Lung Precancer Study

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

Department of Biomedical Engineering
Ulsan National Institute of Science and Technology

jwlee230@unist.ac.kr

2021-10-04

Overview

1. Introduction

1. Introduction

1.1. Lung Cancer

Lung Cancer?

The most common cancer

The most common form of cancer:
12.3 % of all cancers (Minna, Roth, & Gazdar, 2002)

The most important factor

Tobacco

Cancer Survival Rate in Korea

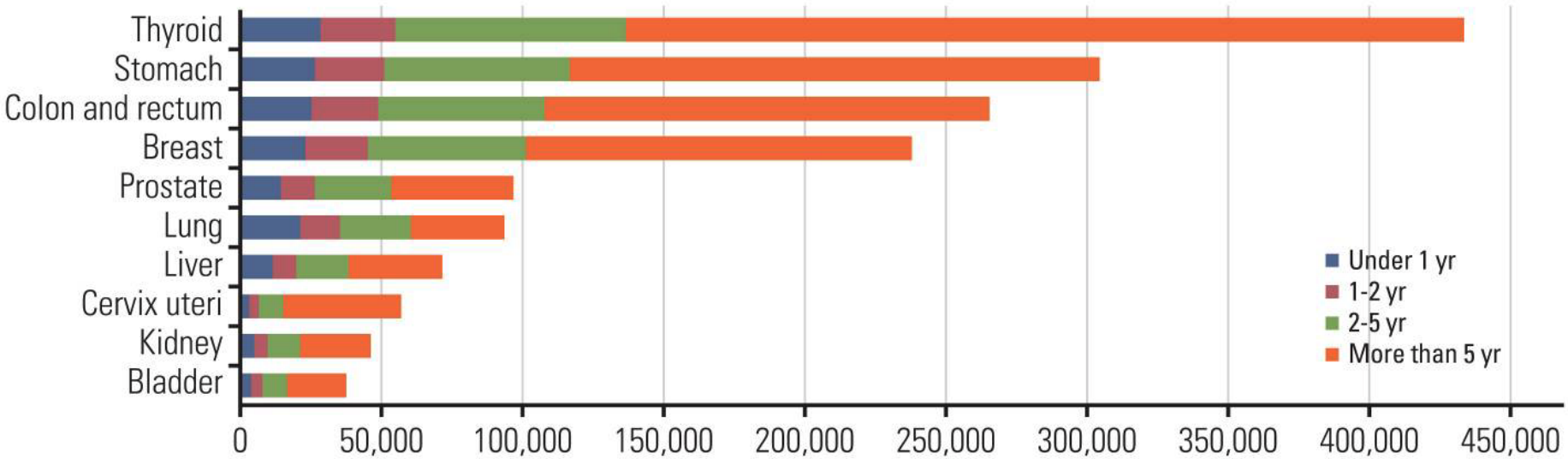


Figure: Common cancer survival rates (Hong et al., 2021)

Survival rate (More than 5 yr)

- Tyroid: 68.4 %
- Lung: 35.4 %

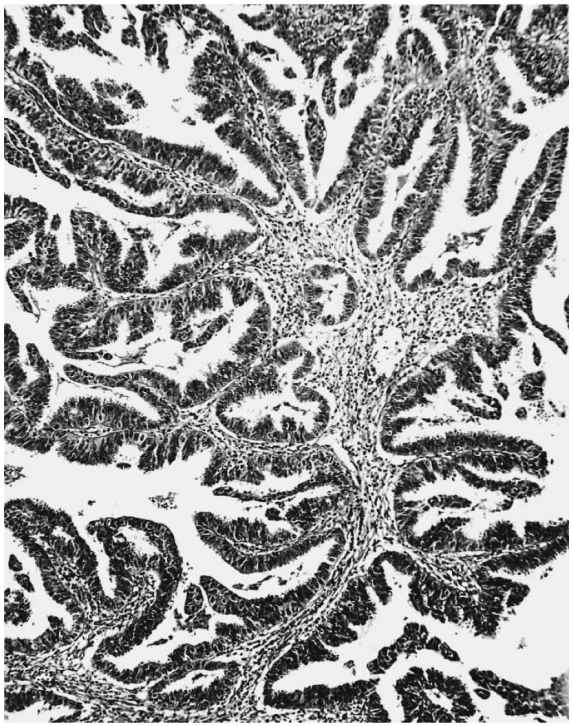
Type of Lung Cancer

Types of lung cancer:

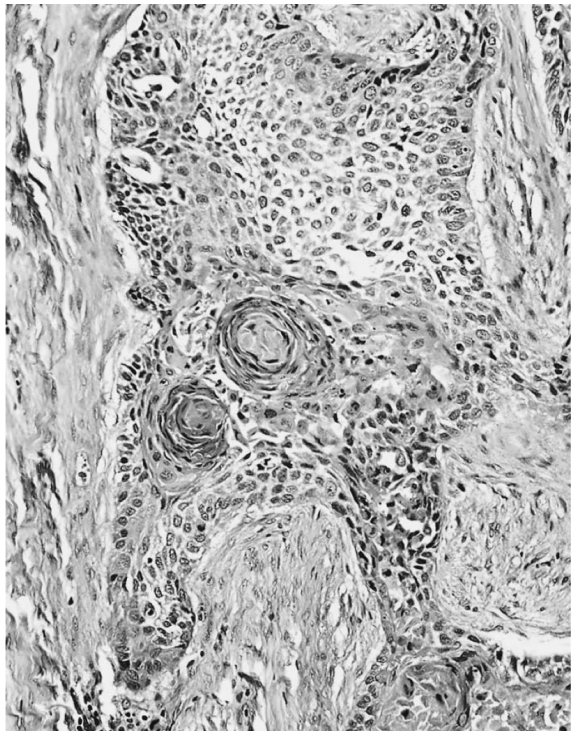
1. Adenocarcinoma (LUAD) (40 %) ★
2. Squamous cell carcinoma (LUSC) (25 %) ★
3. Small cell carcinoma (20 %)
4. Large cell carcinoma (10 %)
5. Adenosquamous carcinoma (< 5 %)
6. Carcinoid (< 5 %)
7. Bronchioalveolar (Bronchial gland carcinoma)

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

LUAD vs. LUSC I



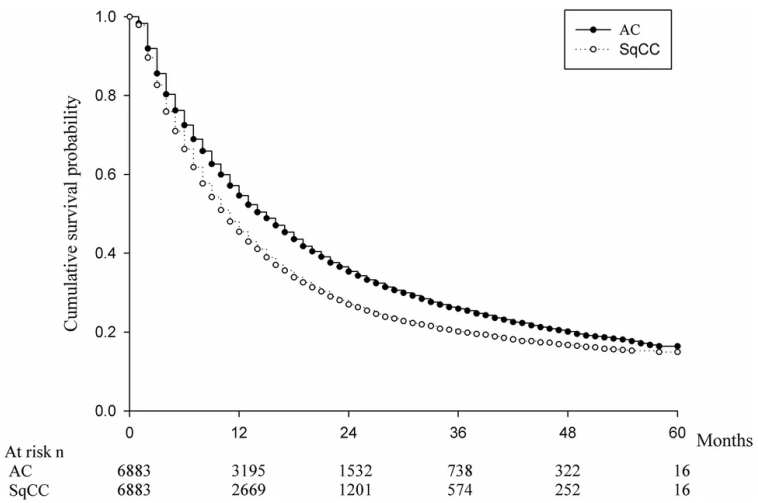
(a) LUAD



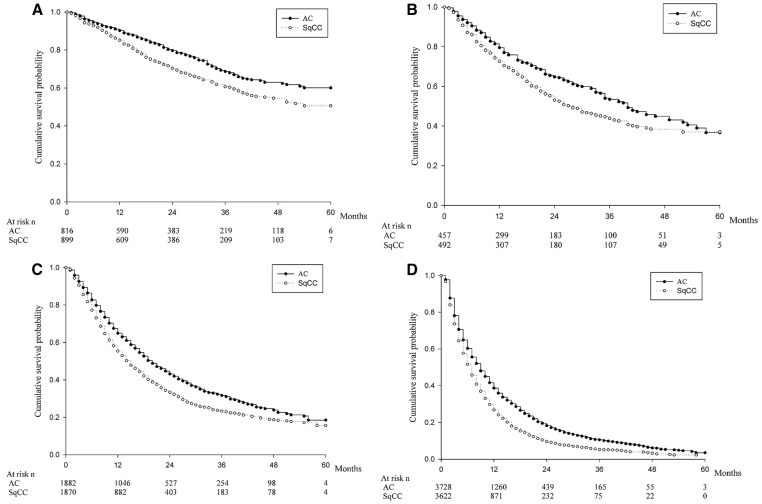
(b) LUSC

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

LUAD vs. LUSC II



(a) All patients



(b) By cancer stages

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

Findings

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

1. Introduction

1.2. Study Objectives

Study Objectives

Find different mutations

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

Pathway examine

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

Ultra-deep sequencing

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

- from blood
- from urine
- from bronchus

2. Materials

Lung Cancer Data

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

3. Methods

3. Methods

3.1. Workflows

Data pre-processing for variant discovery

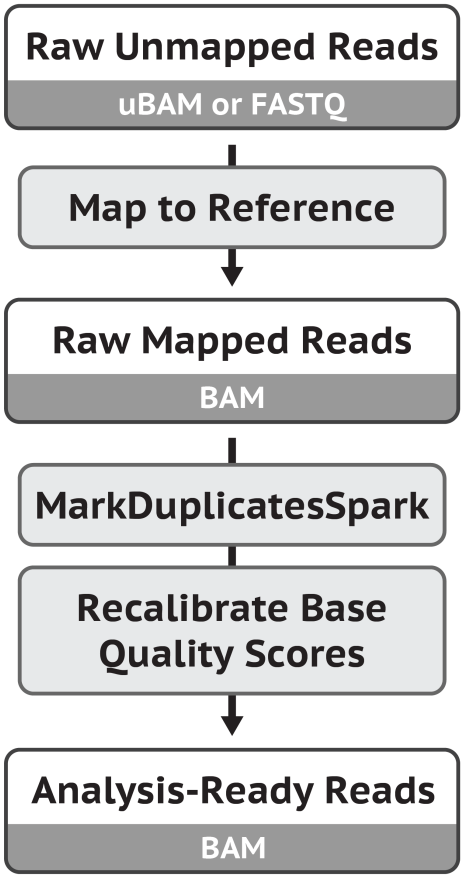


Figure: Data pre-processing for variant discovery (Van der Auwera et al., 2013; DePristo et al., 2011)

Somatic short variant discovery

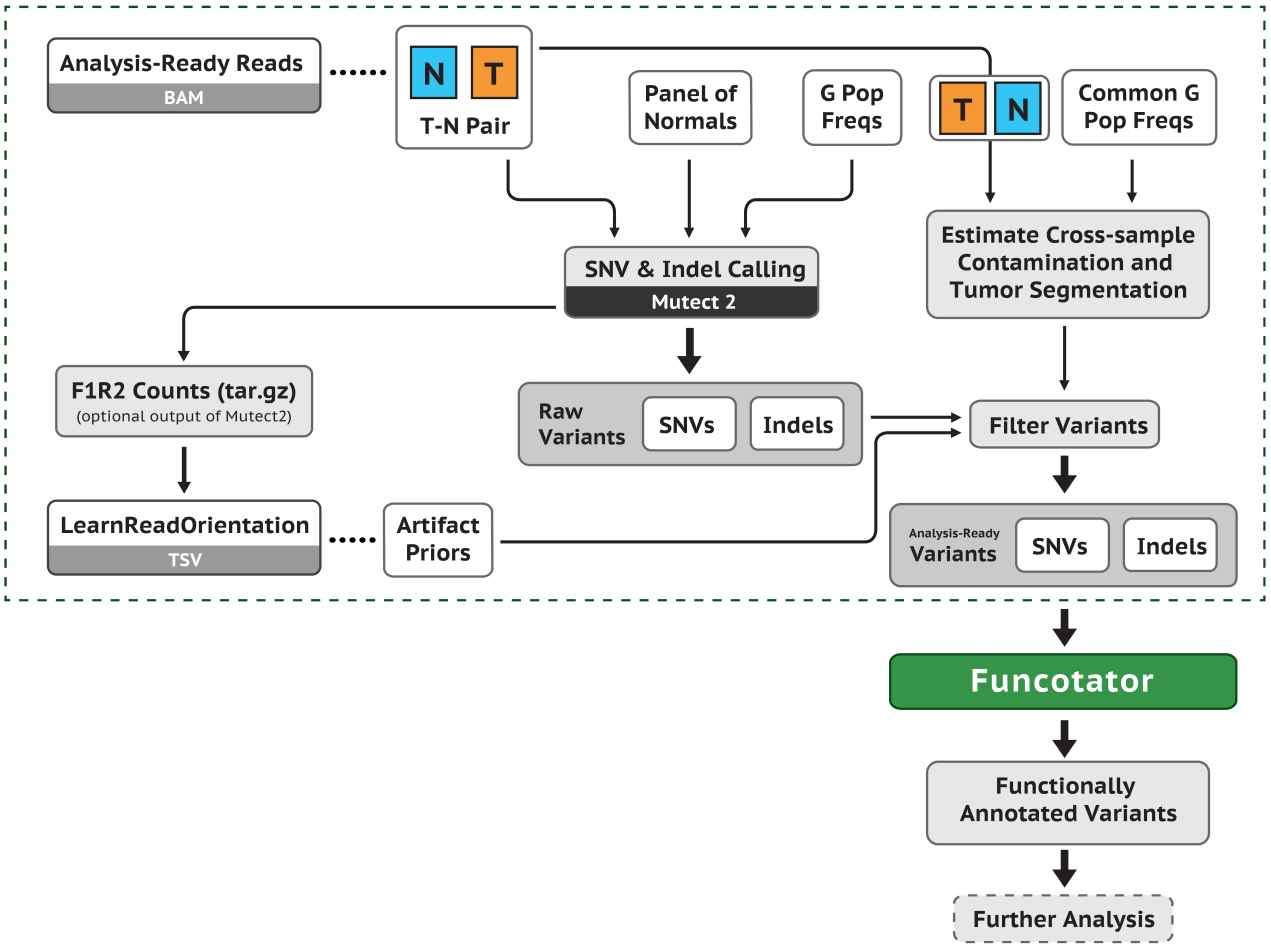


Figure: Somatic short variant (SNVs + Indels) discovery workflow (Van der Auwera et al., 2013; DePristo et al., 2011)

Germline short variant discovery

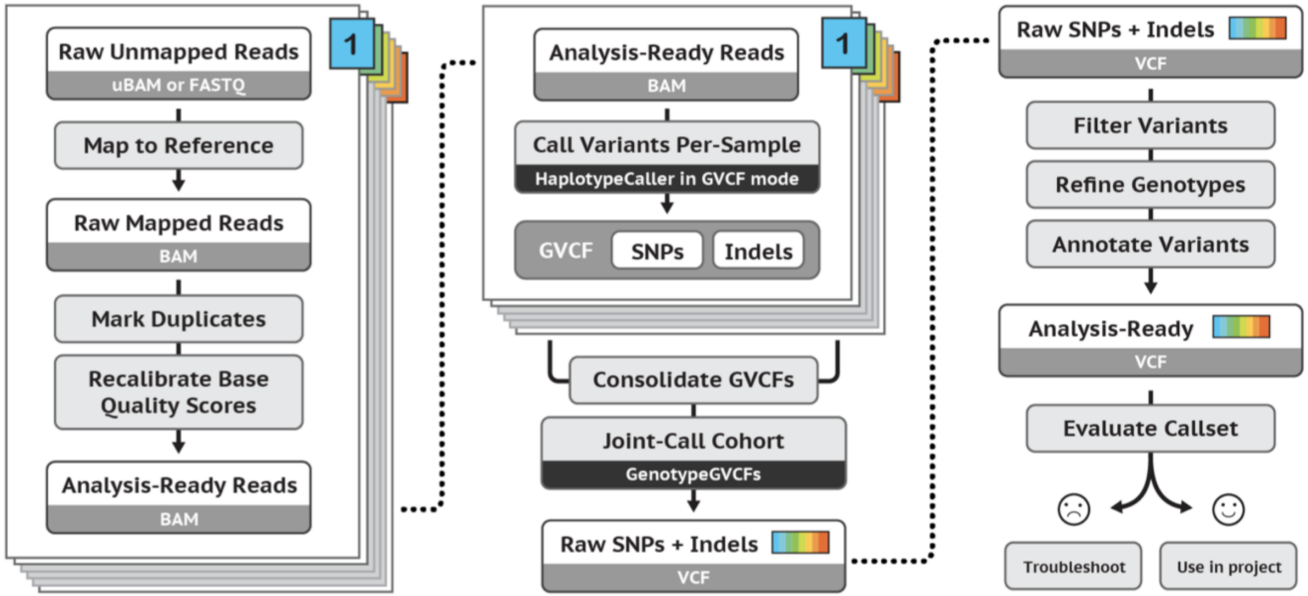


Figure: Germline short variant (SNVs + Indels) discovery workflow (Van der Auwera et al., 2013; DePristo et al., 2011)

RNA-seq short variant discovery

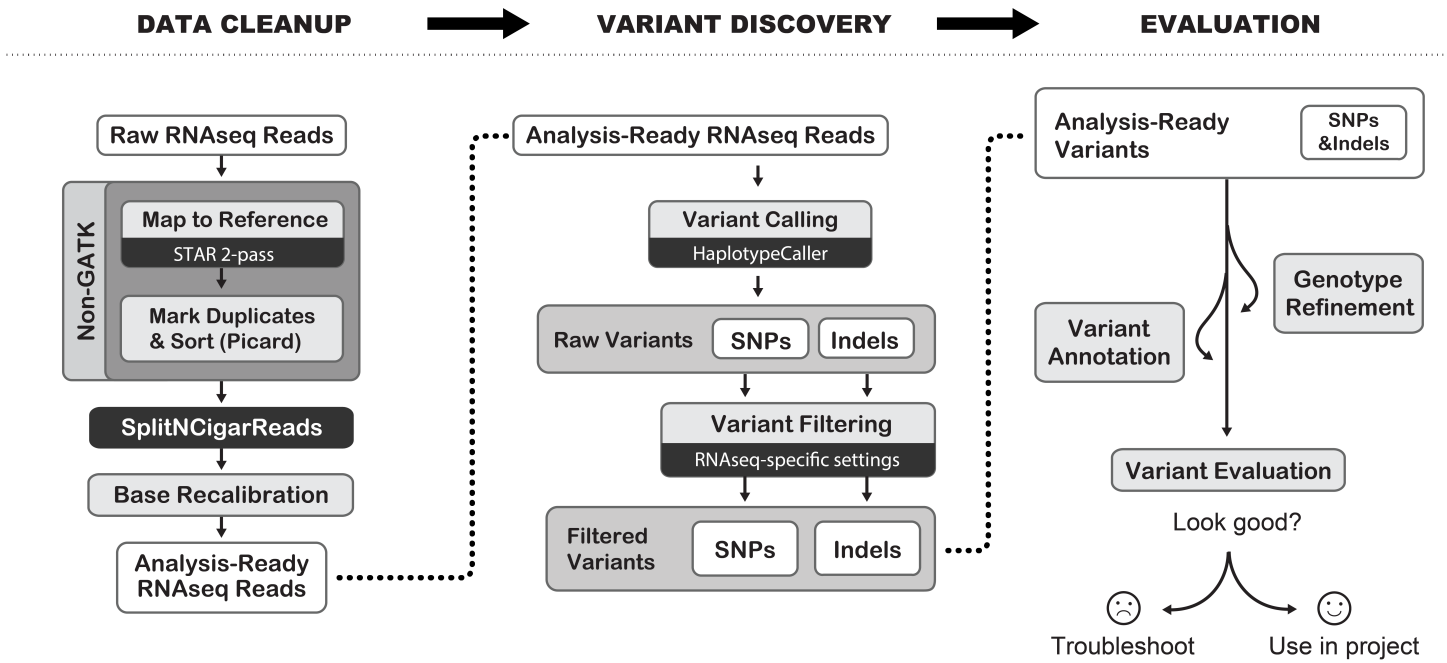


Figure: RNA-seq short variant (SNVs + Indels) discovery workflow (Van der Auwera et al., 2013; DePristo et al., 2011)

4. Results

4. Results

4.1. Quality Checks

FastQC?

