Tumor Mutational Burden Quantification Using Small and Moderate Sized Targeted Anchored Multiplex PCR Panels and Next Generation Sequencing

ST092

Glen Bjerke, Devin Tauber, Ryan Rogge | Archer, Boulder, CO. USA.

Summary

- VARIANT*Plex*™ Pan Solid Tumor (185 gene) and Complete Solid Tumor (430 gene) panels now also measure TMB
- VARIANT*Plex* targeted NGS panels have high concordance with alternative NGS TMB methods
- Novel somatic / germline classifier powers TMB calculation without a paired normal sample

Introduction

Tumor Mutational Burden (TMB) is rapidly gaining acceptance in oncology research as a key biomarker and indicator of tumor response to immune checkpoint inhibitors (ICIs). Whole exome sequencing (WES) of tumor-normal pairs is the reference standard for calculation of TMB; however, given the cost and other workflow implications of WES, TMB is frequently derived from targeted next generation sequencing (NGS) panels. Here we describe measurement of TMB using VARIANT*Plex* Pan Solid Tumor and VARIANT*Plex* Complete Solid Tumor panels. Utilization of Anchored Multiplex PCR (AMP™) chemistry with molecular barcoding allows identification of PCR and sequencing duplicates, enabling powerful error correction. In addition to TMB measurement, VARIANT*Plex* panels allow simultaneous detection of single nucleotide variations (SNVs), insertions/deletions (indels), copy number variations (CNVs), and microsatellite instability (MSI) from formalin-fixed, paraffin-embedded (FFPE) tissue samples. Measurement of TMB using targeted panels has 3 main sources of bias: 1) targeted panels are significantly smaller than the human exome, resulting in lower signal, 2) targeted panels are usually designed around common tumor mutations, resulting in selection bias and 3) targeted panels are normally run without a paired normal sample, necessitating the removal of germline variants. To mitigate the first two sources of bias we developed a linear model that takes as input both nonsynonymous and synonymous mutations and predicts a numeric TMB score that emulates the WES TMB score. Additionally, due to the uncertainty in sampling a limited amount of the exome (an issue inherent to TMB calling using any targeted panel)[1], a user settable intermediate zone was implemented within the TMB call, allowing higher confidence in TMB low/high calls. Finally, to enable TMB measurement of tumor only samples, we developed a machine learning algorithm that classifies variants as somatic or germline with high accuracy, outperforming somatic assignment using only a gnomAD[2] frequency cutoff.

Methods

VARIANT*Plex* Pan Solid Tumor and VARIANT*Plex* Complete Solid Tumor cover genomic regions of 669 kb and 1.42 Mbp, including 185 and 430 genes, respectively. Sequencing libraries were produced with the indicated VARIANT*Plex* panels using 50-100 ng of purified DNA from either cell lines or FFPE samples and sequenced on Illumina NextSeq™ or NovaSeq™ instruments. Bioinformatic analyses were performed using Archer™ Analysis 7.2 which uses bespoke algorithms to optimize for TMB quantification on small, targeted panels. After variant filtering and removal of common sequencing errors, somatic variants were identified using a machine learning algorithm that was initially trained on WES data from 779 tumor-normal paired samples (comprising more than 130,000 total variants). This in-house developed somatic / germline classifier allows for TMB quantification without a paired normal sample. TMB scores were then calculated using somatic variant counts (nonsynonymous and synonymous) by a linear model derived from The Cancer Genome Atlas (TCGA) data[3] to approximate WES based TMB scores, mitigating subsampling and selection biases inherent to targeted panel TMB. Normal samples are assumed to have a TMB of 0.

