

# Lung Pre-cancer

Jaewoong Lee

Yeonsong Choi

Sabin Park

Ilsun Yun

Semin Lee

2021-08-24

# Contents

<b>1</b>	<b>Introduction</b>	<b>4</b>
1.1	Lung Cancer . . . . .	4
1.2	Precancer . . . . .	4
1.3	Study Objectives . . . . .	4
<b>2</b>	<b>Materials</b>	<b>4</b>
2.1	List of IPNs . . . . .	4
2.1.1	Carcinoma <i>in situ</i> . . . . .	4
2.1.2	Adenocarcinoma <i>in situ</i> . . . . .	4
2.1.3	Atypical Adenomatous Hyperplasia . . . . .	4
2.1.4	Dysplasia . . . . .	4
2.1.5	Minimally Invasive Adenocarcinoma . . . . .	4
2.2	Data Structure & Count . . . . .	16
<b>3</b>	<b>Methods</b>	<b>16</b>
3.1	Workflows . . . . .	16
<b>4</b>	<b>Results</b>	<b>16</b>
4.1	Quality Check with FastQC . . . . .	16
4.2	Copy Number Variations . . . . .	16
4.3	Somatic Single Nucleotide Variations . . . . .	16
4.4	Variant Allele Frequencies . . . . .	16
4.5	Differences in Gene Expression levels . . . . .	16
4.6	Bulk Cell Deconvolution . . . . .	16
<b>5</b>	<b>Discussion</b>	<b>16</b>
<b>6</b>	<b>References</b>	<b>16</b>

## List of Tables

## List of Figures

1	Common cancer survival rates (Hong et al., 2021) . . . . .	4
2	Workflow for data pre-processing for variant discovery (Van der Auwera et al., 2013; DePristo et al., 2011) . . . . .	5
3	Somatic short variant discovery workflow (Van der Auwera et al., 2013; DePristo et al., 2011) . . . . .	5
4	Germline short variant discovery workflow (Van der Auwera et al., 2013; DePristo et al., 2011) . . . . .	6
5	RNA-seq short variant discovery workflow (Van der Auwera et al., 2013; DePristo et al., 2011) . . . . .	6
6	FastQC results with WES data . . . . .	6
7	FastQC results with WTS data . . . . .	6
8	Cellularities and Ploidies by BWA with ADC . . . . .	7
9	Cellularities and Ploidies by BWA with SQC . . . . .	7
10	Cellularities and Ploidies by Bowtie2 with ADC . . . . .	8
11	Cellularities and Ploidies by Bowtie2 with SQC . . . . .	8
12	CNV plot by BWA with ADC . . . . .	9
13	CNV plot by BWA with SQC . . . . .	10
14	CNV plot by Bowtie2 with ADC . . . . .	11
15	CNV plot by Bowtie2 with SQC . . . . .	11
16	Simple CNV plot by BWA with ADC . . . . .	12
17	Simple CNV plot by BWA with SQC . . . . .	12
18	Simple CNV plot by Bowtie2 with ADC . . . . .	12
19	Simple CNV plot by Bowtie2 with SQC . . . . .	12
20	CoMut plot by BWA with ADC . . . . .	13
21	CoMut plot by BWA with SQC . . . . .	13
22	DEG volcano plots by Bowtie2 in ADC . . . . .	14
23	DEG volcano plots by Bowtie2 in SQC . . . . .	14
24	DEG Venn Diagram by Bowtie2 in ADC . . . . .	15
25	DEG Venn Diagram by Bowtie2 in SQC . . . . .	15

26	Comprehensive dissection and clustering of 208,506 single cells from LUAD patients (Kim et al., 2020) .	16
27	Cell deconvolution clustermap by Bowtie2 in ADC . . . . .	17
28	Cell deconvolution clustermap by Bowtie2 in SQC . . . . .	17

# 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).

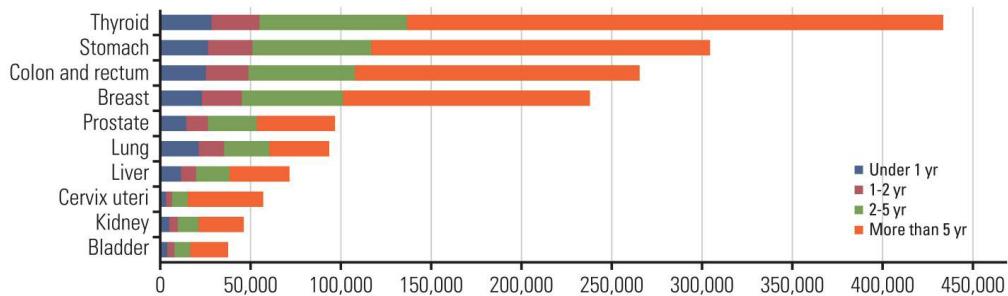


Figure 1: Common cancer survival rates (Hong et al., 2021)

## 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)

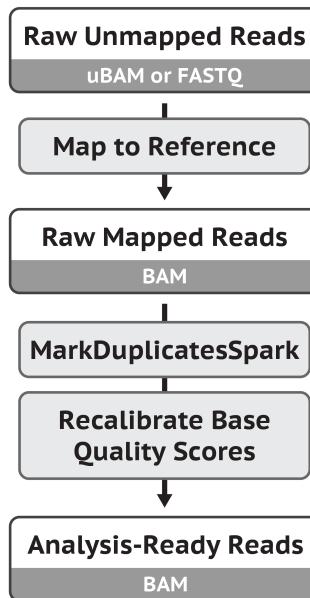


Figure 2: Workflow for data pre-processing for variant discovery (Van der Auwera et al., 2013; DePristo et al., 2011)

