

iSAVE: A Novel Validation Strategy for Next Generation Sequencing Mutation Profiling in FFPE Tissues

Ken CN Chang*, Gladys Arreaza, John Kang, Maureen Maguire, Ping Qiu and Matthew J Marton

Translational Biomarkers, Merck and Co., Inc., Rahway, NJ 07065, USA

Abstract

We previously developed a “concordance calculator” to quantify reproducibility of multi-variant calls among next generation sequencing (NGS) samples and replicates. This tool and a novel replicate approach have been also used to eliminate many different technical artifacts including post tissue collection modifications (PTCM) such as deamination and oxidation artifacts. Here we apply this approach to study the impact of tumor heterogeneity among consecutive FFPE tissue sections across entire tumor blocks and cannot detect any impact of heterogeneity among different sections/regions of tumors in terms of mutation profiles using the NGS AmpliSeq Cancer Panel, even though the tumor was visibly heterogeneous according to the H&E images and pathological review. RNA expression profiling using a NanoString Cancer Panel found significantly different expression patterns among different sections/regions. Additional studies in a different tissue type also found no detectable discrepancies among different tissue sections in terms of their mutation profiles. If confirmed by further studies, these results using FFPE tissue sections would suggest that DNA mutation signatures as novel biomarkers for cancer diagnosis and prognosis might be less sensitive to tumor heterogeneity than RNA-based expression signatures, at least based on the performance and sensitivity of current DNA/RNA profiling technologies. Use of the concordance calculator to quantify reproducibility of multi-variant calls among Next Generation Sequencing replicates and to eliminate many different technical artifacts including PTCM also allowed us to develop an unconventional validation strategy. We call this validation approach “*in situ* analytical validation and evaluation (iSAVE).” As a proof of concept, we evaluated the RainDance ThunderBolts Cancer Panel and demonstrated analytical validation directly on each and every clinical sample. This strategy also comprises using a set of normal FFPE tissue samples in the validation process to eliminate platform-, panel-, amplicon-, library-preparation-, and mutation calling pipeline specific artifacts.

Keywords: Concordance calculator; RNA expression profiling; DNA/RNA profiling technologies; Biomarker; Tumor heterogeneity; DNA mutation; Personal genome machine

Introduction

Ever since the publication of “Intratumor Heterogeneity Revealed by Multiregion Sequencing” in NEJM in 2012 by Gerlinger et al. questions such as, “How can we expect precision medicine to be effective if tumors are so heterogeneous?” are raised at almost every biomarker or precision medicine conference [1,2]. Despite this concern, new biomarkers, genetic tests and gene signatures are discovered on a routine basis for both diagnostic and prognostic applications [3-6]. Several of these have become FDA cleared and/or Medicare reimbursed. Clearly, there is a disconnection between those people who have concerns about the impact of tumor heterogeneity and those who do not. Many publications focus on a very specific example and use that to speculate on implications and potential future applications in the discussion section. It is often up to the readers to interpret the results and decide how to take the advantage of this information even if the authors already set the boundary and limitation of how such results might be interpreted. After trying to interpret the results obtained from Gerlinger’s publication, we initiated a study to investigate the impact of tumor heterogeneity on both RNA expression and DNA mutation profiling in different sections and regions of a tumor.

Through the effort, an Excel tool called concordance calculator was developed to quantify multi-variant call reproducibility among triplicates. The detail of this tool has been described in a recent publication [1]. Every variant call that passes a basic quality filter is evaluated for its reproducibility in this tool. The reproducibility is defined as the number of reproducible calls divided by the number of all variant calls above a pre-specified variant frequency (VF, such as 5%) among three or more replicates. In that study, the calculator was used to evaluate different NGS methods, and the conclusion was the Ion Torrent AmpliSeq cancer panel showed a very good reproducibility

from library prep to library prep (~95%) while the Illumina TruSeq cancer panel did not. On the other hand, Illumina MiSeq instrument provided better run-to-run reproducibility than those from Ion Torrent PGM (Personal Genome Machine). Since technology is rapidly changing and both manufacturers have made upgrades to their instruments, reagents and library preparation protocols, the above conclusion was only applied to that particular study.

Very interestingly, we found that the Illumina TruSeq cancer panel generated many more C to T variant calls than did the Ion Torrent platform. The dominant variant call in the Illumina TruSeq Cancer Panel was clearly observed to be C to T, while data generated on Ion Torrent AmpliSeq Cancer Panel with the same gDNA did not have this pattern. It is clear that most of those C to Ts were artifacts or what we called Post Tissue Collection Modifications (PTCM, such as deamination or oxidation, which depends on which tissue type and fragmentation procedure was used) [1]. In addition, the Illumina TruSeq library preparation protocol increased the chance of bringing these PTCMs above background level. One way to eliminate these PTCM artifacts is to run triplicate library preparations and to use the concordance calculator to exclude them, since the overwhelming majority of these PTCM artifacts are not reproducible because they

***Corresponding author:** Ken CN Chang, Clinical Development Laboratory, Translational Biomarkers, Merck and Co., Inc., Rahway, NJ 07065, USA, Tel: +19084231000; E-mail: ken.chang@merck.com

Received September 24, 2016; **Accepted** November 10, 2016; **Published** November 12, 2016

Citation: Chang KCN, Arreaza G, Kang J, Maguire M, Qiu P, et al. (2016) iSAVE: A Novel Validation Strategy for Next Generation Sequencing Mutation Profiling in FFPE Tissues. Next Generat Sequenc & Applic 3: 139. doi: [10.4172/2469-9853.1000139](https://doi.org/10.4172/2469-9853.1000139)

Copyright: © 2016 Chang KCN, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

occur randomly like noise [1]. This tool can also be used to evaluate the effectiveness of bioinformatics filters, as a good filter will keep more concordant calls and eliminate more discordant calls.

It turned out that the library preparation protocol is the key for the reproducibility of variant calls with low VF. We called the Ion Torrent AmpliSeq protocol the “Amplify First” protocol, and the Illumina TruSeq protocol the “Amplify Later” protocol because Ion Torrent AmpliSeq protocol starts with 20 cycles of PCR amplification, and the amplification step in the Illumina TruSeq protocol is toward the very end of the library preparation [7-10]. In addition to the longer size of amplicon, this “Amplify Later” protocol is responsible for the big variation to the variants with low VF. Also, a plausible explanation is that the primers designed to hybridize to only one strand in the TruSeq method will double the probability of C to T artifacts compared to protocols that copy from both strands like PCR amplification [1].

Based on all the above findings, we concluded, in order to get reproducible results for mutations with variant frequency between 5% and 25% (which is the range where somatic mutations are most likely going to be found in FFPE tissue) it is recommended to use gene panels with short amplicon sizes (preferably <120 bp) and an “Amplify First” protocol like those in the Ion Torrent AmpliSeq method. Furthermore, one should consider running triplicate library preparations for every single sample in order to optimize the NGS bioinformatics pipeline and generate reproducible mutation profiling clinical data, since we also found that the same mutation in different samples or different mutations in the same sample could all be measured differently in terms of their detection sensitivity [1,11-13].

