Supplemental Table 1.

SUMMARY OF BFAST ALIGNMENTS AGAINST TARGET REFERENCE					
Sample	**Total Aligned Reads**	**Total Aligned Bases (Mb)**	**Reads Aligned Against Amplicon Sequence**	**Bases Aligned Against Amplicon Sequence**	**Percent Aligned Against Amplicon Sequence**
RA37N	10672738	534	9224675	461	86.43
RA37T	3773678	189	3059045	153	81.06
RA45AN	8056178	403	6540648	327	81.19
RA45M	10613505	531	9259689	463	87.24
RA45RM	6590092	330	4953946	248	75.17
RA45AT	8295082	415	6593130	330	79.48
RA48RN	11928664	596	10826228	541	90.76
RA48RM	10328977	516	9180672	459	88.88
RA49NS	9641061	482	7707022	385	79.94
RA49LT	7292523	365	5613472	281	76.98
RA53M	10977661	549	9706908	485	88.42
RA53T	8433215	422	7209541	360	85.49
SUMMARY OF BFAST ALIGNMENTS AGAINST WHOLE GENOME*					
RA37N	18182314	909	9525943	476	52.39
RA37T	5493203	275	3157205	158	57.47
RA45AN	14720289	736	6788528	339	46.12
RA45M	16619153	831	9638236	482	57.99
RA45RM	14774691	739	5055057	253	34.12
RA45AT	15697659	785	6769725	338	43.13
RA48RN	16962721	848	11217146	561	66.13
RA48RM	16225405	811	9468204	473	58.35
RA49NS	18361830	918	8024030	401	43.70
RA49LT	16412566	821	5528827	276	33.69
RA53M	17078441	854	10158923	508	59.48
RA53T	14548824	727	7548495	377	51.88

* - Reference genome was NCBI Build 36, hg18

Supplemental Methods

Sequence Alignment

SOLiD NGS csfasta and quality files were used for alignment against a human genome reference (Build 36, hg18) fasta file using BFAST (Homer et al. 2009). This alignment tool relies on indexing of reads to rapidly map each read to its most likely position within the reference, while allowing for read errors and mismatches to insure detection of possible sequence variants (Homer et al. 2009). A final local alignment using a Smith-Waterman method is performed for each read, allowing for gaps to detect potential small indels. The BFAST output files are in the binary alignment format (BAM) (Li et al. 2009). These final BAM files are input files for variant detection algorithms. To assess alignment performance of SOLiD sequenced RainDance PCR products, we calculated percent uniquely mappable reads and percent reads mapping to the target regions. These results are provided in Supplemental Table 1.

Variant Detection

The PCR duplicate removed BAM files for each sample were used for variant detection using two independent algorithms. We used the algorithm SolSNP version 1.0 (Sinari et al. 2010) to detect variants. SolSNP is an individual sample variant detector (classifier) implemented in Java. The variant calling is based on a modified Kolmogorov-Smirnov like statistic. The algorithm is non-parametric and makes no assumptions on the nature of the data. It compares the discrete sampled distribution, the pileup on each strand, to the expected distributions (according to ploidy). An important aspect of SolSNP that reduces overcalling

inherent to the K-S statistic algorithm is that filters are included to reduce false positive rates, among of which is that both strands must provide evidence for the variation. Zero quality bases are trimmed off the pileup before the comparisons. It is a standalone program that can be run from the command line and is general enough to work with any next generation sequencing data with high (~9X) coverage. Furthermore, a Call Bias filter is available, which allows a heterozygous variant to be classified even if it falls outside of the normal distribution for a heterozygous variant within a pileup. This is very important in the case of cancer, where heterozygous alleles fall far outside of the normal diploid distribution based on the amount of normal contaminating stroma. Therefore, we modified the Call Bias filter among our tumor sample alignments to account for differing percentages or tumor cellularity within each sample. A Call Bias of (0) is used when assuming a normal diploid distribution of alleles. However the Call Bias can be increased to allow for variant detection when alleles might fall outside of the normal diploid distribution, such as instances when tumor DNA contains contaminating normal stromal DNA. The parameters used in SolSNP variant detection are provided here. SolSNP documentation and descriptions of these parameters can be found at solsnp.sourceforge.net.

For Normal BAMs:
STRAND_MODE=GenotypeConsensus
MINIMUM_MAPQ=10
FILTER=0.0
MINIMUM_BASE_QUALITY=1
PLOIDY=Diploid
MINIMUM_COVERAGE=3
MAX_MATE_DISTANCE=2147483647
MIN_MATE_DISTANCE=0
CALL_BIAS=0.0

For Tumor BAMs:
STRAND_MODE=VariantConsensus
MINIMUM_MAPQ=10
FILTER=0.0
MINIMUM_BASE_QUALITY=1
PLOIDY=Diploid
MINIMUM_COVERAGE=3
MAX_MATE_DISTANCE=2147483647
MIN_MATE_DISTANCE=0

The Call Bias parameters changed across Tumor BAMs:
RA37 Call Bias=0.2
RA53 Call Bias=0.5
RA45 Call Bias=0.2
RA49 Call Bias=0.5
RA48 Call Bias=0.2

We also utilized a second variant detection algorithm to serve as an independent and redundant pipeline to provide additional support for variants. For this purpose, we employed the variant detection algorithm within the GATK suite (McKenna et al. 2010). The GATK variant detection tool uses a Bayesian method to compute the posterior probability for genotypes at a given base within the NGS sequence read pileup with an expected heterozygosity for the sample (McKenna et al. 2010). The parameters used for the GATK Unified Genotyper analysis in our study are provided here.

UG_base_model=EMPIRICAL
UG_cap_base_quality_by_mapping_quality=false
UG_genotype=false
UG_genotype_model=JOINT_ESTIMATE
UG_heterozygosity=0.0010
UG_max_deletion_fraction=0.05
UG_max_mismatches_in_40bp_window=3
UG_min_base_quality_score=10
UG_min_mapping_quality_score=10

UG_noSLOD=false
UG_output_all_callable_bases=false
UG_platform=SOLID
UG_standard_min_confidence_threshold_for_calling=30.0
UG_standard_min_confidence_threshold_for_emitting=10.0
UG_trigger_min_confidence_threshold_for_calling=30.0
UG_trigger_min_confidence_threshold_for_emitting=30.0

The description for each each parameter is available at the following web-page: http://www.broadinstitute.org/gsa/wiki/index.php/Unified_genotyper.

Sequenome Mass Array

We genotyped position chr1:11104624C>A (FRAP1 6809G>T mutation) and chr17:7517878C>T (TP53 785G>A mutation) in our sample set using the Sequenom MassARRAY™ platform with iPLEX™ chemistry. Briefly, iPLEX™ assays were designed utilizing the Sequenom Assay Design software, allowing for single base extension (SBE) designs used for multiplexing. PCR and SBE primer sequences are available upon request. Briefly, multiplex polymerase chain reactions (PCR) were performed to amplify 5-10 nanagrams (ng) of genomic DNA. PCR reactions were treated with shrimp alkaline phosphatase (SAP) to neutralize unincorporated dNTPs. Subsequently, a post-PCR single base extension reaction was performed for each multiplex reaction using concentrations of 0.625 micromolar (mM) for low mass primers and 1.25 mM for high mass primers. Reactions were diluted with 16 microliters (ml) of H_2O and fragments were purified with resin, spotted onto Sequenom SpectroCHIP™ microarrays and scanned by MALDI-TOF mass spectrometry. Individual SNP genotype calls were generated using Sequenom TYPER™ software, which automatically calls allele specific peaks according to their expected masses.