Date: October 14th, 2019

Department of Genomic Medicine
Cancer Immune Monitoring and Analysis Center
Division of Cancer Medicine
The University of Texas MD Anderson Cancer Center

Director:
Andy Futreal, BS, Ph.D., Professor, and Chair
Curtis Gumbs, Scientific Manager
Jianhua (John) Zhang, Ph.D., Director, Computational Genomics, Genomic Medicine

Whole Exome Sequencing Analytical Validation

Version 4.0

This report describes the analytical validation parameters for Whole Exome Sequencing assay performed at MD Anderson Cancer Center CIMAC.

Whole Exome Sequencing	
Analytical Performance	DNA isolated from tumors (FFPE and FF) and PBMC with A260/280 ratio 1.8-2.0 was used for library preparation for WES. Mean target coverage for FFPE and FF tissues were 108.55 and 133.75 respectively. Consistent coverage was observerd between samples replicates within a sequencing reaction or between independent sequencing reactions.
Analytical Precision and Reproducibility	Mutect and Pindel were used for making Single Nucleid Variants (SNVs) and Indel calls using 50X coverage and Variant Allele Frecuency (VAF) of 0.1. Reproducibility between each sample type was evaluated comparing replicate libraries in each sequencing reaction or between separate sequencing reactions using same library sets. Reproducibility between FF and FFPE was compared by pairwise alignment and calculating the positive percent agreement (PPA), the Jaccard Index and the statistical agreement (ISA) index.
Analytical sensitivity and specificity	Sequencing data from two pools XX and YY of HapMap cell lines were used for determining the sensitivity and specificity of the assay using a benchmark data, The sensitivity and Specifcity of the assays were more than 80 in all cases.
Pilot assay	A pilot assay using 7 lung cancer cases with FF, FFPE and PBMC were sequenced as sets of two independent samples. Sequencing data analysis show high concordance between FF and FFPE samples for each tumor case in a set. Jaccard, ISA and PPA values were also calculated showing reproducibility between the sample tissue types. Pairwise alignment shows clustering between FF and FFPE samples from each tumor in a set as well as duplicate samples in the second set.
Any other performance characteristics required for assay performance	All of the required equipment have annual service contracts with regular Preventive Maintenance performed to maintain optimal calibration and performance. All other small equipment such as multi-channel pipettes and laboratory material have calibration performed by certified vendors.

The analytical validation assay was performed in three steps. In the first step the inter and intra-assay reproducibility was performed using 3 lung tumor cases with Fresh Frozen (FF), Formalin-Fixed Paraffin Embedded (FFPE) tissues and Peripheral Blood Mononucleated Cells (PBMCs). One sample showed poor tumor content and was eliminated from further processing. In the second step the specificity and the sensitivity of the assay was analyzed using HapMap cell line pools and a benchmark cell line from

the pool. Finally in the last step a pilot experiment concordance between FF and frozen tissue was analyzed with addition 7 tumor cases to extend the number of samples.

Step 1

1. **Samples:** Fresh Frozen (FF), Formalin-Fixed Paraffin Embedded (FFPE) tissues and Peripheral Blood Mononucleated Cells (PBMCs) representing two lung cancer cases (MDA-5760: Adenocarcinoma and MDA-5812, MDA-5971: Squamous Cell Carcinoma) were chosen for the first step.

   Sample Quality Control (QC): Histological and cytological examination and assessment of FF and FFPE tissue specimens were done by a reference pathologist. Tissue quality was assessed before extraction of DNA. All H&E stained histological samples used for QC were scanned and digital images are available for review.

2. **DNA extraction:** DNA from the nine samples representing the three patients were isolated by the following methods.
   (i) DNA from buffy coat (PBMC) was extracted using the QiAmp DNA Blood mini kit (Qiagen).

   (ii) DNA from FFPE tissues was extracted using the QIAmp DNA FFPE kit (Qiagen). The FFPE tissue blocks were from the years 2015/2016 and have been stored at room temperature. Slides from these blocks were freshly cut for DNA extraction.

   (iii) DNA from FF tissues was extracted using the QIAmp DNA mini kit (Qiagen).

   DNA Quality Control (QC): DNA quality was assessed by using Qubit/Pico Green assay. The 260/280 ratio was within 1.8 - 2.0 for all samples except MDA-5971-C which was borderline at 1.79 (**Table1**). Tumor content of FF sample from MDA-5812 was very low.

   The integrity of DNA after isolation was assessed using Agilent Genomic DNA Screen Tape assay (DIN score). DIN score was used to determine input modifications.

Table 1. Study Sample Information and Quality Control (QC).

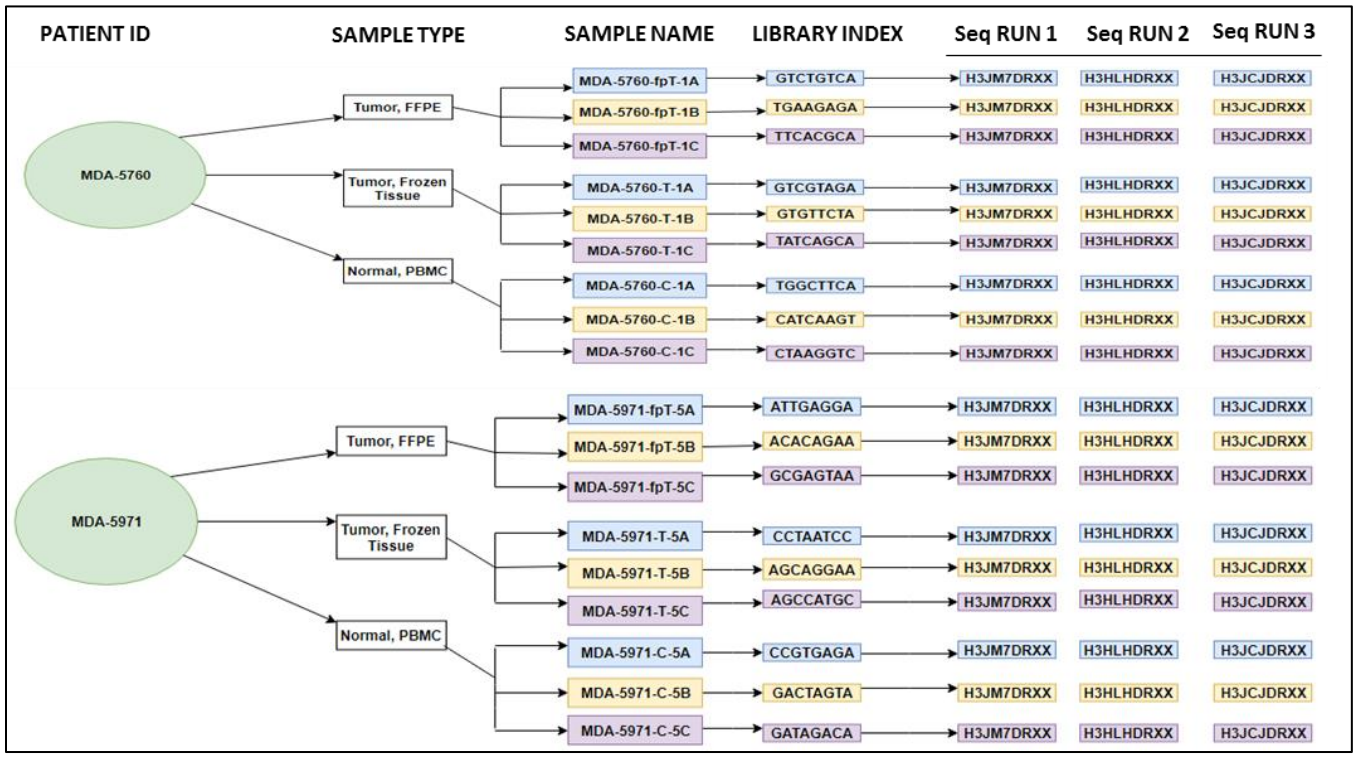
MDA Sample ID	Tissue Type	Tumor Content (%)*	Malignant Cell (%)**	DNA Conc. (ng/uL)	DNA Quality (Ab_{260}/Ab_{280} ratio)	Histology Diagnosis
MDA-5760-T	Frozen	80	80	61.4	1.86	Adenocarcinoma
MDA-5760-fpT	FFPE	90	90	263.7	1.89	Adenocarcinoma
MDA-5760-C	WBC	N/A	N/A	216	1.88	Adenocarcinoma
MDA-5812-T***	Frozen	0	0	81	1.86	Squamous Cell carcinoma
MDA-5812-fpT	FFPE	80	60	372.4	1.87	Squamous Cell carcinoma
MDA-5812-C	WBC	N/A	N/A	450	1.86	Squamous Cell carcinoma
MDA-5971-T	Frozen	90	20	81.6	1.81	Squamous Cell carcinoma
MDA-5971-fpT	FFPE	40	40	501.7	1.87	Squamous Cell carcinoma
MDA-5971-C	WBC	N/A	N/A	35.4	1.79	Squamous Cell carcinoma