All these above mentioned previous study results as well as the interest of studying the impact of tumor heterogeneity among FFPE tissue sections led us to design an unconventional analytical validation approach for an NGS targeted cancer gene panel to support the exploratory objectives of an on-going clinical trial. Many laboratories view validation as one time event that is summarized in a validation report. However, since it's impossible to validate every possible mutation there is a need for ongoing performance verification or validation. In this manuscript we provide an example of how this in situ analytical validation and evaluation (iSAVE) strategy works. Each clinical sample will be run in triplicate, which might typically be 7-10 samples plus normal control and standards all in one MiSeq run. Our data show that by starting with basic QC filters and only including reproducible calls, and then removing those reproducible variant calls that show up across all samples (including normal tissue samples), we are then able to eliminate most artifact mutation calls from these clinical samples, resulting in low false positive and false negative rates.

Materials and Methods

FFPE tissue source and sample preparation

Colorectal cancer (CRC) FFPE tissue blocks that had been profiled with an NGS-based cancer panel and their corresponding 5 µm sectioned slides were purchased from BioChain Institute, Inc. (Newark, CA). Genomic DNA (gDNA) was isolated using Qiagen DNA FFPE Tissue Extraction Kit (Qiagen, Germantown, MD). Qubit (Life Technologies, Carlsbad, CA) and Nanodrop (Thermo Fisher, Waltham, MA) quantification as well as Bioanalyzer (Agilent Technologies, Santa Clara, CA) DNA quality analyses were done according to the standard protocols provided by the manufacturers.

Renal cell carcinoma (RCC) FFPE tissue slides/sections (5 µm) were purchased from BioChain Institute Inc. (Newark, CA). Isolation

of genomic DNA from these slides was done using Norgen FFPE DNA/ RNA plus isolation kit (Norgen Biotek Coro., Thorold, ON, catalog # 54300). Normal (non-tumor) kidney FFPE tissue blocks were also purchased from BioChain.

Ion Torrent AmpliSeq library preparation procedures

All standard library preparation protocols for FFPE tissue samples were followed according to the manufacturer's instruction. A standard Qubit-quantified gDNA input amount of 10 ng for Ion Torrent Cancer Panel (Life Technologies, Carlsbad, CA) was used for all the experiments unless otherwise specified.

RainDance ThunderBolts cancer panel library preparation procedures

The standard RainDance ThunderBolts (RainDance Technologies, Inc, Billerica, MA) cancer panel library preparation protocol was followed with a few modifications. Briefly, 20 ng of purified gDNA from RCC FFPE tissue samples were used for the first PCR reaction using the ThunderBolts cancer panel set 1 and set 2. AmPure beads (Beckman Coulter, Beverly, MA) were used for the purification, followed by the 2nd PCR reaction for Illumina adapters and index addition. After purification using AmPure beads, amplified libraries were quantified by Agilent Bioanalyzer and were normalized to 2 nM prior to sequencing.

Illumina MiSeq and Ion Torrent Ion Personal Genome Machine (PGM) sequencing procedures

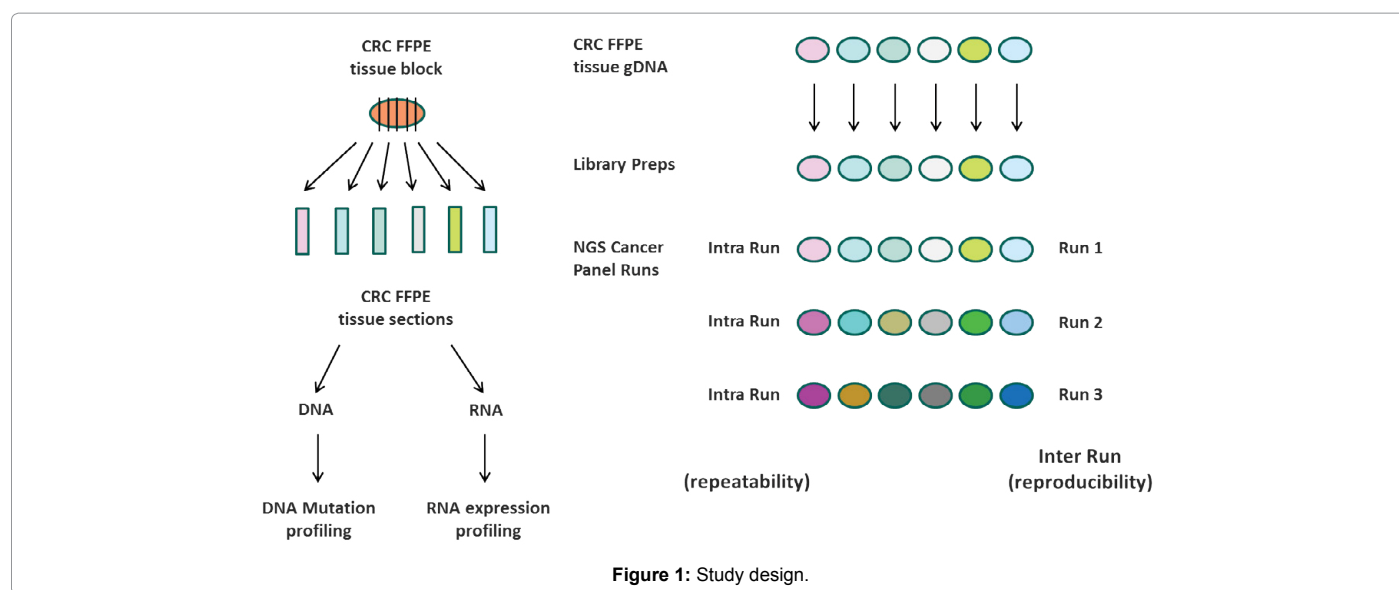
MiSeq Reagent Kit v2 and Ion 318 Chip Kit were used for all the NGS runs with MiSeq and Ion PGM, respectively. Standard protocols were followed for all experiments and instrument runs [14,15], and the reagents were freshly prepared for each run. Seven samples (6 target samples plus a control sample) were multiplexed per chip or flow cell using unique index/barcodes.

Data analysis process and sequence analysis

Data analysis presented was performed using either MiSeq Reporter (Illumina, San Diego, CA), Torrent Suite (Ion Torrent, Life Technologies), or OmicSoft (OmicSoft Corporation, Cary, NC) (for both platforms) as mentioned in each result sub-section. The default setting of Q score cutoff for OmicSoft was Q13 unless otherwise specified.

Definition of concordant calls among replicates (The concordance calculator Excel tool)

The concordance calculator (and Microsoft Excel-based tool) was designed and developed to identify the range of variant frequency (VF) for which variant calls between replicates are concordant for a particular data set. The details of this procedure were recently published [1]. Here is a brief summary of the unique features of this Excel tool. The tool can be applied to any particular data set containing replicate data by 1) starting with a list of minimally filtered variant calls from one replicate, 2) calculating the acceptable VF range using the equation of $VF \pm (\%VF \times \text{acceptable } \%CV + \% \text{ background variation})$, then 3) looking for the same ID of the variant on the other replicate data sets to determine if such variant call can be found, and if yes, 4) determining whether the corresponding VF is within the calculated acceptable range. If both answers are yes, this variant call is considered repeatable or reproducible. The tool permits one to report % repeatability in the context of specific filters, which then should be included in the validation report or clinical sample testing report to specify the limitation of the data.



RainDance digital PCR procedure and primer design

Primer design and sequences for dPCR confirmation are summarized in the dPCR Excel sheets (Supplementary Table 1). Taqman Genotyping Master Mix (Life Technologies) was used in the dPCR reaction setup. A RainDance RainDrop digital PCR instrument was used to perform sample preparation (RainDrop Source) and analysis (RainDrop Sense). RainDance dPCR standard protocols were used for all these studies. RainDrop Analyst Software from RainDance was employed for data analysis.

