

Lung Pre-cancer

Jaewoong Lee

Yeonsong Choi

Sabin Park

Ilsun Yun

Semin Lee

2021-08-19

Contents

1	Introduction	3
1.1	Lung Cancer	3
1.2	Precancer	3
1.3	Study Objectives	3
2	Materials	3
2.1	List of IPNs	3
2.1.1	Carcinoma <i>in situ</i>	3
2.1.2	Adenocarcinoma <i>in situ</i>	3
2.1.3	Atypical Adenomatous Hyperplasia	3
2.1.4	Dysplasia	3
2.1.5	Minimally Invasive Adenocarcinoma	3
2.2	Data Structure & Count	9
3	Methods	9
3.1	Workflows	9
4	Results	9
4.1	Quality Check with FastQC	9
4.2	Copy Number Variations	9
4.3	Somatic Single Nucleotide Variations	9
4.4	Variant Allele Frequencies	9
4.5	Differences in Gene Expression levels	9
4.6	Bulk Cell Deconvolution	9
5	Discussion	9
6	References	9

List of Tables

List of Figures

1	Common cancer survival rates (Hong et al., 2021)	3
2	Workflow for data pre-processing for variant discovery (Van der Auwera et al., 2013; DePristo et al., 2011)	4
3	Somatic short variant discovery workflow (Van der Auwera et al., 2013; DePristo et al., 2011)	4
4	Germline short variant discovery workflow (Van der Auwera et al., 2013; DePristo et al., 2011)	5
5	RNA-seq short variant discovery workflow (Van der Auwera et al., 2013; DePristo et al., 2011)	5
6	FastQC results with WES data	5
7	FastQC results with WTS data	5
8	Cellularities and Ploidies with ADC	6
9	Cellularities and Ploidies with SQC	6
10	CNV plot with ADC	7
11	CNV plot with SQC	8
12	Simple CNV plot with ADC	9
13	Simple CNV plot with SQC	9
14	CoMut plot with ADC	10
15	CoMut plot with SQC	10
16	DEG volcano plots in ADC	11
17	DEG volcano plots in SQC	11
18	DEG Venn Diagram in ADC	12
19	DEG Venn Diagram in SQC	12
20	Comprehensive dissection and clustering of 208,506 single cells from LUAD patients (Kim et al., 2020)	12
21	Cell deconvolution clustermap in ADC	13
22	Cell deconvolution clustermap in SQC	13

1 Introduction

1.1 Lung Cancer

Lung cancer is the most common form of cancer as 12.3 % of all cancers (Minna, Roth, & Gazdar, 2002).

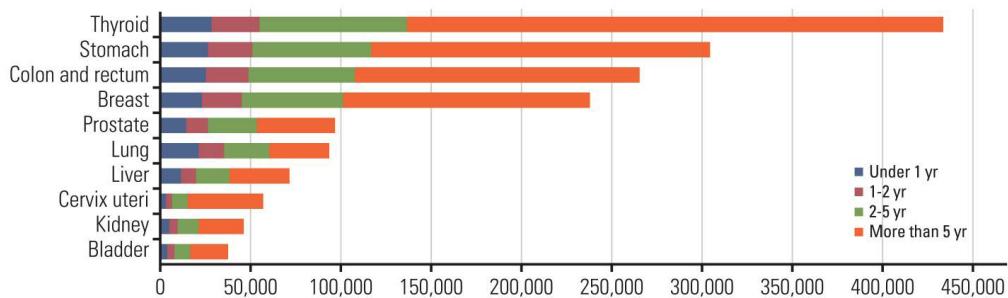


Figure 1: Common cancer survival rates (Hong et al., 2021)

1.2 Precancer

1.3 Study Objectives

2 Materials

2.1 List of IPNs

2.1.1 Carcinoma *in situ*

Carcinoma *in situ* (CIS)

2.1.2 Adenocarcinoma *in situ*

Adenocarcinoma *in situ* (AIS)

2.1.3 Atypical Adenomatous Hyperplasia

Atypical adenomatous hyperplasia (AAH)

2.1.4 Dysplasia

2.1.5 Minimally Invasive Adenocarcinoma

Minimally invasive adenocarcinoma (MIA)

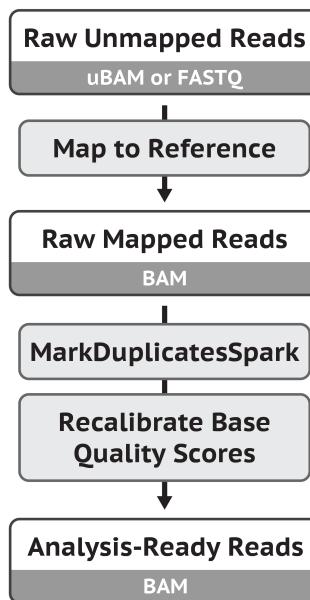


Figure 2: Workflow for data pre-processing for variant discovery (Van der Auwera et al., 2013; DePristo et al., 2011)