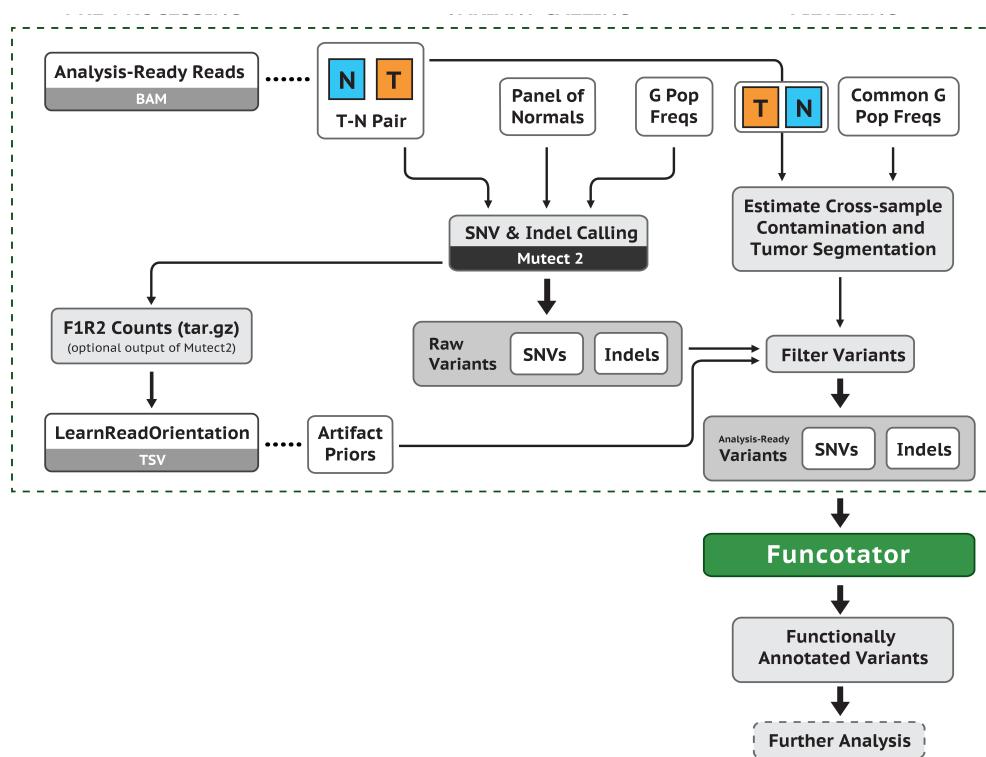


Figure 3: Somatic short variant discovery workflow (Van der Auwera et al., 2013; DePristo et al., 2011)

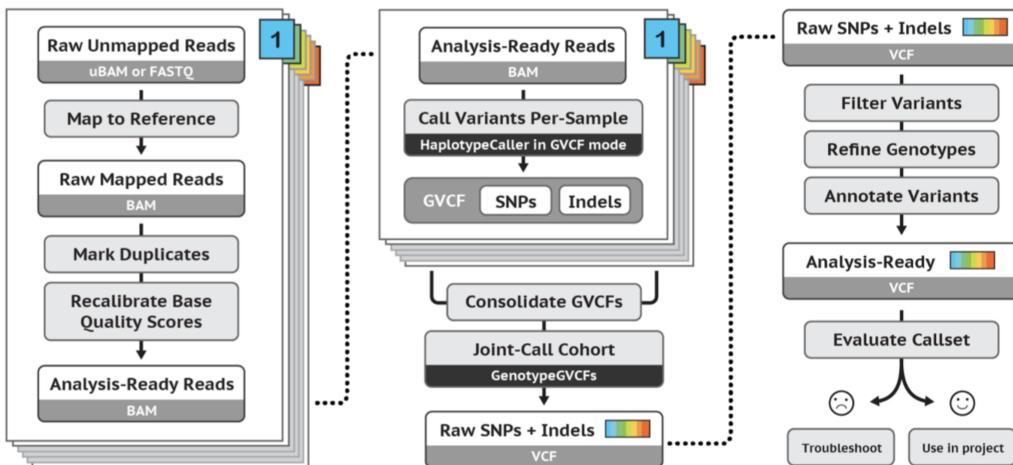


Figure 4: Germline short variant discovery workflow (Van der Auwera et al., 2013; DePristo et al., 2011)

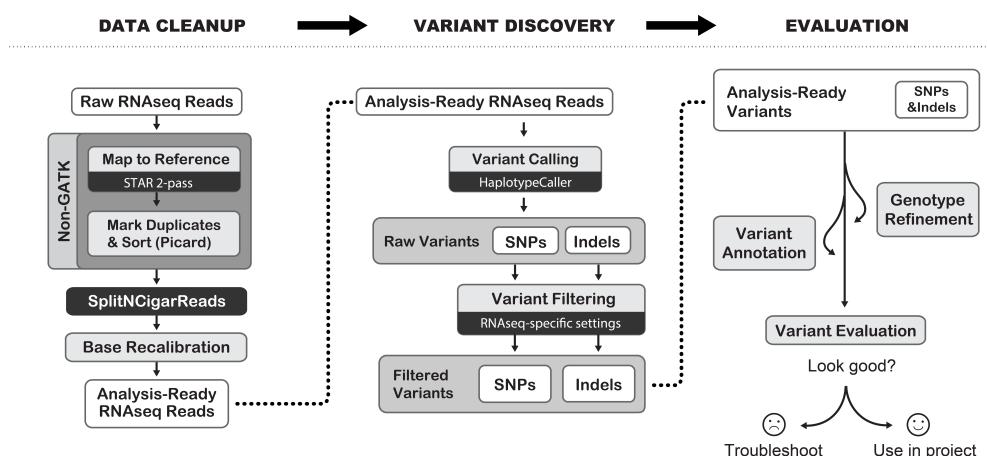


Figure 5: RNA-seq short variant discovery workflow (Van der Auwera et al., 2013; DePristo et al., 2011)