Tumor heterogeneity experimental design

The experimental design for the study of the impact of tumor heterogeneity on different FFPE tissue sections from same tissue block was done using one single CRC tissue block and divided into 6 consecutive regions (20 micron each). The basic design of the experiment is the independent evaluation of 6 different regions/sections starting from gDNA isolation, followed by independent library preparation and analysis in a multiplexed run in the same chip. The Nanostring GX Human Cancer Reference Kit was purchased from NanoString directly and 50 ng of RNA was used for all samples according to the standard protocol. A hierarchical clustering heat map was used to analyze the RNA profiles among different regions/sections of FFPE tissue. A design diagram that depicts how the experiments were conducted is shown in Figure 1.

The experimental design for ThunderBolts Cancer Panel variant calls analytical validation involved an un-conventional approach. Two independent runs were carried out by using the ThunderBolt Cancer Panel with RCC samples analyzed on the MiSeq instrument. The first run was used as the training dataset, serving the purpose of tuning the pipeline parameters, and the second run was considered the testing dataset as part of the validation.

Data analysis procedure and NGS variant-calling pipeline

According to the results obtained from previous studies [1] the vast majority of false positive mutation calls are not reproducible and therefore can be identified and eliminated by the use of replicate experiments. In addition, the default settings of many of the existing NGS software packages are not optimized for targeted somatic

mutation detection because the defaults impose filters that are very stringent in order to reduce false positive rate, and therefore increase the risk of false negatives.

Taking these findings into consideration, and following the general procedures recommended by Broad Institute's "DNA-Seq Best Practices" (<https://www.broadinstitute.org/gatk/guide/best-practices?bpm=DNaseq>), we developed a customized quality-based variant caller that integrates basic quality filters and the reproducibility of mutations across triplicate experiment when making mutation calls.

Identification of quality control parameters

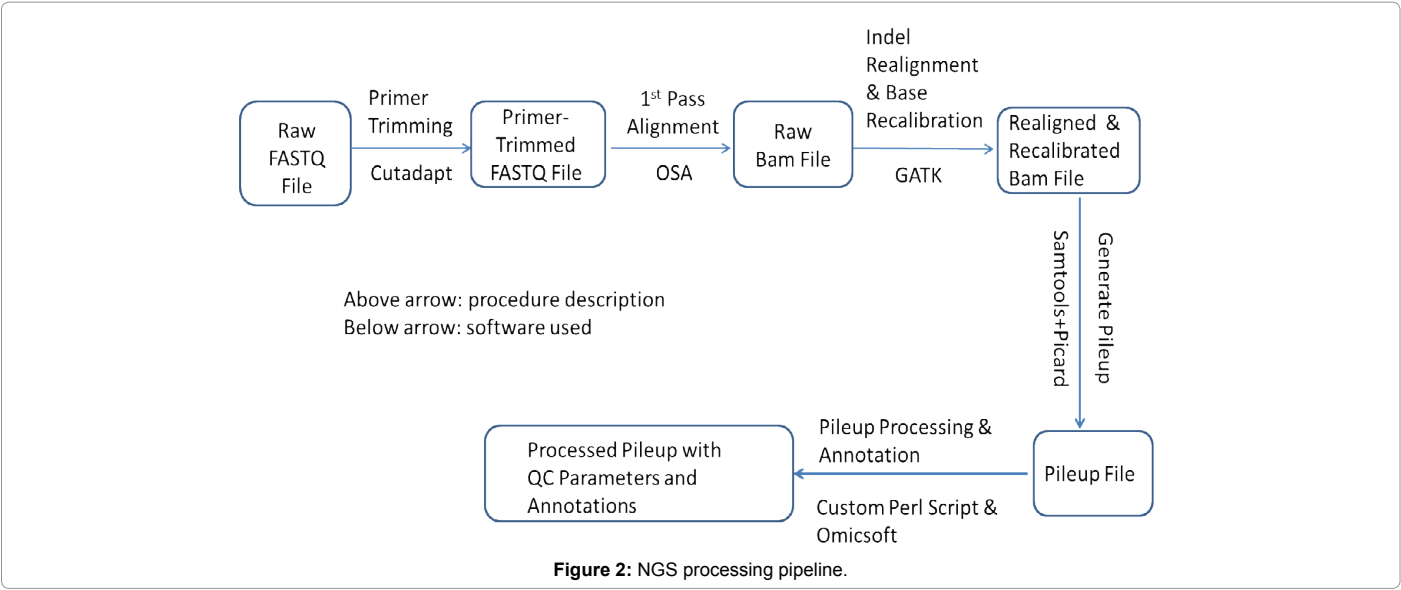
From the NGS mutation calling pipeline outlined above (Figure 2), the following quality control (QC) parameters were obtained: coverage and the number of mutation reads, strand bias, homo-polymer length, the median map quality score of all reads at a chromosome position (Median MQ), and the median base quality score of all reads at a chromosome position (Median BQ).

The strand bias score was calculated as, $\frac{\max(VpCm, VmCp)}{VpCm + VmCp}$ which

is the formula used by Ion Torrent (Torrent Variant Detection Algorithms 3.4). In more detail, Vp= # of alt reads on positive strand, Cp= # of total reads on positive strand, Vm=# of alt reads on minus strand, and Cm= # of total reads on minus strand. By examining this strand bias formula, one would expect that when there is no strand bias, the strand bias score will take the value of 0.5, whereas when there is a strand bias, the strand bias score will tend to 1.

Median BQ and median MQ are indicators of the quality of a chromosome position, because if most of the reads at a chromosome position have low base or map quality scores, we will have less confidence in the mutation call produced at this position. Homo-polymers are stretches of repeating nucleotides in the reference genome; very long homo-polymers form repeats and are difficult to sequence. Variants in homo-polymer runs exceeding eight bases were filtered out (R8 filter as implemented by the Illumina VCF QC filter).

Additional data analysis methods can be found in the Supplemental Methods including filtering out reproducible calls from triplicate data for each sample followed by visual inspection of difficult sequencing regions.



Results

Experimental design to evaluate the impact of tumor heterogeneity on RNA and DNA profiles from FFPE tissue consecutive regions

A colorectal cancer (CRC) FFPE tissue block was randomly selected from the previous platform evaluation study [1]. This FFPE tissue block was evenly divided into six different regions (20 micron each). H&E-stained slides for different regions or sections of FFPET block were reviewed by two pathologists (Supplementary Figures S1-S6 for tissue H&E stains). Although there were no dramatic differences among different sections in terms of gross histology, some degree of heterogeneity from area to area was observed. Each of the six sections was subjected to DNA and RNA preparation. A NanoString commercial GX Human Cancer Reference Kit (230 cancer genes) was used for RNA expression profiling and an Ion Torrent AmpliSeq Cancer Panel (which includes most well documented oncogenes and tumor suppressor genes) was used to profile DNA mutations. Similar to the previous study, intra-run and inter-runs were performed [1]. All samples from 6 regions were processed at the same time throughout the entire profiling process.

Comparison of RNA expression profiles from different regions/sections of a FFPE tissue block using NanoString GX Human cancer panel

NanoString RNA expression profiling data (an average of two replicates) showed some significant differences among different regions or sections as shown in Figure 3 using unsupervised two-dimensional hierarchical clustering. To prove that the observed differences were not due to the variation coming from NanoString cancer panel RNA expression platform, individual replicates for clustering analysis were used. All replicates remained in the same cluster (Supplementary Figures S7 and S8), showing that the replicate variability was much lower than those among different regions/sections, and that the different profiles in the two different clusters among some regions that were statistically significantly different were not experimental variations.