VARIANT*Plex* Pan Solid Tumor / VARIANT*Plex* Complete Solid Tumor

↓

Variant Filtering

↓

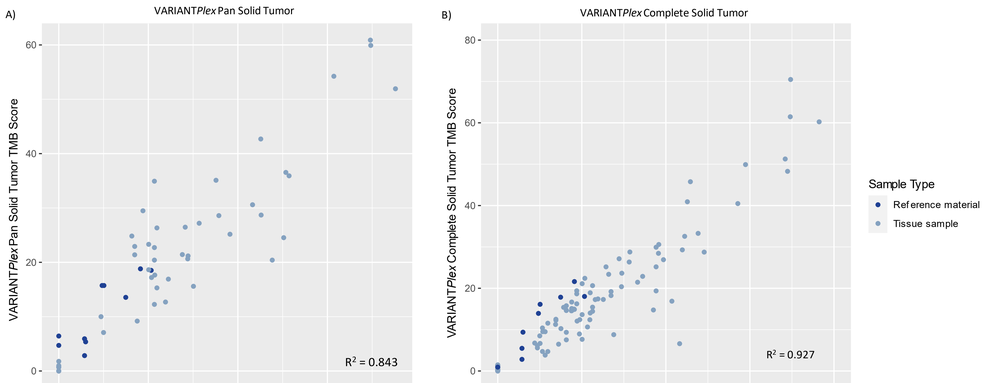
Somatic / Germline Variant Classification

↓

Predict TMB Score Using Linear Model

Results

Figure 1 – VARIANT*Plex* Pan Solid Tumor and Complete Solid Tumor concordance with comparator NGS TMB methods

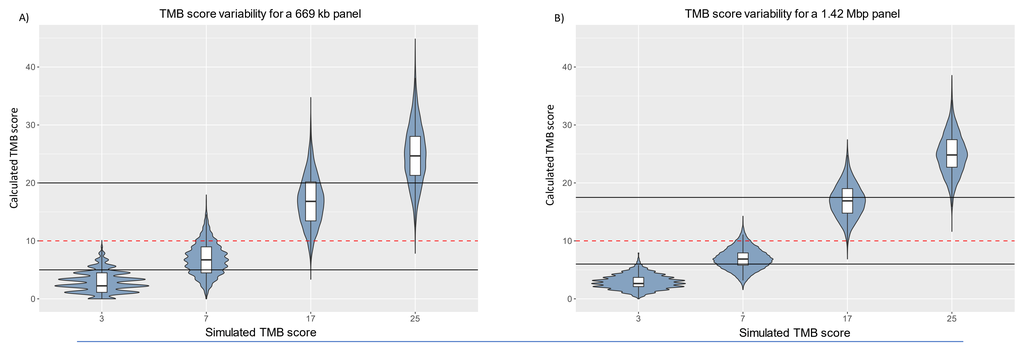


	10 Mutations/Mbp	(-) TMB-Intermediate: 5-20 Mutations/Mbp
Concordance	95.0%	100%
Positive % agreement	93.2%	100%
Negative % agreement	100%	100%

	10 Mutations/Mbp	(-) TMB-Intermediate: 6-17.5 Mutations/Mbp
Concordance	85.2%	100%
Positive % agreement	81.5%	100%
Negative % agreement	100%	100%

Figure 1. VARIANT*Plex* TMB performance versus orthogonal methods.
A) VARIANT*Plex* Pan Solid Tumor Performance. Reference materials used: 11 SeraCare® TMB Standards and 4 WES-characterized reference cell lines (ATCC or GIAB). Tissue sample inputs used: 39 de-identified FFPE tumor tissue samples characterized using a 500+ gene hybrid capture NGS assay, 2 de-identified FFPE normal adjacent tissue samples, 4 de-identified peripheral blood samples.
B) VARIANT*Plex* Complete Solid Tumor Performance. Reference materials used: 11 SeraCare® TMB Standards and 5 WES-characterized reference cell lines (ATCC or GIAB). Tissue sample inputs used: 87 de-identified FFPE tumor tissue samples characterized using a 500+ gene hybrid capture NGS assay, 2 de-identified FFPE normal adjacent tissue samples, 10 de-identified peripheral blood samples.
(-) TMB-Intermediate results exclude samples within the indicated default range for Archer Analysis across all NGS assays evaluated.

Figure 2 – Simulated TMB score variability due to genomic region subsampling for targeted TMB calculations versus WES based TMB



0.0002%	16.7%	3.4%	0.0004%	Percent crossing threshold of 10	0%	7%	0.5%	0%

Figure 2. Variability for targeted panel based TMB versus WES based TMB. Shown on the y axis is the distribution of calculated TMB scores from a panel of the given size simulated for 5,000 samples. The x axis shows the simulated TMB score (what would be the correct score by WES). Calculated TMB scores were computed by first calculating the number of variants to simulate per sample: the simulated TMB score * the size of the exome (30 Mbp) * 1.33 (to mimic the inclusion of synonymous and nonsynonymous mutations). Given the number of mutations to simulate, each mutation was randomly assigned to be covered by the panel or not by the ratio of panel size / size of the exome. This process was repeated 5,000 times for each simulated TMB score to give a distribution. The dotted red line shows the TMB low / high threshold of 10. The percent of the simulated scores that cross the TMB threshold of 10 is shown below each figure. **A) Variability in targeted TMB scores given a panel of 669kb.** The black lines denote the lower and upper bounds for an intermediate zone of 5-20. **B) Variability in targeted TMB scores given a panel of 1.42 Mbp.** The black lines denote the lower and upper bounds for an intermediate zone of 6-17.5.