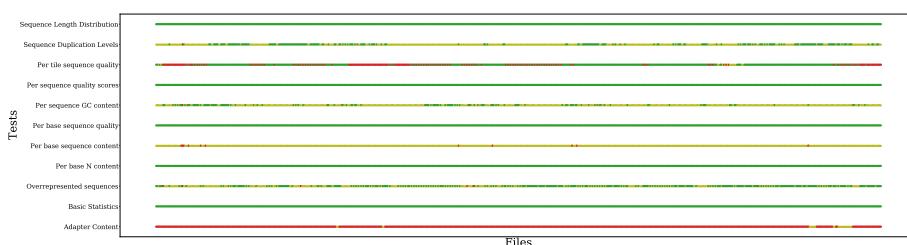


Figure 6: FastQC results with WES data

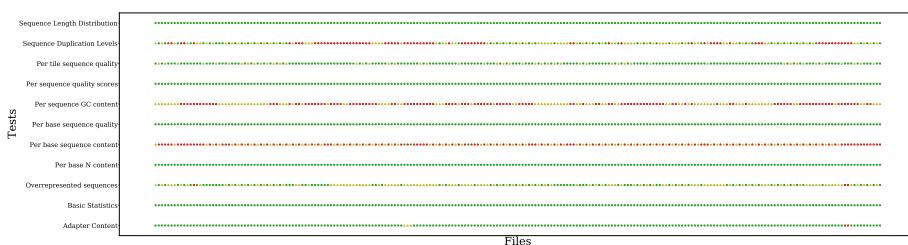


Figure 7: FastQC results with WTS data

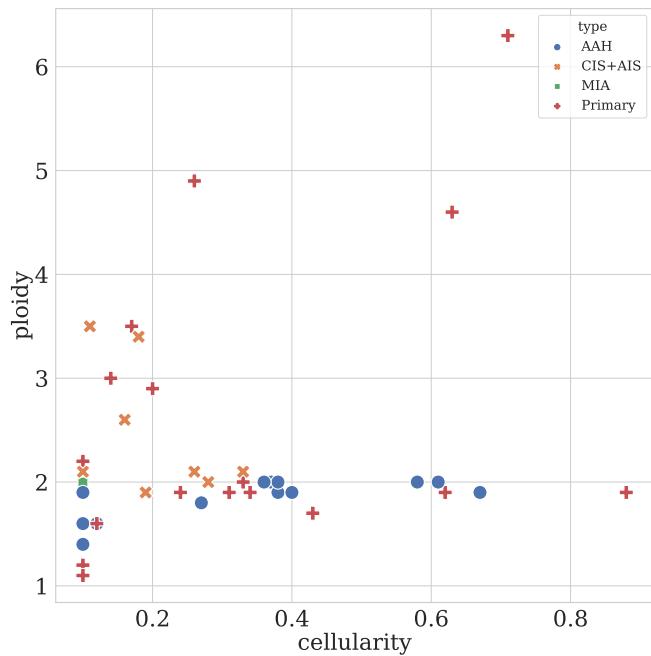


Figure 8: Cellularities and Ploidies by BWA with ADC

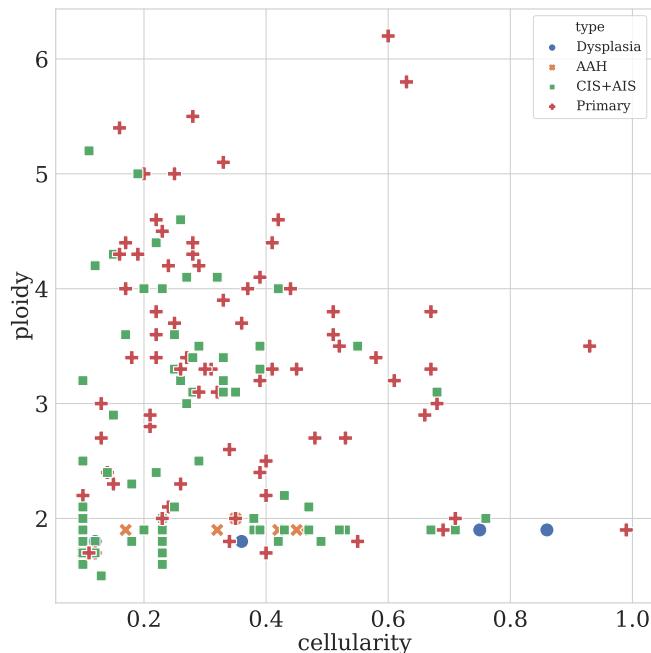


Figure 9: Cellularities and Ploidies by BWA with SQC