***Tumor content (%): Percentage of tumor in tissue; **Malignant cell (%): Percentage of viable tumor cells; ***MDA-5812-T, a sample without tumor; N/A: not applicable*

3. **Study Design.**

Due to the low tumor content of MDA-5812, it was not included in further processing. Details of the Whole Exome Sequencing (WES) workflow are described in MDACC WES SOP. WES was performed on DNA extracted from FF, FFPE and PBMC samples corresponding to the two lung cancer cases using a 2x3x3 experimental design. DNA isolated from the nine samples (**Table 1** and **Figure 1 Column 2**) were used for library generation in triplicate (6X3 libraries). Thus, the same DNA was used to generate the triplicates. The 18 libraries generated from 6 samples were sequenced 3 times independently. Details of study samples and experimental study design are shown in **Table 1** and **Figure 1.**

Figure 1. Experimental Study Design



4. **Library Preparation:** Individual libraries were prepared using Agilent NGS/SureSelect V4 Automated Library Prep and Capture System (SureSelect Library prep kit) for the FF and PBMCs samples. Agilent SureSelect V4 XT Low Input multiplex indexes were used for the FFPE tissues. This library preparation is compatible with high-quality lower-quantity DNA from FFPE samples with a DNA input range of 10-200 ng. The NovaSeq Illumina platform was used for sequencing of the DNA libraries. The sequence length of PE76, coverage of 100X for normal samples and 200X for tumor samples. Quality of raw data conformed to NovaSeq's standards.

Library Quality Control (QC): Pre-capture library quality and quantity were assessed with the Agilent D1000 DNA analyses kit on an Agilent 4200 TapeStation instrument. Final library quality and quantity were assessed with the Agilent High Sensitivity D1000 DNA analyses kit on an Agilent 4200 TapeStation instrument and with the KAPA Biosystems library quantification kit and qPCR on Applied Biosystems Quant Studio 6 system. The QC metrics performed during the sequencing runs were

generated by the sequencers and all metrics exceed Illumina's requirements for performance.

5. Data Analyses and Bioinformatics.

(i) BCL (raw output of NovaSeq) files were processed using Illumina's CASAVA (Consensus Assessment of Sequence And Variation) tool for de-multiplexing/conversion to FASTQ format, which is the standard input for most aligners and downstream analytic tools.

(ii) The FASTQ files were aligned to the reference genome (e.g. human Hg19) using BWA (Li and Durbin 2009).

(iii) The aligned BAM files were subjected to mark duplication, re-alignment, and re-calibration using Picard and GATK (DePristo et. al. 2011) prior to downstream analyses.

(iv) The generated BAM files were subject to germline mutation calls using Platypus (Rimmer et. al. 2014), somatic mutation calls using MuTect version 1.1.7 (Gibulskis et. al. 2013), and Indel calls using Pindel version 0.2.5a8 (Ye et. al. 2009).

6. Analytical Performance.

The highest average target coverage was found in FF tumor samples when compared to FFPE tissues and germline DNA from PBMCs (133.5 vs 108.55). Each sample represents libraries prepared in triplicate from FF or FFPE or PBMC and sequenced in triplicate (3X3, in three independent sequencing reactions) from each of the three tumors (**Table 2**). High reproducibility and consistency in target coverage was observed in specimens across replicate libraries as well as across replicate sequencing reactions (**Supplemental Figure 1)**.

Table 2. Average Coverage in FFPE and FF tumors (in millions).

Sample ID	FFPE					FF				
	Run 1 (0030)	Run 2 (0031)	Run 3 (0032)	Mean	Sample Mean	Run 1 (0030)	Run 2 (0031)	Run 3 (0032)	Mean	Sample Mean
MDA-5760-1A	99.61	100.82	106.64	**102.36**	**97.20**	89.22	90.16	93.35	90.91	**120.59**
MDA- 5760-1B	92.13	93.59	98.57	**94.76**		121.40	123.19	129.98	124.86	
MDA-5760-1C	91.85	93.22	98.34	**94.47**		142.61	143.65	151.72	145.99	
MDA-5971-5A	121.08	122.68	129.94	**124.57**	**121.93**	139.06	141.05	149.15	143.09	**143.33**
MDA-5971-5B	117.55	118.83	125.88	**120.75**		146.95	149.43	157.38	151.25	
MDA-5971-5C	116.78	118.85	125.82	**120.48**		131.93	133.74	141.32	135.66	
TOTAL FFPE MEAN					**109.56**	**TOTAL FF MEAN**				**131.95**

7. Analytical Evaluation.

The coverage parameters applied were 200x for FFPE and FF tumor tissue and 100X for PBMCs. This high coverage depth allows for higher detection sensitivity of genomic sequence variations. The data were analyzed to call for both synonymous and non-synonymous single nucleotide variants (SNVs) and indels in the coding exons using PBMC as the baseline. SNVs were identified by using MuTect algorithm and indels were detected with Pindel using the following criteria: Total read counts

in tumor sample >= 50, in normal sample >= 50 and VAF =0.1. Analyses showed high reproducibility and consistency across sample triplicates in a sequencing run and also across replicate sequencing runs. This was observed in the case of both FFPE and FF samples. Sample MDA-5812-T was an exception with the absence of base calls explained by its very low malignant cell (See **Table 1**) content (**Table 3A**).

Table 3A. Coverage for Somatic Exonic SNVs in FFPE and FF.

Sample ID	FFPE (fpT)					FF (T)				
	Run 1	Run 2	Run 3	Mean	Sample Mean	Run 1	Run 2	Run 3	Mean	Sample Mean
MDA-5760-1A	133	124	149	135		156	151	153	154	
MDA-5760-1B	150	135	121	135	137	193	185	197	191	182
MDA-5760-1C	151	121	148	140		198	195	211	201	
MDA-5971-5A	135	135	139	136		184	187	194	188	
MDA-5971-5B	131	128	136	131	135	196	190	202	196	188
MDA-5971-5C	134	134	141	136		179	178	187	181	
TOTAL FFPE MEAN					136	TOTAL FF MEAN				185

Table 3B. Concordance rates of SNV calls between replicates within a run or between runs (Percentage and Standard Error)

Sample ID	5760				5971			
Tissue type	FFPE (fpT)		FF (T)		FFPE (fpT)		FF (T)	
Replicates	50%	5.62	69%	6.14	82%	1.78	90%	0.37
Runs	35%	4.5	56%	1.77	77%	1.91	88%	1.85
Mean Concordance	43%	7.23	63%	6.15	80%	2.42	89%	1.12

8. **Inter-assay and intra-assay Reproducibility.** To test for the precision of the library preparation and reproducibility of the independent sequencing runs, concordance was evaluated between sample replicates within and across sequencing runs. For this the Jaccard values were obtained through pairwise alignment of each sample. Then the following method was used to calculate the mean **(Table 3B)**. For comparison between replicates, the mean Jaccard between each pair of samples (eg. MDA-5760-1A-fpT vs MDA-5760-1B fpT, MDA-5760-1A fpT vs MDA-5760-1C fpT and MDA-5760-1C fpT vs MDA-5760-1B fpT) within a sequencing run was calculated **(Supplemental Figure 3A and 3B)**. All values from each sequencing run was then used for calculating the final mean for 5760-FFPE replicates. This was done for the rest in a similar manner. For comparison between triplicates in different sequencing runs average of the Jaccard values between of samples across different sequencing runs were evaluated (e.g. MDA-5760-1A-run1 vs MDA-5760-1A-run2, MDA-5760-1A-run1 vs MDA-5760-1A-run3 and MDA-5760-1A-run2 vs MDA-5760-1A-run3) **(Supplemental Figure 3A and 3B)**. All values from each replicate across sequencing runs were then used for calculating the final mean for 5760-FFPE. This was done for the rest in a similar manner. For all SNV calls, the concordance rate per replicated sample ranged from 43% to 89.0% **(Table 3B)**. The mean rate of concordance, for mutations, between runs ranged from 35% to 77 % in the FFPE sample and from

56 to 88% in the FF (excluding sample 5812 that didn't' contain tumor cells). The high range of concordance rate observed in the FFPE samples is due to inherent pre-analytical condition variables of the FFPE tissues like fixation and other parameters. This analysis gives the reproducibility of the assay between the replicate samples of each kind.

