## Lung Cancer

#### Jaewoong Lee

Ulsan National Institute of Science and Technology jwlee230@unist.ac.kr

2021-01-22

#### Overview

- Introduction
- 2 Materials
- Methods
- 4 Results
- 6 Proceedings

## Introduction

# Objectives

#### Materials

## Lung Cancer Data

- WES + WTS
- $\bullet \ \mathsf{Normal} + \{\mathsf{Primary}, \ \mathsf{CIS} \ \mathsf{AIS}, \ \mathsf{AAH}, \ \mathsf{Dysplasia}, \ \mathsf{MIA}\}$
- Total 112 samples

## Methods

### Somatic short variant discovery

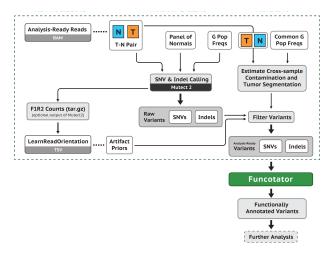


Figure: Somatic short variant (SNVs + Indels) discovery workflow (Van der Auwera et al., 2013; DePristo et al., 2011)

 ✓ □ → ○ □ → ○ □

#### Results

### **FastQC**



Figure: FastQC with WES data

```
Nagaro Ford Strikelin

Formal Read Ford

Ford Strikelin

Ford
```

Figure: FastQC with WTS data

:. All PASS at per base sequence quality!

# Proceedings

#### To-do Lists I

- Alignment WES with BWA
- Alignment WES with Bowtie2

#### References I

- 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.
- 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.