Supplementary Note

TITLE: Integrated molecular drivers coordinate biological and clinical states in melanoma

AUTHORS

Jake R. Conway[1,2,3], Felix Dietlein[2,3], Amaro Taylor-Weiner[1,2,3], Saud AlDubayan[2,3,4,5], Natalie Vokes[2,3], Tanya Keenan[2,3], Brendan Reardon[2,3], Meng Xiao He[2,3,6], Claire A. Margolis[2,3], Jason Weirather[7], Rizwan Haq[7], Bastian Schilling[8, 9], F. Stephen Hodi[7], Dirk Schadendorf[9], David Liu[2,3], Eliezer M. Van Allen[2,3,7]

AFFILIATIONS

[1]Division of Medical Sciences, Harvard University, Boston, Massachusetts 02115, USA
[2]Cancer Program, Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA
[3]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts 02215, USA
[4]Division of Genetics, Brigham and Women's Hospital, Boston, MA 02115, USA
[5]Department of Medicine, King Saud bin Abdul-Aziz University for Health Sciences, Riyadh, Saudi Arabia
[6]Harvard Graduate Program in Biophysics, Boston, MA 02115
[7]Center for Cancer Precision Medicine, Dana-Farber Cancer Institute, Boston, MA 02215, USA
[8]Department of Dermatology, University Hospital Würzburg, Würzburg, Germany
[9]Department of Dermatology, University Hospital, Essen, Germany and German Cancer Consortium of Translational Cancer Research (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany

MuTect1.1.6

Parameters:
fractionContamination: fraction of cross-sample contamination output from ContEst
downsampleToCoverage: 99999
dbSNPVCF: can be downloaded from dbSNP; used to exclude regions around known polymorphisms
cosmicVCF: can be downloaded from COSMIC; catalogue of somatic mutations
readgroupBlacklist: list of read groups to exclude

--out, --coverage_file, and --power_file are all outputs

Command line code

```
java -jar -Xmx4g /muTect-1.1.6.jar --analysis_type MuTect -L ${targetsIntervalList} \
--normal_sample_name ${normalSampleID} -I:normal ${normalBam} \
--tumor_sample_name ${tumorSampleID} -I:tumor ${tumorBam} \
--reference_sequence ${refFasta} --normal_panel ${normalPanel} \
--fraction_contamination ${fractionContamination} \
--downsample_to_coverage ${downsampleToCoverage} \
--dbsnp ${dbSNPVCF} --cosmic ${cosmicVCF} \
--read_group_black_list ${readgroupBlacklist} \
--enable_extended_output \
--out Mutect1.call_stats.txt \
--coverage_file Mutect1.coverage.wig.txt \
--power_file Mutect1.power.wig.txt
```

MutSig2CV

The code for MutSig2CV is available at https://github.com/getzlab/MutSig2CV.

Parameters (not explicitly included in GitHub repository):
mergedMaf: Somatic mutation file containing somatic mutations for all samples in the cohort
mutationBlacklist: list of loci in genome to blacklist from analysis (columns include: gene, chr, start, end, type, classification, ref_allele, newbase, Protein_Change)
genomeBuild: hg19
paramsFile: additional parameters for running MutSig2CV in tab separated text file (impute_full_cov_when_promotes_significance: false, max_neighbors: 1000, num_neighbor_patients: 1, qual_min: 0.1). Documentation on how to generate this file and other available parameter options are included in the GitHub repository.

Parameters (included as a download link in GitHub repository):

The other parameters files can be downloaded from the link listed in the “Installing” section of the GitHub repository (http://software.broadinstitute.org/cancer/cga/sites/default/files/data/tools/mutsig/MutSig2CV.tar.gz).
mutationTypeDictionaryFile: mutation_type_dictionary.v4.txt
coverageModelsFile: coverage_models.v5a.mat
basewiseCoverageFile: coverage_basewise.fwb
targetListFile: target_list.hg19.v1a.txt
contextAndEffectFwbFile: context_and_effect.c65e29b.fwb
contextAndEffectCatsegsFile: context_and_effect.c65e29b.txt
covariatesFile: covariates_transformed.v5a.txt
conservationFwbFile: conservation46.fwb
fixedWidthBinaryJarFile: FixedWidthBinary.jar

Command line code

```
/bin/MutSig2CV \
${mergedMaf} . \
${mutationBlacklist} ${mutationTypeDictionaryFile} \
${coverageModelsFile} ${basewiseCoverageFile} \
${targetListFile} ${contextAndEffectFwbFile} \
${contextAndEffectCatsegsFile} ${covariatesFile} \
${conservationFwbFile} ${fixedWidthBinaryJarFile} \
${genomeBuild} ${paramsFile}
```

GISTIC2.0

Parameters:

Both

arm_peel: 1
broad_length_cutoff: 0.7
cap: 1.5
conf_level: 0.99
do_gene_gistic: 0
gene_collapse_method: mean
join_segment_size: 10
markers_file: NA (permuted markers)
max_sample_segs: 2000
max_window_var: 10000
refgene_file:
remove_X: 0

refgene_file: hg19_GENCODE_v18_20140127.mat

Tumor
amp_thresh: 0.1
del_thresh: 0.1
qv_thresh: 0.25
cnv_files: no regions

Normal
amp_thresh: 0.3
del_thresh: 0.3
qv_thresh: 0.1
cnv_files: all peaks from normal with q < 0.25

Sample Realignment

Realignment to the hg19 genome (ftp://ftp.broadinstitute.org/pub/seq/references/Homo_sapiens_assembly19.fasta) for all samples originally aligned to GRCh37 was performed using the best practices from the GATK and Picard realignment pipeline (https://software.broadinstitute.org/gatk/best-practices/, https://broadinstitute.github.io/picard/) on FireCloud. The syntax for FireCloud variables are ${variable_name}.

The following tasks were run for each sample that underwent realignment: Revert to SAM, Collect Quality Yield Metrics, SAM To FASTQ, BWA Alignment, Merge BAM Alignment, Mark Duplicates, Collect Multiple Metrics, Validate SAM File, Collect HS Metrics, Realigner Target Creator, Indel Realigner, GATK Base Recalibrator, GATK Print Reads.

Revert To SAM
(https://broadinstitute.github.io/picard/command-line-overview.html#RevertSam)

```
java -Dsamjdk.buffer_size=131072 -Dsamjdk.compression_level=1 \
-XX:GCTimeLimit=50 \
-XX:GCHeapFreeLimit=10 \
-Xmx4000m -jar /usr/gitc/picard.jar RevertSam \
TMP_DIR=. \
VALIDATION_STRINGENCY=SILENT \
OUTPUT=${sampleName}.reverted.bam \
INPUT=${inputBam} \
SORT_ORDER=queryname \
```

```
RESTORE_ORIGINAL_QUALITIES=true \
REMOVE_DUPLICATE_INFORMATION=true \
REMOVE_ALIGNMENT_INFORMATION=true \
SANITIZE=true \
MAX_DISCARD_FRACTION=0.01 \
ATTRIBUTE_TO_CLEAR=X0 \
ATTRIBUTE_TO_CLEAR=X1 \
ATTRIBUTE_TO_CLEAR=XA \
ATTRIBUTE_TO_CLEAR=XC \
ATTRIBUTE_TO_CLEAR=XG \
ATTRIBUTE_TO_CLEAR=XM \
ATTRIBUTE_TO_CLEAR=XN \
ATTRIBUTE_TO_CLEAR=XO \
ATTRIBUTE_TO_CLEAR=XT \
ATTRIBUTE_TO_CLEAR=AM \
ATTRIBUTE_TO_CLEAR=AS \
ATTRIBUTE_TO_CLEAR=BQ \
ATTRIBUTE_TO_CLEAR=CC \
ATTRIBUTE_TO_CLEAR=CP \
ATTRIBUTE_TO_CLEAR=E2 \
ATTRIBUTE_TO_CLEAR=H0 \
ATTRIBUTE_TO_CLEAR=H1 \
ATTRIBUTE_TO_CLEAR=H2 \
ATTRIBUTE_TO_CLEAR=HI \
ATTRIBUTE_TO_CLEAR=IH \
ATTRIBUTE_TO_CLEAR=MF \
ATTRIBUTE_TO_CLEAR=NH \
ATTRIBUTE_TO_CLEAR=OC \
ATTRIBUTE_TO_CLEAR=OP \
ATTRIBUTE_TO_CLEAR=PQ \
ATTRIBUTE_TO_CLEAR=R2 \
ATTRIBUTE_TO_CLEAR=S2 \
ATTRIBUTE_TO_CLEAR=SM \
ATTRIBUTE_TO_CLEAR=SQ \
ATTRIBUTE_TO_CLEAR=U2 \
ATTRIBUTE_TO_CLEAR=XQ
```

Collect Quality Yield Metrics

(https://broadinstitute.github.io/picard/command-line-overview.html#CollectQualityYieldMetrics)

```
java -Dsamjdk.buffer_size=131072 -XX:GCTimeLimit=50 -XX:GCHeapFreeLimit=10 -Xmx128m -jar /usr/gitc/picard.jar CollectQualityYieldMetrics \
TMP_DIR=. \
I=${inputBam} \
O=${sampleName}_quality_yield_metrics.txt
```