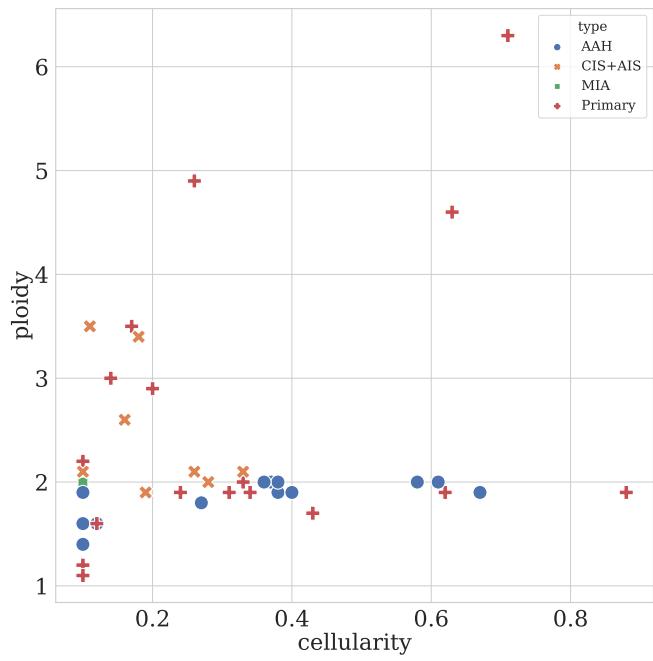


Figure 10: Cellularities and Ploidies by Bowtie2 with ADC

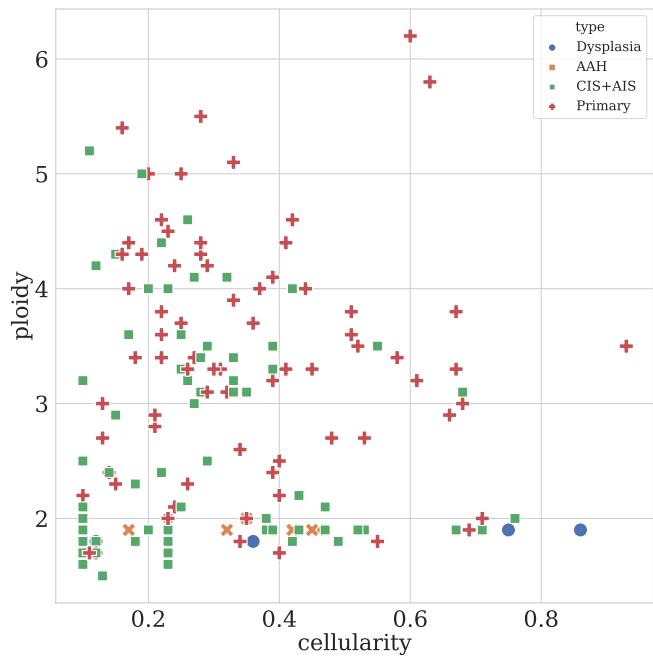


Figure 11: Cellularities and Ploidies by Bowtie2 with SQC

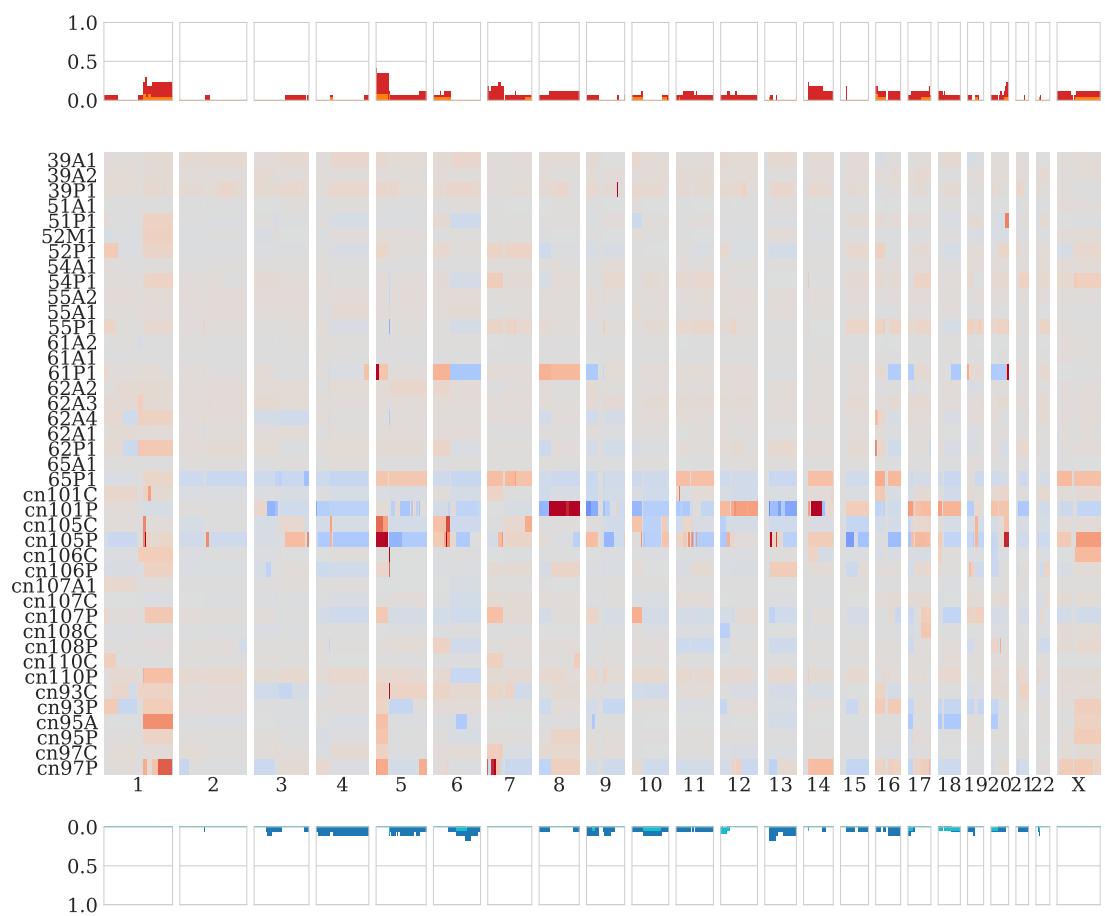


Figure 12: CNV plot by BWA with ADC

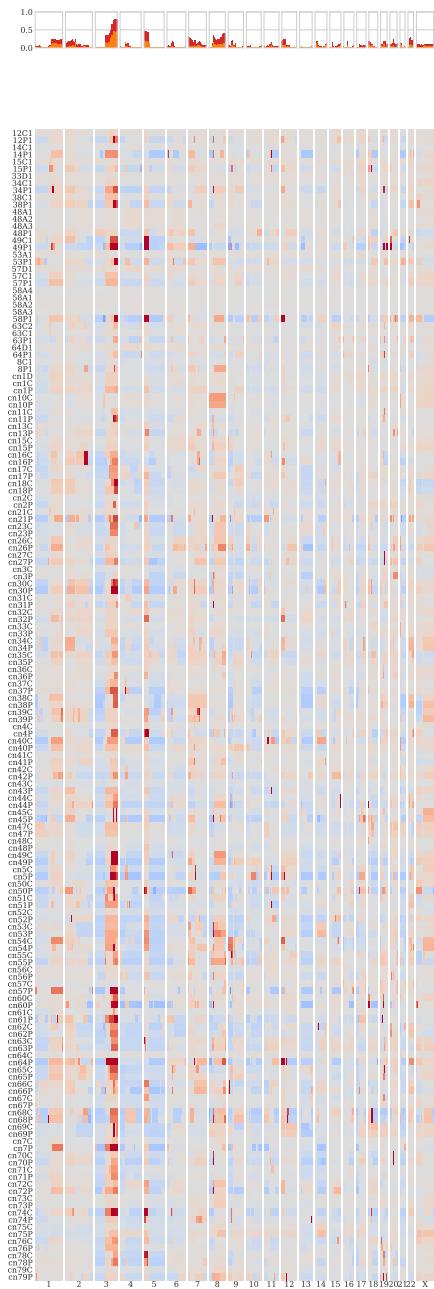


Figure 13: CNV plot by BWA with SQC

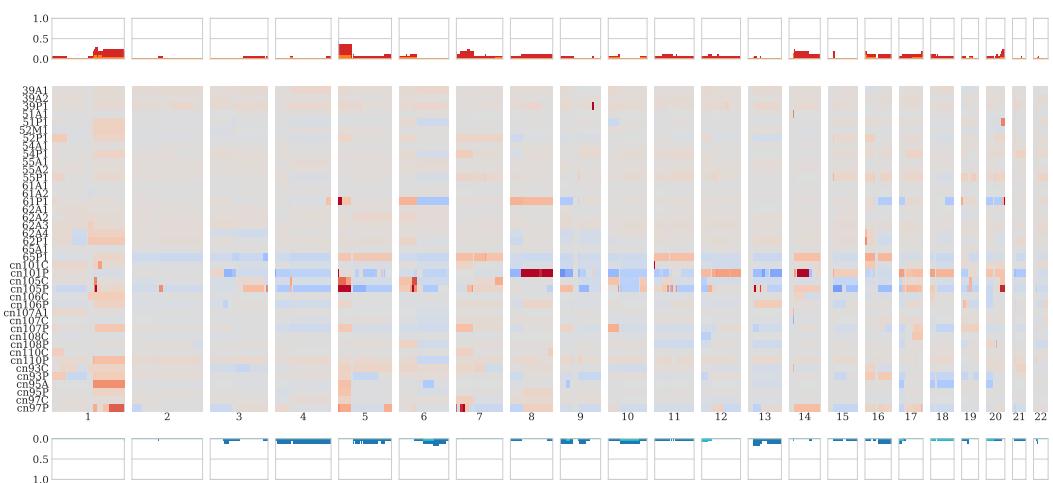