9. **Concordance between FFPE and FF mutation calls.** The Positive Percent Agreement (PPA), a metric similar to sensitivity was determined by pairwise alignment of each sample (3 libraries from same DNA, sequenced 3 times for FFPE and FF from each of the 3 lung tumors). Tumors from MDA-5971 showed higher concordance between FFPE and FF tumors (Jaccard ~ 0.6) compared to MDA-5760 (Jaccard ~ 0.2) while MDA-8612 had no overlapping mutation calls due to lack of tumor cells in frozen samples. PPA values with reference to FF was aligned with values obtained for Jaccard. There was very little spread in the values obtained for Jaccard, PPA_{refFF} and ISA for a particular tumor sample **(Figure 2A, 2B and Table 4)**.

Figure 2A. Positive percent agreement (PPA) by pairwise alignment of samples.

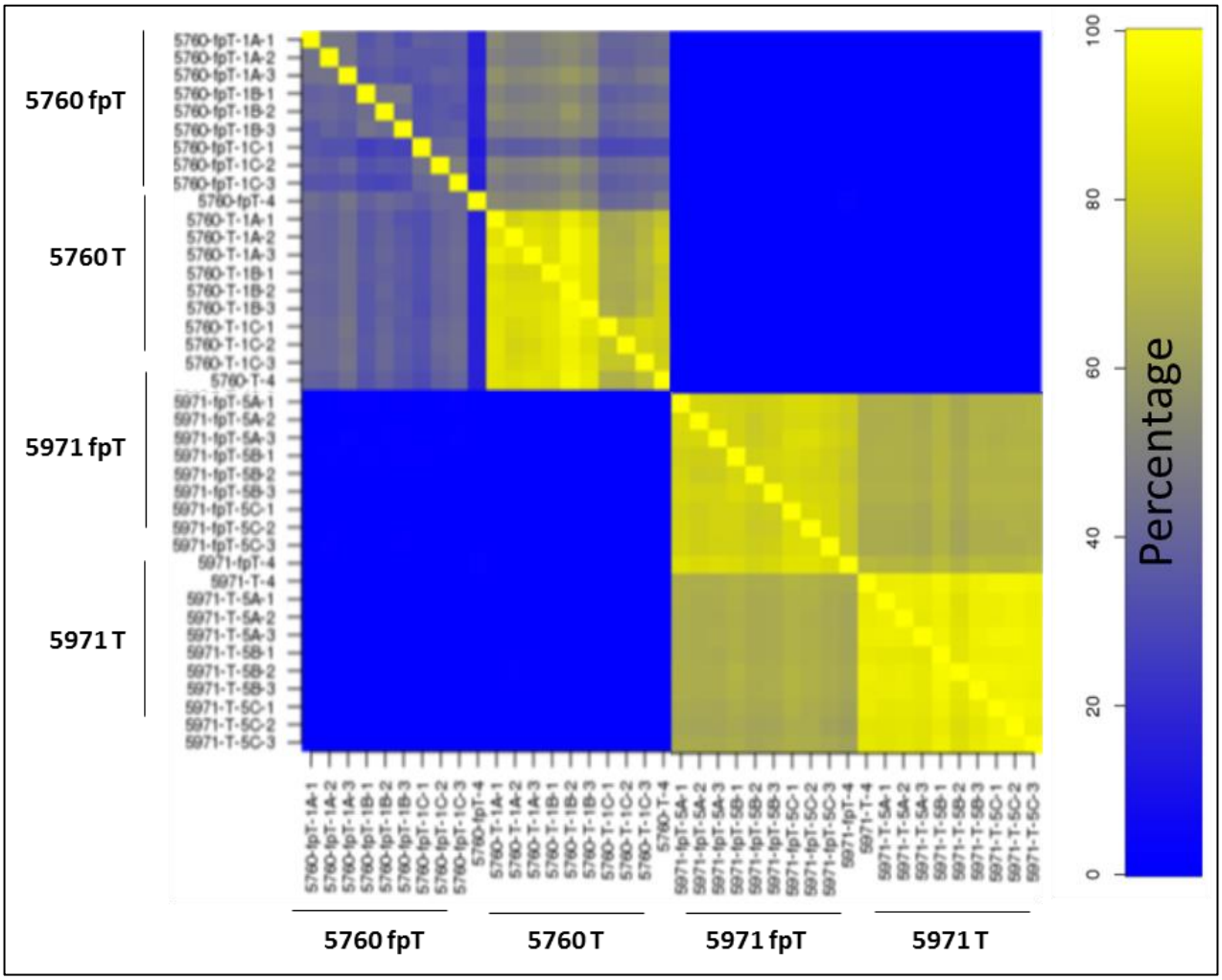
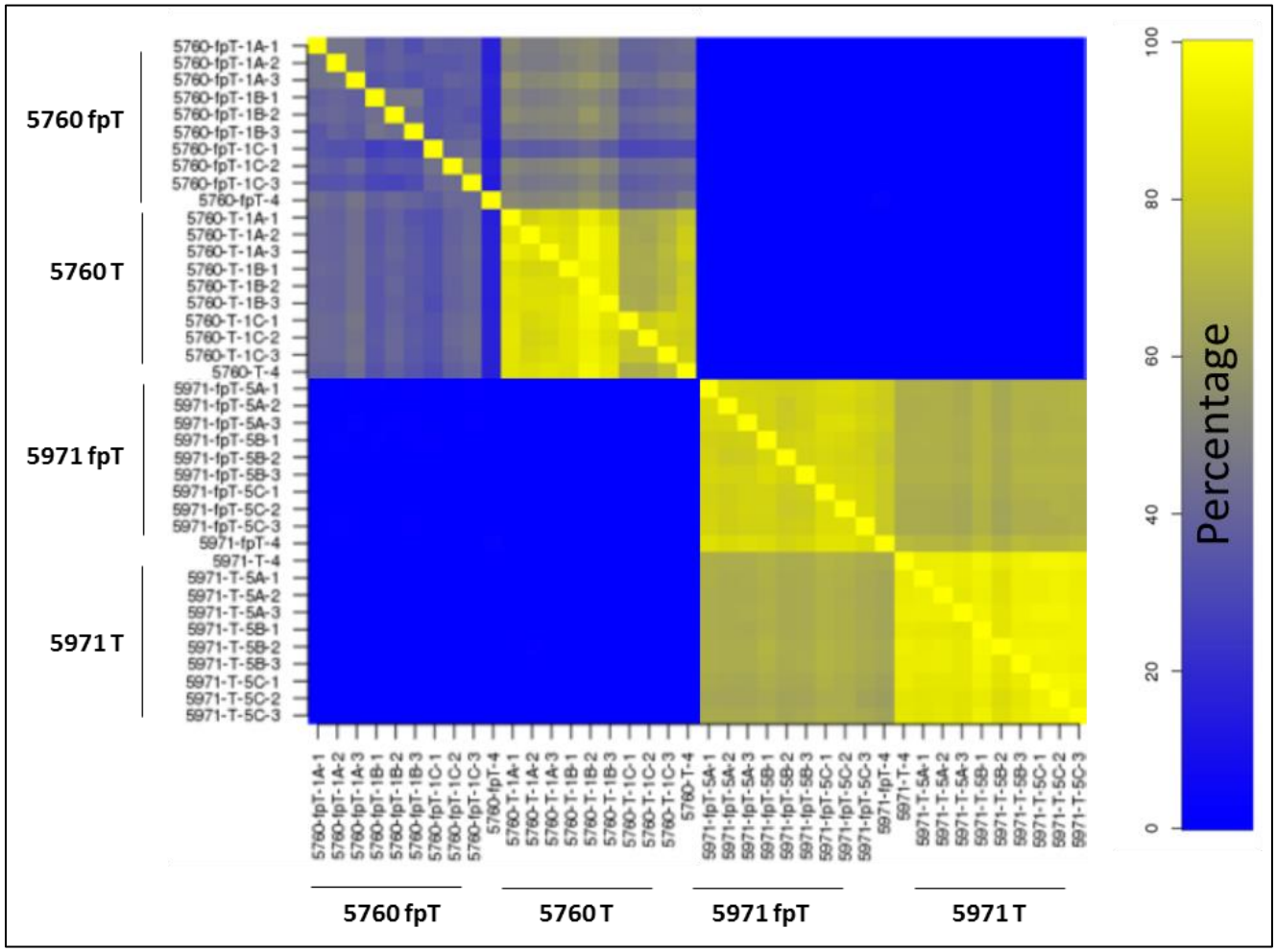


Figure 2B. Jaccard Index by pairwise alignment of samples.