In general, approximately 36 genes (Table 1) out of 230 total genes had gene expression changes greater than 2-fold among six different tissue regions/sections. Although these are not big changes, they are statistically significant (p-values between the two main clusters were less than 0.05) sufficient to illustrate that intra tumor heterogeneity might cause some impact on RNA expression profiles among different sections.

GENE	Fold Change vs Section 1										
	Section 2	Section 3	Section 4	Section 5	Section 6	GENE	Section 2	Section 3	Section 4	Section 5	Section 6
COL1A1	2.8	2.6	0.9	2	0.8	IGF1	1.9	2.1	1	2	0.9
JUNB	3.2	3.3	0.8	2.3	0.8	EGF	2	1.9	0.8	1.8	0.8
TNFRSF1B	2.4	2.5	0.9	1.9	0.9	CDK2	2.1	1.8	1	1.8	0.9
EPS8	2.2	2.3	0.9	1.9	1	FANCG	2	1.8	1	1.7	1.1
TGFB1	2.9	2.7	0.9	2.3	0.9	NPM1	3.2	3.2	0.8	2.5	0.8
FGFR4	2	2	0.8	1.6	0.8	LYN	1.9	1.5	1.1	2	1.1
TIMP1	0.5	0.4	1.2	0.5	1	NGFR	2.6	3	1.1	2.6	1.1
MMP9	1.8	2	1.2	1.7	1	CDKN2C	2.3	2.2	0.9	2	1
BCL3	2	1.5	0.9	1.5	0.9	BCL2	2.1	2.5	0.9	1.7	0.8
ETS1	2.3	2.2	1	2	0.9	KIT	2.6	2.4	1	2.3	0.9
IGFBP6	2.4	2.1	1	1.8	0.9	FLT1	2.3	1.8	0.8	1.8	0.7
TGFBR3	2.2	1.8	0.9	1.9	1	TFE3	1.9	2	1.1	2.2	0.8
DEK	0.3	0.3	1.1	0.4	1	FOLR1	2.8	1.9	1	2.1	1.1
E2F3	2	1.8	0.8	1.7	0.9	PTGS2	3.1	2.4	1.2	2.4	1
SIAH1	2.1	2.3	0.9	1.9	0.9	AR	2.5	3.3	0.8	2.5	0.6
SOD1	2.4	2.2	0.9	2	1	TNF	1.4	2.5	0.7	2.2	0.5
FYN	2.6	2.3	1.3	1.8	1.2	TAL1	2.9	2.6	0.8	2.1	0.6
RARA	2.1	2.1	0.9	2	0.8	GNAS	2.6	3.2	0.8	2.8	0.5

Table 1: Genes impacted by the different sections of FFPE tissue block.

Comparison of DNA mutation profiles on different regions/sections of a FFPE tissue block on Ion Torrent AmpliSeq cancer panel

Using the custom-designed concordance calculator we calculated the reproducibility of variant calls among different regions/sections profiled on Ion Torrent AmpliSeq Cancer Panel and found it was greater than 96% (for those variants with VF between 5 and 25%), which is at least as reproducible as that from six different replicate library preparations using same gDNA [1]. The Concordance Calculator uses very stringent outlier rejection and concordance acceptance criteria based on the degree of VF differences. If only the variant calls detected or not detected were counted, the reproducibility for the two sets of 3-region comparisons (the current version of the concordance calculator only allows analysis of a set of 3 replicates) was both 100%. Therefore, the result clearly shows, although there are some significant impacts on common cancer gene RNA expression profiles from region to region (likely due to tumor heterogeneity using FFPE slides) there is no detectable impact on DNA mutation profiles among frequently encountered cancer genes (based on the data generated using AmpliSeq Cancer Panel).

In addition, other randomly picked tissue samples were subjected to a similar evaluation (e.g., different sections from same renal cell carcinoma FFPE tissue block, unpublished internal study), and we found no significant differences in different regions of the FFPE tissue blocks DNA variant calls that the Calculator called reproducible.

Randomly selected variant calls from Sections 1-3 and Sections 3-5 from the concordance calculator variant call reproducibility evaluation are shown in Tables 2 and 3. Table 2 shows that variant calls with variant frequency (VF) 5% to 25% for sections 1-3 have an overall reproducibility of 96.8%. Variant calls with VF 5% to 25% for sections 3-5 also have a reproducibility of greater than 96% (Table 3). Several identical variant calls across all 5 sections under evaluation were observed (such as chromosome 5 position 170837513 and chromosome 7 position 55221794). Several examples of non-reproducible calls are also included (e.g., chromosome 2 position 29443617 in Table 2 and chromosome 18 position 48584629 in Table 3). The grey boxes show the final tally of reproducible calls divided by the total variant calls

(including non-reproducible calls) to generate percent reproducibility. Furthermore, all variant calls with VF higher than 30% among these sections were 100% reproducible according to the concordance calculator (data not shown).

Analytical performance and validation strategy of RainDance ThunderBolts cancer panel mutation profiling using FFPE tissue samples

Since the results obtained above using a commercial NGS cancer panel suggest that the impact of intra-tumor heterogeneity for mutation profiling studies might be minimal across different sections of an individual FFPE tissue block, and the cost effectiveness of running replicates is important, we decided to evaluate and validate the RainDance ThunderBolts Cancer Panel to support exploratory objectives of renal cell carcinoma (RCC) clinical trial. An example of variant calls from replicate library preparations using the ThunderBolts cancer panel is shown in Table 4. All consistent variant calls (based on the Concordance Calculator) with VF greater than 5% among replicates from the same run were reproducible run-to-run (Supplemental data), and even at 3%, most of the consistent variant calls were also reproducible between runs (Supplemental data). We also included triplicate normal (non-tumor) FFPE tissue in the analytical validation strategy to eliminate panel-/amplicon-/platform-/library preparation-as well as pipeline specific artifacts besides those PTCM artifacts (data not shown). It is interesting to note that in our initial analysis of reproducibility of variant calls among all replicate samples, out of approximately 350 variant calls there were more than 100 variant calls that were present across all samples including non-tumor control FFPE tissue samples (Supplemental information and Training run Omicsoft Pass 2). This observation indicates that there is a large portion of reproducible variant calls that are likely false positives, which further validated the use of non-tumor control FFPE tissue samples in these analyses. A triplicate strategy is a prerequisite of using commercial ThunderBolts Cancer Panel in this case, since without it the run-to-run variant calls will not be as reproducible. Table 4 shows that random high quality score artifacts (Illumina quality score of 100 with very high coverages as shown in the parenthesis following the variant frequency; most likely those PTCM) with mainly C to T or G to A calls can effectively be removed using triplicate strategy. One could set a

