MCCRD-SOP0053: Annotation of Variants from Whole Exome Assay with OncoKB

Effective Date: 1/13/2020

Please check for revision status of the SOP at

https://pdmr.cancer.gov/sops/

PDMR NCI Patient-Derived Models Repository
An NCI Precision Oncology Initiative℠ Resource

TABLE OF CONTENTS

VERSION INFORMATION

1. Change History

Revision	Description
	Internal SOP used by MOCHA Laboratory
1/30/2020	Standardize SOP for posting to PDMR internal site for use by designated NCI intramural laboratories

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

2.1 https://www.oncokb.org/

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

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

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oncoKB annotates variants with the level of evidence, that information is not exposed in the output, as it may become outdated with time.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

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

FGFR2	HLA-A	KMT2A	MSH2	NTRK3
FGFR3	HLA-B	KMT2B	MSH3	P2RY8
FGFR4	HNF1A	KMT2C	MSH6	PAK5
FH	HOXB13	KMT2D	MST1	PALB2
FLCN	HRAS	KNSTRN	MTOR	PARK2
FLT3	ID3	KRAS	MUTYH	PARP1
FOXA1	IDH1	LATS1	MYC	PAX5
FOXL2	IDH2	LATS2	MYCN	PBRM1
FOXO1	IGF1R	LTB	MYD88	PDGFRA
FOXP1	IKZF3	LZTR1	MYOD1	PDGFRB
FUBP1	IL7R	MAF	NBN	PDS5B
GATA2	INHA	MAGOH	NCOR1	PHF6
GATA3	INPP4B	MAP2K1	NF1	PHOX2B
GLI1	INPPL1	MAP2K2	NF2	PIGA
GNA11	IRF1	MAP2K4	NFE2L2	PIK3CA
GNA13	IRF8	MAP3K1	NFKBIA	PIK3CB
GNAQ	JAK1	MAPK1	NKX2-1	PIK3CD
GNAS	JAK2	MAX	NKX3-1	PIK3R1
GNB1	JAK3	MDM4	NOTCH1	PIK3R2
GPS2	JARID2	MED12	NOTCH2	PIK3R3
GRIN2A	KAT6A	MEF2B	NOTCH3	PIM1
GTF2I	KDM5C	MEN1	NOTCH4	PLCG1
H3F3A	KDM6A	MET	NPM1	PLCG2
HDAC1	KDR	MGA	NRAS	PMAIP1
HDAC4	KEAP1	MITF	NRG1	PMS1
HIST1H1B	KIT	MLH1	NSD1	PMS2
HIST1H1C	KLF2	MOB3B	NT5C2	POLD1
HIST1H1D	KLF4	MPL	NTHL1	POLE
HIST1H3B	KLF5	MRE11A	NTRK1	POT1

PPP2R1A	RECQL	SETD3	SPOP	TNFRSF14
PPP6C	RECQL4	SETDB1	SPRED1	TOP1
PRDM1	RELN	SETDB2	SPRTN	TP53
PTCH1	RET	SF3B1	SRC	TP53BP1
PTEN	RHEB	SH2B3	SRSF2	TP63
PTPN1	RHOA	SH2D1A	STAG1	TRAF3
PTPN11	RIT1	SHOC2	STAG2	TRAF5
PTPN2	RNF43	SHQ1	STAT3	TRIP13
PTPRD	ROBO1	SLFN11	STAT5B	TSC1
PTPRS	ROS1	SLX4	STK11	TSC2
PTPRT	RRAS2	SMAD2	SUFU	TYK2
RAB35	RTEL1	SMAD3	SUZ12	U2AF1
RAC1	RUNX1	SMAD4	TBL1XR1	USP8
RAD21	RYBP	SMARCA4	TBX3	VAV1
RAD50	SAMHD1	SMARCB1	TCF3	VHL
RAD51	SDHA	SMC1A	TCF7L2	WHSC1
RAD51B	SDHAF2	SMC3	TERT	WT1
RAD51C	SDHB	SMG1	TET1	XPO1
RAD51D	SDHC	SMO	TET2	XRCC2
RAD54L	SDHD	SOCS1	TET3	YAP1
RAF1	SESN1	SOS1	TFE3	ZFHX3
RASA1	SESN2	SOX17	TGFBR1	ZNRF3
RB1	SESN3	SOX9	TGFBR2	
RBM10	SETBP1	SP140	TMEM127	
RBM15	SETD2	SPEN	TNFAIP3	