

Lung Cancer

Jaewoong Lee

Semin Lee

2021-04-06

Contents

1	Introduction	3
1.1	Lung Cancer	3
1.2	Precancer	3
1.3	Study Objectives	3
2	Materials	3
2.1	Carcinoma <i>in situ</i>	3
2.2	Adenocarcinoma <i>in situ</i>	3
2.3	Atypical Adenomatous Hyperplasia	3
2.4	Dysplasia	3
2.5	Minimally Invasive Adenocarcinoma	3
2.6	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
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 Carcinoma *in situ*

Carcinoma *in situ* (CIS)

2.2 Adenocarcinoma *in situ*

Adenocarcinoma *in situ* (AIS)

2.3 Atypical Adenomatous Hyperplasia

Atypical adenomatous hyperplasia (AAH)

2.4 Dysplasia

2.5 Minimally Invasive Adenocarcinoma

Minimally invasive adenocarcinoma (MIA)

2.6 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

5 Discussion

6 References

- Andrews, S., Krueger, F., Segonds-Pichon, A., Biggins, L., Krueger, C., & Wingett, S. (2012, January). *FastQC*. Babraham Institute. Babraham, UK.
- 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.
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., ... Gingeras, T. R. (2013). Star: ultrafast universal rna-seq aligner. *Bioinformatics*, 29(1), 15–21.
- Favero, F., Joshi, T., Marquard, A. M., Birkbak, N. J., Krzystanek, M., Li, Q., ... Eklund, A. C. (2015). Sequenza: allele-specific copy number and mutation profiles from tumor sequencing data. *Annals of Oncology*, 26(1), 64–70.
- Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with bwa-mem. *arXiv preprint arXiv:1303.3997*.
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with burrows–wheeler transform. *bioinformatics*, 25(14), 1754–1760.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... Durbin, R. (2009). The sequence alignment/map format and samtools. *Bioinformatics*, 25(16), 2078–2079.

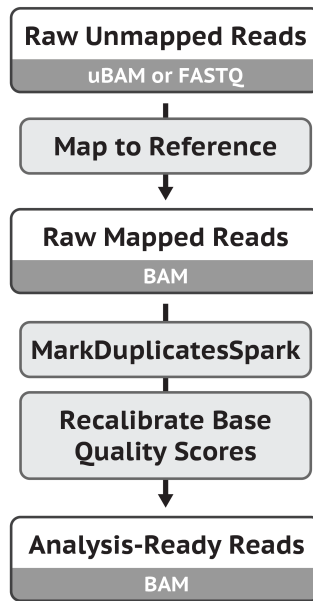


Figure 1: Workflow for data pre-processing for variant discovery (Van der Auwera et al., 2013; DePristo et al., 2011)

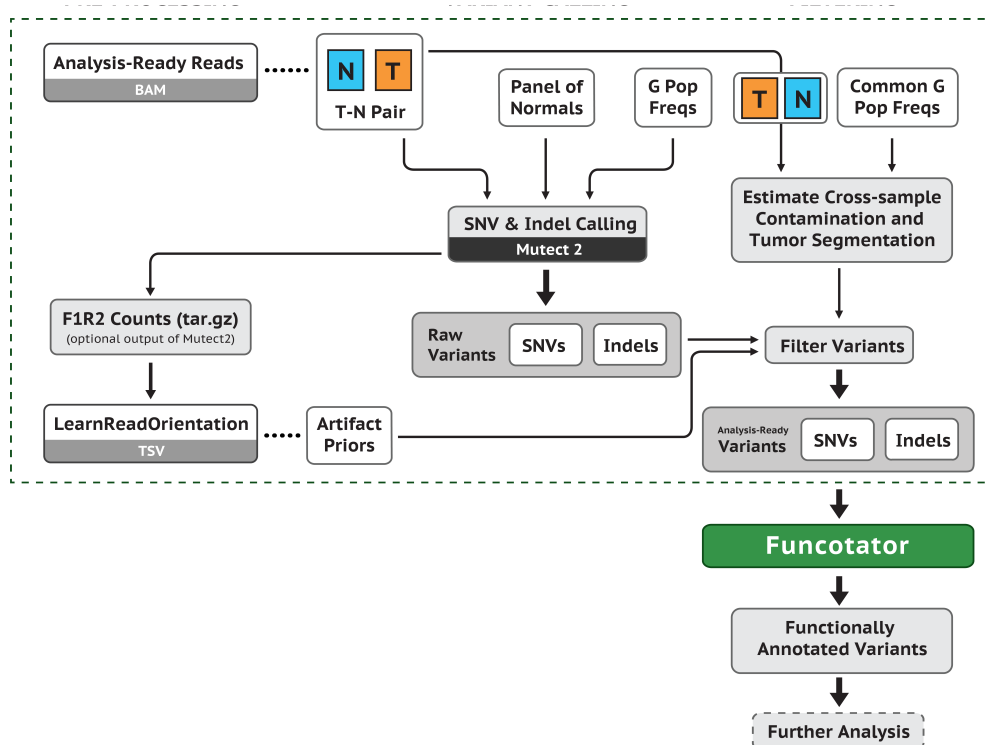


Figure 2: Somatic short variant discovery workflow (Van der Auwera et al., 2013; DePristo et al., 2011)