Figure 14: CNV plot by Bowtie2 with ADC

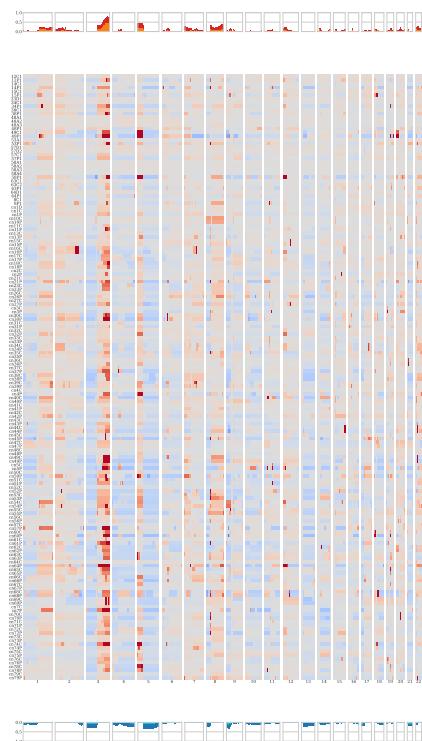


Figure 15: CNV plot by Bowtie2 with SQC

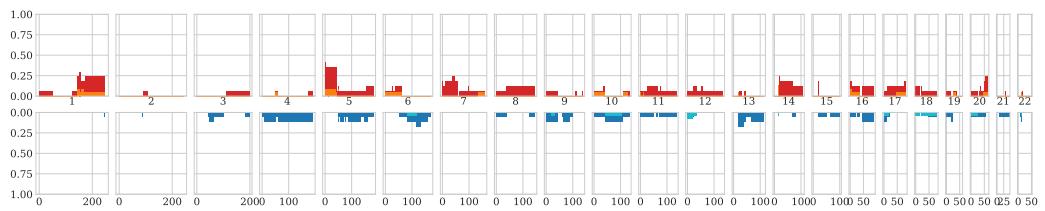


Figure 16: Simple CNV plot by BWA with ADC

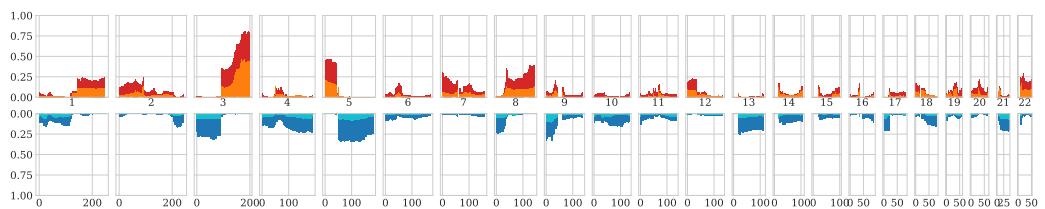


Figure 17: Simple CNV plot by BWA with SQC

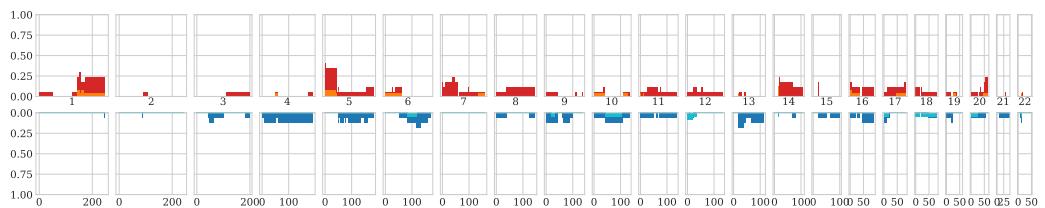