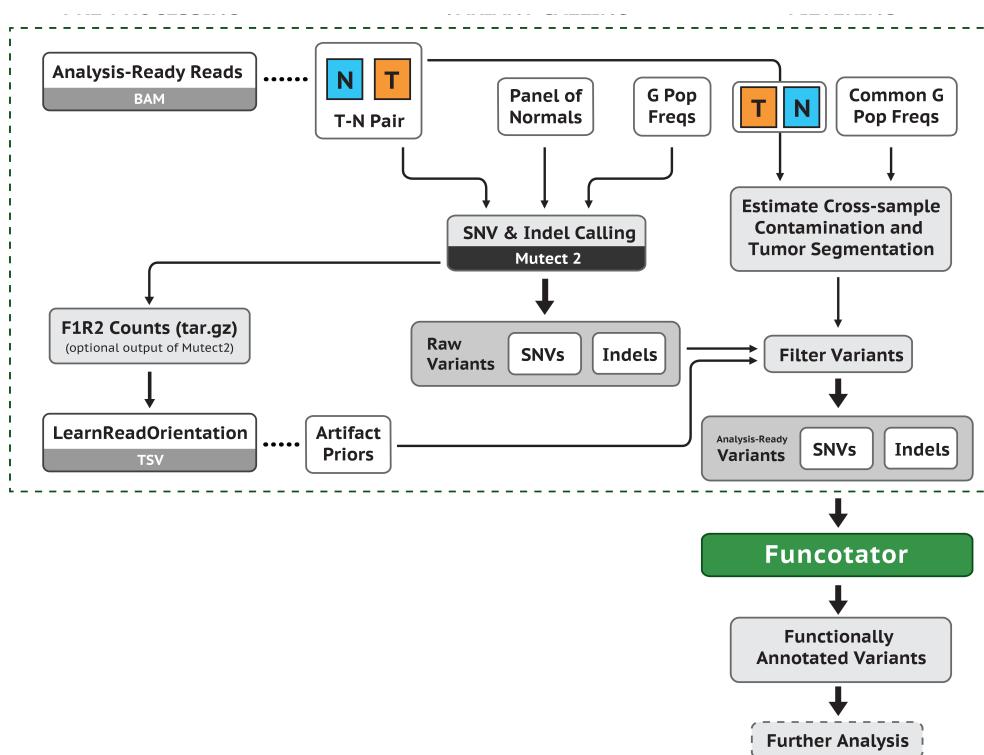


Figure 3: Somatic short variant discovery workflow (Van der Auwera et al., 2013; DePristo et al., 2011)

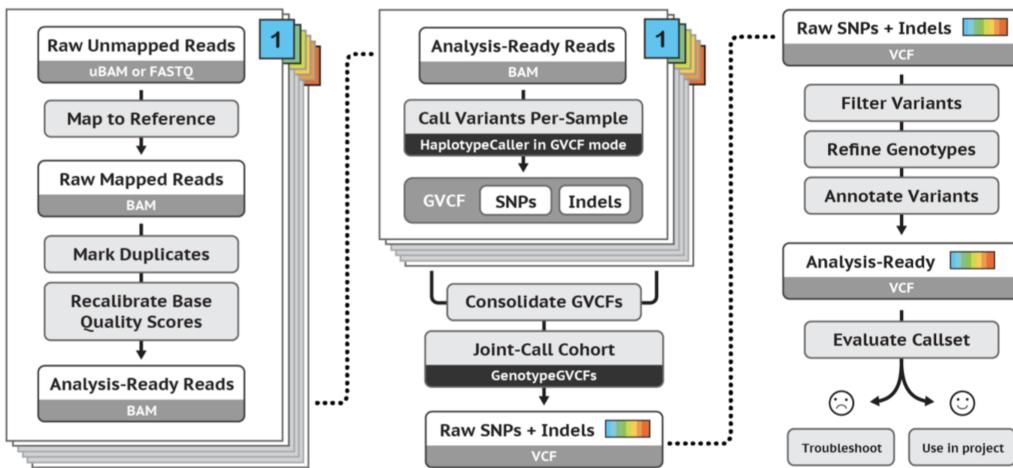


Figure 4: Germline short variant discovery workflow (Van der Auwera et al., 2013; DePristo et al., 2011)

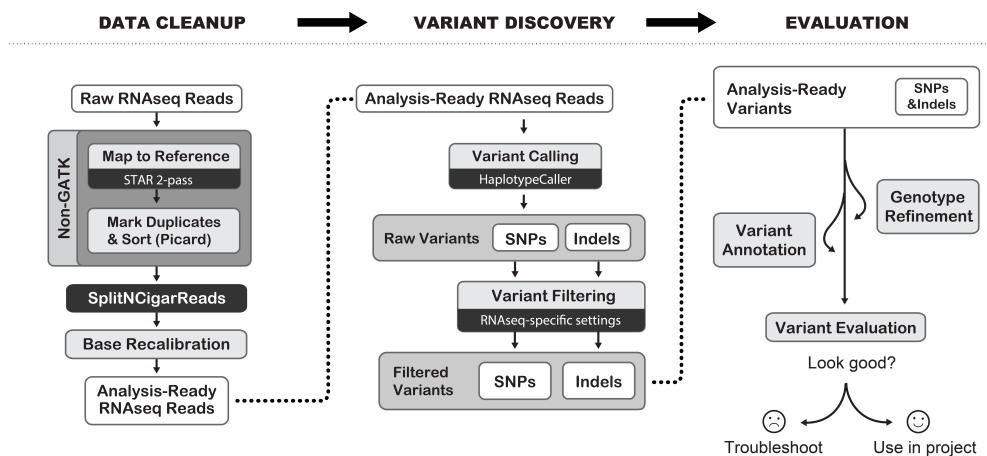


Figure 5: RNA-seq short variant discovery workflow (Van der Auwera et al., 2013; DePristo et al., 2011)

