

Generating NCI Cancer Gene Panel Data Output from Whole Exome Sequence		
Laboratory:	Molecular Characterization Laboratory	
Effective Date:	1/23/2018	Page 1 of 4

SOP Outlines NCI Cancer Gene Panel **Version: 2.0.0**

## 1.0 PURPOSE/SCOPE

This Standing Operating Procedure (SOP) describes procedures for generating NCI Cancer Gene Panel Data Output for reporting in the NCI Patient-Derived Models database as performed by the Molecular Characterization Laboratory (MoCha) at the Frederick National Laboratory for Cancer Research. **This SOP is for research-use purposes only; do not use for clinical sample analysis.**

## 2.0 CAVEATS

**2.1** Reported NCI Cancer Gene Panel data should be considered representative of the patient-derived models provided by the NCI Patient-Derived Models Repository and should not be considered to represent the entire model since intra-model heterogeneity in early-passage patient-derived models is expected.

**2.2** Variants may or may not play a role in cancer predisposition. Reported variants include all non-synonymous variants present for the genes listed in Table 1, including hotspot regions for oncogenes and most of the protein coding regions for tumor suppressor genes.

**2.3** Common variants present in the population (e.g., germline) have not been removed from the reported non-synonymous variants. Population frequency databases (e.g., ExAC, 1000 genomes, NHLBI Exome Sequencing Project) should be used to remove variants present at high frequencies in the population.

## 3.0 DESCRIPTION OF NCI CANCER GENE PANEL

**3.1** The NCI Cancer Gene Panel data are generated using whole exome sequence (WES) \*.VCF files generated following the WES data analysis pipeline, version 2.0 (MCCRD\_SOP0011).

**3.2** The NCI Cancer Gene Panel is currently comprised of 62 genes implicated in cancer (gene list in [Table 1](#)). The regions interrogated are available as a BED file upon request.

**3.3** All non-synonymous variants present in the selected genes are reported, including hotspot regions for oncogenes and most of the protein coding regions of tumor suppressor genes.

## 4.0 PROTOCOL

**4.1** VCF files generated from the WES data analysis pipeline are intersected with the NCI Cancer Gene Panel BED file to extract the NCI cancer genes related variants.

`intersectBed -a Sample_WES.vcf -b NCI_cancer_gene_region.bed -u -header > Sample_NCI_cancer_gene.vcf`

**4.2** The extracted variants are annotated by a research-use copy of GeneMed [PMID: 25861217], an NCI informatics system that automatically performs annotation of each variant and makes functional predictions for each nonsynonymous mutation.

**4.3** GeneMed takes an input file in VCF format (version TVC 3.2) and annotates the variants using the SnpEff annotation tool (version 3.4).

Generating NCI Cancer Gene Panel Data Output from Whole Exome Sequence		
Laboratory:	Molecular Characterization Laboratory	
Effective Date:	1/23/2018	Page 2 of 4

- 4.4** SnpEff marks each variant as either synonymous or non-synonymous, and adds the annotation of the predicted amino acid change.
- 4.5** All non-synonymous variants are further annotated with the Sift prediction algorithm (version 5.0.2). Sift is incorporated into GeneMed to allow the system to perform functional predictions not only for known polymorphisms, but also for the de novo somatic mutations detected by the assay. Sift annotates the variants if they are present in dbSNP and includes the matching rs ID. Non-synonymous variants are also annotated if they are found in the BIOBASE human gene mutation database (HGMD).
- 4.6** Next, homopolymer regions within the amplicons are annotated and marked with the length of the homopolymer region if the reported position is within 1 bp of any homopolymer region.
- 4.7** The variants are finally compared to a table of NCI actionable mutations of interest (aMOIs) currently containing 380 unique mutations in 20 genes that have been curated for the NCI-MPACT trial (NCT01827384). This list will be expanded in future versions of the Gene Panel report.
- 4.7.1** The aMOIs, including SNVs and insertion/deletion variants, are known to be somatically altered in solid tumors based on recent scientific and clinical literature.
- 4.7.2** If an observed variant matches a variant in the aMOI list, then the exact COSMIC ID is annotated.
- 4.8** GeneMed generates an annotated output file (details in [Table 2](#)) for each gene variant identified in each model meeting the above criteria for upload to the NCI Patient-Derived Models database.

**TABLE 1: 62-GENE NCI CANCER GENE PANEL**

ABL1	FGFR3	NF1
AKT1	FLT3	NOTCH1
AKT2	GABRA6	NPM1
AKT3	GABRG2	NRAS
ALK	GNAQ	PARP1
APC	GNAS	PARP2
ARHGAP5	HNF1A	PDGFRA
ATM	HRAS	PIK3CA
ATR	IDH1	PTEN
BRAF	JAK2	PTPN11
CDH1	JAK3	RAD51
CDKN2A	KDR	RB1
CSF1R	KIT	RET
CTNNB1	KRAS	SMAD4
EGFR	MET	SMARCB1
ERBB2	MGMT	SMO
ERBB4	MLH1	SRC
ERCC1	MPL	STK11
FBXW7	MSH2	TP53
FGFR1	MTOR	VHL
FGFR2	NBN	

**TABLE 2: GENEMED OUTPUT FIELDS FOR PUBLIC DATABASE**

**1. Description of Data Fields**

Field	Description
Gene	Gene symbol for location of variant. Links to COSMIC website on that gene
Sift Prediction	Damaging/Tolerated: SIFT prediction for non-synonymous variants
Sift Score	The score is the normalized probability that the amino acid change is tolerated. The amino acid substitution is predicted damaging if the score is $\leq 0.05$ , and tolerated if the score is $> 0.05$ .
Impact	Functional impact of variant
dbSNP id	If available, dbSNP ID
COSMIC id	Is the variant in COSMIC, if so, link to the variant in COSMIC
Chromosome	Chromosome based on hg19
Position	Position of variant in chromosome based on hg19
REF (hg19)	Reference base (VCF format, from hg19)
ALT (tumor)	Variant base (VCF format)
Allele Freq	Variant allele frequency estimate based on present of reads supporting the variant
Read Depth	Total number of reads at the position
Codon Change	Codon change from reference
AA Change	Amino Acid change from reference

**2. Example data output for one gene variant in the PDMR database**

<u>Gene</u>	ATM	<u>Chromosome</u>	11
<u>REF (hg19)</u>	G	<u>Position</u>	108236150
<u>ALT (tumor)</u>	A	<u>Sift Prediction</u>	Damaging
<u>Allele Frequency</u>	0.64	<u>Sift Score</u>	.01
<u>Read Depth</u>	210	<u>dbSNP id</u>	novel
<u>Codon Change</u>	gGt/gAt	<u>Cosmic ID</u>	
<u>AA Change</u>	G3029D	<u>Assay Version</u>	2.0.0
<u>Impact</u>	nonsynonymous		