DeepSV pseudo code

February 5, 2018

1 Prediction

function DEEPSV_CALLER($reads, reference, representation, model$)
    $calls \leftarrow$ SET
    // We need the mean and standard deviation of the read and insert size
    $stats \leftarrow$ READ_STATISTICS($reads$)
    // Classical algorithm family used in Pindel, GenomeSTRIP, Delly. . .
    $candidates \leftarrow$ ANOMALOUS_READ_PAIRS_CLUSTERING($reads, stats$)
    for $candidate$ in $candidates$ **do**
        $processed\_reads \leftarrow$ REPRESENTATION($candidate, reads, stats$)
        $proba \leftarrow$ MODEL($processed\_reads$)
        if $proba \geq 0.5$ **then**
            ADD($calls, candidate$)
        end if
    end for
    $located\_callss \leftarrow$ SET
    for $call$ in $calls$ **do**
        $located\_call \leftarrow$ SPLIT_READ_LOCATION($call, reads, reference$)
        // If we cannot use Split Read to locate the breakpoint, we approximate with the supporting reads
        if $located\_call == NULL$ **then**
            $located\_call \leftarrow$ SUPPORTING_READS_BOUND($call$)
        end if
        ADD($located\_calls, located\_call$)
    end for
    return FORMAT_TO_VCF($located\_calls$)
end function

function REPRESENTATION(candidate, reads, stats)

 $supporting_reads \leftarrow$ GET_SUPPORTING_READS($candidate$)

 $mean\_insert \leftarrow$ MEAN_INSERT_SIZE($stats$)

 $mean\_read \leftarrow$ MEAN_READ_SIZE($stats$)

 $window_size \leftarrow 2 \times (mean\_insert + mean\_read)$

 $leftmost \leftarrow$ LEFTMOST($supporting\_reads$)

 $left\_window \leftarrow$ EXTRACT_READS($leftmost, window\_size, reads$)

 $rightmost \leftarrow$ RIGTHMOST($supporting\_reads$)

 $right\_window \leftarrow$ EXTRACT_READS($rightmost, window\_size, reads$)

 return UNION($left\_wondow, right\_window$)

end function