Figure 18: Simple CNV plot by Bowtie2 with ADC

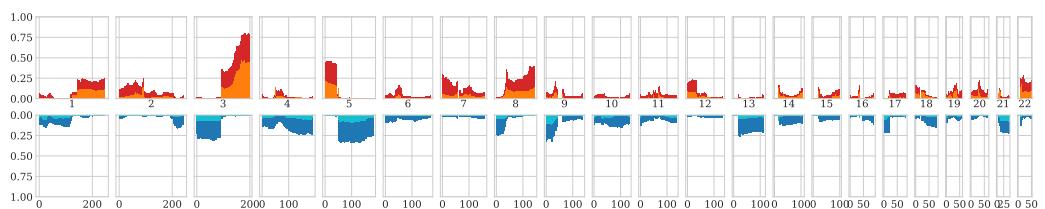


Figure 19: Simple CNV plot by Bowtie2 with SQC

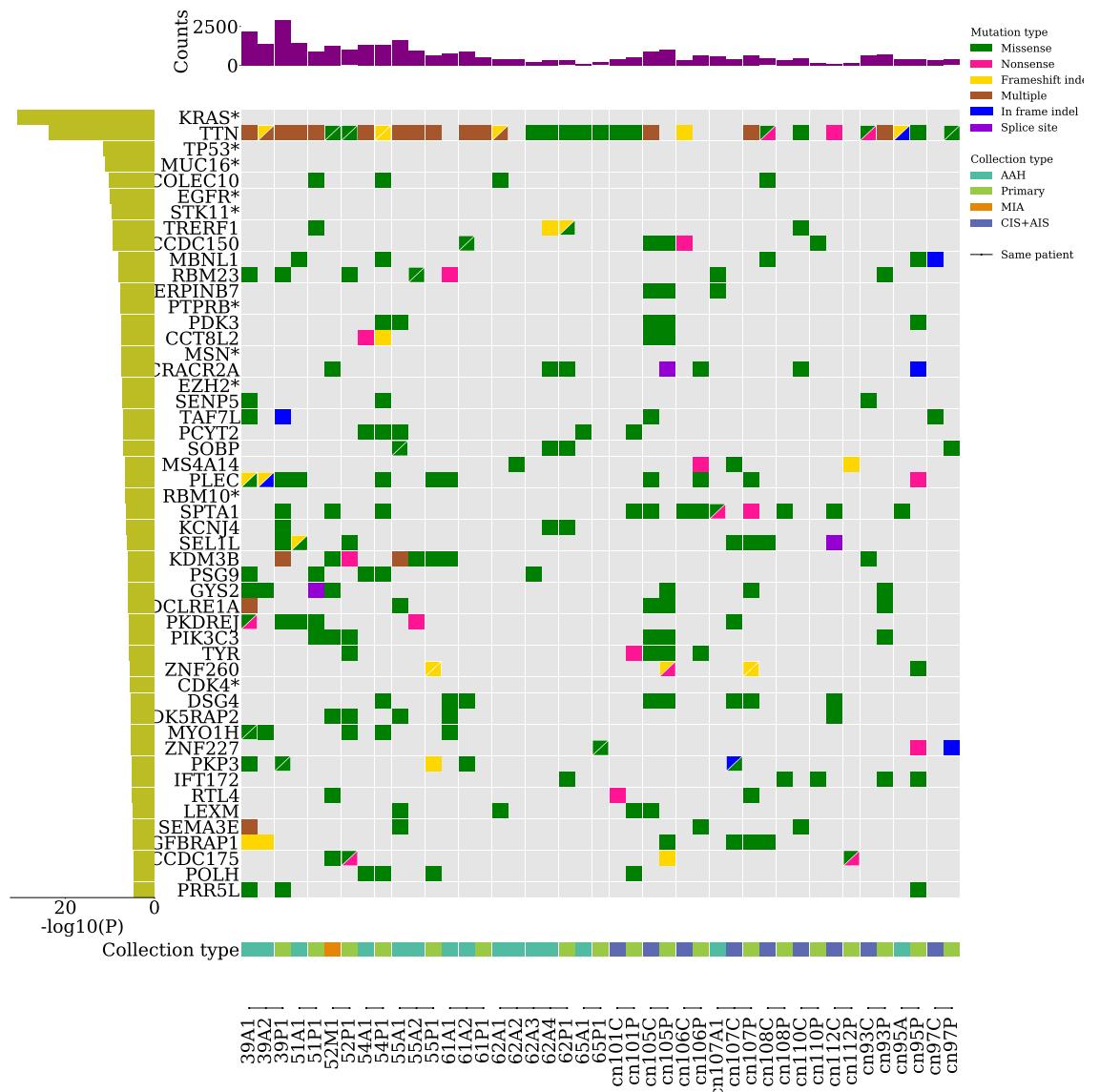


Figure 20: CoMut plot by BWA with ADC

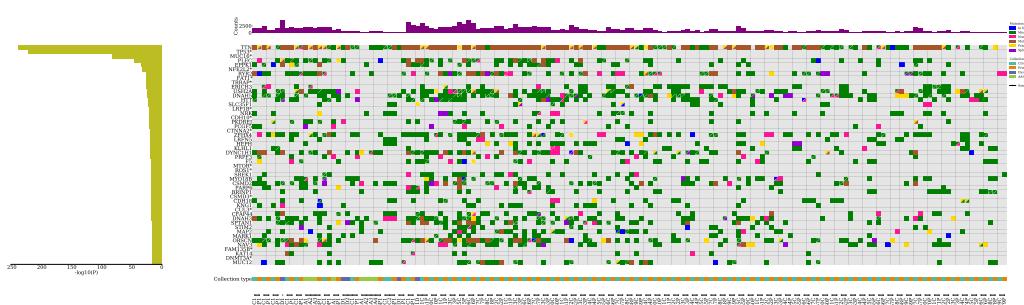


Figure 21: CoMut plot by BWA with SQC

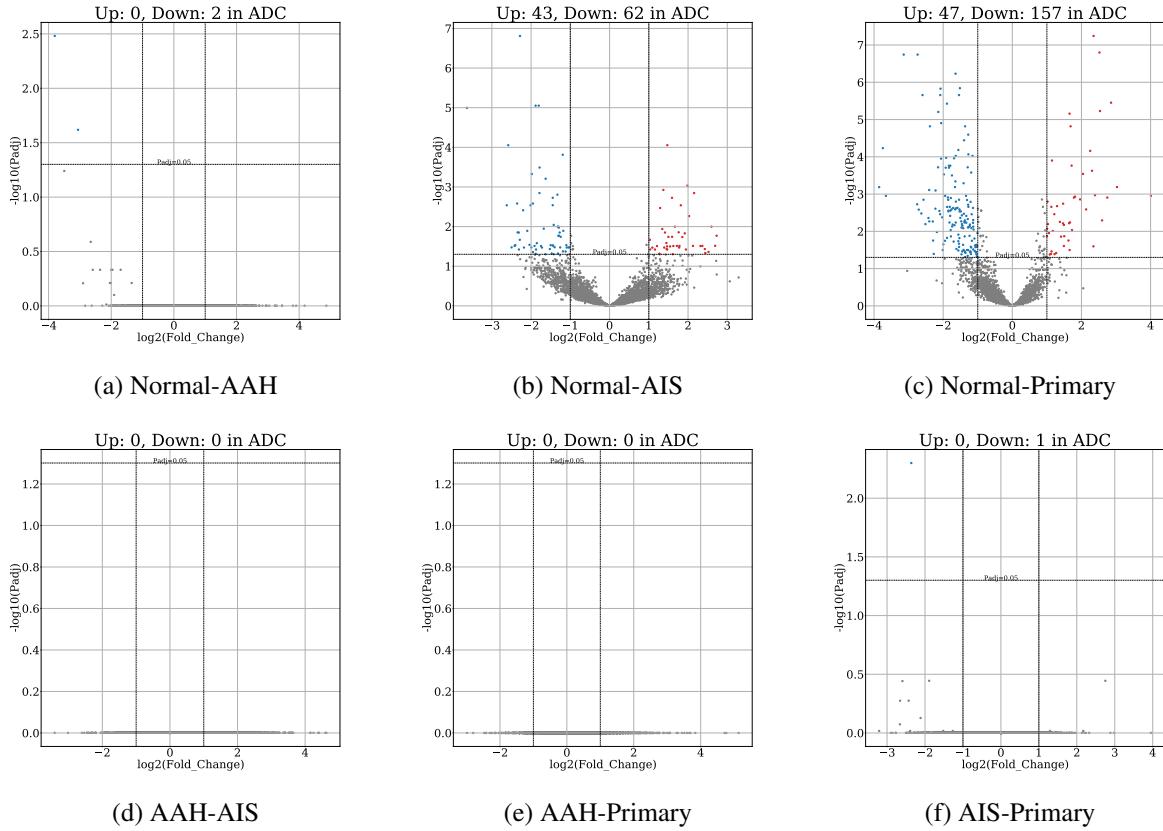


Figure 22: DEG volcano plots by Bowtie2 in ADC