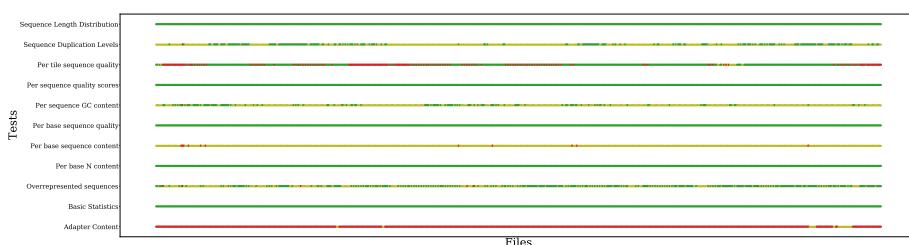


Figure 6: FastQC results with WES data

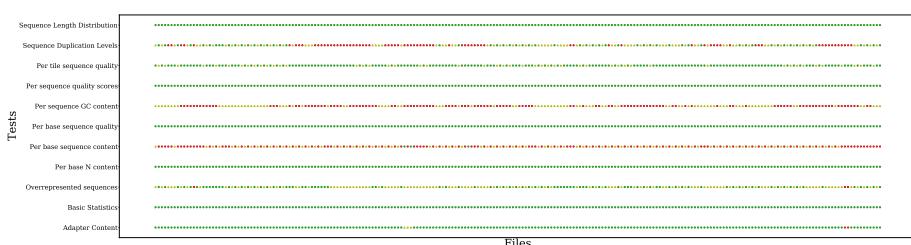


Figure 7: FastQC results with WTS data

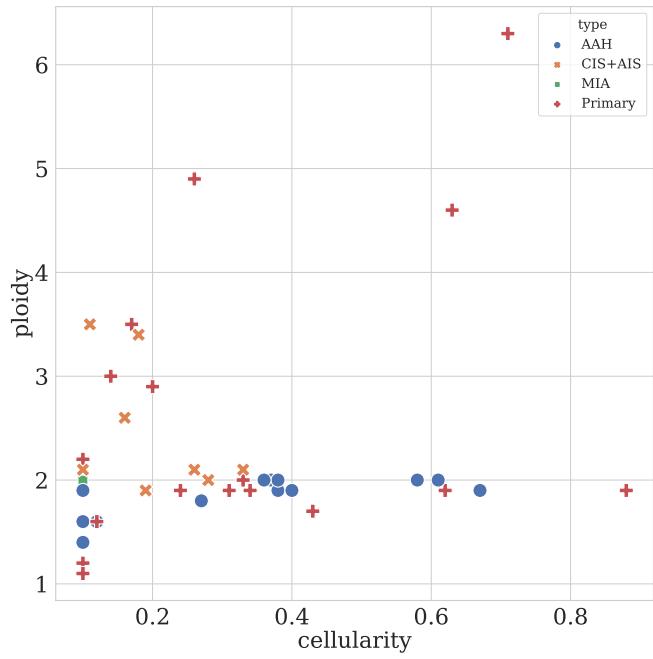


Figure 8: Cellularities and Ploidies with ADC

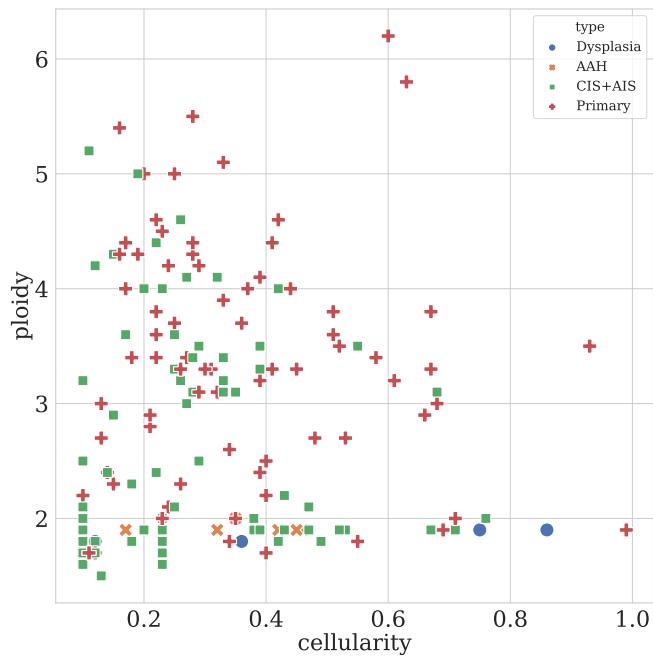