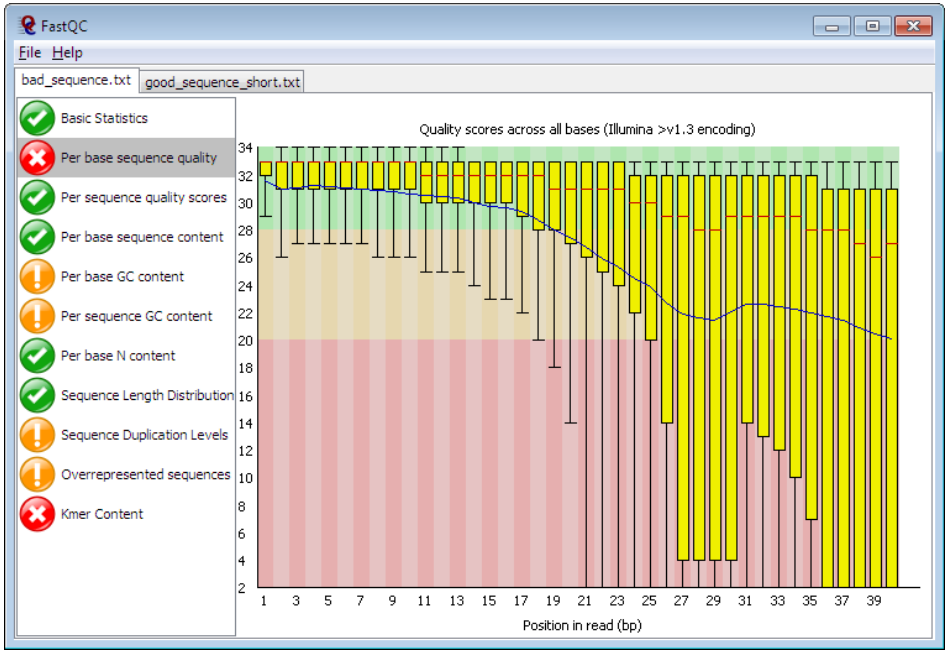


Figure: Example of FastQC Result (Andrews et al., 2012)

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

FastQC on WES

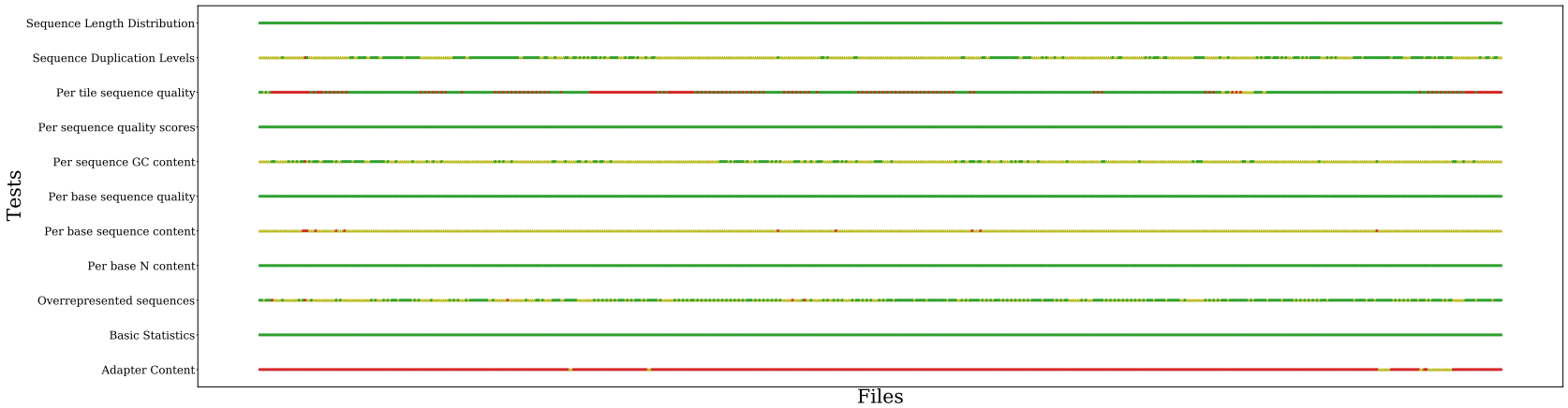
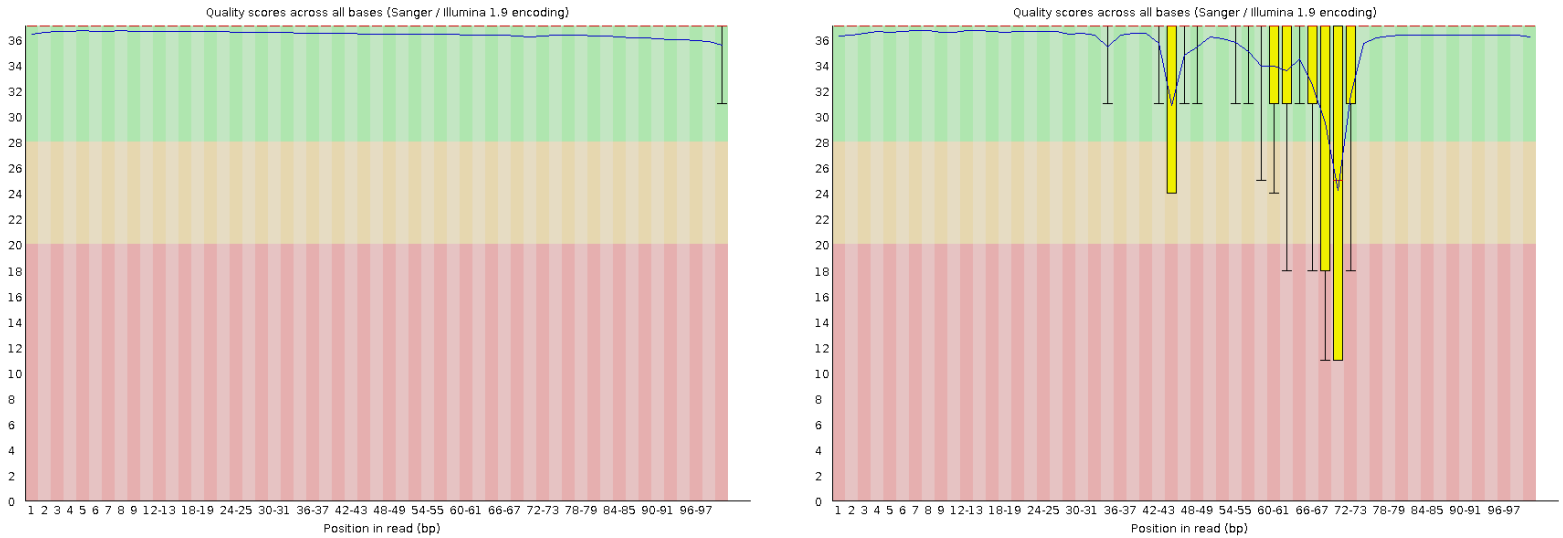


Figure: FastQC with WES data

Failure on 33P1 sample

33P1 is excluded at further analysis.

Failure on 33P1 I



(a) 33N (b) 33P1

Figure: Per Base Sequence Quality Results

Failure on 33P1 II

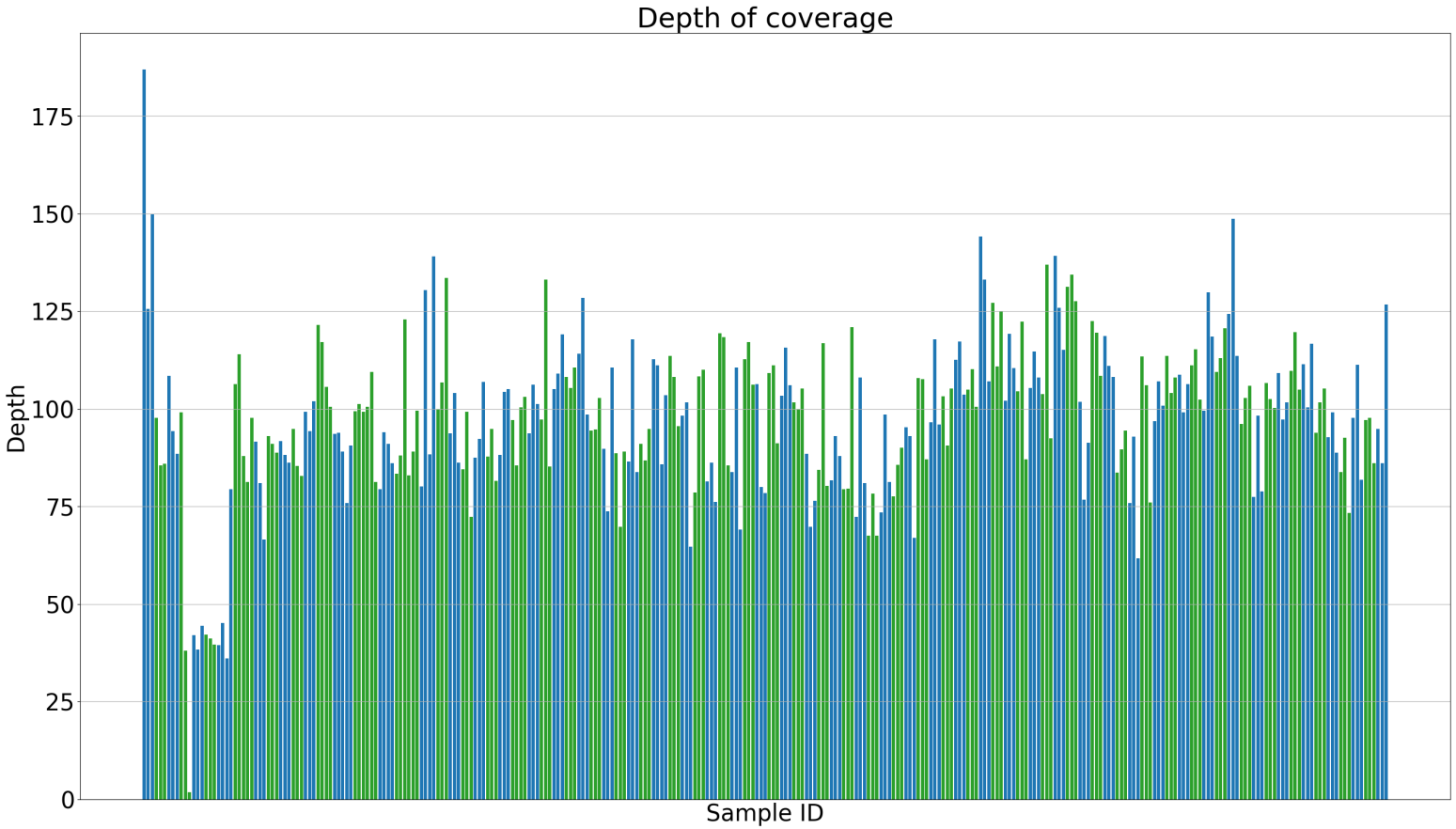


Figure: Coverage Depth Plot

FastQC on WTS

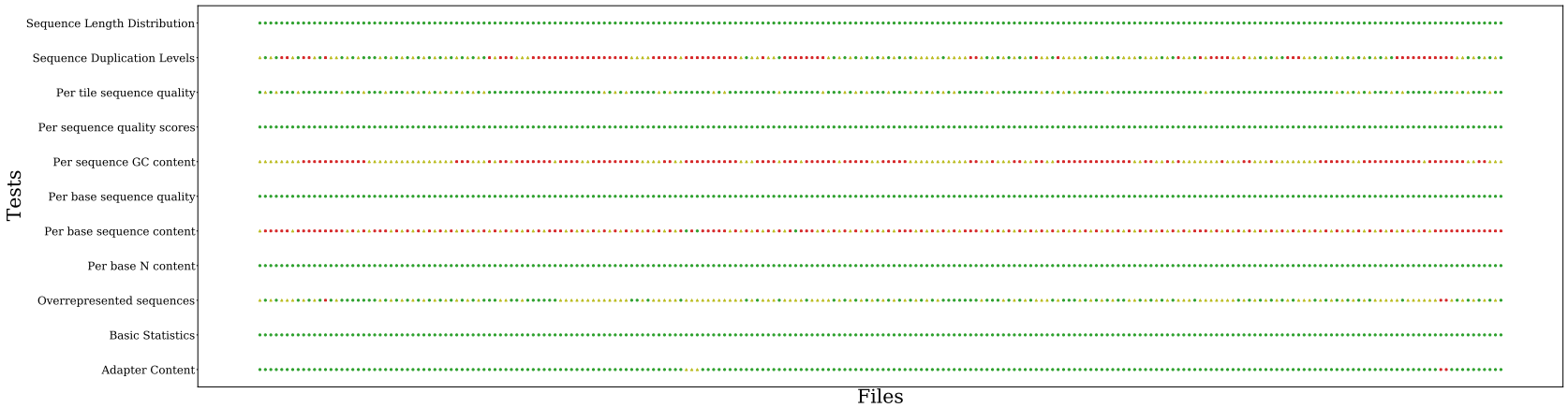


Figure: FastQC with WTS data

All sample are good to analysis

∵ No sample has more than 5 failures.

4. Results

4.2. Copy Number Variations

Sequenza?

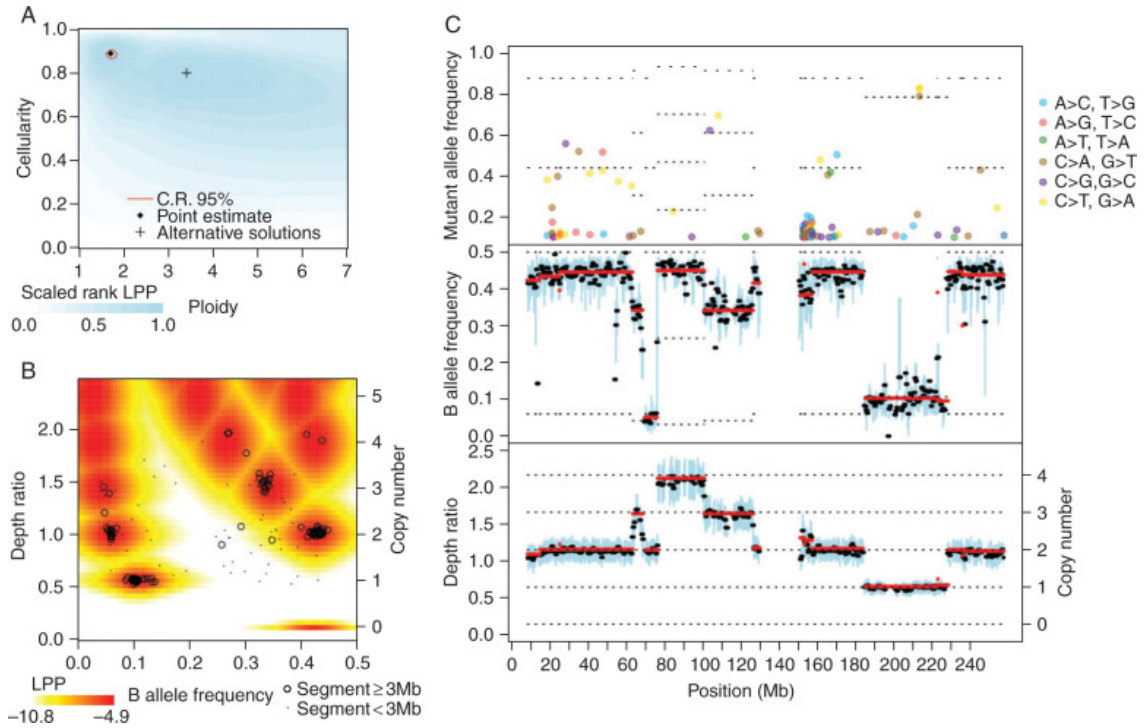
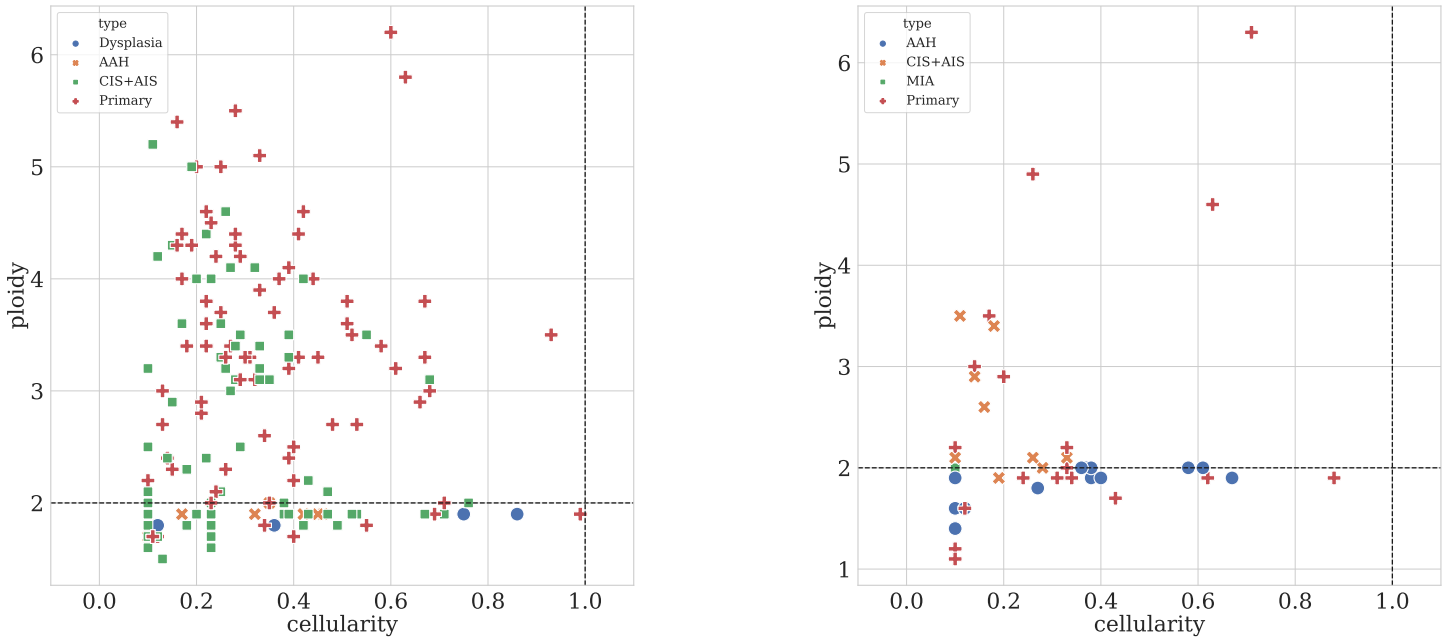