	id	lonXpress_001 user_SN2-30-Tumor_hetero- genity_Run1.bam.Mutation- Frequency	40%	30%	2%	Ref- er- ence	Mutation	To- tal- Hit	btw 5-25%	if Y, pull 1st file freq	define freq range	Calculated Acceptable +/- Range	Low Value	High Value	if Y, look into 2nd file freq	if Y, look into 3rd file freq	Yes=3rd file within range	Y=Yes in 2nd & 3rd files
1-43815036- G+C	MUT000000004	0.1119		1	4E+07	G	G+C	1242	Y	11.19%	B	6.48%	4.71%	17.67%	9.72%	9.62%	Y	Y
1-43815082- G-C	MUT000000006	0.1105		1	4E+07	G	G-C	1185	Y	11.05%	B	6.42%	4.63%	17.47%	11.88%	10.06%	Y	Y
2-29443683- G-T	MUT000000015	0.1412		2	3E+07	G	G-T	510	Y	14.12%	B	7.65%	6.47%	21.77%	12.92%	13.37%	Y	Y
3-178916856- A-T	MUT000000732	0.0503	rs141098973	3	2E+08	A	A-T	398	Y	5.03%	A	4.52%	1.00%	9.55%	5.16%	5.45%	Y	Y
4-55602700- A+T	MUT000000855	0.2319		4	6E+07	A	A+T	1259	Y	23.19%	C	8.96%	14.23%	32.15%	25.29%	20.23%	Y	Y
4-55979608- T-GG	MUT000000893	0.0568		4	6E+07	T	T-GG	634	Y	5.68%	A	4.84%	1.00%	10.52%	5.99%	6.36%	Y	Y
5-170837513- C-T	MUT000000980	0.2256	rs200860746, rs374302777, rs34323200	5	2E+08	C	C-T	634	Y	22.56%	C	8.77%	13.79%	31.33%	25.42%	21.25%	Y	Y
7-55221794- G	MUT000000996	0.069		7	6E+07	A	G	986	Y	6.90%	A	5.45%	1.45%	12.35%	6.57%	6.67%	Y	Y
7-128845111- A-C	MUT000001084	0.2214		7	1E+08	A	A-C	1283	Y	22.14%	C	8.64%	13.50%	30.78%	23.91%	22.75%	Y	Y
9-133747507- C+T	MUT000001166	0.0556		9	1E+08	C	C+T	989	Y	5.56%	A	4.78%	1.00%	10.34%	5.75%	5.17%	Y	Y
11- 108172365-G	MUT000000159	0.0531		11	1E+08	T	G	603	Y	5.31%	A	4.66%	1.00%	9.97%	4.69%	4.74%	Y	Y
11- 108172370- T-T	MUT000000160	0.0899		11	1E+08	T	T-T	901	Y	8.99%	A	6.50%	2.50%	15.49%	7.99%	7.31%	Y	Y
13-48941615- C+T	MUT000000277	0.1232		13	5E+07	C	C+T	349	Y	12.32%	B	6.93%	5.39%	19.25%	14.36%	15.02%	Y	Y
14- 105241497- G-A	MUT000000321	0.0676	rs139904675	14	1E+08	G	G-A	1361	Y	6.76%	A	5.38%	1.38%	12.14%	5.05%	6.68%	Y	Y
17-7579372- G-C	MUT000000389	0.1302		17	8E+06	G	G-C	453	Y	13.02%	B	7.21%	5.81%	20.23%	11.46%	12.41%	Y	Y
2-29443617- C+C	MUT000000610	0.0623	rs56247462	2	3E+07	C	C+C	257	Y	6.23%	A	5.12%	1.12%	11.35%	9.19%	11.54%	N	N
9-21970956- G	MUT000001135	0.1333		9	2E+07	C	G	270	Y	13.33%	B	7.33%	6.00%	20.66%	10.99%	4.15%	N	N
																91	Y	
																3	N	
																0	#N/A	
																94	Total	
																96.8%		

1. The equation used for calculating the acceptable variant frequency range is VF +/- (%VF x acceptable %CV + % background variation), where the acceptable %CV could be customized or empirically determined to be 50%, 40% or 30% shown on the top of this table. The % background variation is empirically determined to be 2%.

2. Only variant frequencies between 5 and 25% are shown in this table; frequency range A = 5-10%, B = 10-20%, C = 20-25%.

Table 2: Selected examples of variant calls and overall reproducibility for sections 1-3 using concordance Calculator.

	id	IonXpress_003 .R_2019_09_04_07_09_23_03- user_SN2-30-Tumor_03- heterogeneity_Run1.bam. MutationFrequency	40%	30%	2%	Position	Reference	Mutation	Total Hit	btw 5-25%	if Y, pull 1st file freq	define freq range	Calculated Acceptable +/- Range	Low Value	High Value	if Y, look into 2nd file freq	if Y, look into 3rd file freq	Yes=3rd file within range	Y=Yes in 2nd & 3rd files
20-57484626- G-C	MUT000000604	0.1439			20	6E+07	G	G-C	1362	Y	14.39%	B	7.76%	6.63%	22.15%	17.17%	16.54%	Y	Y
19-3118933- T-G	MUT000000494	0.1495			19	3E+06	T	T-G	823	Y	14.95%	B	7.98%	6.97%	22.93%	12.06%	12.75%	Y	Y
19-3118893- G-G	MUT000000489	0.0875			19	3E+06	G	G-G	686	Y	8.75%	A	6.38%	2.38%	15.13%	11.72%	8.61%	Y	Y
19-122057- A-G	MUT000000470	0.0909			19	1E+06	A	A-G	682	Y	9.09%	A	6.55%	2.55%	15.64%	10.22%	11.66%	Y	Y
17-7579372- G-C	MUT000000340	0.1241			17	8E+06	G	G-C	419	Y	12.41%	B	6.96%	5.45%	19.37%	12.33%	12.15%	Y	Y
17-7578184- G-G	MUT000000319	0.1657	rs146340390		17	8E+06	G	G-G	670	Y	16.57%	B	8.63%	7.94%	25.20%	19.78%	15.40%	Y	Y
13-48941615- C-T	MUT000000246	0.1502			13	5E+07	C	C-T	426	Y	15.02%	B	8.01%	7.01%	23.03%	10.54%	12.40%	Y	Y
11- 108123568- G+A	MUT000000132	0.0522			11	1E+08	G	G+A	594	Y	5.22%	A	4.61%	1.00%	9.83%	5.47%	5.23%	Y	Y
10- 123279663- T-G	MUT000000108	0.0595			10	1E+08	T	T-G	1261	Y	5.95%	A	4.98%	1.00%	10.93%	5.63%	5.87%	Y	Y
10-89711855- C-T	MUT000000066	0.1165			10	9E+07	C	C-T	601	Y	11.65%	B	6.66%	4.99%	18.31%	11.75%	10.37%	Y	Y
7-128850328- G-C	MUT000001057	0.2011	rs370484649		7	1E+08	G	G-C	761	Y	20.11%	C	8.03%	12.08%	28.14%	22.25%	18.59%	Y	Y
7-116411971- G-C	MUT000001030	0.1917			7	1E+08	G	G-C	1090	Y	19.17%	B	9.67%	9.50%	28.84%	20.55%	20.17%	Y	Y
7-55221794- G	MUT000000958	0.0667			7	6E+07	A	G	915	Y	6.67%	A	5.34%	1.34%	12.01%	6.09%	5.96%	Y	Y
5-170837513- C-T	MUT000000943	0.2125	rs200860746, rs374302777, rs34323200		5	2E+08	C	C-T	626	Y	21.25%	C	8.38%	12.88%	29.63%	25.99%	21.91%	Y	Y
18-48584629- T-A	MUT000000403	0.1416	rs374687785		18	5E+07	T	T-A	685	Y	14.16%	B	7.66%	6.50%	21.82%	22.83%	14.01%	Y	N
12- 112926837- G+A	MUT000000204	0.0597			12	1E+08	G	G+A	586	Y	5.97%	A	4.99%	1.00%	10.96%	21.85%	15.20%	N	N
																	81	Y	
																	3	n	
																	0	#N/A	
																	84	T	
																	96.4%		

1. The equation used for calculating the acceptable variant frequency range is $VF \pm (\%VF \times \text{acceptable \%CV} + \% \text{background variation})$, where the acceptable %CV could be customized or empirically determined to be 50%, 40% or 30% shown on the top of this table. The % background variation is empirically determined to be 2%.