Figure 9: Cellularities and Ploidies with SQC

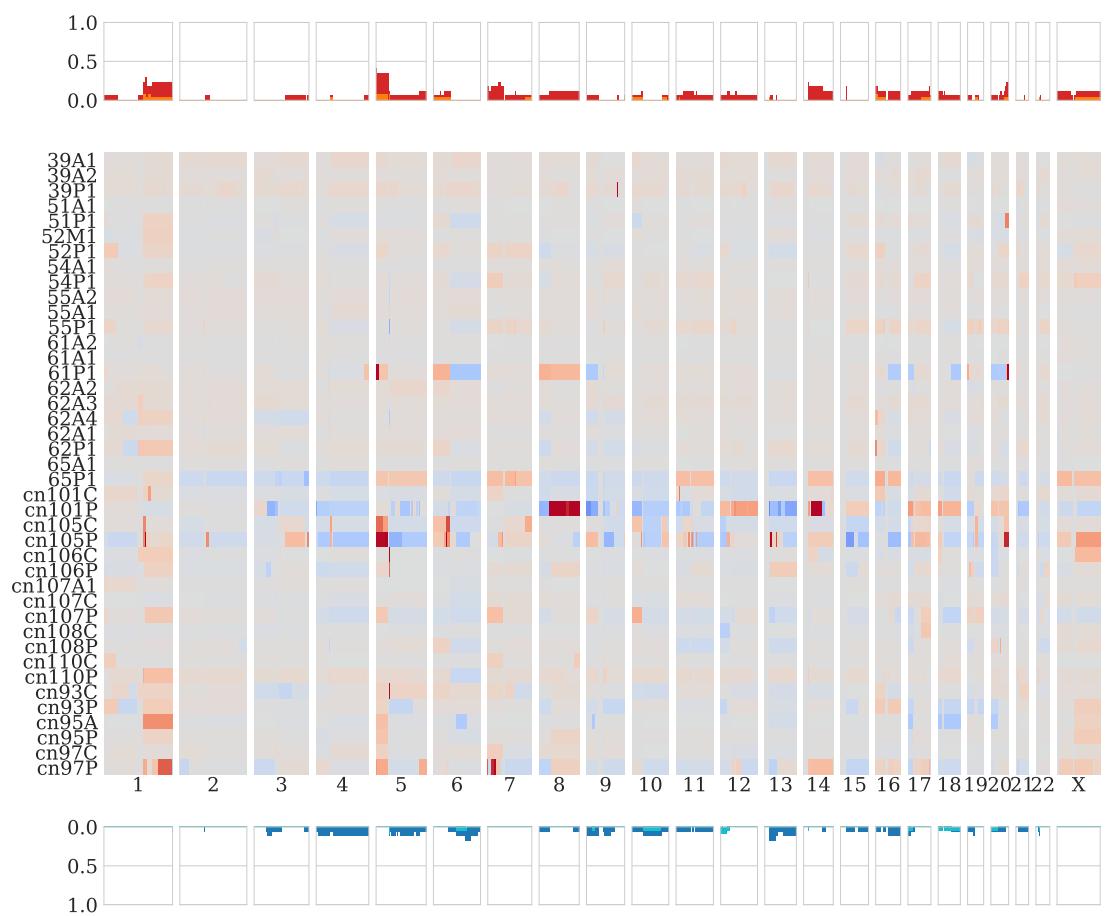


Figure 10: CNV plot with ADC

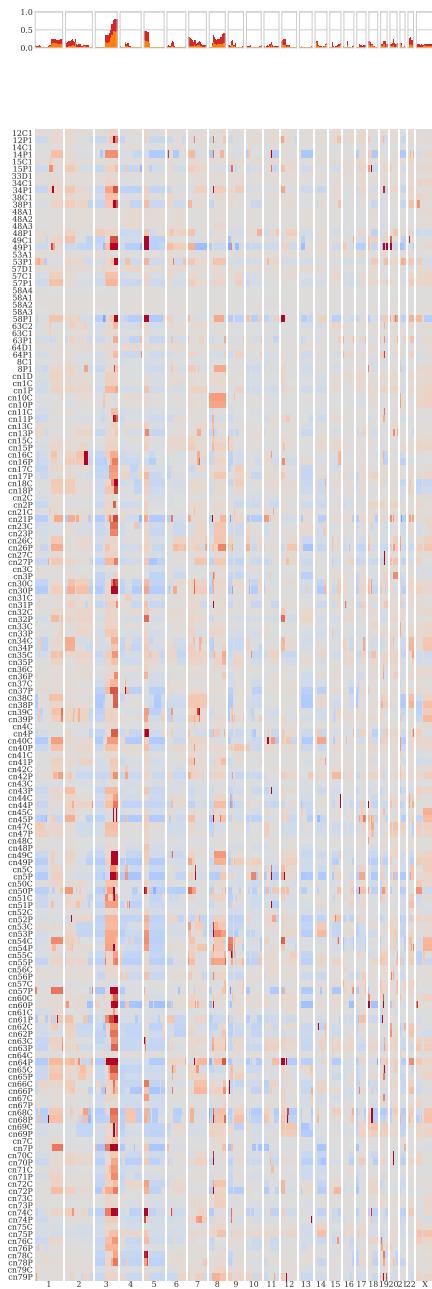


Figure 11: CNV plot with SQC

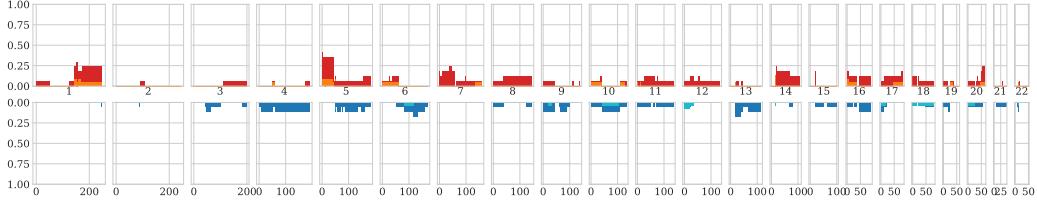


Figure 12: Simple CNV plot with ADC

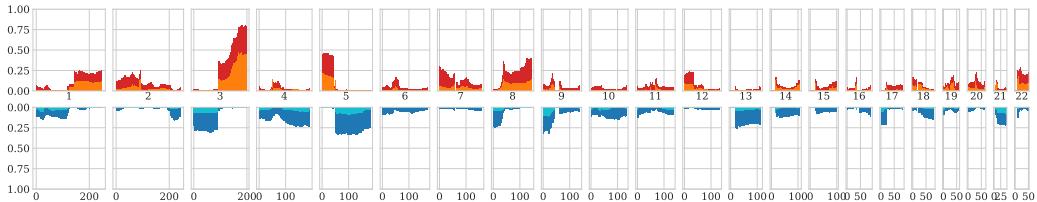


Figure 13: Simple CNV plot with SQC

