Methods

The MBCproject

The MBCproject is a patient-driven genomics initiative that enables metastatic breast cancer patients to enroll through an online portal to provide consent for acquisition of a saliva sample, a blood sample, medical records, and archived FFPE tissue samples. These data are used to create a clinically-annotated genomic dataset that can be shared across multiple platforms. Refer to mbcproject.org/data-release for links to data on public platforms. The following describes the methods used to generate the current dataset (157 samples, 129 patients).

Methods

Design

The Metastatic Breast Cancer Project website (MBCproject.org) enables patients to enroll in the study remotely. The website and all associated messaging were developed through iterative feedback from metastatic breast cancer patients. Patients registered with first name, last name, email address, and confirmation of a metastatic breast cancer diagnosis. Registrants were asked to complete a sixteen question survey, with all questions optional, about their experiences with metastatic breast cancer (see link on mbcproject.org/data-release). Registrants acknowledged that responses would be stored in a secure database. Included in this acknowledgement is an understanding that patients may be re-contacted and that they can withdraw their information as indicated in the pre-submission statement:

"I understand that the information I entered here will be stored in a secure database and may be used to match me to one or more research studies conducted by the Metastatic Breast Cancer Project. If the information that I entered matches a study being conducted by the Metastatic Breast Cancer Project, either now or in the future, I agree to be contacted about possibly participating. I understand that if I would like my information deleted from our database, now or in the future, I can email info@mbcproject.org and my information will be removed from the database."

All patients who submitted a written request to withdraw from the study were immediately exited by study staff.

A link to an electronic informed consent document for formal enrollment in the study (see link on mbcproject.org/data-release) was sent to registrants who completed the survey. Registrants could direct any questions to study staff throughout the enrollment process. Email reminders were sent weekly for three weeks, and again at six weeks, to registrants who had not completed the consent process. Registrants who provided informed consent were then asked to complete a medical release form in order to provide their contact information, as well as a list of physicians and hospitals that provided their care for metastatic breast cancer. Email reminders were sent weekly for three weeks, and again at six weeks, to registrants who had not completed the medical release form. Upon completion, signed copies of consent and medical release forms were sent electronically to the registrants. All data collected online were stored in a secure database hosted by Heroku. Data were transferred from the online database to a secure DatStat database for analysis.

Saliva kits were mailed to registrants who provided informed consent who lived in The United States or Canada. Saliva kits were labeled with a unique two-dimensional barcode and a pre-paid Business Reply Label addressed to the Broad Institute Genomics Platform prior to shipment. Each barcode identifier was assigned to a participant prior to shipment. Participants provided a saliva sample by following the included instructions (see link on mcbproject.org/data-release) and returned the kits free of charge. Saliva kits received at the Broad Institute were logged by their unique barcodes, and stored at room temperature until they were advanced to Whole Exome Sequencing (WES).

Upon receiving saliva kits, study staff called the hospitals and physicians' offices listed in each participant's medical release form to confirm the fax number for the medical records department. A detailed request for medical records including clinic notes from treating providers, breast cancer treatment data (including radiation, chemotherapy and hormonal therapy), pathology reports, operative reports, referrals, MD to MD exchange, and genetic testing reports from the date of diagnosis through the date of the request was faxed to each facility. Medical records were received by fax, mail, or secure electronic message. All medical records were separated into categories (clinic notes, pathology reports, imaging, genetic testing reports, other), and were scanned and uploaded to a secure drive to facilitate abstraction. The date of each procedure (e.g. biopsy, resection, mastectomy, etc.), type of procedure, histology, and facility which performed the procedure were abstracted from all pathology reports in order to prioritize tissue samples for request. Breast cancer samples were flagged for request. Study staff were informed not to request the most recent biopsy in order to avoid exhausting samples necessary for clinical care.

Study staff called the pathology departments associated with each tissue sample to confirm the fax number for tissue requests. A form requesting a minimum of 5 and maximum of 20 5-micron unstained slides, or one tissue block, and one Hematoxylin and Eosin stain (H&E) slide was faxed to each pathology department. Requests explicitly stated that no sample should be exhausted in order to fulfill the request. Tissue samples were received at the Broad Institute by mail. Tissue samples received as blocks were labeled with unique numerical identifiers and sent to the DF/HCC Specialized Histopathology Services (SHS) Core to be cut into three 30-micron scrolls per block. These scrolls were then labeled with unique barcode identifiers and submitted to the Broad Institute Genomics Platform for WES and Transcriptome Capture (RNA Seq). Tissue samples received as unstained slides were logged, labeled with unique barcode identifiers, and submitted to the Broad Institute Genomics Platform for WES and RNA Seq.

Medical Record Abstraction

A data dictionary (see Appendix 1, pg. 7) comprising 22 clinical fields from pathology reports and 67 fields from medical records was developed. A clinical data abstraction specialist reviewed all available records for each patient and abstracted the 89 fields into a table. The date of primary diagnosis with breast cancer was defined as the date of first confirmation either by imaging or biopsy, whichever came first. Dates were abstracted to the greatest level of detail available in the record. Dates reported in the medical record only as a month and year were abstracted as the first of the month while dates reported only as a year were abstracted as the first of January of that year. For all other fields, only data explicitly reported in the medical record were abstracted; no data were inferred. In order to protect patient confidentiality, all dates were reported in shared data sets as elapsed time relative to the date of primary diagnosis and ages were grouped into 5-year increments.

Patient-Reported Data Cleaning

Patient-reported data from two survey questions were cleaned to standardize format as well as protect confidentiality. Registrants who completed the survey had the option to list therapies that they received for two years or longer. Each question was answerable using a free-text field. Responses for therapies received for greater than two years were standardized to reflect only generic drug names, correct misspellings, and to exclude drugs listed with supporting information that contradicted the intent of the question (e.g. patients reported being on a therapy for more than 2 years, but that listed in their free text that they were on that drug for only 6 months. Responses were also classified into the following categories: aromatse inibitors, platinums, tamoxifen, fulvestrant, trastuzumab, pertuzumab, capecitabine, and PARP inhibitors. In order to protect confidentiality, race categories that were reported fewer than five times in the data set were reclassified as “Other.”

Biological Sample Processing

Methods for DNA Isolation from Saliva

DNA was extracted via the Chemagic MSM I with the Chemagic DNA Blood Kit-96 from Perkin Elmer. This kit combines a chemical and mechanical lysis with magnetic bead-based purification.

Saliva samples were incubated at 50°C for 2 hours. The saliva was then transferred to a deep well plate placed on the Chemagic MSM I. The following steps were automated on the MSM I.

M-PVA Magnetic Beads were added to the saliva. Lysis buffer was added to the solution and mixed. The bead-bound DNA was then removed from solution via a 96-rod magnetic head and washed in three Ethanol-based wash buffers. The beads were then washed in a final water wash buffer. Finally, the beads were dipped in elution buffer to resuspend the DNA sample in solution. The beads were then removed from solution, leaving purified DNA eluate.

DNA samples were quantified using a fluorescence-based PicoGreen assay.

Exome Express ICE Methods

Library Construction

Library construction was performed as described in Fisher et al., with the following modifications: initial genomic DNA input into shearing was reduced from 3µg to 10-100ng in 50µL of solution. For adapter ligation, Illumina paired end adapters were replaced with palindromic forked adapters, purchased from Integrated DNA Technologies, with unique dual-indexed molecular barcode sequences to facilitate downstream pooling. With the exception of the palindromic forked adapters, the reagents used for end repair, A-base addition, adapter ligation, and library enrichment PCR were purchased from KAPA Biosciences in 96-reaction kits. In addition, during the post-enrichment SPRI cleanup, elution volume was reduced to 30µL to maximize library concentration, and a vortexing step was added to maximize the amount of template eluted.

In-solution hybrid selection
After library construction, hybridization and capture were performed using the relevant components of Illumina's Nextera Rapid Capture Exome Kit and following the manufacturer's suggested protocol, with the following exceptions: first, all libraries within a library construction plate were pooled prior to hybridization. Second, the Midi plate from Illumina's Nextera Rapid Capture Exome Kit was replaced with a skirted PCR plate to facilitate automation. All hybridization and capture steps were automated on the Agilent Bravo liquid handling system.

Preparation of libraries for cluster amplification and sequencing
After post-capture enrichment, library pools were quantified using qPCR (automated assay on the Agilent Bravo), using a kit purchased from KAPA Biosystems with probes specific to the ends of the adapters. Based on qPCR quantification, libraries were normalized to 2nM, then denatured using 0.1 N NaOH on the Hamilton Starlet. After denaturation, libraries were diluted to 20pM using hybridization buffer purchased from Illumina.

Cluster amplification and sequencing
Cluster amplification of denatured templates was performed according to the manufacturer's protocol (Illumina) using HiSeq 4000 cluster chemistry and HiSeq 4000 flowcells. Flowcells were sequenced on v1 Sequencing-by-Synthesis chemistry for HiSeq 4000 flowcells. The flowcells are then analyzed using RTA v.1.18.64 or later. Each pool of whole exome libraries was run on paired 76bp runs, reading the dual-indexed sequences to identify molecular indices and sequenced across the number of lanes needed to meet coverage for all libraries in the pool.

References
Fisher S, Barry A, Abreu J, Minie B, Nolan J, Delorey TM, Young G, Berlin AM, Blumenstiel B, Cibulskis K, Friedrich D, Johnson R, Juhn F, Reilly B, Shammas R, Stalker J, Sykes SM, Thompson J, Walsh J, Zimmer A, Zwirk Z, Gabriel S, Nicol R, Nusbaum C. A scalable, fully automated process for construction of sequence-ready human exome targeted capture libraries. Genome Biology 2011, 12:R1.

Sequencing Data Analysis

Whole exome sequences were captured using Illumina technology and the sequence data processing and analysis was performed using the Picard and Firehose pipelines at the Broad Institute. The Picard pipeline (http://picard.sourceforge.net) was used to produce a BAM file with aligned reads. This includes alignment to the GRCh37 human reference sequence using the BWA aligner[1] and estimation and recalibration of base quality score with the Genome Analysis Toolkit (GATK)[2]. All sample pairs passed through the Firehose pipeline were subjected to QC testing to test for any tumor/normal and inter-individual contamination as previously described [3,4]. The MuTect algorithm was used to identify somatic mutations[4]. To reduce false positive calls, we additionally analyzed reads covering sites of a putative somatic mutation and realigned them with NovoAlign (www.novocraft.com) and performed additional iteration of MuTect inference on newly aligned BAM files. Furthermore, we filtered for false-positive somatic mutation calls using a panel of normals (PoN), generated from the TCGA dataset, as well as an in-house PoN generated from locally-derived normals, and a FFPE filter and oxoG filter.[5]

Small somatic insertions and deletions were detected using Strelka algorithm[6], MuTect 2 and Snowman (https://github.com/walaj/svaba). Insertions and deletions which were called by two out of the three methods mentioned above were used for analysis.

