Googling The CAncer Genome

UMC Utrecht

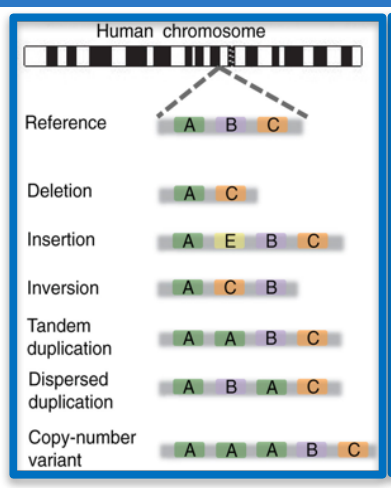
Carl Shneider[1], Jeroen de Ridder[1], Wigard Kloosterman[1], Marleen Nieboer[1], Sonja Georgievska[2], Arnold Kuzniar[2]

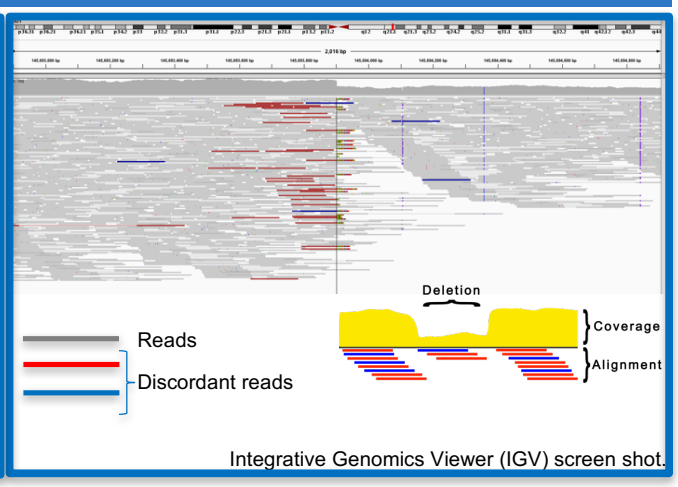
[1]Department of Genetics, UMC Utrecht [2]Netherlands eScience Center, Amsterdam Science Park

Introduction

Cancer affects millions of people worldwide. With the advent of novel DNA sequencing technologies, whole-genome sequencing (WGS) has become part of cancer diagnostics workflow that can potentially enable tailored treatments of individual patients. Processing WGS data from thousands of cancer patients is a major eScience challenge that has not been attempted before.

Structural Variations (SVs) are variants > 50 bp and occur in many forms and sizes. The analysis of SVs in cancer genome sequencing data is the next frontier in cancer genomics and our methods will serve as an important component in future genome-first-based clinical-decision making for cancer patients. Moreover, SVs underlie other human diseases.



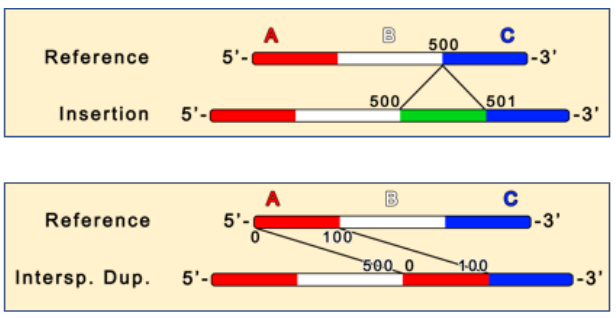


Integrative Genomics Viewer (IGV) screen shot.

M. Baker, Structural variation: the genome's hidden architecture, Nature Methods, 2012.

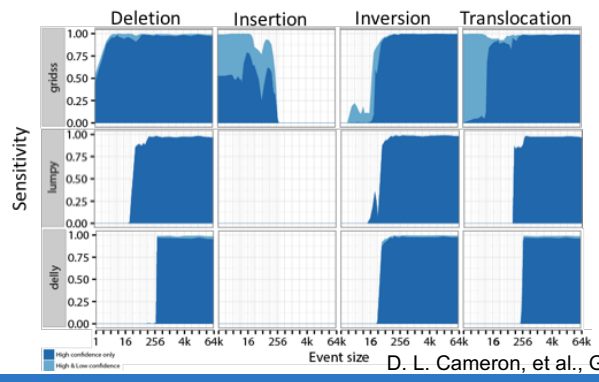
Benchmarking SV Callers

Break Point Junctions



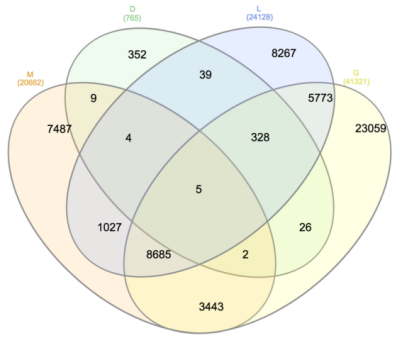
Courtesy Alessio Marcozzi

Sensitivity



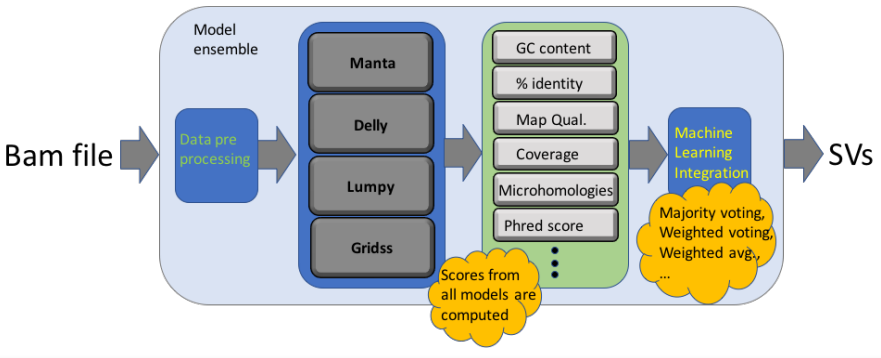
D. L. Cameron, et al., GRIDSS, bioRxiv, 2017

