|  |  |
| --- | --- |
| **UH LAB NGS SOP-3012 Overview BioInformatics** | |
| **Responsible Department: Clinical Laboratory Services - NGS** | **Effective Date: Not Set** |
| **Policy Basis for Procedure**  This SOP describes the processes automatically launch once sequencing runs are complete. There are no deviations from this workflow. | |
| |  | | --- | | **Applicability**  This applies to the bioinformatics personnel of the NGS Lab. | | **Description of Standard Procedure**  **PROCEDURE**    Figure 1: Procedure Overview  1) Demultiplexing with Bcl2Fastq  Demutliplexing DNA   |  |  | | --- | --- | | executable | bcl2fastq/2.17.1.14 | | input | Bcl files, Sequencing Run Folder, RunInfo.xml  RunInfo.xml will be changed to contain the following line if ReadNumber =”2”:  Read Number="2" NumCycles="9" IsIndexedRead="Y"  RunInfo.xml will be changed to contain the following line if ReadNumber =”3”:  Read Number="3" NumCycles="83" IsIndexedRead="N"  Samplesheet with have the following lines added to file  [Reads]  76  75  [Settings]  ReverseComplement,0  Read2UMILength,8 | | output | Demultiplexed R1 and R2 fastq files | | command | bcl2fastq --barcode-mismatches 0 -o /project/PHG/PHG\_Clinical/illumina/$prjid --no-lane-splitting --runfolder-dir $seqdatadir --sample-sheet $newss &> $seqdatadir\/bcl2fastq\_$prjid\.log |   Demutliplexing RNA   |  |  | | --- | --- | | executable | bcl2fastq/2.17.1.14 | | input | Bcl files, Sequencing Run Folder, RunInfo.xml  Samplesheet with have the following lines added to file  [Reads]  76  75 | | output | Demultiplexed R1 and R2 fastq files | | command | bcl2fastq --barcode-mismatches 0 -o /project/PHG/PHG\_Clinical/illumina/$prjid --no-lane-splitting --use-bases-mask Y76,I6N8,Y76 --runfolder-dir $seqdatadir --sample-sheet $newss &> $seqdatadir\/bcl2fastq\_$prjid\.log |   2) DNA Alignment  Align Reads to Reference Genome and Collect Alignment and Sequence QM. Described in more detail in the DNA Analysis SOP Alignment procedure   |  |  | | --- | --- | | executable | nextflow/0.31.0 | | input | Sample sheet, Illumina output directory, reference genome | | output | Bam and quality metrics | | command | nextflow -C nextflow.config run -w $workdir alignment.nf --design design.txt --capturedir $capturedir --capture $capture --input $outdir --output $outnf --markdups $mdup > nextflow\_alignment.log |   3) Tumor Only Variant Detection  Determines variants from tumor samples, where no normal control is available. Described in more detail in the DNA Analysis SOP Tumor Only Variant Detection Procedure   |  |  | | --- | --- | | executable | nextflow/0.31.0 | | input | Output folder from the alignment workflow and design file describing mapping between samples and files | | output | VCF | | command | nextflow -C nextflow.config run -w $workdir tumoronly.nf --design design\_tumor\_only.txt --projectid ${prjid} --input $outnf --output $outnf $pon\_opt > nextflow\_tumoronly.log |   4) Somatic Variant Detection  Determines variants from tumor samples, where a normal control is available. Described in more detail in the Somatic Workflow SOP   |  |  | | --- | --- | | executable | nextflow/0.31.0 | | input | Output folder from the alignment workflow and design file describing mapping between samples and files | | output | VCF | | command | nextflow -C nextflow.config run -w $workdir somatic.nf --design design\_tumor\_normal.txt --projectid ${prjid} --input $outnf --output $outnf $pon\_opt > nextflow\_somatic.log |   5) RNASeq Gene Fusion and Gene Counts  Align reads to Reference Genome, determines variants gene fusion events, variants and relative abundance of genes in RNASeq datasets. Described in more detail in the RNASeq Workflow SOP.   |  |  | | --- | --- | | executable | nextflow/0.31.0 | | input | Sample sheet, Illumina output directory, reference genome | | output | BAM, Gene-Fusions, and Gene Counts | | command | nextflow -C nextflow.config run -w $workdir rnaseq.nf --design design.txt --input $outdir --output $outnf $bamct --markdups skip > nextflow.log |   6) RNASeq Variant Detection  Determines variants from tumor samples, where no normal control is available. Described in more detail in the Tumor-Only Workflow SOP   |  |  | | --- | --- | | executable | nextflow/0.31.0 | | input | Output folder from the alignment workflow and design file describing mapping between samples and files | | output | VCF | | command | nextflow -C nextflow.config run -w $workdir tumoronly.nf --design design\_tumor\_only.txt --genome $index\_path --nuctype rna --callsvs skip --projectid ${prjid} --input $outnf --output $outnf > nextflow\_tumoronly.log |   **IMPLEMENTATION**  The Workflow used in this SOP can be downloaded here: <https://git.biohpc.swmed.edu/brandi.cantarel/clinseq_workflows>  **USAGE**  check\_new\_run.pl  This script checks for new sequencer sample sheet files to identify new sequencer run, using the seuquencer run id (prjid), this script involves the initiation of the workflow described in this SOP by:  init\_workflow.sh -p $prjid -b $execdir -c $prodir -r $refdir  an sbatch script with the run commands will begin with the following script:  sbatch –p 32GB run\_$prjid.sh  Unify data:  1. unify\_case.sh  bash unify\_case.sh -n $caseid -r $Indexpath -a -b $targetbed -k $nucliatoken  Output files: $subject.vcf.gz, $subject.TMB.csv, $subject.cnv.answer.txt, $subject.translocations.answer.txt  $caseid is the project name and $targetbed is the target capture enrichment bed file  Scripts used to create output files  Integration of variants  perl integrate\_vcfs.pl -s $subject -t $tumor\_id -n $normal\_id -r $index\_path -v $rnaseq\_vcf -c $rna\_ntcs  where $index\_path is the directory path to the reference files, $subject is the subject name, $tumor\_id is the internal id of the tumor, $normal\_id is the internal id of the normal, $rnaseq\_vcf is the rnaseq vcf file $rnaseq\_ntct is the rna bam read count file.  Final files are located in /archive/PHG/PHG\_Clinical/cases.  **RESULT REPORTING CRITERIA**  *Organism Identification*  Sequences meeting default parameters for alignment to human reference meets the threshold for inclusion in downstream analysis.  *SNVs and Indels*  SNVs and Indels will be reported if they meet the following criteria: (1) There are greater than 20 reads that cover the genomic region; (2) > 5% MAF for SNV site or >10% MAF for Indels; (3) if a known variation site associated with CANCER (COSMIC Hotspot) >3 supporting reads or if a novel site >8 supporting reads and called by at least two callers; (4) for somatic mutations, the MAF in the tumor samples must be 5 times higher than the MAF of the normal control sample(5) if variant is reported to be in <1% of the general population as defined by EXAC or gnomAD and (6) Clinical relevance will be determined using the Philips Intellispace Knowledgebase software.  *Fusion/Translocation*  Translocation and gene fusion will be reported if they meet the following criteria: (1) breakpoint are on different chromosomes or are on the same chromosome > 10MB in linear distance; (2) if a known fusion then has > 3 RNA supporting reads (3) if a novel fusion > 5 RNA supporting reads (4) is in a gene that is in a known list of known gene fusion partners.  *Structural Variants*  Structural variants identified by delly are kept if they have a depth greater than or equal to 20 and a mapping quality greater than 30. Those identified by svaba are kept if alternate allele depth is greater than or equal to 20 and for indels a minimum length of 20. Pindel keeps structural variants with an alternative allele depth of 10 and a depth of 20 and for indels a minimum length of 30.  *Limits of Detection*  Sensitivity and specificity analysis are based on SNVs >5% MAF, Indels >10% MAF, CNVs >10kb and all variants including fusions in samples with >30% tumor meeting all laboratory quality metrics.  **CALCULATIONS**  N/A  **INTERPRETATION/RESULTS/ALERT VALUES**  N/A  **REFERENCE INTERVALS (NORMAL REFERENCE RANGE)**  N/A  **ANALYTICAL MEASURING RANGE (LINEARITY)**  **N/A** | | **Definitions**  N/A | | **Applicable Forms**   |  |  | | --- | --- | | Document ID# | Title of Document | | **UH LAB NGS SOP-3012** | Overview\_Bioinformatics\_SOP | |  |  | |  |  |   **Related Documents**   |  |  | | --- | --- | | Document ID# | Title of Document | | **UH LAB NGS SOP-3011** | DNA\_Analysis | | **UH LAB NGS SOP-3010** | RNA\_Analysis | |  |  | | | **References** | | **Review, Revision, and Approval History**   |  |  |  |  | | --- | --- | --- | --- | | **History of Document** | | | | | Version | Date | Description of Change | Authored/Revised by | | 1 | 2017/06 | Policy Tech upload, SOP update for UMI | Brandi Cantarel, Erika Villa | | 2 | 2018/05 | Remove filters for off-target variants and overlapping variants – GIT tag version\_0.2.9 | Brandi Cantarel, Erika Villa | | 3 | 2018/08 | Policy Tech upload, demultiplexing changes and snp/indel lod | Brandi Cantarel, Erika Villa | | 4 | 2019/09 | Freebayes, starfusion update, Pipeline update | Brandi Cantarel, Erika Villa | | 5 | 2020/04 | Policy Tech upload, adding structural variant tools, changes made for idt bait panels | Brandi Cantarel, Erika Villa | | | |
| **Contact for Further Information** | |

**Title**: UH LAB NGS SOP-3012 Overview BioInformatics **Version:** 5 **Status:** Approval

**Writer(s)**: Erika Villa (COMPUTATIONAL BIOLOGIST II)

**Approver(s):** Yan Xu (MGR LAB) (04/27/2020)

**Effective Date:** Not Set

**Next Review Date**: Not Set