et al., 2011)

# Somatic short variant discovery

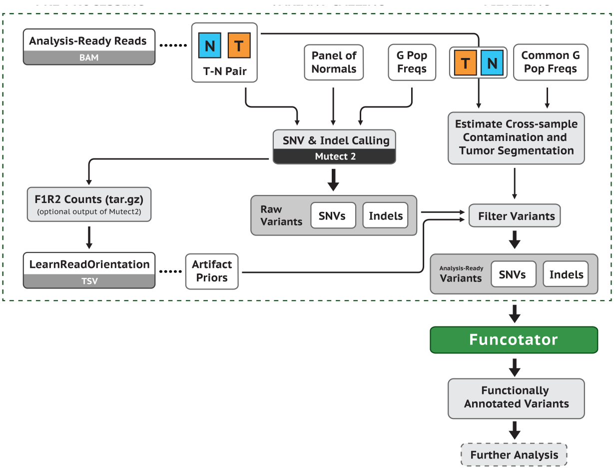



Figure: Somatic short variant (SNVs + Indels) discovery workflow (Van der Auwera et al., 2013; DePristo et al., 2011)

# Germline short variant discovery

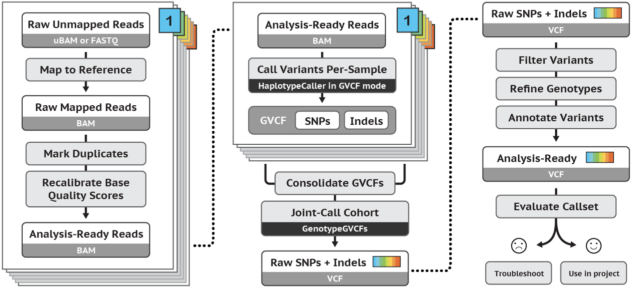



Figure: Germline short variant (SNVs + Indels) discovery workflow (Van der Auwera et al., 2013; DePristo et al., 2011)

# RNA-seq short variant discovery

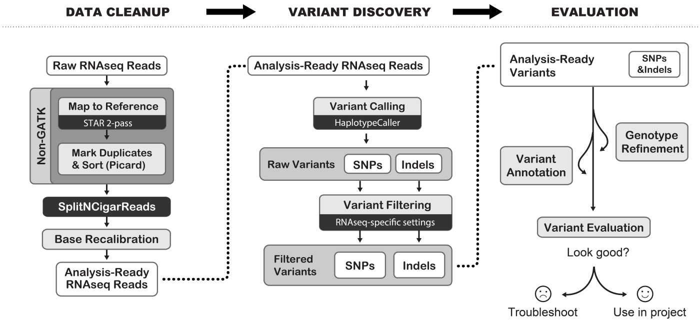



Figure: RNA-seq short variant (SNVs + Indels) discovery workflow (Van der Auwera et al., 2013; DePristo et al., 2011)

# 4. Results

# 4. Results

## 4.1. Quality Checks with Size

# File Size in WES Data

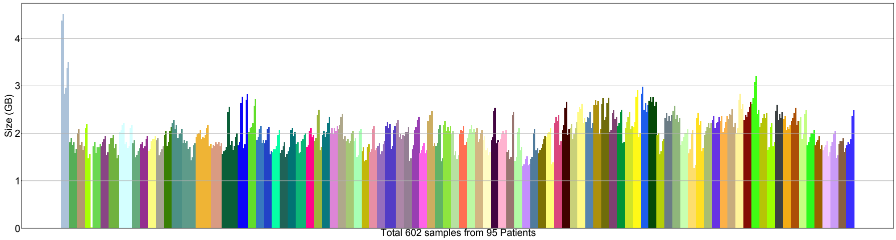



Figure: File Size Distribution in WES Data

# File Size in WTS Data

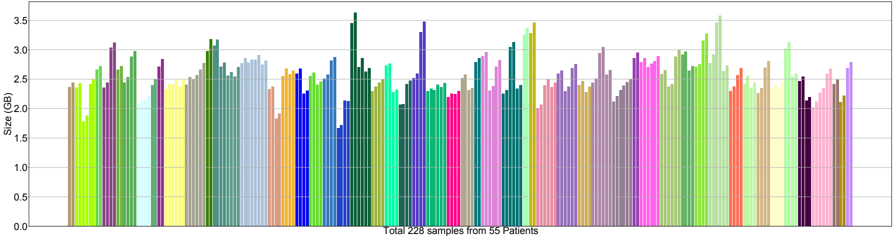



Figure: File Size Distribution in WTS Data

# 4. Results

## 4.2. Quality Checks with FastQC

# FastQC?

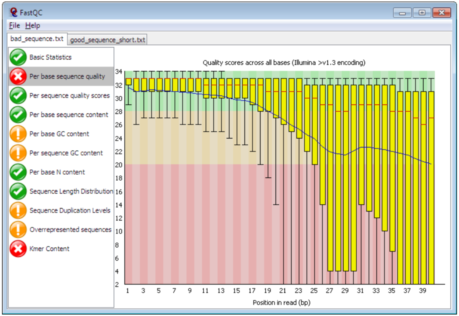



Figure: Example of FastQC Result (Andrews et al., 2012)

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

# FastQC on WES

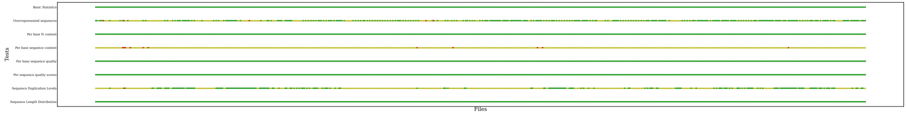
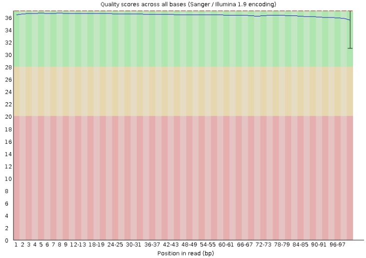



Figure: FastQC with WES data

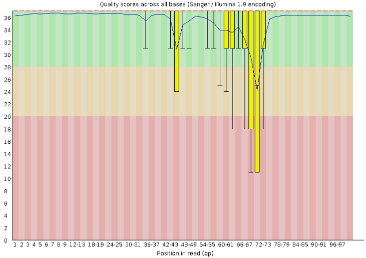
**Failure on 33P1 sample**

33P1 is excluded from further analyses.

# Failure on 33P1 I




(a) 33N




(b) 33P1

Figure: Per Base Sequence Quality Results

# Failure on 33P1 II

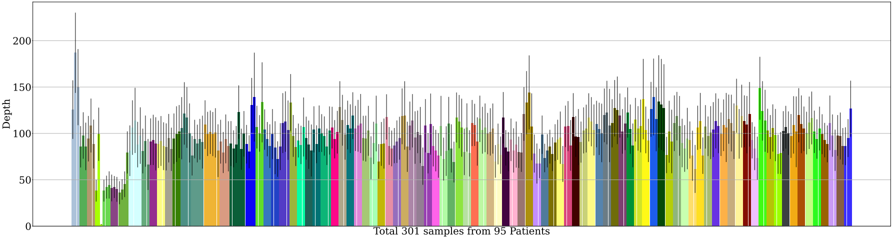



Figure: Coverage Depth Plot

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

# FastQC on WTS

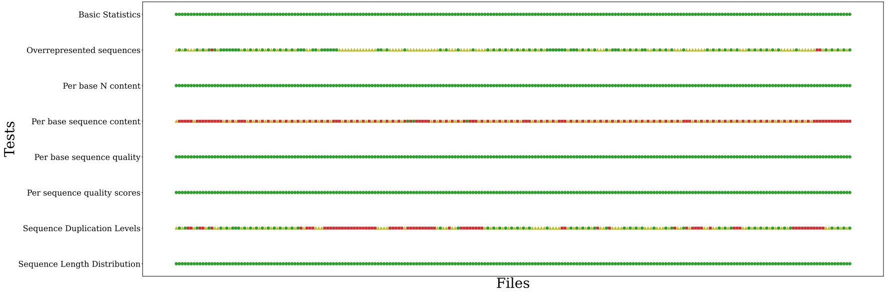



Figure: FastQC with WTS data

## All sample are good to analysis

∵ No sample has more than 5 failures.

# 4. Results

## 4.3. Quality Checks with Picard

# Picard?

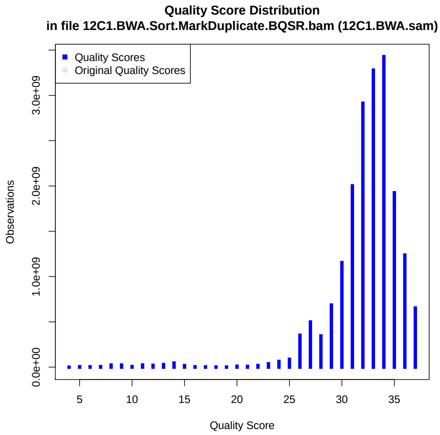



Figure: Quality Distribution of 12C1 sample

# Quality Distribution Plot

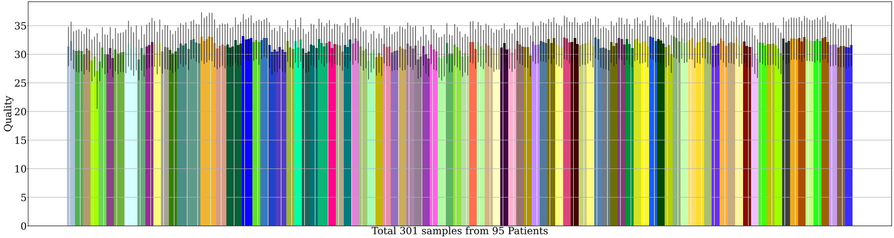



Figure: Quality Distribution by Samples

# Findings in Picard

# 4. Results

## 4.4. Copy Number Variation Analysis with PureCN

# PureCN?

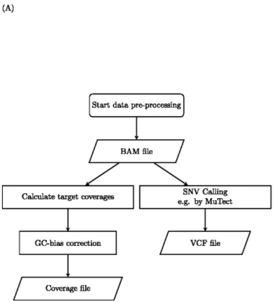
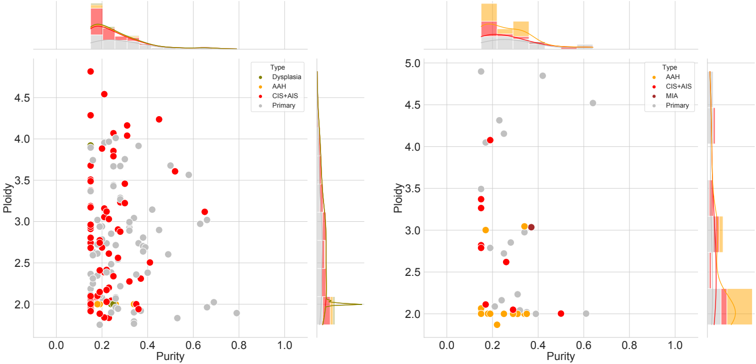



Figure: Flowchart of the PureCN data pre-processing pipeline (Riester et al., 2016)

# Purity & Ploidy on WES




(a) LUSC Samples (b) LUAD Samples

Figure: Cellularity and Ploidy from PureCN

# LUSC in CNV Plot I

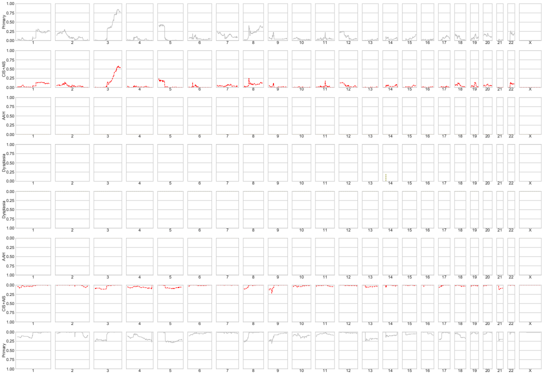


Figure: LUSC in CNV Plot

# LUSC in CNV Plot II

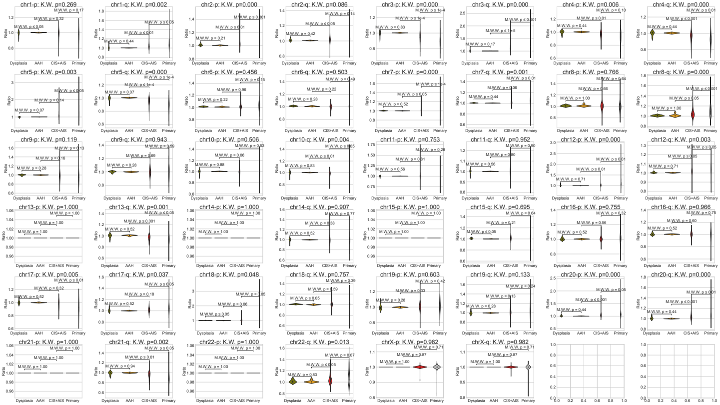



Figure: LUSC in Violin Plots

# LUSC with Recurrence in CNV Plot I

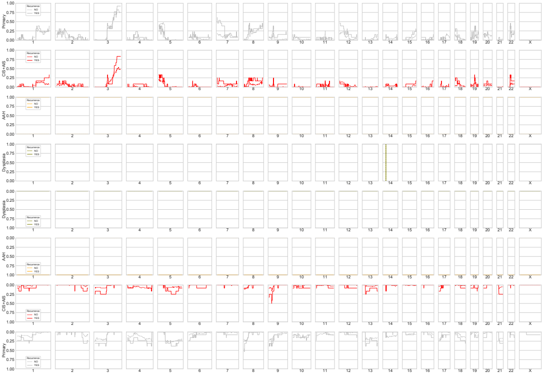


Figure: LUSC with Recurrence in CNV Plot

# LUSC with Recurrence in CNV Plot II

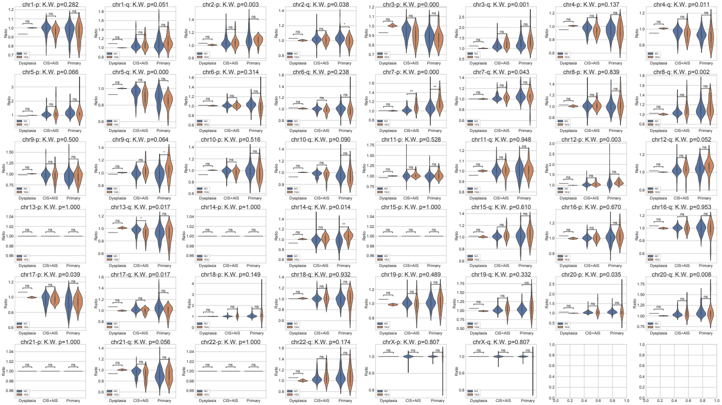



Figure: LUSC with Recurrence in Violin Plots

# LUSC with Smoking in CNV Plot I

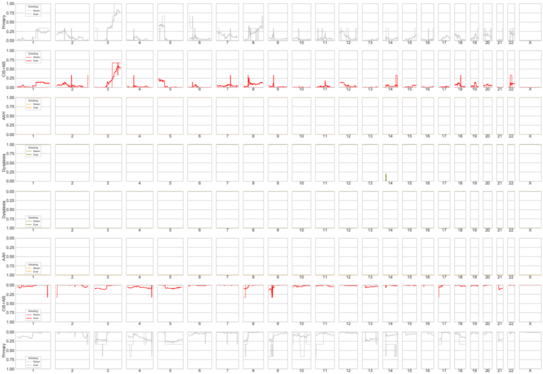


Figure: LUSC with Smoking in CNV Plot

# LUSC with Smoking in CNV Plot II

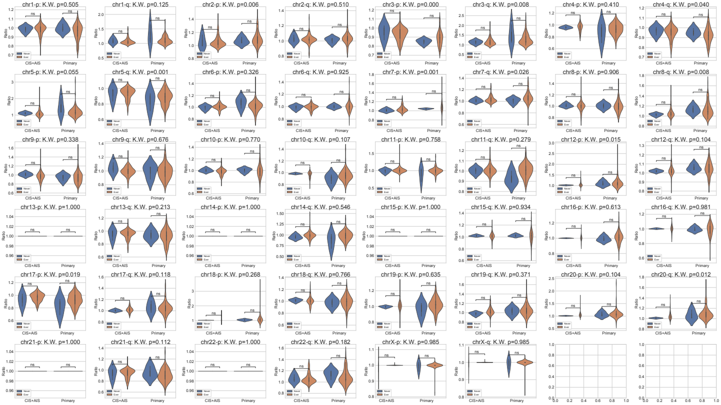



Figure: LUSC with Smoking in Violin Plots

# Findings in PureCN with LUSC I

## Chr2 p-arm gain

1. 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 |

# Findings in PureCN with LUSC II

## 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

1. 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 ho... |
| BRD3 | bromodomain containing 3 |
| CNTRL | centriolin |
| FANCC | Fanconi anemia, complementation group C |
| GNAQ | guanine nucleotide binding protein (G protein),... |

# 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 p-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

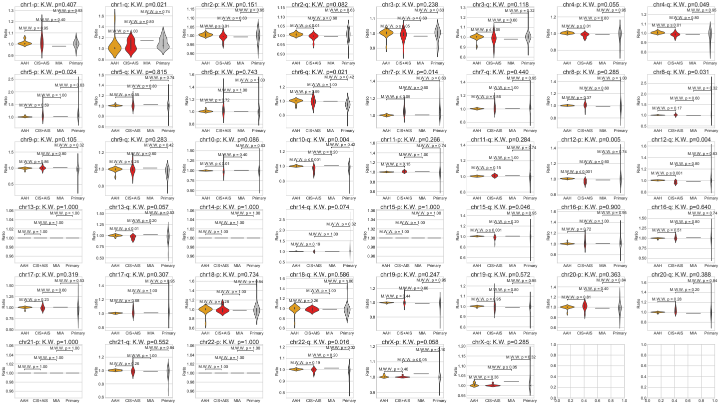



Figure: LUAD in Violin Plots

Figure: LUAD with Recurrence in CNV Plot

# LUAD with Recurrence in CNV Plot II

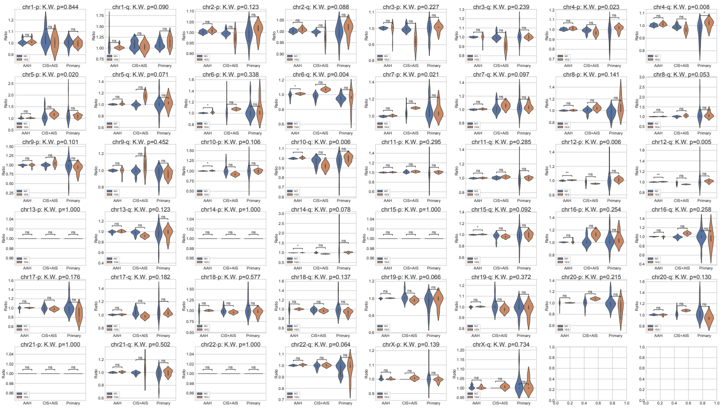



Figure: LUAD with Recurrence in Violin Plots

# LUAD with Smoking in CNV Plot I

Figure: LUAD with Smoking in CNV Plot

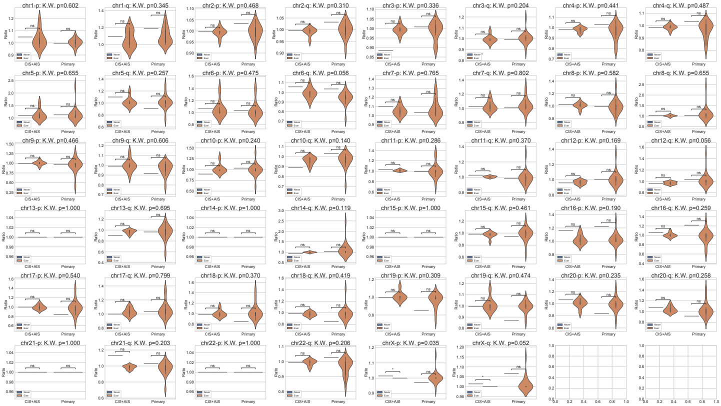



Figure: LUAD with Smoking in Violin Plots

# 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. Chr4 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 h... |
| FIP1L1 | FIP1 like 1 (S. cerevisiae) |

# Findings in PureCN with LUAD III

## Chr5 q-arm gain

1. Chr5 q-arm loss 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

## 4.5. Copy Number Variation Analysis with Gistic

# Gistic?

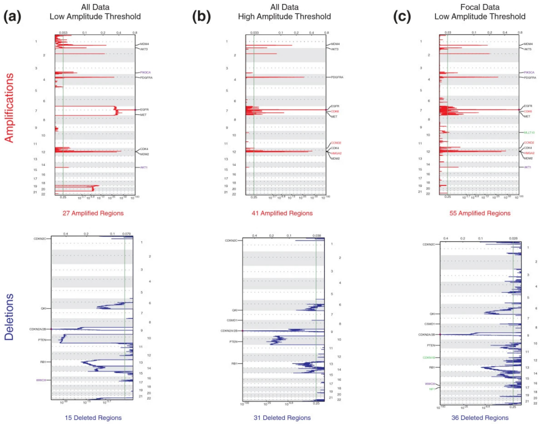



Figure: Effects of arm-level events on GISTIC results (Mermel et al., 2011)

# 4. Results

## 4.5. Copy Number Variation Analysis with Gistic

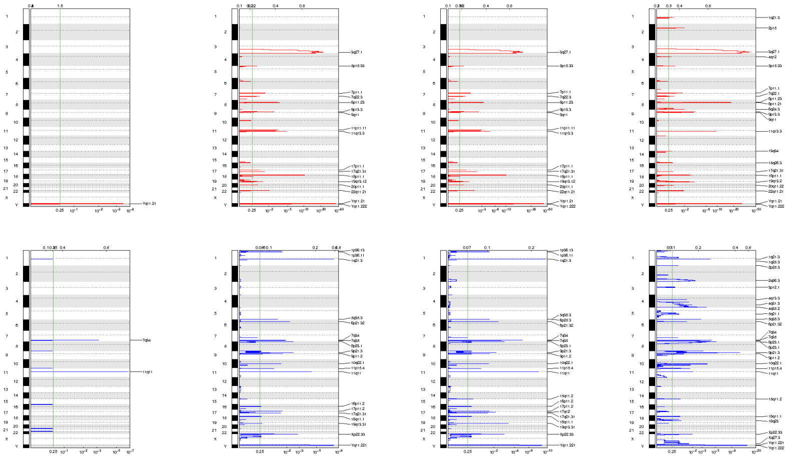
### 4.5.1. Gistic in LUSC

# LUSC 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 |

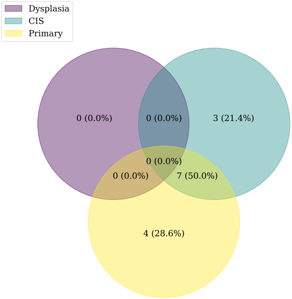
# Gistic in LUSC



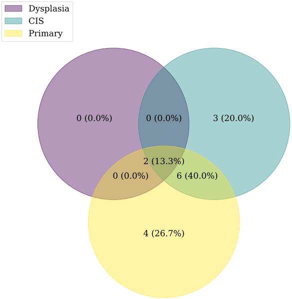
(a) Dysplasia (b) CIS (c) Precancer (d) Primary

Figure: Gistic results in LUSC

# Peaks in LUSC I




(a) Amplification




(b) Deletion

Figure: Venn Diagram among Peaks in LUSC – stage

# Peaks in LUSC II

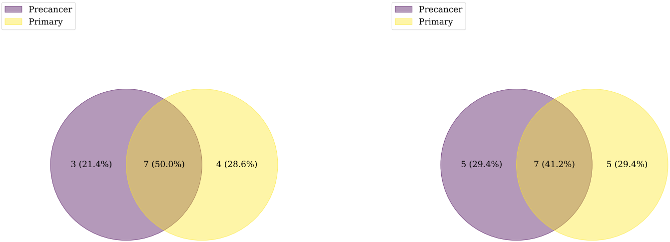
Table: Amplification Peaks in LUSC – stage

| Peaks | Dysplasia | CIS | Primary | CGC Genes |
|---|---|---|---|---|
| 1q21.3 | | | * | ARNT,MLLT11,S100A7,SETDB1,TPM3 |
| 2p15 | | | * | XPO1 |
| 3q27.1 | | * | * | |
| 4q12 | | | * | CHIC2,FIP1L1,KDR,KIT,PDGFRA |
| 5p15.33 | | * | * | SDHA,TERT |
| 7p11.1 | | * | * | |
| 7q22.1 | | | * | CUX1,TRRAP |
| 7q22.3 | | * | | |
| 8p11.21 | | | * | ANK1,HOOK3,IKBKB,KAT6A |
| 8p11.23 | | * | * | FGFR1,NSD3 |
| 8q24.3 | | | * | RECQL4 |
| 9p13.3 | | * | * | FANCG |
| 9q11 | | * | * | |
| 11p11.11 | | * | | |
| 11q13.3 | | * | * | CCND1 |
| 13q34 | | | * | |
| 15q26.3 | | | * | |
| 17p11.1 | | * | | |
| 17q21.31 | | * | * | BRCA1,ETV4 |
| 18p11.1 | | * | * | |
| 19q13.2 | | | * | AKT2,CD79A,CIC |
| 19q13.12 | | * | | |
| 20p11.1 | | * | | |
| 20q11.22 | | | * | |
| 22q11.21 | | * | * | CLTCL1,DGCR8,LZTR1,SEPT5 |

# Peaks in LUSC III

Table: Deletion Peaks in LUSC – stage

| Peaks | Dysplasia | CIS | Primary | CGC Genes |
|---|---|---|---|---|
| 1p36.11 | | * | | ARID1A,MDS2 |
| 1p36.13 | | * | | ARHGEF10L,PAX7,SDHB,SPEN |
| 1q21.3 | | * | * | ARNT,MLLT11,S100A7,SETDB1,TPM3 |
| 1q25.3 | | | * | |
| 2p25.3 | | | * | |
| 2q36.3 | | | * | |
| 3p12.1 | | | * | |
| 4q13.3 | | | * | |
| 4q31.3 | | | * | FBXW7 |
| 4q35.2 | | | * | DUX4L1,FAT1 |
| 5q21.1 | | | * | |
| 5q35.3 | | * | * | FLT4,NSD1 |
| 6p21.32 | | * | * | DAXX |
| 7q34 | * | * | * | BRAF,FAM131B,KIAA1549,TRIM24 |
| 7q35 | | * | * | CNTNAP2 |
| 8p23.1 | | * | * | |
| 9p11.2 | | * | * | |
| 9p21.3 | | * | * | CDKN2A,MLLT3 |
| 10q22.1 | | * | * | PRF1 |
| 11p15.4 | | * | * | CARS,LMO1,NUP98 |
| 11q11 | * | * | * | |
| 15q11.2 | | | * | |
| 16p11.2 | | * | | FUS |
| 17p11.2 | | * | | FLCN,NCOR1,SPECC1 |
| 17q21.31 | | * | | BRCA1,ETV4 |
| 18q11.1 | | * | * | |
| 18q23 | | | * | |
| 19q13.31 | | * | | |




(a) Amplification (b) Deletion

Figure: Venn Diagram among Peaks in LUSC – PRE vs. PRI

# Peaks in LUSC V

Table: Amplification Peaks in LUSC – PRE vs. PRI

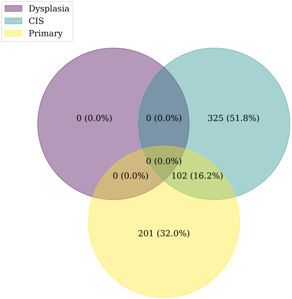
| Peaks | Precancer | Primary | CGC Genes |
|---|---|---|---|
| 1q21.3 | | * | ARNT,MLLT11,S100A7,SETDB1,TPM3 |
| 2p15 | | * | XPO1 |
| 3q27.1 | * | * | |
| 4q12 | | * | CHIC2,FIP1L1,KDR,KIT,PDGFRA |
| 5p15.33 | * | * | SDHA,TERT |
| 7p11.1 | * | * | |
| 7q22.1 | | * | CUX1,TRRAP |
| 7q22.3 | * | | |
| 8p11.21 | | * | ANK1,HOOK3,IKBKB,KAT6A |
| 8p11.23 | * | * | FGFR1,NSD3 |
| 8q24.3 | | * | RECQL4 |
| 9p13.3 | * | * | FANCG |
| 9q11 | * | * | |
| 11p11.11 | * | | |
| 11q13.3 | * | * | CCND1 |
| 13q34 | | * | |
| 15q26.3 | | * | |
| 17p11.1 | * | | |
| 17q21.31 | * | * | BRCA1,ETV4 |
| 18p11.1 | * | * | |
| 19q13.2 | | * | AKT2,CD79A,CIC |
| 19q13.12 | * | | |
| 20p11.1 | * | | |
| 20q11.22 | | * | |
| 22q11.21 | * | * | CLTCL1,DGCR8,LZTR1,SEPT5 |

# Peaks in LUSC VI

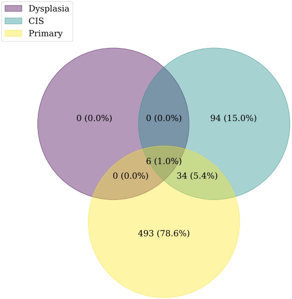
Table: Deletion Peaks in LUSC – PRE vs. PRI

| Peaks | Precancer | Primary | CGC Genes |
|---|---|---|---|
| 1p36.11 | * | | ARID1A,MDS2 |
| 1p36.13 | * | | ARHGEF10L,PAX7,SDHB,SPEN |
| 1q21.3 | * | * | ARNT,MLLT11,S100A7,SETDB1,TPM3 |
| 1q25.3 | | * | |
| 2p25.3 | | * | |
| 2q36.3 | | * | |
| 3p12.1 | | * | |
| 4q13.3 | | * | |
| 4q31.3 | | * | FBXW7 |
| 4q35.2 | | * | DUX4L1,FAT1 |
| 5q21.1 | | * | |
| 5q35.3 | * | * | FLT4,NSD1 |
| 6p21.32 | * | * | DAXX |
| 6p25.3 | * | | IRF4 |
| 7q34 | * | * | BRAF,FAM131B,KIAA1549,TRIM24 |
| 7q35 | * | * | CNTNAP2 |
| 8p23.1 | * | * | |
| 9p11.2 | * | * | |
| 9p21.3 | * | * | CDKN2A,MLLT3 |
| 10q22.1 | * | * | PRF1 |
| 11p15.4 | * | * | CARS,LMO1,NUP98 |
| 11q11 | * | * | |
| 15q11.2 | * | * | |
| 16p11.2 | * | | FUS |
| 17p11.2 | * | | FLCN,NCOR1,SPECC1 |
| 17q12 | * | | CDK12,ERBB2,LASP1,MLLT6,TAF15 |
| 17q21.31 | * | | BRCA1,ETV4 |
| 18q11.1 | * | * | |
| 18q23 | | * | |
| 19q13.31 | * | | |

# Genes in LUSC I




(a) Amplification




(b) Deletion

Figure: Venn Diagram among Genes in LUSC – stage

# Genes in LUSC II

Table: Amplification Genes in LUSC – stage

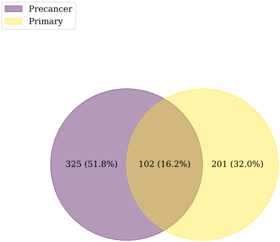
| CGC Genes | Dysplasia | CIS | Primary |
|---|---|---|---|
| AKT2 | | * | |
| ANK1 | | | * |
| BCL11A | | | * |
| CCND1 | | * | |
| CEBPA | | * | |
| CEP89 | | * | |
| DGCR8 | | * | * |
| FANCG | | * | |
| FGFR1 | | * | * |
| HOOK3 | | | * |
| IKBKB | | | * |
| KAT6A | | | * |
| LSM14A | | * | |
| LZTR1 | | * | * |
| PAX5 | | * | |
| PIK3CA | | * | |
| RECQL4 | | | * |
| REL | | | * |
| SOX2 | | * | |
| XPO1 | | | * |

# Genes in LUSC III

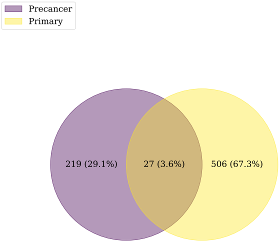
Table: Deletion Genes in LUSC – stage

| CGC Genes | Dysplasia | CIS | Primary |
|---|---|---|---|
| ACKR3 | | | * |
| ACSL3 | | | * |
| ARHGEF10 | | | * |
| ATIC | | | * |
| BARD1 | | | * |
| CASP3 | | | * |
| CDKN2A | | * | * |
| CREB1 | | | * |
| CUL3 | | | * |
| ERBB4 | | | * |
| FAT1 | | | * |
| FEV | | | * |
| IDH1 | | | * |
| MLLT3 | | * | |
| PAX3 | | | * |

# Genes in LUSC IV




(a) Amplification




(b) Deletion

Figure: Venn Diagram among Genes in LUSC – PRE vs. PRI

# Genes in LUSC V

Table: Amplification Genes in LUSC – PRE vs. PRI

| CGC Genes | Precancer | Primary |
| --- | --- | --- |
| AKT2 | * | |
| ANK1 | | * |
| BCL11A | | * |
| CCND1 | * | |
| CEBPA | * | |
| CEP89 | * | |
| DGCR8 | * | * |
| FANCG | * | |
| FGFR1 | * | * |
| HOOK3 | | * |
| IKBKB | | * |
| KAT6A | | * |
| LSM14A | * | |
| LZTR1 | * | * |
| PAX5 | * | |
| PIK3CA | * | |
| RECQL4 | | * |
| REL | | * |
| SOX2 | * | |
| XPO1 | | * |

# Genes in LUSC VI

Table: Deletion Genes in LUSC – PRE vs. PRI

| CGC Genes | Precancer | Primary |
|---|---|---|
| ACKR3 | | * |
| ACSL3 | | * |
| ARHGEF10 | | * |
| ATIC | | * |
| BARD1 | | * |
| CASP3 | | * |
| CD274 | * | |
| CDKN2A | * | * |
| CREB1 | | * |
| CUL3 | | * |
| ERBB4 | | * |
| FAT1 | | * |
| FEV | | * |
| IDH1 | | * |
| JAK2 | * | |
| MLLT3 | * | |
| NFIB | * | |
| PAX3 | | * |
| PDCD1LG2 | * | |
| PSIP1 | * | |
| PTPRD | * | |

# Findings in LUSC Gistic Results

# 4. Results

## 4.5. Copy Number Variation Analysis with Gistic

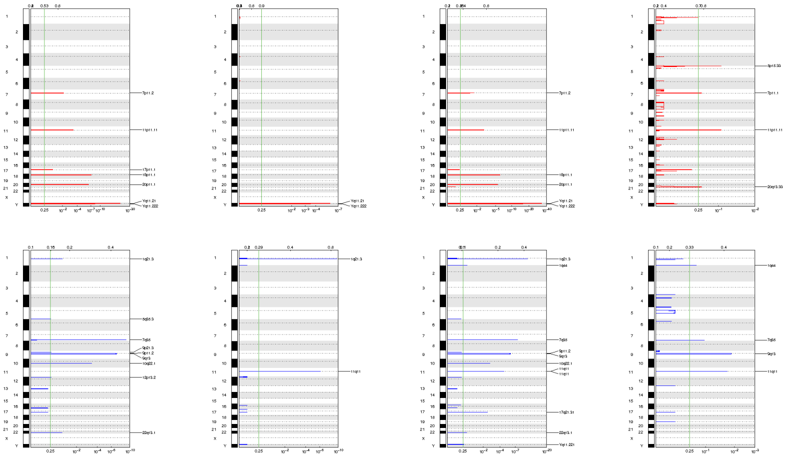
### 4.5.2. Gistic in LUAD

# LUAD 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 |

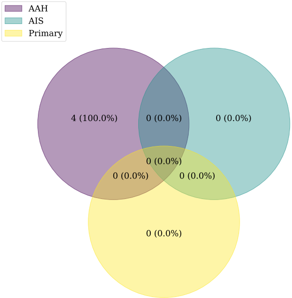
# Gistic in LUAD



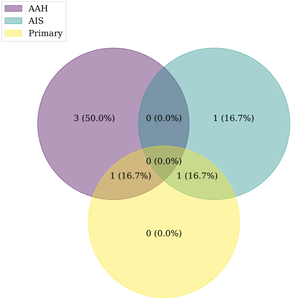
(a) AAH (b) AIS (c) Precancer (d) Primary