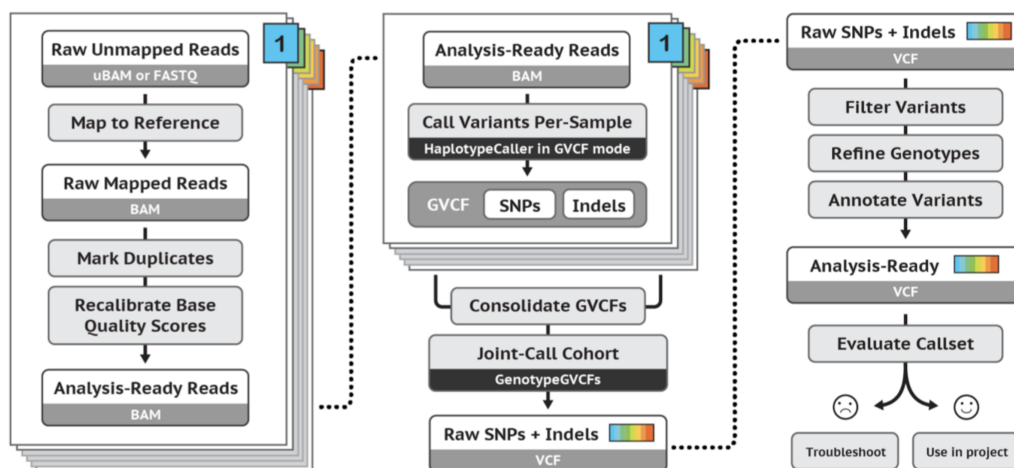


Figure 3: Germline short variant discovery workflow (Van der Auwera et al., 2013; DePristo et al., 2011)

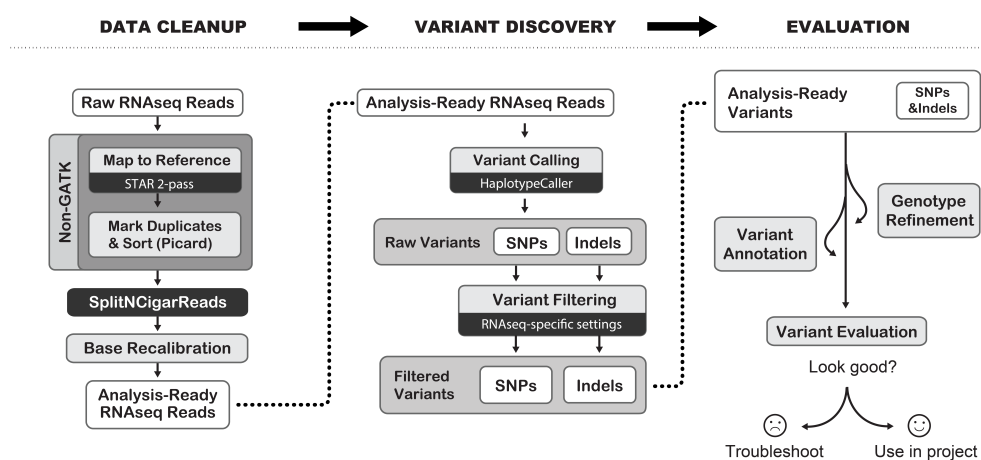


Figure 4: RNA-seq short variant discovery workflow (Van der Auwera et al., 2013; DePristo et al., 2011)

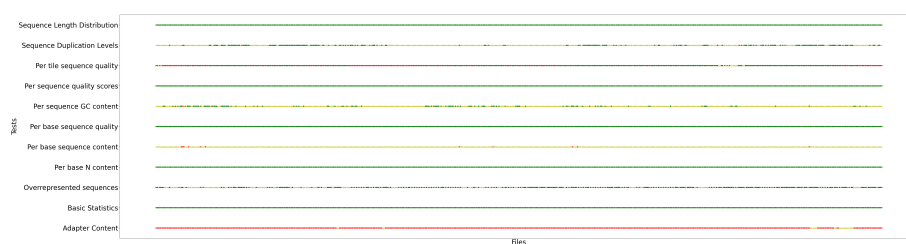


Figure 5: FastQC results with WES data

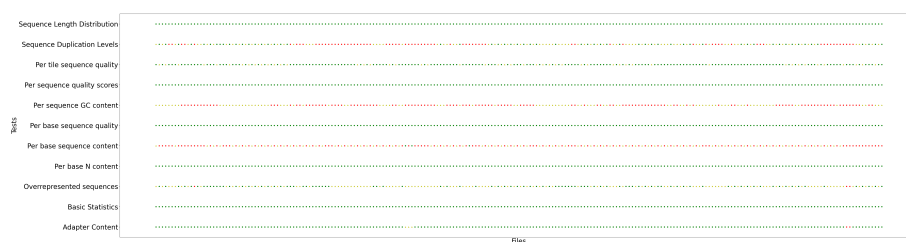


Figure 6: FastQC results with WTS data

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.