2. Only variant frequencies between 5 and 25% are shown in this table; frequency range A= 5-10%, B= 10-20%, C= 20-25%.

Table 3: Selected examples of variant calls and overall reproducibility for sections 3-5 using concordance Calculator.

chromosome	position	dbSNP	ref	alt	Sample 1 replicate 1	Sample 1 replicate 2	Sample 1 replicate 3	Sample 1 replicate 4
chr12	1.21E+08	.	C	T		13.39 (161173,24930,100)		
chr12	1.21E+08	.	G	A		13.57 (161582,25363,100)		
chr12	1.21E+08	.	G	A		2.45 (181777,4576,100)		
chr12	1.21E+08	.	A	G		1.2 (185168,2251,100)		
chr12	1.21E+08	.	T	C		2.54 (182270,4753,100)		
chr12	1.21E+08	.	G	A	1.29 (39808,520,78)			
chr12	1.21E+08	.	A	G	1.12 (39888,451,20)		1.19 (41458,499,40)	1.22 (37660,466,48)
chr13	28592605	.	T	C		1.29 (19051,249,42)		
chr13	28592625	.	G	A		23.97 (14673,4627,100)		
chr13	28592626	.	G	A	1.22 (32366,400,42)			1.23 (30174,377,44)
chr13	28592627	.	A	G			4.66 (45343,2216,100)	
chr13	28592634	.	C	T	1.36 (32450,449,97)			
chr13	28592640	.	A	G		23.01 (14901,4454,100)		
chr13	28592661	.	A	G		1.21 (19107,234,26)	1.18 (46809,561,43)	
chr13	28592674	.	A	G		1.27 (19063,246,38)		
chr13	28592705	.	C	T		15.59 (16342,3019,100)		
chr13	28592708	.	C	T	1.25 (32295,410,51)		1.27 (46611,600,79)	1.25 (30102,381,47)
chr13	28602209	.	A	G	1.41 (26974,385,99)	1.62 (16110,266,100)	1.52 (34426,531,100)	1.5 (26856,409,100)
chr13	28602211	.	A	G		1.42 (16203,234,67)		
chr13	28602226	.	AAGAG	A	5.95 (25582,1619,100)	1.25 (16083,204,30)	4.39 (33276,1529,100)	5.28 (25733,1435,100)
chr13	28602226	rs67947208	AAG	A	34.49 (17820,9381,100)	29.48 (11485,4802,100)	40.3 (20779,14026,100)	36.47 (17261,9907,100)
chr13	28602227	.	AGAG	A	11.89 (3343,451,100)	18.4 (1867,421,100)	9.86 (5073,555,100)	12.3 (3479,488,100)
chr13	28602227	.	AGAGA	A	2.11 (3714,80,87)	1.53 (2253,35,20)		
chr13	28602227	rs76178416	AG	A	12.57 (3317,477,100)	10.1 (2057,231,100)	10.31 (5048,580,100)	10.69 (3543,424,100)
chr13	28602228	.	GA	G	4.83 (2603,132,100)	2.9 (1509,45,90)	4.68 (4134,203,100)	4.33 (2825,128,100)
chr13	28602228	.	GAGA	G	3.58 (2637,98,100)	4.44 (1485,69,100)	2.24 (4240,97,100)	2.47 (2880,73,100)
chr13	28602229	.	AG	A	4.45 (11970,557,100)	4.26 (6294,280,100)	4.47 (18029,844,100)	3.98 (12697,526,100)
chr13	28602231	.	AG	A	2.16 (24495,540,100)	3.16 (13995,457,100)	2.55 (31735,830,100)	2.22 (24681,561,100)
chr13	28602237	.	AG	A				1.15 (26105,303,20)
chr13	28602247	.	A	C			1.2 (33855,413,39)	
chr13	28602250	.	A	G	1.3 (27034,356,59)	1.5 (16160,247,90)	1.32 (34507,461,81)	1.16 (26962,318,24)
chr13	28602256	rs61944200	C	T	53.11 (12701,14405,100)	59.08 (6627,9576,100)	42.71 (19776,14777,100)	49.28 (13671,13318,100)
chr13	28602265	.	C	T			13.13 (30453,4602,100)	

Note: 1. Reproducible variant calls are highlighted in grey across the entire row. 2. Data format: variant frequency (total reads, variant call reads, quality score). 3. High quality artifacts (Illumina quality score of 100 but not reproducible) are also highlighted in grey in individual cells.

Table 4: Selective examples of high quality artifacts and reproducible variant calls.

specific LOD (limit of detection, 5% or as low as 3%), then use triplicate FFPE replicates to filter out a set of reproducible variant calls followed by locking down the optimized pipeline after confirming a subset of variant calls using an orthogonal method. The confirmation rate can then be reported and used for clinical sample testing in triplicate. In this case, the analytical validation is incomplete until a set of reproducible calls are filtered out right on the spot (or “*in situ*”) using real clinical samples. We name this the “*in situ* analytical validation and evaluation (iSAVE)” strategy.

Precision

Analysis of inter and intra-day variability

This novel analytical validation strategy involves an unconventional approach to determine the reproducibility of all-variant calls. Since only reproducible calls were included in the final data analysis the variant call reproducibility within run is 100%. Based on the two-run results (optimization run and validation/confirmation run), we determined the run-to-run (inter-day) reproducibility is approximately 95% without manual inspection and filtering, and 100% with manual inspection and the elimination of apparent false positives.

Sensitivity

The sensitivity in this validation process is defined as the lowest variant frequency (using this triplicate strategy) that is able to achieve reproducible run-to-run variant calls. Although more gDNA input might reduce PTCM artifacts, we do not expect a further reduction of false positive or negative rates as only reproducible variant calls will be used for data analysis. Based on the data the detection sensitivity was greater than 3% for most within-run reproducible variant calls (5% for all within-run reproducible variant calls). To test whether calls were accurate, we analyzed samples of different quality then confirmed accuracy of the calls by digital PCR (dPCR) analysis (Supplemental Table S1). Thus, the current input amount of 20 ng Qubit-quantified gDNA is expected to accommodate various ranges of FFPE tissue sample quality since true signals should remain reproducible while artifacts such as PTCM (deamination or oxidation) and noise should remain non-reproducible [16,17].

Accuracy

All variant calls selected from the optimization run from four different RCC FFPE tissue samples were confirmed by dPCR analysis.

Therefore, the accuracy was 100%, although several variant calls that dPCR could not be clearly resolve were not included in the calculation as they were inconclusive due to difficult sequence regions (Supplemental Table S1).

***In situ* analytical validation and evaluation (iSAVE) strategy for DNA mutation profiling in clinical FFPE tissue samples**

Table 5 shows an example of how this *in situ* analytical validation strategy works if applied to clinical sample testing. Each set of triplicates (as shown in three consecutive columns of data) mimics a clinical sample analyzed in triplicate; 7-10 samples including normal control and standards in triplicate profiled on the ThunderBolts Cancer Panel can be accommodated in one MiSeq run (the cost is about \$100-150 per replicate). We started with basic QC filters and only included reproducible calls, then removed those reproducible variant calls that were observed across all samples including normal tissue samples (these likely represent panel-specific or amplicon-specific artifacts). To approximate the false positive rate (FPR) and false negative rate (FNR), we performed RainDrop dPCR confirmation on all variant calls for which PCR primers could be successfully designed (Supplementary Table S1). When each sample was then analyzed in triplicate as another validation run using very low stringent QC filters to minimize the FNR, we obtained an identical list of mutations. This approach allowed us to eliminate most artifact mutation calls from these clinical samples with very low FPR and FNR.