Somatic mutations including single-nucleotide variants, insertions, and deletions were annotated using Oncotator[7]. The germline somatic variants were analysed using the HaplotypeCaller module of GATK[2]. Significance of identified somatic mutations was analyzed using MutSig2CV[8],which uses patient and gene-specific mutation rates to estimate a background model of predicted mutation incidence across the genome. MutSig2CV then factors in biological co-variates such as replication timing and gene-expression level on a gene-by-gene basis to account for the increased mutational rate of certain classes of genes.
To analyze somatic copy number alterations (SCNA) from whole exome data, we used ReCapseg, which assesses homolog-specific copy ratios from segmental estimates of multipoint allelic copy ratios at heterozygous loci incorporating the statistical phasing software (BEAGLE) and population haplotype panels (HAPMAP3)[9,10].

For copy number alteration significance analysis, segmented copy number data was analyzed by GISTIC 2.0, to identify significantly recurring focal and arm-level amplification/deletion peaks.[11] Allele-specific SCNAs and tumor ploidy/purity status were assessed using ABSOLUTE.[12]
Tumor samples which had a purity of 10% or more were submitted to cBioPortal (http://www.cbioportal.org/index.do).

See Appendix 2, pg.11 for detailed pipeline information.

References

1. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25,** 1754–60 (2009).
2. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20,** 1297–303 (2010).
3. Berger, M. F. *et al.* The genomic complexity of primary human prostate cancer. *Nature* **470,** 214–20 (2011).
4. Cibulskis, K. *et al.* Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* **31,** 213–9 (2013).
5. Costello, M. *et al*. Discovery and characterization of artifactual mutations in deep coverage targeted capture sequencing data due to oxidative DNA damage during sample preparation. *Nucleic Acids Res.* **41**, e67– (2013).
6. Saunders, C. T. *et al.* Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics* **28,** 1811–7 (2012).
7. Ramos, A. H. *et al.* Oncotator: cancer variant annotation tool. *Hum. Mutat.* **36,** E2423-9 (2015).
8. Lawrence, M. et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature **499**, 214-218 (2013)

9. Browning, B. L. & Yu, Z. Simultaneous genotype calling and haplotype phasing improves genotype accuracy and reduces false-positive associations for genome-wide association studies. *Am. J. Hum. Genet.* 85, 847–861 (2009).
10. International HapMap, C. et al. Integrating common and rare genetic variation in diverse human populations. *Nature* 467, 52–58 (2010).
11. Beroukhim, R. *et al.* Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma. *Proc. Natl Acad. Sci. USA* **104**, 20007–20012(2007)
12. Carter, S.L. *et al*. Absolute quantification of somatic DNA alterations in human cancer. *Nat. Biotechnol.* **30**, 413–421 (2012).

Appendix 1: Data Dictionary

Pathology Data	
Biopsy date	MM/DD/YYYY
Biopsy location	Open text field- location where procedure was performed
Procedure sidedness	Left Right (Applies to breast, axillary lymph node, regional lymph node)
Biopsy procedure type	Core Biopsy Punch Biopsy Blood Biopsy Mastectomy Resection Excision Other
Histology	IDC ILC Mixed IDLC Carcinoma Adenocarcinoma Unknown
DCIS present	Yes No
LCIS present	Yes No
Grade	I Low Grade II Intermediate Grade III High Grade Unknown
Estrogen receptor status	Positive Negative Unknown
Estrogen receptor percentage	0-100% Unknown
Progesterone receptor status	Positive Negative Unknown
Progesterone receptor percentage	0-100% Unknown
HER2 status overall	Positive Negative Unknown

HER2 IHC score	3+ 2+ 1+ Unknown
HER2 FISH	Positive Negative Unknown
HER2 Ratio	Numeric value
HER2 copy number	Numeric value
CEP17 cop number	Numeric value
Ki67 score	0-100%
Medical Record Data	
DOB	Date of birth (used to calculate age at dx)
Date of diagnosis	Date of diagnosis with primary breast cancer
Diagnosis stage	I II III IV
T stage	1 2 3
N stage	1 2 3
Diagnosis location	Open text field- location on body where diagnostic procedure was performed
Diagnosis procedure type	Core Biopsy Punch Biopsy Blood Biopsy Mastectomy Resection Excision Other
Diagnosis procedure sidedness	Left Right (Applies to breast, axillary lymph node, regional lymph node)
Diagnosis histology	IDC ILC Mixed IDLC Carcinoma Adenocarcinoma

Diagnosis grade	I Low Grade II Intermediate Grade III High Grade Unknown
Estrogen receptor status at diagnosis	Positive Negative Unknown
Estrogen receptor percentage at diagnosis	0-100% Unknown
Progesterone receptor status at diagnosis	Positive Negative Unknown
Progesterone receptor percentage at diagnosis	0-100% Unknown
HER2 status overall at diagnosis	Positive Negative Unknown
HER2 IHC score at diagnosis	3+ 2+ 1+ Unknown
HER2 FISH at diagnosis	Positive Negative Unknown
HER2 ratio at diagnosis	Numeric value
HER2 copy number at diagnosis	Numeric value
CEP17 cop number at diagnosis	Numeric value
Surgery at diagnosis	Mastectomy Lumpectomy None
Adjuvent radiation	Yes No
Date of diagnosis with metastatic disease	Date of confirmed metastatic disease
Metastatic sites at diagnosis	List- location(s) of metastatic disease at diagnosis
Metastatic sites (all)	List- location(s) where metastatic disease was ever present
Germline mutation- BRCA1	Yes No
Germline mutation- BRCA2	Yes No
Germline mutation- PALB2	Yes

	No
Germline mutation- other	List
Treatment	Drug name (generic)
Treatment start date	MM/DD/YYYY
Treatment stop date	MM/DD/YYYY

