Software requirement specification v1.0

For: QuicK-mer2

By: Anthony Nguyen

1. Introduction

This is an extension of the tool QuicK-mer2, developed by Dr. Jeffrey Kidd (<https://github.com/KiddLab/QuicK-mer2>). This extension has two phases: human genome comparison and implementation onto Cavatica, a browser used specifically by the Gabriella Miller Kids First data project.

1. Overall Description

In the first phase, a new extension of QuicK-mer2 will be developed, in which human samples processed by QuicK-mer2 using GrCH38 will be compared against samples derived from the 1000 Genome Project for comparison to determine rare duplications and deletions.

Input: a BAM or CRAM file of a human sample into QuicK-mer2, converted into the BED file produced by QuicK-mer2

Output: A BED file describing each determined duplication and deletion (based on CN that either exceed 2.5 or are below 1.5 in copy number) and how many 1000 genomes samples contain the same deletion/duplication (as a %)

In the second phase, QuicK-mer2 will be implemented onto CAVATICA, a data browser used by the Gabriella Miller Kids First Data Resource Center. This will allow non-bioinformatics users to access QuicK-mer2 straight from their browser to determine copy number for personal projects, rather than requiring the download of both samples and tool onto a local Linux workspace.

1. Requirements

Input: A BAM or CRAM file of a human sample, or the BED file produced by said sample after already processing through QuicK-mer2.