Figure: Representative Output of the Sequenza (Favero et al., 2015)

Cellularity & Ploidy on WES



(a) LUSC Samples (b) LUAD Samples

Figure: Cellularity and Ploidy from Sequenza

Genome View on Patient #57

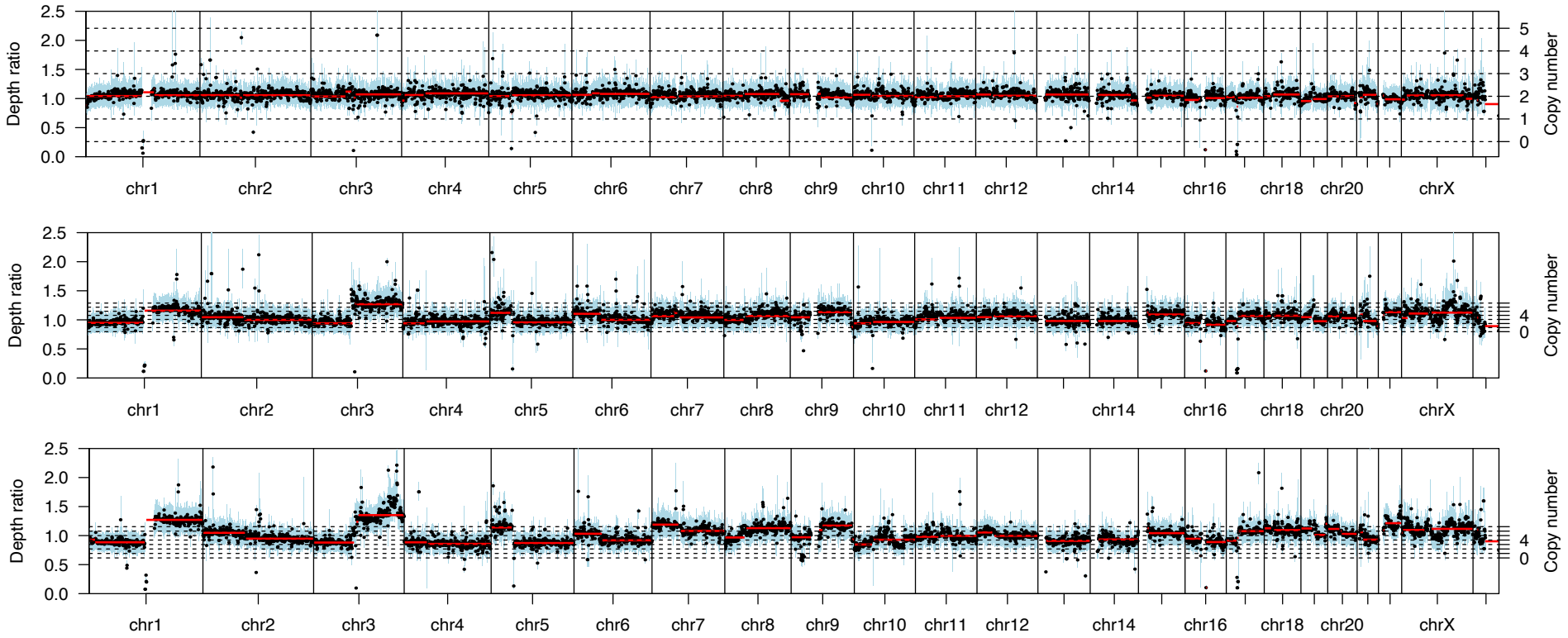


Figure: Dysplasia-CIS-Primary Tumor on Patient #57

CNVs of LUSC

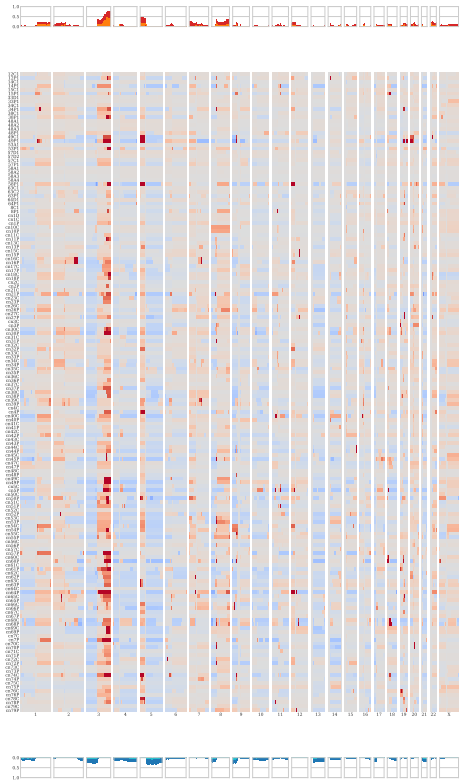


Figure: CNV Plot with LUSC Patients

CNVs of LUAD

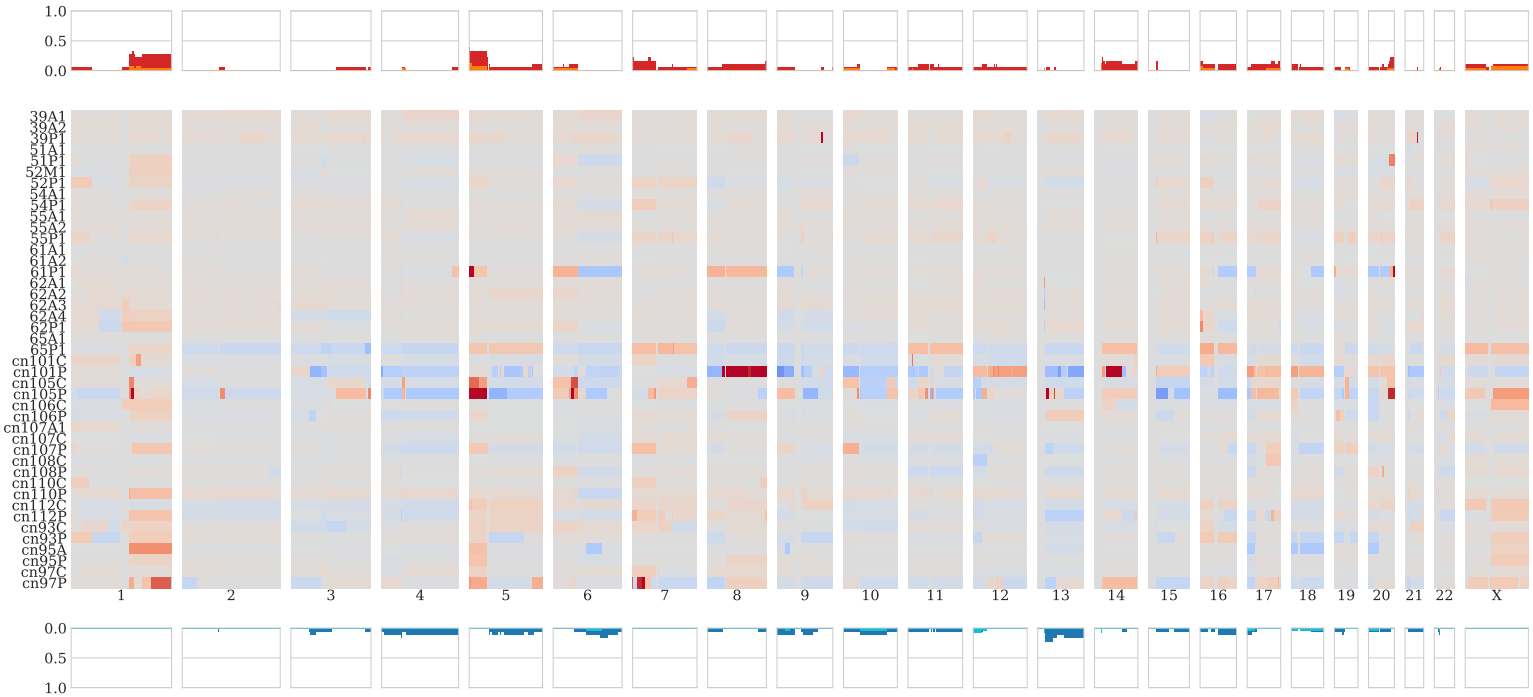


Figure: CNV Plot with LUAD Patients

LUSC vs. LUAD in CNV Plot

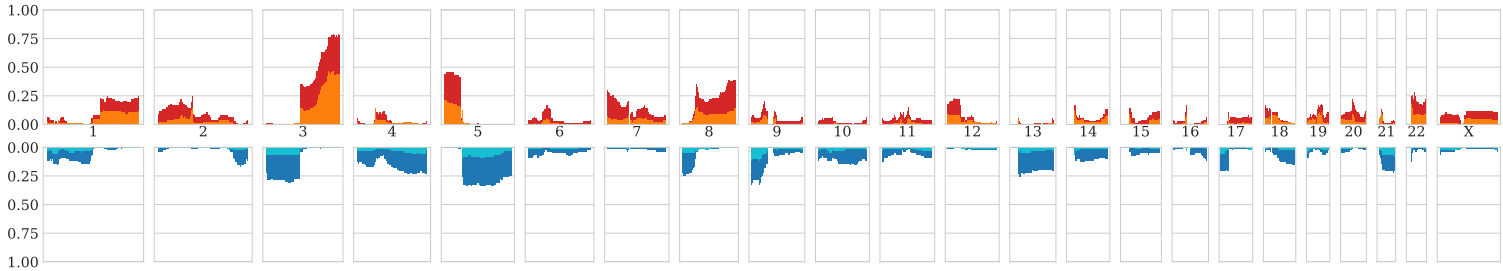


Figure: Simple CNV Plot with LUSC Patients

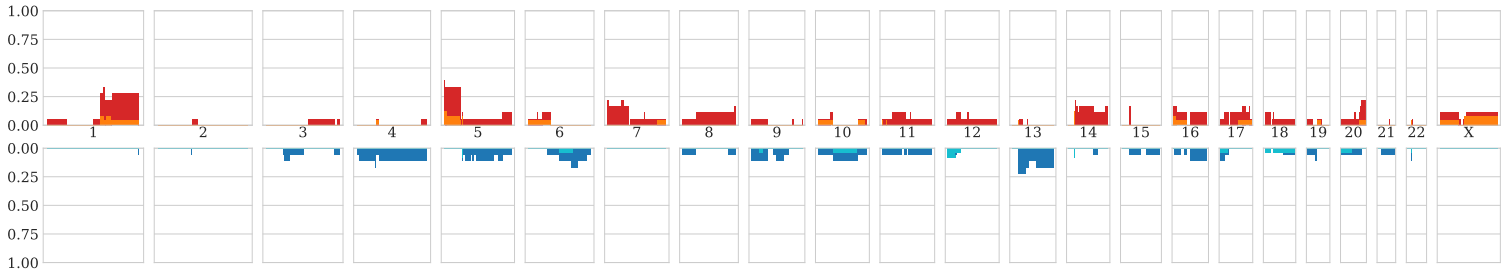


Figure: Simple CNV Plot with LUAD Patients

Findings in Sequenza

4. Results

4.3. SNVs Analysis

Mutect2?

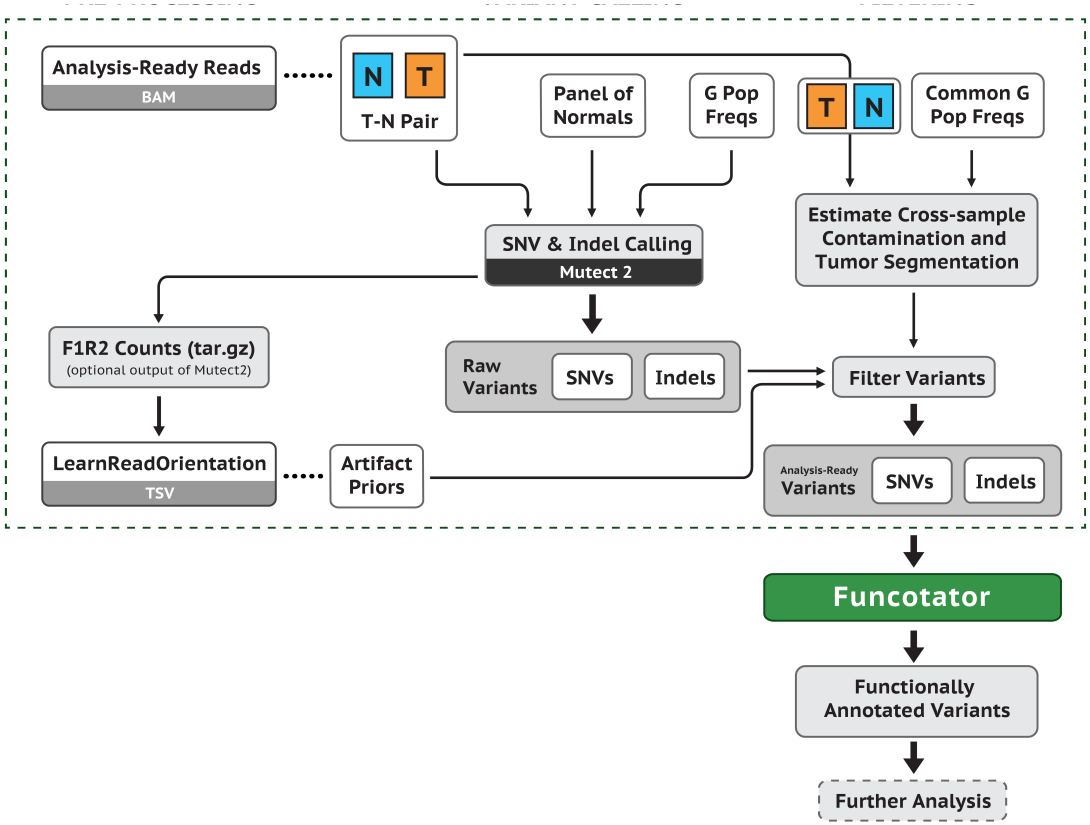


Figure: Somatic short variant discovery workflow (Van der Auwera et al., 2013; DePristo et al., 2011)

MutEnricher?

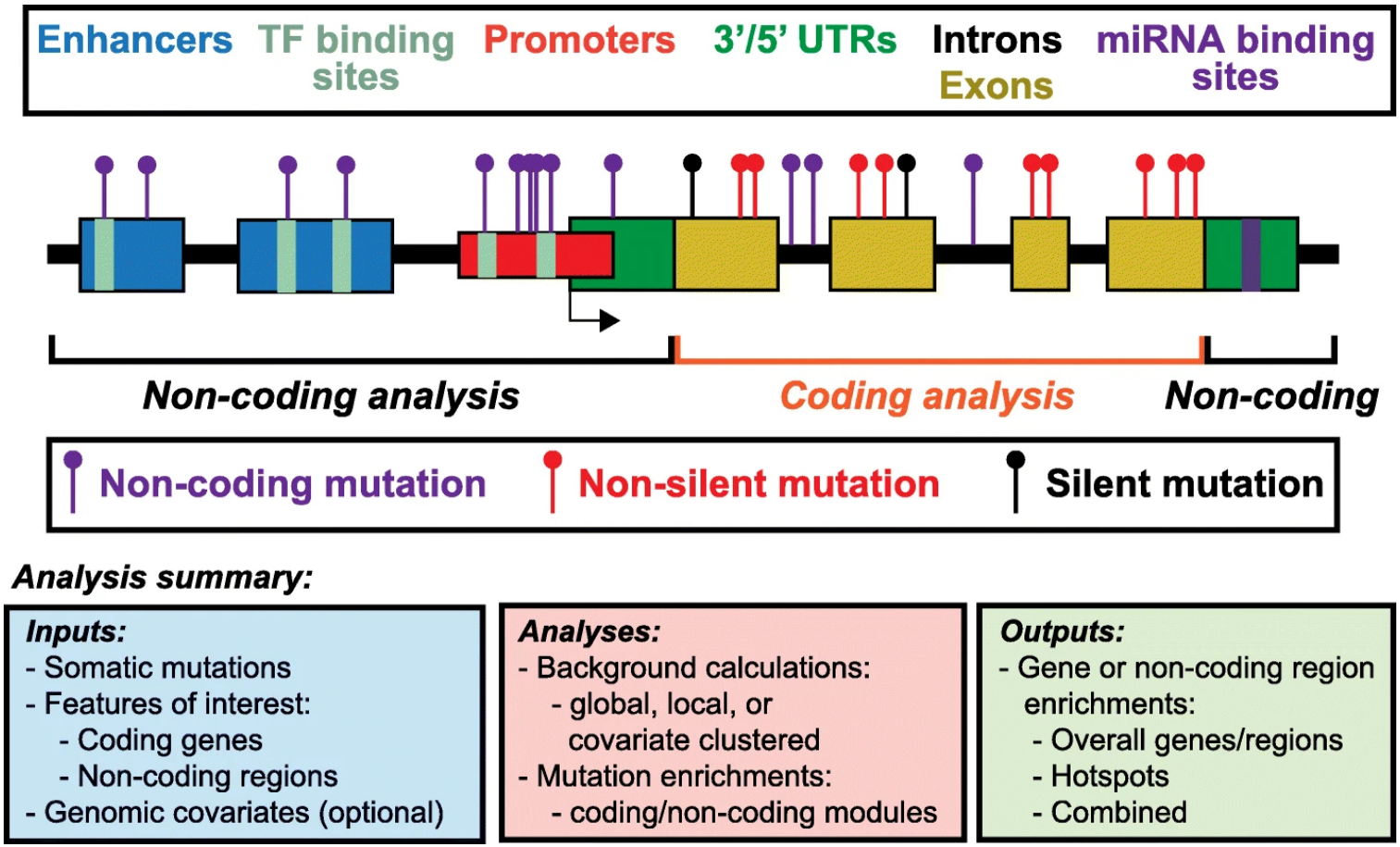


Figure: Schematic representation of MunEnricher's analysis procedures (Soltis et al., 2020)