Figure: Gistic results in LUAD

# Peaks in LUAD I




(a) Amplification




(b) Deletion

Figure: Venn Diagram among Peaks in LUAD – stage

# Peaks in LUAD II

Table: Amplification Peaks in LUAD – stage

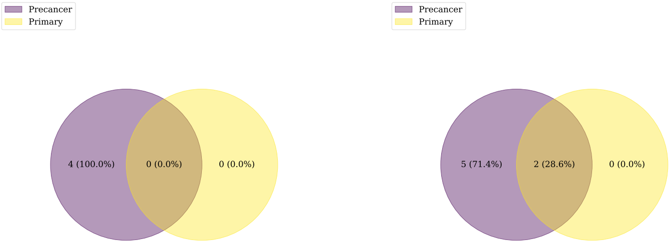
| Peaks | AAH | AIS | Primary | CGC Genes |
|---|---|---|---|---|
| 5p15.33 | | | * | SDHA,TERT |
| 7p11.1 | | | * | |
| 7p11.2 | * | | | EGFR,ZNF479 |
| 11p11.11 | * | | * | |
| 17p11.1 | * | | | |
| 18p11.1 | * | | | |
| 20p11.1 | * | | | |
| 20q13.33 | | | * | PTK6,SS18L1 |

# Peaks in LUAD III

Table: Deletion Peaks in LUAD – stage

| Peaks | AAH | AIS | Primary | CGC Genes |
|---|---|---|---|---|
| 1q21.3 | * | * | | ARNT,MLLT11,S100A7,SETDB1,TPM3 |
| 1q44 | | | * | AKT3 |
| 5q35.3 | * | | | FLT4,NSD1 |
| 7q35 | * | | * | CNTNAP2 |
| 9p11.2 | * | | | |
| 9p21.3 | * | | | CDKN2A,MLLT3 |
| 9q13 | * | | * | |
| 10q22.1 | * | | | PRF1 |
| 11q11 | | * | * | |
| 12p13.2 | * | | | ETV6 |
| 22q13.1 | * | | | APOBEC3B,MRTFA,PDGFB |

# Peaks in LUAD IV




(a) Amplification (b) Deletion

Figure: Venn Diagram among Peaks in LUAD – PRE vs. PRI

# Peaks in LUAD V

Table: Amplification Peaks in LUAD – PRE vs. PRI

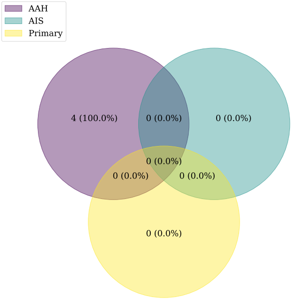
| Peaks | Precancer | Primary | CGC Genes |
|---|---|---|---|
| 5p15.33 | | * | SDHA,TERT |
| 7p11.1 | | * | |
| 7p11.2 | * | | EGFR,ZNF479 |
| 11p11.11 | * | * | |
| 18p11.1 | * | | |
| 20p11.1 | * | | |
| 20q13.33 | | * | PTK6,SS18L1 |

# Peaks in LUAD VI

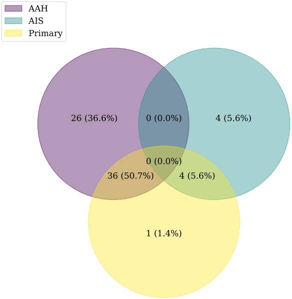
Table: Deletion Peaks in LUAD – PRE vs. PRI

| Peaks | Precancer | Primary | CGC Genes |
|---|---|---|---|
| 1q21.3 | * | | ARNT,MLLT11,S100A7,SETDB1,TPM3 |
| 1q44 | * | * | AKT3 |
| 7q35 | * | * | CNTNAP2 |
| 9p11.2 | * | | |
| 9q13 | * | * | |
| 10q22.1 | * | | PRF1 |
| 11q11 | * | * | |
| 17q21.31 | * | | BRCA1,ETV4 |
| 22q13.1 | * | | APOBEC3B,MRTFA,PDGFB |

# Genes in LUAD I




(a) Amplification




(b) Deletion

Figure: Venn Diagram among Genes in LUAD – stage

# Genes in LUAD II

Table: Amplification Genes in LUAD – stage

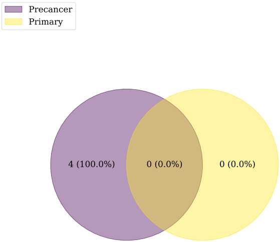
| Genes | AAH | AIS | Primary |
|---|---|---|---|
| FRG1CP | * | | |
| LOC644669 | * | | |
| LOC646813 | * | | |
| ZNF716 | * | | |

# Genes in LUAD III

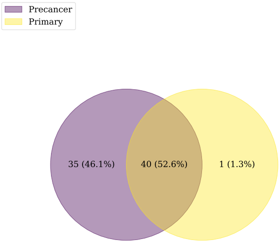
Table: Deletion Genes in LUAD – stage

| Genes | AAH | AIS | Primary |
|---|---|---|---|
| ANKRD20A1 | * | | * |
| ANKRD20A2 | * | | |
| ANKRD20A3 | * | | * |
| ANKRD20A4 | * | | * |
| AQP7P1 | * | | * |
| AQP7P3 | * | | * |
| ARHGEF34P | * | | |
| C10orf35 | * | | |
| CBWD5 | * | | * |
| CBWD6 | * | | |
| CNTNAP3B | * | | |
| COL13A1 | * | | |
| CTAGE4 | * | | |
| CTAGE8 | * | | |
| FAM27C | * | | * |
| FAM27E2 | * | | |
| FAM74A1 | * | | * |
| FAM74A3 | | | * |
| FAM74A4 | * | | * |
| FAM74A7 | * | | |
| FAM95B1 | * | | * |
| FGF7P3 | * | | |
| FGF7P6 | * | | * |
| FLJ43315 | * | | * |
| FOXD4L4 | * | | * |
| FOXD4L5 | * | | * |
| FOXD4L6 | * | | |
| FRG1HP | * | | |
| FRG1JP | * | | * |
| GLIDR | * | | |
| GXYLT1P3 | * | | * |
| LCE3B | | * | |
| LCE3C | | * | |
| LCE3D | | * | |
| LCE3E | | * | |
| LINC01189 | * | | * |
| LINC01410 | * | | * |
| LOC100132249 | * | | * |
| LOC101927827 | * | | |
| LOC101928195 | * | | * |
| LOC101928381 | * | | * |
| LOC101929583 | * | | * |
| LOC102723709 | * | | * |
| LOC102724238 | * | | * |
| LOC102724580 | * | | |
| LOC102725126 | * | | |
| LOC103908605 | * | | * |
| LOC105376064 | * | | |
| LOC286297 | * | | * |
| LOC403323 | * | | * |
| LOC440896 | * | | |
| LOC554249 | * | | * |
| LOC642929 | * | | * |
| LOC728673 | * | | * |
| MIR1299 | * | | |
| MIR4477A | * | | * |
| OR2A1 | * | | |
| OR2A20P | * | | |
| OR2A7 | * | | |
| OR2A9P | * | | |
| OR4C11 | | * | * |
| OR4C6 | | * | * |
| OR4P4 | | * | * |
| OR4S2 | | * | * |
| PGM5P2 | * | | |
| PTGER4P2-CDK2AP2P2 | * | | * |
| SPATA31A5 | * | | * |
| SPATA31A6 | * | | |
| SPATA31A7 | * | | * |
| XLOC_007697 | * | | * |
| ZNF658 | * | | * |

# Genes in LUAD IV




(a) Amplification




(b) Deletion

Figure: Venn Diagram among Genes in LUAD – PRE vs. PRI

Table: Amplification Genes in LUAD – PRE vs. PRI

| Genes | Precancer | Primary |
|---|---|---|
| FRG1CP | * | |
| LOC644669 | * | |
| LOC646813 | * | |
| ZNF733P | * | |

# Genes in LUAD VI

Table: Deletion Genes in LUAD – PRE vs. PRI

| Genes | Precancer | Primary |
|---|---|---|
| ANKRD20A1 | * | * |
| ANKRD20A2 | * | |
| ANKRD20A3 | * | * |
| ANKRD20A4 | * | * |
| AQP7P1 | * | * |
| AQP7P3 | * | * |
| ARHGEF34P | * | |
| ARL17A | * | |
| ARL17B | * | |
| C10orf35 | * | |
| CBWD5 | * | * |
| CBWD6 | * | |
| CNTNAP3B | * | |
| COL13A1 | * | |
| CTAGE4 | * | |
| CTAGE8 | * | |
| FAM27C | * | * |
| FAM27E2 | * | |
| FAM74A1 | * | * |
| FAM74A3 | | * |
| FAM74A4 | * | * |
| FAM74A7 | * | |
| FAM95B1 | * | * |
| FGF7P3 | * | |
| FGF7P6 | * | * |
| FLJ43315 | * | * |
| FOXD4L4 | * | * |
| FOXD4L5 | * | * |
| FOXD4L6 | * | |
| FRG1HP | * | |
| FRG1JP | * | * |
| GLIDR | * | |
| GXYLT1P3 | * | * |
| LCE3B | * | |
| LCE3C | * | |
| LCE3D | * | |
| LCE3E | * | |
| LINC01189 | * | * |
| LINC01410 | * | * |
| LOC100132249 | * | * |
| LOC101927827 | * | |
| LOC101928195 | * | * |
| LOC101928381 | * | * |
| LOC101929583 | * | * |
| LOC102723709 | * | * |
| LOC102724238 | * | * |
| LOC102724580 | * | |
| LOC102725126 | * | |
| LOC103908605 | * | * |
| LOC105376064 | * | |
| LOC286297 | * | * |
| LOC403323 | * | * |
| LOC440896 | * | |
| LOC554249 | * | * |
| LOC642929 | * | * |
| LOC728673 | * | * |
| LRRC37A | * | |
| LRRC37A2 | * | |
| MIR1299 | * | |
| MIR4477A | * | * |
| NSFP1 | * | |
| OR2A1 | * | |
| OR2A20P | * | |
| OR2A7 | * | |
| OR2A9P | * | |
| OR4C11 | * | * |
| OR4C6 | * | * |
| OR4P4 | * | * |
| OR4S2 | * | * |
| PGM5P2 | * | |
| PTGER4P2-CDK2AP2P2 | * | * |
| SPATA31A5 | * | * |
| SPATA31A6 | * | |
| SPATA31A7 | * | * |
| XLOC_007697 | * | * |
| ZNF658 | * | * |

# Findings in LUAD Gistic Results

# 4. Results

## 4.5. Copy Number Variation Analysis with Gistic

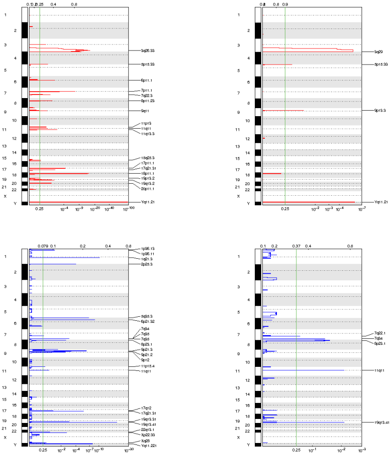
### 4.5.3. Gistic in Recurrence & LUSC

# LUSC Data Composition with Recurrence

Table: LUSC WES Data with Recurrence

| Recurrence? | Stage | Number of Samples |
|---|---|---|
| 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 |

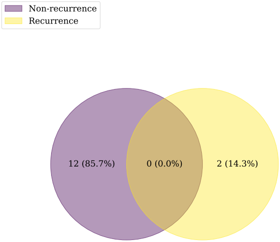
# Gistic in Recurrence & LUSC – CIS



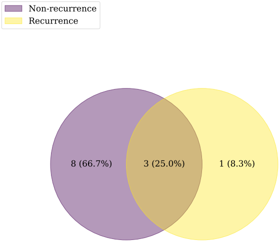
(a) Non-recurrence (b) Recurrence

Figure: Gistic results in Recurrence & LUSC – CIS

# Peaks in Recurrence & LUSC – CIS I




(a) Amplification




(b) Deletion

Figure: Venn Diagram among Peaks in Recurrence & LUSC – CIS

# Peaks in Recurrence & LUSC – CIS II

Table: Amplification Peaks in Recurrence & LUSC – CIS

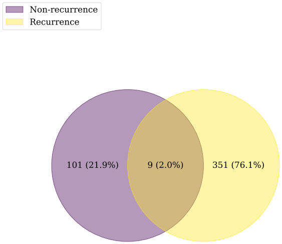
| Peaks | Non-recurrence | Recurrence | CGC Genes |
|---|---|---|---|
| 3q26.33 | * | | SOX2 |
| 3q29 | | * | MB21D2,MUC4,TFRC |
| 5p15.33 | * | * | SDHA,TERT |
| 6p11.1 | * | | |
| 7p11.1 | * | | |
| 7q22.3 | * | | |
| 8p11.23 | * | | FGFR1,NSD3 |
| 9p13.3 | | * | FANCG |
| 9q11 | * | | |
| 11p13 | * | | LMO2,WT1 |
| 11q11 | * | | |
| 11q13.3 | * | | CCND1 |
| 15q25.3 | * | | NTRK3 |
| 17p11.1 | * | | |
| 17q21.31 | * | | BRCA1,ETV4 |
| 18p11.1 | * | | |
| 19p13.2 | * | | CD209,DNM2,KEAP1,MUC16,SMARCA4 |
| 19q13.2 | * | | AKT2,CD79A,CIC |
| 20p11.1 | * | | |

# Peaks in Recurrence & LUSC – CIS III

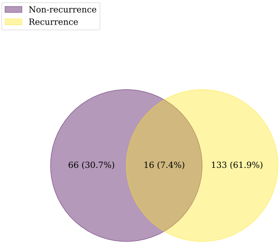
Table: Deletion Peaks in Recurrence & LUSC – CIS

| Peaks | Non-recurrence | Recurrence | CGC Genes |
|---|---|---|---|
| 1p36.11 | * | | ARID1A,MDS2 |
| 1p36.13 | * | | ARHGEF10L,PAX7,SDHB,SPEN |
| 1q21.3 | * | | ARNT,MLLT11,S100A7,SETDB1,TPM3 |
| 2p25.3 | * | | |
| 5q35.3 | * | | FLT4,NSD1 |
| 6p21.32 | * | | DAXX |
| 7q22.1 | | * | CUX1,TRRAP |
| 7q34 | * | * | BRAF,FAM131B,KIAA1549,TRIM24 |
| 7q35 | * | | CNTNAP2 |
| 8p23.1 | * | * | |
| 9p12 | * | | |
| 9p21.2 | * | | |
| 9p21.3 | * | | CDKN2A,MLLT3 |
| 11p15.4 | * | | CARS,LMO1,NUP98 |
| 11q11 | * | * | |
| 17q12 | * | | CDK12,ERBB2,LASP1,MLLT6,TAF15 |
| 17q21.31 | * | | BRCA1,ETV4 |
| 19q13.31 | * | | |
| 19q13.41 | * | * | PPP2R1A |
| 22q13.1 | * | | APOBEC3B,MRTFA,PDGFB |

# Genes in Recurrence & LUSC – CIS I




(a) Amplification




(b) Deletion

Figure: Venn Diagram among Genes in Recurrence & LUSC – CIS

# Genes in Recurrence & LUSC – CIS II

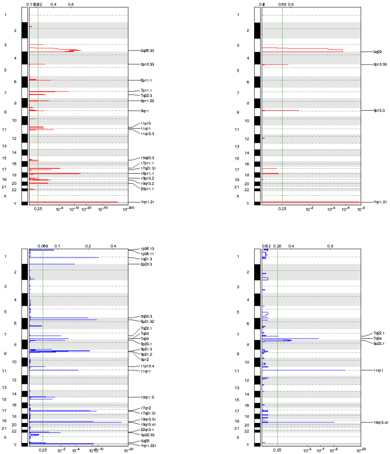
Table: Amplification Genes in Recurrence & LUSC – CIS

| CGC Genes | Non-recurrence | Recurrence |
|---|---|---|
| BCL6 | | * |
| CCND1 | * | |
| EIF4A2 | | * |
| ETV5 | | * |
| FANCG | | * |
| FGFR1 | * | |
| IGF2BP2 | | * |
| LPP | | * |
| MAP3K13 | | * |
| MB21D2 | | * |
| MECOM | | * |
| MUC4 | | * |
| PIK3CA | | * |
| SOX2 | * | * |
| TBL1XR1 | | * |
| TFRC | | * |
| TP63 | | * |

# Genes in Recurrence & LUSC – CIS III

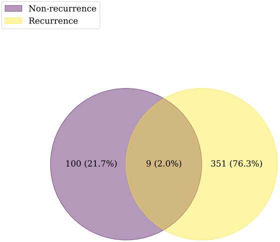
Table: Deletion Genes in Recurrence & LUSC – CIS

| CGC Genes | Non-recurrence | Recurrence |
|---|---|---|
| ARHGEF10 | | * |
| CDKN2A | * | |



(a) Non-recurrence (b) Recurrence

Figure: Gistic results in Recurrence & LUSC – Precancer




(a) Amplification (b) Deletion

Figure: Venn Diagram among Peaks in Recurrence & LUSC – Precancer

# Peaks in Recurrence & LUSC – Precancer II

Table: Amplification Peaks in Recurrence & LUSC – Precancer

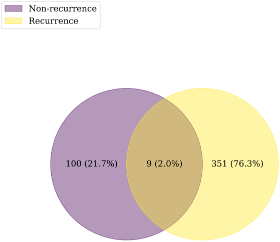
| Peaks | Non-recurrence | Recurrence | CGC Genes |
|---|---|---|---|
| 3q26.33 | * | | SOX2 |
| 3q29 | | * | MB21D2,MUC4,TFRC |
| 5p15.33 | * | * | SDHA,TERT |
| 6p11.1 | * | | |
| 7p11.1 | * | | |
| 7q22.3 | * | | |
| 8p11.23 | * | | FGFR1,NSD3 |
| 9p13.3 | | * | FANCG |
| 9q11 | * | | |
| 11p13 | * | | LMO2,WT1 |
| 11q11 | * | | |
| 11q13.3 | * | | CCND1 |
| 15q25.3 | * | | NTRK3 |
| 17p11.1 | * | | |
| 17q21.31 | * | | BRCA1,ETV4 |
| 18p11.1 | * | | |
| 19p13.2 | * | | CD209,DNM2,KEAP1,MUC16,SMARCA4 |
| 19q13.2 | * | | AKT2,CD79A,CIC |
| 20p11.1 | * | | |

# Peaks in Recurrence & LUSC – Precancer III

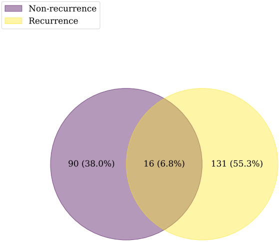
Table: Deletion Peaks in Recurrence & LUSC – Precancer

| Peaks | Non-recurrence | Recurrence | CGC Genes |
|---|---|---|---|
| 1p36.11 | * | | ARID1A,MDS2 |
| 1p36.13 | * | | ARHGEF10L,PAX7,SDHB,SPEN |
| 1q21.3 | * | | ARNT,MLLT11,S100A7,SETDB1,TPM3 |
| 2p25.3 | * | | |
| 5q35.3 | * | | FLT4,NSD1 |
| 6p21.32 | * | | DAXX |
| 7q22.1 | * | * | CUX1,TRRAP |
| 7q34 | | * | BRAF,FAM131B,KIAA1549,TRIM24 |
| 7q35 | * | | CNTNAP2 |
| 8p23.1 | * | * | |
| 9p12 | * | | |
| 9p21.2 | * | | |
| 9p21.3 | * | | CDKN2A,MLLT3 |
| 11p15.4 | * | | CARS,LMO1,NUP98 |
| 11q11 | * | * | |
| 15q11.2 | * | | |
| 17q12 | * | | CDK12,ERBB2,LASP1,MLLT6,TAF15 |
| 17q21.31 | * | | BRCA1,ETV4 |
| 19q13.31 | * | | |
| 19q13.41 | * | * | PPP2R1A |
| 22q13.1 | * | | APOBEC3B,MRTFA,PDGFB |

# Genes in Recurrence & LUSC – Precancer I




(a) Amplification




(b) Deletion

Figure: Venn Diagram among Genes in Recurrence & LUSC – Precancer

# Genes in Recurrence & LUSC – Precancer II

Table: Amplification Genes in Recurrence & LUSC – Precancer

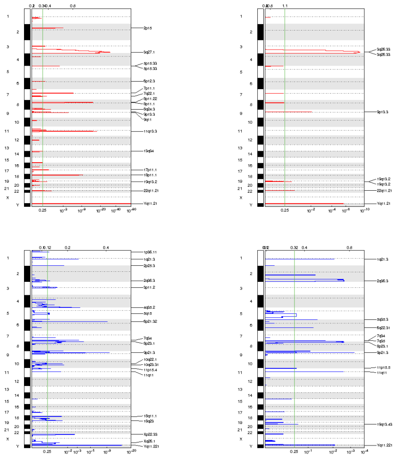
| CGC Genes | Non-recurrence | Recurrence |
|---|---|---|
| BCL6 | | * |
| CCND1 | * | |
| EIF4A2 | | * |
| ETV5 | | * |
| FANCG | | * |
| FGFR1 | * | |
| IGF2BP2 | | * |
| LPP | | * |
| MAP3K13 | | * |
| MB21D2 | | * |
| MECOM | | * |
| MUC4 | | * |
| PIK3CA | | * |
| SOX2 | * | * |
| TBL1XR1 | | * |
| TFRC | | * |
| TP63 | | * |

# Genes in Recurrence & LUSC – Precancer III

Table: Deletion Genes in Recurrence & LUSC – Precancer

| CGC Genes | Non-recurrence | Recurrence |
|---|---|---|
| ARHGEF10 | | * |
| CDKN2A | * | |

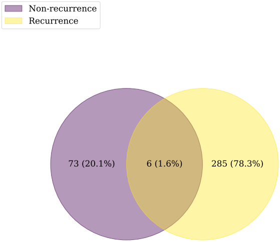
# Gistic in Recurrence & LUSC – Primary



(a) Non-recurrence (b) Recurrence

Figure: Gistic results in Recurrence & LUSC – Primary

# Peaks in Recurrence & LUSC – Primary I




(a) Amplification (b) Deletion

Figure: Venn Diagram among Peaks in Recurrence & LUSC – Primary

# Peaks in Recurrence & LUSC – Primary II

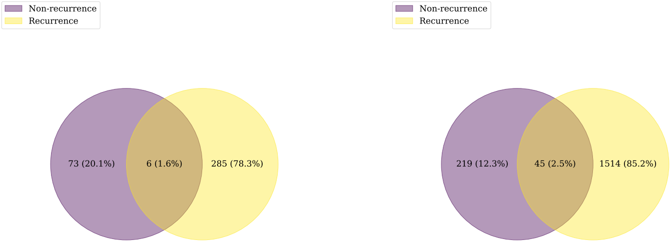
Table: Amplification Peaks in Recurrence & LUSC – Primary

| Peaks | Non-recurrence | Recurrence | CGC Genes |
|---|---|---|---|
| 2p15 | * | | XPO1 |
| 3q26.33 | | * | SOX2 |
| 3q27.1 | * | | |
| 5p15.33 | * | | SDHA,TERT |
| 6p12.3 | * | | |
| 7p11.1 | * | | |
| 7q22.1 | * | | CUX1,TRRAP |
| 8p11.1 | * | | |
| 8p11.22 | * | | |
| 8q24.3 | * | | RECQL4 |
| 9p13.3 | * | * | FANCG |
| 9q11 | * | | |
| 11q13.3 | * | | CCND1 |
| 13q34 | * | | |
| 17p11.1 | * | | |
| 18p11.1 | * | | |
| 19q13.2 | * | * | AKT2,CD79A,CIC |
| 22q11.21 | * | * | CLTCL1,DGCR8,LZTR1,SEPT5 |

# Peaks in Recurrence & LUSC – Primary III

Table: Deletion Peaks in Recurrence & LUSC – Primary

| Peaks | Non-recurrence | Recurrence | CGC Genes |
|---|---|---|---|
| 1p36.11 | * | | ARID1A,MDS2 |
| 1q21.3 | * | * | ARNT,MLLT11,S100A7,SETDB1,TPM3 |
| 2p25.3 | * | | |
| 2q36.3 | * | * | |
| 3p11.2 | * | | |
| 4q35.2 | * | | DUX4L1,FAT1 |
| 5q15 | * | | |
| 5q35.3 | | * | FLT4,NSD1 |
| 6p21.32 | * | | DAXX |
| 6q22.31 | | * | |
| 7q34 | * | * | BRAF,FAM131B,KIAA1549,TRIM24 |
| 7q35 | | * | CNTNAP2 |
| 8p23.1 | * | * | |
| 9p21.3 | * | * | CDKN2A,MLLT3 |
| 10q22.1 | * | | PRF1 |
| 10q23.31 | * | | FAS,PTEN |
| 11p15.4 | * | | CARS,LMO1,NUP98 |
| 11p15.5 | | * | HRAS |
| 11q11 | * | * | |
| 18q11.1 | * | | |
| 18q23 | * | | |
| 19q13.43 | | * | |




(a) Amplification (b) Deletion

Figure: Venn Diagram among Genes in Recurrence & LUSC – Primary

# Genes in Recurrence & LUSC – Primary II

Table: Amplification Genes in Recurrence & LUSC – Primary

| CGC Genes | Non-recurrence | Recurrence |
|---|---|---|
| BCL11A | * | |
| BCL6 | | * |
| EIF4A2 | | * |
| ETV5 | | * |
| FGFR1 | * | |
| IGF2BP2 | | * |
| LPP | | * |
| MAP3K13 | | * |
| MB21D2 | | * |
| MECOM | | * |
| MUC4 | | * |
| PIK3CA | | * |
| REL | * | |
| SOX2 | | * |
| TBL1XR1 | | * |
| TFRC | | * |
| TP63 | | * |
| XPO1 | * | |

# Genes in Recurrence & LUSC – Primary III

Table: Deletion Genes in Recurrence & LUSC – Primary

| CGC Genes | Non-recurrence | Recurrence |
|---|---|---|
| ACKR3 | | * |
| ACSL3 | | * |
| ACSL6 | | * |
| AFF4 | | * |
| APC | | * |
| ARHGAP26 | | * |
| ARHGEF10 | | * |
| ATIC | | * |
| BARD1 | | * |
| BCL2 | * | |
| CASP8 | | * |
| CD28 | | * |
| CD74 | | * |
| CDKN2A | * | * |
| COL3A1 | | * |
| CREB1 | | * |
| CSF1R | | * |
| CUL3 | | * |
| DCC | * | |
| EBF1 | | * |
| ERBB4 | | * |
| FAT1 | * | |
| FEV | | * |
| FGFR4 | | * |
| FLT4 | | * |
| HOXD11 | | * |
| HOXD13 | | * |
| IDH1 | | * |
| ITGAV | | * |
| ITK | | * |
| KDSR | * | |
| MALT1 | * | |
| NFE2L2 | | * |
| NPM1 | | * |
| NSD1 | | * |
| PAX3 | | * |
| PDGFRB | | * |
| PMS1 | | * |
| PWWP2A | | * |
| SETBP1 | * | |
| SF3B1 | | * |
| SMAD2 | * | |
| SMAD4 | * | |
| TLX3 | | * |

# Findings in Recurrence & LUSC Gistic Results

# 4. Results

## 4.5. Copy Number Variation Analysis with Gistic

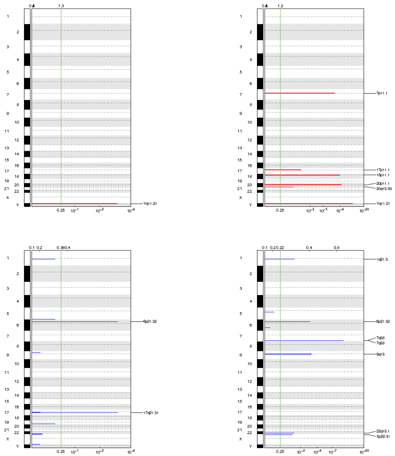
### 4.5.4. Gistic in Recurrence & LUAD

# LUAD Data Composition with Recurrence

Table: LUAD WES Data with Recurrence

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

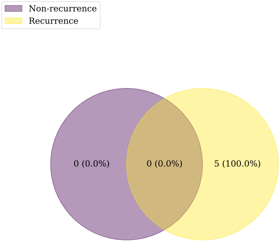
# Gistic in Recurrence & LUAD – AAH



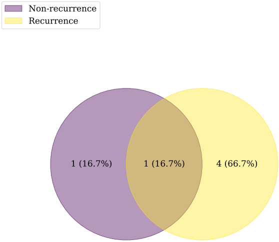
(a) Non-recurrence (b) Recurrence

Figure: Gistic results in Recurrence & LUAD – AAH

# Peaks in Recurrence & LUAD – AAH I




(a) Amplification




(b) Deletion

Figure: Venn Diagram among Peaks in Recurrence & LUAD – AAH

# Peaks in Recurrence & LUAD – AAH II

Table: Amplification Peaks in Recurrence & LUAD – AAH

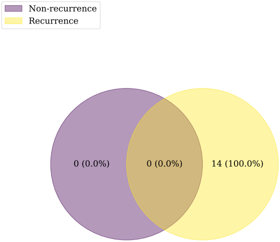
| Peaks | Non-recurrence | Recurrence | CGC Genes |
|---|---|---|---|
| 7p11.1 | | * | |
| 17p11.1 | | * | |
| 18p11.1 | | * | |
| 20p11.1 | | * | |
| 20q13.33 | | * | PTK6,SS18L1 |

# Peaks in Recurrence & LUAD – AAH III

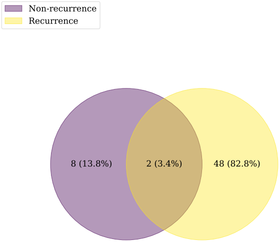
Table: Deletion Peaks in Recurrence & LUAD – AAH

| Peaks | Non-recurrence | Recurrence | CGC Genes |
|---|---|---|---|
| 1q21.3 | | * | ARNT,MLLT11,S100A7,SETDB1,TPM3 |
| 6p21.32 | * | * | DAXX |
| 7q35 | | * | CNTNAP2 |
| 9q13 | | * | |
| 17q21.31 | * | | BRCA1,ETV4 |
| 22q13.1 | | * | APOBEC3B,MRTFA,PDGFB |

# Genes in Recurrence & LUAD – AAH I




(a) Amplification




(b) Deletion

Figure: Venn Diagram among Genes in Recurrence & LUAD – AAH

# Genes in Recurrence & LUAD – AAH II

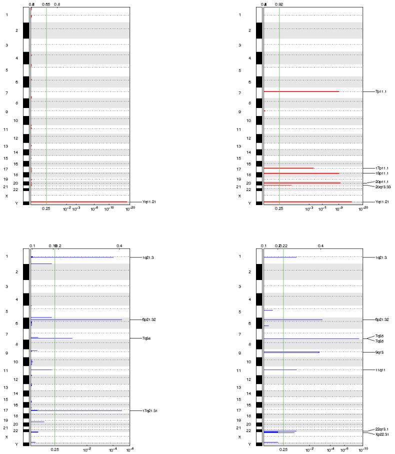
Table: Amplification Genes in Recurrence & LUAD – AAH

| Genes | Non-recurrence | Recurrence |
|---|---|---|
| C20orf197 | | * |
| CDH26 | | * |
| CDH4 | | * |
| FRG1CP | | * |
| LOC100506470 | | * |
| LOC101928048 | | * |
| LOC644669 | | * |
| LOC729296 | | * |
| MIR4533 | | * |
| MIR548AG2 | | * |
| MIR646 | | * |
| MIR646HG | | * |
| MTRNR2L1 | | * |
| ZNF716 | | * |

# Genes in Recurrence & LUAD – AAH III

Table: Deletion Genes in Recurrence & LUAD – AAH

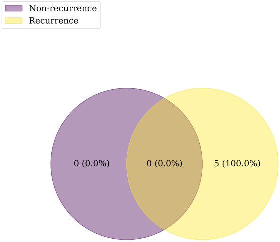
| Genes | Non-recurrence | Recurrence |
|---|---|---|
| ANKRD20A1 | | * |
| ANKRD20A3 | | * |
| ANKRD20A4 | | * |
| APOBEC3A | | * |
| APOBEC3B-AS1 | | * |
| AQP7P1 | | * |
| AQP7P3 | | * |
| ARHGEF34P | | * |
| ARL17A | * | |
| ARL17B | * | |
| BTNL2 | * | |
| C6orf10 | * | * |
| CBWD5 | | * |
| CTAGE4 | | * |
| CTAGE8 | | * |
| FAM27C | | * |
| FAM74A1 | | * |
| FAM74A3 | | * |
| FAM74A4 | | * |
| FAM95B1 | | * |
| FGF7P6 | | * |
| FLJ43315 | | * |
| FOXD4L4 | | * |
| FOXD4L5 | | * |
| FRG1JP | | * |
| GXYLT1P3 | | * |
| HCG23 | * | |
| LCE3A | | * |
| LCE3B | | * |
| LINC01189 | | * |
| LINC01410 | | * |
| LOC100132249 | | * |
| LOC101928195 | | * |
| LOC101928381 | | * |
| LOC101929163 | * | * |
| LOC101929583 | | * |
| LOC102723709 | | * |
| LOC102724238 | | * |
| LOC103908605 | | * |
| LOC286297 | | * |
| LOC403323 | | * |
| LOC554249 | | * |
| LOC642929 | | * |
| LOC728673 | | * |
| LRRC37A | * | |
| LRRC37A2 | * | |
| MIR4477A | | * |
| NSF | * | |
| NSFP1 | * | |
| OR2A1 | | * |
| OR2A20P | | * |
| OR2A7 | | * |
| OR2A9P | | * |
| PTGER4P2-CDK2AP2P2 | | * |
| SPATA31A5 | | * |
| SPATA31A7 | | * |
| XLOC_007697 | | * |
| ZNF658 | | * |



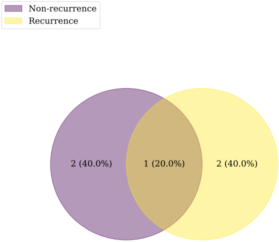
(a) Non-recurrence (b) Recurrence

Figure: Gistic results in Recurrence & LUAD – Precancer

# Peaks in Recurrence & LUAD – Precancer I




(a) Amplification




(b) Deletion

Figure: Venn Diagram among Peaks in Recurrence & LUAD – Precancer

# Peaks in Recurrence & LUAD – Precancer II

Table: Amplification Peaks in Recurrence & LUAD – Precancer

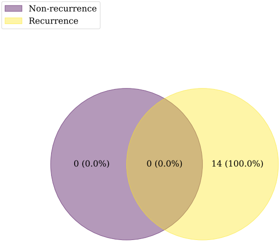
| Peaks | Non-recurrence | Recurrence | CGC Genes |
|---|---|---|---|
| 7p11.1 | | * | |
| 17p11.1 | | * | |
| 18p11.1 | | * | |
| 20p11.1 | | * | |
| 20q13.33 | | * | PTK6,SS18L1 |

# Peaks in Recurrence & LUAD – Precancer III

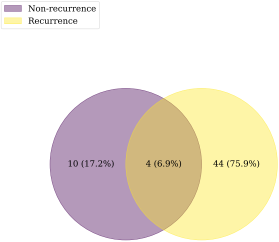
Table: Deletion Peaks in Recurrence & LUAD – Precancer

| Peaks | Non-recurrence | Recurrence | CGC Genes |
|---|---|---|---|
| 1q21.3 | * | * | ARNT,MLLT11,S100A7,SETDB1,TPM3 |
| 6p21.32 | * | * | DAXX |
| 7q34 | * | | BRAF,FAM131B,KIAA1549,TRIM24 |
| 7q35 | | * | CNTNAP2 |
| 9q13 | | * | |
| 11q11 | | * | |
| 17q21.31 | * | | BRCA1,ETV4 |
| 22q13.1 | | * | APOBEC3B,MRTFA,PDGFB |

# Genes in Recurrence & LUAD – Precancer I




(a) Amplification




(b) Deletion

Figure: Venn Diagram among Genes in Recurrence & LUAD – Precancer

# Genes in Recurrence & LUAD – Precancer II

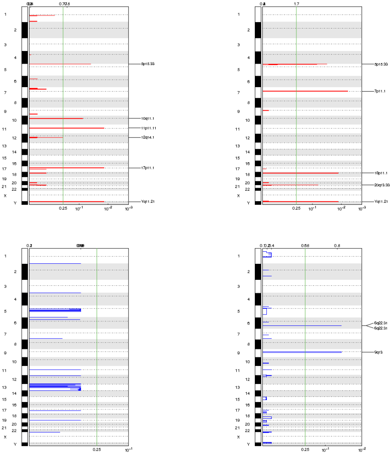
Table: Amplification Genes in Recurrence & LUAD – Precancer

| Genes | Non-recurrence | Recurrence |
|---|---|---|
| C20orf197 | | * |
| CDH26 | | * |
| CDH4 | | * |
| FRG1CP | | * |
| LOC100506470 | | * |
| LOC101928048 | | * |
| LOC644669 | | * |
| LOC729296 | | * |
| MIR4533 | | * |
| MIR548AG2 | | * |
| MIR646 | | * |
| MIR646HG | | * |
| MTRNR2L1 | | * |
| ZNF716 | | * |