2.2 Data Structure & Count

3 Methods

3.1 Workflows

4 Results

4.1 Quality Check with FastQC

4.2 Copy Number Variations

4.3 Somatic Single Nucleotide Variations

4.4 Variant Allele Frequencies

4.5 Differences in Gene Expression levels

4.6 Bulk Cell Deconvolution

5 Discussion

6 References

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

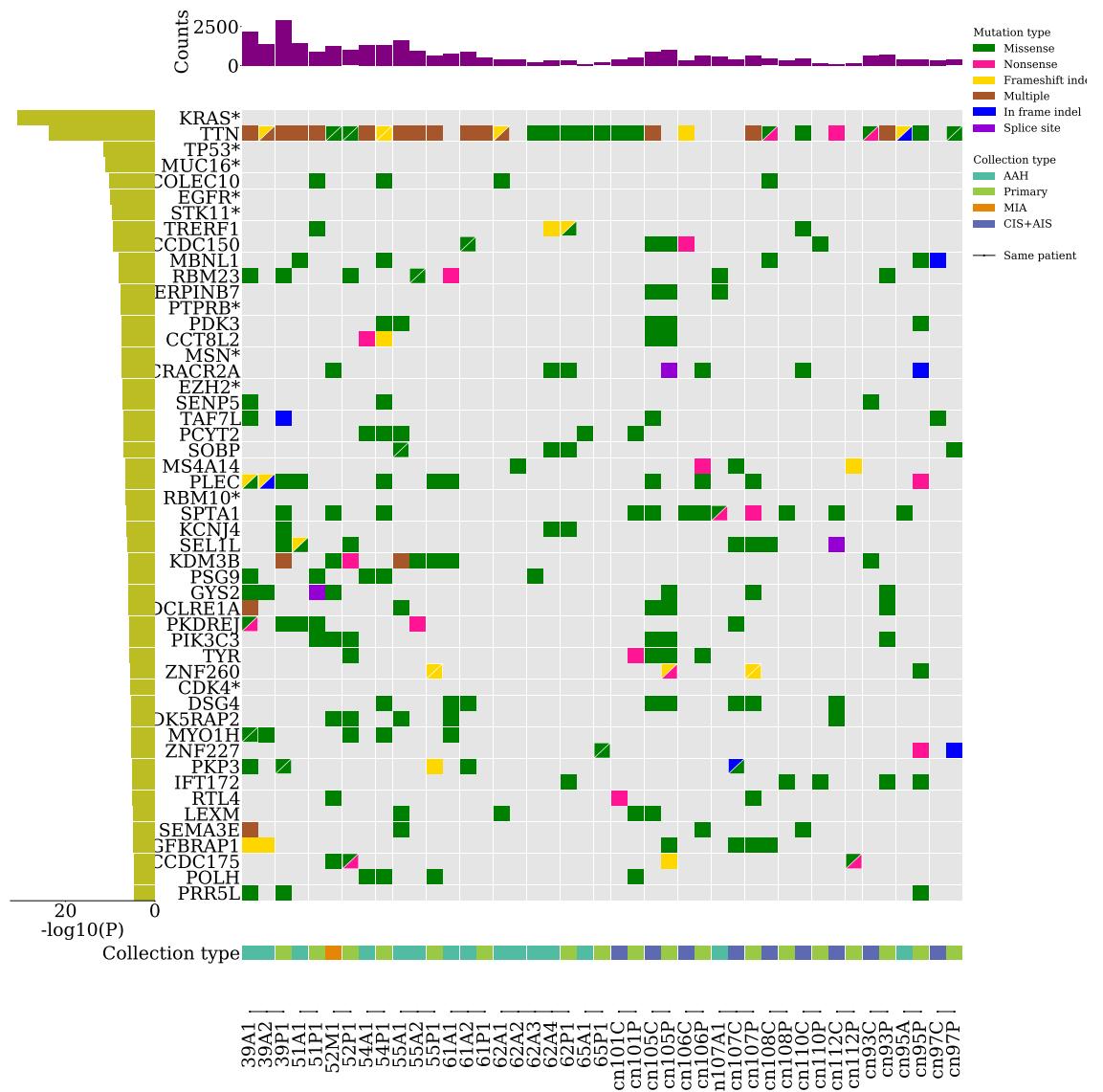


Figure 14: CoMut plot with ADC



Figure 15: CoMut plot with SQC

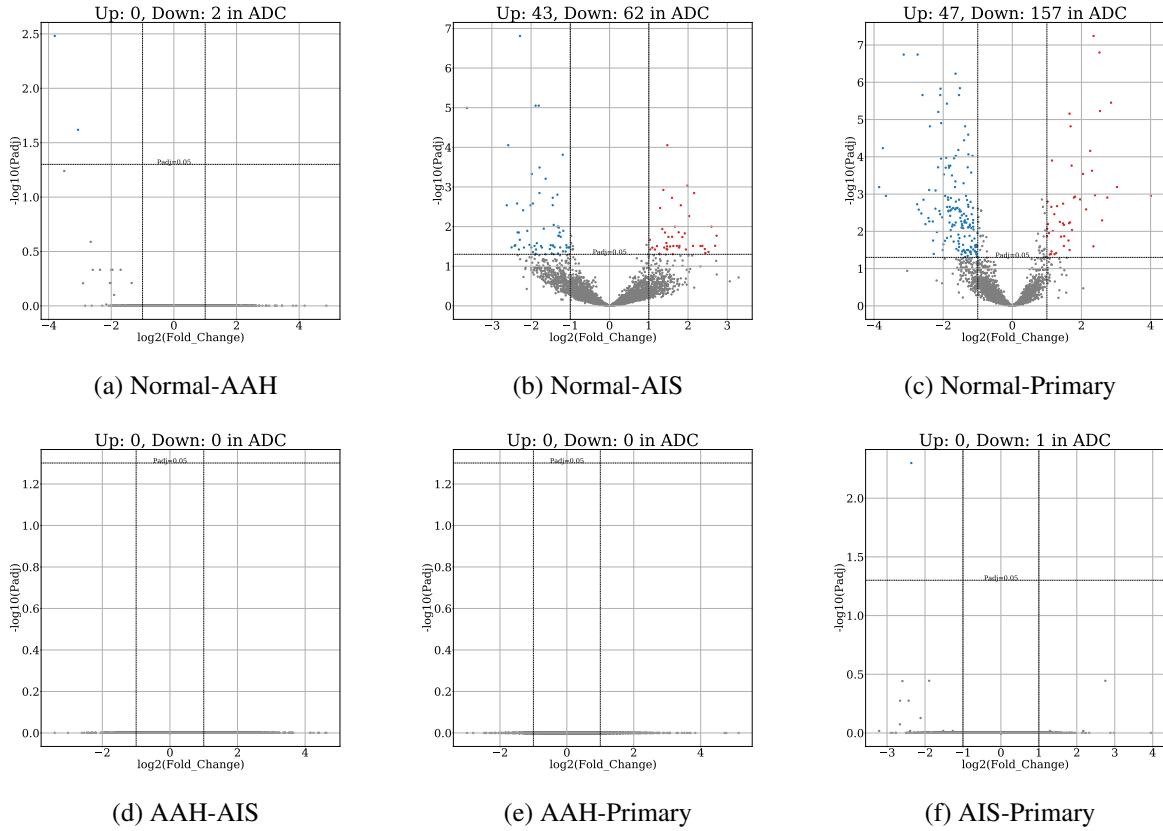


Figure 16: DEG volcano plots in ADC

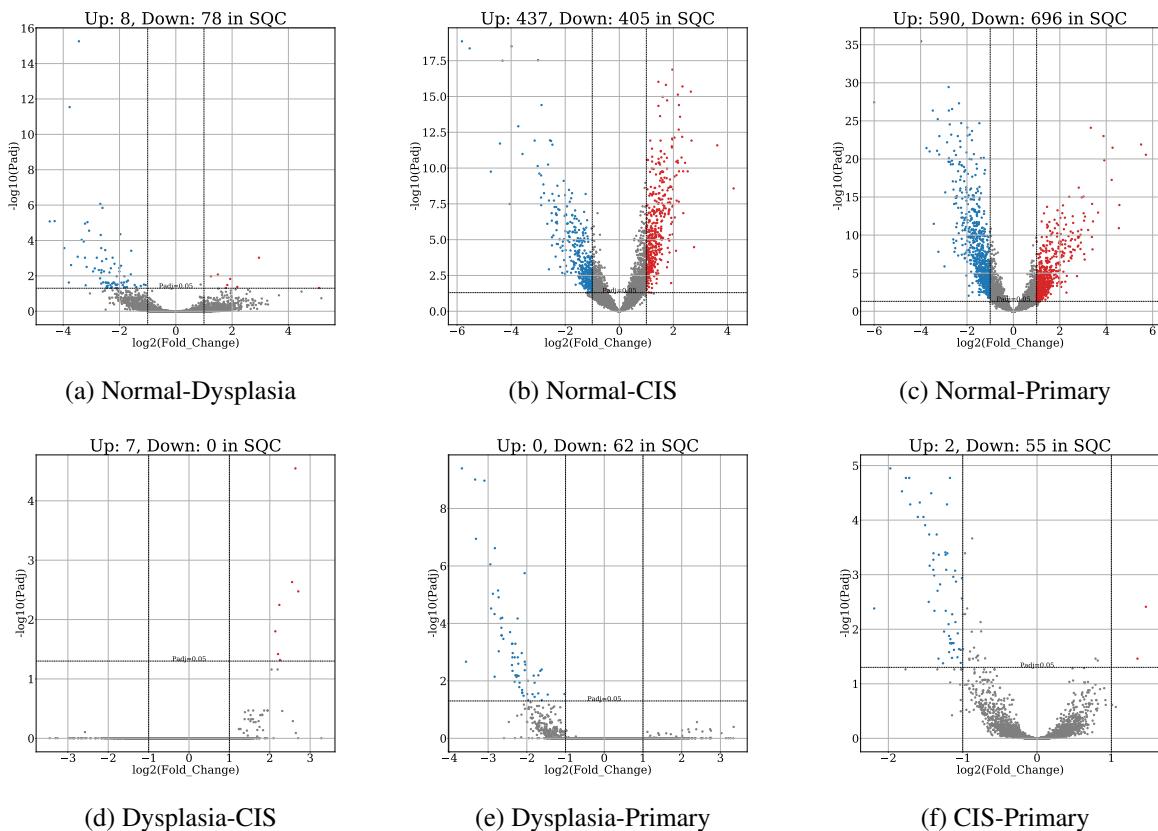


Figure 17: DEG volcano plots in SQC