Driver Gene Selection Strategy

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

Gene $\in$ CGC Tier 1 set

Fisher FDR

Fisher FDR < 0.05

Fisher P-value

Fisher P-value < 0.05

Gene P-value

Gene P-value < 0.05

Somatic Variant in LUSC

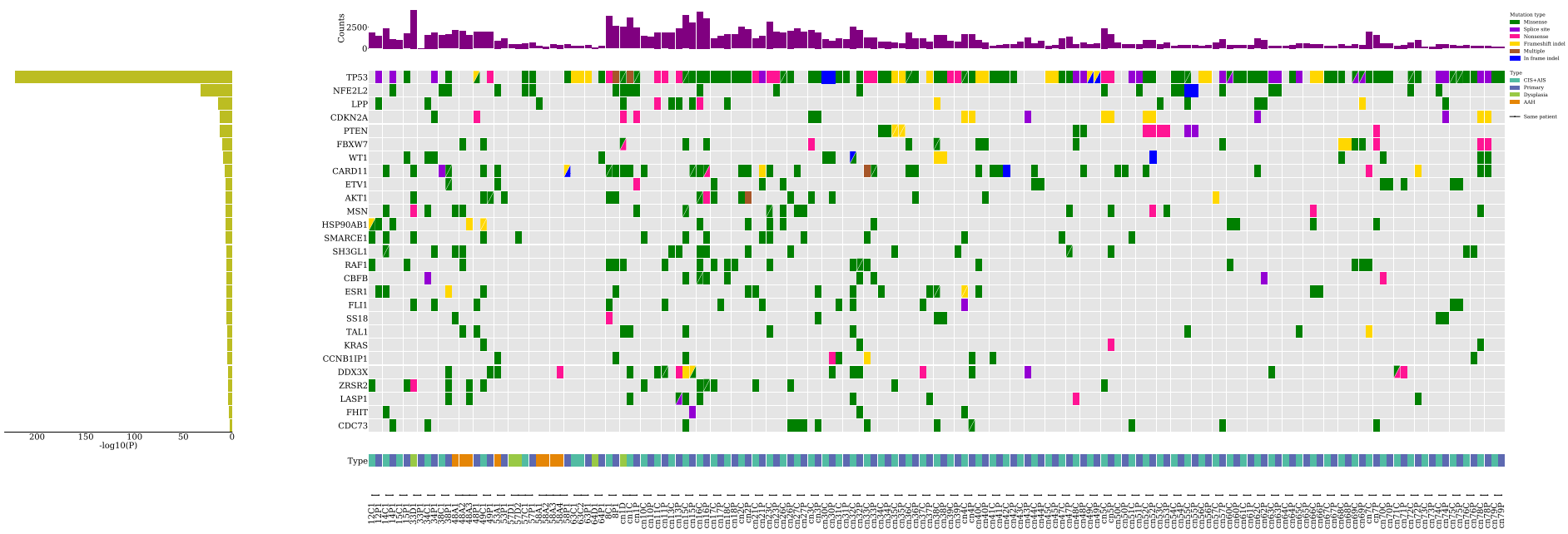


Figure: CoMut Plot with LUSC Patients

Somatic Variant in LUAD

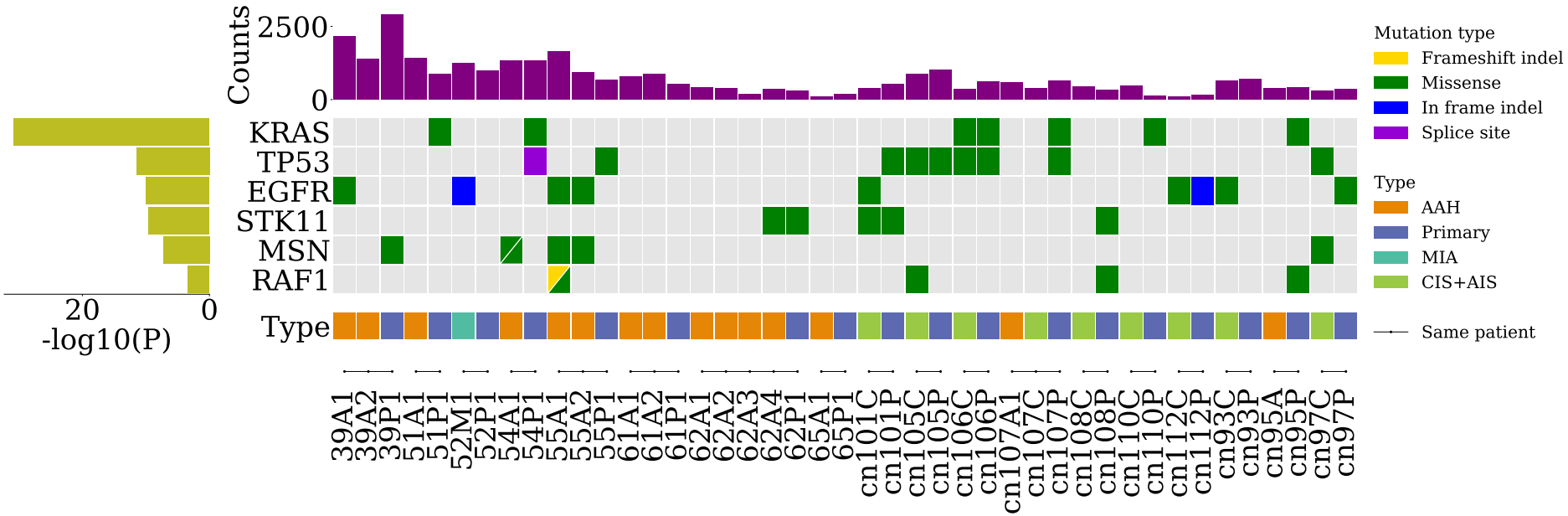


Figure: CoMut Plot with LUAD Patients

Findings in SNVs Analysis

4. Results

4.4. VAF Analysis

VAF?

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

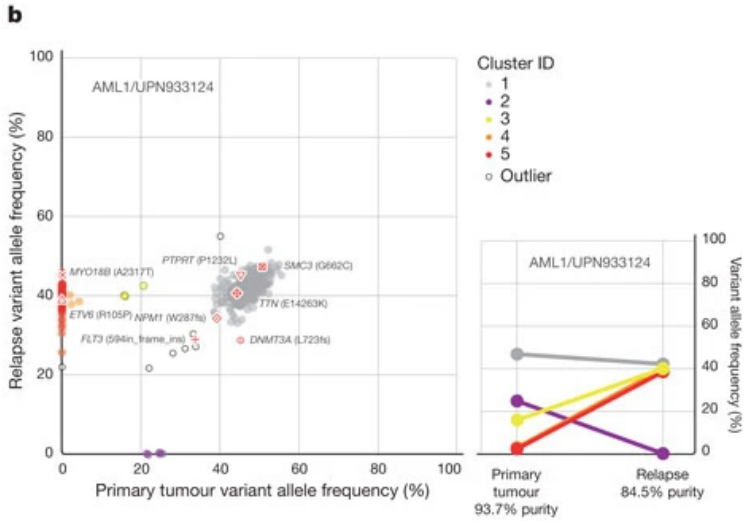


Figure: VAF distribution of validated mutations (Ding et al., 2012)

VAF Plots I

PyClone?

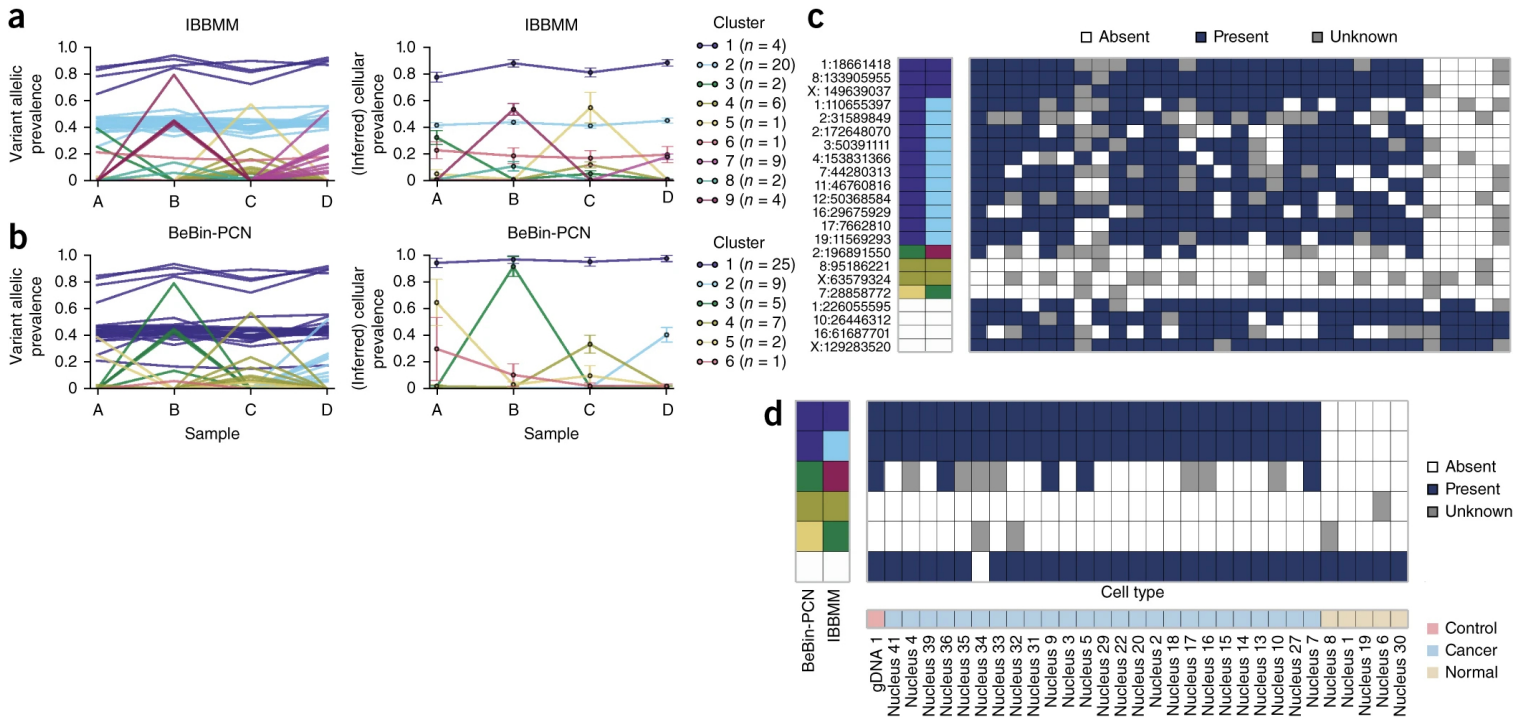


Figure: Analysis of multiple samples by PyClone (Roth et al., 2014)

PyClone Plots I

Findings in VAF Analysis

4. Results

4.5. Tumor Evolution Trajectories Analysis

Revolver?

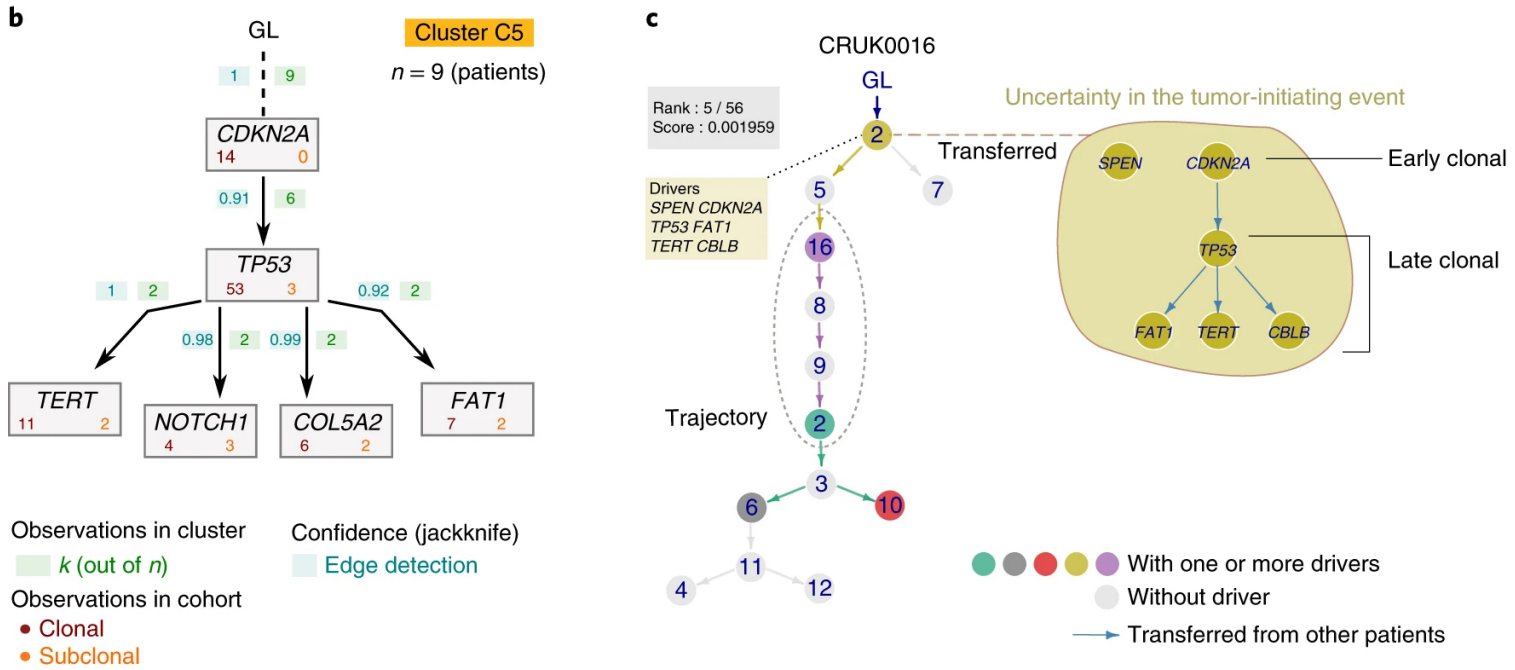


Figure: Repeated Evolutionary Trajectories (Caravagna et al., 2018)

Findings in Tumor Evolution Trajectories Analysis

4. Results

4.6. Differences in Gene Expression Levels

RSEM?

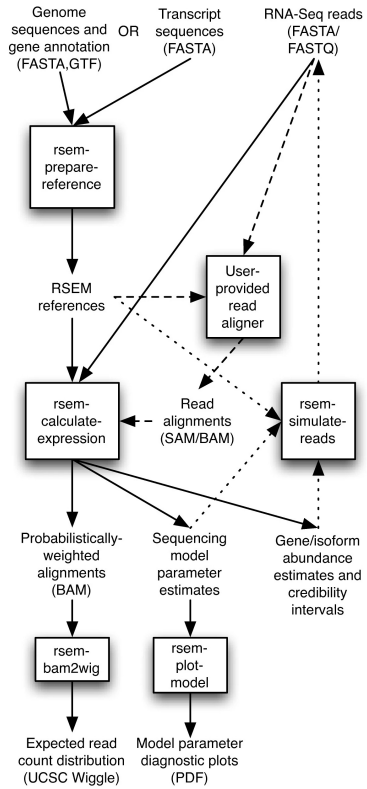


Figure: RSEM workflow (Li & Dewey, 2011)

DESeq2?

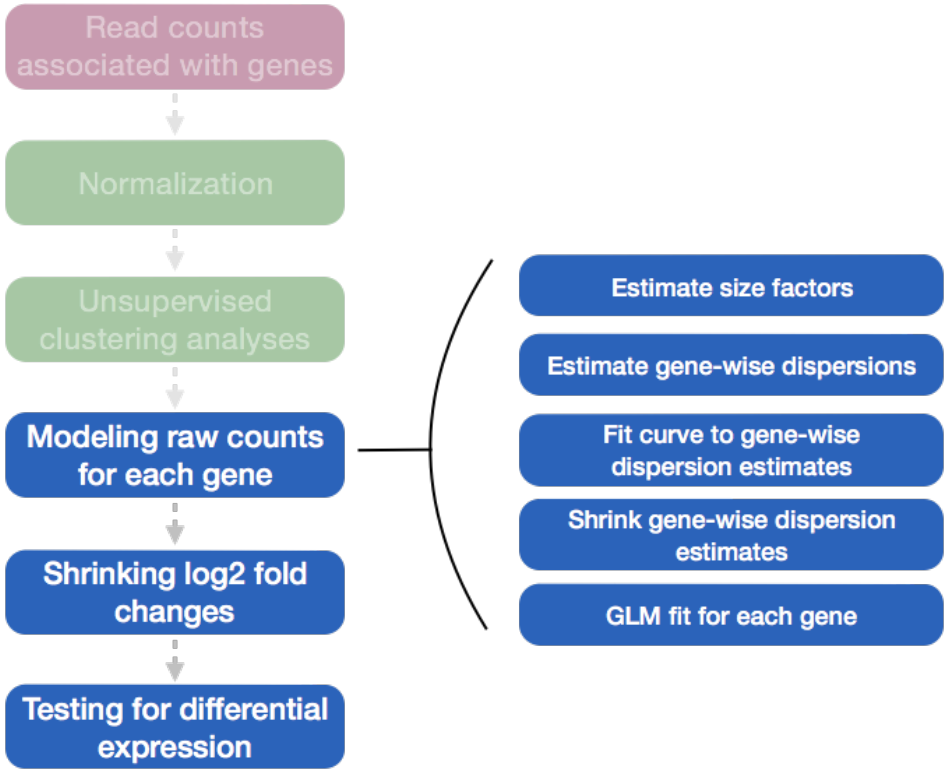


Figure: DESeq2 workflow (Love, Huber, & Anders, 2014)

DEG Selection Strategy

DEG: differentially expressed genes

Fold Change

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

P-value

P-value < 0.05

Adjusted P-value

Padj < 0.05

Enrichr?