# Genes in Recurrence & LUAD – Precancer III

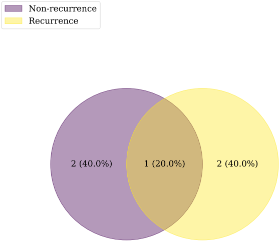
Table: Deletion Genes in Recurrence & LUAD – Precancer

| Genes | Non-recurrence | Recurrence |
|---|---|---|
| ANKRD20A1 | | * |
| ANKRD20A3 | | * |
| ANKRD20A4 | | * |
| AQP7P1 | | * |
| AQP7P3 | | * |
| ARHGEF34P | | * |
| ARL17A | * | |
| ARL17B | * | |
| BTNL2 | * | * |
| C6orf10 | * | * |
| CBWD5 | | * |
| CTAGE4 | | * |
| CTAGE8 | | * |
| FAM27C | | * |
| FAM74A1 | | * |
| FAM74A3 | | * |
| FAM74A4 | | * |
| FAM95B1 | | * |
| FGF7P6 | | * |
| FLJ43315 | | * |
| FOXD4L4 | | * |
| FOXD4L5 | | * |
| FRG1JP | | * |
| GXYLT1P3 | | * |
| HCG23 | * | * |
| LCE3A | * | |
| LCE3B | * | |
| LCE3C | * | |
| LCE3D | * | |
| LCE3E | * | |
| LINC01189 | | * |
| LINC01410 | | * |
| LOC100132249 | | * |
| LOC101928195 | | * |
| LOC101928381 | | * |
| LOC101929163 | * | * |
| LOC101929583 | | * |
| LOC102723709 | | * |
| LOC102724238 | | * |
| LOC103908605 | | * |
| LOC286297 | | * |
| LOC403323 | | * |
| LOC554249 | | * |
| LOC642929 | | * |
| LOC728673 | | * |
| LRRC37A | * | |
| LRRC37A2 | * | |
| MIR4477A | | * |
| NSFP1 | * | |
| OR2A1 | | * |
| OR2A20P | | * |
| OR2A7 | | * |
| OR2A9P | | * |
| PTGER4P2-CDK2AP2P2 | | * |
| SPATA31A5 | | * |
| SPATA31A7 | | * |
| XLOC_007697 | | * |
| ZNF658 | | * |

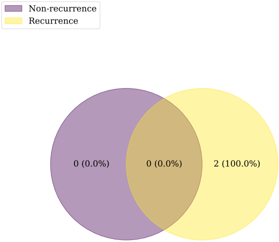



(a) Non-recurrence (b) Recurrence

Figure: Gistic results in Recurrence & LUAD – Primary




(a) Amplification




(b) Deletion

Figure: Venn Diagram among Peaks in Recurrence & LUAD – Primary

# Peaks in Recurrence & LUAD – Primary II

Table: Amplification Peaks in Recurrence & LUAD – Primary

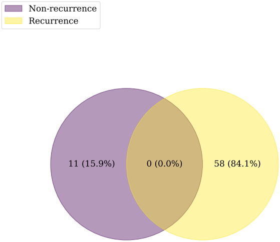
| Peaks | Non-recurrence | Recurrence | CGC Genes |
|---|---|---|---|
| 5p15.33 | * | * | SDHA,TERT |
| 7p11.1 | | * | |
| 10q11.1 | * | | |
| 11p11.11 | * | | |
| 12q14.1 | * | | CDK4,LRIG3 |
| 17p11.1 | * | | |
| 18p11.1 | | * | |
| 20q13.33 | | * | PTK6,SS18L1 |

# Peaks in Recurrence & LUAD – Primary III

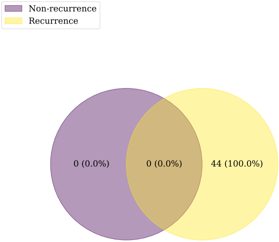
Table: Deletion Peaks in Recurrence & LUAD – Primary

| Peaks | Non-recurrence | Recurrence | CGC Genes |
|---|---|---|---|
| 6q22.31 | | * | |
| 9q13 | | * | |

# Genes in Recurrence & LUAD – Primary I




(a) Amplification




(b) Deletion

Figure: Venn Diagram among Genes in Recurrence & LUAD – Primary

# Genes in Recurrence & LUAD – Primary II

Table: Amplification Genes in Recurrence & LUAD – Primary

| CGC Genes | Non-recurrence | Recurrence |
|---|---|---|
| CTNND2 | | * |
| TERT | * | |

Table: Deletion Genes in Recurrence & LUAD – Primary

| Genes | Non-recurrence | Recurrence |
|---|---|---|
| ANKRD20A1 | | * |
| ANKRD20A3 | | * |
| ANKRD20A4 | | * |
| AQP7P1 | | * |
| AQP7P3 | | * |
| CBWD5 | | * |
| CNTNAP3P2 | | * |
| FAM27B | | * |
| FAM27C | | * |
| FAM27E3 | | * |
| FAM74A1 | | * |
| FAM74A3 | | * |
| FAM74A4 | | * |
| FAM95B1 | | * |
| FGF7P6 | | * |
| FLJ43315 | | * |
| FOXD4L4 | | * |
| FOXD4L5 | | * |
| FRG1JP | | * |
| GXYLT1P3 | | * |
| HRAT13 | | * |
| LINC01189 | | * |
| LINC01410 | | * |
| LOC100132249 | | * |
| LOC101928195 | | * |
| LOC101928381 | | * |
| LOC101929583 | | * |
| LOC102723709 | | * |
| LOC102724238 | | * |
| LOC103908605 | | * |
| LOC105379450 | | * |
| LOC286297 | | * |
| LOC403323 | | * |
| LOC554249 | | * |
| LOC642929 | | * |
| LOC728673 | | * |
| MIR4477A | | * |
| PTGER4P2-CDK2AP2P2 | | * |
| SPATA31A3 | | * |
| SPATA31A5 | | * |
| SPATA31A7 | | * |
| TRDN | | * |
| XLOC_007697 | | * |
| ZNF658 | | * |

# Findings in Recurrence & LUAD Gistic Results

# 4. Results

## 4.5. Copy Number Variation Analysis with Gistic

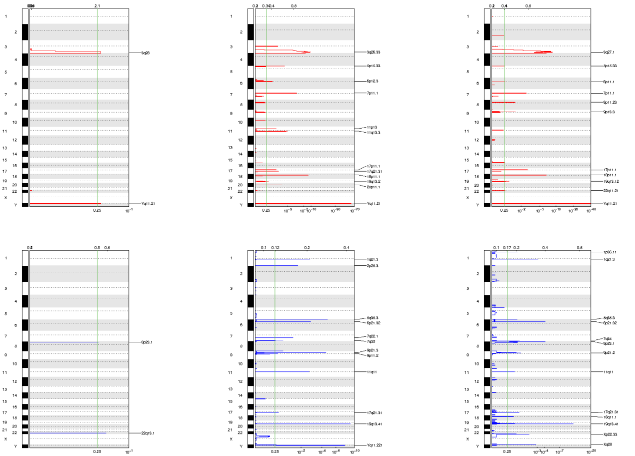
### 4.5.5. Gistic in Smoking & LUSC

# LUSC Data Composition with Smoking

Table: LUSC WES Data with Smoking

| Smoking? | Stage | Number of Samples |
|---|---|---|
| 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 |

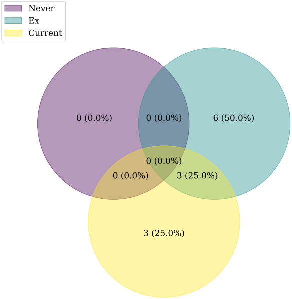
# Gistic in Smoking & LUSC – CIS



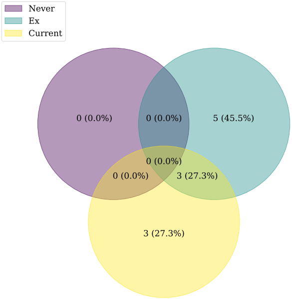
(a) Never Smoker (b) Ex-Smoker (c) Current Smoker

Figure: Gistic results in Smoking & LUSC – CIS

# Peaks in Smoking & LUSC – CIS I




(a) Amplification




(b) Deletion

Figure: Venn Diagram among Peaks in Smoking & LUSC – CIS

# Peaks in Smoking & LUSC – CIS II

Table: Amplification Peaks in Smoking & LUSC – CIS

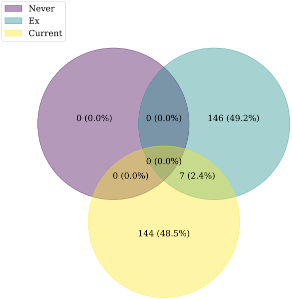
| Peaks | Never | Ex | Current | CGC Genes |
|---|---|---|---|---|
| 3q26.33 | | * | | SOX2 |
| 3q27.1 | | | * | |
| 3q28 | * | | | LPP,TP63 |
| 5p15.33 | | * | * | SDHA,TERT |
| 6p11.1 | | | * | |
| 6p12.3 | | * | | |
| 7p11.1 | | * | * | |
| 8p11.23 | | | * | FGFR1,NSD3 |
| 9p13.3 | | | * | FANCG |
| 11p13 | | * | | LMO2,WT1 |
| 11q13.3 | | * | | CCND1 |
| 17p11.1 | | * | * | |
| 17q21.31 | | * | | BRCA1,ETV4 |
| 18p11.1 | | * | * | |
| 19q13.2 | | * | | AKT2,CD79A,CIC |
| 19q13.12 | | | * | |
| 20p11.1 | | * | | |
| 22q11.21 | | | * | CLTCL1,DGCR8,LZTR1,SEPT5 |

# Peaks in Smoking & LUSC – CIS III

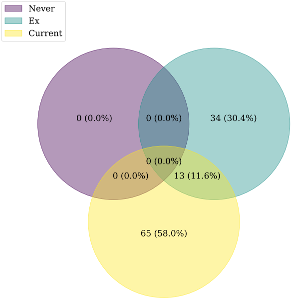
Table: Deletion Peaks in Smoking & LUSC – CIS

| Peaks | Never | Ex | Current | CGC Genes |
|---|---|---|---|---|
| 1p36.11 | | | * | ARID1A,MDS2 |
| 1q21.3 | | * | * | ARNT,MLLT11,S100A7,SETDB1,TPM3 |
| 2p25.3 | | * | | |
| 5q35.3 | | * | * | FLT4,NSD1 |
| 6p21.32 | | * | * | DAXX |
| 7q22.1 | | * | | CUX1,TRRAP |
| 7q34 | | | * | BRAF,FAM131B,KIAA1549,TRIM24 |
| 7q35 | | * | | CNTNAP2 |
| 8p23.1 | * | | * | |
| 9p11.2 | | * | | |
| 9p21.2 | | | * | |
| 9p21.3 | | * | | CDKN2A,MLLT3 |
| 11q11 | | * | * | |
| 17q21.31 | | * | * | BRCA1,ETV4 |
| 18q11.1 | | | * | |
| 19q13.41 | | * | * | PPP2R1A |
| 22q13.1 | * | | | APOBEC3B,MRTFA,PDGFB |

# Genes in Smoking & LUSC – CIS I




(a) Amplification




(b) Deletion

Figure: Venn Diagram among Genes in Smoking & LUSC – CIS

# Genes in Smoking & LUSC – CIS II

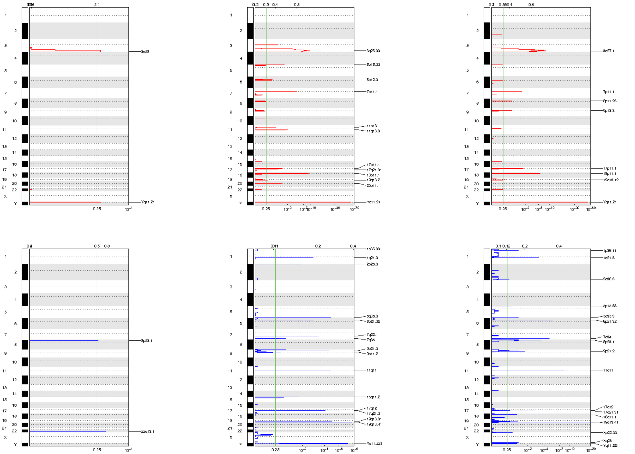
Table: Amplification Genes in Smoking & LUSC – CIS

| CGC Genes | Never | Ex | Current |
|---|---|---|---|
| ANK1 | | | * |
| CCND1 | | * | |
| EXT2 | | * | |
| FANCG | | | * |
| FGFR1 | | | * |
| HOOK3 | | | * |
| IKBKB | | | * |
| KAT6A | | | * |
| LMO2 | | * | |
| PAX5 | | | * |
| SOX2 | | * | |
| WT1 | | * | |

# Genes in Smoking & LUSC – CIS III

Table: Deletion Genes in Smoking & LUSC – CIS

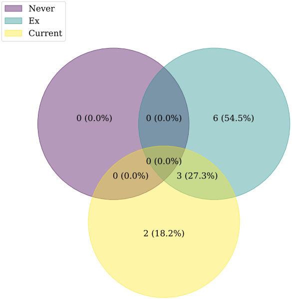
| CGC Genes | Never | Ex | Current |
|---|---|---|---|
| CDKN2A | | * | * |
| MLLT3 | | | * |



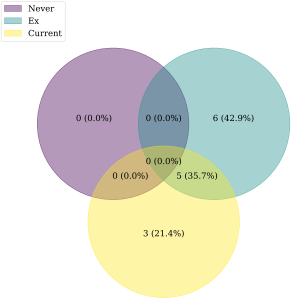
(a) Never Smoker (b) Ex-Smoker (c) Current Smoker

Figure: Gistic results in Smoking & LUSC – Precancer

# Peaks in Smoking & LUSC – Precancer I




(a) Amplification




(b) Deletion

Figure: Venn Diagram among Peaks in Smoking & LUSC – Precancer

# Peaks in Smoking & LUSC – Precancer II

Table: Amplification Peaks in Smoking & LUSC – Precancer

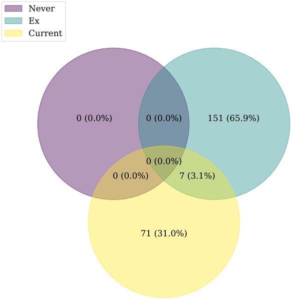
| Peaks | Never | Ex | Current | CGC Genes |
|---|---|---|---|---|
| 3q26.33 | | * | | SOX2 |
| 3q27.1 | | | * | |
| 3q28 | * | | | LPP,TP63 |
| 5p15.33 | | * | | SDHA,TERT |
| 6p12.3 | | * | | |
| 7p11.1 | | * | * | |
| 8p11.23 | | | * | FGFR1,NSD3 |
| 9p13.3 | | | * | FANCG |
| 11p13 | | * | | LMO2,WT1 |
| 11q13.3 | | * | | CCND1 |
| 17p11.1 | | * | * | |
| 17q21.31 | | * | | BRCA1,ETV4 |
| 18p11.1 | | * | * | |
| 19q13.2 | | * | | AKT2,CD79A,CIC |
| 19q13.12 | | | * | |
| 20p11.1 | | * | | |

# Peaks in Smoking & LUSC – Precancer III

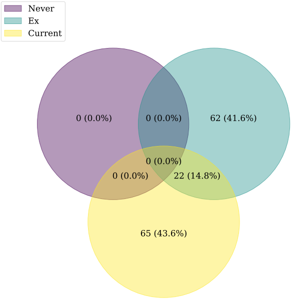
Table: Deletion Peaks in Smoking & LUSC – Precancer

| Peaks | Never | Ex | Current | CGC Genes |
|---|---|---|---|---|
| 1p36.11 | | | * | ARID1A,MDS2 |
| 1p36.33 | | * | | SKI |
| 1q21.3 | | * | * | ARNT,MLLT11,S100A7,SETDB1,TPM3 |
| 2p25.3 | | * | | |
| 2q36.3 | | | * | |
| 5p15.33 | | | * | SDHA,TERT |
| 5q35.3 | | * | * | FLT4,NSD1 |
| 6p21.32 | | * | * | DAXX |
| 7q22.1 | | * | | CUX1,TRRAP |
| 7q34 | | | * | BRAF,FAM131B,KIAA1549,TRIM24 |
| 7q35 | | * | | CNTNAP2 |
| 8p23.1 | * | | * | |
| 9p11.2 | | * | | |
| 9p21.2 | | | * | |
| 9p21.3 | | * | | CDKN2A,MLLT3 |
| 11q11 | | * | * | |
| 15q11.2 | | * | | |
| 17q12 | | * | * | CDK12,ERBB2,LASP1,MLLT6,TAF15 |
| 17q21.31 | | * | * | BRCA1,ETV4 |
| 18q11.1 | | | * | |
| 19q13.31 | | * | | |
| 19q13.41 | | * | * | PPP2R1A |
| 22q13.1 | * | | | APOBEC3B,MRTFA,PDGFB |

# Genes in Smoking & LUSC – Precancer I




(a) Amplification




(b) Deletion

Figure: Venn Diagram among Genes in Smoking & LUSC – Precancer

# Genes in Smoking & LUSC – Precancer II

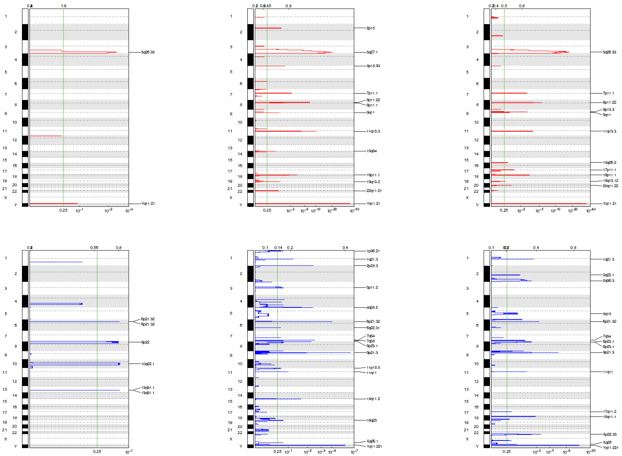
Table: Amplification Genes in Smoking & LUSC – Precancer

| CGC Genes | Never | Ex | Current |
|---|---|---|---|
| CCND1 | | * | |
| EXT2 | | * | |
| FANCG | | | * |
| LMO2 | | * | |
| PAX5 | | | * |
| SOX2 | | * | |
| WT1 | | * | |

Table: Deletion Genes in Smoking & LUSC – Precancer

| CGC Genes | Never | Ex | Current |
|---|---|---|---|
| CDKN2A | | * | * |
| MLLT3 | | | * |

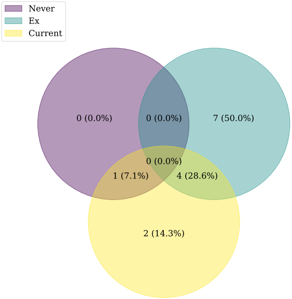
# Gistic in Smoking & LUSC – Primary



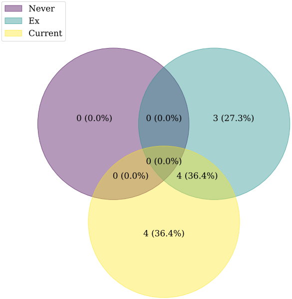
(a) Never Smoker (b) Ex-Smoker (c) Current Smoker

Figure: Gistic results in Smoking & LUSC – Primary

# Peaks in Smoking & LUSC – Primary I




(a) Amplification




(b) Deletion

Figure: Venn Diagram among Peaks in Smoking & LUSC – Primary

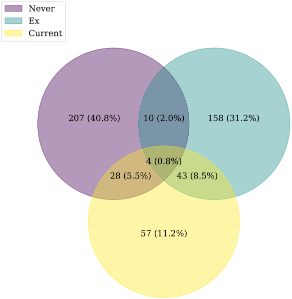
# Peaks in Smoking & LUSC – Primary II

Table: Amplification Peaks in Smoking & LUSC – Primary

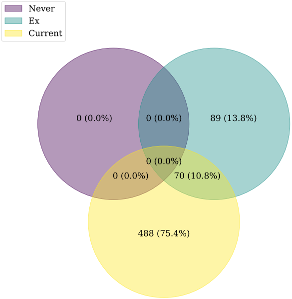
| Peaks | Never | Ex | Current | CGC Genes |
|---|---|---|---|---|
| 2p15 | | * | | XPO1 |
| 3q26.33 | * | | * | SOX2 |
| 3q27.1 | | * | | |
| 5p15.33 | | * | | SDHA,TERT |
| 7p11.1 | | * | * | |
| 8p11.1 | | * | | |
| 8p11.22 | | * | * | |
| 9p13.3 | | | * | FANCG |
| 9q11 | | * | * | |
| 11q13.3 | | * | * | CCND1 |
| 13q34 | | * | | |
| 15q26.2 | | | * | |
| 17p11.1 | | | * | |
| 18p11.1 | | * | * | |
| 19q13.2 | | * | | AKT2,CD79A,CIC |
| 19q13.12 | | | * | |
| 20q11.22 | | | * | |
| 22q11.21 | | * | | CLTCL1,DGCR8,LZTR1,SEPT5 |

Table: Deletion Peaks in Smoking & LUSC – Primary

| Peaks | Never | Ex | Current | CGC Genes |
|---|---|---|---|---|
| 1p36.21 | | * | | CASP9,PRDM2,SPEN |
| 1q21.3 | | * | * | ARNT,MLLT11,S100A7,SETDB1,TPM3 |
| 2p25.3 | | * | | |
| 2q22.1 | | | * | CXCR4,LRP1B |
| 2q36.3 | | | * | |
| 3p11.2 | | * | | |
| 4q35.2 | | * | | DUX4L1,FAT1 |
| 5q15 | | | * | |
| 6p21.32 | * | * | * | DAXX |
| 6q22.31 | | * | | |
| 7q34 | | * | * | BRAF,FAM131B,KIAA1549,TRIM24 |
| 7q35 | | * | | CNTNAP2 |
| 8p22 | * | | | PCM1 |
| 8p23.1 | | * | * | |
| 9p21.3 | | * | * | CDKN2A,MLLT3 |
| 10q22.1 | * | | | PRF1 |
| 11p15.5 | | * | | HRAS |
| 11q11 | | * | * | |
| 13q31.1 | * | | | |
| 15q11.2 | | * | | |
| 17q11.2 | | | * | NF1,SUZ12 |
| 18q11.1 | | | * | |
| 18q23 | | * | | |




(a) Amplification




(b) Deletion

Figure: Venn Diagram among Genes in Smoking & LUSC – Primary

# Genes in Smoking & LUSC – Primary II

Table: Amplification Genes in Smoking & LUSC – Primary

| CGC Genes | Never | Ex | Current |
|---|---|---|---|
| AKT2 | | * | |
| ANK1 | | | * |
| BCL11A | | * | |
| BCL6 | * | | |
| CCND1 | | | * |
| DGCR8 | | * | |
| EIF4A2 | * | | |
| ETV5 | * | | |
| FGFR1 | | * | * |
| HOOK3 | | | * |
| IGF2BP2 | * | | |
| IKBKB | | | * |
| KAT6A | | | * |
| LPP | * | | |
| LZTR1 | | * | |
| MAP3K13 | * | | |
| MB21D2 | * | | |
| MUC4 | * | | |
| PIK3CA | * | | * |
| REL | | * | |
| SOX2 | * | * | * |
| TBL1XR1 | * | | |
| TFRC | * | | |
| TP63 | * | | |
| XPO1 | | * | |

# Genes in Smoking & LUSC – Primary III

Table: Deletion Genes in Smoking & LUSC – Primary

| CGC Genes | Never | Ex | Current |
|---|---|---|---|
| ACKR3 | | | * |
| ACSL3 | | | * |
| ARHGEF10 | | * | * |
| ATIC | | | * |
| BARD1 | | | * |
| CASP3 | | * | |
| CDKN2A | | * | * |
| CREB1 | | | * |
| CUL3 | | | * |
| ERBB4 | | | * |
| FAT1 | | * | |
| FEV | | | * |
| IDH1 | | | * |
| PAX3 | | | * |

# Findings in Smoking & LUSC Gistic Results

# 4. Results

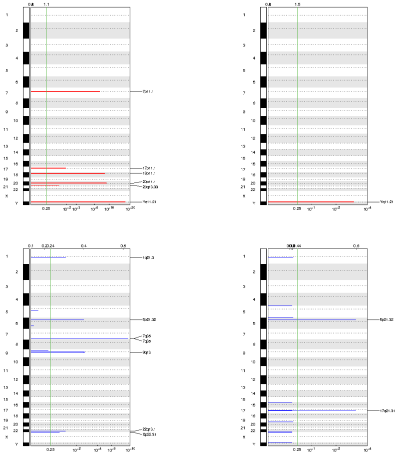
## 4.5. Copy Number Variation Analysis with Gistic

### 4.5.6. Gistic in Smoking & LUAD

# LUAD Data Composition with Smoking

Table: LUAD WES Data with Recurrence

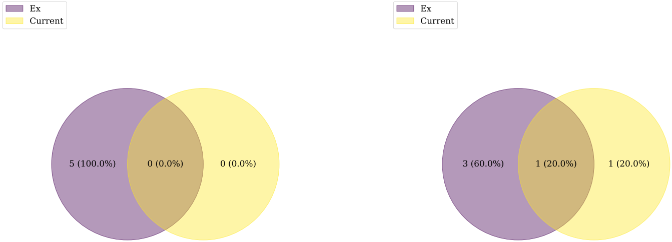
| Smoking? | Stage | Number of Samples |
|---|---|---|
| Never | Normal | 1 |
| | CIS+AIS | 1 |
| | Primary | 1 |
| | Total | 3 |
| Ex | Normal | 10 |
| | AAH | 9 |
| | CIS+AIS | 6 |
| | Primary | 10 |
| | Total | 35 |
| Current | Normal | 7 |
| | AAH | 6 |
| | CIS+AIS | 2 |
| | MIA | 1 |
| | Primary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




(a) Ex-Smoker (b) Current Smoker

Figure: Gistic results in Smoking & LUAD – AAH

# Peaks in Smoking & LUAD – AAH I




(a) Amplification (b) Deletion

Figure: Venn Diagram among Peaks in Smoking & LUAD – AAH

# Peaks in Smoking & LUAD – AAH II

Table: Amplification Peaks in Smoking & LUAD – AAH

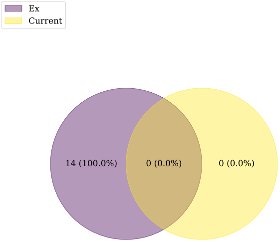
| Peaks | Ex | Current | CGC Genes |
|---|---|---|---|
| 7p11.1 | * | | |
| 17p11.1 | * | | |
| 18p11.1 | * | | |
| 20p11.1 | * | | |
| 20q13.33 | * | | PTK6,SS18L1 |

# Peaks in Smoking & LUAD – AAH III

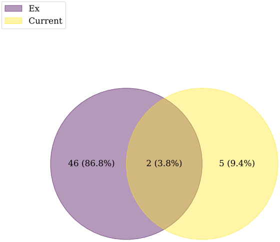
Table: Deletion Peaks in Smoking & LUAD – AAH

| Peaks | Ex | Current | CGC Genes |
|---|---|---|---|
| 1q21.3 | * | | ARNT,MLLT11,S100A7,SETDB1,TPM3 |
| 6p21.32 | * | * | DAXX |
| 7q35 | * | | CNTNAP2 |
| 9q13 | * | | |
| 17q21.31 | | * | BRCA1,ETV4 |
| 22q13.1 | * | | APOBEC3B,MRTFA,PDGFB |

# Genes in Smoking & LUAD – AAH I




(a) Amplification




(b) Deletion

Figure: Venn Diagram among Genes in Smoking & LUAD – AAH

# Genes in Smoking & LUAD – AAH II

Table: Amplification Genes in Smoking & LUAD – AAH

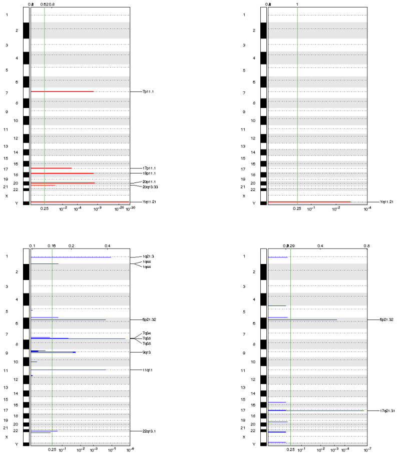
| Genes | Ex | Current |
|---|---|---|
| C20orf197 | * | |
| CDH26 | * | |
| CDH4 | * | |
| FRG1CP | * | |
| LOC100506470 | * | |
| LOC101928048 | * | |
| LOC644669 | * | |
| LOC729296 | * | |
| MIR4533 | * | |
| MIR548AG2 | * | |
| MIR646 | * | |
| MIR646HG | * | |
| MTRNR2L1 | * | |
| ZNF716 | * | |

# Genes in Smoking & LUAD – AAH III

Table: Deletion Genes in Smoking & LUAD – AAH

| | Ex | Current |
|---|---|---|
| Genes | | |
| ANKRD20A1 | * | |
| ANKRD20A3 | * | |
| ANKRD20A4 | * | |
| AQP7P1 | * | |
| AQP7P3 | * | |
| ARHGEF34P | * | |
| ARL17A | | * |
| ARL17B | | * |
| C6orf10 | * | * |
| CBWD5 | * | |
| CTAGE4 | * | |
| CTAGE8 | * | |
| FAM27C | * | |
| FAM74A1 | * | |
| FAM74A3 | * | |
| FAM74A4 | * | |
| FAM95B1 | * | |
| FGF7P6 | * | |
| FLJ43315 | * | |
| FOXD4L4 | * | |
| FOXD4L5 | * | |
| FRG1JP | * | |
| GXYLT1P3 | * | |
| LCE3A | * | |
| LCE3B | * | |
| LINC01189 | * | |
| LINC01410 | * | |
| LOC100132249 | * | |
| LOC101928195 | * | |
| LOC101928381 | * | |
| LOC101929163 | * | * |
| LOC101929583 | * | |
| LOC102723709 | * | |
| LOC102724238 | * | |
| LOC103908605 | * | |
| LOC286297 | * | |
| LOC403323 | * | |
| LOC554249 | * | |
| LOC642929 | * | |
| LOC728673 | * | |
| LRRC37A | | * |
| LRRC37A2 | | * |
| MIR4477A | * | |
| NSFP1 | | * |
| OR2A1 | * | |
| OR2A20P | * | |
| OR2A7 | * | |
| OR2A9P | * | |
| PTGER4P2-CDK2AP2P2 | * | |
| SPATA31A5 | * | |
| SPATA31A7 | * | |
| XLOC_007697 | * | |
| ZNF658 | * | |

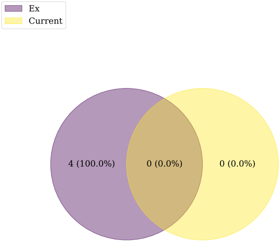
# Gistic in Smoking & LUAD – Precancer



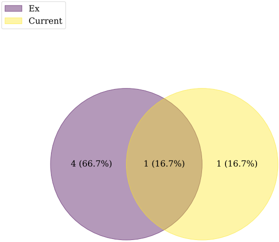
(a) Ex-Smoker (b) Current Smoker

Figure: Gistic results in Smoking & LUAD – Precancer

# Peaks in Smoking & LUAD – Precancer I




(a) Amplification




(b) Deletion

Figure: Venn Diagram among Peaks in Smoking & LUAD – Precancer

# Peaks in Smoking & LUAD – Precancer II

Table: Amplification Peaks in Smoking & LUAD – Precancer

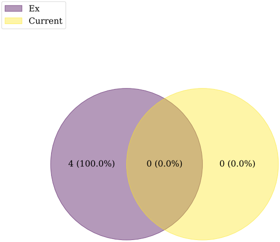
| Peaks | Ex | Current | CGC Genes |
|---|---|---|---|
| 7p11.1 | * | | |
| 17p11.1 | * | | |
| 18p11.1 | * | | |
| 20p11.1 | * | | |
| 20q13.33 | * | | PTK6,SS18L1 |

# Peaks in Smoking & LUAD – Precancer III

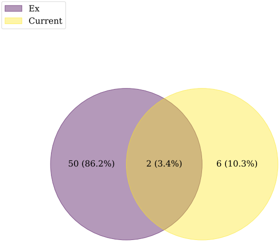
Table: Deletion Peaks in Smoking & LUAD – Precancer

| Peaks | Ex | Current | CGC Genes |
|---|---|---|---|
| 1q21.3 | * | | ARNT,MLLT11,S100A7,SETDB1,TPM3 |
| 1q44 | * | | AKT3 |
| 6p21.32 | * | * | DAXX |
| 7q34 | * | | BRAF,FAM131B,KIAA1549,TRIM24 |
| 7q35 | * | | CNTNAP2 |
| 9q13 | * | | |
| 11q11 | * | | |
| 17q21.31 | | * | BRCA1,ETV4 |
| 22q13.1 | * | | APOBEC3B,MRTFA,PDGFB |

# Genes in Smoking & LUAD – Precancer I




(a) Amplification




(b) Deletion

Figure: Venn Diagram among Genes in Smoking & LUAD – Precancer

# Genes in Smoking & LUAD – Precancer II

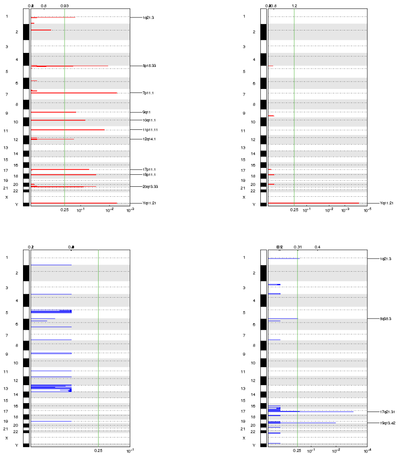
Table: Amplification Genes in Smoking & LUAD – Precancer

| Genes | Ex | Current |
|---|---|---|
| FRG1CP | * | |
| LOC644669 | * | |
| MTRNR2L1 | * | |
| ZNF716 | * | |

# Genes in Smoking & LUAD – Precancer III

Table: Deletion Genes in Smoking & LUAD – Precancer

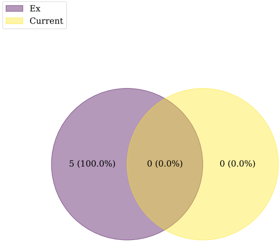
| Genes | Ex | Current |
|---|---|---|
| ANKRD20A1 | * | |
| ANKRD20A3 | * | |
| ANKRD20A4 | * | |
| AQP7P1 | * | |
| AQP7P3 | * | |
| ARHGEF34P | * | |
| ARL17A | | * |
| ARL17B | | * |
| C6orf10 | * | * |
| CBWD5 | * | |
| CTAGE4 | * | |
| CTAGE8 | * | |
| FAM27C | * | |
| FAM74A1 | * | |
| FAM74A3 | * | |
| FAM74A4 | * | |
| FAM95B1 | * | |
| FGF7P6 | * | |
| FLJ43315 | * | |
| FOXD4L4 | * | |
| FOXD4L5 | * | |
| FRG1JP | * | |
| GXYLT1P3 | * | |
| HCG23 | | * |
| LCE3A | * | |
| LCE3B | * | |
| LINC01189 | * | |
| LINC01410 | * | |
| LOC100132249 | * | |
| LOC101928195 | * | |
| LOC101928381 | * | |
| LOC101929163 | * | * |
| LOC101929583 | * | |
| LOC102723709 | * | |
| LOC102724238 | * | |
| LOC103908605 | * | |
| LOC286297 | * | |
| LOC403323 | * | |
| LOC554249 | * | |
| LOC642929 | * | |
| LOC728673 | * | |
| LRRC37A | | * |
| LRRC37A2 | | * |
| MIR4477A | * | |
| NSFP1 | | * |
| OR2A1 | * | |
| OR2A20P | * | |
| OR2A7 | * | |
| OR2A9P | * | |
| OR4C11 | * | |
| OR4C6 | * | |
| OR4P4 | * | |
| OR4S2 | * | |
| PTGER4P2-CDK2AP2P2 | * | |
| SPATA31A5 | * | |
| SPATA31A7 | * | |
| XLOC_007697 | * | |
| ZNF658 | * | |



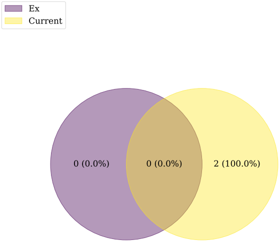
(a) Ex-Smoker (b) Current Smoker

Figure: Gistic results in Smoking & LUAD – Primary

# Peaks in Smoking & LUAD – Primary I




(a) Amplification




(b) Deletion

Figure: Venn Diagram among Peaks in Smoking & LUAD – Primary

# Peaks in Smoking & LUAD – Primary II

Table: Amplification Peaks in Smoking & LUAD – Primary

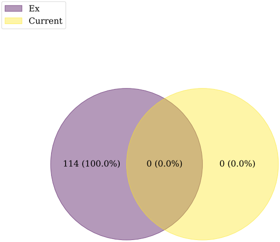
| Peaks | Ex | Current | CGC Genes |
|---|---|---|---|
| 1q21.3 | * | | ARNT,MLLT11,S100A7,SETDB1,TPM3 |
| 5p15.33 | * | | SDHA,TERT |
| 7p11.1 | * | | |
| 9q11 | * | | |
| 10q11.1 | * | | |
| 11p11.11 | * | | |
| 12q14.1 | * | | CDK4,LRIG3 |
| 17p11.1 | * | | |
| 18p11.1 | * | | |
| 20q13.33 | * | | PTK6,SS18L1 |