Appendix 2: Commands used to execute the methods

Key	Workflow/Task	Commands
Contest	**ContEst for Capture Array-Free**	java -Djava.io.tmpdir=/fh/subscription-XRC/ContaminationAnalysis/MBC-MBCProject_Normal_Sample/27115945/tmp -Xmx512m -jar /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/ContaminationAnalysis/broadinstitute.org/cancer.genome.analysis/00262/107//Queue-1.4-437-g6b8a9e1-svn-35362.jar -S /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/ContaminationAnalysis/broadinstitute.org/cancer.genome.analysis/00262/107//ContaminationPipeline.scala -reference /seq/references/Homo_sapiens_assembly19/v1/Homo_sapiens_assembly19.fasta -interval /xchip/cga/reference/hg19/whole_exome_agilent_1.1_refseq_plus_3_boosters_plus_10bp_padding_minus_mito.Homo_sapiens_assembly19.targets.interval_list -out MBC-MBCProject_Normal_Sample.contamination.txt -bam /seq/picard_aggregation/RP-1156/Exome/MBCProject_pt_id_SALIVA/v2/MBCProject_pt_id_SALIVA.bam -nbam /seq/picard_aggregation/RP-1156/Exome/MBCProject_pt_id_SALIVA/v2/MBCProject_pt_id_SALIVA.bam -array none -pop /xchip/cga/reference/hg19/hg19_population_stratified_af_hapmap_3.3.fixed.vcf -faf true -run -array_interval /xchip/cga/reference/hg19/SNP6.hg19.interval_list
SNV	**Run Mutect with Realignment Filter Pairs NovoAlign and Agilent & ICE & CCPM PoN filter (Capture-Pair) for FFPE or FF without ExAC CSQ**	**Call Somatic Mutations for Capture Workflow without ExAC CSQ** **1.Call Somatic Mutations for Capture (configure) -** sh /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/CallSomaticMutations/broadinstitute.org/cancer.genome.analysis/00004/131/runBroadJava7.sh java -Xmx2g -jar /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/CallSomaticMutations/broadinstitute.org/cancer.genome.analysis/00004/131/muTect-1.1.6.jar --read_group_black_list /xchip/cga/reference/read_group_blacklist.v1.txt --analysis_type MuTect --intervals /fh/subscription-XRC/CallSomaticMutations/MBC-MBCProject_Tumor_Normal/27003052/iteration1/./gatk-scatter.0000000043.interval_list --normal_sample_name MBC-MBCProject_Normal_Sample -I:normal /seq/picard_aggregation/RP-1156/Exome/MBCProject_pt_id_SALIVA/v2/MBCProject_pt_id_SALIVA.bam --tumor_sample_name MBC-MBCProject_Tumor_Sample-I:tumor /seq/picard_aggregation/RP-1156/Exome/MBCProject_pt_id_T1/v2/MBCProject_pt_id_T1.bam --reference_sequence /seq/references/Homo_sapiens_assembly19/v1/Homo_sapiens_assembly19.fasta --dbsnp /xchip/cga/reference/hg19/dbsnp_134_b37.leftAligned.vcf --cosmic /xchip/cga/reference/hg19/hg19_cosmic_v54_120711.vcf --normal_panel /xchip/cga/reference/hg19/refseq_exome_10bp_hg19_300_1kg_normal_panel.vcf --out MBC-MBCProject_Tumor_Normal.call_stats.txt --coverage_file MBC-MBCProject_Tumor_Normal.coverage.wig.txt --power_file MBC-MBCProject_Tumor_Normal.power.wig.txt --downsample_to_coverage 10000 --enable_extended_output --fraction_contamination 0.007 **2.CallStats to MAFLite for Capture (configure) -** /usr/bin/perl /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/CallstatsToMaflite/broadinstitute.org/cancer.genome.analysis/00162/14/call_stats_to_maflite.pl /fh/subscription-XRC/CallSomaticMutations/MBC-MBCProject_Tumor_Normal/27003052/iteration1/MBC-MBCProject_Tumor_Normal.call_stats.txt 37 FSTAR MBC-MBCProject_Tumor_Normal.maf tumor_f,init_t_lod,t_lod_fstar,t_alt_count,t_ref_count,judgement **3.Summarize Somatic Coverage for Capture (configure) -** perl /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/SummarizeWigFile/broadinstitute.org/cancer.genome.analysis/10246/1/summarizeWigFile.pl /fh/subscription-XRC/CallSomaticMutations/MBC-MBCProject_Tumor_Normal/27003052/iteration1/MBC-MBCProject_Tumor_Normal.coverage.wig.txt.gz MBC-MBCProject_Tumor_Normal.somatic_coverage_summary.txt **Oncotate_Capture_SNP_Workflow_without_ExAC_CSQ** **4.Apply SNP Maflite Validation for Capture (configure)** /broad/software/free/Linux/redhat_5_x86_64/pkgs/sun-java-jdk_1.6.0-21_x86_64/bin/java -Xmx1g -jar /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/ApplyMAFValidation/broadinstitute.org/cancer.genome.analysis/00163/23/ApplyMAFValidation.jar M=/fh/subscription-XRC/CallstatsToMaflite/MBC-MBCProject_Tumor_Normal/27022277/MBC-MBCProject_Tumor_Normal.maf OUTPUT_MAF=MBC-MBCProject_Tumor_Sample.maf.annotated MATCH_MODE=Sample V=/cga/tcga-gsc/svnreference/validation **5.Oncotate SNP for Capture (configure)** sh /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/Oncotator_v1/broadinstitute.org/cancer.genome.analysis/10202/68/oncotator.sh MAFLITE TCGAMAF /fh/subscription-XRC/ApplyMAFValidation/MBC-MBCProject_Tumor_Normal/27022407/MBC-MBCProject_Tumor_Sample.maf.annotated MBC-MBCProject_Tumor_Normal.snp.capture.maf.annotated hg19 /xchip/cga/reference/annotation/db/oncotator_v1_ds_Sept292015/ /xchip/cga/reference/annotation/db/tcgaMAFManualOverrides2.4.config /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/Oncotator_v1/broadinstitute.org/cancer.genome.analysis/10202/68/EFFECT /xchip/tcga/Tools/python_fh/python_fh_env_March302015 --log_name oncotator_firehose.log --prepend **6.Oncotator - Drop ExAC CSQ field (configure)** python /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/SubsetMafColumns/broadinstitute.org/cancer.genome.analysis/10917/16//subset_maf_by_columns.py --maf /fh/subscription-XRC/Oncotator_v1/MBC-MBCProject_Tumor_Normal/27022437/MBC-MBCProject_Tumor_Normal.snp.capture.maf.annotated --columns /home/unix/dkim/reference/firehose/oncotator_whitelist_no_ExAC_CSQ.txt --out MBC-MBCProject_Tumor_Normal.snp.capture.maf.annotated.no_exac_csq **7.PoN filter (configure)** /broad/software/free/Linux/redhat_5_x86_64/pkgs/python_2.7.1-sqlite3-rtrees/bin/python /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/PoN_filter_py/broadinstitute.org/cancer.genome.analysis/11041/16//PON_filter_fh.py -m /fh/subscription-XRC/Oncotator_v1/MBC-MBCProject_Tumor_Normal/30662860/MBC-MBCProject_Tumor_Normal.snp.capture.maf.annotated -p /cga/fh/pancan_data/pon/pon12/final_summed_tokens.hist.bin -g 0.005 -t 5 -wc 0.5 -pc 0.5 -a 0.2 -r TRUE -o MBC-MBCProject_Tumor_Normal.PoNfiltered.maf -n 1 -l -2.5 -w_c 0.5 --use_static **Capture_OxoG_Filter_v3_Workflow_without_ExAC_CSQ** **8.annotate Picard OxoQ for Capture (configure)** sh /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/annotatePicardOxoQ/broadinstitute.org/cancer.genome.analysis/10360/6//annotatePicardOxoQ.sh -i MBC-MBCProject_Tumor_Normal -b /seq/picard_aggregation/RP-1156/Exome/MBCProject_pt_id_T1/v2/MBCProject_pt_id_T1.bam -r /seq/references/Homo_sapiens_assembly19/v1/Homo_sapiens_assembly19.fasta -d /xchip/cga/reference/hg19/dbsnp_134_b37.leftAligned.vcf -c CCG -o . **9.Create OxoG Interval List for Capture (configure)** sh /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/createOxoGIntervalList/broadinstitute.org/cancer.genome.analysis/10067/22//createOxoGIntervals.sh /fh/subscription-XRC/Oncotator_v1/MBC-MBCProject_Tumor_Normal/27022437/MBC-MBCProject_Tumor_Normal.snp.capture.maf.annotated MBC-MBCProject_Tumor_Normal.oxoG.interval_list

10.Append Picard OxoQ to Maf for Capture without ExAC CSQ (configure)
sh
/xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/AppendAnnotation2MAF/broadinstitute.org/cancer.genome.analysis/10363/16//AppendAnnotation2MAF.sh -i MBC-MBCProject_Tumor_Normal -m /fh/subscription-XRC/SubsetMafColumns/MBC-MBCProject_Tumor_Normal/27023157/MBC-MBCProject_Tumor_Normal.snp.capture.maf.annotated.no_exac_csq.lite.maf -f picard_oxoQ -v 52.52 -o .
11.Create OxoG Metrics for Capture (configure)
java -Xmx2g -jar
/xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/oxoGMetrics/broadinstitute.org/cancer.genome.analysis/10069/8//GenomeAnalysisTK.jar --analysis_type OxoGMetrics -R /seq/references/Homo_sapiens_assembly19/v1/Homo_sapiens_assembly19.fasta -I /seq/picard_aggregation/RP-1156/Exome/MBCProject_pt_id_T1/v2/MBCProject_pt_id_T1.bam -L /fh/subscription-XRC/createOxoGIntervalList/MBC-MBCProject_Tumor_Normal/27072480/MBC-MBCProject_Tumor_Normal.oxoG.interval_list -o MBC-MBCProject_Tumor_Normal.oxoG.metrics.txt
12.Append OxoG Information to MAF for Capture (configure)
sh
/xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/appendOxoGInfo/broadinstitute.org/cancer.genome.analysis/10138/8//appendOxoGInfo.sh --onlyAddColumnsToCopy /fh/subscription-XRC/oxoGMetrics/MBC-MBCProject_Tumor_Normal/27072528/MBC-MBCProject_Tumor_Normal.oxoG.metrics.txt /fh/subscription-XRC/AppendAnnotation2MAF/MBC-MBCProject_Tumor_Normal/27072507/MBC-MBCProject_Tumor_Normal.picard_oxoQ.maf.annotated MBC-MBCProject_Tumor_Normal.oxoGInfo.maf.annotated
13.Filter OxoG Artifact Third Incarnation for Capture (configure)
sh /xchip/tcga/gdac_prod/source/analysis_pipeline/scripts/choose_run_matlab2.sh -v .matlab-2013a
/xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/oxoGFilter_v3/broadinstitute.org/cancer.genome.analysis/10139/69/ --with-display
/xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/oxoGFilter_v3/broadinstitute.org/cancer.genome.analysis/10139/69/startFilterMAFFile /fh/subscription-XRC/appendOxoGInfo/MBC-MBCProject_Tumor_Normal/27115529/MBC-MBCProject_Tumor_Normal.oxoGInfo.maf.annotated MBC-MBCProject_Tumor_Normal.oxoG3.maf.annotated ./ 0 1 0.96 0.01 -1 36 1.5
14.Generate OxoGv3 Report for Capture (configure)
Rscript
/xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/oxoGFilter_v3Report/broadinstitute.org/cancer.genome.analysis/10153/41//createReport.R -L/xchip/tcga/Tools/Nozzle/v1.current -C/fh/subscription-XRC/oxoGFilter_v3/MBC-MBCProject_Tumor_Normal/27115552/allCases.txt -F/fh/subscription-XRC/oxoGFilter_v3/MBC-MBCProject_Tumor_Normal/27115552/figures -T/fh/subscription-XRC/oxoGFilter_v3/MBC-MBCProject_Tumor_Normal/27115552/caseTableData.tsv -M/fh/subscription-XRC/oxoGFilter_v3/MBC-MBCProject_Tumor_Normal/27115552/MBC-MBCProject_Tumor_Normal.oxoG3.maf.annotated -R/fh/subscription-XRC/oxoGFilter_v3/MBC-MBCProject_Tumor_Normal/27115552/MBC-MBCProject_Tumor_Normal.oxoG3.maf.annotated.all.maf.annotated

Capture_Filter_FFPE_Workflow
15.annotate FFPE bias for Capture (configure)
sh
/xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/annotateFFPEbias/broadinstitute.org/cancer.genome.analysis/10653/7//annotateFFPEbias.sh -i MBC-MBCProject_Tumor_Normal -b /seq/picard_aggregation/RP-1156/Exome/MBCProject_pt_id_T1/v2/MBCProject_pt_id_T1.bam -r
/seq/references/Homo_sapiens_assembly19/v1/Homo_sapiens_assembly19.fasta -d /xchip/cga/reference/hg19/dbsnp_134_b37.leftAligned.vcf -f G -t A -c .CG -x ffpe_metrics -s 40000000 -o .
16.Append ffpeQ to Maf for Capture (configure)
sh
/xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/AppendAnnotation2MAF/broadinstitute.org/cancer.genome.analysis/10363/16//AppendAnnotation2MAF.sh -i MBC-MBCProject_Tumor_Normal -m /fh/subscription-XRC/oxoGFilter_v3/MBC-MBCProject_Tumor_Normal/27115552/MBC-MBCProject_Tumor_Normal.oxoG3.maf.annotated -f ffpe_Q -v 31.50 -o .
17.Filter FFPE orientation bias for capture (configure)
sh /xchip/tcga/gdac_prod/source/analysis_pipeline/scripts/choose_run_matlab2.sh -v .matlab-2012a
/xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/orientationBiasFilter/broadinstitute.org/cancer.genome.analysis/10884/13/ --with-display
/xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/orientationBiasFilter/broadinstitute.org/cancer.genome.analysis/10884/13/orientationBiasFilter /fh/subscription-XRC/AppendAnnotation2MAF/MBC-MBCProject_Tumor_Normal/27115594/MBC-MBCProject_Tumor_Normal.ffpe_Q.maf.annotated MBC-MBCProject_Tumor_Normal.ffpeBias.maf.annotated ./ 0 1 0.96 0.01 -1 30 1.5 G A i_ffpe
18.ffpe_filter_Report (configure)
Rscript
/xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/ffpeReport/broadinstitute.org/cancer.genome.analysis/10661/6//ffpeReport.R -L/xchip/tcga/Tools/Nozzle/v1.current -C/fh/subscription-XRC/orientationBiasFilter/MBC-MBCProject_Tumor_Normal/27115864/allCases.txt -F/fh/subscription-XRC/orientationBiasFilter/MBC-MBCProject_Tumor_Normal/27115864/figures/ -T/fh/subscription-XRC/orientationBiasFilter/MBC-MBCProject_Tumor_Normal/27115864/caseTableData.tsv -M/fh/subscription-XRC/orientationBiasFilter/MBC-MBCProject_Tumor_Normal/27115864/MBC-MBCProject_Tumor_Normal.ffpeBias.maf.annotated -R/fh/subscription-XRC/orientationBiasFilter/MBC-MBCProject_Tumor_Normal/27115864/MBC-MBCProject_Tumor_Normal.ffpeBias.unfiltered.maf.annotated