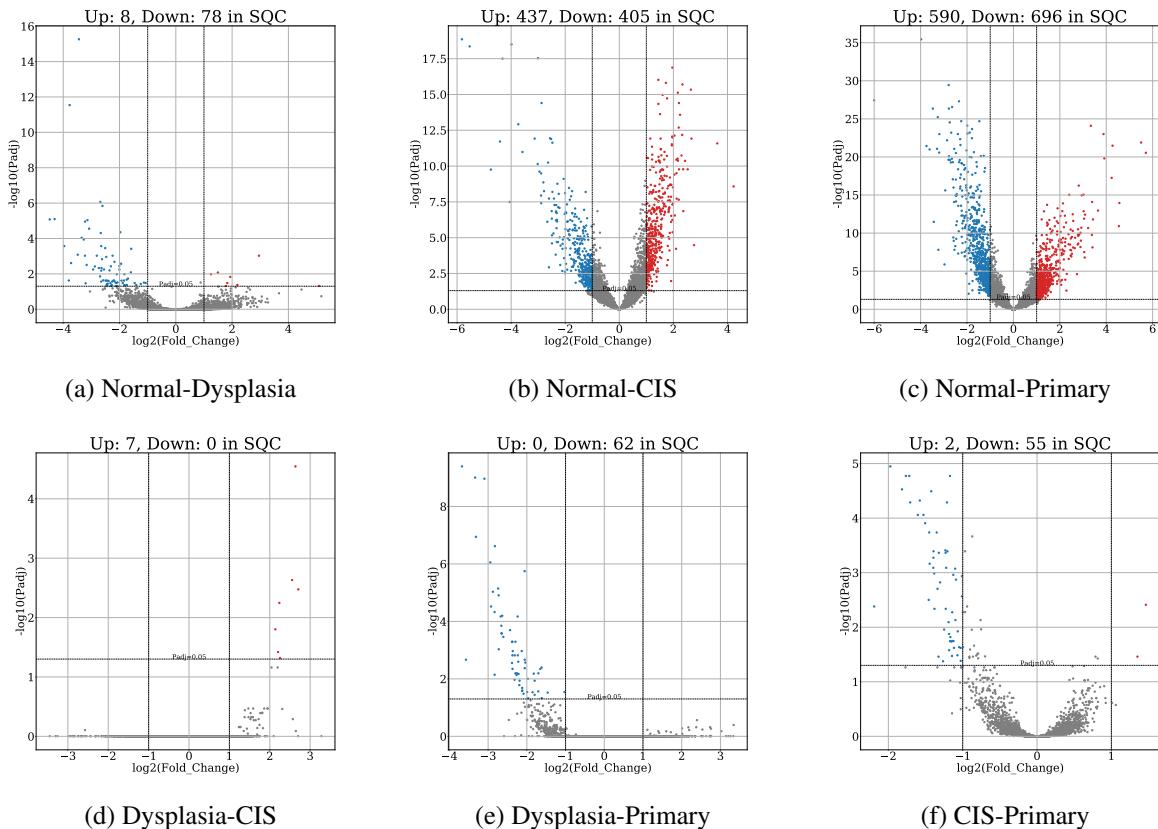


Figure 23: DEG volcano plots by Bowtie2 in SQC

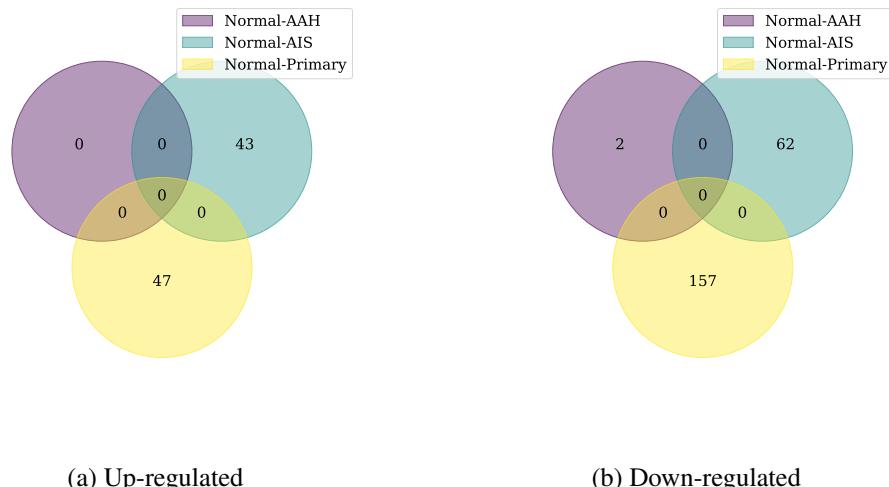


Figure 24: DEG Venn Diagram by Bowtie2 in ADC

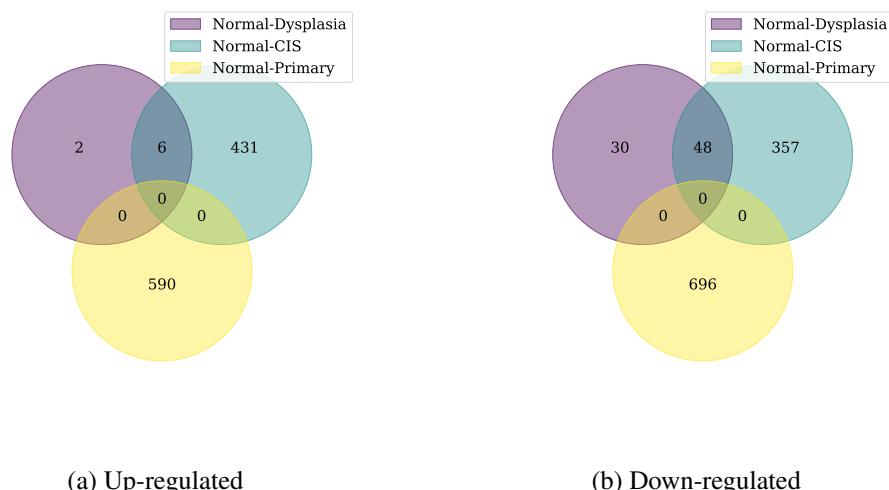


Figure 25: DEG Venn Diagram by Bowtie2 in SQC

## 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

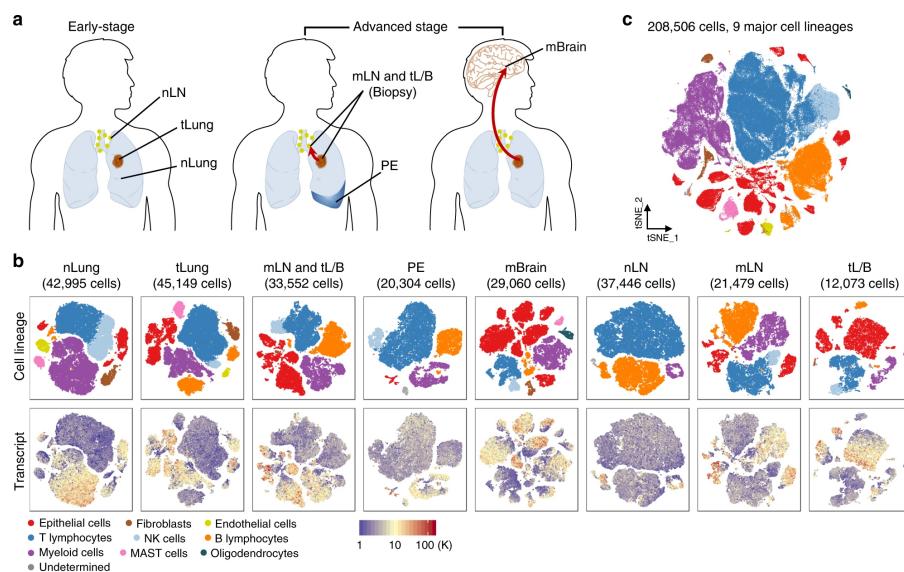