Discussion

As mentioned earlier in the Introduction, often an influential publication causing a lot of discussion and debate results more from the authors' interpretation than from the data itself. This could be exemplified by the following direct quotes from the 2012 New England Journal of Medicine (NEJM) paper published by Gerlinger et al. First, in reference to their experiment involving NGS analysis of nine biopsies of a tumor, the authors stated "Such spatially separated somatic mutations altering pathway activity suggest that multiregional analyses may be required to predict the therapeutic outcome." In other words, the authors raised a concern without proposing a practical solution. How do we know which region is going to give us the correct diagnosis or yield data predictive of treatment response if biopsies from nine different regions are expected to give nine different answers? The other direct quote from the same paper is "Identification of common mutations located in the trunk of the phylogenetic tree may contribute to more robust biomarkers and therapeutic approaches". Basically, the first statement is unrealistic (at least based on current clinical practice) but the second is the reality, so the authors can be commended for presenting both views. However, in this paper the authors showed different regions have different private mutations or shared mutations, but their conclusion is only based on a single experiment without replicates. If we use one DNA stock solution to generate different library preps, and focus only on the highest quality variant calls (such as those with Illumina quality score of 100) and consider each of our independent library preps like one of their individual nine biopsies, then we too would see many private and shared mutations in each of the library preps. In other words, it is only by doing replicates you can distinguish artifacts from genuine private and shared mutations. The non-tumor control identifies the artifactual shared mutations; the replicates identify the artifactual private mutations. This is the primary reason that we studied the impact of tumor heterogeneity among different FFPE tissue sections/regions. That is, we were planning to perform mutation profiling studies using FFPE tissue sections of patient

tumor samples and wanted to investigate the concern that "if tumors are so heterogeneous, how could you use one or a couple consecutive FFPE tissue sections to make clinical decisions or treatment response predictions?"

The impact of tumor heterogeneity on RNA expression and DNA mutation profiling

In our study (Figure 1), although the RNA expression levels appeared to be impacted significantly by tumor heterogeneity, we were unable to detect evidence that tumor heterogeneity impacted the DNA mutation profiles among different regions for the genes and targeted sequence regions included in the cancer panel. In other words, no so-called "private mutations" were identified among different regions or sections from same FFPE tumor block. We have also evaluated RCC (renal cell carcinoma) FFPE tissue block sections and, likewise, no reproducible "private mutations" were identified (unpublished results). We will continue to evaluate different FFPE tissue samples to see if we are able to identify reproducible private mutations, and if we eventually find one, it may only represent the exception rather than the norm. Whether the example presented in the NEJM publication is just an exception or not is unclear. Because of PTCM and related effects, one thing that is very clear is that using FFPE tissue sections (which has been the clinical practice for decades) is likely going to enrich common mutations ("shared mutations") and dilute private mutations if private mutations do exist; and in order to claim the status of private mutations we need to make sure these so-called private mutations are reproducible from library preparation to library preparation. Of course if we perform whole exome sequencing (WES) the chance of finding private mutations in different regions from any given tumor or FFPE tissue block is likely to be higher, however, this is yet to be directly demonstrated and confirmed using orthogonal methods [18]. We are currently applying these methodologies to evaluate the reproducibility of mutations identified through whole exome sequencing (WES) using three independent library preparations.

Key points related to the novel NGS gene panel analytical validation strategy

Based on all the information we learned previously we crafted an unconventional analytical validation strategy. Here are some key points we considered based on our recent results. We found that the same mutation on different samples (such as samples with different degree of fragmentation) or different mutations in a given sample could all have different detection sensitivities. Also, the same mutation on different platforms, different panels or even on FFPE cell lines versus FFPE tissues could have different detection sensitivities. Furthermore, hot spot mutations, rare mutations as well as mutations with different variant frequency ranges could all perform differently in terms of detection sensitivity and reproducibility. The analytical validation is incomplete (and if there is a way to ensure analytical performance we ought to explore it) until the reproducibility of each individual mutation is evaluated. Therefore, we decided to use ThunderBolts Cancer Panel from RainDance to execute this novel analytical validation strategy because it combines an "Amplify First" library preparation protocol (similar to the one used in Ion Torrent AmpliSeq, the more reproducible protocol) and uses the MiSeq instrument (the platform with the higher run-to-run reproducibility). The only further optimization we would recommend is to shorten the amplicon design [19]: the current commercially available design did result in the observation of a high number of PTCMs. This suggests that unless replicates are run, it may not be possible to get reproducible results on the ThunderBolt platform either.

	RCC 4.4_1	RCC 4.4_2	RCC 4.4_3	RCC 3.5_1	RCC 3.5_2	RCC 3.5_3	RCC 2_1	RCC2_2	RCC2_3	RCC1.5_1	RCC1.5_2	RCC1.5_3	Nor- malKid- ney-2_1	Nor- malKid- ney-2_2	Nor- malKid- ney-2_3	Nor- malKid- ney-1_1	Nor- malKid- ney-1_2	Nor- mal
CHR11~534242~A~G	0	0	0	0	0	0	0	0	0	0.16	0.19	0.17	0	0	0	0	0	0
CHR13~28602226~A~Deletion_AG	0.36	0.33	0.34	0.46	0.46	0.45	0.47	0.46	0.44	0.46	0.46	0.45	0	0	0	0	0	0
CHR18~48575196~A~G	0.34	0.33	0.31	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CHR19~17947915~G~A	0.05	0.05	0.05	0.05	0.03	0.06	0	0	0	0.14	0.09	0.07	0.05	0.04	0.02	0.03	0.04	0.02
CHR19~17947925~G~A	0.09	0.05	0.03	0.05	0.05	0.05	0.05	0.09	0.05	0.04	0.05	0.1	0	0	0	0.04	0.02	0.05
CHR2~132181660~A~T	0.06	0.12	0.4	0.05	0.12	0.11	NA	0.67	1	0	0	0	0	0	0	0	0	0
CHR2~132181743~C~T	0	0	0	0	0	0	0	0	0	0.02	0.07	0.05	0	0	0	0	0	0
CHR2~212578380~A~T	0	0	0	0	0	0	0	0	0	0.34	0.48	0.49	0	0	0	0	0	0
CHR2~29432776~T~C	0	0	0	0	0	0	0.39	0.59	0.23	0	0	0	0.53	0.51	0.52	0	0	0
CHR22~24145727~C~A	0	0	0	0	0	0	0	0	0	0	0	0	0.15	0.1	0.16	0	0	0
CHR4~55152040~C~T	0	0	0	0.46	0.45	0.46	0.35	0.54	0.48	0	0	0	0	0	0	0	0	0
CHR4~55593481~A~G	0.32	0.34	0.3	0	0	0	0	0	0	0.47	0.71	0.36	0	0	0	0	0	0
CHR4~55972974~T~G	0.51	0.53	0.7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CHR5~112176020~G~A	0	0	0	0	0	0	0	0	0	0.03	0.05	0.07	0	0	0	0	0	0
CHR4~55972974~T~C	NA	0.05	0.09	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CHR7~128846469~A~T	0	0	0	NA	0.15	0.06	0.02	0.14	0.2	0	0	0	0	0	0	0	0	0
CHR9~139397707~G~C	0	0	0	NA	0.06	0.05	0.09	NA	0.08	0.05	NA	0.25	0	0	0	0	0	0
CHR2~212578381~A~T	0	0	0	0	0	0	0	0	0	0.1	0.15	0.12	0	0	0	0	0	0
CHR2~132181640~G~A	0	0	0	0.13	NA	0.09	0	0	0	0	0	0	0	0	0	0	0	0
CHR3~10188223~A~INSERTION_G	0	0	0	0.14	0.15	0.15	0	0	0	0	0	0	0	0	0	0	0	0
CHR4~1803568~C~G	0	0	0	0	0	0	0	0	0	0.72	0.69	0.64	0	0	0	0	0	0

Table 5: Final reproducible variant calls from the training run.