10. **Inter-run Reproducibility:** Inter-run reproducibility was assessed by sequencing the samples at different time points. As part of the initiation of the multi-center WES validation effort, sequencing was performed on a cohort of five lung carcinoma patients. This sequencing run was performed three months prior to the current study. Samples MDA-5760, MDA-5812, and MDA-5971, of the present study, were part of this previous cohort. **Table 5** shows the sequenced DNA quality and the coverage achieved in the previous sequencing assay. Analysis of the sequenced data along with samples from the current study show significant similarity in the number of mutation calls (**Figure 3A**). Additionally, the Jaccard, ISA PPA_{refFF} for these tumor FFPE and FF samples exhibited values similar to that observed in the analytical validation experiment (**Table 4**). Variation in SNV calls in MDA-5760 FFPE were observed. The number of mutation calls from the previous sequencing assay (MDA-5760fpT-T4**)** was similar to that observed in case of frozen tumor samples from the current and previous sequencing assay. The variation between FFPE samples from current and previous sequencing may be attributed to the significantly different coverage obtained during sequencing (**Figure 3A** Unsupervised clustering of Jaccard values further show all FFPE and FF samples from same tumors cluster together irrespective of the time of their sequencing (**Figure 3B**).

Table 4. Summary table of Jaccard, PPA and ISA describing concordance between FFPE and FF samples. *Highlighted in bold data f in FFPE and FF tumors from the previous sequencing assay*

Sample.ID	Both Present	FFPE Only	FF Only	FFPE Only Unable to Call	FF Only Unable to Call	Jaccard Index	PPA Ref FFPE	PPA Ref FF	ISA
MDA-5760-4	**13**	**17**	**43**	**0**	**0**	**17.81**	**43.33**	**23.21**	**30.23**
MDA-5760-1A-1	7	3	30	0	0	17.5	70	18.92	29.79
MDA-5760-1A-2	8	5	29	0	0	19.05	61.54	21.62	32
MDA-5760-1A-3	8	5	29	0	0	19.05	61.54	21.62	32
MDA-5760-1B-1	9	1	29	0	0	23.08	90	23.68	37.5
MDA-5760-1B-2	9	0	32	0	0	21.95	100	21.95	36
MDA-5760-1B-3	9	5	35	0	0	18.37	64.29	20.45	31.03
MDA-5760-1C-1	9	0	44	0	0	16.98	100	16.98	29.03
MDA-5760-1C-2	13	4	41	0	0	22.41	76.47	24.07	36.62
MDA-5760-1C-3	12	5	36	0	0	22.64	70.59	25	36.92
MDA-5971-4	**140**	**28**	**57**	**0**	**0**	**62.22**	**83.33**	**71.07**	**76.71**
MDA-5971-5A-1	114	18	45	1	1	64.41	86.36	71.7	78.35
MDA-5971-5A-2	108	18	46	0	0	62.79	85.71	70.13	77.14
MDA-5971-5A-3	114	16	48	1	1	64.04	87.69	70.37	78.08
MDA-5971-5B-1	113	23	52	0	1	60.11	83.09	68.48	75.08
MDA-5971-5B-2	114	25	52	0	0	59.69	82.01	68.67	74.75
MDA-5971-5B-3	118	24	50	0	0	61.46	83.1	70.24	76.13
MDA-5971-5C-1	116	17	43	0	0	65.91	87.22	72.96	79.45
MDA-5971-5C-2	105	18	47	0	1	61.76	85.37	69.08	76.36
MDA-5971-5C-3	115	18	42	0	0	65.71	86.47	73.25	79.31

Table 5. Tumor DNA Quality Control and Average Coverage in FFPE and FF tumors from the previous sequencing assay.

MDA Sample ID	Tissue Type	DNA Conc. (ng/uL)	DNA Quality (Ab_{260}/Ab_{280} ratio)	Target Coverage	Read Counts for SNV calls
MDA-5760-T-4	Frozen	61.4	1.86	237.55	104
MDA-5760-fpT-4	FFPE	263.7	1.89	177.84	330
MDA-5971-T-4	Frozen	81.6	1.81	243.18	306
MDA-5971-fpT-4	FFPE	501.7	1.87	234.74	323

Figure 3A. Number of Somatic Exonic SNVs in FFPE (FpT) and FF (T). Samples labelled 5760fpT-T4, 5760-T-4, 5971fp-T4, 5971-T4 were sequenced at a different time point.

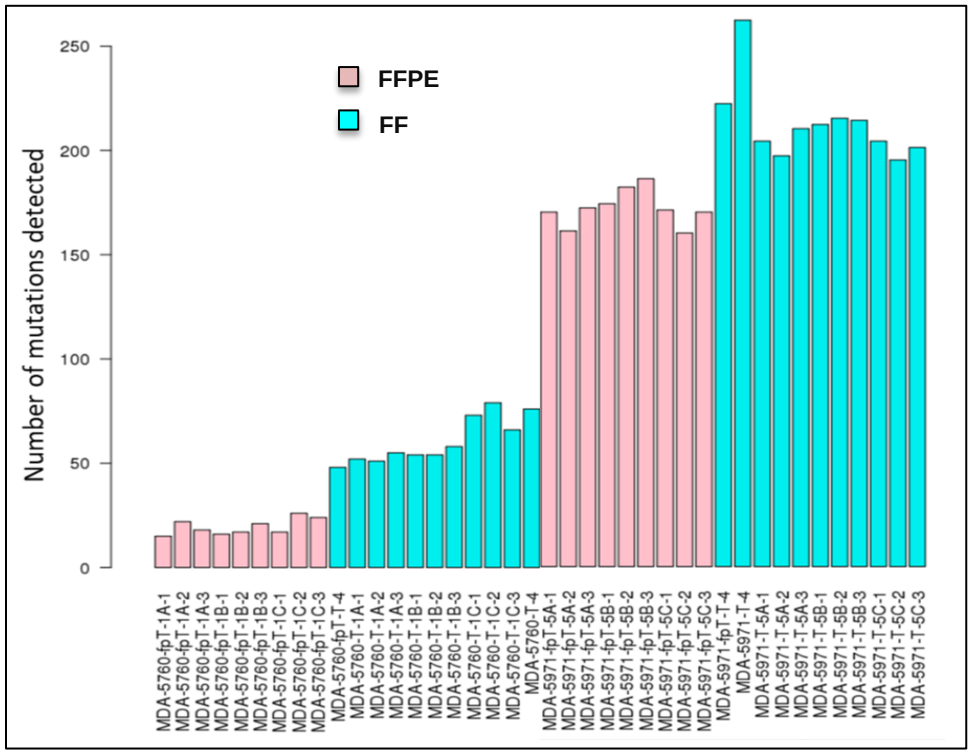
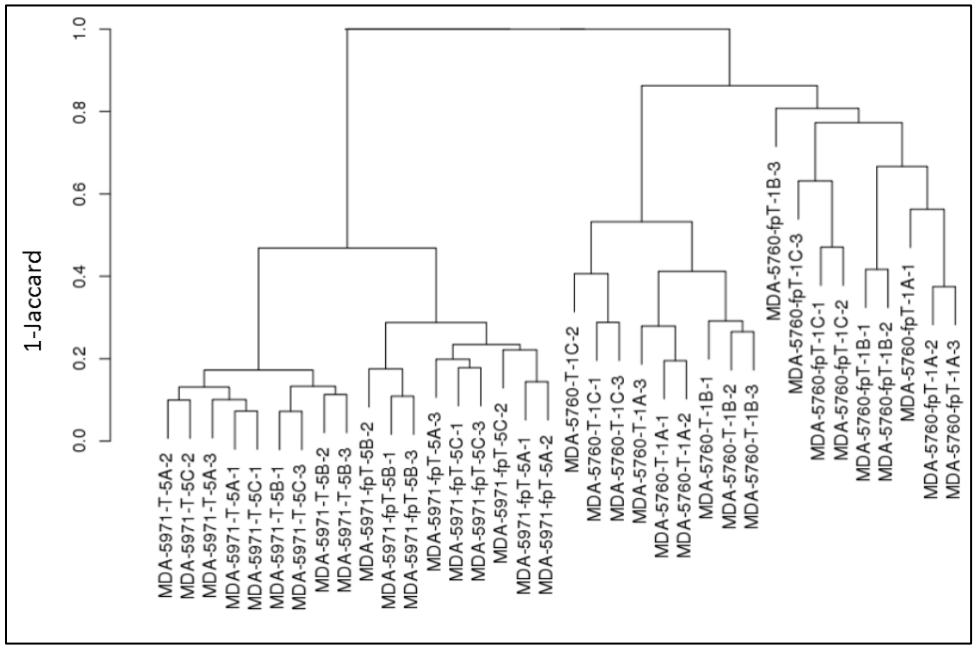


Figure 3B. Unsupervised clustering of Jaccard indices.