Call Somatic Mutations for Capture Workflow Realignment Filter Pairs NovoAlign
19.Realignment_Filter Paired Reads novoalign (FFPE or FF)
sh
/xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/ReadAnalysis_QC_PairedReads_bwamem/broadinstitute.org/cancer.genome.analysis/10845/21/readalignment_wrapper.sh
/xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/ReadAnalysis_QC_PairedReads_bwamem/broadinstitute.org/cancer.genome.analysis/10845/21/ /fh/subscription-XRC/orientationBiasFilter/MBC-MBCProject_Tumor_Normal/27115864/MBC-MBCProject_Tumor_Normal.ffpeBias.maf.annotated /seq/picard_aggregation/RP-1156/Exome/MBCProject_pt_id_T1/v2/MBCProject_pt_id_T1.bam /seq/picard_aggregation/RP-1156/Exome/MBCProject_pt_id_SALIVA/v2/MBCProject_pt_id_SALIVA.bam
/seq/references/Homo_sapiens_assembly19/v1/Homo_sapiens_assembly19.fasta
/xchip/cga_home/mara/projects/readalignment/novoalign//hg19.decoy.nix MBC-MBCProject_Tumor_Sample MBC-MBCProject_Normal_Sample MBC-MBCProject_Tumor_Normal realigned_bam_mutation_interval_list_length_wex_pairs_novo bamfile_mutect_realignreads_control_wex_pairs_novo bamfile_mutect_realignreads_case_wex_pairs_novo novoalign debug
20.Call Somatic Mutations for WEX Pairs Novoalign Realign
sh
/xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/CallSomaticMutations/broadinstitute.org/cancer.genome.analysis/00004/131/runBroadJava7.sh java -Xmx2g -jar
/xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/CallSomaticMutations/broadinstitute.org/cancer.genome.analysis/00004/131/muTect-1.1.6.jar --analysis_type MuTect --intervals /fh/subscription-XRC/CallSomaticMutations/MBC-MBCProject_Tumor_Normal/30661616/iteration1/./gatk-scatter.0000000009.interval_list --normal_sample_name MBC-MBCProject_Normal_Sample -I:normal /fh/subscription-XRC/ReadAnalysis_QC_PairedReads_bwamem/MBC-MBCProject_Tumor_Normal/30657886/MBC-MBCProject_Normal_Sample-final_filtered_control.bam --tumor_sample_name MBC-MBCProject_Tumor_Sample-I:tumor /fh/subscription-XRC/ReadAnalysis_QC_PairedReads_bwamem/MBC-MBCProject_Tumor_Normal/30657886/MBC-MBCProject_Tumor_Sample-final_filtered_case.bam --reference_sequence
/seq/references/Homo_sapiens_assembly19/v1/Homo_sapiens_assembly19.fasta --dbsnp /xchip/cga/reference/hg19/dbsnp_134_b37.leftAligned.vcf --cosmic /xchip/cga/reference/hg19/hg19_cosmic_v54_120711.vcf --out MBC-MBCProject_Tumor_Normal.call_stats.txt --coverage_file MBC-MBCProject_Tumor_Normal.coverage.wig.txt --power_file MBC-MBCProject_Tumor_Normal.power.wig.txt --downsample_to_coverage 9999999 -

		-enable_extended_output --fraction_contamination 0.007 --force_output **21.Realignment Filter Call Stats WEX Pairs novo** sh /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/ReadAnalysis_QC_CallStatsFilter/broadinstitute.org/cancer.genome.analysis/10215/9/callstats_filter_wrapper.sh /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/ReadAnalysis_QC_CallStatsFilter/broadinstitute.org/cancer.genome.analysis/10215/9/ /fh/subscription-XRC/CallSomaticMutations/MBC-MBCProject_Tumor_Normal/27003052/iteration1/MBC-MBCProject_Tumor_Normal.call_stats.txt /fh/subscription-XRC/CallSomaticMutations/MBC-MBCProject_Tumor_Normal/30661616/iteration1/MBC-MBCProject_Tumor_Normal.call_stats.txt MBC-MBCProject_Tumor_Normal **22.CallStats to MAFLite for Capture Realign Novo Pairs** /usr/bin/perl /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/CallstatsToMaflite/broadinstitute.org/cancer.genome.analysis/00162/14/call_stats_to_maflite.pl /fh/subscription-XRC/ReadAnalysis_QC_CallStatsFilter/MBC-MBCProject_Tumor_Normal/30662397/MBC-MBCProject_Tumor_Normal.filtered.annotated.callstats.txt 37 FSTAR MBC-MBCProject_Tumor_Normal.maf tumor_f,init_t_lod,t_lod_fstar,t_alt_count,t_ref_count,judgement **23.Apply SNP Maflite Validation for Capture Realign Pairs Novo** /broad/software/free/Linux/redhat_5_x86_64/pkgs/sun-java-jdk_1.6.0-21_x86_64/bin/java -Xmx1g -jar /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/ApplyMAFValidation/broadinstitute.org/cancer.genome.analysis/00163/23/ApplyMAFValidation.jar M=/fh/subscription-XRC/CallstatsToMaflite/MBC-MBCProject_Tumor_Normal/30662533/MBC-MBCProject_Tumor_Normal.maf OUTPUT_MAF=MBC-MBCProject_Tumor_Sample.maf.annotated MATCH_MODE=Sample V=/cga/tcga-gsc/svnreference/validation **24.Oncotate SNP for Capture Realign Pairs Novo** sh /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/Oncotator_v1/broadinstitute.org/cancer.genome.analysis/10202/70/oncotator.sh MAFLITE TCGAMAF /fh/subscription-XRC/ApplyMAFValidation/MBC-MBCProject_Tumor_Normal/30662698/MBC-MBCProject_Tumor_Sample.maf.annotated MBC-MBCProject_Tumor_Normal.snp.capture.maf.annotated hg19 /xchip/cga/reference/annotation/db/oncotator_v1_ds_Sept292015/ /xchip/cga/reference/annotation/db/tcgaMAFManualOverrides2.4.config /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/Oncotator_v1/broadinstitute.org/cancer.genome.analysis/10202/70/ EFFECT /xchip/tcga/Tools/python_fh/python_fh_env_March302015 --log_name oncotator_firehose.log --prepend **25.PoN Filter (Agilent) - Capture Realignment Filter Pairs NovoAlign** /broad/software/free/Linux/redhat_5_x86_64/pkgs/python_2.7.1-sqlite3-rtrees/bin/python /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/PoN_filter_py/broadinstitute.org/cancer.genome.analysis/11041/22//PON_filter_fh.py -m /fh/subscription-XRC/Oncotator_v1/MBC-MBCProject_Tumor_Normal/30662860/MBC-MBCProject_Tumor_Normal.snp.capture.maf.annotated -p /cga/fh/pancan_data/pon/pon9/final_summed_tokens.hist.bin -g 0.005 -t 5 -wc 0.5 -pc 0.5 -a 0.2 -r TRUE -o MBC-MBCProject_Tumor_Normal.novoalign.PoNfiltered_agilent.maf -n 1 -l -2.5 -w_c 0.5 --use_static **26.PoN Filter (ICE from Agilent) - Capture Realignment Filter Pairs NovoAlign** /broad/software/free/Linux/redhat_5_x86_64/pkgs/python_2.7.1-sqlite3-rtrees/bin/python /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/PoN_filter_py/broadinstitute.org/cancer.genome.analysis/11041/24//PON_filter_fh.py -m /fh/subscription-XRC/PoN_filter_py/MBC-MBCProject_Tumor_Normal/30688082/MBC-MBCProject_Tumor_Normal.novoalign.PoNfiltered_agilent.maf -p /cga/fh/pancan_data/pon/pon12/final_summed_tokens.hist.bin -g 0.005 -t 5 -wc 0.5 -pc 0.5 -a 0.2 -r TRUE -o MBC-MBCProject_Tumor_Normal.novoalign.PoNfiltered_agilent_ICE.maf -n 1 -l -2.5 -w_c 0.5 --use_static --no_pon_columns
Indel (1)	**Run Strelka with MafMaster Filter (Capture-Pair)**	**1.Run Strelka on Capture for Pairs -** sh /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/RunStrelka/broadinstitute.org/cancer.genome.analysis/10439/11/runStrelka.sh /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/RunStrelka/broadinstitute.org/cancer.genome.analysis/10439/11/ /seq/picard_aggregation/RP-1156/Exome/MBCProject_pt_id_T1/v2/MBCProject_pt_id_T1.bam /seq/picard_aggregation/RP-1156/Exome/MBCProject_pt_id_SALIVA/v2/MBCProject_pt_id_SALIVA.bam /seq/references/Homo_sapiens_assembly19/v1/Homo_sapiens_assembly19.fasta MBC-MBCProject_Tumor_Normal strelka_config_bwa_cga_exome.ini **2.Annotate Pass Strelka Indel VCF for Pairs -** python /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/OncotatorForStrelka/broadinstitute.org/cancer.genome.analysis/10543/8//oncotator/Oncotator.py -v --input_format=VCF --output_format=TCGAMAF --db-dir=/xchip/cga/reference/annotation/db/oncotator_v1_ds_April052016 --no-multicore --default_config /xchip/cga/reference/annotation/db/tcgaMAFManualOverrides2.4.config /fh/subscription-XRC/FixStrelkaVCFForAnnotation/MBC-MBCProject_Tumor_Normal/27011318/tumorOnly.fixed.vcf MBC-MBCProject_Tumor_Normal.strelka.pass.somatic.indels.maf hg19 -a center:broad.mit.edu -a individual_barcode:MBC-MBCProject_pt_id -d normal_barcode:MBC-MBCProject_Normal_Sample -a platform:illumina -a normal_uuid:NA -a tumor_barcode:MBC-MBCProject_Tumor_Sample -a tumor_uuid:NA -a tumor_subtype:Metastatic **3.Annotate Pass Strelka SNV VCF for Pairs -** python /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/OncotatorForStrelka/broadinstitute.org/cancer.genome.analysis/10543/8//oncotator/Oncotator.py -v --input_format=VCF --output_format=TCGAMAF --db-dir=/xchip/cga/reference/annotation/db/oncotator_v1_ds_April052016 --no-multicore --default_config /xchip/cga/reference/annotation/db/tcgaMAFManualOverrides2.4.config /fh/subscription-XRC/FixStrelkaVCFForAnnotation/MBC-MBCProject_Tumor_Normal/27011317/tumorOnly.fixed.vcf MBC-MBCProject_Tumor_Normal.strelka.pass.somatic.snvs.maf hg19 -a center:broad.mit.edu -a individual_barcode:MBC-MBCProject_pt_id -d normal_barcode:MBC-MBCProject_Normal_Sample -a platform:illumina -a normal_uuid:NA -a tumor_barcode:MBC-MBCProject_Tumor_Sample -a tumor_uuid:NA -a tumor_subtype:Metastatic **4.Strelka SNV fix INDEL allele counts -** sh /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/strelkamaf2tcgamaf/broadinstitute.org/cancer.genome.analysis/11325/1//strelkamaf2tcgamaf.sh -i MBC-MBCProject_Tumor_Normal -m /fh/subscription-XRC/AnnotateStrelkaVCF/MBC-MBCProject_Tumor_Normal/31666346/OncotatorForStrelka_31666749/MBC-MBCProject_Tumor_Normal.strelka.pass.somatic.indels.maf -x strelka.fix.indel.maf -o . **5.Strelka SNV fix SNV allele counts -** sh /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/strelkamaf2tcgamaf/broadinstitute.org/cancer.genome.analysis/11325/1//strelkamaf2tcgamaf.sh -i MBC-MBCProject_Tumor_Normal -m /fh/subscription-XRC/AnnotateStrelkaVCF/MBC-MBCProject_Tumor_Normal/31666342/OncotatorForStrelka_31666768/MBC-MBCProject_Tumor_Normal.strelka.pass.somatic.snvs.maf -x strelka.fix.snv.maf -o . **6.Combine SNP and Indel Strelka MAF for Capture -** python /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/tsvCatFiles/broadinstitute.org/cancer.genome.analysis/10155/3//tsvConcatFiles.py /fh/subscription-XRC/strelkamaf2tcgamaf/MBC-MBCProject_Tumor_Normal/31667385/output/MBC-MBCProject_Tumor_Normal.strelka.fix.snv.maf /fh/subscription-XRC/strelkamaf2tcgamaf/MBC-MBCProject_Tumor_Normal/31667383/output/MBC-MBCProject_Tumor_Normal.strelka.fix.indel.maf --outputFilename=MBC-MBCProject_Tumor_Normal.combined.maf **7.PoN_filter_indels -**