Practical use of “in situ analytical validation and evaluation” strategy

As presented in the result section (Table 4), because all consistent variant calls among replicates from the same run with VF greater than 5% were reproducible run-to-run, we were able to use a triplicate strategy and include normal FFPE tissue in the analytical validation run to eliminate panel-, amplicon-, platform-, even library preparation- and pipeline-specific artifacts in addition to those PTCM artifacts. We have confirmed all of the variant calls derived from the training run using dPCR as orthogonal method for variants for which it was possible to design PCR primers. The validation run yielded an identical list of mutations using very low stringent QC filters. This in situ analytical validation strategy can be applied to WES assay validation as our unpublished WES studies suggest similar artifacts are observed.

Conclusion

The current advantage of NGS is its ability to detect unknown or rare mutations. Otherwise, cheaper, more effective and faster turn-around time options are readily available. However, often the strategy of eliminating false positives in diagnostic NGS panels or even WES data analysis is to apply aggressive QC filters [18]. This approach could potentially increase the false negative rates of variant detection, and many of the false negatives are likely to be those very important rare or unknown somatic mutations. Interestingly, strategies or data analysis tools for developing reliable low false positive rate diagnostic NGS panels were also adapted by scientists in the field for developing NGS mutation profiling prognostic panels even though in WES that both low false positive and negative rates are equally important. One could argue that in this case, low false negative rate should be prioritized higher than achieving a low false positive rate since many false positives could be eliminated through the confirmation of hypothesis driven biomarker discovery process and will not withstand any type of confirmation. It is rare that any QC filters could accomplish the goal of reducing false positive rate without impacting the false negative rate. Hence, a set of relatively low stringency QC filters should be used in the beginning of the data analysis process. A triplicate strategy is a very effective way to eliminate false positives or technical artifacts without the need to depend on high stringency QC filters. The authors understand the challenge of the high cost associated with running triplicate NGS samples. However, using whole exome sequencing as an example, if a mutation assay is not able to deliver a highly reproducible list of identified somatic mutations, we may need to ask ourselves if this technology is ready for prime time. Hence, we propose the following concept: for each clinical lab to include at least one unique randomly selected clinical sample from a previously run in each and every new batch of NGS clinical assay/test, starting from nucleic acid [13]. Therefore, over a period of time and through the entire sample testing process, some degree of real-world intra-laboratory reproducibility of clinical samples would be generated in terms of the total number and the identity of mutations reported. If the results show the reported mutation calls are highly reproducible we can at least have confidence that the decisions made for the patients are done in a responsible manner. Even with only reproducible calls are reported, there is no guarantee that all the reproducible calls will be true positives (especially if no normal control samples are run in replicates to remove the artifactual “shared mutations;” see Results), but at least we know if the test is done again the same results should be obtained. Finally, if the results show a significant number of variant calls are not reproducible, such tests should be re-evaluated and analytical validation re-designed to address this shortfall.

Acknowledgments

The authors would like to thank Bo Wei for her active involvement of data analysis discussion. We also thank Diane Levitan for her comments and suggestions on this novel validation strategy.

References

1. Chang KCN, Zhao Y, Kang J, Pant S, Qiu P, et al. (2014) Validation of Next Generation Sequencing Cancer Panels for Clinical Somatic Mutation Profiling: Identification of Source of Variations and Artifacts using FFPE Tissues. *Next Generat Sequenc & Applic* 1: 109.
2. Gerlinger M, Rowan AJ, Horswell S, Larkin J, Endesfelder D, et al. (2012) Intratumor Heterogeneity and Branched Evolution Revealed by Multiregion Sequencing. *New England J Med* 366: 883-892.
3. Nielsen T, Wallden B, Schaper C, Ferree S, Liu S, et al. (2014) Analytical validation of the PAM50-based Prosigna Breast Cancer Prognostic Gene Signature Assay and nCounter Analysis System using formalin-fixed paraffin-embedded breast tumor specimens. *BMC Cancer* 14: 177.
4. Klein EA, Cooperberg MR, Magi-Galluzzi C, Simko JP, Falzarano SM, et al. (2014) 17-Gene assay is a significant predictor of prostate cancer aggressiveness in the context of Gleason grade heterogeneity, tumor multifocality, and biopsy undersampling. *Eur Urol* 66: 550-560.
5. Chen DT, Davis-Yadley AH, Huang PY, Husain K, Centeno BA, et al. (2015) Prognostic fifteen-gene signature for early stage pancreatic ductal adenocarcinoma. *PLoS ONE* 10: e0133562.
6. Ayers M, Lunceford J, Nebozhyn M, Murphy E, Loboda A, et al. (2015) Relationship between immune gene signatures and clinical response to PD-1 blockade with pembrolizumab (MK-3475) in patients with advanced solid tumors. *J Immunother Cancer* 3: P80.
7. Illumina (2016) Base Space Sequence Hub: Specification sheet software. Illumina. Accessed on: November 09, 2016.
8. Ion AmpliSeq™ Cancer Hotspot Panel. Catalog number: 4475346. Thermo Fisher scientific. Accessed on November 14, 2016.
9. Ion Torrent (2012) Ion AmpliSeq Cancer Panel: APPLICATION NOTE Accessed on: November 09, 2016.
10. Chang KCN, Galuska S, Weiner R, and Marton MJ (2013) Development and validation of a clinical trial patient stratification assay that interrogates 27 mutation sites in MAPK pathway genes. *PLOS ONE* 8: e72239.
11. Genomics Clinical Trial Assay Development: Issues and Lesson Learned. Published by OMICS Group eBooks, edited by: Ken CN Chang and Matthew J Marton ISBN: 978-1-63278-040-9.
12. Chang KCN, Marton MJ (2016) Next Generation Sequencing and Its Clinical Applications: The Growing Pains. *Next Generat Sequenc & Applic* 3: 131.
13. http://supportres.illumina.com/documents/documentation/system_documentation/miseq/miseq-system-user-guide-15027617-1.pdf
14. Ion PGM™ Sequencing 200 Kit v2. 2012. Life technologies. Accessed on November 14, 2016.
15. Ilsey D (2014) Functional DNA Quality Analysis Improves the Accuracy of Next Generation Sequencing from Clinical Specimens. Asuragen White Paper.
16. Costello M, Pugh TJ, Fennell TJ, Stewart C, Lichtenstein L, et al. (2013) 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.
17. Qiu P, Pang L, Arreaza G, Maguire M, Chang KC, et al. Data interoperability of whole exome sequencing (WES) based mutational burden estimates from different laboratories. *Int J Mol Sci* 17: 651.
18. Thunder bolts, cancer panel. Rain dance technologies. Accessed on November, 14, 2016.