OICR Genomics
REVOLVE cfDNA Accredited Assay
Analysis Pipeline Deliverables

Sequencing data generated from Targeted Sequencing (TarSeq) libraries (tumour + matched normal) are processed through the consensusCruncherWorkflow. Sequencing data generated from tumour shallow whole genome (sWG) libraries are processed through the ichorCNA workflow.

Pipeline Steps

- **Sequence Data Generation**
- **ConsensusCruncherWorkflow (TarSeq)**
- **ichorCNA (shallow WG)**

Details for each pipeline step are provided below including

- Description
- Output files
- Resources
    - GSI workflows repositories. This provides precise information on how the tools were run. Analysis is described and implemented in wdl (workflow description language) for processing under cromwell. Software and data resources are installed on our system as environmental modules.
    - Software pages. Links to manuals and help pages for the various software tools run in our workflows
    - File Type descriptions. Links to pages describing the format of regular file types. This information may also be available on pages for each particular tool

OICR Genomics
REVOLVE cfDNA Accredited Assay
Analysis Pipeline Deliverables

Sequence Data Generation (shallow whole genome and targeted seqeuncing libraries)

Generation of demultiplexed fastq records from illumina run folders for sequence data generated at OICR. If data had been generated elsewhere, then fastq files will be injected into our analysis system

Output Files:

1. Raw sequence data, paired end (.fastq.gz)

Resources:

1. Workflow : https://github.com/oicr-gsi/bcl2fastq
2. bcl2fastq software : https://support.illumina.com/sequencing/sequencing_software/bcl2fastq-conversion-software.html
3. Fastq format : https://support.illumina.com/bulletins/2016/04/fastq-files-explained.html

OICR Genomics
REVOLVE cfDNA Accredited Assay
Analysis Pipeline Deliverables

ConsensusCruncherWorkflow

Raw sequence data (fastq) is aligned to the hg38 genomic reference with consensusCruncher fastq2bam. This first extracts UMIs from the reads, capturing information into the read names, prior to mapping with bwa. The aligned sequence is then processed through ConsensusCruncher consensus to collapse reads and partition into different subsets producing DCS, SSCS and singleton bam files. (see documentation). Variant calls are generated from each bam using mutect2 in single sample, tumour-only mode, then processed through gatk FilterMutectCalls to populate the FILTER column. The resulting vcf files are merged using gatk combineVariants then annotated with variantEffectPredictor.

Output Files:

1. umi-extracted, uncollapsed aligned reads (ID.bam) + index (.bai)
2. duplex-consensus variant calls (ID.dcs.sc.sorted.mutect2.tumor_only.filtered.vcf.gz) + index (.tbi)
3. single-strand-consensus variant calls (ID.sscs.sc.sorted.mutect2.tumor_only.filtered.vcf.gz) + index (.tbi)
4. all-unique-consensus variant calls (ID.all.unique.dsc.sorted.mutect2.tumor_only.filtered.vcf.gz) + index (.tbi)
5. combineVariant merged , vep annotated variant calls (merged.vep.vcf.gz) + index (.tbi)
6. combineVariant merged , vep annotated maf file (maf.vcf.gz)
7. consensuscruncher.tar.gz, archive of consensusCruncher output directories, containing
    a. partitioned/collapsed mapped reads (bam files + indices)
    b. consensus family size frequency table (read_families.txt)
    c. collapse metrics (stats.txt)

Resources:

Workflow : https://github.com/oicr-gsi/consensusCruncherWorkflow
Reference Sequence : https://hgdownload.soe.ucsc.edu/goldenpath/hg38/bigZips/hg38.fa.gz
samtools : http://www.htslib.org/
bwa : **http://bio-bwa.sourceforge.net**
gatk : https://gatk.broadinstitute.org/hc/en-us
consensusCruncher :
mutect2 : https://gatk.broadinstitute.org/hc/en-us/articles/360037593851-Mutect2
sam/bam specifications : https://samtools.github.io/hts-specs/SAMv1.pdf

OICR Genomics
REVOLVE cfDNA Accredited Assay
Analysis Pipeline Deliverables

ichorCNA Workflow

Raw sequence data (fastq) from shallow whole genome sequencing is aligned to the hg38 genomic reference with bwa mem. Read counts are extracted from the mapped sequence and analyzed with ichorCNA for estimation of tumour fraction and detection of large-scale copy number alterations.

Output Files:

1. aligned reads (ID.bam) + index (.bai)
2. Segmentation File (.seg), IGV compatible
3. Segmentation file (.seg.txt), including subclonal status, not IGV compatible
4. Segmentation File (.cna.seg), estimated copy number, log ration and subclone status
5. Converged parameters for optimal solution. (.params.txt)
6. Corrected Depth (.correctedDepth.txt), log2 ratio for each window after GC correction
7. Genome wide plot (_genomeWide.pdf) for optimal solution
8. Genome wide plots (_genomeWide_all_sols.pdf) for all solutions
9. Additional plots bundled (_plots.tar.gz)
10. Saved R image, all data (.RData)

Resources:

1. Workflow : https://github.com/oicr-gsi/ichorCNA
2. Reference Sequence : https://hgdownload.soe.ucsc.edu/goldenpath/hg38/bigZips/hg38.fa.gz
3. samtools : http://www.htslib.org/
4. hmmcopy utils : https://shahlab.ca/projects/hmmcopy_utils
5. bwa : **http://bio-bwa.sourceforge.net**
6. ichorCNA : https://github.com/broadinstitute/ichorCNA
7. sam/bam specifications : https://samtools.github.io/hts-specs/SAMv1.pdf