ion torrent

Shannon entropy of mutational signatures predicts sensitivity of signature detection in targeted sequencing

Chellappan, Ajithavalli; Vora, Chintan; Nair, Shilpa; Patro, Jagannath; Kanap, Rushikesh S.; Hyland, Fiona C.
Thermo Fisher Scientific, 200 Oyster Point Blvd, South San Francisco, CA; Bengaluru, India.

1515

Background

Cancer genomes are subject to diverse mutational processes that generate recognizable mutational signatures. Some processes are driven by defects in specific DNA repair pathways whereas others are characteristic of environmental mutagens. Mutation signatures are generally mined from Whole Genome Sequencing or Whole Exome Sequencing data. We demonstrate that we can identify mutation signatures using amplification-based Targeted Sequencing panels, a method especially robust for sequencing FFPE samples.

Methods

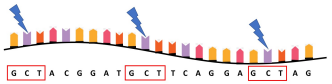
2050 FFPE Samples were identified from a pan solid tumor cohort, amplified with a targeted panels (Oncomine Tumor Mutation Load Assay - TMB or Onocmine Comprehensive Assay Plus - OCAPlus), and sequenced with Ion GeneStudio™ S5 System.. These panels use AmpliSeq-based enrichment that is robust with 20ng of input DNA. We filtered out germline mutations, removing variants in population databases, to generate a set of somatic SNVs. Single base change substitution (SBS) matrix for these somatic mutations was constructed and normalized based on the panel composition. Cancer signatures described in COSMIC Mutational Signatures v3.1, were characterized and the cosine similarity between the normalized sample and SBS COSMIC signatures was measured. Signatures with a strong match (> 0.7) to the normalized sample were shortlisted. We also used an orthogonal approach to impute the signatures using a reduced candidate set. In this approach, the DeconstructSigs R package was used to determine the weights of each mutational signature contributing to an individual tumor sample

Content

Ion Torrent Oncomine Tumor Mutation Load Assay (TML)
The Oncomine Tumor Mutation Load Assay™ is a targeted next-generation sequencing (NGS) assay that provides an assessment of tumor mutation load and mutation signatures in a simple workflow. The assay measures TMB (from 1.2Mb of coding region) and detects mutations in 409 cancer genes.

Ion Torrent Oncomine Comprehensive Assay Plus (OCAPlus)
The Oncomine Comprehensive Assay Plus™ is a targeted next-generation sequencing (NGS) assay that provides a comprehensive genomic profiling solution appropriate for formalin-fixed paraffin-embedded (FFPE) tissues. The assay addresses multiple biomarkers covering over 500 genes, including targets that are relevant in cancer. This assay enables analysis of variants across 500+ genes and detection of SNVs, CNVs, In-Dels, TMB, MSI, and gene fusions.

Figure 1. Differential rates of mutations within trinucleotides is a characteristic of mutational signatures.



Sample has a mutation in MLH1 and SBS44 signature

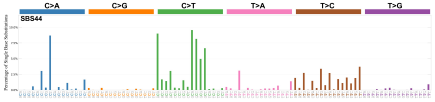


Figure 2a. Mutational profile of Cosmic Signature SBS44

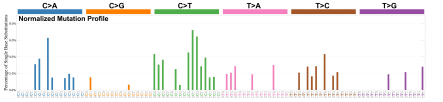


Figure 2b. Mutational profile of a cancer sample with ignature SBS44 sequenced using TML panel

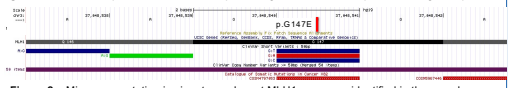


Figure 2c. Missense mutation in signature relevant MLH1 gene was identified in the sample

Examples of a high and low entropy signature

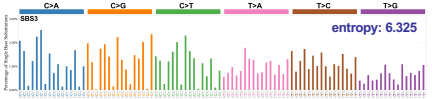


Figure 3a. Mutational profile of high entropy COSMIC signature for HRD : SBS3

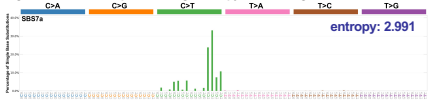


Figure 3b. Mutational profile of low entropy COSMIC signature for UV : SBS7a

Shannon entropy of a Cosmic signature, definition

$$H(x) = -\sum_{i=1}^{96} p_i \log_2 p_i$$

, where p is the proportion of mutations that are observed for the trinucleotide base change i (out of possible 96) in the signature. H(s) is the entropy of signature x.