Step 2

11. **Sensitivity and Specificity:** The sensitivity and specificity of the sequencing assay were derived from the two pools of HapMap cell lines (XX and YY) data that are pools of 10 HapMap cell lines mixed equally. A bench mark mutation list of the pooled samples was provided by MOCHA. Each pool was sequenced twice in two independent library preparation and sequencing run denoted here by R1 and R2.To minimize the effects of differences in sequencing platform and mutation calling pipeline, the bench mark mutation lists was filtered by retaining mutations that are called by any of the three mutation calling algorithms (MuTect, MuTect2, and MuSE) on each pooled sample. These filtered mutation list was used as the bench mark for the sequencing assay validation. Results from the regular pipeline at MDACC were then evaluated for sensitivity and specificity. Table 6 below shows the values obtained for when no filtering vs MuTect, filtering vs MuTect, and MDACC specific (0.2 VAF) filtering was applied.

Table 6. Analysis of assay sensitivity and specificity using HapMap cell lines.

Filter	Number of mutations called	Sample	Sensitivity	specificity
Unfiltered	129852	XX-R1	99.97	50.29
Unfiltered	127869	XX-R2	94.62	48.34
Unfiltered	131735	YY-R1	99.96	30.85
Unfiltered	133700	YY-R2	96.35	30.81
Filter by Mutect	63877	XX-R1	88.27	90.26
Filter By Mutect	60731	XX-R2	82.93	89.2
Filter by mutect	64627	YY-R1	87.33	87.99
Filter by mutect	62746	YY-R2	83.3	86.45
Filter by Mutect and center	60362	XX-R1	86.24	90.34
Filter by Mutect and center	60234	XX-R2	82.41	89.38
Filter by Mutect and center	63039	YY-R1	85.22	88.03
Filter by Mutect and center	62539	YY-R2	83.14	86.57

Sensitivity = True positive/(True positive + False negative)
Specificity = True negative/(True negative + False positive)

Step 3

12. **Pilot experiment:** A pilot test run of the samples were performed using 8 different lung cancer cases with FF and FFPE samples representing each case. The DNA extracted from each sample had a high tumor content and 260/280 ratio between 1.8–2.0. as shown in **Table 7**.

Table 7. Study Sample Information for Step 3 and Quality Control (QC).

CIMAC2 Candidate Case	Tissue Type	T : Frozen tumor, fpT: FFPE tumor, C: WBC or PBMC	% Malignant Cell	DNA Conc (ng/ul, Qubit)	DNA Quality (Ab 260/Ab 280 ratio)
3	Frozen	T	56	724	1.88
3	FFPE	fpT	40	288	1.89
3	PBMC/WBC	C		154.3	1.88
7	Frozen	T	49	538	1.86
7	FFPE	fpT	46	252	1.9
7	PBMC/WBC	C		201.2	1.89
8	Frozen	T	70	598	1.86
8	FFPE	fpT	40.7	118	1.87
8	PBMC/WBC	C		541.9	1.87
9	Frozen	T	80	726	1.88
9	FFPE	fpT	79	366	1.86
9	PBMC/WBC	C		358	1.9
11	Frozen	T	70	860	1.89
11	FFPE	fpT	45	150	1.86
11	PBMC/WBC	C		186.5	1.89
13	Frozen	T	80	730	1.89
13	FFPE	fpT	38	186	1.86
13	PBMC/WBC	C		406	1.88
14	Frozen	T	50	606	1.87
14	FFPE	fpT	almost 100	414	1.87
14	PBMC/WBC	C		836	1.86

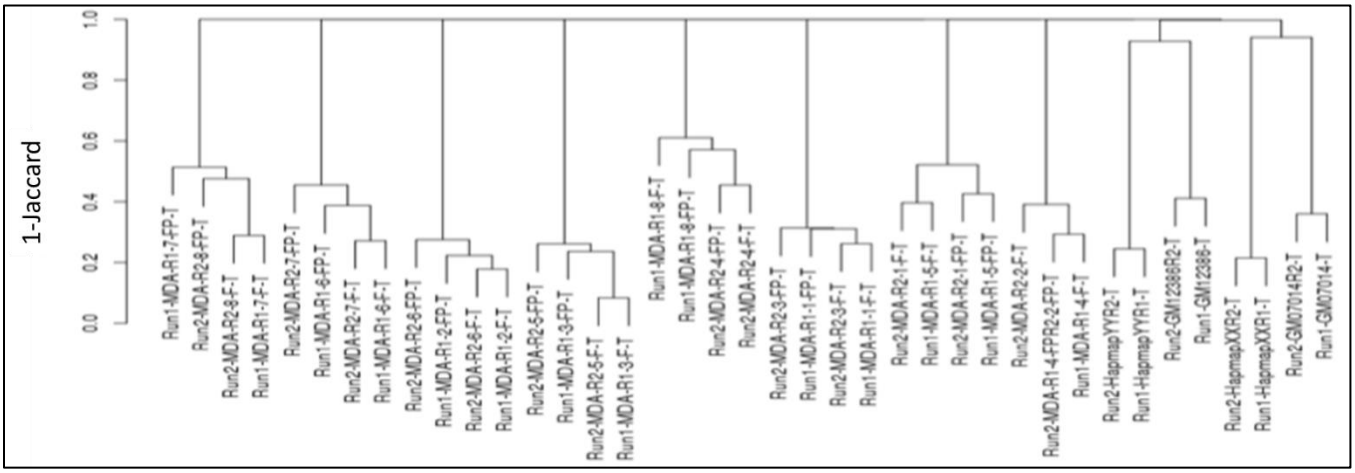
Extracted DNA was aliquoted in duplicates and each set of aliquots were labeled double blinded. Sequencing libraries were generated from each aliquot generating 7 (tumor cases) x 2 (FFPE or FF) x 2 (duplicates) libraries. **Supplemental Table 2** shows the coverage and the total reads obtained from sequencing of each library.

The sequenced data was analyzed as in the prior reproducibility experiment and ISA, PPA and Jaccard values were obtained to determine concordance between FF and FFPE samples. **Table 9** below shows the values obtained for each tumor. Sequencing runs 1 and 2 are shown separately. The Jaccard values and ISA and PPA show good concordance between FFPE and FF samples. An unsupervised clustering (**Figure 4**) of the samples show all duplicates cluster together. Clustering also observed between FF and FFPE samples. A very low concordance was observed between FF and FFPE for MDA-R1-3 and MDA-R2-5 which actually represent the same sample sequenced in two different runs. Possible reasons for this low concordance could be that the different sections of the tumor that contributed to the source of DNA had significantly different mutation profiles. We also noticed significant difference in tumor content between the two samples (50% vs~100%) as well which may also be a contributing factor. These differences could contribute to significantly high mutation identified only in the FF tissue for the sample.

Table 9. Summary table of Jaccard, PPA and ISA describing concordance between FFPE and FF samples.

