

Lung Pre-cancer

Jaewoong Lee

Semin Lee

2021-04-23

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 | 3 |
| 3 | Methods | 3 |
| 3.1 | Workflows | 3 |
| 3.2 | List of Bioinformatics Tools | 3 |
| 3.3 | Python Packages | 3 |
| 4 | Results | 3 |
| 4.1 | Quality Check with FastQC | 3 |
| 4.2 | Quality Check with Sequenza | 3 |
| 4.3 | Progressive Evolution at Single Nucleotide Level | 3 |
| 4.4 | Macro-evolution at Chromosomal Level | 3 |
| 4.5 | Selective Clonal Sweep during Neoplastic Evolution | 3 |
| 4.6 | Cancer Gene Mutation during Cancer Evolution | 3 |
| 4.7 | Distinct Drivers and Genetic Constraints in Multifocal IPNs | 3 |
| 5 | Discussion | 3 |
| 6 | References | 3 |

List of Tables

List of Figures

| | | |
|---|---|---|
| 1 | Workflow for data pre-processing for variant discovery (Van der Auwera et al., 2013; DePristo et al., 2011) | 4 |
| 2 | Somatic short variant discovery workflow (Van der Auwera et al., 2013; DePristo et al., 2011) | 4 |
| 3 | Germline short variant discovery workflow (Van der Auwera et al., 2013; DePristo et al., 2011) | 5 |
| 4 | RNA-seq short variant discovery workflow (Van der Auwera et al., 2013; DePristo et al., 2011) | 5 |
| 5 | FastQC results with WES data | 5 |
| 6 | FastQC results with WTS data | 5 |

1 Introduction

1.1 Lung Cancer

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)

2.2 Data Structure & Count

3 Methods

3.1 Workflows

3.2 List of Bioinformatics Tools

3.3 Python Packages

4 Results

4.1 Quality Check with FastQC

4.2 Quality Check with Sequenza

4.3 Progressive Evolution at Single Nucleotide Level

4.4 Macro-evolution at Chromosomal Level

4.5 Selective Clonal Sweep during Neoplastic Evolution

4.6 Cancer Gene Mutation during Cancer Evolution

4.7 Distinct Drivers and Genetic Constraints in Multifocal IPNs

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.

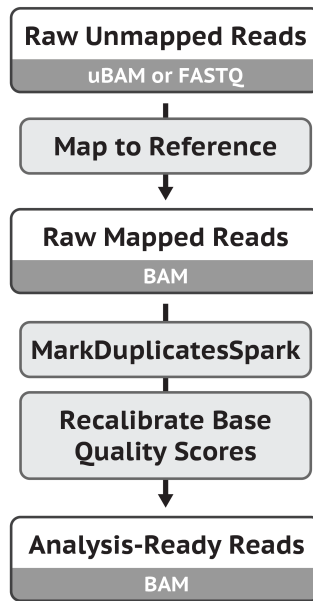


Figure 1: Workflow for data pre-processing for variant discovery (Van der Auwera et al., 2013; DePristo et al., 2011)

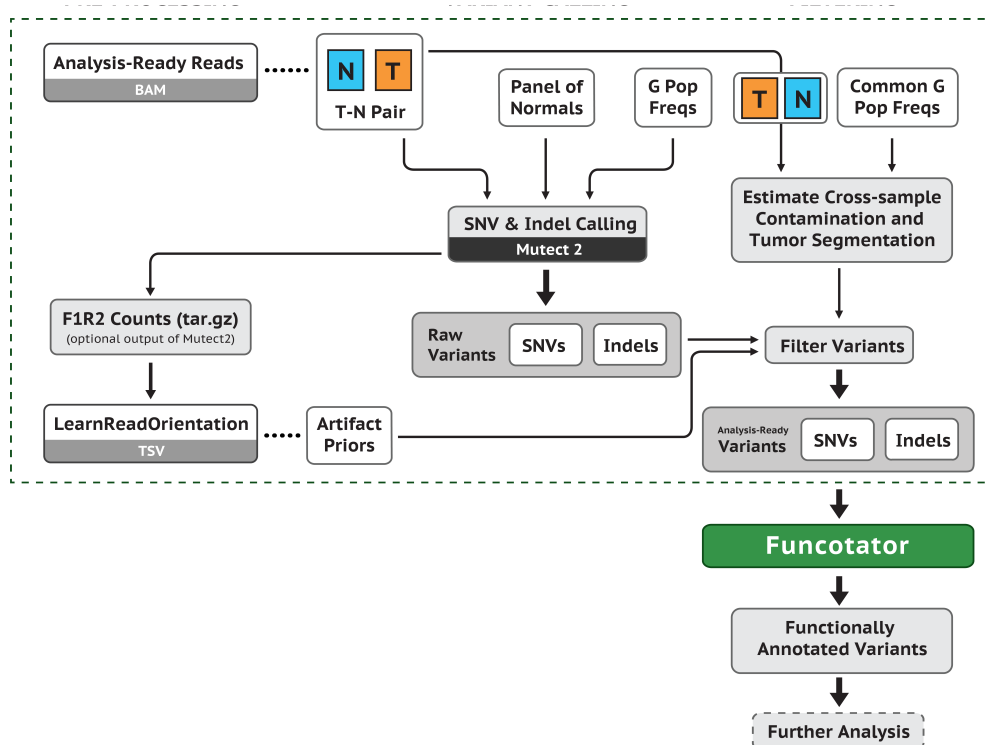


Figure 2: Somatic short variant discovery workflow (Van der Auwera et al., 2013; DePristo et al., 2011)