SAM To FASTQ
(https://broadinstitute.github.io/picard/command-line-overview.html#SamToFastq)

```
java -Dsamjdk.buffer_size=131072 -Dsamjdk.use_async_io=true \
-Dsamjdk.compression_level=1 -XX:GCTimeLimit=50 -XX:GCHeapFreeLimit=10 -Xmx256m \
-jar /usr/gitc/picard.jar SamToFastq \
TMP_DIR=. \
INPUT=${inputBam} \
FASTQ=${sampleName}.1.fastq.gz \
INTERLEAVE=false \
SECOND_END_FASTQ=${sampleName}.2.fastq.gz \
INCLUDE_NON_PF_READS=true \
CLIPPING_ATTRIBUTE=XT \
CLIPPING_ACTION=2 \
UNPAIRED_FASTQ=${sampleName}.unpaired.fastq.gz
```

BWA Alignment
0.5.9-r16 (patched by AW, and multithread samse/sampe patch applied - https://sourceforge.net/p/bio-bwa/mailman/bio-bwa-help/thread/50B8FD2D.7030608@mail.mcgill.ca/)

```
/usr/gitc/bwa-0.5.9/bwa index -a bwtsw ${refFasta}

/usr/gitc/bwa-0.5.9/bwa aln ${refFasta} -q 5 -l 32 -k 2 -o 1 -t ${cpu} -f ${sampleName}.Homo_sapiens_assembly19.1.sai ${firstEndFastq}

/usr/gitc/bwa-0.5.9/bwa aln ${refFasta} -q 5 -l 32 -k 2 -o 1 -t ${cpu} -f ${sampleName}.Homo_sapiens_assembly19.2.sai ${secondEndFastq}
```

/usr/gitc/bwa-0.5.9/bwa aln ${refFasta} -q 5 -l 32 -k 2 -o 1 -t ${cpu} -f
${sampleName}.Homo_sapiens_assembly19.unpaired.sai ${unpairedFastq}

/usr/gitc/bwa-0.5.9/bwa sampe -P -f
${sampleName}.Homo_sapiens_assembly19.aligned_bwa.sam ${refFasta}
${sampleName}.Homo_sapiens_assembly19.1.sai
${sampleName}.Homo_sapiens_assembly19.2.sai ${firstEndFastq} ${secondEndFastq}

/usr/gitc/bwa-0.5.9/bwa samse -f
${sampleName}.Homo_sapiens_assembly19.unpaired.aligned_bwa.sam ${refFasta}
${sampleName}.Homo_sapiens_assembly19.unpaired.sai ${unpairedFastq}

Merge BAM Alignment
(https://broadinstitute.github.io/picard/command-line-overview.html#MergeBamAlignment)

java -Dsamjdk.buffer_size=131072 -Dsamjdk.compression_level=1 -XX:+UseStringCache -
XX:GCTimeLimit=50 -XX:GCHeapFreeLimit=10 -Xmx5000m -jar /usr/gitc/picard.jar
MergeBamAlignment \
TMP_DIR=. \
VALIDATION_STRINGENCY=SILENT \
CREATE_INDEX=true \
ALIGNED_BAM=${alignedBam} \
ALIGNED_BAM=${unpairedAlignedBam} \
EXPECTED_ORIENTATIONS=FR \
ATTRIBUTES_TO_RETAIN=X0 \
UNMAPPED_BAM=${unmappedBam} \
OUTPUT=${sampleName}.merged.aligned.bam \
REFERENCE_SEQUENCE=${refFasta} \
PAIRED_RUN=true \
IS_BISULFITE_SEQUENCE=false \
ALIGNED_READS_ONLY=false \
CLIP_ADAPTERS=false \
MAX_RECORDS_IN_RAM=2000000 \
ADD_MATE_CIGAR=true

Mark Duplicates
(https://broadinstitute.github.io/picard/command-line-overview.html#MarkDuplicates)

```
java -Dsamjdk.buffer_size=131072 -XX:GCTimeLimit=50 -XX:GCHeapFreeLimit=10 -
Xmx4000m -jar /usr/gitc/picard.jar MarkDuplicates \
TMP_DIR=. \
CREATE_INDEX=true \
CREATE_MD5_FILE=true \
INPUT=${inputBam} \
OUTPUT=${sampleName}.aligned.duplicates_marked.bam \
METRICS_FILE=${sampleName}.duplicate_metrics \
VALIDATION_STRINGENCY=SILENT \
OPTICAL_DUPLICATE_PIXEL_DISTANCE=2500 \
READ_ONE_BARCODE_TAG=RX
```

Collect Multiple Metrics
(https://broadinstitute.github.io/picard/command-line-overview.html#CollectMultipleMetrics)

```
java -Dsamjdk.buffer_size=131072 -XX:GCTimeLimit=50 -XX:GCHeapFreeLimit=10 \
-Xmx4000m -jar /usr/gitc/picard.jar CollectMultipleMetrics \
TMP_DIR=. \
METRIC_ACCUMULATION_LEVEL=null \
METRIC_ACCUMULATION_LEVEL=ALL_READS \
PROGRAM=null \
PROGRAM=MeanQualityByCycle \
PROGRAM=QualityScoreDistribution \
PROGRAM=CollectInsertSizeMetrics \
PROGRAM=CollectAlignmentSummaryMetrics \
INPUT=${inputBam} \
REFERENCE_SEQUENCE=${refFasta} \
ASSUME_SORTED=true \
OUTPUT=${sampleName}.multiple_metrics.bam
```