		sh /xchip/tcga/gdac_prod/source/analysis_pipeline/scripts/choose_run_matlab2.sh -v .matlab-2012a /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/MafMasterFilter/broadinstitute.org/cancer.genome.analysis/10822/1/ /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/MafMasterFilter/broadinstitute.org/cancer.genome.analysis/10822/1/master_filter_wrapper /fh/subscription-XRC/AnnotateStrelkaVCF/MBC-MBCProject_Tumor_Normal/31666346/OncotatorForStrelka_31666749/MBC-MBCProject_Tumor_Normal.strelka.pass.somatic.indels.maf /cga/fh/pancan_data/pon/pon10/final_summed_tokens.hist.bin MBC-MBCProject_Tumor_Normal **8.Strelka PoN Filter (ICE from Agilent) -** sh /xchip/tcga/gdac_prod/source/analysis_pipeline/scripts/choose_run_matlab2.sh -v .matlab-2012a /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/MafMasterFilter/broadinstitute.org/cancer.genome.analysis/10822/1/ /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/MafMasterFilter/broadinstitute.org/cancer.genome.analysis/10822/1/master_filter_wrapper /fh/subscription-XRC/MafMasterFilter/MBC-MBCProject_Tumor_Normal/31668062/MBC-MBCProject_Tumor_Normal.pon_filtered.txt /cga/fh/pancan_data/pon/pon12/final_summed_tokens.hist.bin MBC-MBCProject_Tumor_Normal **9.Strelka PoN Filter (CCPM from ICE from Agilent) -** sh /xchip/tcga/gdac_prod/source/analysis_pipeline/scripts/choose_run_matlab2.sh -v .matlab-2012a /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/MafMasterFilter/broadinstitute.org/cancer.genome.analysis/10822/1/ /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/MafMasterFilter/broadinstitute.org/cancer.genome.analysis/10822/1/master_filter_wrapper /fh/subscription-XRC/MafMasterFilter/MBC-MBCProject_Tumor_Normal/31668510/MBC-MBCProject_Tumor_Normal.pon_filtered.txt /cga/fh/pancan_data/pon/pon13/final_summed_tokens.hist.bin MBC-MBCProject_Tumor_Normal **10.Aggregate Strelka Indels - CCPM/ICE/Agilent filtered -** python /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/tsvCat/broadinstitute.org/cancer.genome.analysis/10083/10//tsvConcatListFile.py /fh/subscription-XRC/tsvCat/PR_Wagle_MBC_Pairs_Capture_All_Pairs/33627161/inputListFile.input.tsv PR_Wagle_MBC_Pairs_Capture_All_Pairs.strelka_indels_after_ccpm_ice_agilent
Indel (2)	**M2**	**1. M2 Scatter -** sh /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/M2/broadinstitute.org/cancer.genome.analysis/10855/28/runBroadJava7.sh java -Xmx4g -jar /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/M2/broadinstitute.org/cancer.genome.analysis/10855/28/GenomeAnalysisTK_latest_unstable.jar --analysis_type M2 --intervals /fh/subscription-XRC/M2/MBC-MBCProject_Tumor_Normal/32671507/iteration1/./gatk-scatter.0000000020.interval_list -I:normal /seq/picard_aggregation/RP-1156/Exome/MBCProject_pt_id_SALIVA/v2/MBCProject_pt_id_SALIVA.bam -I:tumor /seq/picard_aggregation/RP-1156/Exome/MBCProject_pt_id_T1/v2/MBCProject_pt_id_T1.bam --reference_sequence /seq/references/Homo_sapiens_assembly19/v1/Homo_sapiens_assembly19.fasta -contamination 0.007 --dbsnp /xchip/cga/home/kcibul/analysis/dream/mutect_ref/dbsnp_138.b37.vcf --cosmic /xchip/cga/reference/hg19/hg19_cosmic_v54_120711.vcf --normal_panel /cga/tcga-gsc/mutect/panel_of_normals/panel_of_normals_m2_paad/agilent_hg19_m2_paad_149_normal_panel.vcf --out MBC-MBCProject_Tumor_Normal.full.vcf **2.VCF Filter to Pass -** sh /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/VCF_filterpass/broadinstitute.org/cancer.genome.analysis/10856/8/wrapper.sh /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/VCF_filterpass/broadinstitute.org/cancer.genome.analysis/10856/8/ /fh/subscription-XRC/M2/MBC-MBCProject_Tumor_Normal/32671507/MBC-MBCProject_Tumor_Normal.full.vcf MBC-MBCProject_Tumor_Normal /seq/picard_aggregation/RP-1156/Exome/MBCProject_pt_id_T1/v2/MBCProject_pt_id_T1.bam /seq/references/Homo_sapiens_assembly19/v1/Homo_sapiens_assembly19.fasta **3.Oncotator for M2 -** sh /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/OncotatorForStrelka/broadinstitute.org/cancer.genome.analysis/10543/8/oncotator.sh VCF TCGAMAF /fh/subscription-XRC/VCF_filterpass/MBC-MBCProject_Tumor_Normal/32684786/MBC-MBCProject_Tumor_Normal.pass.vcf MBC-MBCProject_Tumor_Normal.m2_pass.maf hg19 /xchip/cga/reference/annotation/db/oncotator_v1_ds_Sept292015/ /xchip/cga/reference/annotation/db/tcgaMAFManualOverrides2.4.config /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/OncotatorForStrelka/broadinstitute.org/cancer.genome.analysis/10543/8/ center:broad.mit.edu individual_barcode:MBC-MBCProject_pt_id normal_barcode:MBC-MBCProject_Normal_Sample platform:illumina normal_uuid:MBC-MBCProject_Normal_Sample tumor_barcode:MBC-MBCProject_Tumor_Sampletumor_uuid:MBC-MBCProject_Tumor_Sampletumor_subtype:NA /xchip/tcga/Tools/python_fh/python_fh_env_March302015/ **4.Add Allelic Fraction Columns to MAF -** python /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/M2_reformat_maf/broadinstitute.org/cancer.genome.analysis/10997/1/make_oxoG_compatible.py --maf /fh/subscription-XRC/OncotatorForStrelka/MBC-MBCProject_Tumor_Normal/32684814/MBC-MBCProject_Tumor_Normal.m2_pass.maf --out MBC-MBCProject_Tumor_Normal
Indel (3)	**Snowman**	**1.SnowmanWithIndel_Capture -** /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/Snowman/broadinstitute.org/cancer.genome.analysis/10982/134/snow.sh /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/Snowman/broadinstitute.org/cancer.genome.analysis/10982/134/ /seq/picard_aggregation/RP-1156/Exome/MBCProject_pt_id_T1/v2/MBCProject_pt_id_T1.bam MBC-MBCProject_Tumor_Normal 0 -n /seq/picard_aggregation/RP-1156/Exome/MBCProject_pt_id_SALIVA/v2/MBCProject_pt_id_SALIVA.bam -p 4 -D/xchip/gistic/Jeremiah/Projects/SnowmanFilters/dbsnp_138.b37_indel.vcf -B/xchip/gistic/Jeremiah/Projects/SnowmanFilters/HengLiMask/snowman_blacklist.bed -Y/xchip/gistic/Jeremiah/Projects/SnowmanFilters/viral.1.1.genomic_ns.fna **2.Oncotate Snowman Indel for Capture -** sh /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/OncotateVCFGZ/broadinstitute.org/cancer.genome.analysis/10977/4/oncotator.sh VCF TCGAMAF /fh/subscription-XRC/Snowman/MBC-MBCProject_Tumor_Normal/31865571/MBC-MBCProject_Tumor_Normal.snowman.somatic.indel.vcf MBC-MBCProject_Tumor_Normal.indel.capture.maf.annotated hg19 /xchip/cga/reference/annotation/db/oncotator_v1_ds /xchip/cga/reference/annotation/db/tcgaMAFManualOverrides2.4.config /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/OncotateVCFGZ/broadinstitute.org/cancer.genome.analysis/10977/4/CANONICAL /xchip/tcga/Tools/python_fh/python_fh_env_Dec122014/ --log_name oncotator_firehose.log --prepend --infer-onps -c /xchip/cga/reference/annotation/db/tx_exact_uniprot_matches.txt **3.Oncotate Snowman Indel for Capture unfiltered -** sh /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/OncotateVCFGZ/broadinstitute.org/cancer.genome.analysis/10977/4/oncotator.sh VCF TCGAMAF /fh/subscription-XRC/Snowman/MBC-MBCProject_Tumor_Normal/31865571/MBC-MBCProject_Tumor_Normal.snowman.unfiltered.somatic.indel.vcf MBC-MBCProject_Tumor_Normal.indel.capture.unfiltered.maf.annotated hg19 /xchip/cga/reference/annotation/db/oncotator_v1_ds /xchip/cga/reference/annotation/db/tcgaMAFManualOverrides2.4.config /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/OncotateVCFGZ/broadinstitute.org/cancer.genome.analysis/10977/4/CANONICAL /xchip/tcga/Tools/python_fh/python_fh_env_Dec122014/ --log_name oncotator_firehose.log --prepend --infer-onps -c /xchip/cga/reference/annotation/db/tx_exact_uniprot_matches.txt **4.Snowman Annotate Capture -** sh /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/SnowmanAnnotate/broadinstitute.org/cancer.genome.analysis/11109/25

