SOP Outlines **Version: 2.0**

1.0 PURPOSE/SCOPE

This Standing Operating Procedure (SOP) describes procedures for generating oncoKB Annotated Variants 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 REFERENCES

1 https://www.oncokb.org/

2 https://github.com/oncokb/oncokb-annotator

3 https://github.com/mskcc/vcf2maf

4 https://github.com/FNL-MoCha/nextgenseq_pipeline

3.0 CAVEATS

3.1 Reported oncoKB annotated variants 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.

3.2 The variants reported on a sample from this SOP are bound to the oncoKB database version and with newer version less or more variants will be reported based on new information available.

3.3 Although oncoKB annotates variants with the level of evidence, that information is not exposed in the output, as it may become outdated with time.

4.0 DESCRIPTION OF ONCOKB ANNOTATED VARIANTS

4.1 The variants are generated using whole exome sequence (WES) *.VCF files generated following the WES data analysis pipeline, version 2.0 (MCCRD_SOP0011) and annotated using oncoKB pipeline.

4.2 The oncoKB annotated variants are currently reported based on oncoKB annotation pipeline version 1.1.0

5.0 PROCEDURE

5.1 VCF files generated from the WES data analysis pipeline are converted to MAF format using vcf2maf version 1.6.16, using VEP version 92.

- vcf2maf.pl --input-vcf Sample.merged.vcf --output-maf Sample.merged.maf --tumor-id Sample --ref-fasta <input.ref> --filter-vcf ExAC.r0.3.1.sites.vep.vcf.gz --vep-path $VEP_HOME/ --vep-data $VEP_CACHEDIR --custom-enst <ISOFORM Mapping file>

5.2 Non-Synonymous variants which are either novel or present in gnomAD at Allele frequency <=0.01 are filtered in using a custom perl script

- filterMAF.pl Sample.merged.maf > Sample.clean.merged.maf

5.3 The Variants from step 4.2 are annotated using oncokb-annotator-1.1.0 using a locally hosted URL for API calling to keep the results consistent over time (database version June 21, 2019).

- python /data/MoCha/patidarr/oncokb-annotator-1.1.0/MafAnnotator.py -I Sample.clean.merged.maf -o Sample.clean.merged.oncoKB.maf -t <Diagnosis> -u <local oncoKB URL>

6.0 REVISION HISTORY:

Document No.	Version	Description of Revision	Effective Date
MCCRD-SOP0053	1.0	Original Release	1/13/2020

TABLE 1: ONCOKB OUTPUT FIELDS FOR PUBLIC DATABASE

Field	Description
Hugo Gene Symbol	Hugo Gene symbol harboring the variant
Chr	Chromosome based on hg19
Chr Start	Position on chromosome where variant allele starts
Chr End	Position on chromosome where variant allele ends
Ref Allele	Reference Allele
Alt Allele	Alternate based observed in the patients WES sequencing data
HGVS cDNA Change	cDNA Change in HGVS format
HGVS Protein Change	Protein Change notation in HGVS format
Existing Variant	dbSNP or COSMIC Id if this variant is present in corresponding database
Variant Class	Variant Class: • Missense_Mutation • Nonsense_Mutation • Nonstop_Mutation • Splice_Site • Frame_Shift_Del • Frame_Shift_Ins • In_Frame_Del • In_Frame_Ins
Total Reads	Total Reads sequenced at the position
Variant Allele Frequency	Variant Allele Frequency for the Alt Allele
SIFT	SIFT Prediction and Score
PolyPhen	PolyPhen Prediction and Score
Oncogenicity	Oncogenicity as annotated by oncoKB
Predicted Functional Effect	Functional Effect as annotated by oncoKB

EXAMPLE DATA OUTPUT FOR ONE GENE VARIANT IN THE PDMR DATABASE

Hugo Gene Symbol	BRAF
Chr	chr7
Chr Start	140453136
Chr End	140453137
Ref Allele	AC
Alt Allele	TT
HGVS cDNA Change	c.1798_1799delinsAA
HGVS Protein Change	p.V600K
Existing Variant	rs121913227,COSM1583011,COSM249889,COSM26504,COSM4166148,COSM473,COSM474,COSM5985086,COSM6005496
Variant Class	Missense_Mutation
Total Reads	84
Variant Allele Frequency	0.7024
SIFT	deleterious(0)
PolyPhen	probably_damaging(0.995)
Oncogenicity	Oncogenic
Predicted Functional Effect	Gain-of-function