Validate SAM File
(https://broadinstitute.github.io/picard/command-line-overview.html#ValidateSamFile)

```
java -Dsamjdk.buffer_size=131072 -XX:GCTimeLimit=50 -XX:GCHeapFreeLimit=10 -
Xmx4000m -jar /usr/gitc/picard.jar ValidateSamFile \
TMP_DIR=. \
CREATE_MD5_FILE=false \
I=${inputBam} \
```

```
O=${sampleName}.validation_metrics \
REFERENCE_SEQUENCE=${refFasta} \
MODE=SUMMARY \
IS_BISULFITE_SEQUENCED=false \
INDEX_VALIDATION_STRINGENCY=LESS_EXHAUSTIVE
```

Collect HS Metrics
(https://broadinstitute.github.io/picard/command-line-overview.html#CollectHsMetrics)

```
java -Dsamjdk.buffer_size=131072 -XX:GCTimeLimit=50 -XX:GCHeapFreeLimit=10 \
-Xmx1500m -jar /usr/gitc/picard.jar CollectHsMetrics \
TMP_DIR=. \
INPUT=${inputBam} \
OUTPUT=${sampleName}.hybrid_selection_metrics \
REFERENCE_SEQUENCE=${refFasta} \
TARGET_INTERVALS=${targetIntervalList} \
BAIT_INTERVALS=${baitIntervalList}
```

Realigner Target Creator
(https://software.broadinstitute.org/gatk/documentation/tooldocs/3.8-0/org_broadinstitute_gatk_tools_walkers_indels_RealignerTargetCreator.php)

```
java -Djava.io.tmpdir=/directory -XX:GCTimeLimit=50 -XX:GCHeapFreeLimit=10 -
Xmx4000m -jar /usr/GenomeAnalysisTK.jar -T RealignerTargetCreator \
--disable_auto_index_creation_and_locking_when_reading_rods \
-R ${refFasta} \
-o ${sampleName}.indelrealigner.target.intervals \
-I ${nameBase}.bam -dcov 250 -nt 1 \
--known ${dbSnpVcf} \
--known ${knownIndelVcf} \
--known ${goldStandardVcf}
```

Indel Realigner
(https://software.broadinstitute.org/gatk/documentation/tooldocs/3.8-0/org_broadinstitute_gatk_tools_walkers_indels_IndelRealigner.php)

```
java -jar /usr/GenomeAnalysisTK.jar -T IndelRealigner \
-R ${refFasta} \
-I ${nameBase}.bam \
```

-known ${dbSnpVcf} \
-known ${knownIndelVcf} \
-known ${goldStandardVcf} \
-targetIntervals ${indelRealignmentTargets} \
-o ${sampleName}.indel_realigned.bam \
--consensusDeterminationModel KNOWNS_ONLY \
-LOD 0.4

GATK Base Recalibrator
(https://software.broadinstitute.org/gatk/documentation/tooldocs/3.8-0/org_broadinstitute_gatk_tools_walkers_bqsr_BaseRecalibrator.php)

java -Djava.io.tmpdir=/directory -XX:GCTimeLimit=50 -XX:GCHeapFreeLimit=10 -Xmx4000m -jar /usr/GenomeAnalysisTK.jar -T BaseRecalibrator \
--disable_auto_index_creation_and_locking_when_reading_rods -U \
-R ${refFasta} \
-I ${nameBase}.bam \
--useOriginalQualities \
-o ${sampleName}.recal_data.csv \
-knownSites ${dbSnpVcf} \
-knownSites ${knownIndelVcf} \
-knownSites ${goldStandardVcf}

GATK Print Reads
(https://software.broadinstitute.org/gatk/documentation/tooldocs/3.8-0/org_broadinstitute_gatk_tools_walkers_readutils_PrintReads.php)

java -Djava.io.tmpdir=/directory -Dsamjdk.use_async_io=true -XX:GCTimeLimit=50 -XX:GCHeapFreeLimit=10 -Xmx8000m -jar /usr/GenomeAnalysisTK.jar -T PrintReads \
--disable_auto_index_creation_and_locking_when_reading_rods \
--generate_md5 -U \
-R ${refFasta} \
-I ${nameBase}.bam \
--useOriginalQualities \
-o ${sampleName}.realigned.hg19.bam \
--disable_indel_quals \
-BQSR ${recalData} \
--emit_original_quals