# Peaks in Smoking & LUAD – Primary III

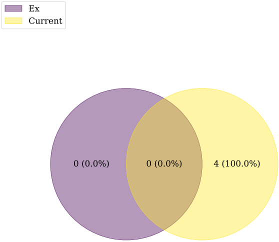
Table: Deletion Peaks in Smoking & LUAD – Primary

| Peaks | Ex | Current | CGC Genes |
|---|---|---|---|
| 1q21.3 | | * | ARNT,MLLT11,S100A7,SETDB1,TPM3 |
| 5q35.3 | | * | FLT4,NSD1 |
| 17q21.31 | | * | BRCA1,ETV4 |
| 19q13.42 | | * | CNOT3,TFPT,ZNF331 |

# Genes in Smoking & LUAD – Primary I




(a) Amplification




(b) Deletion

Figure: Venn Diagram among Genes in Smoking & LUAD – Primary

# Genes in Smoking & LUAD – Primary II

Table: Amplification Genes in Smoking & LUAD – Primary

| CGC Genes | Ex | Current |
|---|---|---|
| PTK6 | * | |
| SS18L1 | * | |

Table: Deletion Genes in Smoking & LUAD – Primary

| Genes | Ex | Current |
|---|---|---|
| ARL17A | | * |
| ARL17B | | * |
| LILRB4 | | * |
| MIR8061 | | * |

# Findings in Smoking & LUAD Gistic Results

# Findings in Gistic

# 4. Results

## 4.6. Single Nucleotide Variations Analysis

# Mutect2?

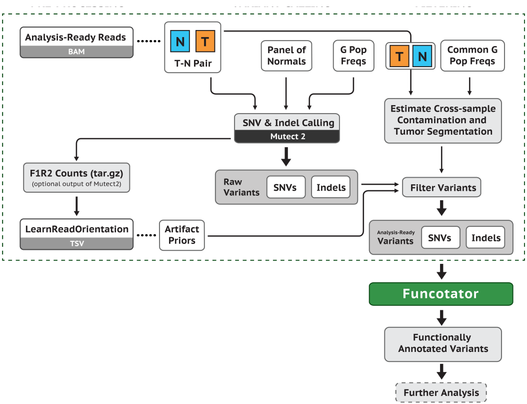



Figure: Somatic short variant discovery workflow (Van der Auwera et al., 2013; DePristo et al., 2011)

# MutEnricher?

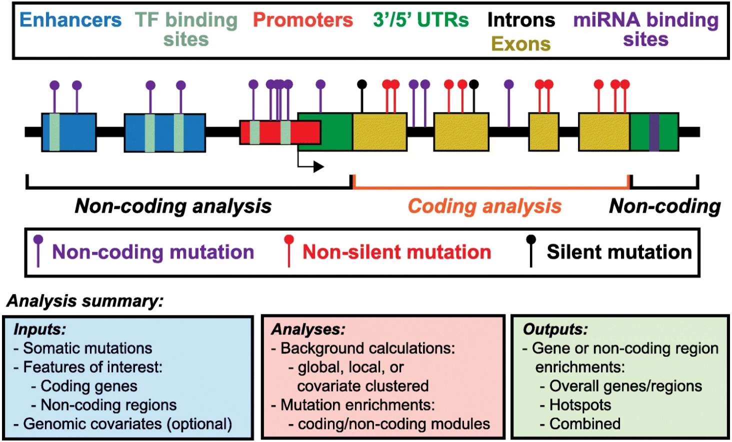



Figure: Schematic representation of MunEnricher's analysis procedures (Soltis et al., 2020)

# CoMut?

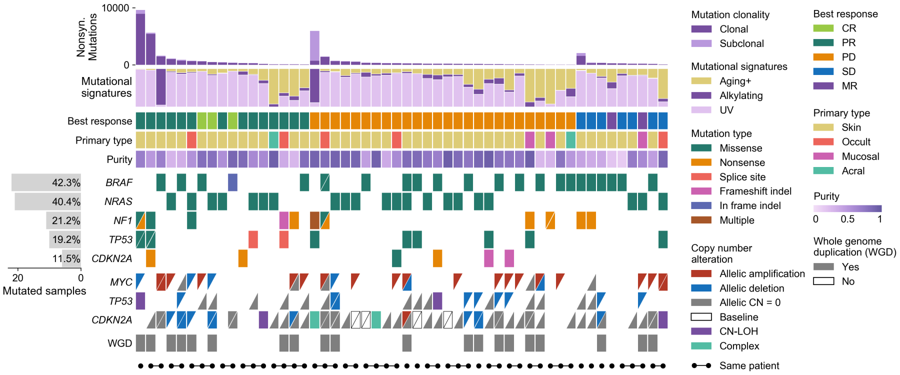



Figure: A comutation plot generated with CoMut (Crowdis et al., 2020)

# 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 $< 0.05$

# 4. Results

## 4.6. Single Nucleotide Variations Analysis

### 4.6.1. Somatic Variant


Figure: CoMut Plot with LUSC Patients

# Somatic Variant in LUAD

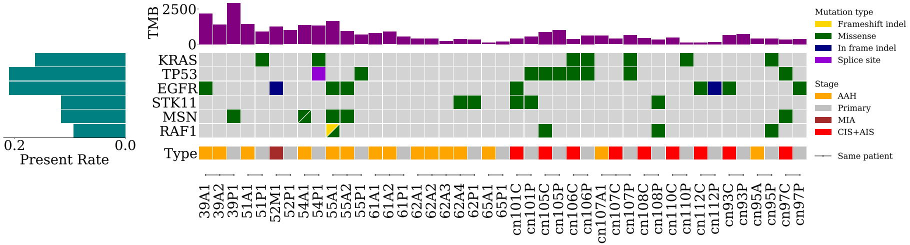



Figure: CoMut Plot with LUAD Patients

# 4. Results

## 4.6. Single Nucleotide Variations Analysis

### 4.6.2. Somatic Variant with Recurrence

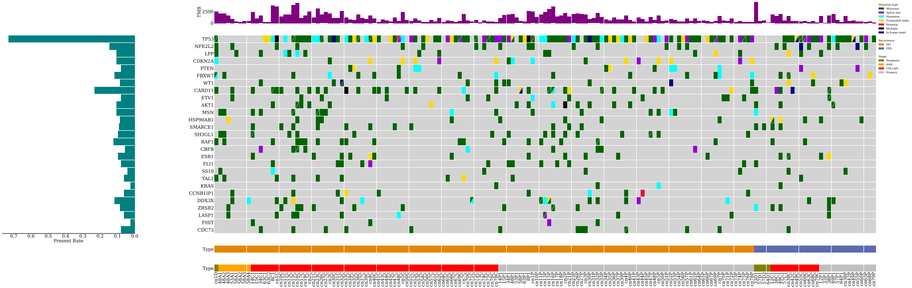



Figure: CoMut Plot in LUSC Patients with Recurrence

# Somatic Variant in LUAD with Recurrence

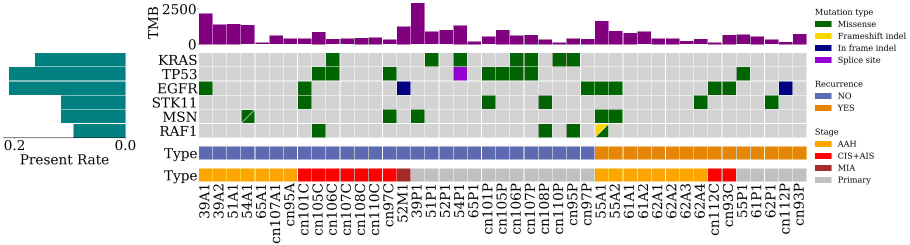



Figure: CoMut Plot in LUAD Patients with Recurrence

# 4. Results

## 4.6. Single Nucleotide Variations Analysis

### 4.6.3. Somatic Variant with Smoking

# Somatic Variant in LUSC with Smoking

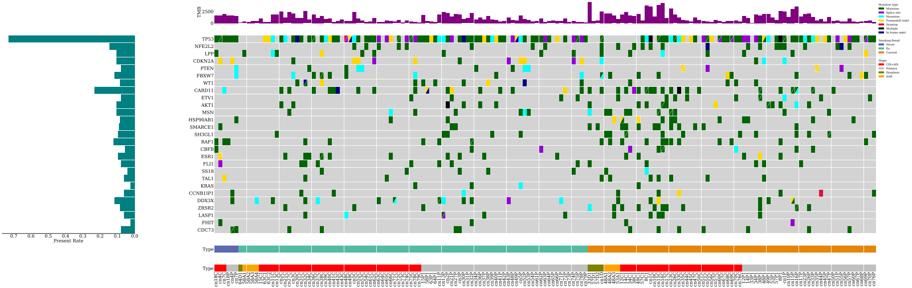



Figure: CoMut Plot in LUSC Patients with Smoking

# Somatic Variant in LUAD with Smoking

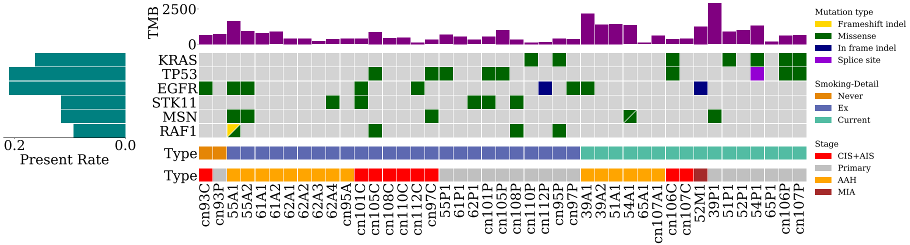



Figure: CoMut Plot in LUAD Patients with Smoking

# Findings in SNVs Analysis

# 4. Results

## 4.7. VAF Analysis

# VAF?

- Variant allele frequency
- $\mathrm{VAF} = \mathrm{Alternative\ allele\ read\ count/Total\ read\ count}$
- To find tumor evolution

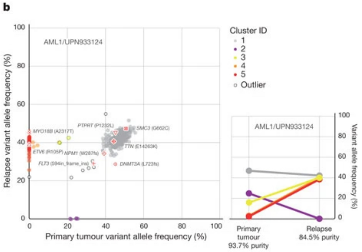
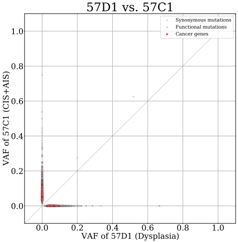
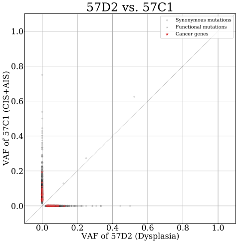



Figure: VAF distribution of validated mutations (L. Ding et al., 2012)

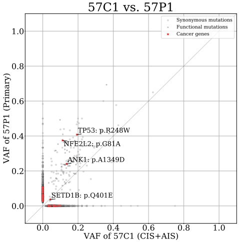
# VAF Plots




(a) Dysplasia + CIS




(b) Dysplasia + CIS




(c) CIS + Primary

Figure: VAF plots in patient #57

# PyClone?

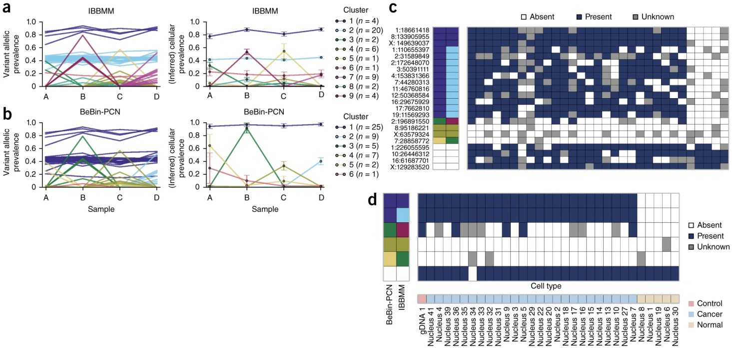
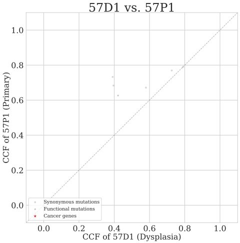
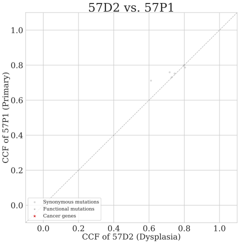


Figure: Analysis of multiple samples by PyClone (Roth et al., 2014)

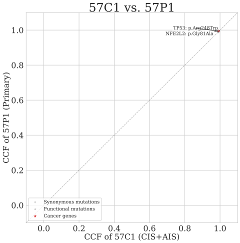
# PyClone Plots I




(a) 57D1



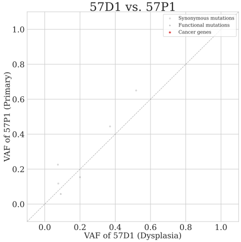

(b) 57D2



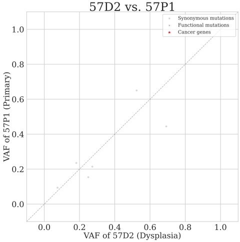

(c) 57C1

Figure: CCF plot in patient #57

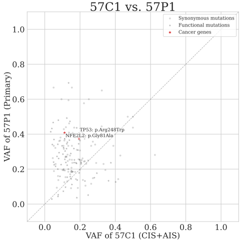
# PyClone Plots II




(a) 57D1




(b) 57D2




(c) 57C1

Figure: VAF plot in patient #57

# Findings in VAF Analysis

# 4. Results

## 4.8. Tumor Evolution Trajectories Analysis

# Mobster?

# Findings in Tumor Evolution Trajectories Analysis

# 4. Results

## 4.9. Bulk Cell Deconvolution

# BisqueRNA?

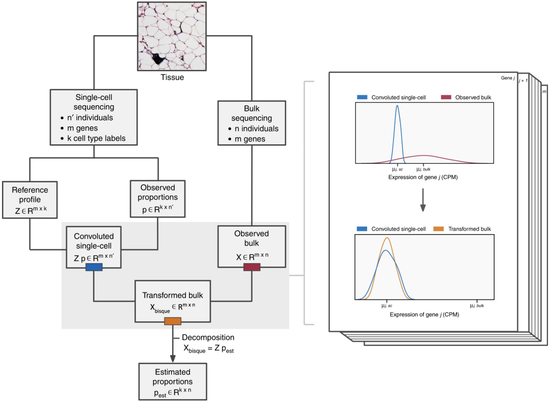



Figure: Workflow for BisqueRNA (Jew et al., 2020)

# 4. Results

## 4.9. Bulk Cell Deconvolution

### 4.9.1. Reference by N. Kim et al. (2020)

# Reference Single-cell Data

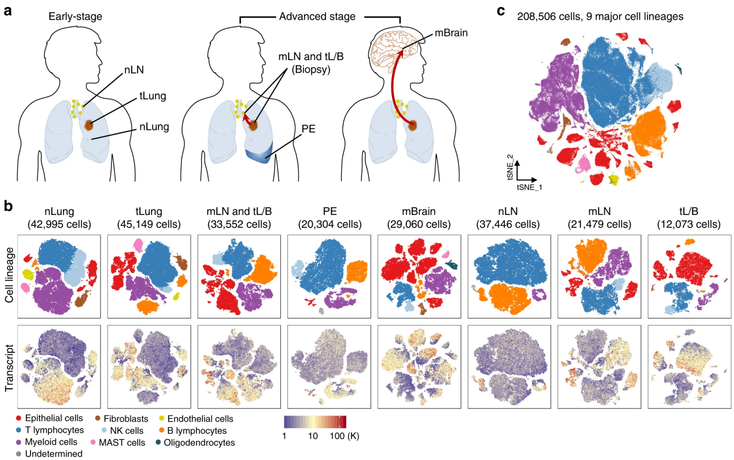



Figure: Comprehensive dissection and clustering of 208,506 single cells from LUAD patients (N. Kim et al., 2020)

# Cluster Plot in LUSC

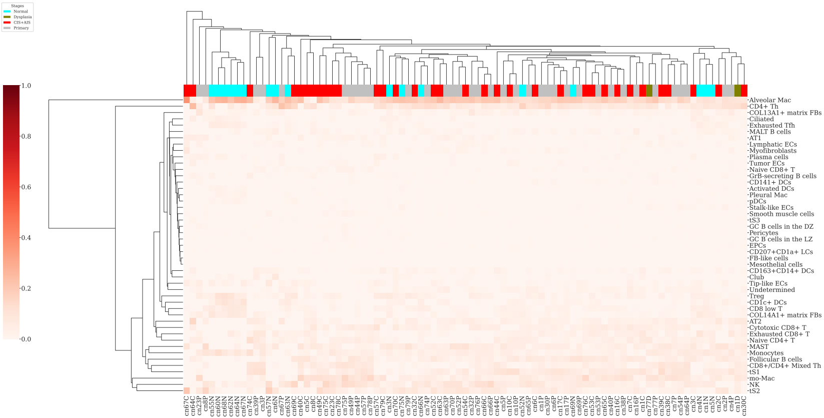
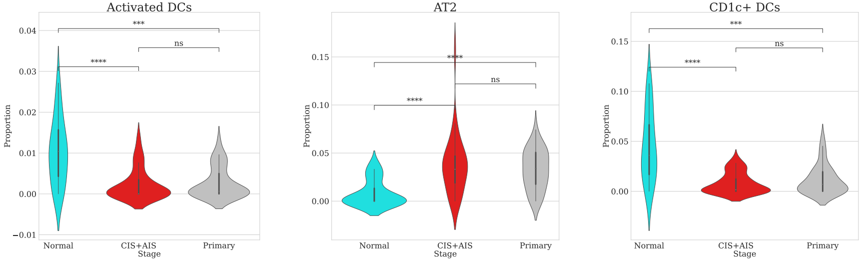



Figure: Cluster Plot in LUSC

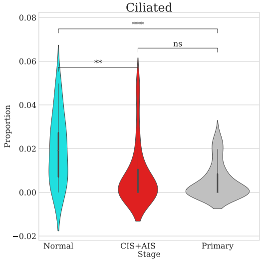
# Violin Plots in LUSC I



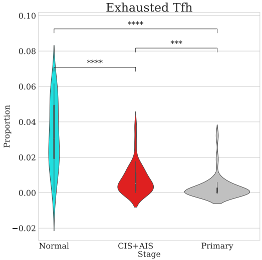

(a) Activated DCs (b) Alveolar type II (c) Langerhans cells

Figure: Violin Plots in LUSC

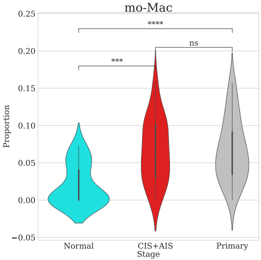
# Violin Plots in LUSC II




(d) Ciliated cells



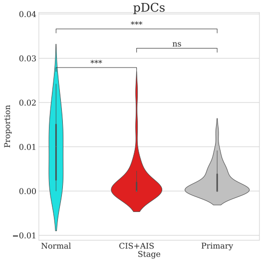

(e) Exhausted T follicular helper



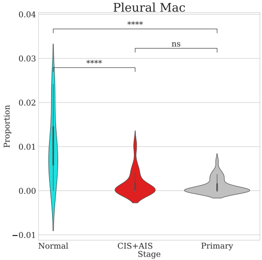

(f) Mo & Mac

Figure: Violin Plots in LUSC

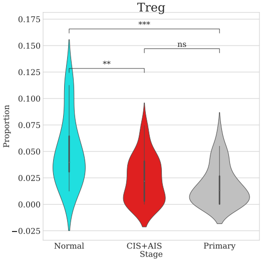
# Violin Plots in LUSC III




(g) Plasmacytoid DCs



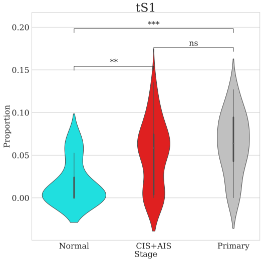

(h) Pleural Mac




(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).

# Findings in Bulk Cell Deconvolution with LUSC IV

## 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 $\Rightarrow$ 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 $\Uparrow$ in solid tumors & hematologic malignancies (Beyer & Schultze, 2006).
3. Increasing Tregs $\Uparrow$ 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 (N. Kim et al., 2020).

# Cluster Plot in LUAD

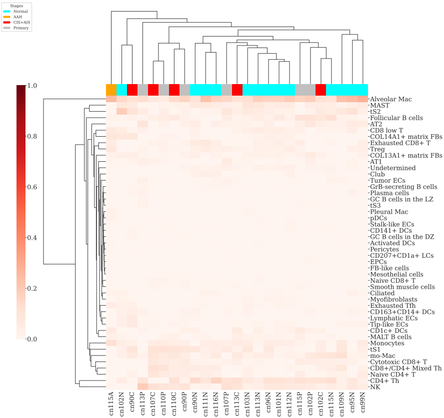
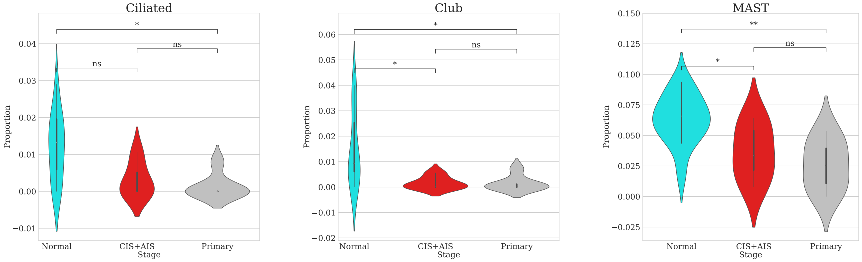



Figure: Cluster Plot in LUAD

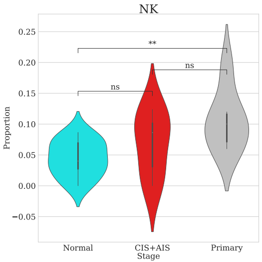
# Violin Plots in LUAD I



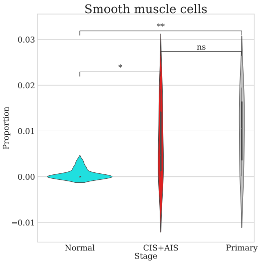

(a) Ciliated cells (b) Club Cell (c) Mast cell

Figure: Violin Plots in LUAD

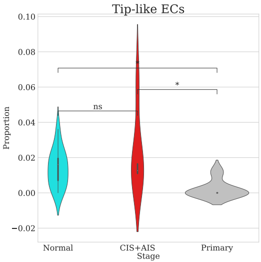
# Violin Plots in LUAD II




(d) NK cells




(e) Smooth muscle cells




(f) Tip-like ECs

Figure: Violin Plots in LUAD

# Findings in Bulk Cell Deconvolution with LUAD I

## Cilited cells

1. Cilited 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 $\Rightarrow$ Good indicator of advanced bronchopulmonary dysplasia (Rokicki, Rokicki, Wojtacha, & Dżeljijli, 2016).

## Mast cells

1. Mast cells have higher proportion in Normal than Primary samples.
2. Mast cells activated by lung cancer-derived extracellular vesicles (H. Xiao et al., 2019).
3. Mast cell promote $\Uparrow$ tumor metastasis (Salamon, Mekori, & Shefler, 2020).

# Findings in Bulk Cell Deconvolution with LUAD III

## Natural Killer cells (NK cells)

1. NK cells have higher proportion in Primary than Normal samples.
2. NK cells play a major role in innate immune system (Shin et al., 2020).
3. NK cells can induce immune response against tumor cells (Shin et al., 2020).
4. NK cells may induce tumor regression in lung cancer (Aktaş et al., 2018) ??.

# Findings in Bulk Cell Deconvolution with LUAD IV

## Smooth muscle cells

1. Smooth muscle cells have higher proportion in Primary than Normal samples.
2. Hypoxia is a characteristic feature of solid tumors (Brahimi-Horn, Chiche, & Pouysségur, 2007; Vaupel & Mayer, 2007).
3. $\therefore$ Smooth muscle cells pathway is up-regulated in cancer (Kyotani, Takasawa, & Yoshizumi, 2019; T.-T. Zhu et al., 2019).

## Tip-like endothelial cells (ECs)

1. Tip-like ECs have lower proportion in Primary than Normal samples.
2. Tip-like ECs were determined into migratory & basement-membrane remodeling phenotypes (Goveia et al., 2020).
3. Tip-like ECs replaced with immature ones in NSCLC (E. Y. Kim et al., 2022).

# 4. Results

## 4.9. Bulk Cell Deconvolution

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

# Reference Single-cell Data

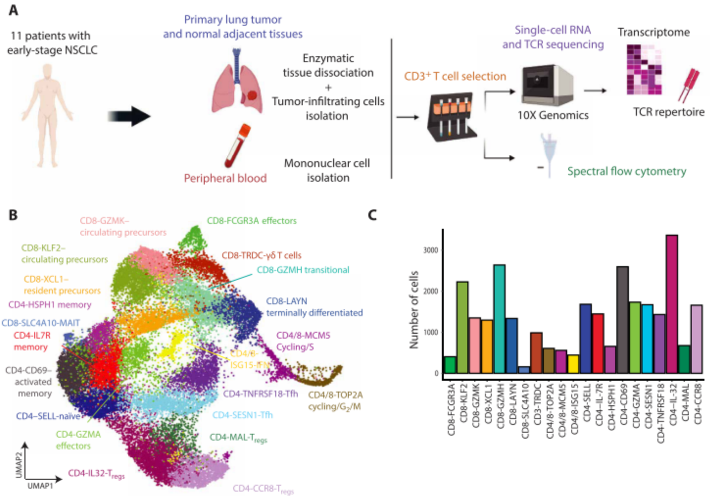



Figure: Characterization of $CD3^+$ TILs in NSCLC (Gueguen et al., 2021)

# Cluster Plots in LUSC

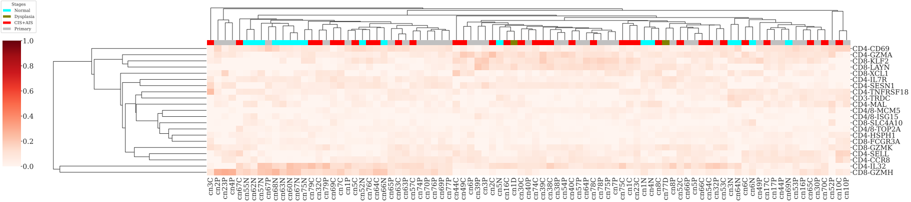
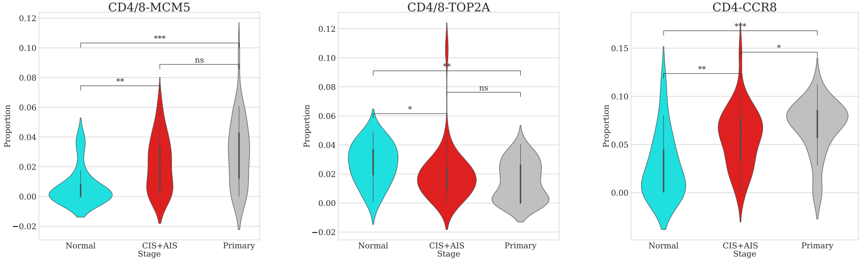


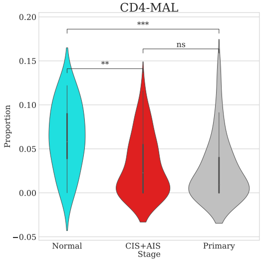

Figure: Cluster Plot in LUAD



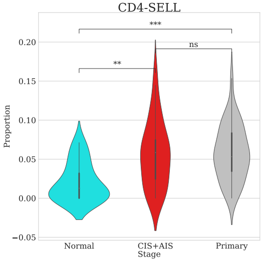

(a) CD4/8-MCM5 (b) CD4/8-TOP2A (c) CD4-CCR8

Figure: Violin Plots in LUSC

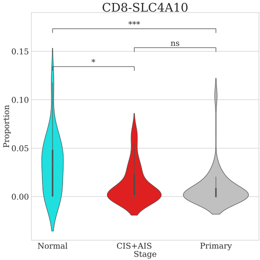
# Violin Plots in LUSC II




(d) CD4-MAL




(e) CD4-SELL




(f) CD8-SLC4A10

Figure: Violin Plots in LUSC

# Findings in Bulk Cell Deconvolution with LUSC I

## CD4/8-MCM5

1. CD4/8-MCM5 have lower proportion in Normal.
2. MCM5, mini-chromosome maintenance protein 5, acts as component of MCM complex (Tsuji, Ficarro, & Jiang, 2006).
3. MCM5, thus, play a major role in replication and cell cycle progression (Paul, Hu, Musahl, Hameister, & Knippers, 1996).
4. MCM5 could be adverse prognostic marker for NSCLC (Grzegrzolka et al., 2021) and lung cancer (Y.-Z. Liu et al., 2017).

# Findings in Bulk Cell Deconvolution with LUSC II

## CD4/8-TOP2A

1. CD4/8-TOP2A have higher proportion in Normal.
2. TOP2A, topoisomerase IIA, have an essential role for modulating DNA topology & cell division (Wyles, Wu, Mirski, & Cole, 2007).
3. TOP2A over-expressed $\Uparrow$ in bladder cancer (Zeng et al., 2019), LUAD (Kou et al., 2020) and NSCLC (W. Ma et al., 2019) ??

## CD4-CCR8

1. CD4-CCR8 is gradually increased along tumor progression.
2. CCR8, C-C chemokine receptor type 8, might modulate monocyte chemotaxis and tymic cell line apoptosis (Tiffany et al., 1997).
3. CCR8 up-regulated along tumor progression in bladder (X. Liu et al., 2019), colon (Villarreal et al., 2018), and breast cancer (Plitas et al., 2016).

# Findings in Bulk Cell Deconvolution with LUSC III

## CD4-MAL

1. CD4-MAL have lower proportion in Primary.
2. MAL, myelin and lymphocyte protein, play a role in indirect route for egress of transcytosing cargo (de Marco et al., 2002).
3. Over-expression of MAL was correlated with worse prognostic factors in uterine carcinoma (D. Li et al., 2021).
4. MAL was highly methylated in gastric cancer (Choi et al., 2017).
5. MAL acts as a tumor suppressor or a tumor progression factor among cancer types (Lara-Lemus, 2019).

# Findings in Bulk Cell Deconvolution with LUSC IV

## CD4-SELL

1. CD4-SELL have higher proportion in Primary.
2. SELL, a calcium-dependent lectin, controls cell adhesion with neighboring cells (Bernimoulin et al., 2003; Wedepohl et al., 2017).
3. SELL over-expressed in breast cancer (Kumari et al., 2021).

## CD8-SLC4A10

1. CD8-SLC4A10 have lower proportion in Primary.
2. SLC4A10, sodium-driven chloride bicarbonate exchanger, have an essential role in regulating intracellular pH (C.-Z. Wang, Yano, Nagashima, & Seino, 2000).
3. SLC4A10 distruption leads to an extreme change in the cellular phenotype (Christensen et al., 2020).

# Cluster Plots in LUAD

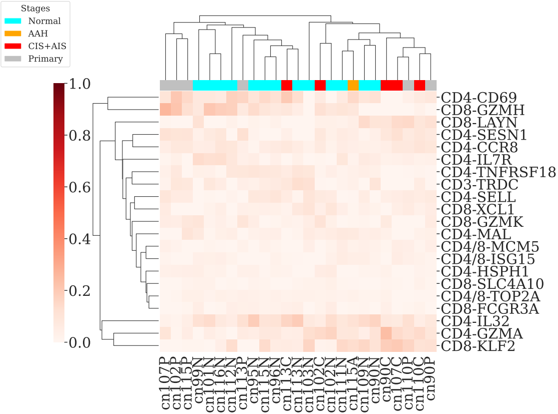
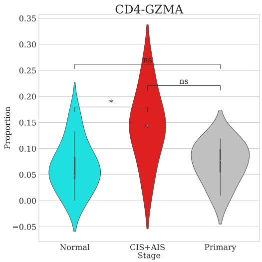
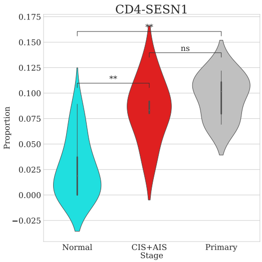



Figure: Cluster Plot in LUAD

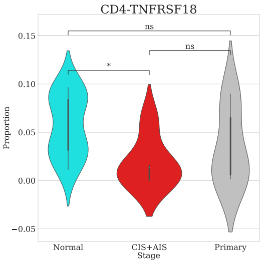
# Violin Plots in LUAD




(a) CD4-GZMA




(b) CD4-SESN1




(c) CD4-TNFRSF18

Figure: Violin Plots in LUAD

# Findings in Bulk Cell Deconvolution with LUAD I

## CD4-GZMA

1. CD4-GZMA have higher proportion in AIS than Normal.
2. GZMA, granzyme A, activates caspase-independent pyroptosis through the immunological synapse (Gershenfeld, Hershberger, Shows, & Weissman, 1988; Hameed, Lowrey, Lichtenheld, & Podack, 1988; Krähenbühl et al., 1988).
3. GZMA promotes many cancers.
    - colorectal cancer (Santiago et al., 2020; Narayanan et al., 2018)
    - breast cancer (Fisler, Sikaria, Yavorski, Tu, & Blanck, 2018)
    - NSCLC (Jia et al., 2018)

# Findings in Bulk Cell Deconvolution with LUAD II

## CD4-SESN1

1. CD4-SESN1 have higher proportion in AIS than Normal.
2. SESN1, sestrin-1, acts as an intracellular *leucine* sensor that controls the TORC1 signaling pathway (Chantranupong et al., 2014; Wolfson et al., 2016).
3. SESN1 is controlled by p53 tumor suppressor, and thus affects in cell growth regulation (Budanov & Karin, 2008).
4. Inhibitor of SESN1 implicates to the pro-oxidant and oncogenic effects of mutant p53 (Cordani et al., 2018)?
5. SESN1 plays opposite role in *early* and *late* stage of lung carcinogenesis (B. Ding et al., 2019).

# Findings in Bulk Cell Deconvolution with LUAD III

## CD4-TNFRSF18

1. CD4-TNFRSF18 have lower proportion in AIS than Normal.
2. TNFRSF18, tumor necrosis factor receptor super-family member 18, is a receptor for TNFSF18.
3. TNFSF18 regulates T-cell responses, and promotes $\Uparrow$ leukocyte adhesion to endothelial cells (Lacal et al., 2013).
4. TNFRSF18 was negatively correlated with survival in endometrial cancer (Zhou, Zhang, Li, Chen, & Cheng, 2020)?