Sample ID	CIMAC case #	Both Present	FFPE Only	FF Only	FFPE Only Unable to Call	FF Only Unable to Call	Jaccard Index	PPA Ref FFPE	PPA Ref FF	ISA
Pilo2-Run1-MDA-R1-1	3	148	18	28	0	0	76.29	89.16	84.09	86.55
Pilo2-Run1-MDA-R1-2	9	325	8	30	0	0	89.53	97.6	91.55	94.48
Pilo2-Run1-MDA-R1-3	14	270	3	1256	0	2	17.66	98.9	17.69	30.02
Pilo2-Run1-MDA-R1-5	8	100	28	10	0	0	72.46	78.12	90.91	84.03
Pilo2-Run1-MDA-R1-6	7	67	6	72	0	1	46.21	91.78	48.2	63.21
Pilo2-Run1-MDA-R1-7	11	145	20	48	0	0	68.08	87.88	75.13	81.01
Pilo2-Run1-MDA-R1-8	13	58	28	122	0	0	27.88	67.44	32.22	43.61
Pilo2-Run2-MDA-R2-1	8	93	11	17	0	0	76.86	89.42	84.55	86.92
Pilo2-Run2-MDA-R2-3	3	141	20	30	0	0	73.82	87.58	82.46	84.94
Pilo2-Run2-MDA-R2-4	13	56	3	76	0	0	41.48	94.92	42.42	58.64
Pilo2-Run2-MDA-R2-5	14	266	22	1271	0	1	17.06	92.36	17.31	29.15
Pilo2-Run2-MDA-R2-6	9	326	24	20	0	0	88.11	93.14	94.22	93.68
Pilo2-Run2-MDA-R2-7	7	55	7	64	0	0	43.65	88.71	46.22	60.77
Pilo2-Run2-MDA-R2-8	11	152	12	67	0	1	65.8	92.68	69.41	79.37

Figure 4. Unsupervised clustering of Jaccard indices.



Conclusion: The analytical validation of WES using FF and FFPE from lung tumors have been successfully performed. Data analysis using the set parameters show reproducible data that can be utilized for further interrogation.

References:

1. Li and Durbin 2009 Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009 Jul 15; 25(14):1754-60.
2. Cibulskis K, McKenna A, Fennell T, Banks E, DePristo M, Getz G. ContEst: estimating cross-contamination of human samples in next-generation sequencing data. Bioinformatics. 2011 Sep 15;27(18):2601-2.
3. Rimmer A, Phan H, Mathieson I, Iqbal Z, Twigg SRF; WGS500 Consortium, Wilkie AOM, McVean G, Lunter G. Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications.Nat Genet. 2014 Aug;46(8):912-918.
4. Cibulskis K1, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, Gabriel S, Meyerson M, Lander ES, Getz G. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnol. 2013 Mar;31(3):213-9.
5. Ye K1, Schulz MH, Long Q, Apweiler R, Ning Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. Bioinformatics. 2009 Nov 1;25(21):2865-71.
6. Olshen AB1, Venkatraman ES, Lucito R, Wigler M. Circular binary segmentation for the analysis of array-based DNA copy number data. Biostatistics. 2004 Oct;5(4):557-72.

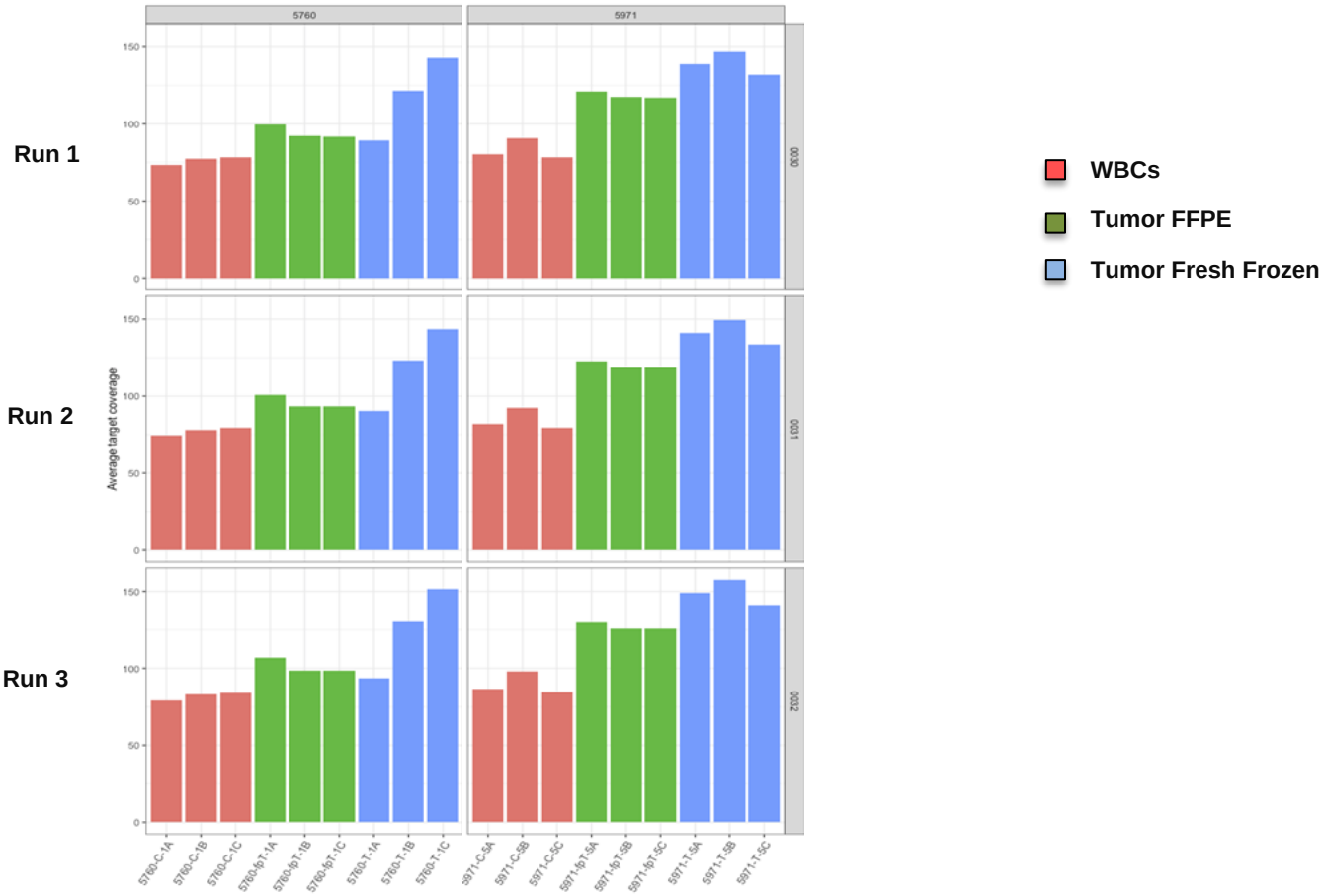
Andy Futreal, BS, Ph.D., Professor, and Chair; Genomic Medicine
Ignacio Ivan Wistuba, MD, Professor, and Chair, Translational Molecular Pathology
Curtis Gumbs, Scientific Manager, Genomic Medicine
Jianhua (John) Zhang, Ph.D., Director, Computational Genomics, Genomic Medicine

[Signatures] and [dates]