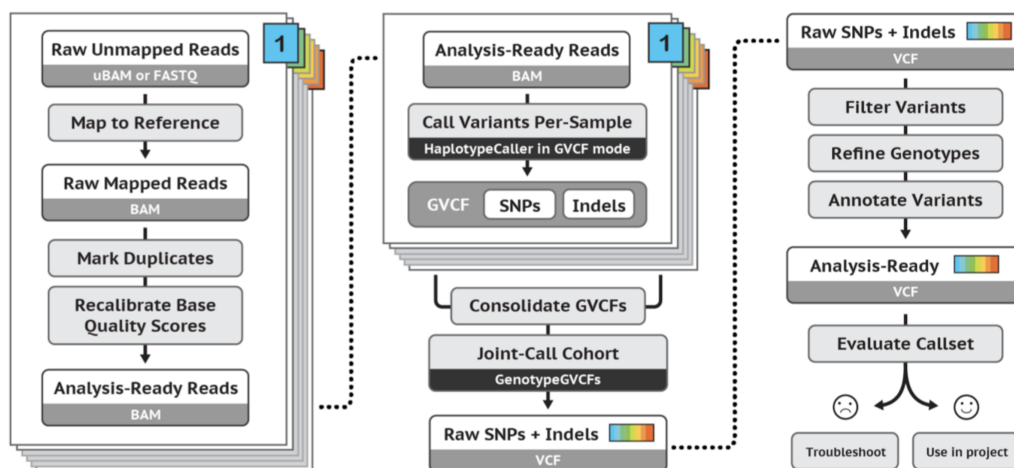


Figure 3: Germline short variant discovery workflow (Van der Auwera et al., 2013; DePristo et al., 2011)

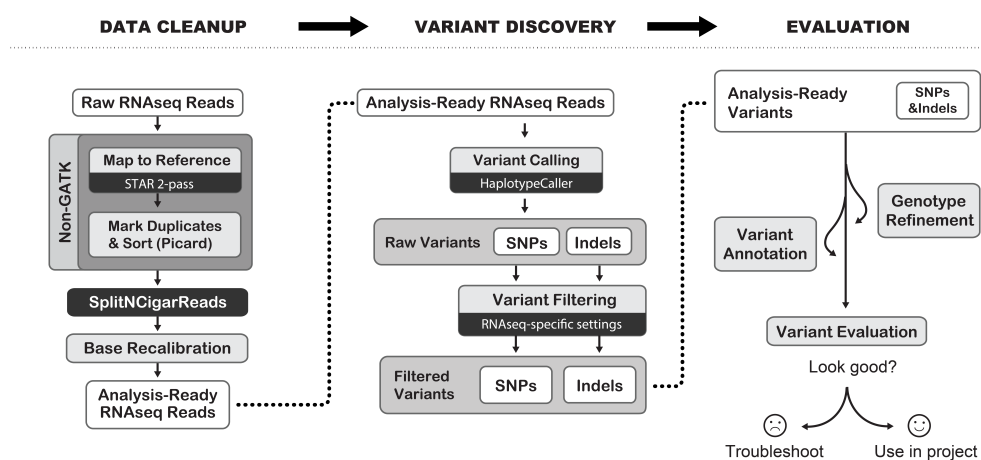


Figure 4: RNA-seq short variant discovery workflow (Van der Auwera et al., 2013; DePristo et al., 2011)

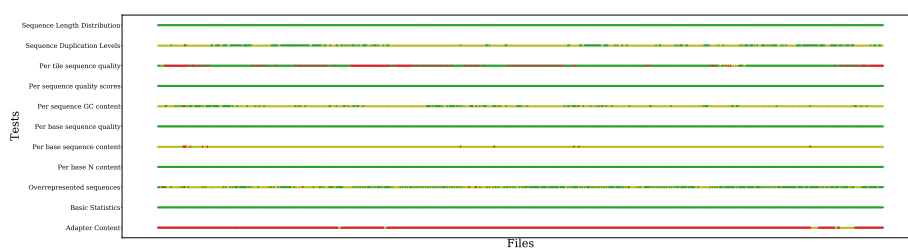


Figure 5: FastQC results with WES data

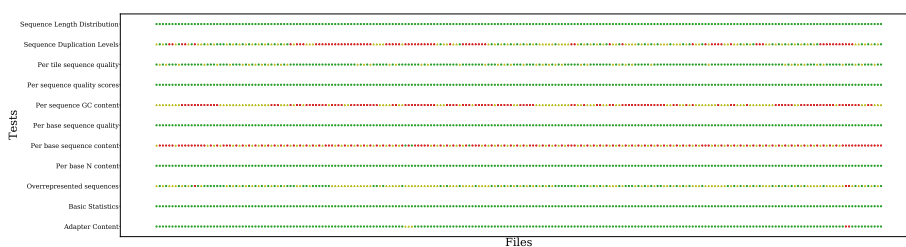


Figure 6: FastQC results with WTS data

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.