APPENDIX: LIST OF HUGO GENE SYMBOLS ASSOCIATED WITH VERSION 2.0

ABL1	AXIN2	CD79B	CYSLTR2	ERF
ACTG1	AXL	CDC73	DAXX	ERRFI1
ACVR1	B2M	CDH1	DDR2	ESCO2
AKT1	BACH2	CDK12	DDX3X	ESR1
AKT2	BAP1	CDK4	DICER1	ETAA1
AKT3	BARD1	CDK6	DIS3	ETNK1
ALK	BBC3	CDKN1A	DNMT3A	ETV6
AMER1	BCL10	CDKN1B	DNMT3B	EZH1
ANKRD11	BCL11B	CDKN2A	DTX1	EZH2
APC	BCL2	CDKN2B	DUSP22	FAM175A
AR	BCL2L11	CDKN2C	DUSP4	FAM58A
ARAF	BCOR	CEBPA	ECT2L	FANCA
ARID1A	BCORL1	CHEK1	EED	FANCC
ARID1B	BIRC3	CHEK2	EGFR	FANCD2
ARID2	BLM	CIC	EGR1	FAS
ARID3A	BMPR1A	CIITA	EIF1AX	FAT1
ARID4A	BRAF	CMTR2	ELF3	FBXO11
ARID4B	BRCA1	CRBN	EP300	FBXW7
ARID5B	BRCA2	CREBBP	EP400	FGFR1
ASXL1	BRIP1	CRLF2	EPAS1	FGFR2
ASXL2	BTG1	CSF1R	EPCAM	FGFR3
ATM	BTK	CSF3R	EPHA3	FGFR4
ATP6AP1	CARD11	CTCF	EPHA7	FH
ATP6V1B2	CASP8	CTLA4	ERBB2	FLCN
ATR	CBFB	CTNNB1	ERBB3	FLT3
ATRX	CBL	CTR9	ERBB4	FOXA1
ATXN2	CCND1	CUX1	ERCC2	FOXL2
AURKA	CCND3	CXCR4	ERCC3	FOXO1
AXIN1	CD58	CYLD	ERCC4	FOXP1

FUBP1	INPP4B	MAP2K4	NKX2-1	PIK3R2
GATA2	INPPL1	MAP3K1	NKX3-1	PIK3R3
GATA3	IRF1	MAPK1	NOTCH1	PIM1
GLI1	IRF8	MAX	NOTCH2	PLCG1
GNA11	JAK1	MDM4	NOTCH3	PLCG2
GNA13	JAK2	MED12	NOTCH4	PMAIP1
GNAQ	JAK3	MEF2B	NPM1	PMS1
GNAS	JARID2	MEN1	NRAS	PMS2
GNB1	KAT6A	MET	NRG1	POLD1
GPS2	KDM5C	MGA	NSD1	POLE
GRIN2A	KDM6A	MITF	NT5C2	POT1
GTF2I	KDR	MLH1	NTHL1	PPP2R1A
H3F3A	KEAP1	MOB3B	NTRK1	PPP6C
HDAC1	KIT	MPL	NTRK3	PRDM1
HDAC4	KLF2	MRE11A	P2RY8	PTCH1
HIST1H1B	KLF4	MSH2	PAK5	PTEN
HIST1H1C	KLF5	MSH3	PALB2	PTPN1
HIST1H1D	KMT2A	MSH6	PARK2	PTPN11
HIST1H3B	KMT2B	MST1	PARP1	PTPN2
HLA-A	KMT2C	MTOR	PAX5	PTPRD
HLA-B	KMT2D	MUTYH	PBRM1	PTPRS
HNF1A	KNSTRN	MYC	PDGFRA	PTPRT
HOXB13	KRAS	MYCN	PDGFRB	RAB35
HRAS	LATS1	MYD88	PDS5B	RAC1
ID3	LATS2	MYOD1	PHF6	RAD21
IDH1	LTB	NBN	PHOX2B	RAD50
IDH2	LZTR1	NCOR1	PIGA	RAD51
IGF1R	MAF	NF1	PIK3CA	RAD51B
IKZF3	MAGOH	NF2	PIK3CB	RAD51C
IL7R	MAP2K1	NFE2L2	PIK3CD	RAD51D
INHA	MAP2K2	NFKBIA	PIK3R1	RAD54L

RAF1	SDHAF2	SMAD4	STAT5B	TP63
RASA1	SDHB	SMARCA4	STK11	TRAF3
RB1	SDHC	SMARCB1	SUFU	TRAF5
RBM10	SDHD	SMC1A	SUZ12	TRIP13
RBM15	SESN1	SMC3	TBL1XR1	TSC1
RECQL	SESN2	SMG1	TBX3	TSC2
RECQL4	SESN3	SMO	TCF3	TYK2
RELN	SETBP1	SOCS1	TCF7L2	U2AF1
RET	SETD2	SOS1	TERT	USP8
RHEB	SETD3	SOX17	TET1	VAV1
RHOA	SETDB1	SOX9	TET2	VHL
RIT1	SETDB2	SP140	TET3	WHSC1
RNF43	SF3B1	SPEN	TFE3	WT1
ROBO1	SH2B3	SPOP	TGFBR1	XPO1
ROS1	SH2D1A	SPRED1	TGFBR2	XRCC2
RRAS2	SHOC2	SPRTN	TMEM127	YAP1
RTEL1	SHQ1	SRC	TNFAIP3	ZFHX3
RUNX1	SLFN11	SRSF2	TNFRSF14	ZNRF3
RYBP	SLX4	STAG1	TOP1	
SAMHD1	SMAD2	STAG2	TP53	
SDHA	SMAD3	STAT3	TP53BP1	