Consensus Problem

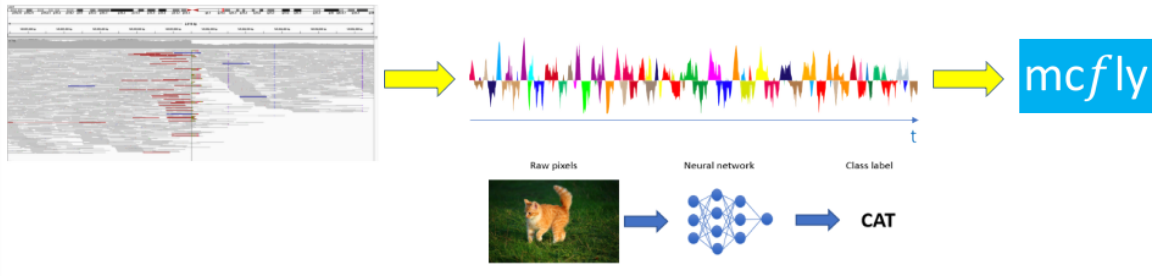


Two Complementary SV Caller Integration Strategies

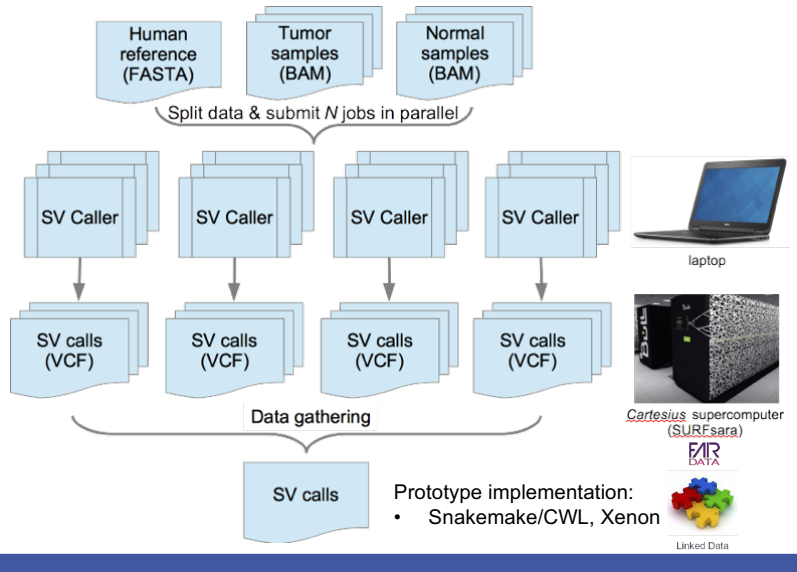
Ensemble Learning



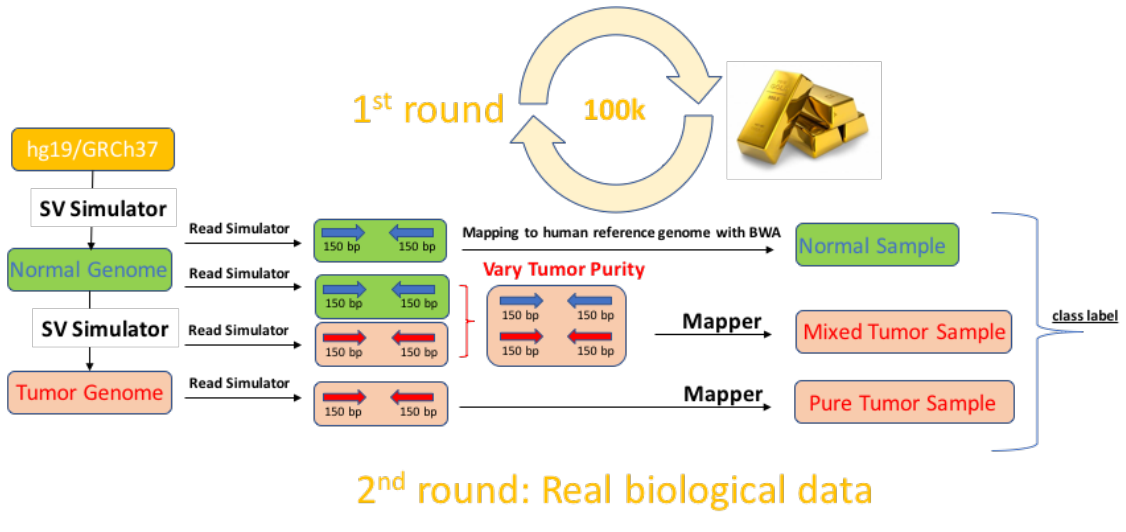
Deep Learning Methodology



SV Calling Workflow



Genomic Variants Simulation Workflow



c.shneider@umcutrecht.nl