# Findings in Bulk Cell Deconvolution

# 4. Results

## 4.10. Discovery of Mutational Signature

# Mutational Signature?

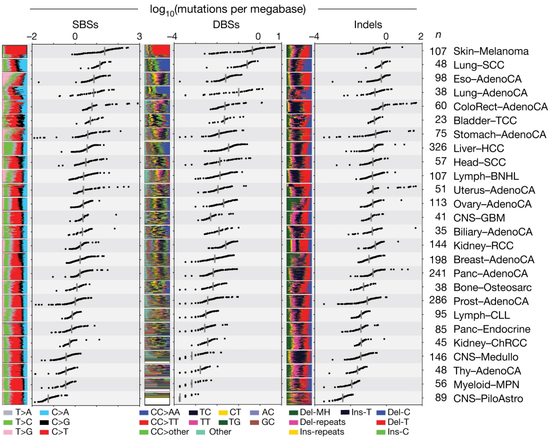



Figure: Mutation Burdens across PCAWG tumor types (Alexandrov et al., 2020)

# 4. Results

## 4.10. Discovery of Mutational Signature

### 4.10.1. SigProfiler

# SigProfiler?

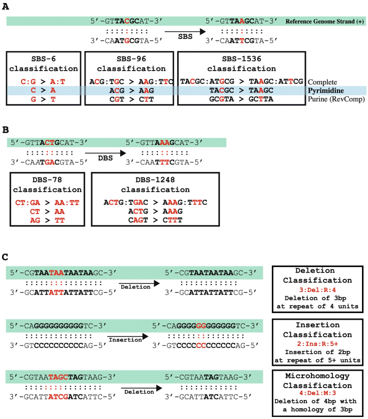



Figure: Classification of mutation signatures by SigProfiler (Bergstrom et al., 2019; Islam et al., 2021; Bergstrom et al., 2020)

# 4. Results

## 4.10. Discovery of Mutational Signature

### 4.10.2. Single Base Substitutions (SBS)

# SBS Signatures I

## SBS1

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

## SBS2

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

## SBS Signatures II

### SBS4

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

### SBS5

- Unknown (Alexandrov et al., 2013)
- SBS5 $\Uparrow$ in bladder cancer
- SBS5 $\Uparrow$ in many cancer types $\because$ Tobacco smoking

### SBS10b

- Polymerase $\epsilon$ exonuclease domain mutations (Alexandrov et al., 2020)

# SBS Signatures III

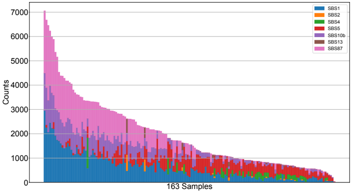
## SBS13

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

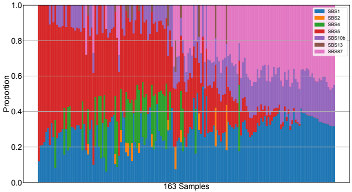
## SBS87

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

# SBS in LUSC I



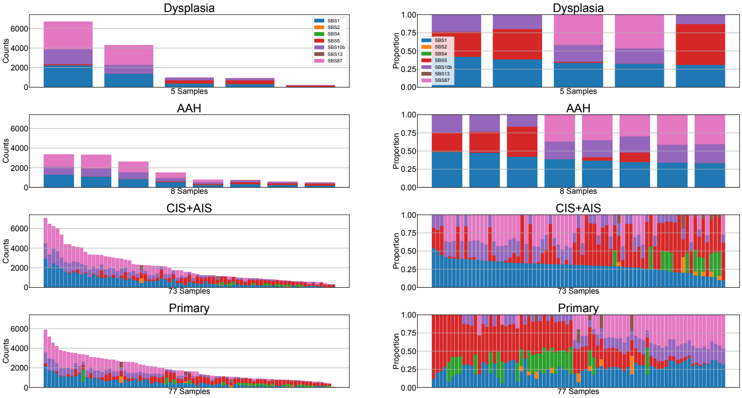

(a) Absolute




(b) Relative

Figure: SBS Bar Plot in LUSC

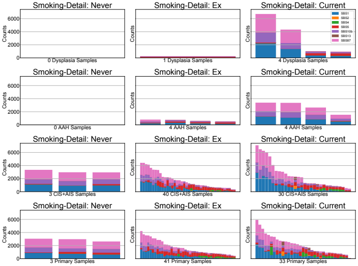
# SBS in LUSC II



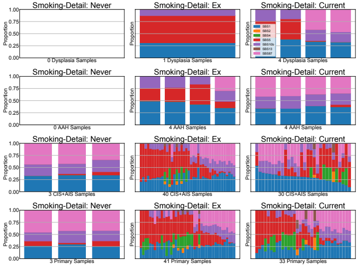

(a) Absolute (b) Relative

Figure: SBS Bar Plot by Cancer Subtype in LUSC

# SBS in LUSC with Smoking I



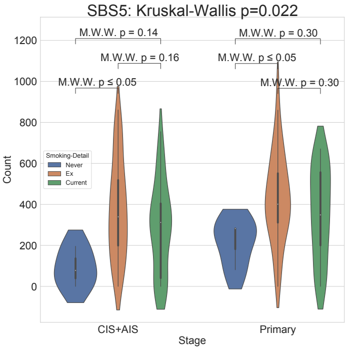

(a) Absolute



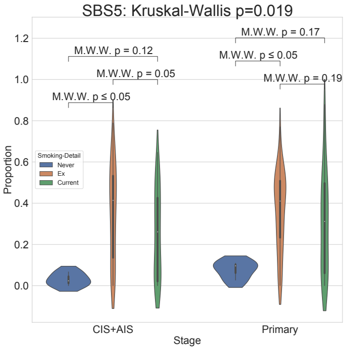

(b) Relative

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

# SBS in LUSC with Smoking II



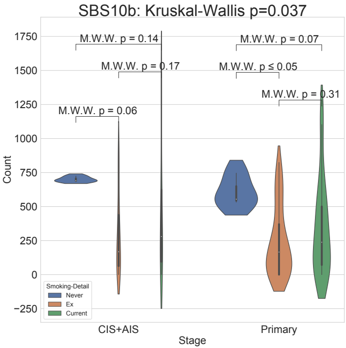

(a) Absolute



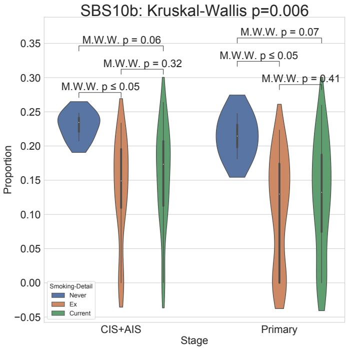

(b) Relative

Figure: SBS5 Signature in LUSC with Smoking

# SBS in LUSC with Smoking III



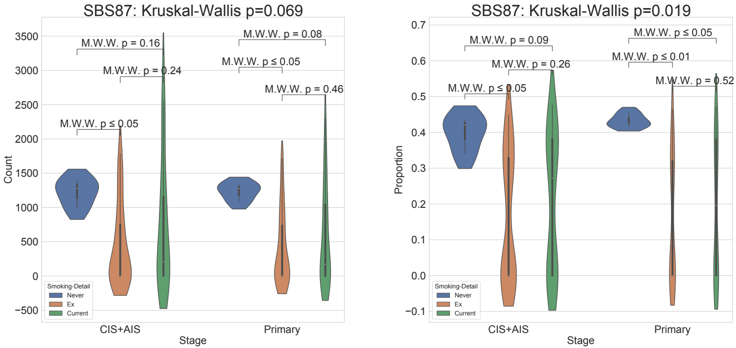

(a) Absolute




(b) Relative

Figure: SBS10b Signature in LUSC with Smoking

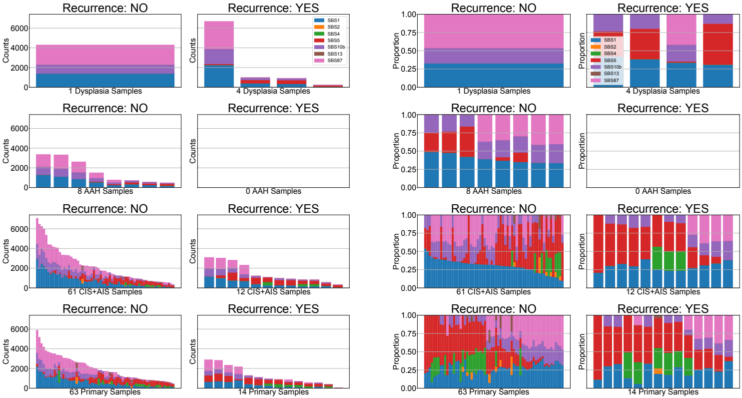
# SBS in LUSC with Smoking IV




(a) Absolute (b) Relative

Figure: SBS87 Signature in LUSC with Smoking

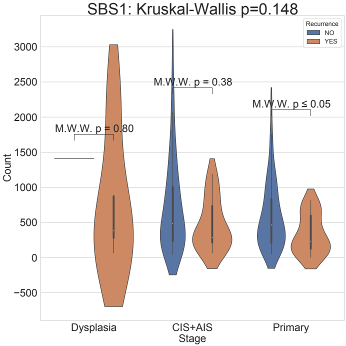
# SBS in LUSC with Recurrence I



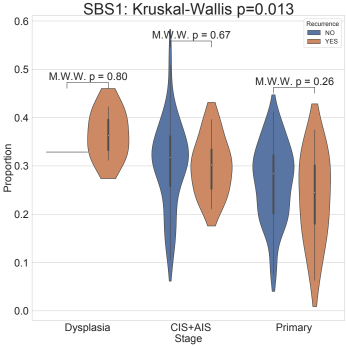

(a) Absolute (b) Relative

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

# SBS in LUSC with Recurrence II



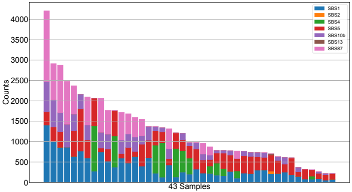

(a) Absolute



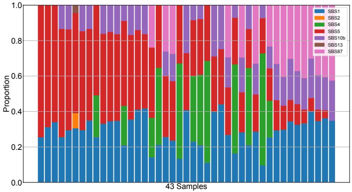

(b) Relative

Figure: SBS1 Signature in LUSC with Recurrence

# SBS in LUAD I



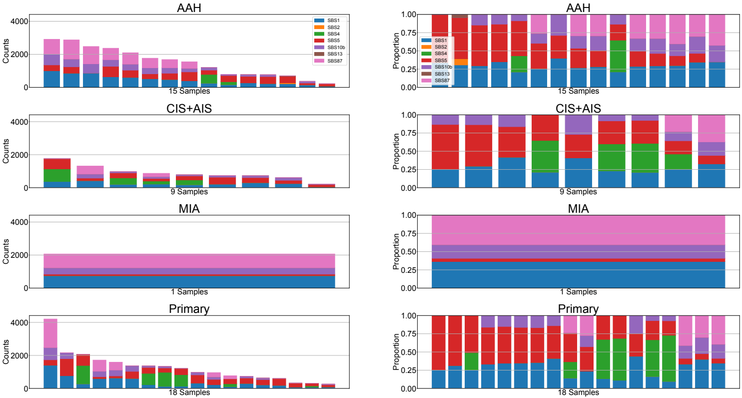

(a) Absolute




(b) Relative

Figure: SBS Bar Plot in LUSC

# SBS in LUAD II

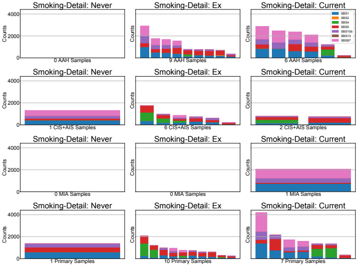



(a) Absolute

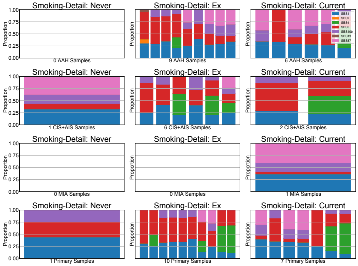
(b) Relative

Figure: SBS Bar Plot by Cancer Subtype in LUSC

# SBS in LUAD with Smoking I



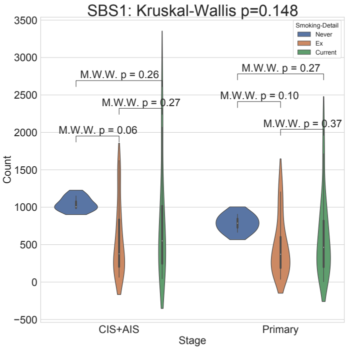

(a) Absolute



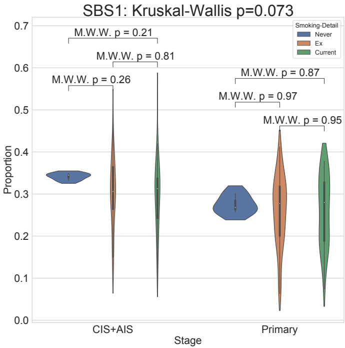

(b) Relative

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

# SBS in LUAD with Smoking II



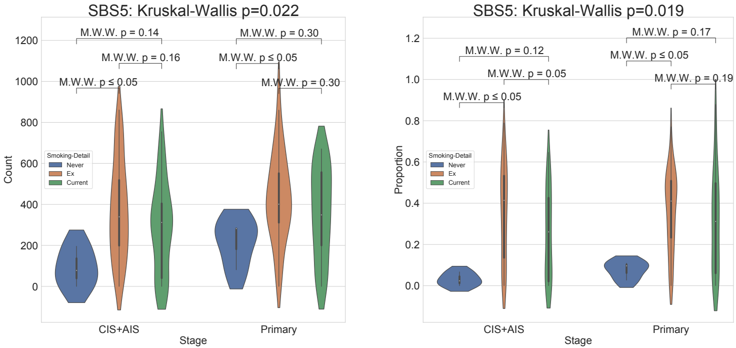

(a) Absolute




(b) Relative

Figure: SBS1 Signature in LUAD with Smoking

# SBS in LUAD with Smoking III

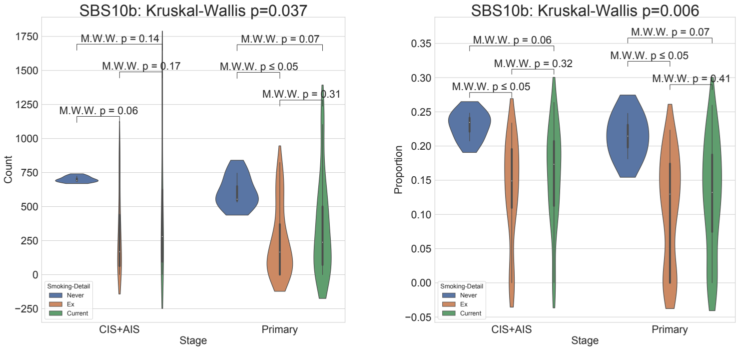



(a) Absolute

(b) Relative

Figure: SBS5 Signature in LUAD with Smoking

# SBS in LUAD with Smoking IV

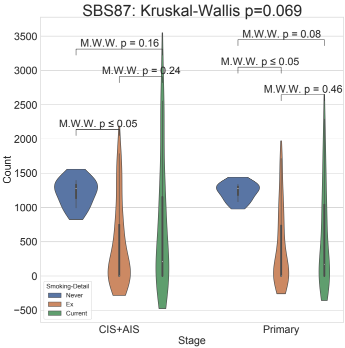



(a) Absolute

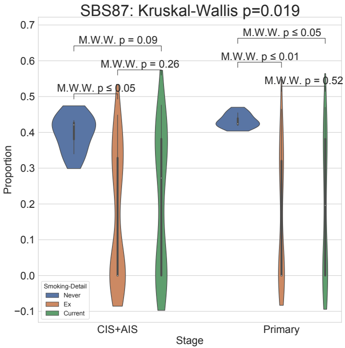
(b) Relative

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

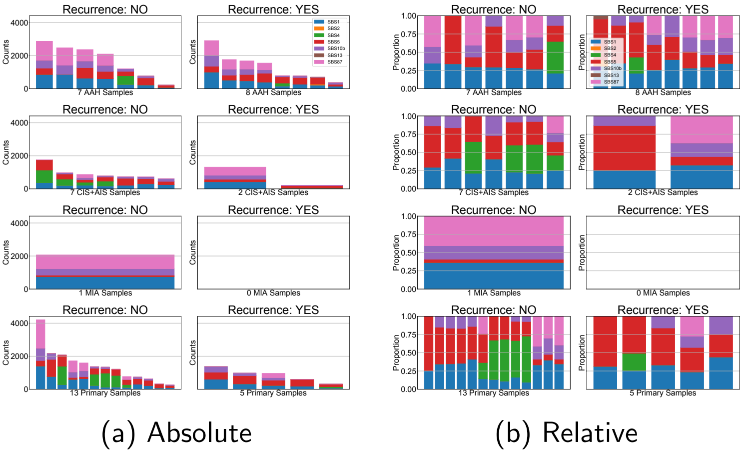
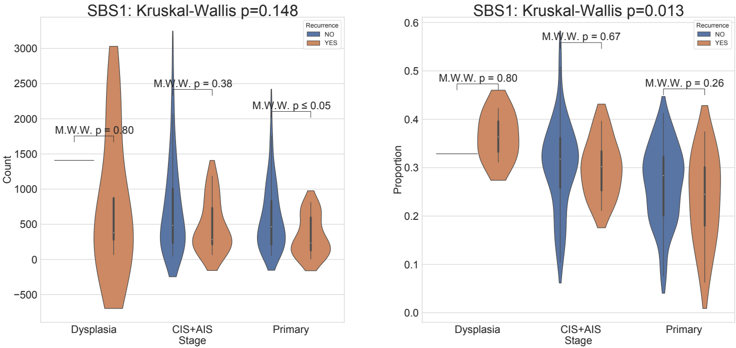



Figure: SBS Bar Plot by Cancer Subtype & Recurrence in LUAD

# SBS in LUAD with Recurrence II




(a) Absolute (b) Relative

Figure: SBS1 Signature in LUAD with Recurrence

# 4. Results

## 4.10. Discovery of Mutational Signature

### 4.10.3. Double Base Substitutions (DBS)

# DBS Signatures I

## DBS2

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

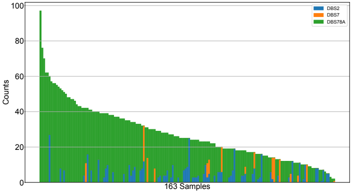
## DBS7

- Defective $\Downarrow$ DNA mismatch repair (Alexandrov et al., 2020)

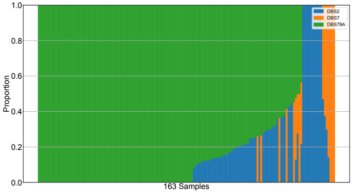
## DBS78A

- Unknown

# DBS in LUSC I



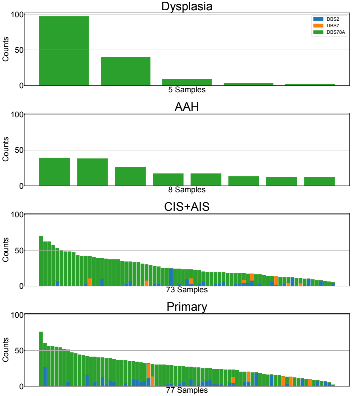

(a) Absolute



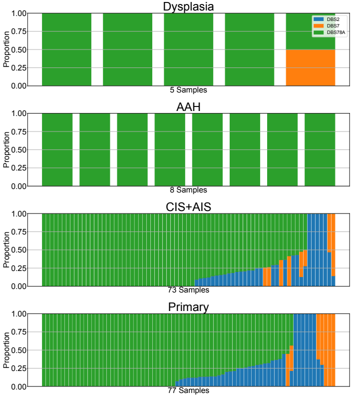

(b) Relative

Figure: DBS Bar Plot in LUSC

# DBS in LUSC II



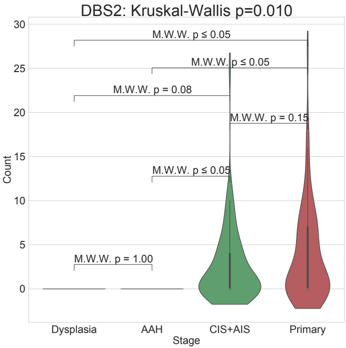

(a) Absolute



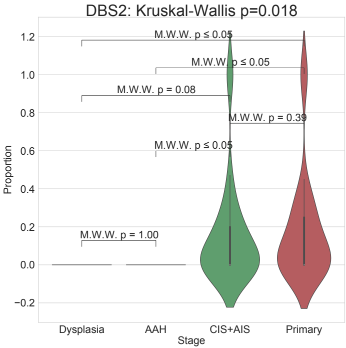

(b) Relative

Figure: DBS Bar Plot by Cancer Subtype in LUSC

# DBS in LUSC III



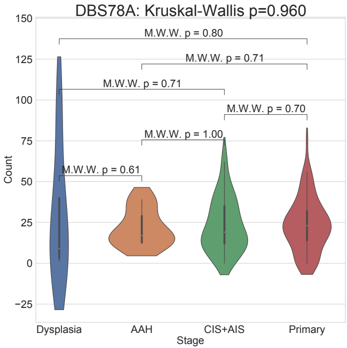

(a) Absolute



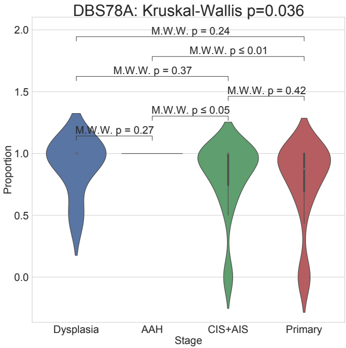

(b) Relative

Figure: DBS2 Signature in LUSC

# DBS in LUSC IV

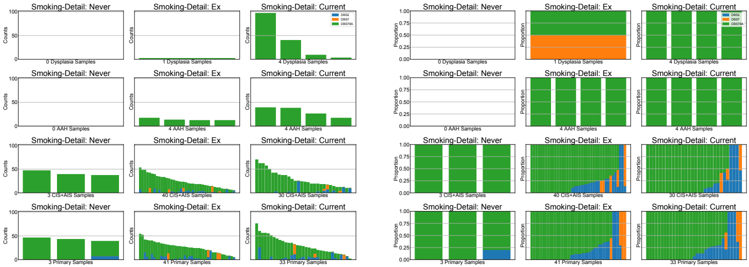



(a) Absolute




(b) Relative

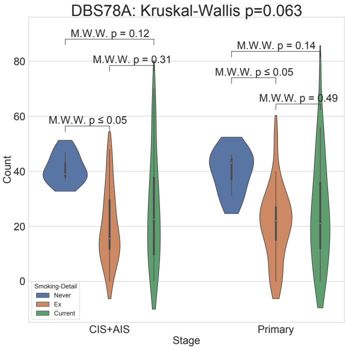
Figure: DBS78A Signature in LUSC



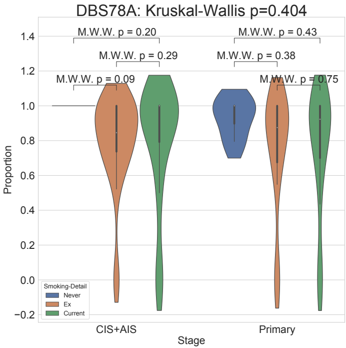

(a) Absolute (b) Relative

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

# DBS in LUSC with Smoking II

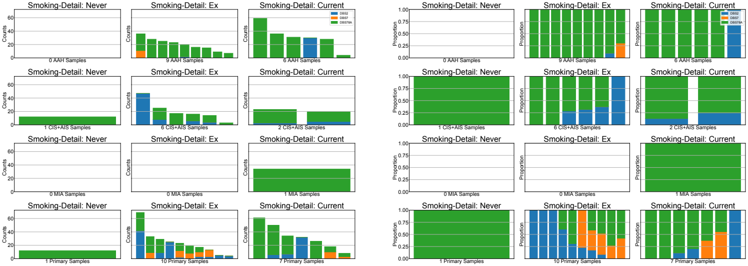



(a) Absolute




(b) Relative

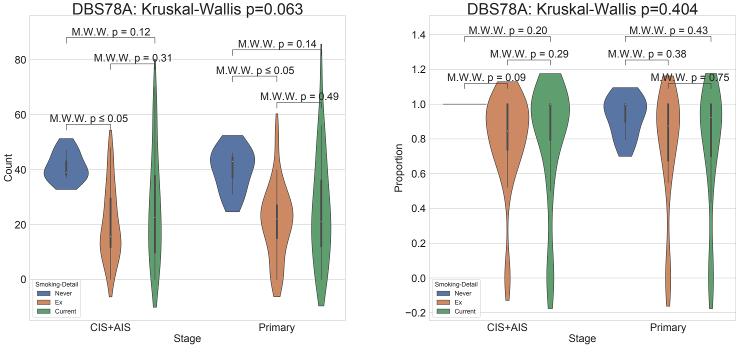
Figure: DBS78A Signature in LUSC with Smoking




(a) Absolute (b) Relative

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

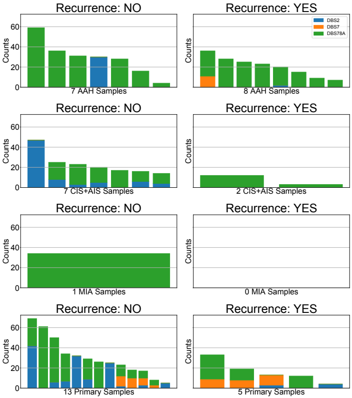
# DBS in LUAD with Smoking II



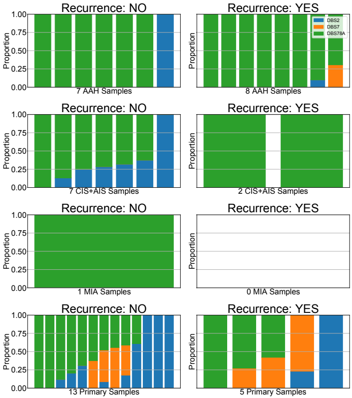

(a) Absolute (b) Relative

Figure: DBS78A Signature in LUSC in Smoking

# DBS in LUAD with Recurrence




(a) Absolute




(b) Relative

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

# 4. Results

## 4.10. Discovery of Mutational Signature

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

# Indel signatures I

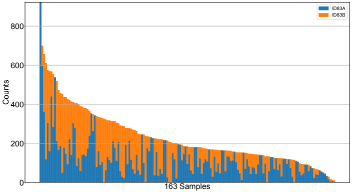
## ID83A

- Unknown

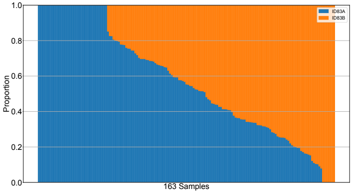
## ID83B

- Unknown

# Indels in LUSC I



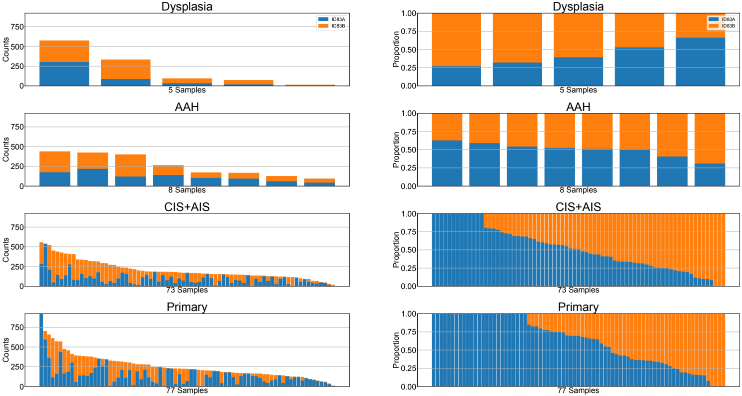

(a) Absolute




(b) Relative

Figure: Indel Bar Plot in LUSC

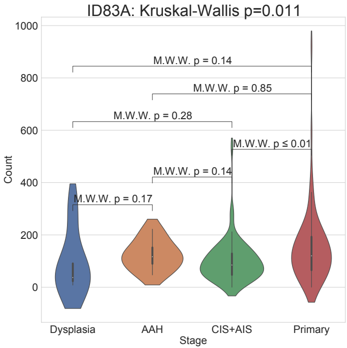
# Indels in LUSC II



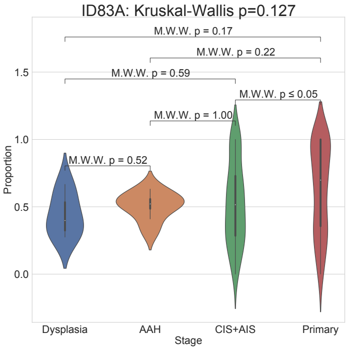

(a) Absolute (b) Relative

Figure: Indel Bar Plot by Cancer Subtype in LUSC

# Indels in LUSC III



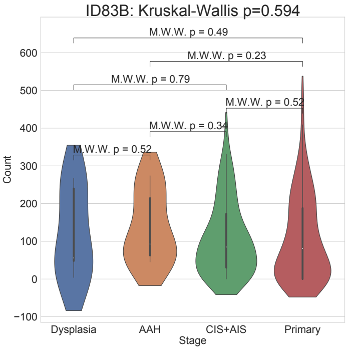

(a) Absolute



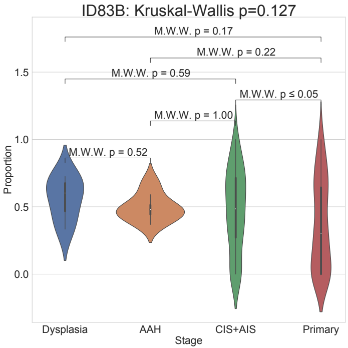

(b) Relative

Figure: Indel83A Signature in LUSC

# Indels in LUSC IV

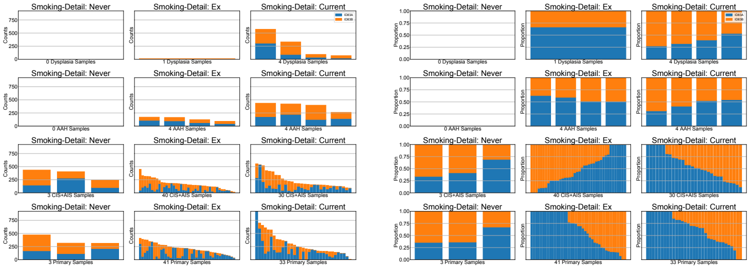



(a) Absolute




(b) Relative

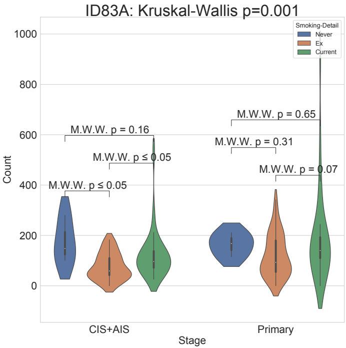
Figure: Indel83B Signature in LUSC



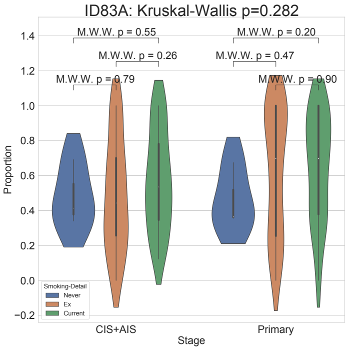

(a) Absolute (b) Relative

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

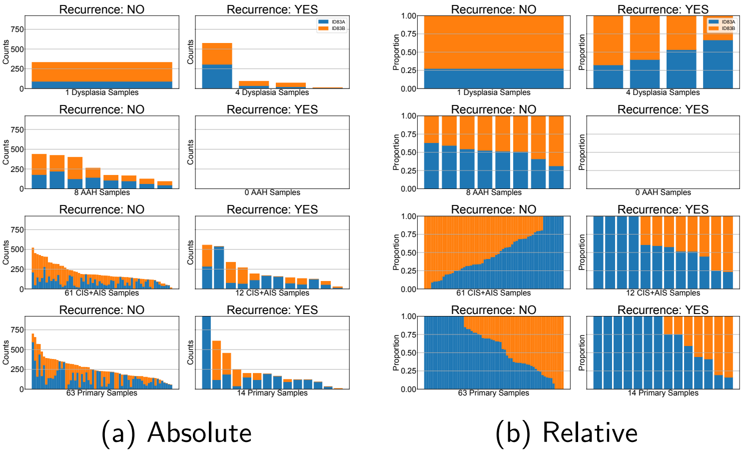
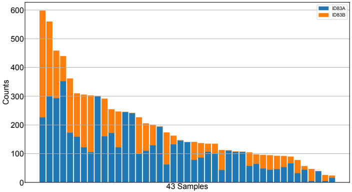
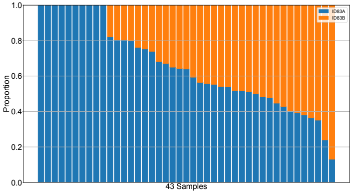



Figure: Indel Bar Plot by Cancer Subtype & Recurrence in LUSC

# Indels in LUAD I

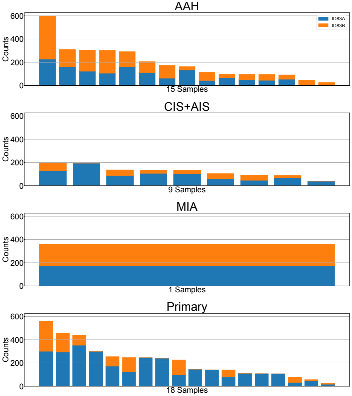



(a) Absolute

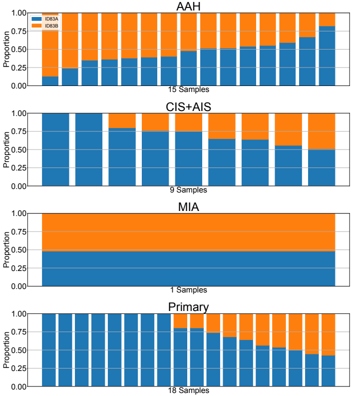



(b) Relative

Figure: Indel Bar Plot in LUAD

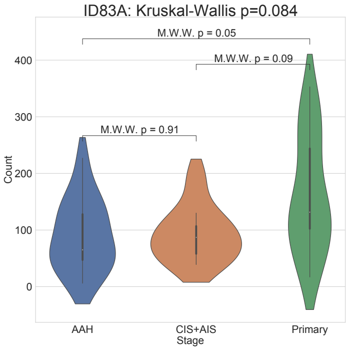



(a) Absolute

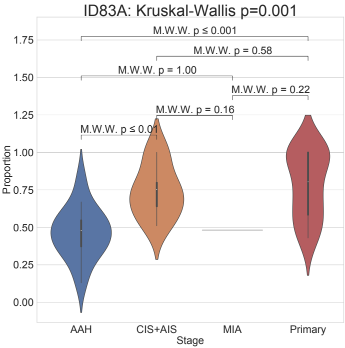



(b) Relative

Figure: Indel Bar Plot by Cancer Subtype in LUAD



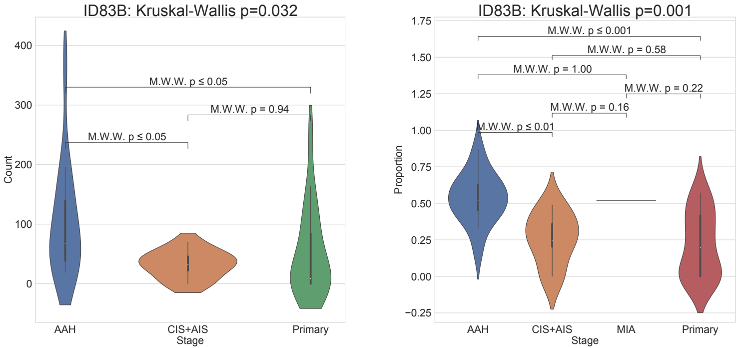

(a) Absolute




(b) Relative

Figure: Indel83A Signature in LUAD

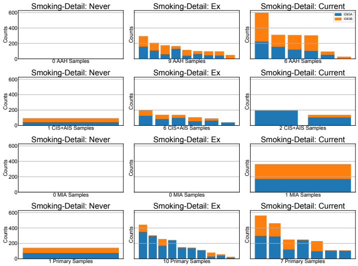
# Indels in LUAD IV



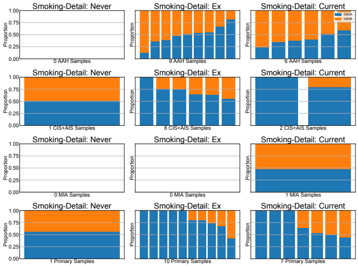

(a) Absolute (b) Relative

Figure: Indel83B Signature in LUAD

# Indel in LUAD with Smoking



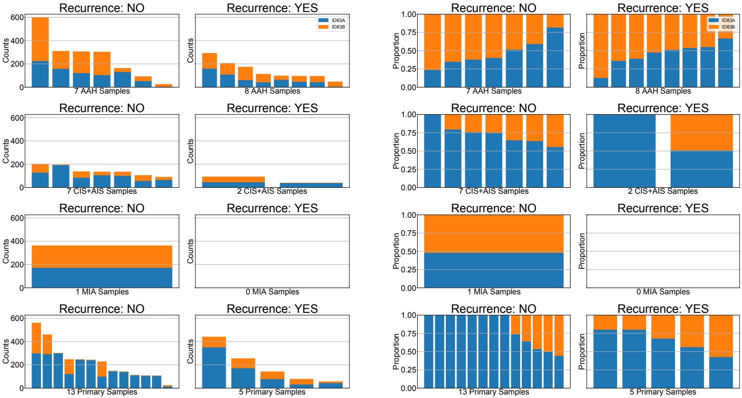

(a) Absolute




(b) Relative

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

# Indel in LUAD with Recurrence




(a) Absolute (b) Relative

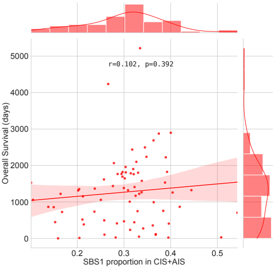
Figure: Indel Bar Plot by Cancer Subtype & Recurrence in LUAD

# 4. Results

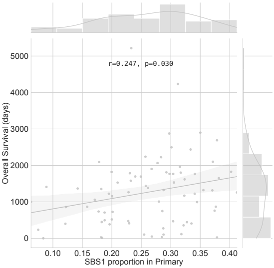
## 4.10. Discovery of Mutational Signature

### 4.10.5. Mutational Signatures vs. Clinical Data

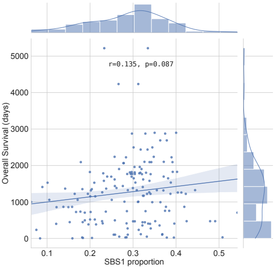
# Mutational Signatures vs. Overall Survival – SBS I




