Take-Home Assignment – Research Engineering

Objective

Your task is to implement a module for a genomics data pipeline that performs fragment end analysis on cell-free DNA (cfDNA). This module will help extract meaningful fragmentomic signals relevant to cancer detection.

You may use any resources you find useful, including AI and coding assistants, but aim to spend no more than 2 hours on the core implementation. Any questions at any time hannah@cfdx.io and will@cfdx.io.

Scenario

Recent studies have shown that fragment end patterns in cfDNA can provide valuable insights for cancer detection. As a new computational biologist in an interdisciplinary engineering team, you need to add a module to their genomics pipeline that performs fragment end analysis.

Requirements

- Your module must take a BAM file as input and output a processed BAM file.
- Your module must perform fragment end analysis using a published tool or algorithm.
- Your module must generate a basic statistical summary of the processed data, including:
    - Fragment length distribution
    - Start and end position distributions
    - End Motif distribution
    - Any other useful statistics

Guidelines

- You can use any language or framework.
- Ensure that we can run the code easily. If special dependencies or installation steps are needed, include a README.md with clear instructions.
- Please provide access to the code and any test data used, either via a GitHub repository or an attached ZIP file.
- If you are unable to fully complete the task in the suggested time, document how you would extend the solution if given more time.