Discovery of Microsatellite Instability (MSI) in Cancer Genome

Chuheng Chen, Judy Wang and Jessica Wu
Department of Biomedical Engineering, Johns Hopkins University

March 16, 2018

I. Introduction

DNA mismatch repair (MMR) is the process by which erroneous insertions, deletions, and mis-incorporation of nitrogenous bases are recognized and repaired during DNA replication and recombination. The deficiency of this MMR process may accumulate DNA mutations in the genome, leading to onset of cancer. Due to mutations in the mismatch repair genes such as MSH2 and MLH1, mismatch repair deficiency is characterized by the loss of function of the MMR pathway and can occur in many different cancer types. Dysfunctional repair proteins may cause a significant amount of somatic mutations that will be unrepaired and accumulate in the cancer genome. As a direct result, the errors caused by DNA replication will accumulate in tracts of repetitive genomic material with high mutation rates and high sensitivity to replication errors. These short tandem repeats, also termed *microsatellites*, mutate at rates between 10^3 and 10^6 per cell generation, which is up to 10 orders of magnitude greater than point mutations[3]. The accumulation of replication errors due to lack of a functional repair system is known as microsatellite instability (MSI)[4].

The genome of MSI-induced cancer is usually hypermutated. As a result, MSI cancer cells are easier to be detected and destroyed by the immune system. DNA mismatch repair (MMR) deficiency is one of the best understood forms of genetic instability in colorectal cancer (CRC)[1,2], having been reported to be responsible for nearly 12~15% cases of the CRC[4,5]. Colorectal tumors with MSI have distinctive features, including an increased tendency to appear in the proximal colon, lymphocytic infiltrate, as well as acquire a mucinous ring appearance[6]. The prognosis of patients with MSI tumor is normally better than that of patients with microsatellite stable (MSS) tumor[5]. For example, Dung et al. demonstrated in a 2015 publication that patients with MSI tumor are more responsive to PD-1 blockade therapy than patients with MMR proficient tumors[1].

According to the National Cancer Institute, there are a total of five microsatellite markers necessary to determine MSI - BAT25/26, D2S123, D5S346, and D17S250[7]. Based on their microsatellite characteristics, colorectal cancers are classified into three categories: high-frequency MSI (MSI-H) have 2 or more of these microsatellite markers show instability, low-frequency MSI (MSI-L) shows instability in 1 of these markers, and MSS show instability in none of its markers[7]. As MSI is present in 80-90% of tumors, it represents a marked phenotypic characteristic of colorectal cancer and a significant marker.

Objective The objective of this project is to develop a statistical null model that can detect MSI in tumor genomes. Using TCGA cohorts, we will align sequencing reads from matched normal and tumor genomes with standard tools and identify indels, which refers to an insertion or deletion of bases in a genome with respect to its matched genome. We will then find indels present in at least two reads and compare the distribution of these indels in MSI and MSS tumor genomes to build the null model. Finally, we will compare our outputs to published literature to make conclusions about the significance of our results.

II. Methods

1. Obtain whole exon CRC cancer and matched normal tissue data from the TCGA database, as is done in the paper of Kawakami et al[2].
    a. Use existing packages such as Bowtie to read and align the sequence, in the process converting BAM to FASTQ files.
    b. Use Tcgabiolinks to identify whether the acquired sequence is MSS or MSI
    c. Identify highly repetitive regions (mononucleotide tracts.)
    d. Find indels in microsatellite unstable sequences with respect to microsatellite stable sequences.
    e. This alignment process is done to find locations that do not match the reference genome, which will be our indel regions. These indel regions will be reported back to us by Bowtie2, after which we will quality check the indels by only examining those that occur in at least two reads.
2. Compare the distributions of indels in MSI and MSS genome.
    a. Statistical significance for indel counts in each group will be compared: *Are they distributed differently?*
3. Develop a null model of MSI and a statistical test
    a. Train the model with the data acquired above
    b. Test the results on new genome data

III. References

[1] Dung T. Le, Jennifer N. Uram, Hao Wang, Bjarne R. Artlett, Holly Kemberling, Aleksandra D. Eyring, Andrew D. Skora, Brandon S. Luber, Nilofer S. Azad, Dan Laheru, Barbara Biedrzycki, Ross C. Donehower, et al (2015). PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. *New England Journal of Med*, 372, 2509-2520 DOI: 0.1056/NEJMoa1500596.

[2] Percesepe A, Kristo P, Aaltonen LA, Ponz de Leon M, de la Chapelle A Peltomäki P. (1998). Mismatch repair genes and mononucleotide tracts as mutation targets in colorectal tumors with different degrees of microsatellite instability. *Oncogene*, 17(2):157-63.

[3] C. Richard Boland, Stephen N. Thibodeau, Stanley R. Hamilton, David Sidransky, James R. Eshleman, Randall W. Burt, Stephen J. Meltzer, Miguel A. Rodriguez-Bigas, Riccardo Fodde, G. Nadia Ranzani & S.Srivastava (1998). A National Cancer Institute Workshop on Microsatellite Instability for Cancer Detection and Familial Predisposition: Development of International Criteria for the Determination of Microsatellite Instability in Colorectal Cancer. *Cancer Res*, 58(22), 5248-5257

[4] Kawakami, H., Zaanan, A., & Sinicrope, F. A. (2015). MSI testing and its role in the management of colorectal cancer. *Current Treatment Options in Oncology*, 16(7), 30.

[5] C. Richard Boland and Ajay Goel (2010). Microsatellite Instability in Colorectal Cancer. *Gastroenterology*, 2010 June; 138(6): 2073-2087.e3.

[6] Ashktorab, H., Ahuja, S., Kannan, L., Llor, X., Ellis, N.A., Xicola, R.M., Laiyemo A.O., Carethers J.M., Brim H., & Nouraie, M. (2016). A meta-analysis of MSI frequency and race in colorectal cancer. *Oncotarget*, 7(23), 34546–34557.

[7] Graziele Moraes Losso, Roberto da Silveira Moraes, Arthur C. Gentili, Iara Taborda Messias-Reason (2012). Microsatellite instability – msi markers (BAT26, BAT25, D2S123, D17S250) in rectal cancer. *ABCD. Arquivos Brasileiros de Cirurgia Digestiva (Sao Paulo).* Print version ISSN 0102-6720.