Distribution of the Entropy of Cosmic signatures

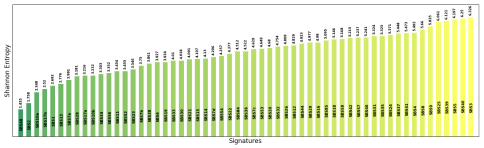


Figure 4 : Bar plot showing the distribution of the Entropy of COSMIC signatures.

Simulated Mutation profiles of high and low entropy signatures: low entropy signatures better recapitulate mutation pattern with fewer somatic mutations

MutSigSim – a tool for simulating mutation profiles

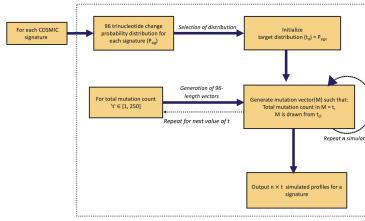


Figure 5 : We developed a simulation tool, MutSigSim. It provides a randomized list of mutations drawn from a mutational signature profile and a total number of mutations. This flowchart depicts the pipeline in the simulation workflow

Simulated profiles from signature distributions show that identification of high entropy signatures requires more mutations than low entropy signatures

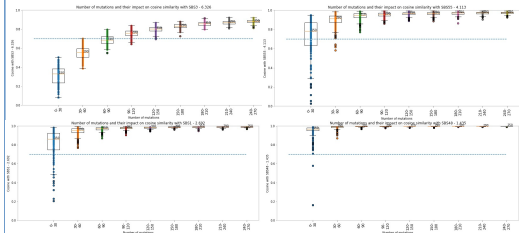


Figure 6 : Box plot of four signatures showing the impact of number of mutations on cosine similarity with the signature. Each dot on the box plot is one sample. Clockwise: SBS3 (entropy = 6.326), SBS55 (entropy = 4.113), SBS48 (entropy = 1.435) and SBS1 (entropy 2.692). Blue dotted line on each plot shows the necessary cosine threshold to call a sample as having the signature.

Ovarian cancer samples with suspected SBS3 signatures (entropy = 6.326) showed increasing cosine score with increase in number of mutations

Sample Name	Total Variants	# trinucleotide base change peaks	Cosine Similarity with SBS3
Ovarian_Prostate_cancer_sample_G04	528	37	0.426
Ovarian_Prostate_cancer_sample_G03	480	63	0.559
Ovarian_Prostate_cancer_sample_D05	264	56	0.492
Ovarian_Prostate_cancer_sample_A05	154	37	0.471
Ovarian_Prostate_cancer_sample_F08	67	25	0.218
Ovarian_Prostate_cancer_sample_B11	9	8	0.471
Ovarian_Prostate_cancer_sample_B07	9	9	0.442
Ovarian_Prostate_cancer_sample_G12	6	6	0.312
Ovarian_Prostate_cancer_sample_A10	2	2	0.269
Ovarian_Prostate_cancer_sample_B08	1	1	0.209

Table 1a: Table showing the total variants, number of non-zero trinucleotide change peaks and cosine similarity with SBS3.

MSI-H and MSS samples with suspected MMR signatures (with medium entropy) showed high specificity

MMR Signature (entropy)	152 MSI-H samples	106 MSS samples	Sensitivity TP / (TP+FN)	Specificity TN / (TN + FP)	False Positivity Rate FP / (FP+TN)
SBS6 (3.927)	5	8	3%	92%	7.5%
SBS44 (4.923)	28	0	18%	100%	0%
SBS26 (4.809)	27	0	18%	100%	0%
SBS20 (4.680)	1	0	1%	100%	0%
SBS14 (4.130)	2	0	1%	100%	0%

Table 1b: Table showing the sensitivity and specificity with MSI-H and MSS samples for signatures associated with MMR.

Conclusions

This research demonstrates that mutation signatures can be identified using amplification-based targeted sequencing data with two Oncomine panels, and that mutation signatures with low entropy are easier to detect than are mutation signatures with high entropy.

ThermoFisher SCIENTIFIC

Thermo Fisher Scientific • 200 Oyster Point Boulevard, South San Francisco, CA 94080• thermofisher.com