Development of duplicated-indel based biomarker for AML diagnosis

Seo Hyein and **Cho Dong-ho**
Korean Advanced Institute of Science and Technology, South Korea

Statement of the Problem: Since the development of the Next Generation Sequencing (NGS), mutations in whole genome of human are investigated and the relationship between mutations and diseases, especially cancers, is studied. Insertion and deletion (Indel) is one of the important mutations. However, analysis result about the relationship between indel and cancer using NGS data is not sufficient. In this research, we want to derive the indel based biomarker that can distinguish whether the sequence is extracted from the cancer cell or not. Among indels, we use the duplicated-indel, which has similar property to microsatellite and exits in gene.

Methodology & Theoretical Orientation: We obtained mutations from all genes of 100 sequences using GATK tool. Sequences were provided by the TCGA project and were sequenced from the cancer cell and the normal cell for 50 Acute Myeloid Leukemia (AML) patients. Among mutations, duplicated-indels, which are frequently found at the cancer cell but not in the normal cell, were listed. By counting the number of listed duplicated-indels, we could generate the biomarker of the cancer.

Findings: In Fig. 1, we used 50 frequently existing duplicated-indels for the cancer distinction. By selecting the diagnosis threshold as 9 (Th1), we could accurately find 12 cancer sequences. Otherwise, if the threshold was 5 (Th2), 38 sequences were rightly extracted as the cancer but there was one wrong diagnosis.

Conclusion & Significance: We showed the probability of the cancer biomarker using the duplicated-indel from NGS data. With optimization of threshold, more accurate biomarker can be derived. Also, we can extend our research to the analysis of the relationship between microsatellite and diseases.

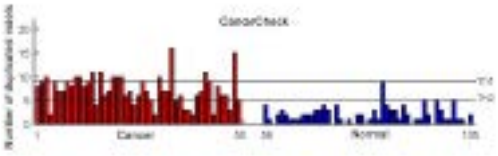

Figure1: *Number of duplicated-indels in NGS data*

Recent Publications

- Sun Q Y, et al. (2017) Ordering of mutations in acute myeloid leukemia with partial tandem duplication of MLL (MLL-PTD). Leukemia 31(1): 1-10.
- Papaemmanuil E, et al. (2016) Genomic classification and prognosis in acute myeloid leukemia. New England Journal of Medicine 374(23): 2209-2221.
- Rabbani B, Nakaoka H, Akhondzadeh S, Tekin M and Mahdieh N (2016) Next generation sequencing: Implications in personalized medicine and pharmacogenomics. Molecular BioSystems 12(6): 1818-1830.
- Carethers J M and Jung B H (2015) Genetics and genetic biomarkers in sporadic colorectal cancer. Gastroenterology 149(5): 1177-1190.
- Ahmed D, et al. (2013) Epigenetic and genetic features of 24 colon cancer cell lines. Oncogenesis 2(9): e71.

Biography

Seo Hyein received Her BS and MS degrees from the School of Electrical Engineering (EE), KNU, Daegu, Korea in 2014 and KAIST, Daejeon, Korea in 2016. Her research interests include WLAN, sensor network and bioinformatics. She is now a PhD student of the School of EE, KAIST.

hiseo2@kaist.ac.kr