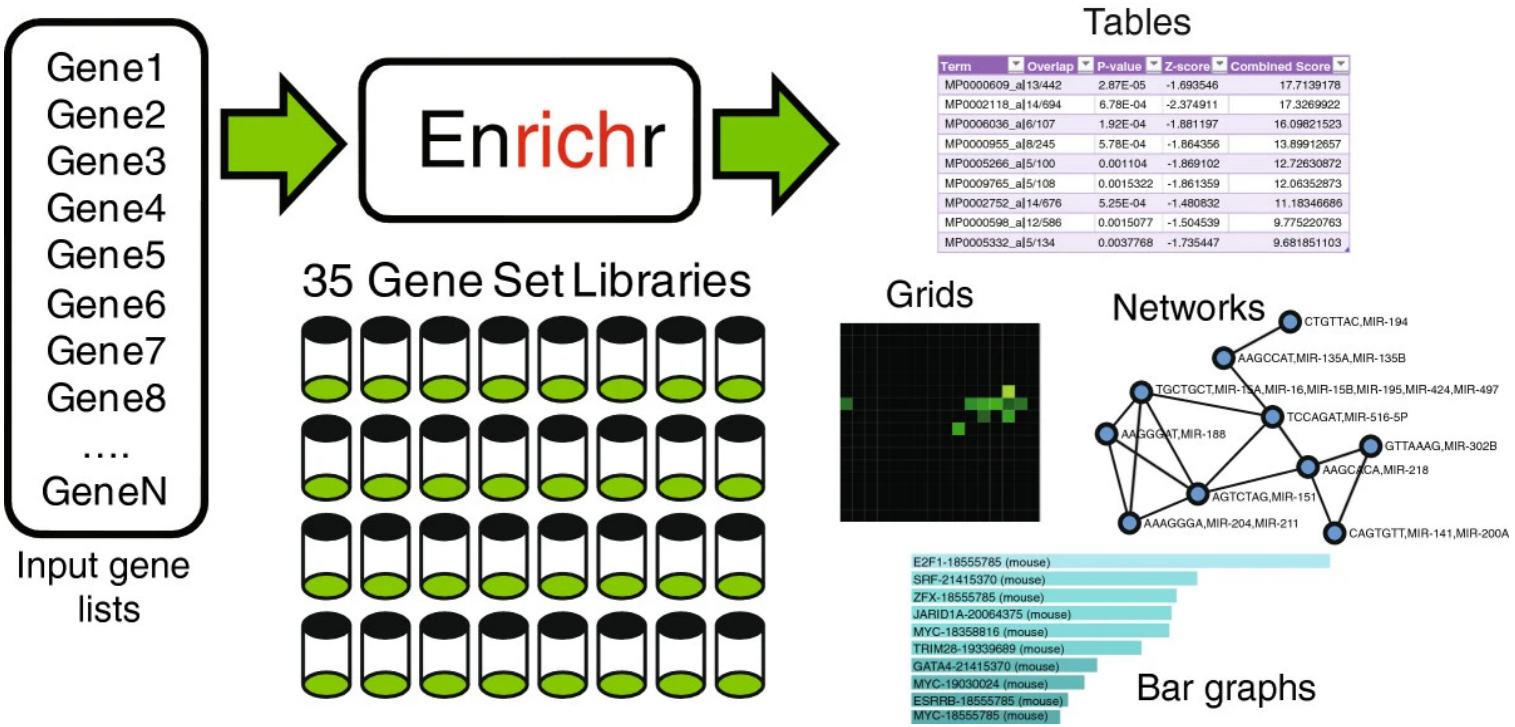


Figure: Enrichr workflow (Chen et al., 2013; Kuleshov et al., 2016)

Gene-set Library

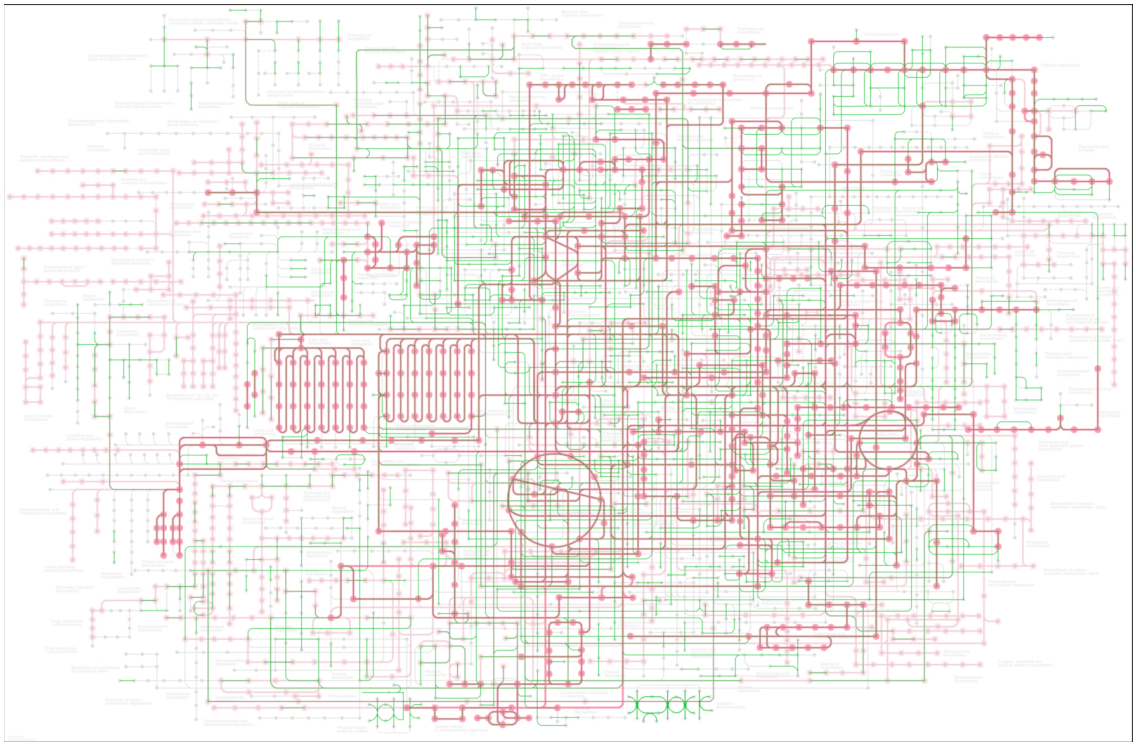


Figure: The global map of metabolic pathways by KEGG (Kanehisa et al., 2021)

KEGG

KEGG 2021 Human

WTS Data Composition I

Table: Number of WTS samples

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

WTS Data Composition II

Table: Number of WTS LUSC samples

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

WTS Data Composition III

Table: Number of WTS LUAD samples

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

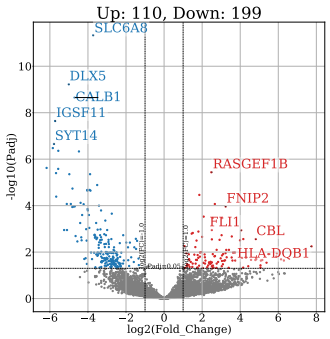
4. Results

4.6. Differences in Gene Expression Levels

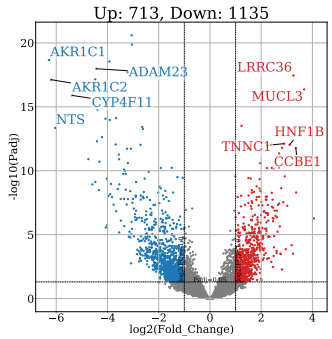
4.6.1. Comparing cancer stage in LUSC

DEG Volcano Plots in LUSC

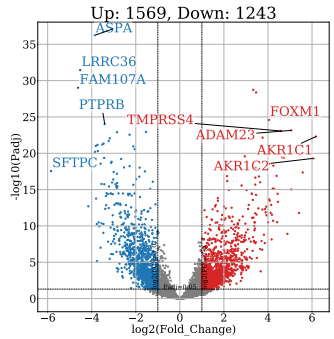
Normal → Dysplasia → CIS → Primary (LUSC)



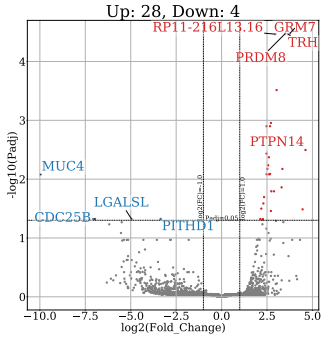
(a) Normal-Dysplasia



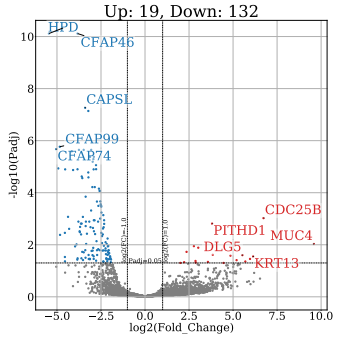
(b) Normal-CIS



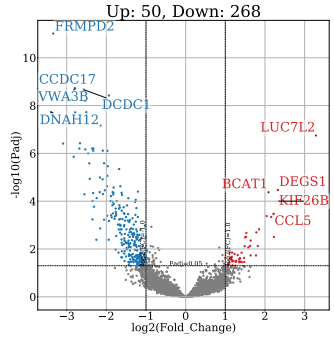
(c) Normal-Primary



(d) Dysplasia-CIS



(e) Dysplasia-Primary

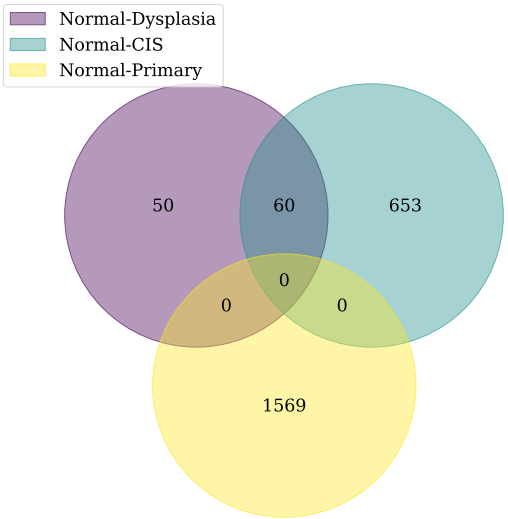


(f) CIS-Primary

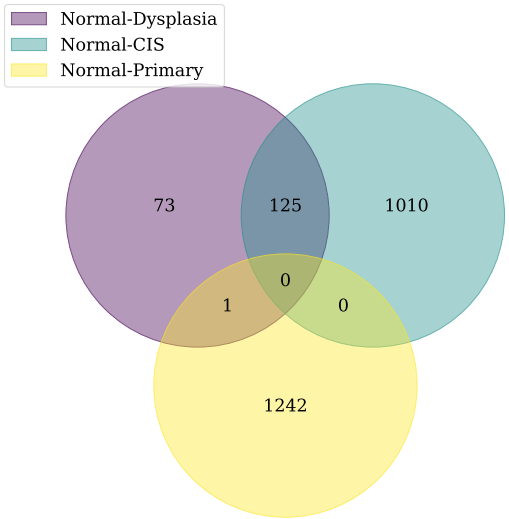
Figure: DEG Volcano Plots in LUSC

DEG Venn Diagram in LUSC

Normal → Dysplasia → CIS → Primary (LUSC)



(a) Up-regulated



(b) Down-regulated

Figure: DEG Venn Diagram in LUSC

Enrichment test with Normal vs. Dysplasia in LUSC

Table: Up-regulated Pathways on Normal vs. Dysplasia

Term name	Adjusted p-value
Leishmaniasis	6.72e-03
Lysosome	6.72e-03
Phagosome	1.15e-02

Table: Down-regulated Pathways on Normal vs. Dysplasia

Term name	Adjusted p-value
NaN	NaN

Enrichment test with Normal vs. CIS in LUSC

Table: Up-regulated Pathways on Normal vs. CIS

Term name	Adjusted p-value
Hematopoietic cell lineage	7.22e-08
Malaria	1.16e-06
Cell adhesion molecules	1.16e-06

Table: Down-regulated Pathways on Normal vs. CIS

Term name	Adjusted p-value
Metabolism of xenobiotics by cytochrome P450	9.34e-06
Drug metabolism	9.06e-05
Cell cycle	1.68e-04

Enrichment test with Normal vs. Primary in LUSC

Table: Up-regulated Pathways on Normal vs. Primary

Term name	Adjusted p-value
Cell cycle	1.53e-04
Glutathione metabolism	1.53e-04
DNA replication	1.72e-04

Table: Down-regulated Pathways on Normal vs. Primary

Term name	Adjusted p-value
Hematopoietic cell lineage	7.33e-09
Malaria	7.33e-09
Hypertrophic cardiomyopathy	1.24e-08

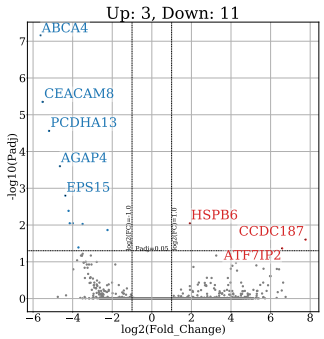
4. Results

4.6. Differences in Gene Expression Levels

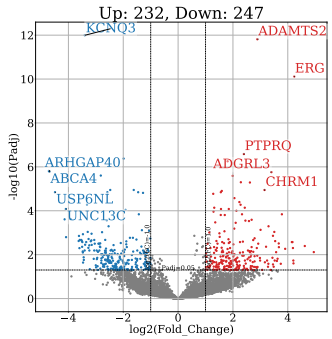
4.6.2. Comparing cancer stage in LUAD

DEG Volcano Plots in LUAD

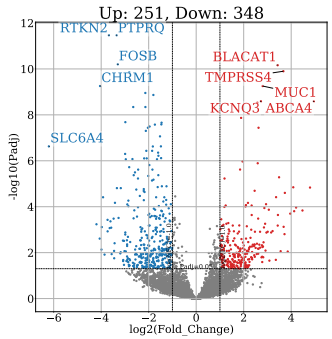
Normal → AAH → AIS → Primary (LUAD)



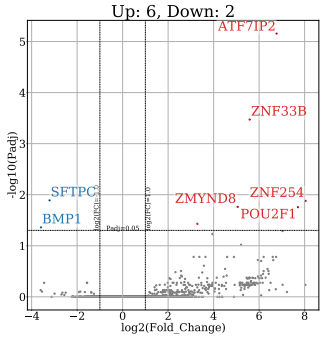
(a) Normal-AAH



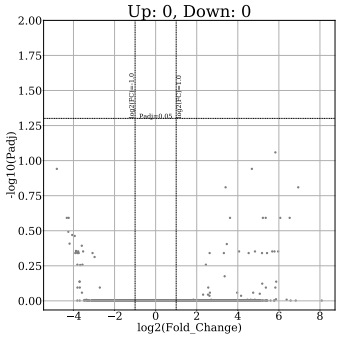
(b) Normal-AIS



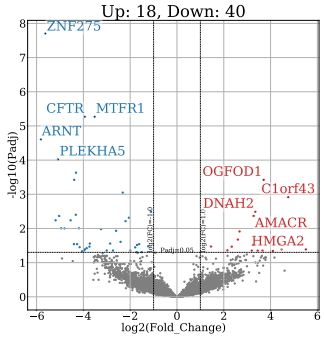
(c) Normal-Primary



(d) AAH-AIS



(e) AAH-Primary

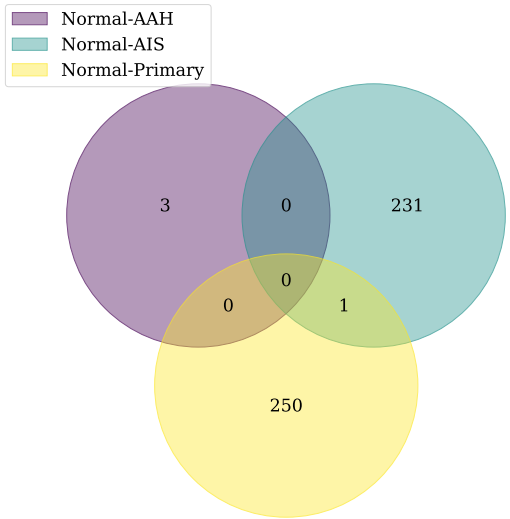


(f) AIS-Primary

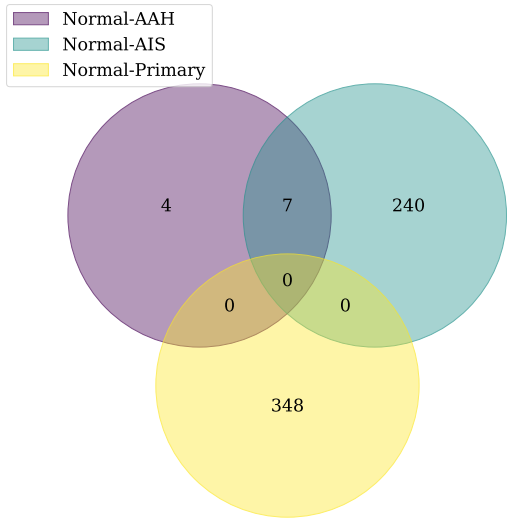
Figure: DEG Volcano Plots in LUAD

DEG Venn Diagram in LUAD

Normal $\rightarrow$ AAH $\rightarrow$ AIS $\rightarrow$ Primary (LUAD)



(a) Up-regulated



(b) Down-regulated

Figure: DEG Venn Diagram in LUAD

Enrichment test with Normal vs. AAH in LUAD

Table: Up-regulated Pathways on Normal vs. AAH

Term name	Adjusted p-value
NaN	NaN

Table: Down-regulated Pathways on Normal vs. AAH

Term name	Adjusted p-value
NaN	NaN

Enrichment test with Normal vs. AIS in LUAD

Table: Up-regulated Pathways on Normal vs. AIS

Term name	Adjusted p-value
Calcium signaling pathway	2.49e-02
Cell adhesion molecules	3.55e-02

Table: Down-regulated Pathways on Normal vs. AIS

Term name	Adjusted p-value
NaN	NaN

Enrichment test with Normal vs. Primary in LUAD

Table: Up-regulated Pathways on Normal vs. Primary

Term name	Adjusted p-value
NaN	NaN

Table: Down-regulated Pathways on Normal vs. Primary

Term name	Adjusted p-value
Vascular smooth muscle contraction	1.38e-04
ECM-receptor interaction	3.58e-04
Calcium signaling pathway	4.03e-04

4. Results

4.6. Differences in Gene Expression Levels

4.6.3. Recur vs. Non-recur in LUSC

LUSC Data Composition

Table: Number of WTS LUSC samples

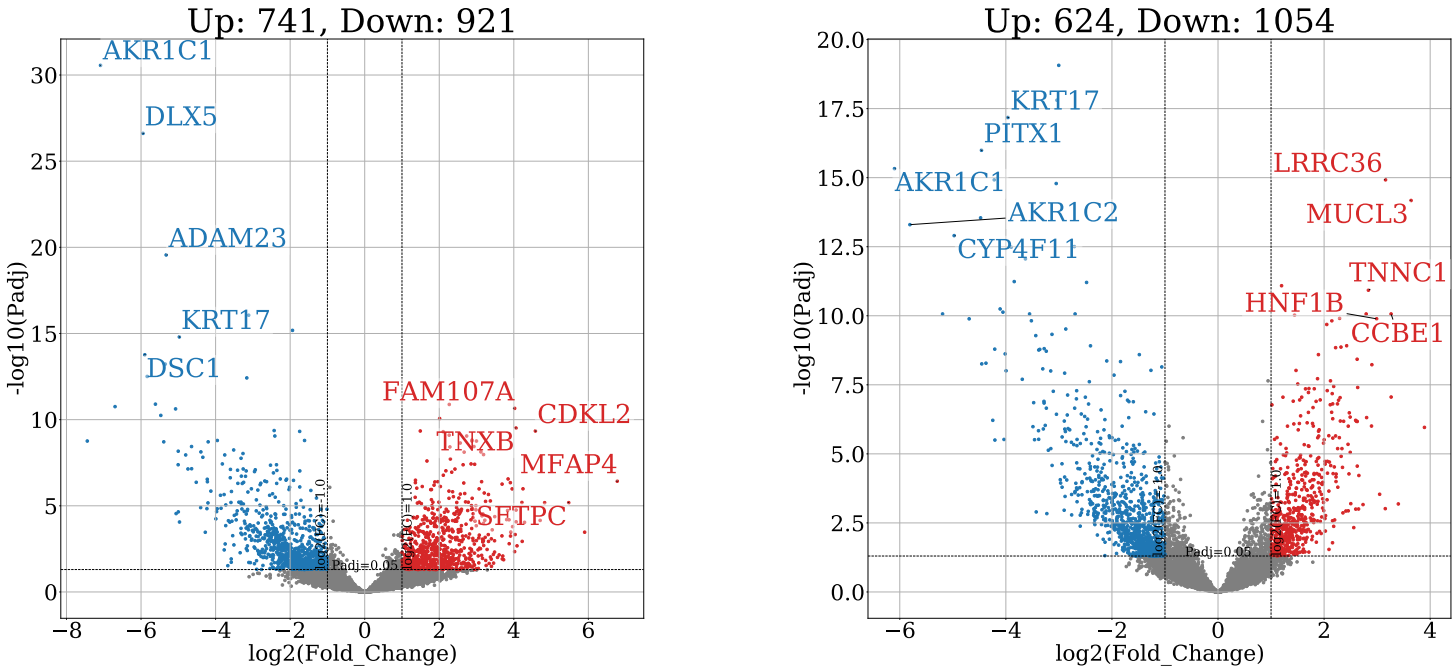
Recurrence?	Stage	Number of Samples
Recurrence (n=13)	Normal	1
	Dysplasia	1
	CIS	5
	Primary	6
Non-recurrence (n=74)	Normal	16
	Dysplasia	1
	CIS	28
	Primary	29

Normal samples

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

$\because$ Insufficient number of Normal samples in Recur.