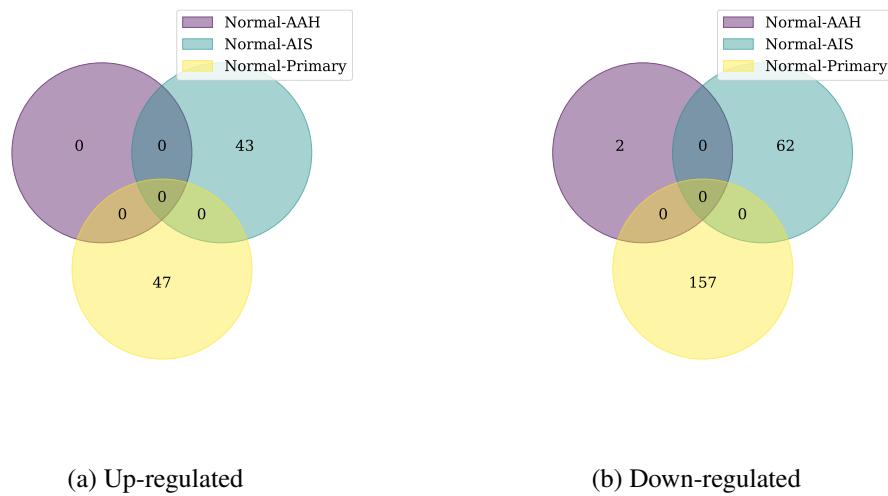


Figure 18: DEG Venn Diagram in ADC

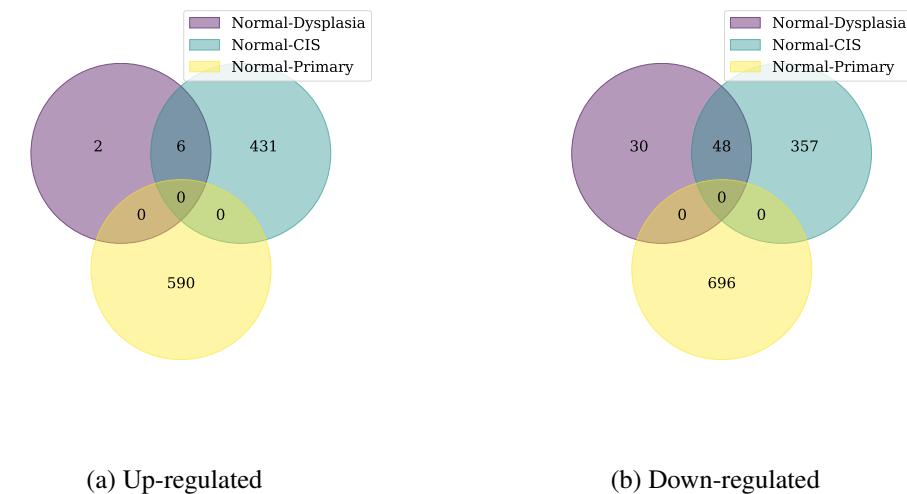


Figure 19: DEG Venn Diagram in SQC

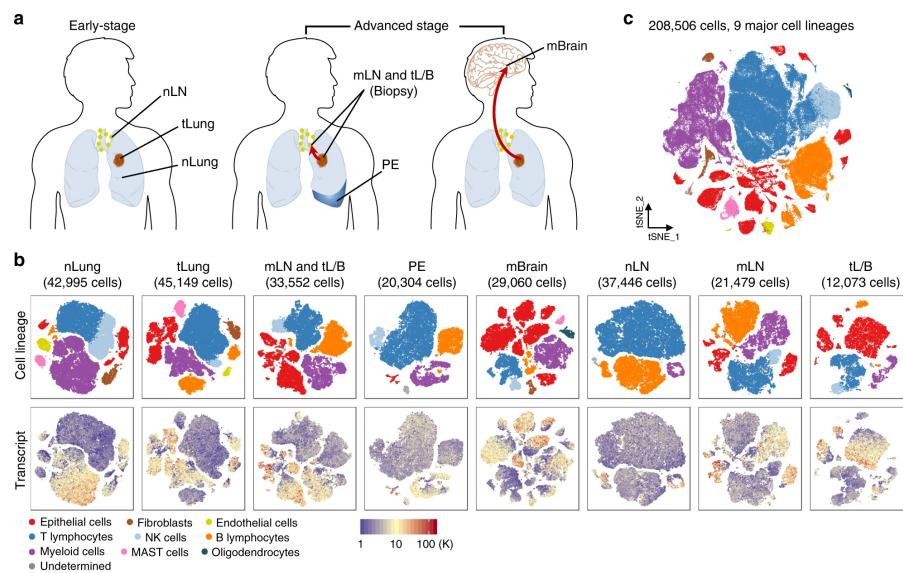


Figure 20: Comprehensive dissection and clustering of 208,506 single cells from LUAD patients (Kim et al., 2020)