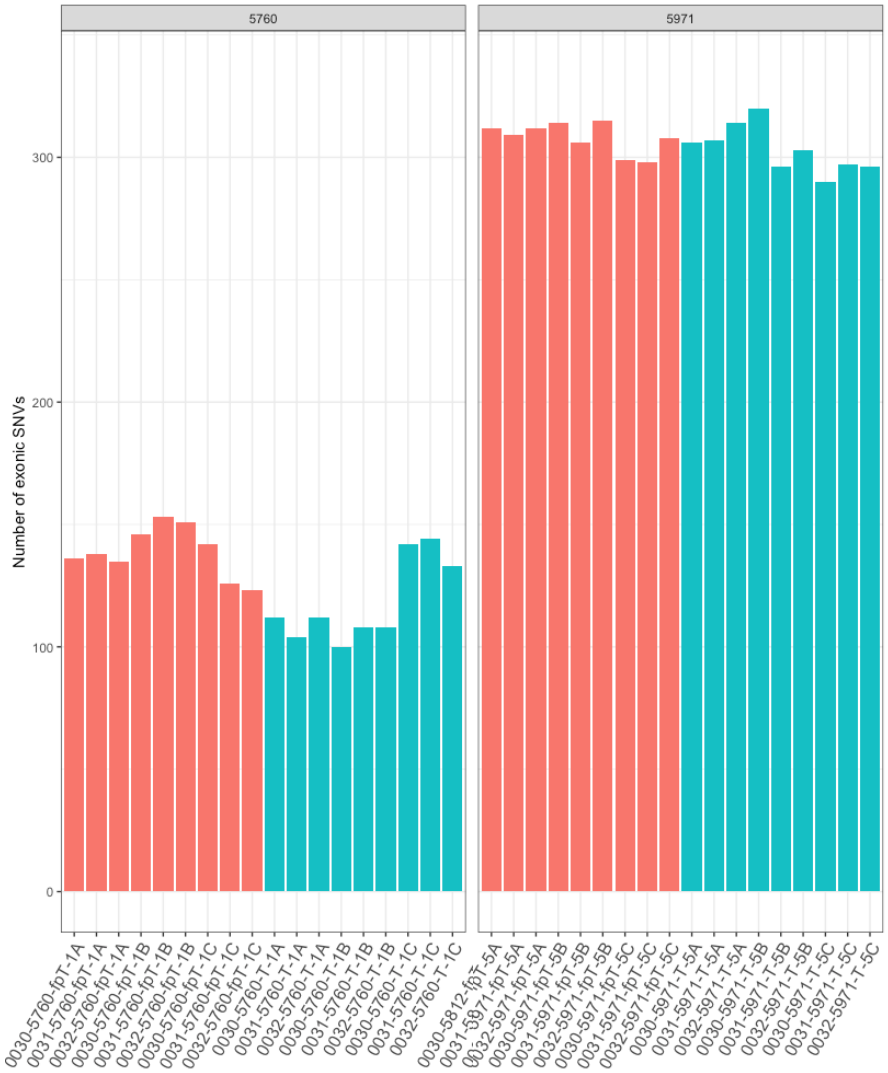
10/14/2019

Appendix

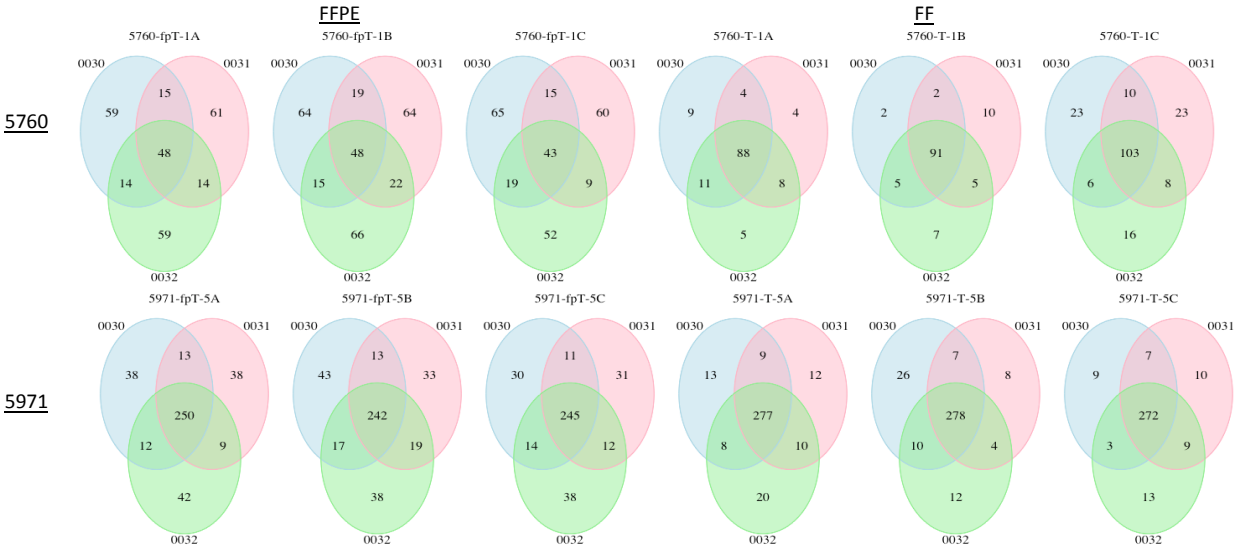
Supplemental Figure 1: DNA Target Coverage per replicate and run.



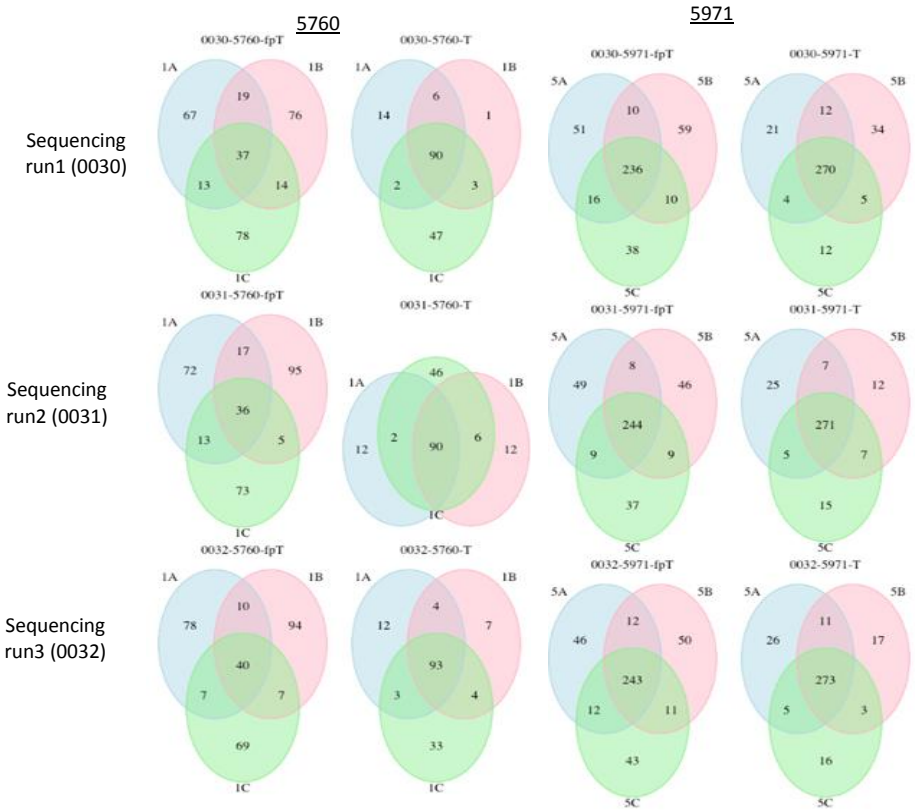
Supplemental Figure 2: Number of Somatic Exonic SNVs in FFPE (FpT) and FF (T).



Supplemental Figure 3A. Venn diagram representing exonic SNVs calls in replicate libraries <u>across</u> sequencing runs (0030, 0031, 0032).



Supplemental Figure 3B. Venn diagram representing exonic SNV calls in replicates <u>within</u> a sequencing run.



Supplemental Table 1: Sequencing coverage

Coverage data for samples sequenced at an earlier time point corresponding to MDA-5760fpT, MDA-5760T, MDA-5971fpT, and MDA- 5971T.

External ID	Total reads (million)	Mapping rate (%)	Duplicate rate mapped reads (%)	Mean Target COV	Median Target COV	100X (% Targets)	75X (% Targets)	50X (% Targets)	OFF BAIT Base Rate
MDA-5760-C-T4	293.68	98.76	44.55	237.27	187.82	81.26	88.4	93.91	0.052668
MDA-5760fpT-T4	326.16	96.95	52.22	177.84	148.56	73.26	84.09	92.31	0.240523
MDA-5760-T-4	317.36	98.63	48.66	237.55	186.28	81.25	88.58	94.12	0.047311
MDA-5971-C-T4	301.94	98.8	45.74	238.09	192.53	82.99	89.63	94.59	0.044646
MDA-5971-fpT-T4	271.86	98.71	43.39	234.74	194.33	83.04	89.8	94.73	0.07465
MDA-5971-T-T4	279.64	98.91	41.28	243.18	181.98	78.52	86.58	92.99	0.045877

Coverage data for analytical validation samples from sequencing assay # 0030.