DEG Volcano Plots Recur vs. Non-recur with CIS in LUSC

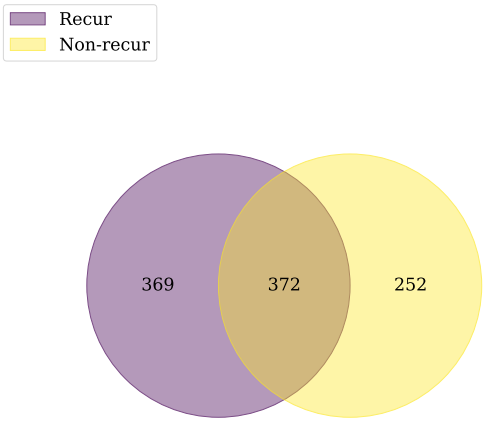


(a) Recurrence

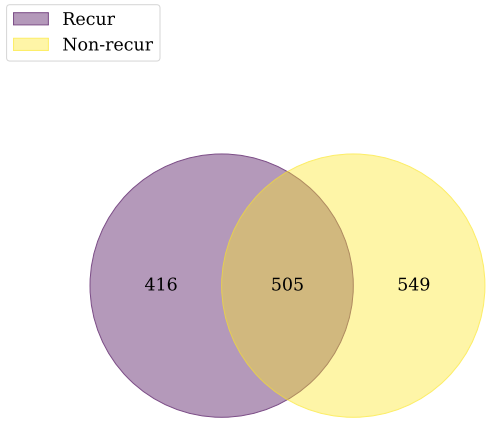
(b) Non-recurrence

Figure: DEG Volcano Plots Recur vs. Non-recur with CIS

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



(a) Up-regulated



(b) Down-regulated

Figure: DEG Venn Diagram Recur vs. Non-recur with CIS

Enrichment test for Recur with CIS in LUSC

Table: Up-regulated Pathways on Recur with CIS in LUSC

Term name	Adjusted p-value
NaN	NaN

Table: Down-regulated Pathways on Recur with CIS in LUSC

Term name	Adjusted p-value
Huntington disease	6.36e-06
Amyotrophic lateral sclerosis	1.62e-05
Parkinson disease	1.62e-05

Enrichment test for Non-recur with CIS in LUSC

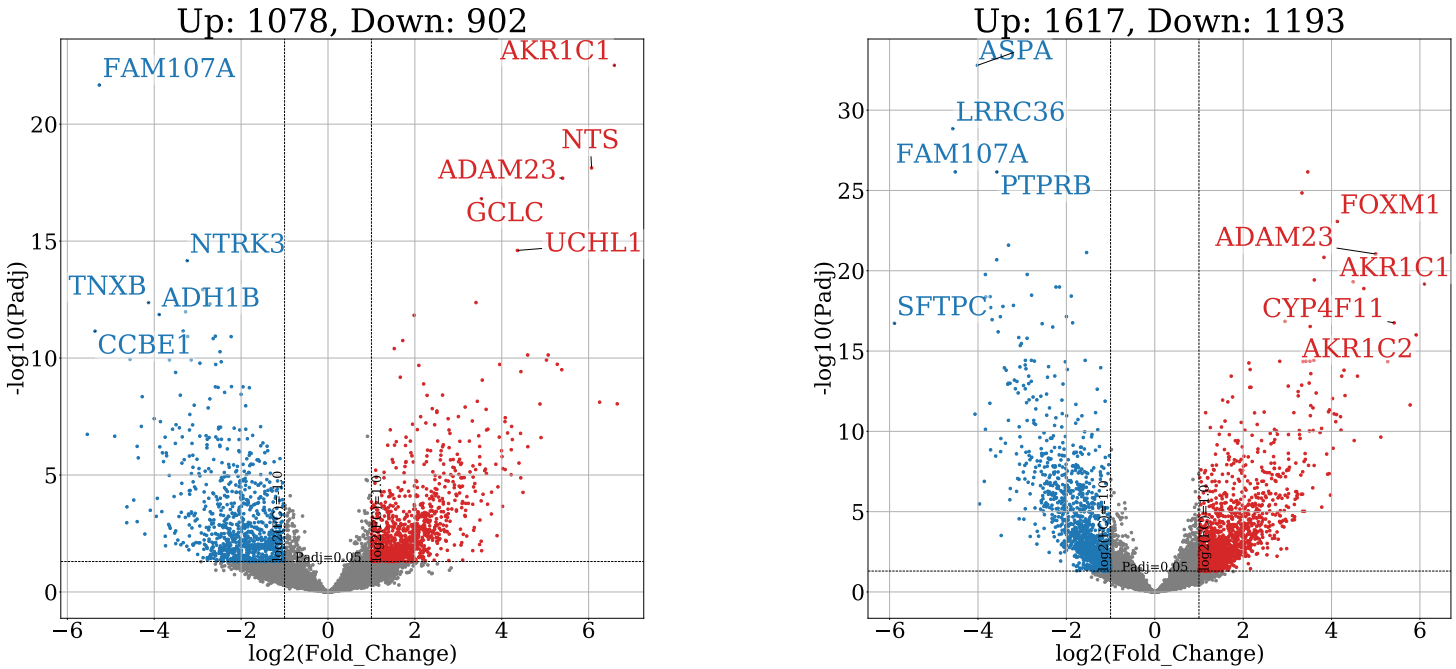
Table: Up-regulated Pathways on Non-recur with CIS in LUSC

Term name	Adjusted p-value
Malaria	7.76e-03
Th1 and Th2 cell differentiation	1.15e-02
Transcriptional misregulation in cancer	1.15e-02

Table: Down-regulated Pathways on Non-recur with CIS in LUSC

Term name	Adjusted p-value
NaN	NaN

DEG Volcano Plots R vs. NR with Primary in LUSC

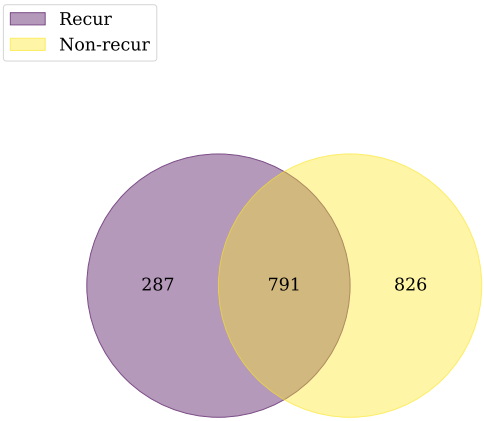


(a) Recurrence

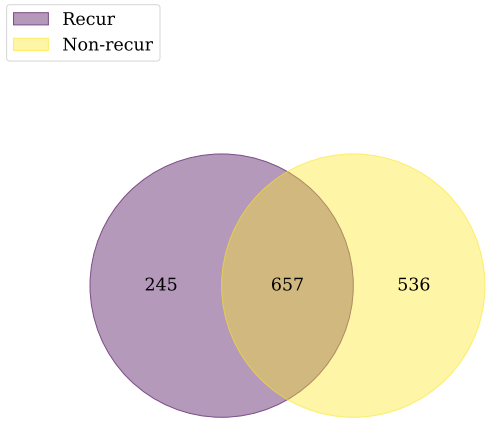
(b) Non-recurrence

Figure: DEG Volcano Plots Recur vs. Non-recur with Primary

DEG Venn Diagram R vs. NR with Primary in LUSC



(a) Up-regulated



(b) Down-regulated

Figure: DEG Venn Diagram Recur vs. Non-recur with Primary

Enrichment test for Recur with Primary in LUSC

Table: Up-regulated Pathways on Recur with Primary in LUSC

Term name	Adjusted p-value
Amyotrophic lateral sclerosis	4.85e-03
RNA transport	6.11e-03
mRNA surveillance pathway	6.11e-03

Table: Down-regulated Pathways on Recur with Primary in LUSC

Term name	Adjusted p-value
NaN	NaN

Enrichment test for Non-recur with Primary in LUSC

Table: Up-regulated Pathways on Non-recur with Primary in LUSC

Term name	Adjusted p-value
Homologous recombination	1.00e-02

Table: Down-regulated Pathways on Non-recur with Primary in LUSC

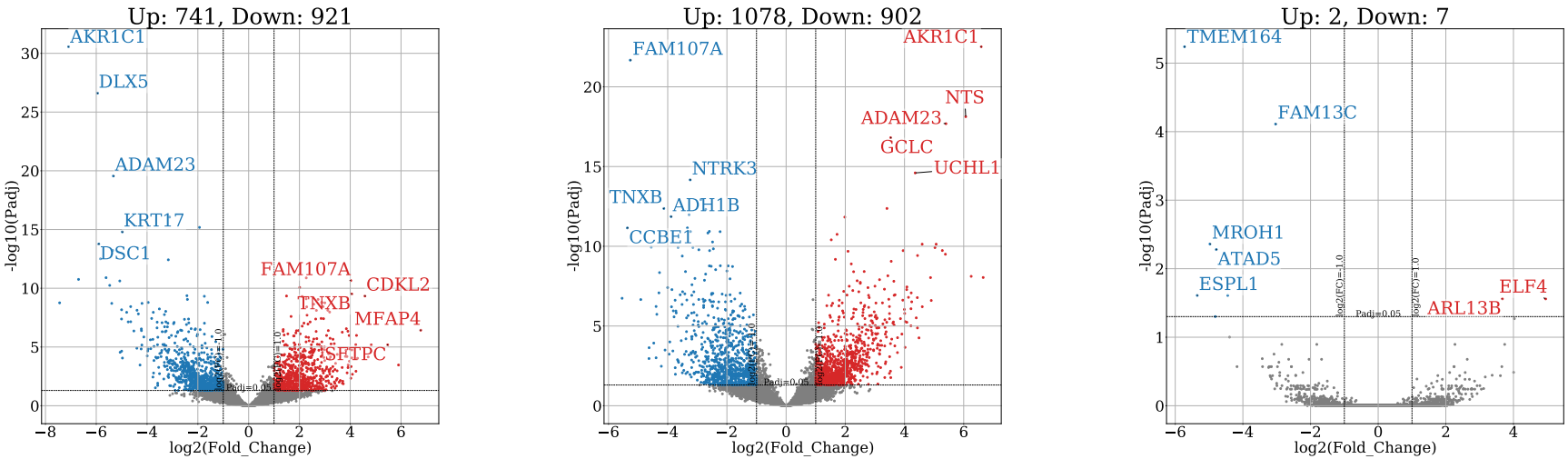
Term name	Adjusted p-value
Staphylococcus aureus infection	5.37e-05
Hematopoietic cell lineage	5.37e-05
Leishmaniasis	4.30e-04

4. Results

4.6. Differences in Gene Expression Levels

4.6.4. Within Recur in LUSC

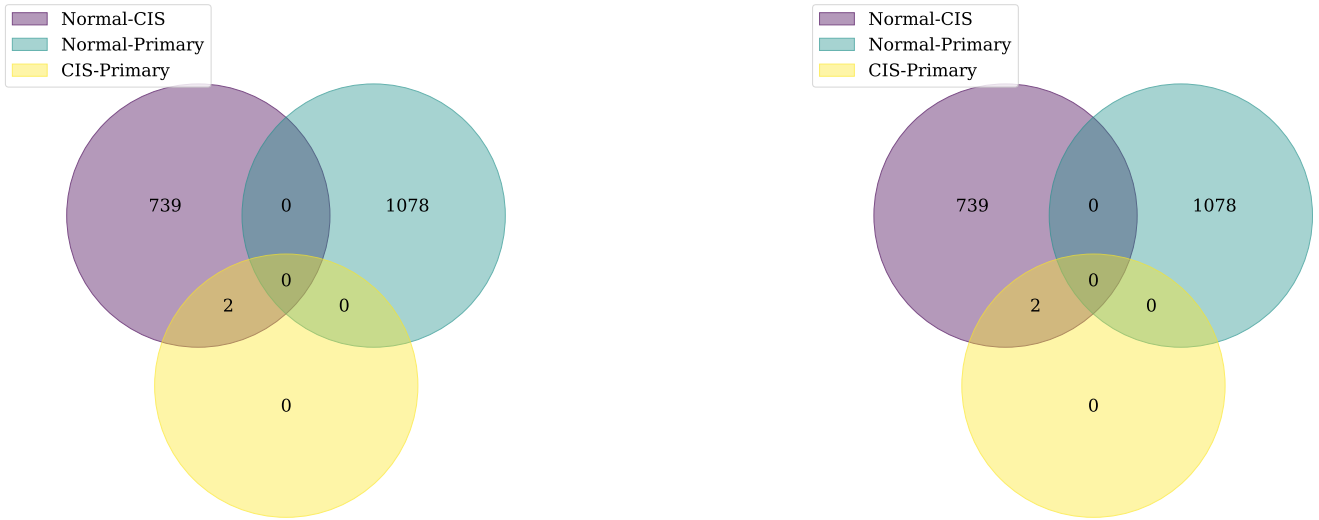
DEG Volcano Plots with Recur in LUSC



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

Figure: DEG Volcano Plots with Recur samples in LUSC

DEG Venn Diagram with Recur in LUSC



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

Figure: DEG Venn Diagram with Recur samples in LUSC

Enrichment test with Normal vs. CIS for Recur

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

Term name	Adjusted p-value
Hematopoietic cell lineage	1.87e-05
Cell adhesion molecules	1.87e-05
Hypertrophic cardiomyopathy	9.66e-05

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

Term name	Adjusted p-value
Parkinson disease	2.11e-05
Alzheimer disease	2.11e-05
Huntington disease	2.11e-05

Enrichment test with Normal vs. Primary for Recur

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

Term name	Adjusted p-value
Glycolysis / Gluconeogenesis	1.90e-05
RNA transport	2.66e-05
Drug metabolism	2.66e-05

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

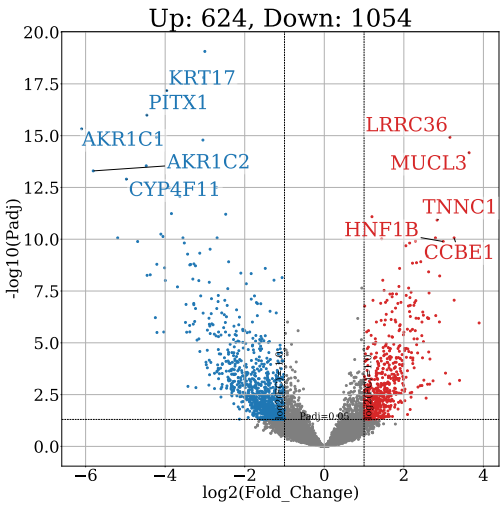
Term name	Adjusted p-value
Dilated cardiomyopathy	2.19e-06
Hypertrophic cardiomyopathy	2.19e-06
Arrhythmogenic right ventricular cardiomyopathy	4.12e-06

4. Results

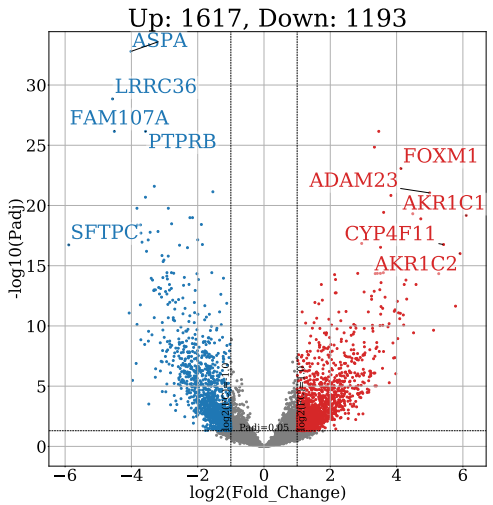
4.6. Differences in Gene Expression Levels

4.6.5. Within Non-recur in LUSC

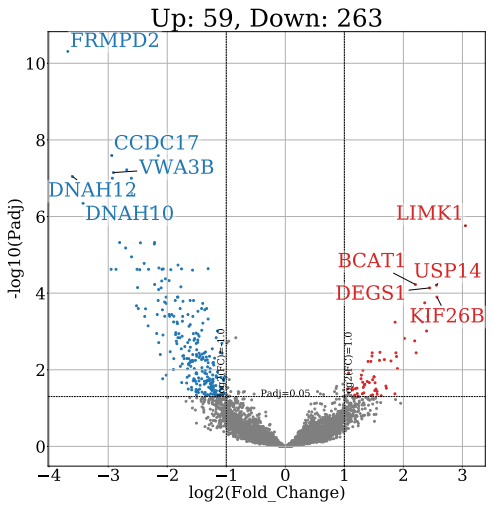
DEG Volcano Plots with Non-recur in LUSC