		/run.sh /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/SnowmanAnnotate/broadinstitute.org/cancer.genome.analysis/11109/25/snow-annotate.R -g TRUE -W 5 -H 5 -s 0 -i /fh/subscription-XRC/Snowman/MBC-MBCProject_Tumor_Normal/31865571/MBC-MBCProject_Tumor_Normal.snowman.somatic.sv.vcf -o MBC-MBCProject_Tumor_Normal -d 8 -t 4 **5.Aggregate Individual MAF Indel Snowman -** python /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/tsvCat/broadinstitute.org/cancer.genome.analysis/10083/10//tsvConcatListFile.py /fh/subscription-XRC/tsvCat/PR_Wagle_MBC_Pairs_Capture_All_Pairs/32757023/inputListFile.input.tsv PR_Wagle_MBC_Pairs_Capture_All_Pairs.indel.snowman.maf.annotated **6.Aggregate Individual MAF Indel Snowman unfiltered -** python /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/tsvCat/broadinstitute.org/cancer.genome.analysis/10083/10//tsvConcatListFile.py /fh/subscription-XRC/tsvCat/PR_Wagle_MBC_Pairs_Capture_All_Pairs/32757022/inputListFile.input.tsv PR_Wagle_MBC_Pairs_Capture_All_Pairs.indel.snowman.unfiltered.maf.annotated

ReCapSeq	**ReCapSeg_workflow**	1.**Picard Target Mapper -** python /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/PicardTargetMapper/broadinstitute.org/cancer.genome.analysis/11037/3//picard_target_mapper.py MBC-MBCProject_Normal_Sample /xchip/cga/reference/seq_target_to_annotations.v4.json /seq/picard_aggregation/RP-1156/Exome/MBCProject_pt_id_SALIVA/v2/MBCProject_pt_id_SALIVA.bam output.txt 2.**ReCapSeg Coverage** - sh /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/ReCapSegCoverage/broadinstitute.org/cancer.genome.analysis/10608/20/wrapper.sh /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/ReCapSegCoverage/broadinstitute.org/cancer.genome.analysis/10608/20/ --jar /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/ReCapSegCoverage/broadinstitute.org/cancer.genome.analysis/10608/20//coverage.jar --sample_id MBC-MBCProject_Normal_Sample --bam /seq/picard_aggregation/RP-1156/Exome/MBCProject_pt_id_SALIVA/v2/MBCProject_pt_id_SALIVA.bam --bed /xchip/cga/reference/hg19/recapseg/target_beds/CRSP_ICE_hg19_wex_illumina_v1.no_X_Y_MT.bed --reference /seq/references/Homo_sapiens_assembly19/v1/Homo_sapiens_assembly19.fasta 3.**ReCapSeg Read Group Collapse and Proportional Coverage** - sh /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/recapseg_pc_collapse/broadinstitute.org/cancer.genome.analysis/11060/1/recapseg_pc_collapse.sh /fh/subscription-XRC/ReCapSegCoverage/MBC-MBCProject_Normal_Sample/27010468/MBC-MBCProject_Normal_Sample.coverage /xchip/cga/reference/hg19/recapseg/target_beds/CRSP_ICE_hg19_wex_illumina_v1.no_X_Y_MT.bed /seq/picard_aggregation/RP-1156/Exome/MBCProject_pt_id_SALIVA/v2/MBCProject_pt_id_SALIVA.bam MBC-MBCProject_Normal_Sample /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/recapseg_pc_collapse/broadinstitute.org/cancer.genome.analysis/11060/1/ /xchip/tcga/Tools/recapseg/releases/recapseg_env_FH_May182015/ 1.0 4.**ReCapSeg Filter TN Segment** - sh /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/recapseg_tumor_pcov/broadinstitute.org/cancer.genome.analysis/10665/34/recapseg_tumor_pcov.sh /xchip/cga/reference/hg19/recapseg/pon/ICE_Combined_EEW_EWS_v1.pon /fh/subscription-XRC/recapseg_pc_collapse/MBC-MBCProject_Normal_Sample/27010964/MBC-MBCProject_Normal_Sample.pcov /fh/subscription-XRC/recapseg_pc_collapse/MBC-MBCProject_Normal_Sample/27010964/MBC-MBCProject_Normal_Sample.pcov.cr.stat MBC-MBCProject_Normal_Sample recapseg /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/recapseg_tumor_pcov/broadinstitute.org/cancer.genome.analysis/10665/34/ /xchip/tcga/Tools/recapseg/releases/recapseg_env_FH_May182015/ /xchip/cga/reference/hg19/recapseg/target_beds/CRSP_ICE_hg19_wex_illumina_v1.no_X_Y_MT.bed --is-plotting 5.**ReCapSeg Caller** - sh /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/recapseg_caller/broadinstitute.org/cancer.genome.analysis/10797/10/recapseg_caller.sh /fh/subscription-XRC/recapseg_tumor_pcov/MBC-MBCProject_Normal_Sample/27010966/recapseg.tn.MBC-MBCProject_Normal_Sample.tsv /fh/subscription-XRC/recapseg_tumor_pcov/MBC-MBCProject_Normal_Sample/27010966/recapseg.MBC-MBCProject_Normal_Sample.seg recapseg.MBC-MBCProject_Normal_Sample.called /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/recapseg_caller/broadinstitute.org/cancer.genome.analysis/10797/10/ /xchip/tcga/Tools/recapseg/releases/recapseg_env_FH_May182015/ 6.**ReCapSeg Oncotate SCNA Seg File** - sh /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/Oncotator_v1/broadinstitute.org/cancer.genome.analysis/10202/60/oncotator.sh SEG_FILE GENE_LIST /fh/subscription-XRC/recapseg_caller/MBC-MBCProject_Normal_Sample/27010973/recapseg.MBC-MBCProject_Normal_Sample.called MBC-MBCProject_Normal_Sample.called.seg.annotated hg19 /xchip/cga/reference/annotation/db/oncotator_v1_ds_June112014/ /xchip/cga/reference/annotation/db/tcgaMAFManualOverrides2.4.config /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/Oncotator_v1/broadinstitute.org/cancer.genome.analysis/10202/60/CANONICAL /xchip/tcga/Tools/python_fh/python_fh_env_June252014/ --log_name oncotator_firehose_seg.log
Gistic2	**Gistic2_From_ReCapSeg_Workflow**	**1.Aggregate_ReCapSeg_Segs -** python /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/tsvCat/broadinstitute.org/cancer.genome.analysis/10083/10//tsvConcatListFile.py /fh/subscription-XRC/tsvCat/PR_Wagle_MBC_Pairs_Capture_All_Pairs/33645478/inputListFile.input.tsv PR_Wagle_MBC_Pairs_Capture_All_Pairs.aggregated_case_sample.seg **2.Gistic2_Analysis_ReCapSeg -** sh /xchip/tcga/gdac_prod/source/analysis_pipeline/scripts/choose_run_matlab2.sh -v .matlab-2010a /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/Gistic2_Analysis/broadinstitute.org/cancer.genome.analysis/00264/129/ /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/Gistic2_Analysis/broadinstitute.org/cancer.genome.analysis/00264/129/call_gistic2 . 4 /fh/subscription-XRC/tsvCat/PR_Wagle_MBC_Pairs_Capture_All_Pairs/33645478/PR_Wagle_MBC_Pairs_Capture_All_Pairs.aggregated_case_sample.seg /xchip/gistic/variables/WES_CRSP_ICE_hg19_wex_illumina_v1_naX_Y_MT1.markers.txt /xchip/gistic/variables/hg19/hg19_with_miR_20120227.mat /xchip/gistic/CNV/blood_normals/CNV.hg19_111204/CNV.hg19.bypos.111213.txt 0.3 0.3 2 0.5 1 0.99 10 1 10000 1 mean /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/Gistic2_Analysis/broadinstitute.org/cancer.genome.analysis/00264/129/version.txt **3.Gistic2_Report_ReCapSeg -** java -Dr_flags=--vanilla -cp /xchip/tcga/gdac_prod/applications/genepattern/RunR/ RunR /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/GDAC_Gistic2Report/broadinstitute.org/cancer.genome.analysis/00269/63/Gistic2Report.R main -L/xchip/tcga/Tools/Nozzle/v1.current -g/xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/GDAC_Gistic2Report/broadinstitute.org/cancer.genome.analysis/00269/63/geneList.txt -a/fh/subscription-XRC/Gistic2_Analysis/PR_Wagle_MBC_Pairs_Capture_All_Pairs/33646697/amp_genes.conf_99.txt -d/fh/subscription-XRC/Gistic2_Analysis/PR_Wagle_MBC_Pairs_Capture_All_Pairs/33646697/del_genes.conf_99.txt -A/fh/subscription-XRC/Gistic2_Analysis/PR_Wagle_MBC_Pairs_Capture_All_Pairs/33646697/amp_qplot.png -D/fh/subscription-XRC/Gistic2_Analysis/PR_Wagle_MBC_Pairs_Capture_All_Pairs/33646697/del_qplot.png -S/fh/subscription-XRC/Gistic2_Analysis/PR_Wagle_MBC_Pairs_Capture_All_Pairs/33646697/raw_copy_number.png -V/fh/subscription-XRC/Gistic2_Analysis/PR_Wagle_MBC_Pairs_Capture_All_Pairs/33646697/gisticVersion.txt -T/fh/subscription-