Figure 26: Comprehensive dissection and clustering of 208,506 single cells from LUAD patients (Kim et al., 2020)

# 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.

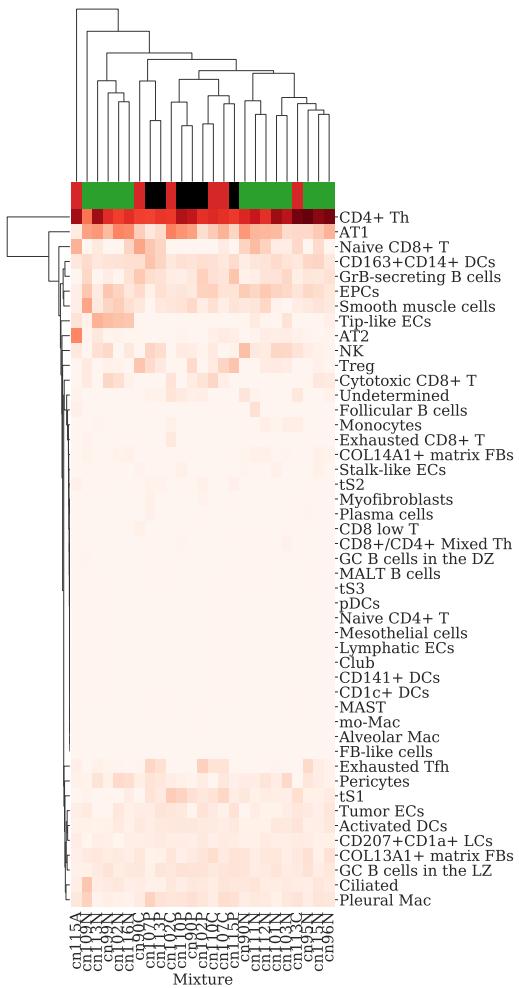


Figure 27: Cell deconvolution clustermap by Bowtie2 in ADC

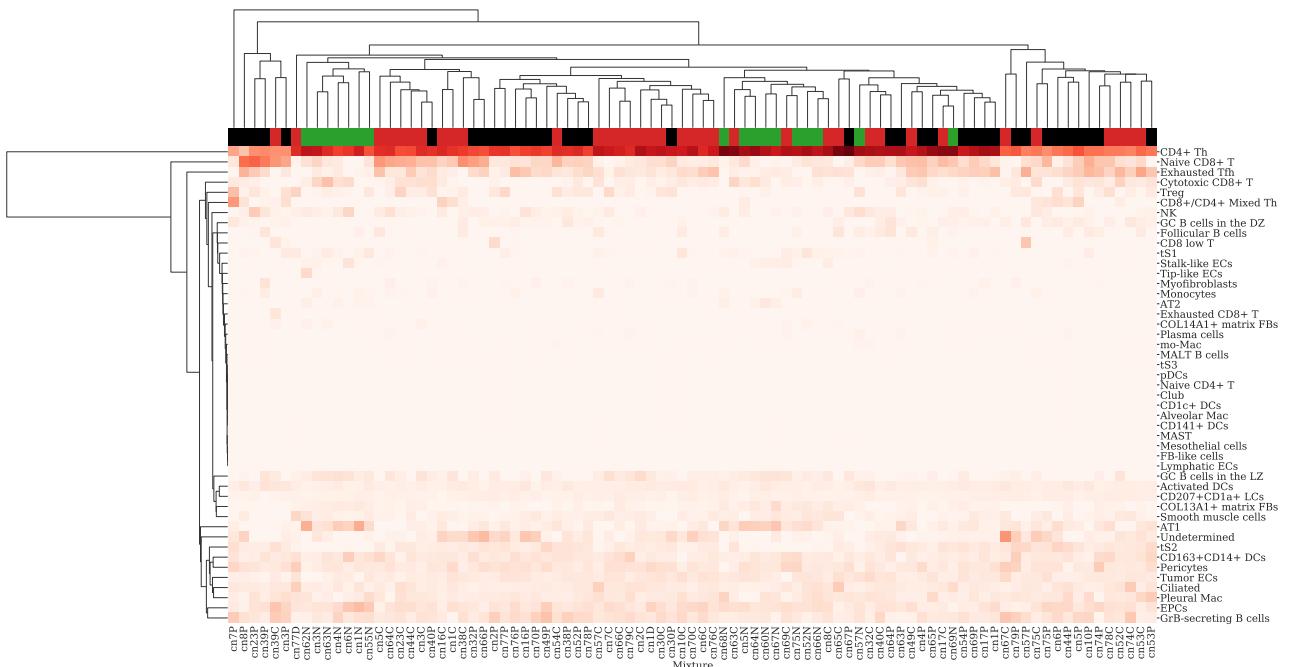


Figure 28: Cell deconvolution clustermap by Bowtie2 in SQC