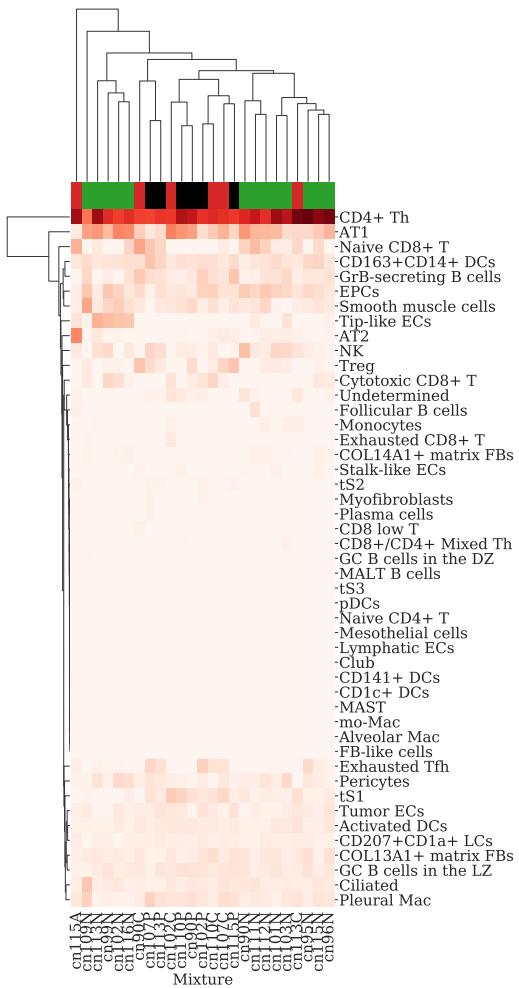


Figure 21: Cell deconvolution clustermap in ADC

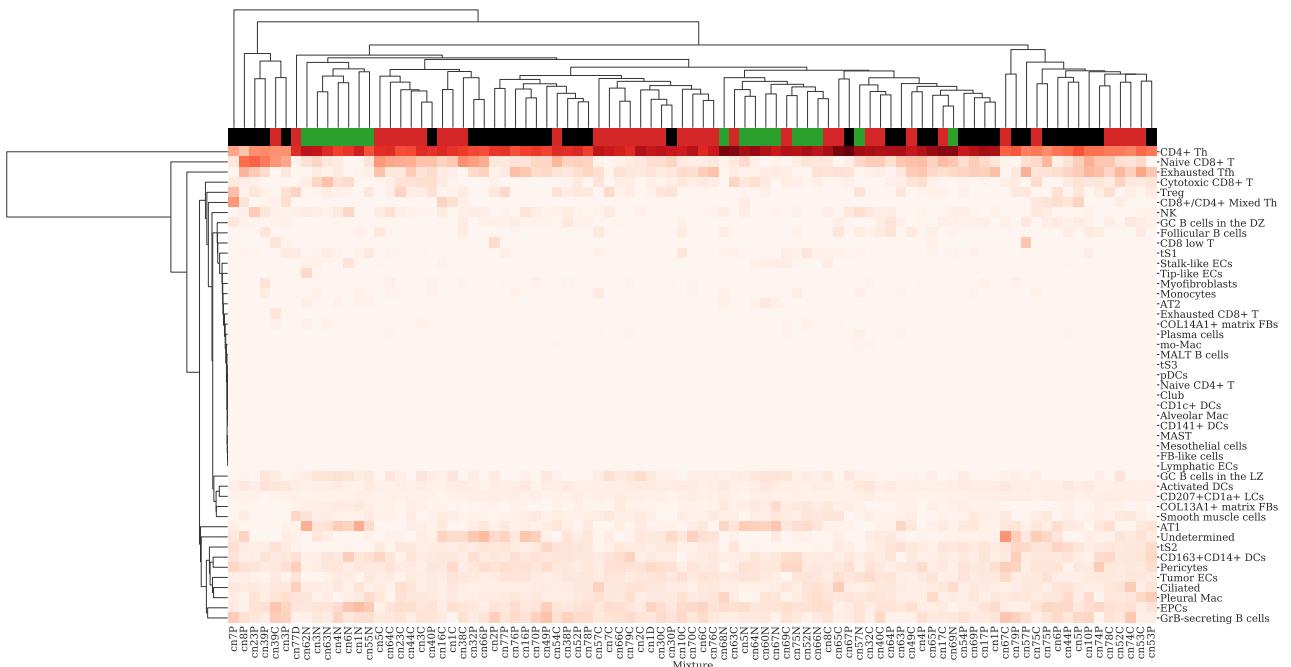


Figure 22: Cell deconvolution clustermap in SQC