(a) CIS



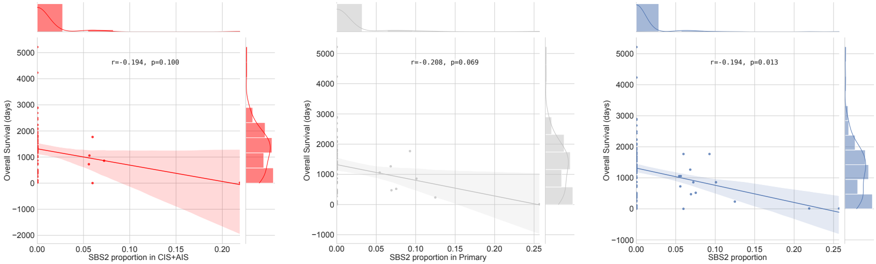

(b) Primary




(c) All

Figure: SBS1 vs. Overall Survival in LUSC

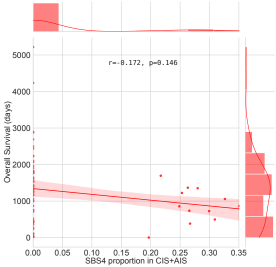
# Mutational Signatures vs. Overall Survival – SBS II



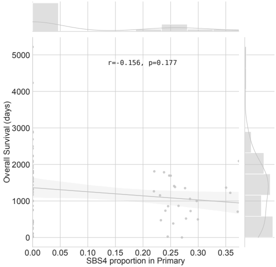

(a) CIS (b) Primary (c) All

Figure: SBS2 vs. Overall Survival in LUSC

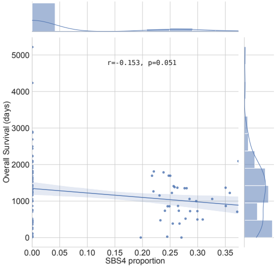
# Mutational Signatures vs. Overall Survival – SBS III




(a) CIS



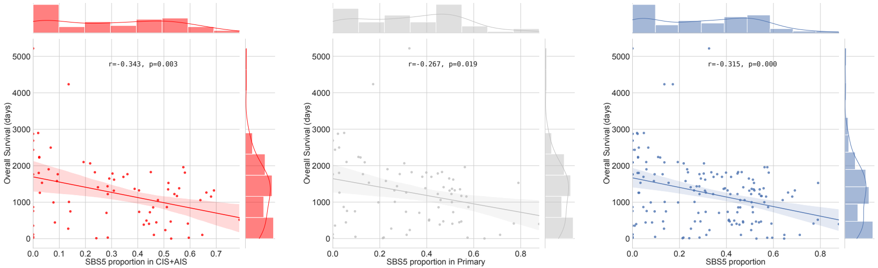

(b) Primary




(c) All

Figure: SBS4 vs. Overall Survival in LUSC

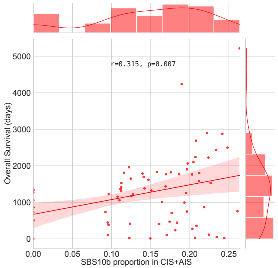
# Mutational Signatures vs. Overall Survival – SBS IV



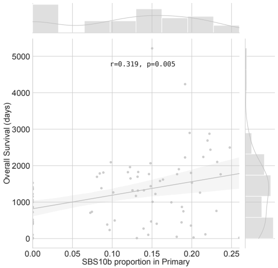

(a) CIS (b) Primary (c) All

Figure: SBS5 vs. Overall Survival in LUSC

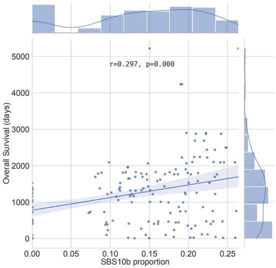
# Mutational Signatures vs. Overall Survival – SBS V




(a) CIS

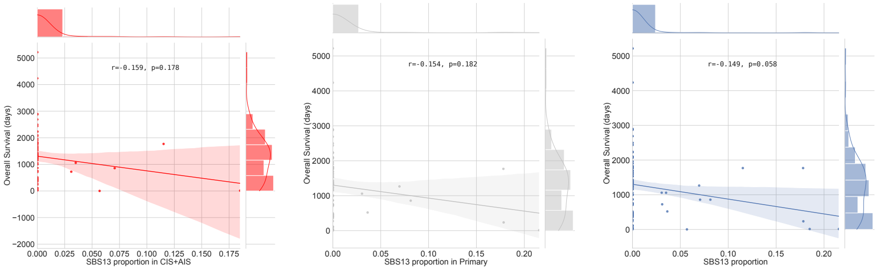



(b) Primary




(c) All

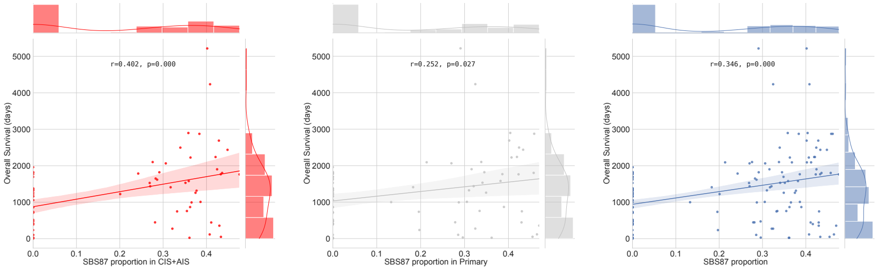
Figure: SBS10b vs. Overall Survival in LUSC




(a) CIS (b) Primary (c) All

Figure: SBS13 vs. Overall Survival in LUSC

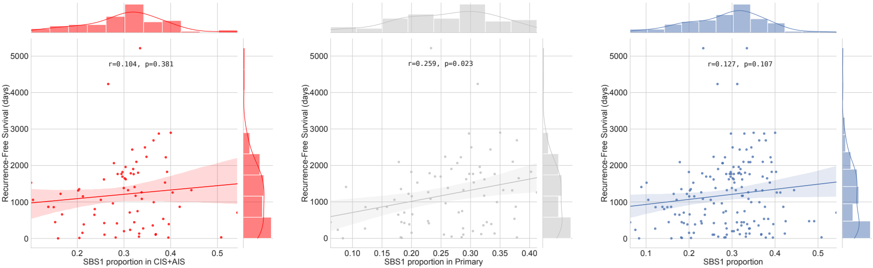
# Mutational Signatures vs. Overall Survival – SBS VII




(a) CIS (b) Primary (c) All

Figure: SBS87 vs. Overall Survival in LUSC

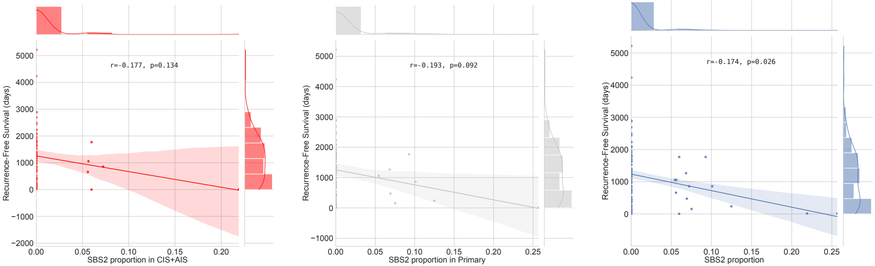
# Mutational Signatures vs. Recurrence-Free Survival I




(a) CIS (b) Primary (c) All

Figure: SBS1 vs. Recurrence-Free Survival in LUSC

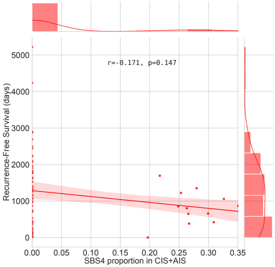
# Mutational Signatures vs. Recurrence-Free Survival II



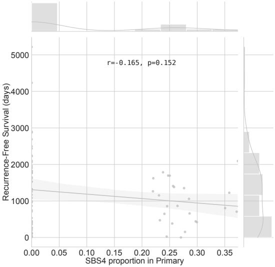

(a) CIS (b) Primary (c) All

Figure: SBS2 vs. Recurrence-Free Survival in LUSC

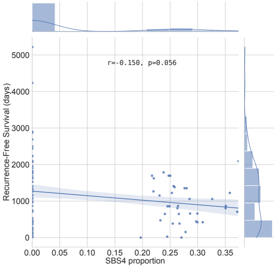
# Mutational Signatures vs. Recurrence-Free Survival III




(a) CIS

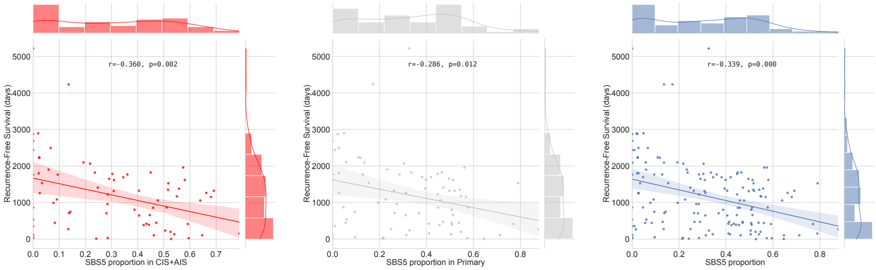



(b) Primary



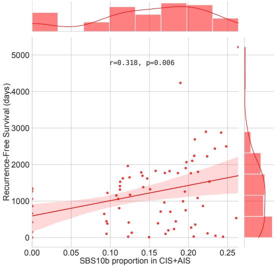

(c) All

Figure: SBS4 vs. Recurrence-Free Survival in LUSC

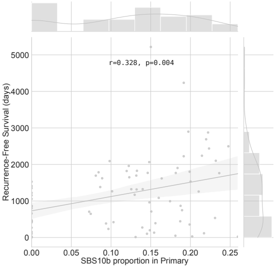



(a) CIS (b) Primary (c) All

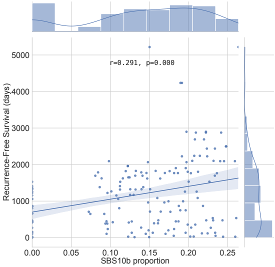
Figure: SBS5 vs. Recurrence-Free Survival in LUSC




(a) CIS



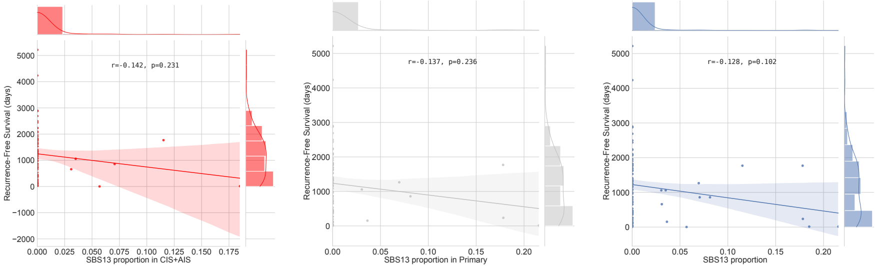

(b) Primary




(c) All

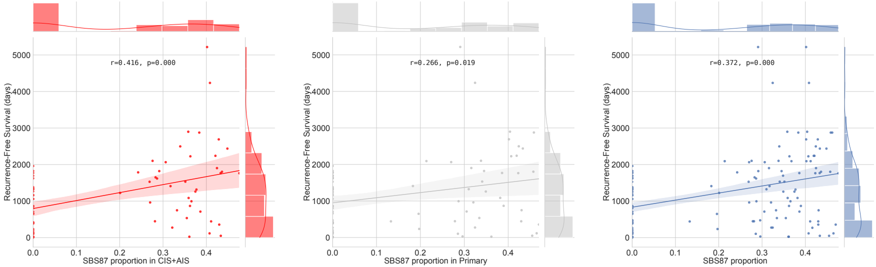
Figure: SBS10b vs. Recurrence-Free Survival in LUSC

# Mutational Signatures vs. Recurrence-Free Survival VI



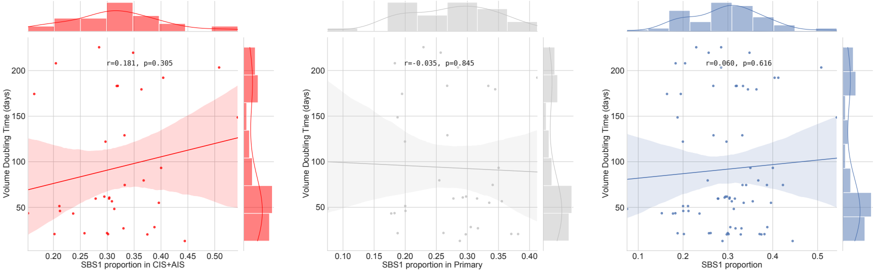

(a) CIS (b) Primary (c) All

Figure: SBS13 vs. Recurrence-Free Survival in LUSC




(a) CIS (b) Primary (c) All

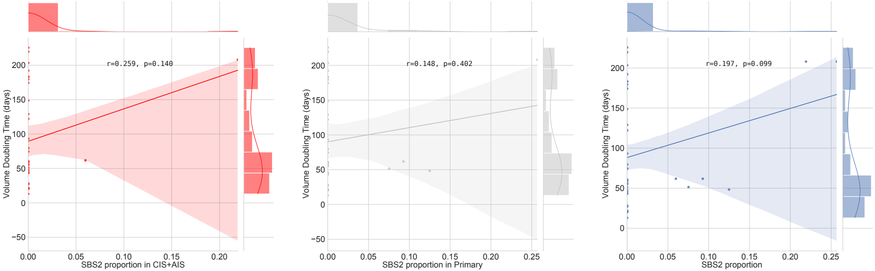
Figure: SBS87 vs. Recurrence-Free Survival in LUSC




(a) CIS (b) Primary (c) All

Figure: SBS1 vs. Volume Doubling Time in LUSC

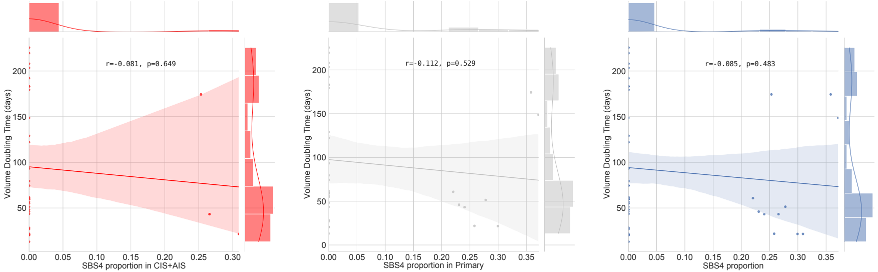
# Mutational Signatures vs. Volume Doubling Time II




(a) CIS (b) Primary (c) All

Figure: SBS2 vs. Volume Doubling Time in LUSC

# Mutational Signatures vs. Volume Doubling Time III




(a) CIS (b) Primary (c) All

Figure: SBS4 vs. Volume Doubling Time in LUSC

# Mutational Signatures vs. Volume Doubling Time IV

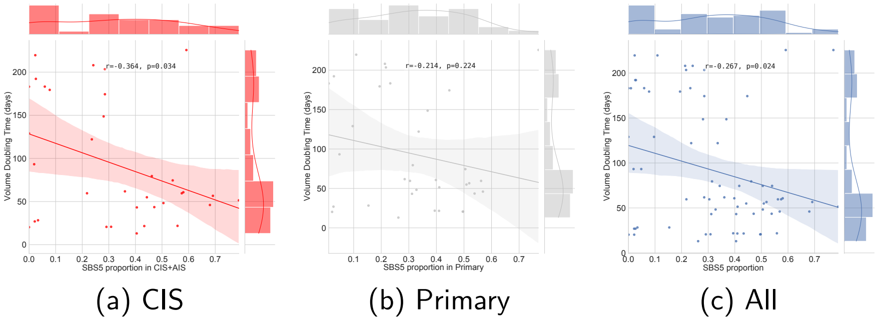
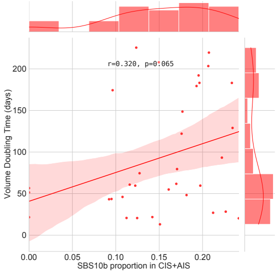
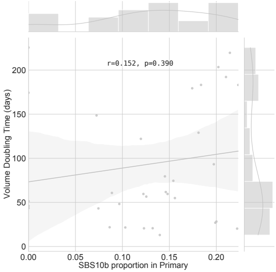


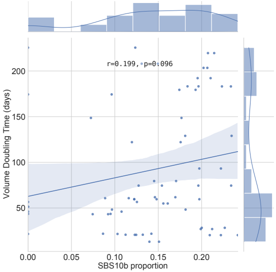

Figure: SBS5 vs. Volume Doubling Time in LUSC




(a) CIS

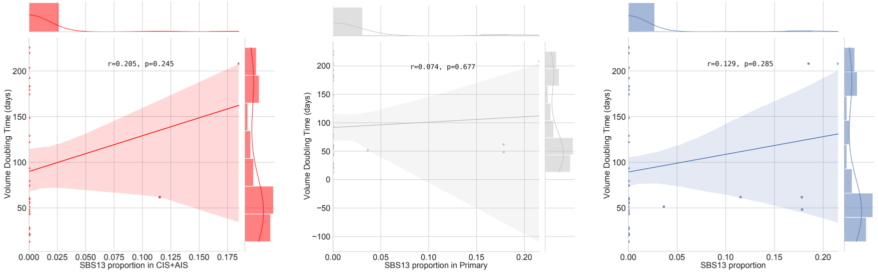



(b) Primary



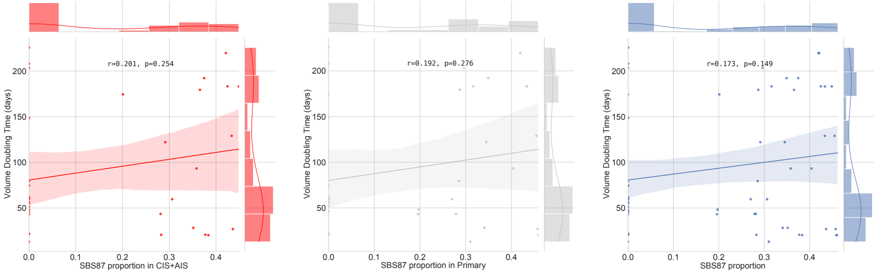

(c) All

Figure: SBS10b vs. Volume Doubling Time in LUSC




(a) CIS (b) Primary (c) All

Figure: SBS13 vs. Volume Doubling Time in LUSC




(a) CIS (b) Primary (c) All

Figure: SBS87 vs. Volume Doubling Time in LUSC

# Findings in Mutation Signature

# 4. Results

## 4.11. Clinical Data with Point Mutation

# Mutect2?

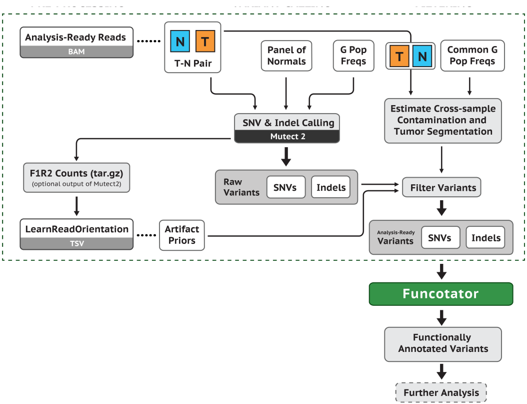



Figure: Somatic short variant discovery workflow (Van der Auwera et al., 2013; DePristo et al., 2011)

# 4. Results

## 4.11. Clinical Data with Point Mutation

### 4.11.1. For Smoking

# LUSC with Smoking

Table: LUSC WES Data with Smoking

| Smoking? | Stage | Number of Samples |
|---|---|---|
| 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

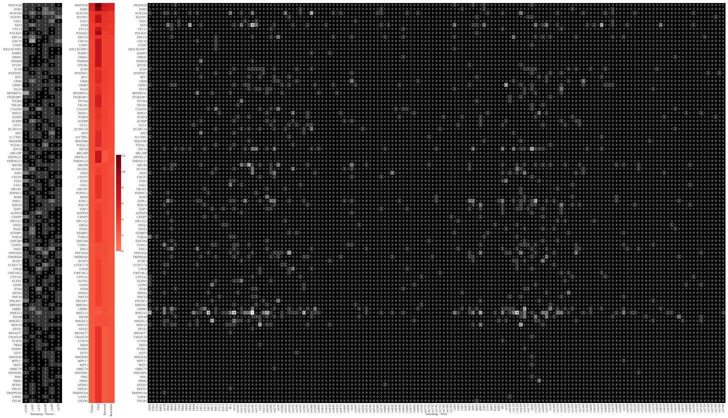
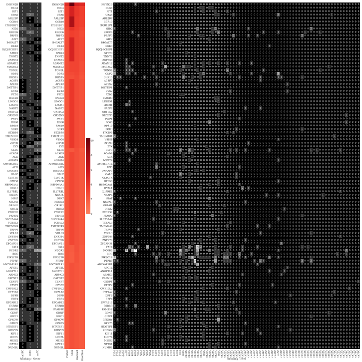
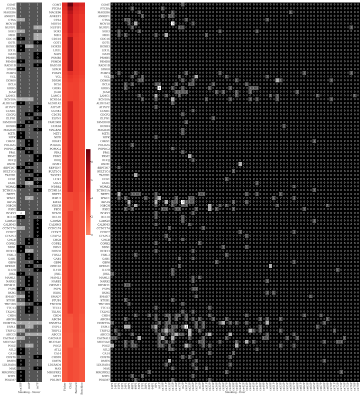


Figure: Clinical Data about LUSC for Smoking

# Clinical Data about LUSC for Smoking II



(a) Precancer



(b) Primary Cancer

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.
2. COMT catalyzes the O-methylation, and inactivates of neurotransmitters and hormones (Dawling, Roodi, Mernaugh, Wang, & Parl, 2001; J. Chen et al., 2011).

# 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

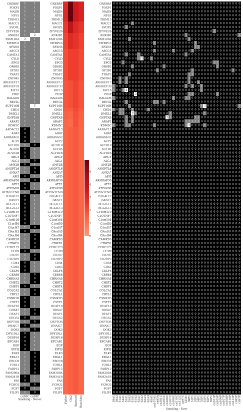
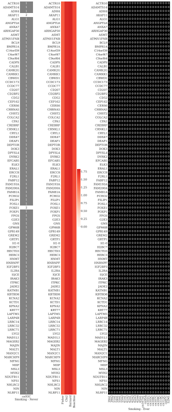
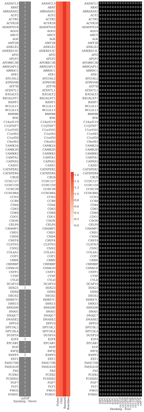


Figure: Clinical Data about LUAD for Smoking

# 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

## 4.11. Clinical Data with Point Mutation

### 4.11.2. For Recurrence

# LUSC with Recurrence

Table: LUSC WES Data with Recurrence

| Recurrence? | Stage | Number of Samples |
|---|---|---|
| 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

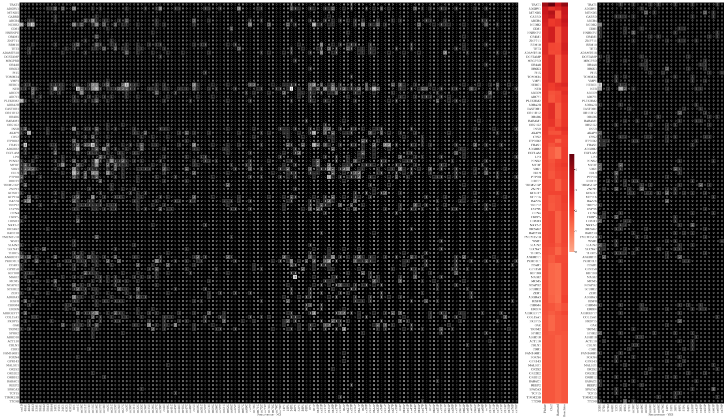
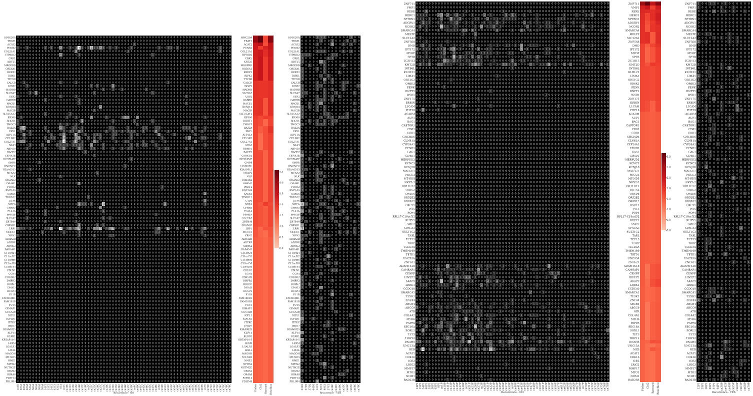


Figure: Clinical Data about LUSC for Recurrence

# Clinical Data about LUSC for Recurrence II



(a) Precancer (b) Primary Cancer

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 |
|---|---|---|
| Recurrence | Normal | 5 |
| | AAH | 8 |
| | CIS+AIS | 2 |
| | Primary | 5 |
| | Total | 20 |
| Non-recurrence | Normal | 13 |
| | AAH | 7 |
| | CIS+AIS | 7 |
| | MIA | 1 |
| | Primary | 13 |
| | Total | 41 |

# Clinical Data about LUAD for Recurrence I

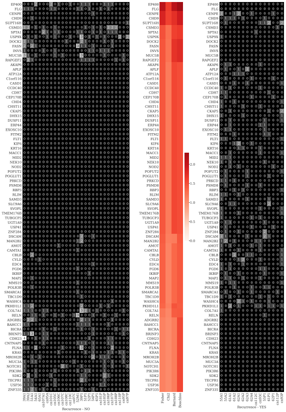
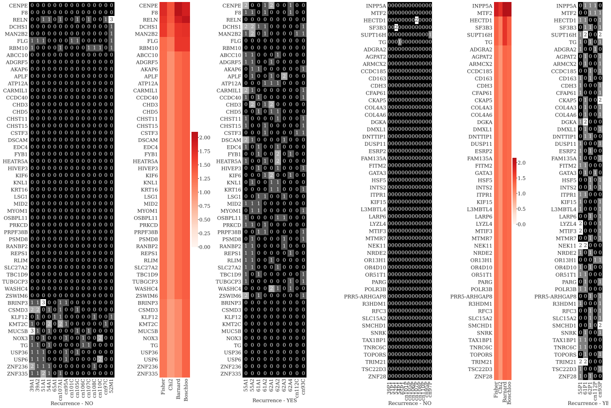



Figure: Clinical Data about LUAD for Recurrence

# 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

# 4. Results

## 4.12. Differences in Gene Expression Levels

# RSEM?

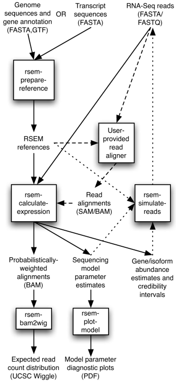



Figure: RSEM workflow (B. Li & Dewey, 2011)

# DESeq2?

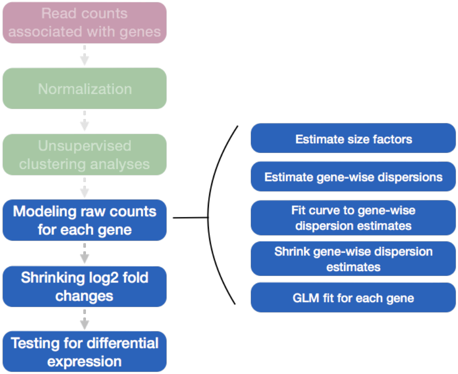



Figure: DESeq2 workflow (Love, Huber, & Anders, 2014)

# DEG Selection Strategy

DEG: differentially expressed genes

## Fold Change

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

## P-value

P-value $< 0.05$

## Adjusted P-value

Padj $< 0.05$

# Enrichr?

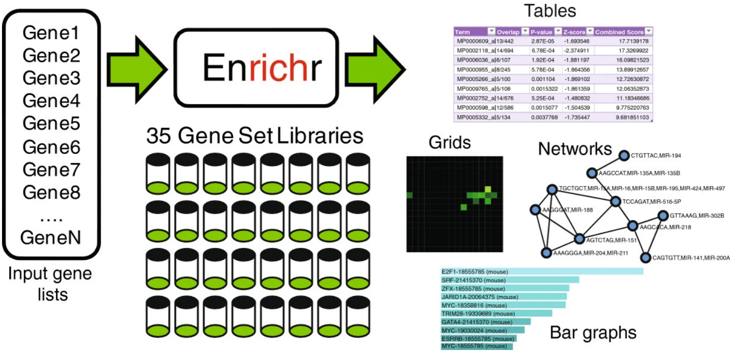



Figure: Enrichr workflow (E. Y. Chen et al., 2013; Kuleshov et al., 2016)

# Gene-set Library

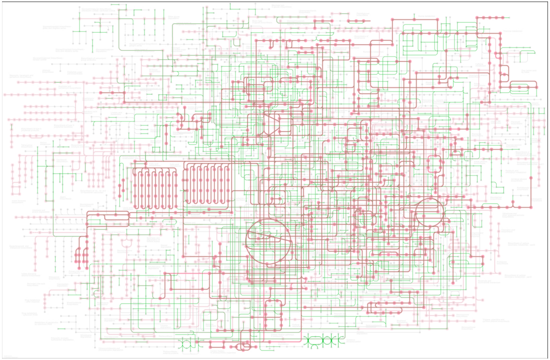


Figure: The global map of metabolic pathways by KEGG (Kanehisa et al., 2021)

## KEGG

KEGG 2021 Human

# WTS Data Composition

Table: Number of WTS samples

| Cancer Subtype | Stage | Number of Samples |
|---|---|---|
| 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 I

Table: Number of WTS LUSC samples

| Recurrence? | Stage | Number of Samples |
|---|---|---|
| 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 |
|---|---|---|
| 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 |
|---|---|---|
| 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 |

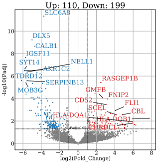
# 4. Results

## 4.12. Differences in Gene Expression Levels

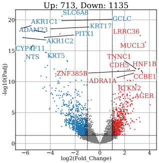
### 4.12.1. Comparing cancer stage in LUSC

# DEG Volcano Plots in LUSC

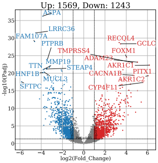
Normal → Dysplasia → CIS → Primary (LUSC)



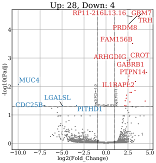

(a) Normal-Dysplasia



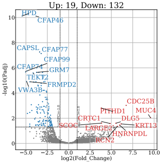

(b) Normal-CIS



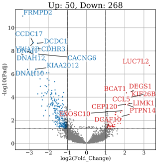

(c) Normal-Primary




(d) Dysplasia-CIS




(e) Dysplasia-Primary

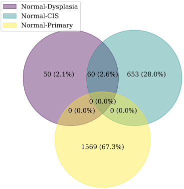



(f) CIS-Primary

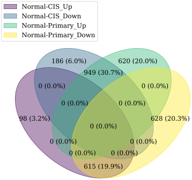
Figure: DEG Volcano Plots in LUSC

# DEG Venn Diagram in LUSC

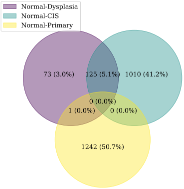
Normal $\rightarrow$ Dysplasia $\rightarrow$ CIS $\rightarrow$ Primary (LUSC)




(a) Up-regulated




(b) Both




(c) Down-regulated

Figure: DEG Venn Diagram in LUSC

# Enrichment test with Normal vs. Dysplasia in LUSC

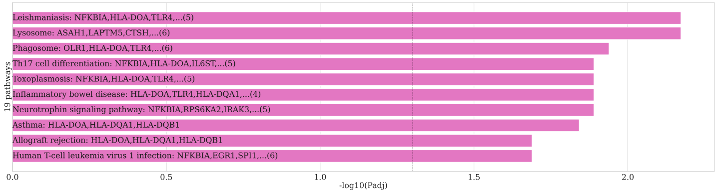



Figure: Up-regulated Pathways on Normal vs. Dysplasia

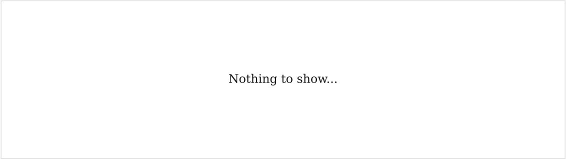



Figure: Down-regulated Pathways on Normal vs. Dysplasia

# Enrichment test with Normal vs. CIS in LUSC

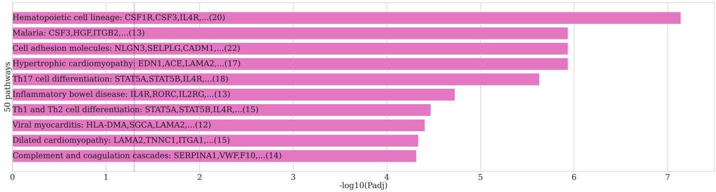



Figure: Up-regulated Pathways on Normal vs. CIS

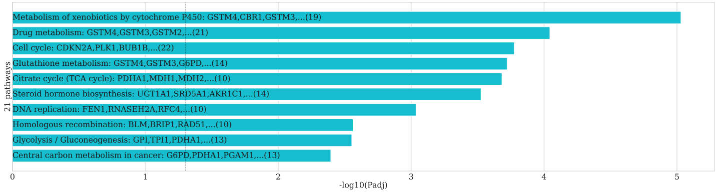



Figure: Down-regulated Pathways on Normal vs. CIS

# Enrichment test with Normal vs. Primary in LUSC

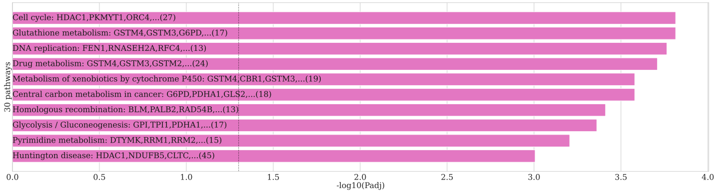



Figure: Up-regulated Pathways on Normal vs. Primary

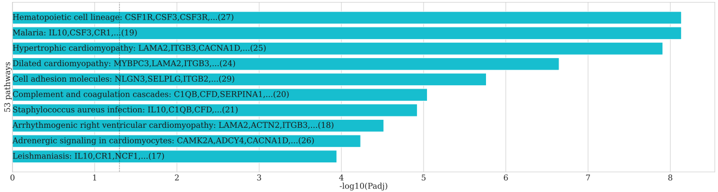



Figure: Down-regulated Pathways on Normal vs. Primary

# 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 (B. Li et al., 2019).
4. Associated with lung disease in adult (Henderson et al., 2013) and baby (Brasch et al., 2004).

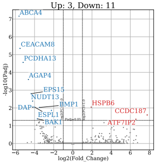
# 4. Results

## 4.12. Differences in Gene Expression Levels

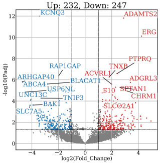
### 4.12.2. Comparing cancer stage in LUAD

# DEG Volcano Plots in LUAD

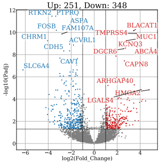
Normal → AAH → AIS → Primary (LUAD)



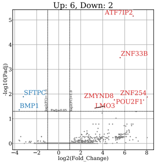

(a) Normal-AAH



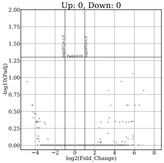

(b) Normal-AIS



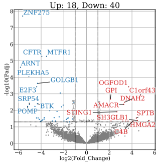

(c) Normal-Primary




(d) AAH-AIS




(e) AAH-Primary

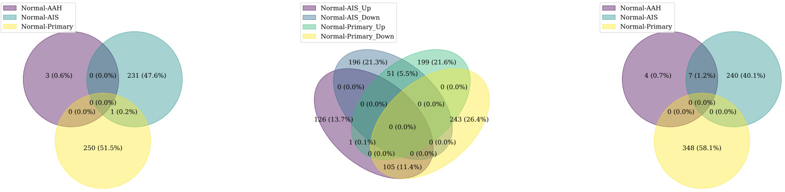



(f) AIS-Primary

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

Nothing to show...

Figure: Up-regulated Pathways on Normal vs. AAH

Nothing to show...

Figure: Down-regulated Pathways on Normal vs. AAH

# Enrichment test with Normal vs. AIS in LUAD

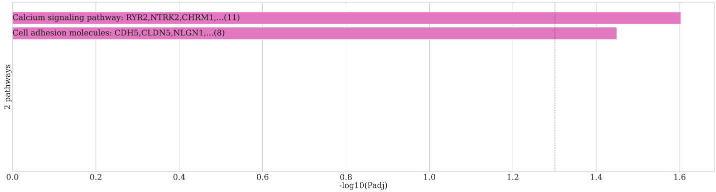



Figure: Up-regulated Pathways on Normal vs. AIS

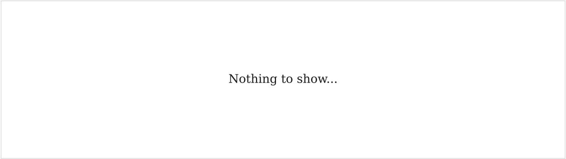



Figure: Down-regulated Pathways on Normal vs. AIS

# Enrichment test with Normal vs. Primary in LUAD

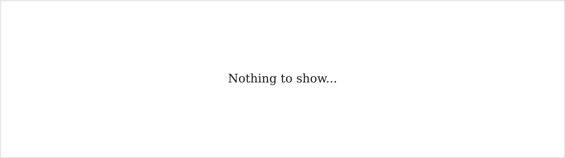



Figure: Up-regulated Pathways on Normal vs. Primary

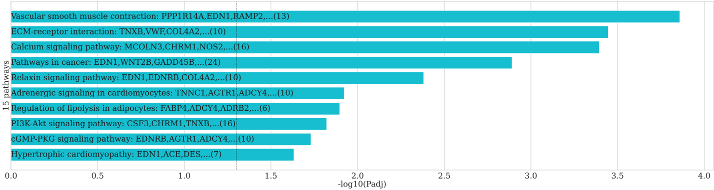



Figure: Down-regulated Pathways on Normal vs. Primary

# 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 ⇒ 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).

# Finding in Comparing cancer stage in LUAD III

## CHRM1

1. Up-regulated in AIS, but down-regulated in Primary.
2. Various cellular responses $\Rightarrow$ 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 (G. Ma et al., 2019).