		XRC/Gistic2_Analysis/PR_Wagle_MBC_Pairs_Capture_All_Pairs/33646697/arraylistfile.txt -b/fh/subscription-XRC/Gistic2_Analysis/PR_Wagle_MBC_Pairs_Capture_All_Pairs/33646697/broad_significance_results.txt -I/fh/subscription-XRC/Gistic2_Analysis/PR_Wagle_MBC_Pairs_Capture_All_Pairs/33646697/gisticInputs.txt
CoMut	**Mutation_Significance_2CV**	1.**MutSigRun2CV-** /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/MutSig_2CV/broadinstitute.org/cancer.genome.analysis/10594/36/choose_dotkit Python-2.7 /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/MutSig_2CV/broadinstitute.org/cancer.genome.analysis/10594/36/MutSig_Wrapper.sh MutSig_2CV_standalone_v3_1 /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/MutSig_2CV/broadinstitute.org/cancer.genome.analysis/10594/36/Merge_MAFs.py --maf1 /fh/subscription-XRC/MutSig_2CV/PR_Wagle_MBC_Pairs_Capture_All_Pairs/33639571/maf1.input.tsv --maf2 /fh/subscription-XRC/MutSig_2CV/PR_Wagle_MBC_Pairs_Capture_All_Pairs/33639571/maf2.input.tsv PR_Wagle_MBC_Pairs_Capture_All_Pairs sh /xchip/tcga/gdac_prod/source/analysis_pipeline/scripts/choose_run_matlab2.sh -v .matlab-2013a /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/MutSig_2CV/broadinstitute.org/cancer.genome.analysis/10594/36/ /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/MutSig_2CV/broadinstitute.org/cancer.genome.analysis/10594/36/MutSig_2CV_standalone_v3_1 . /xchip/cga/reference/mutsig_params/pancan_mutation_blacklist.v14.hg19.add.ERpos.CCPM.final.txt /xchip/cga/reference/mutsig_params/mutation_type_dictionary.v4.txt /xchip/cga/reference/mutsig_params/coverage_models.v5a.mat /xchip/cga/reference/mutsig_params/coverage_basewise.fwb /xchip/cga/reference/mutsig_params/target_list.hg19.v1a.txt /xchip/cga/reference/mutsig_params/context_and_effect.c65e29b.fwb /xchip/cga/reference/mutsig_params/context_and_effect.c65e29b.txt /xchip/cga/reference/mutsig_params/covariates_transformed.v5a.txt /xchip/cga/reference/mutsig_params/conservation46.fwb /xchip/cga/reference/mutsig_params/FixedWidthBinary.jar hg19 /xchip/cga/reference/mutsig_params/MutSig2CV.hg19.params.txt 2. **GenerateStickFiguresMutSig2CV** - /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/GenerateStickFigures/broadinstitute.org/cancer.genome.analysis/00497/16/choose_dotkit Python-2.7 /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/GenerateStickFigures/broadinstitute.org/cancer.genome.analysis/00497/16/generate_figs.py -s 0.99 -e True /fh/subscription-XRC/MutSig_2CV/PR_Wagle_MBC_Pairs_Capture_All_Pairs/33639571/PR_Wagle_MBC_Pairs_Capture_All_Pairs.final_analysis_set.maf /fh/subscription-XRC/MutSig_2CV/PR_Wagle_MBC_Pairs_Capture_All_Pairs/33639571/sig_genes.txt /xchip/cga/reference/annotation/db/uniprot/Uniprot_SWISS_human_Release_2010_12.shelf sh /xchip/tcga/gdac_prod/source/analysis_pipeline/scripts/choose_run_matlab2.sh -v .matlab-2012a /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/GenerateStickFigures/broadinstitute.org/cancer.genome.analysis/00497/16/ /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/GenerateStickFigures/broadinstitute.org/cancer.genome.analysis/00497/16/mutfig 3. **CoMut2CV** - /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/CoMut/broadinstitute.org/cancer.genome.analysis/00488/17/choose_dotkit R-2.15 Rscript /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/CoMut/broadinstitute.org/cancer.genome.analysis/00488/17/coMut.R -v -o . --firehose.mode --png --sort.by.mutation.status -a PR_Wagle_MBC_Pairs_Capture_All_Pairs -s /fh/subscription-XRC/MutSig_2CV/PR_Wagle_MBC_Pairs_Capture_All_Pairs/33639571/sig_genes.txt -c /fh/subscription-XRC/MutSig_2CV/PR_Wagle_MBC_Pairs_Capture_All_Pairs/33639571/patient_counts_and_rates.txt -m /fh/subscription-XRC/MutSig_2CV/PR_Wagle_MBC_Pairs_Capture_All_Pairs/33639571/PR_Wagle_MBC_Pairs_Capture_All_Pairs.final_analysis_set.maf --firehose.mutsig.mutcategs /fh/subscription-XRC/MutSig_2CV/PR_Wagle_MBC_Pairs_Capture_All_Pairs/33639571/mutcategs.txt -q 0.75 4.**MutSigNozzleReport2CV** - /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/GDAC_MutSig_Report/broadinstitute.org/cancer.genome.analysis/00412/47/choose_dotkit R-2.15 Rscript --slave --vanilla /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/GDAC_MutSig_Report/broadinstitute.org/cancer.genome.analysis/00412/47/Mutsig_Report.R -k /fh/subscription-XRC/MutSig_2CV/PR_Wagle_MBC_Pairs_Capture_All_Pairs/33639571/patient_counts_and_rates.txt -l /fh/subscription-XRC/CoMut/PR_Wagle_MBC_Pairs_Capture_All_Pairs/33640421/PR_Wagle_MBC_Pairs_Capture_All_Pairs_coMut.png -m /fh/subscription-XRC/CoMut/PR_Wagle_MBC_Pairs_Capture_All_Pairs/33640421/PR_Wagle_MBC_Pairs_Capture_All_Pairs_coMut.pdf -n /fh/subscription-XRC/GenerateStickFigures/PR_Wagle_MBC_Pairs_Capture_All_Pairs/33640314/generatedStickFigurePaths.txt -o /fh/subscription-XRC/MutSig_2CV/PR_Wagle_MBC_Pairs_Capture_All_Pairs/33639571/PR_Wagle_MBC_Pairs_Capture_All_Pairs.final_analysis_set.maf /xchip/tcga/Tools/Nozzle/v1.current PR_Wagle_MBC_Pairs_Capture_All_Pairs /fh/subscription-XRC/MutSig_2CV/PR_Wagle_MBC_Pairs_Capture_All_Pairs/33639571/sig_genes.txt /fh/subscription-XRC/MutSig_2CV/PR_Wagle_MBC_Pairs_Capture_All_Pairs/33639571/MutSig_version.txt
Germline SNV	**Germline Pipeline Simplified**	**1.HaplotypeCallerSingleSampleGVCF for Pairs on Capture -** sh /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/HaplotypeCallerSingleSampleGVCF/broadinstitute.org/cancer.genome.analysis/10614/9/script.sh /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/HaplotypeCallerSingleSampleGVCF/broadinstitute.org/cancer.genome.analysis/10614/9/ /seq/picard_aggregation/RP-1156/Exome/MBCProject_pt_id_SALIVA/v2/MBCProject_pt_id_SALIVA.bam /fh/subscription-XRC/HaplotypeCallerSingleSampleGVCF/MBC-MBCProject_Tumor_Normal/27003021/tmp MBC-MBCProject_Normal_Sample 10 /xchip/cga/reference/hg19/whole_exome_agilent_1.1_refseq_plus_3_boosters_plus_10bp_padding_minus_mito.Homo_sapiens_assembly19.targets.interval_list /xchip/cga/reference/hg19/dbsnp_134_b37.leftAligned.vcf /seq/references/Homo_sapiens_assembly19/v1/Homo_sapiens_assembly19.fasta **2.Combine and Genotype GVCFs -** sh /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/CombineGenotypeGVCFs/broadinstitute.org/cancer.genome.analysis/10674/1/script.sh /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/CombineGenotypeGVCFs/broadinstitute.org/cancer.genome.analysis/10674/1/ /fh/subscription-XRC/CombineGenotypeGVCFs/PR_Wagle_MBC_Pairs_Capture_All_Pairs/33627528/gvcfs_list.input.tsv 300 PR_Wagle_MBC_Pairs_Capture_All_Pairs 20 /seq/references/Homo_sapiens_assembly19/v1/Homo_sapiens_assembly19.fasta TRUE **3.VariantRecalibrator for Germline SNPs-** sh /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/GATKVariantRecalibrator/broadinstitute.org/cancer.genome.analysis/10677/3/script.sh /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/GATKVariantRecalibrator/broadinstitute.org/cancer.genome.analysis/10677/3/ /seq/references/Homo_sapiens_assembly19/v1/Homo_sapiens_assembly19.fasta /fh/subscription-XRC/GATKVariantRecalibrator/PR_Wagle_MBC_Pairs_Capture_All_Pairs/33638806/tmp /fh/subscription-XRC/CombineGenotypeGVCFs/PR_Wagle_MBC_Pairs_Capture_All_Pairs/33627528/PR_Wagle_MBC_Pairs_Capture_All_Pairs.FinalMergedGVCF.gvcf.gz /xchip/cga/reference/hg19/whole_exome_agilent_1.1_refseq_plus_3_boosters_plus_10bp_padding_minus_mito.Homo_sapiens_assembly19.targets.interval_list PR_Wagle_MBC_Pairs_Capture_All_Pairs SNP **4.ApplyRecalibration for Germline SNPs -** sh /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/GATKApplyRecalibration/broadinstitute.org/cancer.genome.analysis/10678/1/script.sh /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/GATKApplyRecalibration/broadinstitute.org/cancer.genome.analysis/1