(a) Normal-CIS



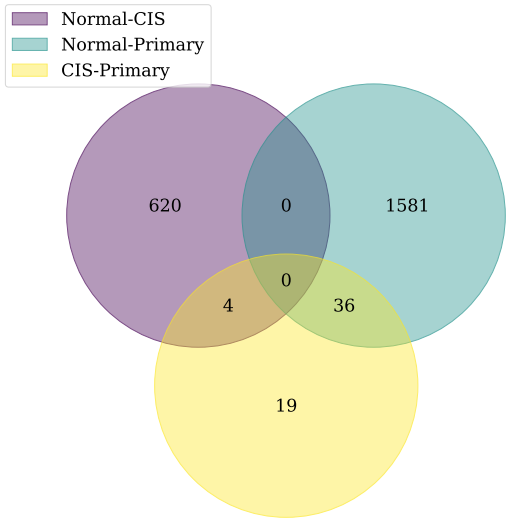
(b) Normal-Primary



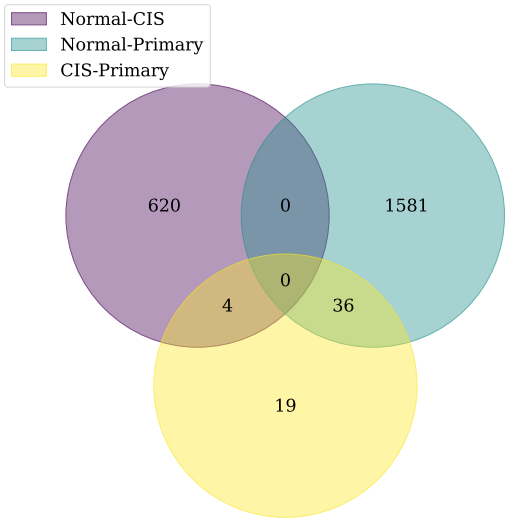
(c) CIS-Primary

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

DEG Venn Diagram with Non-recur in LUSC



(a) Up-regulated



(b) Down-regulated

Figure: DEG Venn Diagram with Non-recur in LUSC

Enrichment test with Normal vs. CIS for Non-recur

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

Term name	Adjusted p-value
Malaria	6.53e-08
Hematopoietic cell lineage	2.01e-07
Hypertrophic cardiomyopathy	1.53e-06

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

Term name	Adjusted p-value
Metabolism of xenobiotics by cytochrome P450	9.67e-05
Drug metabolism	1.18e-04
Cell cycle	1.89e-04

Enrichment test with Normal vs. Primary for Non-recur

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

Term name	Adjusted p-value
Cell cycle	3.04e-06
DNA replication	6.47e-06
Homologous recombination	3.33e-05

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

Term name	Adjusted p-value
Hematopoietic cell lineage	6.65e-10
Malaria	3.57e-09
Hypertrophic cardiomyopathy	5.12e-09

4. Results

4.6. Differences in Gene Expression Levels

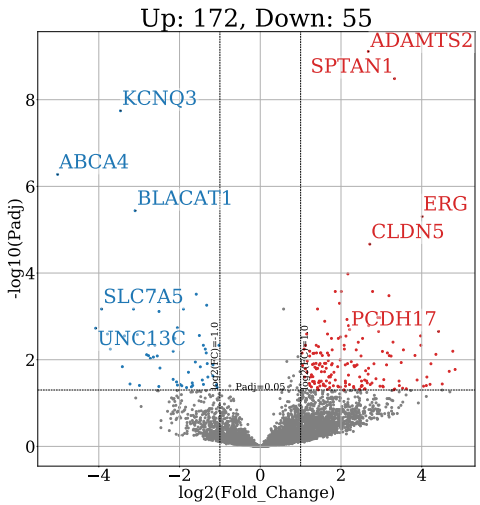
4.6.6. Within Non-recur in LUAD

LUAD Data Composition

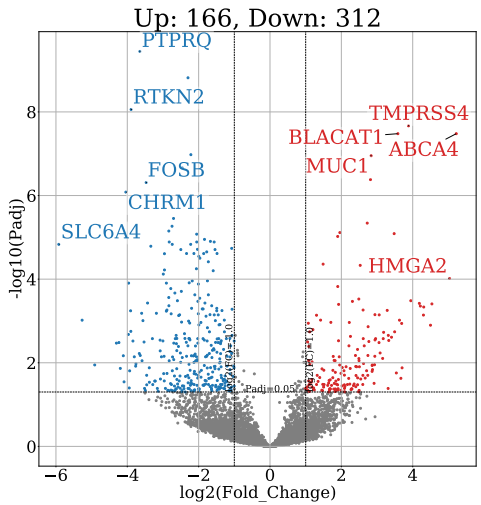
Table: Number of WTS LUAD samples

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

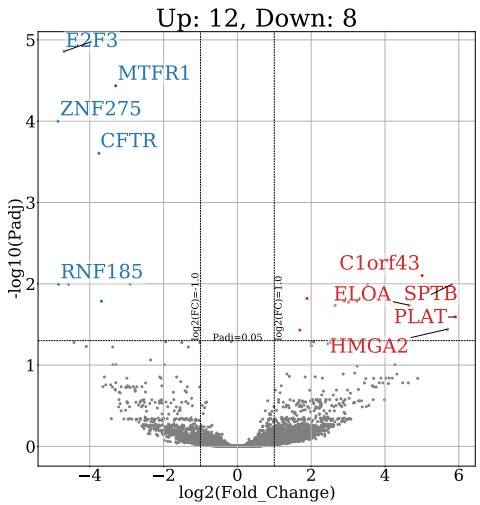
DEG Volcano Plots with Non-recur in LUAD



(a) Normal-AIS



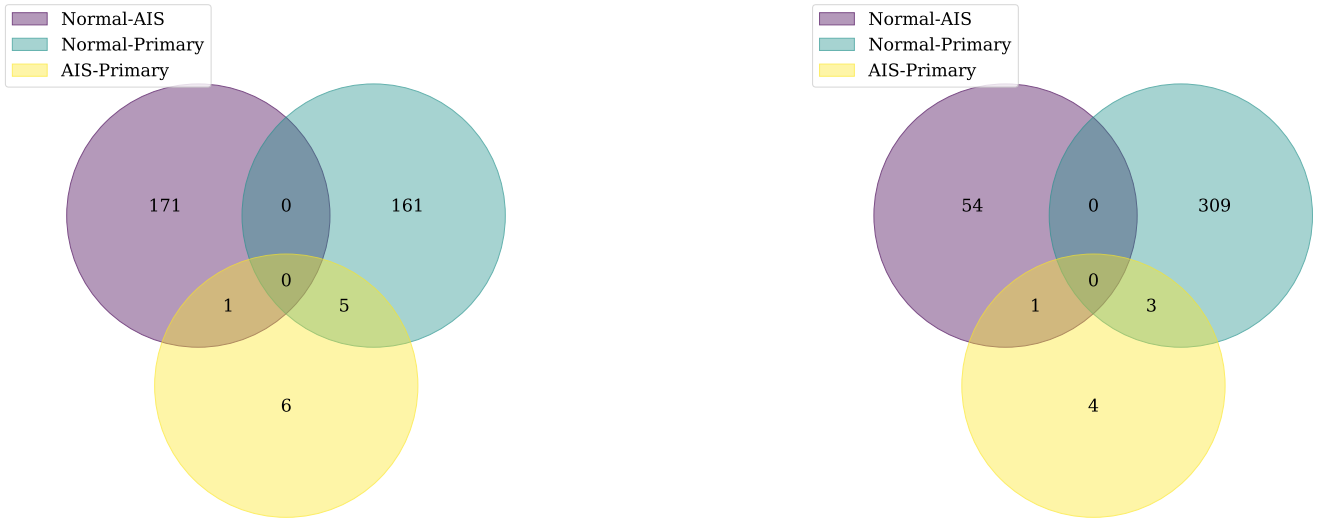
(b) Normal-Primary



(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) Down-regulated

Figure: DEG Venn Diagram with Non-recur in LUAD

Enrichment test with Normal vs. AIS in LUAD

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

Term name	Adjusted p-value
Calcium signaling pathway	3.90e-02

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

Term name	Adjusted p-value
NaN	NaN

Enrichment test with Normal vs. Primary in LUAD

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

Term name	Adjusted p-value
NaN	NaN

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

Term name	Adjusted p-value
ECM-receptor interaction	2.05e-03
Vascular smooth muscle contraction	4.98e-03
Calcium signaling pathway	7.82e-03

Findings in DEG Analysis

4. Results

4.7. Bulk Cell Deconvolution

Single-cell data as Reference

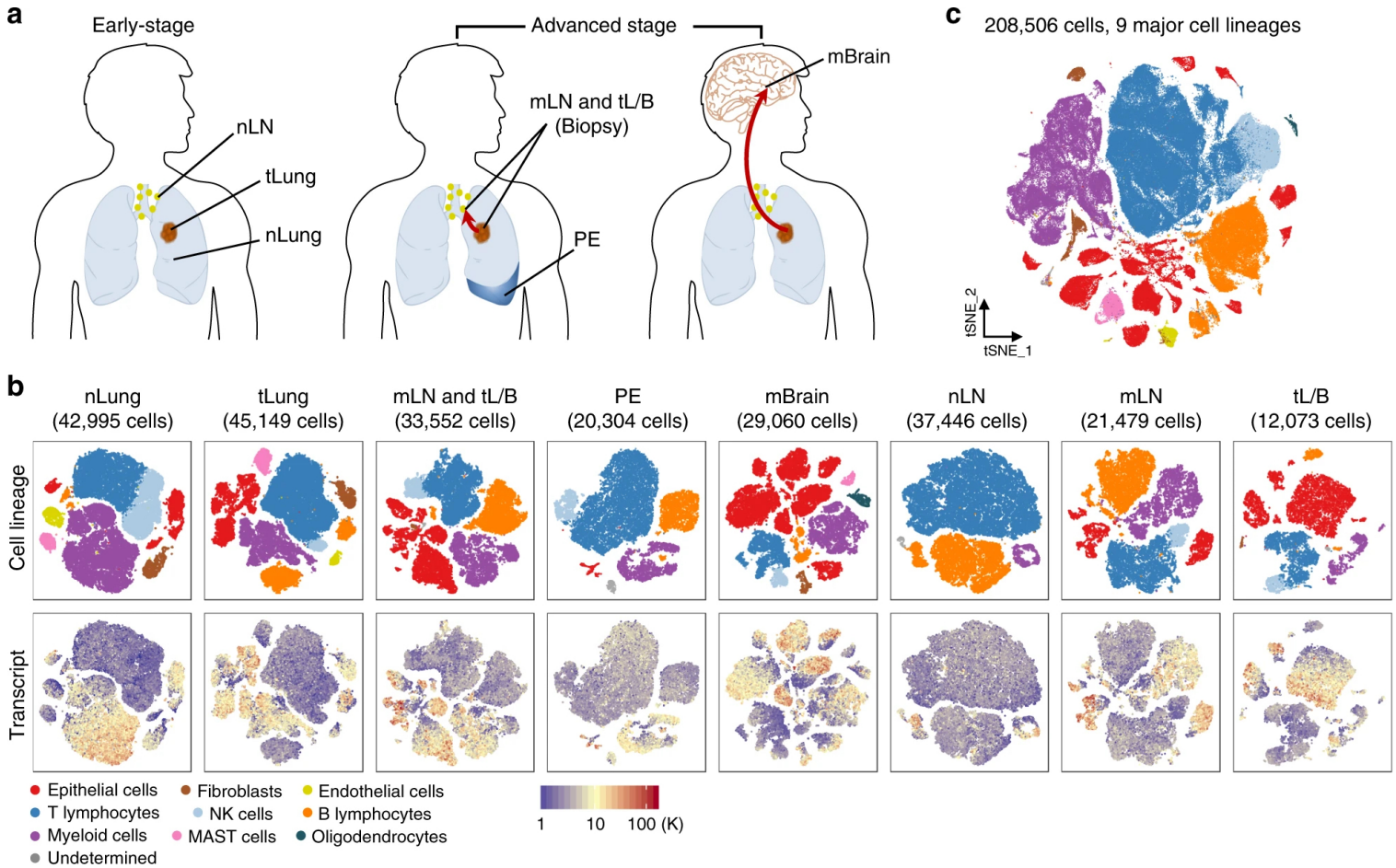


Figure: Comprehensive dissection and clustering of 208,506 single cells from LUAD patients (Kim et al., 2020)

BisqueRNA?

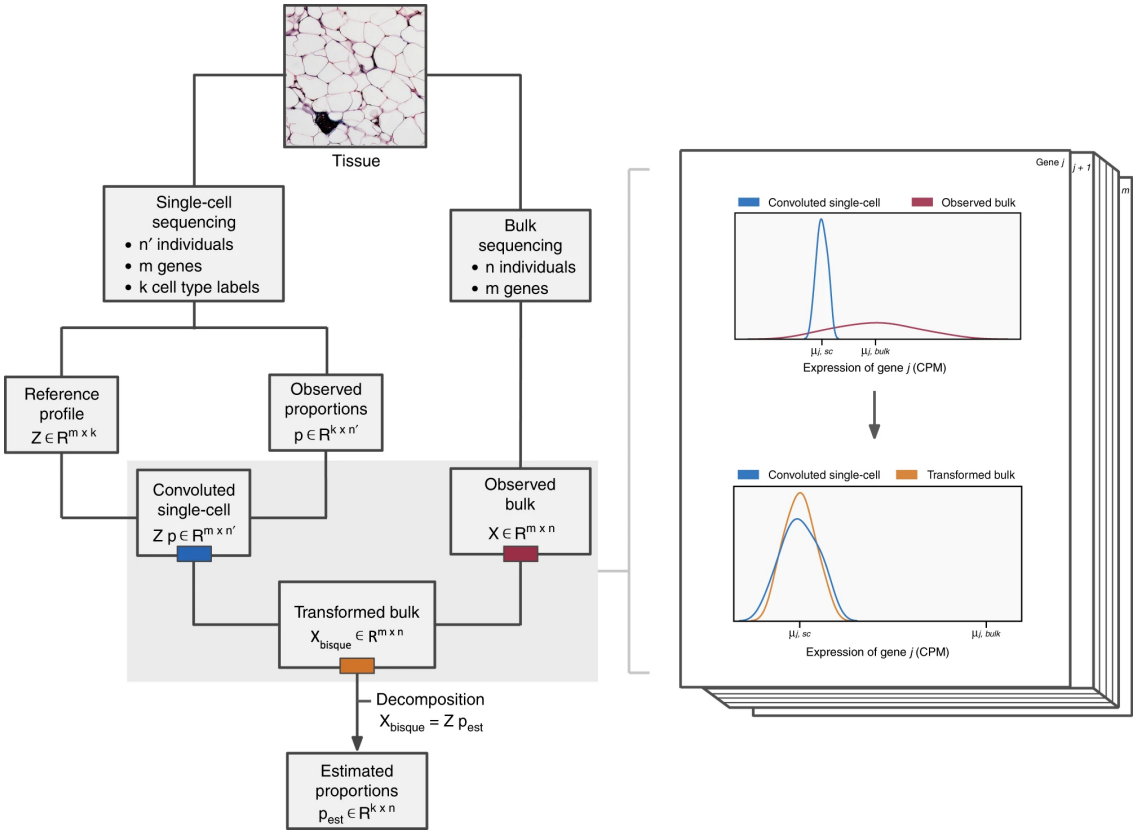


Figure: Workflow for BisqueRNA (Jew et al., 2020)