External ID	Total reads (million)	Mapping rate (%)	Duplicate rate mapped reads (%)	Mean Target COV	Median Target COV	100X (% Targets)	75X (% Targets)	50X (% Targets)	OFF BAIT Base Rate
MDA-5760-C-1A	94.43	99.01	23.59	79.16	62.12	21.4	38.25	62.55	0.063766
MDA-5760-C-1B	103.49	98.99	26.53	83.08	67.25	23.87	42.76	67.53	0.055375
MDA-5760-C-1C	112.43	98.88	30.67	84.2	67.27	24.86	43.02	67.06	0.063525
MDA-5760-fpT-1A	186.09	89.85	42.68	106.64	88.19	41.98	59.79	78.68	0.127776
MDA-5760-fpT-1B	167.19	92.2	44.55	98.57	81.36	36.72	55.11	75.54	0.103044
MDA-5760-fpT-1C	177.66	89.13	43.22	98.34	80.94	36.34	54.73	75.34	0.138759
MDA-5760-T-1A	189.09	98.98	53.31	93.35	75.25	30.19	50.22	73.6	0.069505
MDA-5760-T-1B	166.03	98.89	29.26	129.98	100.09	50.05	66.32	82.3	0.055176
MDA-5760-T-1C	199.86	99.04	29.78	151.72	119.05	60.8	75.24	87.51	0.056545
MDA-5971-C-5A	122.82	98.41	22.51	86.71	69.36	26.45	44.92	68.79	0.203903
MDA-5971-C-5B	119.19	99.08	24.61	97.92	77.44	33.7	52	73.67	0.056925
MDA-5971-C-5C	101.64	98.97	23.98	84.59	67.82	25.01	43.43	67.61	0.058883
MDA-5971-fpT-5A	192.35	95.23	40.53	129.94	109.36	55.86	71.76	85.86	0.075908
MDA-5971-fpT-5B	185.66	94.97	40.3	125.88	105.4	53.47	69.98	84.78	0.076067
MDA-5971-fpT-5C	184.55	94.93	39.12	125.82	106.91	54.38	70.92	85.58	0.082501
MDA-5971-T-5A	199.9	98.97	31.83	149.15	115.65	58.84	73.44	86.45	0.060758
MDA-5971-T-5B	217.45	98.97	33.35	157.38	123.63	62.78	76.39	88.06	0.061846
MDA-5971-T-5C	190.95	98.59	28.92	141.32	108.03	54.57	69.73	83.97	0.114592

Coverage data for analytical validation samples from sequencing assay # 0031.

External ID	Total reads (million)	Mapping rate (%)	Duplicate rate mapped reads (%)	Mean Target COV	Median Target COV	100X (% Targets)	75X (% Targets)	50X (% Targets)	OFF BAIT Base Rate
MDA-5760-C-1A	92.31	99	26.2	74.55	58.42	18.21	34.47	59.05	0.063827
MDA-5760-C-1B	100.63	98.99	28.81	78.09	63.13	20.24	38.37	64.07	0.055448
MDA-5760-C-1C	109.52	98.88	32.59	79.58	63.35	21.58	39.22	63.83	0.063457
MDA-5760-fpT-1A	177.98	89.93	43.41	100.82	83.29	38.17	56.44	76.38	0.127265
MDA-5760-fpT-1B	160.79	92.27	45.3	93.59	77.18	33.33	51.78	73.27	0.102621
MDA-5760-fpT-1C	170.49	89.23	43.99	93.22	76.57	32.78	51.34	72.93	0.138198
MDA-5760-T-1A	184.9	98.98	53.87	90.16	72.42	27.75	47.57	71.72	0.068946
MDA-5760-T-1B	162.55	98.89	31.35	123.19	94.56	46.55	63.38	80.6	0.055329
MDA-5760-T-1C	195.02	99.04	31.71	143.65	112.36	57.36	72.68	86.21	0.056558
MDA-5971-C-5A	120.09	98.41	25.25	81.73	65.18	22.99	40.85	65.6	0.202544
MDA-5971-C-5B	116.34	99.07	27.05	92.22	72.7	29.96	48.05	70.88	0.057148
MDA-5971-C-5C	99.15	98.96	26.47	79.6	63.61	21.43	39.24	64.34	0.059035
MDA-5971-fpT-5A	183.95	95.27	41.32	122.68	103.01	51.95	68.85	84.28	0.075433
MDA-5971-fpT-5B	177.98	95.03	41.26	118.83	99.25	49.53	66.93	83.13	0.075493
MDA-5971-fpT-5C	177.53	94.98	40.25	118.85	100.81	50.56	67.8	83.89	0.082
MDA-5971-T-5A	194.38	98.97	33.57	141.05	109.09	55.28	70.76	85.01	0.060675
MDA-5971-T-5B	212.94	98.95	35.23	149.43	117.03	59.64	74.13	86.85	0.061816
MDA-5971-T-5C	186.07	98.6	30.9	133.74	101.92	51.18	66.93	82.36	0.113869

Coverage data for analytical validation samples from sequencing assay # 0032.

External ID	Total reads (million)	Mapping rate (%)	Duplicate rate mapped reads (%)	Mean Target COV	Median Target COV	100X (% Targets)	75X (% Targets)	50X (% Targets)	OFF BAIT Base Rate
MDA-5760-C-1A	90.06	99.05	25.44	73.51	57.6	17.61	33.68	58.34	0.063931
MDA-5760-C-1B	98.55	99.03	28.14	77.22	62.39	19.7	37.69	63.41	0.055551
MDA-5760-C-1C	106.21	98.92	31.68	78.27	62.37	20.65	38.16	62.92	0.063445
MDA-5760-fpT-1A	172.17	90.03	42.28	99.61	82.22	37.46	55.68	75.9	0.127124
MDA-5760-fpT-1B	154.5	92.37	44.04	92.13	75.99	32.34	50.81	72.62	0.102526
MDA-5760-fpT-1C	164.04	89.33	42.73	91.85	75.48	31.92	50.39	72.24	0.138024
MDA-5760-T-1A	179.6	99.03	53.04	89.22	71.71	27.04	46.89	71.16	0.068791
MDA-5760-T-1B	158.34	98.94	30.59	121.4	93.23	45.63	62.64	80.1	0.055209
MDA-5760-T-1C	191.84	99.08	31.12	142.61	111.63	56.9	72.32	86	0.056568
MDA-5971-C-5A	116.35	98.47	24.33	80.2	63.9	21.99	39.68	64.58	0.202538
MDA-5971-C-5B	113.39	99.11	26.31	90.83	71.67	28.98	47.13	70.13	0.057138
MDA-5971-C-5C	96.69	99.01	25.68	78.49	62.77	20.72	38.32	63.53	0.059041
MDA-5971-fpT-5A	178.1	95.35	40.23	121.08	101.63	51.05	68.12	83.88	0.07541
MDA-5971-fpT-5B	172.78	95.1	40.19	117.55	98.2	48.78	66.3	82.85	0.0755
MDA-5971-fpT-5C	170.87	95.07	39.06	116.78	99.06	49.33	66.93	83.37	0.081974
MDA-5971-T-5A	189.1	99.01	32.73	139.06	107.5	54.42	70.13	84.55	0.060651
MDA-5971-T-5B	206.13	99.01	34.25	146.95	115.04	58.62	73.34	86.45	0.061782
MDA-5971-T-5C	181.39	98.65	30.11	131.93	100.68	50.43	66.26	81.98	0.1139

Supplemental Table 2. Average Coverage in samples (in millions).

	Run1					Run2				
CIMAC2 Candidate Case #	External ID	Total Reads (million)	Mean Target Coverage (millions)	100X (% Targets)	OFF BAIT Base Rate	External ID	Total Reads (million)	Mean Target Coverage (millions)	100X (% Targets)	OFF BAIT Base Rate
3	MDA-R1-1-FP	354.42	297.08	65.48	0.027	MDA-R2-3-F	452.92	344.08	72.84	0.029
3	MDA-R1-1-F	359.42	295.33	68.08	0.033	MDA-R2-3-FP	358.09	293.14	64.05	0.028
9	MDA-R1-2-FP	425.6	310.52	62.11	0.032	MDA-R2-6-F	405.49	317.41	64.68	0.03
9	MDA-R1-2-F	417.92	309.7	62.2	0.029	MDA-R2-6-FP	343.48	267.91	57.31	0.028
14	MDA-R1-3-FP	349.92	253.43	61.48	0.031	MDA-R2-5-F	398.37	301.49	66.37	0.028
14	MDA-R1-3-F	410.75	315.94	67.21	0.036	MDA-R2-5-FP	337.03	252.78	59.02	0.035
8	MDA-R1-5-FP	389.94	295.55	63.66	0.03	MDA-R2-1-F	420.08	313.16	66.08	0.031
8	MDA-R1-5-F	395.62	314.07	69.49	0.028	MDA-R2-1-FP	366.02	274.81	62.44	0.032
7	MDA-R1-6-FP	376.63	319.48	62.43	0.028	MDA-R2-7-F	453.36	338.98	69.11	0.034
7	MDA-R1-6-F	442.43	341.72	68.36	0.031	MDA-R2-7-FP	412.4	332.35	59.6	0.027
11	MDA-R1-7-F	392.98	294.97	64.99	0.032	MDA-R2-8-F	411.91	291	65.03	0.033
13	MDA-R1-8-FP	381.43	279.25	62.79	0.031	MDA-R2-4-F	383.28	292.2	62.14	0.03
13	MDA-R1-8-F	406.86	317.53	67.38	0.029	MDA-R2-4-FP	374.05	268.21	60.92	0.047