# 4. Results

## 4.12. Differences in Gene Expression Levels

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

# LUSC Data Composition

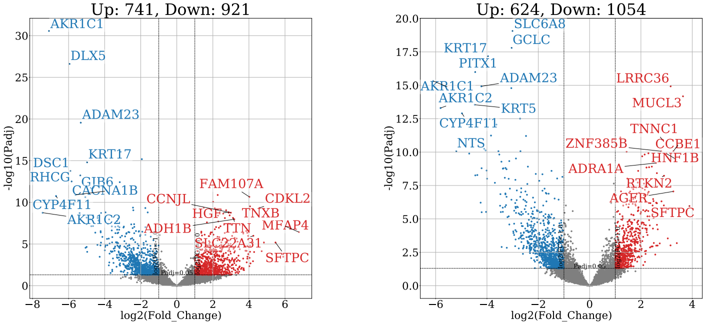
Table: Number of WTS LUSC samples

| Recurrence? | Stage | Number of Samples |
|---|---|---|
| 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

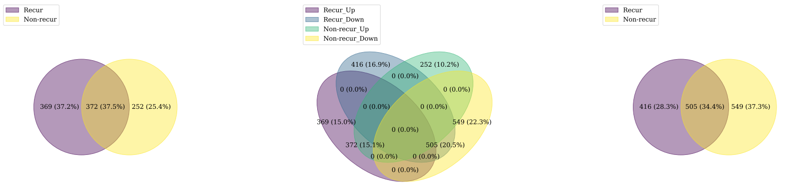
In order to compare with Normal stage, merging Normal samples.
$\because$ Insufficient number of Normal samples in Recur.

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




(a) Recur (b) Non-recur

Figure: DEG Volcanot Plot with CIS in LUSC




(a) Up-regulated (b) Both (c) Down-regulated

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

# Enrichment test for Recur-specific with CIS in LUSC

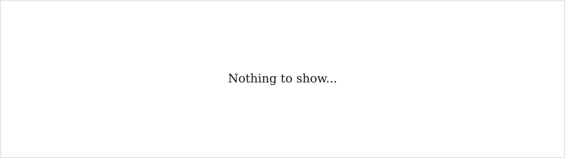



Figure: Up-regulated Pathways for Recur-specific

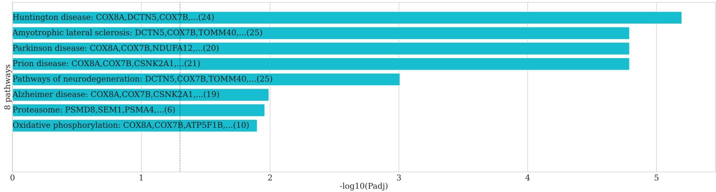



Figure: Down-regulated Pathways for Recur-specific

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

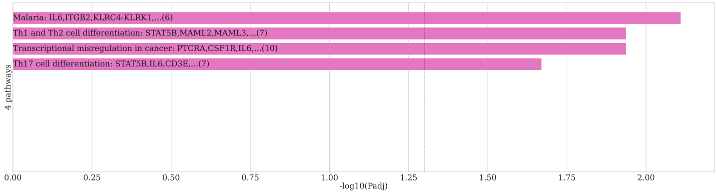



Figure: Up-regulated Pathways for Non-recur-specific

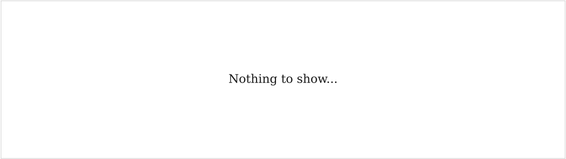



Figure: Down-regulated Pathways for Non-recur-specific

# Enrichment test for Intersected with CIS in LUSC

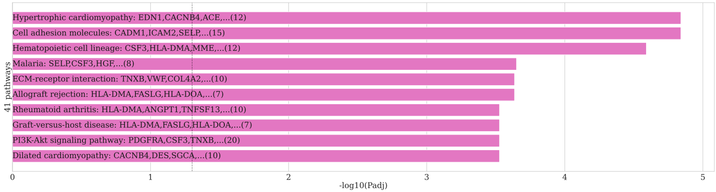



Figure: Up-regulated Pathways for Intersected

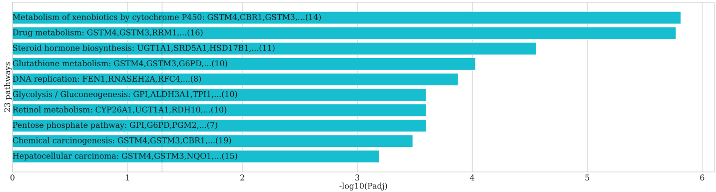
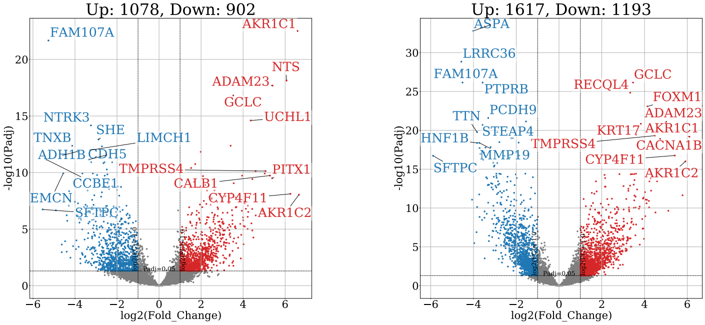



Figure: Down-regulated Pathways for Intersected

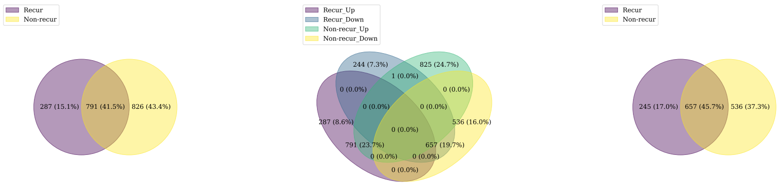
# DEG Volcano Plots for R vs. NR with Primary in LUSC




(a) Recur (b) Non-recur

Figure: DEG Volcanot Plot with Primary in LUSC

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




(a) Up-regulated (b) Both (c) Down-regulated

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

# Enrichment test for Recur-specific with Primary in LUSC

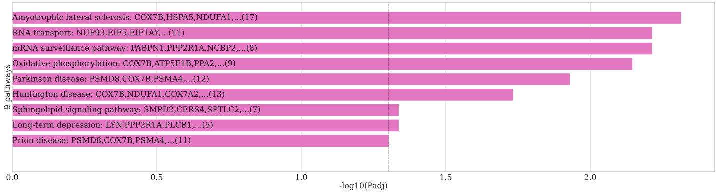



Figure: Up-regulated Pathways for Recur-specific

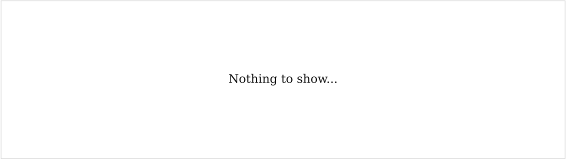



Figure: Down-regulated Pathways for Recur-specific

# Enrichment test for NR-specific with Primary in LUSC

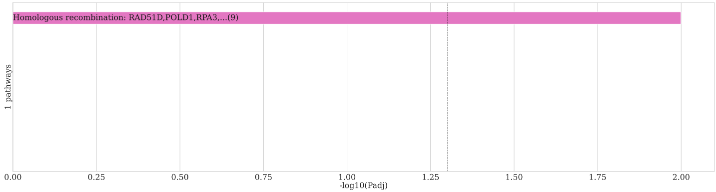



Figure: Up-regulated Pathways for Non-recur-specific

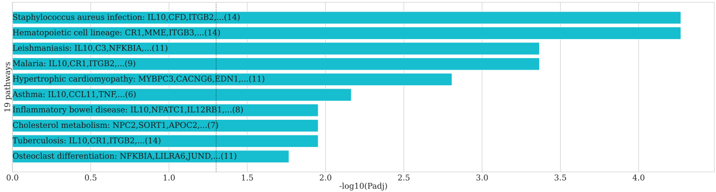



Figure: Down-regulated Pathways for Non-recur-specific

# Enrichment test for Intersected with Primary in LUSC

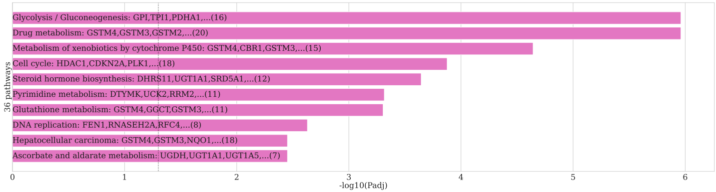



Figure: Up-regulated Pathways for Intersected

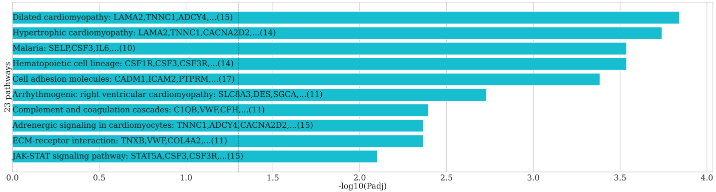



Figure: Down-regulated Pathways for Intersected

# 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

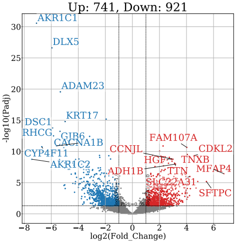
1. Highly up-regulated in Non-recur patients.
2. DNA-dependent ATPase (Yin, Kwon, Varshavsky, & Wang, 2004)
3. RECQL4 modulate chromosome segregation (Yin et al., 2004)
4. RECQL5 promotes metastasis & resistance in NSCLC (Xia, Zhang, Yuan, & Niu, 2021)
5. RECQL4 $\Uparrow \Rightarrow$ Poor prognosis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

# 4. Results

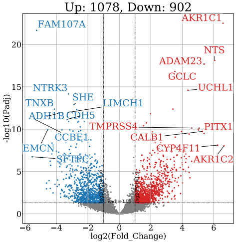
## 4.12. Differences in Gene Expression Levels

### 4.12.4. Within Recur in LUSC

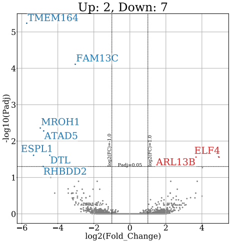
# DEG Volcano Plots with Recur in LUSC




(a) Normal-CIS



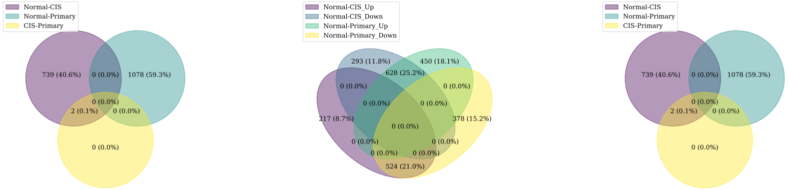

(b) Normal-Primary




(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

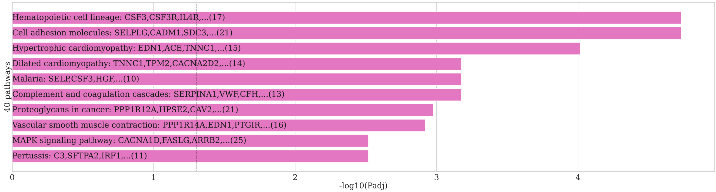



Figure: Up-regulated Pathways on Normal vs. CIS for Recur in LUSC

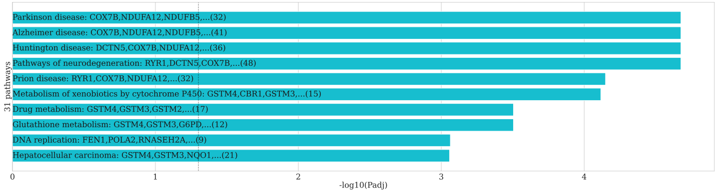



Figure: Down-regulated Pathways on Normal vs. CIS for Recur in LUSC

# Enrichment test with Normal vs. Primary for Recur

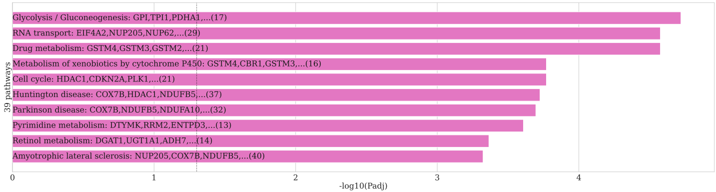



Figure: Up-regulated Pathways on Normal vs. Primary for Recur in LUSC

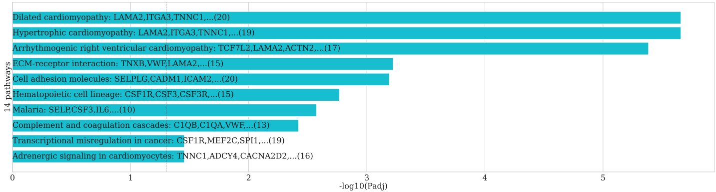



Figure: Down-regulated Pathways on Normal vs. Primary for Recur in LUSC

# 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-Otın, 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)

# Finding in Comparing within Recur in LUSC II

## FAM107A

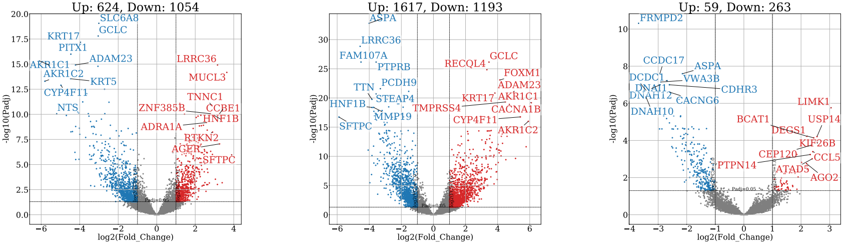
1. Up-regulated in CIS, but down-regulated in Primary.
2. May play a role in tumor development (L. Wang et al., 2000)
3. Negatively regulates focal adhesion assembly (Le et al., 2010)

# 4. Results

## 4.12. Differences in Gene Expression Levels

### 4.12.5. Within Non-recur in LUSC

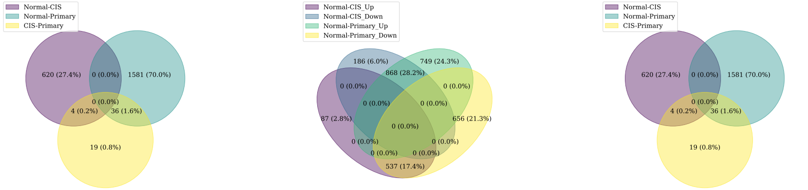
# DEG Volcano Plots with Non-recur in LUSC




(a) Normal-CIS (b) Normal-Primary (c) CIS-Primary

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

# DEG Venn Diagram with Non-recur in LUSC




(a) Up-regulated (b) Both (c) Down-regulated

Figure: DEG Venn Diagram with Non-recur in LUSC

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

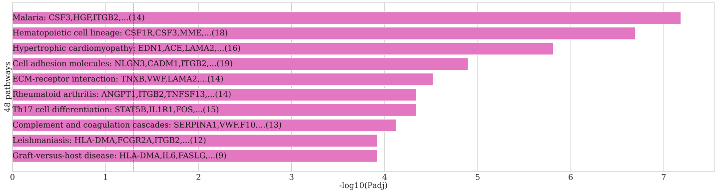



Figure: Up-regulated Pathways on Normal vs. CIS for Non-recur in LUSC

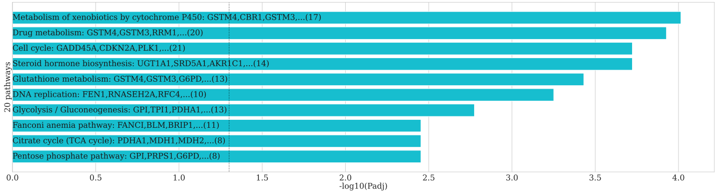



Figure: Down-regulated Pathways on Normal vs. CIS for Non-recur in LUSC

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

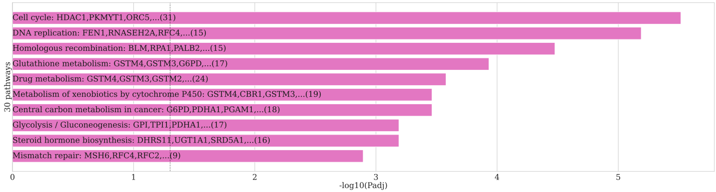



Figure: Up-regulated Pathways on Normal vs. Primary for Non-recur in LUSC

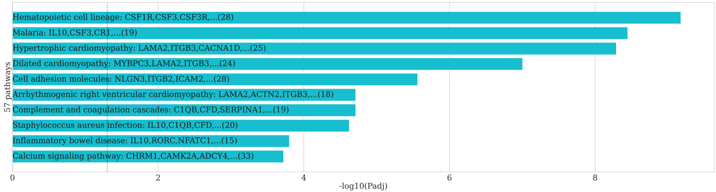



Figure: Down-regulated Pathways on Normal vs. Primary for Non-recur in LUSC

# 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).

# Finding in Comparing within Non-recur in LUSC II

## 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

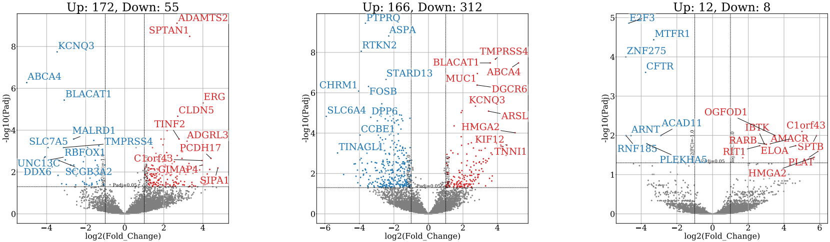
## 4.12. Differences in Gene Expression Levels

### 4.12.6. Within Non-recur in LUAD

# LUAD Data Composition

Table: Number of WTS LUAD samples

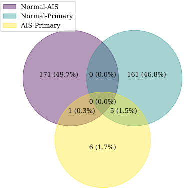
| Recurrence? | Stage | Number of Samples |
|---|---|---|
| Recurrence | Normal | 2 |
| | CIS+AIS | 1 |
| | Primary | 1 |
| | Total | 4 |
| Non-recurrence | Normal | 11 |
| | AAH | 1 |
| | CIS+AIS | 4 |
| | Primary | 5 |
| | Total | 21 |



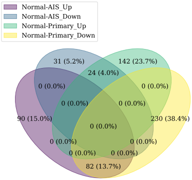

(a) Normal-AIS (b) Normal-Primary (c) AIS-Primary

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

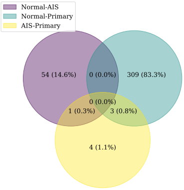
# DEG Venn Diagram with Non-recur in LUAD




(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

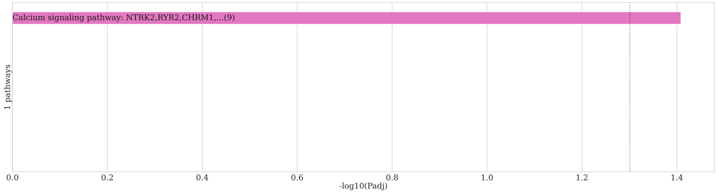



Figure: Up-regulated Pathways on Normal vs. AIS for Non-recur in LUA

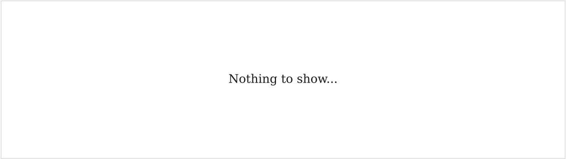



Figure: Down-regulated Pathways on Normal vs. AIS for Non-recur in LUAD

# Enrichment test with Normal vs. Primary in LUAD

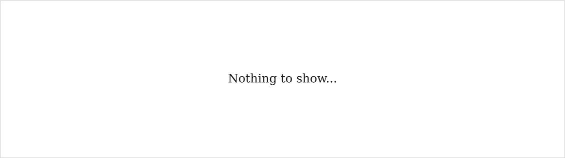



Figure: Up-regulated Pathways on Normal vs. Primary for Non-recur in LUA

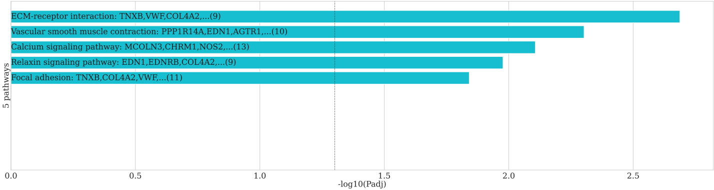



Figure: Down-regulated Pathways on Normal vs. Primary for Non-recur in LUAD

# Finding in Comparing within Non-recur in LUAD I

## 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)

# Finding in Comparing within Non-recur in LUAD II

## 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

## 4.13. Mutation Shared Proportion

# Mutation Shared Proportion?

## Mutation Shared Proportion

$$\text{Mutation Shared Proportion} = (\text{Precancer} \cap \text{Primary})/\text{Primary}$$

## Selection Strategy

1. Non-synonymous: point mutation, frame-shift, splice site, etc.
2. Same site: chromosome, start position, & end position
3. Exact reference allele
4. Exact tumor allele

# 4. Results

## 4.13. Mutation Shared Proportion

### 4.13.1. BWA

# Mutation Shared Proportion Distribution

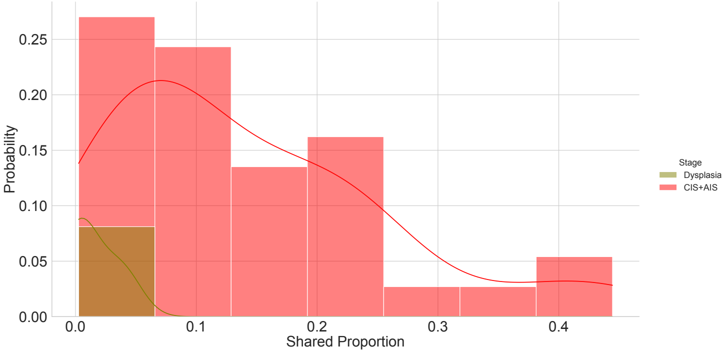
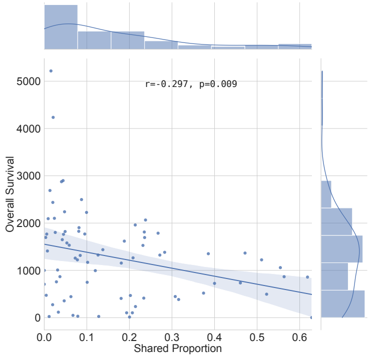
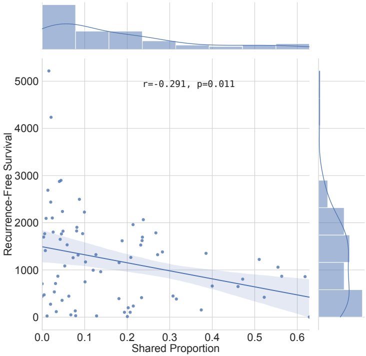



Figure: Mutation Shared Proportion Distribution in LUSC

# Mutation Shared Proportion with Clinical Data I



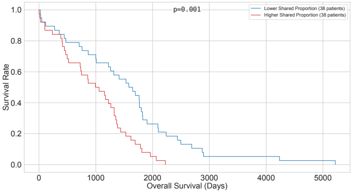

(a) Overall Survival



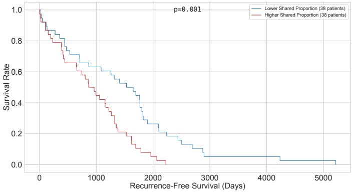

(b) Recurrence-Free Survival

Figure: Mutation Shared Proportion with Clinical Data from LUSC

# Mutation Shared Proportion with Clinical Data II



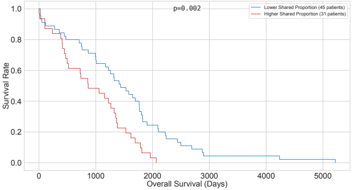

(a) Overall Survival



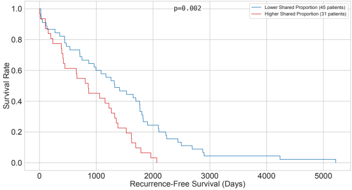

(b) Recurrence-Free Survival

Figure: K-M Survivial Plot with Median separation from LUSC

# Mutation Shared Proportion with Clinical Data III




(a) Overall Survival




(b) Recurrence-Free Survival

Figure: K-M Survivial Plot with Mean separation from LUSC

# Mutation Shared Genes

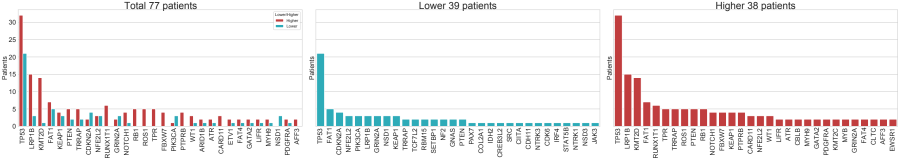



Figure: Mutation Shared Genes with Median separation from LUSC

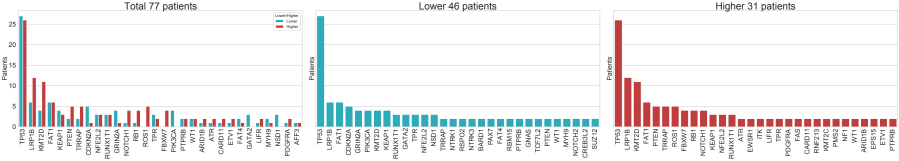



Figure: Mutation Shared Genes with Mean separation from LUSC

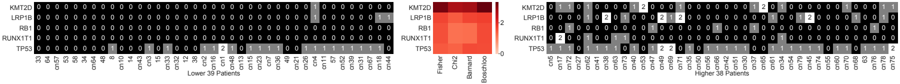



Figure: Exact Test on Mutation Shared Genes with Median separation from LUSC

## Notable genes

1. KMT2D
2. LRP1B
3. RB1
4. RUNX1T1
5. TP53

# Mutation Shared Genes – Exact test II

## KMT2D

1. Lysine Methyltransferase 2D
2. KMT2D have a role in chromatin remodeling
   ⇒ transcription & DNA repair (Shinsky, Monteith, Viggiano, & Cosgrove, 2015; Cho et al., 2007)
3. KMT2D deficiency impairs super-enhancers
   ∴ To confer a glycolytic vulnerability (Alam et al., 2020)
4. KMT2D mutation is associated with poor prognosis in NSCLC (Ardeshir-Larijani et al., 2018)

# Mutation Shared Genes – Exact test III

## LRP1B

1. LDL Receptor Related Protein 1B
2. LRP1B mutation is associated with favorable outcome to ICB across multiple cancer types (Brown et al., 2021)
3. LRP1B mutation is associated with tumor mutational burden in Lung cancer (Lan et al., 2019), especially NSCLC (H. Chen et al., 2019)
4. Higher prevalence of LRP1B mutation in LUAD with COPD (D. Xiao et al., 2017)

# Mutation Shared Genes – Exact test IV

## RB1

1. Retinoblastoma Transcriptional Co-repressor 1
2. Tumor suppressor $\Leftarrow$ G1/S transition of cell cycle (Harbour, Luo, Dei Santi, Postigo, & Dean, 1999)
3. Common RB1 re-arrangements associated with histopathologic transformation
   in non-smoking-related lung cancer (Pros et al., 2020)
4. Nicotine
   $\Rightarrow$ up-regulates FGFR3 & RB1 expression
   $\Rightarrow$ promote NSCLC cell proliferation & EMT transition (Du, Qi, Lu, Li, & Han, 2018)
5. RB1 mutation predicts poor outcomes in NSCLC (Bhateja et al., 2019)

# Mutation Shared Genes – Exact test V

## RUNX1T1

1. RUNX1 Partner Transcriptional Co-Repressor 1
2. RUNX1T1 facilitates transcriptional repression
   $\Leftarrow$ via DNA-binding transcription factors & histone-modifying enzymes (Davis, McGhee, & Meyers, 2003; Rossetti, Hoogeveen, & Sacchi, 2004; Melnick et al., 2000)
3. RUNX1T1 has a role in SCLC (He et al., 2021)
4. RUNX1T1 is associated with the brain metastatic process in lung cancer (Tomasini et al., 2020)

# Mutation Shared Genes – Exact test VI

## TP53

1. Tumor Protein P53
2. TP53 mutation is associated with response & longer survival under ICB in NSCLC (Assoun et al., 2019)
3. TP53 interacts with human lung cancer microbiome (Greathouse et al., 2018)

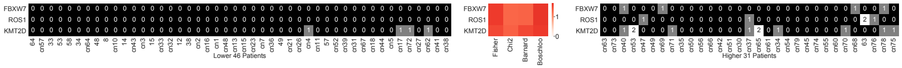



Figure: Exact Test on Mutation Shared Genes with Mean separation from LUSC

## Notable genes

1. FBXW7
2. ROS1
3. KMT2D

# Mutation Shared Genes – Exact test VIII

## FBXW7

1. F-Box and WD Repeat Domain Containing 7
2. A part of E3 ubiquitin-protein ligase complex
   $\Rightarrow$ mediates the ubiquitination and subsequent proteasomal degradation (Duda et al., 2012; Hao, Oehlmann, Sowa, Harper, & Pavletich, 2007; Yalla et al., 2018)
3. FBXW7 is a critical tumor suppressor of human cancers, including NSCLC (Yeh, Bellon, & Nicot, 2018)
4. FBXW7 mediates chemotherapeutic sensitivity & prognosis in NSCLC (Yokobori et al., 2014)
5. FBXW7-mediated ERK3 degradation proliferates lung cancer cells (An et al., 2022)

# Mutation Shared Genes – Exact test IX

## ROS1

1. ROS Proto-Oncogene 1
2. Crizontinib treatment in ROS1-rearranged NSCLC (Shaw et al., 2014)
3. Gene fusion (RET, ROS1, & ALK) in lung cancer (Takeuchi et al., 2012)
4. Resistance mechanisms of ROS1-positive lung cancer

# 4. Results

## 4.13. Mutation Shared Proportion

### 4.13.2. Bowtie2

# Mutation Shared Proportion Distribution

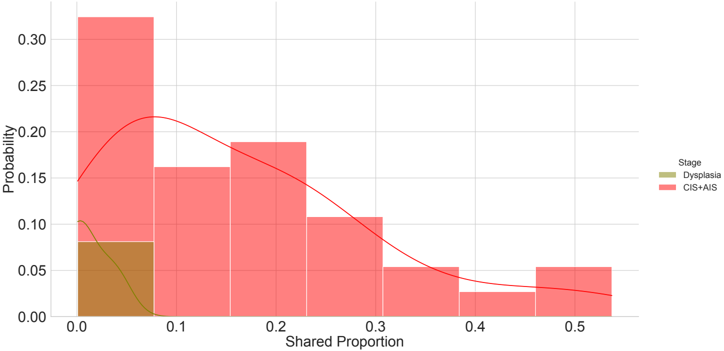
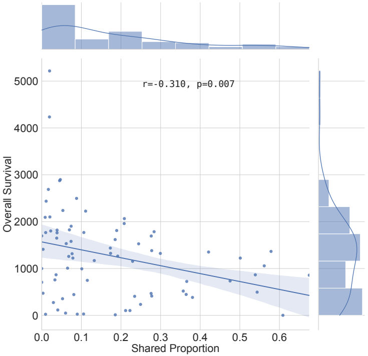



Figure: Mutation Shared Proportion Distribution in LUSC

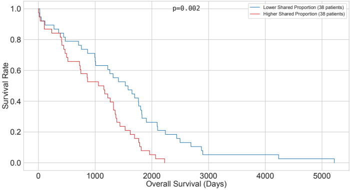
# Mutation Shared Proportion with Clinical Data I



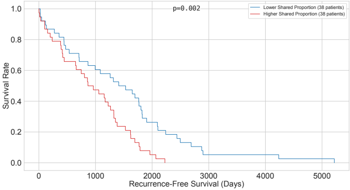

(a) Overall Survival (b) Recurrence-Free Survival

Figure: Mutation Shared Proportion with Clinical Data from LUSC

# Mutation Shared Proportion with Clinical Data II



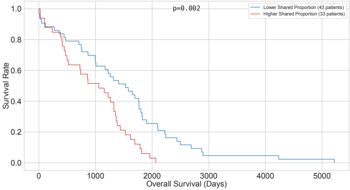

(a) Overall Survival



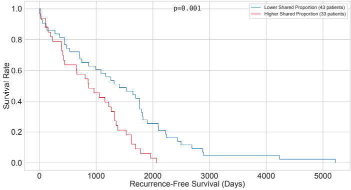

(b) Recurrence-Free Survival

Figure: K-M Survivial Plot with Median separation from LUSC

# Mutation Shared Proportion with Clinical Data III




(a) Overall Survival




(b) Recurrence-Free Survival

Figure: K-M Survivial Plot with Mean separation from LUSC

# Mutation Shared Genes

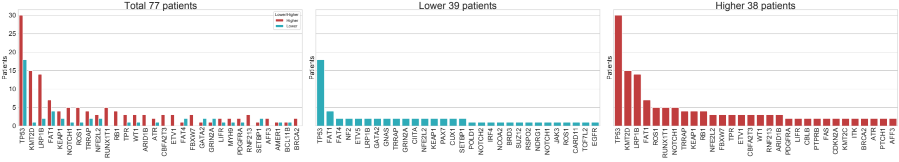



Figure: Mutation Shared Genes with Median separation from LUSC

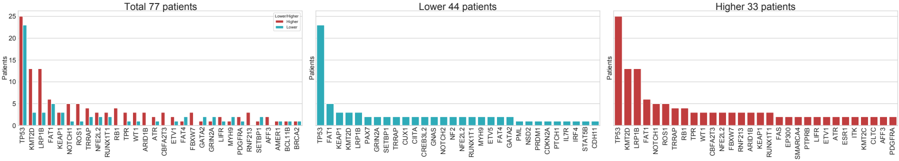



Figure: Mutation Shared Genes with Mean separation from LUSC

# Mutation Shared Genes – Exact test I

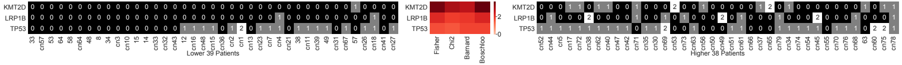



Figure: Exact Test on Mutation Shared Genes with Median separation from LUSC

## Notable genes

1. KMT2D
2. LRP1B
3. TP53

# Mutation Shared Genes – Exact test II

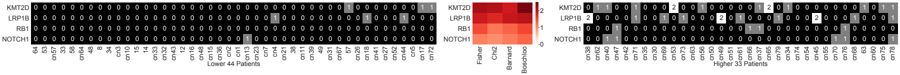



Figure: Exact Test on Mutation Shared Genes with Mean separation from LUSC

## Notable genes

1. KMT2D
2. LRP1B
3. RB1
4. NOTCH1

# Mutation Shared Genes – Exact test III

## NOTCH1

1. Neurogenic locus notch homolog protein 1
2. A receptor for membrane-bound ligands to regulate cell-fate determination (Brütsch et al., 2010)
3. NOTCH1 controls cell proliferation, apoptosis, & differentiation in lung carcinoma (Wael et al., 2014)
4. NOTCH1 enhances $\Uparrow$ EMT transition in lung cancer cell (Xie et al., 2012)
5. NOTCH1 mutation increased in lung cancer patient plasma (Liao et al., 2019)

# Findings in Mutation Sharing Proportion

# 4. Results

## 4.14. Mutation Shared Proportion vs. Mutational Signature

# Mutational Signatures I

## SBS1

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

## SBS2

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

# Mutational Signatures II

## SBS4

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

## SBS5

- Unknown (Alexandrov et al., 2013)
- SBS5 $\Uparrow$ in bladder cancer
- SBS5 $\Uparrow$ in many cancer types $\because$ Tobacco smoking

## SBS10b

- Polymerase $\epsilon$ exonuclease domain mutations (Alexandrov et al., 2020)