Cluster Plot in LUSC

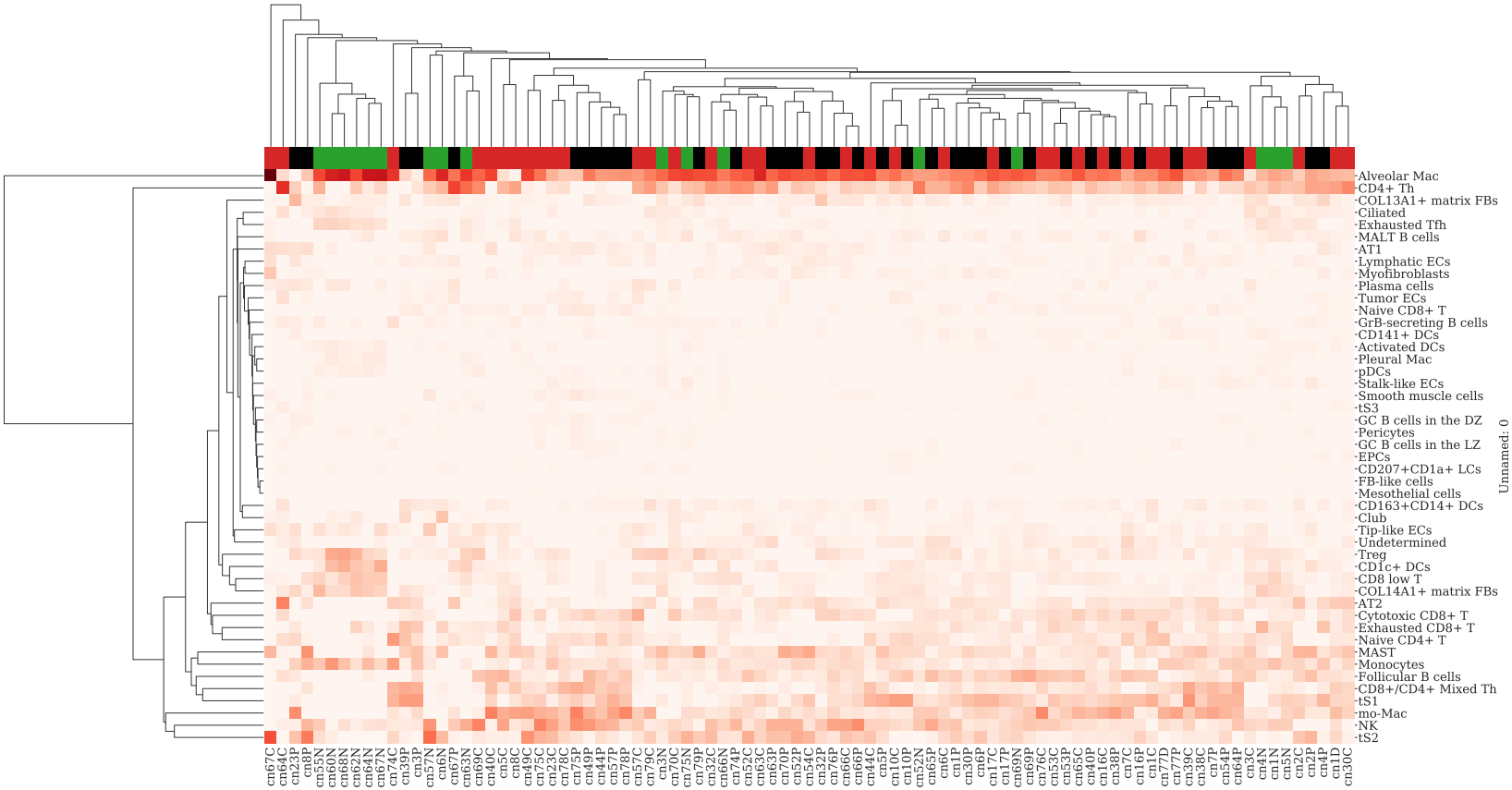


Figure: Cluster Plot in LUSC

Violin Plots in LUSC I

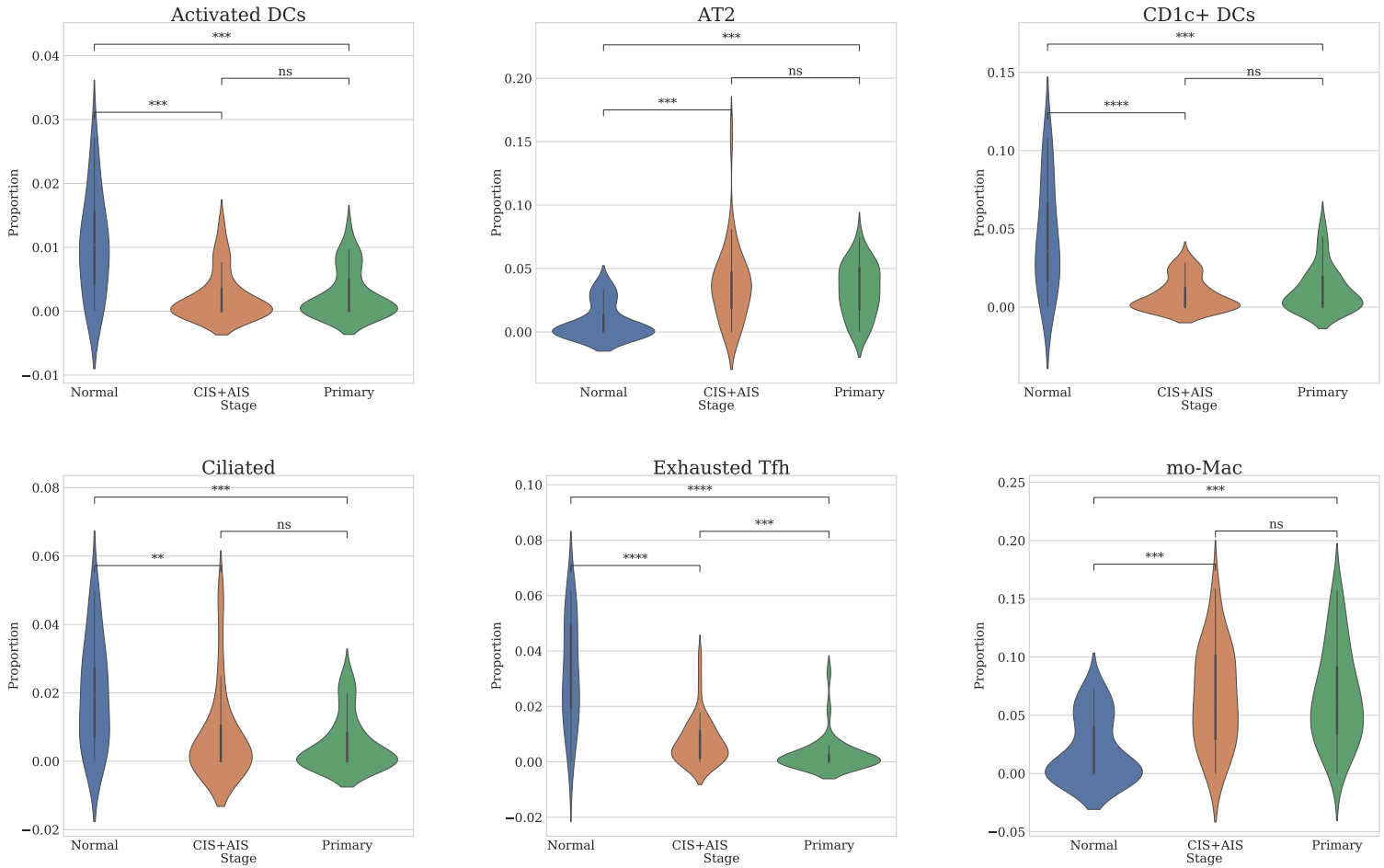


Figure: Violin Plots in LUSC

Violin Plots in LUSC II

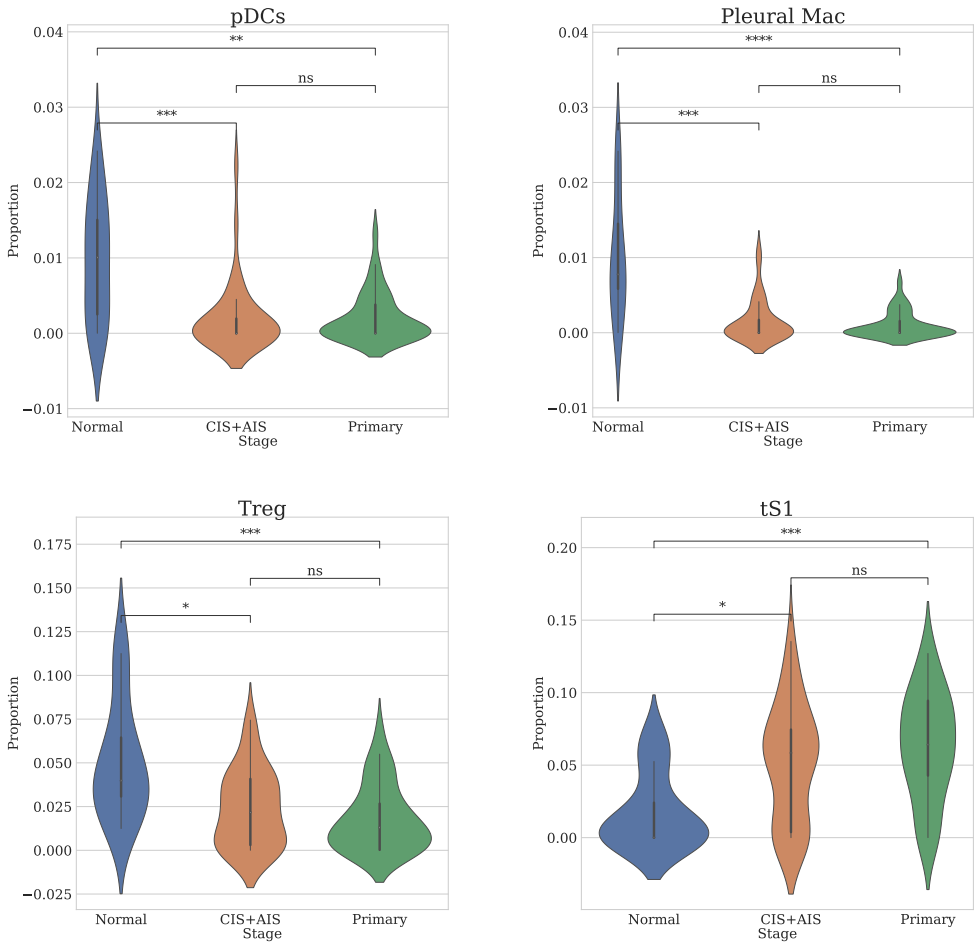


Figure: Violin Plots in LUSC

Cluster Plot in LUAD

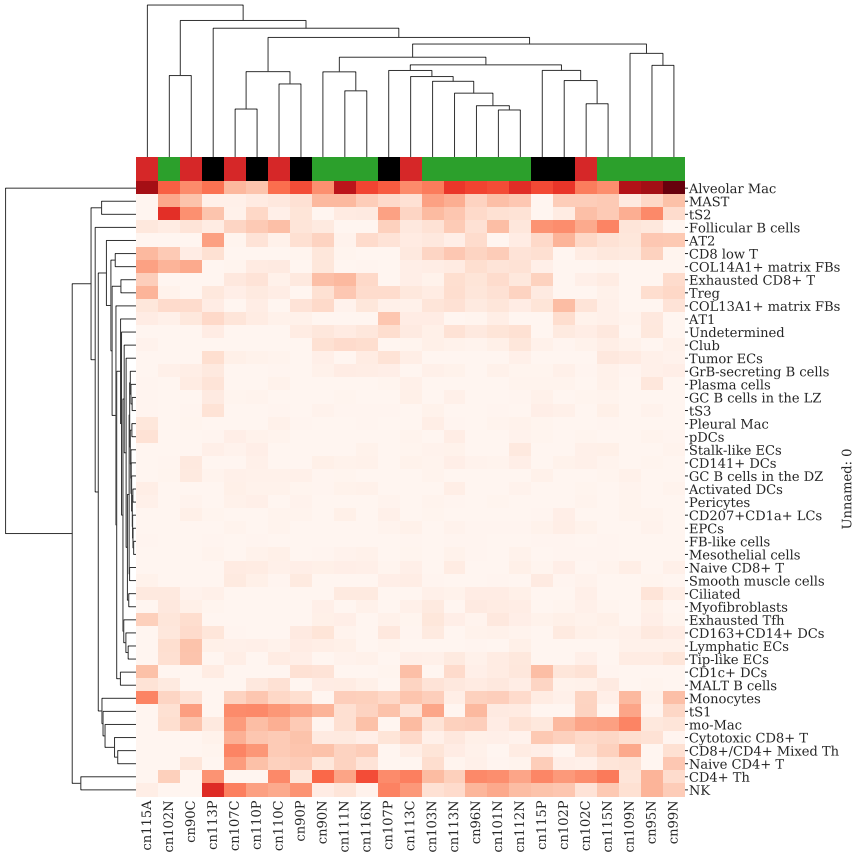


Figure: Cluster Plot in LUAD

Violin Plots in LUAD

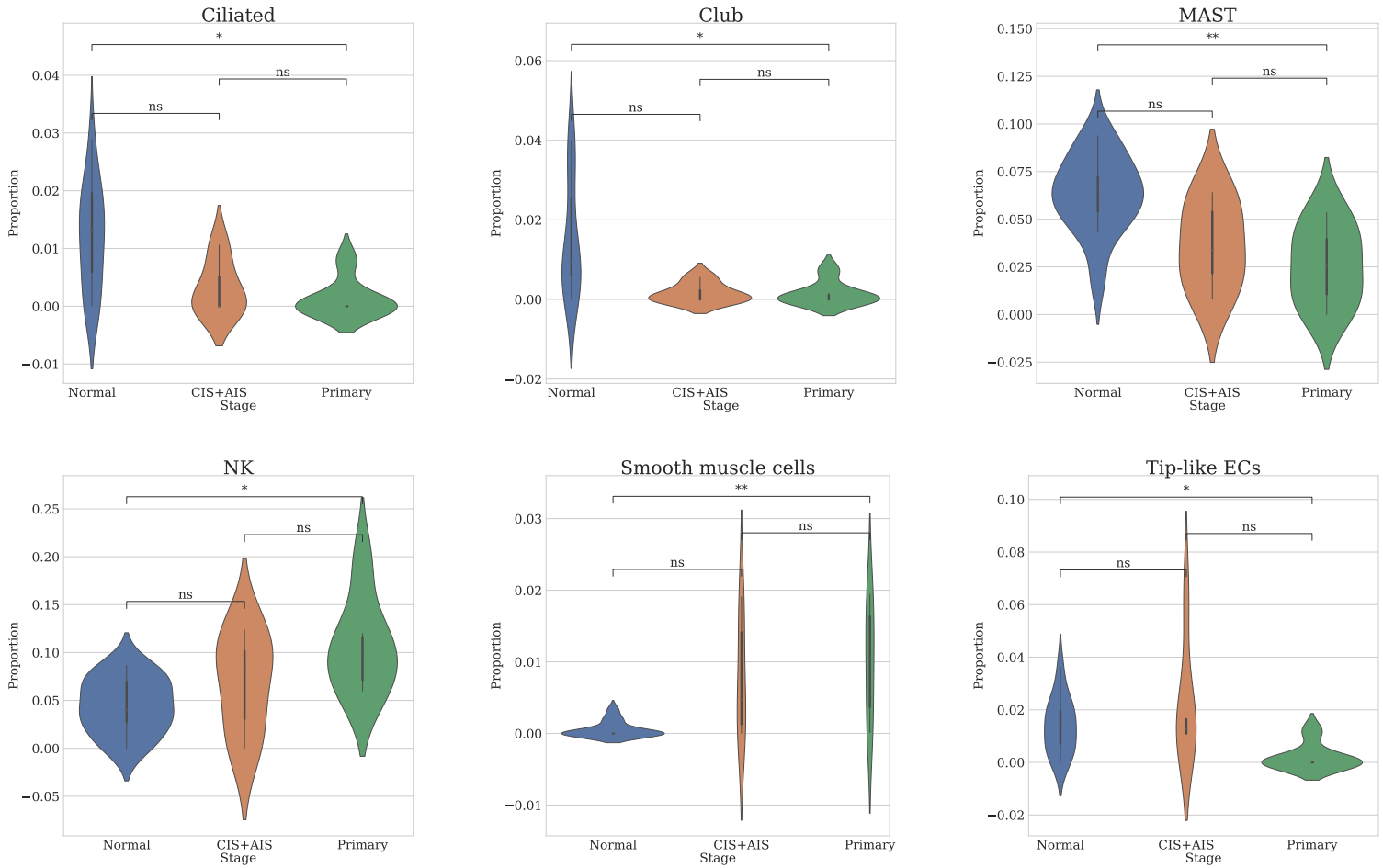


Figure: Violin Plots in LUAD

Findings in Bulk Cell Deconvolution

4. Results

4.8. Discovery of Gene Fusion

Arriba?

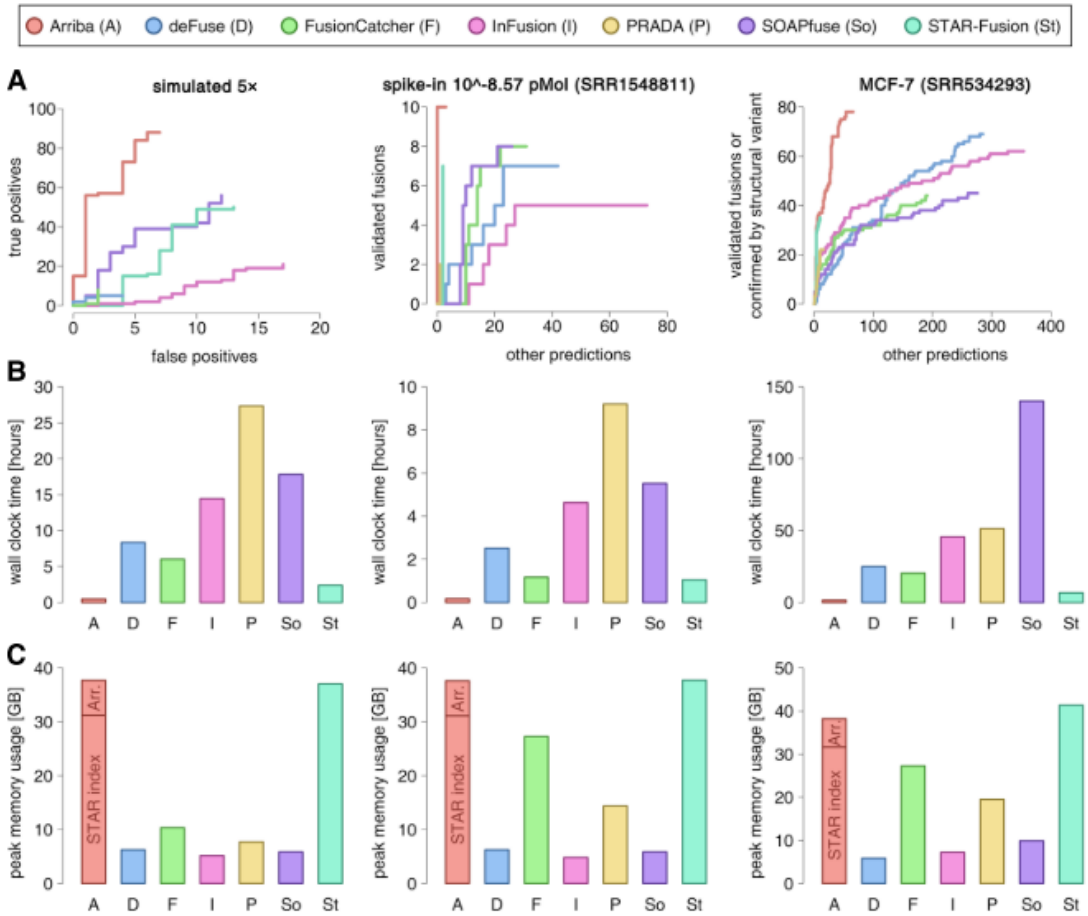


Figure: Benchmark of Arriba versus alternative methods (Uhrig et al., 2021)

Findings in Gene Fusion Discovery

5. Discussion

6. References

References I

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

Caravagna, G., Giarratano, Y., Ramazzotti, D., Tomlinson, I., Graham, T. A., Sanguinetti, G., & Sottoriva, A. (2018). Detecting repeated cancer evolution from multi-region tumor sequencing data. *Nature methods*, *15*(9), 707–714.

Chen, E. Y., Tan, C. M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G. V., ... Ma'ayan, A. (2013). Enrichr: interactive and collaborative html5 gene list enrichment analysis tool. *BMC bioinformatics*, *14*(1), 1–14.

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

References II

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., ... others (2011). A framework for variation discovery and genotyping using next-generation dna sequencing data. *Nature genetics*, *43*(5), 491.

Ding, L., Ley, T. J., Larson, D. E., Miller, C. A., Koboldt, D. C., Welch, J. S., ... others (2012). Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature*, *481*(7382), 506–510.

Favero, F., Joshi, T., Marquard, A. M., Birkbak, N. J., Krzystanek, M., Li, Q., ... Eklund, A. C. (2015). Sequenza: allele-specific copy number and mutation profiles from tumor sequencing data. *Annals of Oncology*, *26*(1), 64–70.

References III

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.

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.

Kanehisa, M., Furumichi, M., Sato, Y., Ishiguro-Watanabe, M., & Tanabe, M. (2021). Kegg: integrating viruses and cellular organisms. *Nucleic acids research*, *49*(D1), D545–D551.

Kim, N., Kim, H. K., Lee, K., Hong, Y., Cho, J. H., Choi, J. W., ... others (2020). Single-cell rna sequencing demonstrates the molecular and cellular reprogramming of metastatic lung adenocarcinoma. *Nature communications*, *11*(1), 1–15.

References IV

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.

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.

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.

Minna, J. D., Roth, J. A., & Gazdar, A. F. (2002). Focus on lung cancer. *Cancer cell*, *1*(1), 49–52.

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 V

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.

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.

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

Uhrig, S., Ellermann, J., Walther, T., Burkhardt, P., Fröhlich, M., Hutter, B., ... others (2021). Accurate and efficient detection of gene fusions from rna sequencing data. *Genome research*, *31*(3), 448–460.

References VI

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.

Vincent, R. G., Pickren, J. W., Lane, W. W., Bross, I., Takita, H., Houten, L., ... Rzepka, T. (1977). The changing histopathology of lung cancer. a review of 1682 cases. *Cancer*, *39*(4), 1647–1655.

Wang, B.-Y., Huang, J.-Y., Chen, H.-C., Lin, C.-H., Lin, S.-H., Hung, W.-H., & Cheng, Y.-F. (2020). The comparison between adenocarcinoma and squamous cell carcinoma in lung cancer patients. *Journal of cancer research and clinical oncology*, *146*(1), 43–52.