Fragmentomics: An R Package For Integrating Cell-Free Dna Fragment Features With Mutational Status To Support Liquid Biopsy Interpretation

Abstract Submitter: Killian Maudet, France*

Co-Authors: Juliette Samaniego, Yoann Pradat, Elsa Bernard

*UMR 981, Gustave Roussy, Villejuif, France and IHU PRISM National PRecISion Medecine Center in Oncology, Gustave Roussy, Villejuif, France

Abstract

Background

Plasma circulating cell-free DNA (cfDNA) analysis has transformed cancer care by enabling the detection of circulating tumor DNA (ctDNA), facilitating cancer diagnosis, treatment selection, and disease monitoring. The majority of cfDNA originates from hematopoietic cells, complicating the interpretation of ctDNA results in clinical practice, in the absence of matched white blood cells sequencing.

Objective

Recent work has demonstrated that ctDNA fragments have distinct size distribution profiles and 5'/3' end sequences compared to healthy cfDNA fragments. We aimed to develop a unified computational framework that integrate features of cfDNA fragments (size, end sequences) with their mutational status (tumor drivers or clonal hematopoiesis mutations).

Methods

We developed fRagmentomics, a user-friendly R package, that enables the characterization of each cfDNA fragment overlapping one or multiple mutations of interest, starting from a sequencing file containing aligned reads (BAM file). It supports multiple mutation input formats (e.g., VCF, TSV, or string "chr:pos:ref:alt" representation), accommodates one-based and zero-based genomic conventions, handles mutation representation ambiguities, and accepts any reference file and species in FASTA format. For each cfDNA fragment, fRagmentomics outputs its size, its 3' and 5' sequences, and its mutational status. The package also enables pipeline integration, ensuring scalability and ease of high-throughput analysis.

Results

We applied fRagmentomics to plasma cfDNA samples profiled with a clinical capture-based gene panel (FoundationOne Liquid CDx) from an institutional cohort comprising thousands of samples. We observed shorter fragment size profiles when comparing cfDNA fragments carrying tumor driver mutations (e.g. PIK3CA, KRAS) compared to cfDNA fragments carrying clonal hematopoiesis mutations (e.g. DNMT3A, TET2). At the patient level, fRagmetomics allowed to distinguish cfDNA fragments that were mutated or wildtype for tumor driver mutations within the same sample according to their respective size distribution.

Conclusion

fRagmentomics provides a robust and user-friendly analytical framework to explore cfDNA fragment features such as size and 3'/5' end sequences, integrating their genotype at user-specified positions. This represents a valuable resource to further support the inference of the tumoral or hematopoietic origin of cfDNA variants identified in clinical routine. Ongoing work includes an in-depth investigation of TP53 hotspot mutations, for which the tumor or hematopoietic origin remains ambiguous.

Do you have any conflicts of interest?

No, I do not have a conflict of interest.