

Lung Cancer

Jaewoong Lee

Ulsan National Institute of Science and Technology

jwlee230@unist.ac.kr

2021-03-19

- 1 Introduction
- 2 Materials
 - Cancer Types
- 3 Methods
- 4 Results
 - Quality Checks
- 5 Proceedings

Introduction

Lung Cancer

Study Objectives

- Find different mutations
 - between WES
 - between WTS
 - from cancer
 - from pre-cancer
- Pathway examine from the mutations
 - of WES
 - of RNA-seq
- Ultra-deep sequencing to find an *infinitesimal* quantity of Non-Circulating Tumor DNA
 - from blood
 - from urine
 - from bronchus
- Diagnostic performance

Materials

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

Materials

Cancer Types

- Carcinoma *in situ* + Adenocarcinoma *in situ*

- Atypical adenomatous hyperplasia

Dysplasia



- Minimally invasive adenocarcinoma

Methods

Data pre-processing for variant discovery

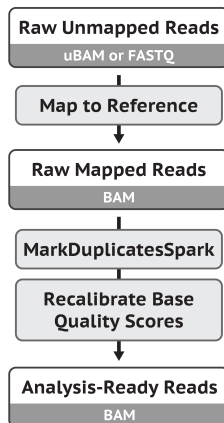


Figure: Data pre-processing for variant discovery (Van der Auwera et al., 2013; DePristo et al., 2011)

Somatic short variant discovery

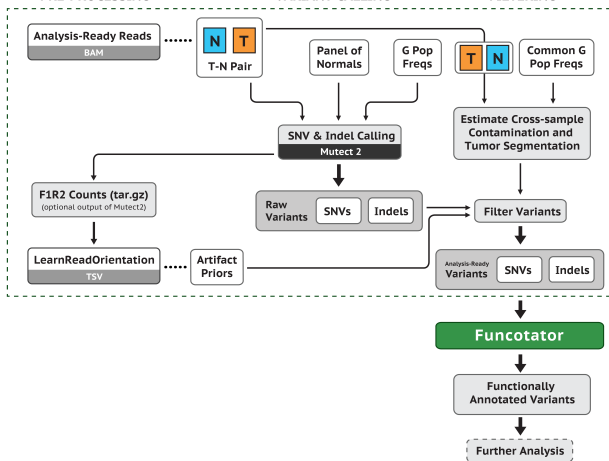


Figure: Somatic short variant (SNVs + Indels) discovery workflow (Van der Auwera et al., 2013; DePristo et al., 2011)

Germline short variant discovery

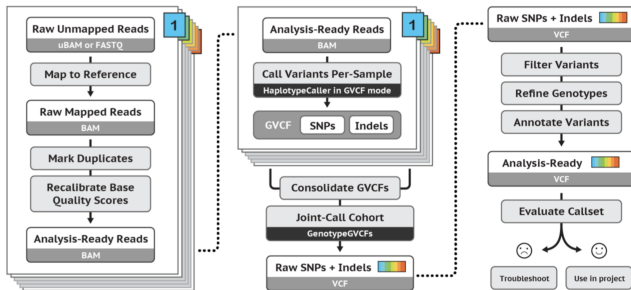


Figure: Germline short variant (SNVs + Indels) discovery workflow (Van der Auwera et al., 2013; DePristo et al., 2011)

RNA-seq short variant discovery

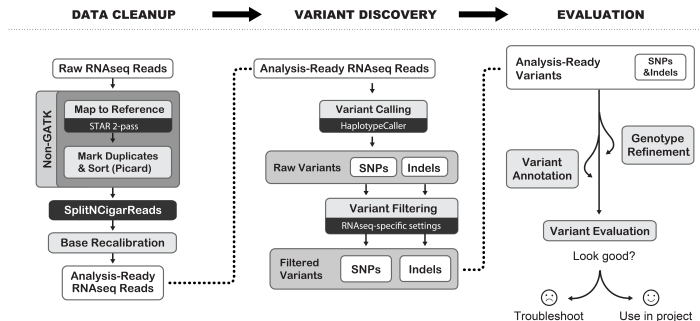


Figure: RNA-seq short variant (SNVs + Indels) discovery workflow (Van der Auwera et al., 2013; DePristo et al., 2011)

Results

Results

Quality Checks

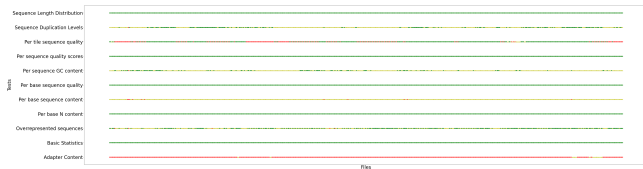


Figure: FastQC with WES data

∴ Only 33P1 has more than 3 failures: 6 FAILs.

∴ 33P1 is excluded at further analysis.



Figure: FastQC with WTS data

- ∴ No sample has more than 5 failures.
- ∴ All sample are good to analysis.

Proceedings

To-do Lists I

- Alignment WES with BWA
- Alignment WES with Bowtie2

- 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.