

NYGC SV/CNV Pipeline

Manta ([GitHub](#),[Paper](#)) is a structural variant (SV) calling tool developed by Illumina. Manta utilizes discordant read-pair and/or split-read evidence during the discovery phase to call a variety of structural variants including insertions, deletions, translocations, inversions and tandem duplications. Manta also attempts local assembly of insertions in order to provide sequenced resolved calls. Called variations have a minimum size of 50bp and fully assembled insertions have an approximate maximum of twice the read-pair fragment size. Very large insertions are reported when the breakend signature of events support such calls.

Canvas ([GitHub](#),[Paper](#)), also developed by Illumina, is a read-depth based copy number variant (CNV) calling tool that calls large (10kbp+) amplifications and deletions. Canvas relies on a input ploidy VCF to determine the base ploidy for the autosomes and sex chromosomes. We use an internal tool to determine sample sex, which is then translated to the appropriate ploidy VCF in order to correctly call CNVs on the X and Y chromosomes in both female and male samples.

I. Software versions and parameters

- Manta
 - Version: 1.5
 - Mode: single sample
 - Parameters: default
- Canvas
 - Version: 1.40.0.1613
 - Mode: SmallPedigree-WGS (single sample)
 - Parameters: CanvasPartition,-m=CBS ; CanvasBin,-m=0
 - * CanvasPartition - allows more sensitive copy number calling via the Circular Binary Segmentation algorithm
 - * CanvasBin - Calibrates read counting for better visualization
 - Custom Filters:
 - * Filter calls with less than 8 depth bins (BC field in genotype field)

II. DELIVERED FILES

1. Manta annotated vcf (bgzip & tabix indexed)
 - <sampleID>.manta.annotated.vcf.gz
 - <sampleID>.manta.annotated.vcf.gz.tbi
2. Canvas annotated vcf (bgzip & tabix indexed)
 - <sampleID>.canvas.annotated.vcf.gz
 - <sampleID>.canvas.annotated.vcf.gz.tbi

III. Manta VCF

Contains four structural variant types: break ends (BND), deletions (DEL), duplications (DUP), and insertions (INS).

IV. Canvas VCF

Contains two copy number variant types: amplifications (GAIN) and deletions (LOSS). In addition to these, Canvas also provides reference copy number calls where confidence is high.

V. Annotation

Annotations included in the INFO field of delivered VCFs are:

- GeneOverlap - RefSeq gene IDs that the variant overlaps
- ExonOverlap - RefSeq gene exons <geneID:exon#> that the variant overlaps
- Known - SV Databases with an overlapping hit (1000G, DGV, both)
- Mappability - Overlap with UCSC mappability track. Manta - two values, one for each breakpoint. Canvas - single value for complete CNV.
- SimpleRepeat - Overlap with UCSC simple repeat track. Manta - two values, one for each breakpoint. Canvas - single value for complete CNV.
- SegDuplication - Overlap with UCSC segmental duplication track. Manta - two values, one for each breakpoint. Canvas - single value for complete CNV.

Abbreviations

1000G => 1000 Genomes Project call set (phase3)

DGV => Database of Genomic Variants