		0678/1/ /seq/references/Homo_sapiens_assembly19/v1/Homo_sapiens_assembly19.fasta /fh/subscription-XRC/GATKApplyRecalibration/PR_Wagle_MBC_Pairs_Capture_All_Pairs/33639082/tmp /fh/subscription-XRC/CombineGenotypeGVCFs/PR_Wagle_MBC_Pairs_Capture_All_Pairs/33627528/PR_Wagle_MBC_Pairs_Capture_All_Pairs.FinalMergedGVCF.gvcf.gz 99.0 /fh/subscription-XRC/GATKVariantRecalibrator/PR_Wagle_MBC_Pairs_Capture_All_Pairs/33638806/PR_Wagle_MBC_Pairs_Capture_All_Pairs.SNP.tranches /fh/subscription-XRC/GATKVariantRecalibrator/PR_Wagle_MBC_Pairs_Capture_All_Pairs/33638806/PR_Wagle_MBC_Pairs_Capture_All_Pairs.SNP.recal PR_Wagle_MBC_Pairs_Capture_All_Pairs SNP **5.VariantRecalibrator for Germline Indels -** sh /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/GATKVariantRecalibrator/broadinstitute.org/cancer.genome.analysis/10677/4/script.sh /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/GATKVariantRecalibrator/broadinstitute.org/cancer.genome.analysis/10677/4/ /seq/references/Homo_sapiens_assembly19/v1/Homo_sapiens_assembly19.fasta /fh/subscription-XRC/GATKVariantRecalibrator/PR_Wagle_MBC_Pairs_Capture_All_Pairs/33639233/tmp /fh/subscription-XRC/GATKApplyRecalibration/PR_Wagle_MBC_Pairs_Capture_All_Pairs/33639082/PR_Wagle_MBC_Pairs_Capture_All_Pairs.recalibrated.vcf /xchip/cga/reference/hg19/whole_exome_agilent_1.1_refseq_plus_3_boosters_plus_10bp_padding_minus_mito.Homo_sapiens_assembly19.targets.interval_list PR_Wagle_MBC_Pairs_Capture_All_Pairs INDEL 6.ApplyRecalibration for Germline Indels - sh /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/GATKApplyRecalibration/broadinstitute.org/cancer.genome.analysis/10678/1/script.sh /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/GATKApplyRecalibration/broadinstitute.org/cancer.genome.analysis/10678/1/ /seq/references/Homo_sapiens_assembly19/v1/Homo_sapiens_assembly19.fasta /fh/subscription-XRC/GATKApplyRecalibration/PR_Wagle_MBC_Pairs_Capture_All_Pairs/33639324/tmp /fh/subscription-XRC/GATKApplyRecalibration/PR_Wagle_MBC_Pairs_Capture_All_Pairs/33639082/PR_Wagle_MBC_Pairs_Capture_All_Pairs.recalibrated.vcf 99.0 /fh/subscription-XRC/GATKVariantRecalibrator/PR_Wagle_MBC_Pairs_Capture_All_Pairs/33639233/PR_Wagle_MBC_Pairs_Capture_All_Pairs.INDEL.tranches /fh/subscription-XRC/GATKVariantRecalibrator/PR_Wagle_MBC_Pairs_Capture_All_Pairs/33639233/PR_Wagle_MBC_Pairs_Capture_All_Pairs.INDEL.recal PR_Wagle_MBC_Pairs_Capture_All_Pairs INDEL **7.Subset Filtered VCF to Original Set** sh /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/GermlinePipelineSubsetVCF/broadinstitute.org/cancer.genome.analysis/10683/5/script.sh /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/GermlinePipelineSubsetVCF/broadinstitute.org/cancer.genome.analysis/10683/5/ /seq/references/Homo_sapiens_assembly19/v1/Homo_sapiens_assembly19.fasta /fh/subscription-XRC/GermlinePipelineSubsetVCF/PR_Wagle_MBC_Pairs_Capture_All_Pairs/33639325/squid_id_list.input.tsv /fh/subscription-XRC/GATKApplyRecalibration/PR_Wagle_MBC_Pairs_Capture_All_Pairs/33639324/PR_Wagle_MBC_Pairs_Capture_All_Pairs.recalibrated.vcf PR_Wagle_MBC_Pairs_Capture_All_Pairs **8.Subset Filtered VCF to Single Sample -** sh /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/GermlinePipelineSubsetVCFSingleSample/broadinstitute.org/cancer.genome.analysis/10684/10/script.sh /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/GermlinePipelineSubsetVCFSingleSample/broadinstitute.org/cancer.genome.analysis/10684/10/ /seq/references/Homo_sapiens_assembly19/v1/Homo_sapiens_assembly19.fasta /fh/subscription-XRC/GermlinePipelineSubsetVCFSingleSample/PR_Wagle_MBC_Pairs_Capture_All_Pairs/33639326/list_of_squid_ids.input.tsv /fh/subscription-XRC/GermlinePipelineSubsetVCF/PR_Wagle_MBC_Pairs_Capture_All_Pairs/33639325/PR_Wagle_MBC_Pairs_Capture_All_Pairs.vcf 9.Single Sample VCFtoMAF - sh /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/VCFtoMAF/broadinstitute.org/cancer.genome.analysis/10649/3/vcf_to_maf.sh /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/VCFtoMAF/broadinstitute.org/cancer.genome.analysis/10649/3/ /fh/subscription-XRC/GermlinePipelineSubsetVCFSingleSample/PR_Wagle_MBC_Pairs_Capture_All_Pairs/33639326/MBCProject_pt_id_SALIVA.GATK_PrimitiveAlleles.vcf MBC-MBCProject_Normal_Sample TRUE
Allelic CapSeg	**AllelicCapseg**	1.**VCFtoHetIntervals** - sh /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/VCFtoHetIntervals/broadinstitute.org/cancer.genome.analysis/10733/5/VCFtoHetIntervals.sh /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/VCFtoHetIntervals/broadinstitute.org/cancer.genome.analysis/10733/5/ /fh/subscription-XRC/GermlinePipelineSubsetVCFSingleSample/PR_Wagle_MBC_Pairs_Capture_All_Pairs/33639326/MBCProject_pt_id_SALIVA.GATK_PrimitiveAlleles.vcf 2.**HetSitePullDown** - sh /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/CallSomaticMutations/broadinstitute.org/cancer.genome.analysis/00004/131/runBroadJava7.sh java -Xmx2g -jar /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/CallSomaticMutations/broadinstitute.org/cancer.genome.analysis/00004/131/muTect-1.1.6.jar --analysis_type MuTect --intervals /fh/subscription-XRC/CallSomaticMutations/MBC-MBCProject_Tumor_Normal/33640808/iteration1/./gatk-scatter.0000000015.interval_list --normal_sample_name MBC-MBCProject_Normal_Sample -I:normal /seq/picard_aggregation/RP-1156/Exome/MBCProject_pt_id_SALIVA/v2/MBCProject_pt_id_SALIVA.bam --tumor_sample_name MBC-MBCProject_Tumor_Sample-I:tumor /seq/picard_aggregation/RP-1156/Exome/MBCProject_pt_id_T1/v2/MBCProject_pt_id_T1.bam --reference_sequence /seq/references/Homo_sapiens_assembly19/v1/Homo_sapiens_assembly19.fasta --dbsnp /xchip/cga/reference/hg19/dbsnp_134_b37.leftAligned.vcf --cosmic /xchip/cga/reference/hg19/hg19_cosmic_v54_120711.vcf --normal_panel /xchip/cga/reference/hg19/refseq_exome_10bp_hg19_300_1kg_normal_panel.vcf --out MBC-MBCProject_Tumor_Normal.call_stats.txt --coverage_file MBC-MBCProject_Tumor_Normal.coverage.wig.txt --power_file MBC-MBCProject_Tumor_Normal.power.wig.txt --downsample_to_coverage 100000 --enable_extended_output --fraction_contamination 0.007 --force_output 3. **HetPullDownPostProcess** - sh /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/HetPullDownPostprocess/broadinstitute.org/cancer.genome.analysis/10621/10//CallStatsToCov.sh /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/HetPullDownPostprocess/broadinstitute.org/cancer.genome.analysis/10621/10/ /fh/subscription-XRC/CallSomaticMutations/MBC-MBCProject_Tumor_Normal/33640808/iteration1/MBC-MBCProject_Tumor_Normal.call_stats.txt 4. **AllelicCapseg** - /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/AllelicCapseg/broadinstitute.org/cancer.genome.analysis/10622/38/choose_dotkit R-2.15 Rscript /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/AllelicCapseg/broadinstitute.org/cancer.genome.analysis/10622/38/All

		elicCapseg_cli.R --SID=MBC-MBCProject_Tumor_Sample --capseg.probe.fn=/fh/subscription-XRC/recapseg_tumor_pcov/MBC-MBCProject_Tumor_Sample/27011228/recapseg.tn.MBC-MBCProject_Tumor_Sample.tsv --capseg.seg.fn=/fh/subscription-XRC/recapseg_tumor_pcov/MBC-MBCProject_Tumor_Sample/27011228/recapseg.MBC-MBCProject_Tumor_Sample.seg --germline.het.fn=/fh/subscription-XRC/HetPullDownPostprocess/MBC-MBCProject_Tumor_Normal/33641000/Tumor.cov --drop.x=FALSE --drop.y=TRUE --seg.merge.thresh=0.5 --min.seg.size=3 --verbose=TRUE --base.output.dir=. --initial.merge=TRUE --split.merge=TRUE --outlier.thresh=0.005 --working.dir=/xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/AllelicCapseg/broadinstitute.org/cancer.genome.analysis/10622/38/
ABSOLUTE	**ABSOLUTE _v1.5_WES**	/broad/software/free/Linux/redhat_5_x86_64/pkgs/r_2.15.3/bin/Rscript /xchip/tcga/gdac_prod/applications/process_mgmt/firehose_task_registry/cga/ABSOLUTE_v1.5/broadinstitute.org/cancer.genome.analysis/10971/23//ABSOLUTE_cli_start.R --seg_dat_fn /fh/subscription-XRC/AllelicCapseg/MBC-MBCProject_Tumor_Normal/33641022/results/MBC-MBCProject_Tumor_Sample.tsv --maf_fn /fh/subscription-XRC/PoN_filter_py/MBC-MBCProject_Tumor_Normal/30688317/MBC-MBCProject_Tumor_Normal.novoalign.PoNfiltered_agilent_ICE.maf --indelmaf_fn /fh/subscription-XRC/MafMasterFilter/MBC-MBCProject_Tumor_Normal/31668062/MBC-MBCProject_Tumor_Normal.pon_filtered.txt --sample_name MBC-MBCProject_Tumor_Normal --results_dir raw_results/ --ssnv_skew 0.9568517 --abs_lib_dir /xchip/tcga/Tools/absolute/releases/v1.5/