Table 1 – Archer somatic / germline classifier models' performance metrics

A) VARIANT*Plex* Pan Solid Tumor

	Training data	Validation data
Concordance	93.79	97.3
PPA	93.08	97.47
NPA	97.93	92.73

B) VARIANT*Plex* Complete Solid Tumor

	Training data
Concordance	95.42
PPA	94.96
NPA	98.02

Table 1. Concordance, Positive Percent Agreement (PPA) and Negative Percent Agreement (NPA) for Archer's somatic / germline classifier models. Somatic / germline classifier models were trained using a machine learning model on 779 tumor-normal WES pairs to differentiate between somatic and germline variants using 10-fold repeated cross validation.
A) Somatic / germline classifier metrics for VARIANTPlex Pan Solid Tumor. Validation data is a comparison of somatic / germline classifier calls from samples run on Archer Analysis versus tumor-normal WES data for the same samples used as the source of truth.
B) Somatic / germline classifier metrics for VARIANTPlex Complete Solid Tumor.

Figure 3 – Comparison of the number of somatic mutations called versus using a gnomAD cutoff alone

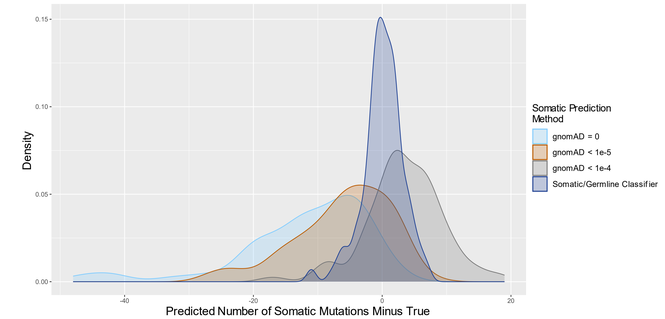


Figure 3. Comparison of somatic / germline predictions using only gnomAD cutoffs vs Archer's somatic/germline classifier. Using the VARIANT*Plex* Complete Solid Tumor region of interest a somatic / germline classifier was trained on 623 paired tumor-normal WES samples. This model was then used to predict somatic mutations using tumor only data for the remaining 156 samples. The total number of somatic mutations called was compared to the true number of mutations (via the paired normal sample) for the somatic / germline classifier and three gnomAD population frequency cutoffs: 0 (i.e. not in gnomAD), less than 1e-5 or less than 1e-4.

Conclusions

VARIANT*Plex* panels offer an alternative to WES based TMB in the form of small (Pan Solid Tumor) or medium (Complete Solid Tumor) sized targeted NGS panels, while simultaneously allowing detection of SNVs, indels, CNVs and MSI from FFPE samples of varying qualities. Archer's methods to correct for the selection bias inherent to panel based TMB provide high concordance with WES based TMB. Finally, we present an innovative new somatic / germline classifier with high accuracy that outperforms using a gnomAD population frequency cutoff to determine somatic variants, enabling TMB measurement on tumor only samples without a paired normal.

References

1. Budczies et al. Optimizing panel-based tumor mutational burden (TMB) measurement. Ann Oncol. 2019 Sep 1;30(9):1496-1506. doi: 10.1093/annonc/mdz205. PMID: 31268125.
2. Chen at al. A genome-wide mutational constraint map quantified from variation in 76,156 human genomes. bioRxiv 2022.03.20.485034 (2022). https://doi.org/10.1101/2022.03.20.485034.
3. Ellrott et al. Cancer Genome Atlas Research Network. Scalable Open Science Approach for Mutation Calling of Tumor Exomes Using Multiple Genomic Pipelines. Cell Syst. 2018 Mar 28;6(3):271-281.e7. doi: 10.1016/j.cels.2018.03.002. PMID: 29596782.

Glen Bjerke
Staff Scientist, Archer (Integrated DNA Technologies)
gbjerke@idtdna.com