Gwangju Institute of Science and Technology
Official Press Release — https://www.gist.ac.kr

Section of Public Relations	Dongsun Cho Section Chief 062-715-2061	Nayeong Lee Senior Administrator 062-715-2062
Contact Person for this Article	Professor Hyunju Lee School of Electrical Engineering and Computer Science 062-715-2213	
Release Date	2021.05.03	

Professor Hyunju Lee's research team develops genetic mutation discovery and genomic restoration algorithms for cancer cells

□ GIST (Gwangju Institute of Science and Technology) School of Electrical Engineering and Computer Science Professor Hyunju Lee's research team developed a new graph-based algorithm that analyzes the whole genome sequencing data* to discover genetic variations and restores the genome structure to the level of a single nucleotide sequence.

◦ The research team succeeded in accurately discovering genetic mutations from whole genome sequencing data with many mutation detection errors and identifying rearranged genome structures that were not found in cancer patients by conventional methods.

* whole genome sequencing data: data providing the base sequence of an individual's entire DNA

□ The human genome is composed of 3 billion nucleotide sequences, and, in the case of cancer cells, genetic mutations different from those of normal cells exist. It is important for personalized treatment to accurately identify different genetic variations according to individual cancers.

- However, it is a very difficult task to accurately identify mutations in cancer cells by analyzing 3 billion nucleotide sequences. In particular, the rearranged chromosome structure, which has been microscopically observed in cancer cells in the past, has yet to be identified at a single-sequence level. Therefore, an algorithm that can analyze the full-length genome and understand it is needed.

- The research team developed Integrative Framework for Genome Reconstruction (InfoGenomeR), an algorithm for discovering genetic mutations and reconstructing genomes and converting base sequences with structural mutations into graphs. The detection error was reduced by reconstructing the graph so that the structural variation and the copy number variation have consistent values.

  - Subsequently, after constructing a haplotype graph using heterozygous single nucleotide polymorphism information, the genome arrangement was restored by finding the Euler path with the minimum entropy value.

- The InfoGenomeR developed by the research team significantly reduced the detection error of genetic mutations (InfoGenomeR's structural mutation detection accuracy was 98.1% and F-measure was 94.9%). The genomic arrangement of cancer cell lines was restored to the level of a single nucleotide sequence.

  - The detection accuracy of genetic mutations was significantly improved compared to Manta's algorithm of Illumina, an international genome analysis company (Manta's structural mutation detection accuracy was 94.2% and F-measure was 90.4%).

- By applying the developed algorithm to breast and brain cancer patient data, the research team found that the circular genome structure (double minute) was amplified dozens of times in patients. This revealed that certain chromosomes were rearranged for each type of cancer. In addition, it was discovered that recurrent or metastatic cancers showed newly re-arranged genomes that were not present in the existing cancer site.

- Professor Hyunju Lee said, "It is a challenging problem to restore the genome sequence of cancer cells to the level of a single nucleotide sequence using only full-length genome data. Although not possible with existing algorithms, InfoGenomeR is the first algorithm to successfully do this. It is expected that the expression regulation of cancer-related genes for personalized medicine can be identified based on the genomic arrangement of cancer cells resulting from this algorithm."

- This research was led by GIST Professor Hyunju Lee and conducted by integrated student Yeonghun Lee with support from the Institute of Information & communications Technology Planning & Evaluation (IITP) and the National Research Foundation of Korea and was published online on April 29, 2021, in the world-renowned journal, *Nature communications*.

GIST
Since 1993