# Mutational Signatures III

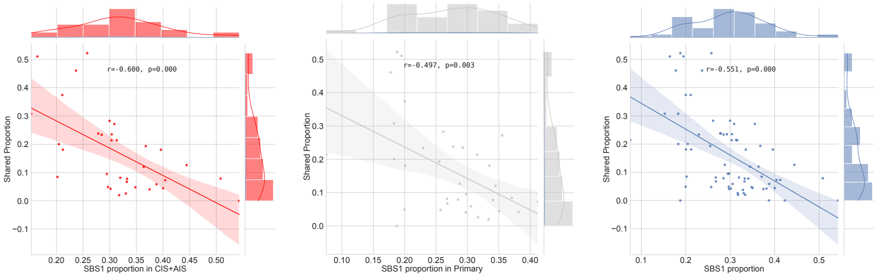
## SBS13

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

## SBS87

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

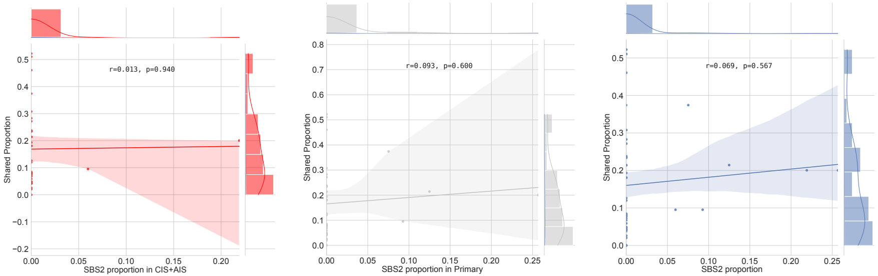
# SBS vs. Mutation Shared Proportion I




(a) CIS (b) Primary (c) All

Figure: Shared Proportion vs. SBS1 with BWA in LUSC

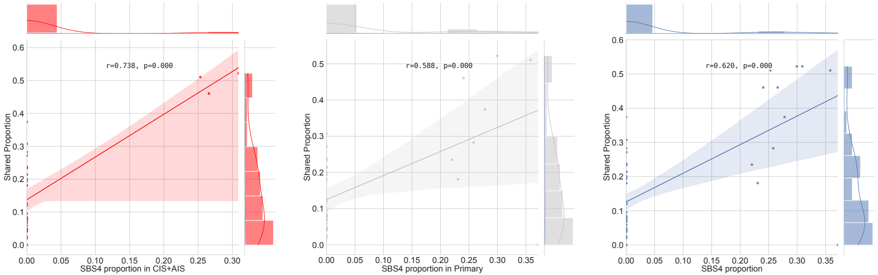
# SBS vs. Mutation Shared Proportion II




(a) CIS (b) Primary (c) All

Figure: Shared Proportion vs. SBS2 with BWA in LUSC

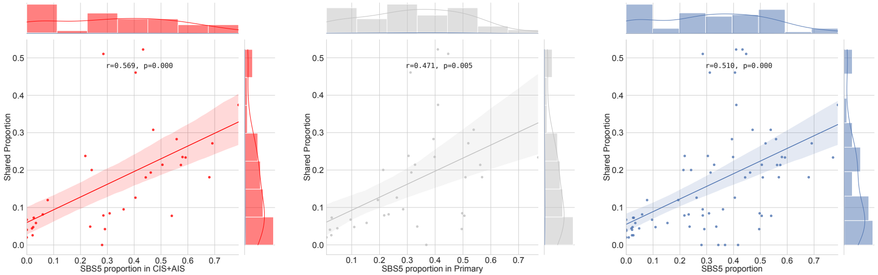
# SBS vs. Mutation Shared Proportion III




(a) CIS (b) Primary (c) All

Figure: Shared Proportion vs. SBS4 with BWA in LUSC

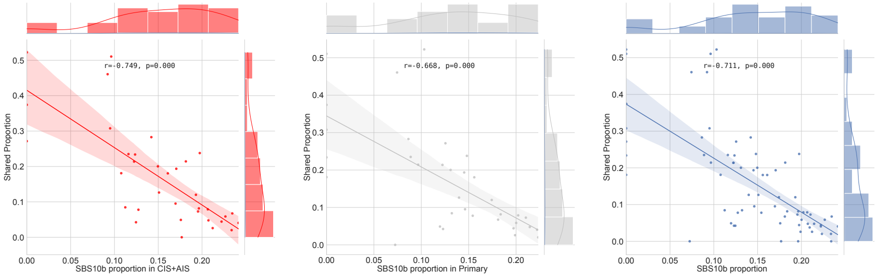
# SBS vs. Mutation Shared Proportion IV




(a) CIS (b) Primary (c) All

Figure: Shared Proportion vs. SBS5 with BWA in LUSC

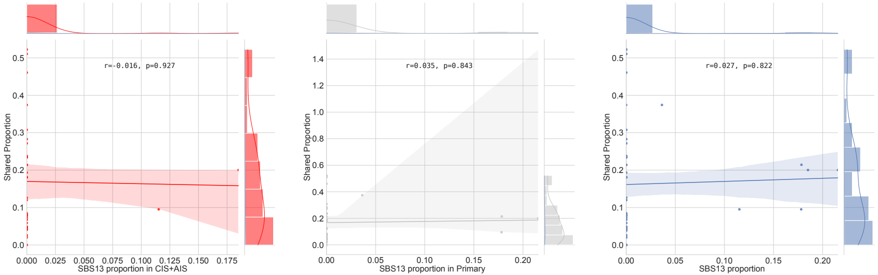
# SBS vs. Mutation Shared Proportion V




(a) CIS (b) Primary (c) All

Figure: Shared Proportion vs. SBS10b with BWA in LUSC

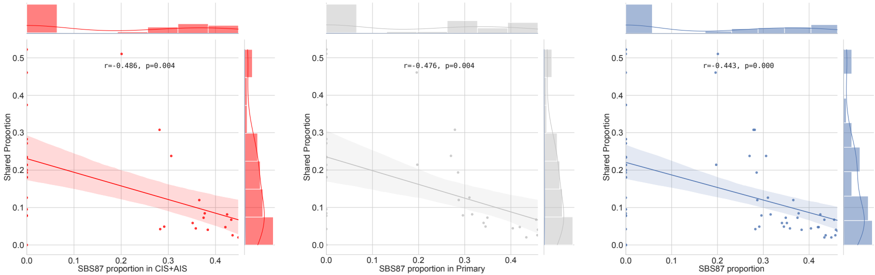
# SBS vs. Mutation Shared Proportion VI




(a) CIS (b) Primary (c) All

Figure: Shared Proportion vs. SBS13 with BWA in LUSC

# SBS vs. Mutation Shared Proportion VII




(a) CIS (b) Primary (c) All

Figure: Shared Proportion vs. SBS87 with BWA in LUSC

# Findings in Mutation Shared Proportion vs. Mutational Signature

# 4. Results

## 4.15. Identification of Microbial Sequences

# PathSeq?

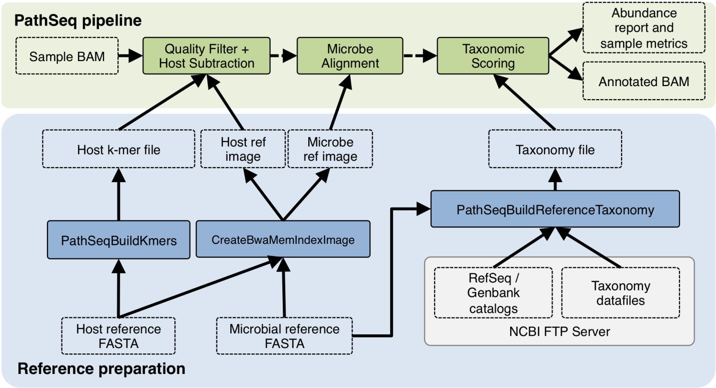



Figure: PathSeq Pipeline Diagram (Kostic et al., 2011; Walker et al., 2018)

# Taxonomy Class

## Taxonomy Class

1. Kingdom
2. Phylum
3. Class
4. Order
5. Familty
6. Genus
7. Species

# 4. Results

## 4.15. Identification of Microbial Sequences

### 4.15.1. WES – BWA

# Taxonomy Distribution

# 4. Results

## 4.15. Identification of Microbial Sequences

### 4.15.2. WTS – STAR

# Taxonomy Distribution

# Diversity Indices

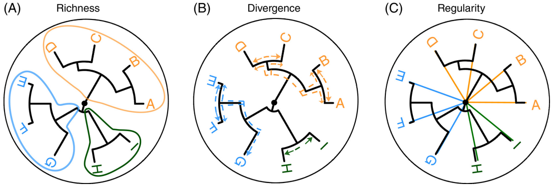



Figure: Three dimensions of phylogenic information (Tucker et al., 2017)

- A quantitative measure that shows richness, divergence, and regularity (Tucker et al., 2017)
- Alpha diversity indices: the richness of taxa **at a single community**
- Beta diversity indices: the taxonomic differentiation **between communities**

# Findings in Identification of Microbial Sequences

# 4. Results

## 4.16. Discovery of Gene Fusion

# Arriba?

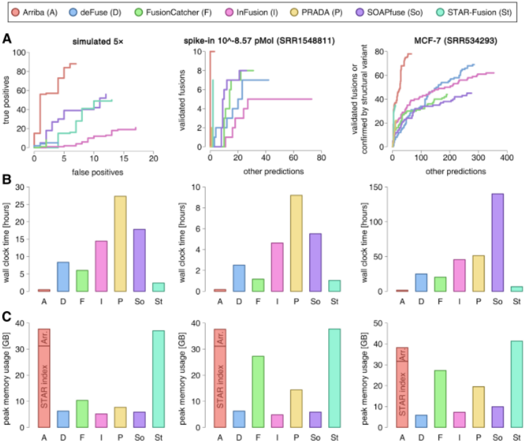



Figure: Benchmark of Arriba versus alternative methods (Uhrig et al., 2021)

# Findings in Gene Fusion Discovery

# 5. Discussion

# 6. References

# References I

Aktaş, O. N., Öztürk, A. B., Erman, B., Erus, S., Tanju, S., & Dilege, Ş. (2018). Role of natural killer cells in lung cancer. *Journal of cancer research and clinical oncology*, *144*(6), 997–1003.

Alam, H., Tang, M., Maitituoheti, M., Dhar, S. S., Kumar, M., Han, C. Y., ... others (2020). Kmt2d deficiency impairs super-enhancers to confer a glycolytic vulnerability in lung cancer. *Cancer cell*, *37*(4), 599–617.

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.

# References II


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.

An, H.-J., Lee, C.-J., Lee, G.-E., Choi, Y., Jeung, D., Chen, W., . . . others (2022). Fbxw7-mediated erk3 degradation regulates the proliferation of lung cancer cells. *Experimental & molecular medicine*, *54*(1), 35–46.

Andrews, S., Krueger, F., Segonds-Pichon, A., Biggins, L., Krueger, C., & Wingett, S. (2012, January). *FastQC.* Babraham Institute. Babraham, UK.

Ardeshir-Larijani, F., Bhateja, P., Lipka, M. B., Sharma, N., Fu, P., & Dowlati, A. (2018). Kmt2d mutation is associated with poor prognosis in non–small-cell lung cancer. *Clinical lung cancer*, *19*(4), e489–e501.

# References III


Assoun, S., Theou-Anton, N., Nguenang, M., Cazes, A., Danel, C., Abbar, B., ... others (2019). Association of tp53 mutations with response and longer survival under immune checkpoint inhibitors in advanced non-small-cell lung cancer. *Lung cancer*, *132*, 65–71.

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.

# References IV


Bernimoulin, M. P., Zeng, X.-L., Abbal, C., Giraud, S., Martinez, M., Michielin, O., . . . Spertini, O. (2003). Molecular basis of leukocyte rolling on psgl-1: predominant role of core-2 o-glycans and of tyrosine sulfate residue 51. *Journal of Biological Chemistry*, *278*(1), 37–47.

Beyer, M., & Schultze, J. L. (2006). Regulatory t cells in cancer. *Blood*, *108*(3), 804–811.

Bhateja, P., Chiu, M., Wildey, G., Lipka, M. B., Fu, P., Yang, M. C. L., . . . Dowlati, A. (2019). Retinoblastoma mutation predicts poor outcomes in advanced non small cell lung cancer. *Cancer medicine*, *8*(4), 1459–1466.

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.

Brahimi-Horn, M. C., Chiche, J., & Pouysségur, J. (2007). Hypoxia and cancer. *Journal of molecular medicine*, *85*(12), 1301–1307.

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.

Brown, L. C., Tucker, M. D., Sedhom, R., Schwartz, E. B., Zhu, J., Kao, C., ... others (2021). Lrp1b mutations are associated with favorable outcomes to immune checkpoint inhibitors across multiple cancer types. *Journal for immunotherapy of cancer*, *9*(3).

Brütsch, R., Liebler, S. S., Wüstehube, J., Bartol, A., Herberich, S. E., Adam, M. G., ... Fischer, A. (2010). Integrin cytoplasmic domain–associated protein-1 attenuates sprouting angiogenesis. *Circulation research*, *107*(5), 592–601.

# References VI


Budanov, A. V., & Karin, M. (2008). p53 target genes sestrin1 and sestrin2 connect genotoxic stress and mtor signaling. *Cell*, *134*(3), 451–460.

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.

Cal, S., Freije, J. M., López, J. M., Takada, Y., & Lopez-Otın, 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.

# References VII


Chantranupong, L., Wolfson, R. L., Orozco, J. M., Saxton, R. A., Scaria, S. M., Bar-Peled, L., ... Sabatini, D. M. (2014). The sestrins interact with gator2 to negatively regulate the amino-acid-sensing pathway upstream of mtorc1. *Cell reports*, *9*(1), 1–8.

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, H., Chong, W., Wu, Q., Yao, Y., Mao, M., & Wang, X. (2019). Association of lrp1b mutation with tumor mutation burden and outcomes in melanoma and non-small cell lung cancer patients treated with immune check-point blockades. *Frontiers in immunology*, *10*, 1113.

# References VIII


Chen, J., Song, J., Yuan, P., Tian, Q., Ji, Y., Ren-Patterson, R., . . . Weinberger, D. R. (2011). Orientation and cellular distribution of membrane-bound catechol-o-methyltransferase in cortical neurons: implications for drug development. *Journal of biological chemistry*, *286*(40), 34752–34760.

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.

Cho, Y.-W., Hong, T., Hong, S., Guo, H., Yu, H., Kim, D., . . . others (2007). Ptip associates with mll3-and mll4-containing histone h3 lysine 4 methyltransferase complex. *Journal of Biological Chemistry*, *282*(28), 20395–20406.

Choi, B., Han, T.-S., Min, J., Hur, K., Lee, S.-M., Lee, H.-J., . . . Yang, H.-K. (2017). Mal and tmem220 are novel dna methylation markers in human gastric cancer. *Biomarkers*, *22*(1), 35–44.

Christensen, I. B., Wu, Q., Bohlbro, A. S., Skals, M. G., Damkier, H. H., Hübner, C. A., . . . Praetorius, J. (2020). Genetic disruption of slc4a10 alters the capacity for cellular metabolism and vectorial ion transport in the choroid plexus epithelium. *Fluids and Barriers of the CNS*, *17*(1), 1–18.

Collins, L. G., Haines, C., Perkel, R., & Enck, R. E. (2007). Lung cancer: diagnosis and management. *American family physician*, *75*(1), 56–63.

# References X


Cordani, M., Butera, G., Dando, I., Torrens-Mas, M., Butturini, E., Pacchiana, R., . . . others (2018). Mutant p53 blocks sesn1/ampk/pgc-1$\alpha$/ucp2 axis increasing mitochondrial o $2^{-}\cdot$ production in cancer cells. *British journal of cancer*, *119*(8), 994–1008.

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.

Davis, J. N., McGhee, L., & Meyers, S. (2003). The eto (mtg8) gene family. *Gene*, *303*, 1–10.

# References XI


Dawling, S., Roodi, N., Mernaugh, R. L., Wang, X., & Parl, F. F. (2001). Catechol-o-methyltransferase (comt)-mediated metabolism of catechol estrogens: comparison of wild-type and variant comt isoforms. *Cancer research*, *61*(18), 6716–6722.

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.

Deckers, J., Hammad, H., & Hoste, E. (2018). Langerhans cells: sensing the environment in health and disease. *Frontiers in immunology*, *9*, 93.

# References XII


de Marco, M. C., Martín-Belmonte, F., Kremer, L., Albar, J. P., Correas, I., Vaerman, J. P., . . . Alonso, M. A. (2002). Mal2, a novel raft protein of the mal family, is an essential component of the machinery for transcytosis in hepatoma hepg2 cells. *Journal of Cell Biology*, *159*(1), 37–44.

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.

# References XIII


Ding, B., Haidurov, A., Chawla, A., Parmigiani, A., van de Kamp, G., Dalina, A., . . . others (2019). p53-inducible sestrins might play opposite roles in the regulation of early and late stages of lung carcinogenesis. *Oncotarget*, *10*(65), 6997.

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.

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.

# References XIV


Du, X., Qi, F., Lu, S., Li, Y., & Han, W. (2018). Nicotine upregulates fgfr3 and rb1 expression and promotes non-small cell lung cancer cell proliferation and epithelial-to-mesenchymal transition via downregulation of mir-99b and mir-192. *Biomedicine & Pharmacotherapy*, *101*, 656–662.

Duda, D. M., Olszewski, J. L., Tron, A. E., Hammel, M., Lambert, L. J., Waddell, M. B., . . . Schulman, B. A. (2012). Structure of a glomulin-rbx1-cul1 complex: inhibition of a ring e3 ligase through masking of its e2-binding surface. *Molecular cell*, *47*(3), 371–382.

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.

# References XV

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.

Erfani, N., Mehrabadi, S. M., Ghayumi, M. A., Haghshenas, M. R., Mojtahedi, Z., Ghaderi, A., & Amani, D. (2012). Increase of regulatory t cells in metastatic stage and ctla-4 over expression in lymphocytes of patients with non-small cell lung cancer (nsclc). *Lung cancer*, *77*(2), 306–311.

Fisler, D. A., Sikaria, D., Yavorski, J. M., Tu, Y. N., & Blanck, G. (2018). Elucidating feed-forward apoptosis signatures in breast cancer datasets: Higher fos expression associated with a better outcome. *Oncology letters*, *16*(2), 2757–2763.

# References XVI


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.

Gershenfeld, H. K., Hershberger, R. J., Shows, T. B., & Weissman, I. L. (1988). Cloning and chromosomal assignment of a human cdna encoding a t cell-and natural killer cell-specific trypsin-like serine protease. *Proceedings of the National Academy of Sciences*, *85*(4), 1184–1188.

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 XVII


Goveia, J., Rohlenova, K., Taverna, F., Treps, L., Conradi, L.-C., Pircher, A., ... others (2020). An integrated gene expression landscape profiling approach to identify lung tumor endothelial cell heterogeneity and angiogenic candidates. *Cancer cell*, *37*(1), 21–36.

Greathouse, K. L., White, J. R., Vargas, A. J., Bliskovsky, V. V., Beck, J. A., von Muhlinen, N., ... others (2018). Interaction between the microbiome and tp53 in human lung cancer. *Genome biology*, *19*(1), 1–16.

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

Gridelli, C., Rossi, A., Carbone, D. P., Guarize, J., Karachaliou, N., Mok, T., ... Rosell, R. (2015). Non-small-cell lung cancer. *Nature reviews Disease primers*, *1*(1), 1–16.

# References XVIII


Grzegrzolka, J., Olbromski, M., Gomulkiewicz, A., Piotrowska, A., Glatzel-Plucinska, N., Ratajczak, K., ... others (2021). Role of tesmin expression in non-small cell lung cancer. *Oncology Letters*, *21*(1), 1–1.

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.

Hameed, A., Lowrey, D., Lichtenheld, M., & Podack, E. (1988). Characterization of three serine esterases isolated from human il-2 activated killer cells. *The Journal of Immunology*, *141*(9), 3142–3147.

# References XIX


Hao, B., Oehlmann, S., Sowa, M. E., Harper, J. W., & Pavletich, N. P. (2007). Structure of a fbw7-skp1-cyclin e complex: multisite-phosphorylated substrate recognition by scf ubiquitin ligases. *Molecular cell*, *26*(1), 131–143.

Harbour, J. W., Luo, R. X., Dei Santi, A., Postigo, A. A., & Dean, D. C. (1999). Cdk phosphorylation triggers sequential intramolecular interactions that progressively block rb functions as cells move through g1. *Cell*, *98*(6), 859–869.

He, T., Wildey, G., McColl, K., Savadelis, A., Spainhower, K., McColl, C., ... others (2021). Identification of runx1t1 as a potential epigenetic modifier in small-cell lung cancer. *Molecular Oncology*, *15*(1), 195–209.

# References XX

Henderson, L. B., Melton, K., Wert, S., Couriel, J., Bush, A., Ashworth, M., & Nogee, L. M. (2013). Large abca3 and sftpc deletions resulting in lung disease. *Annals of the American Thoracic Society*, *10*(6), 602–607.

Hong, S., Won, Y.-J., Lee, J. J., Jung, K.-W., Kong, H.-J., Im, J.-S., ... others (2021). Cancer statistics in korea: Incidence, mortality, survival, and prevalence in 2018. *Cancer Research and Treatment: Official Journal of Korean Cancer Association*, *53*(2), 301.

Hong, Z., Chang, L.-L., Fang-Jie, Y., Yan, H., Chen-Ming, Z., Tian-Yi, Z., ... others (2018). Akr1c1 activates stat3 to promote the metastasis of non-small cell lung cancer. *Theranostics*, *8*(3), 676.

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.

# References XXI

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.

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.

Jia, Q., Wu, W., Wang, Y., Alexander, P. B., Sun, C., Gong, Z., . . . others (2018). Local mutational diversity drives intratumoral immune heterogeneity in non-small cell lung cancer. *Nature communications*, *9*(1), 1–10.

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.

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.

# References XXIII

Kim, E. Y., Cha, Y. J., Lee, S. H., Jeong, S., Choi, Y. J., Moon, D. H., ... Chang, Y. S. (2022). Early lung carcinogenesis and tumor microenvironment observed by single-cell transcriptome analysis. *Translational oncology*, *15*(1), 101277.

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.

Kostic, A. D., Ojesina, A. I., Pedamallu, C. S., Jung, J., Verhaak, R. G., Getz, G., & Meyerson, M. (2011). Pathseq: software to identify or discover microbes by deep sequencing of human tissue. *Nature biotechnology*, *29*(5), 393–396.

# References XXIV


Kou, F., Sun, H., Wu, L., Li, B., Zhang, B., Wang, X., & Yang, L. (2020). Top2a promotes lung adenocarcinoma cells' malignant progression and predicts poor prognosis in lung adenocarcinoma. *Journal of Cancer*, *11*(9), 2496.

Krähenbühl, O., Rey, C., Jenne, D., Lanzavecchia, A., Groscurth, P., Carrel, S., & Tschopp, J. (1988). Characterization of granzymes a and b isolated from granules of cloned human cytotoxic t lymphocytes. *The Journal of Immunology*, *141*(10), 3471–3477.

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.

Kumari, S., Arora, M., Singh, J., Chauhan, S. S., Kumar, S., & Chopra, A. (2021). L-selectin expression is associated with inflammatory microenvironment and favourable prognosis in breast cancer. *3 Biotech*, *11*(2), 1–13.

Kyotani, Y., Takasawa, S., & Yoshizumi, M. (2019). Proliferative pathways of vascular smooth muscle cells in response to intermittent hypoxia. *International journal of molecular sciences*, *20*(11), 2706.

Lacal, P. M., Petrillo, M. G., Ruffini, F., Muzi, A., Bianchini, R., Ronchetti, S., . . . Nocentini, G. (2013). Glucocorticoid-induced tumor necrosis factor receptor family-related ligand triggering upregulates vascular cell adhesion molecule-1 and intercellular adhesion molecule-1 and promotes leukocyte adhesion. *Journal of Pharmacology and Experimental Therapeutics*, *347*(1), 164–172.

Lan, S., Li, H., Liu, Y., Ma, L., Liu, X., Liu, Y., . . . Cheng, Y. (2019). Somatic mutation of lrp1b is associated with tumor mutational burden in patients with lung cancer. *Lung cancer*, *132*, 154–156.

Lara-Lemus, R. (2019). On the role of myelin and lymphocyte protein (mal) in cancer: a puzzle with two faces. *Journal of Cancer*, *10*(10), 2312.

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.

# References XXVII


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.

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.

Li, D., Zhang, J., Wu, L., Yang, X., Chen, Z., & Yuan, J. (2021). Myelin and lymphocyte protein (mal): A novel biomarker for uterine corpus endometrial carcinoma. *Cancer Management and Research*, *13*, 7311.

Liao, Y., Ma, Z., Zhang, Y., Li, D., Lv, D., Chen, Z., . . . others (2019). Targeted deep sequencing from multiple sources demonstrates increased notch1 alterations in lung cancer patient plasma. *Cancer medicine*, *8*(12), 5673–5686.

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.

Liu, X., Xu, X., Deng, W., Huang, M., Wu, Y., Zhou, Z., . . . others (2019). Ccl18 enhances migration, invasion and emt by binding ccr8 in bladder cancer cells. *Molecular medicine reports*, *19*(3), 1678–1686.

# References XXIX


Liu, Y.-Z., Wang, B.-S., Jiang, Y.-Y., Cao, J., Hao, J.-J., Zhang, Y., ... Wang, M.-R. (2017). Mcms expression in lung cancer: implication of prognostic significance. *Journal of Cancer*, *8*(18), 3641.

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for rna-seq data with deseq2. *Genome biology*, *15*(12), 1–21.

Ma, G., Ji, D., Qu, X., Liu, S., Yang, X., Wang, G., ... Du, J. (2019). Mining and validating the expression pattern and prognostic value of acetylcholine receptors in non-small cell lung cancer. *Medicine*, *98*(20).

Ma, W., Wang, B., Zhang, Y., Wang, Z., Niu, D., Chen, S., ... others (2019). Prognostic significance of top2a in non-small cell lung cancer revealed by bioinformatic analysis. *Cancer cell international*, *19*(1), 1–17.

# References XXX


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.

Melnick, A. M., Westendorf, J. J., Polinger, A., Carlile, G. W., Arai, S., Ball, H. J., . . . Licht, J. D. (2000). The eto protein disrupted in t (8; 21)-associated acute myeloid leukemia is a corepressor for the promyelocytic leukemia zinc finger protein. *Molecular and cellular biology*, *20*(6), 2075–2086.

Mermel, C. H., Schumacher, S. E., Hill, B., Meyerson, M. L., Beroukhim, R., & Getz, G. (2011). Gistic2. 0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome biology*, *12*(4), 1–14.

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.

# References XXXII


Mondejar-Parreño, G., Perez-Vizcaino, F., & Cogolludo, A. (2020). Kv7 channels in lung diseases. *Frontiers in Physiology*, *11*, 634.

Narayanan, S., Kawaguchi, T., Yan, L., Peng, X., Qi, Q., & Takabe, K. (2018). Cytolytic activity score to assess anticancer immunity in colorectal cancer. *Annals of surgical oncology*, *25*(8), 2323–2331.

Nikolaou, S., Qiu, S., Fiorentino, F., Simillis, C., Rasheed, S., Tekkis, P., & Kontovounisios, C. (2020). The role of neurotensin and its 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 (2012). 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.

Paul, R., Hu, B., Musahl, C., Hameister, H., & Knippers, R. (1996). Coding sequence and chromosome mapping of the human gene (cdc46) for replication protein hcdc46/mcm5. *Cytogenetic and Genome Research*, *73*(4), 317–321.

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.

Plitas, G., Konopacki, C., Wu, K., Bos, P. D., Morrow, M., Putintseva, E. V., … Rudensky, A. Y. (2016). Regulatory t cells exhibit distinct features in human breast cancer. *Immunity*, *45*(5), 1122–1134.

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.

Pros, E., Saigi, M., Alameda, D., Gomez-Mariano, G., Martinez-Delgado, B., Alburquerque-Bejar, J., . . . others (2020). Genome-wide profiling of non-smoking-related lung cancer cells reveals common rb1 rearrangements associated with histopathologic transformation in egfr-mutant tumors. *Annals of Oncology*, *31*(2), 274–282.

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.

# References XXXVI


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.

Rosenthal, R., McGranahan, N., Herrero, J., Taylor, B. S., & Swanton, C. (2016). Deconstructsigs: delineating mutational processes in single tumors distinguishes dna repair deficiencies and patterns of carcinoma evolution. *Genome biology*, *17*(1), 1–11.

Rossetti, S., Hoogeveen, A. T., & Sacchi, N. (2004). The mtg proteins: chromatin repression players with a passion for networking. *Genomics*, *84*(1), 1–9.

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 XXXVII


Salamon, P., Mekori, Y. A., & Shefler, I. (2020). Lung cancer-derived extracellular vesicles: a possible mediator of mast cell activation in the tumor microenvironment. *Cancer Immunology, Immunotherapy*, *69*(3), 373–381.

Santiago, L., Castro, M., Sanz-Pamplona, R., Garzón, M., Ramirez-Labrada, A., Tapia, E., ... others (2020). Extracellular granzyme a promotes colorectal cancer development by enhancing gut inflammation. *Cell reports*, *32*(1), 107847.

Schroeder, B. C., Kubisch, C., Stein, V., & Jentsch, T. J. (1998). Moderate loss of function of cyclic-amp-modulated kcnq2/kcnq3 k+ channels causes epilepsy. *Nature*, *396*(6712), 687–690.

Shaw, A. T., Ou, S.-H. I., Bang, Y.-J., Camidge, D. R., Solomon, B. J., Salgia, R., ... others (2014). Crizotinib in ros1-rearranged non–small-cell lung cancer. *New England Journal of Medicine*, *371*(21), 1963–1971.

Shin, M. H., Kim, J., Lim, S. A., Kim, J., Kim, S.-J., & Lee, K.-M. (2020). Nk cell-based immunotherapies in cancer. *Immune network*, *20*(2).

Shinsky, S. A., Monteith, K. E., Viggiano, S., & Cosgrove, M. S. (2015). Biochemical reconstitution and phylogenetic comparison of human set1 family core complexes involved in histone methylation. *Journal of Biological Chemistry*, *290*(10), 6361–6375.

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.

# References XXXIX


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.

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.

Takeuchi, K., Soda, M., Togashi, Y., Suzuki, R., Sakata, S., Hatano, S., . . . others (2012). Ret, ros1 and alk fusions in lung cancer. *Nature medicine*, *18*(3), 378–381.

Tate John, G., Sally, B., Jubb Harry, C., Zbyslaw, S., Beare David, M., Nidhi, B., . . . Elisabeth, D. (2018). Stefancsik ray, thompson sam l, 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.

Tiffany, H. L., Lautens, L. L., Gao, J.-L., Pease, J., Locati, M., Combadiere, C., . . . Murphy, P. M. (1997). Identification of ccr8: a human monocyte and thymus receptor for the cc chemokine i-309. *The Journal of experimental medicine*, *186*(1), 165–170.

# References XLI


Tomasini, P., Barlesi, F., Gilles, S., Nanni-Metellus, I., Soffietti, R., Denicolai, E., . . . Metellus, P. (2020). Comparative genomic analysis of primary tumors and paired brain metastases in lung cancer patients by whole exome sequencing: a pilot study. *Oncotarget*, *11*(50), 4648.

Travis, W. D. (2002). Pathology of lung cancer. *Clinics in chest medicine*, *23*(1), 65–81.

Tsuji, T., Ficarro, S. B., & Jiang, W. (2006). Essential role of phosphorylation of mcm2 by cdc7/dbf4 in the initiation of dna replication in mammalian cells. *Molecular biology of the cell*, *17*(10), 4459–4472.

Tucker, C. M., Cadotte, M. W., Carvalho, S. B., Davies, T. J., Ferrier, S., Fritz, S. A., . . . others (2017). A guide to phylogenetic metrics for conservation, community ecology and macroecology. *Biological Reviews*, *92*(2), 698–715.

# References XLII


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.

Vaupel, P., & Mayer, A. (2007). Hypoxia in cancer: significance and impact on clinical outcome. *Cancer and Metastasis Reviews*, *26*(2), 225–239.

Vermi, W., Soncini, M., Melocchi, L., Sozzani, S., & Facchetti, F. (2011). Plasmacytoid dendritic cells and cancer. *Journal of leukocyte biology*, *90*(4), 681–690.

# References XLIII


Villarreal, D. O., L'Huillier, A., Armington, S., Mottershead, C., Filippova, E. V., Coder, B. D., . . . Princiotta, M. F. (2018). Targeting ccr8 induces protective antitumor immunity and enhances vaccine-induced responses in colon cancer. *Cancer research*, *78*(18), 5340–5348.

Wael, H., Yoshida, R., Kudoh, S., Hasegawa, K., Niimori-Kita, K., & Ito, T. (2014). Notch1 signaling controls cell proliferation, apoptosis and differentiation in lung carcinoma. *Lung Cancer*, *85*(2), 131–140.

Walker, M. A., Pedamallu, C. S., Ojesina, A. I., Bullman, S., Sharpe, T., Whelan, C. W., & Meyerson, M. (2018). Gatk pathseq: a customizable computational tool for the discovery and identification of microbial sequences in libraries from eukaryotic hosts. *Bioinformatics*, *34*(24), 4287–4289.

# References XLIV


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, C.-Z., Yano, H., Nagashima, K., & Seino, S. (2000). The na+-driven cl-/hco3- exchanger: Cloning, tissue distribution, and functional characterization* 210. *Journal of Biological Chemistry*, *275*(45), 35486–35490.

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 XLV


Wedepohl, S., Dernedde, J., Vahedi-Faridi, A., Tauber, R., Saenger, W., & Bulut, H. (2017). Reducing macro-and microheterogeneity of n-glycans enables the crystal structure of the lectin and egf-like domains of human l-selectin to be solved at 1.9 å resolution. *ChemBioChem*, *18*(13), 1338–1345.

Wolfson, R. L., Chantranupong, L., Saxton, R. A., Shen, K., Scaria, S. M., Cantor, J. R., & Sabatini, D. M. (2016). Sestrin2 is a leucine sensor for the mtorc1 pathway. *Science*, *351*(6268), 43–48.

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.

Wyles, J. P., Wu, Z., Mirski, S. E., & Cole, S. P. (2007). Nuclear interactions of topoisomerase ii $\alpha$ and $\beta$ with phospholipid scramblase 1. *Nucleic acids research*, *35*(12), 4076–4085.

# References XLVI


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.

Xiao, D., Li, F., Pan, H., Liang, H., Wu, K., & He, J. (2017). Integrative analysis of genomic sequencing data reveals higher prevalence of lrp1b mutations in lung adenocarcinoma patients with copd. *Scientific reports*, *7*(1), 1–8.

Xiao, H., He, M., Xie, G., Liu, Y., Zhao, Y., Ye, X., . . . Zhang, M. (2019). The release of tryptase from mast cells promote tumor cell metastasis via exosomes. *BMC cancer*, *19*(1), 1–9.

Xie, M., Zhang, L., He, C.-s., Xu, F., Liu, J.-l., Hu, Z.-h., . . . Tian, Y. (2012). Activation of notch-1 enhances epithelial–mesenchymal transition in gefitinib-acquired resistant lung cancer cells. *Journal of cellular biochemistry*, *113*(5), 1501–1513.

Yalla, K., Elliott, C., Day, J. P., Findlay, J., Barratt, S., Hughes, Z. A., ... others (2018). Fbxw7 regulates disc1 stability via the ubiquitin-proteosome system. *Molecular psychiatry*, *23*(5), 1278–1286.

Ye, T., Yang, X., Liu, H., Lv, P., & Ye, Z. (2020). Long non-coding rna blacat1 in human cancers. *OncoTargets and therapy*, *13*, 8263.

Yeh, C.-H., Bellon, M., & Nicot, C. (2018). Fbxw7: a critical tumor suppressor of human cancers. *Molecular cancer*, *17*(1), 1–19.

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.

Yokobori, T., Yokoyama, Y., Mogi, A., Endoh, H., Altan, B., Kosaka, T., ... others (2014). Fbxw7 mediates chemotherapeutic sensitivity and prognosis in nsclcs. *Molecular Cancer Research*, *12*(1), 32–37.

Zeng, S., Liu, A., Dai, L., Yu, X., Zhang, Z., Xiong, Q., . . . others (2019). Prognostic value of top2a in bladder urothelial carcinoma and potential molecular mechanisms. *BMC cancer*, *19*(1), 1–12.

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.

Zhou, H., Zhang, C., Li, H., Chen, L., & Cheng, X. (2020). A novel risk score system of immune genes associated with prognosis in endometrial cancer. *Cancer cell international*, *20*(1), 1–12.

Zhu, T.-T., Sun, R.-L., Yin, Y.-L., Quan, J.-P., Song, P., Xu, J., . . . Li, P. (2019). Long noncoding rna uca1 promotes the proliferation of hypoxic human pulmonary artery smooth muscle cells. *Pflügers Archiv-European Journal of Physiology*, *471*(2), 347–355.

# References XLIX


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.